Source, propagation, and effects of lightning in the Earth-ionosphere system

Michael L. Hutchins

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Robert H. Holzworth, Chair

Michael P. McCarthy

Abram R. Jacobson

John M. Wallace

John D. Sahr

Program Authorized to Offer Degree:
UW Department of Earth and Space Sciences

University of Washington

Abstract

Source, propagation, and effects of lightning in the Earth-ionosphere system

Michael L. Hutchins

Chair of the Supervisory Committee:
Professor Robert H. Holzworth
Department of Earth and Space Sciences

The capabilities of the World Wide Lightning Location Network (WWLLN) are expanded to enable research of the source, propagation, and effects of lightning in the Earth-ionosphere system. The main expansion of the network capability is the measurement of the very low frequency radiated energy from lightning; the radiated stroke energy is one to one related to the canonical peak current measurements of other ground based networks. Stroke energy is used to develop a model of the network relative detection efficiency; this internal model rates the coverage capabilities of the network compared to the networks best regional coverage. The last dataset developed and discussed is the clustering of the lightning locations into both flashes and the active lightning regions of thunderstorms. These three capabilities of the network allow tracing the effects of lightning and thunderstorms from their source, to a proxy for the global electric circuit and to the magnetosphere.

The source of lightning is investigated in two regimes: within thunderstorms and between thunderstorms. Within thunderstorms the time between flashes is found to be proportional to the resulting flash energy for differing thunderstorms, regions, and seasons. Between thunderstorms the lightning energy is shown to differ between land and ocean, with oceanic thunderstorms producing stronger and fewer strokes. The propagation of the radiated energy is measured using the lightning as a probe of attenuation along the different propagation paths. Attenuation is seen to have an asymmetry with magnetic azimuth: eastward moving waves are attenuated less than westward moving waves. The attenuation asymmetry is

complimentary to the observed asymmetry in whistler and radio energy emitted through the ionosphere into the magnetosphere. Thunderstorm clusters are used to estimate the total upward current contribution of thunderstorms to the global electric circuit. It is shown that WWLLN can provide one of the first continuous global measurements of this current to the global electric circuit.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

DEDICATION

To Leah

Chapter 1

INTRODUCTION

This thesis explores the science and capabilities of ground based lightning detection networks. What are the measurement limits of lightning networks? How can these networks be expanded and what are the results of the expansions? Can choosing the right algorithm reveal larger structures from individual lightning locations?

With these questions lightning, thunderstorms, and the Earth-Ionosphere system are explored. The electrical energy of thunderstorms is investigated in factors of production, how it discharges, where the energy goes, and how it contributes to the Earth system as a whole.

1.1 Background

1.1.1 Lightning and Thunderstorms

Lightning is a process of discharging disparate charged regions in a thunderstorm. When electrical charge is separated in a thunderstorm it can discharge, through lightning flashes, in order to neutralize the separated charge.

As deep convection develops in a thunderstorm, moist air is lifted from below the thunderstorm through the cloud where it condenses then freezes. Figure 1.1 shows the basic components of an active thunderstorm. The water begins to freeze as it passes through the freezing level of the thunderstorm; collisions at this freezing level occur with downwelling graupel pellets. These collisions cause charge exchange between the ice particles and graupel; in the upper portion of the cloud the smaller ice particles gain a positive charge and the graupel becomes negatively charged while the exchange is reversed in the lower portion of the thundercloud (*Saunders*, 1995). This leads to the gross charge structure shown in Figure 1.1, with net positive charge in the top portion of the thunderstorm and net negative charge in the bottom portion. There are smaller charge layers in the thunderstorm, such as screening layers, but the overall charge structure is that depicted in Figure 1.1.

The lightning discharge begins with the stepped-leader process. While a voltage on the order of 1 MV builds up between the charge centers (e.g. the thunderstorm and ground) it is still far less than the breakdown electric field of air (on the order of 1 MV/m). However the air can breakdown in steps between 3 – 200 m, creating a continuous ionized plasma
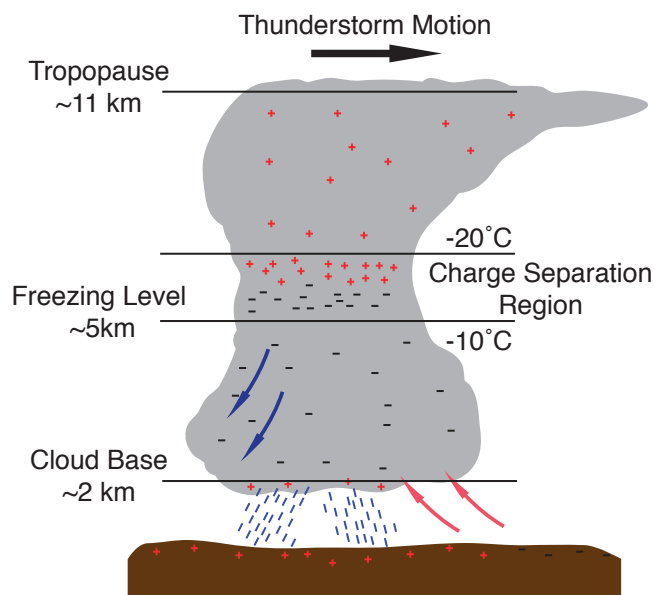
Figure 1.1: Overall thunderstorm charge structure, showing a typical freezing level and tropopause hight.

channel; the steps occur until an oppositely charged region is connected (*Ogawa*, 1995). An example of stepped leader breakdown is shown at the beginning of Figure 1.2.

For the first 10 ms the leader streamer steps down from the cloud creating an ionized channel to ground (or to another charge region). Once the channel is established the return stroke discharges the charge from to ground from the cloud over the course of $\sim$70 μs (the return stroke labeled in Figure 1.2). After the return stroke occurs the top of the ionized channel can connect to other charge regions through subsequent positive streamer processes; producing multiple return strokes along the same ionized channel. The total process, from step-leader through multiple return strokes, is considered a single lightning flash.

Lightning can discharge in several configurations with the most common shown in Figure 1.3. Lightning predominately occurs either between clouds or within the same cloud: cloud-to-cloud lightning, in-cloud lightning, or cloud lightning (Figure 1.3c and 1.3f). Lightning between the thunderstorm and ground is more powerful, easier to detect, and less common. It can occur in 4 combinations of positive or negative charge in the cloud dis-
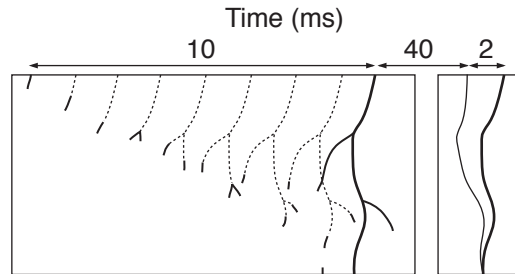
Figure 1.2: The step-leader and discharge process of cloud to ground lightning with typical times of each stage (adapted from *Ogawa* (1995)).

charging to ground with the discharge beginning in the cloud or on the ground (Figure 1.3a, 1.3b, 1.3d, and 1.3e). Negative cloud to ground lightning (moving negative charge from the cloud to ground, Figure 1.3a) is the most common cloud-to-ground discharge at a negative : positive ratio of 10 : 1. It is estimated that there are 4 in-cloud (IC) strokes (or flashes) for every cloud-to-ground (CG) stroke (*Uman*, 1969).

There are other sources for large electrical discharges that produce electromagnetic waves similar to thunderstorm lightning. For example: terrestrial gamma ray flashes, terrestrial luminous events, volcanic lightning, and compact intra-cloud discharges. However these events are less common than typical lightning and will not be discussed in this work.

1.1.2   The Ionosphere

The ionosphere is the plasma environment located between 80 km and 1000 km altitude; it is a highly conductive region within the exponentially decreasing atmospheric neutral density. It forms from the ionization of the background neutral particles by extreme ultraviolet sunlight. During the day the ionospheric extends farther down in altitude with the D- and E-regions; during night the ions in these regions recombine, seen in Figure 1.4. At higher altitudes the F-region remains through the night as the neutral particle density is much lower, preventing significant recombination.

For very low frequency waves (VLF, 1 kHz to 24 kHz) the conductive ionosphere and
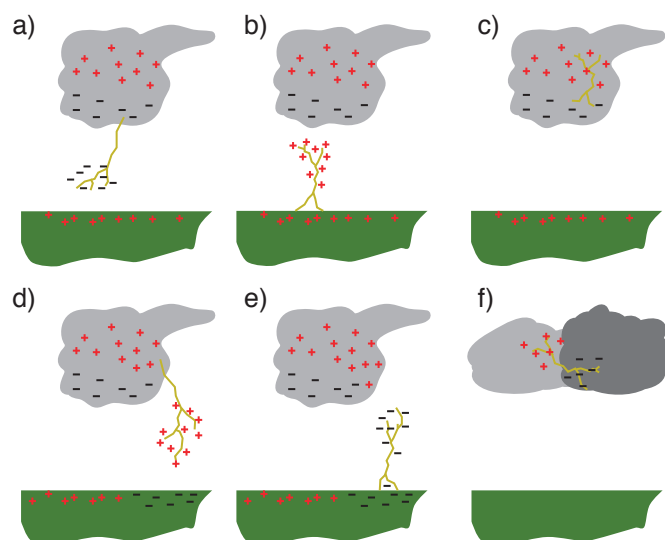
Figure 1.3: Different types of lightning discharges: a) negative cloud-to-ground, b) positive ground-to-cloud, c) in-cloud, d) positive cloud-to-ground, e) negative ground-to-cloud, f) cloud-cloud. Note that a) and b) produce the same overall charge movement (negative charge to ground) and as does d) and e). (Based on *Uman* (1969))

ground are good reflectors and form the Earth-Ionosphere Waveguide (EIWG). The EIWG allows for propagation of VLF waves from natural sources (e.g. lightning) and artificial sources (e.g. Navy VLF transmitters) to propagate large distances (> 5000 km). VLF waves above 5 kHz propagating in the waveguide undergo loss on the order of 1 dB/Mm with the lost energy heating the waveguide and generating plasma waves in the ionosphere (Figure 1.5). VLF wave attenuation is greatest over ice, then continents, with the lowest attenuation over the oceans.

1.1.3 Global Electric Circuit

The global electric circuit is the multiply connected current system formed by thunderstorms charging the ionosphere which leaks through the conductive atmosphere like a leaky spherical capacitor to ground, as shown in Figure 1.6a, with a very simple equivalent circuit shown in Figure 1.6b. Thunderstorms are the primary drivers of the global electric cir-
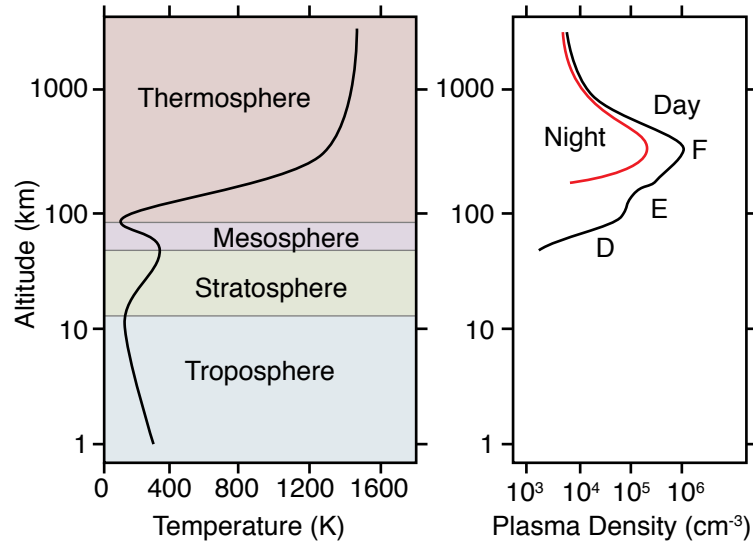
Figure 1.4: Atmospheric temperature profile and ionospheric plasma density profile (day ionosphere in black, night ionosphere in red).
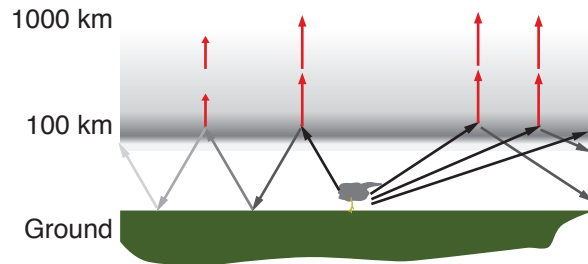
Figure 1.5: Schematic of the Earth-Ionosphere Waveguide.

cuit with the charged ionosphere discharging through fair weather atmosphere. The global circuit was first proposed after observing diurnal variation in global thunderstorm activity along with fair weather electric field measurements; originally observed by *Wilson* (1921) and *Whipple* (1929). Strong correlations between thunderstorm activity and fair weather return current led to the present model of the global electric circuit. The global electric circuit activity changes on short time scales that are not resolved with past models or with

long term averaged observations (*Holzworth et al.*, 1984). The global electric circuit is an important component to the solar-terrestrial system creating a link between solar activity, the ionosphere, aerosols, cloud microphysics, thunderstorms, weather, and climate (*Tinsley et al.*, 2007; *Holzworth and Volland*, 1986).
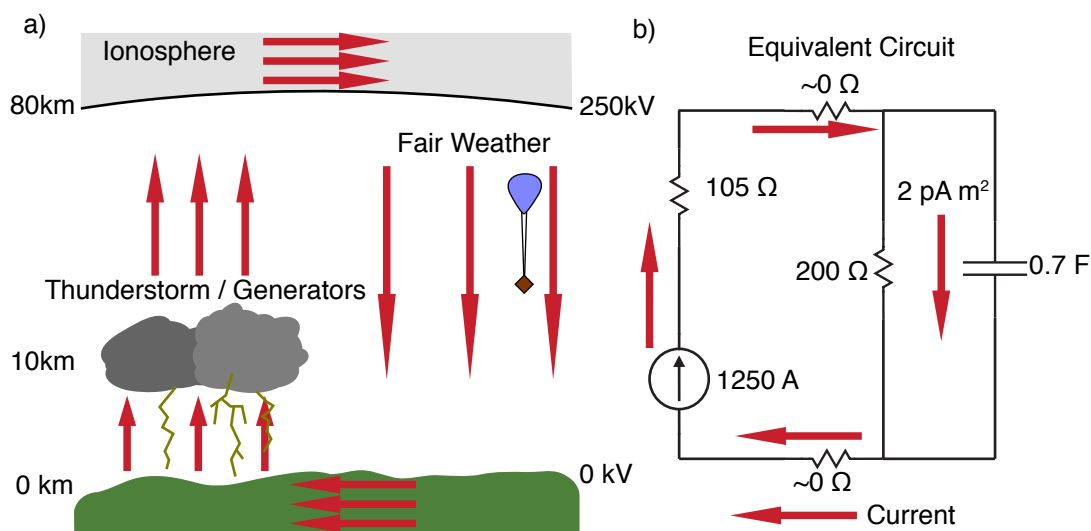
Figure 1.6: a) Global electric circuit model, arrow correspond to currents. b) Very simple circuit equivalent model with representative values.

1.2 Lightning Detection Systems

There is a growing importance, both scientifically and operationally, of ground based lightning detection networks. Lightning detection networks are being used in a larger gamut of research areas including: terrestrial gamma ray flashes (*Dwyer*, 2012; *Gjesteland et al.*, 2011; *Connaughton et al.*, 2010), lightning climatology (*Virts et al.*, 2013, 2011; *Bürgesser et al.*, 2012), ionospheric disturbances and probing (*Jacobson et al.*, 2010; *Singh et al.*, 2011b), transient luminous events (*Soula et al.*, 2011), global electric circuit (*Holzworth et al.*, 2005), and whistler observation (*Collier et al.*, 2010, 2011; *Burkholder et al.*, 2013). This is in conjunction with the extended usage of lightning networks operationally in weather

prediction and tracking (*Fierro et al.*, 2012; *Pan et al.*, 2010; *Thomas et al.*, 2010), volcano monitoring (*Doughton*, 2010), and hazard estimation (*Altaratz et al.*, 2010). With growing usage it is necessary to understand and improve the capabilities and efficiencies of the various available lightning networks.

Ground based total lightning networks distinguish themselves from other ground based networks and satellites by detecting and identifying in-cloud (IC) discharges as well as cloud to ground (CG) strokes. Lightning type is critical in understanding thunderstorm dynamics (*Williams et al.*, 1989), with real time monitoring of sudden increases of IC activity able to predict severe weather events (*Rudlosky and Shea*, 2013; *Darden et al.*, 2010; *Metzger and Nuss*, 2013; *Schultz et al.*, 2009, 2011). The higher operational frequencies of total lightning networks are also useful for researching narrow bipolar events (*Suszcynsky*, 2003) and large scale lightning behavior (*Hutchins et al.*, 2013a). Lightning mapping arrays are able to locally detect, locate, and distinguish IC activity, however total lightning networks have the advantage of much larger spatial coverage.

1.2.1 *World Wide Lightning Location Network*

The World Wide Lightning Location Network (WWLLN, see http://wwlln.net) determines the location for nearly all lightning producing storms around the globe in real time (*Jacobson et al.*, 2006). WWLLN has been generating global lightning locations starting in 2004 (*Rodger et al.*, 2006, 2009); since then, the network has grown from 18 stations to over 70 as of April 2014. Knowledge of individual stroke locations, with high temporal accuracy and within a fraction of a wavelength, is beneficial for both scientific and technical uses. WWLLN lightning location data have recently been used for advances in space science (*Lay et al.*, 2007; *Kumar et al.*, 2009; *Collier et al.*, 2009; *Holzworth et al.*, 2011; *Jacobson et al.*, 2011), meteorology (*Price et al.*, 2009; *Thomas et al.*, 2010), detailed lightning physics (*Connaughton et al.*, 2010), and volcanic eruption monitoring (*Doughton*, 2010). The observed global lightning stroke density for 2011 – 2012 is shown in Figure 1.7.

The network uses a time of group arrival (TOGA) technique, originally developed by *Dowden et al.* (2002), to locate strokes by analyzing the sferic waveforms at each station
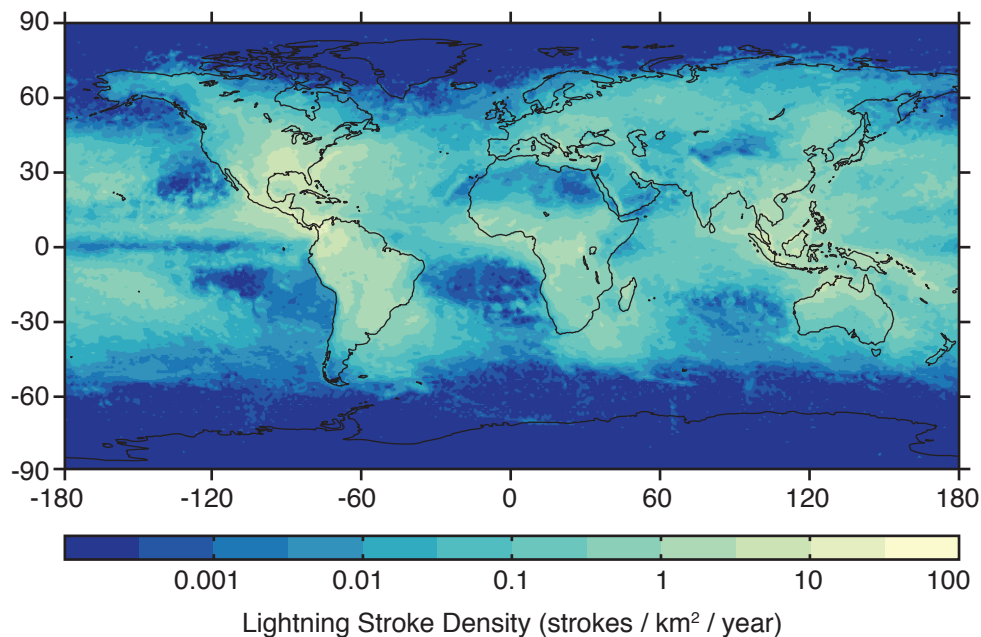
Figure 1.7: Global WWLLN lightning stroke density for 2011 – 2012.

using the Stroke_B algorithm as discussed by *Rodger et al.* (2006, 2009). WWLLN locates strokes by analyzing the TOGA of the sferic wave packet in the 6 – 18 kHz band (*Dowden and Brundell*, 2000). The TOGA of the VLF wave packet is used rather than the trigger time to produce more uniform arrival times across the network. A recent upgrade to the network allows for the measurement of the radiated VLF energy of located strokes within the 8 – 18 kHz VLF band (see Chapter 2). The stroke energy is calculated from the square root of the time-integrated squared electric field of the VLF sferic at each WWLLN station; it uses the Long Wave Propagation Capability code (*Ferguson* (1998), see Section 1.3) to estimate the sferic attenuation and calculate the radiated energy (*Hutchins et al.*, 2012b).

As stations are added the accuracy and detection efficiency of the network improves. In 2010 the network locates most strokes to within 5 km and <10 μs with an estimated detection efficiency of about 11% for all strokes and >30% for more powerful strokes (*Abarca et al.*, 2010; *Rodger et al.*, 2009). The network improves in accuracy and detection efficiency with added stations; for example an increase in the number of WWLLN stations from 11

in 2003 to 30 in 2007 led to a $\sim$165% increase in the number of lightning strokes located (*Rodger et al.*, 2009). However the WWLLN does not observe lightning with the same detection efficiency everywhere. This is due to variable WWLLN station coverage and the strong effect on VLF radio propagation from surface electrical conductivity and ionospheric conditions along the great-circle path of the wave.

1.2.2   Earth Networks Total Lightning Network

The Earth Networks Total Lightning Network (ENTLN) is a ground based network that began in 2009 as the Weatherbug Total Lightning Network (WTLN). It has two operational regimes: short range using broadband sferic waveforms (5 kHz – 12 MHz) and long range with only VLF/LF waveforms (1 Hz – 256 kHz) (*Heckman and Liu*, 2010). Correlations of the stroke waveform and amplitude from multiple stations determines the time, location, altitude, peak current, polarity, and type of the located stroke (*Liu and Heckman*, 2011). The network utilizes a time of arrival method to determine the location of each stroke; where a minimum of 8 stations is required to produce a valid solution. To compress the broadband waveforms each station removes the necessary amount of low-amplitude signal to reach the requisite packet size. In the continental United States (CONUS) the network has approximately 530 operational stations in September 2013. A comparison to the Oklahoma Lightning Mapping Array (OKLMA) in 2010 demonstrated the ability to discriminate between CG and IC strokes (*Beasley et al.*, 2010). The observed stroke density over North America for 2011 – 2012 is shown in Figure 1.8.

1.2.3   TRMM/LIS

The Lightning Imaging Sensor (LIS, 1997-present) is a satellite-based lightning detector flown onboard the Tropical Rainfall Measurement Mission (TRMM) satellite orbiting at a 35° inclination and 402 km altitude (*Christian et al.*, 1999). In low earth orbit it observes the total lightning activity from individual thunderstorms for 90 sec in each 0.5° × 0.5° viewtime granule. LIS is useful as it is an optical lightning detection system without the spatial dependency inherent with ground based networks, and the sensor performance has
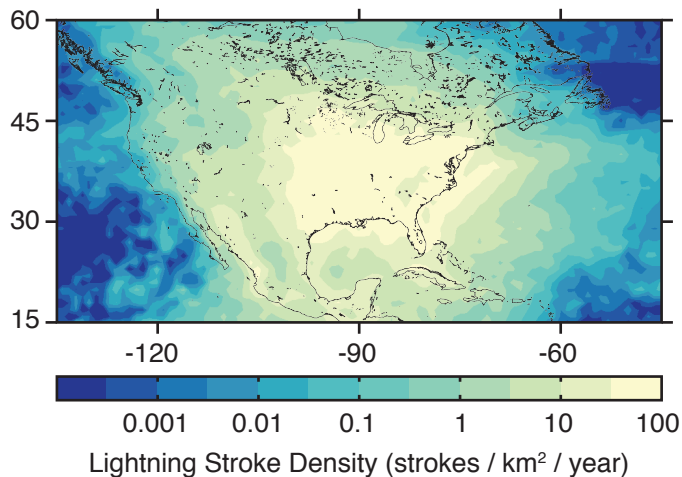
Figure 1.8: North American ENTLN lightning stroke density for the 2011 − 2012 dataset. Note same range as Figure 1.7

not changed over time. The LIS data are available at several processed levels; throughout this work the flash level data are used. The LIS flash density for 2011 − 2012 is shown in Figure 1.9.
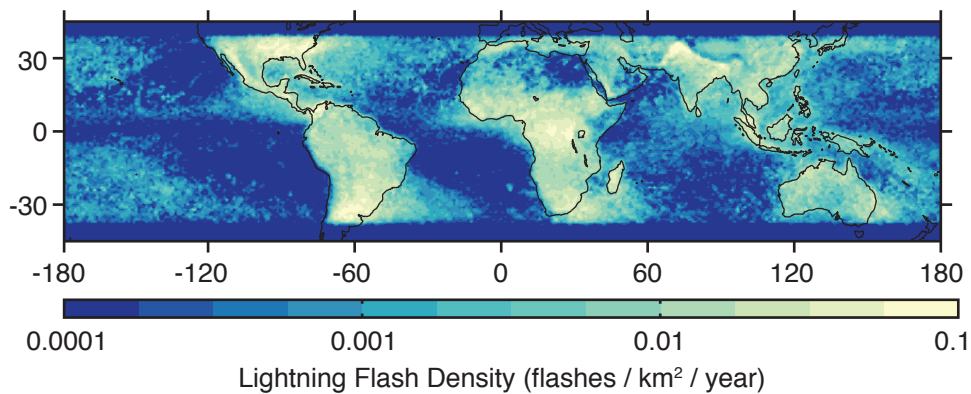
Figure 1.9: LIS lightning flash density for 2011 − 2012. Note the reduced range from Figures 1.7 and 1.8

1.3 Long Wave Propagation Capability Code

The Long Wave Propagation Capability code (LWPC) is used to model the VLF attenuation between WWLLN located lightning and the network stations. The LWPC code was developed by the Space and Naval Warfare Systems Center by *Ferguson* (1998) and has most recently been validated by *McRae and Thomson* (2000) and *Thomson et al.* (2011). In this research we made use of an adapted LWPC version 2.1 (available online, see Appendix A).

LWPC can be used to model the attenuation along a mixed day/night ionospheric path, however due to computational limitations a lookup table is used instead. The lookup tables model the received electric field at a given station for a 100 kW transmitter in each grid cell. A $1°$ by $1°$ grid is used for either an all day ($\beta = 0.3$ km^{-1} and $h' = 74$ km) or an all night ($\beta(f) = 0.3 - 0.8$ km^{-1} and $h' = 87$ km) ionospheric model, where β and h' are the slope of the conductivity (β is frequency dependent at night) and the reference height. The ionospheric models are the default models of LWPC and fully described in *Ferguson* (1998).

For each grid cell the electric field is averaged over the $8 - 18$ kHz band, capturing the frequencies of the peak radiated power from lightning (*Volland*, 1995). An example of the day ionosphere lookup table for the Dunedin station is shown in Figure 1.10. The discontinuity of electric field over Greenland and in the South Atlantic is caused by the high attenuation rate of VLF propagating over ice. Using the lookup tables assumes that the attenuation rates do not vary greatly within a given grid cell and that the frequency response of a sferic is relatively flat within the band considered.

1.4 Outline

An outline of this work is provided in Table 1.1 along with the corresponding publications adapted for each chapter. Chapters $2 - 4$ cover the technical improvements and analysis of the networks used, work that is necessary for Chapter $5 - 8$.

The appendices discuss the code implementation of the techniques described in a few of the chapters along with the construction, design, and operation of the WWLLN stations. The descriptions are given in Table 1.2. Notably the code for the energy calculations (Chapter 2), relative detection efficiency model (Chapter 3), and thunderstorm clustering
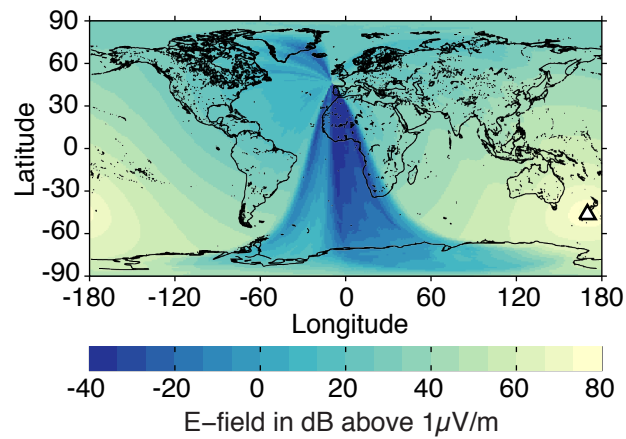
Figure 1.10: LWPC generated lookup table for Dunedin station (white triangle) using an all day ionospheric model ($\beta = 0.3$ km^{-1} and $h' = 74$ km) averaged over $8 - 18$ kHz. Each $1°$ by $1°$ bin shows the electric field seen at Dunedin if a 100kW transmitter is centered on that bin.

(Chapter 7) are available online as described in Appendix A. The same appendix contains the schematics, EAGLE files, parts list, and software for the WWLLN Service Units.

Table 1.1: Chapter outline.

Chapter 2	Describes the process for calculating the far-field radiated VLF energy of lightning with WWLLN.	*Hutchins et al.* (2012b)
Chapter 3	Discusses the relative detection efficiency model of WWLLN.	*Hutchins et al.* (2012a)
Chapter 4	Examines the detection efficiency between the ENTLN and LIS lightning data over North America.	Under Review
Chapter 5	Examines the energy contrast between oceanic and continental lightning.	*Hutchins et al.* (2013a)
Chapter 6	Estimates the VLF attenuation rates over ocean under the day and night ionosphere.	*Hutchins et al.* (2013b).
Chapter 7	Develops WWLLN thunderstorm clusters to create a prediction of the global electric circuit activity.	*Hutchins et al.* (2014).
Chapter 8	The thunderstorm clustering is expanded to flash clustering and explored in the context of thunderstorm accuracy.	
Chapter 9	Explores the future work that can be carried forward from this work.	

Table 1.2: Appendix outline.

Appendix A	Describes the different types of WWLLN data files available, their location, and the common code used in this thesis.
Appendix B	Describes the WWLLN stroke energy processing code and operations, enabling reproduction or reprocessing of the energies.
Appendix C	Describes the development, design, and construction of the WWLLN Service Unit v4.
Appendix D	Describes the software and operations of the WWLLN Service Unit v4 computers.
Appendix E	Outlines the new WWLLN website structure and operation along with the realtime lightning display.

Chapter 2

STROKE ENERGY

2.1   Overview

The World Wide Lightning Location Network (WWLLN) is a long range network capable of locating lightning strokes in space and time. While able to locate lightning to within a few kilometers and ten microseconds, the network currently does not measure any characteristics of the strokes themselves. The capabilities of the network are expanded to allow for measurements of the far-field energy from the root mean square electric field of the detected strokes in the 6 – 18 kHz band. This is accomplished by calibrating the network from a single well calibrated station using a bootstrapping method. With this technique the global median stroke energy seen by the network is 1.3 x 10^3 J with an average uncertainty of 17%. The results are validated by comparing to return stroke peak current as measured by the New Zealand Lightning Detection Network and to previous energy measurements in the literature. The global median stroke energy is found to be 2% of past estimates of the average radiated electromagnetic energy. Accounting for the different observational distances, we find our far-field observations of the waveguide mode are consistent with the previous literature. This study demonstrates that the WWLLN determined energies can be used to estimate the return stroke peak currents of individual lightning strokes occurring throughout the globe.

As only the spectral variations through the sferic wave packet are needed for determining the TOGA, the absolute electric field amplitude is unnecessary to accurately locate lightning. Due to this, the network does not currently report other characteristics of strokes such as peak field. However, each WWLLN station does record the root mean square (RMS) electric field value of the sferic waveform used in the TOGA calculation, but these values need to first be calibrated.

The station operated by the University of Otago near New Zealand's Antarctic station, Scott Base, was calibrated by a field team in December 2009. The field team injected a series of calibration signals, ramped progressively in frequency, through the crossed magnetic loops of the Scott Base antenna and computed the calibration from the equivalent electric field. This calibration of a single station allows for the calculation of stroke energy as seen by the network.
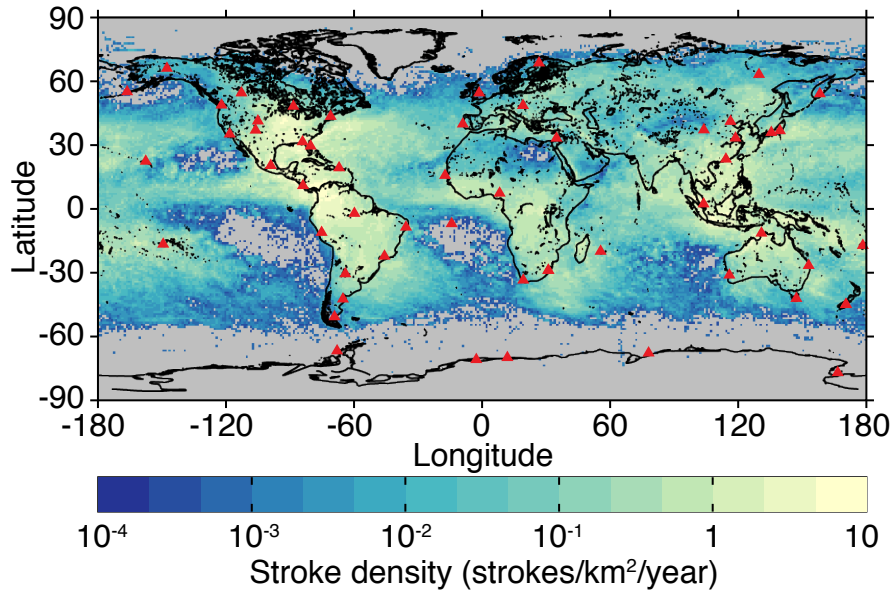
Figure 2.1: WWLLN 2010 global stroke density on $1° \times 1°$ grid, station locations shown with black triangles. Data processed with the Stroke_B algorithm.

A previous study by *Rodger et al.* (2006) attempted to calibrate the network through observations of narrow-band VLF communication transmitters at each WWLLN station. These observations were combined with the U.S. Navy Long Wave Propagation Capability (LWPC) code, described by *Ferguson* (1998), to predict what the received amplitude should be in order to calibrate each station. However the study assumed that the frequency response and calibration of the sound cards was the same across the network, assumptions that are false. In fact, the sound cards, preamplifier, and antenna can vary in their sensitivity by a factor of 10 . The current study utilizes a broader range of frequencies and calibrates each station such that differing frequency responses are accounted for in the calibration.

Measuring stroke energy is an important step forward as it allows the network to make real time measurements of the strength of lightning worldwide. Measuring characteristics of the strokes will allow for insights into thunderstorm evolution, large scale storm phenomena, and global effects of lightning. For example, stroke energy values could help current research on terrestrial gamma ray flashes by *Briggs et al.* (2011) constraining efficiencies and source

mechanisms, and tropical hurricanes by *Thomas et al.* (2010) could utilize the energy in analyzing eyewall replacement.

As we will show in this Chapter, the network measures a median global VLF energy in the far-field waveguide mode of 1.3×10^3 J with an average uncertainty of 17%. Previous measurements have shown the energy radiated by strokes is often near 7.0×10^4 J (*Taylor*, 1963). Past measurements were measured much closer to that strokes (100 km) and normalize to closer distances, resulting in a measurement of both the sky- and ground-wave of the sferic. WWLLN measures the RMS electric field at distances where only the waveguide mode of the sferic remains. When these factors are accounted for the median energy from WWLLN located strokes is comparable to the previously reported value of 7.0×10^4 J electromagnetic energy.

2.2   Instrumentation and Data Processing

In order to calculate the stroke energy from WWLLN three steps are necessary: calibrate each station in the network, measure RMS electric field of a stroke at each station, and calculate the stroke energy needed to produce the electric field at each station.

2.2.1   Station Electric Field

Each WWLLN station consists of four main components: the antenna, a preamplifier, a service unit for signal and power management, and the sound card that digitizes the measured fields. The difficulty in making an energy measurement with the network arises in calibrating for the coupling between a short (2m) station antenna to a signal with an approximately ∼10 km wavelength. When a station digitizes the electric field waveform it stores it in uncalibrated sound card units (SCU). Additionally, the effective gain and calibration differs at each station due to the preamplifier, antenna construction, soundcard sensitivity, and local environmental conditions. This results in the RMS electric field being reported in station specific sound card units.

The average power spectra from 194 stroke waveforms recorded at the Tallahassee, Florida station are shown in figure 2.2 along with the frequency response of a typical preamplifier. The strokes used were located at distances of 5000 km to 10000 km from the station

and were recorded between 18:00 and 21:00 UTC on May 3 and May 9 2011. As can be seen in the figure the average power peaks between $6 - 18$ kHz with the analog response remaining relatively flat through the entire frequency range. The spikes in the power spectra are a result of manmade VLF communication transmitters.
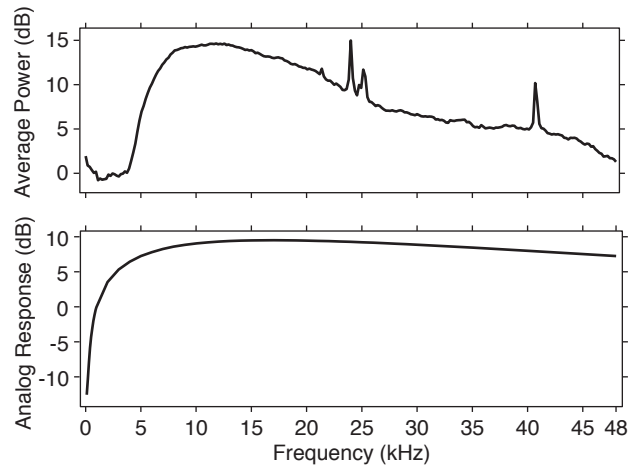
Figure 2.2: The top panel shows average power spectra from 194 stroke waveforms recorded at the Tallahassee, FL station between 0 and 48 kHz. The strokes were located 5 Mm – 10 Mm away from the station on May 3 2011 and May 9 2011 between 18:00 and 21:00 UTC. The bottom panel shows the frequency response of a preamplifier.

The waveforms used are 1.33 ms long with 0.33 ms pre-trigger samples and 1 ms post-trigger samples. Prior to processing the waveform is put through a $6 - 18$ kHz 16 point finite impulse response (FIR) bandpass filter. The RMS value of the resultant waveform is stored in uncalibrated sound card units. After being calibrated to the ∼10 km wavelength signal the SCU measurement can be converted into the RMS electric field of the sferic.

2.2.2 LWPC and Energy

To calculate the stroke energy based on the RMS electric field at a station the LWPC code is used to model the attenuation, as a function of frequency, between a transmitter at the stroke location and the receiving station. The LWPC code was developed by the Space and

Naval Warfare Systems Center by *Ferguson* (1998) and has most recently been validated by *Thomson et al.* (2011). In this research we made use of LWPC version 2.1.

With a known conversion from a station's SCU value to V m^{-1}, A_{local}, of the lightning waveform, power is calculated using equation 2.1. The ratio from LWPC between a 100 kW transmitter and the modeled field (given in dB above 1 μV m^{-1}) is used to account for the sferic attenuation, α, along the path for every grid location for every detector. Since energy of the stroke is proportional to the time integrated square electric field, and the power measurements are from the RMS electric field value, the energy of the stroke can be found from the size of the recording window: $U_{stroke} = P_{stroke} * t_{record}$, with the current recording window set at 1.33 ms. WWLLN stroke strength values measured and recorded in terms of radiated energy.

$$P_{stroke} = \frac{E_{scu}^2}{A_{local}^2} * \frac{100kW}{(10^{\alpha/20}\mu V/m)^2} \tag{2.1}$$

$$U_{stroke} = P_{stroke} * t_{record} = \frac{E_{scu}^2}{A_{local}^2} * \frac{100kW}{(10^{\alpha/20}\mu V/m)^2} * t_{record} \tag{2.2}$$

Due to computing limitations running the LWPC code we cannot conduct a full run for every stroke-station pair in real time. Instead a lookup table is used which breaks stroke locations into 1° by 1° bins and uses either an all day ($\beta = 0.3$ km^{-1} and $h' = 74$ km) or an all night ($\beta(f) = 0.3 - 0.8$ km^{-1} and $h' = 87$ km) ionospheric model, where β and h' are the slope of the conductivity (β is frequency dependent at night) and the reference height in the ionospheric model. The ionospheric models are the default models of LWPC code and fully described in *Ferguson* (1998). To account for transitions across the terminator a weighted average of the day and night electric field values are used. The lookup tables give the electric field averaged over the 8 – 18 kHz band which captures the frequencies of the peak radiated power from lightning (6 – 7 kHz omitted due to code limitations) (*Volland*, 1995). An example of the day ionosphere lookup table for the Dunedin station is shown in figure 2.3; the discontinuity of electric field over Greenland and in the South Atlantic is caused by the high attenuation rate of VLF propagating over ice.
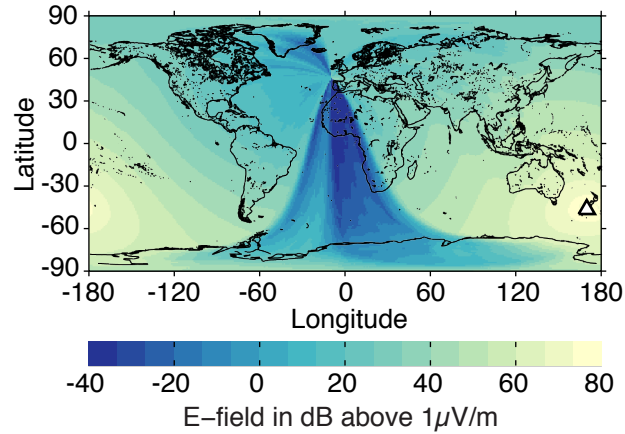
Figure 2.3: LWPC generated lookup table for Dunedin station (white triangle) using an all day ionospheric model ($\beta = 0.3$ km^{-1} and $h' = 74$ km) averaged over $8 - 18$ kHz. Each $1°$ by $1°$ bin shows the electric field seen at Dunedin if a 100 kW transmitter is centered on that bin.

2.2.3 Calibration and Bootstrapping

With one calibrated station it is possible to find the calibration of other nearby stations; the process is shown in figure 2.4. Using a well calibrated station (on the right) the stroke energy of a given stroke is found using LWPC for the same stroke. The uncalibrated station also finds the energy using LWPC, however instead of an energy in joules, it measures the energy in sound card energy, SCE. The ratio between the two energy values will give the calibration factor, A^2_{local}, of the second station. This is repeated for many strokes with the median of the conversion factors used as the conversion factor between the two stations.

