Synthetic Lightning Generation Employing Autoregressive-Moving-Average (ARMA) Models

THESIS

Seth R. Powers, 1st Lt, USAF

AFIT-ENP-MS-21-M-131

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-MS-21-M-131

SYNTHETIC LIGHTNING GENERATION EMPLOYING AUTOREGRESSIVE-MOVING-AVERAGE (ARMA) MODELS

THESIS

Presented to the Faculty

Department of Engineering Physics

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 Atmospheric Science

Seth R. Powers, B.S.

1st Lt, USAF

March 2021

AFIT-ENP-MS-21-M-131

SYNTHETIC LIGHTNING GENERATION EMPLOYING AUTOREGRESSIVE-MOVING-AVERAGE (ARMA) MODELS

THESIS

Seth R. Powers, B.S.
1st Lt, USAF

Committee Membership:

Maj Peter Saunders, Ph.D.
Chair

Lt Col Robert Tournay, Ph.D.
Member

Abstract

Lightning is a meteorological phenomenon that occurs frequently in numerous locations every single day on Earth. Comprehending the way it materializes spatially and chronologically is imperative to develop a realistic environmental scene. Live lightning data can be fed into a scene, but that process is costly. Therefore, this work explores if lightning data can be generated synthetically using vector autoregressive-moving-average (VARMA) models. Geostationary Lightning Mapper (GLM) data is used as the basis for the study. Lightning climatology is examined and compared to previous research to gain insight into the targeted areas. Individual lightning flashes are analyzed to inspect how well the process works on a smaller scale. Then, entire regions are evaluated to simulate lightning creation in a larger setting. The results of each step suggest that VARMA models are sufficient at lightning generation up to a certain degree. However, the techniques used in this study have the potential to be improved and allow the models to mirror expansive scenes containing many lightning strikes.

Acknowledgements

I would like to express my sincere appreciation to my faculty advisor, Major Peter Saunders, for his patience and guidance throughout this journey. His words of encouragement and direction aided me greatly and I don't think I could have finished this study without his support.

I would like to express gratitude toward Dr. Bryan Steward and Stephen Hinton from the Center for Technical Intelligence Studies Research. Their time and energy spent assisting me with this study was greatly appreciated.

I would also like to thank Anna for her motivation, words of encouragement, and support.

Lastly, I would like to thank my family, friends, classmates, and AFIT professors for their unwavering emotional and technical support in this endeavor.

Seth R. Powers

Table of Contents

Page

Page

List of Figures

List of Tables

List of Abbreviations

2D Two Dimensional

3D Three Dimensional

AIC Akaike Information Criterion

AR Autoregressive

ARMA Autoregressive Moving-Average

ASSET AFIT Sensor and Scene Emulator for Testing

BLM Bulk Lightning Model

CCD Cold Cloud Depth

CONUS Contiguous United States

CTH Cloud Top Height

CTP Cloud Top Pressure

DF Direction Finder

EO/IR Electro-optical and Infrared

FHEE Flash Consisting of High Energy Event

fJ Femtojoule

GLM Geostationary Lightning Mapper

GOES Geostationary Operational Environmental Satellite

GPM Global Precipitation Measurement

IR Infrared

km Kilometer

LD Lightning Density

LIS Lightning Imaging Sensor

LPI Lightning Potential Index

MA Moving-Average

NASA National Aeronautics and Space Administration

NLDN National Lightning Detection Network

NOAA National Oceanic and Atmospheric Administration

TLN Total Lightning Network

TRMM Tropical Rainfall Measurement Mission

US United States

UTC Coordinated Universal Time

VARMA Vector Autogressive Moving-Average

WFOV Wide Field-of-View

I. Introduction

Lightning is one of the most prominent and recognizable meteorological phenomenon on Earth. Lightning is deadly, as from a period of 1940-1991, 8,316 deaths in the United States (US) were attributed to its occurrences, a number that was higher than combined deaths from hurricanes and flooding (NLSI, 1995). The weather feature is also economically destructive, as residential claims in the US from lightning damage surpassed one billion dollars in the year 1990 alone (NLSI, 1995).

Lightning is defined as a "giant spark of electricity in the atmosphere between clouds, the air, or the ground" (NSSL, 2020). Charge builds in the thundercloud as hydrometeors collide with each other in the updrafts and downdrafts found within the cloud. When the separation of charge becomes sufficient, the insulating capacity of the air breaks down and there is a rapid discharge of electricity (NSSL, 2020).

Typically, the thundercloud from which the strike originates creates a negative charge relative to its outside environment (NSSL, 2020). These charges can be negated by opposite charges on the ground, in the air, or within the same cloud. The charge that initially descends from the cloud to either the ground, air, or another cloud is referred to as the stepped leader, and the transfer of charge from the positive source to the negative source is called the return stroke, which is the vivid flash of light that is visually associated with the phenomenon.

Despite the fact that millions of lightning strikes occur each day (Cecil et al., 2014), observing these lightning strikes in depth is a daunting task. Lightning occurs on a time scale that is too small to fully comprehend when observed by the human eye. Cameras exist that have the ability to fully detect the temporal and spatial nature of a flash, but the equipment required to analyze these intricacies is not advanced enough to observe the event over a broad range; thus the physical understanding of lightning within the atmospheric science community is incomplete.

Rakov and others completed studies at Camp Blanding, Florida and Fort Rucker, Alabama utilizing rockets with trailing copper wires to induce artificial lightning strikes (Rakov et al., 1998). An observation made by the team was that the electric leader was able to alter the electric field and induce effects on subsequent strikes. This proved that return strokes are not the only facet of lightning that have the ability to influence other strikes within the vicinity. Marshall and others theorized that lightning initiation in thunderclouds is a result of electron avalanches in which the particles are accelerated by thunderstorm electric fields (Marshall et al., 1995). These charges are hypothesized by Petersen et al. to be generated by either polarized hydrometeors or cosmic rays (Petersen et al., 2008).

The spatial mapping of lightning is another aspect that the atmospheric science community is continually trying to perfect. The most comprehensive database and detection networks are currently ground-based (Zhu et al., 2020), which utilize direction finder (DF) sensors that are able to respond to the waveform produced by the return strokes of lightning. Vaisala operates the US National Lightning Detection Network (NLDN), which offers reliable real-time lightning detection within the continental US (CONUS) (NASA, 2020). This company has been collecting data to the millisecond (ms) since 1995 (NASA, 2020). Earth Networks runs the Total Lightning Network (TLN), which provides lightning tracking around the world (Earth Networks, 2020), utilizing technology similar to the NLDN.

The Geostationary Lightning Mapper (GLM) is a tool that collects lightning data with a "near-infrared optical transient detector that can detect the momentary changes in an optical scene, indicating the presence of lightning" (NOAA, 2017). The tool is found on both the Geostationary Operational Environmental Satellite-16 (GOES-16) and GOES-17. It is the successor of the Lightning Imaging Sensor

(LIS), which operated aboard the Tropical Rainfall Measurement Mission (TRMM) satellite (Goodman et al., 2012). The TRMM mission was designed to understand precipitation patterns in the tropics, thus offering the LIS a field of view (FOV) from 35° S to 35° N. The LIS had a detection rate of 69% near noon and 88% at night (GHRC, 2020). The GLM is the first operational lightning mapper flown in geostationary orbit, allowing constant lightning surveillance throughout a fixed environment, and making it the optimal source of lightning data for this study. The capabilities and specifications of the GLM are highlighted in further detail in Section 2.1.

The forecasting of lightning is an additional aspect that must be understood to fully grasp the phenomenon. Like other features of lightning, it currently is not completely understood by the science community. Different methods exist for how lightning is forecasted. One is to rely solely on cloud micro-physics properties to predict a binary field of lightning. Yair and others (2010) utilized this procedure when they introduced the Lightning Potential Index (LPI). The LPI is defined as the "kinetic energy of the updraft in the developing thundercloud, scaled by the potential for charge separation based on ratios of ice and liquid water within the main 'charging zone' (0° C to -20° C) of the cloud" (Yair et al., 2010). This range of temperatures is associated with the optimal zone for the presence of supercooled water. The LPI consists of variables of the maximum mixing ratios of ice, graupel, snow, and supercooled water, as well as the upward motion within a column.

A second approach to lightning prognosis involves the creation of a lightning parameter that initiates lightning strikes when the "electric field magnitude is above a threshold value" (MacGorman et al., 2001).This helps to both improve the forecasts of individual strikes, and observe how the electric and magnetic fields change within the thundercloud. Similarly, the Weather Research and Forecasting (WRF) Model

also accounts for these explicit lightning physics by analyzing the electric field, electric potential, and discharge physics of lightning (Fierro et al., 2013). However, the WRF lightning model is a bulk lightning model (BLM). The BLM is not able to forecast individual strikes, but instead predicts if an area or swath within the model domain will contain convection.

A third approach to forecasting lightning involves the application of statistical procedures. With this method, lightning is forecast based upon physical, atmospheric parameters that are used as candidate predictors. The area of analysis can be separated into different regions, for example along coastlines (Shafer and Fuelberg, 2006), over which regression techniques or other statistical schemes are applied to predict lightning.

The main objective of this research is to synthetically produce lightning signatures in a user-defined location for the AFIT Sensor and Scene Emulator for Testing (ASSET) model. This objective is accomplished by applying vector autoregressive moving average (VARMA) models to GLM lightning data and assuring that the output of energy values and structure of those values within both individual flashes and areas is congruent with the original GLM data while also not being restricted by the source GLM resolution. Additionally, the climatology of significantly bright event energies is investigated.

This chapter introduced the thesis problem and broadly discussed the processes used to solve it. Chapter 2 provides background information for the types of data and statistical methodologies implemented, as well as a discussion of the ASSET model. Chapter 3 describes the methodology used for this work in further detail. Chapter 4 discusses the results of this study, including the algorithms utilized for the different lightning generation scenarios, and some implementations of those programs, and finally, chapter 5 provides a summary and discussion of the impact of this research,

its limitations, and future research.

II. Background

2.1 Geostationary Lightning Mapper

The GLM operates by detecting the light emitted from lightning at the tops of thunderclouds, during both day and night (NOAA, 2019). The instrument provides a field of view of almost 18,000 kilometers (Goodman et al., 2012) throughout the Western Hemisphere via the GOES-16 and GOES-17. The positions of these two satellites and GLM instruments provide 8 km pixel resolution at nadir and 14 km pixel resolution towards the outer edges of the field of view (Goodman et al., 2012).

The GLM images the landscape at a Near-Infrared (IR) wavelength of 777.4 nanometers (nm). This wavelength is optimal for observing not only lightning strikes that occur during the night, which are displayed clearly against the backdrop of primarily dark clouds, but also during the day, during which the background consists of predominantly brightly lit clouds (Goodman et al., 2012). Even at this optimal wavelength, the lightning strikes are not always clearly visible during daylight hours. Thus, the GLM uses four different filtering techniques to detect diurnal lightning strikes. that are performed sequentially until a lightning signature is located (Goodman et al., 2012).

First, the image is broadly surveyed by the sensor to search for cloud tops that are being illuminated by strikes. Second, that area is enhanced by utilizing narrow band filtering, of which the imager attempts to detect ionized oxygen at the wavelength of 777.4 nanometers, indicating the presence of lightning (Goodman et al., 2012). Third, the GLM filters the signal on a temporal scale to compare it to the backdrop of the clouds. The backdrop of the clouds will be illuminated at a constant rate, while lightning pulses will typically alter the frame for a period of less than 400 ms due to the brief nature of light emitted from lightning strikes (Goodman et al.,

2012). Fourth, if the algorithm still cannot decipher a lightning pulse, it will subtract the background of the signal, frame by frame, in order to decrease the ratio of the backdrop of the lit clouds to the lightning signal (Goodman et al., 2012).

The GLM depicts lightning pulses as events, groups, and flashes. An event occurs when a single pixel meets the criteria to be considered a lightning pulse, and is the basic unit of data for the GLM (Goodman et al., 2012). The sensor is able to detect events on a two ms time scale, so the possibility exists that multiple events can occur within that time frame. A group is a collection of events that occur within adjacent pixels; groups can be composed of just one event, but normally include many events. A flash is defined as "a set of groups sequentially separated in time" (Goodman et al., 2012). Specifically, the groups that are composed of a flash can occur with no more than 330 ms of a hiatus between them, and must be located within 16.5 km from the centers of the groups. The placement of the flash's latitude and longitude are located within the center of every group that the flash is composed of Goodman et al., 2012). Figure 1 shows a snapshot of a typical GLM data scenario. Two separate lightning flashes are found in the figure. Each are broken down into its components of events and groups to demonstrate how the GLM distinguishes flashes, events, and groups.

In addition to the spatial depiction of lightning, the GLM also provides physical characteristics of lightning features. The optical energy of events, groups, and flashes is measured in femtojoules (fJ), as well as the precise time from the first event of a group or flash to the last event of the same features (Rudlosky et al., 2019). Since events are the core data source for the GLM, and are the raw depiction of lightning signatures, they serve as the primary reference to lightning in this study.

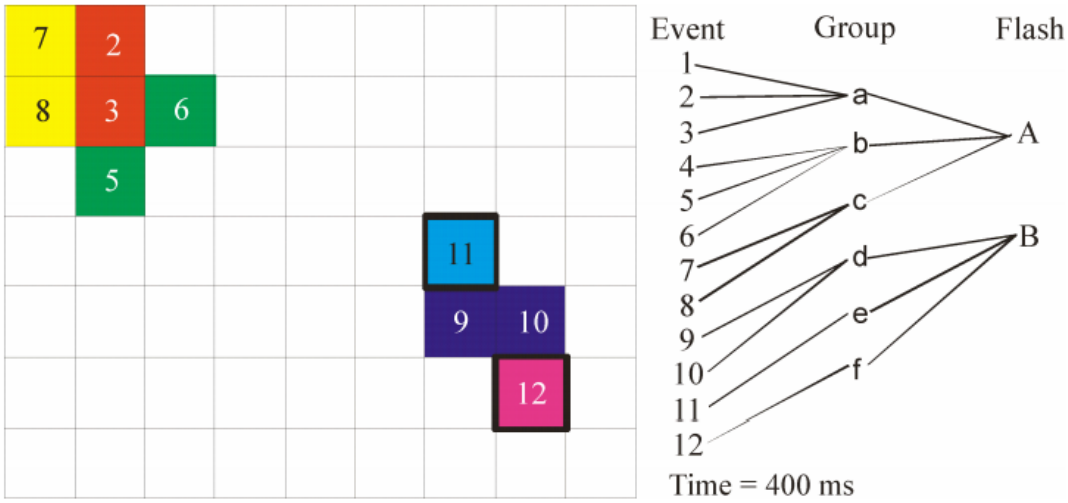


Figure 1. A snapshot of a typical GLM data scenario. Each pixel grid is assumed to be 0.1° wide in latitude and longitude within the figure. Numbered events are colored to match per their parent group. Derived from Goodman and others (Goodman et al., 2012).

2.2 ASSET

ASSET is a physics-based model that has the ability to simulate the real-time depiction of Wide Field-of-View (WFOV) images through the generation of synthetic data sets (Young et al., 2017). The tool emulates the properties, characteristics, and properties of electro-optical sensors. The ASSET Configuration Parameters are presented in Table 1 from Young and others (Young et al., 2017). ASSET allows clouds to be inserted into the scene at the request of the user. The scene for ASSET is a 2D image that can be found on the Earth's surface within the arbitrary WFOV, determined by the user. As such, many of the environmental characteristic parameters are related to the properties of clouds. Clouds present a challenge to sensor collecting efforts, as their motion and changes in shape can lead to errors in the algorithm (Young et al., 2017). These characteristics are necessary to consider when considering the generation of synthetic lightning as clouds provide the necessary charge build-up to produce lightning (NSSL, 2020).

Currently, lightning is not an environmental characteristic that ASSET users

can configure. Simulating lightning on the tool is crucial to provide a realistic Field-of-View (FOV) for many locations on the Earth, as it is a phenomena that occurs in abundance (Cecil et al., 2014). This work serves to provide a substitute for live lightning data, which would be impractical and computationally expensive to obtain and insert into ASSET.

