3-2022

Comparison of Lightning Warning Radii Distributions

Michael M. Maestas

Comparison of Lightning Warning Radii Distributions

THESIS

Michael M. Maestas, 2d Lt, USAF

AFIT-ENS-MS-22-M-149

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENS-MS-22-M-149

COMPARISON OF LIGHTNING WARNING RADII DISTRIBUTIONS

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Michael M. Maestas, BS

2d Lt, USAF

March 24, 2022

AFIT-ENS-MS-22-M-149

COMPARISON OF LIGHTNING WARNING RADII DISTRIBUTIONS

THESIS

Michael M. Maestas, BS
2d Lt, USAF

Committee Membership:

Dr. Edward D. White
Chair

Lt Col Timothy W. Holzmann, PhD
Reader

Capt Dawn L. Sanderson
Reader

AFIT-ENS-MS-22-M-149

Abstract

Previous research investigating lightning warning radii about the Cape Canaveral space launch facilities have focused on reducing these radii from either 5 nautical miles (NM) to 4 NM or from 6 NM to 5 NM depending on the structures being protected. Some of these findings have suggested the possibility of both a seasonal difference (warm versus cold) and lightning detection events (cloud-to-ground lightning (CG) or total lightning (TL)) impacting these radii and associated risk levels. Utilizing the 2017-2020 data provided by the 45th Weather Squadron at Patrick Space Force Base via the Mesoscale Eastern Range Lightning Information System (MERLIN), this thesis investigates these seasonal and data collection impacts. Our findings indicate that there is a substantial increased risk to warning radii's when just utilizing CG data in comparison to TL data. For the years studied, the mean risk for just using CG data was 5.94% in comparison to just 0.015% for the TL data for the 5 NM radii safety buffer. There were negligible seasonal differences between the warm season (May through September months) in comparison to the cold season (the remaining months) for using TL (0.015% versus 0.014%). In contrast, utilizing just CG data resulted in mean risk for the warm season of 4.86% compared to 11.56% for the cold season. Our recommendations are to utilize TL data where available. If using just CG data, then the risk of a lightning event occurring outside of a warning radii approximately doubles during the cold season in comparison to the warm season.

Table of Contents

Page

List of Figures

List of Tables

List of Algorithms

COMPARISON OF LIGHTNING WARNING RADII DISTRIBUTIONS

I. Introduction

1.1 Motivation and Background

With private companies, such as SpaceX and Blue Origin, and the emergence of the United States Space Force conducting space operations, it is important for contractors to understand the amount of risk they would undergo conducting an operation in extreme weather conditions, specifically lightning strikes, while reducing the amount of potential loss for postponing or damaging a spacecraft. According to the Federal Aviation Administration, calendar year 2020 witnessed 39 licensed launches, the highest number of licensed launches recorded since 1989, with most of the launches occurring within Florida (*Commercial Space Data*, 2021). Florida is often referred to the "lightning capital" of the North America given the high concentration of thunderstorms (Darack, 2007). The total cost of delay in the United States was roughly 15.4 billion dollars in 2007 alone (Ball et al., 2010), with 75% of flight system delays being due to weather conditions (Rosenberger et al., 2002). However, weather delays could be avoided if weather systems created more accurate depictions of where pre-existing storms occur.

Sanderson (2019) analyzed the distribution of lightning strike distances outside of a cluster of previous lightning strikes that can be seen as a storm. Within her analysis, she evaluated the Lightning Detection and Ranging (LDAR) II data collected from a previous system. The 45th Weather Squadron at Patrick Space Force Base collects this type on data, and in 2016, the Mesoscale Eastern Range Lightning Information

System (MERLIN) was introduced due to the previous system being unsustainable since the manufacturer stopped producing sensors for the maintenance (Roeder and Saul, 2012). Although the system is currently in use, analysis needs to be conducted to maximize operational gain while minimizing failures within the data set collected by the MERLIN system. Considering that the 45th Weather Squadron provides weather safety for 25,000 personnel and billions of dollars in resources (Roeder and Saul, 2016), maximizing operational gain is important while also minimizing failures within the MERLIN system.

Analysis on the data produced by the MERLIN system is needed since lightning within 5 nautical miles of a base cancels operations in that area. Although 5 nautical miles is the prevalent system to cancel operations, research to decrease the number has been conducted in order to continue operations since cancelling such operations within the military is extremely costly, while also protecting the airmen in the surrounding area. Captain Sanderson, in her research, found that a reduction from 5 nautical miles to 4 nautical miles incurred a risk of 0.277%, a number that seems to indicate the reduction is safe for operations using the LDAR II system (Sanderson, 2019).

1.2 Problem Statement

In order to analyze the MERLIN system, cluster analysis on the storms is needed in order to analyze where the storms are at any given time around the surrounding areas. Once these clusters are present, analysis on the distance between the edge of the storm and the new lightning strike will need to be conducted, while also gathering the probability of how far a new strike will occur outside of the boundary of the storm. Once the new strike occurs, the strike will be added to the cluster by creating a new elliptical cluster to include recent strikes. This should provide a Weibull distribution with 4 nautical miles being the cutoff point. This analysis will be conducted on the different types of MERLIN data: the cloud-to-ground data, the lightning aloft data, and the total lightning data that combined both the cloud-to-ground and lightning aloft data set. The purpose is to see if cloud-to-ground data is sufficient to provide accurate storm data or would the lightning aloft data set be needed in order to minimize failures. The hypothesis is that the cloud-to-ground data alone is insufficient and that lightning aloft data would need to be included in some way to minimize potential failures.

1.3 Organization of the Thesis

The structure of this thesis has the following layout. Chapter II covers the literature review of storm patterns and previous studies conducted, with Chapter III containing the methodology for this study, such as the types of analysis being used in this report. Chapter IV presents the results and analysis, while Chapter V contains the conclusion for this study as well as future research opportunities that should be conducted.

II. Literature Review

2.1 Overview

The use of lightning detection is important since it provides valuable information with regards to the evaluation of lightning launch commit criteria (Roeder and McNamara, 2006). With that in mind, it is important to look at the different types of lightning events and how lightning forms, as well as how lightning changes throughout the seasons. It is also important to look at the current system and how it stores the data collected. Previous algorithms are examined and goodness of fit tests are explained within this Chapter.

2.2 Types of Lightning Events

There are different types of lightning events and can be broken down into two different types of events: cloud-to-ground lightning events and lightning aloft events. These two types of events make up the total lightning data set captured by the MERLIN system (Roeder and Saul, 2016). The main difference between them is that the cloud-to-ground lightning events contact the surface of the planet while lightning aloft lightning events remain in the clouds. The difference can be seen in the lightning area when lightning events begin to form.

2.2.1 How Lightning Forms

While there is still some debate around the creation of lightning events, recent theories propose that an electric field is created within a thunderstorm, as well as a separation of electric charges through electrification is present within the storm (Saunders, 1993). When hot air moves upward as an updraft in the storm, it carries with it water droplets with it to higher altitudes (35,000 to 70,000 feet), creating super-cooled water droplets. At the same time, cold downdrafts push hail and ice from the higher portion of the cloud, and when these two items collide in the clouds, a graupel forms. These graupels either continue downwards in the downdraft or remain suspended in the cloud, usually around the center of the cloud.

As the upward moving super-cooled water droplets collide with the falling or suspended graupels, an electron is exchanged from the water droplet to the graupel, causing the water droplet to become positively charged while the negatively charged graupel to continue falling or remain suspended. The water droplet moves towards the top of the cloud, becoming ice or hail, along with other positively charged ice droplets, while the center of the thunderstorm becomes negatively charged. Figure 1 will show the process and the charges of the thunderstorm.

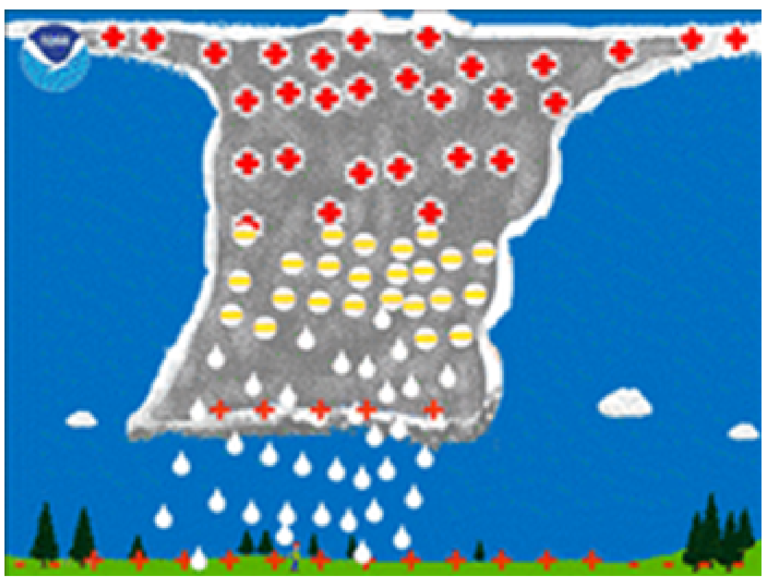

Figure 1. The Electrification of a Storm (NOAA)

With the electric field in the cloud or storm being stronger than the electric field between the surface and the base of the storm, the vast majority of lightning events occur above the surface of the planet, with about 70 percent of all lightning events remaining in the sky (Roeder and Rockledge, 2010). The other 30 percent of lightning events occur as cloud-to-ground lightning strikes, where the transfer of electricity can be devastating for the humans, as well as equipment, on the surface. The leading cause of deaths from storms in Florida from 1959 through 1999 were lightning strikes, which accounted for 49% of storm deaths (Roeder et al., 2012).

2.2.2 Cloud-to-Ground Lightning

As the thunderstorm moves over the surface of the planet, a group of positively charged particles start to accumulate along the surface and move in the same direction of the thunderstorm. These charged particles begin to rise onto tall objects such as metal poles, trees, and buildings. As the charged particles are moving upwards, a negative charge is sent outwards known as a "stepped leader" from the thunderstorm (US Department of Commerce, 2018*a*). The stepped leader moves out of the cloud and tries to sense a positive charge within 50 meters from its leader. The leader continues to surge for about 50 meters based on the air around the tip of the leading strike, producing small flashes of light that it hard to see within special equipment (weather.gov/safety). Figure 2 illustrates how the stepped leader appears coming from a thunderstorm.

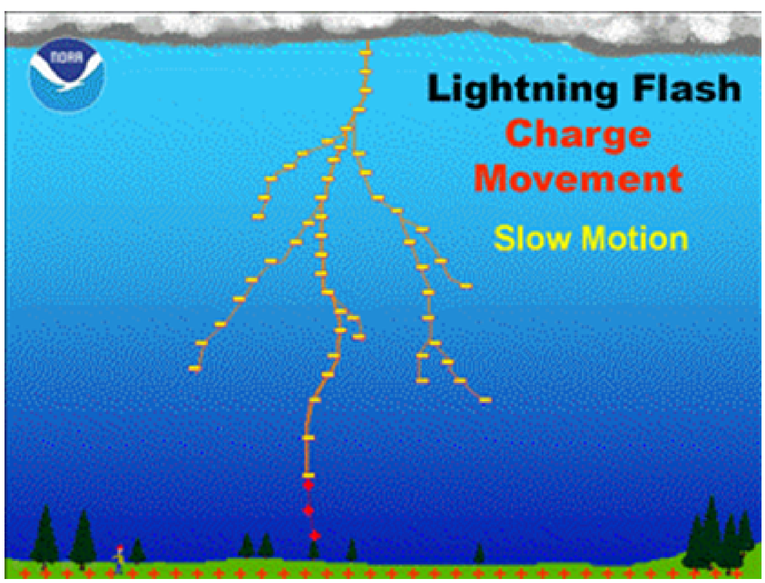

Figure 2. Stepped Leader Illustration (NOAA)

2.2.3 Lightning Aloft

Lightning aloft lightning events occur when the lightning strike occurs inside of the cloud itself or strikes a nearby cloud. Cloud-to-cloud lightning events occur when a lightning strike contacts another cloud nearby. Intra-cloud lightning events occur when a lightning strike occurs within the thunderstorm. Of the two types of lightning aloft lightning events, intra-cloud lightning events occur most frequently. Intra-cloud lightning events usually occur between the top of the anvil-shape of the thunderstorm and the base of the cloud. Figure 3 illustrates the different types of lightning events.

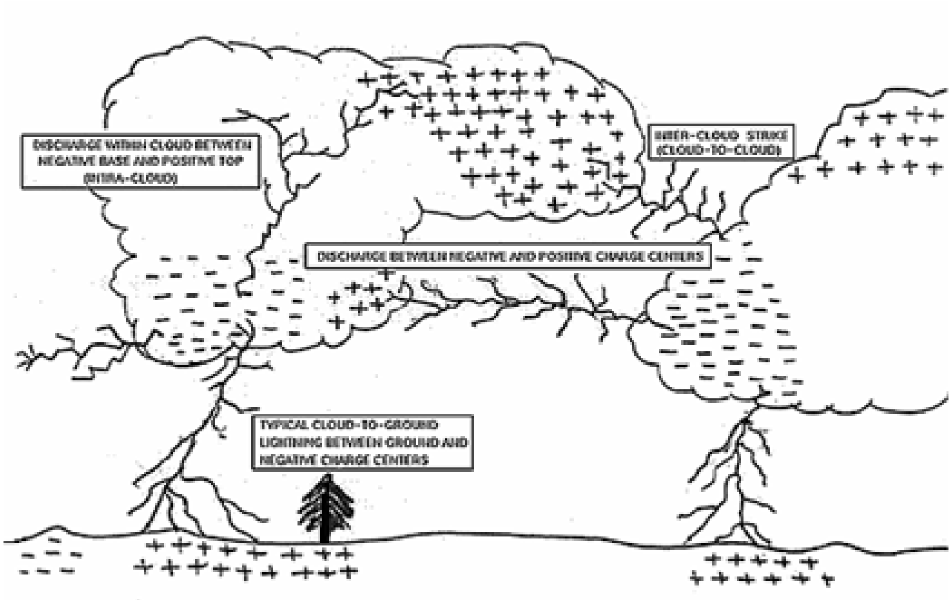

Figure 3. Different Types of Lightning Events (NASA, n.d.)

2.3 Summer vs Winter Storms

It can be shown that different seasons yield different types of storms or have an impact on the distribution of lightning storms. Lericos et al. (2002) defined warm season days between 1 May and 30 September, with the cold season days between 1 October and 30 April. Although this is a rough estimate of when warm seasons and cold seasons are defined in Florida, it is a great starting point to do analysis on the distributions of the distance between new lightning events and pre-existing lightning storms. With regards to previous research, Tello (2021) and Sanderson (2019) both observed lightning strikes between the months of May and September. Holland (2021) examined lightning events throughout the full year, with the warm season being between May and September, with the cold season being the months from October through April. Gold et al. (2020) conducted studies on lightning events and rain patterns, and considered the months of June through September to be the warm season and the months of October through May to be the cold season. In order to maintain consistency with previous research, the months of May through September will be considered the warm season and the months from October to April will be the cold season.

2.3.1 Seasonal Change

A set of requirements are needed to form a thunderstorm: moisture, atmospheric stability, and an event to trigger motion in the atmosphere (Price, 2009). As stated earlier, moisture is needed to produce precipitation, which is needed with conjunction with upward drafts to create electrification within the storm. During the warmer seasons, the United States has enough moisture to create and develop thunderstorms, given that the other requirements are also present. For atmospheric stability, upward drafts are needed to push cooled water particles upwards to contact falling ice and graupel to create electrification towards the center of the cloud. An unstable atmosphere allows air close to the surface of the planet to become lighter and more buoyant, allowing the air to rise quickly upwards. This is why warmer seasons typically see more unstable atmospheres, and hence more lightning events. As for the triggering motion event in the atmosphere, heating from the sun or cooling from water fronts or sea breezes can cause lightning events to occur.

2.4 Data Collection: MERLIN

With the introduction of the improved system, MERLIN provides more accurate information with regards to cloud-to-ground data compared to the 4DLSS, with better accuracy when it comes to stroke detection, flash detection, location accuracy and peak current (Hill et al., 2016). With regards to the lightning aloft data accuracy, the MERLIN system provides better accuracy for event detection and false detections, along with a higher location accuracy. The increase in model accuracy provides better tracking of storms and new lightning strikes, allowing analysts the ability to conduct thorough studies on lightning strikes and the distances between new strikes. The performance of MERLIN with regards to distances between strikes is not well documented (Roeder and Saul, 2016).

2.4.1 Lightning Aloft Data

The MERLIN data set collects both lightning aloft data as well as cloud-to-ground lightning events. The data is split by the day they occur as well as by lightning type: CC for lightning aloft and CG for cloud-to-ground lightning events. Figure 4 shows what the collected and finalized data appears using the MERLIN system for lightning aloft collected by the 45th Weather Squadron at Patrick Space Force Base in Florida.

```
2018-05-13 15:04:17.573  26.7003  -80.8866
2018-05-13 15:11:58.436  26.4919  -81.2643
2018-05-13 16:24:15.502  26.9545  -81.3624
2018-05-13 17:13:37.160  27.2630  -80.9225
2018-05-13 17:21:44.182  27.5171  -79.5095
2018-05-13 17:21:44.184  27.4380  -79.4189
2018-05-13 17:21:44.185  27.4546  -79.4752
2018-05-13 17:22:31.120  27.7732  -81.2090
```

Figure 4. Unedited Lightning Aloft Data for May 13, 2018

From Figure 4, the following data is collected: date, exact time, latitude, and longitude respectively. The MERLIN data collects a large portion of the lightning events within Florida, and therefore a sampling technique will be used to shrink the number of observations to make the data more manageable. An algorithm is displayed in Chapter III that explains the sampling method.

2.4.2 Cloud-to-Ground Data

For the cloud-to-ground lightning data collected using the MERLIN system, more information is collected compared to the lightning aloft data. Figure 5 is an example of one of the days collected for May 14, 2018.

2018-05-14	09:33:08.452	26.8168	-80.2478	-232.4	8.5	0.4	163	10	0.2	G
2018-05-14	10:38:02.683	27.1186	-80.4771	-43.3	8.4	0.4	168	6	0.2	G
2018-05-14	10:41:43.017	27.1573	-80.3147	-54.3	5.9	0.3	162	10	0.2	G
2018-05-14	10:45:01.351	27.0651	-80.3139	-109.3	6.5	0.4	163	9	0.3	G
2018-05-14	10:45:01.411	27.0418	-80.3040	-66.5	6.9	0.4	163	10	0.3	G
2018-05-14	10:45:01.536	27.0528	-80.3096	-57.9	7.0	0.4	163	10	0.1	G
2018-05-14	10:46:14.607	27.1042	-80.4161	-56.2	6.3	0.3	166	10	0.4	G
2018-05-14	10:46:14.680	27.0706	-80.4086	-64.8	6.5	0.4	166	10	0.3	G
2018-05-14	10:46:14.820	27.0884	-80.4121	-53.9	6.5	0.4	166	10	0.2	G
2018-05-14	10:48:40.977	27.0587	-80.3282	-33.1	9.6	0.5	163	6	0.1	G
2018-05-14	10:49:29.830	27.0562	-80.2868	-162.1	6.2	0.4	162	10	0.2	G
2018-05-14	10:51:10.491	27.0942	-80.2738	-77.7	6.3	0.4	161	10	0.2	G
2018-05-14	10:51:56.021	27.0994	-80.2827	-160.7	5.9	0.3	161	10	0.2	G
2018-05-14	10:51:56.096	27.0744	-80.2726	-73.9	6.6	0.4	161	10	0.1	G
2018-05-14	10:51:56.128	27.1147	-80.2900	-39.3	6.9	0.4	161	10	0.3	G
2018-05-14	10:53:01.167	27.0913	-80.2749	-114.5	6.1	0.4	161	10	0.4	G
2018-05-14	10:53:01.208	27.0921	-80.2762	-82.7	6.3	0.4	161	10	0.2	G
2018-05-14	10:53:01.370	27.0942	-80.3152	-38.3	9.6	0.5	162	5	0.4	G
2018-05-14	10:53:01.438	27.1201	-80.3249	-38.7	6.8	0.4	163	10	0.3	G
2018-05-14	10:53:01.687	27.1195	-80.3257	-33.7	7.9	0.4	162	8	0.1	G
2018-05-14	10:55:41.756	27.0856	-80.2881	-74.8	6.4	0.4	162	10	0.1	G
2018-05-14	11:07:45.712	27.1354	-80.3051	-47.9	6.6	0.4	162	9	0.2	G
2018-05-14	11:31:09.555	27.5948	-80.1880	+62.6	2.9	0.3	148	10	0.4	G
2018-05-14	11:33:11.559	27.6006	-80.2725	+43.1	2.8	0.2	151	10	1.3	C

Figure 5. Unedited Cloud-to-Ground Data for May 14, 2018

The first four columns are the same as lightning aloft data: date, exact time, latitude, and longitude. The next column is the stroke kiloamperes with polarity, which shows how much electrical current was produced and distributed for the lightning strike. The sixth and seventh columns are the semi-major and semi-minor axis of error ellipse measures using kilometers. The eighth column shows the error ellipse angle of orientation that coincides with columns six and seven for the lightning event.

The ninth column tracks the number of reporting sensors that tracked this lightning event. The last column is mostly "G" to illustrate it is a cloud-to-ground lightning event. Although a good amount of data is presented in the MERLIN data set, only the first four columns will be used in this thesis.

2.5 Clustering Method

Previous research in the study of creating a boundary around pre-existing storms have differing opinions on the best way to model the boundary of the storm. Sanderson (2019) fit the boundary of the storm by testing different ellipse fitting algorithms for LDAR II data. Hinkley (2019) fit the boundary of pre-existing storms by creating a convex hull algorithm which takes the outer most points of a storm, identifies them as vertices, and creates a boundary by fitting lines to each of the vertices. Tello (2021) used the an algorithm called Clustering of Online Data Streams (CODAS), an algorithm that cluster lightning events quickly based on the density of the data, as well as shapes depending on the number of dimensions. Holland (2021) uses an ellipse fitting algorithm to examine total lightning events from the MERLIN data set.