To calibrate the entire network off of a single station a bootstrapping method is used. Station to station calibrations are done using strokes that have all day paths to both stations and are within $1000 - 8000$ km of both stations. All day paths were chosen as the daytime ionosphere is modeled more accurately by LWPC than the night ionosphere (*McRae and Thomson*, 2000).

The first set of calibrations are done between the well calibrated station and those with common strokes that have the desired path characteristics. Once calibrated these stations
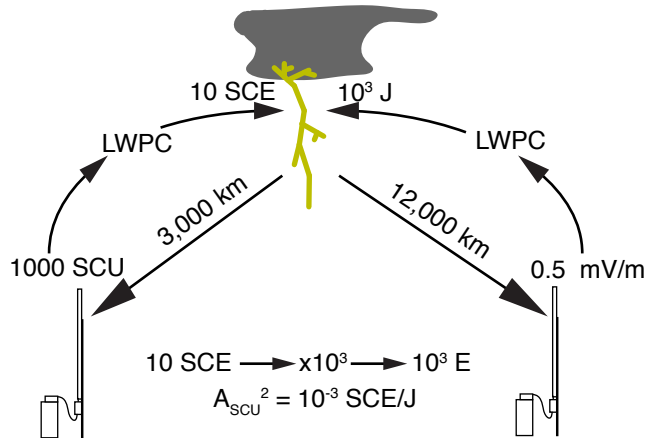
Figure 2.4: Method of calibrating one station to another using LWPC.

are used to calibrate the next set of stations, and these newly calibrated stations are used to find the next set. This process is repeated until no further stations can be calibrated; Figure 2.5 is an example of this process. Not all stations are calibrated for each day; they may not be calibrated if they do not see any common strokes with another station that match the path requirements or if their calibration to the next set of stations does not match direct calibrations. For example, if station A calibrates B and then B calibrates C then if the calibration path ABC does not match the well calibrated path of AC it is determined that B is not well calibrated, so it is not used for the day.

2.2.4 Energy Calculation

The fully calibrated WWLLN network is used to calculate the stroke energy for each station participating in a TOGA event using equation 2.2 with the A_{local} values known for a majority of the stations. Of the participating stations, the median of their energy measurements is used as the final energy value for the event. The uncertainty in the energy is the median absolute deviation (MAD) of the participating station energy measurements, the MAD is the method of getting standard deviation of median values ($\mathrm{MAD} = \mathrm{median}(|U_i - \mathrm{median}(U_i)|)$). On average 96.7% of WWLLN strokes have an energy value, even with only an average of 2/3 of stations being well calibrated and participating in energy calculations. The 3% of
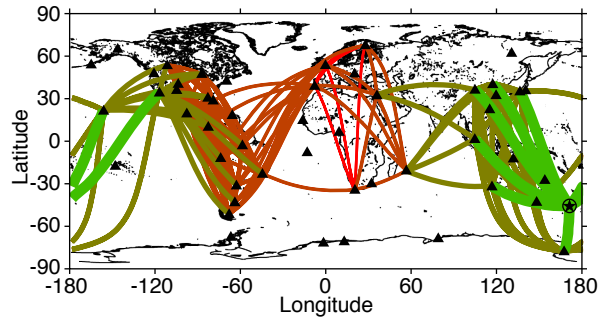
Figure 2.5: An example of the bootstrapping technique, showing calibration distance from the main Dunedin station. Thick green lines are the first calibration stage and the thin red lines the last. Stations may be unconnected due to not having common strokes, being poor intermediary stations or being down for the day.

strokes without energy values either were not reported by any well calibrated stations or had an uncertainty greater than 100% and are removed.

The last step in calculating the energy per stroke is an iterative technique to improve accuracy. The first set of energy values are used as a basis to recalibrate all of the stations in the network. These new values are used to recalibrate again with this process repeating several times until the station calibrations converge to a stable and final value. Currently the station calibrations along with this iterative technique are performed once a day for calculating the stroke energies of that day.

The bootstrap method of calibration is an internally consistent method of determining the conversion factors at each station. The iterative technique allows for a convergence of calibrations to reduce the uncertainty in the energy measurements. Even the original calibrated station is updated in each iteration, removing errors that may be introduced by local effects since the initial ground truth calibration. However the internal consistency could be improved and monitored through the introduction of a second well-calibrated station; such a station would enable additional diagnostics and error correction not possible with a single starting station.

2.3 Results

2.3.1 Validation

A study was conducted comparing the stroke energy values determined by WWLLN to the ground based NZLDN estimates of return stroke peak current (NZLDN described in *Rodger et al.* (2006)). The comparison was done using three periods of high lightning activity over New Zealand: $25 - 27$ August 2009, $26 - 27$ September 2009 and 21 October 2009. WWLLN strokes were considered to match NZLDN strokes if they occurred within 0.5 ms and 400 km of a NZLDN detector, the same criteria used by *Rodger et al.* (2006). From the comparison the empirical relation between return stroke peak current and $6 - 18$ kHz radiated energy was found to be:

$$U_{stroke} = 2.23 * |I_{peak}|^{1.62} \qquad (2.3)$$

where U_{stroke} has units of joules and I_{peak} has units of kA. In figure 2.6 the WWLLN peak current (using the inverse of equation 2.3) is shown against the NZLDN absolute peak current. When taking the uncertainties of the energy values (converted to peak current) and an assumed 30% uncertainty in the NZLDN data 84% of the matched strokes have equivalent peak currents. The peak current values fit close to the unity line with a robust linear fit of $I_{WWLLN} = 0.93 * I_{NZLDN} + 1.93$ with an R^2 value of 0.92; a robust fit is used due to the lognormal behavior of the WWLLN energy data. Of the matched strokes 86.5% are shown in figure 2.6 with the remaining 13.5% out of the plotted bounds. This strong relationship confirms that the energy values measured are directly related to the physical properties of the stroke (i.e. the return stroke peak current).

2.3.2 Error and Uncertainty

The model used in relating the stroke energy to the peak current uses the 2010 average stroke uncertainty of 17% from the median of the MAD distribution of all strokes in the year. If a stroke has an uncertainty greater than 100% that value is thrown out, doing so only decreases the number of stroke energies by 3%.
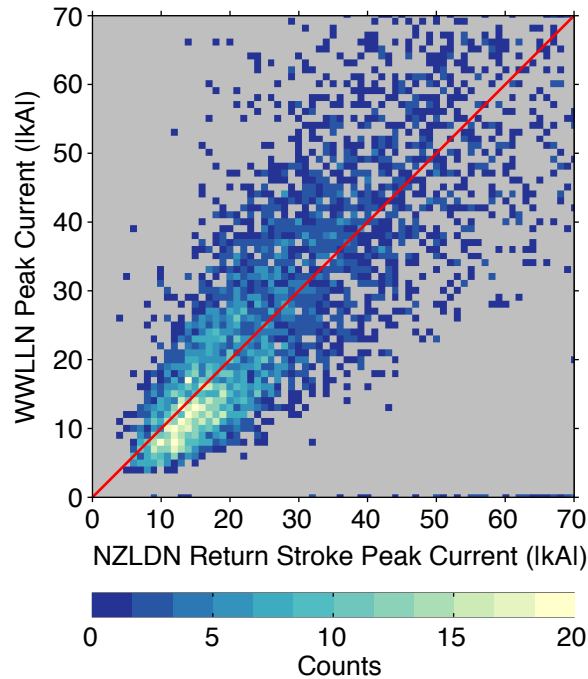
Figure 2.6: WWLLN peak current versus NZLDN return stroke peak current for three time periods in 2009 using 5260 matches. WWLLN peak current derived from $E_{stroke} = 2.23 * |I_{peak}|^{1.62}$, 84% of strokes are within range of the unity line (red solid line) with uncertainty taken into account. 86.5% of NZLDN-WWLLN matched strokes shown (others out of range).

The largest source of uncertainty in the calculations arises from the assumptions that are made using the LWPC code and in the calibration process. The lookup tables used to calculate attenuation on a given stroke are gridded into $1°$ by $1°$ bins and averaged over the $8 - 18$ kHz frequency range. This assumes that the attenuation rates do not vary greatly within a given grid and that the frequency spectra of a sferic is relatively flat within the band considered. The ionospheric models used within the code assume a perfectly smooth day or night ionosphere, paths crossing the terminator are weighted averages of these values. These issues arise from the inherent speed limitations of the LWPC code.

A secondary source of uncertainty manifests as variations in the WWLLN station cali-

bration factors. These are not caused by drift in the electronics, but from variations in the local station environment and the ionosphere along the path. Local weather could change the gain at a station (for example through water or ice on an antenna) or if there are significant changes in conductivity of the ground or ionosphere. Since the LWPC code results are fixed, the only free variable in the bootstrapping process are the station conversion factors, so variations in the propagation path manifests within these factors. These uncertainties are mitigated by performing a seven day running average of the calibration values but they are still present as daily variations in the calibrations.

Uncertainties that arise through the use of the LWPC model and through the station calibrations are seen through the MAD of each stroke. Without these uncertainties each stations should agree on the detected stroke energy, however the variations between each station, due to changes in the propagation path or in the calibration, appear as differences in reported stroke energy. Even with these differences the overall 17% uncertainty of the WWLLN energy measurements is comparable to the 13% uncertainty in the peak current estimates of the U.S. National Lightning Detection Network (NLDN) (*Nag et al.*, 2011).

2.4 Discussion

2.4.1 Comparison to Literature

The current method of measuring stroke energies using WWLLN results in a global median energy of 1.33×10^3 J for 2010. However it has been shown in past literature that the average radiated energy for a typical stroke is 6.9×10^4 J (*Taylor*, 1963). A 17.2 dB increase is needed to bring the WWLLN stroke energies into the range of those in the literature. The main difference in the measurement method is the difference in attenuation between nearby measurements of the sferic with measurements made at a larger distance.

The analysis of this difference was done using the raw waveform data from three stations: the secondary Seattle station, the Canaveral station and the Tallahassee station. Data was taken from between 20 April 2011 and 9 May 2011. In this interval 198 events were selected that had a clear waveform at the network trigger time.

The biggest effect on the received stroke energy is caused by the distances involved in

the measurements. Particularly the difference between the ground and sky wave near the stroke and the wave propagating in the waveguide. Most of the VLF energy measurements in the literature have measured waveforms near 100 km distance from the stroke. At larger distances only the lowest order mode is propagating in the Earth-ionosphere waveguide; at near distances other modes have not been attenuated by the structure of the waveguide.

To measure the effect that the nearby high attenuation will have compared to a signal in the waveguide the difference in received power of the VLF transmitter near Seattle, NLK (250 kW radiated power at a frequency of 24.8 kHz, i.e. *Clilverd et al.* (2009)), was compared to the same signal seen at two stations in Florida. A second transmitter in Hawaii, NPM (500 kW radiated power at a frequency of 21.4 kHz), was used as a reference for both stations. The two Florida stations were chosen as they are far from both VLF transmitters with a sampling frequency of 96 kHz (Nyquist frequency of 48 kHz), well above the frequencies of the lightning peak power and the VLF radio transmitter frequencies.

The LWPC code estimate for the transmitter signal at the WWLLN stations is compared to the measured transmitter signal to determine the conversion factor for each event. The conversions are found to be consistent with calculations and calibrations reported in this study for broadband lightning produced signals. Based on the verification of the LWPC code performed by *Thomson* (2010), LWPC is confidently used as a ground truth for this comparison.

The conversion factors from the NLK and NPM signals present in the selected waveforms were used to calculate the power of the two transmitters using the same method as the WWLLN energy calculations (Equation 2.2). To determine the importance of the ground wave, the ratio of estimated NLK to NPM powers at Seattle were divided by the NLK to NPM ratio at the two Florida stations. This ratio is necessary, instead of just the Seattle NLK to actual NLK power ratio, in order to normalize out any intrinsic errors in the calculation process. Measuring the waveguide signal causes a loss of 17.9 dB to 23.8 dB on the signal power as seen by WWLLN compared to the corresponding nearby ground and sky wave measurements. This loss is near the 17.2 dB difference between the measured WWLLN stroke energies and those made by *Taylor* (1963).

This analysis leads to the conclusion that the energy being measured by WWLLN is

not the total radiated energy of the stroke, rather it is the energy radiated into the Earth-ionosphere waveguide in the 6 – 18 kHz band.

2.4.2 Energy Distribution

Stroke energy in a given thunderstorm, region, or time span closely follows a lognormal distribution (*Golde*, 1977). Figure 2.7 shows the stroke energy distribution for all strokes seen by WWLLN in 2010 along with the distributions for the three major chimney regions of the Americas, Africa/Europe and Asia/Australia. The Americas and Asia/Australia show similar distributions, likely a result of having similar WWLLN coverage, while Africa/Europe differ, possibly due to uneven network coverage.
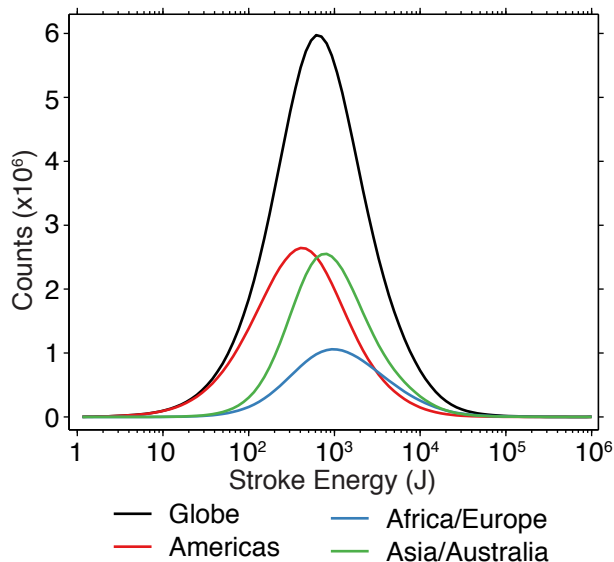
Figure 2.7: Histogram of stroke energies for 2010 with 100 logarithmically spaced bins, the histogram for the globe (1.4 x 10^8 strokes) is shown in black, the Americas (6.1 x 10^7 strokes) in blue, Africa/Europe (2.4 x 10^7 strokes) in red and Asia/Australia (5.0 x 10^7 strokes) in green. Error bars are too small to display.

When the energy distribution of figure 2.7 is converted to peak current using equation 2.3 it can be compared to earlier measurements of peak current distributions. We can compare

the WWLLN peak current distribution to that of *Popolansky* (1972) as shown in *Golde* (1977). That study used peak current data from 624 return strokes to create a cumulative probability distribution of stroke currents as shown in Figure 2.8. A similar frequency distribution was created using the WWLLN peak current estimates and is shown in the same figure. As can be seen, the WWLLN distribution is shifted by a factor of two compared to the previous distribution, which is expected as WWLLN is known to preferentially detect higher current strokes when compared to what is seen by regional ground based networks (*Abarca et al.*, 2010). In a comparison to the NLDN those authors found that the probability distribution function of the WWLLN coincident strokes was similarly of an order two higher than the distribution for all NLDN strokes. These comparisons further validate the stroke energy to peak current relation of equation 2.3 and demonstrates this relationship is likely valid for the entire global network, and not just the New Zealand region.
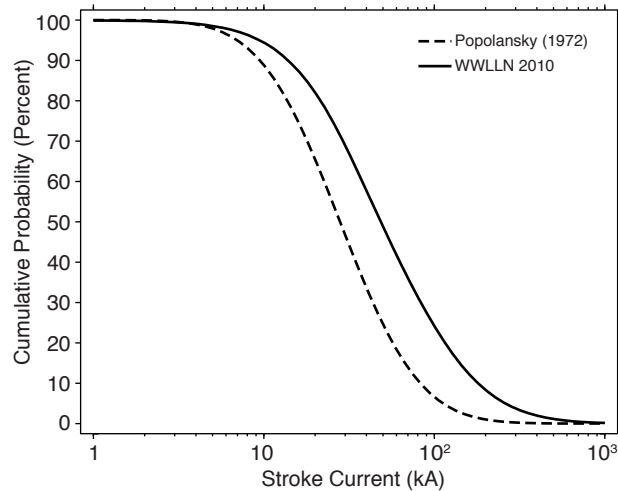
Figure 2.8: The cumulative probability distributions of stroke current for the best fit of the *Popolansky* (1972) data (dashed) and the 2010 WWLLN dataset (solid).

Africa has a lower detection efficiency in the network due to the small number of stations on the continent. With a lower detection efficiency only the strongest strokes are seen by the more distant stations while the weaker strokes are not seen by these stations due to the stronger attenuation of VLF over continents than for propagation paths over water (*Wait*,

1970). An initial effort to model the detection efficiency of WWLLN showed this effect very strongly (*Rodger et al.* (2006), Fig. 11). This results in the distribution of figure 2.7, with fewer low energy strokes occurring over Africa compared to the other regions.

Without a full understanding of the regional detection efficiency of the network it cannot be determined whether every region follows the same lognormal energy distribution or whether local environments affect the energy of strokes in the region. However, the bootstrap calibration process described in the current study allows new estimates of the global variation in WWLLN detection efficiency in Chapter 3.

2.5 Conclusion

A new method of measuring the VLF waveguide mode energy radiated from lightning using the WWLLN has been shown and validated. While not the total radiated energy of the strokes, the energy measured is directly related to the peak current and therefore to inherent properties of the strokes. Our study shows that WWLLN observations can provide realistic return stroke peak current measurements in addition to the timing and location of global lightning activity.

The method developed here can be applied to similar ground based lightning detection networks. The key components in applying this method are the availability of a propagation model in the frequency range of the sensors (LWPC in the case of WWLLN), a measure of stroke strength (e.g. peak electric field), and at least one well-calibrated station. Finally the station density needs to be high enough such that the average inter-station distance is within the applicable range of the propagation model used.

An enhanced WWLLN allows for a global real time view of lightning with the ability to distinguish between weak and strong strokes. This will allow for future research into areas such as the long range evolution of thunderstorms over oceans, wider surveys of terrestrial gamma ray flashes, and with a more complete understanding of detection efficiency an analysis of global lightning energy output.

32

Acknowledgments for this Chapter

Chapter 3

WWLLN DETECTION EFFICIENCY

3.1 Overview

Using the detected energy per strokes of the World Wide Lightning Location Network (WWLLN) we calculate the relative detection efficiency for the network as if it had a uniform detection efficiency. The model uses the energy statistics of located strokes to determine which stations are sensitive to what stroke energies. We are then able to estimate the number of strokes that may be missing from any given regions as compared to the best, most sensitive regions of the WWLLN network. Stroke density maps can be corrected with the knowledge of how sensitive various regions of the network are operating.

This new model for the relative WWLLN detection efficiency compensates for the uneven global coverage of the network sensors as well as variations in very low frequency (VLF) propagation. The model gives a way to represent the global distribution of strokes as if observed by a globally uniform network. The model results are analyzed in spatial and temporal regimes, and the effects of a single VLF detector going offline are investigated in areas of sparse and dense detector coverage. The results are also used to show spatial, temporal, and energy distributions as seen by the detection efficiency corrected WWLLN.

The WWLLN network does not observe lightning with the same detection efficiency everywhere. This is due to variable WWLLN station coverage and the strong effect on VLF radio propagation from surface electrical conductivity and ionospheric conditions along the great circle path of a wave. This chapter demonstrates a technique which only uses data collected by the WWLLN network itself to estimate the relative detection efficiency of each $1° \times 1°$ pixel over the earth compared to the best average WWLLN detection efficiency. For instance, the lightning stroke density over central Africa, where WWLLN station density is sparse, can now be compared to the region of the Earth with the best detection efficiency, such as North America. This chapter does not provide an absolute detection efficiency calculation.

A concern for all VLF networks is the non-uniform propagation of VLF waves due to changing ionospheric and surface conditions; this is true for networks monitoring lightning produced VLF signals like WWLLN, or those monitoring fixed-frequency communication transmitters like AARDDVARK (*Clilverd et al.*, 2009). During the day there is a larger

ionospheric electron density at lower D-region altitudes; this lowers the effective reflection height of the sferics, increasing the attenuation rate of the sferics. This increase in electron number density is also seen in the change of the reference ionospheric height, h' (*Wait and Spies*, 1960), during the day ($h' = 74$ km) compared to during the night ($h' = 87$ km). There is a similar change in attenuation over the path of the sferic from the differences in the conductivity of the oceans (4 S/m), continents ($10^{-2} - 10^{-4}$ S/m), and Antarctic/Arctic ice (10^{-5} S/m). The many path parameters for a given sferic result in a highly variable attenuation (*Volland*, 1995).

Thus, independently determining the real-time detection efficiency has always been a challenging topic. Several studies have been conducted comparing WWLLN to other ground based networks or satellite measurements (*Lay et al.*, 2004; *Jacobson et al.*, 2006; *Rodger et al.*, 2009; *Abarca et al.*, 2010; *Abreu et al.*, 2010). These studies tend to be limited in either scope or in time due to the availability of data from other lightning detection networks. Past work by *Rodger et al.* (2006) attempted to determine the global detection efficiency of WWLLN using a theoretical model linked to observations from a ground based commercial lightning network in New Zealand. In this chapter a new method is developed for determining the relative detection efficiency of WWLLN based upon the recent network expansion of measuring the radiated energy of detected strokes (Chapter 2, *Hutchins et al.* (2012b)).

Developing a model of detection efficiency expands the capabilities and uses for WWLLN. In particular a model that does not rely on external comparisons to other networks or sensors is critical for obtaining a dynamic global view of network performance. Such a view will enable the network to be used with more confidence in areas of lower coverage and enable the network to be utilized with uniform detection efficiency in work requiring lightning rates and densities. This uniform performance will allow for more accurate studies of global phenomena such as the short time (<10 minute) variability of the global electric circuit, comparative lightning climatology between regions, and production rate estimations of transient luminous events and terrestrial gamma ray flashes. The detection efficiency model can combine with the measurements of stroke energy and regional absolute detection efficiency studies to advance research in global effects of lightning such as estimating the

total sferic energy transferred to the magnetosphere in the form of whistler waves.

3.1.1 Calculating the Radiated Stroke Energy

Every WWLLN sferic packet includes the TOGA and a measure of the root mean square (RMS) electric field of the triggered waveform. The RMS electric field is taken in the 6 – 18 kHz band over the triggering window of 1.33 ms. The U.S. Navy Long Wave Propagation Capability (LWPC) code described by *Ferguson* (1998) is utilized to model the VLF propagation from each located stroke to determine the necessary stroke energy to produce the measured RMS electric field (in the VLF band) at each WWLLN station. Using the measured RMS field at each station, the radiated energy of each detected stroke is found. In 2010 WWLLN observed a global median stroke energy of 630 J, with a 25% average uncertainty in the measured energy. The global and regional distribution of energy is shown in Figure 3.1a. Of all the detected strokes 97% have corresponding energy values. (*Hutchins et al.*, 2012b)

In Figure 3.1a the statistical error bars (Poisson statistics) are not plotted as they would be on the order, or smaller than, the line width. It is important to note that the distribution of strokes in each region is lognormal (*Hutchins et al.*, 2012b) with the main differences in the total strokes detected and the median energy, which is 399 J, 1101 J, and 798 J for the Americas ($-180°$ E to $-60°$ E), Africa ($-60°$ E to $60°$ E), and Asia ($60°$ E to $180°$ E) respectively. An overall lower detection efficiency over Africa, particularly for low energy strokes, causes median energy to be higher than the other regions. Along with each region the energy distribution is lognormal from an hourly time scale to the annual distribution. In Figure 3.1b the annual lognormal distribution is shown with a monthly, daily, and hourly distribution. It is not until the hourly distribution that the errors are noticeable, and the distribution is still fairly lognormal.

3.2 Minimum Detectable Energy

The first step in calculating the relative detection efficiency for the entire network is working out the minimum stroke energy that WWLLN can detect at a given location and time. This process starts by finding the detection threshold at each station, converting it to
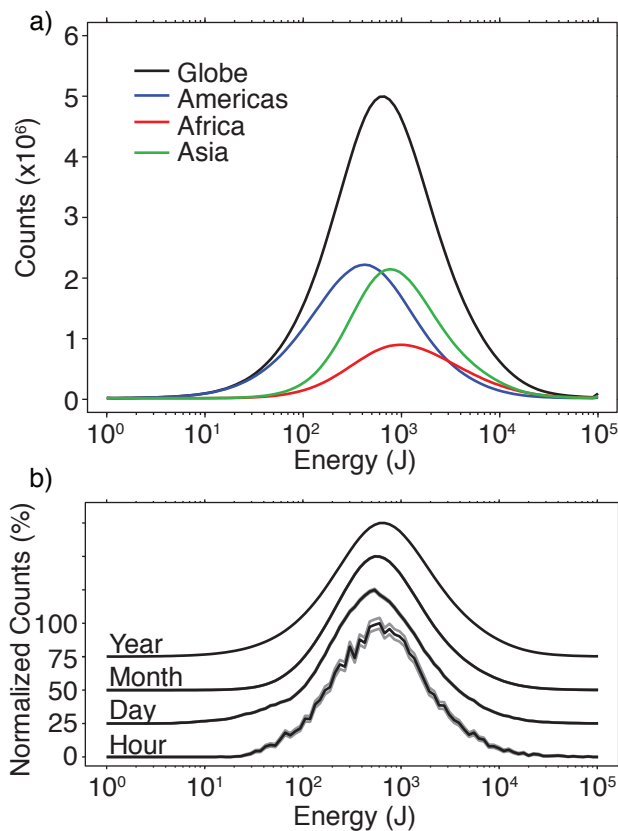
Figure 3.1: (a) WWLLN stroke energy distribution for the globe (black), the Americas (blue), Asia (green) and Aftica/Europe (red). (b) WWLLN global stroke energy distribution for a year (2010), month (June 2010), day (15 June 2010), and hour (09 UTC 15 June 2010). Grey lines are statistical count errors.

an energy value at each location in the world, and then selecting the minimum detectable network energy at every location based on the minimum observable energy from each station. Detailed examples of how this works are given next for single stations and for the network as a whole.

3.2.1 Station Threshold

At each WWLLN station the threshold for triggering on an event (and calculating the TOGA at that station) is dynamically selected depending on observed activity at that

station as described in Section 5.3 of *Rodger et al.* (2006). Presently every WWLLN station automatically adjusts the triggering threshold to send an average of 3 packets per second to the central processor. For instance, when a station is detecting many strokes, the trigger threshold at that station is raised to maintain a steady flow of sferic packets. Since a station can only measure the electric (or magnetic) field of an event it cannot accurately discern whether a sferic comes from a nearby weak stroke or a strong distant stroke; for the case of the strongest lightning strokes the discharge could be on the other side of the Earth from the WWLLN station and still be detected.

The effect of the variable trigger threshold can be seen in Figure 3.2a which is a 2-D histogram of number of strokes with specific RMS field and UT values on 15 June 2010 for the Dunedin, New Zealand, WWLLN station ($-45.864°$ N, $170.514°$ E). In Figure 3.2b the threshold can be seen as the lower cutoff of the triggered RMS field strength distribution, the station threshold is reconstructed hourly as the 5th percentile value (red line) of the distribution. The threshold value varies relatively slowly over the course of the day.
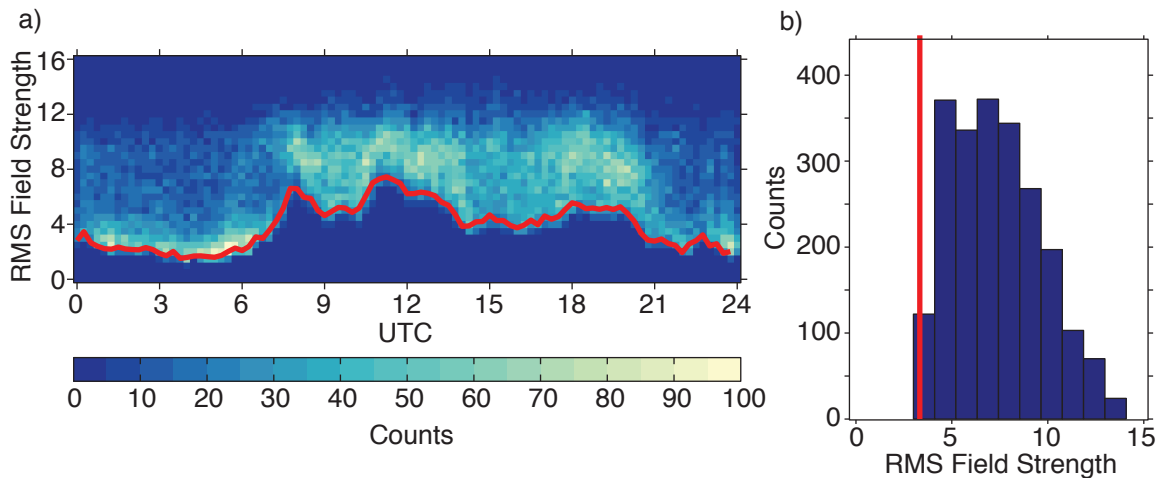
Figure 3.2: (a) shows the evolution of the triggered RMS field strength distribution (in arbitrary units) for the Dunedin WWLLN station with the red line showing the 5th percentile value. (b) shows the 9 UTC slice of the distribution, with the 5th percentile value marked (red line).

3.2.2   Station Minimum Detectable Energy

The minimum detectable energy (MDE) is the minimum energy a lightning stroke must radiate in the VLF to be detected by WWLLN or a WWLLN station (denoted network MDE and station MDE respectively). The MDE is a function of space, time and station threshold. Each station has a variable threshold which varies slowly during the day. Slow ionospheric variations can also affect the MDE by changing the VLF attenuation and detected RMS field.

Every hour the reconstructed minimum RMS field necessary to trigger an event is calculated and converted to a stroke energy. To make this conversion the same method as calculating the radiated energy per stroke is used as described in *Hutchins et al.* (2012b). This results in a station MDE for every point on a 1° x 1° global grid, which is the stroke energy necessary at that location to trigger a TOGA calculation at the given station. As an example the map of the MDE for our Dunedin station (data shown in Figure 3.2) is shown in Figure 3.3a. Figure 3.3a applies only to strokes detected at this one station in Dunedin, a similar map can be generated for every WWLLN station. The high MDEs in Figure 3.3a over the Antarctic, Western Africa, and Greenland are due to the high VLF attenuation over ice, and imply that Dunedin is very unlikely to detect strokes with energy less than the MDE if they were to occur in these regions.

In order to locate a stroke, WWLLN requires TOGA values from at least five stations in order to conduct adequate fit error analysis. For every 1° x 1° grid cell all of the minimum stroke energies from currently active WWLLN stations are ordered. An example for one cell is shown in Table 3.1. The 5th lowest from this list is used as the network MDE, because at least five stations can trigger on that energy value. In other words, WWLLN cannot detect a stroke until it has a radiated energy which is above the trigger threshold at five or more WWLLN stations. A map of the network MDE is shown in Figure 3.3b. Similar to the station MDE map for our Dunedin station, Figure 3.3a, there are higher MDE values above the Arctic and Antarctic ice regions.

Regions of the network with higher MDE, from either increased VLF attenuation, station thresholds or sparse coverage, preferentially detect a higher ratio of energetic strokes to all
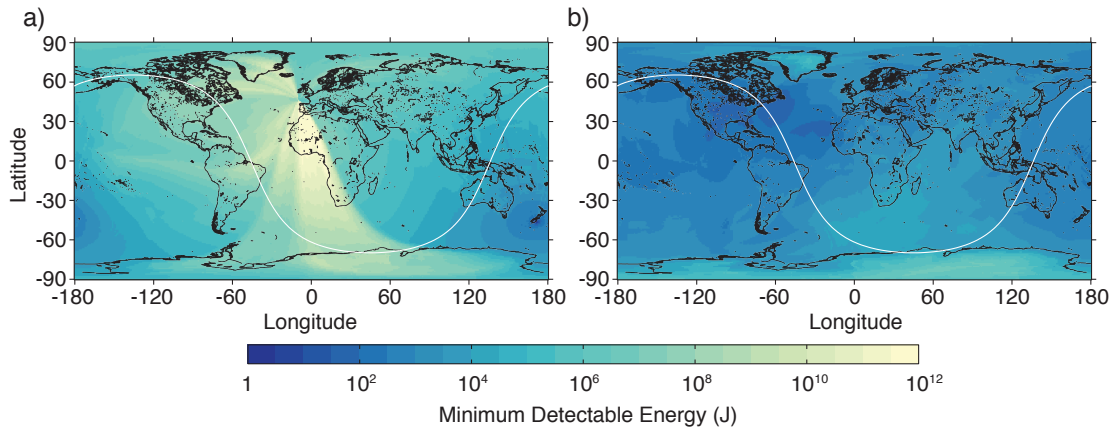
Figure 3.3: (a) the minimum detectable energy (MDE) for the Dunedin station at 9 UTC on 15 June 2010. The regions of high MDE are due to poor VLF propagation over ice from those regions to Dunedin station. (b) the minimum detectable energy (MDE) for the entire WWLLN network. The white line shows the terminator.

Table 3.1: Ordered list of station MDE values at $-25°$N, $20°$E and 09 UTC on 15 June 2010. The fifth lowest value (in bold) is the network MDE at this location.

Station Name	MDE (J)
Davis, Antarctica	34
Ascension Island	170
SANAE Base, Antarctica	190
Perth, Australia	2300
Rothera, Antarctica	**2400**
Tel Aviv, Isreal	4700
. . .	. . .
Honolulu, Hawaii	1.4×10^8
Dunedin, New Zealand	5.1×10^8

strokes. For example southern Africa has a higher MDE than other regions and the median energy, shown in Figure 3.1a is correspondingly higher. Conversely regions with low MDE, such as the Americas, show a lower median energy.

3.3   Relative Detection Efficiency

The next important step in calculating the relative detection efficiency is to establish the relationship between the network MDE and relative detection efficiency. The relative detection efficiency is a measure of how well a given location in the network is being observed relative to the best region in the network. In a given grid cell the network MDE is compared to the total WWLLN energy distribution of the past seven days. For a given network MDE value the fraction of total strokes above the network MDE gives the relative detection efficiency. The past seven day distribution is used as the base distribution in order to average over diurnal and station performance variations. This lognormal base distribution is assumed to be representative of a single universal distribution of stroke energies that could be detected globally by a uniform WWLLN.

For example, if a location has an network MDE of 100 J, then the number of strokes in the past seven days above 100 J (grey area, Figure 3.4a) is compared to the total number of strokes which were located in that location in those seven days. In this case the grey area has a count of 2.6×10^6 strokes and the total number of WWLLN strokes is 2.9×10^6 strokes, so for this network MDE of 100 J the relative detection efficiency is 90%. Similarly if a location has a high network MDE value there will be few strokes with energy above it, so it will have a low relative detection efficiency.

This calculation is done for a range of hypothetical network MDE values which produces a curve shown in Figure 3.4b, giving the relationship between MDE and relative detection efficiency. This relationship is established once per day, and it is used to produce hourly maps of relative detection efficiency for that day. This is done by taking the hourly maps of network MDE and applying this relation to every $1° \times 1°$ point on the globe for every hour to convert the network MDE to the relative detection efficiency.

The relative detection values given by this process are only in reference to the energy distribution of the past seven days as seen by WWLLN. If a region has a relative detection
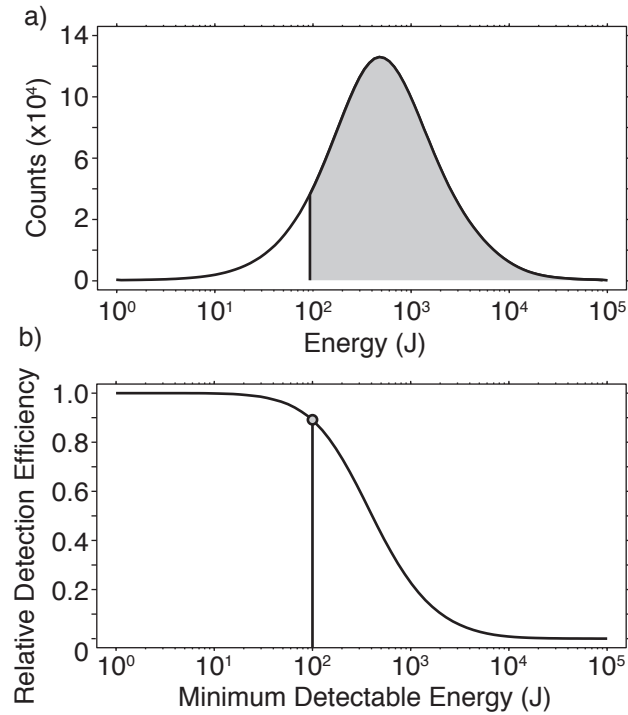
Figure 3.4: (a) The seven day energy distribution with the strokes above the MDE of 100 J shown in grey. The fraction of strokes above 100 J to total strokes gives a relative detection efficiency of 0.9, shown as a circle in (b). The fraction for all possible MDE values is shown as the curve in (b).

efficiency of 100% then the region is able to detect all of the detected stroke energies present in the 7-day network energy distribution. The corrections from the relative detection efficiency maps can be used to generate lightning density distributions as though WWLLN had global uniform coverage at the same level as that of the best parts of the network. This is because the method does not correct the network to absolute stroke counts, just to a globally uniform performing WWLLN.

3.3.1 Hourly Maps

A set of four hourly maps from 15 June 2010 showing the networks relative detection efficiency every 6 hours from 00 UTC to 18 UTC is presented in Figure 3.5. Stations that

were operational for the hour shown are displayed in white and stations that were not operational are black (operational taken to triggering > 500 strokes/hour). The four major competing effects on the detection efficiency are the day/night terminator, local stroke activity, station density, and station performance. The day/night terminator effect can be seen as it moves from 00 UTC (Figure 3.5a) through 18 UTC (Figure 3.5d). An increase in local stroke activity in North American afternoon (Figure 3.5a) causes a decrease in detection efficiency as nearby stations raise their triggering thresholds. Station density is coupled with station performance, since when a station is not operating optimally it has a similar effect as removing that station; the effect of station performance is discussed in Section 3.4.3.

Figure 3.6 shows the daily relative detection efficiency from the average of the hourly maps, here grey stations were only operational part of the day. This average map is more representative of the relative detection efficiency for the day and it shows behavior that is expected based on the distribution of stations: lower detection efficiency over most of Africa with higher detection efficiency over and around the Pacific and North America. The low detection efficiency over Antarctica, parts of Siberia, and Greenland are due to the high attenuation of VLF propagating subionospherically over ice. Conversely the high detection efficiency over North America, Western Europe, and Oceania, are due to the high station density and low attenuation of VLF over ocean. In order to prevent unphysical overcorrections, a minimum relative detection efficiency of 5% has been set for all of the relative detection efficiency maps.

3.4 Analysis

3.4.1 Distribution Changes

As shown in the previous sections the relative detection efficiency values in a given day are derived from the WWLLN observed stroke energy distribution from the previous seven days; this allows for direct comparisons within a day and for nearby days, but it does not take into account the changing distribution from changes in the network. As more stations are added to the network additional low-energy strokes will be detected and the overall
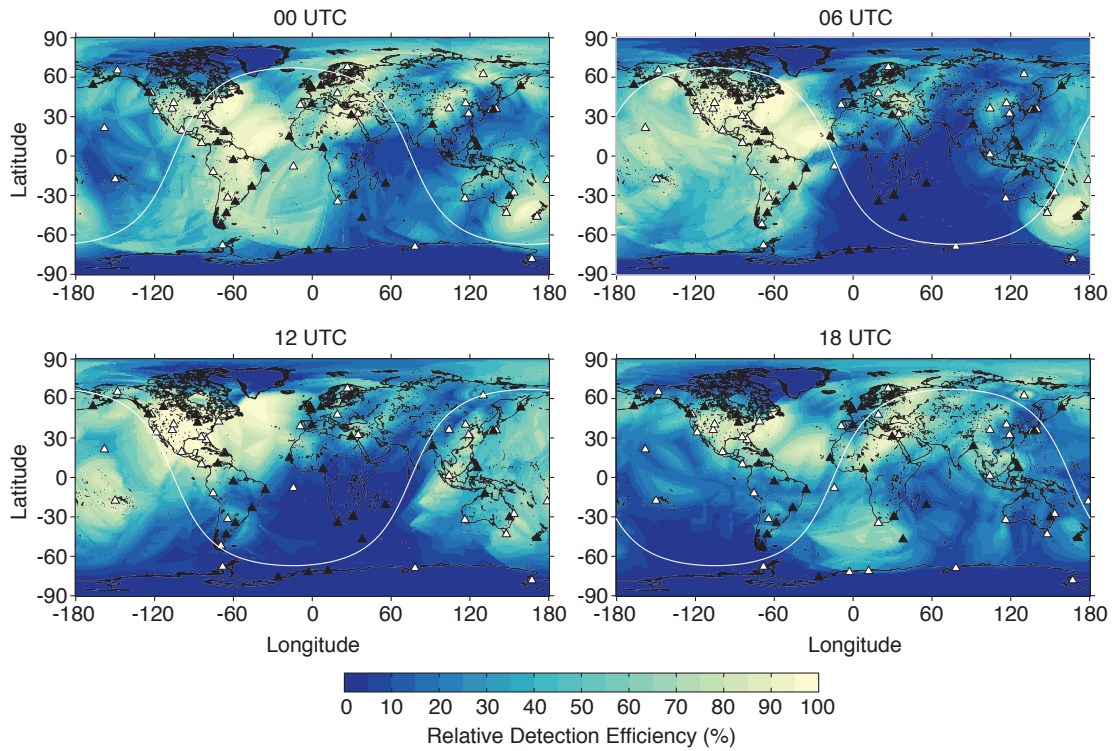
Figure 3.5: Relative detection efficiency maps for 00, 06, 12, and 18 UTC on 15 June 2010. Stations are shown as triangles with operational stations in white and non-operational in black. The minimum value of detection efficiency is set at 5% to prevent unphysical corrections.

energy distribution will shift towards lower values. When the overall network distribution changes between years, then for a given region the relative detection efficiency can change even if that region of the network has detected the same distribution of strokes.