Table 1. Applicable ASSET Configuration Parameters derived from Young et al. (2017). The tool also has other variables to help it emulate scenes, though most would not be pertinent for a potential variable of lightning.

Source Image Properties	Units	Environmental Characteristics	Units
Scene dimensions	(km x km)	Cloudiness	(%)
Pixel dimensions	(m x m)	Cloud velocity	(km/hr)
Radiance range	(W/m^2 x sr)	Cloud warp rate	(m/s)
		Cloud base	(m)
		Cloud depth	(m)
		Cloud albedo	(%)
		Location	(lat,lon)

2.3 Lightning Climatology

2.3.1 Lightning Flash Climatology

This work models lightning on a scale that is much smaller than that of its climatology. However, because of the statistical nature of the study, climatology plays a crucial role in accurately mapping the phenomenon on sub-second timescales. A study by Rudlosky and others summarized GOES-16 GLM data for the first nine months of its existence (Rudlosky et al., 2019). They found that 83% of lightning flashes occur over the land as opposed to 17% over the ocean. During December, January, and February, 86% of lightning flashes were found in the southern hemisphere. This number dropped down to 16% during the months of

June, July, and August, displaying the seasonality of the phenomenon. Rudlosky and colleagues stated that one of the greatest amounts of flash densities during the months of March, April, and May was found in the foothills of the Sierra Madre mountains over northwestern Mexico, exhibiting the effects that terrain can have on lightning (Rudlosky et al., 2019).

Additionally, the team reported that the duration of flashes (time between first and last event) that occur over the ocean is longer on average than that over land. Similarly, flash energy over the ocean exhibits an average flash energy of 420 fJ compared to 230 fJ for those over the land (Rudlosky et al., 2019). Koshak and others studied the same area as Peterson's team from a period of January 11, 2018 to March 11, 2018 (Koshak et al., 2018). They discovered that the average flash had an optical output of 502.1 fJ while the max was 100,004.4 fJ (Figure 2). This contrast in values created a standard deviation of 2947.8 fJ for the data set (Koshak et al., 2018).

Per Figure 2, the GLM can occasionally observe flashes that have large flash energy values. These flashes can be called superbolts. Superbolt is a term coined by Turman (Turman, 1977). He defined superbolts as lightning flashes that are at least 100 times more intense than the average lightning observation (Turman, 1977). Peterson and Lay analyzed superbolts using GLM data and discovered that a superbolt threshold of 100 times more intense optical output resulted in the average energy of all GLM groups to be 1,167 fJ or greater, or a flash percentile of 99.68% (Peterson and Lay, 2020). Table 2 shows other thresholds discovered by Peterson and Lay (2020) and the percentiles associated with each of them.

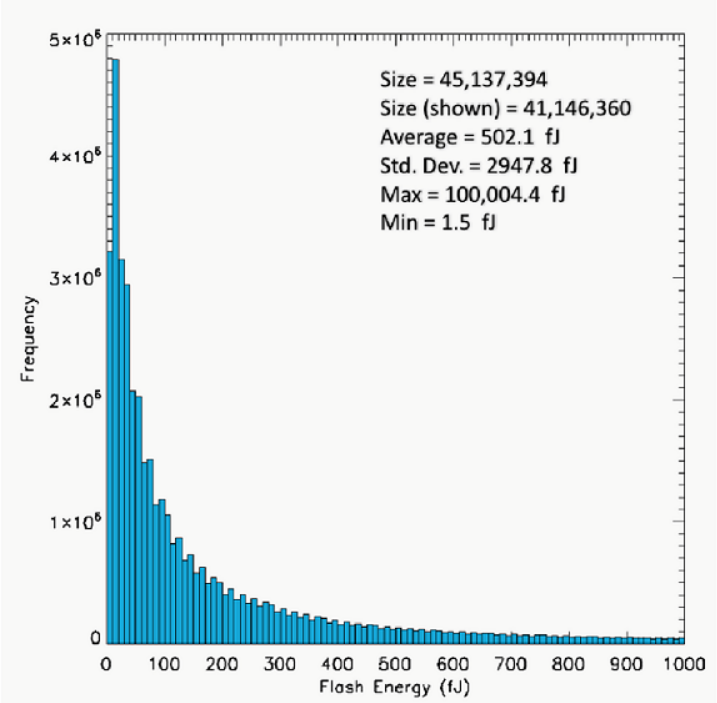


Figure 2. The flash energy distribution in fJ recorded by Koshak et al. (2018) during the 60 day period of January 11, 2018 to March 11, 2018. As this occurred during the boreal winter, a higher concentration of the flashes were found in the southern hemisphere compared to the flashes collected by Peterson et al. (2017).

2.3.2 Lightning Event Climatology

As highlighted in Section 2.1, events are the core of the GLM data structure, and their layout serves as the most realistic portrayal of how lightning is observed from the view of a satellite. The GLM uses the optical energy as a threshold for determining the presence of lightning via filtering techniques, as emphasized in Section 2.1 (Goodman et al., 2012). The sensor uses a measurement of Joules, though the average values are often quite minuscule, and are often on the order of fJ (Rudlosky et al., 2019). The average lightning flash contains 1 million Joules of energy (Jensen, 2015) and the distance from the sensor to the source results in a wide discrepancy between actual and sensed energy.

The GLM also provides the optical energies of groups and flashes. These values are found by taking the sum of all events which compose the flash (Koshak et al.,

Table 2. The threshold for how much brighter a superbolt is than an average flash and the percentile for flash brightness associated with that level. Table derived from Peterson and Lay (2020).

Multiplicity of Intensity	Flash Percentile
100 x	99.68%
250 x	99.98%
500 x	99.9994%
1000 x	99.99998%

2018). Rudlosky et al. (2019) analyzed a nine month period ranging from December 2017 to August 2018 across the entire FOV of the GOES-16 sensor and reported that the average flash, event, and group produced 261 fJ, 16.4 fJ, and 6.2 fJ of energy respectively. Additionally, the team noted that optical energies rendered over water were higher than those over land, with an average event energy of 7.1 fJ over the ocean compared to 6 fJ over land (Rudlosky et al., 2019).

2.4 Lightning Prediction via Cloud Parameterization

As reported in section 2.2, ASSET is a tool that utilizes parameters in infrared through visible wavelengths (Young et al., 2017); thus, lightning forecasts need to align with specifications shown in Table 1. Of the environmental characteristics offered by the ASSET model (Table 1), cloud height is one that was found to be largely correlated to lightning occurrence in a study by Price and Rind (Price and Rind, 1992). The team adapted a formula to predict lightning flashes per minute based on data from Florida, New Mexico, and New England. The data for the study was compiled by Williams (Williams, 1985). The logic behind the comparison was the notion that lightning activity is positively correlated with updraft intensity, which itself is positively correlated with cloud height. Table 3 shows the relationship between flashes per minute and cloud height in km from Williams (1985).

Karagiannidis and others made further observations and improvements related

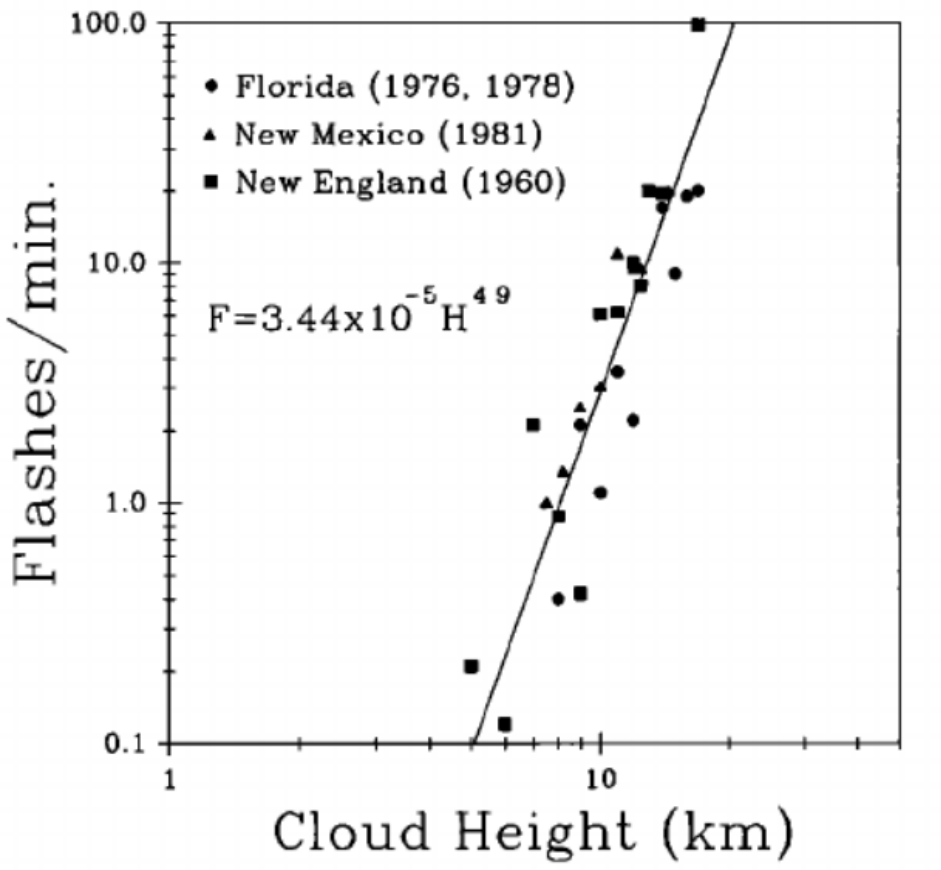


Figure 3. Price and Rind (1992) utilized data from Williams (1985) to ascertain an empirical relationship between cloud height and the number of flashes we observe in those clouds. Though the data from three different locations were utilized, a logarithmic relationship can be observed between the two parameters. Forging a formula between cloud properties and the presence of lightning will be necessary to generate realistic scenes.

to the Price-Rind 1992 parameterization (Karagiannidis et al., 2019). They developed two additional parameters, cloud top pressure (CTP) and cold cloud depth (CCD), and found empirical formulas for both relating to Lightning Density (LD). The CCD is defined as the height of the freezing level subtracted from the cloud top height. Karagiannidis et al. (2019) applied a Generalized Linear Regression methodology to an analysis domain in Central Europe and derived the following equations to relate the two aforementioned variables and an improvement of the cloud top height (CTH) parameter:

$$LD = 1.1 * 10^{-5} * \ CTH^{7.4} \tag{1}$$

$$LD = 1.9 * 10^{13} * \ CTP^{-4.2} \tag{2}$$

$$LD = 2.8 * 10^{-9} * \ CCD^{4.4} \tag{3}$$

The equation relating CTH to lightning density (Equation 1) is likely the best option available to ASSET to assure that generated lightning aligns with the cloud environment. Though it is not a variable offered explicitly in ASSET, it can be formed by the addition of the cloud base and the cloud depth variables (Young et al., 2017).

2.5 VARMA

The GLM has mapped tens of billions of lightning events since its inception in 2017 (Rudlosky et al., 2019). Therefore, providing a realistic statistical depiction of the phenomenon would prove challenging utilizing standard approaches and manual techniques. This study utilizes the statistical technique VARMA to discover an optimal pattern for lightning in different regions, weather events, seasons, and times of day. VARMA is applied to time series, which is a series of data points indexed by a certain timestamp (Box et al., 1994). Time series analysis is the method by which each statistical dependence of each variable is calculated relative to time (Box et al., 1994). No previous research utilizing VARMA models for scene generation is discovered.

VARMA is broken up into three different components: the autoregressive model piece, the moving-average model piece, and a multivariate vector (Wilms et al., 2019). Autoregressive models are stochastic models in which the current value of the process is expressed as a "finite, linear aggregate of previous values of the process" (Box et al., 1994). Autoregressive processes alone can be both stationary or nonstationary (Box et al., 1994). A stationary process is one in which the probability distribution,

mean, and variance do not change as you step through the time steps of the model. Moving average models are similar to nonstationary autoregressive processes, but they allow the previous average and variance of the function to be finite and fluid (Box et al., 1994). The combination of these functions is an autoregressive moving average (ARMA) function, which is univariate by nature. A VARMA process is one in which the parameters are matrices instead of scalars. This is represented statistically as:

$$y_t = \sum_{l=1}^{p} \Phi_l * y_t - l + \sum_{m=1}^{q} \Theta_m * a_t - m + a_t \tag{4}$$

In Equation 4, Φ is given as the autoregressive parameters fitted to all matrices and Θ is a similar parameter, except in its case fitted to the moving average parameters (Wilms et al., 2019). (p) and (q) determine the process of the order for each of their respective processes. Each VARMA model steps through a time series based on an arbitrary amount of iterations. The (p) summation is calculated for each time step by forecasting values for each variable of y_t predicted by previous numbers generated by the model itself. The order establishes the number of lagged values which are accounted for in the vector time series (Wilms et al., 2019). The (q) summation is calculated by finding the error in each of the lagged value terms compared to the real data. The addition of this term assures that the disparities between real and model generated values for each step of the time series are similar (Wilms et al., 2019).

The selection process is resolved by minimizing the Akaike Information Criterion (AIC) (Date, 2019). The AIC determines the quality of the model compared to the data by testing how well it can be fitted to the original (Date, 2019). The formula for the AIC is as follows:

$$AIC = 2k - 2ln(L) \tag{5}$$

In Equation 5, k is the number of parameters and L is the maximum value of the

likelihood function (Date, 2019). The parameter term for this study was the number of GLM events that the selected datasets were composed of. Essentially, lower AIC numbers represent models that are more representative of the original data.

III. Methodology

3.1 Data Selection

3.1.1 GLM

The backbone of this research was the archived data from the GOES-16 GLM. The first reason for why GLM data is used as opposed to other lightning detection sources (NLDN, TLN) is the ease of access. GLM data is readily attainable via National Oceanic and Atmospheric Association (NOAA) and is free to use (NOAA, 2020). The second reason that GLM data is utilized is its accuracy and availability over ocean water. Sensors from the NLDN and TLN are constrained to the land, limiting their effectiveness over open ocean waters compared to the satellite derived data from the GLM (Bui et al., 2015). The third reason that GLM data is employed over the ground-based methods of lightning discovery is that each lightning flash is broken down and classified as groups and events. The events and groups are given unique identification numbers which correspond to their respective flash (Goodman et al., 2012). The TLN is able to detect the multiple pulses that occur within a typical lightning strike, but the output produced is based on the probability of the source detection and centered at that location (Bui et al., 2015). The NLDN provides the available data only as flashes (NASA, 2020). ASSET is a tool that is meant to provide a realistic WFOV of the Earth; therefore, GLM event information is pertinent in representing the spatial depiction of lightning occurrence.

GOES-16 GLM data is used as opposed to either GOES-17 data or a combination of the two due to a longer record of recording strikes. The GOES-16 GLM began collecting data in 2017 while the GOES-17 GLM did not start until 2019 (Goodman et al., 2012). Additionally, there is a lack of terrain variety within the GOES-17's FOV, as it consists predominantly of open ocean. (Goodman et al.,

2012, Figure 4). Events are preferred over flashes and groups, as they depict the best representation of the actual presence of the lightning signatures while groups and flashes are passed through an algorithm and provide the central latitude and 16 longitude relative to the events that compose them (Goodman et al., 2012).

