3-2021

CCSFS/KSC Total Lightning Warning Radii Optimization for MERLIN Using Preexisting Lightning Areas

Kimberly G. Holland

CCSFS/KSC Total Lightning Warning Radii Optimization for MERLIN Using Preexisting Lightning Areas

THESIS

Kimberly G. Holland, Captain, USAF

AFIT-ENS-MS-21-M-168

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-21-M-168

CCSFS/KSC TOTAL LIGHTNING WARNING RADII OPTIMIZATION FOR MERLIN USING PREEXISTING LIGHTNING AREAS

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

Kimberly G. Holland, B.S.
Captain, USAF

March 26, 2021

AFIT-ENS-MS-21-M-168

CCSFS/KSC TOTAL LIGHTNING WARNING RADII OPTIMIZATION FOR MERLIN USING PREEXISTING LIGHTNING AREAS

THESIS

Kimberly G. Holland, B.S.
Captain, USAF

Committee Membership:

Edward D. White, Ph.D
Chair

Timothy W. Holzmann, Lt Col, Ph.D
Member

Dawn Sanderson, Capt
Member

AFIT-ENS-MS-21-M-168

Abstract

The purpose of this research is to optimize lightning warning radii specifications for the 45th Space Wing (45 SW), thus reducing the number of unnecessary warnings that delay ground processing needed for space launch execution at Kennedy Space Center and Cape Canaveral Space Force Station. This thesis sought to answer two key research questions addressing: 1) What radius reduction effectively balances both safety and operations and do reduction recommendations from previous research align with results from the new detection system? 2) What insights can be gained from comparing measurement results for seasonal lightning events as well as lightning types? This study focused on recoil/dart leader events for both lightning types, cloud to ground and lightning aloft, occurring around the Cape Canaveral space launch facilities. Location information for these events are collected by the Mesoscale Eastern Range Lightning Information Network during 2017-2019. In this research, a clustering approach based on spatial and temporal criteria was applied to establish storm groupings. The Minimum Volume Enclosing Ellipsoid method was then utilized to generate the distance distribution for analysis on proposed radii reduction alternatives that would achieve the optimal balance between productivity and risk. Findings will show that a 1NM reduction from the radii baseline provides the optimum balance between operations and safety, resulting in reducing 5NM warning circles to 4NM, and 6NM warning circles to 5NM, comparable to results in previous research.

Acknowledgements

I would like to thank Mr. William Roeder from the 45th Operations Group for his invaluable insights. I would also like to thank my faculty research advisor Dr. Edward White, and committee members, Lt Col Holzmann and Capt Dawn Sanderson, for their guidance and encouragement. Finally, I would like to thank my husband, for his unrelenting patience and support.

Kimberly G. Holland

Table of Contents

Page

List of Figures

List of Tables

CCSFS/KSC TOTAL LIGHTNING WARNING RADII OPTIMIZATION FOR MERLIN USING PREEXISTING LIGHTNING AREAS

I. Introduction

1.1 Problem Background

Equally fascinating and destructive, lightning stands to be one of the most concerning weather occurrences that tests the safety and success of space launches at Cape Canaveral Space Force Station and Kennedy Space Center (CCSFS/KSC) [2]. Situated on the east central coast of Florida, a location known for having the most lightning strikes in North America [2], the CCSFS/KSC places lightning safety as one of their top priorities [6] amid the growing endeavor of space exploits. Specifically, CCSFS/KSC conducts its mission with careful consideration of lightning proximity to the launch installation, as safety for installation personnel and resources are paramount to the mission.

With this priority in mind, the 45 Space Wing (SW) continually strives to improve its effectiveness in meeting their core objective, executing successful launching of spacecraft. This prompts leadership and experts from or in support of the 45th Weather Squadron (WS) to place significant effort in developing new methods that refine the recognition and prediction of lightning strikes as well as mitigate the operational impact of those lightning strikes while maintaining personnel safety and resource protection. The most recent and well improved lightning locating system employed by the 45th WS is called the Mesoscale Eastern Range Lightning Information Network (MERLIN), capable of collecting lightning strikes with higher accuracy than its predecessor, Four Dimensional Lightning Surveillance System (4DLSS) [2].

Alongside these developments, prior research maintained the standing guidance that lightning is imminent when it is within a 5 nautical mile (NM) margin, but the assessments fixated on measuring from the center of the lightning origin [4, 7, 8]. This caused perhaps too wide of a warning area, which has negatively affected space launch operations by decreasing operational availability. With the advent of new lightning systems continually being updated over time at 45 WS, recent research has explored a different lightning point of origin for measurement that would better align with the basis of measurement from the more recent systems. The studies by the American Meteorology Unit (AMU) and Sanderson et al. (2020) sought to determine if measuring the distance from the edge of a preexisting lightning area to a lightning strike better reflected the true risk. With the realization that the measurement of lightning warning radii using the edge aligned better with 45WS lightning location systems, as opposed to using the origin of the flashes or the core of the thunderstorm, this suggested that a smaller warning radii could be implemented. Using data collected from the lightning aloft component of the 4DLSS system called Lightning Detection and Ranging II (LDAR II), both of these studies concluded that the edge of the lightning group, instead of its center, provided a more accurate assessment of proposed radii reduction options that would maintain minimal risk. Additionally, both suggested a lightning radii of 4 NM as a viable option [3, 9]. While this advancement serves to ensure alignment with LDAR II, further assessment is needed to ensure it also aligns with the current lightning location system, MERLIN.

1.2 Research Objectives

This research focuses on determining whether measuring the distance from the edge of a preexisting lightning area using the most relevant data and method, yields a similar or more improved lightning warning radii result than the confirmed 4NM radii distance reduction from the recent LDAR II studies (a 1NM reduction in comparison to the status quo of 5NM). The effort strives to build on current processes and research insights that have been suggested for application with focus on the newest lightning location method employed, the MERLIN. The main objectives of this research seek to answer the following key questions:

1. What optimal lightning warning radius preserves minimal risk to loss of life and resources while increasing space launch productivity, based on MERLIN data collected during the summer months of 2017-2019 (synchronous with the LDAR II study pertaining to summer months of 2013-2016)?

2. What is the comparison between distance distribution and risk findings with regard to the MERLIN assessment and results from the recent LDAR II studies?

A secondary objective to take into consideration is the inclusion of present data to refine the accuracy of results, with focus on the following:

3. Does the optimized radius for the summer season lightning (May-September) apply to the relatively rarer cold season lightning at CCSFS/KSC (October-April) for 2017-2019?

4. What other insights can be gained by considering other distance distribution comparisons such as types of lightning (cloud to ground and lightning aloft) or occurrences by year?

1.3 Document Overview

Chapter I establishes the research focus and goals to be attained during the study. Chapter II reviews a detailed background and evolution of lightning governance, collection and prediction methods, and research as well as meteorological and mathematical concepts pertaining to lightning detection. Chapter III presents the method of grouping events using spatial and temporal clustering as well as measuring distances using the minimum volume enclosing ellipsoid (MVEE) technique. Chapter IV present analysis and results derived from the methodology's resulting distance distribution. Finally, Chapter V provides conclusions and recommendations for implementation.

II. Literature Review

To provide an accurate and relevant lightning radii assessment that optimizes the safety and productivity of space launches, this study begins with a review of previous work to gain a foundation on the fundamental concepts that would drive the research. This review is initiated by diving into the process of lightning strikes to better understand the behavior of both cloud to ground lightning and lightning aloft. Further review encapsulates the evolution of the lightning tracking system, lightning safety criteria, and methods applied in lightning research. Finally, a review of clustering and ellipse fitting concepts helps to determine lightning event boundaries and the distance of lightning flash leaders from the edge of a pre-existing boundary.

2.1 Lightning Strikes

Lightning is an unpredictable natural phenomenon that is a cause for both wonder and concern due to its magnificence and destructive capability. It can formulate from convective systems of varying width, depth, temperature and humidity and are usually present in storms that produce a variety of meteorological conditions, such as damaging winds, hail, tornadoes, and hurricanes.[1, 10] Lightning's pattern of approach is erratic and much is yet to be learned about the lightning strike as its exact location and timing is unpredictable. What is known about lightning strikes so far is that they accompany thunderstorms that contain the presence of a mixed-phase region, requiring the presence of water vapor, atmospheric instability, vertical cloud buoyancy, and aerosol particles [1].

2.1.1 The Lightning Flash

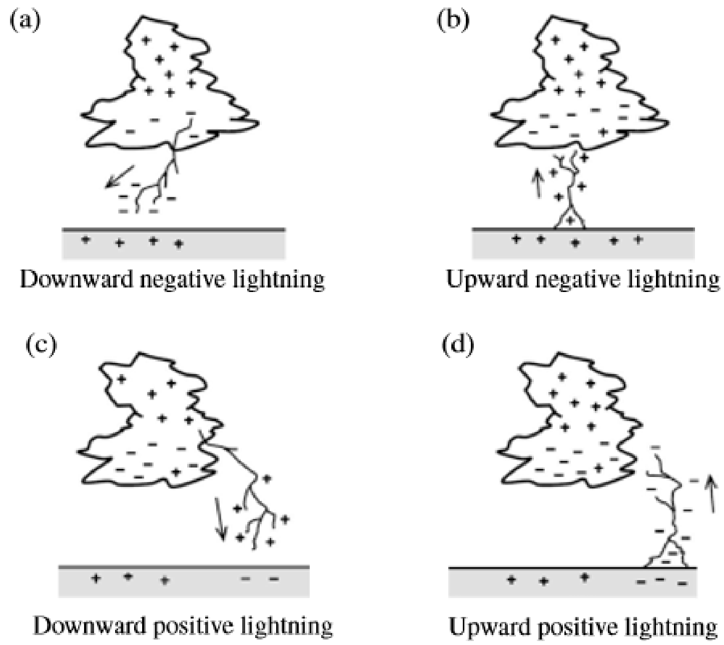
The spark of a lightning flash comes from the electrification of a mix of particles that are nested within a thundercloud. These electric charges are housed in the area of a cumulonimbus with "-a net positive charge near the top, a net negative charge below it, and an additional positive charge at the bottom" [11]. This region in effect allows electron and ion detachment and recombination to occur. The subsequent thermal ionization due to temperature causes a massive electric discharge called an arc discharge, formally observed as the lightning stroke [12].

2.1.2 Types of Lightning

In order for the arc discharge to occur, the electric emissions from a source must meet an opposite charge from a secondary source. There are several combinations of source interactions that can occur to create the lightning strike. The two main categories that define those interactions are called cloud-to-ground lightning (CG) and lightning aloft (CC), which encompasses intra-cloud lighting, inter-cloud lightning, and cloud-to-air lightning. There is also a relatively rare ground-cloud lightning. As a whole, these lightning types make up what is referred to as Total Lightning. Lightning detection systems aim to collect information on Total Lightning to gain a complete picture of lightning behavior for the most accurate predictions.

CG lightning specifically focuses on electrification transference between the cloud source and the ground. In this category, a lightning flash will generally find the path of least electrical resistance by seeking the most conductive way to reach its destination. Once it achieves any form of ground connection, such as making contact with the peak of a land based structure, the lightning flash generated from the interaction will usually last for 0.5 seconds and will incur an average of 4-5 strokes [1]. CG interactions fall under four classifications, as depicted in Figure 1. Each classification is based on the type of discharge emitted (whether positive or negative) and the type of stroke conducted (upward or downward). Of the four, downward negative lightning flashes encompass 90% of CG flashes [11].

Figure 1: CG Classifications



CC on the other hand, involves lightning occurrences within its cloud origin, between two cloud formations, or by cloud to air interactions. Lightning occurrences are more common in this category, encompassing 70% of TL [13], but its data is much harder to collect as its starting and end points are harder to track than the former lightning type [11]. This is because CC's energy output is ten times weaker than CG's [13]. .

2.1.3 Lightning Flash Leaders

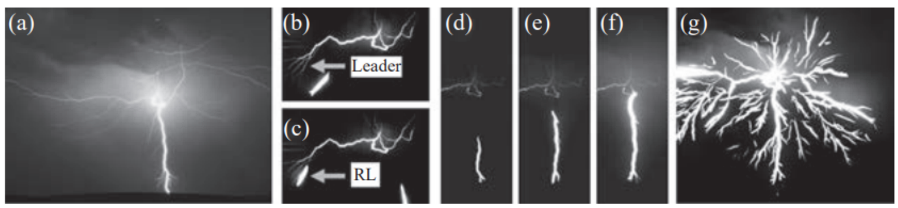
Different lightning location systems detect different parts of the lightning flash. Thus, when a system collects lightning data, a certain component of the lightning flash construct is identified for data collect. During a lightning event, these lightning flash components usually represent different types of leaders. In this study, stepped (or step), recoil, and dart leaders are of interest.

At the onset of a lightning flash, an electric charge (whether positive or negative) attempts to find a path to a source with an opposite charge to create a connection that would allow an electric current to run through, neutralizing the opposite charge. This type of electric charge is referred to as the original step leader. This initial leader branches out to create many subsequent step leaders, each one proceeding in a step and pause manner while racing to find the nearest source for charge pairing. Eventually, one of the step leaders will make the connection, providing the main conduit to achieve the initial lightning stroke. Step leaders were identified and collected by the 4DLSS detection system.

Once a stepped leader establishes connection with an opposite charge source, leading the first strike, an aperture for a return stroke is established. In this part of the lightning progression, source potential can now travel back up the connected path. [1] This interaction gives way for dart leaders to use the established path to create subsequent lightning strokes. Many of the step leaders can become dart leaders as they use the open channel to travel downward instead of continuing to seek an opposite charge, as they opt to travel to the path of least resistance.

Also occurring simultaneously after a return stroke, is the formation of recoil leaders (also known as recoil streamers). Recoil leaders are energy branches protruding from positive channels such as a return stroke (specifically as a positive rebound after a touch down from a negative leader) or from the formation of a positive leader. [1] A recoil leader's purpose is to re-energize one of the gradually decaying step leaders, turning them into dart leaders. Thus, there are significantly less recoil leaders than there are step leaders. With this notion, one recoil leader could serve to represent several step leaders [2]. In the same aspect, a lightning flash may still contain several recoil leaders. Shown in Figure 2, an example of a recoil leader is seen making it's entrance and moving towards a step leader after an initial CG strike to make a re-connection (frame c). This type of charge is identified and collected by the MERLIN detection system.

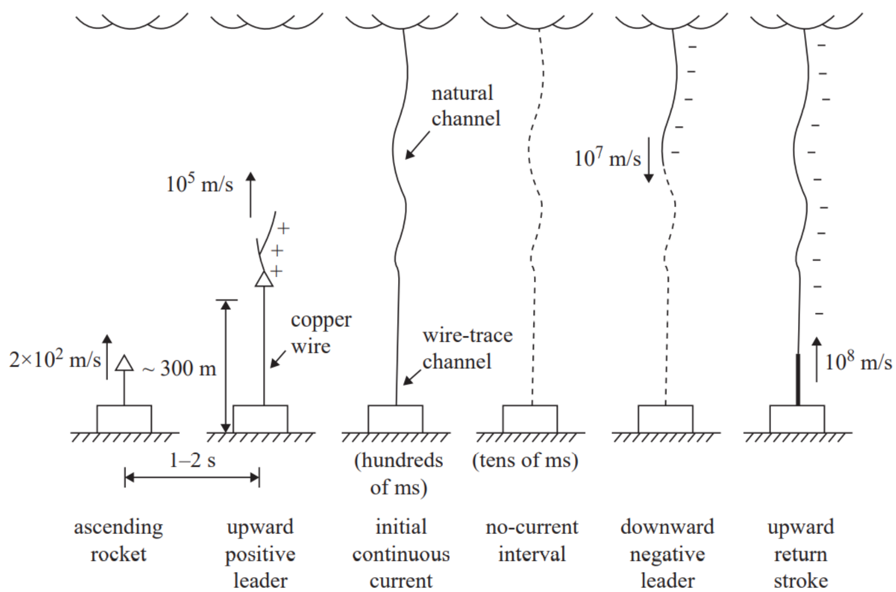
Figure 2: Recoil Leader (RL) Makes Contact with Step Leader [1]



2.1.4 Lightning Propagation

Lightning can be generated by either natural or triggered lightning. Inherent in its name, natural lightning is brought forth by electrical discharges naturally occurring in the presence of favorable atmospheric conditions. These conditions are what has been described of lightning so far, where discharges and return strokes are conducted between a primary source (cloud) meeting a secondary source (cloud, air, or ground). To the contrary, triggered lightning is sparked from an aberrant source, separate from the usual natural occurrence. Rather, a lightning charge in this manner is "triggered" by an object that can give off an electrical discharge. Given the right conditions from a cloud overhead, the object provides a direct path for a discharge to initiate. Figure 3 shows an example of rocket-triggered lightning, where a positive discharge from an object interacts with a cloud overhead, triggering a negative downward stroke. Using an electric field to measure atmospheric conditions in Florida, a cloud that is sensed to provide an electrostatic charge, with an absolute value of 4–10 kV$*m^1$, is determined as having favorable conditions for triggered lightning [1]. This type of condition became a concern for KSC/CCSFS in November 14, 1969, when the Apollo XII space craft triggered two lightning events during its space launch, causing a failure in some of the space craft's components and nearly costing the lives of three astronauts [14]. Knowing that lightning can still be initiated when there is enough electrical charge in the air from a cloud overhead, this paved the way to develop a launch avoidance criteria. The data and focus for this study is centered on natural lightning, but it is important to note that triggered lightning was pivotal to the establishment of safety requirements at KSC/CCSFS.

Figure 3: Rocket-Triggered Lightning [1]



2.2 Lightning Protocols

2.2.1 Establishment of Safety Criteria

The first official safety guidance for space launches, called the Lightning Launch Commit Criteria (LLCC), was established in 1970 shortly after Apollo XII's triggered lightning incident [14]. In this guideline, a 5 nautical mile (NM) radius was established to ensure the safety of ground and flight operations during a space launch event.

As research and collection methods improved, refined protocols were established.

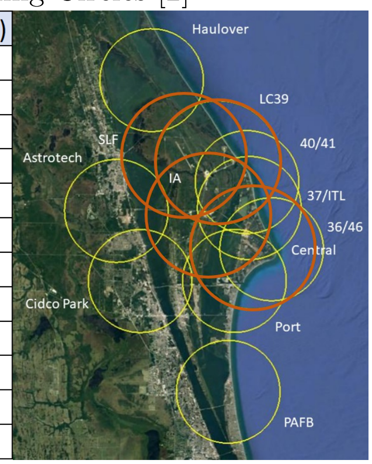
In 1998, members of the Lightning Safety Group convened at the American Meteorological Society Annual Meeting in Arizona to discuss further measures that needed to be established to ensure safety for the general population [7]. During this discussion, the "30-30" rule was established, denoting the first 30 as number of seconds between thunder and lightning and the second 30 as the number of minutes before an all clear signal is sent out once a thunder or lightning occurrence has been observed. Since then, modifications have been made to the guideline, resulting in three safety slogans that were set in place for the public awareness and adherence. The three safety slogans were "No place outside is safe when thunderstorms are in the area!", "When thunder roars, go indoors!", and "Half an hour since thunder roars, now it's safe to go outdoors!" [13]. These slogans are the general guideline for public lightning in the current setting.

2.2.2 Present Day Safety Requirements at CCSFS/KSC

Taking into account lightning safety criteria at the present time, KSC/CCSFS has a two-phase lightning warning policy. In this policy, provided by Roeder [13], Phase I involves issuing a Lightning Watch with a desired lead time of 30 minutes for Total Lightning that is expected within any of the 12 lightning warning circles placed among the Cape Canaveral space launch facilities. The lightning watch is issued when either a preexisting lightning system is approaching, or a locally developing thunderstorm is expected to eventually form inside the warning circle. Next, Phase II involves issuing a Lightning Warning for thundercloud systems that are occurring within the established lightning warning circles [15]. In this phase, the lightning warning is only dismissed after 15 minutes of lightning inactivity has passed. The lightning warning may be left in effect at the discretion of the forecaster if more lightning is possible in the very near future. Some of the conditions that would prompt a forecaster to keep the lightning warning in effect involve three conditions; if the thunderstorm is expected to redevelop, if a new thunderstorm is expected to approach or develop shortly after the first, or if there is evidence that indicates that the risk of lightning remains. As shown in Figure 4, the radii of those lightning warning circles are either 5NM (yellow circles) for a single small facility or several closely located facilities, or 6NM (orange circles) for a single large facility or several widely spaced facilities. The accompanying table references name of each radii, the facility it belongs to, and the center locations of each radius, represented in latitude and longitude.