2.5.1 Ellipse Fitting Algorithm

Sanderson (2019) tested a few different ellipse fitting algorithms in her thesis and reported the results of each of the differing types. The first type of ellipse fitting algorithm hypothesized to work was the least squares best fit ellipse but was later ruled out due to the method trying to minimize the distance of every lightning event from the edge of the ellipse created. Instead, Sanderson decided the ellipse would seek to draw an ellipse around a defined percentage of lightning events, and therefore decided to investigate a principal component analysis ellipse fitting algorithm. After initial research and trials, Sanderson tries to salvage the principal component analysis ellipse fitting algorithm by using a convex hull of the extreme points, therefore allowing a select number of lightning events to establish the ellipse instead of the entire data set. In the end, many of the source points were still beyond the boundary of the ellipse, and therefore decided to use a different ellipse fitting algorithm, namely the minimum volume enclosing ellipsoid. This algorithm was the method chosen in her research, and this choice provided accurate representations of clusters of storms useful in her thesis for calculating the distance between new lightning events and the boundary of a pre-existing storm. Although an ellipse fitting algorithm is not used in the analysis of this thesis, Sanderson's work and histograms were influential in the writing of this paper.

2.6 Goodness of Fit Tests

The use of goodness of fit tests is crucial in statistics due to the nature of ensuring a certain distribution is correct for the data collected. Different distributions will be examined in Chapter IV, and a number of statistical tests will be used in determining the correct type of distribution is selected, namely the probability – probability (P-P) plot, the quantile – quantile (Q-Q) plot, as well as examining how well the empirical and theoretical distributions fit for the cumulative density function as well as the probability density function.

With the large sample of data points being evaluated in this thesis, the goodness of fit tests being examined may suffer from some lack of fit. Johnson and Wichern in their textbook "Applied Multivariate Statistical Analysis" state that "very large samples invariably produce statistically significant lack of fit. Yet the departure from the specified distributions may be very small and technically unimportant to the inferential conclusion" (Johnson and Wichern, 1992). This means that some of the fits may not be the best; however, the features of the theoretical distributions are important.

2.6.1 P-P Plot

The P-P plot is used in determining how well the cumulative density functions for the empirical and theoretical distributions coincide in an easy-to-understand way. The plot consists of points and a straight line, where the points are the difference between the distributions and the straight diagonal line is the theoretical distribution. If the majority of the points lie on the diagonal line, then the distribution fits well in terms of their cumulative density functions. If most of the points do not lie on the diagonal line, then another distribution should be examined, or the data set should be examined for outliers. Figure 6 illustrates this summary by giving two examples, one P-P plot that shows the data fits the selected distribution correctly and one P-P plot that shows the distribution used is incorrect and should be reexamined. Both graphs were created using R and creating random numbers using distributions.

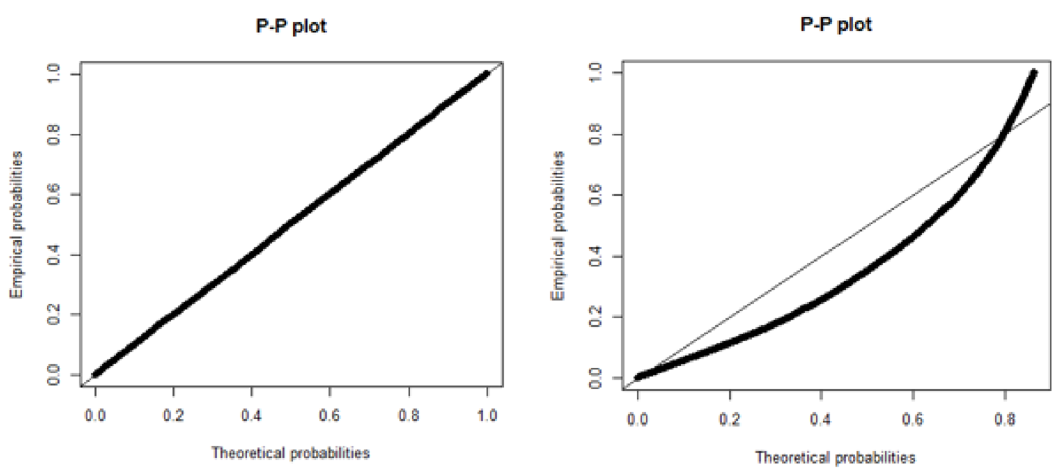

Figure 6. P-P Plot Examples

The left image shows the distribution selected fits the data correctly since the points lie on the diagonal line. The image on the right illustrates an incorrect distribution selected. Another similar plot to define model adequacy is called the Q-Q plot.

2.6.2 Q-Q Plot

The Q-Q plot is used to show how well the quantiles of the data compare to the theoretical distribution created. Both the P-P plot and the Q-Q plot seek to illustrate how well the data and the distributions fit by having similar types of graphs, since the Q-Q plot also has a straight diagonal line and a set of points that should lie on the diagonal line. If the majority of points line on the line, then the selected distribution fits well. If the majority of points do not lie on the line, then another distribution should be examined. Figure 7 shows two examples of Q-Q plots, with the one example fitting the selected distribution well and one example showing an ill-fitting distribution.

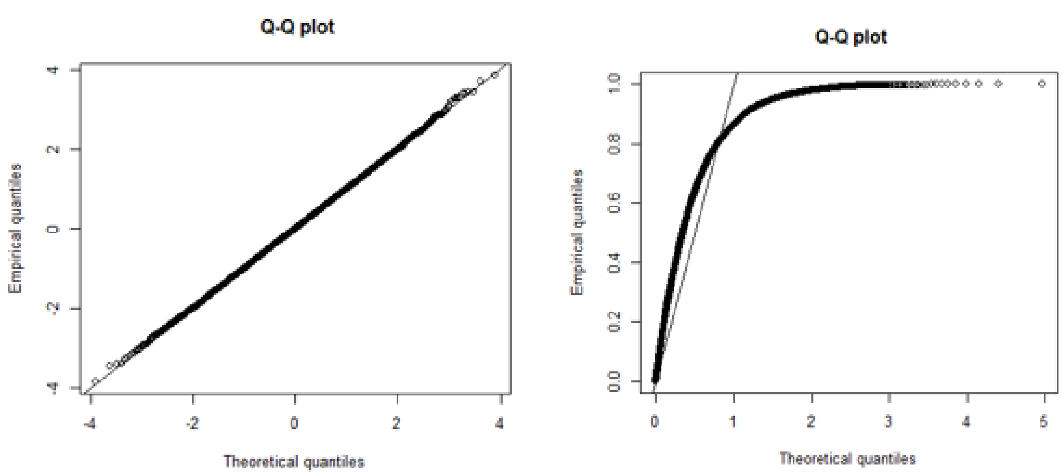

Figure 7. Q-Q Plot Examples

As shown in the left figure, the vast majority of the points lie on the diagonal line, and therefore using the theoretical distribution would accurately explain the histogram. Meanwhile, the graph on the right shows that most of the points do not lie on the diagonal line, and therefore using the theoretical distribution would not accurately describe the histogram being tested.

2.6.3 Empirical vs. Theoretical Density

The empirical vs theoretical density graph displays the histogram created by the data as well as the theoretical distribution being evaluated. This goodness of fit test is helpful for gauging what type of distribution would best fit the shape of the histogram. One drawback is that some distributions look similar, and therefore this goodness of fit test should be used in tandem with other goodness of fit tests, such as the P-P plot as well as the Q-Q plot. Figure 8 gives an example of two empirical vs theoretical density graphs. One graph shows a histogram that is normally distributed along with a theoretical distribution that is also normally distributed, while the other graph shows a uniform histogram that is trying to be explained using an exponential distribution.

The top plot in Figure 8 shows an ill-fitting theoretical exponential distribution to a uniform histogram while the bottom plot shows a well-fitting theoretical distribution to the normally distributed histogram. The purpose of this graph is to give insight on the underlying histogram and distribution while visualizing what the proposed theoretical distribution would look like on the histogram.

2.6.4 Empirical vs. Theoretical CDF

This last plot looks at the cumulative density function using the theoretical distribution and applying it to the cumulative density function of the histogram created by the data. This graph, in conjunction with the other three plots adds another layer of understanding how well the cumulative density functions align. This is a helpful plot considering that using the cumulative density function will be helpful in

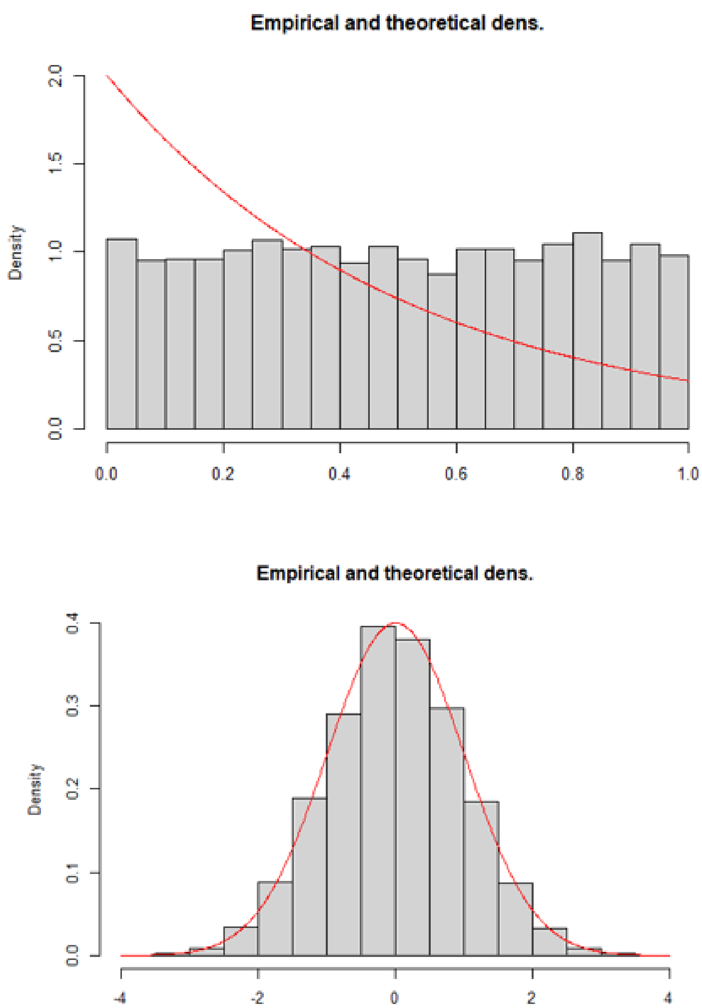

Figure 8. Empirical vs. Theoretical Density Examples

examining how many lightning events occur between certain distances. The goal is for the red line and the points to align. Figure 9 shows two plots for the empirical vs theoretical plots for the cumulative density functions, with one of the plots fitting the distribution well while the other plot does not fit the distribution well.

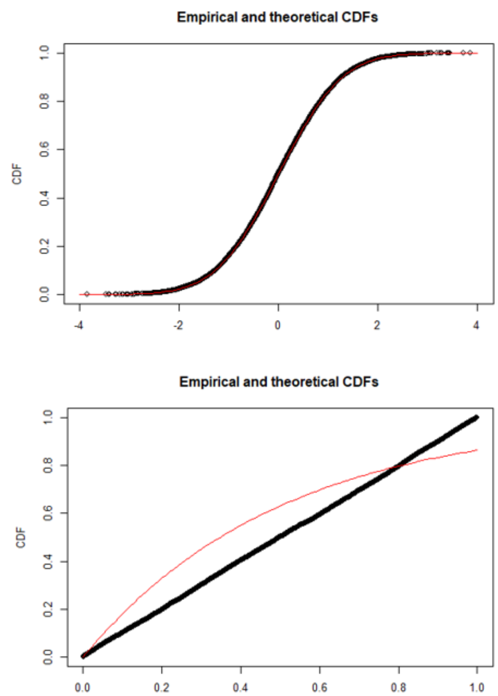

Figure 9. Empirical vs. Theoretical Cumulative Density Function Examples

2.7 Summary

Lightning aloft lightning events stay suspended in the air and do not contact the surface of the planet while cloud-to-ground lightning events make contact with the ground, usually through tall structures or objects. Lightning events are not the same throughout the year, certain events and environments can cause changes to lightning events, such as temperature, wind, and atmospheric instability. On average, these changes usually occur over different seasons on the eastern coast of Florida, with the warmer season, between May and September, having more frequent thunderstorms while the colder season having less frequent lightning events.

With the inclusion of lightning aloft events, goodness of fit tests are utilized to gain an understanding of how far new lightning events occur beyond the border of pre-existing storms, as well as to see if this inclusion to the total lightning data set would provide a more accurate representation of the whereabouts of pre-existing storms in the model compared to relying solely on cloud-to-ground lightning events. Although the program utilized in this research relies heavily on the convex hull, more information about the program and how the new algorithm is written is explained within Chapter III, along with a detailed explanation of the investigation conducted.

III. Modeling and Solution Methodology

3.1 Overview

In this chapter an algorithm is established to create a boundary for lightning events. With the new algorithm, a histogram is created for distances from the new lightning event to the boundary of pre-existing storms. The purpose of the histogram is to see if including lightning aloft into the data set will have more lightning events closer to 0 rather than 5 nautical miles compared to only using cloud-to-ground lightning events. Once the histograms are established, goodness of fit tests will be conducted to see which types of distributions fit the data better. An investigation will then be used to test the safety of certain warning radii for using total lightning or solely cloud-to-ground lightning events.

3.2 Data

The MERLIN system collects two different types of data: cloud-to-ground lightning events and lightning aloft events. Cloud-to-ground, as the name implies, tracks lightning events where a lightning strike makes contact with the Earth. Lightning aloft tracks lightning events that occur but stay in the Earth's atmosphere. From these two data sets a third set can be created called Total Lightning which combines both types of lightning events into one data set. The positions of both types of lightning events are tracked using latitude and longitude, as well as the date and time of the event. Although the latitude and longitude of the events are recorded, the altitude of the events are not logged in the system. This issue is addressed in the distance section of this chapter.

Although the MERLIN system covers most of the state of Florida, not all of the data is important in this investigation. Sanderson (2019) considered data points that are within 25 nautical miles from the points of interest. In the thesis, the lightning events within 25 nautical miles from the points of interest will be used create the histogram of the distance between new lightning events and the boundary of a storm.

3.3 Shape of the Storm

Initially, ellipses were used to estimate shape and length of the storm with more than five lightning events. The goal of the ellipse was to capture all of the lightning events that occurred for a particular storm while minimizing the area of the ellipse so that an accurate picture of the dimension of the storm can be drawn. Lightning areas can take different shapes and polygons can be better suited to estimate the perimeter of a storm rather than an ellipse (Hinkley et al., 2019). Therefore, it could be beneficial for the perimeter of the storm change shape depending on the lightning events within the given cluster for the lightning area.

Convex hulls aim to achieve this goal by creating the perimeter of a cluster by using the outermost points in the system. An example of this is shown in Figure 10 for a particular storm in August of 2018.

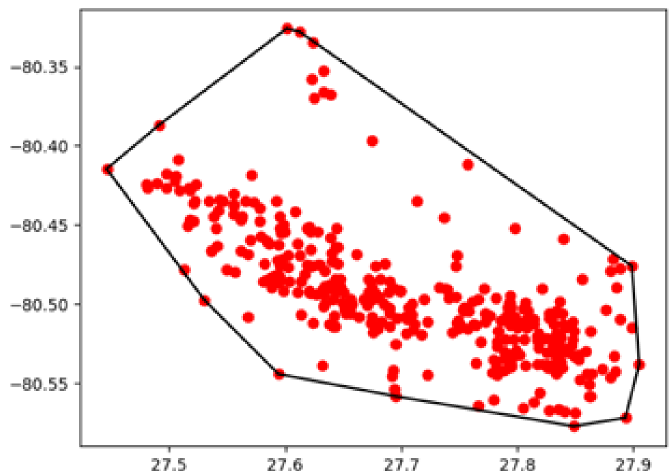

Figure 10. Convex Hull Example

In this example, the perimeter of the storm is clearly defined and is quickly calculated using the convex hull function within the SciPy library. Distances from the perimeter to the new lightning event are estimated using points along the convex hull that are 0.1 nautical miles away from one another. The reason for having points that are 0.1 nautical miles apart is to get a better estimate of the distance from the new lightning event to the boundary of the storm. There will also be a buffer zone of 0.1 nautical miles around the boundary, such that if a lightning event occurs next to the boundary and is less than 0.1 nautical miles from the storm, then it is counted as inside of the storm. This brings into question how distances are calculated in this model.

3.4 Distances

Although one can use the Cartesian plane to model lightning events with the x-axis being the latitude and the y-axis being the longitude, the distances between these events should take into account the curvature of the region. This is because the latitude and longitude represent points on a sphere rather than a two-dimensional plane. To overcome this issue, the haversine function is used to calculate distances between lightning events.

The haversine formula calculates the great-circle distance between points on a sphere using the latitude and longitude of the events, along with the radius of the planet. The great-circle distance is the shortest distance between points on the surface of a sphere. Although Earth is not a perfect sphere, the great-circle distance is close to perfect and can calculate the distance between points on Earth with an error of around 0.5% (Admiralty Manual of Navigation, Volume 1). Now that we understand the shape, the lightning events used in the system and the method for calculating distances between events, now we must address how storms are classified and how lighting events are added to a storm.

Certain cutoff distances are used in the algorithms, namely a 5 nautical mile cutoff as well as a 10 nautical mile limit. The 5 nautical mile cutoff is for the distances from new lightning events to new emerging storms, with a new lightning event being within 5 nautical miles from a cluster of lightning events with less than 5 lightning events to the cluster. This assumption in the algorithm comes from Tello's research on emerging storms, with roughly 3.24% of lightning distances in developing storms exceed 5 nautical miles (Tello, 2021).

The 10 nautical mile limit refers to developed storms, such that a new lightning event could be clustered with a developed storms if the lightning strike is within 10 nautical miles from the boundary of the lightning area. The 10 nautical mile limit comes from the Lightning Launch Commit Criteria, which has a 10 nautical mile threshold (Roeder and Saul, 2016). Not only that, but Holland (2021) found that lightning aloft events only extend to 9 nautical miles, with cloud-to-ground events having a shorter reach (Holland, 2021).

3.5 Sampling Method

When we initially looked at the problem, we wanted to keep as many points as possible but still be able to conduct a preliminary investigation in a meaningful amount of time. For that, we decided to sample the first hour of each storm from 2018 through 2020 for the MERLIN data set. The meaning behind this decision is because after the first hour, most of the distances from the boundary to new lightning events are 0 since most of the lightning events are within a pre-existing storm. Therefore, to minimize the number of samples while still capturing all of the days throughout the years, taking the first hour proved helpful since it took computation time from 10 days to 5 days. Similar to Sanderson (2019), we decided to only include lightning events within 25 nautical miles from the central node for LDAR II. The MERLIN data set is very expansive and includes many lightning events across Florida, but to shorten the data set to conduct the investigation, only lightning events within 25 nautical miles are included to test 8 important locations around the central node.

3.6 Clustering Method

Initially, a prebuilt machine learning clustering algorithm was thought to be the solution to the issue of clustering lighting events into storms. The issue with these methods is that these machine learning algorithms rely on distance between points; and considering that the distance between points in this model are on a sphere instead of a two-dimensional plane, we had to make our own clustering algorithm. The python file created is given in Appendix A.

3.6.1 The Start of a Storm

Hinkley (2019) who uses an algorithm where they used the convex hull to generate polygons to define a pre-existing lightning area using LDAR data. In the paper, Hinkley used lightning strikes within 15 minutes from the current lightning event, a similar idea that will be implemented in this analysis. In the new system, lightning events within 10 minutes from the current lightning strike will be considered for the boundary created by only cloud-to-ground lightning event, and 2 minutes from the current lightning strike for the boundary created by total lightning events. The reason for the decrease in the amount of time for the boundary is for computational reasons, since using 15 minutes for both data sets would take weeks to compute for similar results.

The program restarts and starts calculating the distances after a few options occur: 30 minutes have passed with no lightning events occurring within the 25 nautical mile zone after the first hour has passed, 10 or 2 minutes have passed with no lightning events occurring within the operational zone before the first hour has passed.

3.6.2 Distance from Boundary Algorithm

In the program, there are two types of distances being calculated: convex and nonconvex distances. The distances calculated for convex has gone through a number of iterations, but the pseudo code can be found in Algorithm 1. It is the set of distances calculated from the boundary of the nearby storm with 5 or more lightning events having occurred within the set time limit given before, either 10 or 2 minutes. The new lightning event finds a nearby storm with 5 or more lightning events, finds the distance from the new lightning events to the vertices of the convex hull, and finds the vertex with the minimum distance. The program then creates a line from that vertex to the neighboring vertices, and the new lightning event will then find the minimum distance along the new lines drawn.

Once the minimum distance is found, the algorithm stores the distance along with the cluster number, and then tests the other pre-existing storms in the simulation also within the 10- or 2-minute time frame. To check if a lightning event is within the boundaries of a pre-existing storm, another convex hull is calculated, this time with the new lightning event within the pre-existing storm. If the new lightning event is a vertex, then the new lightning event is outside of the pre-existing boundary. If the new lightning event is not a vertex, then the new lightning event is within the boundary and therefore the distance is no longer calculated. Once a minimum distance is found, or if the new lightning event is within the boundary of a pre-existing storm, then the new lightning event is added to the pre-existing storm with the minimum distance and the distances, if needed, are recorded.

Algorithm 1 Convex Hull Pseudocode

```
Require: time ≤ 600 or 120 seconds            ▷ time interval depends on CG or TL
Ensure: len(Cluster) ≥ 5
  for Cluster with more than 5 Lightning Events do
      Include New Lightning Event into Cluster
      if New Lightning Event is in Convex Hull then
          distance = -1
      else if New Lightning Event is not in Convex Hull then
          Take Convex Hull of Cluster except the New Lightning Event
          Calculate Distances from New Lightning Event to each Vertex
          Find min(distances)
          Connect Min Distance Vertex with Neighboring Vertices  ▷ Creates 2 Lines
          Create Points Along Line at most 0.1 N.M apart
          Find Distances From Points to New Lightning Event
          Store min(distance) from the points, along with cluster id
```

Once all of the clusters are evaluated, the minimum from all of the clusters are recorded and compared with the minimum distance from clusters with less than 5 lightning events. For clusters with less than 5 lightning events, the nonconvex distances are calculated. Algorithm 2 finds the distance from the new lightning event to pre-existing lightning events in clusters with less than 5 lightning events using the haversine function as well. Minimum distances and cluster id are recorded. The algorithm does this for all storms and clusters with less than 5 lightning events and records all distances.