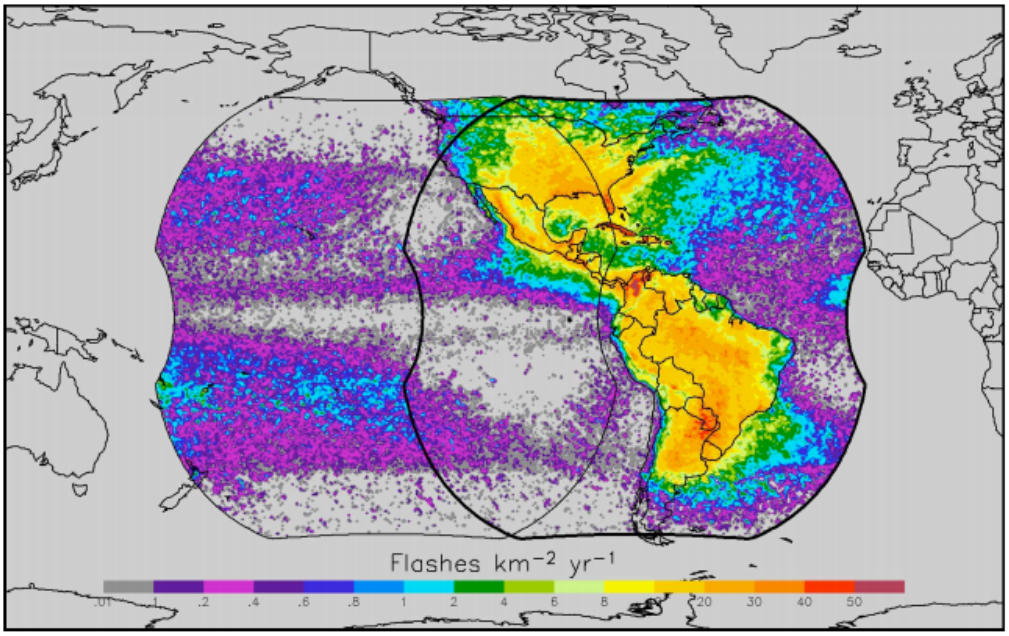


Figure 4. Thin line outlines the FOV for the GOES-17 and thick line outlines the FOV for the GOES-16. Plot shows the density of flashes per year within the entire GOES-R series FOV. Figure courtesy of Goodman et Al. (2012).

The GLM is deemed the best option available for this research; however, there are limitations to the data. Firstly, the GLM is limited by its spatial resolution. This is especially problematic when comparing it to a tool like ASSET, as the pixels of the two dimensional (2D) scene are able to be viewed at a resolution of meters compared to kilometers (Young et al., 2017, Goodman et al., 2012). This limitation is the primary reason that VARMA models are utilized to generate synthetic lightning observations, as the pattern generated through the statistical method is not bound by constraints. The temporal resolution is also a limitation, as the 2 ms integration time is sometimes too long to collect every lightning signature that physically occurred (Goodman et

al., 2012). The integration time for ASSET is on the order of microseconds, and an integration time closer to this value would be ideal, but is currently not available (Young et al., 2017).

Secondly, the GLM provides data files in 20 second intervals. Though this is a file size that allows users flexibility when working with the data, it is sometimes too small of a timeframe for this research, as its goal is to generate synthetic lightning observations up to a timestep in the order of hours. To mitigate this issue, the time is adjusted to arrange the files in chronological order, allowing the timing of the files to be successive. This correction provides the time in seconds from origination.

Thirdly, the event energies of the GLM are presented in Joules. As highlighted in Chapter 2.3.2, this results in minuscule values of event optical energy that are hard to decipher. ASSET is a tool that is designed to emulate scenes focused on electro-optical and infrared (EO/IR) properties; thus, understanding the spatial and temporal layout of the range of optical energies detected by the GLM sensors is important. To remedy this, each event optical energy is converted to fJ, as was accomplished in previous research (Rudlosky et al., 2019, Koshak et al., 2018).

3.2 High Energy Event Climatology

The GLM collects billions of lightning events in a year and most of them are of low optical energies (Rudlosky et al., 2019). Characterizing all of that data would be rigorous and computationally expensive; therefore, the first task of this study was to target locations in the GOES-16 FOV in which lightning observations that would be prominent on ASSET's FOV would be found. As ASSET is an EO/IR model, high energy events are a feature that would stand out on the tool (Young et al., 2017).

No previous research is found on the geographic location of high energy GLM events; however, Rudlosky et al. completed a comprehensive study on GLM flash

climatology (Rudlosky et al., 2019). The team calculated the average flash energy throughout a nine month period (December 2017 to August 2018) and categorized the flash density (regardless of optical energy) into seasons of boreal winter, spring, and fall (Rudlosky et al., 2019). The average flash energy throughout the nine month period is presented in Figure 5 and the seasonal flash densities for the boreal summer and winter are displayed in Figures 6 and 7.

To discover if the locations of high energy flashes discovered by Rudlosky et al. (2019) correspond well with those of high energy events, this work maps out the location of all events that are within the top .32% and top .02%, two gauges that are congruous with the criteria for a superbolt (Turman, 1977) and a threshold defined by Peterson and Lay (2020). This is accomplished for the months of June, July, and August of 2020 for the portion of CONUS found within the FOV of the GOES-16 GLM and the coastal and ocean waters to the east of the land. Multiplicities of intensity for both 100 times an average event (top .32%) and 250 times an average event (top .02%) are plotted. Of the 92 days analyzed, a vast majority of the 20 second files are accounted for; however, some of the data is unable to be obtained; 1,575 files are missing from the month of June, 6 are missing from the month of July, and 924 are missing from the month of August. Additionally, 3 events in each data set are dropped that had event optical energies of 6.5 x 1019 fJ. Though no literature discusses this seemingly arbitrary large value, it is on the order of 1016 fJ times higher than the next closest intensity.

3.3 Scene Selection

3.3.1 Flashes Consisting of High Energy Events

The ultimate goal of this research is to work toward generating synthetic lightning observations within an entire ASSET scene; accordingly, the structure of

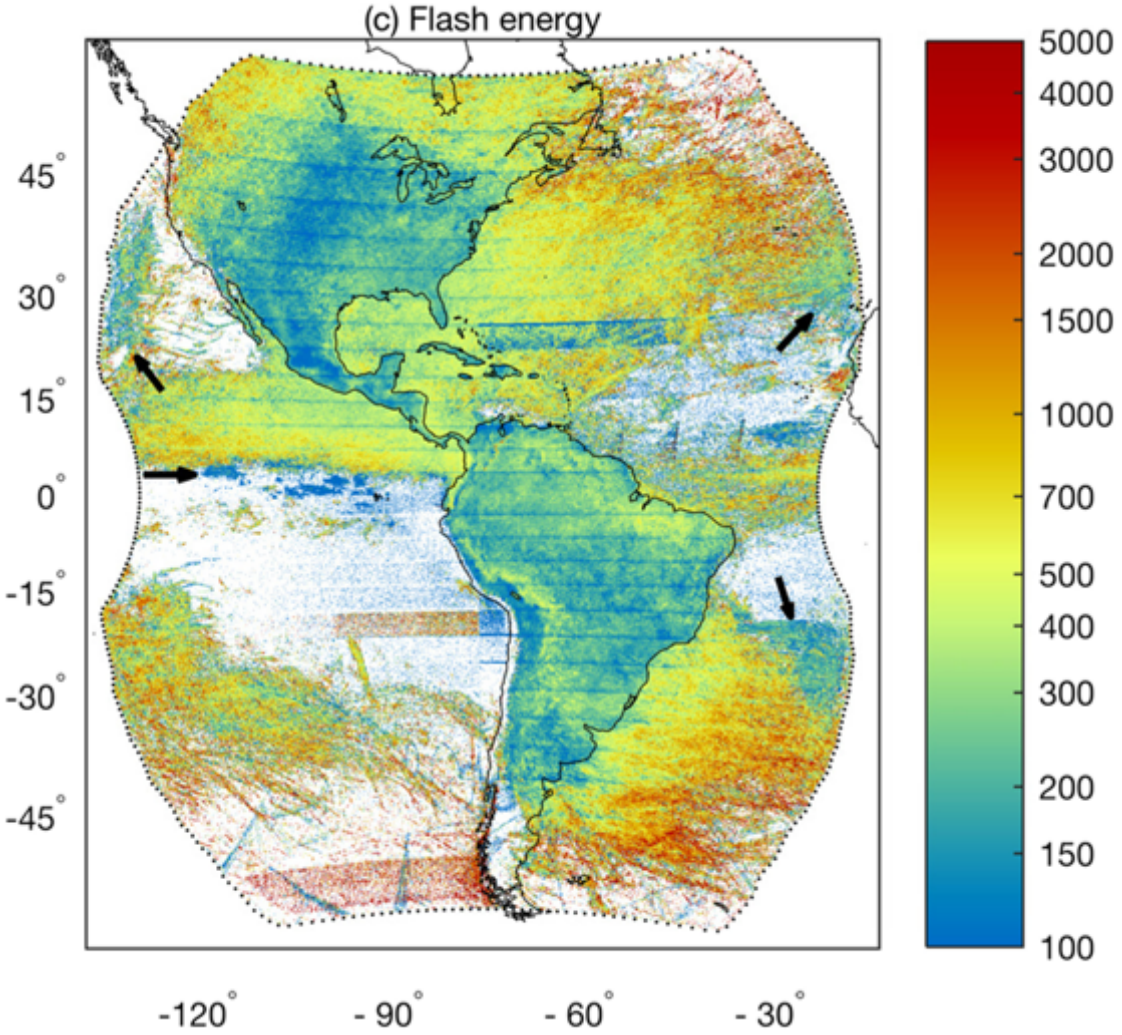


Figure 5. The average flash energy from December 2017-August 2018, the first nine months that the GLM collected data. The optical energy of flashes as a whole were not necessarily the focus of this research, as the events which are the building blocks for these flashes offer a more realistic spatial depiction. Higher flash energies are found over the ocean in the figure, consistent with data from Chapter 2.3.1. Areas of higher terrain such as the Andes show greater energy values compared to all other land areas. Figure courtesy of Rudlosky et al. (2019).

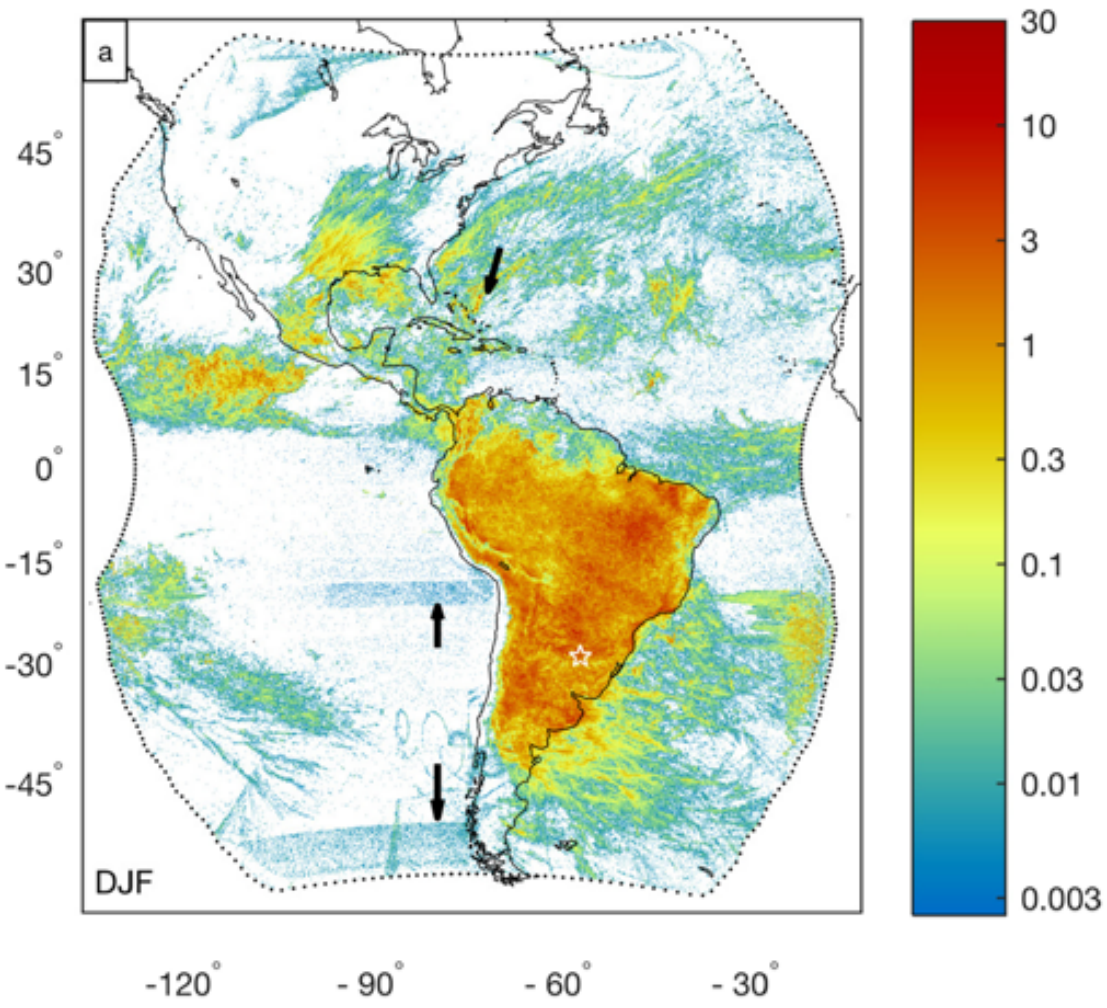


Figure 6. The GOES-16 GLM's flash density from December 2017-February 2018, the first three months that the GLM collected data. The optical energy of flashes as a whole were not necessarily the focus of this research, as the events that compose of these flashes offer a more realistic spatial depiction. Higher flash energies are found over the ocean in the figure, consistent with reportings from Chapter 2.3.1. The black arrows point to discrepancies in the data quality which are not explored in this research. Figure courtesy of Rudlosky et al. (2019).

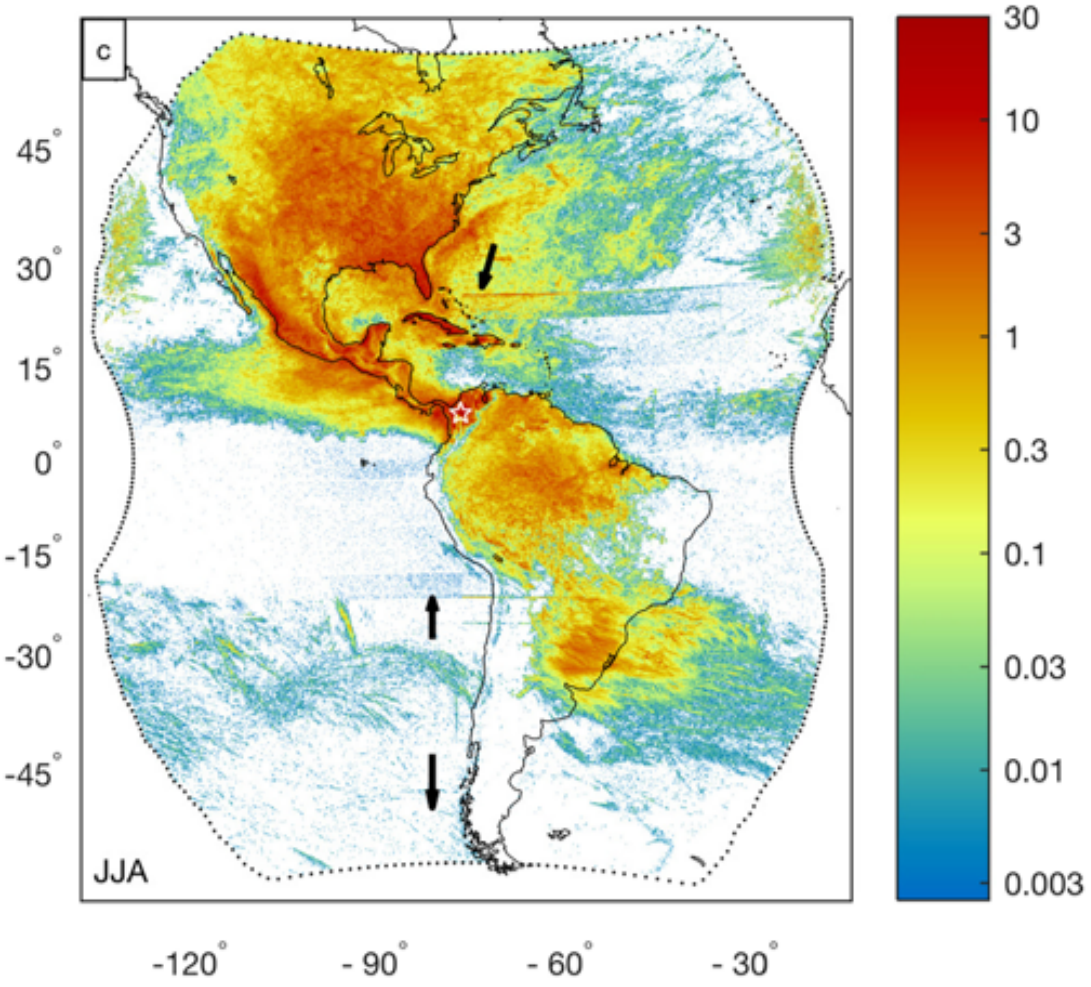


Figure 7. The flash density from June 2018-August 2018 for the GOES-16 GLM. Much higher densities are observed for the CONUS; thus, the summer months were emphasized during the study to effectively collect large amounts of lightning data. The higher densities over the land compared to the ocean and coastal areas are much more prominent as it was in Figure 6. Figure from Rudlosky et al. (2019).

individual flashes generated using VARMA would not represent a natural view of the phenomenon. However, studying the output of a VARMA model applied to a single flash and observing the spatial placement of the generated events within the flash and the numerical optical energies is an efficient way to determine how well VARMA performs. Most flashes that the GLM observes are not complex in nature, and on average are composed of 42.2 events per flash and have an average flash energy of 261 fJ (Rudlosky et al., 2019). Events per flash is the number of events that each lightning observation contains, and the flash energy is the sum of the optical energies of that collection of GLM events (Rudlosky et al., 2019). To analyze flashes that would offer a greater scope of data, this work utilizes flashes that meet two criteria.