Figure 4: Cape Canaveral Lightning Warning Circles [2]

Facility	Name	Center Lat	Center Long	Radius (NM)
KSC	HAULOVER	28.7364	-80.7547	5
CCSFS	40/41	28.572716	-80.579891	5
CCSFS	37/ITL	28.531489	-80.574834	5
CCSFS	36/46	28.464949	-80.533056	5
CCSFS	PORT	28.41336	-80.6	5
PSFB	PSFB	28.23408	-80.60976	5
OFFSITE	ASTROTECH	28.52424	-80.8172	5
OFFSITE	CIDCO PARK	28.411806	-80.772892	5
KSC	LC39	28.60419	-80.6317	6
KSC	SLF	28.61475	-80.69486	6
KSC	KSC INDUSTRIAL AREA	28.52	-80.65	6
CCSFS	CAPE CENTRAL	28.46741	-80.56711	6



2.3 Evolution of Lightning Tracking Systems Used by KSC

Many lightning location systems have been developed to better understand and predict the behaviour of lightning as well as aid in determining lightning warnings in support of operations at CCSFS/KSC. This support includes issuing lightning watches and warnings, evaluating the Lightning Launch Commit Criteria, supporting ground processing operations several months leading to a space launch, and evaluating the risk of induced current damage of electronics due to cloud-to-ground strike occurrences [13]. Partaking in this development, the following lightning systems have contributed greatly to or continues to serve the overall lightning exploration effort at

Cape Canaveral's space launch facilities.

2.3.1 National Lightning Detection Network

The National Lightning Detection Network (NLDN) was established in the 1980's, to continuously gather real-time data on ground flash density (GFD) during CG lightning events [16]. It began with the initiation of a few Direction Finder (DF) sensors that were established in New York, handled by a newly developed team headed by Richard Orville. This emerging network was eventually merged together with two other competing lightning networks and its DF sensor footprint was expanded over the course of time to cover the entirety of the United States and Canada. During this time, NLDN underwent two iterations and its sensors were eventually upgraded with a new model called the Improved Accuracy from Combined Technology (IMPACT) sensor, which had the capability of the previous sensor plus the capacity to record time on arrival (ToA) [12]. In this updated version, 114 sensors are able to detect both CG and CC flashes with 80-90% detection efficiency [17] as well as 0.6 km location accuracy [18, 19]. This network continues to collect lightning data to date and is used to supplement the shorter range higher-performance lightning location systems used by 45WS. NLDN was initially integrated with MERLIN, but it was later discontinued due to occassional lightning solutions that displayed very large location errors and very eccentric location error ellipses. [13]. NLDN does continue to serve as a backup system for the CCSFS/KSC for operations outside the MERLIN's effective range.

2.3.2 Weather Surveillance Doppler Radar

The inception of the Weather Surveillance Radar - 1988 Doppler (WSR-88D) Radar Operations Center (ROC) was established through the Next Generation Weather Radar (NEXRAD) Program in the year 1988 and it maintains its operational use in the present time throughout 164 sites in the U.S. Its employment of Radio Detection and Ranging (RADAR) tracking allows for wide swath surveillance of many atmospheric elements that make up a developing thunderstorm, to include lightning [20]. WSR-88D sensors uses reflectivity to detect the onset of lightning as severe temperatures coincide with the electrification process. Its most recent iteration in 2013 applied polarimetric capabilities in order to eliminate unnecessary noise, such as birds, insects, and ground. While different versions of the WSR have been employed at the Cape Canaveral installations in the past, at the current setting, the 45 WS uses the WSR-88D/Melbourne Radar as a backup to their more recent primary Weather Surveillance Radar, the RadTech 250/43 [13].

2.3.3 Cloud to Ground Lightning Surveillance System (CGLSS)

CGLSS became operational in 1989, utilizing six sensors comparable to the IMPACT sensors from NDLN [16]. This system specifically gathers data on CG flashes. CGLSS was needed to provide better location accuracy and detection efficiency than NLDN in the CCSFS/KSC area. This requirement was primarily required for the daily lightning reports issued by 45WS to help space launch customers assess the risk of induced current damage from nearby lightning on electronics in satellite payloads, rocket avionics, and test equipment [13]. Its first iteration's main limitations involved collection solely on just the initial return stroke and accommodates a central processing unit that relies on an outdated and unsupported system. Thus, a newer central processing unit supplanted the former along with newer sensors that collects on all return strokes, labeled as CGLSS-II. This upgrade allowed for a 98% location accuracy within the sensor perimeter [18]. Additionally, it detects 250% more flashes than the former model, specifically on individual return strokes [21]. However, one of the six sensors have been incapacitated since 2009 [19].

2.3.4 Lightning Detection and Ranging (LDAR)

The Lightning Detection and Ranging (LDAR) system was first established in 1971. For over 40 years, KSC has relied on the LDAR system, which uses Very High Frequency (VHF) signals to pinpoint the location of step leaders to determine the presence of a lightning event [22]. The visual representation of these step leaders allowed forecasters to evaluate events that are specific to lightning aloft. During its utility, two iterations of the system were employed, the LDAR I and II.

2.3.4.1 LDAR I

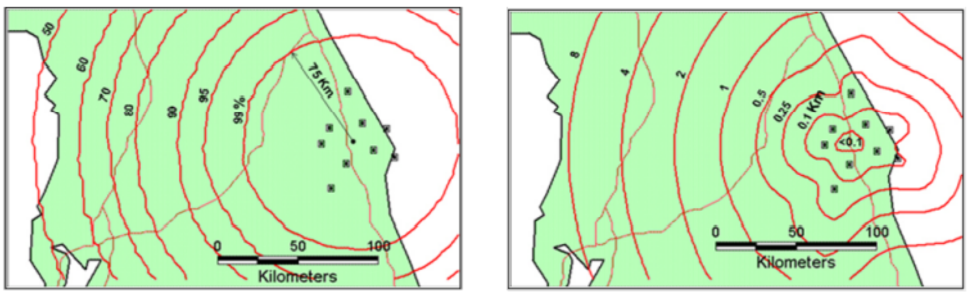
The first of its class, LDAR I was utilized by KSC between 1971 to 2008. The system construct comprised of seven total receivers that detected step leader discharges, with six of the VHF radio antennas surrounding a central VHF receiver to ensure line of sight (LOS). This allowed LDAR to paint a 3-D spatial representation of the lightning data, but only on lightning aloft specifically [16]. For a collective assessment, LDAR was implemented in conjunction with CGLSS to more accurately analyze the behavior of lightning strikes [3].

2.3.4.2 LDAR II

From 2008 to 2016, LDAR II replaced its predecessor, with notable improvements. In this overhaul, nine new sensors replaced the seven former, with capability to gather data beyond line of sight (BLOS). This capability allowed the new sensors to have wider separation, covering more area to attain higher accuracy [23]. Thus, the spacing of the sensors were doubled. With the increase and improvement in sensor capability, LDAR II's detection efficiency surpassed its former model by 140%. [21] Though an improvement from the last iteration, LDAR II still has its limitations to note. Major limitations of concern are its limits on detection rate and location accuracy. LDAR

II's detection rate is at 99% up to 75km from the central site. As lightning detection expands beyond 75km of LDAR II's central site, the detection rate decreases by 5-10% for each estimated 25 km increment. Similarly, LDAR II's location accuracy steadily declines in confidence as its area of detection extends past the sensor groupings. A detailed breakdown of both limitations are depicted in Figure 5.

Figure 5: LDAR Detection Rate and Location Accuracy



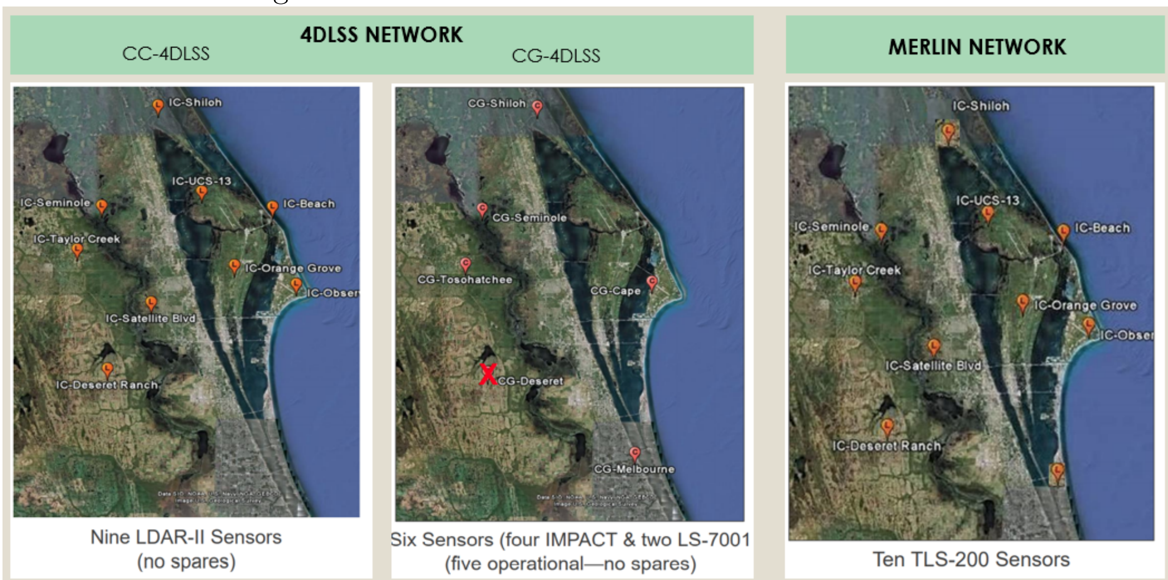
a) LDAR Detection Rate

b) LDAR Location Accuracy (50% confidence)

2.3.5 Four Dimensional Lightning Surveillance System (4DLSS)

4DLSS makes up the framework that employs the most updated central processing system and analyzes LDAR II in tandem with CGLSS II to provide a full picture of lightning strikes at KSC. Sensor placement for all 15 working sensors is shown in Figure 6, where CGLSS II sensors outline the group of LDAR II sensors. Evaluated by Roeder[21], this system achieves at least a 98% and 100% detection efficiency for CG and CC, respectively. Additionally, it achieves a 140% CC and 250% CG relative detection efficiency in comparison to the previous LDAR I and CGLSS I pairing. 4DLSS was also more effective in noise reduction in comparison to the older system, as electrostatic emissions from planes were weak and can be filtered out by the improved system's quality control algorithm [23].

Figure 6: 4DLSS and MERLIN Sensor Locations



2.3.6 Mesoscale Eastern Range Lightning Information Network

The most recent lightning detection system currently employed by 45WS is the Mesoscale Eastern Range Lightning Information Network (MERLIN). Replacing the previous 4DLSS system in place at CCSFS/KSC, this system has the ability to identify Total Lightning (TL); both CG and CC lightning types. To determine the presence of a lightning, it detects electrical activity after the occurrence of the initial stroke, specifically on recoil/dart leaders [2]. MERLIN has the capability to work with integrated NLDN sensors to account for improved CG detection for longer ranges or when sensors are down. It is important to note that since 2018, the NLDN sensors were turned off due to occasional large location error ellipses with excessively large eccentricity [13]. The new system also has exceptional performance for CG lightning with a median location accuracy of 58m and detection efficiency over 99% inside the network of sensors. Additionally, it boasts a 10% increase in stroke detection efficiency in comparison to the system used previously, 4DLSS, within a 30NM radius

[2]. MERLIN has exceptional performance for CG lightning with a median location accuracy of 58m and a detection efficiency over 99% inside its network of sensors [13].

In contrast, the 4DLSS has a higher detection rate than MERLIN for lightning aloft past 30NM. Another limitation to MERLIN is that it is only able to provide a 2D represenation of lightning aloft, in comparison to LDAR II which has the capability to determine the height of lightning aloft in order to present a 3D description [2]. Other significant improvements to MERLIN in comparison to both CGLSS and LDAR II is outlined in Table 1. As the most advanced lightning detection system at CCSFS/KSC to date, MERLIN has collected total lightning data within the past seven years [2].

Table 1: System Comparison Between 4DLSS and MERLIN

COMPARISONS BETWEEN 4DLSS AND MERLIN				
	DETAILS	CG-4DLSS	CC-4DLSS	MERLIN
SPECIFICATIONS	SENSOR MODEL	IMPACT/LS-7001	LDAR II	TLS-I200
	SENSOR COUNT	4/2 (only 5 operational)	9	10
	EVENTS DETECTED	Cloud to Ground - Return Strokes	Lightning Aloft - Step Leaders	TOTAL LIGHTNING: Cloud to Ground (CG) - Return Strokes Lightning Aloft (CC) - Recoil/Dart Leaders
	REPORT CAPABILITIES	CG - 2D location (x,y), date/time, peak current, polarity, error ellipse location	CC - 3D location (x,y), date/time	CG - 2D location (x,y), date/time, peak current, polarity, error ellipse location CC - 2D location (x,y), date/time
	DETECTION METHOD	MDF/TOA	TOA	CG - MDF/TOA CC - INFEROMETRY
	FREQUENCY BAND	LF	VHF	CG - LF/HF CC - VHF
	SIGNAL PROCESSING	ANALOG	ANALOG	DIGITAL
	NLDN CG CAPABILITY	No sensor integration: Long-range CG performance degrades to zero	N/A	10 sensor integration: Long-range CG performance and loss of sensors degrades to NLDN raw observations
	CENTER SITE COORDS			
PERFORMANCE	EVENT DETECTION EFFICIENCY %	STROKE - 82%	LEADERS - 70%	STROKE - 92% LEADERS - ~92%
	FLASH DETECTION EFFICIENCY %	96%	100%	100%
	LOCATION ACCURACY	58 m	100 m	CG - 350m CC - 500 m
	DETECTION FAILURE RATE		5% per 50 km	
	FALSE DETECTIONS	0%	Rare and identifiable	0%
	CG RETURN STROKE MISSES	5% OF STRONG STROKES	N/A	0.5% OF STRONG STROKES
	LIGHTNING ALOFT WITHIN NETWORK	N/A	LOWER DETECTION %	30% MORE DETECTIONS
	LIGHTNING ALOFT BEYOND ~30nmi	N/A	HIGHER DETECTION %	LOWER DETECTION %

2.4 Lightning Studies

2.4.1 Assessment Comparison of Previous Detection Systems

Previous research has been conducted to compare the current lightning radii criteria with the determined distance distribution of a lightning event. Renner [20] and Cox [24] explored the WSR-88D Centroid Method, focusing on the center of the storm as their distance measurement origin. Another method called Distance Between Successive Flashes (DBSF) was applied by Krider [25], Lopez [17], Parsons [8], and Cox [24]. In this method, lightning clusters were predetermined and the center of the clusters served as the distance origin for measurement.

Poehler [26] and McNamara [4] applied yet another method based on Horizontal Ground Strike Distance. In their studies, the center of predetermined flash groups became the basis for measurement. Poehler first introduced the concept, focusing only on one storm. McNamara adapted the methodology, applying four years worth of data from LDAR and NLDN to assess lightning events during all four seasons. Using both data sources, McNamara measured CG distance from a lightning stroke's assessed point of origin to the ground location where the lightning hit. His results, shown in Table 2, demonstrated that the radius requirement needed to be increased with continued use of those detection methods. This was due to a 28% probability that lightning flashes would occur outside the 5NM radius. The similarity in these previous methods is the use of the center of a lightning area as the origin for measuring the lightning distance. A comparison of these methods and the results rendered are displayed in Table 2.

2.4.2 Assessment Comparisons of Recent Detection Systems

As lightning detection systems advanced over the years, the method in collecting lightning information also changed. Additionally, the understanding that lightning warning measurement needed to match how detection systems located lightning was realized. Thus, more recent studies moved to a different approach when defining criteria for the measurement point of origin, in order to accurately reflect the measurement criteria based on 45 WS systems. Instead of the center of the lightning area as a starting point, these works determined that lightning can be more accurately measured from the edge of a predetermined lightning area. Two studies were conducted using LDAR II data to assess the performance of the more recently used 45 WS systems with regard to its effectiveness in predicting lightning, to help determine its impact on current policy. These studies are also shown in Table 2 for comparison.

Table 2: Previous Research [4]

Method Assessed	Distance Measurement Origin	Lightning Detection System	Scientific Analyst	Lightning Data	Location	Scope	Results
WSR-88 D Centroid Method	From center of thunderstorm (based on radar echo return)	WSR-88 D	Renner	CG	Gulf Coast Southern Plains	April - July	75% < 8.69 NM (10mi)
		WSR-88 D	Cox	CG	Southeast Coast	April - July	32-39% > 5 NM
Distance Between Successive Flashes	From center of pre-determined lightning clusters	LLP DF Network	Krider	CG	Florida	3 Storms	Avg 1.9 - 2.5 NM
		NLDN	Lopez, Holle	CG	Florida, Oklahoma, Colorado	1995 1 Storm 1996	75% < 4.9 NM
		WSR-88 D	Cox	CG	Southeast Coast	April - July	30% > 5 NM
		WSR-88 D	Parsons	CG	Continental U.S.	1995 - 1999	65 % < 5 NM
Horizontal Ground Strike Distance	From center of pre-determined flash groupings	LDAR I *(and other sources)	Poehler	CG	Florida	1 Storm 13 CG Flashes	100% < 4.8 NM
		LDAR I and NLDN	McNamara	CG	Florida	Mar - Dec 1997-2000	71.6% < 5NM
Distance Beyond a Pre-existing Lightning Area	From the edge of pre-determined flash groupings	LDAR II	Hinkley, Huddleston, Roeder	Total	Florida	2013	<4 NM radius threshold
		LDAR II	Sanderson	Total	Florida	May - Sept 2013-2016	4 NM radius threshold

*other sources include radar returns, ground strike locations, airborne field strength, ground based field strength

2.4.2.1 LDAR II Assessment by Hinkley

Led by Hinkley, [9] the American Meteorological Unit (AMU) team conducted research to determine lightning distribution beyond a pre-existing boundary. They analyzed lightning events that occurred year-round between 2013-2016 within 50km of LDAR II's central site. This study focused on a new approach to measuring lightning distances. Instead of considering distance from the center of predetermined flash groups, measurement from the edge of flash groups were the analytic focus. Hinkley applied the convex hull approach to flashified data to determine a flash group boundary and the events that lay on the boundary represented each flash.

Flashifying involves taking raw data, referred to as source points to represent step leader events, and grouping them together based on a time and distance criteria of 0.3 seconds and 3km, respectively. Then, the convex hull approach creates a boundary around an initial flash, which contains the first grouping of source points defined by flashified data. Subsequent flashes are included in the boundary based on time and location criteria. If the flash does not meet the spatial and temporal constraints, then it is set aside to be included in the next storm.

As the process iterates, any flash that is 15 minutes old is also discarded from the boundary and the boundary is refitted to reflect movement of the storm. Lightning distances are then gathered and the lightning distribution evaluated. To achieve CCSFS/KSC's safety requirement, the minimum reduction of the lightning warning radii was assessed to ensure it satisfied the national safety standard of 10^{-5} deaths per year. In this evaluation, Hinkley applied a conservative 30m proximity assumption, as the lethality of lightning in close proximity is not well known. The results gathered from this approach showed that the radius required is about 3.99NM. Thus, AMU determined that a 4NM standoff suffices as a recommended radii warning circle, to once more provide another added measure of conservative safety.

2.4.2.2 LDAR II Assessment by Sanderson

Sanderson [3] also conducted research on LDAR II lightning flashes between 2013-2016, but focused only on the summer months (May-Sep), when lightning events more often occur. Similar to AMU, Sanderson also used flashified data and focused on the edge of a preexisting lightning area as a measurement origin. However, her research differed in several areas. First, she reduced the lightning data to exclude any flashes that occurred greater than 25NM from the central site. While she also applied the convex hull approach to isolate a flash boundary from raw source points in order to have less points represent a flash, she also proceeded with applying the ellipse fitting approach instead of a polygon when it came to grouping those newly represented flashes.

Her spatial and temporal parameters also differed slightly. For instance, a flash must be within 16km of the edge of the preexisting lightning area to be measured and included within the boundary of the next ellipse fitting. Further, she discarded flashes that are 10 minutes old. Once all measurements were gathered, she determined that a Weibull distribution was an appropriate fit to distance data, in preparation for evaluation. During analysis, she assessed productivity versus risk values to determine the minimum radii reduction to be considered. Sanderson concluded that a reduction to 4NM radii would be the lowest option to consider, observing that it achieves the optimal balance between safety and operational impact, as the risk rate incurred a sharp incline for radii values less than 4NM while productivity had the steepest incline at 4NM. This radii consideration allows for a 0.277% average failure rate, but generates a 30% productivity increase. Thus, Sanderson confirmed that the radius requirement could indeed be reduced to 4 NM.