Algorithm 2 Nonconvex Pseudocode

Require: $time \leq 600$ or 120 seconds ▷ time interval depends on CG or TL
Ensure: $len(\text{Cluster}) < 5$
  for $Cluster$ with less than 5 Lightning Events **do**
    Find Distances from each Lightning Event in Cluster to New Lightning Event
    Record $min(distance)$ along with cluster id

Once all of the minimum distances are calculated, minimum distance will compare with the minimum convex hull distance. If the minimum distance for the convex hull is greater than 10 nautical miles and the minimum distance for the nonconvex distances is greater than 5 nautical miles, then the new lightning event will have its own unique cluster id and will begin looking for new lightning events close to it. If the distance from the new lightning event to the convex hull is less than 10 nautical miles and the minimum distance for the nonconvex distances is greater than 5 nautical miles, then the new lightning event will be added to the cluster with the minimum convex hull distance.

If the minimum distance from the new lightning event to the convex hull is less than 10 nautical miles and the minimum distance from the new lightning event to the nonconvex distance is less than 5 nautical miles, then the new lightning event will be added to the cluster with the minimum distance. This cycle will continue until 1 hour has passed since the beginning of the thunderstorm, or if a 10 minute break occurs with no lightning events occurring before the hour has passed. If a 10 minute break occurs, then the algorithm will start again and stop once a full hour occurs. After a full hour occurs, distances are no longer recorded and will begin recording again once 30 minutes have passed without a lightning event occurring.

3.6.3 Validation of the Algorithm

Sanderson (2019) found that a Weibull distributed was the best fitting distribution for the distance from the boundary to new lightning events. Although Sanderson used an ellipse to create and fit the boundary of the cluster, the Weibull distribution was also the best fitting distribution when using the convex hull to fit the boundary of the storm. Figure 11 shows the histogram for the distance from the boundary of a pre-existing storm to new lightning events. In the figure, one of the histograms uses only cloud-to-ground lightning events to create the boundary of the storm and the other histogram is created by using both cloud-to-ground and cloud-to-cloud lightning events to create the boundary.

Figure 11 are examples of the goodness of fit tests to show how well the plots fit the distributions.

Figure 11 illustrates different fitness tests, such as the P-P Plots and comparing the Empirical and Theoretical CDF plots for the different histograms in 2018. Both P-P Plots and Empirical and Theoretical CDF plots show that a Weibull distribution fits the data well, even though they have different shape and scale values. This is on par with Sanderson's findings, and after comparing the different histograms with the gamma, exponential and Weibull distribution, the Weibull seemed to perform the best for all of the different histograms.

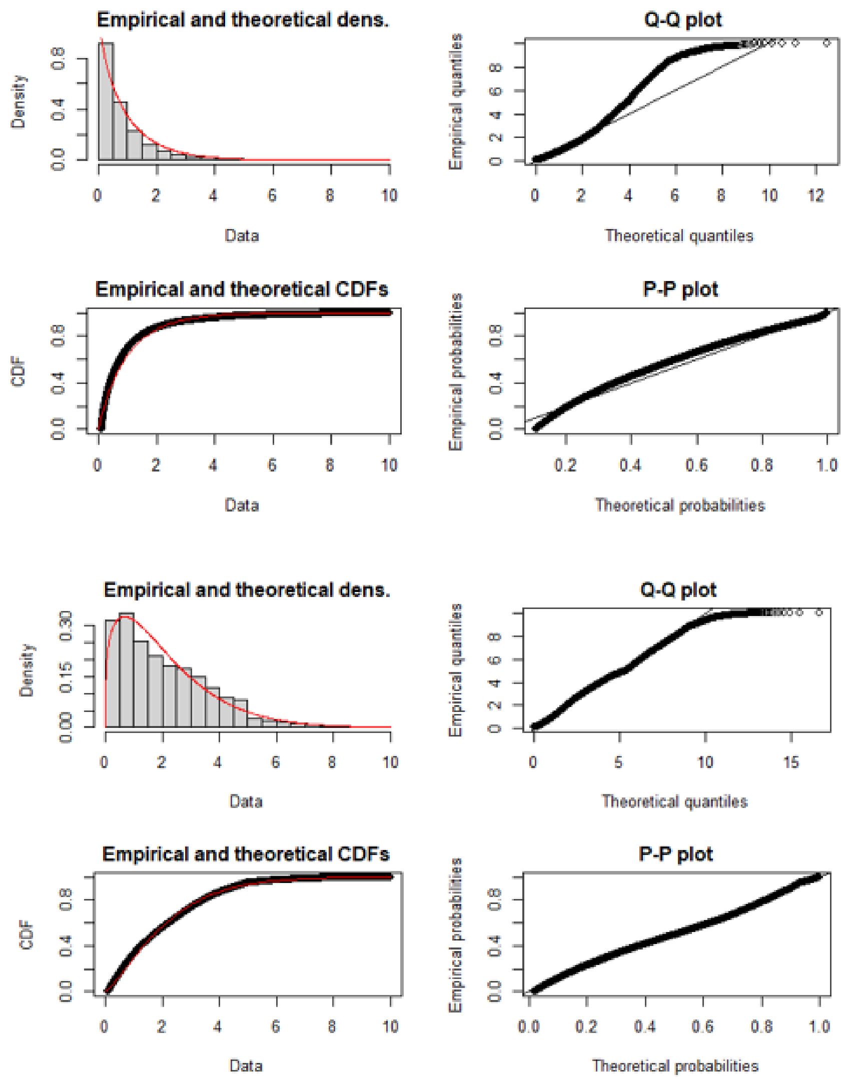

Figure 11. Example of Goodness of Fit Analysis: 2018 Total Lightning and Cloud-to-Ground

3.7 Comparing the Different Types of Boundaries

Similar to Sanderson (2019) and her experiment conducted using LDAR II, the experiment for the MERLIN system uses a set of important locations, all with a lightning warning radius as well as a 1 nautical mile radius around the center of the important locations. The 1 nautical mile radius is to establish a critical area that needs to be safeguarded from lightning strikes, such as a launch site or a space command center. The goal is to see if using both cloud-to-cloud and cloud-to-ground lightning events will produce a more accurate representation of the whereabouts of pre-existing storms compared to only using cloud-to-ground lightning events. Once a lightning event occurs within the warning radius around the place of interest, the location becomes offline and safe until the storm passes. If the 1 nautical mile radius is struck by a lightning event before the area is offline, then it is counted as a failure. Not only is this to test the accuracy and precision of using both types of lightning events, but it is also to see if the warning radius, initially at 5 nautical miles, can be shortened while keeping personnel safe.

The places of interest are launch sites, bases, and facilities in eastern Florida within the 25 nautical mile radius mentioned earlier. Table 1 gives the names and locations of these places of interest, as well as the acronym used within the python program. All locations are found using google maps.

Table 1. Places of Interest for Experiment

Place of Interest	Acronym	Latitude	Longitude
Launch Site 39	LC39	28.573469	-80.651070
Launch Site 40	LC40	28.562091	-80.577385
Launch Site 37	LC37	28.531212	-80.564782
Space-X Central Launch Facility	Spx_central	28.417199	-80.604741
Astrotech Space Operations	Astrotech	28.5256	-80.8209
Patrick Space Force Base	Psfb	28.234332	-80.605998
Kennedy Space Launch Facility	Slf	28.59238	-80.66091
Florida Institute of Technology	IA	28.092399	-80.641466

The investigation will look at a few different metrics: number of base failures, the warning radius used in the investigation, as well as the different seasons and years examined. Different warning radii will be tested to see if the change will keep personnel and the facilities safe.

3.8 Summary

Including cloud-to-cloud lightning events along with the cloud-to-ground events could give a more accurate boundary for storm, as seen in Figure 11. Using the convex hull to create the boundary of pre-existing storms, the histograms in Figure 11 are drawn. From Figure 11, a few distributions, such as a Weibull distribution, could be used to fit the histogram, which concurs with previous research from Hinkley (2019) and Sanderson (2019). With all of this information, an investigation is conducted to see if using both cloud-to-ground and cloud-to-cloud lightning events to create the boundary is safer than only using cloud-to-ground lightning events.

IV. Testing, Results, and Analysis

4.1 Overview

An analysis is conducted to examine the different distributions and histograms created when measuring the distances from boundaries of storms created by using solely cloud-to-ground lightning events as well as total lightning events. From there, the empirical cumulative density functions for the different algorithms are used to find the number of lightning strikes beyond a certain threshold, such as 5 nautical miles for the different algorithms. An investigation is then conducted to examine the number of lightning aloft and cloud-to-ground failures occurred when using solely cloud-to-ground lightning events compared to the number of failures when using total lightning to establish a warning radius.

4.2 Distance Distributions

From 2017 through 2020, the first hour of uninterrupted lightning storms are recorded with the boundary of preexisting storms being created by cloud-to-ground lightning events or total lightning events. Then, both types of boundaries are evaluated to see how far new lightning events, both cloud-to-ground and lightning aloft, are from the boundaries. The initial goal is to see how the global histograms compare when using only cloud-to-ground lightning to create the boundary compared to using total lightning events to create the boundary of the storm. Figures 12 and 13 illustrate the difference in terms of the distance from boundary to new lightning events.

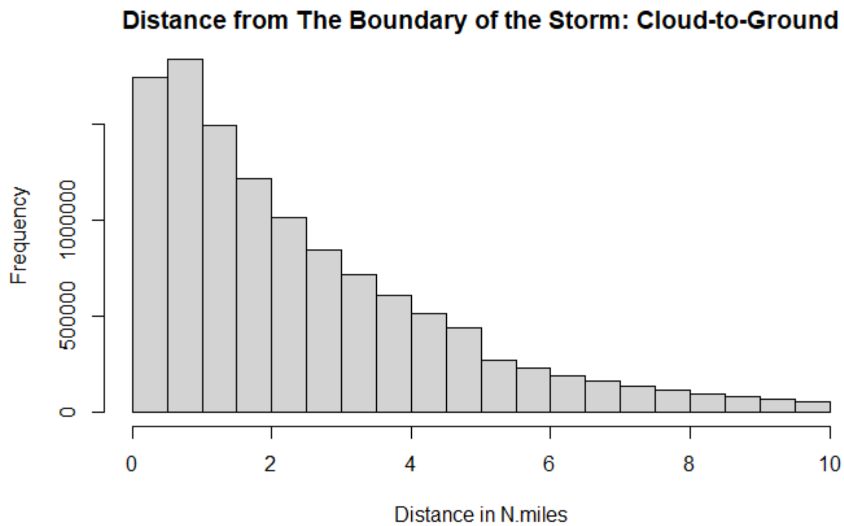

Figure 12. Global Histogram for Distance from the Boundary of the Storm: Cloud-to-Ground

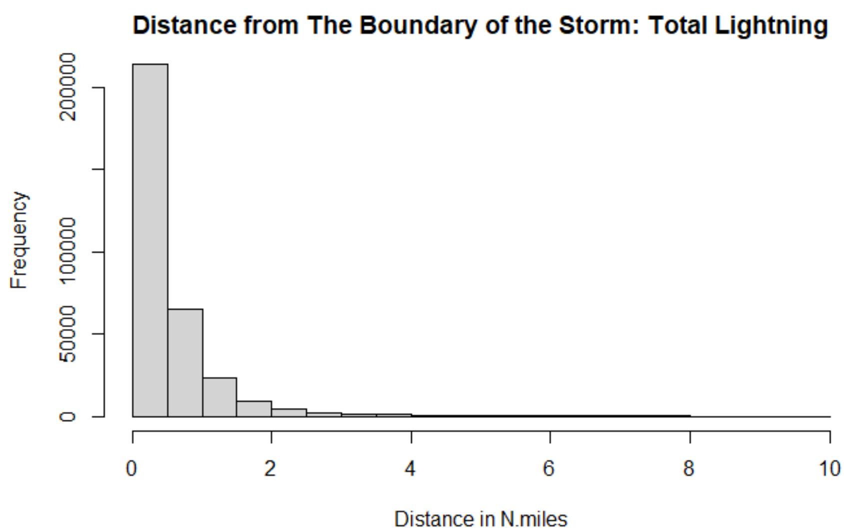

Figure 13. Global Histogram for Distance from the Boundary of the Storm: Total Lightning

As shown in figures 12 and 13, the distances from the boundary to new lightning events when only using cloud-to-ground lightning events to establish a boundary have a much wider distribution compared to the algorithm that uses total lightning to create a boundary. When solely using cloud-to-ground to establish the boundary, the distances from the boundary to new lightning events stretch the entire 10 nautical mile boundary compared to using total lightning to create the boundary. In figure 14, much fewer lightning events occur beyond 3 nautical miles, and practically no lightning events occur beyond 5 nautical miles. This means that total lightning gives a more accurate representation of where pre-existing storms are currently located in the model compared to using solely cloud-to-ground lightning events. Figures 12 and 13 use the first hour of every storm from 2017 through 2020.

4.2.1 Distribution for the Global Models

After looking at Figures 12 and 13, it is important to ask which types of distributions might suit the histogram. Three distributions come to mind: the gamma distribution, the exponential distribution, and the Weibull distribution.

The histogram in Figure 12 will look at two of the distributions, namely the gamma and Weibull distribution. With the initial hump between 0 and 1 nautical miles, the exponential distribution would not describe the model accurately. Figures 14 and 15 show the statistical tests applied to the histogram in Figure 12.

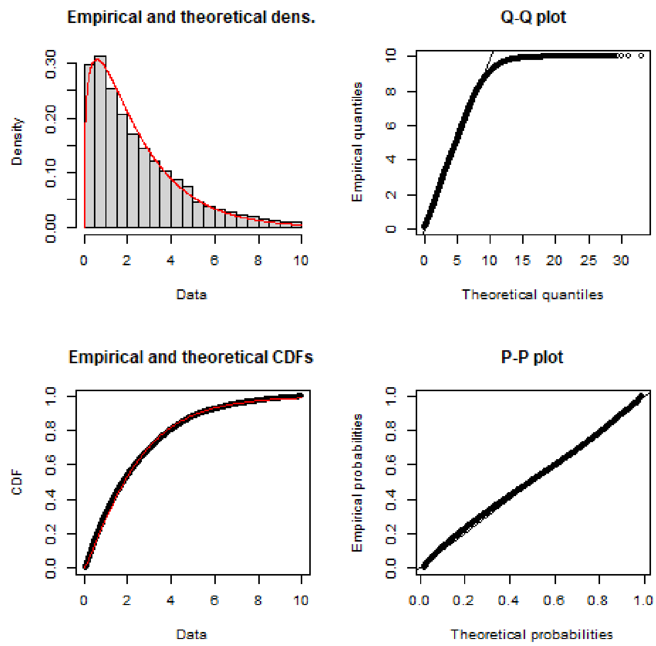

Figure 14. Goodness of Fit Tests for Gamma Distribution of Global Cloud-to-Ground Algorithm

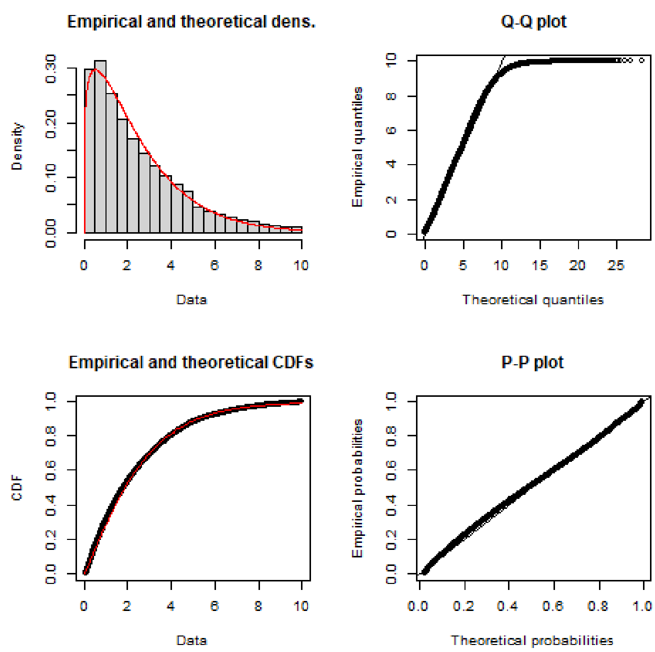

Figure 15. Goodness of Fit Tests for Weibull Distribution of Global Cloud-to-Ground Algorithm

Looking at the P-P plot for both Goodness of Fit Tests, both distributions appear to model the histogram well for the most part. Towards the tail end of the model, the distribution is not perfect, as shown in the Q-Q plot after the 10th quantile for both distributions. However, the vast majority of the points lie on the diagonal line for both the P-P plots and the Q-Q plots, which entails a good fit, along with the empirical vs theoretical cumulative density function graphs for both distributions. Either distribution would suffice to describe the data, but to retain continuity with Sanderson (2019) a Weibull distribution would be preferred.

As for the total lightning distributions, all three distributions would need to be evaluated, namely the gamma, exponential and Weibull distribution. The steep fall in distance from the boundary to new lightning events between 0 and 1 nautical mile could be best explained with an exponential distribution, and thus this distribution will need to be examined. Figures 16, 17 and 18 contain the 4 Goodness of Fit tests for each distribution evaluated.

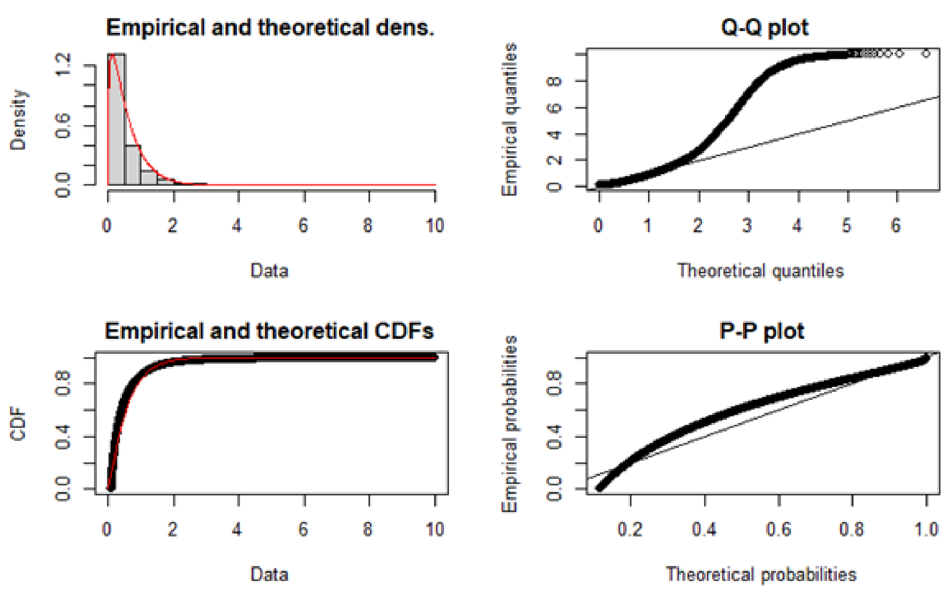

Figure 16. Goodness of Fit Tests for Gamma Distribution of Global Total Lightning Algorithm

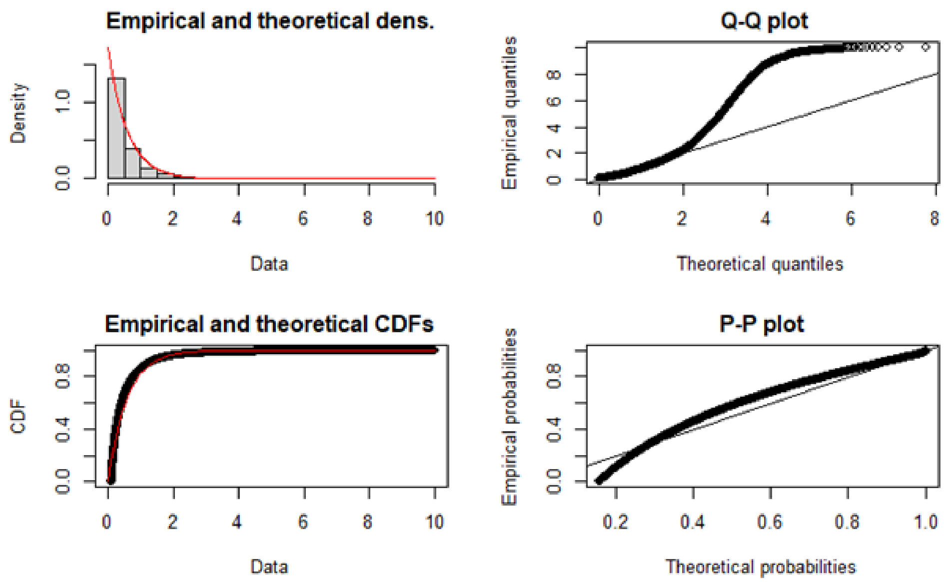

Figure 17. Goodness of Fit Tests for Exponential Distribution of Global Total Lightning Algorithm

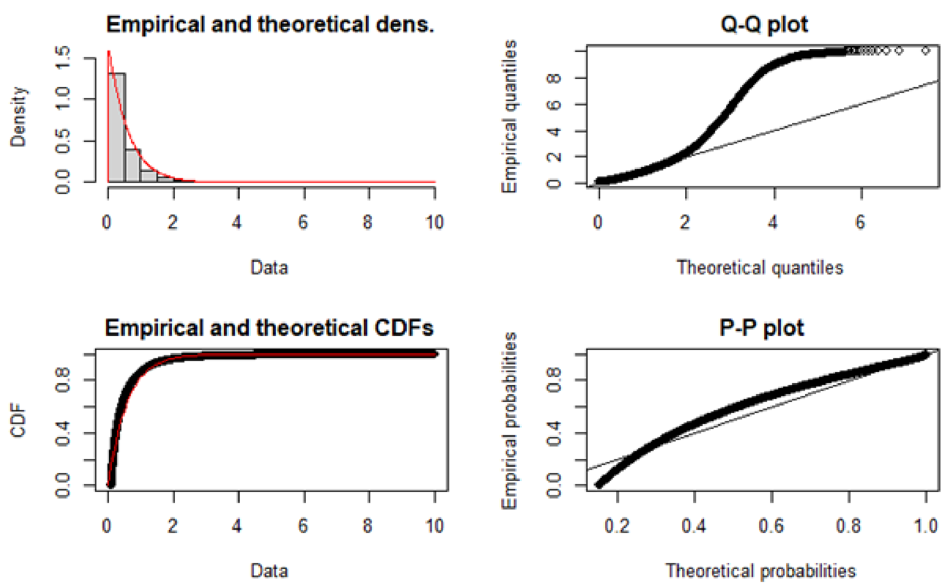

Figure 18. Goodness of Fit Tests for Weibull Distribution of Global Total Lightning Algorithm

Focusing on the Q-Q plots for all of the distributions shown in Figures 16 through 18, it appears that all of the distributions do not fit the histogram well after the 2nd quantile, which is understandable due to there not being many lightning events beyond 3 nautical miles, but still having some lightning events up to 8 nautical miles. Therefore, for this analysis, the Q-Q plot will be ignored after the 2nd quantile for all of the distributions since all of the distribution explain the model well for the most part. With regards to the P-P plots, all three distributions fit the data well, but the exponential distribution fits the data the best, with there being less points off the diagonal line in the middle of the plot, with the Weibull distribution being a close second. For the empirical vs theoretical graphs for the density and cumulative density function, both explain the data well, with the Weibull having an initial hump between 0 and 0.5 nautical miles. Although the distributions explain the global or entire data set well, the seasonal difference between the hot and cold months need to be examined.