Firstly, the flash must contain at least one event that is extremely bright compared to the average GLM event. These flashes are used to gauge how well the statistical model can replicate a diverse array of optical energy values. A multiplicity of intensity of 250 times the average event GLM energy recorded in each scenario is used (see Chapter 3.3.2). Secondly, the flashes analyzed have high events per flash. For those events which meet the brightness criteria, only those that are portions of flashes which have events per flash counts within the top ten percent are collected. These flashes which meet both criteria are individually called a flash consisting of a high energy event (FHEE) for the purpose of this study. One FHEE is analyzed from each scenario (see Chapter 3.3.2) for this research to observe if geographic or seasonal disparities can affect how VARMA performs.

3.3.2 Scenarios

In order to generate synthetic lightning observations for an entire FOV, collections of lightning observations called scenarios are utilized. For the purpose of

this work, a scenario is a database of the latitude, longitude, time, and optical energy of all GLM events that occur within a defined location and timeframe. The research methods allow for any arbitrary location and timeframe to be selected; for the purpose of this research, scenarios are selected over regions and timescales that offer geographical and temporal variety.

Geographical differences are captured by analyzing lightning that occurred over different terrain. Areas of interest that are viewed include mountainous terrain, coastal waters, and coastal land areas. Additionally, areas within different climate zones are considered; some scenarios contain data from tropical locations, while others are found in the mid-latitudes. The dimensions of the geographical areas were also fluid, as different ranges of latitudes and longitudes were studied. Figure 8 contains an outline of the location of the three different scenarios studied.

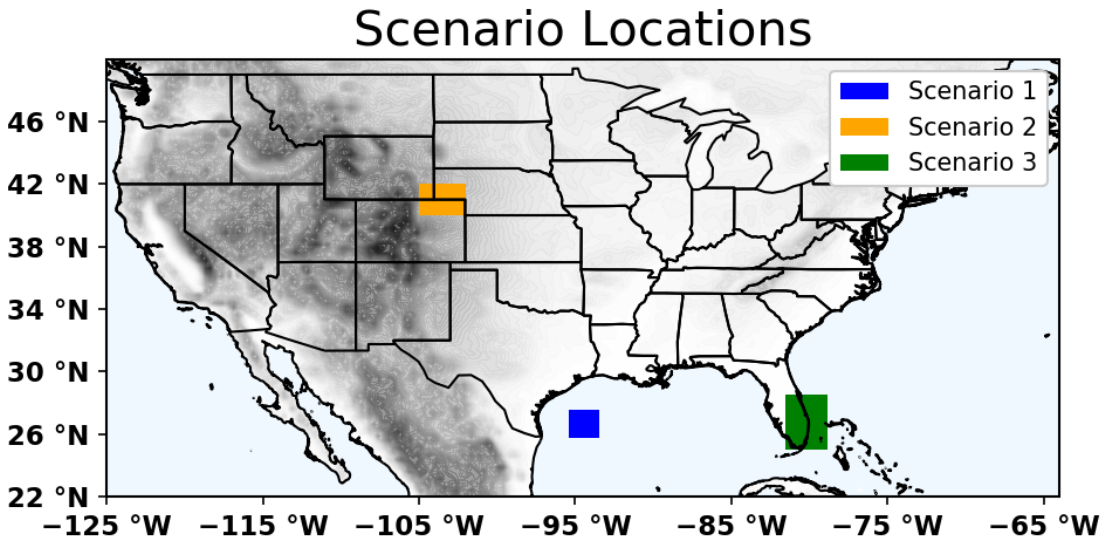


Figure 8. The geographical location of each scenario. The legend pairs up each scenario with a color, and it is presented in the top right corner of the plot. The boundaries are marked by two latitudes and two longitudes in a rectangular shape. Information regarding the features of each scenario can be found in Table 3.

Temporal differences are included within the scenarios in two different ways. Firstly, scenarios are evaluated for both the boreal winter and the boreal summer.

Secondly, scenarios that occur during both night, morning, and afternoon are inspected. The GLM is calibrated to have a higher detection rate at night due to the dark backdrop of the clouds, thus affording the opportunity to explore optical energy discrepancies for the different cases (Goodman et al., 2012). Table 3 below shows the layout detailing the distinctions for all of the scenarios.

Table 3. The location, time, and event count for each scenario. The day each scenario occurs is given for both the Julian and Gregorian calendars. The GLM event count is the number of pixels within the specified area that met the lightning threshold for the timeframe. The locations enclosed by the latitudes and longitudes are plotted in colors of blue, orange, and green in Figure 8, and correspond with the colors of Figure 8.

Scenario 1 - 550,891 GLM Events	
Scenario Location:	Scenario Time:
25.73°N : 27.52°N and -95.43°W : -93.43°W	1200Z - 1400Z 3 Jan 2020 (Day 3)
Scenario 2 - 41,435 GLM Events	
Scenario Location:	Scenario Time:
40°N : 42°N and -105°W : -102°W	0200Z - 0600Z 26 Jun 2020 (Day 178)
Scenario 3 - 85,087 GLM Events	
Scenario Location:	Scenario Time:
25.04°N : 28.55°N and -81.52°W : -79.82°W	1200Z - 1230Z 5 Jul 2020 (Day 187)

3.4 Statistical Application Selection

3.4.1 VARMA

Many methods exist in the realm of time series analysis; however, few are robust enough to handle the tens of thousands of lightning events that can be observed on the order of minutes by the GOES-16 GLM. Therefore VARMA (see Chapter 2.5) is selected as the process used to develop realistic lightning data predicated by the

authentic GLM data. The approach is applied to both the individual FHEEs and the scenarios within the study.

Time is included as both the index and the independent parameter for both FHEEs and scenarios, and the variables that are passed through the model are longitude, latitude, and optical energy of each event. The first index of each output is also dropped, as the model is not able to initialize instantly. Additionally, stationarity is enforced for the autoregressive (AR) portion of the model for both processes. A stationary process (see Chapter 2.5) is applicable for this work as it is necessary to maintain a constant probability distribution throughout. This was especially true for observing lightning patterns for a storm tracking in a certain direction, as not maintaining stationarity would cause the pattern to be abandoned at each iteration of the model. Moreover, invertibility is enforced for the AR term. Invertibility of time series is where the weights of each previous term are restricted and forced to converge (Box et al., 1994). This causes more recent observations to have a higher weight than distant observations within the model.

3.4.2 VARMA for FHEEs

The application of VARMA to individual lightning strikes is studied to observe precisely how well the approach works in generating synthetic lightning observations that mirror the spatial placement and optical energies of GLM events within a single flash. A similar VARMA model is applied to each of the three FHEEs (see Chapter 3.3.1). For the purpose of this study, each FHEE is titled by the scenario database in which it is found, i.e. FHEE 1, FHEE 2, and FHEE 3. When the statistical method is applied, FHEE 1 and 2 utilize an AR and MA of four for each term of the VARMA model, and FHEE 3 utilizes an AR and MA average of five for each term. The mean, standard deviation, range, and maximum event optical energy value are calculated

for each genuine GLM FHEE and synthetic FHEE.

Verification for how well the VARMA model replicates the GLM observations without being restricted by the resolution is evaluated in two ways. Firstly, the generated FHEEs are compared visually to the real FHEE to see how well the latitudes and longitudes of the events mesh. The synthetic FHEEs are evaluated to see if the structure of a flash is maintained (see Chapter 2.1) and the location of the two flashes are in the same general location. The two flashes are also compared chronologically to assure that the general pattern of when and where the high energy events occur within the flashes coincide. Additionally, any abnormal patterns are noted. These include unnatural lightning flash patterns such as lines and breaks within the data. An example of this anomaly is found in Appendix A. Secondly, the numerical values of the mean, range, standard deviation, and maximum optical energy are compared. The AIC and maximum value of the likelihood function are calculated for the synthetic data. The similarities of these values are analyzed to observe how well the VAMRA model replicated the optical energies.

3.4.3 VARMA for Scenarios

In order to discover how well the VARMA technique can replicate an entire FOV, an identical VARMA model is applied to the three aforementioned scenarios (Figure 8, Table 3). The VARMA models for all scenarios use an AR term of one and a MA term of two. The AIC and maximum value of the likelihood function are calculated for the synthetic scenes. Verification of the skill for each model at producing a synthetic lightning scene which emulates that of the original GLM data is performed visually for each scenario. As with the FHEEs (see Chapter 3.3.2), abnormal patterns such as lines and gaps are identified, as they also would appear unnatural compared to the original observations. Furthermore, the event patterns

as a whole for the two scenes are compared to assure that the clusters of lightning flashes mesh.

As identified by Goodman et al. (2012), the GOES-16 GLM has differing resolutions throughout its FOV (see Chapter 2.1). Thus, each GLM event detected could occur at any location within the pixel box. Accordingly, another variable is introduced to each scene: statistical noise. The noise is added onto both the generated latitudes and longitudes of each event. All scenarios are found in a region where the pixel size is roughly eleven kilometers, so the noise values are calculated as a random uniform distribution between -5.5 km and 5.5 km. The visual patterns of the VARMA model-generated data with noise added are then compared to the unadulterated GLM event maps for each scenario. Finally, each scenario is presented chronologically to assure that the added noise does not skew the progression of the synthetic lightning events.

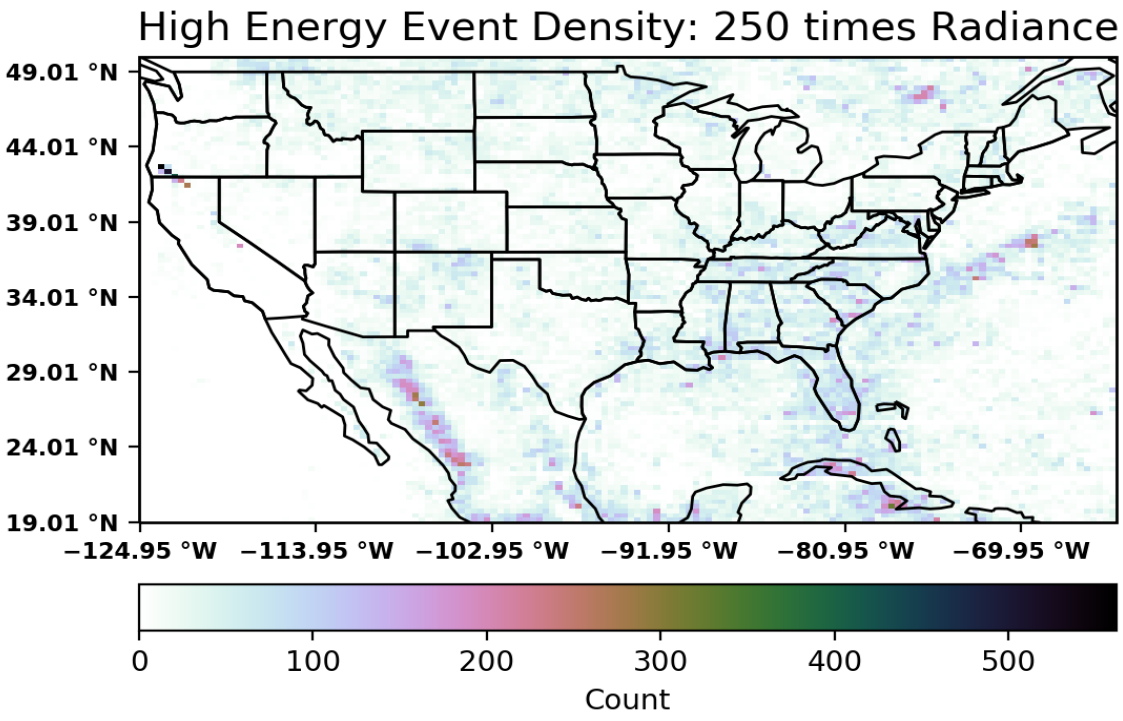
IV. Analysis of Results

4.1 High Energy Event Climatology

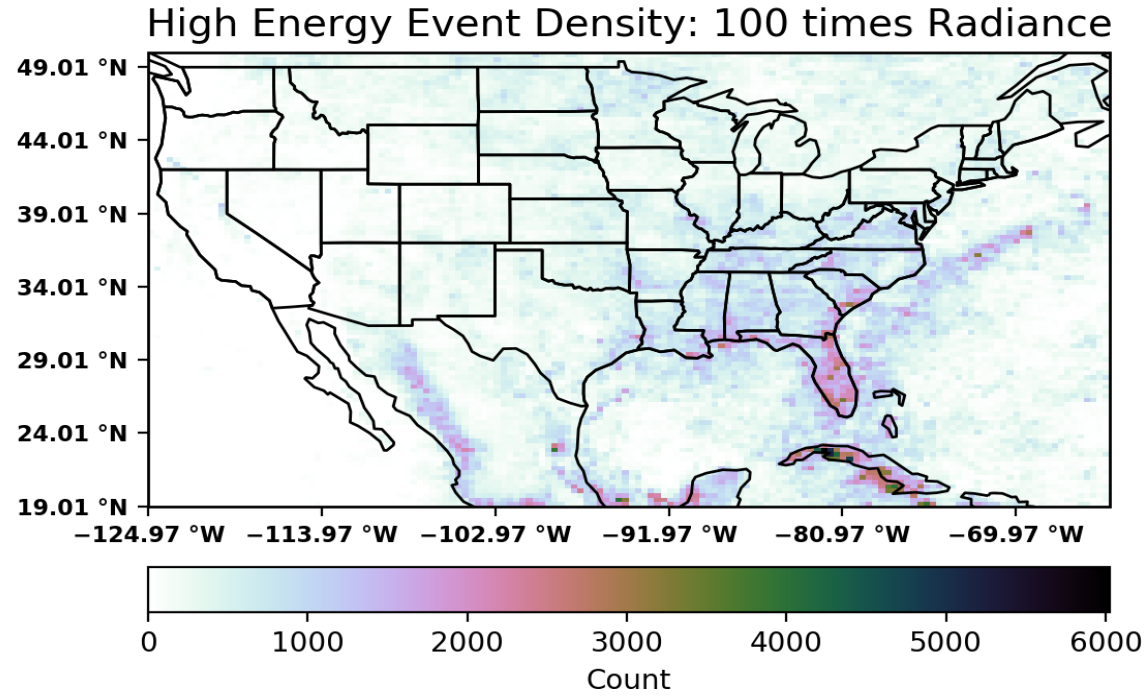
The density of high energy events for two different energy thresholds is compared to the results of both the average flash energy and the summer flash density from Rudlosky et al. (2019). These analyses are performed to reveal if the locations of high energy flashes side with those of high energy events. Previous GLM research has thus far been focused on flashes as a whole and not the events and groups that they are composed of.