Additionally, Sanderson initiated analysis of early lightning from the safety perspective. In this portion of her study, the first few flashes were also considered for locally developing thunderstorms. In this observation, if early lightning struck further than lightning from the aforementioned preexisting area, then the previous optimized warning radii would provide inadequate safety for thunderstorms developing just outside the warning circle. At this time, the expectation is that lightning will not strike further than what has been observed in preexisting lightning areas. Lightning safety is critical at CCSFS/KSC that this case certainly warranted further investigation and confirmation.

2.5 Principles for Insight and Application

2.5.1 Clustering Concepts

In this research, to identify flash point groupings that constitute a lightning event, cluster analysis was considered. This method is an unsupervised learning approach that helps to classify data points into distinct subgroups, based on similarities or dissimilarities. Here, two clustering methods were explored.

2.5.1.1 K Means Clustering

In her latest research, Sanderson suggested the application of K-Means Clustering to group flash points into one lightning event. This method seeks to iteratively partition data into a number of K distinct clusters, determined by minimizing the within sums of squares of all points in each cluster, defined as:

$$\underset{C_1 \ldots C_k}{\text{minimize}} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{P} (x_i j - x'_i j)^2 \right\}$$

Here, $C_1, ..., C_K$ denotes sets of observations in each cluster. These sets satisfy two properties; $C_1 \cup C_2 \cup ... \cup C_K = 1, ..., n$ and $C_k \cap C_{k'} = \emptyset, \forall k \neq k'$. Then, $|C_k|$ denotes the number of observations in the kth cluster. Moreover, the within-cluster variation is the sum of all pairwise Euclidean distances, defined as:

$$\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{P} (x_i j - x'_i j)^2$$

Shown in Algorithm 1, the value k is predetermined by the analyst and it initializes the algorithm by choosing k centroids, concluding once the average distance between cluster points and cluster centroid no longer change.

Algorithm 1 K-Means

1: Randomly assign a number, from 1 to K, to each of the observations. These serve as initial cluster assignments for the observations.
2: For each of the K clusters, compute the cluster centroid. The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster
3: Assign each observation to the cluster whose centroid is closest (where closest is defined using Euclidean distance).
4: Iterate until the cluster assignments stop changing

The use of K-Means clustering is advantageous when handling very large sets data, such as that captured by MERLIN. The disadvantage in this method is that it requires a predetermined k value. This value cannot be determined at the onset, thus the method would require repetition, with different values of k and a comparison of those values to determine which is the best k. Another issue encountered is that, the cluster results can vary even for one k value. This is because the cluster outcome could differ based on the random initial cluster assignment in Step 1 of the algorithm [27].

2.5.1.2 Hierarchical Clustering

Another clustering algorithm that can be used for consideration is Hierarchical Clustering (shown in Algorithm 2). In this method, a bottom-up agglomerative approach is used, where each point starts out as its own cluster and these clusters continue to merge with each iteration until all clusters eventually come together to become one whole cluster. This merging is shown in a dendogram and represents an upturned tree, with its vertical branches representing the distance measured for each merging. This tree represents data partitions X_{NM}. During the process, the pairwise distance between each cluster is calculated and the pair with the smallest distance is fused together into one cluster H_l such that:

$$X, H = H_1, ..., H_Q(Q \leq N), \text{ s.t.}$$

$$1)\ H_q \neq \emptyset, q = 1, ..., Q$$

$$2)\ H_Q = X$$

$$3) \text{ if } C_l \in H_q \text{ and } C_m \in H_r, q > r \Rightarrow$$

$$C_l \subset C_m \text{ or } C_l \cap C_m = \emptyset \forall l, m, \text{ and } m, l = 1, ..., Q \text{ [28]}.$$

This distance is referred to as the clusters' dissimilarity measure [27]. There are several dissimilarity measures that can be applied to hierarchical clustering, shown in Table 3. Then, the clusters continue to be branched together until it eventually turns into one cluster or until a distance cut-off (height of the dendogram) is determined.

Algorithm 2 Hierarchical Clustering

1: Begin with n observations and a measure (such as Euclidean distance) of all the
2: For $i = n, n1, ..., 2$:

1. Examine all pairwise inter-cluster dissimilarities among the i clusters and identify the pair of clusters that are least dissimilar (that is, most similar). Fuse these two clusters. The dissimilarity between these two clusters indicates the height in the dendrogram at which the fusion should be placed.
2. Compute the new pairwise inter-cluster dissimilarities among the i 1 remaining clusters

Table 3: Linkage Methods [5]

Linkage	*Description*
Complete	Maximal intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the *largest* of these dissimilarities.
Single	Minimal intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the *smallest* of these dissimilarities. Single linkage can result in extended, trailing clusters in which single observations are fused one-at-a-time.
Average	Mean intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the *average* of these dissimilarities.
Centroid	Dissimilarity between the centroid for cluster A (a mean vector of length p) and the centroid for cluster B. Centroid linkage can result in undesirable *inversions*.

The benefit of this method is that a number of clusters do not need to be determined initially and is instead determined by the chosen dendogram height. This height is a measurement of the distance between clusters based on the linkage method chosen. The disadvantage of this method is that processing takes significantly longer for larger data to process compared to K-Means. This is because the hierarchical algorithm runs on On^3 time complexity, as opposed to K-Means which runs only on $On(logn)$. Another disadvantage is memory limit. Hierarchical must also run on On^2 complexity, which is requires more memory than a 64-bit desk computer can provide. Thus, the process more often fails due to memory before it can begin to address processing power.

2.5.1.3 Clustering based on Spatial and Temporal Constraints

Another approach to consider is clustering data based on specific distance and time constraints. The spatial and temporal grouping method was applied both by Hinkley and Sanderson. In the method used by Sanderson [3], clustering starts with the first flash based on date and time information, which forms the first cluster. Then, each subsequent flash is checked if it meets the determined time and distance criteria. In her study, the time requirement was for a flash to be within 30 min from the most recent flash in the cluster. For distance, the flash must have no more than a 16km separation from the edge of the cluster. If a flash meets both criteria, it is included in the cluster, otherwise it is set aside. The process iterates until the last flash is checked, and at this point the first flash grouping or cluster is established. The algorithm now checks if any flashes are unassigned and if so, the next cluster will be established using the same process. New clusters will continue to be formed until all flashes are assigned to a cluster. The advantage of this method is that it allows time and distance requirements for what constitutes a storm to be met, based on SME guidance. Additionally, it takes into account the movement of the storm, providing a more accurate clustering overall. With the limitations identified in k-means and hierarchical clustering methods, this method will be used in this study.

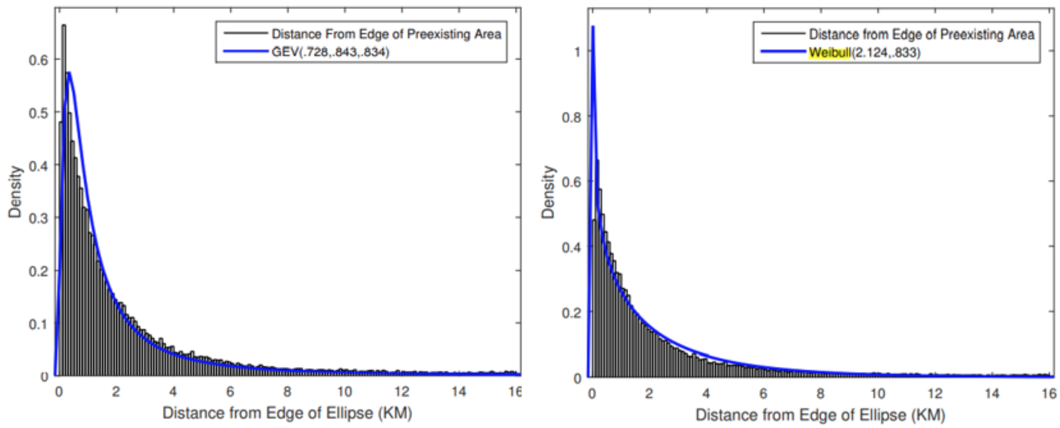
2.5.2 Ellipse Fitting

With the establishment of lightning event groupings, the ellipse fitting approach was explored by Sanderson [3] for use in gathering flash distances to fit to a distribution. In order to most accurately draw an ellipse based on the edge points of a storm, she applied the Minimum Volume Enclosing Ellipsoid (MVEE) method. The MVEE approach serves to identify lightning flash groups based on time steps and calculates distances between those flash groups with respect to a point in the edge of their boundaries. In this method, a tolerance level needed to be determined to find a balance between over-processing (affecting computational speed) and fitting most of the data points within the ellipse. She found that a tolerance of 0.01 was the appropriate level for this balance. The ellipse fitting algorithm was embedded within her spatial and temporal clustering algorithm to gather distances while she defined her storm groupings. To accurately fit an ellipse based on a moving storm, only events that occur within the past 10 minutes were applied to each ellipse fitting. Using the MATLAB program, she applied the Khachiyan algorithm. As MVEE seeks to find an ellipsoid with the smallest volume, optimization can be applied to minimize the volume of the ellipsoid. Thus, the approach to Khachiyan's algorithm is to apply the Lagrangian Dual to effectively compute the optimization by maximizing the objective function's value [3]. As the R program is used for this study, the MVEE approach will be applied using Titterington's algorithm [29]. Titterington's algorithm differs in that he uses D-optimal design based on the approximate theory. Yu [30] mentions in Titterington's design that the D-criterion determines iteration of the algorithm until it reaches convergence. It maximizes the determinant of an $m*m$ matrix with respect to $w = (w_1...w_n)^T \in \Omega$, with closure $\Omega = w : \sum_{i=1}^{n} w_i = 1, w_i > 0$. This algorithm is expressed using the ellipsoidhull() command in R, where the tolerance is set at a default of 0.01.

2.5.3 Distribution Fitting

Upon the collection of flash distances, fitting the results to a distribution is the next step in preparation for analysis. Sanderson [3] determined that the Weibull distribution was the best fit in comparison to another relevant distribution of interest, the Generalized Extreme Value (GEV). While both distributions fit the data fairly well, the Weibull distribution was able to incorporate more of the high number of short distances, as shown in Figure 7. In the AMU study, Hinkley applied the Exponential distribution and verified that with an R^2 value of 99.7%, the distribution model constitutes as a good fit as well. [9].

Figure 7: Sanderson's Comparison of Fit - GEV and Weibull [3]



2.5.4 Validation

During the validation portion of Sanderson's analysis [3], her main focus was to perform a comparative assessment between productivity and risk and determine which radii reduction would achieve an optimal balance with regard to both areas. To accomplish this, she investigated specific values that definitively represented productivity and risk increases for reduced radii alternatives between 3.25NM to 4.75NM. During this process Sanderson applied a conservative assumption, stating that events that arrive within 0.5NM from the radii center would be close enough to warrant a potential lightning related incident. This pertains to 5NM warning circles. For 6NM radii, 1.5NM was the applied assumption equivalent.

Taking these assumptions into account, Sanderson sought to measure two values of interest for productivity, the number of false warnings prevented and the resulting man-hours regained for operational utilization. To gather the false warning count, she defined these warnings as storms that entered the radii, but did not move within the assumed area for a potential surprise incident. Specifically, for the 5NM baseline, false warnings were defined as storms that passed within the radii, but did not arrive within 0.5NM of the radii center. Similarly, in the 6NM baseline, a storm is determined a false alarm if it passed within the radii, but not within 1.5NM of the radii center. To determine the potential number of man-hours saved, she identified and calculated the difference between the storm's time duration within the baseline radii and the time it spent within the reduced radii counterpart.

Sanderson measured risk as well, basing it on additional warning failures that resulted from a chosen radii reduction. In this context, a warning failure is defined as a lightning events of a storm that occured within 0.5NM of the warning radii's center before it was detected between 5NM and 0.5NM. Part of her assumptions involved treating each event as if it were cloud-to-ground strike interactions. Thus, all missed strikes were counted though it represented lightning aloft. She then represented risk as the failure rate generated based on how frequent missed warning occur.

In the AMU study, Hinkley's [9] assessment of risk involved determining whether the percentage of deaths met the allowable threshold of 10^{-5} as per the national permissible death standard. To accomplish this, Hinkley applied the cumulative distribution function for an exponential fit to determine what reduced radii option would result in 99.988% of distances falling within the distribution. This allowed Hinkley to determine what is the smallest radii option that would allow for the national standard to be met. For this study, the validation assessment portion will mirror that of both Sanderson's and Hinkley's work.

2.5.5 Summary

We have now gained comprehension on the different components that make up lightning and the process of a lightning strike to better utilize MERLIN's lightning data for the application of methods needed to validate radii reduction options suggested from previous research. A review of the evolution of 45 WS lightning location systems, the previous methods applied, and the lightning safety criteria allows us to understand platform capabilites and methods employed to ensure the safety of personnel and property at CCSFS/KSC. Lastly, conducting an examination of clustering, ellipse fitting, and statistical methods aids in the formulation of the methodology that will be applied in this research, discussed in detail in the next chapter.

III. Methodology

3.1 Overview

This chapter outlines the methodologies used in the analysis of lightning warning radii. It details the source, preparation, and programs applied to lightning data. Further, it identifies the construction of grouped recoil/dart leader events that constitute a storm and its movement mapped over time. Finally, it describes the processes implemented for analysis of lightning activity, to include the method for distance measure and analysis of the fitted distance distribution.

3.2 Data

3.2.1 Scope

The area of interest for this study is lightning that occurs in vicinity of Cape Canaveral, Florida between 2017-2019. The initial focus of lightning data spans during the summer months, between the months of May to September. 2017 was chosen by 45 WS as the starting year to ensure accurate date was provided for this study. For the years prior, MERLIN had an issue with radial smearing outside the sensor, with the issue resolved in 2016. The ending year of 2019 was chosen by 45 WS as it is the most current year that can provide a full year's worth of data [13].

The secondary focus will be data for the cold season months, October to April. Although about 80% of the lightning at CCSFS/KSC occurs in the summer season (May to September), the optimized lightning warning radii that was determined by Sanderson for those months will still need to be evaluated for the cold season portion of the year. This is important for CCSFS/KSC as thunderstorms in the summer can be different in the winter. The summer thunderstorms are caused by local surface-based boundaries that are weak and shallow, such as sea breeze fronts, river breeze fronts, and convective outflows. However, the winter thunderstorms are caused primarily by broad synoptic forcing and upper air support, such as cold fronts and variations in the jet stream [13]. Although no change in radii reduction is the expected outcome for the cold season, validation of this expectation must be verified. This data, also provided by the 45 WS, is produced by the MERLIN lightning location system as well. MERLIN collects information on recoil/dart leader events for both CG and CC occurrences.

Additionally, the data provided did not undergo the flashification process. Therefore, events were not grouped into flashes. With no flashes identified, the raw data that represents recoil/dart leader events are going to be processed. For the remainder of the paper, recoil/dart leader events will be used interchangeably. The data used for this analysis includes event location (expressed in latitude and longitude), the instance of occurrence (expressed in date and time) and the lightning type (G for cloud-to-ground or C for lightning aloft). Table 4 shows a sample of MERLIN data used, with the first four columns belonging to the raw data provided.

Table 4: MERLIN Lightning Data Loaded in R

Date	Time	Latitude	Longitude	Type	DateTime
<date>	<chr>	<dbl>	<dbl>	<chr>	<S3: POSIXct>
2019-07-01	00:00:05.086	28.4776	-81.0251	C	2019-07-01 00:00:05
2019-07-01	00:00:05.086	28.4786	-81.0381	C	2019-07-01 00:00:05
2019-07-01	00:00:05.086	28.4787	-81.0398	C	2019-07-01 00:00:05
2019-07-01	00:00:05.086	28.4798	-81.0645	C	2019-07-01 00:00:05
2019-07-01	00:00:05.086	28.4829	-81.0543	C	2019-07-01 00:00:05
2019-07-01	00:00:05.087	28.4758	-81.0805	C	2019-07-01 00:00:05
2019-07-01	00:00:05.087	28.4849	-81.0738	G	2019-07-01 00:00:05
2019-07-01	00:00:05.087	28.4924	-81.0613	C	2019-07-01 00:00:05
2019-07-01	00:00:05.087	28.4933	-81.0627	C	2019-07-01 00:00:05
2019-07-01	00:00:05.087	28.4939	-81.0560	C	2019-07-01 00:00:05

641-650 of 1,183,516 rows Previous 1 ... 63 64 65 66 67 ... 100 Next

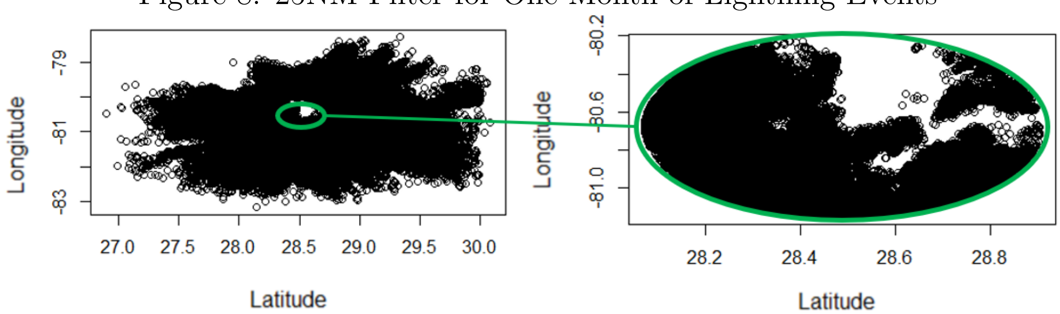
3.2.2 Programming Platforms Applied

The R Studio program was applied to generate the necessary algorithms to screen and cluster lightning data. Additionally, it was utilized to produce the lighting event's distance distribution via ellipse fitting as well generate the necessary measurements between storms and warning circles for validation. The JMP program was a secondary tool employed to gather initial insights on lightning behavior and was also used to assess the fit of the lightning event distance distribution.

3.2.3 Screening

As further distances from the MERLIN system affects the accuracy of lightning detection, the distance from a lightning flash to the center of the detection system needed to be considered. While the MERLIN is able to accurately detect lightning events up to 30NM, a 25NM constraint was considered to mirror Sanderson's scope[3], for an accurate comparison. Shown in Figure 8, only flash points within 25NM of the MERLIN system's central sensor location was assessed in this study.

Figure 8: 25NM Filter for One Month of Lightning Events



3.3 Identifying Events that belong to a Storm

The first portion of the algorithm focuses on identifying the group of flash points that make up each storm. Hereon, we will be using storm, cluster, and storm cluster interchangeably. To generate the algorithm, the steps to consider are identifying lightning event cutoff and clustering events based on location and time. Consultation with SMEs to establish parameters was imperative during this step.

3.3.1 Recoil/Dart Leader Event Cutoff

To ensure that the start and end of a storm is not affected by the end of day (i.e., 12:00am), event cutoffs were defined based on an appropriate time separation between recoil/dart leader groupings that would represent one storm. The 45 WS defined that grouped events are assigned to the next storm if there is 15 minutes of inactivity in between events, thus cutoff groupings will be based on 15 minute separation between recoil/dart leader discharges. A numerical identifier is placed in a new column in the existing data-frame to represent this grouping.

3.3.2 Clustering Based on Spatial and Temporal Constraints

Once the initial cut-off groupings are established, further sub-grouping is applied based on both the time and geographic separation of events. Following Sanderson's concept, the data is sorted chronologically and the first recoil/dart leader event is included in the first cluster. Each subsequent event is evaluated to determine if its minimum distance from the cluster is within 16km and its time difference is within 15 minutes from the most recent event in the cluster. If the subsequent point meets both these conditions, it is included in the cluster. Otherwise, no action is taken to the event and the event is set aside as it awaits inclusion in the next cluster. Once all the events have been evaluated in the first iteration, the first cluster is finalized. While there are unassigned events, another cluster will be formed and that cluster grouping will undergo the same process. This algorithm iterates and new clusters continue being formed until all events are assigned to a cluster. Once all cluster groups are established, a numerical identifier is then placed in a new column in the existing dataframe to represent the storm cluster groupings. The pseudo-code labeled Algorithm 3 outlines the steps that are applied to spatial and temporal clustering.