One noted characteristic from the Q-Q plots is that it diverges after the second quantile and is above the line. This means that skewness is present within the plot. Using the theoretical distributions generated from this study by itself may not tell the full story, and thus it is beneficial to present other goodness of fit tests along side the Q-Q plots.

4.2.2 Yearly Changes

It is important to see how the years compare, to see if there are any important features that occur every year. Figures 19 through 22 illustrate the histogram for every year starting from 2017 through 2020 when using Cloud-to-Ground lightning events to establish a boundary.

For the cloud-to-ground algorithm, a noticeable change occurs at 5 nautical miles, where there is a sudden drop off for year 2018. Although a similar shape occurs every year where there is a large amount of lightning events between 0 and 1 nautical miles and then it begins to drop in an exponential fashion. Although these graphs are for

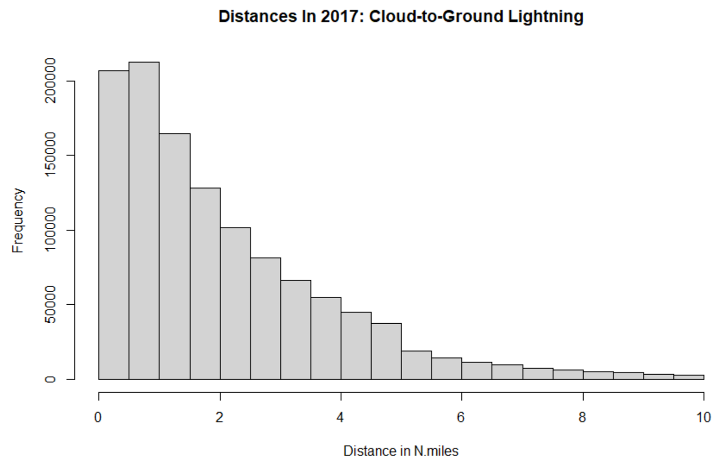

Figure 19. Histogram for 2017 Using Cloud-to-Ground

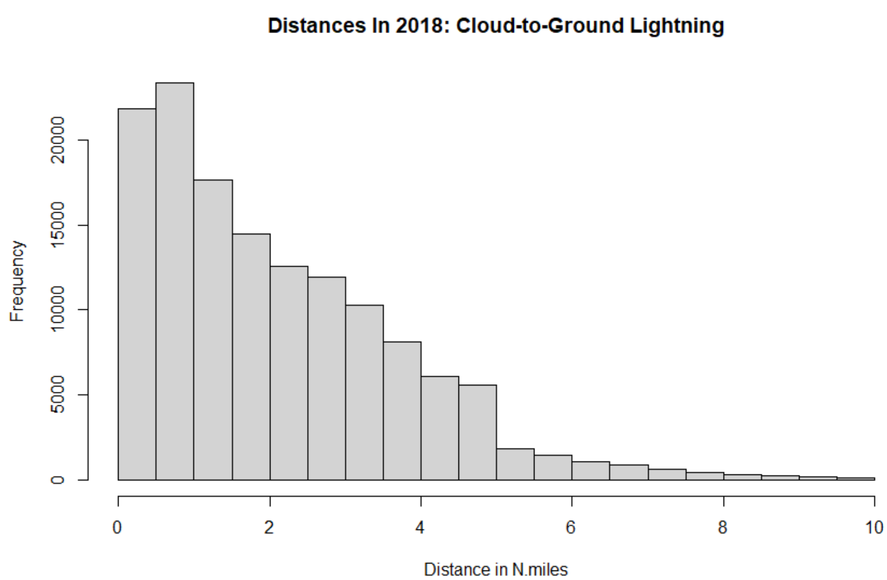

Figure 20. Histogram for 2018 Using Cloud-to-Ground

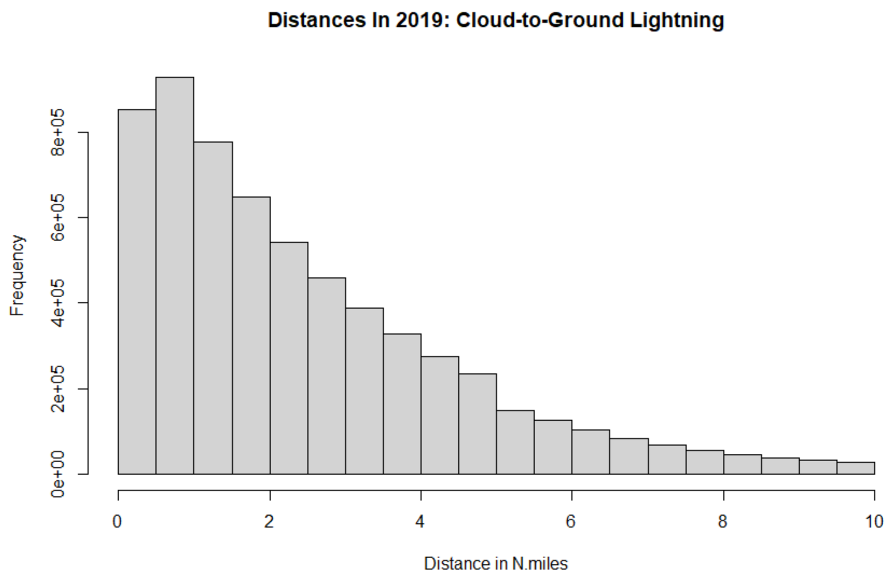

Figure 21. Histogram for 2019 Using Cloud-to-Ground

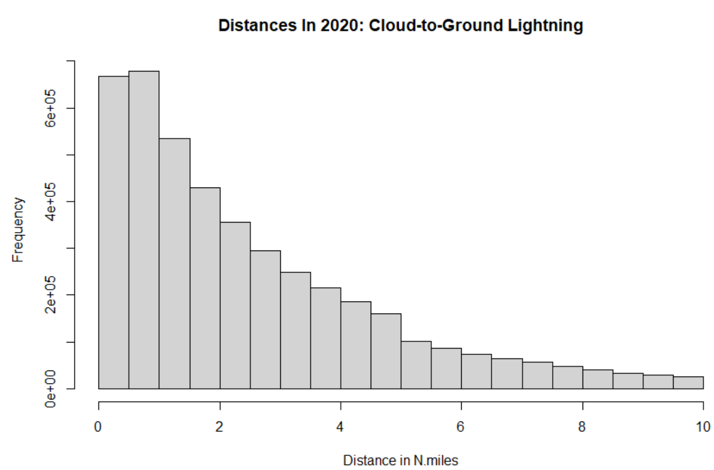

Figure 22. Histogram for 2020 Using Cloud-to-Ground

cloud-to-ground, it is also important to observe the yearly changes for total lightning.

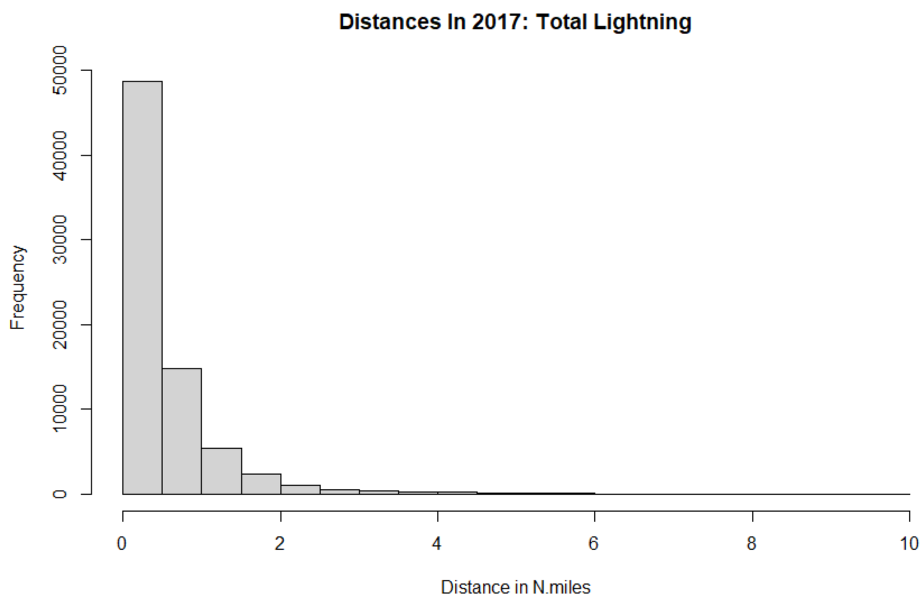

Figure 23. Histogram for 2017 Using Total Lightning

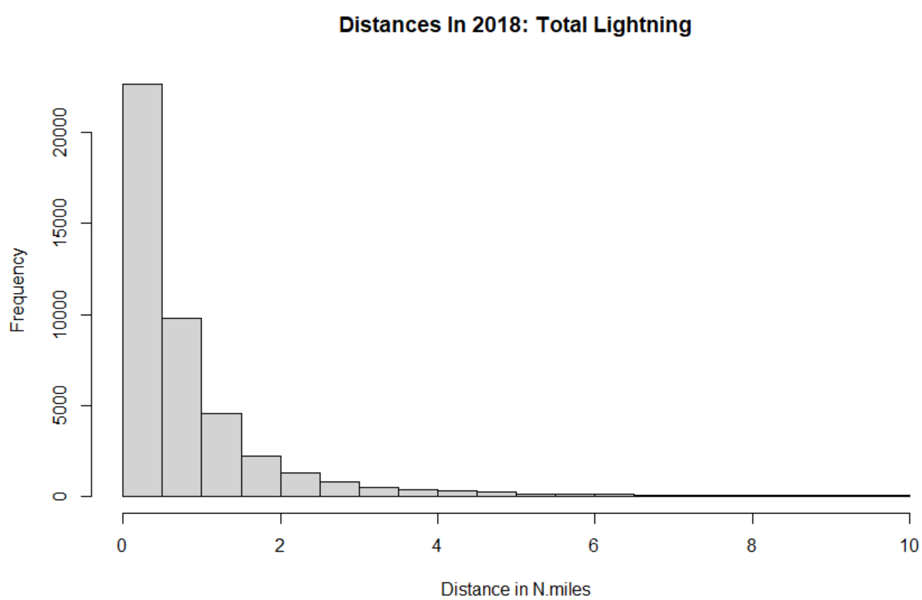

Figure 24. Histogram for 2018 Using Total Lightning

For the utilization of both lightning aloft and cloud-to-ground lightning events, the vast majority of lightning events occur before 5 nautical miles, with very little

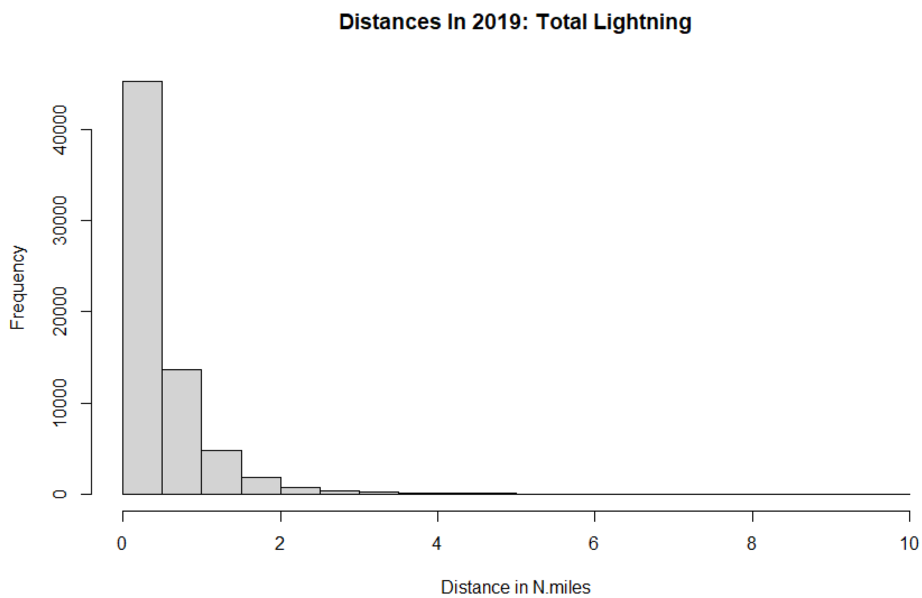

Figure 25. Histogram for 2019 Using Total Lightning

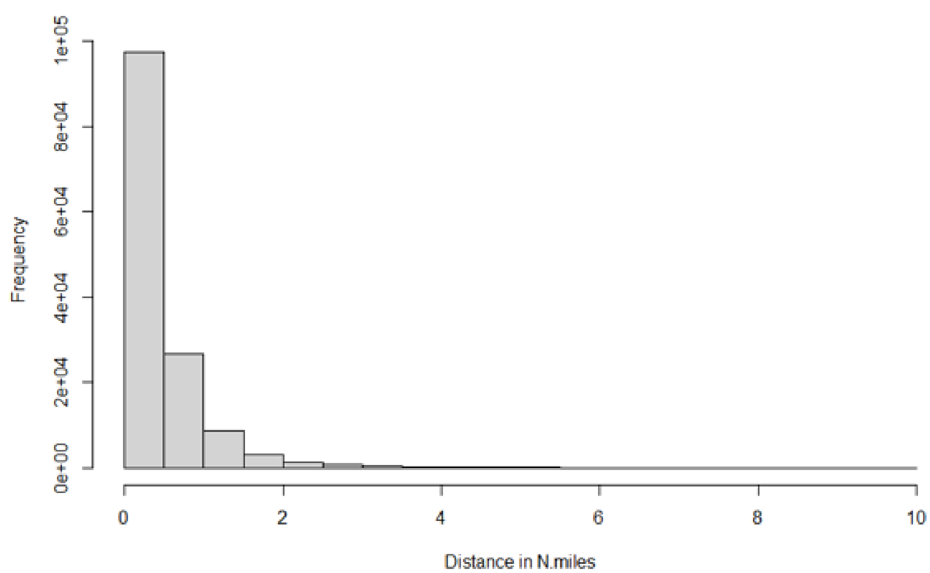

Figure 26. Histogram for 2020 Using Total Lightning

lightning events occurring beyond 4 nautical miles every year. This signifies that total lightning gives a more accurate representation for generating a lightning boundary.

4.2.3 Season Plots: Total Lightning

The warm months are between the months of May and September, as stated earlier, and therefore all lightning events between these months will be compared to its counterparts with regards to the types of lightning events to create the boundary. In other words, Figures 27 and 28 will be split into two separate histograms, one with lightning events having occurred between the months of May and September, and the other with lightning events having occurred outside of the warm months.

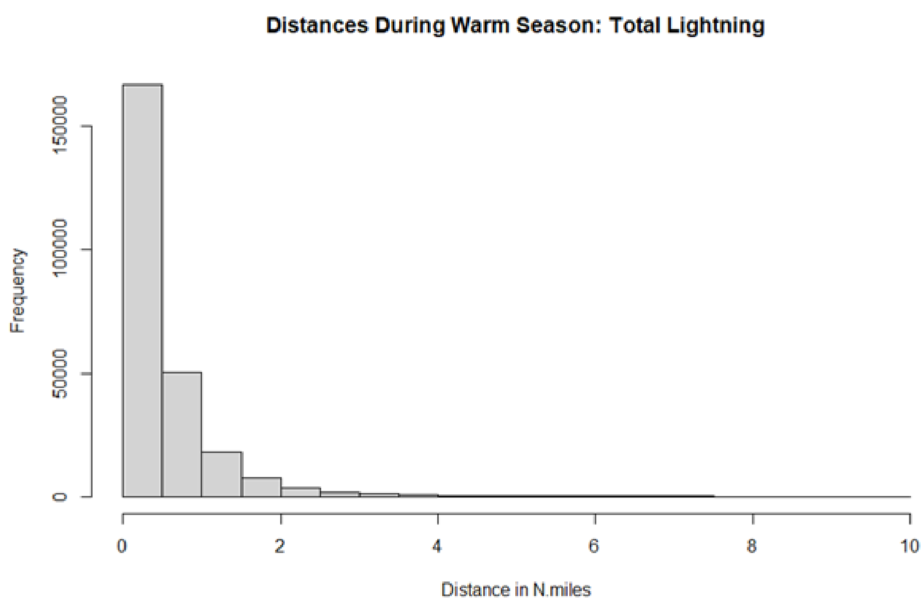

Figure 27. Warm Seasonal Months for Total Lightning

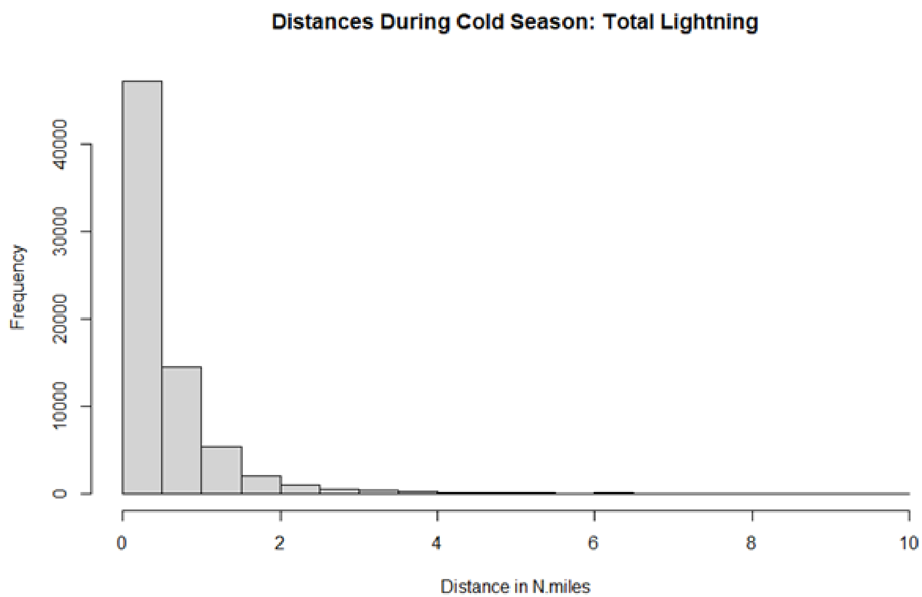

Figure 28. Cold Seasonal Months for Total Lightning

Both figures appear the same, with a few more lightning events occurring beyond 8 nautical miles in the warmer season compared to the colder season. There appears to be more lightning events occurring in the warmer months compared to the colder months, which concurs with previous studies by the NOAA (US Department of Commerce, 2018*b*). However, the distributions of each histogram will need to be examined, and again all 3 types of distributions will need to be examined. Figures 29 through 34 contain the Goodness of Fit tests for the different seasons.