One way to examine the change in the distribution of energy is to examine the temporal variability of the median of the global WWLLN energy distribution. The median of the seven day distribution is shown in Figure 3.7. The median energy varies from the three year median by 52% with the daily median value ranging from 400 J to 2000 J. The variability is caused by ionospheric changes not accounted for in the ionospheric model used. Several jumps in the median energy (e.g., Dec 2009 and Dec 2010) are caused by changes in the
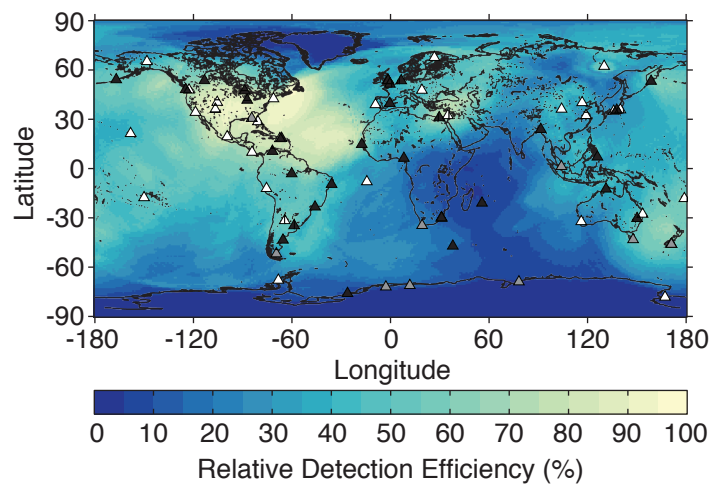
Figure 3.6: Daily average relative detection efficiency for 15 June 2010. Stations are shown as triangles with operational stations in white, non operational in black, and operational for part of the day in grey. The minimum value of detection efficiency is set at 5% to prevent unphysical corrections.

primary calibrated WWLLN station, e.g. gain changes. The slow increase to Aug 2011 was due to a change of the primary calibrated station from the Dunedin, New Zealand station to the Scott Base, Antarctica station. It is important to note that since the detection efficiency is relative to the past seven days, the relatively slow changes in median energy do not strongly affect the detection efficiency and highlight how the relative detection efficiency cannot correct for absolute overall network performance.

3.4.2   Temporal Variability

The evolution of the network can be seen as an increase in the global average relative detection efficiency, calculated by averaging all grid cells of each hourly map for a day. While no region can have a relative detection efficiency over 100%, as regions improve with more stations they will approach 100% and increase the global average detection efficiency. The global average relative detection efficiency from April 2009 through October 2011 is shown as the green line in Figure 3.8. In the Figure the total number of operational stations
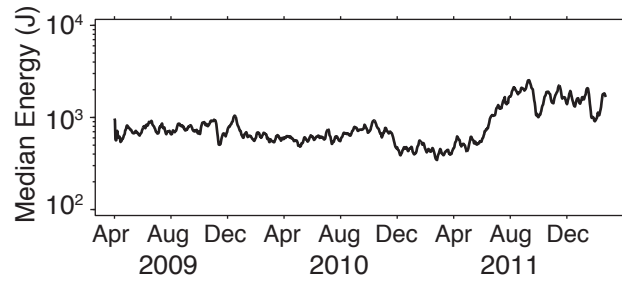
Figure 3.7: Median stroke energy of the 7-day distribution observed by WWLLN. The relative detection efficiency of the network is based on this 7-day energy distribution. Systematic changes in median stroke energy result from unaccounted for gain changes at the primary calibrated WWLLN station (see Chapter 2). Tick marks correspond to the start of the listed month.

is shown as the black line, and it has a strong correlation to the global averaged detection efficiency with a correlation value of 0.86. With more stations strategically added to the network the 7-day energy distribution will also change to include more low energy strokes and increase the average relative detection efficiency.
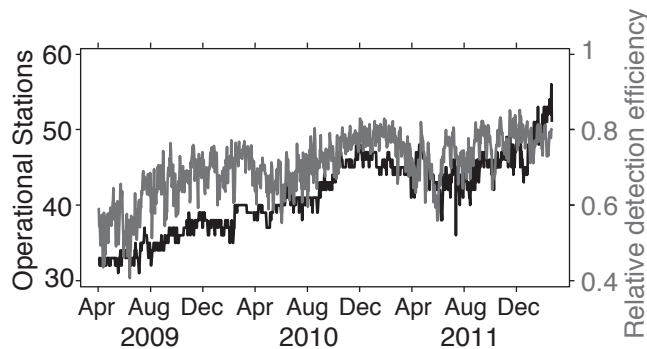
Figure 3.8: The number of WWLLN stations operating (black) and the global average relative detection efficiency (gray) for April 2009 through October 2011. Tick marks correspond to the start of the listed month.

While Figure 3.8 shows an overall increase in the number of network stations and hence

detection efficiency; Figure 3.9 shows similar curves for just low-latitude regions ($-30°$ N to $30°$ N, blue), a single location near Florida (at $-85°$E, $30°$N, red), and a single location near South Africa (at $25°$E, $-20°$N, green). Removing high latitude regions, which have small contributions to lightning, increases the overall detection efficiency but does not change the overall upward trend shown by the blue curve in Figure 3.9. When the region near Florida is examined it can be seen that it remains fairly close to 1.0 for the entire dataset, with downward trends during local summer months due to increased local lightning activity. The region near South Africa has a steady increase in detection efficiency except during a large drop out which occurred in the middle of 2011, caused by one of the African stations going offline. This shows the global detection efficiency tracks the network as a whole, but it cannot be used as an accurate proxy for smaller spatial scales.
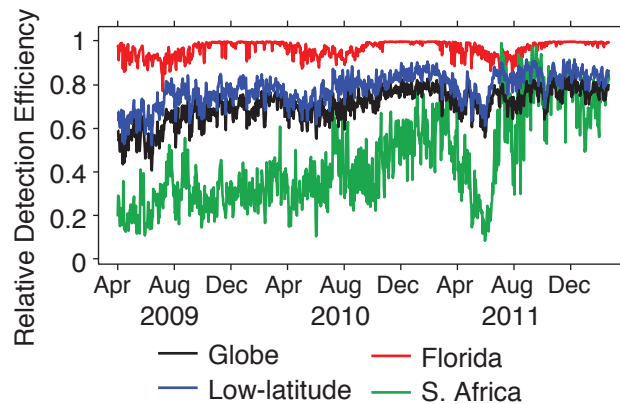
Figure 3.9: Daily variation of average detection efficiency for the globe (black), low-latitudes ($-30°$ N to $30°$ N, blue), over Florida (at $-85°$E, $30°$N, red), and over South Africa (at $25°$E, $-20°$N, green). Tick marks correspond to the start of the listed month.

The local time variability over the region near Florida is shown in black in Figure 3.10 and shows a total variability of about 4.9%. The largest drop in the relative detection efficiency occurs in the afternoon, near the peak in local lightning activity at 3pm. This drop is due to the nearby stations raising their detection threshold in response to detecting more local strokes. For this location the effects of local activity dominates over the expected

day/night effect due to changes in VLF propagation.
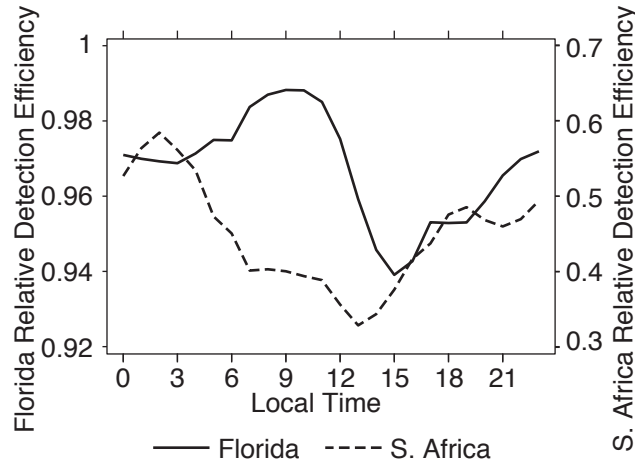
Figure 3.10: Average solar-local time variation of detection efficiency over Florida ($-85°$E, $30°$N, solid) and South Africa ($25°$E, $-20°$N, dashed), from 2009-2011.

The variability for the region near South Africa is shown as the dotted line in Figure 3.10, there is a total variability of 25.5%. There is an overall decrease in relative detection efficiency during the day when the sferics are propagating over the continent. The best relative detection efficiency occurs in the middle of the night when the stations in Africa have less nearby activity and sferics are able to propagate more readily under a night ionosphere. Compared to the Florida region there is a much higher dependence on day and night conditions as well as a much wider range of variability.

3.4.3 Station Outage Effects

While the overall performance of the network trends along with the total number of stations, a single station turning on or off can affect a large region with only a small effect on the network as a whole. To test the influence of single stations a day of data was randomly selected, 16 June 2010, and the entire data were reprocessed with just the Honolulu, Hawaii station ($-158°$E, $21°$N) removed from the raw data and again with just the Maitri, Antarctica station ($12°$E, $-71°$N) removed. The maps of the daily average with and without these

stations are shown in Figure 3.11. For Hawaii the change is fairly local to its region in the Pacific Ocean, but leads to little effect across the entire network. In the case of Maitri there is a larger effect as it is located in a region of sparse detector coverage and covers much of the southern Atlantic.
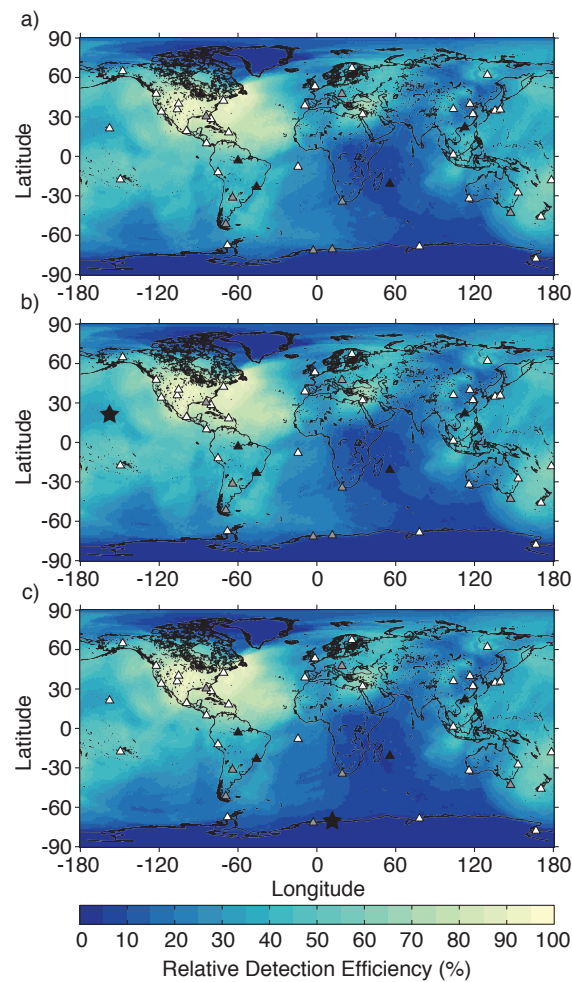
Figure 3.11: Relative detection efficiency map of 16 June 2010 for (a) the complete network, (b) the network with the Hawaii station (black star, $-158°$E, $21°$N) removed, and (c) the network with Maitri station (black star, $12°$E, $-71°$N) removed. Stations are shown as triangles with operational stations in white, non operational in black, and operational for part of the day in grey.

The daily average global relative detection efficiency dropped from 64% to 63% without Hawaii and from 64% to 53% without Maitri. The detection efficiency in the grid cell over Hawaii dropped from 85% to 78% and from 45% to 7.4% in the grid cell over Maitri. A plot of the total change between the daily averages in Figure 3.11 is shown in Figure 3.12.
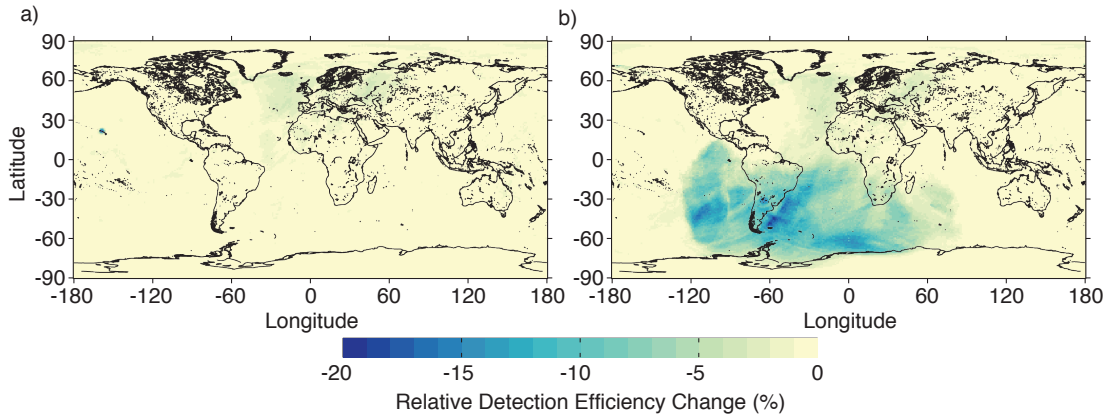
Figure 3.12: The difference in detection efficiency for 16 June 2010 with Hawaii (a) and Maitri (b) stations completely removed from processing.

3.5   Results

The detection efficiency model can be applied to global maps of stroke density to estimate, or correct for the global stroke density which would be seen if WWLLN had a uniform spatial and temporal coverage. This does not correct for the overall absolute detection efficiency (11% for CG flashes in the United States, see *Abarca et al.* (2010)), rather it corrects for the areas with less WWLLN coverage. The hourly stroke density plots are corrected by dividing the counts in each grid cell by the relative detection efficiency of that cell. For example a grid cell with 100 strokes and an efficiency of 80% would be corrected to 125 strokes. The stroke density from 2011, Figure 3.13a, had the model corrections applied hourly with the condition that a 1° x 1° grid cell needed at least two strokes to have a correction applied. A second condition was that a minimum relative detection efficiency of 5% was set for the model.
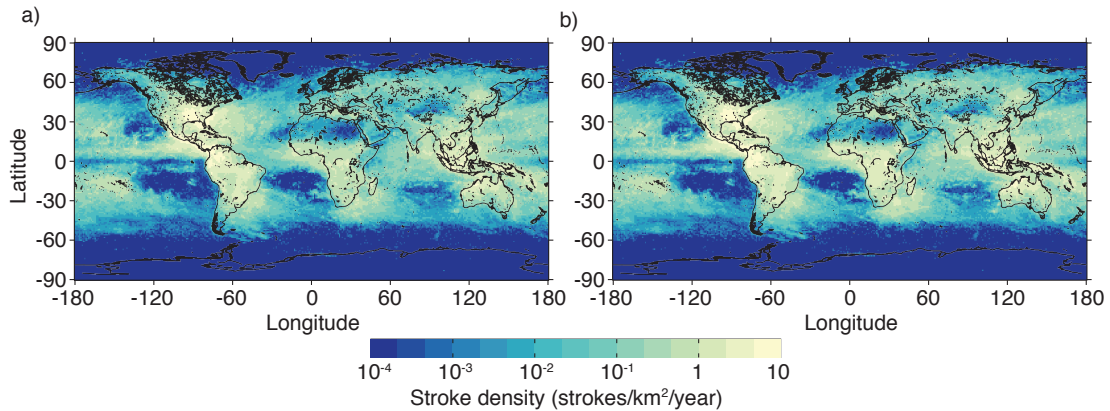
Figure 3.13: a) The raw 2011 global stroke density measured by WWLLN. b) The 2011 global stroke density corrected with the relative detection efficiency model of the network.

The total number of strokes for 2010 was 1.4×10^8 (4.4 strokes/second), and after applying the model the total was 2.0×10^8 strokes (6.3 strokes/second). In 2011 the total number of strokes was 1.5×10^8 (4.8 strokes/second) with a model-corrected value of 1.9×10^8 (6.0 strokes/second). In 2010 63% of the global area between $\pm 60°$ latitude had a relative detection efficiency of at least 80% and in 2011 this area increased from 66% to 72%. If we assume that the global lightning flash rate was a constant 46 flashes/second as determined by satellite measurements using the Optical Transient Detector and Lightning Imaging Sensor (*Cecil et al.*, 2011; *Christian et al.*, 2003) for both years, this would imply a corrected global absolute detection efficiency for cloud to ground and in-cloud flashes of 13.7% for 2010 and 13.0% in 2011. The slight decline in overall detection efficiency in 2011 compared to 2010 is likely caused by the temporay decrease in the number of operational WWLLN stations during 2011, seen in Figure 3.8.

The corrected yearly density is shown in Figure 3.13b, aside from the overall increase in number counts the important feature is the relative count rates over the US, Africa, and Southeast Asia. In the uncorrected Figure 3.13a the peak stroke density in Asia and America are similar while Africa is about ~1-10% of these values (also shown in Figure 3.1a). In the corrected maps we can see that the peak density in Africa is much closer in magnitude to

that seen for America and Asia, and the relative densities match the distributions seen by OTD (see Figure 3.15, from *Christian et al.* (2003) Figure 4). The total increase in stroke counts is shown in Figure 3.14 with the greatest increases occurring over land, in particular central Africa.
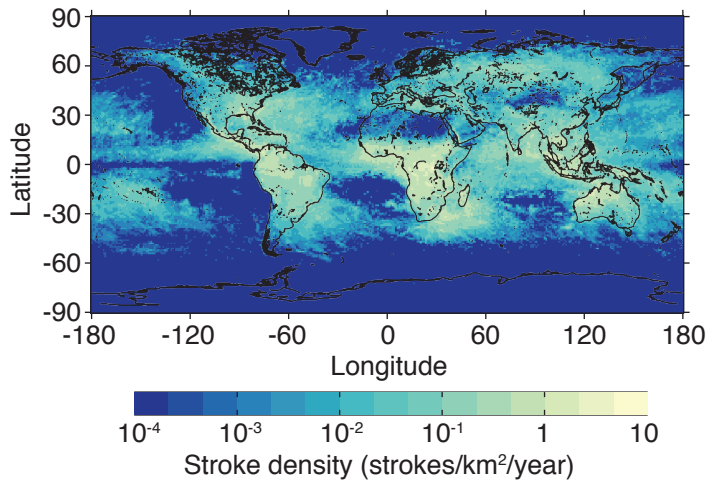
Figure 3.14: The increase in stroke density due to the relative detection efficiency corrections for 2011. Uncorrected and corrected stroke densities shown in Figure 3.13a and 3.13b respectively. The increase is plotted on the same scale as the previous two figures.

3.6 Conclusion

A relative detection efficiency model is developed for WWLLN based on the WWLLN observed stroke energy distribution. The model is examined on various temporal scales as well as performance changes due to station outage effects. The model is applied to the 2011 WWLLN dataset to produce a corrected map of stroke activity, matching the satellite optical flash incidence distribution. Work on comparing distant regions is now possible as the network data can be corrected to a uniform global level of performance.

The model developed can be used for other lightning detection networks given similar operations. Another network would need some way to determine the detection threshold at each stations, an available propagation model, and a measure of the stroke (or flash)
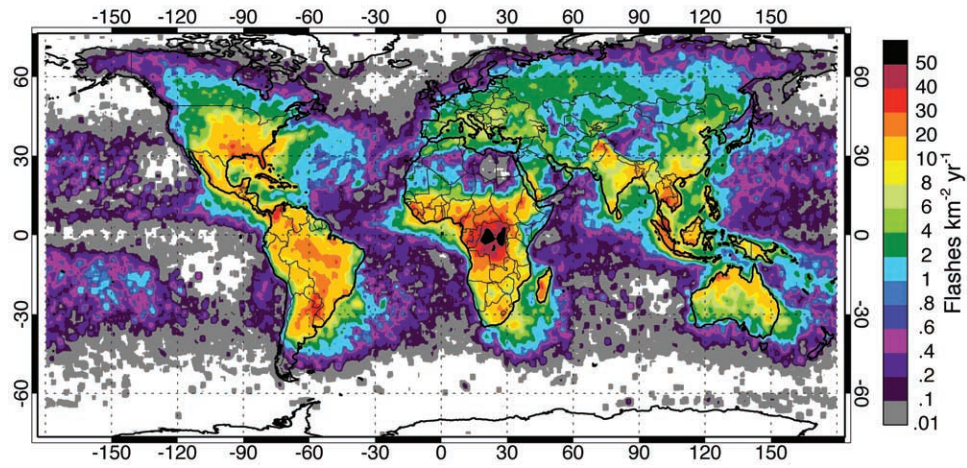
Figure 3.15: The distribution of lightning activity from 5 years of the Optical Transient Detector. Adapted from *Christian et al.* (2003) Figure 4.

strength. WWLLN uses the triggering threshold at each station as reported in the TOGA packets, the LWPC model, and the energy of each stroke.

Chapter 4

ENTLN-LIS DETECTION EFFICIENCY

4.1 Overview

Performance of the Earth Networks Total Lightning Network (ENTLN) is evaluated with respect to the TRMM Lightning Imaging Sensor (LIS). ENTLN lightning strokes are matched to LIS lightning flashes from $2011 - 2013$ over the continental United States and within the $35°$ orbital inclination of LIS. ENTLN matched 1.4×10^5 of the 2.1×10^5 LIS flashes with a total of 4.8×10^5 strokes, giving an average multiplicity of 3.5 and 66% detection efficiency of all flashes; over the course of the evaluation period the average daily detection efficiency was $69 \pm 14\%$. The median timing offset from the first ENTLN stroke was 2.4 ms (LIS leading) with a sharp peak at -1.9 ms; the median distance offset was 7.0 km from the flash centroid. The sharp peak at -1.9 ms is likely caused by inherent differences in the detection techniques: LIS observes light emitted from the tops of the clouds and ENTLN measures the radio signals from the return stroke.

The performance of LIS relative to ENTLN is also evaluated by finding all ENTLN strokes located within the LIS field of view. LIS matched 68% of the 7.1×10^5 ENTLN strokes in the LIS field of view. The $70 \pm 12\%$ daily average LIS detection efficiency of ENTLN strokes demonstrates that neither system detects every flash or stroke located by the other. Assuming a 3.5 multiplicity of ENTLN unmatched strokes, LIS did not detect 6.6×10^4 flashes, or 31% of all flashes. The ENTLN-LIS matched strokes were $20.3 \pm 0.6\%$ cloud to ground and unmatched strokes were $21.8 \pm 0.8\%$ cloud to ground. There is a marginal difference in the lightning populations sampled by both systems; LIS is slightly biased towards cloud flashes.

There is a growing importance, both scientifically and operationally, of ground based lightning detection networks. Lightning detection networks are being used in a larger gamut of research areas including: terrestrial gamma ray flashes (*Dwyer*, 2012; *Gjesteland et al.*, 2011; *Connaughton et al.*, 2010), lightning climatology (*Virts et al.*, 2013, 2011; *Bürgesser et al.*, 2012), ionospheric disturbances and probing (*Jacobson et al.*, 2010; *Singh et al.*, 2011b), transient luminous events (*Soula et al.*, 2011), global electric circuit (*Holzworth et al.*, 2005), and whistler observation (*Collier et al.*, 2010, 2011; *Burkholder et al.*, 2013). This is in conjunction with the extended usage of lightning networks operationally in weather

prediction and tracking (*Fierro et al.*, 2012; *Pan et al.*, 2010; *Thomas et al.*, 2010), volcano monitoring (*Doughton*, 2010), and hazard estimation (*Altaratz et al.*, 2010). With growing usage it is necessary to understand the capabilities and efficiencies of the various available lightning networks.

Ground based total lightning networks distinguish themselves from other ground based networks and satellites by detecting and identifying in-cloud (IC) discharges as well as cloud to ground (CG) strokes. Lightning type is critical in understanding thunderstorm dynamics (*Williams et al.*, 1989), with real time monitoring of sudden increases of IC activity able to predict severe weather events (*Rudlosky and Shea*, 2013; *Darden et al.*, 2010; *Metzger and Nuss*, 2013; *Schultz et al.*, 2009, 2011). The higher frequencies of total lightning networks are also useful for researching narrow bipolar events (*Suszcynsky*, 2003) and large scale lightning behavior (*Hutchins et al.*, 2013a). Lightning mapping arrays are able to locally detect, locate, and distinguish IC activity, however total lightning networks have the advantage of much larger spatial coverage.

To evaluate the ENTLN performance it is compared against the Lightning Imaging Sensor (LIS) onboard the Tropical Rainfall Measurement Mission (TRMM) satellite orbiting at a $35°$ inclination (*Christian et al.*, 1999). LIS is used as a reference system as the sensor performance has not changed over time and it uses a different detection method (optical) than ground networks. The LIS data are available at several processed levels; the 2011 – 2013 flash level data are used for this comparison. The comparison with LIS is made over North America and within the inclination of LIS; the area covered is shown in Figure 4.1. In this region LIS located 2.1×10^5 flashes, with the distribution shown in Figure 4.1a, and ENTLN located 3.1×10^8 strokes shown in Figure 4.1b. ENTLN detected 1.4×10^5 strokes within the field of view of LIS.

Utilizing lightning networks requires a thorough understanding of what the networks detect and their limitations. Examining detection efficiencies is one method to characterize the performance of networks, however such measures need to be used carefully as they often necessitate the assumption that the reference system is uniform, constant, and complete. Using the LIS instrument allows for a robust and standard analysis of the ENTLN efficiency and accuracy.
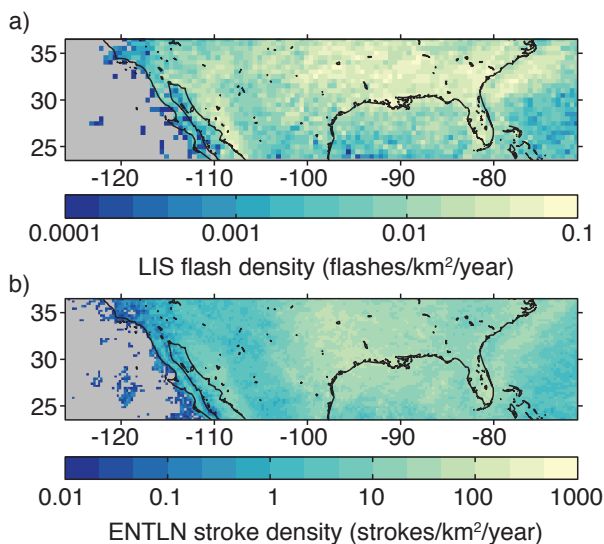
Figure 4.1: a) LIS flash density at 0.5° grid spacing for fully viewed granules and b) ENTLN stroke density at 0.25° grid spacing for 2011 – May 2013. Densities shown in counts/km^2/year; grid points with less than 30 counts are shown in gray.

4.2 Performance

4.2.1 Matching

ENTLN strokes are matched to LIS flashes when the ENTLN stroke is within a $\pm 0.5°$ box around the flash centroid and within ± 10 ms of the flash duration. If multiple strokes are matched to a single LIS flash the timing and location of the stroke closest in time to the start of the flash is used as the best match; for the case of multiple flashes matched to a single stroke only the flash closest in time to the stroke is matched. Matches between ENTLN-LIS had 30% of LIS flashes matched to a single ENTLN stroke, with an overall mean multiplicity of 3.5. Days with less than 30 total LIS flashes over North America are not used in this evaluation.

Evaluation of the effectiveness of LIS at observing ENTLN strokes utilizes the viewpoint granule data available with the flash level data. The viewpoint granules give the start and end times of LIS observation for $0.5° \times 0.5°$ bins. Full coverage for a given viewpoint

granule is determined by the start and end times of the adjacent corner granules, as shown in Figure 4.2. When LIS has at least partial coverage of the four corner viewpoint granules the center viewpoint granule is guaranteed to be in full view of the sensor. The latest start time and earliest end time of the adjacent corner granules determines the time when the viewpoint granule is in full view of LIS. Only ENTLN strokes and LIS flashes within viewpoint granules with full LIS coverage are used in this evaluation. This restriction reduces the total number of LIS flashes by 30% from 3.0×10^5 to 2.1×10^5 flashes; this ensures that all ENTLN strokes are within full view of LIS, removing edge and partial view cases.
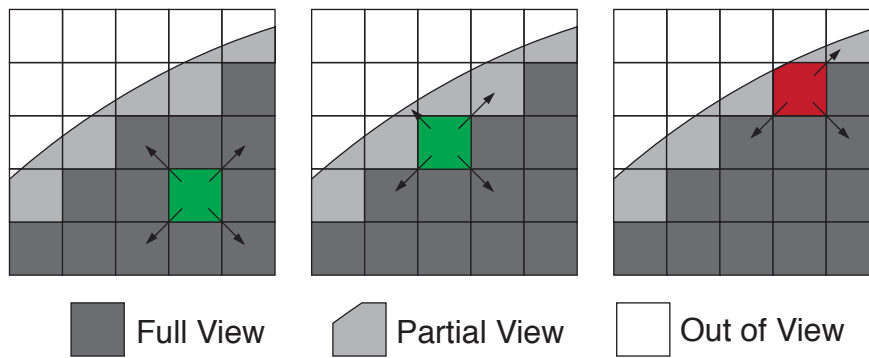
Figure 4.2: Classification of a viewpoint granule ($0.5° \times 0.5°$ bin) of LIS being in full view. If all four corner granules are at least partially viewed (left and center) the granule is in full view, if not it is only partially viewed (right).

An example of a single LIS overpass full coverage field of view, determined by the viewpoint granules, is shown in Figure 4.3. This overpass occurred on 2012 April 21 from 15:20 – 15:31 UTC. Within the field of view LIS observed 220 flashes, of which ENTLN matched 66%, shown in Figure 4.3a. ENTLN strokes are considered within the field of view of LIS if they occur within the bounds of the viewpoint granule and between the start and end time of full LIS coverage. In the pass shown in Figure 4.3b ENTLN detected 773 strokes, where 67% were matched to LIS flashes. In both cases the missed events are shown
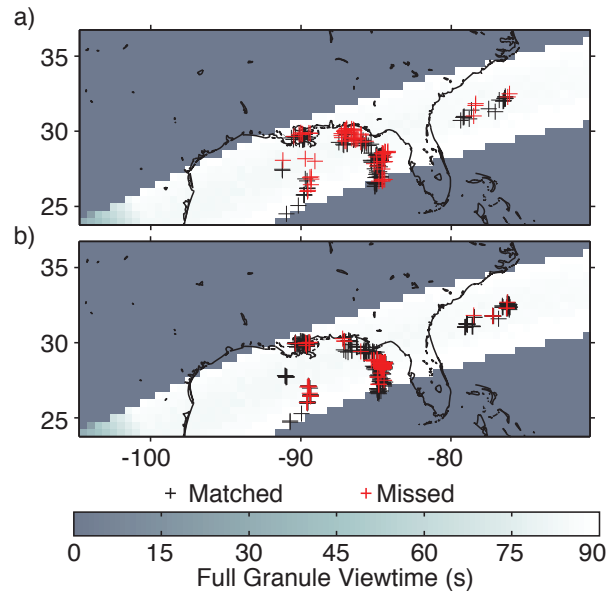
Figure 4.3: LIS observation area for 2012 April 21 from 15:20 – 15:31 UTC (day), seconds within full view of LIS shown in grayscale. a) shows the LIS flashes matched by ENTLN (black crosses) during the overpass, red crosses are LIS flashes missed by ENTLN. b) shows the ENTLN strokes matched by LIS (black crosses) during the overpass, red crosses are ENTLN strokes missed by LIS.

as red crosses.

4.2.2   Accuracy

System accuracy is measured by the timing and spatial offset between the lightning locations. The timing offset ($t_{LIS} - t_{ENTLN}$), Figure 4.4a, is sharply peaked at -1.9 ms, with ENTLN occurring before the LIS flash time; the median timing offset is 2.4 ms. The LIS flash time is the time of the first LIS event in the flash group; 42% of ENTLN strokes occurred $0 - 4$ ms before the start of the LIS flash, the remaining strokes correspond to subsequent discharges of the flash.

The median distance offset from the flash centroid is 7.0 km, as seen Figure 4.4b, slightly above the LIS spatial resolution of $3 - 6$ km (*Christian et al.*, 1999). The location offset
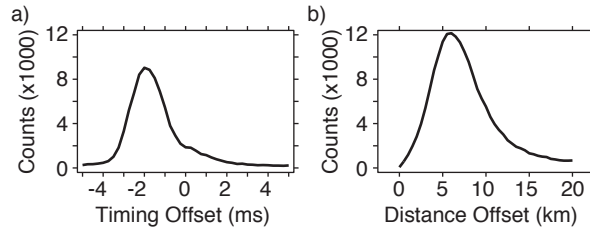
Figure 4.4: a) ENTLN-LIS timing offset ($t_{LIS} - t_{ENTLN}$) for the first matched stroke and b) ENTLN stroke to LIS flash centroid distance for the first matched stroke. Bin spacing set at (a) 0.25 ms and (b) 0.5 km.

can be given in either kilometers from the flash centroid or in units of the estimated flash radius; the radius is calculated by considering the given LIS flash area as a circle. Using the flash radius for distance results in a median offset of 0.80 flash radii; 65% of the strokes are located within the visible extent of the flash. A more refined location accuracy of ENTLN cannot be determined with LIS due to the spatial accuracy of the LIS pixels.

4.2.3 Detection efficiency

During the $2011 - 2013$ evaluation period the average daily detection efficiency of LIS flashes by ENTLN was $69\% \pm 14\%$. The detection efficiency was marginally higher during the day at $72 \pm 14\%$ and lower during the night at $65 \pm 16\%$. The ENTLN decrease at night likely stems from the increased detection efficiency of LIS at observing lightning against a non-sunlit ground. The spatial distribution of detection efficiency is shown in Figure 4.6a, with the daily variability shown in Figure 4.5a. The spatially averaged detection efficiency is $61 \pm 16\%$, due to the uneven distribution of ENTLN stations with fewer stations in the West. ENTLN detected 66% of the 2.1×10^5 LIS flashes; within those flashes ENTLN detected 7.4×10^5 strokes ($< 0.15\%$ of total ENTLN strokes). ENTLN has a spike of low detection efficiency in mid-June 2012 that is reflected in the LIS detection efficiency, this temporary decrease is likely due to a short-term issue with the network.

Within the LIS field of view ENTLN detected 7.1×10^5 strokes; of these strokes 4.8×10^5 were matched to LIS flashes for an overall LIS detection efficiency of ENTLN strokes of 68%.

The spatial distribution of LIS detection efficiency of ENTLN strokes is shown in Figure 4.6b and the daily detection efficiency is shown in Figure 4.5c. The average spatial detection efficiency of LIS is $67\pm13\%$, with an average daily detection efficiency of $70\pm12\%$. Assuming the same multiplicity of 3.5 for the unmatched ENTLN strokes LIS missed 6.6×10^4, or 31%, of all flashes. To test for the possibility that the strokes not detected by LIS were improperly located, the waveforms for a random subset for ENTLN strokes were examined and the locations were found to be correct.

Over land the highest ENTLN detection efficiency can be seen to occur in the Great Plains, Florida, and Texas. There is decreased detection efficiency over both the Rocky and Appalachian Mountains, possibly caused by the lower sensor density in these regions. While the mountain regions show lower detection efficiency with lower sensor density, there is a higher and more uniform detection efficiency within 450 km off the coasts. The increased detection efficiency off of the coasts is a combined effect of fewer and stronger lightning strokes (*Hutchins et al.*, 2013a; *Rudlosky and Fuelberg*, 2010). LIS has a more uniform detection efficiency than ENTLN, as is expected from the imaging system.

On a daily scale the ENTLN-LIS detection efficiency, seen in Figure 4.5a, is fairly high through this time period, with brief periods of decreased performance. The average daily detection efficiency is $69 \pm 15\%$ with a 5th to 95th percentile range of 48% – 87%. Some of the larger decreases in detection efficiency match with the increased LIS flash counts in Figure 4.5b, such as early April 2012, late May 2012, and mid-July 2012. On days with low LIS flash counts, $< 40^{\text{th}}$ percentile, ENTLN detects $71\pm16\%$ of flashes; while on days with high LIS flash counts, $> 70^{\text{th}}$ percentile, ENTLN detects $65 \pm 10\%$ of flashes. During these times of increased lightning activity the ENTLN would need to raise the data compression threshold; causing it to miss weaker strokes that are still detected by LIS, resulting in the suppressed detection efficiency.

4.2.4   Observation bias

Two populations of ENTLN strokes are found with the comparison to LIS: the strokes matched to LIS flashes and the strokes that are unmatched. Characteristics of these popu-

lations should show any bias of the flashes that are observed by LIS. As ENTLN is a total lightning network it has the capability to detect and distinguish between CG and IC strokes.

To get a statistically robust measure of the percentage of cloud to ground strokes observed by LIS the matched LIS-ENTLN stroke data was randomly subdivided into 100 subsets of approximately 4.8×10^3 matched strokes. Averaging these subsamples results in a statistically robust measure of the matched stroke CG percentage. The populations of matched and unmatched LIS-ENTLN strokes are very similar; matched strokes are $20.3 \pm 0.6\%$ CG and unmatched strokes are $21.8 \pm 0.8\%$ CG. The small difference between the populations shows that there is only marginal detection efficiency bias between lightning observed by LIS and ENTLN, with LIS slightly biased towards cloud flashes.

4.3   Conclusion

Characterizing the ENTLN, and other ground based networks, relative to a uniform detection system is critical in enabling the application of these networks in scientific and operational uses. The Earth Networks Total Lightning Network was compared to the TRMM/LIS instrument on a stroke to flash matching level. An average daily detection efficiency of $69 \pm 14\%$ was found for the continental United States for January 2012 – May 2013 for 2.1×10^5 LIS flashes. The relative location accuracy with respect to the flash centroid is 7.0 km and a timing offset to the start of the flash of 2.3 ms. Within the LIS field of view ENTLN located 7.1×10^5 strokes and 3.1×10^8 strokes within the entire evaluation region. Of those located ENTLN strokes LIS had a $70 \pm 12\%$ detection efficiency; leading to an estimated 6.6×10^4 flashes (31% of total flashes) not detected by LIS. Long range ground networks have the advantage of continuous real-time coverage of their encompassed region, resulting in more strokes detected along with capabilities such as stroke type discrimination and peak current estimation.
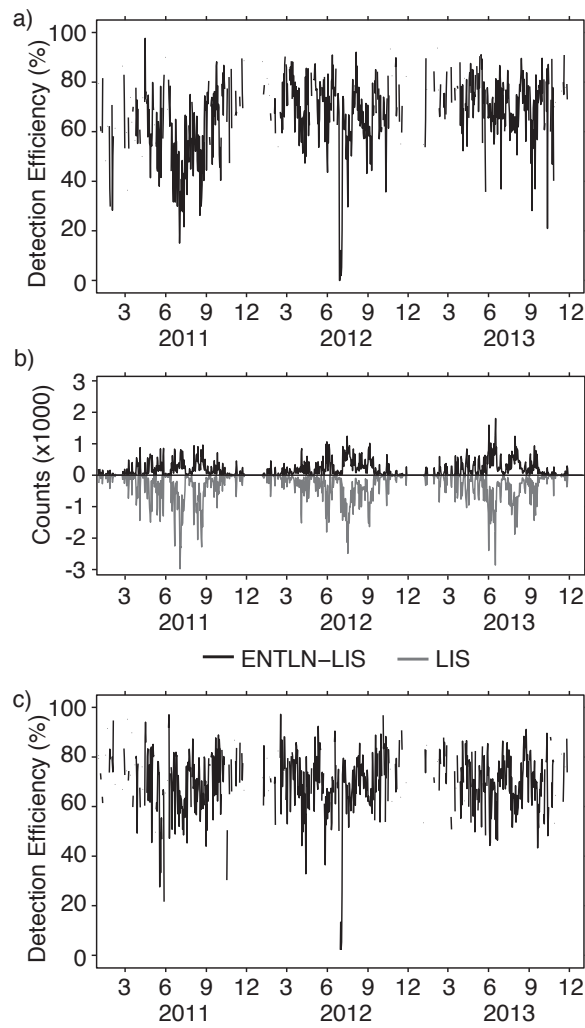
Figure 4.5: a) ENTLN-LIS daily detection efficiency, b) ENTLN-LIS total matches (black) and total LIS flashes (gray), c) LIS daily detection efficiency of ENTLN strokes. Gaps indicate days with less than 30 LIS flashes.
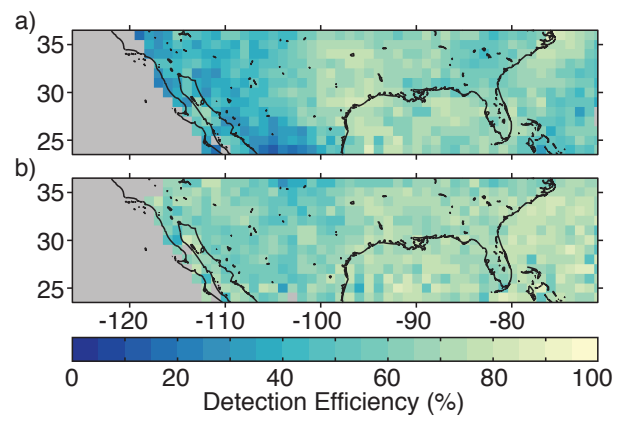
Figure 4.6: a) Spatial distribution of ENTLN detection efficiency of LIS flashes and b) LIS detection efficiency of ENTLN strokes. Gray indicates $1° \times 1°$ bins with fewer than 30 LIS flashes.

Chapter 5

LAND-SEA CONTRAST

5.1 Overview

A global contrast between oceanic and continental lightning very low frequency energy is observed using the World Wide Lightning Location Network (WWLLN). Strokes over the ocean are found to be stronger on average than those over land with a sharp boundary along a majority of coastlines. A linear regression method is developed to account for the spatial and temporal variation of WWLLN in order to perform a multi-year and global analysis of stroke energy distributions.

The results are corroborated with data from the Lightning Imaging Sensor, the Optical Transient Detector, and the Earth Networks Total Lightning Network. These systematic comparisons lead to the conclusion that there exists a strong difference in the energetics between land and ocean thunderstorms that results in a higher fraction of more powerful strokes over the oceans.

Global surveys of lightning climatology have routinely shown more lightning activity over continents than over oceans (*Christian et al.*, 2003). The difference in activity is often attributed to changes in the convective regimes in the clouds. *Williams and Stanfill* (2002) and *Williams et al.* (2005) discuss aerosol concentration, wet bulb temperature, and cloud base height as dominant mechanisms of the difference in cloud electrification. *Zipser* (1994) suggests updraft velocity due to differential surface heating may lead to the difference in observed flash rates. *Boccippio et al.* (2000) shows the total flash counts may be due to a lesser frequency of occurrence of oceanic storms and not a difference in the storms themselves.

Along with the difference in flash rates, there have been several observations suggesting an inherent difference in the lightning peak currents and optical radiance between land and ocean storms (*Seity et al.*, 2001; *Ishii et al.*, 2010). The U.S. National Lightning Detection Network (NLDN) observed higher average peak currents for negative cloud to ground (CG) strokes off of the coast, but the NLDN is limited in range for oceanic strokes near to coastlines (*Rudlosky and Fuelberg*, 2010; *Lyons et al.*, 1998). *Boccippio et al.* (2000) observed with LIS/OTD an increase in the optical radiance and extent of oceanic flashes compared to those over land. It was suggested that either a more energetic lightning generation process

or a reduced cloud optical depth for the oceanic storms could produce the increased optical radiance. However using just the available satellite data, it could not be be determined whether the more radiant flashes were caused by changes in the flashes or in the cloud optical depth.