Algorithm 3 Spatial and Temporal Clustering

1: Begin with n observations in chronological order. The first observation $i = 1$ is assigned to cluster $j = 1$

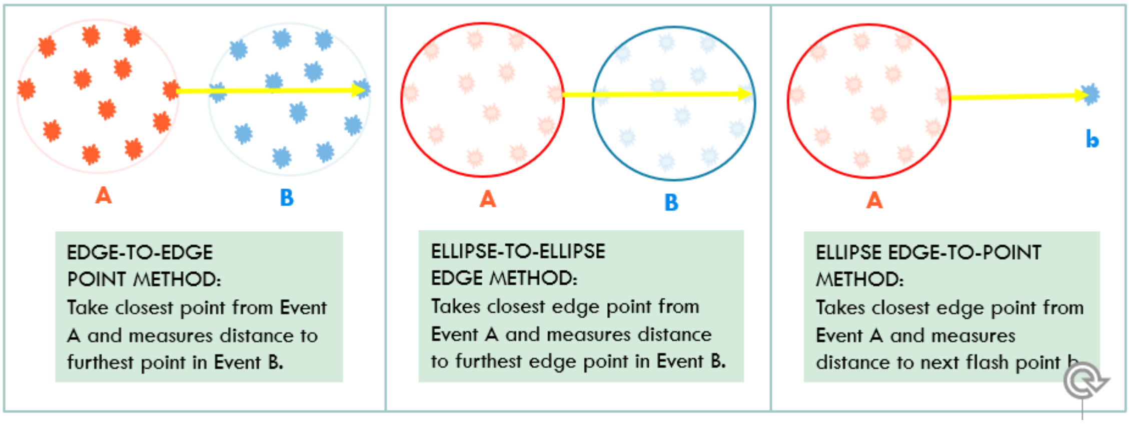
2: **While** any i is unassigned, declare the next unassigned value m

1. **For** $i = m, ..., n$:
   (a) Subset cluster j to only events that occur in an established time increment (i.e. last 10 minutes)
   (b) If i is unassigned to a cluster, check its time and minimum distance criteria from cluster j. If i is within the time and minimum distance criteria, add to cluster j, otherwise, **next** i.
2. Increment $j = j + 1$ and appoint the first unassigned observation m to next cluster j
3. End **While**

3.4 Distance Measurement Criteria for a Storm Movement

Once complete lightning event groupings have been identified, calculating distances with respect to storm movement through time is the next step. Here, clear guidance on what needs to be measured must be defined. A recoil/dart leader event in this instance will also be represented as a point. Three approaches were considered during this step, an edge to edge point method, edge to edge ellipse method, and ellipse edge to point method. Figure 9 provides a snapshot of each concept, visually depicting the criteria that needs to be measured.

Figure 9: Distance Measurement Approaches



3.4.1 Edge to Edge Point Method

The edge to edge point method, discussed with SME guidance, was first explored. This approach suggested placing events on a grid system and applying rectilinear distance as opposed to euclidean as a measurement, to potentially decrease processing time. In this method, a lightning event grouping was separated into 10 minute increments. The intent of the groupings is to measure distance between the closest point from one group to the furthest point of the next group. During this process, it was discovered that this method could only work if the sub-groupings did not overlap, or stay in place while expanding. Unfortunately, the lightning behavior allows for such movements to occur. Thus, it would be a challenge to capture lightning behavior if this method was applied.

3.4.2 Edge to Edge Ellipse Method

The next method explored, an adaptation from the first, is the edge to edge ellipse method. In this method, the same criteria for grouping and distance measurement were considered, then ellipses are fitted to each grouping so that only points on the ellipse edge would be considered for measurement. While there were concerns fairly similar to the first method, the biggest concern for this method is that, the determined time increment could affect the distance measurement. For example, if the lightning event is not overlapping and continues to move, there would be a significant difference between the distance distribution between the sub-groupings if the time increment was changed from 10 minutes to 15 minutes. A secondary concern is that the duration of the event determine the number of time increment groups. This may measure groupings too small of a sample size to truly represent the distribution of a lightning event.

3.4.3 Ellipse to Point Method

Lastly, the ellipse to point method, also applied by Sanderson, was considered. This method involved grouping a set of events in a cluster and fitting an ellipse to the group. Using the ellipsoidhull command in R, a minimum of 3 points is required to form an ellipse. Thus, only clusters with more than 3 events were considered. Then, the minimum distance is measured from the ellipse to the flash point identified. The ellipse is then refitted to include that flash point. The process then iterates for each remaining flash point in the group. During the iteration, only the flash points that occur within 10 minutes of the most current flash point are refitted to an ellipse, to account for the movement of the lightning event. This method was considered the most accurate of the three approaches and therefore applied in this study. The pseudo-code created in Algorithm 4 outlines the steps that are applied to the Ellipse to Point approach using the MVEE method.

Algorithm 4 Ellipse to Point Method using MVEE

For Storm $i = 1, ..., n$:

1. Begin with r events in chronological order. **For** events $l = 4, ..., r$:
   (a) Let event group $j = 1, ..., l-1$
   (b) Subset j to only events that occur in an establisehd time increment (i.e. last 10 minutes)
   (c) Fit ellipse k to event group j
   (d) **If** l is outside ellipse k:
      i. Measure distance between l and k
      ii. **End If**
   (e) **Next** l
2. **Next** i

3.5 Lightning Warning Distance Distribution

Upon collection of distance measurements from identified recoil/dart leader events that occur outside the predetermined boundary, the next step is to fit the distance results to a distribution. With the appropriate distribution identified, cumulative distribution values could be identified to assess how far event distances travel past the edge of the predetermined boundary. Similarly, events that extend past the boundary were analyzed to assess the behavior of the storm and gain insights on comparisons between seasonal events as well as the differences between CG and CC types.

3.6 Validation of Reduced Radii Alternatives

The last portion of the analysis involved observing reduced radii options and determining if these options can be vetted as a safe lightning warning distance alternative, that would result in increased productivity while maintaining an acceptable level of risk. Here, the main objective was to determine if the reduced radii option proposed by both Sanderson and Hinkley was applicable based on the MERLIN system. To validate that a 1NM reduction does achieve optimal balance between productivity and risk, three alternatives were assessed. These alternatives include reducing the radii by 0.5NM, 1NM, and 1.5NM. With these alternatives determined, we applied Sanderson's validation approach by measuring productivity based on prevented false alarms and operational hours given back to the space launch mission. Specifically, measurement of productivity is based on the percentage increase of operational hours from the baseline. In this context, there are two warning circles of interest, 5NM and 6NM, represented as the baseline. Similarly, risk is assessed based on the percentage of risk increase from the baseline. This will be formulated based on the number of additional missed warnings that occur due to the proposed radii reduction. Then, the failure rate will be measured based on the percentage of storms that strike within 0.5NM of the warning radii (for 5NM circles) before a warning outside that strike area is determine. Similarly, the failure rate is applied when within 1.5NM (for 6NM circles). An overall assessment was provided for comparison between previous research recommendations. Additionally, a breakdown analysis for each type of warning circle, 5NM and 6NM, was conducted to provide further insights and recommendations.

3.6.1 Summary

We have now discussed the specific methods we will apply in this research , to include data preparation, storm identification via clustering, distance measurement using MVEE, statistical evaluation of the distance distribution, and validation of reduced lightning warning radii options. We will see in the next chapter the results and insights that were derived from the application of these methods, the conclusions gathered that will answer the main and secondary objectives from our research question, and how the consequent findings compare with the previous studies of Sanderson and Hinkley.

IV. Results and Analysis

4.1 Overview

This chapter provides all relevant results and insights gathered pertaining to grouping the data into storm clusters as well as gathering distances from recoil/-dart leader events that are outside the preexisting lightning boundary with the use of the ellipse fitting method. With the evaluation of the corresponding distance distribution, we also apply the results gathered from the identified lightning distance warning alternatives to validate the feasibility of potential new lightning warning radii and present findings and insights based on productivity and safety implementation.

4.2 Results for Identification of Storm Clusters

After running 2017-2019 data through spatial and temporal clustering, a total of 2,658 storms was identified for observation. Table 5 shows a monthly breakdown of the number of these storms. To note, for each year, the data in the table involves only the months provided by 45WS. Thus, 2019 includes data for all months, while 2017 and 2018 includes only consists of 5 and 11 months, respectively. The month of July in the year 2018 is observed to have the most storms, at a count of 323. Additionally, the amount of data the number of storm clusters with 3 events or less as well as events with more than 3 for total lightning are identified. Only storms in the latter case will run through ellipse fitting. Overall, 22% of the clusters had 3 or less events and was not processed in ellipse fitting. More than half of those storms consisted only of one event (about 12% of total clusters) and these isolated events were observed to consist of only CC lightning.

Results for average and max counts were gathered as well for examination, focusing on TL events as well as isolating CG events. June 2018 is shown to have the storm with the largest number of events for total lightning, while August 2017 includes the largest number of events for CG. While most activity is expected to occur during these summer months, we also see a few heavy storms occur at random cold months. A notable cold season month is Mar 2018, experiencing about 114,000 events on average during this time period.

Table 5: Statistical Results for Storm Clusters

Statistical Results for Storm Clusters								
	Month	Total # of Storms	Storms w/ <=3 Events	Storms w/ >3 Events	Average # of Events in a Storm (TL)	Average # of Events in a Storm (CG)	Largest # of Events in a Storm (TL)	Largest # of Events in a Storm (CG)
2017	Feb	35	8	27	3,369	4	74,765	53
	Mar	23	9	14	53,108	53	1,106,633	1,164
	Apr	3	1	2	904	0	2,058	0
	Jul	114	29	85	62,022	50	1,123,744	2,499
	Aug	212	52	160	32,743	49	1,039,129	1,521
	Sep	174	37	137	79,281	39	2,452,795	1,054
	Oct	54	15	39	1,508	9	75,867	349
2018	Feb	2	2	0	2	0	2	0
	Mar	33	4	29	114,115	8	2,293,785	176
	Apr	119	37	82	21,754	7	1,384,124	325
	May	121	34	87	38,721	16	1,325,056	668
	Jun	192	71	121	126,652	14	2,749,369	1,748
	Jul	323	94	229	26,139	18	1,242,787	820
	Aug	141	25	116	53,711	35	1,667,995	798
	Sep	195	34	161	5,095	18	218,034	1,485
	Oct	16	4	12	1,041	2	10,065	12
	Nov	43	9	34	4,930	9	123,578	129
	Dec	42	7	35	27,875	24	484,913	611
2019	Jan	14	6	8	19,600	6	273,556	83
	Feb	20	5	15	56,862	9	866,218	139
	Mar	24	6	18	66,418	2	1,403,800	29
	Apr	20	3	17	64,042	11	1,001,690	110
	May	59	7	52	159,983	12	2,299,815	418
	Jun	128	21	107	86,154	40	1,181,695	1,239
	Jul	204	40	164	23,667	39	744,551	1,128
	Aug	183	20	163	36,253	24	1,457,980	1,151
	Sep	28	6	22	590	2	8,967	35
	Oct	88	15	73	2,689	19	109,071	804
	Nov	34	9	25	21,138	4	642,556	28
	Dec	14	2	12	27,875	36	484,913	552
TOTAL		**2,658**	**612**	**2,046**	**40,608**	**19**	**2,749,369**	**2,499**

4.3 Results for Events Outside a Predetermined Boundary

With storm clusters identified, we observe results for distance measurements using the ellipse fitting algorithm. Table 6 shows a monthly breakdown of total number of events in comparison to the events that fall outside the predetermined ellipse boundary. Of the storms that had four or more events, 95.5% were applied for ellipse fitting, as the size of extremely large storms affected process time. Based on the time allotted for this portion of the analysis, only this percentage fit the available time. Nevertheless, the bulk of the storms that was processed for ellipse fitting was a sufficient amount to assess a distribution, given the data is significantly large and storms processed are randomized to ensure there is representation for each month. While a small portion was not included in compiling the distance distribution, all storms were applied to validation in the latter portion of the analysis.

Moreover, the distribution needed to represent single storms. With the application of spatial and temporal constraints, a few groupings were observed to have at least two storms instead of one. These storms showed either a parallel structure, or there were evidence of merging or splitting storms. As these groupings did not truly represent single storms and would result in heavily skewing the data, they were set aside for this portion of the analysis. These type of storms represented 4.5% of the total storm count.

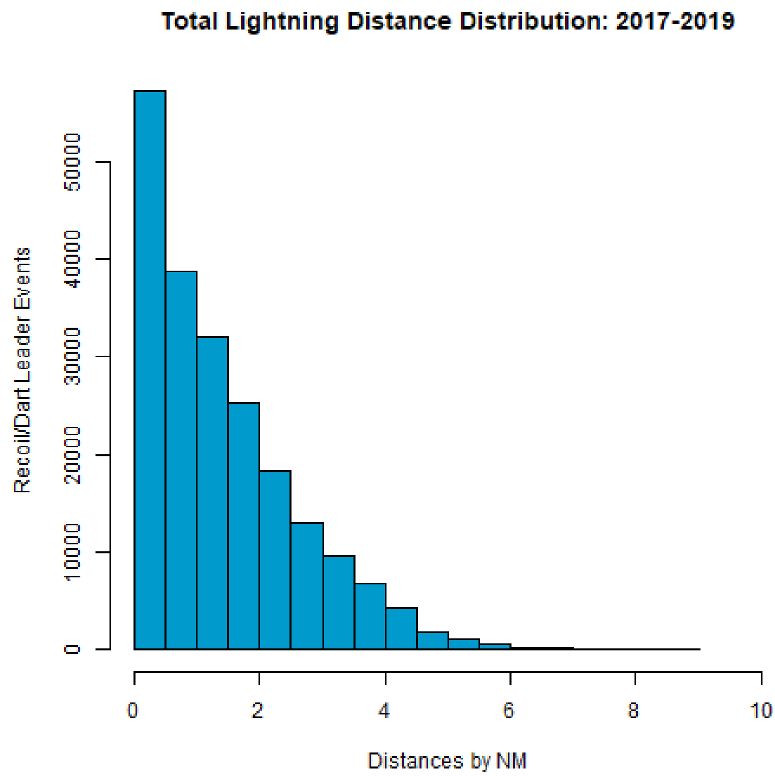
Along with TL events, we also observe statistical results for CG lightning in the ellipse fitting results. Overall, there are 556,901 events that fall outside the boundary of a preexisting lightning area, about 11.6% in total. CG events occur less frequently in general, with only 0.3% representing the total event sample. Thus, the proportion of CG events that fall outside the boundary is 8.4% when compared to the CG population, but only 2.5% of CG events fall beyond the boundary when compared to Total Lightning.

Table 6: Statistical Results for Ellipse Fitting

Year	Month	Total # of TL Events	# of Events Outside Ellipse (TL)	% of Events Outside Ellipse (TL)	Total # of CG Events	# of Events Outside Ellipse (CG)	% of Events Outside Ellipse (CG)
2017	Feb	43,137	3,940	9.1%	55	1	1.8%
	Mar	66,677	2,414	3.6%	47	2	4.3%
	Apr	2,711	283	10.4%	0	0	0.0%
	Jul	123,534	13,032	10.5%	452	134	29.6%
	Aug	1,031,739	30,910	3.0%	5,082	167	3.3%
	Sep	963,108	84,263	8.7%	4,496	448	10.0%
	Oct	34,802	4,105	11.8%	84	7	8.3%
2018	Feb	0	0	0.0%	0	0	0.0%
	Mar	137,900	10,338	7.5%	9	7	77.8%
	Apr	143,270	38,865	27.1%	234	54	23.1%
	May	291,901	13,892	4.8%	853	26	3.0%
	Jun	197,318	5,182	2.6%	424	11	2.6%
	Jul	1,115,503	56,468	5.1%	3,370	190	5.6%
	Aug	610,298	40,210	6.6%	2,376	211	8.9%
	Sep	399,215	35,109	8.8%	1,312	194	14.8%
	Oct	2,949	1,667	56.5%	13	13	100.0%
	Nov	46,706	8,357	17.9%	71	6	8.5%
	Dec	245,897	7,991	3.2%	101	14	13.9%
2019	Jan	828	266	32.1%	0	0	0.0%
	Feb	32,304	2,442	7.6%	16	2	12.5%
	Mar	13,345	2,147	16.1%	4	3	75.0%
	Apr	43,271	2,203	5.1%	54	1	1.9%
	May	53,053	15	0.0%	171	19	11.1%
	Jun	841,830	27,236	3.2%	2,128	44	2.1%
	Jul	1,042,869	57,030	5.5%	3,615	246	6.8%
	Aug	770,202	79,375	10.3%	3,012	338	11.2%
	Sep	3,385	1,042	30.8%	9	2	22.2%
	Oct	224,324	14,316	6.4%	1,603	77	4.8%
	Nov	7,921	2,486	31.4%	67	13	19.4%
	Dec	521,344	11,317	2.2%	606	307	50.7%
TOTAL		**9,011,341**	**556,901**	**11.6%**	**30,264**	**2,537**	**8.4%**

Placing the distance results into a histogram as depicted in Figure 10 shows that the data is positively skewed. Most events travel a few NM past the boundary edge, with almost half (49.4%) only within a 1NM distance from the edge. A small amount (1.9%) of the distribution observed shows that events travel 6NM past the boundary and it increases to 3.6% when traveling past 5NM. It further doubles to 7% when moving past 4NM.

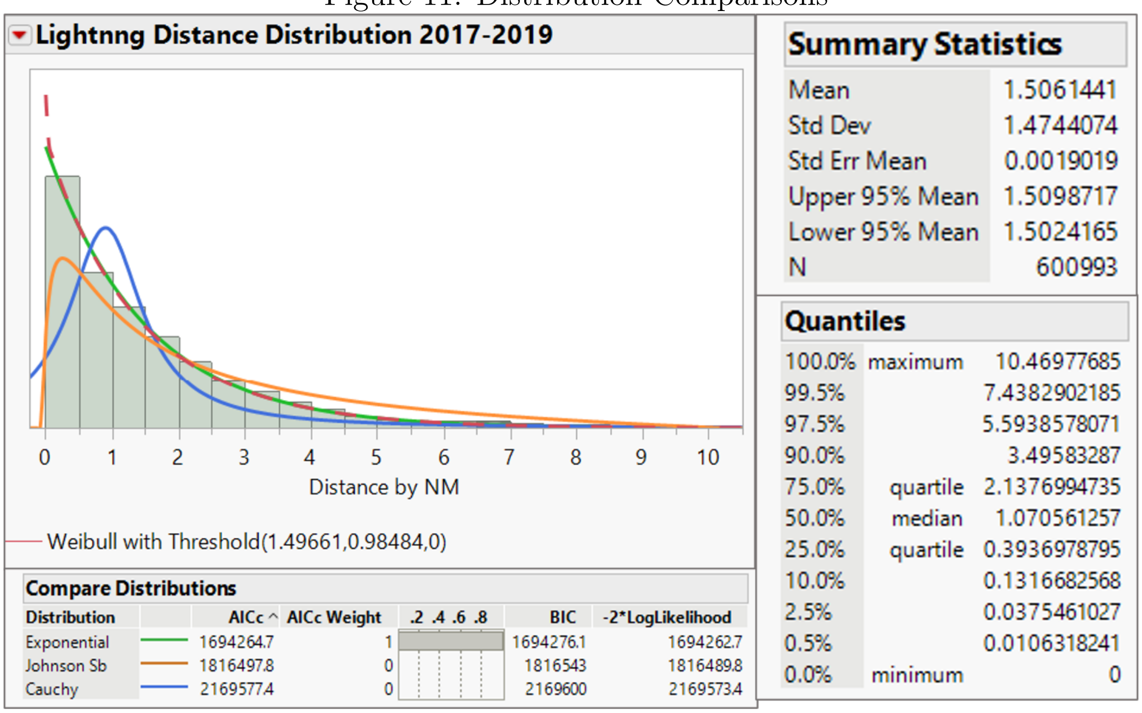
Figure 10: Total Lightning Distribution 2017-2019



The shape and skewness of the data suggests a Weibull disribution would be an appropriate fit, as suggested in Sanderson's research. True to its versatility for many different applications, the Weibull was expected be compatible with data that presented an extreme value distribution (EVD) pattern. A comparison of distributions (reference Figure 11) confirms that Weibull is a very good fit, achieving the lowest Akaike Information Criterion (AIC) at 1693984.5, amongst the three other distributions explored; the exponential, Johnson, and Cauchy distribution. The Weibull is shown to visibly provide the best fit to the data, as represented in dotted red on the graph in Figure 11. Thus, the lightning distance distribution can be represented as a Weibull with scale and shape values of 1.50 and 0.98, respectively. In Sanderson's research[3], she achieved a scale of 2.1 and shape of 0.8. These results are fairly similar and the slight deviation may be due to several differences in the data, such as her application of flashified data composed from step leader events as opposed the application of raw data composed of recoil/dart leader events in this study, as well as the coding structure and parameters applied to the algorithm. Cold season and CG lightning included only in this study also contributed to the difference in values. For comparison, Sanderson's lightning distance distribution fit to a Weibull can be seen in Chapter 2, Figure 7.

Another very close and feasible distribution that can be applied is the Exponential, as suggested by Hinkley. This distribution is represented as the green line on the graph, nearly overlapping the Weibull curve and coming to a close second in AIC at 1694264.7. A definitive reason why this distribution exhibits a fairly close fit is because the Exponential is a special case of the Weibull distribution, specifically when the shape value is equal to 1 [31].