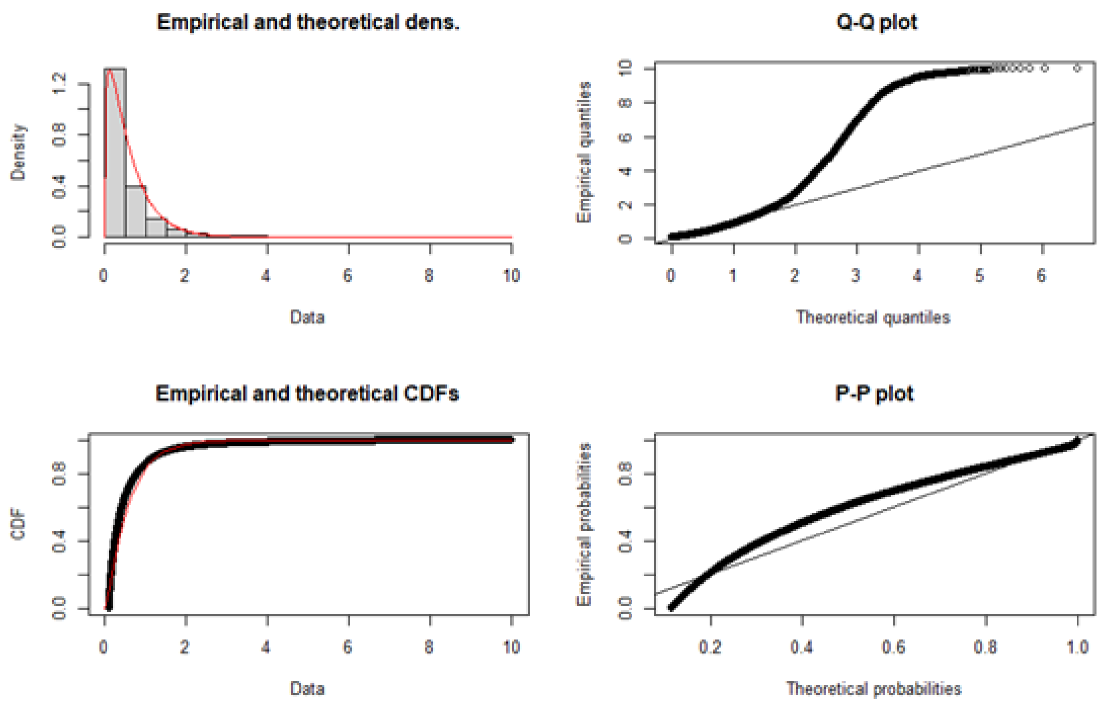

Figure 29. Gamma Distribution for the Warm Season

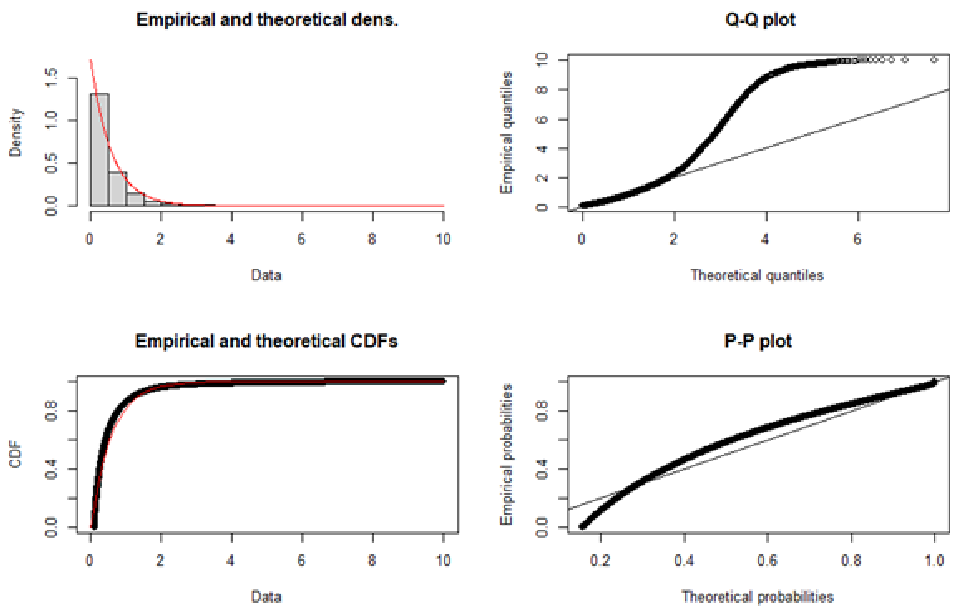

Figure 30. Exponential Distribution for the Warm Season

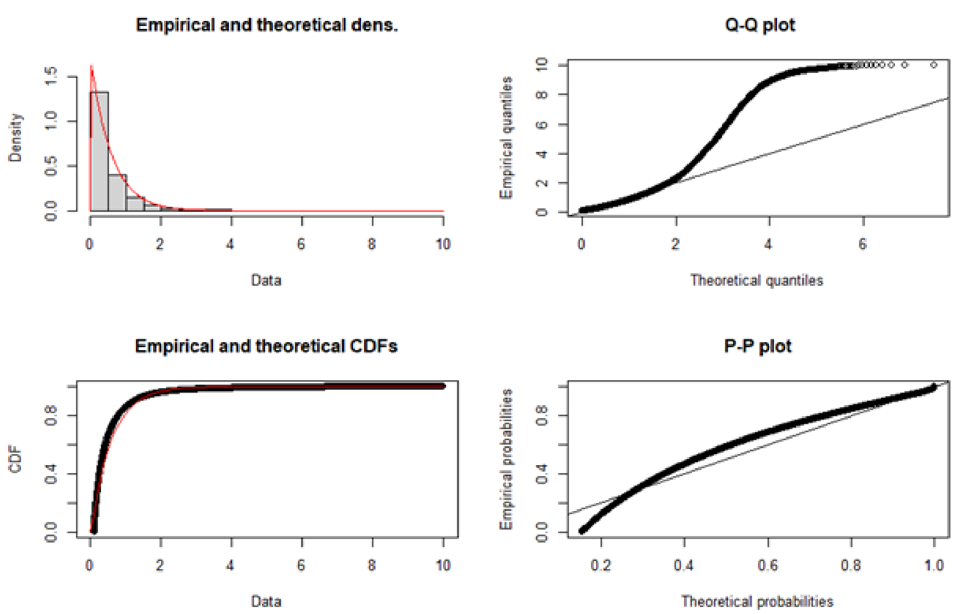

Figure 31. Weibull Distribution for the Warm Season

With the three distributions examined, it appears the Weibull distribution is the best fit for the warm seasons from 2017 through 2020. In terms of the parameter estimates for the distributions, the shape and rate parameters for the gamma distribution are 1.23 and 2.11, with the standard errors of 0.003 and 0.006, respectively. For the exponential distribution tested in Figure 30, the rate parameter is 1.71 with a standard error of 0.003. For the final Weibull distribution in Figure 31, the shape and scale parameters are 1.01 and 0.59, with the standard errors of 0.001 and 0.001, respectively. Figures 32 through 34 will examine the three distributions during the cold seasons.

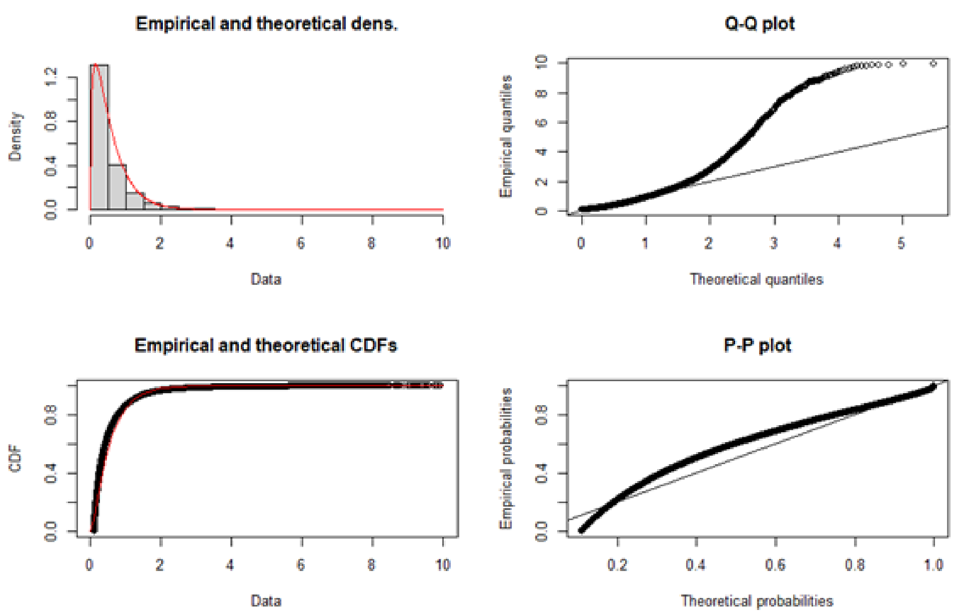

Figure 32. Gamma Distribution for the Cold Season

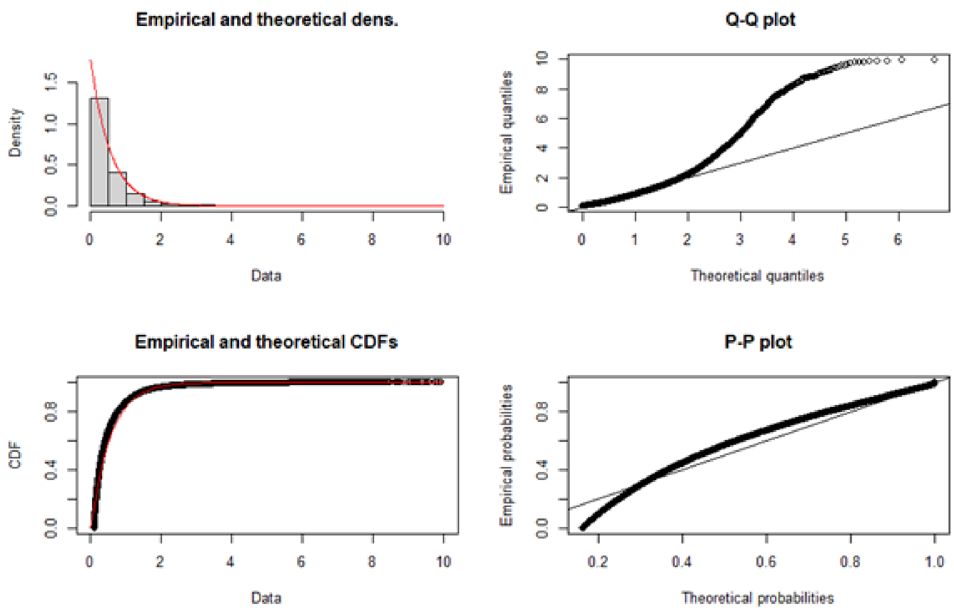

Figure 33. Exponential Distribution for the Cold Season

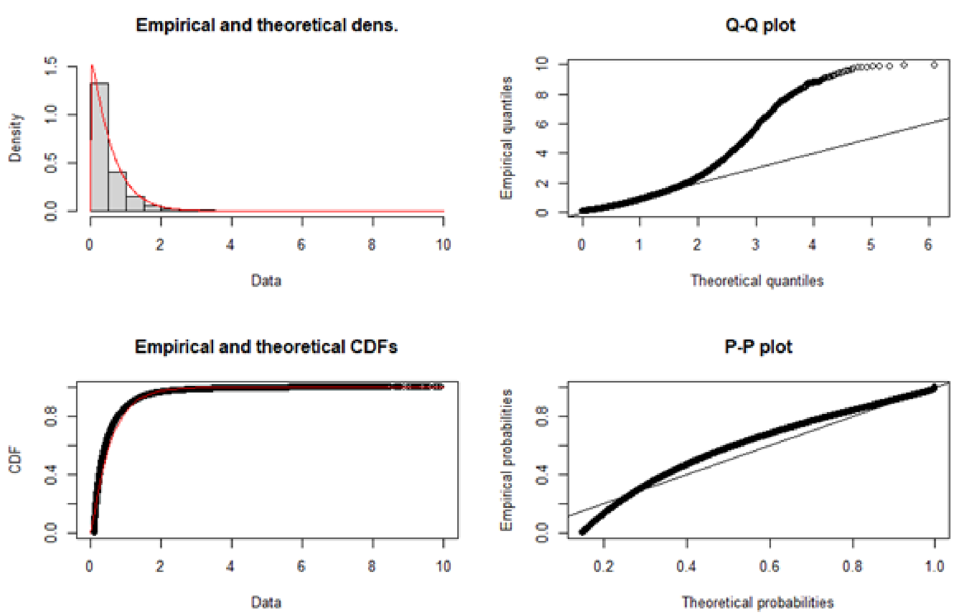

Figure 34. Weibull Distribution for the Cold Season

The cold season appears to have more of a bend at the beginning of the histogram, with the exponential doing slightly better than the Weibull and the gamma distribution. For the shape and rate parameter estimates for the gamma distribution in Figure 32, the shape parameter is 1.31 and the rate parameter is 2.33, with a standard error of 0.01 and 0.01, respectively. The rate parameter for the exponential distribution in Figure 33 is 1.78, with the standard error of 0.01. The shape and scale parameter for the Weibull distribution are 1.05 and 0.58, with a standard error of 0.003 and 0.002, respectively. Since both seasons have similar parameter estimates for all 3 types of distributions, it is safe to sat there is not much of a difference between the warm and cold season when using total lightning.

4.2.4 Seasonal Plots: Cloud-to-Ground

The same process as shown in 4.2.2 will also be applied to the model that created the boundary of pre-existing storms by solely utilizing cloud-to-ground lightning events to shape the storm. Figures 35 and 36 show the entire histograms for the warm and cold seasons from 2017 through 2020 as a way to glance at the differences between the seasons when using cloud-to-ground lightning events to establish a boundary.

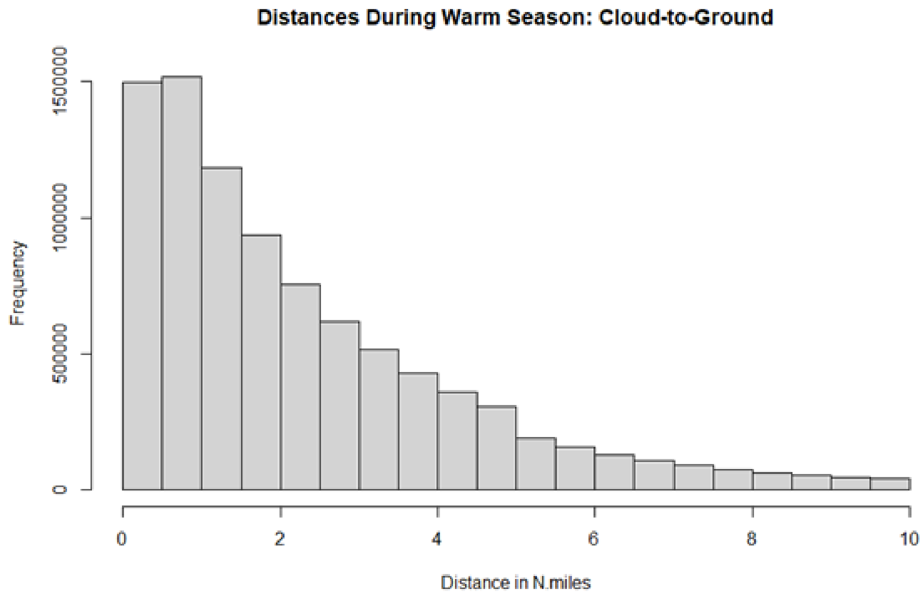

Figure 35. Warm Seasonal Months for Cloud-to-Ground

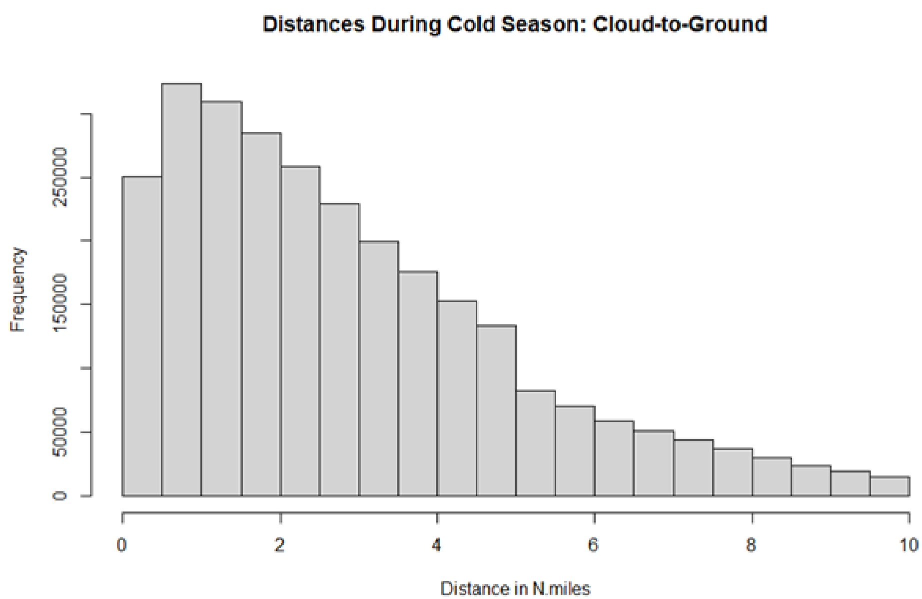

Figure 36. Cold Seasonal Months for Cloud-to-Ground

In term of differences between the two plots, there appears to be more of a spread during the cold season compared to the warm season with regards to lightning strikes between 0.5 and 5 nautical miles.

4.2.5 Yearly Seasonal Plots

It is also important to compare the different seasons between the years for the different seasons as well. The following figures illustrate the differences and similarities between the seasons when using cloud-to-ground lightning events to establish the boundary of the storm.