A comparison is made between the global stroke count climatologies of WWLLN and the 13-year Lightning Imaging Sensor (LIS) and 5-year Optical Transient Detector (OTD) flash count climatologies. The LIS (1997-present) and OTD (1995-2000) are nearly identical satellite-based lightning detectors flown in low earth orbit that observe total lightning activity from individual thunderstorms for 90 sec and 2 min, respectively, as the satellite passes overhead. The LIS observes storms from an inclined orbit of $35°$ at an altitude of 402 km while the OTD observed storms from an inclined orbit of $70°$ at an altitude of 740 km (*Christian et al.*, 1999, 2003). Since WWLLN preferentially detects high-energy strokes (*Hutchins et al.*, 2012a) a direct comparison between the two systems, as described in *Virts et al.* (2013), gives a comparison between high and low-energy strokes because of the detection biases of the two systems.

5.2 Linear Regression Analysis

In order to compare WWLLN energy data over large spatial and temporal scales the data needs to be processed to account for the regional variations in detection efficiency and temporal changes in network performance. To examine the spatial changes in stroke energy while accounting for network variability a linear regression method is developed.

Energy data for each day is binned every $0.5°$ in latitude and longitude. The strokes in each bin are split into 10 energy deciles each containing J strokes, with the mean energy of each decile, i, given by:

$$\overline{E_i} = \frac{\sum_{j=1}^{J} E_{i,j}}{J} \tag{5.1}$$

$\overline{E_i}$ has a power law dependence on decile number due to the lognormal distribution of stroke energies. In order to make a regression between $\overline{E_i}$ and decile number i, the $\log_{10}$ of $\overline{E_i}$ is used. This allows for a simple linear regression between the two variables instead of a

power regression in linear space; this prevents overweighting of the high energy values of the distribution. A linear regression is found between $\log_{10}(\bar{E}_i)$ and i to get an approximation of the mean energy with decile:

$$\log_{10}(\overline{E(i)}) = C + \frac{\partial \log_{10}(\overline{E_i})}{\partial i} i \tag{5.2}$$

The first parameter from the fit, C, is related to the overall mean of the stroke energies in the particular spatial bin. It is not used in this analysis as C greatly depends on the network coverage at the time and location where it calculated. C will be larger where coverage is lower (WWLLN detecting only strong strokes) and lower with high coverage (WWLLN detecting both strong and weak strokes). By using the regression method the network coverage and variable detection efficiency is factored out of the energy distribution into C allowing for the shape of the energy distribution to be examined directly.

The second parameter, $\frac{\partial \log_{10}(\overline{E_i})}{\partial i}$, is the slope of the regression and will be used to study the energy changes between land and ocean regimes. $\frac{\partial \log_{10}(\overline{E_i})}{\partial i}$ is the measure of how much (logarithmically) the average stroke energy changes from one decile to the next. So a value of $\frac{\partial \log_{10}(\overline{E_i})}{\partial i} = 0.1$ at a location will increase the mean stroke energy by $10^{0.1} = 25\%$ per decile, while a value of 0.2 will increase by $10^{0.2} = 58\%$ per decile. The regression slope will be higher for either relatively more high-energy or low-energ strokes, corresponding to greater variance in the distribution. The low-energy tail of the energy distribution is mostly set by the detection efficiency of the network where a higher detection efficiencies result in an overall shift of the distribution to lower energies, while the high end is always well detected even where the network has low station density. As C captures the effects of the network performance, the slope of the regression will be mainly set by the energy of the strokes in the tails of the distribution. Detecting more low energy strokes will shift the distribution but not change the shape, or variance, of energies.

An example of the linear regression method is shown in Figure 5.1 using 15 days of data from $1 - 15$ June 2012, separated into land (black) and ocean (gray) strokes. The stroke counts are split into the ten decile bins outlined in Figure 5.1a. $\overline{E_i}$ is shown in Figure 5.1b with the corresponding regressions plotted on top of the points. Departures

from the regression are acceptable as only the trend of increasing energy is important and not an exact fit.
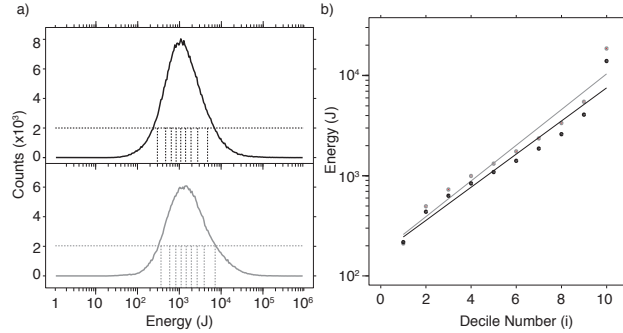
Figure 5.1: WWLLN data from 1-15 June 2012 for global strokes grouped into land (black) and ocean (gray) to demonstrate the linear regression method. a) Shows two energy distributions with the corresponding energy decile bins (dashed lines). b) is the plot of mean energy $\overline{E_i}$ in each bin with the linear regression (solid lines).

5.3 Regression Slope Maps

This technique is applied over 3 years of WWLLN data from May 2009 through May 2012 on a 0.5° grid; the resulting regression slopes are shown in Figure 5.2. For this analysis the calculated regressions must have an R-square value of at least 0.80 to be used. In general higher slopes are seen over oceans and lower slopes over land, except for several regions of low detection efficiency (e.g. off the shore of Madagascar) and regions such as the Andes mountain range. The map is similar to Figure 5.3, adapted from *Virts et al.* (2013), which shows the ratio of the WWLLN normalized stroke climatology to the LIS/OTD normalized flash climatology. The climatologies are normalized by their total stroke and flash counts respectively. Figure 5.3 is the spatial distribution of where WWLLN preferentially detects more strokes than LIS/OTD due to the bias of WWLLN towards detecting the most energetic strokes. Comparing the stroke and flash data directly is possible as a majority of flashes have only the first, and strongest, stroke detected by WWLLN (*Abarca et al.*, 2010).
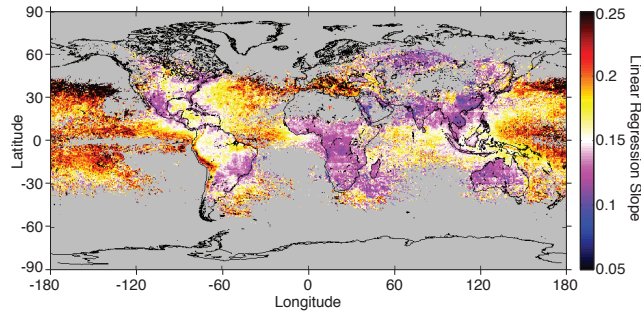
Figure 5.2: Slope of the linear regression used on the energy distribution as described in the text. High slope values corresponds to a more strokes in the tails of the energy distribution.
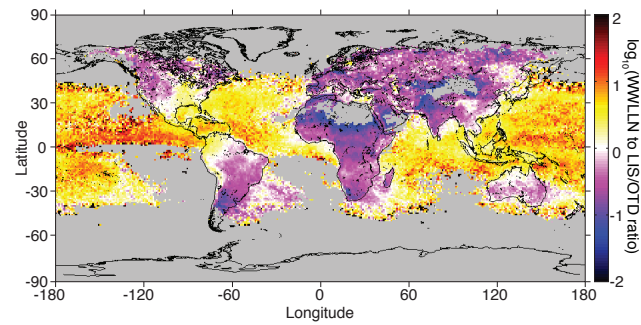
Figure 5.3: Ratio of the WWLLN stroke count density climatology to the LIS/OTD flash count density climatology normalized by their relative total counts, adapted from *Virts et al.* (2013).

The same linear regression was applied to the 2011 ENTLN data for a region over North America. The absolute peak current was used for the regression instead of the stroke energy with the results in Figure 5.4. The land-ocean contrast is seen strongly in the ENTLN dataset, particularly over Mexico, Cuba and Haiti. The difference also exists off the coast of the southeastern United States, but the contrast is not as strong (slope increase on the order of 0.01 instead of 0.1).

For the ENTLN data the linear regression slopes only range from 0.05 to 0.15 compared to 0.05 to 0.25 for the WWLLN regressions. Since energy is related to peak current by
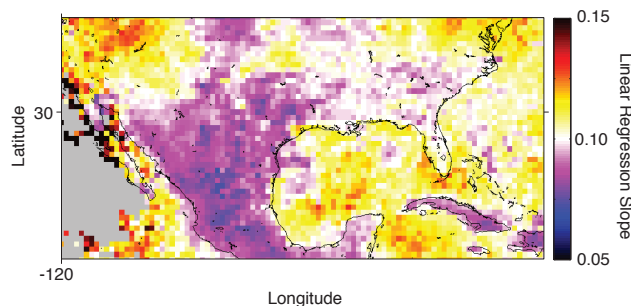
Figure 5.4: Slope of the linear regression used on the 2011 ENTLN absolute peak current distribution as described in the text. High slope values corresponds to a more strokes in the tails of the energy distribution.

$E_{stroke} = 2.23 \times |I_{peak}|^{1.62}$ (*Hutchins et al.*, 2012b), $\log_{10}(\overline{E_i})$ will be 1.62 times higher than $\log_{10}(\overline{I_{peak,i}})$. The range of the slopes for the ENTLN regression should then be correspondingly lower by a factor of 1.62, or from 0.03 to 0.15, matching the range seen by WWLLN.

5.4 Stroke Distributions

The global ratio of the land and ocean stroke distributions clearly shows the prevalence of higher energy strokes over oceans. The energy distribution of the WWLLN data is found for the set of strokes occurring over land and for the strokes occurring over oceans. The ratio of these ocean and land energy distributions is shown in Figure 5.5.

The ocean-land ratio starts increasing quickly with increasing energy at 3000 J, showing there are relatively more high-energy (top 15% of stroke energy) strokes over the oceans than over land. Similarly there is a general decrease in the ratio for decreasing strokes energies, with the downward trend interrupted with a small bump near 10 J.

5.5 Regional Contrast

In Figures 5.2 and 5.3 there is an evident overall difference between land and ocean strokes, and it can be seen to vary sharply across most coastlines. This raises the question of whether
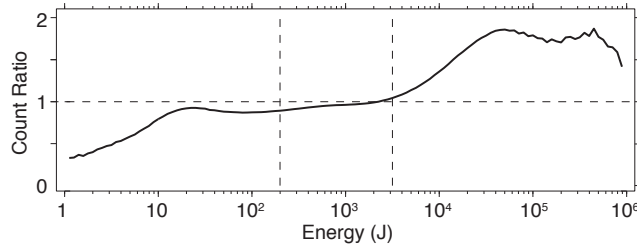
Figure 5.5: The ratio of ocean to land counts for WWLLN within each energy bin. The dashed horizontal line is a ratio of 1. The vertical dashed lines show the 15th and 85th percentile levels for the distribution.

network detection efficiency across the coastline should be considered as a potential cause for the change. Three regions are chosen for closer examination, shown outlined by the white boxes on top of a map of the WWLLN relative detection efficiency in Figure 5.6a: North America, Western Africa, and Northeastern Brazil.

WWLLN has an inherent bias towards more readily detecting low energy strokes over oceans than over land. This is due to lower VLF wave attenuation over oceans; low-energy strokes propagating over the ocean can reach more WWLLN stations compared to the same stroke traveling over land. Hence WWLLN is naturally biased to predominately detect only the highest energy strokes over land and relatively more lower energy strokes over water (*Hutchins et al.*, 2012b). The method of linear regression described in Section 5.2 should remove most of this bias in WWLLN; this can be checked in part by the recent work on relative detection efficiency (*Hutchins et al.*, 2012a). The three regions chosen in Figure 5.6 were chosen such that the gradient of relative detection efficiency is changing parallel to the coastline.

Over North America and the Gulf of Mexico, Figure 5.6b, the strokes over Mexico, Florida, Cuba, and Haiti are all weaker than those over the nearby ocean. This is also seen with the ENTLN in Figure 5.4. The relative detection efficiency (see *Hutchins et al.* (2012a)) is fairly uniform over this region with the largest change occurring over the Atlantic where there is no change to the regression slope. Over the central United States there is
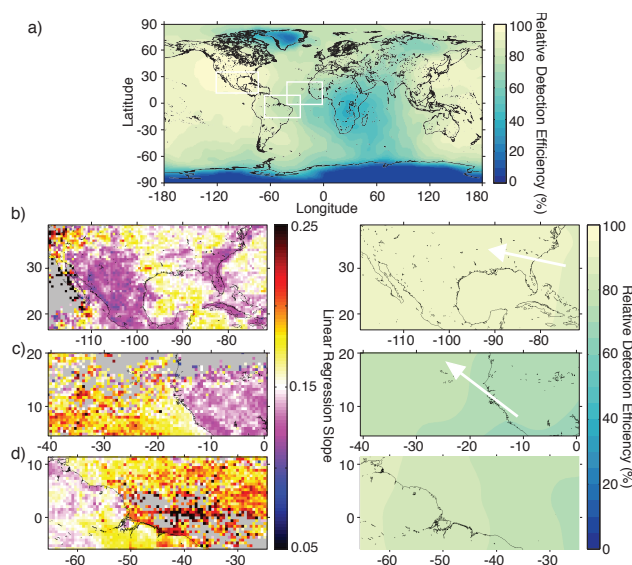
Figure 5.6: Regional maps of the linear regression slopes in the left column with respective WWLLN relative detection efficiency maps on the right. Selected regions outlined in a) on top of the map of the May 2009 through May 2012 average relative detection efficiency. b) shows the Continental United States and Gulf of Mexico, c) Western Africa, and d) Northeast Brazil. The white arrows point in the direction of increasing relative detection efficiency.

a large region of high stroke energies; this is also observed in the ENTLN regression slope (Figure 5.4) and LIS/OTD count ratio (Figure 5.3) data.

In Western Africa, Figure 5.6c, there is a clear difference between the land and the ocean; the coast shows a very sharp change in the stroke strength. The changing detection efficiency in this case is parallel to the coast and would not affect the variation in stroke energy at the coastline.

The difference in Brazil, shown in Figure 5.6d, is similar to that over Western Africa with the exception of the increased stroke energies seen over Amazon River delta. Even with the increase over the delta there is still a contrast off of the coast with the change in regression slope comparable to the coastline northwest of the delta.

5.6 *Conclusion*

A linear regression method is developed and applied to the WWLLN dataset in order to examine global and regional changes of lightning stroke strength over several years of network data. Through comparing WWLLN, ENTLN, and LIS/OTD the difference between stroke strength is seen to be highly dependent on whether the storms occur over land or over ocean with a sharp boundary occurring along most coastlines. Smaller regions are examined to show that the contrast along coastlines is not due to abrupt changes in the detection efficiency of the networks. The sharpness of the coastal changes, less than 100 km, suggests the effect is due to a local phenomena and not be caused by large scale changes in the convective land-ocean regions. Changes exist within continental regions, but these transition were not examined as the underlying change between regimes is not as sharp as for coastlines.

Chapter 6

VLF PROPAGATION

6.1 Overview

The World Wide Lightning Location Network (WWLLN) is used to measure the normalized lightning electric field at three network stations in order to examine the sferic attenuation between the stroke and the station. The electric field measurements are normalized to the square root radiated very low frequency (VLF) stroke energy to allow direct comparisons of the many stroke-station paths seen by WWLLN. WWLLN observes a strong dependence of VLF propagation on magnetic azimuth similar to past work. From WWLLN the average attenuation over the water of eastward-propagating sferics is found to be 1.13 ± 0.35 dB/Mm during the day and 0.71 ± 0.68 dB/Mm at night, with westward-propagating sferics having average attenuation rates of 2.98 ± 0.68 dB/Mm and 2.66 ± 0.39 dB/Mm for day and night respectively.

The propagation dependence on magnetic azimuth stems from the polarization of the VLF wave with respect to the electron motion in the ionosphere . Ionospheric electrons are constrained to gyrate in one direction; for VLF waves propagating eastward the polarization of the wave is in the same direction as the electrons providing more efficient propagation, the opposite in the case for westward propagating waves. There is then no change in propagation for waves aligned with the magnetic field, e.g. northward or southward propagating waves.

Past experimental work has shown a strong dependence of VLF attenuation on the magnetic azimuthal angle of propagation. *Wait and Spies* (1960) provide an in-depth analysis of the theoretical background.present a theoretical background for why the attenuation changes with azimuth. By measuring the signal strength from the same nearly antipodal transmitter along the short and long great circle path, *Crombie* (1958) found less attenuation for the eastward propagating signals. Similarly *Taylor* (1960) showed that eastward propagating VLF waves have 1 – 3 dB/Mm less attenuation than westward propagating waves. Recently *Jacobson et al.* (2012) used negative cloud-to-ground strokes to examine ionospheric reflectance and also found a strong azimuthal dependence to the measurements.

The past work has examined stationary VLF transmitters with either a few receivers or one moving receiver. Recent work by *Burkholder et al.* (2013) utilized the World Wide Lightning Location Network (WWLLN) to examine how eastward and westward propa-

gating VLF sferics couple into the ionosphere. This work motivated us to use the many stroke-receiver paths of WWLLN to study the azimuthal dependence of VLF propagation within the Earth-ionosphere waveguide.

The global coverage of the network allows for many long range stroke-receiver paths with which to estimate VLF attenuation rates. The sferic attenuation for a single station can be found with WWLLN by comparing the stroke energies measured by the network to the RMS electric field measured at a single WWLLN station. Using WWLLN allows for the azimuthal propagation effects to be examined at all magnetic azimuths for both all-night and all-day paths. The measured electric fields and observed attenuation rates can be directly compared to the theoretical predictions of *Wait and Spies* (1960) and *Taylor* (1960). By examining the propagation it is shown that WWLLN captures the effects of propagation and that the theories are applicable for non-stationary transient sources.

6.2 Path Selection

To examine eastward and westward propagation three WWLLN stations were chosen based on their island locations: Suva, Tahiti, and Honolulu. These stations were selected as a majority of their stroke-receiver paths are over water, so the effects of variable ground conductivity can be ignored.

The WWLLN energy calculation uses the LWPC code that accounts for eastward and westward propagation. However each WWLLN energy measurement is the median energy of several WWLLN stations, so by selecting strokes with a similar number of stations to the east and west the azimuthal dependence computed by LWPC can be minimized. There is allowed to be at most 25% more WWLLN stations to the east or west of a located stroke, where this abundance is defined by: $|n_{east} - n_{west}|/n$. For example, a located stroke with 1 station to the north, 2 east, and 3 west will have a westward abundance of 16.7% and would be included in the analysis.

With these three stations the WWLLN data are reduced to only consider sferics that crossed at most 5% land, and propagated in either at least 95% day or night ionospheric conditions. The stroke energy uncertainty (median absolute deviation of contributing stations, see *Hutchins et al.* (2012b)) is limited to a maximum of 10%. Further the data are reduced

by selecting only strokes with a similar number of locating WWLLN stations situated to the east and west as described above. These requirements reduce the stroke-receiver paths in the WWLLN dataset to 0.2% of the total paths for each station; where 87% occur during the day and 13% occur at night. This resulted in over 2×10^6 total stroke-receiver paths used in this Chapter.

The subset of data contains the RMS electric field at each station, the distance to the strokes, and the VLF energy of the stroke. VLF stroke energies and station measured electric fields both vary over several orders of magnitude. All station electric fields (in units of $1\mu V m^{-1}$) are normalized by the square root of the stroke energy (in J) to allow for direct comparisons between differing source energies. The stroke energy normalization gives the electric field in dB above $1\mu V m^{-1}/J^{1/2}$. Changes in the normalized electric field with stroke distance gives the attenuation of the lightning sferic for that distance interval, reported as the dB/Mm decrease.

Normalized field values are grouped into $45°$ azimuth and 500 km distance bins. An example of all of the distance bins for a given azimuth for one station is shown in Figure 6.1 (northwestward-propagation during the day for Honolulu station). Each distance bin has a distribution of normalized field values with the spread a result of differing ionospheric conditions, uncertainty in the energy measurements, and changes in ground conditions at the station. Because of this spread the median value of the normalized field is used for each distance-angle combination (shown as the solid black line). At least 15 strokes are required for each distance bin at a given azimuth in order to calculate a reasonable median. The magnetic azimuth used is the average magnetic azimuth over the path of wave propagation.

6.3   Azimuthal Dependence

For the three stations data were used from May 2009 to May 2013, resulting in 2.1×10^6 stroke-station paths to analyze. The results are split into the 8 azimuth octants for day paths and night paths as shown in Figure 6.2. For the three stations the westward propagating waves (green) show higher attenuation (electric field change per unit distance) than for the eastward propagating waves (red). The attenuation is also seen in the maximum distance that strokes are detected, with night (eastward) paths detectable at farther distances than
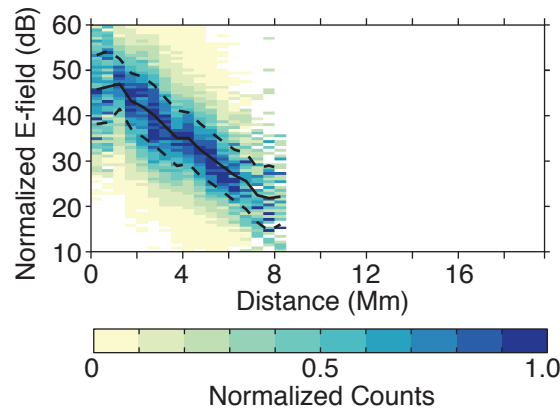
Figure 6.1: Normalized electric field values for the 315° azimuth bin for the Honolulu station during the day. Counts are normalized to the maximum value in each distance bin. The median (solid) and median absolute deviation (dashed) values are plotted on top of the distribution. Distances with less than 15 total strokes are not plotted or used.

day (westward) paths.

For some stations, such as Honolulu in Figure 6.2, the difference between eastward and westward sferics is quite clear during the day at all distance ranges, and between day and night. For other stations, such as Suva, the difference in propagation direction is not distinct for all azimuths. Suva and Tahiti stations do not show a differentiation in attenuation until the waves have propagated some distance from the stroke, for example the night strokes for Suva in Figure 6.2 are indistinguishable until 4 Mm.

In all cases the slopes of the field strength curves in Figure 6.2 show that the eastward sferics exhibit less attenuation than westward sferics away from the stroke. In cases like Honolulu the westward sferics initially have a lower normalized electric field, but the attenuation rate of these sferics is still lower than the eastward sferics.

The overall behavior of the VLF waves observed by WWLLN can be seen in the combined station data. The station-stroke pairs were combined for all stations to give a single dataset of paths, shown in Figure 6.3. The westward sferics in both day and night have similar electric field to the eastward sferics with the distinction developing at greater than 4 Mm
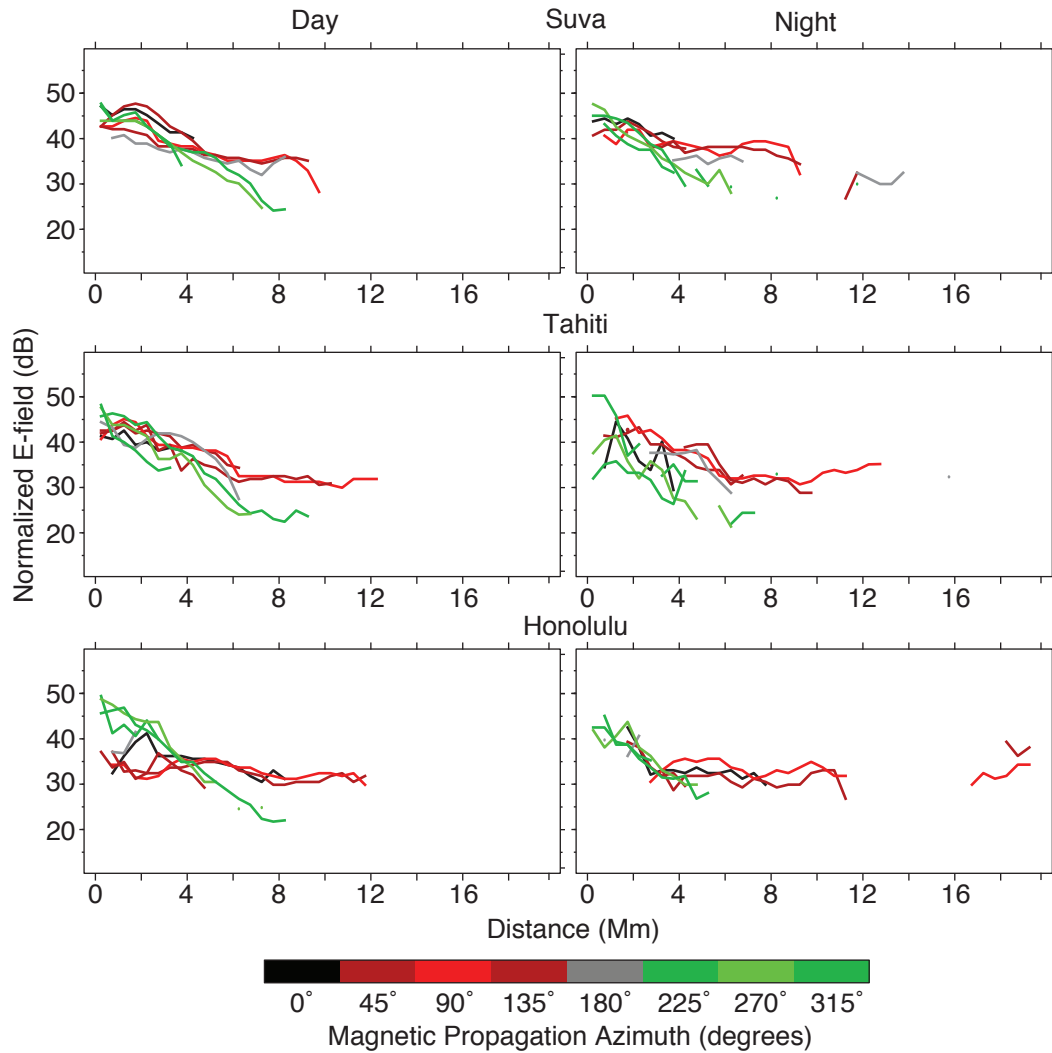
Figure 6.2: Station RMS electric field vs distance for Suva, Tahiti, and Honolulu. Electric field is normalized by the square root stroke energy and given in dB above $1\mu V m^{-1}/J^{1/2}$. Day ionosphere paths are in the left column and night ionosphere paths in the right column.

from the stroke. There is a small upturn in field strength as the sferics propagate past 10 Mm because the waves are re-focusing when they cross the halfway point to their antipode.

The method for estimating the attenuation and the uncertainty is outlined in Figure 6.4 using the station averaged eastward-propagating electric field values as an example (red line
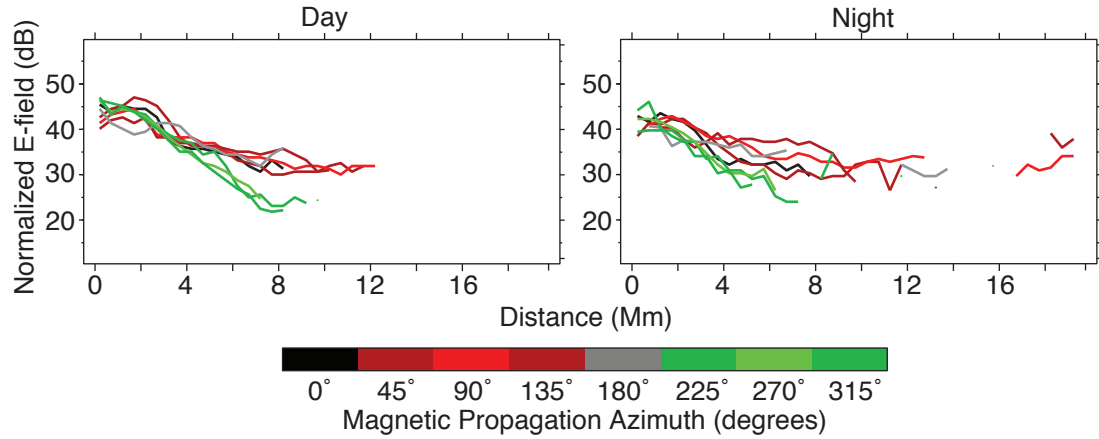
Figure 6.3: Combined station normalized electric field for the three selected stations. Shown in dB above $1\mu V m^{-1}/J^{1/2}$.

in Figure 6.3). The following fitting and smoothing method is used to estimate the attenuation to remove the noise in the direct attenuation (Figure 6.4b). First, the normalized electric field versus distance curves are smoothed and fitted to quadratics for the bins that have data, as shown with the dashed line in Figure 6.4a. Second, the attenuation between each bin of the fit (the slope of Figure 6.4a) is found and shown in Figure 6.4b. Lastly, the attenuation for a given electric field curve is calculated by the mean of the fitted attenuation where positive attenuation corresponds to a signal loss over distance. The standard deviation of the attenuation is taken as the uncertainty in the attenuation measurement. In the example of Figure 6.4 the fitted attenuation is 1.13 ± 0.35 dB/Mm whereas the attenuation directly from the data (solid line in Figure 6.4b) is 0.79 ± 2.34 dB/Mm. The increased uncertainty in the direct attenuation is due to the variations in attenuation between successive 500 km distance bins. The fitted attenuation is used for all of the following attenuation values.

There is an inherent geometric factor to the measured attenuation values. As the wave expands radially out from the lightning source it is moving through the spherical waveguide of the Earth and Ionosphere. This expansion adds in a factor of $\frac{1}{sin(d/R)}$ to the received wave power at a station, or $\frac{1}{sin(d/R)^{1/2}}$ to the measured electric field. In the geometric factor

d is the great circle distance from the wavefront to the lightning source and R is the radius of the Earth. Because past models, such as *Wait and Spies* (1960), include the effects of the geometric factor in their predictions and past measurements do not correct for it, the measured attenuations will not be corrected for the geometric effect.
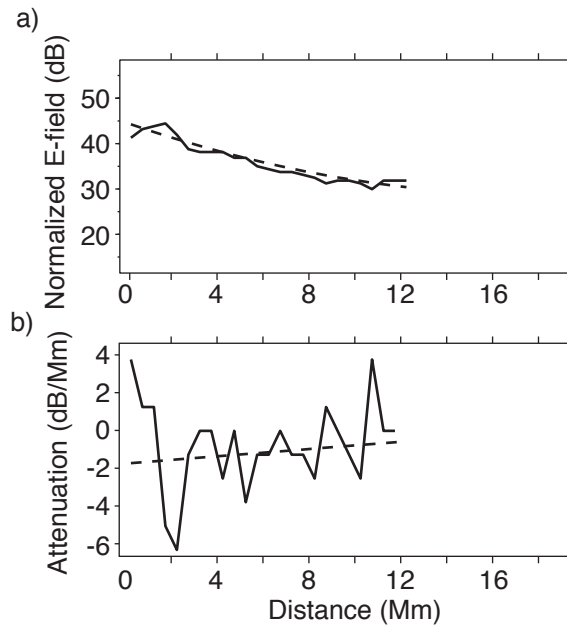
Figure 6.4: Method for calculating the attenuation. In a) the normalized electric field vs distance data (solid line) and the fitted quadratic (dashed line). In b) the change in electric field with distance step (derivative) for the data (solid line) and the fit (dashed line) normalized to dB/Mm.

Overall there is an average attenuation of 1.87 ± 1.06 dB/Mm for all azimuths and times. Within the first 4 Mm of the stroke the attenuation is fairly high for both day and night, with 2.08 ± 0.90 dB/Mm during the day and 2.09 ± 1.02 dB/Mm at night. Beyond 4 Mm the attenuation decreases to 1.29 ± 1.09 dB/Mm for day and 0.24 ± 0.48 dB/Mm for night. The 4 Mm cutoff was chosen based on Figure 6.2 and 6.3 where the eastward and westward-propagating sferics start to clearly differentiate. The increased attenuation near the stroke is likely due to the fast decay of higher order modes that cannot propagate far from their

source (*Wait*, 1970).

For eastward sferics the average attenuation is 1.13±0.35 dB/Mm and 0.71±0.68 dB/Mm for day and night; for westward sferics attenuation is higher with rates of 2.98±0.68 dB/Mm and 2.66±0.39 dB/Mm for day and night paths. The difference in attenuation between eastward and westward-propagating sferics is an increase of 1.9 dB/Mm for day and 2.0 dB/Mm for night.

6.4 Comparisons to Theory

The azimuthal variability of the WWLLN normalized electric-field attenuation is directly compared to the azimuthal variance predicted by *Wait and Spies* (1960) in Figure 6.5. The azimuthal variation, P(ϕ), is found by comparing the attenuation with the presence of a magnetic field to the attenuation without. For the WWLLN data the average attenuation of all azimuths is taken as representative of the attenuation with no magnetic field. The variability of the attenuation rate (in dB/Mm) with magnetic azimuth from *Wait and Spies* (1960) is approximated as $P(\phi) = -0.3 \times \sin(\phi) + 1$.
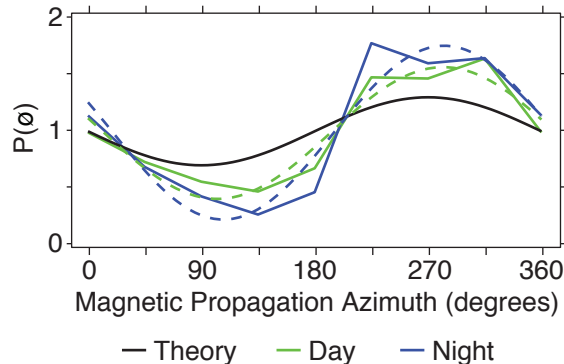
Figure 6.5: Dependence of attenuation with magnetic azimuth shown as P(ϕ), the attenuation normalized to attenuation with no magnetic field. Shown for *Wait and Spies* (1960) (black), day paths (green), and night paths (blue). The best fit curves are shown as dashed lines. Day and night paths are normalized by their mean.

The average attenuation azimuthal variation was fit to the same form as the theory

$(P(\phi) = -a \times \sin(\phi + b) + 1)$ with the resulting fits shown in Figure 6.5. The leading coefficient, a, gives the relative variation in attenuation with azimuth; how much attenuation rates will change between northward/southward and eastward/westward propagating sferics. The day attenuation best fit is $P(\phi) = -0.58 \times \sin(\phi + 348°) + 1$ and the night attenuation best fit is $P(\phi) = -0.76 \times \sin(\phi - 344°) + 1$. Both day and night paths show twice the amplitude relative to the theory ($a_{theory} = 0.3$), with a weaker day dependence and a stronger night dependence, $a = 0.58 \pm 0.18$ and $a = 0.76 \pm 0.26$ respectively. On average the day measurements vary from the theory by 19% and the night measurements by 34%.

In previous measurements of $3 - 30$ kHz VLF attenuation *Taylor* (1960) observed westward VLF paths to exhibit 1-3 dB/Mm more attenuation than eastward paths. This is in line with the WWLLN measured increase of 1.9– 2.0 dB/Mm from eastward to westward-propagating sferics. Similarly the LWPC model shows an attenuation increase of 2 dB/Mm from eastward to westward-propagating sferics for equatorial day paths over the Pacific Ocean. The model also gives a 28% to 37% variability of attenuation between propagation directions, relative to northward propagation, compared to 30% for *Wait and Spies* (1960) and 58% to 77% for WWLLN.

The measured attenuation rates of WWLLN are within the previously measured ranges of attenuation and within the same difference between propagation directions. However, the total azimuthal dependence of VLF propagation has been observed to be greater than that of the theoretical model of *Wait and Spies* (1960) and the propagation model LWPC. The inconsistency may stem from the WWLLN measurements being the average over a range of frequencies while past measurements and models are for specific narrow frequencies, usually VLF transmitters.

6.5   Conclusion

Four years of WWLLN data were used to analyze the normalized VLF electric field from lightning at three island stations, with the variation with magnetic azimuth compared to theoretical results. The electric fields were used to calculate the average attenuation in the $8 - 18$ kHz band at different propagation azimuths. The stroke-receiver paths were selected for sferics propagating over at least 95% water and under either 90% day or 90% night

ionospheric conditions.

It was found that compared to day propagation, night propagating sferics have higher attenuation close to the stroke (2.09 ± 1.02) with less attenuation farther out (0.24 ± 0.48). Similarly attenuation of night sferics have a higher dependence on magnetic azimuth compared to day sferics. Variations with magnetic azimuth showed that westward propagation had $1.9 - 2.0$ dB/Mm more attenuation than eastward propagation for both day and night ionospheric conditions.

Combining three optimally placed WWLLN stations allowed for this examination of the azimuthal dependence of VLF attenuation. Utilizing more of the 70 stations will allow for further investigation of VLF attenuation rates with other path parameters such as ocean salinity, ice, and ground conductivity.

Chapter 7

GLOBAL ELECTRIC CIRCUIT

7.1 Overview

The diurnal variation of the global electric circuit is investigated using the World Wide Lightning Location Network (WWLLN), which has been shown to identify nearly all thunderstorms (*Jacobson et al.* (2006), using WWLLN data from 2005). To create an estimate of global electric circuit activity a clustering algorithm is applied to the WWLLN dataset to identify global thunderstorms from 2010 – 2013. Annual, seasonal, and regional thunderstorm activity is investigated in this new WWLLN thunderstorm dataset in order to estimate the source behavior of the global electric circuit. Through the clustering algorithm, the total number of active thunderstorms are counted every 30 minutes creating a measure of the global electric circuit source function. The thunderstorm clusters are compared to precipitation radar data from the Tropical Rainfall Measurement Mission satellite and with case studies of thunderstorm evolution.

The clustering algorithm reveals an average of 660 ± 70 thunderstorms active at any given time with a peak-to-peak variation of 36%. The highest number of thunderstorms occurs in November (720 ± 90) and the lowest number occurs in January (610 ± 80). Thunderstorm cluster and electrified storm cloud activity are combined with thunderstorm overflight current measurements to estimate the global electric circuit thunderstorm current contribution to be 1090 ± 70 A with a variation of 24%. By utilizing the global coverage and high time resolution of WWLLN, the total active thunderstorm count and current is shown to be less than previous estimates based on compiled climatologies.

Diurnal variation in global thunderstorm activity was original observed by *Wilson* (1921) and *Whipple* (1929) through a combination of thunderstorm day and electric field measurements. Strong correlations between thunderstorm activity and fair weather return current led to the model of the global electric circuit, a system of ionospheric charging and discharging through thunderstorms and fair weather return currents. Studies to date have estimated that globally there are $1000 – 2000$ thunderstorms active at any one time; most concentrated over tropical landmasses, covering $1 – 10\%$ of Earth's surface (*Markson*, 1978; *Rycroft and Harrison*, 2011; *Singh et al.*, 2011a). Previous work on diagnosing the generator source of the global electric circuit used several methods: long time scales (*Tinsley et al.*,

2007; *Liu et al.*, 2010), mathematical models (*Kasemir*, 1977; *Hays and Roble*, 1979; *Roble*, 1991), engineering models (*Ogawa*, 1985; *Kartalev et al.*, 2004; *Rycroft*, 2006), parameterization (*Price and Rind*, 1992; *Williams*, 1985), and thunderstorm overflight estimates (*Mach et al.*, 2011). Most work with the global electric circuit uses long term averaging to recreate the known Carnegie curve of *Whipple* (1929), yet short time scales do not match the long term averaging (*Holzworth et al.*, 1984). The variation of the global electric circuit changes on short time scales that is not resolved in past models or with long term averaged observations. The global electric circuit is an important component to the solar-terrestrial system creating a link between solar activity, the ionosphere, aerosols, cloud microphysics, thunderstorms, weather, and climate (*Tinsley et al.*, 2007; *Holzworth and Volland*, 1986).

Individual lightning strokes and flashes are not yet a reliable method of characterizing the source of the global electric circuit, but global thunderstorm activity is a reliable measure. *Ruhnke* (1969) proposes that the conduction and displacement currents above thunderstorms are the global electric circuit generators source as they vary slowly through the evolution of a thunderstorm and the currents are fairly independent from impulsive events like lightning. *Krider and Blakeslee* (1985) looked at displacement currents below a thunderstorm to find them steady with abrupt, but insignificant, changes due to lightning. *Rycroft et al.* (2000) created an engineering model with three different regions for the the return current, concluding that sprites, lightning, and other transients will have little direct effect on the global electric circuit. *Stergis et al.* (1957) showed that cloud to ground lightning is not necessary for upward current from thunderstorms. With a numerical model of a dipolar thunderstorm *Tzur and Roble* (1985) estimated the average upward current contribution of a thunderstorm to the global electric circuit to be 0.7 A per thunderstorm. Similarly, with a combination of numerical and analytical models of dipolar thunderstorm, *Driscoll et al.* (1992) estimated a total contribution of 0.4 A to the global circuit per thunderstorm. Unlike the other models, *Mareev et al.* (2008) estimated a 50 – 400 A current contribution directly from global lightning; with a similar model *Mallios and Pasko* (2012) investigated the efficiency of lightning and found lightning only contributed 1% (3 A) of the total 300 A contribution from thunderstorms. These models show that thunderstorms contribute an appreciable upward current to the global electric circuit with a wide range of

estimated current contributions.

Balloon and aircraft overflights have been used to estimate the total current contributions from thunderstorms to the global electric circuit. *Stergis et al.* (1957) found thunderstorm currents ranging from 0.6 – 4.3 A, with an average of 1.3 A; these estimates are the lower bound with uncertainties of up to 50%. *Blakeslee et al.* (1989) shows the upward current generated by a thunderstorm to range between 0.1 – 6 A with an average current between 0.5 – 1 A. Other studies show current density over thunderstorms ranging from 10 – 40 pA/m^2 (*Holzworth*, 1981) to -20 – 33 nA/m^2 (*Mach et al.*, 2009). The overflights in previous research found no consistent parameterization between lightning rates and fair weather return current, but recent balloon work found strong correlation between global lightning activity and the fair weather return current on short time scales (*Holzworth et al.*, 2005). Lightning stroke activity and locations cannot directly provide estimates of the global circuit, however they can be used for directly locating and defining active thunderstorm areas and relative intensities.

Compared to satellite and balloon observations, ground based lightning networks have the advantage of continuous observation of large regions. Global very low frequency networks, such as the World Wide Lightning Location Network (WWLLN), are capable of locating lightning around the entire globe. *Holzworth et al.* (2005) compared the stroke counts of a nascent WWLLN to the fair weather return current and found strong temporal correlation between the measurements. A better measure of global circuit activity is the total number of active thunderstorms around the globe; applying clustering algorithms to lightning network data enables the network to locate, track, and monitor global thunderstorm activity.

The WWLLN data are clustered into thunderstorms with the Density-Based Spatial Clustering of Application with Noise (DBSCAN) algorithm (*Ester et al.*, 1996; *Kriegel et al.*, 2011). DBSCAN was chosen as the clustering algorithm for several key features: the capability to handle noise, no requirement to specify the number of clusters, arbitrary cluster shapes, and the insensitivity to the ordering of the data. Clustering lightning strokes into thunderstorms cannot require a designated number of clusters before clustering, as the total number of thunderstorms is not known before clustering. Similar algorithms, such as Ward's

Method, cluster into large unphysical thunderstorms due to noise (*Ward*, 1963). In another approach *Mezuman* (2013) used a connected components methods to cluster the WWLLN data. They found global thunderstorm activity to average near 1000 thunderstorms with significant daily variability. The resulting WWLLN lightning clusters are representative of the lightning active stage of the thunderstorm. Even though the electrically active portion of a thunderstorm extends beyond the active lightning stage (*Jacobson and Krider*, 1976; *Stolzenburg et al.*, 2010), the clustered lightning active stage in this work will be referred to as the clustered thunderstorm or thunderstorm clusters.