Figure 11: Distribution Comparisons



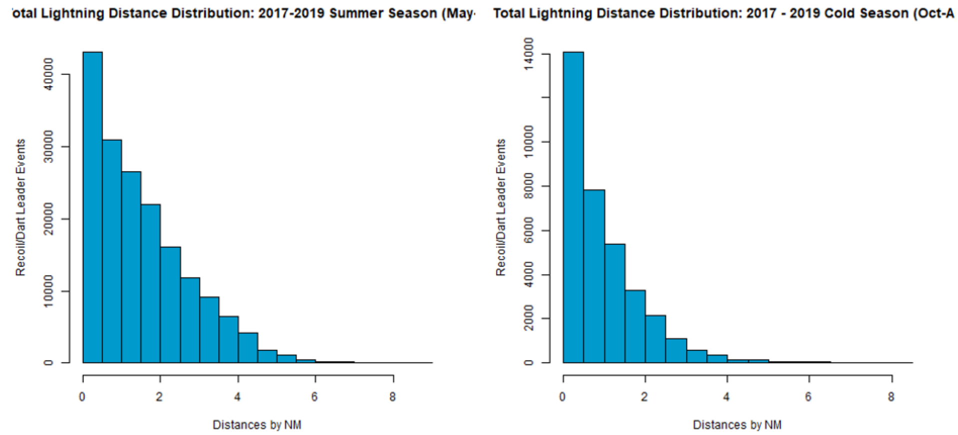
In comparison to the assessment considered in the AMU study, we also determined the Weibull cumulative distribution values that would include 99.981% of lightning. Shown in Figure 12), based on the given shape and scale parameters, a 4.1NM reduction achieves the lightning percentage values defined in the AMU study that would fall within the desired safety threshold. To note, the 4NM radii reduction option achieves a distribution value of 99.974%, fairly close to the desired threshold as well.

Figure 12: Radii Reduction Comparison - Desired Safety Threshold

Radius (X)	Weibull CDF
4NM	99.974%
4.1NM	99.981%
4.5NM	99.995%

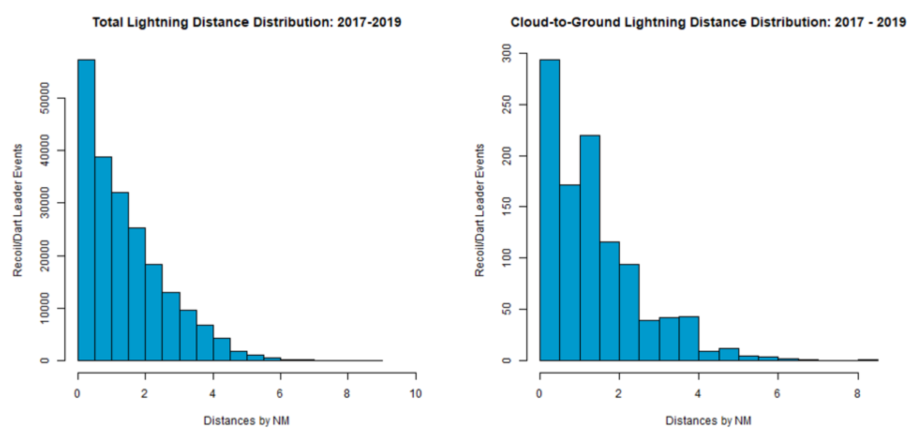
Distribution results were also separated into on-season and off-season categories, to gain insights as a secondary objective. In Figure 13, we see that in the cold seasons, there are shorter distances traveled by lightning events. Only 0.3% of cold season events travel past the preexisting storm area boundary and about 98.6% travel within 5NM of the edge in comparison to the summer seasons, where 0.6% extend past the edge and only 95.9% fall within 5NM. Based on subject matter expert insight, this behaviour is due to winter thunderstorms being driven by approaching weather fronts which organizes mature thunderstorms with well-developed lightning areas [13], The development causes events to be more compactly saturated than widely disbursed, affecting the distance that lightning travels.

Figure 13: Distribution Comparison between Summer and Cold Seasons



The second objective also involves assessing comparisons between TL and CG lightning. To note, both charts in Figure 14 are not scaled similarly due to CG encompassing a very small amount of TL events. CG exhibits a rough, but similar pattern of distribution to TL, with only 3.3% extending 5NM past the ellipse edge and 1.4% extending 6NM past in cumulative distribution. We also observe that CC lightning only extends to 9NM, while CG lightning has a shorter reach, extending only to about 8NM in totality.

Figure 14: Distribution Comparison: TL vs CG Only



4.4 Productivity and Risk Validation

The storms were assessed based on their interaction with the 12 warning circles established by KSC/CCSFS to determine what alternate distance warning radii options can be considered with respect to safety requirements. In the initial results gathered, we see that 62% of storms occurring in the Cape Canaveral area pass through at least one or more of the warning circles. A breakdown of the number of storms can be seen in Table 7, where storm activity is mostly evenly disbursed amongst the 12 radii. During the 2017-19 time frame, storm activity generated a total of 2,564 warnings. The 37/ITL warning radii had the least activity at 181 storms, while the warning radii at Patrick Space Force Base (PSFB) had the maximum number of storm activity, at a count of 244.

Table 7: Overview of Warning Radii Storm Counts

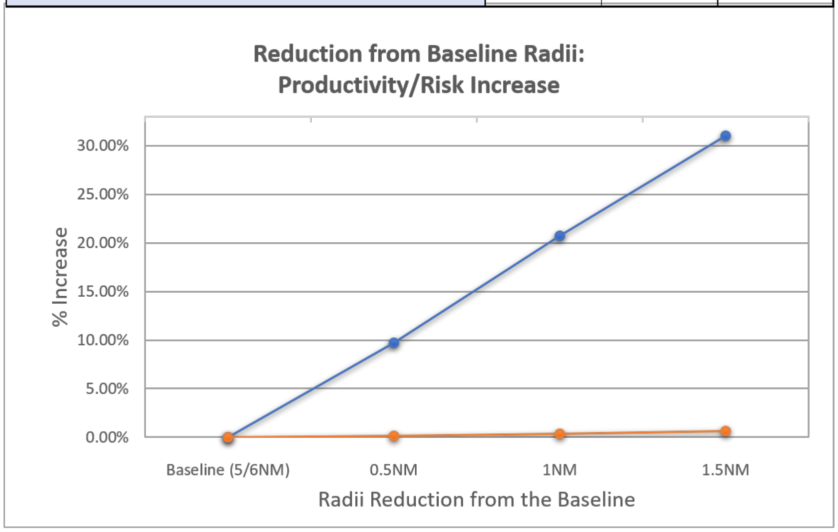
2017-2019 Storm Counts	
Total Storms	2658
Storms Passing Inside a Warning Radii	1646
Total Warnings	2564
5NM Warning Circle Locations	# of Storms in 5NM Radii
HAULOVER	231
40/41	181
37/ITL	166
36/46	194
PORT	198
PSFB	244
ASTROTECH	205
CIDCO PARK	223
6NM Warning Circle Locations	# of Storms in 6NM Radii
LC39	228
SLF	239
KSC INDUSTRIAL AREA	227
CAPE CENTRAL	228

Based on the two areas of interest, productivity and safety, the main focus is to see if the general guidance of 5NM can be reduced to 4NM. To generate comparative results that center around the 4NM option, the three alternatives considered were reducing the baseline radii by 0.5NM, 1NM, and 1.5NM. Thus, a 5NM baseline would be reduced to 4.5NM, 4NM, and 3.5NM, respectively. The same reduction for 6NM radii would result in the observation of 5.5NM, 5NM, and 4.5NM. From previous research, the recommendation by both Sanderson and Hinkley was a reduction to 4NM for the optimal safety and productivity balance. The objective in this research is to determine if similar results are achieved with the most recent system, MERLIN. In the overall findings shown in Figure 15, the risk increase overall is miniscule when reducing the baseline radii to any of the alternatives, in comparison to productivity increase. The smallest risk increase incurred (0.11%) is from reducing the baseline by 0.5NM. For each consecutive reduction alternative, the risk increases by 0.21% and 0.32%, respectively.

In contrast, productivity increase is very significant when compared to risk. Even with the first 0.5NM reduction, productivity increase is 88 times greater than the risk incurred. Depicted in the graph, the steepest incline in productivity increase is shown at the 1NM reduction, while maximum productivity is realized at 1.5NM reduction. While reducing by 1.5 NM achieves the most productivity, it also incurs the maximum risk. On the other hand, while 0.5NM keeps the risk increase to a minimum, it also does not generate much productivity. Overall, a balance in productivity and risk is shown to be achieved at the 1NM reduction, as it exhibits the sharpest incline in productivity, but does not achieve the highest amount of risk. This alternative reduces the 5NM baseline to 4NM and the 6NM baseline to 5NM.

Figure 15: Productivity and Risk Comparison

Overall Warning Circle Evaluation (2017-19)			
Warning Radii Reduced By	0.5NM	1NM	1.5NM
Productivity % Increase from Baseline	9.70%	20.74%	31.00%
Risk % Increase from Baseline	0.11%	0.32%	0.64%



Radii reduction results for both 5NM and 6NM warning circles were also separately evaluated and observed in detail for further insight. For the 5NM warning radii, the steepest incline in productivity increase (12.2%) is again shown at the 4NM option, allowing the prevention of 205 false warning and saving 234.6 operational hours. On risk, there is overall a very small amount of missed warning resulting from radii reduction resulting in only about a 1% risk of failure. 4NM also exhibited no incline in risk in comparison to the 4.5NM option, but at 3.5NM, risk doubles.

Table 8: 5NM Warning Radii Reduction Results

5NM Warning Circle Evaluation (2017-19)			
Radii Reduction Options	4.5NM	4NM	3.5NM
PRODUCTIVITY			
# of False Warnings Prevented	130	205	301
# of Time Saved (in hrs)	105.2	234.6	349
Productivity % Increase from Baseline	9.9%	22.1%	32.8%
SAFETY			
# of Missed Warnings	5	5	6
% Risk of Failure	0.9%	0.9%	1.1%
% Missed Warning Resulting in CG Strike	0.0%	0.0%	0.0%
Risk % Increase from Baseline	0.2%	0.2%	0.4%

Looking at 6NM radii reduction results in Table 9, we see that each consecutive reduced radii option saves an additional 50 operational hours, resulting in a linear increase in productivity through each reduction option. In terms of risk, the 6NM baseline results in 20 missed warnings and a failure rate of 5.2%. There is no increase in risk when reducing the baseline to 5.5NM, but the failure rate and resulting risk increase from the baseline also rises linearly. While a reduction to 5NM radii achieves a balance between productivity and safety, the other two alternatives could also be considered, depending on whether the least amount of risk is most important or maximum productivity is paramount, due to linearity in the results.

Table 9: 6NM Warning Radii Details

6NM Warning Circle Evaluation (2017-19)			
Radii Reduction Options	5.5NM	5NM	4.5NM
PRODUCTIVITY			
# of False Warnings Prevented	75	154	213
# of Time Saved (in hrs)	54.2	105.4	159.2
Productivity % Increase from Baseline	9.4%	18.3%	27.6%
SAFETY			
# of Missed Warnings	20	22	24
% Risk of Failure	5.2%	5.8%	6.3%
% Missed Warning Resulting in CG Strike	0.0%	0.0%	0.0%
Risk % Increase from Baseline	0.0%	0.5%	1.1%

In an overall comparison of the two evaluations, we see that the failure rate for missed warnings is significantly higher for 6NM Warning radii than it is for its 5NM counterpart. At its baseline, 5NM radii only incurs a failure rate of 0.7% with 4 missed warnings, while 6NM radii incur 5.2% with 20 missed warnings. We also see that none of the missed warnings resulted in a CG strike for both 5NM and 6NM outcomes, showing that while it is not impossible for a CG strike to occur, it will indeed be very rare.

4.4.1 Summary

We have now presented the results derived from this study's application of clustering, ellipse fitting, statistical analysis, and validation. We determined that the Weibull is the best fit to present the distance distribution for lightning outside a pre-existing area. We were able to confirm that aside from lightning aloft in the summer season, a 4NM radii reduction was feasible for all conditions, including cold season and CG type lightning, as expected. We also determined that results in this research is comparable to the previous studies conducted by Sanderson and Hinkley. In the next chapter, we discuss the culmination of these insights, provide recommendations, and suggest several avenues for future work.

V. Conclusions

Lightning stands to be one of the most influential factors that dictate the safety and productivity of space launches at CCSFS/KSC. The currently established lightning warning radii criteria of 5NM has more than ensured safety of personnel, but has also resulted in a significant number in false alarms that has caused degradation in space launch productivity and mission success. This thesis takes part in ongoing research that examines radii reduction alternatives to safely reduce the impact of lightning warnings to increase mission productivity, with its focus specific to the latest lightning warning detection system, MERLIN. This study sought to compare and determine if results from the new lightning detection system are concurrent with previous research findings, providing detailed insights in reducing radii by 0.5NM, 1NM, and 1.5NM alternatives. As a secondary objective this research also examined seasonal and CG comparisons.

To meet these objectives, clustering with spatial and temporal constraints was first utilized to define groupings that represent storms. In order to determine how far a lightning event travels from a preexisting storm boundary, the MVEE fitting method was applied to each storm to get the overall lightning distance distribution. This allowed further examination of distance distribution comparisons based on CG and seasonal subsets. A validation assessment was conducted in the final stages of the analysis to determine balance between productivity and risk with respect to the proposed radii reduction alternatives. Here, storms passing through any of the established warning circles at KSC/CCSFS are assessed based on false alarms prevented, operational hours saved, number of missed warnings based on radii redcution, failure rate, and likelihood of a CG strike in vicinity of facilities in the center of the established radii.

5.1 Research Insights

With defined storms established, 22% of storms consisted of less than 3 events, with half of the count made up of only single events. No CG lightning was found in these single events. While the summer season marked heavy storm activity each month, the cold season also presented a few instances where certain months encountered heavy activity comparable to the summer season. Thus, we can expect lightning behavior in the cold season to potentially display the same amount of activity and prepare for the cold season activity in the same manner as summer.

70% of the storms identified were processed for ellipse fitting. Of the 30% that was set aside, 4.5% were found to consist of at least two storms that moved in a parallel, merged, or split structure. This is more likely a result of the clustering model in place, and as these storms did not represent a single storm, they were set aside and more insight can be gained in this area with future research.

Majority of the events observed are from CC lightning, with only 0.3% representing CG lightning. During ellipse fitting, 11.6% of TL events overall and 8.6% of CG events travelled outside the predetermined storm boundary. Half of the events that travelled only moved 1NM past the boundary.

As part of the main objective, lightning distances based on events detected by the MERLIN displayed a similar distribution as the previous study by Sanderson, which was appropriately fit to a Weibull distribution. Observing the Weibull fit, the percentage of events doubles for each NM when comparing the reach between 6NM to 4NM for which lightning travels past the boundary edge. At a reach past 6NM, 1.9% of events are observed. A reach of 5NM results in 3.6% of events, and at 4NM, 7% of events travel 4NM outside the predetermined boundary. These results are also comparable to Sanderson's work, where CDF values include 1.9% for 6NM, 3.3% for 5NM, and 5.9% for 4NM.

Focusing on the secondary objective, in comparison to the summer season (May-Sep), the cold season (Oct-Apr) displayed storms with heavier saturation situated within the predetermined boundary and events that travelled shorter distances. The cold season obtained double the percentage of events occurring outside the boundary (0.6%) and 98.6% remained within a 5NM reach past the boundary edge.

Also included in the second objective, CG was separately assessed from the TL structure. Comparing cloud-to-ground lightning to the overall distribution, CG displays a smaller overall reach in comparison to CC. While CC event distances extend to 9NM, CG events only travel to about 8NM.

A comparison of results between 5NM and 6NM warning circles shows that the failure rate for missed warning is generally five times higher with 6NM radii than it is with 5NM. With respect to ground operations, none of the missed warnings resulted in CG strike for either radii type. While it may not be impossible for a CG strike to occur, it appears very rare.

5.2 Recommendations for Action

The prevalent part of the main objective involves validating whether reducing the current warning radii (baseline of 5NM) to 4NM, based on previous research, remains relevant and applicable to the new MERLIN detection system. We see evidence that this recommendation can be maintained. Looking at a productivity versus risk comparison, the overall risk increase is insignificant compared to productivity. At the smallest radii reduction, 0.5NM, productivity increase is nearly 90 times greater than risk incurred. A 1NM reduction shows the steepest incline in productivity, while maintaining fairly minimal risk. With the smallest option, a 0.5NM reduction does achieve the lowest incline in risk, but also the lowest incline in productivity. On the other end, a 1.5NM reduction provides the maximum amount of productivity, but also the highest amount of risk. While all options appear feasible, a 1NM radii reduction achieves the balance between productivity and risk. This suggests a reduction of 5NM warning radii to 4NM and similarly, a reduction of 6NM radii to 5NM.

5.3 Recommendations for Future Work

This study identified three potential areas of research. The first area involves taking a closer look at storms that exhibited a parallel, merged, or split structure. This study was able to identify these types by the evidence of their distance distribution, but more can be gained by taking a closer look at these storms constructs. Shown in both Figure 16 and Figure 17, these storms do not have an extreme value distribution that is positively skewed. Rather, they will lean closer to a normal, uniform, or even a negatively skewed construct. The corresponding plotted graph shows that while they meet the spatial and time constraint of 16km and 15minutes, they still exhibit separation between event groups, causing the distributions to take on a different shape. To provide a cleaner separation for these storm, a hierarchical clustering with complete linkage method can be applied. Examining this behavior could draw further insights on where lightning can be expected to occur and whether lightning warnings need to take into account the different behavior in these occurrences.

Figure 16: Storms - Parallel Construct

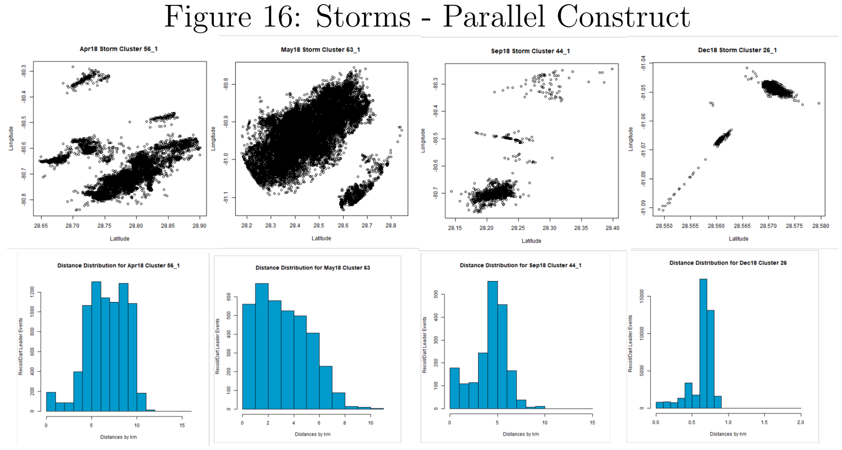
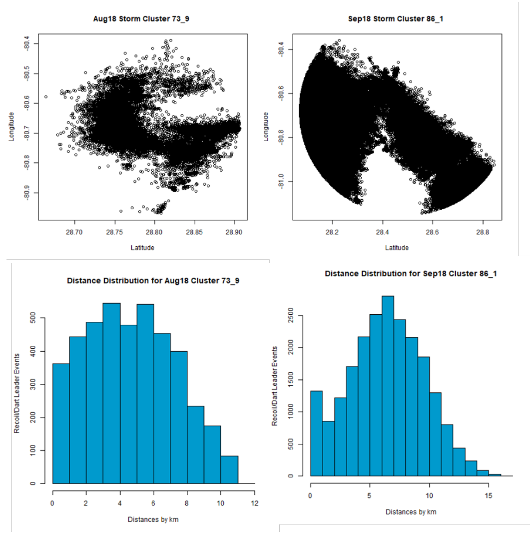
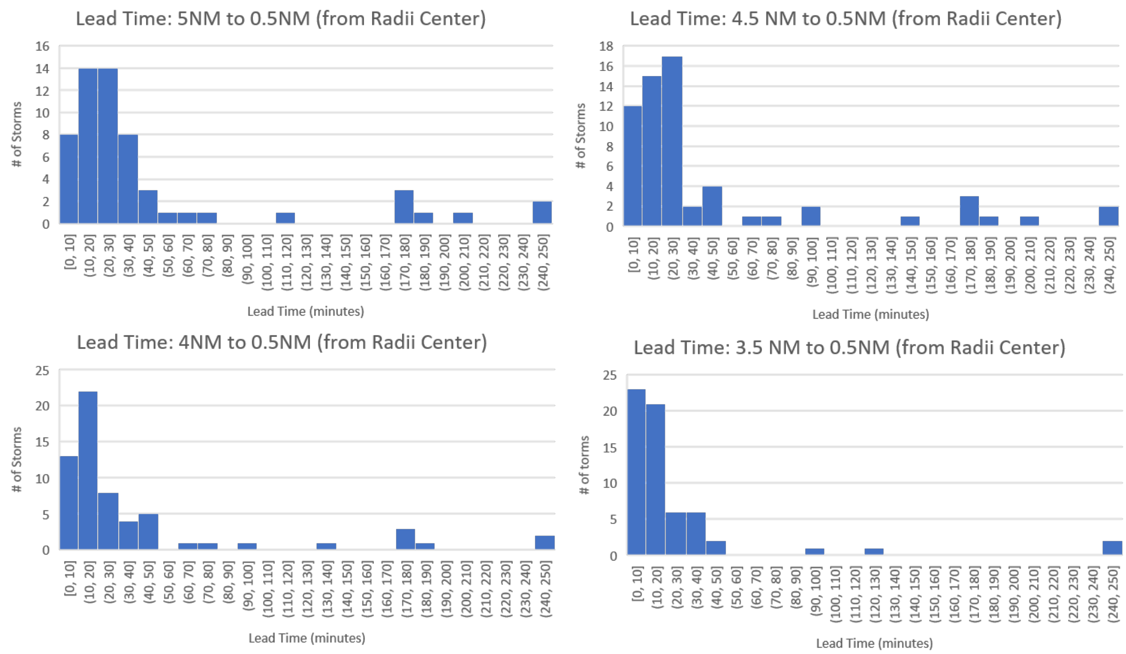


Figure 17: Storms - Merge/Split Construct



The second area for potential research encompasses examining lightning event lead time. In this context, lead time represents the amount of time it takes for lightning to arrive within 0.5NM of the radii center, when its starting point is at the lightning warning radii distance (for example, the 5NM baseline). Shown in Figure 18, for the case of a 5NM warning circle, the lead time falls mainly around the 30 minute mark. There are cases where lead time can take as long as 4 hours. Factors that may affect longer lead time length could include seasonal, temperature (hottest or coldest month), and storm type, to name a few. There are also a few cases where lead time falls under 10 minutes. This can be due to a large cloud overhead that allows lightning aloft to travel internally through its mass at greater speed or two different storms occurring within established spatial and temporal constraints as mentioned in the first potential area of research. Further study in this area can help determine what effect lead time length has for awareness and to aid operations in adjusting space launch progress accordingly. Additionally, we see that when lightning radii is reduced based on the alternatives proposed, lightning lead time on average also decreases. Further analysis regarding lead time can be pivotal to determining whether reducing lightning radii will allow for enough reaction time to pause operations and take shelter.