Figure 9 displays the densities of high energy GLM events that were either 250 times (9a) or 100 times (9b) brighter than the mean optical energy each day during the months of June, July, and August of 2020. When comparing the subfigures to each other, the layout overall of the higher dimension events correlates well. Increased densities can be observed over the Sierra Madre Occidental in Mexico, the general location of the Gulf Stream, the Southeastern CONUS (particularly the panhandle of Florida), and regions towards the edge of the FOV of the GOES-16. Figure 9b contains a higher count, to be expected due to the lower baseline energy to meet the criteria. The densities found towards the boundaries of the FOV for Figure 9a are higher relative to Figure 9b.

Less similarities are found when comparing Figure 9 to Figure 5. In general, flash energies are lesser over land compared to over the ocean. Lightning energies in mountainous regions such as the Rocky Mountains and the Andes do consist of higher values, but not to the same degree as flash energies over the water. However, the increase in average flash energies towards the northern edge of the FOV noted in Figure 9a and Figure 9b for the event energies are also detected.



(a) Multiplicity of Intensity of 250 times the average energy



(b) Multiplicity of Intensity of 100 times the average energy

Figure 9. Subplots of the density of events having optical energies over a certain brightness threshold from June 1 to August 31, 2020. Figure 9a contains energies that were over 250 times brighter (99.98 percentile) than all other GLM events on each day, and Figure 9b contains event energies that were over 100 times brighter (99.68 percentile) than all others for each day throughout the month.

4.2 Synthetic FHEE Generation

Individual lightning flashes are generated using VARMA models. VARMA models are applied to FHEEs that occurred at some point within the timeframe of each scenario. Specifically, these were flashes that have at least one event within them that is 250 times or more brighter than the average event brightness and that also is in the top 10% of event count per flash (see Chapter 3.2.1).

FHEE 1 is created from a lightning flash that occurred during Scenario 1. The GLM observation occurred in the Gulf of Mexico at a location of approximately 26.6° N and -93.67° W and had a duration of 1.216 seconds (Figure 10). The flash consists of 335 events and the lightning event observations have an average optical energy of 13.18 fJ. The highest event energy within the flash is 398.67 fJ, a value that is within the top .02% of brightest events within the scenario. The range of the optical energies for this flash is 397.34 fJ and the standard deviation is 32.70 fJ.

The VARMA model applied to FHEE 1 uses an AR order of four and a MA order of four. This results in a maximum value of the likelihood function of -707.66 and an AIC of 1571.32. The generated optical energy values for the events have an average energy of 13.23 fJ, a range of 76.79 fJ, and a standard deviation of 8.31 fJ. The synthetic lightning observation verifies well spatially with the original lightning observation (Figure 11). The location of the generated flash is within the vicinity criteria of the actual flash, and no abnormal patterns are observed amongst the events. The fitting of each synthetic variable from this VARMA model is compared to the values of the GLM flash in Figure 12. The chronological evaluation of the synthetic lightning observation also appears visually realistic, as the progressions resemble that of the actual flash. Table 4 catalogues the numerical comparisons of the two flashes.

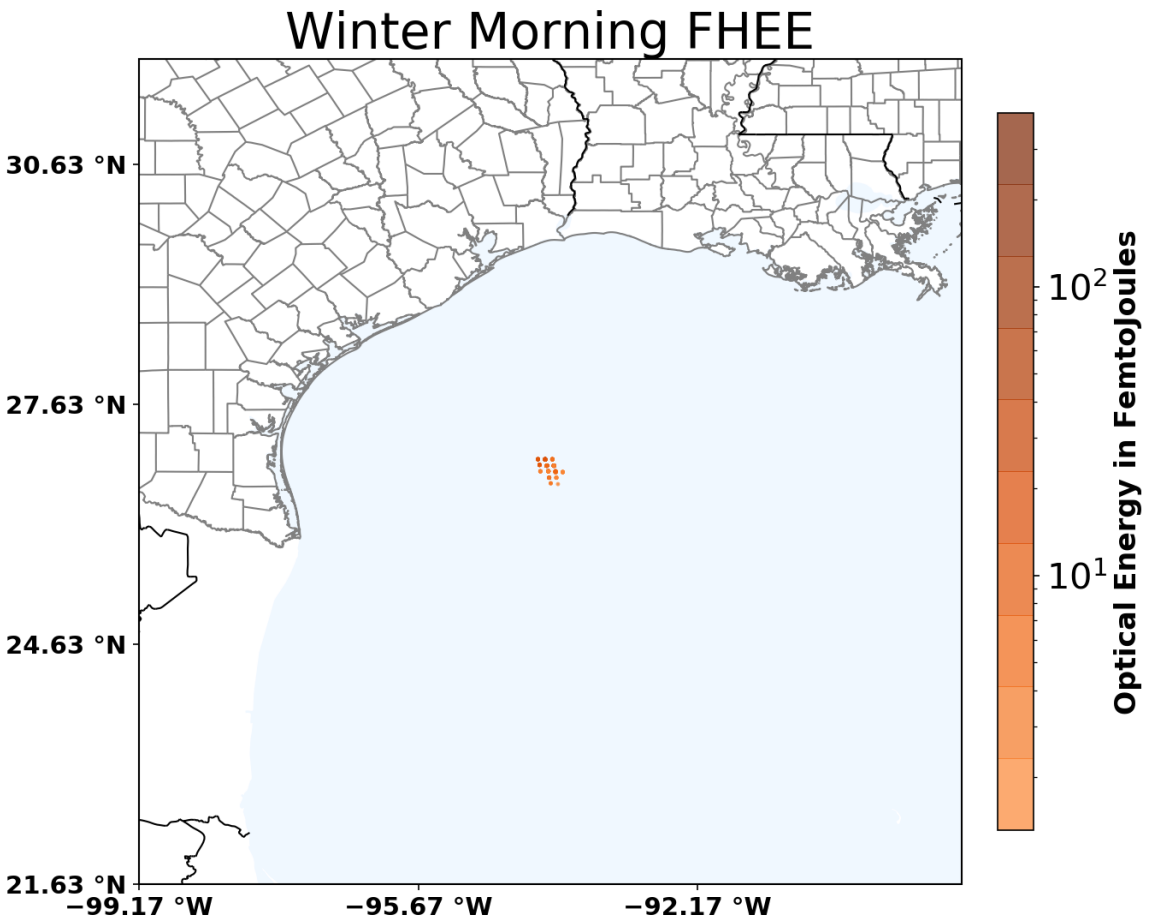
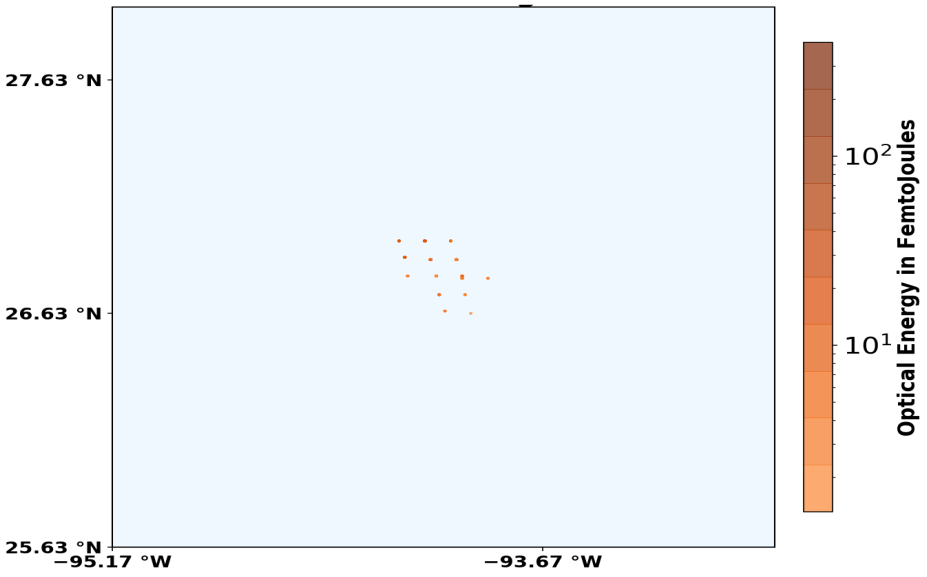


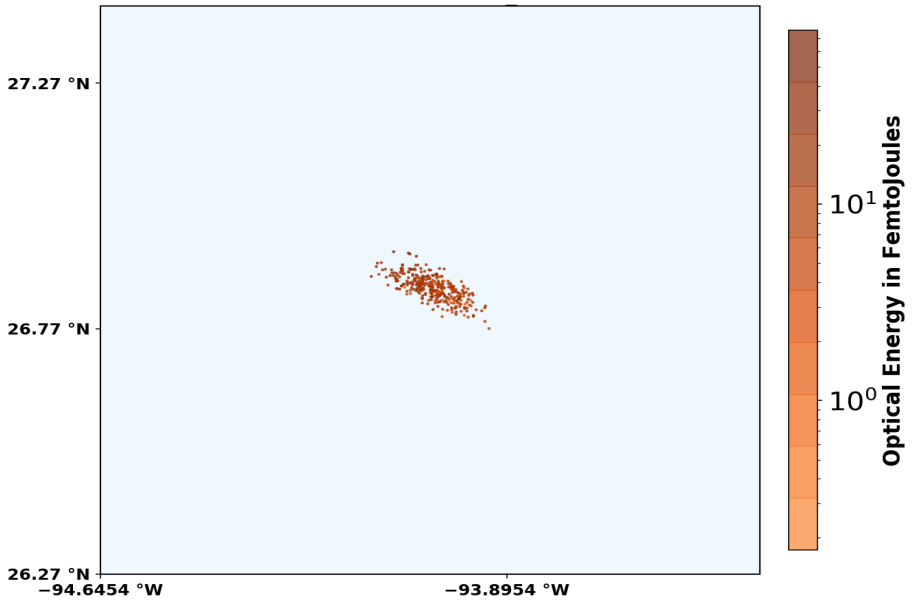
Figure 10. This figure shows the location of FHEE 4, which occurred at some point during Scenario 1. The figure is focused on the western region of the Gulf of Mexico. The colorbar is scaled logarithmically to show the contrast between the events with high and low optical energy. All events are adjacent to each other, consistent with the flash definition from Goodman et al. (2012).

FHEE 2 is formed from a lightning flash observation that occurred during Scenario 2. The GLM observation took place in Western Nebraska at a location of approximately 41.5° N and -102° W and had a time length of 2.149 seconds (Figure 13). The flash consists of 351 events. Those events have an average optical energy of 9.46 fJ. The highest event energy within the flash is 209.92 fJ, a value that is in the 99.98 percentile for brightest events within the scenario. The range of the optical energies is 204.19 fJ and the standard deviation is 17.86 fJ.

The VARMA model implemented for FHEE 2 applies an AR order of four and a

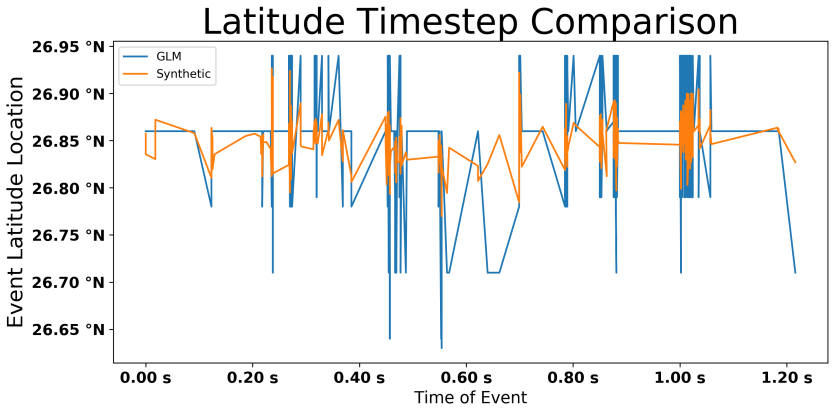


(a) The same plot as in Figure 10 centered on the FHEE to better display its structure. The limited resolution of the GLM is apparent in this plot.

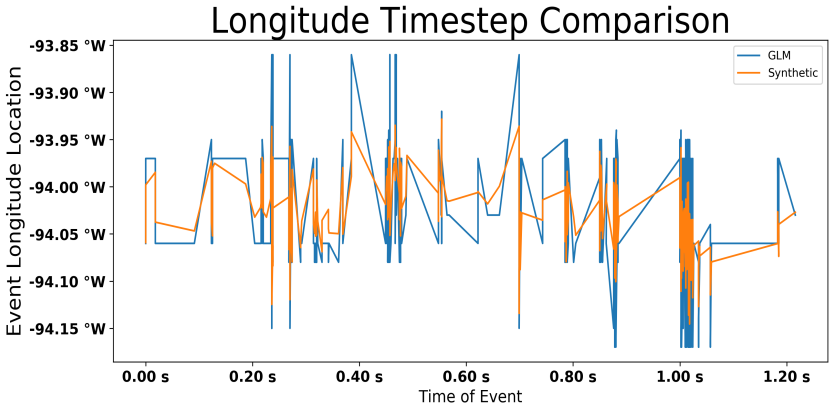


(b) The synthetic flash generated using a tailored VARMA model for FHEE1. The specifications of the model are found in Table 4. The location of the flash in Figure 11a is mimicked, and the limited resolution is no longer observed.

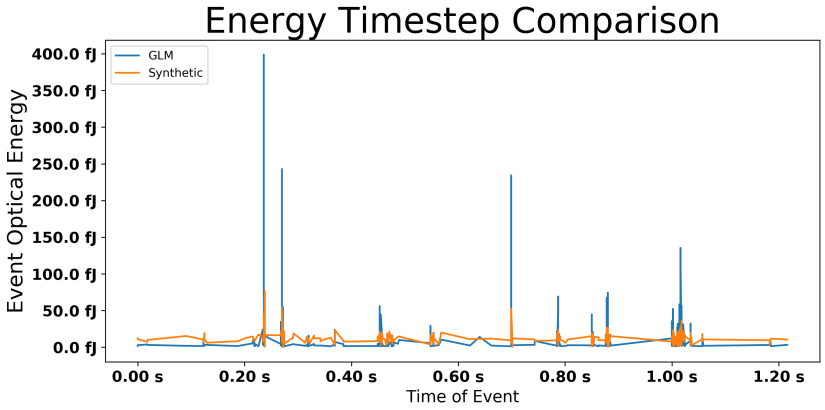
Figure 11. The visual results of the VARMA model application for FHEE 1. All latitudes, longitudes, and events observed are generated using the model.



(a) This subplot displays how the VARMA model fits data to the latitude values for this particular flash. Isolating the latitude variable does not offer any absolute clarity on VARMA model performance, as its location is dependent on the location of each corresponding event longitude.



(b) The relationship between synthetic and GLM longitudes for FHEE 1. As was the case with the latitude variable, they are dependent on each other and the 2D combination of the variables offers more insight into how the VARMA model performed.



(c) The comparison of event energies for FHEE 1. Though the VARMA model is able to detect that certain events had increased brightness, the magnitude of those escalations was not replicated well by the VARMA model. This is factor causing the standard deviation and range of the optical energies to be dissimilar to those of the real values.

Figure 12. The individual comparison for all variables generated through the VARMA model (synthetic data) and the real observations of events from FHEE 1 (GLM data).