7.2   Clustering

7.2.1   DBSCAN

DBSCAN clusters n-dimensional points based on the distance between the points, a length scale ϵ, and the minimum number of points necessary to form a cluster, $minPts$ (*Kriegel et al.*, 2011). Points in a cluster are either core points or non-core points; a point is a core point of a cluster if there are $minPts - 1$ other points within ϵ distance of that point (resulting in $minPts$ points within distance ϵ). In Figure 7.1 the distance ϵ is represented by the circle around each point, lines connect points within ϵ of each other, and core points are shown as filled symbols. Points that are within ϵ of one core point, but less than two core points ($minPts = 3$), are added to the cluster as non-core points and cannot be used to add more points into the cluster. For example the unfilled triangle symbol in Group 1 of Figure 7.1 is within ϵ of one core point and clustered into the group as a non-core point, but cannot connect further unclustered points (crosses) into the cluster.

WWLLN lightning strokes are separated by three dimensions: latitude, longitude, and time. With time as a consideration two clusters that appear to overlap in Figure 7.1a, group 2 (blue) and group 3 (green), are separated by ϵ in time and are distinct groups as seen in Figure 7.1b. DBSCAN is a physically realistic algorithm for clustering lightning data, such as WWLLN data, as the core points of a thunderstorm are the intense lightning centers while edge points are clustered but do not connect disparate lightning centers.
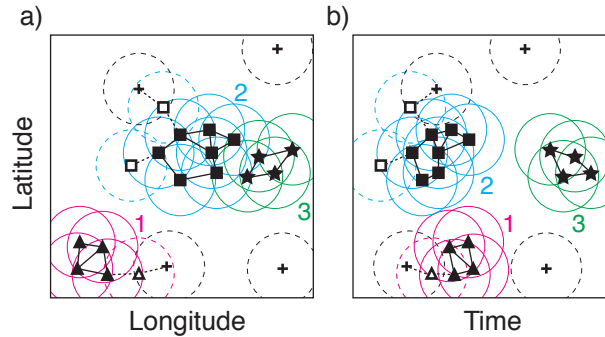
Figure 7.1: DBSCAN clustering example, with $minPts = 3$, showing the same clusters located in a) latitude and longitude and b) latitude and time. Solid rings show the ϵ distance from core points (filled), dashed rings are for non-core points (unfilled). Triangles (1), squares (2), and stars (3) show clustered points, crosses are non-clustered points.

7.2.2   Clustering WWLLN

Accurate thunderstorm clustering of WWLLN data requires optimizing the clustering parameters ϵ and $minPts$. WWLLN requires a second ϵ parameter, ϵ_{time}, for clustering in time. The parameter ϵ corresponds to the physical extent of an average thunderstorm, ϵ_{time} the duration, and $minPts$ the number of lightning strokes necessary to consider a thunderstorm electrically active.

As the clustering parameters are varied the total number of thunderstorms, the average thunderstorm area, and the average thunderstorm duration change. In Figure 7.2a it can be seen that ϵ has a high degree of control over thunderstorm area and total thunderstorms. To balance the total number of thunderstorms, average area (320 km^2), and percentage of strokes clustered (89%, not shown) the best value is found to be $\epsilon = 0.12°$ or about 13 km. To prevent consecutive thunderstorms from being clustered (e.g. one thunderstorm occurring in the same area as a previous distinct thunderstorm storm) ϵ_{time}, Figure 7.2b, needs to be smaller than average duration of a thunderstorm. With this requirement the value of $\epsilon_{time} = 18$ minutes is found; giving an average thunderstorm cluster duration of 16 minutes. The minimum number of strokes needed to produce a cluster is set at $minPts = 2$, as two

detected WWLLN strokes confirms the presence of a thunderstorm. A sharp drop in the total number of thunderstorms can be seen in Figure 7.2c when $minPts$ moves from 2 to 3. To get a majority of strokes included in the thunderstorm clusters while retaining reasonable physical attributes of the thunderstorms the parameters $\epsilon = 0.12°$, $\epsilon_{time} = 18$ minutes, and $minPts = 2$ strokes were chosen.
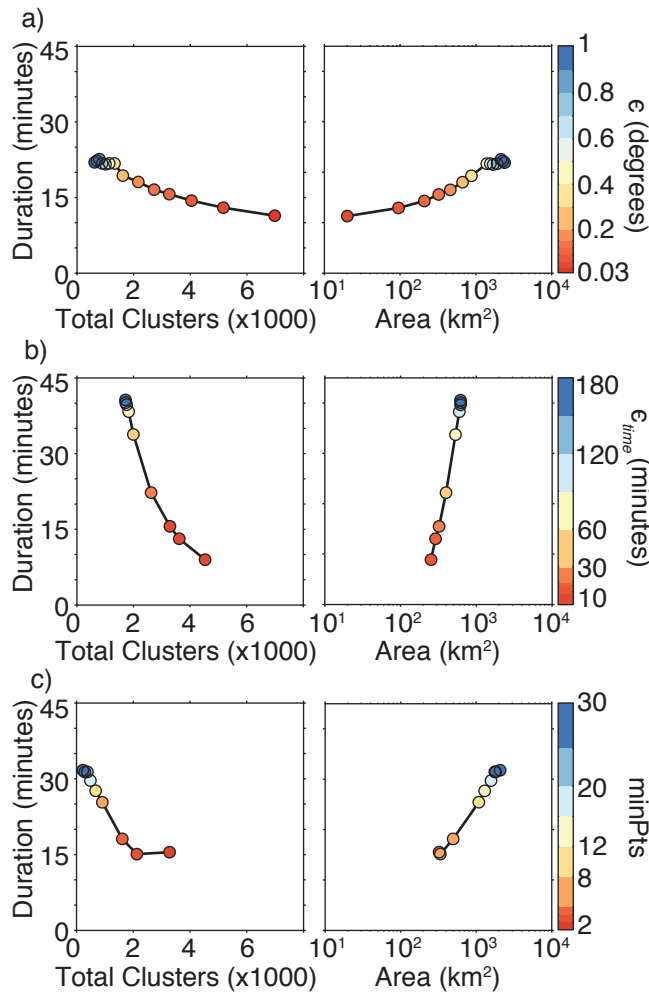
Figure 7.2: Variation in average thunderstorm duration (rows), counts (left column) and area (right column) through varying one clustering parameter with the others held constant (constant set: $\epsilon = 0.12°$, $\epsilon_{time} = 18$ minutes, and $minPts = 2$). a) ϵ varies from $0.03° - 1°$. b) ϵ_{time} varies from $10 - 180$ minutes. c) $minPts$ varies from $2 - 30$ strokes.

7.3   WWLLN thunderstorm clusters

Using the DBSCAN algorithm the individual WWLLN lightning strokes from 2010 June - 2013 June are clustered into active thunderstorms. Clustered WWLLN strokes allow for simple thunderstorm tracking as shown in Figure 7.3; here the strokes comprising a thunderstorm are outlined with polygons every hour, and plotted with opacity increasing with time. Figure 7.3 shows several active thunderstorms on 2013 May 21, 12 – 23 UTC. For clarity clustered thunderstorms with less than 50 strokes were removed from the plot. The thunderstorm opacity increases at 8% per hour with each color corresponding to a single thunderstorm cluster.
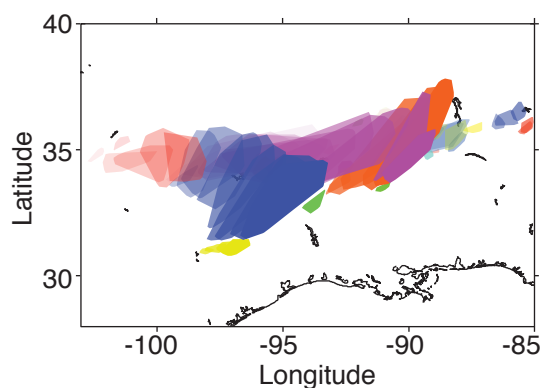
Figure 7.3: Thunderstorm evolution from 2013 May 21, 12 – 23 UTC. Polygons outline active lightning regions, colors correspond to thunderstorm cluster, opacity increases 8%/hour. For clarity, thunderstorms with less than 50 strokes were removed.

The cluster results can be directly compared to active precipitation regions as seen by the TRMM Precipitation Radar (*Kawanishi et al.*, 2000). Rainfall rates from the TRMM data product 2A25 were used as binned rates on a 0.25° grid; in Figure 7.4 the TRMM rainfall data are shown as the background image, gray areas are not in view of the satellite. WWLLN strokes are shown in Figure 7.4 if they occur between the start and end times of the TRMM regional overpass.

In Figure 7.4 two sets of consecutive overpasses are used, the first on 2013 May 06 from
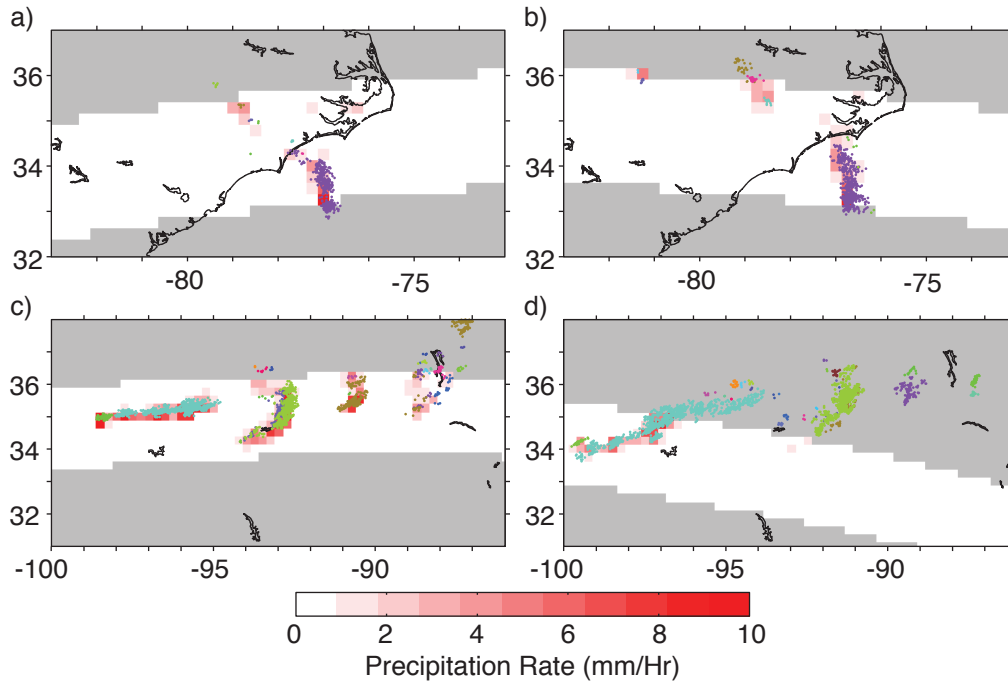
Figure 7.4: WWLLN thunderstorm clusters (identified by color) over TRMM precipitation rate (mm/Hr) for 2013 May 06, 15:49 – 15:59 UTC (a), 17:28 – 17:37 UTC (b), 2013 May 21, 10:04 – 10:13 UTC (c), 11:42 – 11:52 UTC (d). (a) and (b) are successive passes as are (c) and (d), cluster colors are contiguous between passes. Gray areas were outside the range of the TRMM radar.

15:49 – 15:59 UTC (7.4a) and 17:28 – 17:37 UTC (7.4b), the second in on 2013 May 21 from 10:04 – 10:13 UTC (7.4c) to 11:42 – 11:52 UTC (7.4d). These passes were selected as they passed over the same thunderstorms twice with an appreciable amount of lightning activity in view of the satellite. Unlike Figure 7.3, the clustered WWLLN strokes are shown as individual strokes in the thunderstorm with many of the strokes overlapping each other in the center of the clustered regions. The clustered thunderstorms match up well with the TRMM precipitation regions and clearly track the same thunderstorm between the consecutive overpasses. Thunderstorms continue, and are clustered by WWLLN, well after TRMM no longer observes the area (green thunderstorm in Figure 7.4c and 7.4d). There

were no previous or later TRMM overpasses of these thunderstorms.

7.4   *Global thunderstorm activity*

Original estimates of global thunderstorm activity show afternoon peaks in lightning activity for each of the major lightning chimney regions: the Americas, Africa/Europe, and Asia (*Wilson*, 1921). The global averaged WWLLN thunderstorm clusters show the previously measured long term averaged diurnal behavior of the global electric circuit activity, with a diurnal variation of 36%. The three year average of thunderstorm activity is shown in Figure 7.5 for different regions (7.5a), for thunderstorm type (7.5b), and for seasons (7.5c), with the global average displayed in each panel (black). Averages of thunderstorm activity are calculated from the total number of unique thunderstorms every 30 minutes.

Each chimney region is separated in Figure 7.5a with peaks in thunderstorm activity occurring in the local afternoon (Americas 19 UTC, Africa 15 UTC, and Asia 8 UTC) then a slow decrease during the night until a minimum in the early hours of the morning. A strong diurnal variation is evident for thunderstorms over land, seen in Figure 7.5b, with little diurnal variation in oceanic thunderstorm activity. On average WWLLN observes a total of 660 ± 70 thunderstorms on any given day of the year; while the total is lower than previous estimates in the $1000 - 2000$ thunderstorm range, those estimates were made using partial or extrapolated datasets. For each chimney region the average number of thunderstorms are 280 ± 80 for the Americas, 140 ± 60 for African and Europe, and 240 ± 50 for Asia and the Maritime Continent. The long-term thunderstorm behavior observed by WWLLN resembles the previous measurement of global electric circuit behavior. Of the clustered thunderstorms 350 ± 70 were continental and 300 ± 10 were oceanic. This is in contrast to the disparity in the distribution of individual lightning strokes where a majority are continental; the continental thunderstorms tend to be larger with higher stroke rates than those over the oceans.

There are slight changes in the overall diurnal behavior between each season shown Figure 7.5c. The contribution of each chimney region, divided by northern and southern hemisphere, is shown in Figure 7.6. The overall seasonal change in thunderstorm activity is clearly seen with the change in dominant contributor from the northern hemisphere in

a)
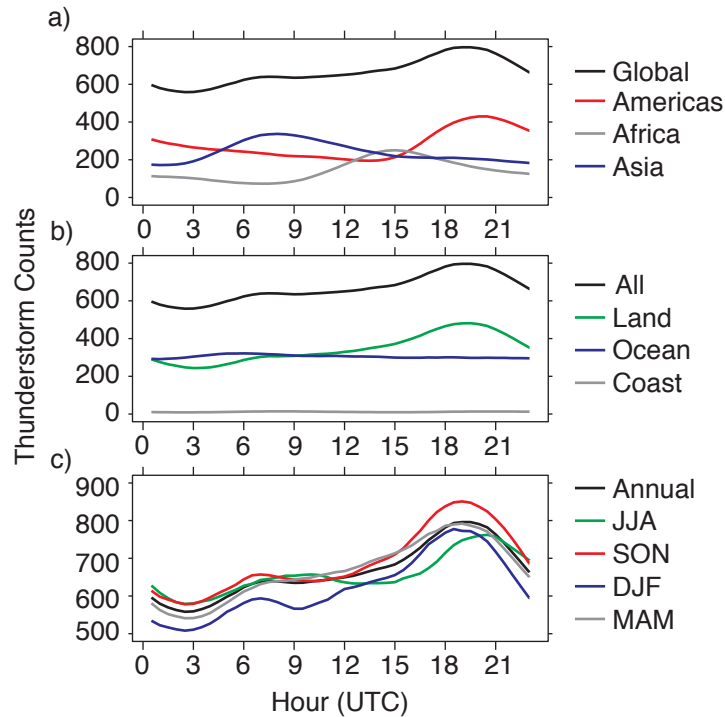
Figure 7.5: Diurnal variation of WWLLN thunderstorm 30 minute counts for 2010 – 2013 obtained using the DBSCAN clustering algorithm. For a) thunderstorms over each major lightning chimney region (divided between longitudes $-180°$, $-30°$, and $60°$), b) all, land, ocean, and coastal thunderstorms (coastal thunderstorms have strokes over land and ocean), and c) the full year and each season.

May – August to the southern hemisphere in November – February. In the shoulder seasons (March – April and September – October) the different chimneys are closer in activity levels to each other. The shift in peak activity times for each chimney region between northern and southern summer reflects the difference in longitudinal landmass distribution of each region. For example the North America peak occurs at 23:30 UTC while the South America peak occurs at 18:30 UTC since the South American lightning regions are, in general, farther east than the North American ones. This time change in regional contributions has been seen in other global circuit measurements; in the Vostok, Antarctica electric field measurements

of the fair weather field, the peak changes from 18:00 UTC in January to 21:00 UTC in August (*Burns et al.*, 2005, 2012).
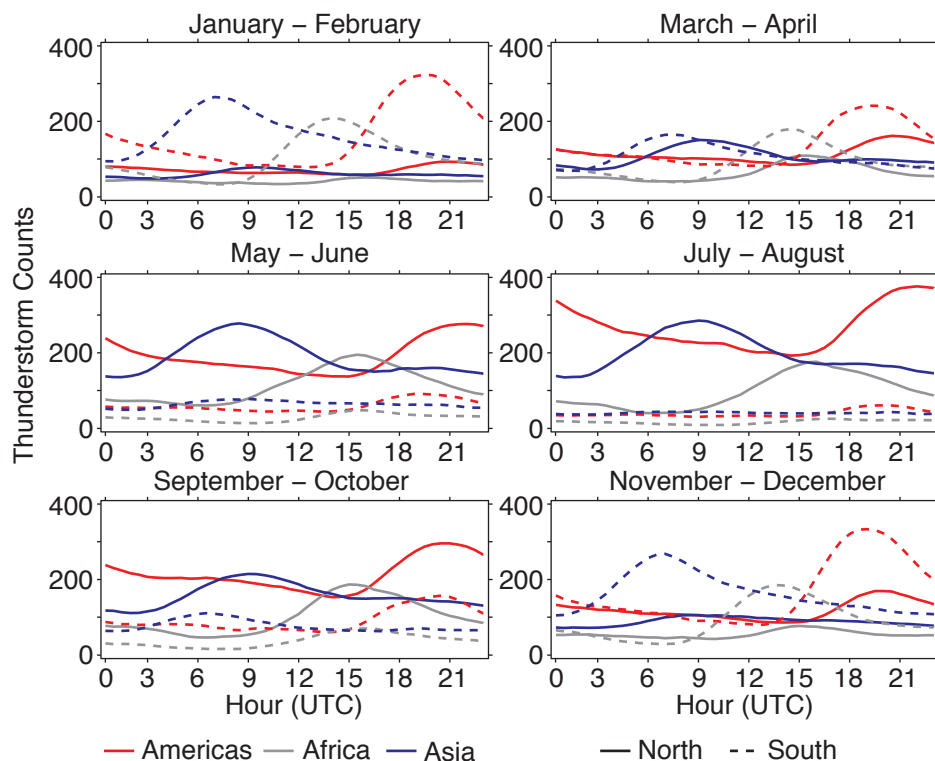
Figure 7.6: Diurnal variation of WWLLN thunderstorms for each major chimney region (colors) divided by hemisphere (Northern Hemisphere solid line, Southern Hemisphere dashed line). Each panel shows two months of thunderstorm clusters averaged over 2010 – 2013.

7.5   *Temporal thunderstorm activity*

As a result of a constantly growing number of WWLLN stations, the number of WWLLN strokes detected increases with time due to improvements in the network and the detection efficiency. As a result long term tracking of stroke rate cannot be used without deconvolving detection efficiency improvements. However thunderstorm counts have remained relatively constant while stroke rate has increased; WWLLN has been capable of detecting almost every thunderstorm since 2005 (*Jacobson et al.*, 2006). In Figure 7.7a the daily average

30 minute thunderstorm counts (black) are plotted alongside the daily average stroke rate (red). It can been seen that the stroke rate has increased relatively steadily while the thunderstorm count has remained relatively stable.
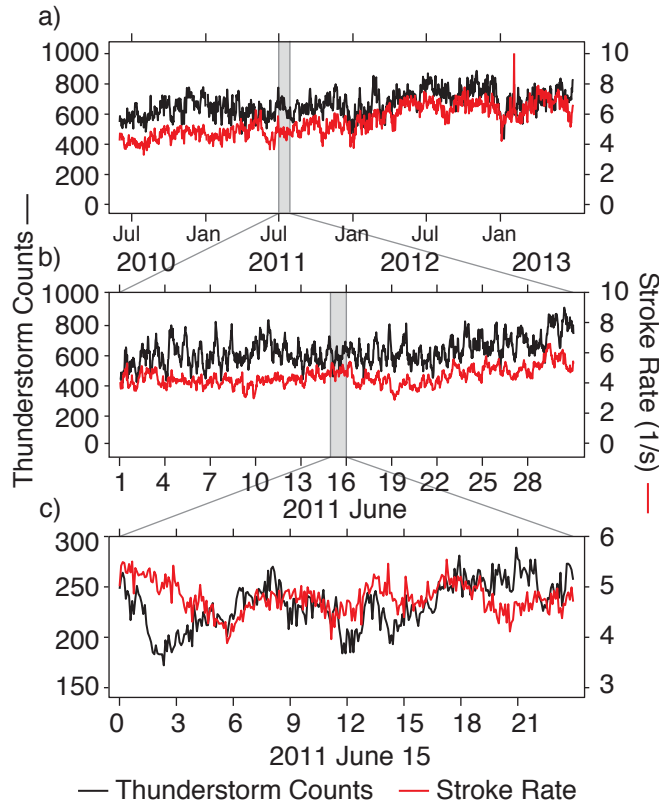
Figure 7.7: Variation in WWLLN thunderstorm count (black) and stroke rate (red) for: a) daily averages 30 minute counts from 2010 June – 2013 June, b) 30 minute counts from 2011 June 01 – 30, c) 5 minute counts from 2011 June 15, 00 – 23 UTC.

The daily average thunderstorm count remains relatively constant during 2010 – 2013, while the stroke rate increases during the same time span. The thunderstorm count increased an average of 3% per year, while the stroke had a much higher yearly increase of 13%. Similarly 90% of the daily thunderstorm averages are within 20% of the mean for the three years. Previous work has suggested that changes in climate will cause a change in global thunderstorm behavior (*Williams*, 2005; *Price*, 2009). During 2010 – 2013 the average

global surface temperature and thunderstorm count have both remained relatively constant (*Hansen et al.*, 2013), but future changes in global temperature may be reflected in the global average thunderstorm count.

Figure 7.8 shows that the typical diurnal behavior of Figure 7.5 emerges only after long term averaging. When examining the thunderstorm counts on a monthly scale (Figure 7.8, dot-dash line) the long term average begins to emerge, while over a single day (Figure 7.8, dashed line), the expected diurnal behavior is missing. Similarly, on the daily scale of Figure 7.7c it can be seen clearly that stroke rate does not follow thunderstorm counts as located by WWLLN. Despite the short time scale variation, the overall average of the data still accurately reproduces the expected global thunderstorm activity in Figure 7.5. With such short time scale variations global observations of the global electric circuit source mechanism, be it with ground lightning networks or several geostationary satellites, need to occur along with an accurate measure of the return current to better understand the charging of the global electric circuit.
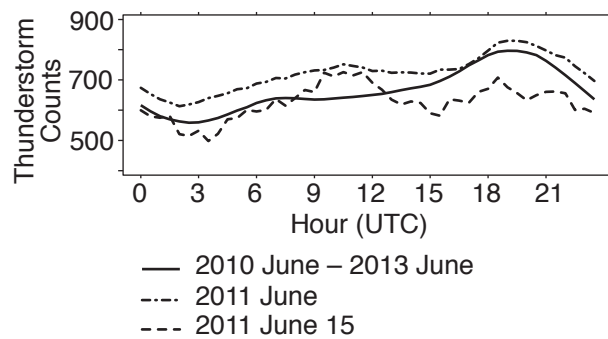
Figure 7.8: Diurnal UTC variation in WWLLN 30 minute thunderstorm count for: multi-year average of Figure 7.7a (solid line), monthly average of Figure 7.7b (dot-dash line), and 30 minute averages of Figure 7.7c (dashed line).

7.6 *Global Electric Circuit Thunderstorm Contribution*

With the WWLLN thunderstorm clusters a simple model is made to estimate the total contribution of thunderstorms and electrified storm clouds to the global electric circuit.

Mach et al. (2011) considered LIS/OTD observed land thunderstorms with less than 1.7 flashes min^{-1} and ocean thunderstorms with less than 0.33 flashes min^{-1} electrified storm clouds. For an average storm duration of 15 minutes these cutoffs are 26 flashes for land and 5 flashes for oceanic electrified storm clouds (ESCs). With a WWLLN-LIS/OTD detection efficiency of 6.4% over land and 17% over oceans (*Rudlosky and Shea*, 2013), WWLLN would expect cutoffs of 2 strokes per thunderstorm over land and 1 stroke per thunderstorm over oceans. Given the clustering parameters used in this work requires a minimum of 2 strokes to be considered a thunderstorm, every un-clustered WWLLN stroke is considered a single ESC. The daily average of thunderstorms and ESCs observed by WWLLN are shown in Figure 7.9a. Since not all ESC will produce lightning the counts shown in Figure 7.9a should be considered a low estimate.
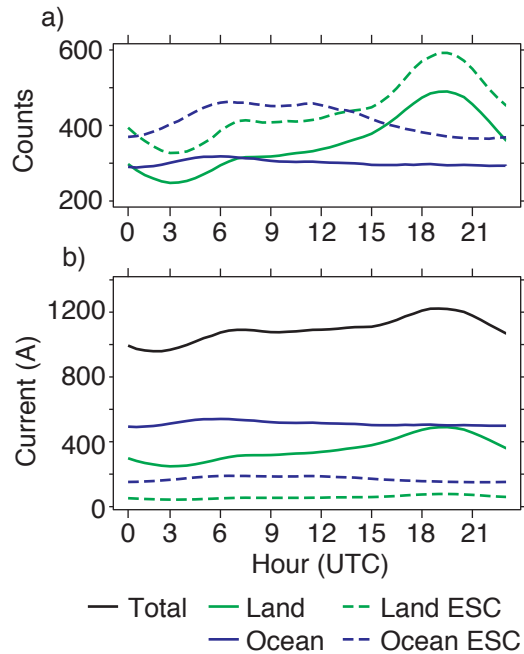
Figure 7.9: A simple model of the total global electric circuit current with contributions from land thunderstorms (green solid), oceanic thunderstorms (blue solid), land electrified storm clouds (green dashed), and oceanic electrified storm clouds (blue dashed). a) shows the counts for each group and b) the current contribution to the total (black line).

The average thunderstorm and ESC current contribution is taken from the overflight data of *Mach et al.* (2010). Average current contribution for land thunderstorms is 1.0 A, for oceanic thunderstorms 1.7 A, for land ESCs 0.41 A, and for oceanic ESCs 0.13 A. The total current for each contributor and the total current is shown in Figure 7.9b. The average thunderstorm current is found to be 1090±70 A with a total peak-to-peak variability of 24%. The largest contributor are oceanic thunderstorms (47%) with 510 ± 10 A, followed by land thunderstorms (32%) with 350 ± 70 A; overall ESCs contribute 21% to the thunderstorm global circuit current.

With a similar model based on the TRMM LIS/OTD data, *Mach et al.* (2011) found ocean thunderstorms contribute 32% and land thunderstorms 55%, with a total ESC contribution of 13% to the total mean current of 2.04 kA. Their difference in current and contributing fraction stems from an increased count of land thunderstorms. This highlights a shortcoming of this simple model: land thunderstorms tend to be larger than oceanic storms; thunderstorm area should be taken into account along with overall counts. However to validate any global circuit model a comparison needs to be made with simultaneous fair weather return current measurements in order to constrain the models.

7.7 Conclusion

Global thunderstorm count is a good measure of global electric circuit activity and for a simple model of the circuit; the location and size of thunderstorms is necessary along with totals to create a more accurate model of the real time fair weather return current. WWLLN strokes are successfully clustered into thunderstorms using the DBSCAN clustering algorithm with appropriately chosen clustering parameters. The clustered thunderstorms were compared against the TRMM Precipitation Radar and a case study of thunderstorm tracking and evolution. When the three years of WWLLN data were averaged the diurnal behavior of global thunderstorm activity aligned with the expected behavior of both thunderstorms and the fair weather return current. The results of global thunderstorm and electrified storm cloud activity are combined with upward thunderstorm current averages to create an estimate of the fair weather return current. The model found an average thunderstorm current contribution of 1090 ± 70 A. This, and future, estimates of the current can

be validated against an accurate fair weather return current measurement.

Acknowledgments for this Chapter

Chapter 8

THUNDERSTORMS AND FLASHES

8.1 Overview

Application of clustering algorithms to ground based lightning detection networks expands the real time global observations of lightning from strokes to flashes and strokes to thunderstorms. Lightning detection networks, such as WWLLN or ENTLN, are then able to identify, locate, and analyze nearly every active thunderstorms within their operational range. Global thunderstorm information allows for research into the climatological structures of thunderstorm behavior on large spatial and temporal scales. Flash clustering allows for new network diagnostics, such as flash multiplicity and thunderstorm detection efficiency.

The DBSCAN algorithm is used to cluster strokes to flashes and thunderstorms for both the WWLLN and ENTLN. Cross validation of the networks is performed with the located thunderstorms and comparisons of their inferred areas and duration. Overall WWLLN detects 61% of all ENTLN thunderstorm clusters and 80% of thunderstorms larger than 10^3 km^2. In the reverse analysis, ENTLN detection of WWLLN thunderstorms, ENTLN detects 86% of all WWLLN thunderstorms over North America. On average WWLLN observes thunderstorm clusters lasting 10 minutes and spanning 66 km^2; ENTLN observes averages of 10 minutes and 60 km^2. Within thunderstorms the average time between flashes is 21 seconds as seen by WWLLN and 10 seconds by ENTLN, with a strong dependence on season. Clustering algorithms applied to lightning detection networks allow for a new range of analysis from thunderstorm effects, network performances, to the links between lightning and thunderstorms properties.

Lightning network detection efficiency is often measured in terms of relative stroke or flash performance through pair-wise comparisons of different systems. For example comparing two ground based networks (*Abarca et al.*, 2010), a ground based network with a satellite (Chapter 4, *Rudlosky and Shea* (2013)), or a network to an actual ground truth such as rocket-and-wire triggered lightning (*Nag et al.*, 2011). In some cases, such as examining the global electric circuit (*Hutchins et al.* (2014), Chapter 7), it is the location and properties of the thunderstorms generating the lightning that are important. The thunderstorm detection efficiency for a lightning location system is then important in these applications, both the location, duration, and extent of the thunderstorm.

Within known thunderstorms the time between successive flashes can be used to examine further properties of the thunderstorm such as charging rate, controls on lightning strength, life cycle, and dynamics of the thunderstorm. *Zoghzoghy et al.* (2013) examined the effect of the time between flashes as a measure of thunderstorm discharging in terms of the supression of future flashes. *Zipser* (1994) examined the relation between the flash rate in a thunderstorm with the updraft velocities, which in turn are controlled by the differential surface heating below the thunderstorm. Only after considering and validating the performance of lightning detection networks can the effects of a flash be directly related to properties and behaviors of the parent thunderstorm. Inferring thunderstorm properties based on the observed lightning flash rates can benefit regions with low availability of direct or continuous observations (e.g. radar over oceans).

8.2 Data

The global WWLLN data will be used everywhere in this work unless specifically restricted to a given region, the ENTLN data used will only be the North American subset of the data. The data are from 2011 and 2012, with the exception of Section 8.5 which extends through June 2013.

8.3 Methods

Clustering the located strokes of both networks provides the thunderstorm and flash data. The clustering is performed with the DBSCAN algorithm (*Ester et al.*, 1996; *Kriegel et al.*, 2011) following the application methodology discussed in Chapter 7 (*Hutchins et al.*, 2014). DBSCAN is used over other clustering methods as it clusters based on the spatial and temporal distance between strokes with robust handling of noise (e.g. nearby thunderstorms). The flash clustering uses the same algorithm as the thunderstorm clustering, with adjustment of the spatial and temporal clustering parameters.

For thunderstorms, WWLLN strokes are clustered together if they occur within 0.5° and 18 minutes; for ENTLN it is 0.25° and 15 minutes. The tighter restraints on the ENTLN clustering stem from the increased network performance of by ENTLN compared to WWLLN, notably a better location accuracy ($<$ 1 km compared to 5 km). For flash

clustering the strokes of both networks are clustered if they are within $0.12°$ and 1 second of each other; the 1 second timing is based on the natural inflection point observed in the WWLLN data (see Figure 8.6 in Section 8.5).

The thunderstorm area is estimated by the best fit ellipse encapsulating all of the strokes in the cluster. This will not be the areal extent of the thunderstorm system, rather just the electrified lightning region of the thunderstorm. Similarly, thunderstorm duration is taken to be the time from the first stroke in a thunderstorm cluster to the last. With these thunderstorm clustering parameters both networks observe similar average thunderstorm size and durations over North America. WWLLN observes a median area and duration of 66 km^2 and 10 minutes, while ENTLN observes an average of 60 km^2 and 10 minutes.

Two parameters used throughout this chapter are the times between successive strokes or flashes. This interstroke time will refer to the time since the previous stroke in the parent thunderstorm cluster; interflash time will refer to the time since the start of the previous flash in the thunderstorm cluster. When discussing interflash time in relation to flash energy or peak current the values refer to the strength of the second flash.

8.4   Detection Efficiency

The WWLLN thunderstorm detection efficiency is examined by comparing the located thunderstorm clusters with those located by ENTLN. ENTLN does not have 100% detection efficiency of strokes, it is used as a readily available baseline to compare to as a nominal ground truth for WWLLN. Two thunderstorm clusters are considered to be matches if the convex hull of their constituent strokes overlap to within 5 km and ± 5 minutes. Using a convex hull method requires at least three strokes in a thunderstorm for it be considered a match; this creates an artificially lower detection efficiency for WWLLN for the cases where only $1 - 2$ strokes are detected in a thunderstorm.

Over the $2011 - 2012$ period in North America the average thunderstorm detection efficiency of WWLLN compared to ENTLN was 61% while ENTLN has a detection efficiency of WWLLN of 86%. Similar to how detection efficiency varies with stroke strength, WWLLN has a higher detection efficiency of larger thunderstorms than smaller ones. The spatial distribution of the detection efficiency is shown in Figure 8.1. In the southern regions,

where there is typically higher thunderstorm activity, WWLLN has a detection efficiency above 70%. In regions with few thunderstorms (e.g. the West) there is a lower detection efficiency, possibly due to a lower detection efficiency of both networks in that region.
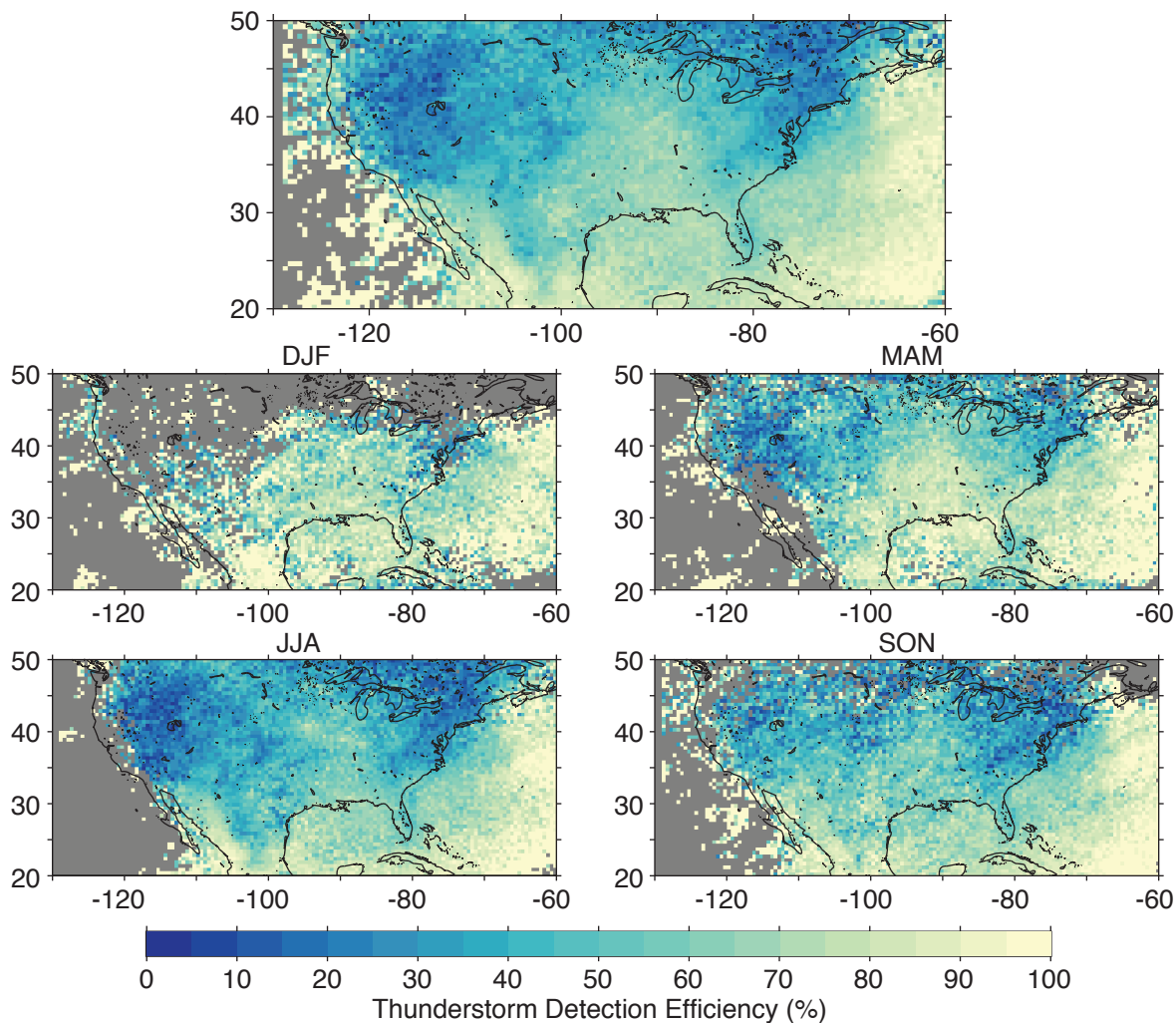
Figure 8.1: WWLLN detection efficiency of ENTLN thunderstorms over North America for 2011 – 2012 (top panel), with each season broken out in the lower panels.

The seasonal behavior shows a similar overall distribution of detection efficiency, with an overall increase in winter (DJF) and spring (MAM) compared to summer (JJA) and fall (SON). A winter increase, 68% compared to 56% in summer, is tied to the decreased occur-

rence of thunderstorms during the cold season. On a regional scale a decrease in activity allows for the networks to lower their detection thresholds and increase the sensitivity of the nearby stations. In all seasons WWLLN has the highest detection efficiency over the oceans. Increased detection efficiency at the limits of the ETNLN detection range (e.g. in the center of the Atlantic Ocean) is convolved with the decreased efficiency of ENTLN and not solely an increase in the WWLLN detection efficiency.

The daily spatially averaged detection efficiency is shown in Figure 8.2. Here the seasonal variation is evident; the high variation in the winter reflects the lower overall thunderstorm counts during these months. With fewer lightning strokes and thunderstorms during the winter, both networks are able to lower the detection threshold at the nearby stations and increase their detection sensitivity. The annual variation in detection efficiency is imprinted on the dependence of detection efficiency with thunderstorm area, size, and duration; the relative scale and behavior of the annual signal is seen in every other detection efficiency relation.
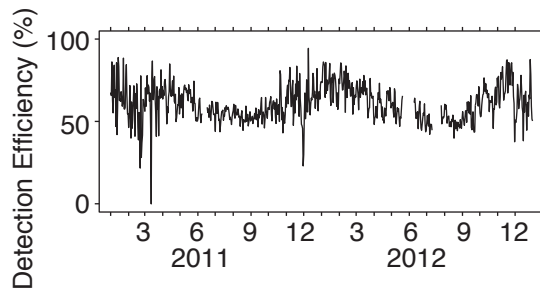
Figure 8.2: Spatially averaged detection efficiency of ENTLN thunderstorms over North America.

The temporal distribution of thunderstorm detection efficiency can be split into different thunderstorm sizes (Figure 8.3a), total counts (Figure 8.3b), and durations (Figure 8.3c). The daily detection efficiency increases for larger, productive, and longer thunderstorms. For thunderstorms larger than 10^2 km^2 the detection efficiency is 53%, while for those larger than 10^3 km^2 it is 80%. In a similar vein, thunderstorms with more than 10^2 strokes are

detected with 75% efficiency and for more than 10^3 strokes at 94% efficiency. Almost all thunderstorms observed are less than 1 hour in duration (Figure 8.3c, right); for those lasting more than 2 hours the detection efficiency increases to 93%.
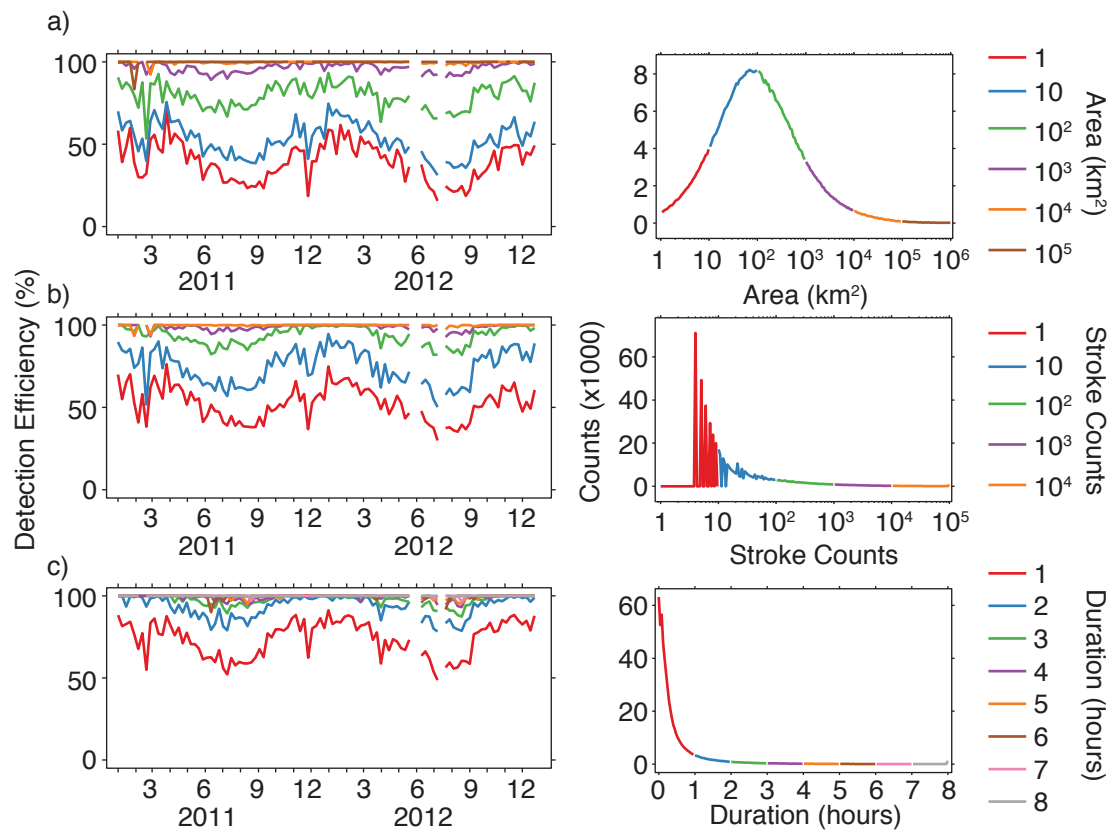
Figure 8.3: Temporal variation in thunderstorm detection efficiency for different thunderstorm sizes: a) thunderstorm area, b) thunderstorm stroke counts, and c) thunderstorm duration. The distribution of thunderstorms for each parameter is shown in the right panels.