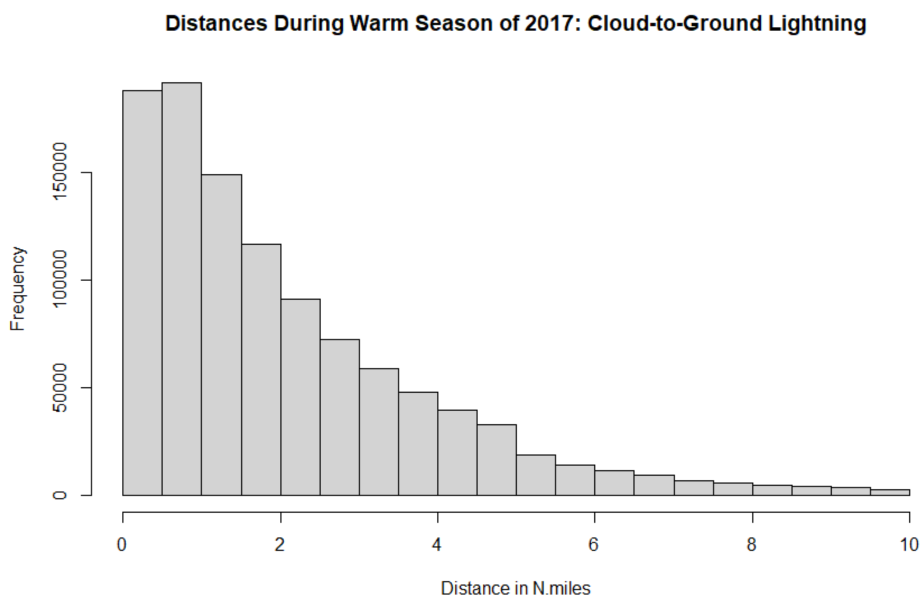

Figure 37. Warm Season of 2017 for Cloud-to-Ground

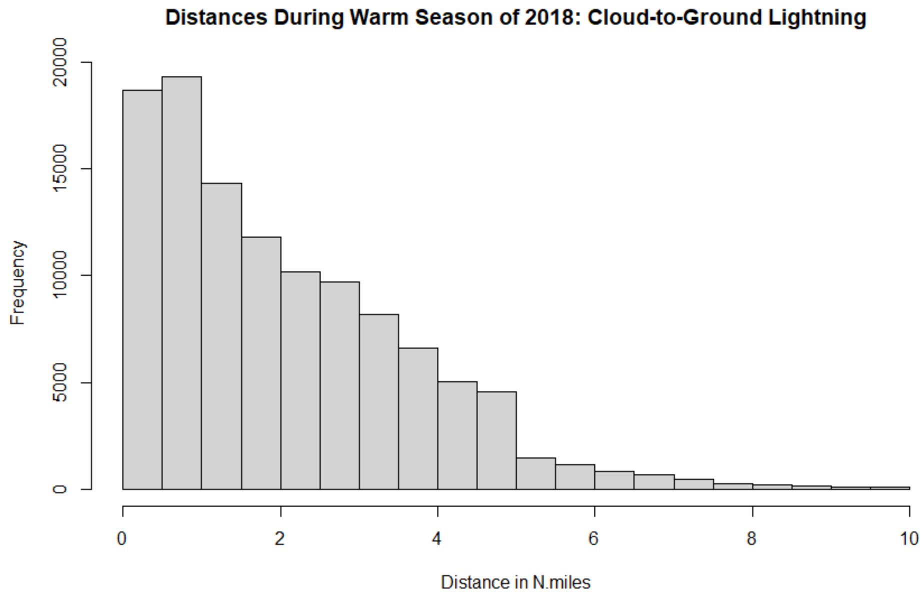

Figure 38. Warm Season of 2018 for Cloud-to-Ground

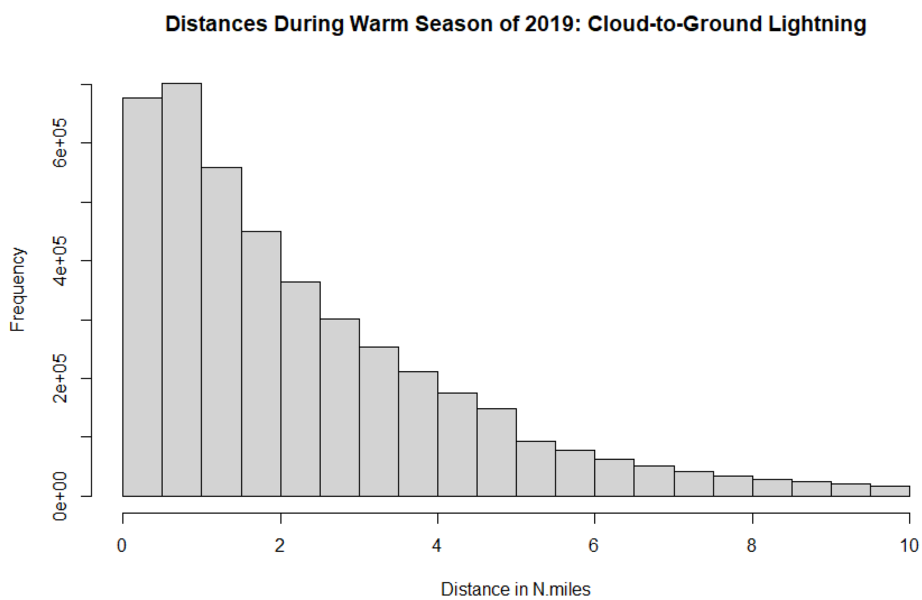

Figure 39. Warm Season of 2019 for Cloud-to-Ground

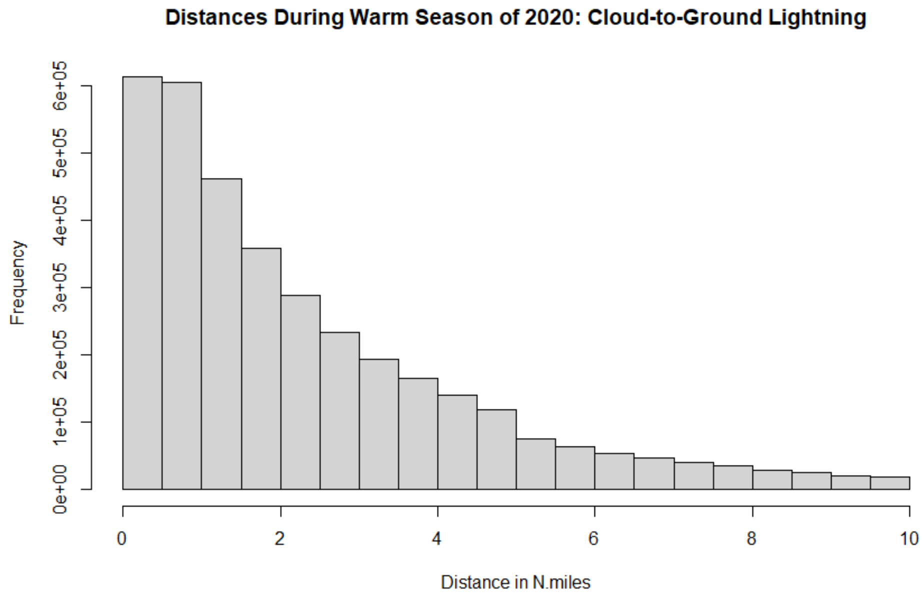

Figure 40. Warm Season of 2020 for Cloud-to-Ground

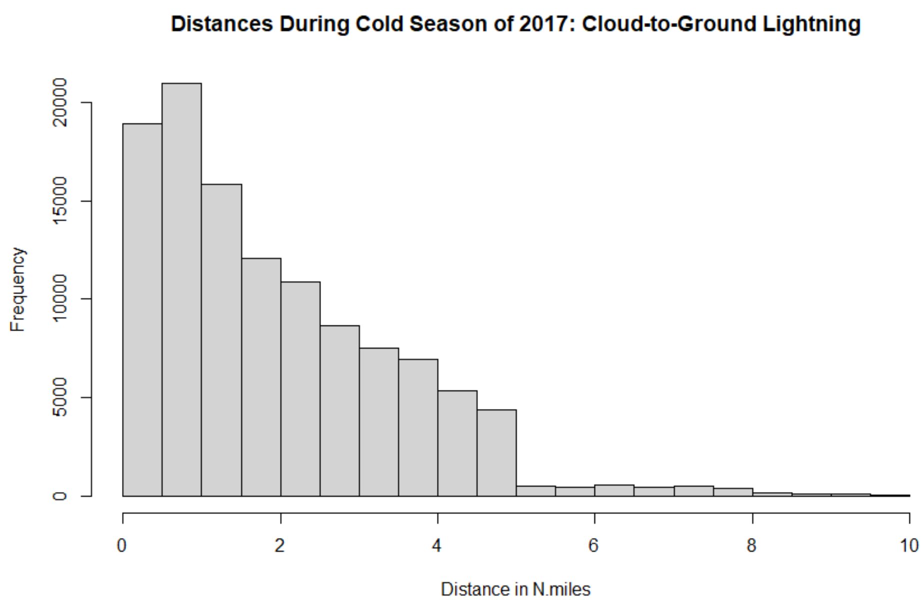

Figure 41. Cold Season of 2017 for Cloud-to-Ground

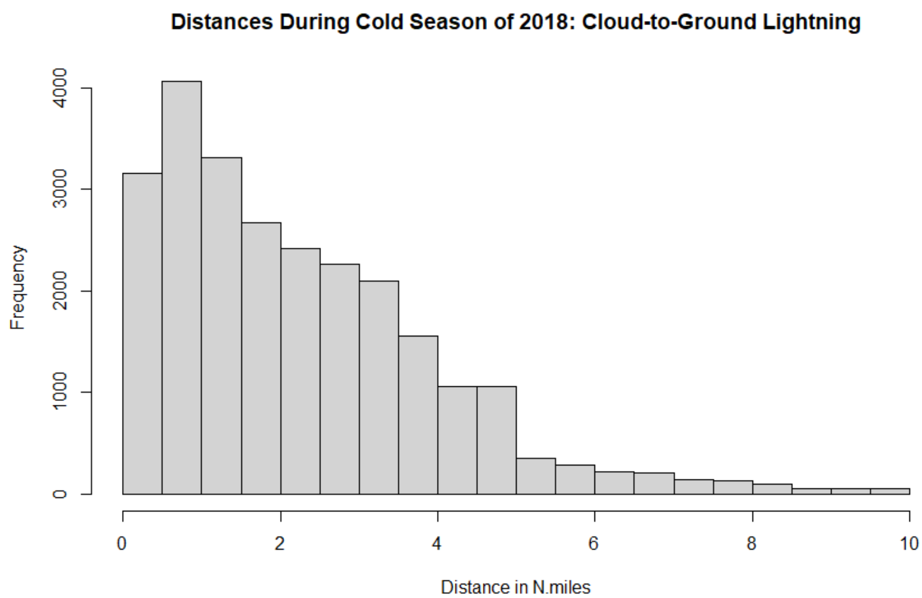

Figure 42. Cold Season of 2018 for Cloud-to-Ground

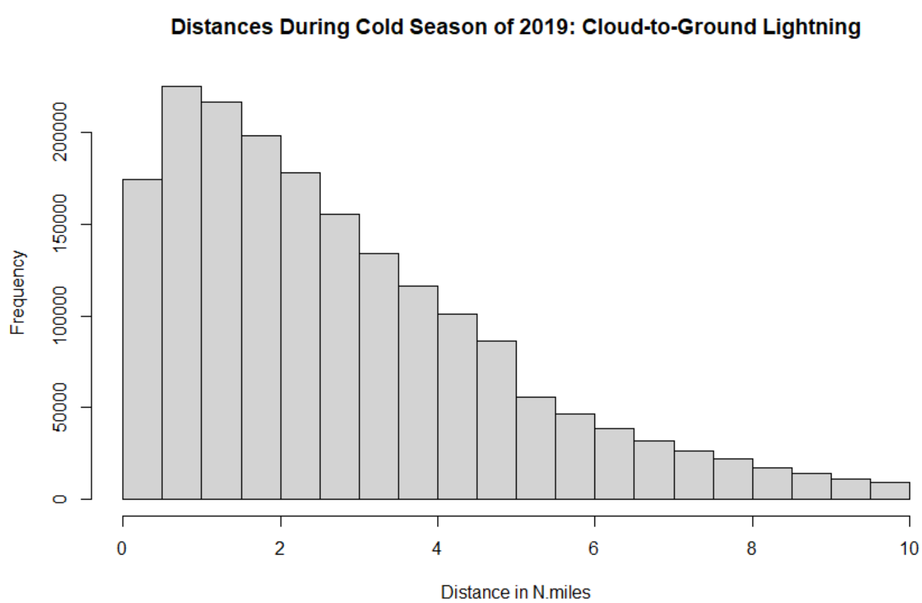

Figure 43. Cold Season of 2019 for Cloud-to-Ground

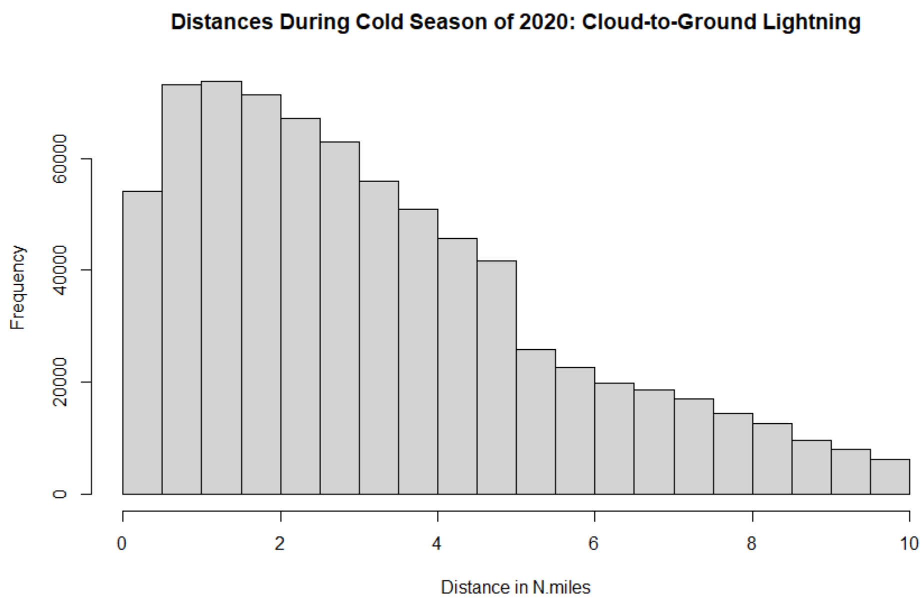

Figure 44. Cold Season of 2020 for Cloud-to-Ground

The warm seasons throughout the years appear to have similar amount of lightning strikes between 0 and 0.5 nautical miles compared to the number of lightning strikes between 0.5 and 1 nautical mile. From there the distribution seems to follow a Weibull or an exponential distribution. For the cold seasons throughout the years, there seems to be a large number of lightning strikes between 0 and 3 nautical miles for most of the years, with a noticeable drop off of lightning strikes beyond 5 nautical miles. With the following plots examined, it is also important to evaluate the total lightning distributions throughout the years.

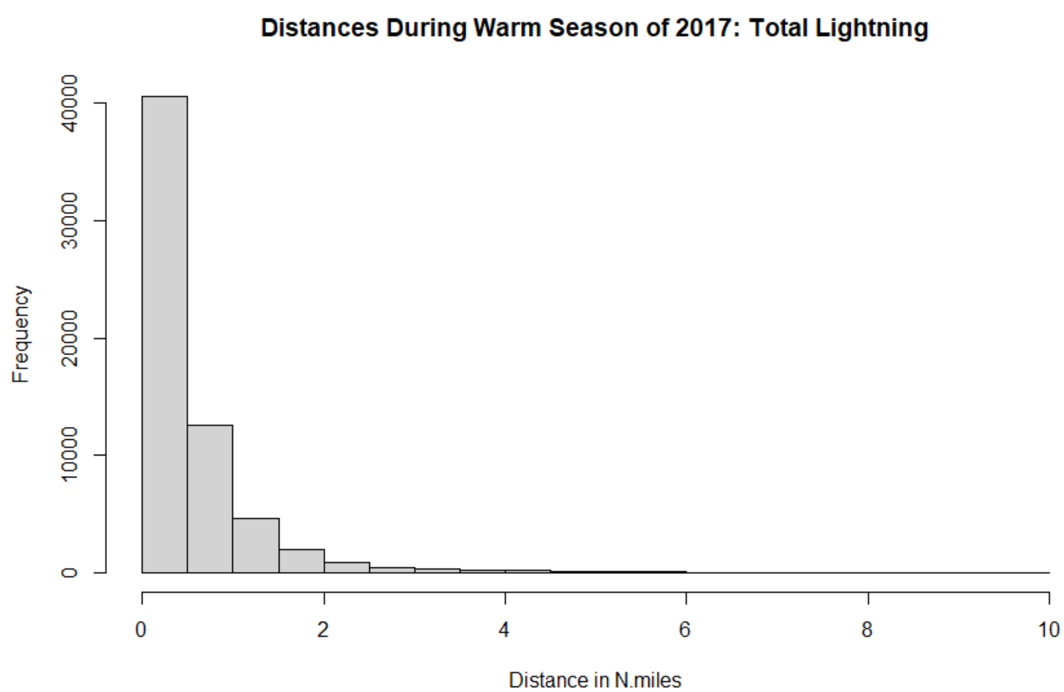

Figure 45. Warm Season of 2017 for Total Lightning

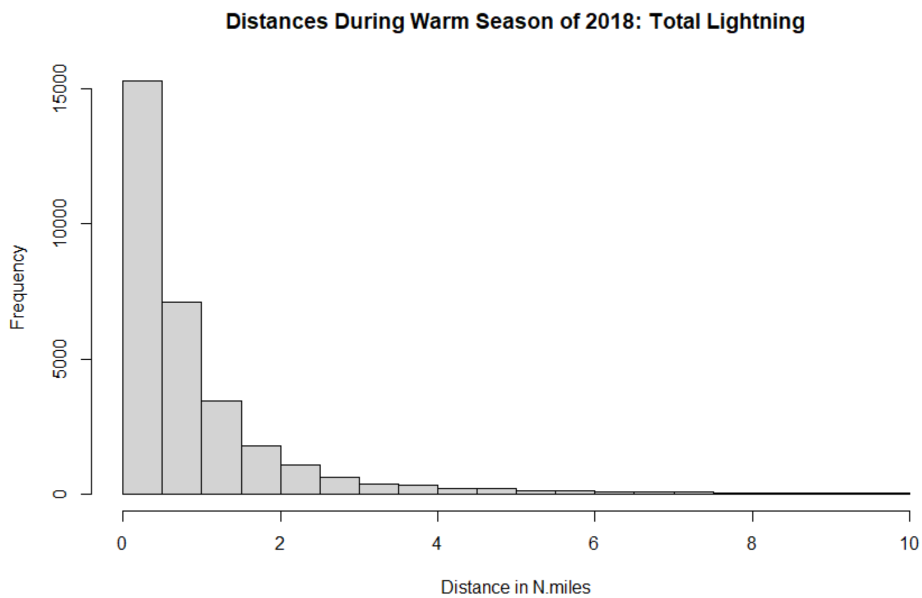

Figure 46. Warm Season of 2018 for Total Lightning

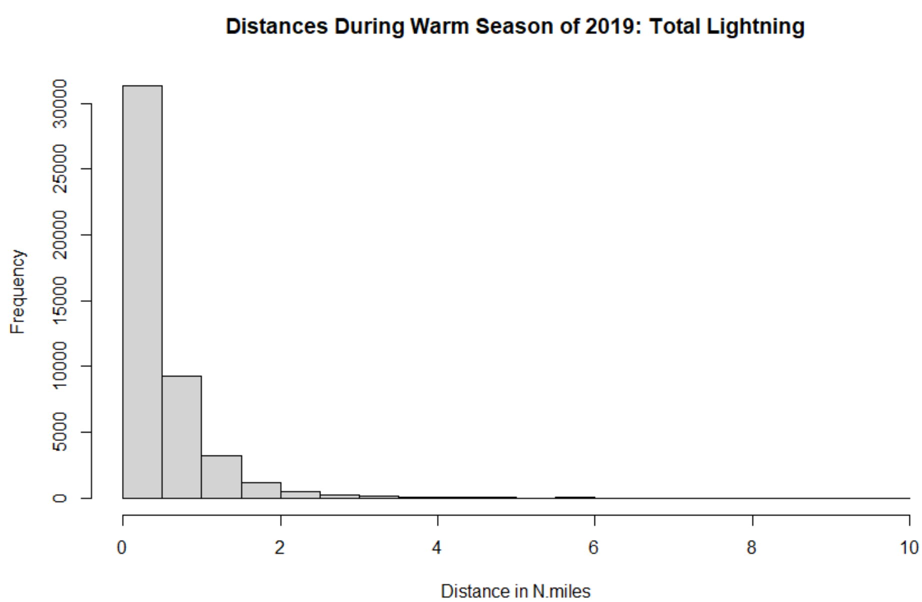

Figure 47. Warm Season of 2019 for Total Lightning

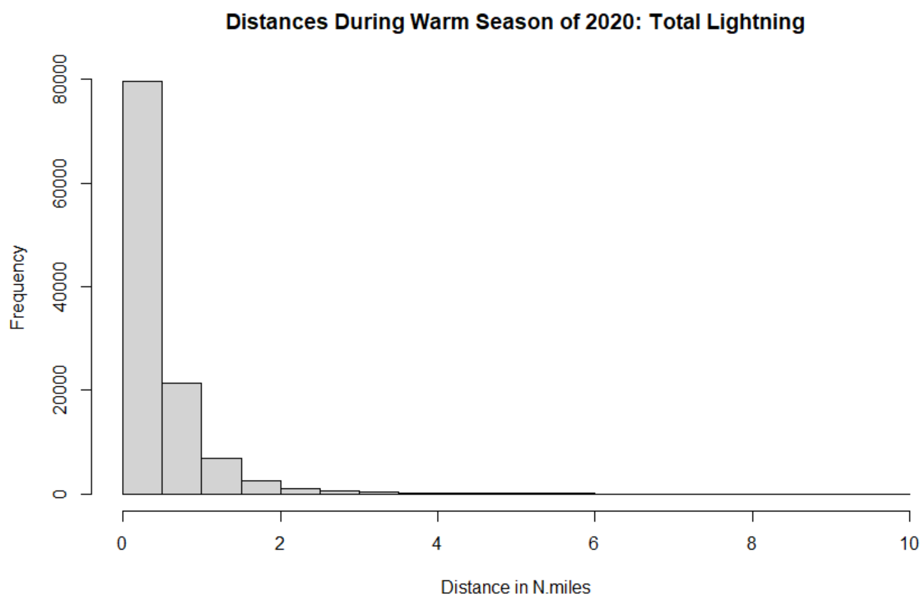

Figure 48. Warm Season of 2020 for Total Lightning

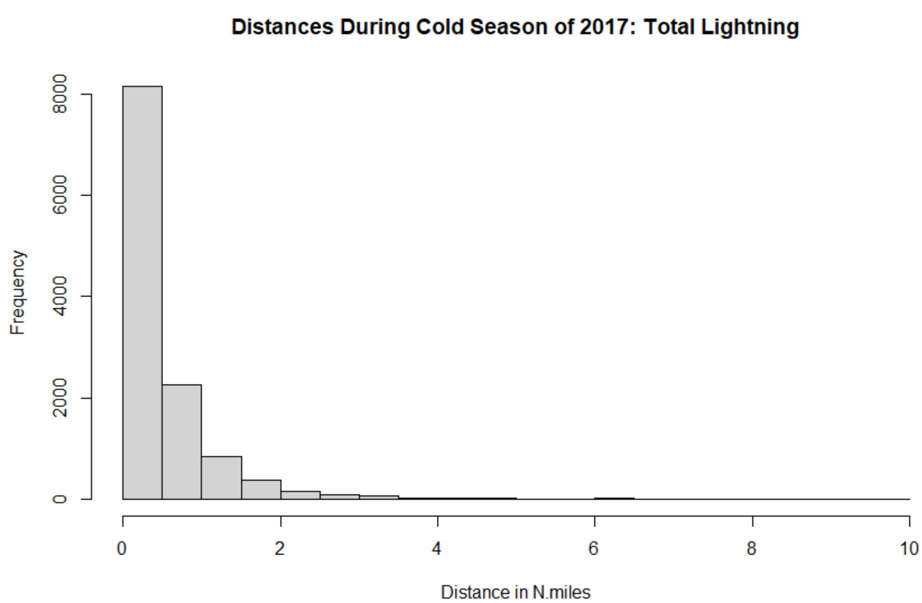

Figure 49. Cold Season of 2017 for Total Lightning

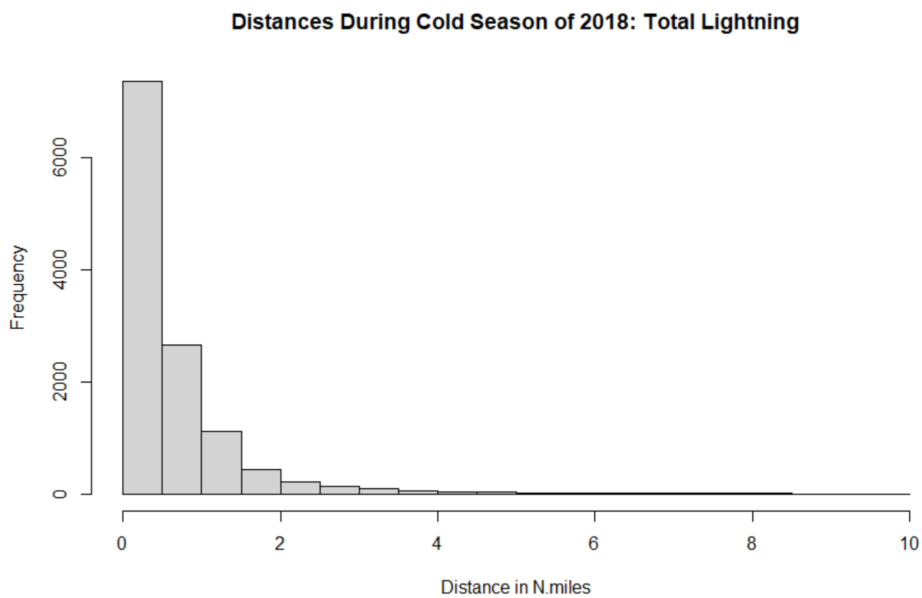

Figure 50. Cold Season of 2018 for Total Lightning

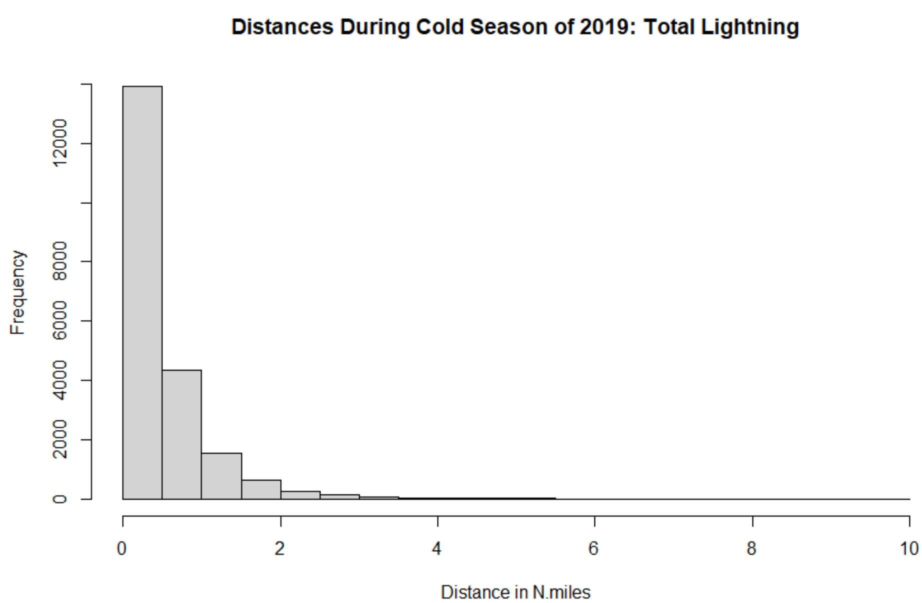

Figure 51. Cold Season of 2019 for Total Lightning

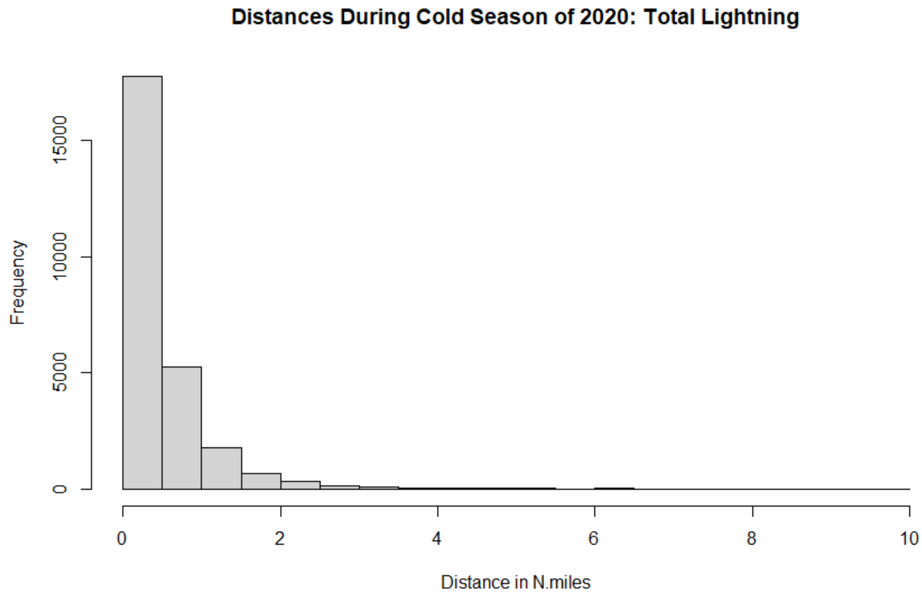

Figure 52. Cold Season of 2020 for Total Lightning

There does not seem to be an apparent difference between the warm and cold seasons for total lightning. In terms of characteristics, all of the graphs have a large amount of lightning strikes between 0 and 0.5 nautical miles with not many if any lightning strikes beyond 3.5 nautical miles.