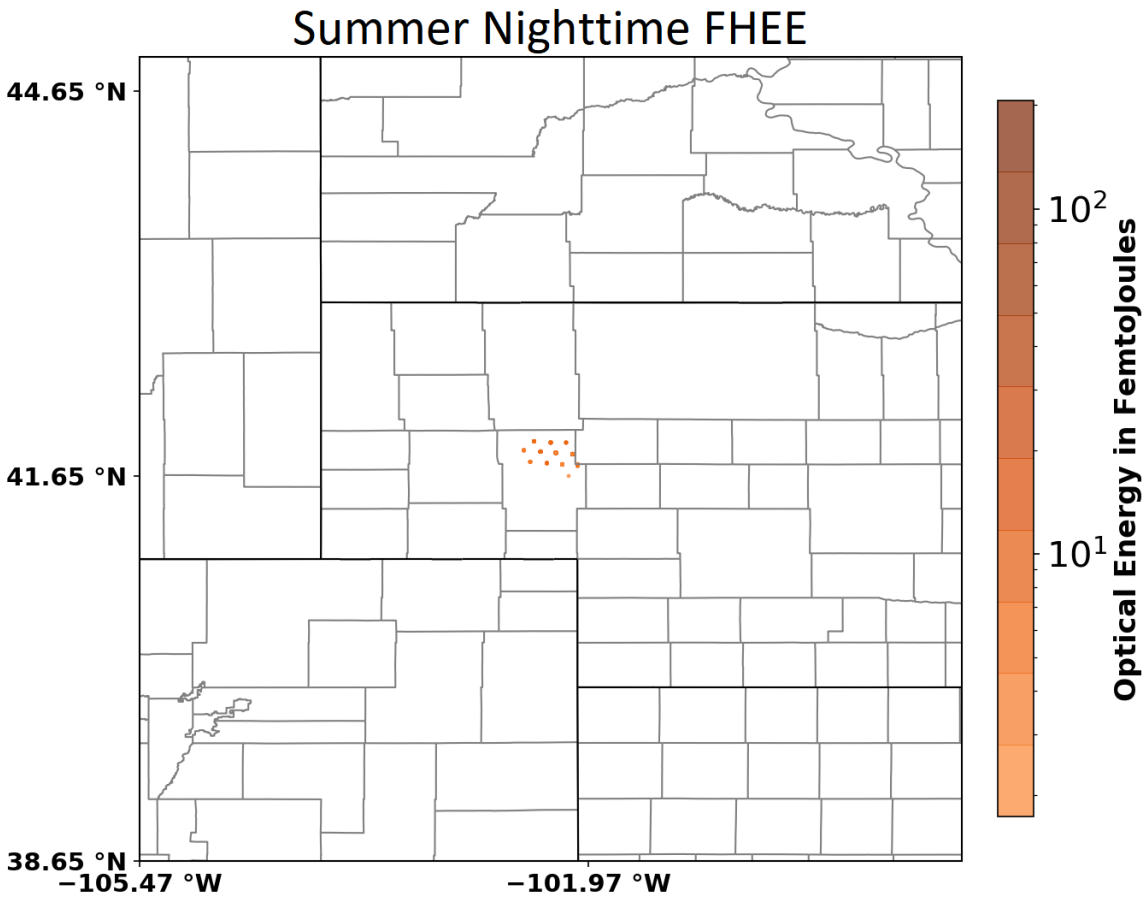


Figure 13. Presented is the location of FHEE 2, which occurred at some point during Scenario 2. The colorbar is once again scaled logarithmically. Counties are displayed with gray outlines while US states are edged in black. The area displayed is centered in on western Nebraska.

MA order of four, similar to the model structure for the lightning flash analyzed from Scenario 1. This application results in a maximum value of the likelihood function of -1895.37 and an AIC of 3946.75. The optical energy values that the VARMA model created have an average energy of 10.31 fJ, a range of 96.23 fJ, and a standard deviation of 9.36 fJ. A majority of the synthetic lightning events are within the area of the genuine flash; however, some of the events stray many degrees of latitude and/or longitude away from the aforementioned cluster (Figure 14b). No abnormal patterns are detected in the main cluster of the synthetic events, though the outlying events

themselves cause the geographical properties of the events to not appear realistic compared to the original lightning observation. The fitting of each synthetic variable from this VARMA model is compared to the values of the GLM flash in Figure 15. The temporal sequence of the generated lightning flash does also not resemble that of the GLM observation for this reason. Table 5 displays the analytical comparisons of the synthetic and actual lightning flash.

FHEE 6 is created from a lightning flash observation that occurred during Scenario 3. The GLM observation occurred near the coastline of southern Florida at a location close to 20° N and -80.6° W. The flash observation had a duration of .478 seconds, and consisted of 164 events (Figure 16). One of the 164 events had an optical brightness of 315.40 fJ, which places it in the top .02% in terms of brightest GLM events within the scenario. The mean of the optical energies within this flash is 20.90 fJ, the range is 313.54 fJ, and the standard deviation is 46.53 fJ.

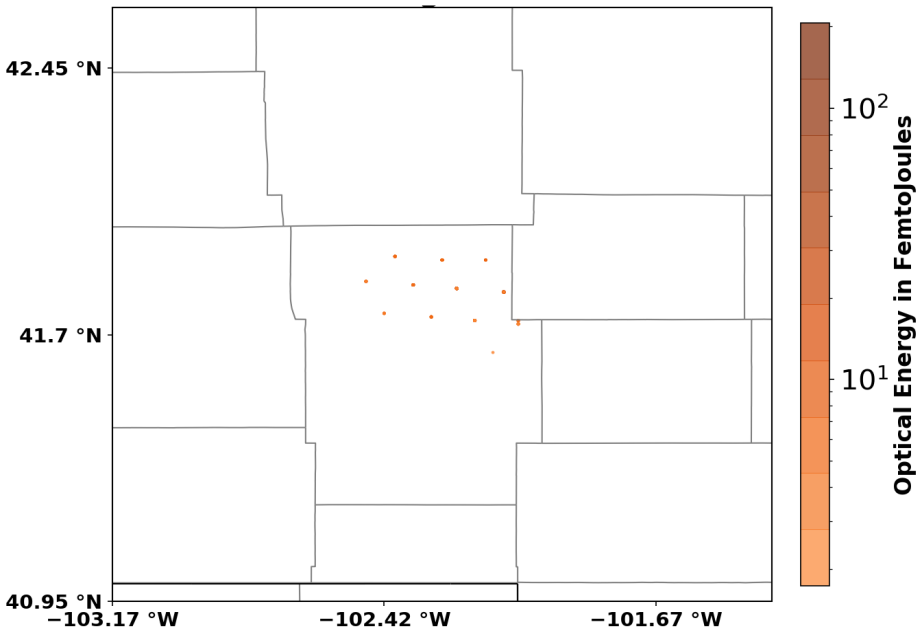
The VARMA model applied to FHEE 6 uses an AR order of five and a MA order of five. This statistical operation results in a maximum value of the likelihood function of -519.55 and an AIC of 1231.09. The generated optical energy values for the events have an average energy of 25.39 fJ, a range of 108.23 fJ, and a standard deviation of 17.28 fJ. The synthetic lightning observation verifies well spatially with the GLM observation, similar to what is observed for the FHEE 4 (Figure 17b). No abnormal patterns are found in the generated data points, and the location of the created flash is close to that of the original observation. The fitting of each synthetic variable from this VARMA model is compared to the values of the GLM flash in Figure 18. The chronological evaluation of the synthetic lightning observation resembles that of the original flash. Table 6 contains the numerical comparisons of the two flashes.

Table 4. The numerical analysis for FHEE 1. *L* is given as the maximum value of the likelihood function, and the flash duration is the time difference between the first and last event of the flash. Time is treated as the independent variable throughout this research, so the flash duration is identical for the physical and synthetic observations.

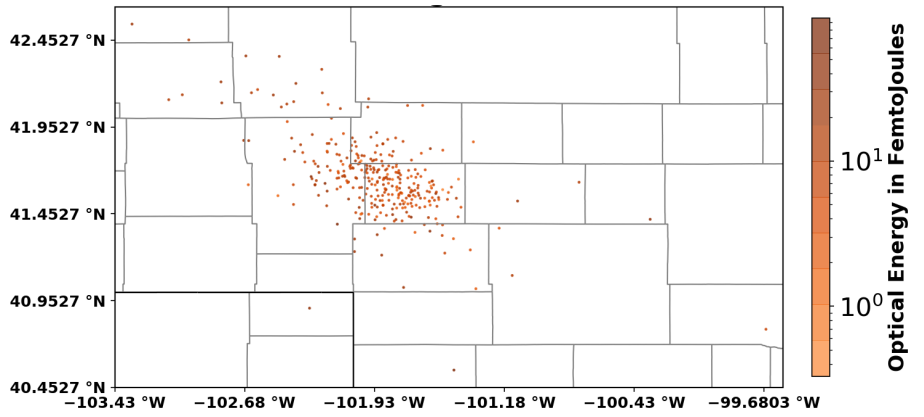
FHEE 1 - 335 Observations		
	Physical Observation	Synthetic Observation
Flash Duration	1.216 seconds	1.216 seconds
Mean	13.18 fJ	13.23 fJ
Standard Deviation	32.70 fJ	8.31 fJ
Range	397.34 fJ	79.76 fJ
Maximum	398.67 fJ	79.59 fJ
(AR,MA) Order		(4,4)
AIC		1571.32
L		-707.66

Table 5. The numerical analysis for FHEE 2. *L* is given as the maximum value of the likelihood function, and the flash duration is the time difference between the first and last event of the flash. This is the longest flash of the three, as it persists for 2.149 seconds.

FHEE 2 - 351 Observations		
	Physical Observation	Synthetic Observation
Flash Duration	2.149 seconds	2.149 seconds
Mean	9.46 fJ	10.31 fJ
Standard Deviation	17.86 fJ	9.36 fJ
Range	204.19 fJ	96.23 fJ
Maximum	205.92 fJ	96.56 fJ
(AR,MA) Order		(4,4)
AIC		3946.75
L		-1895.37

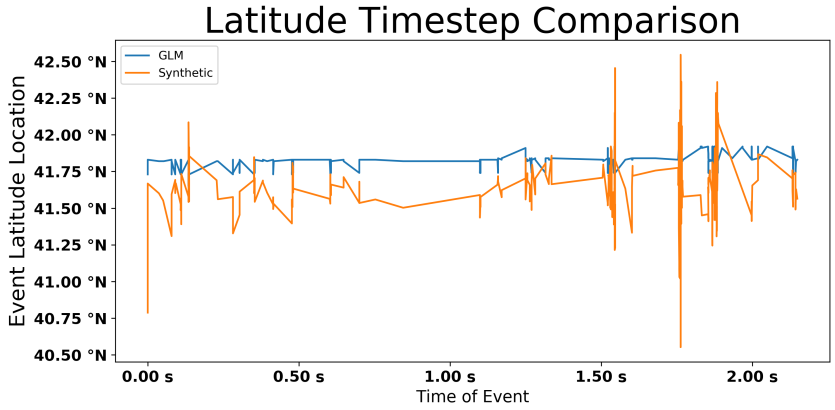


(a) The same flash that is displayed in Figure 13 centered in on the flash to better show its structure All events that occur during the flash are found in Garden County, Nebraska.

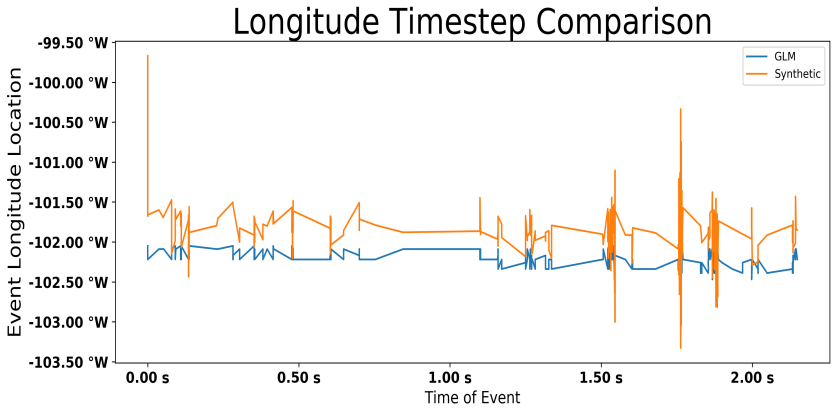


(b) The synthetic flash predicated by FHEE2. The arrangement of the created events no longer is restricted by the approximately 11 km resolution; however, several of the event deviate many kilometers away from Garden County, Nebraska, the location where they are expected to materialize.

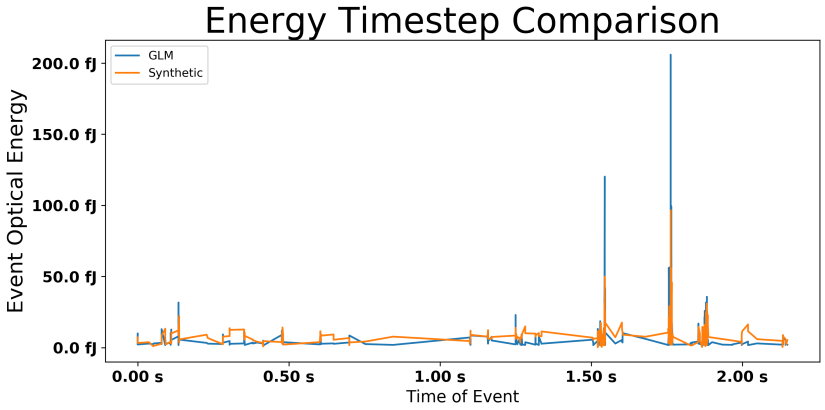
Figure 14. The plot of the transition from GLM data to synthetic data. All latitudes, longitudes, and events examined in Figure 14b are generated using a model.



(a) The comparison of event latitudes for FHEE 2. The isolated view of the chronological progression of this variable displays how this particular VARMA model initialized poorly and was unable to replicate the general latitude, and thus the location, for a majority of events.



(b) The relationship between synthetic and GLM longitudes for FHEE 2. An analogous pattern is observed in this progression as was found in the generated longitudes, as location of the synthetic event longitudes is not in the general vicinity for most events.



(c) The comparison of event energies for FHEE 3. The spikes that occur during the high energy events can also be found in the synthetic data, but not to the extent that was detected in the GLM variables.

Figure 15. The individual comparison for all variables generated through the VARMA model (synthetic data) and the real observations of events from FHEE 2 (GLM data).