Figure 18: 2017 Storms - Lead Time



The third area for potential research involves examining and applying the concept in this study towards the LLCC. At this time, LLCC guidance requires a 10NM stand-off for all facets of launch operations. This study provided the insight that lightning aloft travels 9NM outside the predetermined boundary, while cloud to ground lightning travels to 8NM. This baseline insight along with results from this study may be expanded upon to determine its applicability to launch operations as a whole, potentially providing further increase in space launch productivity in other areas.

Appendix A. Data Processing Code

```
---
title: "Data Processing"
author: "Kimberly Holland"
date: "11 Oct 20"
output: html_notebook
---
Install necessary packages
```{r}
library(MASS)
library(tictoc)
library(lubridate)
library(data.table)
library(Imap)

#Clear workspace
gc()
rm(list = ls())
```
Upload lightning aloft data

```{r}
#Start recording process time
tic()


#Import Data Files
ltng_aloft_files <- list.files(path = "C:\\Users\\queen\\Desktop\\
                               OPER Thesis\\45 OG Data\\Coding\\
                               Raw Data\\CC", recursive = TRUE,
                               pattern = "\\.txt$", full.names = TRUE)
ltng_aloft_files

#Combine files into one dataframe
CC <- rbindlist(sapply(ltng_aloft_files, fread, simplify = FALSE),
                use.names = TRUE, idcol = "FileName")
CC$FileName <- NULL

#Label headers
names(CC) = c("Date","Time","Latitude","Longitude")

#Change to date format
CC$Date = as.Date(CC$Date, format="%Y-%m-%d")
CC$Type = "C"

#Combine date and time
CC$DateTime = as.POSIXct(paste(CC$Date, CC$Time),
                         format="%Y-%m-%d %H:%M:%S", tz = "EST")

#Stop recording process time
toc()
```
```

````
```
Upload cloud to ground data

```{r}
#Start recording process time
tic()

#Import Data Files
ctg_files <- list.files(path = "C:\\Users\\queen\\Desktop\\
                        OPER Thesis\\45 OG Data\\Coding\\
                        Raw Data\\CG", recursive = TRUE,
                        pattern = "\\.txt$", full.names = TRUE)
#ctg_files

#Combine files into one dataframe
CG <- rbindlist(sapply(ctg_files, fread, simplify = FALSE),
                use.names = TRUE, idcol = "FileName")
CG <- CG[,c(2:5,12)]
names(CG) = c("Date","Time","Latitude","Longitude","Type")
CG$Date = as.Date(CG$Date, format="%Y-%m-%d")

#Combine date and time
CG$DateTime = as.POSIXct(paste(CG$Date, CG$Time),
                         format="%Y-%m-%d %H:%M:%S", tz = "EST")

#Compile CG and CC
ltngCG_CC <- rbindlist(list(CC,CG))

#Print results
ltngCG_CC
#plot(ltngCG_CC$Latitude,ltngCG_CC$Longitude)

#Stop recording process time
toc()
```
Screen flashes outside 25NM
```{r}
#Start recording process time
tic()

#Subset data within 25NM
ltngCG_CC_Filtered <- subset(ltngCG_CC,
                             gdist(ltngCG_CC$Longitude,
                                   ltngCG_CC$Latitude,
                                   -80.6783,28.489,
                                   units = "nm")<25)
ltngCG_CC_Filtered


#Sort by date and time
class(ltngCG_CC_Filtered$DateTime)
ltngCG_CC_Filtered <- ltngCG_CC_Filtered[do.call(order, ltngCG_CC_Filtered), ]
````

````
#2d plot of Events
plot(ltngCG_CC_Filtered$Latitude,ltngCG_CC_Filtered$Longitude)

#Stop recording process time
toc()
```
Cluster lightning by > 15 min of inactivity
- Groups together consecutive flash activity continuing from Day 1 to Day 2
- Creates subgroups for smaller matrix calculations
```{r}
#Start recording process time
tic()

#Partition a new group every time there is >15 min inactivity
ltngCG_CC_Filtered$Rollover_Cluster <-
  cumsum(c(TRUE, difftime(ltngCG_CC_Filtered$DateTime[-1],
                 ltngCG_CC_Filtered$DateTime[-nrow(ltngCG_CC_Filtered)],
                 units = 'mins') > 15))
ltngCG_CC_Filtered

#Stop recording process time
toc()
```
Check if last cluster of last month belongs with first cluster
of next month. If so, add this grouping to end of last month.
```{r}
#Add last cluster of last month
lastEvent <- read.csv("C:\\Users\\queen\\Desktop\\OPER Thesis\\
                      45 OG Data\\Coding 2\\Last Clusters of Month\\
                      LastClusterJun19.csv",stringsAsFactors = FALSE)

#Label headers
names(lastEvent) = c("X","Date","Time","Latitude","Longitude",
                     "Type","DateTime","Rollover_Cluster")

#Change to date and time format
lastEvent$Date = as.Date(lastEvent$Date, format="%Y-%m-%d")
lastEvent$DateTime = as.POSIXct(paste(lastEvent$Date, lastEvent$Time),
                                format="%Y-%m-%d %H:%M:%S", tz = "EST")

#Assign Rollover ID
lastEvent$Rollover_Cluster <- 1

#Add last cluster to new dataframe
ltngCG_CC_Filtered <- rbind(lastEvent[,2:8], ltngCG_CC_Filtered)
```

Plot sequence by Rollover Cluster
```{r}
#Creates a chart to observe all storms in one Rollover
tic()
````

````
library(ggplot2)
library(plotly)
library(gapminder)

fig1 <- ltngCG_CC_Filtered %>%
  plot_ly(
    x = ~Latitude,
    y = ~Longitude,
    frame = ~Rollover_Cluster,
    color = ~Type,
    text = ~Rollover_Cluster,
    type = 'scatter',
    mode = 'markers'
  )
fig1 <- fig1 %>% layout(
  xaxis = list(
    type = "log"
  )
)
fig1 <- fig1%>%animation_opts(
  frame = 750, transition = 0, easing = "linear-in-out",redraw = FALSE
)
fig1
toc()
```
Save By Rollover
```{r}
#Save rollower groups based on # of storms
Rollover <- flashGrp1 #grp 1, (3 less)
#Rollover <- flashGrp1a #grp1, (4 more)

#Rollover <- rbind(flashGrp1a,flashGrp1b) #grp1 (4 more) & grp 2 (3 less)
#Rollover <- rbind(flashGrp1a,flashGrp2a,flashGrp3a,flashGrp4a,)
              #grp1 (4 more) & grp 2 (4 more)

#Save as a csv file
write.csv(Rollover, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\Jun18Rollover141_onlyVal.csv")

#Save as a txt file
#write.table(Rollover,"Jun19Rollover70.txt",sep="\t",row.names=FALSE)
```
````
````

Appendix B. Spatial and Temporal Clustering Code

```
---
title: "Spatial/Temporal Clustering"
author: "Kimberly Holland"
date: "11 Oct 20"
output: html_notebook
---
Install necessary packages
```{r}
library(MASS)
library(tictoc)
library(lubridate)
library(data.table)
library(Imap)
gc()
rm(list = ls())
```
Identify one Cluster for Location Clustering
```{r}
tic()
flashGrp1 <- ltngCG_CC_Filtered[which(ltngCG_CC_Filtered$Rollover_Cluster=="66"),]
flashGrp1
toc()
```
Save last cluster
```{r}
write.csv(flashGrp1, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\LastClusterJun19_41.csv")
```

Save unprocessed rollover
```{r}
write.csv(flashGrp1, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\Large Storm Rollovers\\
          Feb17LargeRollover14.csv")
```
Group Event by Spatial and Temporal Criteria
```{r}
#Start recording process time
tic()

#Install more packages
library(geosphere)
library(sp)

#Start Storm counter
j <- 1

#Add a blank column to fill storm ID
flashGrp1$DistTime <- NA
```
```

```
#First event belongs to first storm
flashGrp1$DistTime[1] <- j

#Continue to loop while there are events that are unassigned
while (any(is.na(flashGrp1$DistTime))==TRUE) {

  #Identify first event that is unassigned
  start <- which(is.na(flashGrp1[[1]]))

  #Loop from first unassigned event to last event
for (i in start:nrow(flashGrp1)) {

  #Check each next next event that is unassigned,
  #if it is unassigned, assign it to a storm
  if (is.na(flashGrp1$DistTime[i])==TRUE){

   #Subset the rows that belong to a storm so far
  group <- flashGrp1[which(flashGrp1$DistTime==j)]

  #Get the time from the most recent event in the storm
  lastrow <- tail(group, n = 1)
  moment <- lastrow$DateTime

  #Subset the events in the storm that are < 10 minutes old
  flashes <- group[which(difftime(moment,group$DateTime,units = "mins")<10)]

  #Then take the coordinates of the storm
  h <- cbind(flashes$Longitude,flashes$Latitude)
  #And take coordinates of the next unassigned event
  point <- cbind(flashGrp1$Longitude[i],flashGrp1$Latitude[i])
  }

  #Check in next unassigned event meets time/distance criteria
  #If it does, add it to the current storm j,
  #If not, leave it unassigned
  ifelse(min(spDistsN1(h,point,longlat = TRUE))<16 &
           difftime(flashGrp1$DateTime[i],moment,units = "mins")<15,
         flashGrp1$DistTime[i] <-j,flashGrp1$DistTime[i] <-j+1)

}

  #At the end of loop, establish new storm j
  j=j+1
}

#Stop recording process time
toc()
```

Appendix C. Ellipse Fitting Code

```
---
title: "Ellipse Fitting"
author: "Kimberly Holland"
date: "11 Oct 20"
output: html_notebook
---
Install necessary packages
```{r}
library(MASS)
library(tictoc)
library(lubridate)
library(data.table)
library(Imap)
gc()
rm(list = ls())
```
Pick a Storm
```{r}
#Choose one storm based on clustering criteria
flashGrp1 <- flashGrp8a
#Print results
flashGrp1

```
Ellipse Fitting Process
```{r}
#Start recording process time
tic()
#Install more packages
library(sp)
library(cluster)

#Identify first 3 flashes
FlashPoints <- cbind(flashGrp1$Latitude[1:3],flashGrp1$Longitude[1:3])
FlashPoints
plot(FlashPoints)

#Fit an ellipse and get next flash point
tic()


#Fit an ellipse around flashpoints
EllipseShape <- ellipsoidhull(FlashPoints)
EllipseShape

#Get set of points that make an ellipse
EllipsePoints <- predict(EllipseShape)
h <- cbind(EllipsePoints[,2],EllipsePoints[,1])
#h
```
```

```
# Create function to check if the next flashpoint (xp, yp)
# belongs to the ellipse with parameters a,b,... with tolerance eps
onEllipse <- function (xp, yp, a, b, x0, y0, alpha, eps=1e-3) {
  return(abs((cos(alpha)*(xp-x0)+sin(alpha)*(yp-y0))^2/a^2+
               (sin(alpha)*(xp-x0)-cos(alpha)*
                  (yp-y0))^2/b^2 - 1) <= eps)
}

# Create function to check if the point (xp, yp) is inside
# the ellipse with parameters a,b,...
insideEllipse <- function (xp, yp, a, b, x0, y0, alpha) {
  return((cos(alpha)*(xp-x0)+sin(alpha)*(yp-y0))^2/a^2+
           (sin(alpha)*(xp-x0)-cos(alpha)*(yp-y0))^2/b^2 <= 1)
}

#Iterate Next Flash, Ellipse Fitting, and Distance Measurement
##Establish empty list to fill in distances
flashDistance <- c()
flashGrp1$Distances <- NA

#Establish iteration for one Storm
for (i in 4:nrow(flashGrp1)){
#for (i in 11:250){

#Get next flash
nextFlash <- cbind(flashGrp1$Longitude[i],flashGrp1$Latitude[i])


#Establish/Re-establish parameters to check where point falls
xp <- flashGrp1$Latitude[i]
yp <- flashGrp1$Longitude[i]
x0 <- EllipseShape$loc[1] # centroid locations
y0 <- EllipseShape$loc[2]
eg <- eigen(EllipseShape$cov)
axes <- sqrt(eg$values)
alpha <- atan(eg$vectors[1,1]/eg$vectors[2,1]) #angle of major axis with x axis
a <- sqrt(EllipseShape$d2) * axes[1]  # major axis length
b <- sqrt(EllipseShape$d2) * axes[2]  # minor axis length
ifl <- 1

#Check to see if outside the ellipse then get distance.
#If it is, add to flash distance list
if(insideEllipse(xp, yp, a, b, x0, y0, alpha)==FALSE &&
   onEllipse(xp, yp, a, b, x0, y0, alpha)==FALSE){
  nextDist <- min(spDistsN1(h, nextFlash, longlat = TRUE))
  flashDistance <- rbind(flashDistance, nextDist)
  flashGrp1$Distances[i] <- nextDist

  #Omit flashes 10 minutes old
  movingTstorm <- subset(flashGrp1[1:i,],
                         difftime(flashGrp1$DateTime[1:i],
                                  flashGrp1$DateTime[i],
```

```
                                  units = "mins")<10)
  FlashPoints <- cbind(movingTstorm$Latitude[1:i],
                       movingTstorm$Longitude[1:i])
  FlashPoints

  #Fit an ellipse around flashpoints
  EllipseShape <- ellipsoidhull(FlashPoints)
  EllipseShape

  #Get set of points that make an ellipse
  EllipsePoints <- predict(EllipseShape)
  h <- cbind(EllipsePoints[,2],EllipsePoints[,1])
  h

}

next(i)
}

#Recording process time ends
toc()

#Create a Histogram for Lightning Distances
hist(flashDistance,
main="Distance Distribution for Cluster",
xlab="Distances by km",
#xlim=c(50,100),
col="deepskyblue3"
)
```
```

---
```

Appendix D. Validation Code

```
---
title: "Validation"
author: "Kimberly Holland"
date: "20 Dec 20"
output: html_notebook
---

Install necessary packages & clear workspace
```{r}
library(MASS)
library(tictoc)
library(lubridate)
library(data.table)
library(dplyr)
```
Upload lightning data
```{r}
#clear workspace
gc()
rm(list = ls())

#Get data from folder and put in R dataframe
Storms <- read.csv("C:\\Users\\queen\\Desktop\\OPER Thesis\\45 OG
                   Data\\Coding 2\\Rollovers\\17_07\\
                   Jul19Rollover45.csv",stringsAsFactors = FALSE)

#Label dataframe headers
names(Storms) = c("X","Date","Time","Latitude","Longitude","Type",
                  "DateTime","Rollover_Cluster","DistTime")

#Change these columns to date and time format
Storms$Date = as.Date(Storms$Date, format="%Y-%m-%d")
Storms$DateTime = as.POSIXct(paste(Storms$Date, Storms$Time),
                             format="%Y-%m-%d %H:%M:%S", tz = "EST")

#Print results
Storms
```
Get Distances
```{r}
#start recording process time
tic()

#install more packages
library(sp)
library(geosphere)