Removing the temporal variation demonstrates the trend of increasing detection efficiency for each thunderstorm parameter. In Figure 8.4 the black lines represent all thunderstorms and the colored lines split the thunderstorms by total counts detected by WWLLN in these thunderstorms. Thunderstorm area, Figure 8.4a, shows a gradual increase in detection efficiency from 40% to near 100% efficiency above 10^3 km^2; the total number of

strokes does not have an independent effect from area of the thunderstorm. Similarly the longer a thunderstorm is active the more likely it is to be detected by WWLLN, shown in Figure 8.4c. On short time scales the thunderstorm detection efficiency is lower than the previous estimates of *Jacobson et al.* (2006), but for thunderstorms lasting more than 1 hour the average detection efficiency is 91%. Unlike area, duration is convolved with the lightning count of the thunderstorm: smaller thunderstorms have lower detection efficiency regardless of duration and larger thunderstorms are easier to detect.
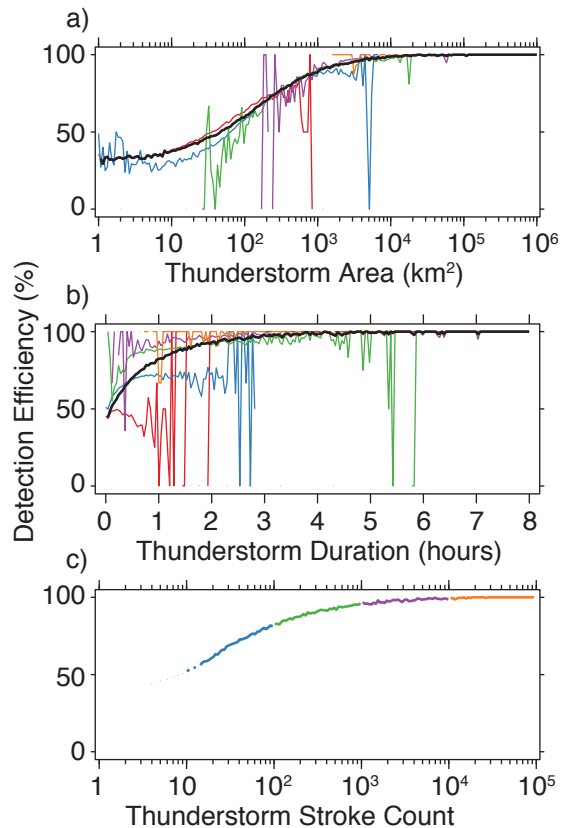
Figure 8.4: Thunderstorm detection efficiency for different thunderstorm a) areas, b) stroke counts, and c) durations. Area and duration plots are shown with the overall average (black) and for different total stroke counts (colors) within the thunderstorms.

Matching thunderstorms between the networks shows the overall performance of the

networks but not how the characteristics of the thunderstorms agree between the two networks. If the thunderstorm clusters are capturing intrinsic properties of the thunderstorms independent of the overall detection efficiency, then the properties should be same between both networks. For example, if WWLLN measures a thunderstorm cluster to have an area of 66 km^2 with 50 strokes and ENTLN measures an area of 60 km^2 with 200 strokes it can be said that both measure the actual estimated area of that thunderstorm cluster.

The properties of the matched thunderstorm between WWLLN and ENTLN can be directly compared for thunderstorm area and duration. Density plots of the two networks estimation for each matched thunderstorm are shown in Figure 8.5; the counts in each bin are displayed on a log scale. Each characteristics in Figure 8.5 shows a strong distribution along a single 1-to-1 line. Area, Figure 8.5a, has good agreement between both networks with 46% of matches within $\pm25\%$ of each other; there is increased deviation for smaller thunderstorms compared to larger ones. The WWLLN overestimation compared to ENTLN thunderstorm area may be caused by incorrect matches or accidental merging between two thunderstorm clusters. The thunderstorm duration has the stronger relation of the two parameters, with 33% of matches falling within $\pm25\%$ of the 1-to-1 line. There is fairly good agreement between the two networks on thunderstorm properties despite the differences in overall detection efficiency.

The overall detection efficiency of WWLLN for ENTLN thunderstorms is lower than previous estimates of the networks thunderstorm detection performance on the order of 90% (*Jacobson et al.*, 2006); in this case it likely stems from only 67% of WWLLN strokes clustered into thunderstorms. The remaining 33% of total strokes would increase the overall detection efficiency of thunderstorms for the cases where only $1-2$ WWLLN strokes match ENTLN thunderstorms. In contrast ENTLN had 96% of strokes clustered into thunderstorms due, in part, to the higher overall efficiency of the network. The identified WWLLN thunderstorms exhibited similar characteristics to their corresponding ENTLN thunderstorms.
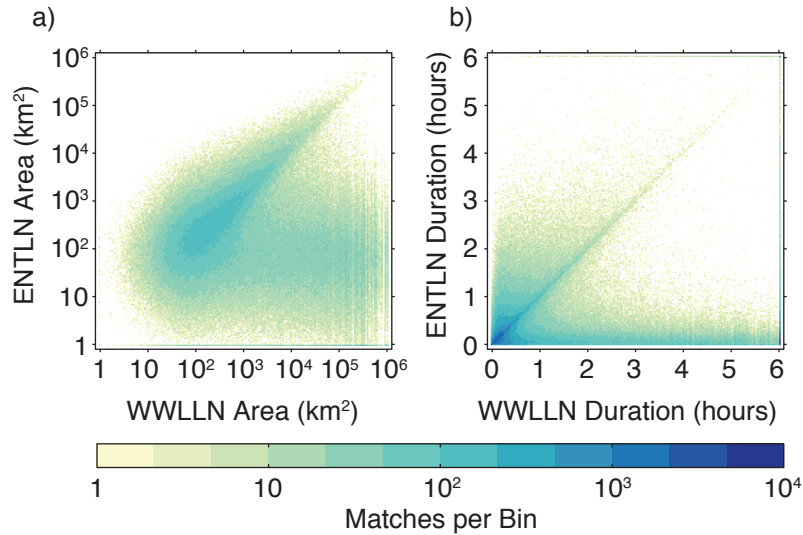
Figure 8.5: Thunderstorm comparison of matched thunderstorms for WWLLN and ENTLN for a) area and b) duration. Note: the density levels are on a log scale.

8.5   Flash Clusters

Within a thunderstorm cluster lightning detection networks are able to measure two additional properties: the interstroke and interflash timing. The full WWLLN interstroke time, Figure 8.6a, shows two distinct peaks: one at 40 ms and another at 100 seconds with a clear minima at 1 second. With ENTLN the interstroke distribution does not show the same distinct peaks (Figure 8.6c); because it is able to locate more strokes in each flash, the tail of the interstroke time distribution overlaps the interflash time distribution. In the WWLLN interevent time distribution there is a spike of events near 10 μs (shown as points) caused by the same event recorded twice by the network, this also occurs less frequently with ENTLN.

After the network events are clustered into flashes the time distributions can be split into the time between strokes in the same flash (black) and time between flashes (blue), shown in Figure 8.6b and 8.6d. The stroke and flash distributions can be fit as lognormal distributions for each day of the year, an example set of fits is shown with the dashed
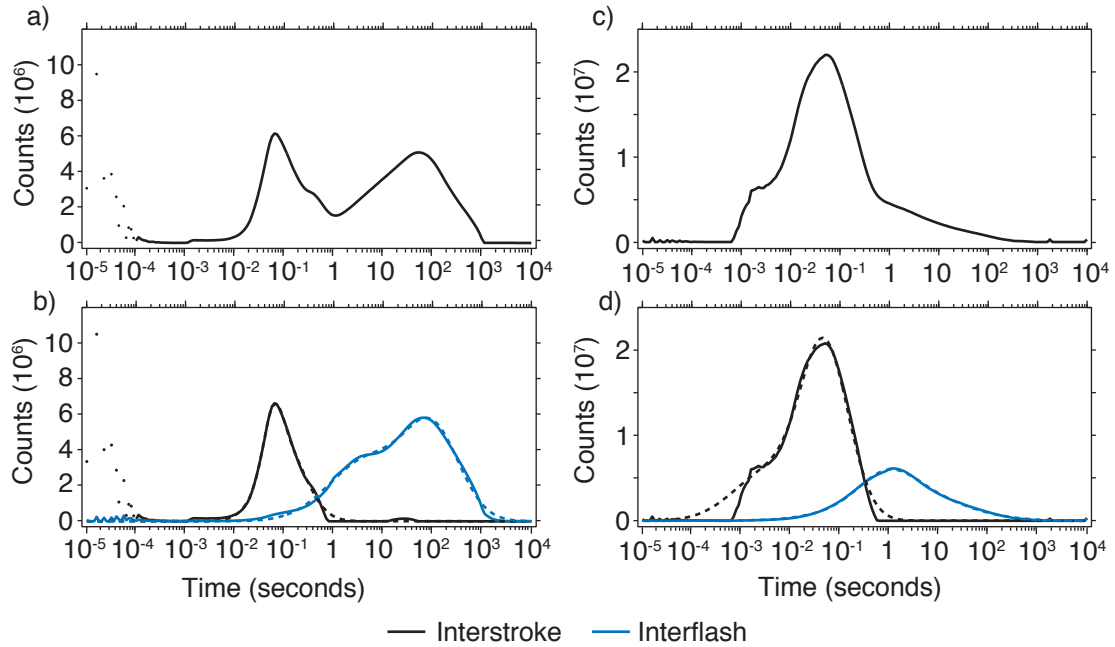
Figure 8.6: The inter-event times for WWLLN (a) and ENTLN (c) and the interstroke (black) and interflash (blue) time distributions for WWLLN (b) and ENTLN (d). The dashed lines correspond to the best lognormal fits of the distributions. There are 20 logarithmically spaced bins per decade.

lines in Figure 8.6b and 8.6d. This fitting allows for the daily tracking of the interstroke (Figure 8.7a) and interflash (Figure 8.7b) times for both networks. With WWLLN the global distributions (black) remain relatively centered at 71 ms and 100 seconds; over North America WWLLN averages (blue) are 60 ms for interstroke and 39 seconds for interflash times. ENTLN (red) has lower daily averages of 53 ms and 17 seconds due to detecting more strokes within each thunderstorm. The North American WWLLN distribution more closely matches the seasonal behavior present in the ENTLN distribution with a small offset in timing; the average daily offset in interflash timing is 14 seconds, with WWLLN detecting flashes further apart than ENTLN. The longer interflash time is directly related to the detection efficiency of both networks, with the longer WWLLN interflash time reflecting the lower detection efficiency of the network.
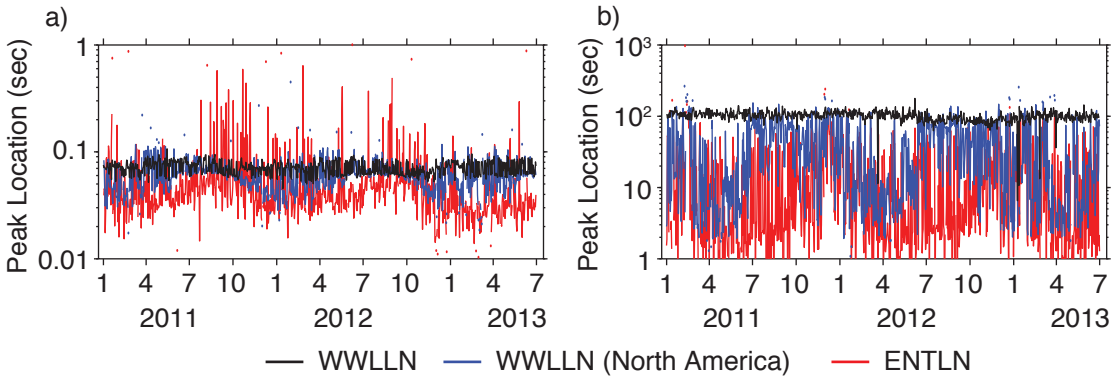
Figure 8.7: Peak of interstroke (a) and interflash (b) times for WWLLN (black), WWLLN over North America (blue), and ENTLN (red). Points beyond 2th and 98th percentiles shown as dots for WWLLN North America and ENTLN.

Directly comparing the time evolution of a thunderstorm will not work with thunderstorms of differing durations. To account for this the thunderstorm life cycle is normalized by the duration of the thunderstorm. Each thunderstorm is broken into 9 time segments where the first flash in the thunderstorm is at time 1 and the last flash at time 9. With this normalization the average interflash time is found at each step for all thunderstorm clusters located by the two networks, shown in black in Figure 8.8. The interflash time can be seen to decrease to the center of a thunderstorms life cycle before increasing at the end; this is consistent with the known life cycle of a thunderstorm (*Peckham et al.*, 1984; *Rakov and Uman*, 2003). The interflash time for different energy (Figure 8.8a) and absolute peak current (Figure 8.8b) deciles are also shown (energy and peak current from the second flash). The interflash values shown are the median values of the interflash distribution (e.g. Figure 8.6c, blue) with median absolute deviations of 74% for WWLLN and 98% for ENTLN.

All flashes follow the same behavior, regardless of energy or peak current, however there are offsets in interflash time for different strength strokes. As seen with WWLLN, Figure 8.8a, the interflash time increases for more energetic strokes; for ENTLN, Figure 8.8b, the interflash time increases for absolute peak current. Between normalized time 1 and 5
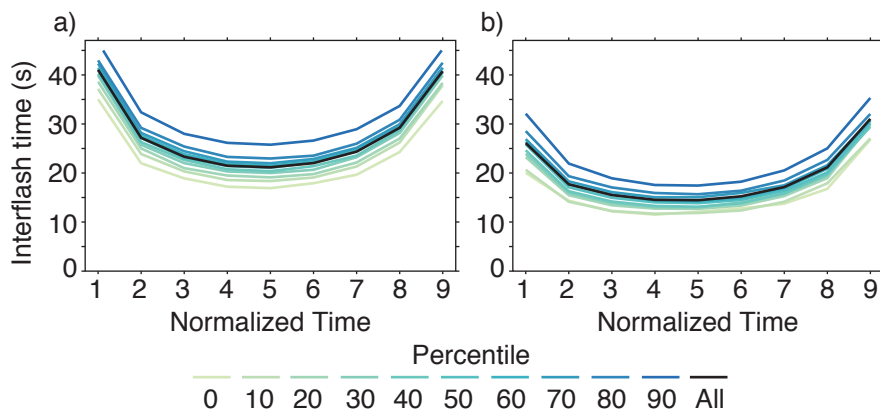
Figure 8.8: The interflash time, since the previous flash, for time-normalized thunderstorms for WWLLN (a) and ENTLN (b), with times shown for all flashes (black) and by stroke strength decile bins. WWLLN strength is divided by flash energy and ENTLN by absolute peak current of the second flash.

the interflash time for the 90th percentile and above stroke strengths decreases by 21% for WWLLN and 11% for ENTLN compared to the median interflash time. This shows that longer times between flashes is associated with stronger flashes; more charge separation can occur in the thunderstorm resulting in the stronger flashes. In Figure 8.8 the separation between different flash strengths is consistent at all points in the thunderstorm life cycle. It is not a constant offset but increases at the weaker stages of the thunderstorm (beginning and end) where longer time is necessary for a stronger flash, caused by a decreased charge separation rate. The time between flashes is a parameter associated with the strength of the resulting flash.

8.6 Conclusion

When compared to ENTLN, WWLLN performs better when detecting large and active thunderstorm regions. The lower performance for smaller thunderstorm clusters may be due to the cutoff in the clustering that requires at least 3 strokes for a cluster, removing 39% of strokes from the analysis. Including these strokes would lead to less robust matches between

the networks and conflate a stroke to thunderstorm detection efficiency and the thunderstorm to thunderstorm detection efficiency. For the thunderstorms WWLLN does detect, the characteristics of the thunderstorm are on par with those of ENTLN. As WWLLN does not detect every ENTLN thunderstorm, ENTLN does not detect every WWLLN thunderstorm. So assuming ENTLN as a ground-truth leads to lower estimates of the WWLLN performance as ENTLN only detects 86% of WWLLN thunderstorms.

On a global scale there is little change in the time between successive strokes and flashes within thunderstorms, but locally there are strong seasonal variations. Both ENTLN and WWLLN observed the same seasonal variation in North America, with a small offset in their times possibly due to the general detection efficiency differences in the networks. The seasonal behavior shows a decrease in interflash times from the end of winter to the beginning of summer; thunderstorms in these times have higher flash rates compared to thunderstorms during the rest of the year.

The charging rate for thunderstorms evolves through the lifecycle of the thunderstorm, with the highest lighting flash rate during the middle of a thunderstorm. The observed change in interflash time reflects the changes in the charging rate of the thunderstorm; the resulting flash strength depends on the charging rate and the total time charging between flashes. If charging rate was constant with time, then flash strength would be constant with interflash time through the life of a thunderstorm. For a given thunderstorm, the flash strength depends in part on the time since the previous flash and the current convective activity of the thunderstorm.

Chapter 9

CONCLUSION AND FUTURE WORK

My work in this dissertation started with the WWLLN producing only lightning stroke locations and uncalibrated RMS energy values of the detected strokes. To enable new and extended analysis with the dataset I developed techniques to measure the energy per stroke, model the relative detection efficiency of the network, and cluster lightning into flashes and thunderstorms. To prove and validate the new techniques they were applied to problems and questions that have only been addressed with limited extent or duration; the new techniques allowed for broader, longer, and more detailed analysis.

The new analysis were: an examination of the difference between continental and oceanic thunderstorm energies, the dependence of VLF propagation on magnetic azimuth, a model of the thunderstorm contribution to the global electric circuit, and an analysis of observed thunderstorm and flash properties. These analysis were made using WWLLN by itself or in conjunction with other lightning detection systems.

Most of the research completed here naturally leads to future projects and research avenues that can be explored in the context of this work. Several of the possibilities for research with WWLLN are: network characterization, thunderstorms, the global electric circuit, and other lightning processes are discussed in this chapter.

9.1 WWLLN Characterization

The relative detection efficiency model developed in Chapter 3 provides one view of the WWLLN network performance but it relies on several assumptions. Either the model can be advanced to not rely on these assumptions or those assumptions themselves can be directly monitored. The main assumption is that the distribution of lightning energy is the same everywhere; this was seen not to hold in Chapter 5 between land and oceanic thunderstorms. The model can take into account the varying energy distributions, or the expected variance, through the variability of the observed regional energy distributions over time.

A promising method for examining the energy detection efficiency is to check how the measured energy values change when the station configuration changes. For example, processing the energy data after artificially removing a few stations to quantify how the energy distributions change. Compare the two energy distributions to the change in detection ef-

ficiency relative to a reference network (e.g. ENTLN or LIS) to better parameterize the model.

Similarly, the accuracy of the energy measurements themselves can be improved by adapting the energy per stroke processing to include multiple well calibrated WWLLN stations. In the processing described in this work the entire network calibration is bootstrapped from one station; the addition of a second well calibrated station can improve the energy measurements and help quantify the regional variation in energy uncertainty. This can address the uncertainty that is introduced in the station calibrations for stations far from the first station.

Finally other network characterizations can be developed based on the clustering results. The number of thunderstorms, the average number of strokes per thunderstorm, average flash multiplicity, or time between flashes within a single thunderstorm can be routinely monitored for network health.

9.2   Thunderstorms

9.2.1   Cluster Validation

The thunderstorm clustering discussed in Chapter 7 and 8 can benefit from additional validation. Two potential methods for validation are large scale comparisons to TRMM precipitation rates and case studies against regional weather radars. The TRMM comparison would require grouping of the precipitation data to compare to the estimated thunderstorm extents. Weather radar comparisons would provide detailed ground truth of the thunderstorm cluster areas and durations; further, weather radar would allow for a measure of thunderstorm tracking accuracy.

Aside from comparisons to other systems the properties measured by the thunderstorm clustering can be tracked over spatial and temporal scales. For example: the average thunderstorm duration interannual variability in a given region or the overall network observed thunderstorm area. This can explore the changes in the clustering values and accuracy as the stroke rate detected by WWLLN changes with time.

9.2.2 Parameterization

Thunderstorm parameterizations are functional relationships between different properties of thunderstorms. Using lightning detection networks allows for the development of these relationships that infer thunderstorm properties based on the observed lightning flash rates; of particular benefit for regions with low availability of direct observations (e.g. radar, satellite). *Zipser* (1994) examined the relation between the flash rate in a thunderstorm with the updraft velocities, which in turn are controlled by the differential surface heating below the thunderstorm. Parameterization of flash rate based on other observations has led to simple models of global lightning behaviour; *Price and Rind* (1992) used cloud height parameterization for flash rate. These empirical parameterizations can help in the development and validation of models (e.g. *Baker et al.* (1999)) of thunderstorm electrification and enable or expand prediction of the global electric circuit, NO_x production, and weather forecasting. Investigating the parameterization of lightning behavior in relation to the properties of a thunderstorm on a large scale can create better and more robust parameterizations. The WWLLN thunderstorm clustering can be compared to other networks and systems in order to develop and validate new parameterizations.

9.2.3 Thunderstorm Properties

The WWLLN thunderstorm clustering can be used to investigate global and regional thunderstorm properties and their distributions. Thunderstorm area and duration are briefly discussed in Chapter 8, but can be explored in more depth and in more regions. These two properties can be measured over different spatial and temporal scales. For example the distribution of winter thunderstorms over the Sea of Japan (*Ishii et al.*, 2010), or the South American "green ocean" phenoma (*Williams and Stanfill*, 2002).

Alternatively more complex measures of thunderstorms can be developed: the average stroke count per thunderstorm, the peak flash rate, or the electrical activity. The electrical activity can be measured as the total flash energy per unit area per unit time for the thunderstorm: the sum of measured flash energies within the thunderstorm. This measure, and others, may be highly dependent on the network performance or may not be absolutely

calibrated, however they can still allow for relative comparisons between thunderstorms in a limited temporal or spatial scope (e.g. North America in July 2013).

The thunderstorm dataset can also be advanced by classifying thunderstorms into different thunderstorm types. While this may only work for small case studies it may be possible to use more advanced machine learning classifiers to automatically assign cloud types to thunderstorm clusters.

9.3 Global Electric circuit

The global electric circuit model presented in Chapter 7 provides an estimate of the thunderstorm contribution to the circuit based on the WWLLN thunderstorm clusters. It is shown that the model is able to reproduce the expected Carnegie curve over long time scales and has the capability to provide current estimates on shorter time scales. The model can be improved for a more accurate measure of the current contribution, but without any validation it cannot be reliably used.

To improve the model there are several underlying assumptions that can be factored in. The main one is that all thunderstorms produce the same upward current to the ionosphere regardless of their size or activity. The estimates used in the model from *Mach et al.* (2010) can be scaled by either the thunderstorm cluster activity or the estimated cluster size. Since the development of the model the thunderstorm cluster area was validated against ENTLN in Chapter 8, giving confidence to the area measurements. The second assumption is that WWLLN detects on the order of 90% of all thunderstorms, an assumption explored in Chapter 8. The third assumption is that all thunderstorm currents are the same everywhere. This assumption is harder to address as it relies on direct measurements of the current made in different regions.

Regardless of the model improvements made, the model needs to be compared against a ground truth measurement. Further, comparison to a ground truth will allow for tuning of the model to better fit the observed measurements; or if adjustments cannot be made then other components of the global electric circuit need to be included in the model. For the ground truth a balloon campaign needs to be conducted to measure the fair weather return current over the span of one to several weeks. With a reliable and clean measurements of

the return current the model can be validated on a very short time scale instead of relying on long term averages.

9.4 Other Lightning Processes

WWLLN is able to measure and detect more than just the main discharge of a lightning stroke. The network is able to directly detect terrestrial gamma-ray flashes (TGF) and their related strokes, it is also able to detect whistler waves at the individual stations. For TGFs WWLLN has been used to geolocate the TGF itself and the originating thunderstorm (*Connaughton et al.*, 2010, 2013). WWLLN can be used to create an estimate of the global TGF production rate by comparing the detected TGFs with the lightning within view of one of the TGF satellites (e.g. FERMI or RHESSI). With whistler waves an automated whistler detector running at the remote WWLLN stations could enable the creation of a large database of whistler waves. Such a database could provide the means for investigations of the ionosphere and magnetosphere, especially if the detector is able to run on multiple WWLLN stations. The advantage of a WWLLN based system compared to existing systems is the potential to use the existing distribution of stations and to develop a new system that can run on unspecialized hardware.

BIBLIOGRAPHY

Abarca, S. F., K. L. Corbosiero, and T. J. Galarneau, An evaluation of the Worldwide Lightning Location Network (WWLLN) using the National Lightning Detection Network (NLDN) as ground truth, *Journal of Geophysical Research*, *115*(D18), 1–11, doi:10.1029/2009JD013411, 2010.

Abreu, D., D. Chandan, R. H. Holzworth, and K. Strong, A performance assessment of the World Wide Lightning Location Network (WWLLN) via comparison with the Canadian Lightning Detection Network (CLDN), *Atmospheric Measurement Techniques*, *3*(4), 1143–1153, doi:10.5194/amt-3-1143-2010, 2010.

Altaratz, O., I. Koren, Y. Yair, and C. Price, Lightning response to smoke from Amazonian fires, *Geophysical Research Letters*, *37*(7), 1–6, doi:10.1029/2010GL042679, 2010.

Baker, M., a.M Blyth, H. Christian, J. Latham, K. Miller, and a.M Gadian, Relationships between lightning activity and various thundercloud parameters: satellite and modelling studies, *Atmospheric Research*, *51*(3-4), 221–236, doi:10.1016/S0169-8095(99)00009-5, 1999.

Beasley, W. H., S. Weiss, and S. Heckman, Comparison of WTLN and OKLMA data, in *Southern Thunder*, Southern Thunder, 2010.

Blakeslee, R. J., H. J. Christian, and B. Vonnegut, Electrical Measurements Over Thunderstorms, *Journal of Geophysical Research*, *94*(89), 13,135–13,140, 1989.

Boccippio, D. J., S. J. Goodman, and S. Heckman, Regional Differences in Tropical Lightning Distributions, *Journal of Applied Meteorology*, *39*(12), 2231–2248, doi:10.1175/1520-0450(2001)040⟨2231:RDITLD⟩2.0.CO;2, 2000.

Briggs, M. S., et al., Electron-positron beams from terrestrial lightning observed with Fermi GBM, *Geophysical Research Letters*, *38*(2), 1–5, doi:10.1029/2010GL046259, 2011.

Bürgesser, R. E., M. G. Nicora, and E. E. Ávila, Characterization of the lightning activity of Relámpago del Catatumbo, *Journal of Atmospheric and Solar-Terrestrial Physics*, pp. 1–7, doi:10.1016/j.jastp.2012.01.013, 2012.

Burkholder, B. S., M. L. Hutchins, M. P. McCarthy, R. F. Pfaff, and R. H. Holzworth, Attenuation of Lightning-Produced Sferics in the Earth-Ionosphere Waveguide and Low-Latitude Ionosphere, *Journal of Geophysical Research - Space Physics*, *118*, 3692–3699, doi:10.1002/jgra.50351, 2013.

Burns, G. B., A. Frank-Kamenetsky, O. A. Troshichev, E. A. Bering, and B. D. Reddell, Interannual consistency of bi-monthly differences in diurnal variations of the ground-level, vertical electric field, *Journal of Geophysical Research*, *110*(D10), 1–14, doi:10.1029/2004JD005469, 2005.

Burns, G. B., B. a. Tinsley, a. V. Frank-Kamenetsky, O. a. Troshichev, W. J. R. French, and a. R. Klekociuk, Monthly Diurnal Global Atmospheric Circuit Estimates Derived from Vostok Electric Field Measurements Adjusted for Local Meteorological and Solar Wind Influences, *Journal of the Atmospheric Sciences*, *69*(6), 2061–2082, doi:10.1175/JAS-D-11-0212.1, 2012.

Cecil, D., D. Buechler, and R. Blakeslee, TRMM-based Lightning Climatology, in *Proceedings of the 14th International Conference on Atmospheric Electricity - ICAE*, ICAE, Rio de Janeiro, 2011.

Christian, H., et al., Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *Journal of Geophysical Research*, *108*(D1), 4005, doi:10.1029/2002JD002347, 2003.

Christian, H. J., et al., The Lightning Imaging Sensor, in *11th International Conference on Atmospheric Electricity*, pp. 746–749, Guntersville, AL, 1999.

Clilverd, M. a., et al., Remote sensing space weather events: the AARDDVARK network, *Space Weather*, *7*(4), 1–15, doi:10.1029/2008SW000412, 2009.

Collier, A., S. Bremner, J. Lichtenberger, J. R. Downs, C. J. Rodger, P. Steinbach, and G. McDowell, Global lightning distribution and whistlers observed at Dunedin, New Zealand, *Annales Geophysicae*, *28*, 499–513, 2010.

Collier, a. B., B. Delport, a. R. W. Hughes, J. Lichtenberger, P. Steinbach, J. Öster, and C. J. Rodger, Correlation between global lightning and whistlers observed at Tihany, Hungary, *Journal of Geophysical Research*, *114*(A7), doi:10.1029/2008JA013863, 2009.

Collier, a. B., J. Lichtenberger, M. a. Clilverd, C. J. Rodger, and P. Steinbach, Source region for whistlers detected at Rothera, Antarctica, *Journal of Geophysical Research*, *116*(A3), 1–13, doi:10.1029/2010JA016197, 2011.

Connaughton, V., et al., Associations between Fermi Gamma-ray Burst Monitor terrestrial gamma ray flashes and sferics from the World Wide Lightning Location Network, *Journal of Geophysical Research*, *115*(A12), 1–14, doi:10.1029/2010JA015681, 2010.

Connaughton, V., et al., Radio signals from electron beams in terrestrial gamma ray flashes, *Journal of Geophysical Research: Space Physics*, *118*(5), 2313–2320, doi:10.1029/2012JA018288, 2013.

Crombie, D., Differences between the east-west and west-east propagation of VLF signals over long distances, *Journal of Atmospheric and Terrestrial Physics*, *12*(2-3), 110–117, 1958.

Darden, C. B., D. J. Nadler, B. C. Carcione, R. J. Blakeslee, G. T. Stano, and D. E. Buechler, Utilizing Total Lightning Information to Diagnose Convective Trends, *Bulletin of the American Meteorological Society*, *91*(2), 167–175, doi:10.1175/2009BAMS2808.1, 2010.

Doughton, S., Faraway volcanic eruptions now detected in a flash, *Seattle Times*, p. 1, 2010.

Dowden, R. L., and J. B. Brundell, Improvements relating to the location of lightning discharges, 2000.

Dowden, R. L., J. B. Brundell, and C. J. Rodger, VLF lightning location by time of group arrival (TOGA) at multiple sites, *Journal of Atmospheric and Solar-Terrestrial Physics*, *64*(7), 817–830, 2002.

Driscoll, K. T., R. J. Blakeslee, and M. E. Baginski, A modeling study of the time-averaged electric currents in the vicinity of isolated thunderstorms, *Journal of Geophysical Research*, *97*(D11), 11,535–11, 1992.

Dwyer, J. R., The relativistic feedback discharge model of terrestrial gamma ray flashes, *Journal of Geophysical Research*, *117*(A2), 1–25, doi:10.1029/2011JA017160, 2012.

Ester, M., H. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise., in *2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996.

Ferguson, J. A., Computer Programs for Assessment of Long- Wavelength Radio Communications, Version 2.0, *Tech. Rep. 3030*, Space and Naval Warfare Systems Center, San Diego, 1998.

Fierro, A. O., E. R. Mansell, C. L. Ziegler, and D. R. MacGorman, Application of a Lightning Data Assimilation Technique in the WRF-ARW Model at Cloud-Resolving Scales for the Tornado Outbreak of 24 May 2011, *Monthly Weather Review*, *140*(8), 2609–2627, doi:10.1175/MWR-D-11-00299.1, 2012.

Gjesteland, T., O. Nikolai, A. B. Collier, B. E. Carlson, M. B. Cohen, and N. G. Lehtinen, Confining the Angular Distribution of Terrestrial Gamma-ray Flash Emission., *Journal of Geophysical Research*, *116*, 1–8, doi:10.1029/2011JA016716, 2011.

Golde, R. H., Lightning currents and related parameters, in *Lightning*, edited by R. H. Golde, chap. 9, pp. 309–350, Academic Press, New York, 1977.

Hansen, J., M. Sato, and R. Ruedy, Global Temperature Update Through 2012, *NASA GISS*, p. website November 1, 2013.

Hays, P. B., and R. G. Roble, A Quasi-Static Model of Global Atmsopheric Electricity 1. The Lower Atmosphere, *Journal of Geophysical Research*, *84*(A7), 3291–3305, 1979.

Heckman, S., and C. Liu, The Application of Total Lightning Detection for Severe Storm Prediction, in *10th European Conference on Applications of Meteorology*, p. Abstract, Zürich, 2010.

Holzworth, R., High latitude stratospheric electrical measurements in fair and foul weather under various solar conditions, *Journal of Atmospheric and Terrestrial Physics*, *43*(11), 1115–1125, 1981.

Holzworth, R., and H. Volland, Do We Need a Geoelectric Index?, *Eos*, *67*(26), 545,547–548, 1986.

Holzworth, R., T. Onsager, P. Kintner, and S. Powerll, Planetary-scale variability of the fair-weather vertical electric field in the stratosphere, *Physical review . . .*, *53*(14), 1398 –1401, 1984.

Holzworth, R. H., M. P. McCarthy, R. F. Pfaff, a. R. Jacobson, W. L. Willcockson, and D. E. Rowland, Lightning-generated whistler waves observed by probes on the Communication/Navigation Outage Forecast System satellite at low latitudes, *Journal of Geophysical Research*, *116*(A6), 1–7, doi:10.1029/2010JA016198, 2011.

Holzworth, R. H., et al., Balloon observations of temporal variation in the global circuit compared to global lightning activity, *Advances in Space Research*, *36*, 2005.

Hutchins, M. L., R. H. Holzworth, J. B. Brundell, and C. J. Rodger, Relative detection efficiency of the World Wide Lightning Location Network, *Radio Science*, *47*(6), RS6005, doi:10.1029/2012RS005049, 2012a.

Hutchins, M. L., R. H. Holzworth, C. J. Rodger, and J. B. Brundell, Far-Field Power of Lightning Strokes as Measured by the World Wide Lightning Location Network, *Journal of Atmospheric and Oceanic Technology*, *29*(8), 1102–1110, doi:10.1175/JTECH-D-11-00174.1, 2012b.

Hutchins, M. L., R. H. Holzworth, K. S. Virts, J. M. Wallace, and S. Heckman, Radiated VLF energy differences of land and oceanic lightning, *Geophysical Research Letters*, *40*(March), 1–5, doi:10.1002/grl.50406, 2013a.

Hutchins, M. L., A. R. Jacobson, R. H. Holzworth, and J. B. Brundell, Azimuthal Dependence of VLF Propagation, *Journal of Geophysical Research: Space Physics*, *118*, 5808–5812, doi:10.1002/jgra.50533, 2013b.

Hutchins, M. L., R. H. Holzworth, and J. B. Brundell, Diurnal variation of the global electric circuit from clustered thunderstorms, *Journal of Geophysical Research - Space Physics*, (119), 620–629, doi:10.1002/2013JA019593.Received, 2014.

Ishii, M., M. Saito, F. Fujii, M. Matsui, and N. Itamoto, High-Current Lightning Discharges in Winter, *Electrical Engineering in Japan*, *170*(1), 291–297, doi:10.1002/eej.20874, 2010.

Jacobson, A. R., R. Holzworth, J. Harlin, R. Dowden, and E. Lay, Performance Assessment of the World Wide Lightning Location Network (WWLLN), Using the Los Alamos Sferic Array (LASA) as Ground Truth, *Journal of Atmospheric and Oceanic Technology*, *23*(8), 1082 – 1092, doi:10.1175/JTECH1902.1, 2006.

Jacobson, A. R., X.-M. Shao, and R. Holzworth, Full-wave reflection of lightning long-wave radio pulses from the ionospheric D region: Comparison with midday observations of broadband lightning signals, *Journal of Geophysical Research*, *115*, 1–12, doi:10.1029/2009JA014540, 2010.

Jacobson, a. R., R. H. Holzworth, R. F. Pfaff, and M. P. McCarthy, Study of oblique whistlers in the low-latitude ionosphere, jointly with the C/NOFS satellite and the World-Wide Lightning Location Network, *Annales Geophysicae*, *29*(5), 851–863, doi:10.5194/angeo-29-851-2011, 2011.

Jacobson, A. R., X.-M. Shao, and E. Lay, Time domain waveform and azimuth variation of ionospherically reflected VLF/LF radio emissions from lightning, *Radio Science*, *47*(4), 1–17, doi:10.1029/2012RS004980, 2012.

Jacobson, E., and E. Krider, Electrostatic field changes produced by Florida lightning, *Journal of the Atmospheric Sciences*, *33*, 103 – 117, 1976.

Kartalev, M., M. Rycroft, and V. Papitashvili, A quantitative model of the effect of global thunderstorms on the global distribution of ionospheric electrostatic potential, *Journal of Atmospheric and Solar-Terrestrial Physics*, *66*(13-14), 1233–1240, doi:10.1016/j.jastp. 2004.05.012, 2004.

Kasemir, H. W., Theoretical Problems of the Global Atmospheric Electric Circuit, in *Electrical Processes in Atmospheres*, edited by H. Dolezalek and R. Reiter, pp. 423–439, Darmstadt: Steinkopff, 1977.

Kawanishi, T., et al., TRMM precipitation radar, *Advances in Space Research*, *25*(5), 969–972, 2000.

Krider, E. P. I., and R. J. Blakeslee, The Electric Currents Produced by Thunderclouds, *Journal of Electrostatics*, *16*, 369–378, 1985.

Kriegel, H.-P., P. Kröger, J. Sander, and A. Zimek, Density-based clustering, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(3), 231–240, doi: 10.1002/widm.30, 2011.

Kumar, S., A. Deo, and V. Ramachandran, Nighttime D-region equivalent electron density determined from tweek sferics observed in the South Pacific Region, *Earth Planets Space*, *3*(2), 905–911, 2009.

Lay, E. H., R. H. Holzworth, C. J. Rodger, J. N. Thomas, O. P. Jr., and R. L. Dowden, WWLL global lightning detection system: Regional validation study in Brazil, *Geophysical Research Letters*, *31*, 2004.

Lay, E. H., A. R. Jacobson, R. H. Holzworth, C. J. Rodger, and R. L. Dowden, Local time variation in land/ocean lightning flash density as measured by the World Wide Lightning Location Network, *Journal of Geophysical Research*, *112*(D13), 1–9, doi:10. 1029/2006JD007944, 2007.

Liu, C., and S. Heckman, The Application of Total Lightning Detection and Cell Tracking for Severe Weather Prediction, in *91st American Meteorological Society Annual Meeting*, pp. 1–10, Seattle, 2011.

Liu, C., E. R. Williams, E. J. Zipser, and G. Burns, Diurnal Variations of Global Thunderstorms and Electrified Shower Clouds and Their Contribution to the Global Electrical Circuit, *Journal of the Atmospheric Sciences*, *67*(2), 309, doi:10.1175/2009JAS3248.1, 2010.

Lyons, W., M. Uliasz, and T. Nelson, Large peak current cloud-to-ground lightning flashes during the summer months in the contiguous United States, *Monthly weather review*, *126*, 2217–2233, 1998.

Mach, D. M., R. J. Blakeslee, M. G. Bateman, and J. C. Bailey, Electric fields, conductivity, and estimated currents from aircraft overflights of electrified clouds, *Journal of Geophysical Research*, *114*(D10), 1–15, doi:10.1029/2008JD011495, 2009.

Mach, D. M., R. J. Blakeslee, M. G. Bateman, and J. C. Bailey, Comparisons of total currents based on storm location, polarity, and flash rates derived from high-altitude aircraft overflights, *Journal of Geophysical Research*, *115*(D3), D03,201, doi:10.1029/2009JD012240, 2010.

Mach, D. M., R. J. Blakeslee, and M. G. Bateman, Global electric circuit implications of combined aircraft storm electric current measurements and satellite-based diurnal lightning statistics, *Journal of Geophysical Research*, *116*(D5), 1–13, doi:10.1029/2010JD014462, 2011.

Mallios, S. a., and V. P. Pasko, Charge transfer to the ionosphere and to the ground during thunderstorms, *Journal of Geophysical Research*, *117*(A8), 1–16, doi:10.1029/2011JA017061, 2012.

Mareev, E. a., S. a. Yashunin, S. S. Davydenko, T. C. Marshall, M. Stolzenburg, and C. R. Maggio, On the role of transient currents in the global electric circuit, *Geophysical Research Letters*, *35*(15), 1–5, doi:10.1029/2008GL034554, 2008.

Markson, R., Solar modulation of atmospheric electrification and possible implications for the Sun-weather relationship, *Nature*, *273*, 103–109, 1978.

McRae, W., and N. Thomson, VLF phase and amplitude: daytime ionospheric parameters, *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(7), 609–618, 2000.

Metzger, E., and W. a. Nuss, The Relationship between Total Cloud Lightning Behavior and Radar-Derived Thunderstorm Structure, *Weather and Forecasting*, *28*(1), 237–253, doi:10.1175/WAF-D-11-00157.1, 2013.

Mezuman, K., Detecting global thunderstorm distribution from lightning clusters, Msc thesis, Tel Aviv Univeristy, 2013.

Nag, A., et al., Evaluation of U.S. National Lightning Detection Network performance characteristics using rocket-triggered lightning data acquired in 20042009, *Journal of Geophysical Research*, *116*(D2), 1–8, doi:10.1029/2010JD014929, 2011.