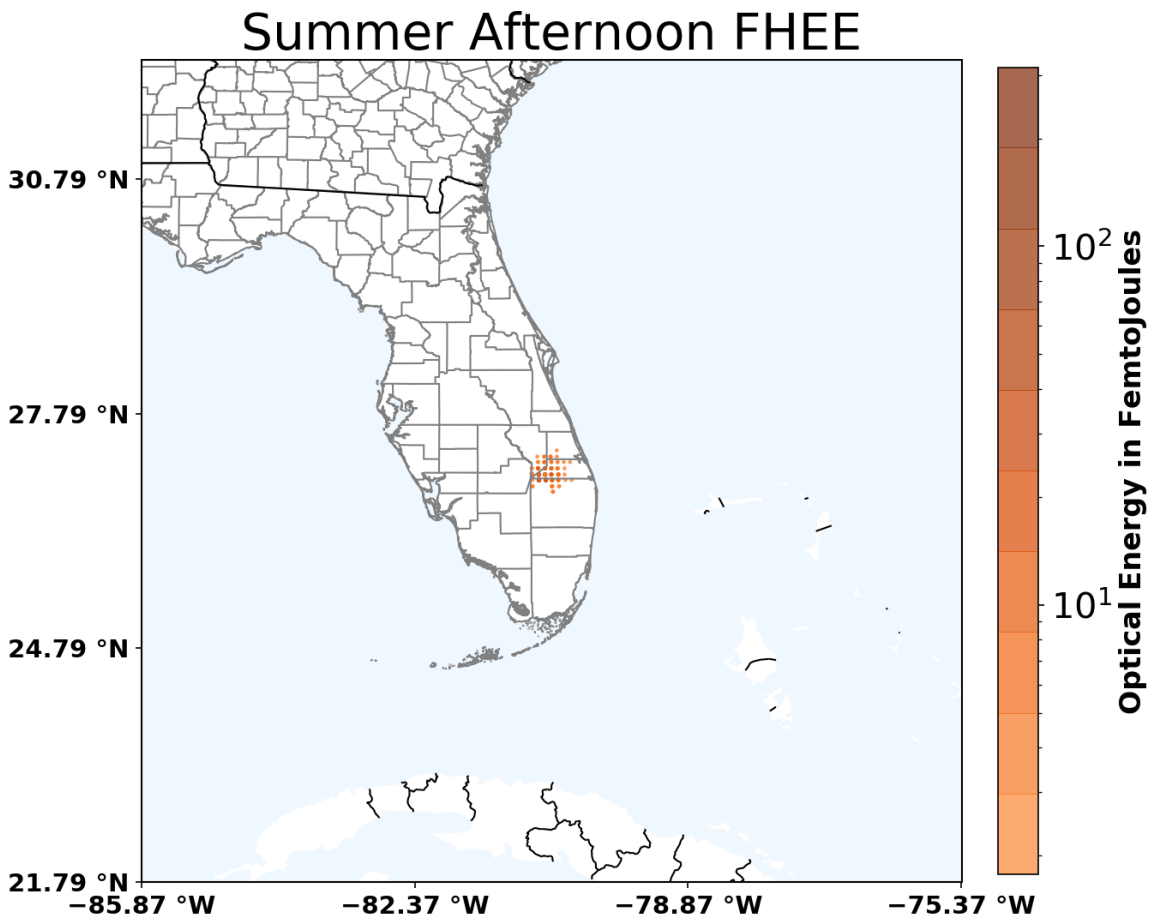
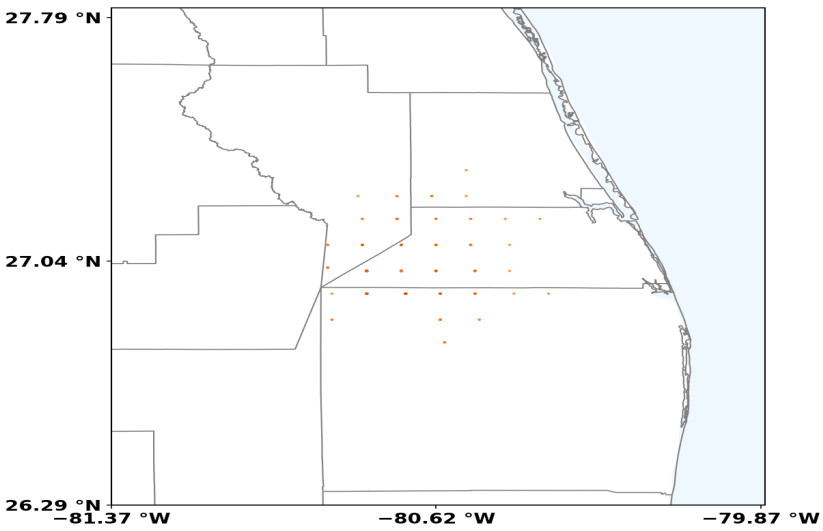


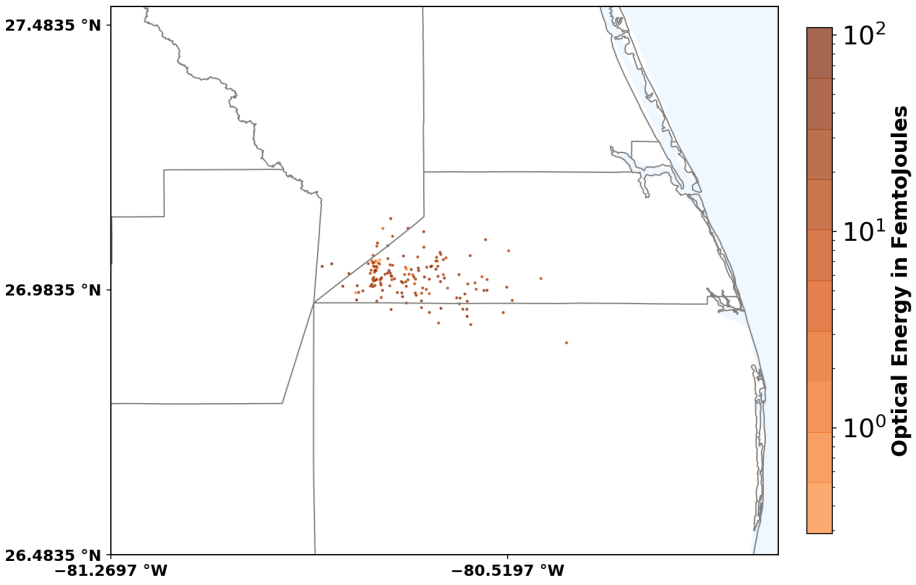
Figure 16. Plotted is the location of FHEE 3, which occurred at some point during Scenario 3. Counties are outlined in black and CONUS states international states are colored in black. The area shown is centered over southern Florida.

FHEE 3 - 164 Observations		
	Physical Observation	Synthetic Observation
Flash Duration	.478 seconds	.478 seconds
Mean	20.90 fJ	25.39 fJ
Standard Deviation	46.53 fJ	17.28 fJ
Range	313.54 fJ	108.23 fJ
Maximum	315.40 fJ	108.51 fJ
(AR,MA) Order		(5,5)
AIC		1231.09
L		-519.55

Table 6. The numerical analysis for FHEE 3. L is given as the maximum value of the likelihood function, and the flash duration is the time difference between the first and last event of the flash. The flash duration is only .478 seconds and there are 164 observations, both the lowest amounts of the three FHEEs.

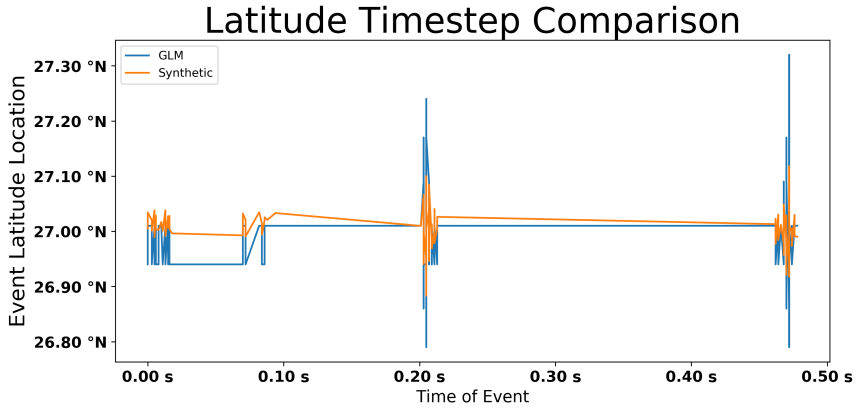


(a) The same data plotted as Figure 16 enhanced to show the flash structure more clearly. Of the three FHEEs, FHEE 3 has the widest flash area, as its events cover a larger area. The events for this flash are found on the border of Palm Beach and Martin County, Florida.

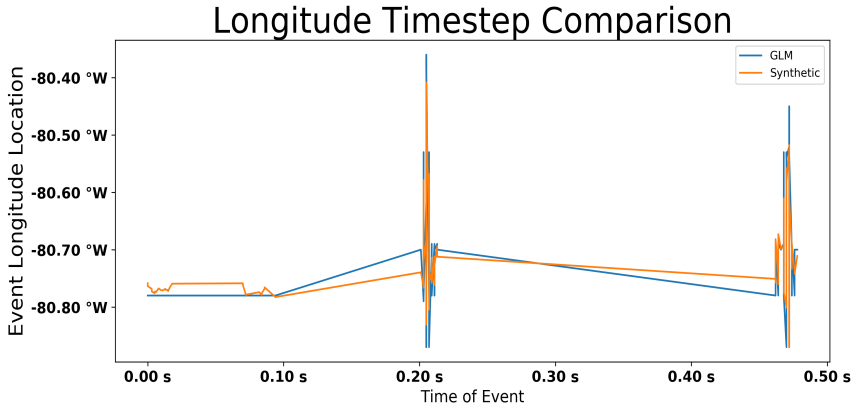


(b) The flash that a custom VARMA model created based on FHEE 3. Similar to the results for FHEE 4, the location is consistent with the original data, and the GLM resolution is not obvious.

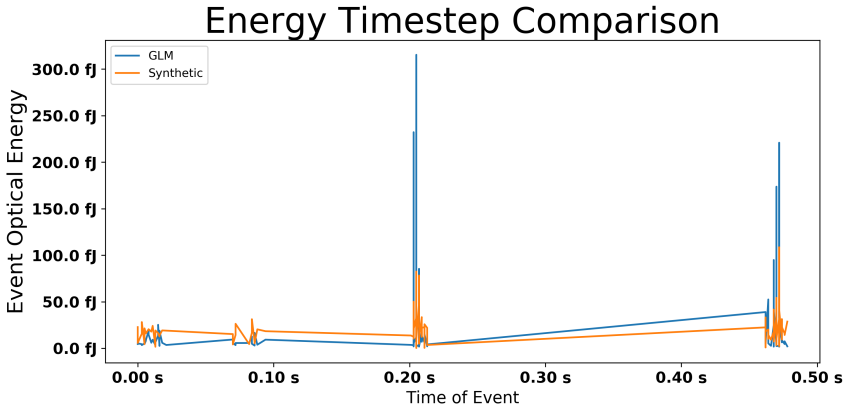
Figure 17. Plots of the GLM events for FHEE 3 (Figure 17a) and the synthetic events predicted by it (Figure 17b).



(a) The comparison of event latitudes for FHEE 3. As was the case with FHEE 1, the general location of the synthetic event latitudes is representative of the GLM data.



(b) The relationship between synthetic and GLM longitudes for FHEE 3. Though the generated longitudes do not mirror the real values, the location and pattern have similar characteristics.



(c) The comparison of event energies for FHEE 3. Two areas of distinct energy spikes occurred during the flash. The VARMA model is able to represent those features, but to less of an extent than was observed naturally. This is the cause for the discrepancies of the event energy standard deviation and range.

Figure 18. The individual comparison for all variables generated through the VARMA model (synthetic data) and the real observations of events from FHEE 3 (GLM data).

4.3 Synthetic Scene Generation

Scenario 1 originates from a batch of lightning events that occurred on January 3 2020 from 12Z-14Z in the Gulf of Mexico (Table 3, Figure 8). All latitudes within Scenario 1 fall between values of 25.73° North and 27.52° North, and all longitude values are in the range of -95.43° West and -93.43° West. The pattern of the collection of events over the two hour period is an oval shape that is oriented southwest to northeast. During the 2 hour period, 550,891 events were detected, and the locations of all of them are plotted in Figure 19.

The VARMA model applied to this dataset of events utilizes an AR order of one and an MA order of two. The model results in an AIC of 2,080,661.40 and a maximum value of the likelihood function of -1,040,297.70. The output of the synthetic lightning events predicted by Scenario 1 is plotted in Figure 20. It has the same outline and the same approximate locations as the real data. However, a striped pattern is observed in the data, with lines running from north to south throughout the entirety of the area where events are generated (Figure 20). Additionally, some of the forged events stray away from the original shape, an occurrence not observed in the original GLM data (Figure 19).

The map of the lightning data created using VARMA from Scenario 1 with statistical noise added is found in Figure 21. The noise added comes from a random uniform distribution and limits the events to stay within 11 km of the beginning location, as this was the approximate resolution of the GLM data from which the data is derived. The striations found in Figure 17 are not observed when noise is added to the data (Figure 21). The pattern of the data also resembles that of the authentic GLM events, as the original oval shape is now detected. The generated data with noise still contains a handful of synthetic events that deviated away from the main cluster.

Figure 22 shows the entire fabricated lightning scene parsed into four different time frames of 30 minutes each. The timeframes of 1200Z-1229Z and 1230Z-1259Z on June 3 contain events throughout the original shape of Scenario 1, while a lesser density is observed in the southeast portion of the shape during the intervals of 1300Z-1329Z and 1330Z-1359Z.

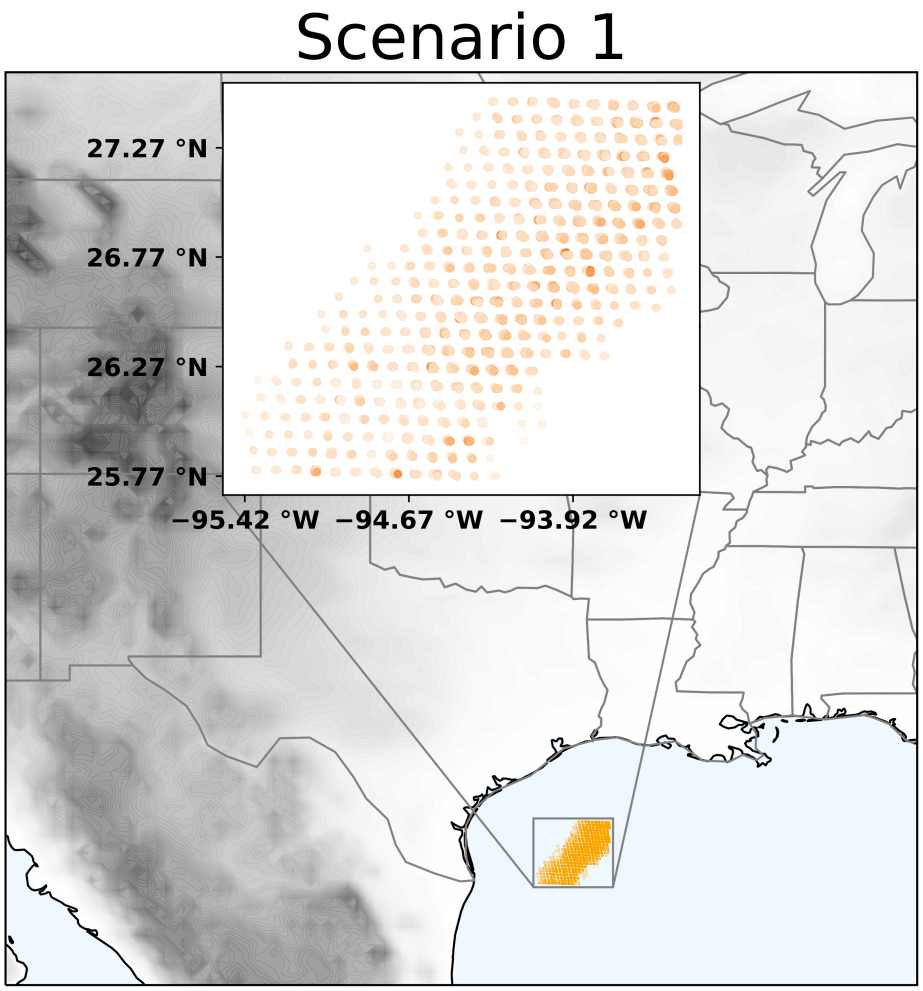
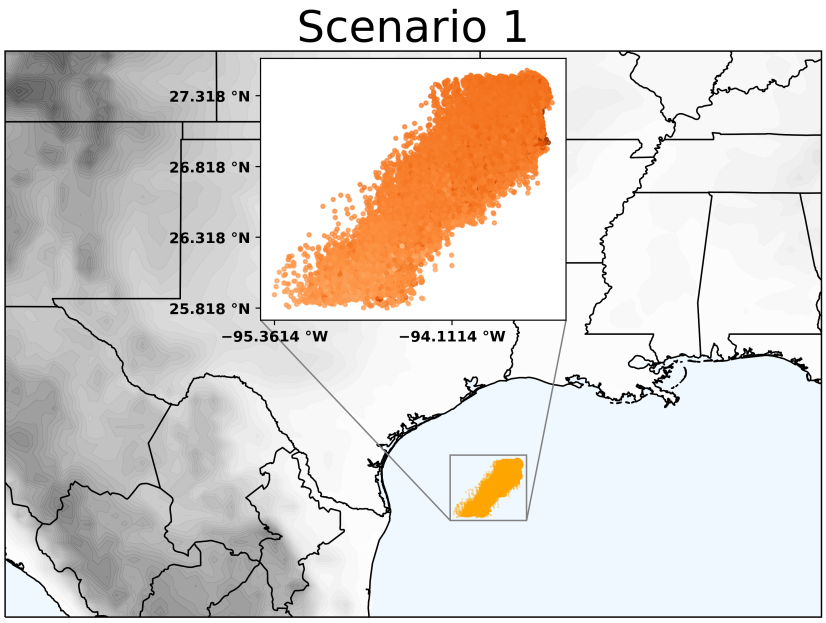
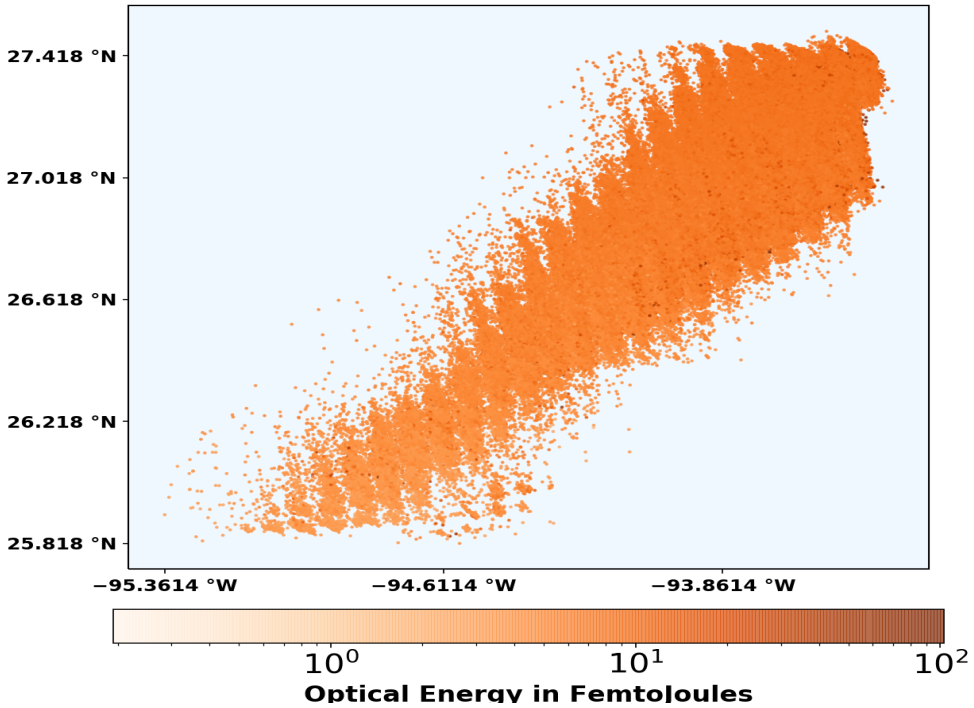