#Create Data Frame and Add Warning Radii Info
##Create vectors with info
circlesFacility <- c("KSC","CCSFS","CCSFS","CCSFS","CCSFS","PSFB",
```
```

```
                     "OFFSITE","OFFSITE","KSC","KSC","KSC","CCSFS")
circlesName <- c("HAULOVER","40/41","37/ITL","36/46","PORT","PSFB",
                 "ASTROTECH","CIDCO PARK","LC39","SLF","KSC
                 INDUSTRIAL AREA","CAPE CENTRAL")
circlesLat <- c(28.736400,28.572716,28.531489,28.464949,28.413360,
                28.234080,28.524240,28.411806,28.604190,28.614750,
                28.520000,28.467410)
circlesLong <- c(-80.754700,-80.579891,-80.574834,-80.533056,
                 -80.600000,-80.609760,-80.817200,-80.772892,
                 -80.631700,-80.694860,-80.650000,-80.567110)
circlesRadius <- c("5","5","5","5","5","5","5","5","6","6","6","6")

##Add vectors to dataframe
circles <- as.data.frame(cbind(circlesFacility,circlesName,
                               circlesLat,circlesLong,circlesRadius))
##Add column headers
names(circles) <- c("Facility","Name","CenterLat",
                    "CenterLong", "Radius NM")
##Print results
circles

#Create subdataframe with radii center coordinates
circleCtr <- cbind(circlesLong,circlesLat)
circleCtr

#Create an empty matrix
mat <- matrix(ncol = 12)

#Measure distances from each event in a Storm to each radii center
##Then populate results in empty matrix
for (i in 1:nrow(Storms)) {
  point <- (cbind(Storms$Longitude[i],Storms$Latitude[i]))
  point
  measures <- spDistsN1(circleCtr,point,longlat = TRUE)

  mat <- rbind(mat,c(measures))
}

#Omit first matrix row, it's empty
mat <- mat[-1,]

#Convert km to NM values
mat <- mat*0.5399568

#Turn matrix into a dataframe
dist2Circles <- data.frame(mat)

#Label column headers
names(dist2Circles) <- c("HAULOVER","40/41","37/ITL","36/46","PORT",
                         "PSFB","ASTROTECH","CIDCO PARK","LC39",
                         "SLF","KSC INDUSTRIAL AREA","CAPE CENTRAL")
```

````
#Add new columns to original dataframe
Storms <- cbind(Storms,dist2Circles)
Storms

#end recording process time
toc()
```
Get Results for 5NM Warning Circles (Loop through several storms)
```{r}
#start recording process time
tic()

#install more packages
library(pracma)

#Create empty vector
vec <- c()

#Create a dataframe
results <- data.frame(matrix(ncol = 18))

#Label dataframe headers
names(results) <- c("Rollover_Cluster","DistTime","Location",
                  "5NM_TimeStart","5NM_DistStart","5NM_TimeEnd",
                  "4.5NM_TimeStart","4.5NM_DistStart","4.5NM_TimeEnd",
                  "4NM_TimeStart","4NM_DistStart","4NM_TimeEnd",
                  "3.5NM_TimeStart","3.5NM_DistStart","3.5NM_TimeEnd",
                  "0.5NM_TimeStart","0.5NM_DistStart","CGStrike")

#Create another dataframe, same size
compile5 <- data.frame(matrix(ncol = 18))

#Label dataframe headers, same labels
names(compile5) <- c("Rollover_Cluster","DistTime","Location",
                 "5NM_TimeStart","5NM_DistStart","5NM_TimeEnd",
                 "4.5NM_TimeStart","4.5NM_DistStart","4.5NM_TimeEnd",
                 "4NM_TimeStart","4NM_DistStart","4NM_TimeEnd",
                 "3.5NM_TimeStart","3.5NM_DistStart","3.5NM_TimeEnd",
                 "0.5NM_TimeStart","0.5NM_DistStart","CGStrike")


#For 5NM Warning Circles
## Loop through all storms in one rollover group
for (h in 1:max(Storms$DistTime)){
  Storm <- Storms[which(Storms$DistTime==h),]

## For every storm, loop through each Warning Radii
## If the storm goes within 5NM of a warning radii,
## Subset the row data of the events that went within the radii
## Pull data such as starting distance and time, end time, radii ID,
## Populate this data in empty vector
## Do the same for 4.5NM, 4NM, 3.5NM, and 0.5NM
````

```
## Also check in any events that fall within 0.5NM are CG (T or F)
for (i in 10:17) {
    if (any(Storm[,i]<5)==TRUE){
      temp1 <- subset(Storm,Storm[,i]<5)
      rowstart1 <- head(temp1,1)
      rowend1 <- tail(temp1,1)
      vec[1:2] <- rowstart1[,8:9]
      vec[3] <- i
      vec[4] <- rowstart1$DateTime
      vec[5] <- rowstart1[,i]
      vec[6] <- rowend1$DateTime
      if (any(Storm[,i]<4.5)==TRUE){
        temp2 <- subset(Storm,Storm[,i]<4.5)
        rowstart2 <- head(temp2,1)
        rowend2 <- tail(temp2,1)
        vec[7] <- rowstart2$DateTime
        vec[8] <- rowstart2[,i]
        vec[9] <- rowend2$DateTime
      }else{
        vec[7] <- NA
        vec[8] <- NA
        vec[9] <- NA
      }
      if (any(Storm[,i]<4)==TRUE){
        temp3 <- subset(Storm,Storm[,i]<4)
        rowstart3 <- head(temp3,1)
        rowend3 <- tail(temp3,1)
        vec[10] <- rowstart3$DateTime
        vec[11] <- rowstart3[,i]
        vec[12] <- rowend3$DateTime
      }else{
        vec[10] <- NA
        vec[11] <- NA
        vec[12] <- NA
      }
      if (any(Storm[,i]<3.5)==TRUE){
        temp4 <- subset(Storm,Storm[,i]<3.5)
        rowstart4 <- head(temp4,1)
        rowend4 <- tail(temp4,1)
        vec[13] <- rowstart4$DateTime
        vec[14] <- rowstart4[,i]
        vec[15] <- rowend4$DateTime
      }else{
        vec[13] <- NA
        vec[14] <- NA
        vec[15] <- NA
      }
      if (any(Storm[,i]<0.5)==TRUE){
        temp5 <- subset(Storm,Storm[,i]<0.5)
        rowstart5 <- head(temp5,1)
        vec[16] <- rowstart5$DateTime
        vec[17] <- rowstart5[,i]
```

```
        #Is strike CG?
        ifelse(any(temp5$Type=="G"), vec[18] <- 1, vec[18] <- 0)
      }else{
        vec[16] <- NA
        vec[17] <- NA
        vec[18] <- NA
      }

      #Each vector of data will be populated in this dataframe
      #and vector gets reused for each radii checked in the loop
      results <- rbind(results,vec)
      j=j+1
    }

}
}
#data frame in loop is transcribed to this dataframe
compile5 <- results

#first row is omitted because it is empty (consequence of
#creating an empty dataframe)
compile5 <- compile5[-c(1),]

#Convert these columns from decimal to date and time format
#(consequence of using vectors to cut process time)
compile5$`5NM_TimeStart` <- as.POSIXct(compile5$`5NM_TimeStart`,
                                       origin ="1970-01-01")
compile5$`5NM_TimeEnd` <- as.POSIXct(compile5$`5NM_TimeEnd`,
                                     origin ="1970-01-01")
compile5$`4.5NM_TimeStart` <- as.POSIXct(compile5$`4.5NM_TimeStart`,
                                         origin ="1970-01-01")
compile5$`4.5NM_TimeEnd` <- as.POSIXct(compile5$`4.5NM_TimeEnd`,
                                       origin ="1970-01-01")
compile5$`4NM_TimeStart` <- as.POSIXct(compile5$`4NM_TimeStart`,
                                       origin ="1970-01-01")
compile5$`4NM_TimeEnd` <- as.POSIXct(compile5$`4NM_TimeEnd`,
                                     origin ="1970-01-01")
compile5$`3.5NM_TimeStart` <- as.POSIXct(compile5$`3.5NM_TimeStart`,
                                         origin ="1970-01-01")
compile5$`3.5NM_TimeEnd` <- as.POSIXct(compile5$`3.5NM_TimeEnd`,
                                       origin ="1970-01-01")
compile5$`0.5NM_TimeStart` <- as.POSIXct(compile5$`0.5NM_TimeStart`,
                                         origin ="1970-01-01")
toc()
```
If there's a dataframe, process results
```{r}
tic()

#Get Time Duration Storm stays within 5NM bubble
compile5$`5NM_Duration` <- difftime(compile5$`5NM_TimeEnd`,compile5$`5NM_TimeStart`,
                                    units = "mins")
```

```
#Get Time Saved if reduced to 4.5NM bubble
compile5$`4.5NM_Duration` <- difftime(compile5$`4.5NM_TimeEnd`,
                                      compile5$`4.5NM_TimeStart`,
                                      units = "mins")
compile5$`4.5NM_Duration`[is.na(compile5$`4.5NM_Duration`)] = 0
compile5$`4.5NM_TimeSaved` <- compile5$`5NM_Duration`-compile5$`4.5NM_Duration`

#Get Time Saved if reduced to 4NM bubble
compile5$`4NM_Duration` <- difftime(compile5$`4NM_TimeEnd`,
                                    compile5$`4NM_TimeStart`,
                                    units = "mins")
compile5$`4NM_Duration`[is.na(compile5$`4NM_Duration`)] = 0
compile5$`4NM_TimeSaved` <- compile5$`5NM_Duration`-compile5$`4NM_Duration`

#Get Time Saved if reduced to 3.5NM bubble
compile5$`3.5NM_Duration` <- difftime(compile5$`3.5NM_TimeEnd`,
                                      compile5$`3.5NM_TimeStart`,
                                      units = "mins")
compile5$`3.5NM_Duration`[is.na(compile5$`3.5NM_Duration`)] = 0
compile5$`3.5NM_TimeSaved` <- compile5$`5NM_Duration`-compile5$`3.5NM_Duration`

#Loop through row in dataframe
for (k in 1:nrow(compile5)) {

#Check if False Alarm 4.5NM
ifelse((isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          compile5$`5NM_DistStart`[k]> 4.5) &&
          isempty(compile5$`4.5NM_TimeStart`[k])==TRUE,
          compile5$FalseAlarm_4.5NM[k] <- 1,
          compile5$FalseAlarm_4.5NM[k] <- 0)

#Check if False Alarm 4NM
ifelse((isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          compile5$`5NM_DistStart`[k]> 4) &&
          isempty(compile5$`4NM_TimeStart`[k])==TRUE,
          compile5$FalseAlarm_4NM[k] <- 1,
          compile5$FalseAlarm_4NM[k] <- 0)

#Check if False Alarm 3.5NM
ifelse(isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
         compile5$`5NM_DistStart`[k]> 3.5 &&
         isempty(compile5$`3.5NM_TimeStart`[k])==TRUE,
         compile5$FalseAlarm_3.5NM[k] <- 1,
         compile5$FalseAlarm_3.5NM[k] <- 0)


#Check if Failure 4.5NM
compile5$Failure_4.5NM[k] <- 0
ifelse(((isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          compile5$`5NM_DistStart`[k]>4.5) &&
          isempty(compile5$`4.5NM_TimeStart`[k])==TRUE &&
```

```
          isempty(compile5$`0.5NM_TimeStart`[k])==FALSE)||
         (isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          isempty(compile5$`4.5NM_TimeStart`[k])==FALSE &&
          isempty(compile5$`0.5NM_TimeStart`[k])==FALSE &&
          compile5$`0.5NM_TimeStart`[k]<compile5$`4.5NM_TimeStart`[k]),
          compile5$FalseAlarm_4.5NM[k] <- 1,
          compile5$FalseAlarm_4.5NM[k] <- 0)


#Check if Failure 4NM
compile5$Failure_4NM[k] <- 0
ifelse(((isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          compile5$`5NM_DistStart`[k]>4) &&
          isempty(compile5$`4NM_TimeStart`[k])==TRUE &&
          isempty(compile5$`0.5NM_TimeStart`[k])==FALSE) ||
         (isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          isempty(compile5$`4NM_TimeStart`[k])==FALSE &&
          isempty(compile5$`0.5NM_TimeStart`[k])==FALSE &&
          compile5$`0.5NM_TimeStart`[k]<compile5$`4NM_TimeStart`[k]),
          compile5$`Failure_4NM`[k] <- 1,
          compile5$`Failure_4NM`[k] <- 0)

#Check if Failure 3.5NM
compile5$Failure_3.5NM[k] <- 0
ifelse(((isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          compile5$`5NM_DistStart`[k]>3.5) &&
          isempty(compile5$`3.5NM_TimeStart`[k])==TRUE &&
          isempty(compile5$`0.5NM_TimeStart`[k])==FALSE) ||
         (isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
          isempty(compile5$`3.5NM_TimeStart`[k])==FALSE &&
          isempty(compile5$`0.5NM_TimeStart`[k])==FALSE &&
          compile5$`0.5NM_TimeStart`[k]<compile5$`3.5NM_TimeStart`[k]),
          compile5$Failure_3.5NM[k] <- 1,
          compile5$Failure_3.5NM[k] <- 0)

#Check duration from 5NM Warning to 0.5NM
compile5$Time_5NM_toFacility[k] <- NA
ifelse(isempty(compile5$`5NM_TimeStart`[k])==FALSE &&
         compile5$`5NM_DistStart`[k]> 4.5 &&
         isempty(compile5$`0.5NM_TimeStart`[k])==FALSE,
         compile5$Time_5NM_toFacility[k] <-
         abs(difftime(compile5$`5NM_TimeStart`[k],
                      compile5$`0.5NM_TimeStart`[k],units = "mins")),
         compile5$Time_5NM_toFacility[k] <- NA)

#check duration from 4.5NM Warning to 0.5NM
compile5$Time_4.5NM_toFacility[k] <- NA
ifelse(isempty(compile5$`4.5NM_TimeStart`[k])==FALSE &&
         compile5$`4.5NM_DistStart`[k]> 4 &&
         isempty(compile5$`0.5NM_TimeStart`[k])==FALSE,
         compile5$Time_4.5NM_toFacility[k] <-
         abs(difftime(compile5$`4.5NM_TimeStart`[k],
```

```
                      compile5$`0.5NM_TimeStart`[k],units = "mins")),
         compile5$Time_4.5NM_toFacility[k] <- NA)

#check duration from 4NM Warning to 0.5NM
compile5$Time_4NM_toFacility[k] <- NA
ifelse(isempty(compile5$`4NM_TimeStart`[k])==FALSE &&
         compile5$`4NM_DistStart`[k]> 3.5 &&
         isempty(compile5$`0.5NM_TimeStart`[k])==FALSE,
         compile5$Time_4NM_toFacility[k] <-
         abs(difftime(compile5$`4NM_TimeStart`[k],
                      compile5$`0.5NM_TimeStart`[k],
                      units = "mins")),
         compile5$Time_4NM_toFacility[k] <- NA)

#check duration from 3.5NM Warning to 0.5NM
compile5$Time_3.5NM_toFacility[k] <- NA
ifelse(isempty(compile5$`3.5NM_TimeStart`[k])==FALSE &&
         compile5$`3.5NM_DistStart`[k]> 3 &&
         isempty(compile5$`0.5NM_TimeStart`[k])==FALSE,
         compile5$Time_3.5NM_toFacility[k] <-
         abs(difftime(compile5$`0.5NM_TimeStart`[k],
                      compile5$`3.5NM_TimeStart`[k],
                      units = "mins")),
         compile5$Time_3.5NM_toFacility[k] <- NA)
}

toc()
```
Get Results for 6NM Warning Circles (Loop through several storms)
```{r}
#Same notes as 5NM Warning Circle process, only difference is it is applied to 6NM radii.

library(pracma)
tic()

j=1

vec <- c()
results <- data.frame(matrix(ncol = 18))
names(results) <- c("Rollover_Cluster","DistTime","Location",
                  "6NM_TimeStart","6NM_DistStart","6NM_TimeEnd",
                  "5.5NM_TimeStart","5.5NM_DistStart","5.5NM_TimeEnd",
                  "5NM_TimeStart","5NM_DistStart","5NM_TimeEnd",
                  "4.5NM_TimeStart","4.5NM_DistStart","4.5NM_TimeEnd",
                  "1.5NM_TimeStart","1.5NM_DistStart","CGStrike")

compile6 <- data.frame(matrix(ncol = 18))
names(compile6) <- c("Rollover_Cluster","DistTime","Location",
                  "6NM_TimeStart","6NM_DistStart","6NM_TimeEnd",
                  "5.5NM_TimeStart","5.5NM_DistStart","5.5NM_TimeEnd",
                  "5NM_TimeStart","5NM_DistStart","5NM_TimeEnd",
                  "4.5NM_TimeStart","4.5NM_DistStart","4.5NM_TimeEnd",
```

```
                  "1.5NM_TimeStart","1.5NM_DistStart","CGStrike")


#6NM Warning Circles
for (h in 1:max(Storms$DistTime)){
  Storm <- Storms[which(Storms$DistTime==h),]

for (i in 18:length(Storm)) {

    if (any(Storm[,i]<6)==TRUE){
      temp1 <- subset(Storm,Storm[,i]<6)
      rowstart1 <- head(temp1,1)
      rowend1 <- tail(temp1,1)
      vec[1:2] <- rowstart1[,8:9]
      vec[3] <- i
      vec[4]<- rowstart1$DateTime
      vec[5] <- rowstart1[,i]
      vec[6]<- rowend1$DateTime
      if (any(Storm[,i]<5.5)==TRUE){
        temp2 <- subset(Storm,Storm[,i]<5.5)
        rowstart2 <- head(temp2,1)
        rowend2 <- tail(temp2,1)
        vec[7]<- rowstart2$DateTime
        vec[8] <- rowstart2[,i]
        vec[9]<- rowend2$DateTime
      }else{
        vec[7] <- NA
        vec[8] <- NA
        vec[9] <- NA
      }
      if (any(Storm[,i]<5)==TRUE){
        temp3 <- subset(Storm,Storm[,i]<5)
        rowstart3 <- head(temp3,1)
        rowend3 <- tail(temp3,1)
        vec[10]<- rowstart3$DateTime
        vec[11] <- rowstart3[,i]
        vec[12]<- rowend3$DateTime
      }else{
        vec[10] <- NA
        vec[11] <- NA
        vec[12] <- NA
      }
      if (any(Storm[,i]<4.5)==TRUE){
        temp4 <- subset(Storm,Storm[,i]<4.5)
        rowstart4 <- head(temp4,1)
        rowend4 <- tail(temp4,1)
        vec[13]<- rowstart4$DateTime
        vec[14] <- rowstart4[,i]
        vec[15]<- rowend4$DateTime
      }else{
        vec[13] <- NA
        vec[14] <- NA
```

```
        vec[15] <- NA
      }
      if (any(Storm[,i]<1.5)==TRUE){
        temp5 <- subset(Storm,Storm[,i]<1.5)
        rowstart5 <- head(temp5,1)
        vec[16]<- rowstart5$DateTime
        vec[17] <- rowstart5[,i]
      #Is strike CG?
        ifelse(any(temp5$Type=="G"), vec[18] <- 1, vec[18] <- 0)
      }else{
        vec[16] <- NA
        vec[17] <- NA
        vec[18] <- NA
      }

      results <- rbind(results,vec)
      j=j+1
    }
}
}

compile6 <- results
#compile6 <- rbind(compile6,results)
compile6 <- compile6[-c(1),]
compile6$`6NM_TimeStart` <- as.POSIXct(compile6$`6NM_TimeStart`,
                                       origin ="1970-01-01")
compile6$`6NM_TimeEnd` <- as.POSIXct(compile6$`6NM_TimeEnd`,
                                     origin ="1970-01-01")
compile6$`5.5NM_TimeStart` <- as.POSIXct(compile6$`5.5NM_TimeStart`,
                                         origin ="1970-01-01")
compile6$`5.5NM_TimeEnd` <- as.POSIXct(compile6$`5.5NM_TimeEnd`,
                                       origin ="1970-01-01")
compile6$`5NM_TimeStart` <- as.POSIXct(compile6$`5NM_TimeStart`,
                                       origin ="1970-01-01")
compile6$`5NM_TimeEnd` <- as.POSIXct(compile6$`5NM_TimeEnd`,
                                     origin ="1970-01-01")
compile6$`4.5NM_TimeStart` <- as.POSIXct(compile6$`4.5NM_TimeStart`,
                                         origin ="1970-01-01")
compile6$`4.5NM_TimeEnd` <- as.POSIXct(compile6$`4.5NM_TimeEnd`,
                                       origin ="1970-01-01")
compile6$`1.5NM_TimeStart` <- as.POSIXct(compile6$`1.5NM_TimeStart`,
                                         origin ="1970-01-01")

toc()
```
If there's a dataframe, process results
```{r}
#Same notes as 5NM Warning Circle process,
#only difference is it is applied to 6NM radii.

tic()
#Get Time Duration Storm stays within 6NM bubble
```

```
compile6$`6NM_Duration` <- difftime(compile6$`6NM_TimeEnd`,
                                    compile6$`6NM_TimeStart`,
                                    units = "mins")

#Get Time Saved if reduced to 5.5NM bubble
compile6$`5.5NM_Duration` <- difftime(compile6$`5.5NM_TimeEnd`,
                                      compile6$`5.5NM_TimeStart`,
                                      units = "mins")
compile6$`5.5NM_Duration`[is.na(compile6$`5.5NM_Duration`)] = 0
compile6$`5.5NM_TimeSaved` <- compile6$`6NM_Duration`-compile6$`5.5NM_Duration`

#Get Time Saved if reduced to 5NM bubble
compile6$`5NM_Duration` <- difftime(compile6$`5NM_TimeEnd`,
                                    compile6$`5NM_TimeStart`,
                                    units = "mins")
compile6$`5NM_Duration`[is.na(compile6$`5NM_Duration`)] = 0
compile6$`5NM_TimeSaved` <- compile6$`6NM_Duration`-compile6$`5NM_Duration`

#Get Time Saved if reduced to 4.5NM bubble
compile6$`4.5NM_Duration` <- difftime(compile6$`4.5NM_TimeEnd`,
                                      compile6$`4.5NM_TimeStart`,
                                      units = "mins")
compile6$`4.5NM_Duration`[is.na(compile6$`4.5NM_Duration`)] = 0
compile6$`4.5NM_TimeSaved` <- compile6$`6NM_Duration`-compile6$`4.5NM_Duration`

for (k in 1:nrow(compile6)) {
#Check if False Alarm 5.5NM
ifelse((isempty(compile6$`6NM_TimeStart`[k])==TRUE ||
          compile6$`6NM_DistStart`[k]< 5.5) &&
         isempty(compile6$`5.5NM_TimeStart`[k])==FALSE,
        compile6$FalseAlarm_5.5NM[k] <- 1,
        compile6$FalseAlarm_5.5NM[k] <- 0)

#Check if False Alarm 5NM
ifelse((isempty(compile6$`6NM_TimeStart`[k])==TRUE ||
          compile6$`6NM_DistStart`[k]< 5) &&
         isempty(compile6$`5NM_TimeStart`[k])==FALSE,
       compile6$FalseAlarm_5NM[k] <- 1,
       compile6$FalseAlarm_5NM[k] <- 0)

#Check if False Alarm 4.5NM
ifelse((isempty(compile6$`6NM_TimeStart`[k])==TRUE ||
          compile6$`6NM_DistStart`[k]< 4.5) &&
         isempty(compile6$`4.5NM_TimeStart`[k])==FALSE,
       compile6$FalseAlarm_4.5NM[k] <- 1,
       compile6$FalseAlarm_4.5NM[k] <- 0)

#Check if Failure 5.5NM
compile6$Failure_5.5NM[k] <- 0
ifelse(((isempty(compile6$`6NM_TimeStart`[k])==FALSE ||
          compile6$`6NM_DistStart`[k]>5.5) &&
          isempty(compile6$`5.5NM_TimeStart`[k])==TRUE &&
```

```
          isempty(compile6$`1.5NM_TimeStart`[k])==FALSE) ||
         (isempty(compile6$`6NM_TimeStart`[k])==FALSE &&
          isempty(compile6$`5.5NM_TimeStart`[k])==FALSE &&
          isempty(compile6$`1.5NM_TimeStart`[k])==FALSE &&
          compile6$`1.5NM_TimeStart`[k]<compile6$`5.5NM_TimeStart`[k]),
       compile6$Failure_5.5NM <- 1,
       compile6$Failure_5.5NM[k] <- 0)


#Check if Failure 5NM
compile6$Failure_5NM[k] <- 0
ifelse(((isempty(compile6$`6NM_TimeStart`[k])==FALSE ||
          compile6$`6NM_DistStart`[k]>5) &&
          isempty(compile6$`5NM_TimeStart`[k])==TRUE &&
          isempty(compile6$`1.5NM_TimeStart`[k])==FALSE) ||
         (isempty(compile6$`6NM_TimeStart`[k])==FALSE &&
          isempty(compile6$`5NM_TimeStart`[k])==FALSE &&
          isempty(compile6$`1.5NM_TimeStart`[k])==FALSE &&
          compile6$`1.5NM_TimeStart`[k]<compile6$`5NM_TimeStart`[k]),
       compile6$Failure_5NM <- 1,
       compile6$Failure_5NM[k] <- 0)

#Check if Failure 4.5NM
compile6$Failure_4.5NM[k] <- 0
ifelse(((isempty(compile6$`6NM_TimeStart`[k])==FALSE ||
          compile6$`6NM_DistStart`[k]>4.5) &&
          isempty(compile6$`4.5NM_TimeStart`[k])==TRUE &&
          isempty(compile6$`1.5NM_TimeStart`[k])==FALSE) ||
         (isempty(compile6$`6NM_TimeStart`[k])==FALSE &&
          isempty(compile6$`4.5NM_TimeStart`[k])==FALSE &&
          isempty(compile6$`1.5NM_TimeStart`[k])==FALSE &&
          compile6$`1.5NM_TimeStart`[k]<compile6$`4.5NM_TimeStart`[k]),
       compile6$Failure_4.5NM <- 1,
       compile6$Failure_4.5NM[k] <- 0)


#Check duration from 6NM Warning to 1.5NM
compile6$Time_6NM_toFacility[k] <- NA
ifelse(isempty(compile6$`6NM_TimeStart`[k])==FALSE &&
         compile6$`6NM_DistStart`[k]> 5.5 &&
         isempty(compile5$`1.5NM_TimeStart`[k])==FALSE,
         compile6$Time_6NM_toFacility[k] <-
         abs(difftime(compile6$`6NM_TimeStart`[k],
                      compile6$`1.5NM_TimeStart`[k],
                      units = "mins")),
       compile6$Time_6NM_toFacility[k] <- NA)


#check duration from 5.5NM Warning to 1.5NM
compile6$Time_5.5NM_toFacility[k] <- NA
ifelse(isempty(compile6$`5.5NM_TimeStart`[k])==FALSE &&
         compile6$`5.5NM_DistStart`[k]> 5 &&
```

```
        isempty(compile6$`1.5NM_TimeStart`[k])==FALSE,
        compile6$Time_5.5NM_toFacility[k] <-
        abs(difftime(compile6$`5.5NM_TimeStart`[k],
                     compile6$`1.5NM_TimeStart`[k],
                     units = "mins")),
        compile6$Time_5.5NM_toFacility[k] <- NA)

#check duration from 5NM Warning to 1.5NM
compile6$Time_5NM_toFacility[k] <- NA
ifelse(isempty(compile6$`5NM_TimeStart`[k])==FALSE &&
        compile6$`5NM_DistStart`[k]> 4.5 &&
        isempty(compile6$`1.5NM_TimeStart`[k])==FALSE,
        compile6$Time_5NM_toFacility[k] <-
        abs(difftime(compile6$`5NM_TimeStart`[k],
                     compile6$`1.5NM_TimeStart`[k],
                     units = "mins")),
        compile6$Time_5NM_toFacility[k] <- NA)


#check duration from 4.5NM Warning to 1.5NM
compile6$Time_4.5NM_toFacility[k] <- NA
ifelse(isempty(compile6$`4.5NM_TimeStart`[k])==FALSE &&
        compile6$`4.5NM_DistStart`[k]> 4 &&
        isempty(compile6$`1.5NM_TimeStart`[k])==FALSE,
        compile6$Time_4.5NM_toFacility[k] <-
        abs(difftime(compile6$`1.5NM_TimeStart`[k],
                     compile6$`4.5NM_TimeStart`[k],
                     units = "mins")),
        compile6$Time_4.5NM_toFacility[k] <- NA)
}

toc()
```