Ogawa, T., Fair-weather electricity, *Journal of Geophysical Research*, *90*(10), 5951–5960, 1985.

Ogawa, T., Lightning Currents, in *Handbook of Atmospheric Electrodynamics*, edited by H. Volland, chap. 4, pp. 93 – 136, CRC Press, 1995.

Pan, L., X. Qie, D. Liu, D. Wang, and J. Yang, The lightning activities in super typhoons over the Northwest Pacific, *Science China Earth Sciences*, *53*(8), 1241–1248, doi:10.1007/s11430-010-3034-z, 2010.

Peckham, D. W., M. a. Uman, and C. E. Wilcox, Lightning phenomenology in the Tampa Bay area, *Journal of Geophysical Research*, *89*(D7), 11,789, doi:10.1029/JD089iD07p11789, 1984.

Popolansky, F., Frequency Distribution of amplitudes of lightning currents, *Electra*, *22*, 139–146, 1972.

Price, C., Will a drier climate result in more lightning?, *Atmospheric Research*, *91*(2-4), 479–484, doi:10.1016/j.atmosres.2008.05.016, 2009.

Price, C., and D. Rind, A simple lightning parameterization for calculating global lightning distributions, *Journal of geophysical research*, *97*(D9), 9919–9933, 1992.

Price, C., M. Asfur, and Y. Yair, Maximum hurricane intensity preceded by increase in lightning frequency, *Nature Geoscience*, *2*(April), 2–5, doi:10.1038/NGEO477, 2009.

Rakov, V. A., and M. A. Uman, *Lightning: Physics and Effects*, Cambridge University Press, Cambridge, 2003.

Roble, R., On modeling component processes in the Earth's global electric circuit, *Journal of Atmospheric and Terrestrial Physics*, *53*(9), 831–847, doi:10.1016/0021-9169(91)90097-Q, 1991.

Rodger, C. J., S. Werner, J. B. Brundell, E. H. Lay, N. R. Thomson, R. H. Holzworth, and R. L. Dowden, Detection efficiency of the VLF World-Wide Lightning Location Network (WWLLN): initial case study, *Annales geophysicae*, *24*, 3197–3214, 2006.

Rodger, C. J., J. B. Brundell, R. H. Holzworth, E. H. Lay, N. B. Crosby, T.-Y. Huang, and M. J. Rycroft, Growing Detection Efficiency of the World Wide Lightning Location Network, *AIP Conference Proceedings*, *1118*, 15–20, doi:10.1063/1.3137706, 2009.

Rudlosky, S. D., and H. E. Fuelberg, Pre- and Postupgrade Distributions of NLDN Reported Cloud-to-Ground Lightning Characteristics in the Contiguous United States, *Monthly Weather Review*, *138*(9), 3623–3633, doi:10.1175/2010MWR3283.1, 2010.

Rudlosky, S. D. S., and D. T. D. Shea, Evaluating WWLLN performance relative to TRMM/LIS, *Geophysical Research Letters*, *40*(March), 1–5, doi:10.1002/grl.50428, 2013.

Ruhnke, L., Area Averaging of Atmospheric Electric Currents, *Tech. Rep. 1*, Atmospheric Physics and Chemistry Lab., Boulder, Colo., 1969.

Rycroft, M. J., Electrical processes coupling the atmosphere and ionosphere: An overview, *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(3-5), 445–456, doi:10.1016/j.jastp.2005.04.009, 2006.

Rycroft, M. J., and R. G. Harrison, Electromagnetic Atmosphere-Plasma Coupling: The Global Atmospheric Electric Circuit, *Space Science Reviews*, *168*(1-4), 363–384, doi:10.1007/s11214-011-9830-8, 2011.

Rycroft, M. J., S. Isrealsson, and C. Price, The global atmospheric electric circuit, solar activity and climate change, *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(17-18), 1563–1576, doi:10.1016/S1364-6826(00)00112-7, 2000.

Saunders, C. P. R., Thunderstorm Electrification, in *Handbook of Atmospheric Electrodynamics*, vol. 1, edited by H. Volland, chap. 3, pp. 61 – 92, CRC Press, 1995.

Schultz, C. J., W. a. Petersen, and L. D. Carey, Preliminary Development and Evaluation of Lightning Jump Algorithms for the Real-Time Detection of Severe Weather, *Journal of Applied Meteorology and Climatology*, *48*(12), 2543–2563, doi:10.1175/2009JAMC2237.1, 2009.

Schultz, C. J., W. a. Petersen, and L. D. Carey, Lightning and Severe Weather: A Comparison between Total and Cloud-to-Ground Lightning Trends, *Weather and Forecasting*, *26*(5), 744–755, doi:10.1175/WAF-D-10-05026.1, 2011.

Seity, Y., S. Soula, and H. Sauvageot, Lightning and precipitation relationship in coastal thunderstorms, *Journal of Geophysical Research*, *106*(D19), 22,801 – 22,816, 2001.

Singh, a. K., D. Siingh, R. P. Singh, and S. Mishra, Electrodynamical Coupling of Earth's Atmosphere and Ionosphere: An Overview, *International Journal of Geophysics*, *2011*, 1–13, doi:10.1155/2011/971302, 2011a.

Singh, R., B. Veenadhari, A. K. Maurya, M. B. Cohen, S. Kumar, R. Selvakumaran, P. Pant, A. K. Singh, and U. S. Inan, D-region ionosphere response to the total solar eclipse of 22 July 2009 deduced from ELF-VLF tweek observations in the Indian sector, *Journal of Geophysical Research*, *116*(A10), 1–9, doi:10.1029/2011JA016641, 2011b.

Soula, S., O. van der Velde, J. Montanya, P. Huet, C. Barthe, and J. Bór, Gigantic jets produced by an isolated tropical thunderstorm near Réunion Island, *Journal of Geophysical Research*, *116*(D19), doi:10.1029/2010JD015581, 2011.

134

Stergis, C. G., G. C. Rein, and T. Kangas, Electric field measurements above thunderstorms, *Journal of Atmospheric and Terrestrial Physics*, *11*, 83–90, 1957.

Stolzenburg, M., T. C. Marshall, and P. R. Krehbiel, Duration and extent of large electric fields in a thunderstorm anvil cloud after the last lightning, *Journal of Geophysical Research*, *115*(D19), D19,202, doi:10.1029/2010JD014057, 2010.

Suszcynsky, D. M., Narrow Bipolar Events as indicators of thunderstorm convective strength, *Geophysical Research Letters*, *30*(17), 1879, doi:10.1029/2003GL017834, 2003.

Taylor, W., Radiation Field Characteristics of Lightning Discharges In the Band 1 kc/s to 100 kc/s, *Journal of Research National Beruau of Standards, Section D Radio Propagation*, *67D*(5), 539 – 550, 1963.

Taylor, W. L., VLF attenuation for east-west and west-east daytime propagation using atmospherics, *Journal of Geophysical Research*, *65*(7), 1933–1938, 1960.

Thomas, J. N., N. N. Solorzano, S. A. Cummer, and R. H. Holzworth, Polarity and energetics of inner core lightning in three intense North Atlantic hurricanes, *Journal of Geophysical Research*, *115*(A00E15), 1–11, doi:10.1029/2009JA014777, 2010.

Thomson, N. R., Daytime tropical D region parameters from short path VLF phase and amplitude, *Journal of Geophysical Research*, *115*(A9), 1–11, doi:10.1029/2010JA015355, 2010.

Thomson, N. R., M. A. Clilverd, and C. J. Rodger, Daytime midlatitude D region parameters at solar minimum from short-path VLF phase and amplitude, *Journal of Geophysical Research*, *116*(A3), 1–12, doi:10.1029/2010JA016248, 2011.

Tinsley, B., G. Burns, and L. Zhou, The role of the global electric circuit in solar and internal forcing of clouds and climate, *Advances in Space Research*, *40*(7), 1126–1139, doi:10.1016/j.asr.2007.01.071, 2007.

Tzur, I., and R. Roble, The interaction of a dipolar thunderstorm with its global electrical environment, *Journal of Geophysical Research*, *90*(D4), 5989–5999, 1985.

Uman, M. A., *Lightning*, Dover books on physics and chemistry, Dover Publications, 1969.

Virts, K. S., J. a. Thornton, J. M. Wallace, M. L. Hutchins, R. H. Holzworth, and A. R. Jacobson, Daily and intraseasonal relationships between lightning and NO 2 over the Maritime Continent, *Geophysical Research Letters*, *38*(19), 2–5, doi:10.1029/2011GL048578, 2011.

Virts, K. S., J. M. Wallace, M. L. Hutchins, and R. H. Holzworth, Highlights of a new ground-based, hourly global lightning climatology, *Bulletin of the American Meteorological Society*, *94*, 1381 – 1391, doi:10.1175/BAMS-D-12-00082.1, 2013.

Volland, H., Longwave Sferic Propagation within the Atmospheric Waveguide, in *Handbook of Atmospherics Volume II*, edited by H. Volland, chap. 3, pp. 65–93, CRC Press, Boca Raton, 1995.

Wait, J., and K. Spies, Influence of earth curvature and the terrestrial magnetic field on VLF propagation, *Journal of Geophysical Research*, *65*(8), 2325–2331, 1960.

Wait, J. R., *Electromagnetic Waves in Stratified Media*, 2nd ed., Pergamon Press, New York, 1970.

Ward, J. H. J., Hierarchical grouping to optimize an objective function, *Journal of the American statistical association*, *58*(301), 236 – 244, 1963.

Whipple, F., On the association of the diurnal variation of electric potential gradient in fine weather with the distribution of thunderstorms over the globe, *Quarterly Journal of the Royal Meteorological Society*, *55*(229), 1–18, 1929.

Williams, E., and S. Stanfill, The physical origin of the land ocean contrast in lightning activity, *Comptes Rendus Physique*, *3*(10), 1277–1292, 2002.

Williams, E., V. Mushtak, D. Rosenfeld, S. Goodman, and D. Boccippio, Thermodynamic conditions favorable to superlative thunderstorm updraft, mixed phase microphysics and lightning flash rate, *Atmospheric Research*, *76*(1-4), 288–306, doi:10.1016/j.atmosres.2004.11.009, 2005.

Williams, E. R., Large-Scale Charge Separation in Thunderclouds, *Journal of Geophysical Research*, *90*(D4), 6013–6025, 1985.

Williams, E. R., Lightning and climate: A review, *Atmospheric Research*, *76*, 272–287, 2005.

Williams, E. R., M. E. Weber, and R. E. Orville, The relationship between lightning type and convective state of thunderclouds, *Journal of Geophysical Research*, *94*(Dll), 213–220, 1989.

Wilson, C. T. R., Investigation on Lightning Discharges and on the Electric Field of Thunderstorms, *Philosophical Transactions of the Royal Society of London*, *221*(1921), 73–115, 1921.

Zipser, E., Deep cumulonimbus cloud systems in the tropics with and without lightning, *Monthly weather review*, *1222*, 1837–1851, 1994.

Zoghzoghy, F. G., M. B. Cohen, R. K. Said, and U. S. Inan, Statistical patterns in the location of natural lightning, *Journal of Geophysical Research: Atmospheres*, *118*(2), 787–796, doi:10.1002/jgrd.50107, 2013.

Appendix A

CODE REPOSITORIES

This appendix discusses the distribution of WWLLN data files on the Flash series of computers along with the storage locations of routine processing code. The code locations are given in terms of both their active locations and as the source control repositories, managed using git (http://git-scm.com).

A.1 File Storage on Flash Machines

There are several types of WWLLN data files, summarized in Table A.1. The example filenames given in the table are for 31 December 2013, 5:10 UTC. Most of the file types, with exception of WB-files, are located on `flashfile` on `/wd1/wwlln/`. Each file type has it's own directory, denoted Afiles, AEfiles, etc., with flat file structures (except for R-files). R-files, due to the large number generated, are broken into yearly and monthly folders of the form: `Rfiles/R[YYYY]/R[YYYY][MM]`. Some, but not all, of the WWLLN data file directories have been copied to the other flashes on a need-basis and so the other flashes should not be considered a backup of the flashfile dataset.

An example line from each of the R, A, AP, and AE files are given below with descriptors for each file. R-files are the simplest with each line containing three columns:

```
10      3000.015563     2668
```

The first number is the station ID, the second is the seconds from the start of the hour, and the last is the root integrated square electric field at the station, in uncalibrated sound card units. For A-files each line if of the form:

```
2013/01/31,16:18:54.103460,  9.0621,-117.8276, 19.9,  9
```

The first six numbers are the date (YYYY/MM/DD) and UTC time of the stroke (hh,mm,ss). The next two are the stroke location in degrees north latitude and east longitude. The final two are the timing uncertainty in microseconds and the number of WWLLN stations that participated in the stroke location. AP-files are similar to A-files with additional information after the stroke location:

```
2013/02/12,00:00:07.142335,-10.4622,  20.0863, 13.4,  5, 17, 600, 19, 2084,
 26, 453, 53, 225, 64, 2263
```

Table A.1: WWLLN Data Types

File Type	Name	Description
R-files	R20121231051000	Raw packets sent by stations
A-files	A20121231.loc	Lightning locations generated in real time
AP-files	AP20121231.loc	Location data with station and E-field data of each stroke solution, generated daily with relocate-B
AE-files	AE20121231.loc	Location and stroke energy data, generated daily by Bootstrap.
DE-files	DE20121231.mat	Relative detection efficiency maps, as Matlab data.
S-files	S201212310510.loc	Waveform data for strokes located near a given station.
WB-files	WB201212310510.dat	Continuos wideband field data output by toga -r
T-files	T20131114_Single.dat	Station packet count record, TCurrent.dat is the current running record

Starting at 17,600 are pairs of numbers that give the station ID (17) and the field strength (600) in sound card units, the same value as in the R-files. Finally AE-files are similar to A-files but with three additional numbers:

2013/2/12,00:00:07.142335, -10.4622, 020.0863, 13.4, 5, 448.86, 114.87, 4

The last three numbers correspond to the radiated VLF stroke energy in joules (448.86), the median absolute deviation of the energy value in joules (114.87), and the subset of stations that participated in the energy value solution (4).

T-files are records of how many packets a given station sent to flash4 during the history of WWLLN. Unlike the other files it is not automatically generated, rather it is written whenever it is updated through the TfileUpdater.m script in the functions.git repository. Each line contains the date in days since 01/01/0000 and the counts for each station,

where column 2 corresponds to station 0.

A.2 Code Repositories

All of my code should be stored on `flashfile` at `/home/mlhutch/`. The operational code running every day is stored on `flash5` at `/home/mlhutch/matlab/bootstrap/` with a non-operational copy on `flashfile`. The exact details of where the energy processing code is located, the necessary files, and similar code for the relative detection efficiency processing is covered in Appendix B. Aside from being stored in plaintext scripts, all of the code is also stored in a Git repository located on `flashfile:/home/mlhutch/Git` and hosted online on `github.com/mlhutchins/`.

The Git repositories have the advantage of multiple back ups, version control, easily portable, and easy to use. A list of the WWLLN processing code and corresponding git repository is given in Table A.2. A basic primer on how to use git is given in Section A.3; online resources are also readily available.

Table A.2: Git Repositories (`flashfile:/home/mlhutch/Git/`)

Code	Repository	Notes
Energy Processing	`bootstrap.git`	Requires `process.git` and `functions.git`
LWPC	`lwpc.git`	LWPC and Matlab implementation
AP Processing	`process.git`	Relocate files (required for energy processing)
MATLAB Functions	`functions.git`	Various Matlab functions required by other scripts.
SU Eagle Files	`eagle.git`	Eagle files for WWLLN SU and Pre-amps, includes Erin Lay's files.
Gumstix	`gumstix.git`	Files for building and configuring Gumstix microSD cards.

A.3 Git Primer

Git (`http://git-scm.com`) is an open source distributed version control system available on all operating systems with very little system resources required. Most often git is accessed through the command line but there are many graphical interfaces available. I chose git for my research and code since it allows easy branching, constant backups (locally and on remote servers), and helps to organize code and projects. A brief tutorial on git is available online: `http://git-scm.com/book/en/Getting-Started` and `http://try.github.com/`.

Once git is installed on a host machine (`http://git-scm.com`, `apt-get install git`, or `yum install git-core`) repositories can be downloaded with the command:

```
git clone [user]@[server]:[repository] .
```

This will copy (clone) the repository to the current directory as a folder of the same name as the repository. For example to copy over the energy processing repository listed in Table A.2 to the `matlab` directory:

```
cd ~/matlab
clone sferix@flash5.ess.washington.edu:/home/mlhutch/Git/bootstrap.git .
```

Where the user can be anyone with permission to access that directory. More information on how to use git for source control, instead of downloading the repositories, can be found in the online resources.

A.4 Other Useful Matlab Code

Some of the functions present in the functions.git repository are very useful and often critical to the routine WWLLN processing. The function terminator.m is used to calculating the percentage of a VLF path that is in daylight and the percentage that is nighttime conditions. It is written to take into a either a vector or scaler initial location, time, and end location. The function `vdist.m` is from the MathWorks file exchange and calculated the great circle distance between two points on the ellipsoidal Earth.

Aside from these three functions there is a directory in the `functions.git` repository dedicated to reading and writing WWLLN file types. In the directory `functions/IO` are a set of import functions to import A, AP, AE, R, T, and DE files. There are also functions to read and write tab delimited data, the LWPC lookup tables (described in Appendix B), and WWLLN `stations.dat` data. All of these functions take in a given date (and possibly time) to find and load the desired file. The file `dataPath.dat` lists locations that the import functions check for the data folders of the particular file type. The default is to look for either `flashfile:/wd1/wwlln/` or `/wd2/` drives, but other locations can be added to this list as needed.

Appendix B

ENERGY PROCESSING

The process for bootstrap calibrating the WWLLN stations and how the energy is calculated is described in *Hutchins et al.* (2012b). This section walks through the code used in the processing along with the decisions behind various checks and adjustments made in the processing.

B.1 Code Summary

There are 8 main sections to the bootstrap processing:

- Process R-files into AP-files

- Calculate LWPC attenuation coefficients for each stroke-station pair

- Bootstrap calibrate the network

- Calculate stroke energy using LWPC and calibration

- Iterative re-calibrate the network

- Final energy calculation

- Run Relative Detection Efficiency Model

- Move files to storage locations

This processing requires several files, locations given relative to ml-hutch@flash5.ess.washington.edu.

- **~/Process/relocate-B31Jan2013.x86-64** – James Brundell's relocate program used to process R-files into AP-files, this version compiled for Linux 64-bit systems

- **stations.dat** – list of current WWLLN stations, copied over from flash4 at the start every processing run.

- **lwpcv21** – directory containing a working compiled version of the LWPC code. Used in generating new lookup tables as stations are added, a parallelized matlab implementation is discussed below. Provided as a sub-module of the Bootstrap directory.

- ~/**matlab/Bootstrap/** – directory containing the necessary matlab files, available as a git directory (see Appendix A).

- ~/**matlab/functions/** – directory of necessary matlab functions used by the Bootstrap code files, also available as a git directory (see Appendix A).

B.2   LWPC

The Long Wave Propagation Capability code is a codebase developed by *Ferguson* (1998) that is used to calculated the electric field at a given location for a VLF transmitter at another location. For the WWLLN energy processing it is used to estimate the attenuation between a stroke (treated as a transmitter) and a station. The original LWPC code has been altered in two ways: first the Windows-compiler specific code has been replaced with GCC compilable code, second all warning have been removed to produce output of a constant shape (for reading into MATLAB). All edits in the source code have been marked by my initials of MH.

B.2.1   Compiling

To recompile LWPC two bash scripts need to be run. The first is **BuildData.cmd**, BuildData recompiles the data files that contain the surface parameters such as ground conductivity and coastlines. The second step is to run **buildlwpc.cmd**, this should compile the program and result in an executable called LWPC. To test a successful compile run the script **run_bench.cmd**, if it runs successfully it is compiled, if not consult the **Readme - Unix.txt** file that has some common troubleshooting steps.

The last step in setting up LWPC is to set the **lwpcDAT.loc** file to point towards the **data** folder.

B.2.2 Running

LWPC is run from the command line with the structure:

```
./LWPC test1.inp
```

Where **test1.inp** is formatted as per the structure in **User_manual.pdf**. The matlab implemtations discussed below automatically generate formatted input files.

The result of running LWPC is an output file the specifies the electric field at various distance along the path from the transmitter to receiver. It is also capable of outputting plots, azimuthal dependence, and many other features not utilized in this research.

B.2.3 Matlab Function

A matlab implementation of LWPC is available as a git repository (see Appendix A). This implantation allows for LWPC to called in matlab with:

```
LWPCpar(freq,lat,long,time,stat_lat,stat_long,model)
```

Where freq is the transmitter frequency, lat/long is the transmitter location, time is the date and time, stat_lat/stat_long are the receiver locations, and model is the ionospheric model used. If model is set to "time" then day and night is considered in the calculation, if it is set to "day" or "night" an all day or all night ionosphere will be used.

LWPCpar as the advantage of being runnable within matlab parallelized loops (parfor). While LWPC itself cannot be run in parallel, this method simply copies the LWPC directory to allow each matlab instance it's own copy.

B.3   Lookup Tables

While LWPC can be parallelized it is still too slow to run faster than realtime for WWLLN processing. As a result two sets of lookup tables are generated for each station with the pre-calculated LWPC electric field values.

The lookup tables are part of the main Bootstrap directory under the names **lookup_day.dat** and **lookup_night.dat**. Each of these files are tab-delimited text with

the format given in Table B.1. The tables list the electric field measured the the station given a 100 kW transmitter in each grid point on the globe. The latest lookup tables are at a resolution of 2°.

Station Name	Station North Latitude	Station East Longitude	
E-field at (-179,89)	E-field at (-178,89)	E-field at (-177,89)	...
E-field at (-179,88)	E-field at (-178,88)	E-field at (-177,88)	...
E-field at (-179,87)	E-field at (-178,87)	E-field at (-177,87)	...
...	...	...	...

Table B.1: Format for **lookup_day.dat** and **lookup_night.dat** data files for a resolution of 1°.

An example of a lookup table for the all day ionosphere for Dunedin station is shown in Figure B.1.
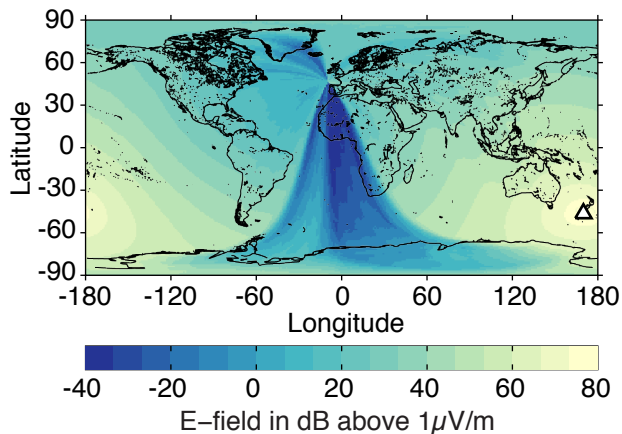
Figure B.1: Lookup table for an all day ionosphere for Dunedin station.

B.3.1 Generation

The LWPC lookup tables are generated automatically with the matlab script **∼/Bootstrap/**
Lookup_automation.m that is called at the start of the daily energy processing. The script checks the current number of stations in the **.dat** files and compares it to the most recent **stations.dat** file. If there are a new entries in the **stations.dat** then the script proceeds to make the lookup tables.

The automatic script calculates the electric field at every grid point for 11 frequencies ranging from 8 kHz to 18 kHz, and averages them together to add to the **.dat** files.

The same script can be run for specific stations with the **∼/matlab/lwpcpar/lwpc_generate.m** script.

Tables generated with either script can be validated with the ∼/**matlab/Bootstrap/**
lookup_validation.m scipt. This script checks each grid point to ensure there is a real number there, sometimes LWPC generates errors for very specific locations. If there is no valid electric field value at that spot it is reprocessed by offsetting the transmitter by a fraction of a degree.

B.4 Matlab Code

The main matlab script to run the entire energy calculation process is **Boostrap_automation.m**. This script:

1. Defines the dates to run (in case past days need to be rerun)

2. Sets the system path parameters from **data_path.m**

3. Updates stations.dat and the lookup tables **Lookup_automation.m**.

4. Generate the current AP-file with **generate_ap.m**

5. Calculate the energies and station calibrations using **wwlln_energy.m**

6. Generates the relative detection efficiency map with **de_mapper.m**

7. Calculates network statistics with **network_statistics.m**

8. Generate and amend T-files to TCurrent.dat

9. Saves and archives all of the resulting data files

In order to rerun lost days the only change necessary for Bootstrap_automation is to change the rundate from:

```
RUNDATE = floor(now - 1);
```

to the day (or dates) that need to be rerun. The LIVE variable and second part of the if clause is for rerunning *previously processed* data. If LIVE is set to false then the script runs and stores in an alternative directory with a copy of the resulting calibration file. For the case of restarting the script on the same day it failed to run, no changes need to be made.

B.4.1   data_path.m

It is important to correctly set up the **data_path.m** script. Aside from the storage paths, it is important to set how the script should obtain the stations.dat and R-files from flash4. The transfer type can either be cp if the computer has flash4 network mounted, or scp if passwordless scp is setup. The r_path_flat variable determines whether the R-files are stored in one directory or with a hierarchal structure (e.g. /wd2/Rfiles/R2013/R201304/R201304*).

Additionally the location of the bootstrap processing folder itself as well as the TOGA relocate script.

B.4.2   Lookup_automation.m

The **Lookup_automation.m** script compares the current number of entries in the lookup_day.dat data file and compares it to the latest stations.dat file. If stations.dat has new entries, then the script generates the corresponding LWPC lookup tables for those new stations.

Using the current lookup spatial resolution LWPCpar is called for each grid point at 11 frequencies (8 – 11 kHz) for both day and night. The 11 values in each grid point (in dB) are averaged together and added to the existing lookup data files.

This code takes about twelve hours to run using a single core due to calling the LWPC code. The LWPCpar.m function is set up to run in parallel (it copies the LWPC code to new locations), however the **Lookup_automation.m** script has not be parallelized.

B.4.3   generate_ap.m

generate_ap.m copies over the R-files to the TOGA relocate folder and processes them with the -e flag. This generate the normal WWLLN A-file structure, with the addition of the participating stations and their RMS electric field values at the end of each line. The RMS electric field (as mentioned elsewhere) is in the uncalibrated sound card units, where the calibration is the main part of the wwlln_energy.m code. The script ends with moving the new AP file and R-files to their storage locations.

B.4.4   wwlln_energy.m

The **wwlln_energy.m** function runs the subfunctions described below in order to calculate the stroke energies. Unless it is set otherwise (which it should be), the script sets the default master station to use to be Scott Base with the most recent calibration. The master station is one with a known calibration that is used to bootstrap the calibrations for the rest of the network. One future upgrade to the code would be the ability to include multiple master stations and combine their results for a better system-wide calibration.

The script also generates and saves two sets of station ratios. The first is the station conversion file, the values in this file give how to convert the sound card units at the master station to the sound card units at the other stations. It is saved in the case that the master calibration requires a retroactive change. The calibration file gives the values to convert from individual station sound card units to RMS electric-field values. The station calibration incorporates both the actual station calibration and any environmental changes near the station.

apply_lwpc.m applies the LWPC lookup tables to each sound card value in each stroke-station pair. The day and night lookup table values are weighted by the percent of the path in daytime and nighttime. There is also a distance range restriction, where only stroke-station values between 1 and 8 Mm are calculated, otherwise they are given a value of zero.

cross_calibration.m finds all strokes common to each pair of stations, and uses them to create a conversion ratio between every pair of stations. Only if the two stations have common strokes in all day paths and are within the same $1 - 8$ Mm distance restriction. The median value of every common stroke value is used as the resulting station-pair conversion ratios.

bootstrap.m solves the connected graph of station-pair conversions to find the valid calibration paths from the master station to every other possible station. There are two ways for a station to be added as a well calibrated station. First it can have common strokes between itself and the master station, this would result in a single hop conversion. The second way is to be calibrated off of one of the previously calibrated stations. In this case a secondary (or tertiary etc.) calibration is validated by comparing it's direct conversion with a conversion using it as an intermediary. So to test if the calibration of A to B is valid, B is used in the A to B to C calibration. If the ABC calibration matches the direct AC calibration to within 75% then B is considered well calibrated. This bootstrapping is conducted for 5 hops, so any converted station is at most 4 intermediary stations away from the master station.

conversion_check.m takes the resulting calibration paths from bootstrap.m and performs the bootstrap calibration using the initial calibration of the master station and the conversion ratios of cross_calibration.m. It also requires that the master station is included in the case where it becomes excluded in the bootstrap.m subfunction.

To prevent sudden changes in gain or wildly varying station from being included **conversion_mean.m** checks the new calibration values against the previous 7 days of calibrations. If the new calibration is not drastically different then the average of the previous 7 days of calibration values is used as the calibration value, otherwise it is excluded. A station is deemed unstable if one of the past seven days is either larger or smaller than the median

by a factor of 10 (normalized by Scott Base, Dunedin, and Seattle).

stroke_energy.m applies the calibration values from conversion_mean.m to the LWPC corrected data from apply_lwpc.m to get the stroke energy. The stroke energy is the median of the individual station values and the uncertainty is the median absolute deviation of the values. The values are stored as the 11th and 12th column in the resulting AE-file. The 13th value is the number of stations that participated in the calculation.

The final step is to iterate the entire process 5 times with **energy_iterate.m**. Here the stroke energies are taken to be ground truths from which all of the calibration values are recalculated. The new calibrations are then used in stroke_energy.m to get a new set of energy calculations. After 5 iterations the values start to converge and the resulting calibrations and values are used as the final values. The iteration is performed to remove far outliers (2-4 order of magnitude larger or smaller than other values) and get a more stable result.

Once the final energy and calibration values are found, the data is written to both .loc and .mat files for the AE-files and the .txt files for the calibration and conversion values. The conversion file is only used within the processing internally (and with reprocessing), for any station level calibration work the calibration file should be used. Then the **de_mapper.m** function is called to get the relative detection efficiency for the day, the results stored as a DE-file. Finally a set of network statistics are stored in a file for quick reference (currently this file is not being ingested or incorporated anywhere, such as the WWLLN management page).

B.5   Relative Detection Efficiency Code

The overall description of the relative detection efficiency model is given in *Hutchins et al.* (2012a).

B.5.1   de_mapper.m

de_mapper.m creates the relative detection efficiency model for each hour in the day, currently the code is written such that this step is parallelized if the parallel toolbox is

available and running. Within each hour the subset of operational stations (> 500 stroke solutions) are used to create the model.

For each station the stroke-station pairs are found that are between $0.5 - 20$ Mm, these are used to determine the station detection threshold for the current hour. The threshold is determined as the 5th percentile value of the local electric field value (in uncalibrated soundcard units) from the stroke-station pairs. This is converted into a minimum stroke energy for an array of potential attenuation rates to serve as a lookup table. Using the current station calibration averaged over the past 7 days.

The **terminator.m** program is used to determine the percent day and night paths between the station and every grid point on the globe. These ratios are combined with the LWPC lookup tables to determine the attenuation at every grid point. Finally the attenuation values are used to index the potential stroke energy lookup to get the minimum detectable stroke energy at each grid point.

After this is performed for each station the 5th lowest value is found at each grid point. The 5th lowest is chosen as at least 5 stations are required to locate a WWLLN stroke, if that requirement ever changes then it would need to be changed in this code as well. The 5th lowest value corresponds to the minimum detectable energy of the network for the given hour.

The minimum detectable network energy is compared to the energy distribution of the past 7 days. For this distribution only strokes with energy solutions using 2 or more stations are totaled. The percentage of strokes above the minimum is called the relative detection efficiency for that grid point.

B.5.2   Output Interpretation

The output of the model is a $n \times m \times 24$ array where n and m are determined by the resolution of the lookup tables. For example if the lookup tables have a resolution of $2°$, then $n = 180$ and $m = 90$. The third dimension of the array corresponds to the UTC hours 0–23.

In each grid cell is the relative detection efficiency of the network on a 0–1 scale. The 1

corresponds to locations that could detect 100% of the past 7 day distribution and are the best performing regions of the network. Other locations will have values below 1%, such as Antarctica. It is recommended that prior to use a minimum relative detection efficiency, such as 5%, is set on the maps to prevent unphysical stroke density counts.

In the automated Boostrap processing the resulting relative detection efficiency maps are expanded to a 1° grid and smoothed to give a standardized size in the case that the lookup tables change their resolution. In the unlikely case that the resolution goes below 1° the maps should then be downsampled to recreate this resolution.

Appendix C

WWLLN SERVICE UNIT V4

C.1 New Design

The WWLLN service unit was redesigned due to the obsolescence of the previous Trimble GPS unit, this redesign allowed for several other major revisions to the board:

- Built in USB-Serial conversion

- On board computer

- Remote controlled pre-amp power system

- Remove inline preamp LEDs

- Increased stereo driving power

- Two-way GPS serial communication

The biggest change is the addition of the onboard computer, eliminating the need for a separate CPU to be installed with every service unit. The Gumstix WaterStormCOM with Tobi breakout board was chosen for the onboard computer. An on board computer is able to toggle the preamp power supply remotely, communicate with the GPS engine via TSIP, and standardize the host computer capabilities. The USB-serial convert is used to output GPS messages to an external computer if desired, and two-way communication can be switched from the GPS to the USB port via the use of an onboard jumper. Finally the system is designed to be operated remotely via SSH, but it does have the ability to use a keyboard, mouse and monitor through a powered USB hub and HDMI capable display.

C.2 Gumstix Selection

There are a few important aspects of the Gumstix WaterStormCOM that makes it suitable as a WWLLN service unit computer. Other compact computers meet most, if not all, of these requirements and they should be satistifed for future upgrades and revisions to the design. The primary requirements are:

- 48 kHz stereo input

- Ethernet connection

- Linux OS

Once these are met the secondary requirements are:

- Serial input (GPS comminication)

- 8 GB of memory

- 512 MB of RAM

- USB

- Video out (VGA/DVI/HDMI)

- Low power (< 1 A)

- Low cost ($<\$300$)

Other alternative systems include BeagleBoards, PandaBoard, Gumstix (other models), and maybe future versions of RaspberryPi. One consideration not listed above is size. The Gumstix was chosen for the final computer since it could fit into the previous form factor for the service units, other systems such as BeagleBoards were too large to fit easily.

Another important note is that some of the requirements can be remedied with USB peripherals if suitable ones are found. RaspberryPi with a USB soundcard could work in lieu of the Gumstix, however no USB soundcard could be reliable found that accurately processed the GPS pulse per second signal.

C.3 Layout and Design

The Eagle files (v5) are available on `flashfile` at `/home/mlhutch/wwlln/eagle/` or through a git repository (see Appendix A).

C.3.1   Design

The board design is shown in Figure C.1 with corresponding component names and values. The Gumstix is represented as a 40 pin header and a matching header needs to be soldered to the Tobi breakout board. One adjustment that may need to be changed in future iterations is the voltage divider on the VLF and PPS signal output to the internal stereo jack (upper left of diagram). Current values are set to reduce the two signals to within the input voltages of the Gumstix sound card.

C.3.2   Layout

Unlike previous versions of the service unit most small components (resistors, capacitors, etc.) are placed on the underside of the board so the same form factor can be used. The overall layout is shown in Figure C.2 with the topside in Figure C.3 and bottom in Figure C.4. Both the USB and ethernet jacks are on the board so the Gumstix does not face directly out of the box, preventing inevitable wear and tear on the Gumstix itself. The HDMI plug is not on the board due to the inability to wire two jacks together (pins are spaced too close) so a wall-mount HMDI extension cord is used instead.

C.4   Construction

Service unit construction steps:

1. Populate board with components

2. Add in socket for USB-Serial converter

3. Add in pins for GPS jumper

4. Add jumper set to Gumstix

5. Screw down voltage regulators

6. Solder connection pins to USB-serial converter
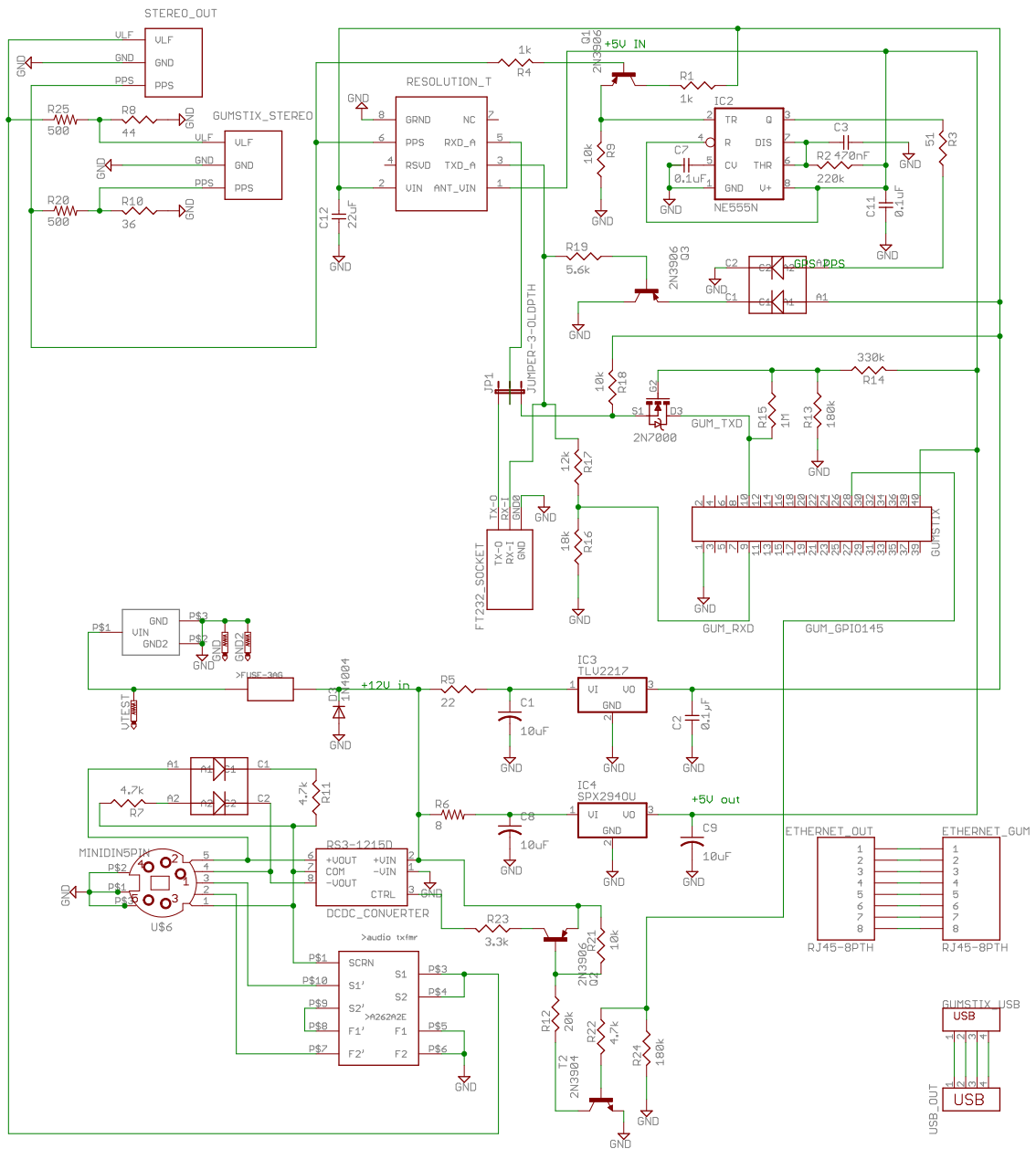
Figure C.1: WWLLN Service Unit v4d design.