4.2.6 Number of Occurrences

With the previous plots and histograms being drawn, it is important to keep in mind the number of lightning occurrences being examined. Table 2 gives light on the number of lightning strikes in each year, as well as by the season, whether it is the warm season or the cold season.

Table 2. Number of Lightning Strikes Beyond 0.1 N.M for Cloud-to-Ground Boundary

Year	Warm Season	Cold Season	Total for Year
2017	1,067,868	114,750	1,182,618
2018	113,687	25,274	138,961
2019	4,292,093	1,588,602	5,880,695
2020	3,282,749	751,235	4,033,984

Table 3. Number of Lightning Strikes Beyond 0.1 N.M for Total Lightning Boundary

Year	Warm Season	Cold Season	Total for Year
2017	62,186	12,040	74,226
2018	31,096	12,263	43,359
2019	46,252	21,088	67,340
2020	113,675	26,264	139,939

As shown in Tables 2 and 3, the number of lightning strikes beyond 0.1 nautical miles from the boundary of a storm is much lower when lightning aloft is also utilized in the algorithm compared to not using lightning aloft such as in Table 2.

4.2.7 Empirical and Theoretical CDF

When finding the risk associated with new lightning events, it is important to evaluate the cumulative density function for the data and the theoretical distribution to see how many lightning events occurred beyond a certain threshold, for instance, how many lightning events occurred beyond 5 nautical miles from the boundary of a pre-existing storm. Tables 4 and 5 investigate how many lightning events occurred outside of the pre-existing storm for different distances, both for the cloud-to-ground and the total lightning algorithms for the warm and cold seasons for each year.

Table 4. Empirical Right-Hand Tail Probabilities for the Lightning Strikes Beyond 5 Nautical Miles for Cloud-to-Ground Boundary

Year	Warm Season	Cold Season	Total for Year
2017	2.43%	1.21%	2.34%
2018	3.13%	4.61%	3.38%
2019	5.47%	10.80%	6.71%
2020	5.30%	17.10%	6.55%
Total	4.86%	11.56%	5.94%

Table 5. Empirical Right-Hand Tail Probabilities for the Lightning Strikes Beyond 5 Nautical Miles for Total Lightning Boundary

Year	Warm Season	Cold Season	Total for Year
2017	0.017%	0.017%	0.017%
2018	0.22%	0.07%	0.16%
2019	0.005%	0.005%	0.005%
2020	0.009%	0.02%	0.01%
Total	0.015%	0.014%	0.015%

As shown in Table 5, the number of lightning strikes beyond 5 nautical miles is extremely small when utilizing lightning aloft to determine a boundary warning. When solely using cloud-to-ground lightning events, a greater number of lightning strikes are not accounted for, and therefore is not as accurate, as shown in Figure 12. However, with the current safety protocol being at 5 nautical miles, when using both lightning aloft and cloud-to-ground lightning events, the warning radius can decrease while remaining safe for personnel outside. Table 6 seeks to find a safe alternative to the current safety warning radius by testing different radii. Table 6 takes into account every year from 2017 through 2020 where both lightning aloft and cloud-to-ground lightning events are taken into account when creating the boundary for pre-existing storms.

Table 6. Empirical Right-Hand Tail Probabilities for Different Potential Warning Radii Using Total Lightning

Warning Radius (N.M.)	Warm Season	Cold Season	Total for Year
5	0.015%	0.014%	0.015%
4.5	0.020%	0.19%	0.19%
4	0.024%	0.025%	0.024%
3.5	0.031%	0.033%	0.032%
3	0.041%	0.046%	0.041%
2.5	0.057%	0.066%	0.058%
2	0.086%	0.106%	0.090%

As shown in Table 6, as the warning radius decreases, the number of lightning events beyond the warning radius increases, with the percentages being extremely small. With that being said, a reasonable warning radius could be at 4 nautical miles, with the percentage of lightning events beyond 4 nautical miles being 0.024%. However, another alternative could be lowering the warning radius to a proposed 3.5 nautical miles with the percentage of lightning events beyond 3.5 nautical miles being 0.032%. Again, this is for the number of lightning events beyond the boundary, most of the lightning events for total lightning resides within the boundary of the storm, and not accounted for in these tables.

4.3 Cloud-to-Ground vs. Total Lightning at Locations of Interest

When it comes to lightning storms, there are often precautions and rules-of-thumb used to keep people safe from destructive acts of nature. One said rule is "lightning in 5," which means that if a lightning event occurs within 5 nautical miles from a location of interest, the location "shuts down" and the personnel go inside to avoid getting injured. Previous research by Sanderson (2019) and Hinkley (2019) showed that decreasing the warning radius from 5 nautical miles to 4 nautical miles can be accomplished without compromising personnel safety. An investigation is conducted to see if utilizing lightning aloft lightning events along with cloud-to-ground lightning events would yield better results compared to only using cloud-to-ground lightning events.

For only using cloud-to-ground lightning events, a location of interest would "shut down" once a cloud-to-ground lightning event occurs within 4 nautical miles from the place of interest. If any type of lightning event occurs within 1 nautical mile from the location, a failure occurs, and if the failure is a cloud-to-ground lightning event, the base will also "shut down," but if the lightning event is a cloud-to-cloud lightning event, the base remains open, and failures continue to be tracked. The location will be up and running once 30 minutes have occurred without a lightning event within 5 nautical miles from the base. For using total lightning events to establish a lightning warning radius, the same rules apply; however, if any lightning event occurs within 4 nautical miles, the location closes, and the personnel are safe.

Currently, the lightning warning radius of 4 nautical miles is fixed, but later investigations will look at potentially shortening the warning radius for lightning aloft lightning events. Table 7 shows the results of this investigation, with the locations of interest located in Table 1. This investigation covers all lightning events from 2017 through 2020 as well as all of the locations of interest. Tables 7 and 8 will take a closer look at where the failures occurred and the circumstances around them.

Table 7. Number of Failures by Year (4 Nautical Miles)

Year	CC Fails CG	CG Fails CG	CC Fails TL	CG Fails TL
2017	4,770	13	3	1
2018	2,350	17	0	17
2019	48,542	7	4	0
2020	30,479	17	7	0

As shown in Table 7, the number of failures dropped drastically, with total lightning closing the places of interest in time before failures could occur more often than only using cloud-to-ground lightning events could give the proper warning. The fascinating part is that of the 15 failures from 2017-2020 (with the exception of 2018 since the data has some days with only cloud-to-ground), 14 failures were only lightning aloft failures with only 1 cloud-to-ground failure throughout the 3 years. For most of the lightning aloft failures in the total lightning, the failure occurred when there were no lightning events within 5 nautical miles. In essence, after 30 minutes have passed without a single lightning event occurring within 5 nautical miles from the center of the location, the first lightning event is within 1 nautical mile from the base. The question now is to see if there is a difference in terms of seasons when it comes to the failures for both using the cloud-to-ground and total lightning events, as well as looking at the breakdown of where the failures occurred.

4.3.1 Location of Failures

Initially, only lightning events within 25 nautical miles from the center point [28.5387, -80.6428] in terms of [latitude, longitude] were considered in order to decrease the size of the total data set. Once the sample size is decreased, only lightning events within 5 nautical miles from the location of interest were considered, with the warning radius for both total lightning and cloud-to-ground lightning remained constant at 4 nautical miles. Table 8 shows how many lightning events were examined for every location of interest from 2017 through 2020, along with how many failures occurred at each location depending on the type of failure as well as if total lightning was utilized or not.

Table 8. Failures at Location of Interest

Location	Strikes	CC Fails CG	CG Fails CG	CC Fails TL	CG Fails TL
Lc39	5,591,014	13,423	7	3	0
Lc40	4,593,987	13,592	6	1	1
Lc37	4,565,283	9,602	5	1	3
Spx_central	4,717,596	12,208	4	3	4
Astrotech	7,693,108	13,576	13	5	6
Psfb	2,236,282	8,260	13	1	2
Slf	5,584,622	15,480	6	0	2
IA	392,745	0	0	0	0

As shown in Table 8, the number of lightning aloft failures dramatically decreases when the total lightning is added to the data set. The number of cloud-to-ground failures decreased or stayed the same, but as shown in table 7, most of the cloud-to-ground failures occurred in 2018 alone, with only one cloud-to-ground failure occurring in 2017 when total lightning is used. The warning radius for both the cloud-to-ground lightning events and the lightning aloft events were set at 4 nautical miles at the start of the investigation, but as the warning radius shrinks, how much does the number of failures increases when using total lightning?

4.3.2 Comparing Different Warning Radii

Table 9 seeks to answer this question by decreasing the radius by 0.1 nautical miles starting from 4 nautical miles to 3.5 nautical miles to see how they compare. Both the lightning aloft failures and the cloud-to-ground lightning events will be tracked separately such as in Table 7; however, the number of failures will not be split into years. Instead, the rows will focus on the warning radius using total lightning, similar to Table 6. The purpose of this investigation is to see if the warning radius can decrease while maintaining similar results as Table 6 to ensure that personnel are safe. For Table 9, the warning radius for lightning aloft and cloud-to-ground lightning will be the same radius, which means that if either a lightning aloft or a cloud-to-ground lightning event occurs within the warning radius, the location of interest will close.

Table 9. Warning Radius Comparison

Warning Radius (N.M.)	CC Fails	CG Fails	Total Fails
4	14	18	32
3.9	16	19	35
3.8	16	22	38
3.7	16	23	40
3.6	17	23	40
3.5	19	23	42

From Table 9, the number of failures does not increase by much as the warning radius decreases, with only 10 more failures occurring as the warning radius drops to 3.5 nautical miles. Looking at the cumulative density function when using total lightning, the amount of risk associated with decreasing the warning radius from 4 to 3.5 nautical miles increases from 0.024% to 0.032%, and when using Table 9, the number of failures from 2017 through 2020 increased from 32 to 42 failures. Thus, utilizing 4 nautical miles could become a great alternative to the “lightning in 5” rule-of-thumb, but decreasing the warning radius to a lower value, such as 3.9 or 3.8 nautical miles when using total lightning can keep personnel outside safe while keeping locations of interest operational.

Now, it is important to look at the different years as well, to see if excluding 2018 data would affect the results from Table 9. Tables 10 through 13 show the different years, all with varying warning radii to see how many failures occur as the warning radius shrinks.

Table 10. Warning Radius Comparison: 2017

Warning Radius (N.M.)	CC Fails	CG Fails	Total Fails
4	3	1	4
3.9	3	1	4
3.8	3	1	4
3.7	3	1	4
3.6	3	1	4
3.5	3	1	4

Table 11. Warning Radius Comparison: 2018

Warning Radius (N.M.)	CC Fails	CG Fails	Total Fails
4	0	17	17
3.9	0	18	18
3.8	0	20	20
3.7	0	21	21
3.6	0	21	21
3.5	0	21	21

Table 12. Warning Radius Comparison: 2019

Warning Radius (N.M.)	CC Fails	CG Fails	Total Fails
4	4	0	4
3.9	5	0	5
3.8	5	0	5
3.7	5	0	5
3.6	5	0	5
3.5	7	0	7

Table 13. Warning Radius Comparison: 2020

Warning Radius (N.M.)	CC Fails	CG Fails	Total Fails
4	7	0	7
3.9	8	0	8
3.8	8	1	9
3.7	8	1	9
3.6	9	1	10
3.5	9	1	10

Tables 10 through 13 show that 2018 was a bit of an anomaly, but the decrease in the warning radius appears to be safe for the decrease from 4 nautical miles to 3.5 nautical miles for most of the years. For 2017 and 2019, the number of cloud-to-ground failures stayed constant as the warning radius decreased to 3.5 nautical miles. For 2020, when the warning radius decreased from 4 to 3.8 nautical miles, a cloud-to-ground failure occurred but remained steady after 3.8 nautical miles. As for cloud-to-cloud failures, the number rose as the warning radius decreased; however, the number of failures did not increase by more than 3 failures in a year.

4.4 Summary

An analysis is conducted to examine the number of lightning strikes beyond boundaries constructed by using only cloud-to-ground lightning events or using both cloud-to-ground and lightning aloft events. The analysis compares the warm and cold seasons for the different types of algorithms, as well as how the algorithm compares throughout years and the seasons within the years. From there, a statistical analysis of the empirical cumulative density functions was conducted to see how many lightning events occur beyond potential warning radii, such as dropping the current standard of 5 nautical miles to 4 nautical miles. An investigation was then constructed where all of the lightning events from 2017 through 2020 were tested with different warning radii for locations of interest, as well as an investigation that seeks to compare the effectiveness of a warning radius of 4 nautical miles for both cloud-to-ground as well as total lightning.

V. Conclusions and Recommendations

5.1 Overview

This chapter provides a comprehensive summary of the results and methodology used within this research, along with alternative methods that could be useful for further research. This chapter also seeks to give insights on future research opportunities.

5.2 Results

With the vast majority of lightning strikes being within the boundary of pre-existing storms when using total lightning, along with less than 0.024% of lightning strikes being beyond 4 nautical miles, the utilization of total lightning would be beneficial to add when accounting for positions of lightning storms. When using total lightning, a warning radius of 4 nautical miles can be a safe distance to adopt, and if some more risk can be taken, the up to 3.5 nautical miles can safely be adopted without much of a risk increase. If the warning radius does not change, then the use of total lightning would be beneficial to give accurate representations of the existence of pre-existing storms within models. When total lightning is utilized, there is not much of a difference in terms of the distance from the boundary to new lightning events when it comes to the warm and cold season. When only cloud-to-ground lightning strikes are utilized to fit the boundary of the storm, there appears to be a difference between the warm and cold seasons. Only using cloud-to-ground lightning events greatly impacts the safety of locations and personnel, with Table 7 illustrating that using total lightning decreases risk substantially. Table 4 also shows that 5.94% of lightning events occur beyond 5 nautical miles when using cloud-to-ground compared to 0.015% when using total lightning to establish a boundary.

5.3 Future Research Opportunities

As for future research in the field, it would be beneficial to experiment with the ellipse fitting methods from previous researchers to speed up the processing time for the clustering algorithm. Bear in mind that this study focused on the central eastern coast of Florida, with lightning storms varying by location around the country. Therefore, testing different regions of Florida or even different parts of the United States could be beneficial in painting a more accurate depiction of lightning events. A place of interest could be Oklahoma, seeing as they have been experiencing an abundance of lightning events in recent years. Another place of interest could be the Houston Spaceport in Texas. In terms of accurate depictions of lightning events and predictions, it could be beneficial to track how the center of the storm moves throughout time, and see if that can help predict where the next few lightning strikes might occur. Such a process would take advantage of modern software and the use of neural networks. Another opportunity would be removing percentages of data from total lightning in order to simulate sensor failure. This could bear insight on how robust the system is when losing 5% of data compared to 10% or even 20% of its data. The investigation within this study looked at various warning radius but did not account for operational hours, so it could be beneficial to see how the warning radius affects working hours.

5.4 Final Remarks

From the results of this study, consider reducing the lightning warning radius on the eastern coast of Florida. The warning radius could drop to 4 or perhaps 3.5 nautical miles while keeping personnel safe from lightning strikes. If the warning radius is not reduced, the utilization of total lightning will provide adequate protection of the whereabouts of pre-existing storms and lightning strikes compared to solely using cloud-to-ground lightning events. Implementing total lightning gives a more accurate representation of the location of lightning areas compared to solely using cloud-to-ground lightning events.

Bibliography

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A. and Zou, B. (2010), 'Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the united states'.

Commercial Space Data (2021).
URL: *https://www.faa.gov/data_research/commercial_space_data/*

Darack, E. (2007), 'Florida's central gulf coast: Lightning capital of north america', *Weatherwise* **60**(1), 14–15.

Gold, S., White, E., Roeder, W., McAleenan, M., Kabban, C. S. and Ahner, D. (2020), 'Probabilistic contingency tables: An improvement to verify probability forecasts', *Weather and Forecasting* **35**(2), 609–621.

Hill, J. D., Mata, C. T., Nag, A. and Roeder, W. P. (2016), Evaluation of the performance characteristics of merlin and nldn based on two years of ground-truth data from kennedy space center/cape canaveral air force station, florida, *in* '24th International Lightning Detection Conference, San Diego, California, USA'.

Hinkley, J. J., Huddleston, L. L. and Roeder, W. P. (2019), 'Lightning strike distance distribution beyond a preexisting lightning area', *NASA Tech. Rep. TM-220183* .

Holland, K. G. (2021), Ccsfs/ksc total lightning warning radii optimization for merlin using preexisting lightning areas, Master's thesis, Air Force Institute of Technology.

Johnson, R. A. and Wichern, D. W. (1992), *Applied multivariate statistical analysis*, 3 edn, Englewood Cliffs, N.J: Prentice Hall.

Lericos, T. P., Fuelberg, H. E., Watson, A. I. and Holle, R. L. (2002), 'Warm season lightning distributions over the florida peninsula as related to synoptic patterns', *Weather and Forecasting* **17**(1), 83–98.

NASA (n.d.).
URL: *https://lightning.nsstc.nasa.gov/primer/primer2.html*

Price, C. (2009), Thunderstorms, lightning and climate change, *in* 'Lightning: principles, instruments and applications', Springer, pp. 521–535.

Roeder, W. P., Holle, R. L., Cooper, M. and Hodanish, S. (2012), Lessons learned in communicating lightning safety effectively, *in* '4th international lightning meteorology conference, Broomfield, CO, Vaisala'.

Roeder, W. P. and McNamara, T. M. (2006), A survey of the lightning launch commit criteria, *in* '2nd Conference on Meteorological Applications of Lightning Data', Vol. 29.

Roeder, W. P. and Rockledge, F. (2010), Lightning safety procedures for the public, *in* '3rd International Lightning Meteorology Conference', pp. 21–22.

Roeder, W. P. and Saul, J. M. (2012), Four dimensional lightning surveillance system: Status and plans, *in* '22nd International Lightning Detection Conference', pp. 2–3.

Roeder, W. P. and Saul, J. M. (2016), 'The new mesoscale eastern range lightning information system', *Sensors* **10**, 9.

Rosenberger, J. M., Schaefer, A. J., Goldsman, D., Johnson, E. L., Kleywegt, A. J. and Nemhauser, G. L. (2002), 'A stochastic model of airline operations', *Transportation science* **36**(4), 357–377.

Sanderson, D. L. (2019), Modeling the distribution of lightning strike distances outside a preexisting lightning area, Master's thesis, Air Force Institute of Technology.

Saunders, C. (1993), 'A review of thunderstorm electrification processes', *Journal of Applied Meteorology and Climatology* **32**(4), 642–655.

Tello, E. A. (2021), Behavior of lightning in developing storms, Master's thesis, Air Force Institute of Technology.