Save files

```{r}
#Save dataframe with distances from each storm to each warning radii
write.csv(Storms, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\Results\\
          Jul19Rollover45_WS.csv")

#Save dataframe for 5NM radii results
write.csv(compile5, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\Results\\
          Jul19Rollover45_5NMResults.csv")

#Save dataframe for 6NM radii results
write.csv(compile6, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\Results\\
          Jul19Rollover45_6NMResults.csv")
```

Appendix E. Charts and Graphs Code

```
---
title: "Display Charts and Graphs"
author: "Kimberly Holland"
date: "20 Dec 20"
output: html_notebook
---
Install necessary packages
```{r}
library(MASS)
library(tictoc)
library(lubridate)
library(data.table)
library(Imap)
gc()
rm(list = ls())
```
Read in Data
```{r}
#Clean Workspace
gc()
rm(list = ls())

#Upload data
monthdata <- read.csv("C://Users//queen//Desktop//OPER Thesis
                      //45 OG Data//Coding 2//2017_19.csv")
#Label headers
names(monthdata) <- c("X","Date","Year","Month","Type","Distances km","nmi")

#Print results
monthdata
```
Histogram of Total Lightning
```{r}
#Create and save histogram of TL
png(filename="C:\\Users\\queen\\Desktop\\OPER Thesis\\
    45 OG Data\\Coding 2\\17_19_TL_NM_v2.png")
hist(monthdata$nmi,
main="Total Lightning Distance Distribution: 2017-2019",
xlab="Distances by NM",
ylab="Recoil/Dart Leader Events",
xlim=c(0,10),
breaks=40,
col="deepskyblue3",
)
dev.off()
```
CG Distr Stats
```{r}
#Subset CG data
justCG <- monthdata[which(monthdata$Type=="G"),]
```
```

```
#Save dataframe to get lightning distribution
write.csv(justCG, "C:\\Users\\queen\\Desktop\\OPER Thesis\
          \45 OG Data\\Coding 2\\17_19_TL_NM_justCG.csv")
```

CG Histogram and Distribution
```{r}
#Install more packages
library(fitdistrplus)
library(MASS)
library(bda)
library(fBasics)

#Create and save histogram of CG
hist(monthdata$nmi,
main="Total Lightning Distance Distribution: 2017-2019",
xlab="Distances by NM",
ylab="Recoil/Dart Leader Events",
xlim=c(0,8),
#breaks=40,
col="skyblue",
border = F)
hist(justCG$nmi,
     #main="Total Lightning Distance Distribution: 2017-2019",
     add = T,
     col="red",
     border = F)

sdev <- sd(monthdata$nmi)
avg <- mean(monthdata$nmi)
shape <- avg/sdev
shape
scale <- avg/(gamma(1+(1/shape)))
scale

descdist(monthdata$nmi, discrete = FALSE)
skewness(monthdata$nmi.wei)
kurtosis(monthdata$nmi.wei)

curve(dweibull(monthdata$nmi, scale=1.5, shape=0.98),
      from = 0,to = 8, main = "Weibull Distribution")

plot(weibull_dist)
```
Histogram of CG and CC Comparison
```{r}
#CG
justCG <- monthdata[which(monthdata$Type=="G"),]
justCG
```

```
#Create and save histogram of CG
png(filename="C:\\Users\\queen\\Desktop\\OPER Thesis\\
    45 OG Data\\Coding 2\\17_19justCG.png")
hist(justCG$nmi,
main="Cloud-to-Ground Lightning Distance Distribution: 2017 - 2019",
xlab="Distances by NM",
ylab="Recoil/Dart Leader Events",
#xlim=c(50,100),
col="deepskyblue3",
)
dev.off()

#CC
justCC <- monthdata[which(monthdata$Type=="C"),]
justCC

#Create and save histogram of CC
png(filename="C:\\Users\\queen\\Desktop\\OPER Thesis\\
    45 OG Data\\Coding 2\\17_19justCC.png")
hist(justCC$nmi,
main="Lightning Aloft Distance Distribution: 2017 - 2019",
xlab="Distances by NM",
ylab="Recoil/Dart Leader Events",
#xlim=c(50,100),
col="deepskyblue3",
)
dev.off()
```

Seasonal Comparisons

```{r}
#Subset summer season
justSeason <- monthdata[which(monthdata$Month!="1" &
                                monthdata$Month!="2" &
                                monthdata$Month!="3" &
                                monthdata$Month!="4" &
                                monthdata$Month!="10" &
                                monthdata$Month!="11" &
                                monthdata$Month!="12"),]

#Save dataframe to get lightning distribution
write.csv(justSeason, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\17_19_justSeason.csv")

#Subset cold season
justOffSeason <- monthdata[which(monthdata$Month!="5" &
                                   monthdata$Month!="6" &
                                   monthdata$Month!="7" &
                                   monthdata$Month!="8" &
                                   monthdata$Month!="9"),]

#Save dataframe to get lightning distribution
```

```
write.csv(justOffSeason, "C:\\Users\\queen\\Desktop\\OPER Thesis\\
          45 OG Data\\Coding 2\\17_19_justOffSeason.csv")
```

Histogram of Seasons
```{r}
#Summer Season
justSeason <- monthdata[which(monthdata$Month!="1" &
                                monthdata$Month!="2" &
                                monthdata$Month!="3" &
                                monthdata$Month!="4" &
                                monthdata$Month!="10" &
                                monthdata$Month!="11" &
                                monthdata$Month!="12"),]
justSeason

#Create and save histogram of summer
png(filename="C:\\Users\\queen\\Desktop\\OPER Thesis\\
    45 OG Data\\Coding 2\\17_19justSeason.png")
hist(justSeason$nmi,
main="Total Lightning Distance Distribution: 2017-2019 Summer Season (May-Sep)",
xlab="Distances by NM",
ylab="Recoil/Dart Leader Events",
#xlim=c(50,100),
col="deepskyblue3",
)
dev.off()

#Lightning Off-Season
justOffseason <- monthdata[which(monthdata$Month!="5" &
                                   monthdata$Month!="6" &
                                   monthdata$Month!="7" &
                                   monthdata$Month!="8" &
                                   monthdata$Month!="9"),]
justOffseason

#Create and save histogram of cold
png(filename="C:\\Users\\queen\\Desktop\\OPER Thesis\\
    45 OG Data\\Coding 2\\2017justOffseason.png")
hist(justOffseason$nmi,
main="Total Lightning Distance Distribution: 2017 - 2019 Cold Season (Oct-Apr): ",
xlab="Distances by NM",
ylab="Recoil/Dart Leader Events",
#xlim=c(50,100),
col="deepskyblue3",
)
dev.off()
```
Histogram by Month
```{r}
#Choose a month
month <- monthdata[which(monthdata$Month=="9"),]
```

```
month

#Create and save histogram of specific month
png(filename="C:\\Users\\queen\\Desktop\\OPER Thesis\\
    45 OG Data\\Coding 2\\2017_09.png")
hist(month$nmi,
main="Total Lightning Distance Distribution for Sep 2017",
xlab="Distances by NM",
ylab="Recoil/Dart Leader Events",
#xlim=c(50,100),
col="deepskyblue3",
)
dev.off()
```
Make a 3D rep of Storm for Cluster Depiction
```{r}
Storm <- rbind(flashGrp1a,flashGrp2a,flashGrp3a)
Storm <- Storm[,c(3,4,6,8)]
Storm

library("scatterplot3d")

colors = c("#999999", "#E69F00", "#56B4E9")
colors <- colors[as.numeric(Storm$DistTime)]
s3d <- scatterplot3d(Storm[,1:3],
                     main = "Clustering Storms by Spatial/Temporal Criteria",
                     pch = 16,
                     color = colors)
```
Observe Storms - For parallel, merge, split
```{r}
observe <- read.csv("C:\\Users\\queen\\Desktop\\OPER Thesis\\
                    45 OG Data\\Coding 2\\Rollovers\\17_09\\
                    Sep17Rollover42.csv",stringsAsFactors = FALSE)
names(observe) = c("X","Date","Time","Latitude","Longitude",
                   "Type","DateTime","Rollover_Cluster","DistTime")

oneStorm <- observe[which(observe$DistTime==2),]
plot(oneStorm$Latitude,oneStorm$Longitude)

plot(observe$Latitude,observe$Longitude)

```

Bibliography

1. Cooray, Vernon. *The Lightning Flash.* The Institution of Engineering and Technology, 2014.

2. Roeder,William P. and Jon M. Saul. The New Mesoscale Eastern Range Lightning Information System. *Sensors*, 10:9, 2016.

3. Sanderson, Dawn L., Edward D. White, Andrew J. Geyer, William P. Roeder, and Alex J. Gutman. Optimizing the Lightning Warning Radii at Spaceport Florida. *Weather and Forecasting*, 35(2):523–536, 2020.

4. McNamara, Todd M. The Horizontal Extent of Cloud-to-Ground Lightning Over the Kennedy Space Center. Master's thesis, Air Force Institute of Technology, Wright Patterson AFB, OH, 1 2002.

5. Johnson, Richard A. and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Pearson, 2007.

6. HQ AFSEC/SEG. *Air Force Occupational Safety, Fire and Health Standards.* Air Force Departmental Publishing Office, 2018.

7. Holle,Ronald L., Raul E. Lopez, and Christoph Zimmerman. Updated Recommendation for Lightning Safety-1998. *National Oceanic and Atmospheric Administration Technical Memorandum ERL NSSL*, 105:1–29, 1999.

8. Parsons, Tamara. Determining the Horizontal Distance Distribution of Cloud-to-Ground Lightning . Master's thesis, Air Force Institute of Technology, Wright Patterson AFB, OH, 3 2000.

9. Hinkley, Jeremy J., Lisa L. Huddleston, and William P. Roeder. Lightning Strike Distance Distribution Beyond a Preexisting Lightning Area. Technical report, NASA Center for AeroSpace Information, 2019.

10. Tippett, Michael K., Chiara Lepore, William J. Koshak, Themis Chronis, and Brian Vant-Hull. Performance of a Simple Proxy for U.S. Cloud-to-Ground Lightning. 10 2018.

11. Baba, Yoshiro and Vladimir A. Rakov. *Electromagnetic Computation Methods for Lightning Surge Protection Studies*. Wiley, 2016.

12. Richard E. Orville. Development of the National Lightning Detection Network. *Bulletin of the American Meteorological Society*, 89(2):180–190, 2008.

13. Roeder,William P. Personal Communication, 2 2021.

14. Merceret, Francis J., John C. Willett, Hugh J. Christian, James E. Dye, E. Phillip Krider, John T. Madura, T. Paul O'Brien, W. David Rust, and Richard L. Walterscheid. A History of the Lightning Launch Commit Criteria and the Lightning Advisory Panel for America's Space Program. *National Aeronautics and Space Administration/Space Program*, 2010.

15. Martin J. Murphy. NASAfacts: Lightning and the Space Program. Technical report, 2005.

16. Cummins, Kenneth L. and Martin J. Murphy. An Overview of Lightning Locating Systems: History, Techniques, and Data Uses, With an In-Depth Look at the U.S. NLDN . *IEEE Transactions on Electromagnetic Compatibility*, 51(3):499–518, 2009.

17. Lopez,Raul E. and Ronald L. Holle. The Distance Between Successive Lightning Flashes. *Bulletin of the American Meteorological Society*, 80:2035–2041, 1999.

18. Ward, Jennifer G., Kenneth L. Cummins, and E. Philip Krider. Comparison of the ksc-er cloud-to-ground lightning surveillance system (cglss) and the u.s. national lightning detection network (nldn). 2008.

19. Mata, Carlos T., Angel. G. Mata, V.A. Rakov, A. Nag, and Jon Saul. Evaluation of the Performance Characteristics of CGLSS II and U.S. NLDN Using Ground-Truth Data from Launch Complex 39B, Kennedy Space Center, Florida. 2012.

20. Renner, Steve L. Analyzing Horizontal Distances Between WSR-88D Thunderstorm Centroids and Cloud-to-Ground Lightning Strikes. Master's thesis, Wright Patterson AFB, OH, 1998.

21. Roeder, William P. The Four Dimensional Lightning Surveillance System. In *21st International Lightning Detection Conference*, pages 19–20, 2010.

22. Boccippio, D.J., S. Heckman, and S.J. Goodman. A Diagnostic Analysis of the Kennedy Space Center LDAR Network: 1. Data characteristics. *Journal of Geophysical Research: Atmospheres*, 106(D5):4769–4786, 2001.

23. Murphy, Martin J, Kenneth L. Cummins, Nicholas W.S. Demetriades, and William P. Roeder. Performance of the New Four-Dimensional Lightning Surveillance System (4DLSS) at the Kennedy Space Center/Cape Canaveral Air Force Station Complex. In *13th Conference on Aviation, Range, and Aerospace Meteorology*, pages 20–24, 2008.

24. Cox, Christopher C. A Comparison of Horizontal Cloud-to-Ground Lightning Flash Distance Using Weather Surveillance Radar and the Distance Between Successive Flashes Method. Master's thesis, Air Force Institute of Technology, Wright Patterson AFB, OH, 3 1999.

25. Krider, E Philip. Spatial Distribution Of Lightning Strikes to Ground During Small Thunderstorms in Florida. 1988.

26. Poehler, Horst A. LDAR Observations of a Developing Thunderstorm Correlated with Field Mill, Ground Strike Location, and Weather Radar Data Including the First Report of the Design and Capabilities of a New, Time-of-Arrival Ground-Strike Location System (GSLS). 1978.

27. James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R.* Springer, 2013.

28. Georgieva, Petia, Lyudmila Mihaylova, and Lakhmi C. Jain. *Advances in Intelligent Signal Processing and Data Mining.* Springer, 2013.

29. Titterington, D. Michael. Algorithms for Computing D-Optimal Design on Finite Design Spaces. In *Proc. of the 1976 Conf. on Information Science and Systems, John Hopkins University*, volume 3, pages 213–216, 1976.

30. Yaming Yu. Strict Monotonicity and Convergence Rate of Titterington's Algorithm for Computing D-Optimal Designs. *Computational Statistics & Data Analysis*, 54(6):1419–1425, 2010.

31. Bain, Lee J. and Max Engelhardt. *Introduction to Probability and Mathematical Statistics.* Brooks/Cole Cengage Learning, 1992.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2021	Master's Thesis	Sept 2019 — Mar 2021

4. TITLE AND SUBTITLE

CCSFS/KSC Total Lightning Warning Radii Optimization for MERLIN Using Preexisting Lightning Areas

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Holland, Kimberly G, Capt

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
WPSFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-21-M-168

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

45 WS/WXT
Bldg 81900, Rm 200
Cape Canaveral SFS, FL 32925-2618
DSN 467-8410, COMM 937-467-8410
Email: william.roeder@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

45WS/WXT

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

The purpose of this research is to optimize lightning warning radii specifications for the 45th Space Wing (45 SW), thus reducing the number of unnecessary warnings that delay ground processing needed for space launch execution at Kennedy Space Center and Cape Canaveral Space Force Station. This thesis sought to answer two key research questions addressing: 1) What radius reduction effectively balances both safety and operations and do reduction recommendations from previous research align with results from the new detection system? 2) What insights can be gained from comparing measurement results for seasonal lightning events as well as lightning types? This study focused on recoil/dart leader events for both lightning types, cloud to ground and lightning aloft, occurring around the Cape Canaveral space launch facilities. Location information for these events are collected by the Mesoscale Eastern Range Lightning Information Network during 2017-2019. In this research, a clustering approach based on spatial and temporal criteria was applied to establish storm groupings. The Minimum Volume Enclosing Ellipsoid method was then utilized to generate the distance distribution for analysis on proposed radii reduction alternatives that would achieve the optimal balance between productivity and risk. Findings will show that a 1NM reduction from the radii baseline provides the optimum balance between operations and safety, resulting in reducing 5NM warning circles to 4NM, and 6NM warning circles to 5NM, comparable to results in previous research.

15. SUBJECT TERMS

lightning, radii, optimization, safety, MERLIN, Cape Canaveral

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 White, AFIT/ENC
					19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	102	(937) 255-3636, ext 4540; edward.white@afit.edu

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18