7. Solder matching header to Gumstix Tobi board (with COM detached!)

8. Test board (see Section C.5.1)
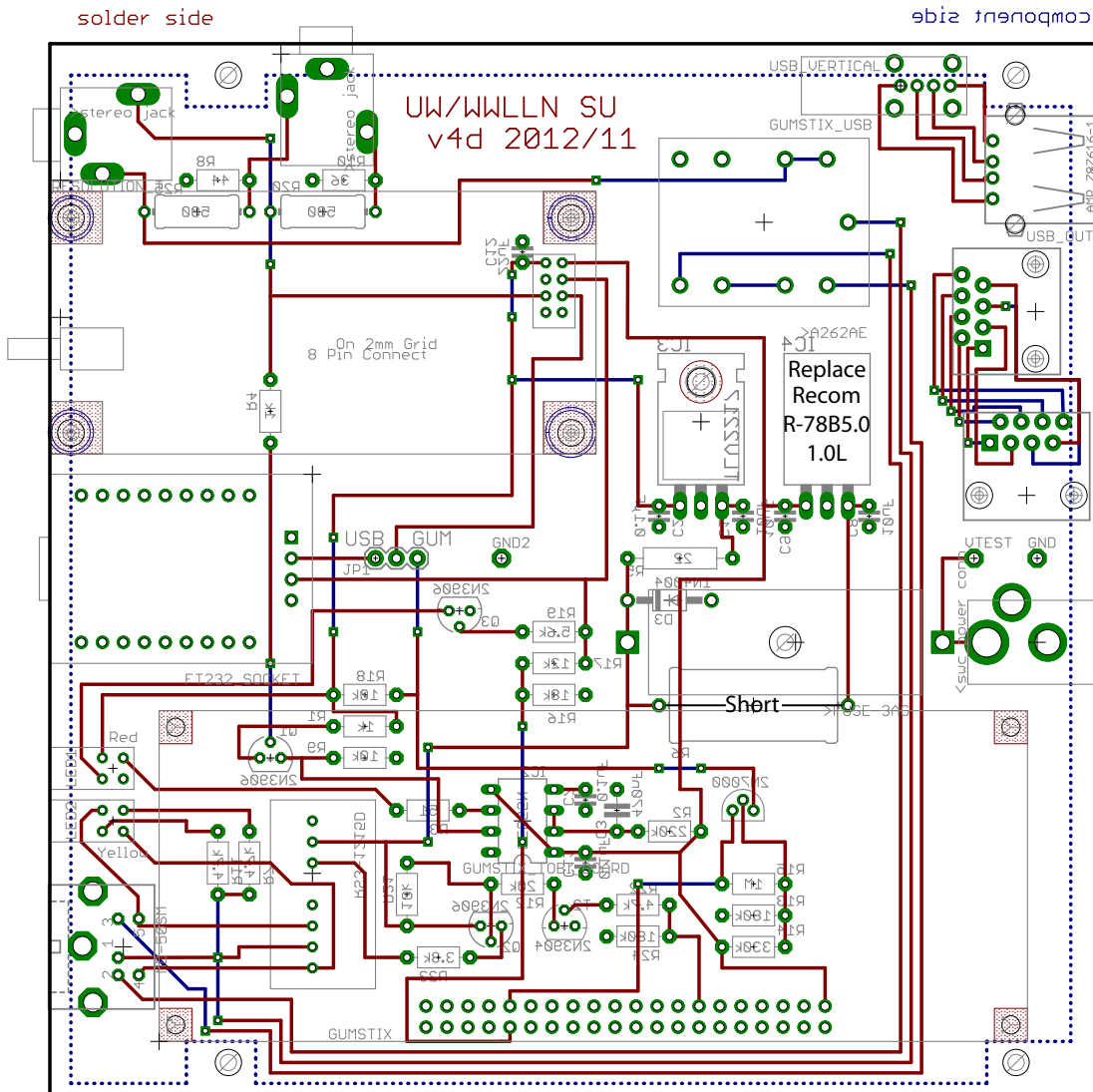
Figure C.2: WWLLN Service Unit v4d Schematic

9. Mount GPS and USB-serial converter

10. Test board (see Section C.5.2)

11. Mount Gumstix to board

12. Plug USB, ethernet, HDMI, and stereo cables into Gumstix
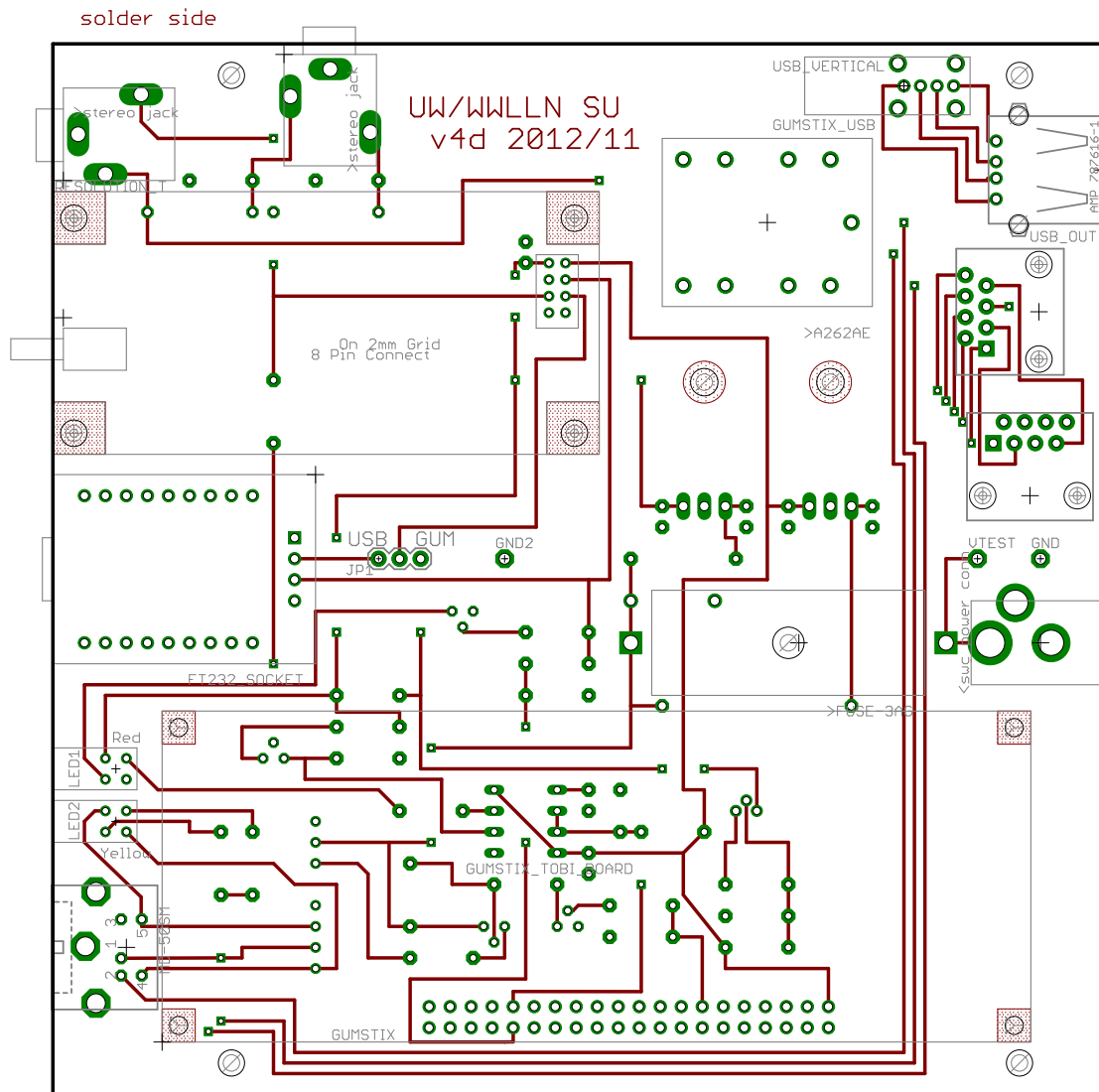
Figure C.3: WWLLN Service Unit v4d Topside

13. Plug USB, ethernet, and stereo into Service Unit board

14. Insert formatted microSD card into Gumstix board

15. Test board and Gumstix (see Section C.5.3)

16. Drill out holes in Service Unit box (Figures C.5 and C.6)
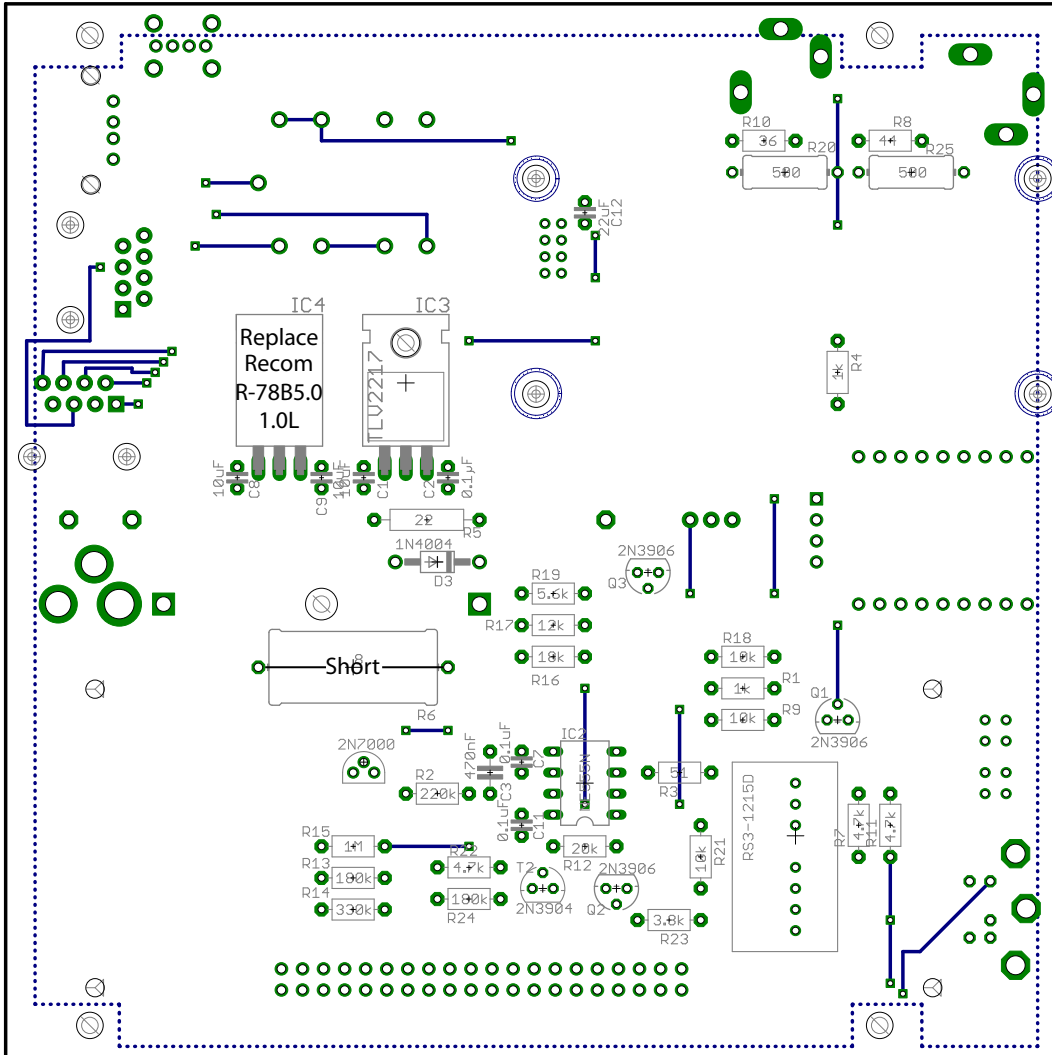
Figure C.4: WWLLN Service Unit v4d Bottomside

17. Attach HDMI wall mount to side of box

18. Mount Service Unit board into box and wrap cables to fit

19. Final Service Unit and Gumstix test (see Section C.5.4)

Front

(0,0)

Ø 1/4"
(1.4375", 0.97")

(2.4375", 1")
Ø 3/8"

(3.5", 1.4375")
Ø 3/8"

(4.6", 1.05")
Ø 1/2"

Ø 1/2"
(5.28", 1.14")

Back

(0,0)

Ø 3/8"
(3.225", 1.095")

Ø 3/8"
(4.74", 1.125")

(4.4375", 1.03125")
Ø 3/8"

(6.5", 1.3125")
Ø 1/8"

(6.5", 0.8125")
Ø 3/8"

(6.5", 0.3125")
Ø 1/8"
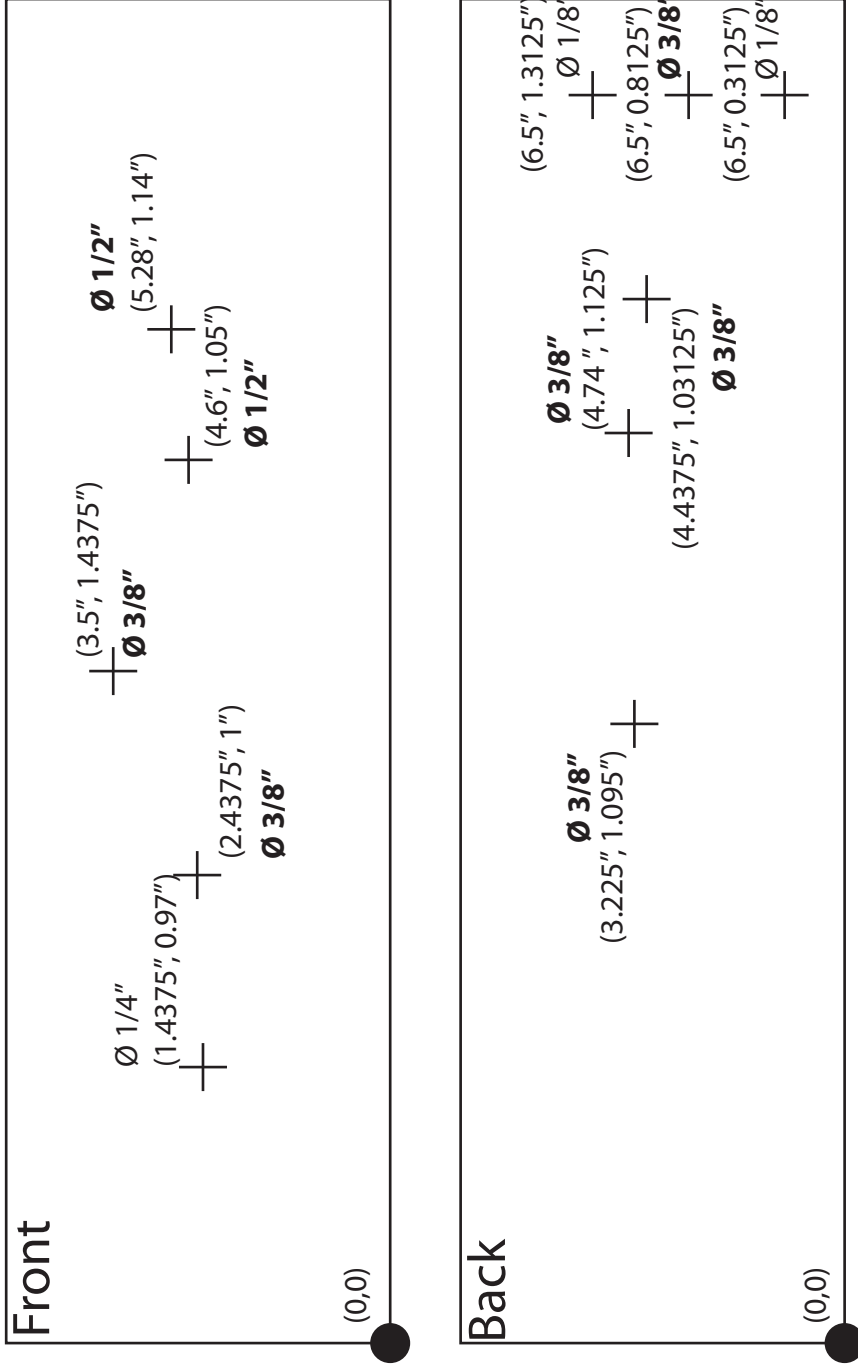
- Ø 1/2" and Ø 3/8" need Ø 1/8" pilots
- Use file and nibbler after test fit
- Trim Service Unit board corners

Figure C.5: Schematic for Service Unit box holes.
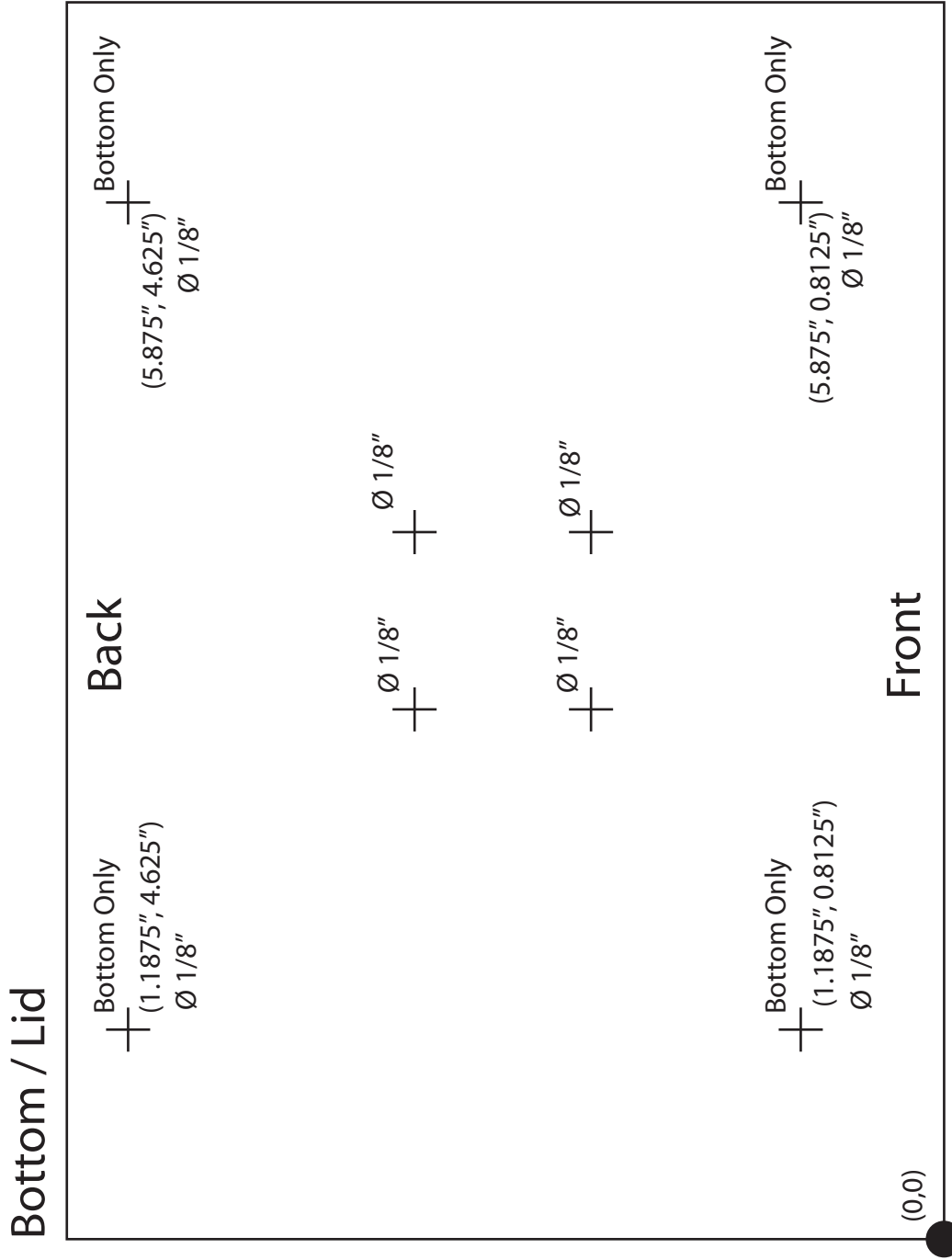
Figure C.6: Schematic for Service Unit mounting holes.

C.5 Testing Procedures

C.5.1 Initial Board Test

Without an external boards (Gumstix, GPS, etc) connected slowly raise the input voltage from 0 to 12 V. At 12 V the board should be drawing near to 55 mA, and both yellow LEDs should be on. The yellow LEDs indicate that the preamp power supply defaults on without the Gumstix plugged in. Test the voltages for the GPS and Gumstx mounts. The GPS should have two voltages, 5 V on pin 1 and 3.3 V on pin 2, all other pins should be 0 V. If the USB-serial jumper it set to Gumstix, then the GPS will have a voltage on pin 5. The Gumstix socket should have 0 V everywhere except on pin 40 which should be 5 V, pin 10 which should be 2.8 V, and pin 9 at 0.4 V.

C.5.2 GPS Power Test

Turn off the board and plug in the GPS engine. At 12 V power the board should be drawing 155 mA. The bottom red LED should be blinking and the top one should be on.

C.5.3 Gumstix Power Test

Plug in to the Gumstix and connect the ethernet (the ethernet LEDs allow for a quick check of the system). Power the GPS with a 12 VDC wall plug, not a variable power supply. After about 30 seconds the ethernet lights should start to blink and flash on the Gumstix ethernet port, indicating it has booted fully. While idling the board should be drawing 290 mA.

C.5.4 Final Service Unit Test

With all of the boards and cables connected, including the preamp and GPS antenna, power the complete service unit box on with the 12 VDC wall supply. While waiting for the GPS to sync up check to see if the Gumstix has started up properly, the yellow LEDs should flash on briefly, turn off and then turn back on when the system has started up. Follow the steps in Section C.6.2 to connect to the Gumstix via a direct ethernet connect to ensure all cable connections. If GPS does not sync up in 10 minutes check the status with either

the Gumstix readTSIP.py script, or connect to the GPS using the Trimble Studio Software via a Windows PC. Allow the box to run for several hours with the lid closed to check for overheating.

C.6 Software Setup

Most of the software is already installed and updated during the process of creating the microSD card, as described in Appendix D. However the GPS engine needs to be programmed directly, and the network settings are unique to each station host.

C.6.1 GPS Engine

There are three ways to configure the GPS Resolution T engine. The first is to utilize the Trimble Studio Configurator tool and load the configuration file that is located in the Gumstix software directory: gumstix/gps/ResolutionT Configuration.dat The second is to use the Trimble Studio tool to manually set the configuration on the GPS itself, if that option is taken the desired settings are listed in Table C.6.1. The third option is to configure the GPS engines from the configured Gumstix itself.

This can be done with the **sendTSIP.py** program included in the normal Gumstix software. After the configure GPS option has been selected within the program the settings are not yet saved onboard the GPS (i.e. a power cycle will reset them). A second command, save to ROM, needs to be sent in order to write the settings to the GPS memory.

The same program also allows for setting the output to be NMEA messages instead of TSIP. However once set to NMEA the GPS can no longer receive TSIP commands, a power cycle is needed to get out of NMEA mode.

	Setting	Value
I/O	Position	LLA
	Altitude	HAE
	Velocity	ENU
	Time	UTC
	Signal Level	AMU
	Satellite Tracking Packet	0x5C
GPS	Mode	O-D Clk
	Dynamics	Stationary
	DGPS	Auto Off
	Datum	WGS-84
	Auto Jamming	Enabled
	Elevation Mask	5°
	Signal Level Mask	4 AMU
	PDOP Mask	12
	PDOP Switch	6
PPS	Output	Enabled
	Polarity	Positive
	Time Base	UTC
	Qualifier	>= 3 Satellites
	Packet Masks	0x8F-AB

Table C.1: Resolution T GPS Settings

C.6.2 Network Configuration

The WWLLN Service Unit v4 runs a customized version of the Angstrom Linux distribution for ARM processors. For additional help contact Michael Hutchins (mlhutch@uw.edu) or Bob Holzworth
(bobholz@ess.washington.edu).

WWLLN Service Unit v4 Initial Setup

Method 1: SSH Setup   The SSH setup method requires:

- Ethernet cable

- SSH capable computer

1. Connect SU to a host computer directly with an ethernet cable

2. Set host computer ethernet network settings to:

   ```
   address: 192.168.10.1
   gateway: 192.168.10.100
   netmask: 255.255.255.0
   ```

3. SSH into the SU from host computer:

   ```
   ssh -p 7777 sferix@192.168.10.2
   password: [ ]
   ```

4. Set desired static ip configuration in file ∼/networkSetup.sh

5. `sudo ./networkSetup.sh`

6. Switch SU to main network ethernet within 1 minute of running networkSetup.sh

7. Test connection by SSH'ing into SU with new IP address

8. (a) If successful: set new IP setting in /etc/network/interfaces

   (b) If unsuccessful: power cycle SU and check settings starting with step 3

9. Reset SU and confirm new settings

Method 2: Workstation Setup   The Workstation setup method requires:

- HDMI Monitor and cable

- **Powered** USB Hub

- USB Keyboard

- USB Mouse

Connect the powered USB hub to the back USB port of the service unit, and attach the keyboard and mouse to the hub. Connect a monitor to the HDMI port, DVI - HDMI adapters work as well. Power on the box, it will take a few minutes for the login screen to show up. Select "Other..." and login with the username **host**. Wait a few more minutes for the graphical display to load.

Adjust the network settings by either changing the files listed in Method 1, or by logging in as root as adjusting them through System → Network in the top menu bar.

1. Connect an HDMI display, keyboard and mouse

2. Set network information through GUI

Method 3: Manual microSD Editing   The file that need to be edited on the rootfs partition are:

```
/etc/network/interfaces
/etc/resolv.conf
/etc/init.d/dropbear
```

The interfaces file lists the IP information of the machine whole the resolv.conf file is for the DNS information. The sshd_config file on line 13 sets the port with which SSH is allowed.

The last step, if a non-standard port is being used, is to also alter the built in firewall of iptables and netfilter. The firewall settings are stored in:

```
/etc/iptable.rules
```

and can be edited as a standard iptables configuration file.

Website Setup

Starting apache2   To get apache2 running only one change needs to be made in the /etc/apache2/httpd.conf file.

```
Line 96: #ServerName www.example.com:80
```

Needs to be uncommented and changed to the hostname of the computer, e.g.:

```
Line96: ServerName gumstix.ess.washington.edu:80
```

Then httpd needs to be restarted:

```
sudo httpd -k restart
```

Setting up the website   All changes to the website need to be made in the /home/sferix/public_html_static folder, this folder is copied to /home/sferix/public_html during start up. Changes to public_html are not saves as the folder is located in system RAM due to SD card read/write limitations. A restart in not necessary if the public_html_static contents are copied to public_html.

C.7 Operations

C.7.1 LED Signals

The two sets of front facing LEDs can be used to diagnose most issues with a given Service Unit. The left two red lights correspond to the status of the GPS receiver, and the right two yellow lights with the Gumstix and Preamp.

The bottom red LED lights up when the GPS engine is sending a serial packet to the Gumstix and USB serial out port on the front. The packet is a TSIP packet (9600 801) that can be read on the Gumstix with **readTSIP.py** python program, on a connected computer with the same program, or on a Windows computer with the Trimble Studio Software.

The red LED lights up for every pulse per second. If the GPS is not synced with three or more satellites the engine will not send a pulse per second.

The two yellow LED's correspond to the ± 15 V preamp power supply output. The default setting for no Gumstix attached is to turn the preamp power supply on. When the Gumstix is booting it first turns the preamp off (default behavior) and then turns it on during the boot process. The power can be manually toggled with the **preampOn.sh** and **preampOff.sh** scripts. If the Gumstix does not turn the preamp on after a minute then there is something wrong with the Gumstix or the operating system.

If none of the LEDs turn on then the fuse has likely blown and needs to be replaced.

Appendix D

GUMSTIX

D.1 Hardware

The WWLLN Service Unit uses a Gumstix WaterStormCOM mounted on a Tobi breakout board as the on-board computer to run the WWLLN software.

D.1.1 Hardware Summary

The Gumstix WaterStormCOM is part of the Overo COM series and has the following features:

- 1 GHz ARM Cortex-A8 CPU

- 512 MB RAM

- microSD Card Slot

- OpenGL POWER SGX graphics Aaccelerator

- C64x Fixed Point DSP (Max: 660,800 MHz)

The Tobi breakout boards adds:

- HDMI video Out

- 1 USB port

- 1 USB console connection

- Stereo in/out

- Ethernet

SV1

V_BATT_5	40	39	ADCIN4
ADCIN3	38	37	AGND
ADCIN5	36	35	ADCIN6
ADCIN2	34	33	ADCIN7
PWM1	32	31	PWM0
GPIO144_PWM9	30	29	GPIO147_PWM8
GPIO145_PWM10	28	27	GPIO146_PWM11
VCC_1.8	26	25	GND
GPIO185_SDA3	24	23	GPIO184_SCL3
GPIO166_IR_TXD3	22	21	GPIO165_IR_RXD3
GPIO163_IR_CTS3	20	19	GPIO170_HDQ_1WIRE
GPIO10_TS_IRQ	18	17	GPIO186_GPS_PPS
VCC_1.8	16	15	GND
POWERON	14	13	GPIO31_WAKEUP
VBACKUP	12	11	SYS_EN
GPIO148_TXD1	10	9	GPIO151_RXD1
GPIO175_SPI1_CS1	8	7	GPIO173_SPI1_MISO
GPIO174_SPI1_CS0	6	5	GPIO172_SPI1_MOSI
GPIO114_SPI1_NIRQ	4	3	GPIO171_SPI1_CLK
VCC_3.3	2	1	GND

Figure D.1: Pinout of the Gumstix Tobi breakout board.

D.1.2 Pinouts

The Gumstix COM has direct pinouts to the processor, however as it is not used without the Tobi breakout board, only the Tobi pinout is shown in Figure D.1.

In the service unit only pins 1 (GND), 9 (RXD1), 10 (TXD1), 28 (GPIO145), and 40 (V_BATT) are used. RXD1 and TXD1 are used for the serial communication with the GPS. GPIO145 is used to control the pre-amp power supply. V_BATT is the 5 V power for the Gumstix.

D.2 Gumstix Operating System v2.0

D.2.1 Gumstix-Yocto

The Gumstix used in the service unit is running a custom Linux distribution created using the Yocto Project build system. This OS runs similar to most unix operating systems, with the main difference being the smart package manager instead of `yum` or `apt-get`.

A useful resource in setting up and configuring the Gumstix software is the Gumstix developer site (`http://gumstix.org`) and the mailing list archive forum (`http://gumstix.8.x6.nabble.com`).

D.2.2   Distribution Location

The operating system used is the Yocto Project Gumstix Layer v1.5, available on `flashfile` or in the Git repository:

    `/home/mlhutch/Git/gumstix.git` repository on `flashfile` in the image folder.

D.2.3   Building

Building the operating system can be done by following the instructions at:

    `https://github.com/gumstix/Gumstix-YoctoProject-Repo`

    With the final configuration in the Git repository on `flashfile` in `image/yocto/local.conf`. If rebuilding use the `local.conf` file in the `build/conf` folder and run the bitbake target `bitbake gumstix-xfce-image`.

    With either a newly made OS image or the one in the `gumstix.git` repository, follow the install instructions in `INSTALL.md` for setting up a new microSD card to run with WWLLN.

D.3   Software

D.3.1   WWLLN Software

The WWLLN software is provided by James Brundell and compiled specifically for the ARM process. The three programs are `toga`, `ntpcheck`, and `GDspectro`. `toga` is the main WWLLN processing programming that reads in the VLF and GPS signals to produce the UDP packets sent on to the main WWLLN processors. It should be always running on the system with a `crontab` entry such as:

`0,5,10,15,20,25,30,35,40,45,50,55 * * * * toga -s 100 -a 3 -j 1 -g -o &`

    This will try to start it every 5 minutes in case it stops for any reason.

    The `ntpcheck` and `GDspectro` and programs used by `toga` but do not need to be called or run on their own.

D.3.2   Hardware Controls

Pin GPIO145 is the pin that controls whether the preamp power supply is turned on or off. When the pin is held low (value of 0) the power supply is on, when it is set high (value of 1) it is turned off. The command to change a GPIO pin value is:

```
echo 0 > /sys/class/gpio/gpio145/value
```

The two scripts `preampOn.sh` and `preampOff.sh` can be used to easily toggle the preamp power supply.

The default value for GPIO pins is to hold them high, so during boot the preamp turns off until the `preampOn.sh` script can be called at the end of the boot sequence.

D.3.3   GPS Interface

The Trimble GPS communicates with the TSIP protocal, compared the NMEA of the previous GPS engine used. The pythons script `readTSIP.py` interprets the TSIP messages and reports the GPS status to the file gps.log and prints them to the console. The console printing can be turned off by changing the variable `print_to_console` to `False`.

The program can be started and run in the background to produce a continues record of GPS activity. The default location for the `gps.log` file is in the `public_html` folder where it can be remotely checked through the service unit website.

D.3.4   RAM Disk

The Linux distribution for Gumstix automatically sets up a RAM disk for users. It is created at `/var/volatile` with half of the available RAM (256 MB). It needs to be used for the running of the WWLLN software as the microSD card is too slow. At start-up the `public_html` folder (logs and spectrograms) and sferics folder are created in the ram disk and symlinked to the main sferics directory.

For this reason all permanent edits to the Service Unit website should be made in the `public_html_static directory`.

D.4   Creating Gumstix microSD Card

There are two methods for configuring a new microSD card for use with the service unit Gumstix computer. Either a card can be formatted and loaded with the latest software following the `INSTALL.md` instructions, or an existing installation disk image can be copied over (located on `flashfile` and `flash5`).

D.4.1   Card Duplication

Once a new card is created it is advised to make an exact copy of the card to allow for duplication onto new cards. Since the partitioning, and bit location on the card, is critical for the Gumstix the direct copy program `dd` should be used to make the copy. To copy a card to a disk image the command will look like:

```
sudo dd if=/dev/sdb of=/Path/To/Target/gumstix.iso bs=512
```

And the command to copy a microSD image back to a new card in the same slot as the previous card:

```
sudo dd if=/Path/To/Target/gumstix.iso of=/dev/sdb bs=512
```

A word of warning: the `dd` command can overwrite and destroy hard drives if they are incorrectly targeted. Always double check the mount point of the microSD card before running the `dd` command.

D.5   Gumstix System Setup

Subsection D.5.1 gives the setup instructions for a new station, namely setting the IP address so it can be added to the network. A more up to date version is available in the `gumstix.git` repository in the `user_manual` folder. Also included in the station setup instructions are how to setup and customize the service unit webpage, as described in Subsection D.5.2.

D.5.1   Connecting to the Gumstix

Method 1: SSH Setup

The SSH setup method requires:

- Ethernet cable

- SSH capable computer

1. Connect SU to a host computer directly with an ethernet cable

2. Set host computer ethernet network settings to:

   ```
   address: 192.168.10.1
   gateway: 192.168.10.100
   netmask: 255.255.255.0
   ```

3. SSH into the SU from host computer:

   ```
   ssh -p 7777 sferix@192.168.10.2
   password: [ ]
   ```

4. Set desired static ip configuration in file ∼/`networkSetup.sh`

5. `sudo ./networkSetup.sh`

6. Switch SU to main network ethernet within 1 minute of running `networkSetup.sh`

7. Test connection by SSH'ing into SU with new IP address

8. (a) If successful: set new IP setting in `/etc/network/interfaces`

   (b) If unsuccessful: power cycle SU and check settings starting with step 3

9. Reset SU and confirm new settings

Method 2: Workstation Setup

The Workstation setup method requires:

- HDMI Monitor and cable

- **Powered** USB Hub

- USB Keyboard

- USB Mouse

Connect the powered USB hub to the back USB port of the service unit, and attach the keyboard and mouse to the hub. Connect a monitor to the HDMI port, DVI - HDMI adapters work as well. Power on the box, it will take a few minutes for the login screen to show up. Select "Other..." and login with the username `host`. Wait a few more minutes for the graphical display to load.

Adjust the network settings by the steps listed in Method 1 or 3.

1. Connect an HDMI display, keyboard and mouse

2. Follow Method 1 or Method 3

Method 3: Manual microSD Editing

The files that need to be edited on the `rootfs` partition are:

```
/etc/network/interfaces
/etc/resolv.conf
/etc/systemd/system/sshd.socket
```

The interfaces file lists the IP information of the machine whole the `resolv.conf` file is for the DNS information. The `sshd.socket` file sets the port with which SSH is allowed.

The last step, if a non-standard port is being used, is to also alter the built in firewall of `iptables` and netfilter. The firewall settings are stored in:

```
/etc/iptable.rules
```

and can be edited as a standard iptables configuration file.

D.5.2   Website Setup

Starting apache2

To get `apache2` running only one change needs to be made in the `/etc/apache2/httpd.conf` file.

```
#ServerName www.example.com:80
```

Needs to be uncommented and changed to the hostname of the computer, e.g.:

```
ServerName gumstix.ess.washington.edu:80
```

Then `httpd` needs to be restarted:

```
sudo httpd -k restart
```

Setting up the website

All changes to the website need to be made in the `/home/sferix/public_html_static` folder, this folder is copied to `/home/sferix/public_html` during start up. Changes to `public_html` are not saves as the folder is located in system RAM due to SD card read/write limitations. A restart in not necessary if the `public_html_static` contents are copied to `public_html`.

D.5.3   Sound Settings

The Gumstix a myriad of analog inputs that are all controlled with `alsamixer`. For the WWLLN service unit the stereo input is controlled by the <Analog> input as shown high-lighted in Figure D.2, and a digital gain through <TX1 Digital>. Included in the installation files is the default alsa profile `asound.state`.
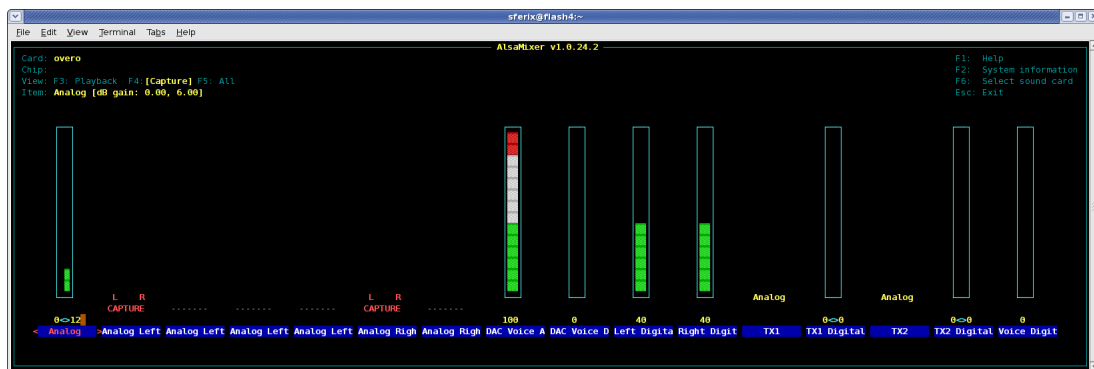
Figure D.2: `alsamixer` settings for Gumstix stereo input, <Analog> controls the stereo gain.

D.6 Common Problems

- The network icon in the top menu bar says the network connections are disabled.

  The GUI network manager is disabled, but the network settings set as above still work.

- The GPS pulse per second is not working with the TOGA program, it lists "PPS bad" for most lines.

  Adjust the gain on the pulse per second with Alsamixer ¡TX1 Digital¿ right channel, usually lowering it will resolve the problem.

Appendix E

WEBSITE

This appendix discusses operations of the updated WWLLN website and the real-time lightning map.

E.1   WWLLN.net

The new WWLLN website uses Wordpress as the content management system for every webpage. The structure of the webpage (e.g. headers, footers, titles, spacing) are stored in the Wordpress theme, while the content is, for the most part, stored in either the MySQL database or in three list files. Editing the files in the theme will change every corresponding page of the website, while content edits will just change that page.

E.1.1   Directory Structure

The website directory looks like:

```
backup.sh
hosts_list.htm
index.php
map/
map_demo/
publications_list.htm
README.txt
spectra_list.htm
volcanoMonitor.html
wordpress/
```

The three files `hosts_list.htm`, `spectra_list.htm`, and `publication_list.htm` are used to generate the list content on their respective pages (hosts, spectra, publications).

`backup.sh` is a script to backup the Wordpress database to a file called `wwlln_site.bak.sql.bz2` that is discussed below.

`index.php` redirects the user to the Wordpress based pages and generally should not be changed.

`map/` and `map_demo/` are the folders containing the scripts and data for the WWLLN lightning maps. The `.htaccess` file in each directory should be used to control who can access them (this data access is completely separate from Wordpress). The `.htaccess` is currently set to use the `.htpasswd` file at `/home/mlhutch/sites/.htpasswd`

`volcanoMonitor.html` is used to redirect users to the new volcano monitor page, other previous permanent links (e.g. `wwlln.net/climate`) were replaced more easily and did not need a redirect file.

Finally there is the `wordpress/` directory. This contains all of the support files used to power the wordpress aspects of the website, including the admin page, the blogs, the MySQL database settings, and WWLLN theme page settings.

Outside of the website directory is the MySQL database. This is stored by the MySQL program running on webflash and does not have an exact file location. How to backup and restore the content from the MySQL database is covered below.

E.1.2 Wordpress

Files

Inside the `wordpress/` directory there are only a few files that will ever need to be changed, or that even should be changed. The main file is the `wordpress/wp-config.php` file, this file contains the information for connecting to the MySQL database including the user name and password *in plain text*. It is important to note that access to this file non-locally (e.g. through the website) is not allowed and it can only be seen with a local connection (e.g. ssh).

The other set of files in the `wordpress/` directory are the WWLLN theme files. These files control how the website looks and feels with some content stored in the files themselves. They are located in the non-intuitive location of `wordpress/wp-content/themes/wwlln`. Editing these files will change how the corresponding pages look. In general only the style files (*.css files) should be changed to alter the website appearance. The other files that may be changed is the `header.php` file, this is where you can change the items and order of the menu bar.

Admin

The admin panel for Wordpress is used to create content alter the site if the user is an administrator. To access the admin panel go to `wp-admin.php` from the main page of the website (e.g. `wwlln.net/wp-admin.php`. From here new pages can be made, new blog posts, website settings, and user control.

Content Creation

Wordpress content comes in two forms: blog posts and static pages. Blog posts show up in reverse chronological order on the blog (newest first) and the first few are listed on the front page of the website. Blog posts are created with the Wordpress tools and text editor and support embedding pictures and videos. Pages are similarly created but use different templates instead (each described below). Both can be drafted, edited, published, deleted, and generally controlled through the admin page. If other users are added (say other management team members) they can add blog posts as well.

Page Templates

There are 6 page templates, depending on the template the content will be placed in different locations. All pages can be written with the text editor provided or HTML.

- Default Template – content is placed inside a white box exactly like the blog posts.

- HTML Content Page – blank page below the header that uses the HTML given to create the page

- Home page – main index page, entirely in HTML. Editing using the visual editor may break the page.

- Lightning map – page specifically designed to show the WWLLN lightning maps. The content is placed below the inset lightning map. Note: there needs to be a folder with the same URL as the listed page with the lightning code in it, see the existing pages/folders for examples.

- List – Uses a [name]_list.htm file to generate a single list of that data below the content, where [name] is the post URL. The file format is one line per list item with a blank line in between.

- List (Two-Column) – same as List but uses two columns. Two lines between spaces where the second line is the second column entry. See hosts_list.htm as an example.

E.1.3   MySQL

The MySQL server can be backed up with the `backup.sh` script or with the command:

```
mysqldump --add-drop-table -h localhost -u wwlln -p wordpress |
       bzip2 -c > wwlln_site.bak.sql.bz2
```

To restore the Wordpress database from a back up file is relatively easy, use the command:

```
mysql -h localhost -u wwlln -p wordpress < wwlln_site.bak.sql
```

Which will then prompt for the password (located in the `wp-config.php` file) and perform the backup. This does not require any special sudo permissions to run.

E.1.4   Moving Wordpress

Moving the Wordpress directory is a more involved process and is not recommended to do casually. First backup the current Wordpress directory along with the `.htaccess` and `index.php` files in the root directory. Then go to the settings page on the websites `wp-admin.php` page, here change the website and Wordpress directories to what they will be after the move. After saving the site will stop working. Move the Wordpress directory to the new location listed on the admin page and move the `.htaccess` and `index.php` files to the root of the website page that was also listed. Now the admin page should work. Back on the admin page go to the Permalinks page on the settings menu and at the bottom will be the updated .htaccess file listed. Propagate the changes over to the root `.htaccess` file and the `wordpress/.htaccess` file. After that the website should be working as before.

E.2 Lightning Maps

A WWLLN (http://wwlln.net) visualization tool to show realtime lightning activity.

The map displays WWLLN data viewed in realtime or at other speeds. The user can pause the display, return to the start of the loaded data, jump around by 30-seconds, resume playing, or return to real time data display (where available). There are options to show a Google Map cloud layer (only shows current cloud conditions), a density map of all loaded strokes, place a selection box for more refined statistics, and an option to clear the loaded strokes in memory.

Users can also load their own WWLLN .loc files for playback and viewing. There is a load limit of 25,000 strokes to prevent slow playback due to too many strokes. Loading a file larger than 25,000 will only load the first 25,000 strokes and ignore the rest. Loaded data files play from the beginning of the file.

The terminator code is taken from `https://github.com/marmat/google-maps-api-addons`.

Login in controlled by the `.htpasswd` file specified in `.htaccess`. An `.htpasswd` file is created with: `htpasswd -c .htpasswd [new_user]` And amended with: `htpasswd .htpasswd [new_user]`

VITA

Michael Hutchins was born in Santa Rosa, California in 1987. He had an uncomplicated childhood set amidst the vineyards of the Sonoma wine country. Without any inherent struggles or tragedies in his life to fuel an artistic career, he left Northern California to attend the University of California Santa Barbara for a BS in Physics. He spent his third year at UCSB abroad studying Astrophysics at the University of Edinburgh, Scotland where a chance assumption directed him to graduate school. In 2009 he eschewed a move away from the West Coast to attain his PhD in Earth and Space Sciences studying lightning at the University of Washington, Seattle.