US Department of Commerce, N. (2018*a*).
URL: *https://www.weather.gov/safety/lightning-science-initiation-stepped-leader: :text=Stepped%20leaders%20develop%20within%20thunderstorm,become%20large%20(Figure%*

US Department of Commerce, N. (2018*b*).
URL: *https://www.weather.gov/safety/lightning-thunderstorm-development*

Appendix . Python Code for Total Lightning Boundary

```
import datetime as dt
import numpy as np
import math
from scipy.spatial import ConvexHull
import matplotlib.pyplot as plt
from haversince import Unit, haversine

in1 = open('2017_combined_txt_file.txt')
next(in1)
next(in1)
next(in1)
input_reader = in1.readlines()
raw_data = []
for row in input_reader:
    split_data = row.split(' ')
    if split_data[0] =='\n' or split_data[0][0] =="c" or split_data[0][0] =='*' or split_data[1].split(':')[0] =="-":
        pass
    else:
        date = split_data[0].split('-')
        clock = split_data[1].split(':')
        sec = int(clock[2].split('.')[0])
        nano = int(clock[2].split('.')[1])
        lat = float(split_data[3])
        if len(split_data) > 6:
            long = float(split_data[5])
        else:
            long = float(split_data[5].split("\\")[0])
        row_data = [dt.datetime(int(date[0]), int(date[1]), int(date[2]), int(clock[0]), int(clock[1]),
            sec, nano), lat, long]
        print(row_data[0])
        raw_data.append(row_data + [0])
in1.close()

raw_data = sorted(raw_data)

view_data = np.array(raw_data)

print(view_data)

center = [28.5387, -80.6428]

def distance_nautical_miles(point_of_interest, line):
    lat1, lon1, = point_of_interest[0], point_of_interest[1]
    lat2, lon2, = line[0], line[1]
    return haversine((lat1, lon1), (lat2, lon2), unit = Unit.NAUTICAL_MILES)

print (len(raw_data))

raw_data = [x for x in raw_data if distance_nautical_miles(x[1:3], center) <= 25]

#raw_data = [x for x in raw_data if x[0] > dt.datetime(2018, 9, 1, 10, 10, 46, 66)]
print(len(raw_data))

#create txt file
"""
with open('unfilteredtxtfull.txt', 'w') as f2:
    for row in raw_data:
        f2.write(str(row))
"""
# initialization

time_diff_goal = 120
current_time = []
distances = []
start_time = dt.datetime.now()

def clean_currenttime(array):
    return [array[row] for row in range(len(array)) if (array[-1][0] - array[row][0]).seconds < time_diff_goal]


def check_unique_len(array):
    unique_class = np.unique(array[:, 3])
    unique_t_f = False
    for unique_value in unique_class:
        if len(np.where(array[:, 3] == unique_value)[0]) >= 5:
            unique_t_f = True
    return unique_t_f


def find_nonconvex_distance(point_interest, sample_data):
    unique_class = np.unique(sample_data[:, 3])
    dist = []
    for unique_value in unique_class:
        if len(np.where(sample_data[:, 3] == unique_value)[0]) < 5:
            for i in sample_data[np.where(sample_data[:, 3] == unique_value)[0], 1:3]:
```

```
                dist.append([distance_nautical_miles(point_interest, i), unique_value])
    if len(dist) < 1:
        dist.append([50,0])
    dist = np.array(dist)
    shortest_distance = min(dist[:, 0])
    return [shortest_distance, int(dist[np.where(dist[:, 0] == shortest_distance)[0], 1][0])]

def find_convex_dist(point_interest, sample_data):
    unique_class = np.unique(sample_data[:, 3])
    dist = []
    for unique_value in unique_class:
        if len(np.where(sample_data[:, 3] == unique_value)[0]) >= 5:
            curr = sample_data[np.where(sample_data[:, 3] == unique_value)[0], 1:3]
            hull = ConvexHull(curr)
            total = np.append(curr, point_interest).reshape(len(curr) + 1, 2)
            test_hull = ConvexHull(total)
            vert = []
            for i in test_hull.vertices:
                vert.append(total[i, 0:2].tolist())
            if point_interest.tolist() not in vert:
                dist.append([-1, unique_value])
                break
            local_min = []
            dist_list = []
            for i in hull.vertices:
                dist_list.append([distance_nautical_miles(curr[i], point_interest)])
            smallest_val = min(dist_list)
            min_vert = hull.vertices[dist_list.index(smallest_val)]
            simplices_list = np.where(hull.simplices == min_vert)[0]
            for j in simplices_list:
                a = np.linspace(curr[hull.simplices[j][0], 0], curr[hull.simplices[j][1], 0],
                    100), np.linspace(curr[hull.simplices[j][0], 1],
                    curr[hull.simplices[j][1], 1], 100)
                for i in range(100):
                    local_min.append(distance_nautical_miles(point_interest, [a[0][i], a[1][i]]))
                dist.append([min(local_min), unique_value])
    dist = np.array(dist)
    #print(dist)
    shortest_distance = min(dist[:, 0])
    return [shortest_distance, int(dist[np.where(dist[:, 0] == shortest_distance)[0], 1][0])]


distance_hist = []
num_storms = []
counter = 1
within_hour = True
for i in range(len(np.array(raw_data))):
    if not within_hour:
        print(raw_data[i][0])
        if (raw_data[i][0] - raw_data[i-1][0]).seconds > 1800: #greater than 30 minutes
            within_hour = True
    if within_hour:
        current_time.append(raw_data[i])
        current_time = clean_currenttime(current_time)
        if len(np.array(current_time)) == 1:
            counter = 1
            current_time[0][3] = counter
            num_storms.append(counter)
            counter += 1
            start = current_time[0][0]
        elif not check_unique_len(np.array(current_time)):
            if (current_time[-1][0] - start).seconds >= 3600:
                print(current_time[-1][0])
                within_hour = False
                pass
            else:
                sample_array = np.array(current_time)
                short_dist = find_nonconvex_distance(sample_array[-1, 1:3], sample_array[0:-1])
                if short_dist[0] < 5:
                    distance_hist.append([sample_array[-1,0],short_dist[0]])
                    current_time[-1][3] = short_dist[1]
                    num_storms.append(short_dist[1])
                else:
                    current_time[-1][3] = counter
                    num_storms.append(counter)
                    counter += 1
                #print(distance_hist[-1])
        elif check_unique_len(np.array(current_time)):
            if (current_time[-1][0] - start).seconds >= 3600:
                print(current_time[-1][0])
                within_hour = False
                pass
            else:
                sample_array = np.array(current_time)
                short_noncon = find_nonconvex_distance(sample_array[-1, 1:3], sample_array[0:-1])
                short_con = find_convex_dist(sample_array[-1, 1:3], sample_array[0:-1])
                if short_con[0] <= 10 or short_noncon[0] <= 10:
                    if short_con[0] <= short_noncon[0]:
                        current_time[-1][3] = short_con[1]
```

```
                        distance_hist.append([sample_array[-1,0], short_con[0]])
                        num_storms.append(short_con[1])
                    else:
                        current_time[-1][3] = short_noncon[1]
                        distance_hist.append([sample_array[-1,0], short_noncon[0]])
                        num_storms.append(short_con[1])
                else:
                    current_time[-1][3] = counter
                    num_storms.append(counter)
                    counter += 1
                print(distance_hist[-1])


with open('testupdatedtxtfile2017.txt', 'w', newline= '') as f1:
    for row in raw_data:
        if row[3] > 0:
            f1.write(str(row))
            f1.write('\n')
        else:
            pass

with open('testfull_distance_hist2017.txt', 'w', newline= '') as f4:
    for row in distance_hist:
        f4.write(str(row))
        f4.write('\n')

end_time = dt.datetime.now()

print(end_time - start_time)
#print(num_storms)
bins = range(0, 10)
plt.hist([x[1] for x in distance_hist if x[1] > 0.1], bins = bins, edgecolor = 'black')
plt.show()
```

Appendix . Python Code for Cloud-to-Ground Boundary

```
import datetime as dt
import numpy as np
import math
from scipy.spatial import ConvexHull
import matplotlib.pyplot as plt
from haversince import Unit, haversine


in1 = open('2017_combined_txt_file.txt')
next(in1)
next(in1)
next(in1)
input_reader = in1.readlines()
raw_data = []
for row in input_reader:
    split_data = row.split(' ')
    if split_data[0] =='\n' or split_data[0][0] =="c" or split_data[0][0] =='*' or
    split_data[1].split(':')[0] =="-":
        pass
    else:
        date = split_data[0].split('-')
        clock = split_data[1].split(':')
        sec = int(clock[2].split('.')[0])
        nano = int(clock[2].split('.')[1])
        lat = float(split_data[3])
        CG_or_CC = None
        if len(split_data) > 6:
            long = float(split_data[5])
            CG_or_CC = "cg"
        else:
            long = float(split_data[5].split("\\")[0])
            CG_or_CC = "cc"
        row_data = [dt.datetime(int(date[0]), int(date[1]), int(date[2]), int(clock[0]), int(clock[1]),
        sec, nano), lat, long]
        print(row_data[0])
        raw_data.append(row_data + [0] + [CG_or_CC])
in1.close()

raw_data = sorted(raw_data)

view_data = np.array(raw_data)

print(view_data)

center = [28.5387, -80.6428]

def distance_nautical_miles(point_of_interest, line):
    lat1, lon1, = point_of_interest[0], point_of_interest[1]
    lat2, lon2, = line[0], line[1]
    return haversine((lat1, lon1), (lat2, lon2), unit = Unit.NAUTICAL_MILES)

print (len(raw_data))

raw_data = [x for x in raw_data if distance_nautical_miles(x[1:3], center) <= 25]

#raw_data = [x for x in raw_data if x[0] < dt.datetime(2018, 6, 15)]
print(len(raw_data))

#create txt file
"""
with open('unfilteredt2019xtcg_full.txt', 'w') as f2:
    for row in raw_data:
        f2.write(str(row))
"""
# initialization

time_diff_goal = 600  # 15 minutes from Tello Paper
current_time = []
distances = []


def clean_currenttime(array, t_time):
    return [array[row] for row in range(len(array)) if (t_time - array[row][0]).seconds <
    time_diff_goal
    and array[row][4] == 'cg']


def check_unique_len(array):
    unique_class = np.unique(array[:, 3])
    unique_t_f = False
    for unique_value in unique_class:
        if len(np.where(array[:, 3] == unique_value)[0]) >= 5:
            unique_t_f = True
    return unique_t_f
```

```
def find_nonconvex_distance(point_interest, sample_data):
    unique_class = np.unique(sample_data[:, 3])
    dist = []
    for unique_value in unique_class:
        if len(np.where(sample_data[:, 3] == unique_value)[0]) < 5:
            for i in sample_data[np.where(sample_data[:, 3] == unique_value)[0], 1:3]:
                dist.append([distance_nautical_miles(point_interest, i), unique_value])
    if len(dist) < 1:
        dist.append([50,0])
    dist = np.array(dist)
    shortest_distance = min(dist[:, 0])
    return [shortest_distance, int(dist[np.where(dist[:, 0] == shortest_distance)[0], 1][0])]


def find_convex_dist(point_interest, sample_data):
    unique_class = np.unique(sample_data[:, 3])
    dist = []
    for unique_value in unique_class:
        if len(np.where(sample_data[:, 3] == unique_value)[0]) >= 5:
            curr = sample_data[np.where(sample_data[:, 3] == unique_value)[0], 1:3]
            hull = ConvexHull(curr)
            total = np.append(curr, point_interest).reshape(len(curr) + 1, 2)
            test_hull = ConvexHull(total)
            vert = []
            for i in test_hull.vertices:
                vert.append(total[i, 0:2].tolist())
            if point_interest.tolist() not in vert:
                dist.append([-1, unique_value])
                break
            local_min = []
            dist_list = []
            for i in hull.vertices:
                dist_list.append([distance_nautical_miles(curr[i], point_interest)])
            smallest_val = min(dist_list)
            min_vert = hull.vertices[dist_list.index(smallest_val)]
            simplices_list = np.where(hull.simplices == min_vert)[0]
            for j in simplices_list:
                a = np.linspace(curr[hull.simplices[j][0], 0], curr[hull.simplices[j][1], 0], 100),
                np.linspace(curr[hull.simplices[j][0], 1], curr[hull.simplices[j][1], 1], 100)
                for i in range(100):
                    local_min.append(distance_nautical_miles(point_interest, [a[0][i], a[1][i]]))
                dist.append([min(local_min), unique_value])
    dist = np.array(dist)
    #print(dist)
    shortest_distance = min(dist[:, 0])
    return [shortest_distance, int(dist[np.where(dist[:, 0] == shortest_distance)[0], 1][0])]

distance_hist = []
num_storms = []
counter = 1

start_time = dt.datetime.now()


for i in range(len(np.array(raw_data))):
    current_time.append(raw_data[i])
    current_time = clean_currenttime(current_time[0:-1], current_time[-1][0])
    current_time.append(raw_data[i])
    if len(np.array(current_time)) == 1:
        counter = 1
        current_time[0][3] = counter
        num_storms.append(counter)
        counter += 1
        start = current_time[0][0]
    elif not check_unique_len(np.array(current_time)):
        if (current_time[-1][0] - start).seconds >= 3600:
            print(current_time[-1][0])
            pass
        else:
            sample_array = np.array(current_time)
            short_dist = find_nonconvex_distance(sample_array[-1, 1:3], sample_array[0:-1])
            if short_dist[0] < 5:
                distance_hist.append([sample_array[-1,0],short_dist[0]])
                current_time[-1][3] = short_dist[1]
                num_storms.append(short_dist[1])
            else:
                current_time[-1][3] = counter
                num_storms.append(counter)
                counter += 1
            #print(distance_hist[-1])
    elif check_unique_len(np.array(current_time)):
        if (current_time[-1][0] - start).seconds >= 3600:
            print(current_time[-1][0])
            pass
        else:
            sample_array = np.array(current_time)
            short_noncon = find_nonconvex_distance(sample_array[-1, 1:3], sample_array[0:-1])
            short_con = find_convex_dist(sample_array[-1, 1:3], sample_array[0:-1])
```

```
            if short_con[0] <= 10 or short_noncon[0] <= 10:
                if short_con[0] <= short_noncon[0]:
                    current_time[-1][3] = short_con[1]
                    distance_hist.append([sample_array[-1,0], short_con[0]])
                    num_storms.append(short_con[1])
                else:
                    current_time[-1][3] = short_noncon[1]
                    distance_hist.append([sample_array[-1,0], short_noncon[0]])
                    num_storms.append(short_con[1])
            else:
                current_time[-1][3] = counter
                num_storms.append(counter)
                counter += 1
            print(distance_hist[-1])

with open('updatedtxtfile2017cg_.txt', 'w') as f1:
    for row in raw_data:
        if row[3] > 0:
            f1.write(str(row))
            f1.write('\n')
        else:
            pass

with open('2017_distance_histcg_.txt', 'w') as f4:
    for row in distance_hist:
        f4.write(str(row))
        f4.write('\n')

#print(num_storms)
end_time = dt.datetime.now()
print(end_time - start_time)


bins = range(0, 10)
plt.hist([x[1] for x in distance_hist if x[1] > 0.1], bins = bins, edgecolor = 'black')
plt.show()
```

Appendix . Python Code for Investigation

```
import datetime as dt
import numpy as np
import math
from scipy.spatial import ConvexHull
import matplotlib.pyplot as plt
from haversine import Unit, haversine
import pandas as pd

def distance_nautical_miles(point_of_interest, line):
    lat1, lon1, = point_of_interest[0], point_of_interest[1]
    lat2, lon2, = line[0], line[1]
    return haversine((lat1, lon1), (lat2, lon2), unit = Unit.NAUTICAL_MILES)

lc39 = [28.573469, -80.651070, 'LC39']
slf = [28.59238, -80.66091, 'SLF']
IA = [28.092399, -80.641466, 'IA']
lc40 = [28.562091, -80.577385, 'LC40']
pafb = [28.234332,  -80.605998, 'PAFB']
lc37 = [28.531212, -80.564782, 'LC37']
spx_central = [28.417199, -80.604741, 'SPX_CENTRAL']
astrotech = [28.5256, -80.8209, 'ASTROTECH']

list_locations = [lc39, slf, IA, lc40, pafb, lc37, spx_central, astrotech]


def input_file(file_name):
    in1 = open(file_name)
    next(in1)
    next(in1)
    next(in1)
    input_reader = in1.readlines()
    raw_data = []
    center = [28.5387, -80.6428]
    for row in input_reader:
        split_data = row.split(' ')
        if split_data[0] =='\n' or split_data[0][0] == "c" or split_data[0][0] == '*' or split_data[1].split(':')[0] ==
            pass
        else:
            date = split_data[0].split('-')
            clock = split_data[1].split(':')
            sec = int(clock[2].split('.')[0])
            nano = int(clock[2].split('.')[1])
            lat = float(split_data[3])
            CG_or_CC = None
            if len(split_data) > 6:
                long = float(split_data[5])
                CG_or_CC = "cg"
            else:
                long = float(split_data[5].split("\\")[0])
                CG_or_CC = "cc"
            row_data = [dt.datetime(int(date[0]), int(date[1]), int(date[2]), int(clock[0]), int(clock[1]), sec, nano),
            if distance_nautical_miles(row_data[1:3], center) <= 25:
                #print(row_data)
                raw_data.append(row_data + [CG_or_CC])
            else:
                pass
    in1.close()
    return raw_data

raw_data = sorted(input_file('2019part1_total_combined_file.txt'))
def experiment_cg(row, center, center_warning, base_fail_cg, base_on):
    if distance_nautical_miles(row[1:3], center[0:2]) <= center_warning:
        base_fail_cg.append(row + [center[2]])
        if row[3] == 'cg':
            base_on = False
    elif row[3] == 'cg' and distance_nautical_miles(row[1:3], center[0:2]) <= 4: # test using 4 nautical miles
        base_on = False
    else:
        pass
    return base_fail_cg, base_on
def run_experiment_cg(raw_data):
    global list_locations
    start = dt.datetime.now()
    base_fail_cg = []
    base_on = True
    for center in list_locations:
        raw_data_part1 = [x for x in raw_data if distance_nautical_miles(x[1:3], center[0:2]) <= 5]
        center_warning = 1
        base_on = True
        for row in raw_data_part1:
            if base_on == True:
                base_fail_cg, base_on = experiment_cg(row, center, center_warning, base_fail_cg, base_on)
            else:
                if (row[0] - current_time).seconds >= 1800: #more than an hour passed last shut down
                    base_on = True
                    base_fail_cg, base_on = experiment_cg(row, center, center_warning, base_fail_cg, base_on)
```

```
            current_time = row[0]
    print(dt.datetime.now() - start)
    return base_fail_cg

def experiment_tl(row, center, center_warning, base_fail_tl, base_on):
    if distance_nautical_miles(row[1:3], center[0:2]) <= center_warning:
        base_fail_tl.append(row + [center[2]])
        if row[3] == 'cg' or row[3] =='cc': #
            base_on = False
    elif distance_nautical_miles(row[1:3], center[0:2]) <= 4:  #
        base_on = False #
    else:
        pass
    return base_fail_tl, base_on

def run_experiment_tl(raw_data):
    global list_locations
    start = dt.datetime.now()
    base_fail_tl = []
    base_on = True
    for center in list_locations:
        raw_data_part1 = [x for x in raw_data if distance_nautical_miles(x[1:3], center[0:2]) <= 5]
        #print([len(raw_data_part1), center[2]])
        center_warning = 1
        base_on = True
        for row in raw_data_part1:
            if base_on == True:
                base_fail_tl, base_on = experiment_tl(row, center, center_warning, base_fail_tl, base_on)
            else:
                if (row[0] - current_time).seconds >= 1800: #more than an hour passed last shut down
                    base_on = True
                    len_before = len(base_fail_tl)
                    base_fail_tl, base_on = experiment_tl(row, center, center_warning, base_fail_tl, base_on)
            current_time = row[0]
    print(dt.datetime.now() - start)
    return base_fail_tl

part2_2019 = sorted(input_file('2019part2_total_combined_file.txt'))
part3_2019 = sorted(input_file('2019part3_total_combined_file.txt'))
part4_2019 = sorted(input_file('2019part4_total_combined_file.txt'))

part2_2019_tl =  run_experiment_tl(part2_2019)
part3_2019_tl =  run_experiment_tl(part3_2019)
part4_2019_tl =  run_experiment_tl(part4_2019)
part1_2019_tl =  run_experiment_tl(raw_data)

part2_2019_cg =  run_experiment_cg(part2_2019)
part3_2019_cg =  run_experiment_cg(part3_2019)
part4_2019_cg =  run_experiment_cg(part4_2019)
part1_2019_cg =  run_experiment_cg(raw_data)

part1_2020 = sorted(input_file('2020part1_total_combined_file.txt'))
part2_2020 = sorted(input_file('2020part2_total_combined_file.txt'))
part3_2020 = sorted(input_file('2020part3_total_combined_file.txt'))
part4_2020 = sorted(input_file('2020part4_total_combined_file.txt'))

part1_2020_tl =  run_experiment_tl(part1_2020)
part2_2020_tl =  run_experiment_tl(part2_2020)
part3_2020_tl =  run_experiment_tl(part3_2020)
part4_2020_tl =  run_experiment_tl(part4_2020)

part1_2020_cg =  run_experiment_cg(part1_2020)
part2_2020_cg =  run_experiment_cg(part2_2020)
part3_2020_cg =  run_experiment_cg(part3_2020)
part4_2020_cg =  run_experiment_cg(part4_2020)

part1_2018 = sorted(input_file('2018_combined_txt_file.txt'))
part1_2017 = sorted(input_file('2017_combined_txt_file.txt'))

part1_2017_cg =  run_experiment_cg(part1_2017)
part1_2017_tl =  run_experiment_tl(part1_2017)
part1_2018_tl =  run_experiment_tl(part1_2018)
part1_2018_cg =  run_experiment_cg(part1_2018)
```

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From — To)*
24–03–2022	Master's Thesis	August 2020 — March 2022

4. TITLE AND SUBTITLE

Comparison of Lightning Warning Radii Distributions

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Maestas, Michael M., 2d Lt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-M-149

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Left Intentionally Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Previous research investigating lightning warning radii about the Cape Canaveral space launch facilities have focused on reducing these radii from either 5 nautical miles (NM) to 4 NM or from 6 NM to 5 NM depending on the structures being protected. Some of these findings have suggested the possibility of both a seasonal difference (warm versus cold) and lightning detection events (cloud-to-ground lightning (CG) or total lightning (TL)) impacting these radii and associated risk levels. Utilizing the 2017-2020 data provided by the 45th Weather Squadron at Patrick Space Force Base via the Mesoscale Eastern Range Lightning Information System (MERLIN), this thesis investigates these seasonal and data collection impacts. Our findings indicate that there is a substantial increased risk to warning radii's when just utilizing CG data in comparison to TL data. For the years studied, the mean risk for just using CG data was 5.94% in comparison to just 0.015% for the TL data for the 5 NM radii safety buffer. There were negligible seasonal differences between the warm season (May through September months) in comparison to the cold season (the remaining months) for using TL (0.015% versus 0.014%). In contrast, utilizing just CG data resulted in mean risk for the warm season of 4.86% compared to 11.56% for the cold season. Our recommendations are to utilize TL data where available. If using just CG data, then the risk of a lightning event occurring outside of a warning radii approximately doubles during the cold season in comparison to the warm season.

15. SUBJECT TERMS

applied statistics, lightning, programming

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Edward D. White III, AFIT/ENC
U	U	U	UU	94	**19b. TELEPHONE NUMBER** *(include area code)* (937) 255-3636, x4540; Edward.White@afit.edu