Figure 19. The location of every event captured during Scenario 1. The gray square outlines the all events that occurred during Scenario 1, and the box with white backdrop shows the enlarged picture of the plot. The resolution of the GLM is evident in the formation of the data. Darker locations signify higher event energies, and the larger dots imply that many events occurred at that specific location during the time span.



(a) The location of synthetic lightning events after a VARMA model has been applied to Scenario 1. The area with the white backdrop is a magnified image of the synthetic events.



(b) An enlarged picture of the events found in Figure 20a. The events seem to follow a linear pattern oriented north-south, which is anomalous based on the GLM data. 550,891 events are plotted, which causes Scenario 1 to be the largest data set. The metrics for the model performance are found in the figure caption.

Figure 20.

Scenario 1 - 550,891 Observations	
AIC : 2,080,661.40	L : -1,040,297.70

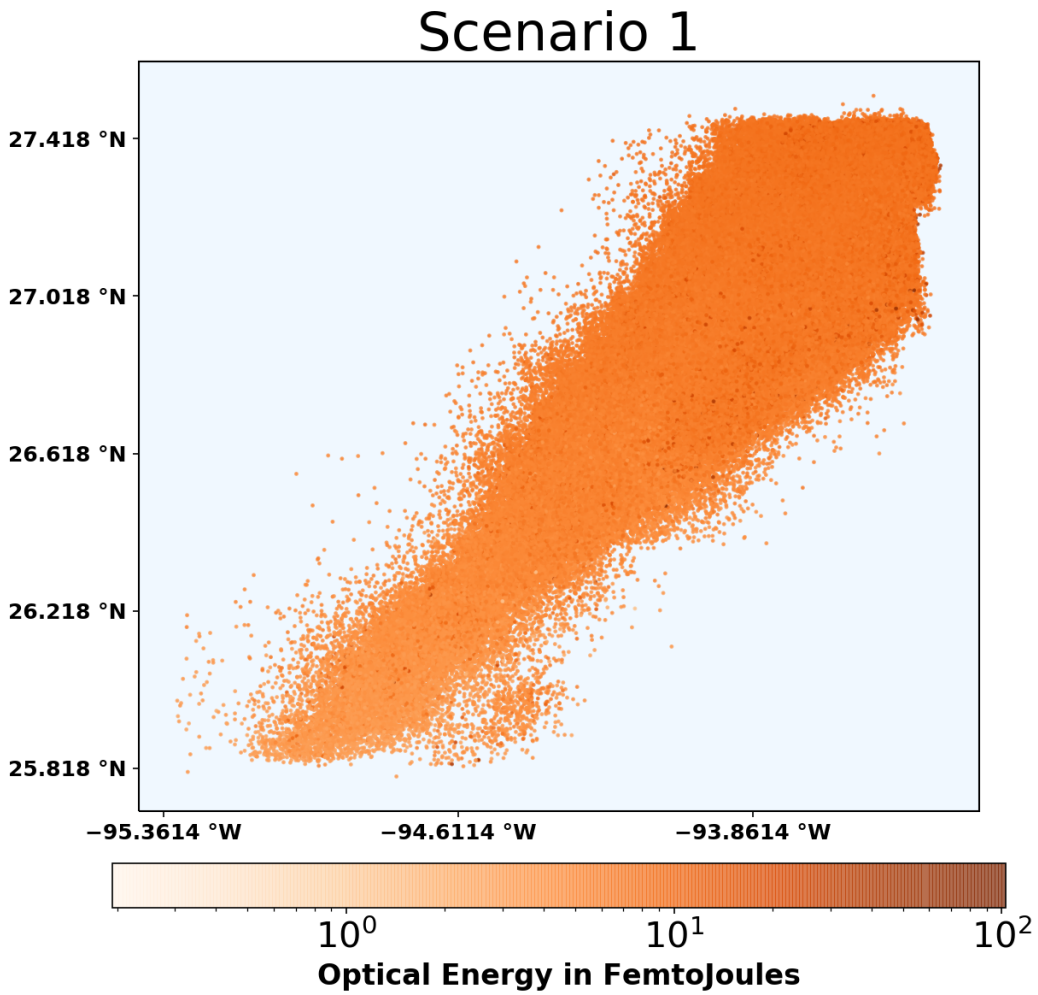
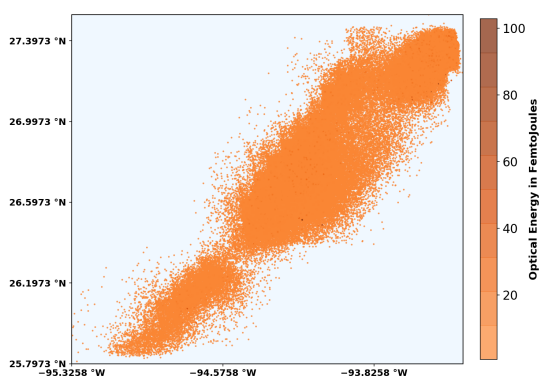
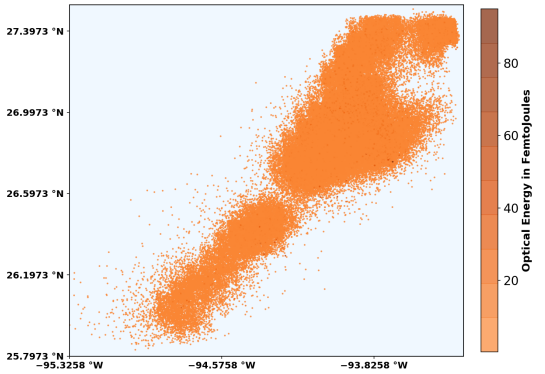


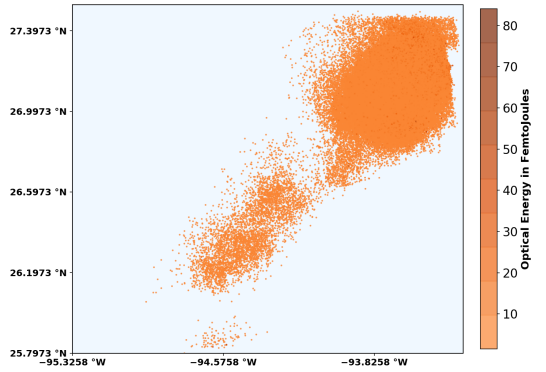
Figure 21. This figure contains the same data as was found in Figure 20b, but this time with statistical spatial noise added to each event. Though no land reference is plotted, the latitudes and longitudes still correspond with the location outlined in Figure 20, just off the east coast of Texas in the Gulf of Mexico. The added statistical noise eliminates the linear pattern observed with just data acquired through a VARMA model.



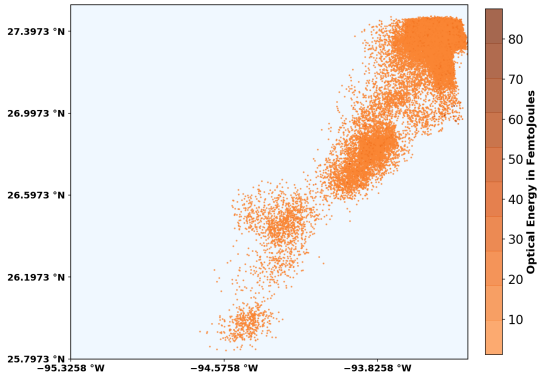
(a) **June 26 2020 1200Z-1230Z**



(b) **June 26 2020 1230Z-1300Z**



(c) **June 26 2020 1300Z-1330Z**



(d) **June 26 2020 1330Z-1400Z**

Figure 22. Displayed is the plot from Figure 21 parsed into four time blocks of 30 minutes each. A snapshot the progression of Scenario 1 is now apparent. The axes of each subplot is locked on ensure that the perspective of the plot remains constant. Each subplot has its own set of event energies; thus, those values oscillate throughout the sequence.

Scenario 2 is formed from the collection of GLM events that occurred on June 26 2020 from 02Z-06Z (Table 3, Figure 8). The data points are found in western Nebraska, northeastern Colorado, and southeastern Wyoming. All latitudes within Scenario 2 fall between values of 40° North and 42°North, and all longitude values are in the range of -105° West and -102° West. When every event from the time frame is plotted, the event plot forms a larger cluster that is oriented west-east with three smaller clusters found south of the predominant one. 41,435 events were observed during the four hour period. A plot of the entirety of the dataset is found in Figure 23.

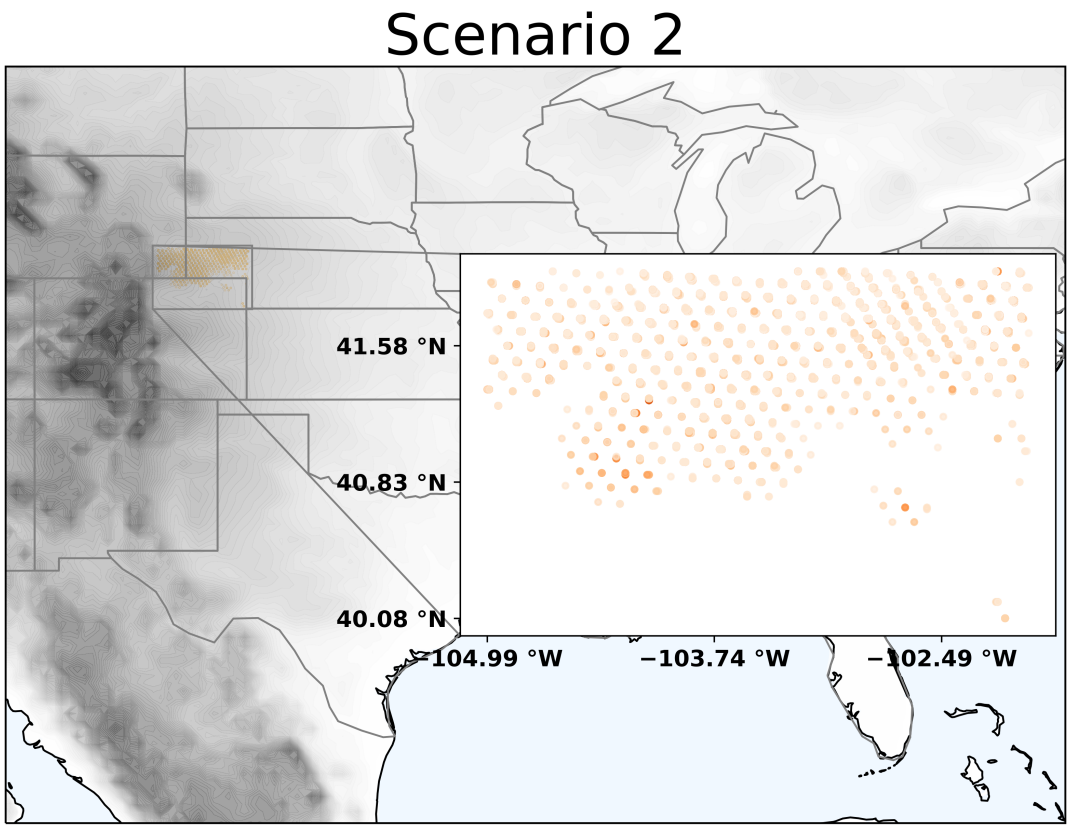
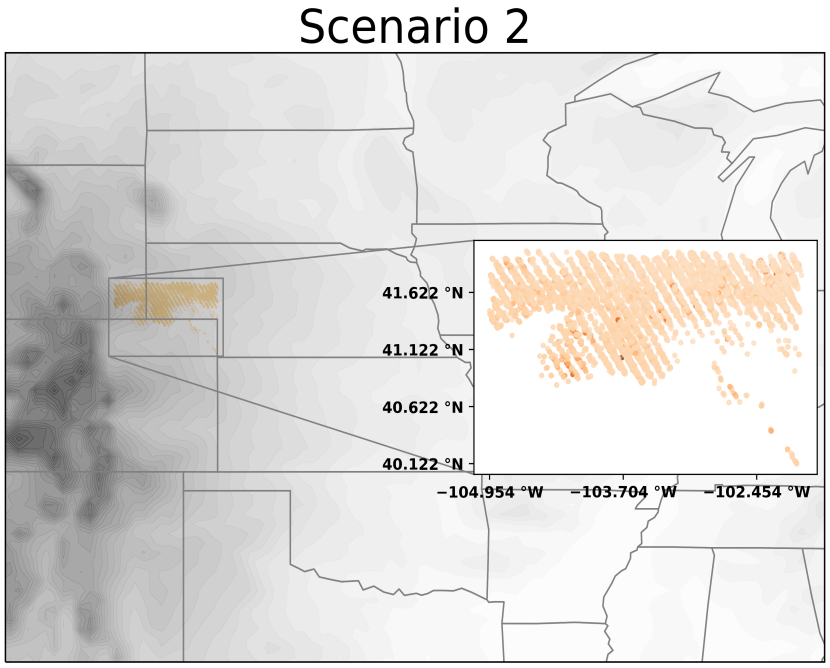


Figure 23. Plotted is every event that occurred during the 4 hour period from 02Z-06Z on Jun 26. This dataset is called Scenario 2 for the purpose of the research. The plot of these latitudes and longitudes is less symmetrical than that of Scenario 1 and also contains far fewer observations.

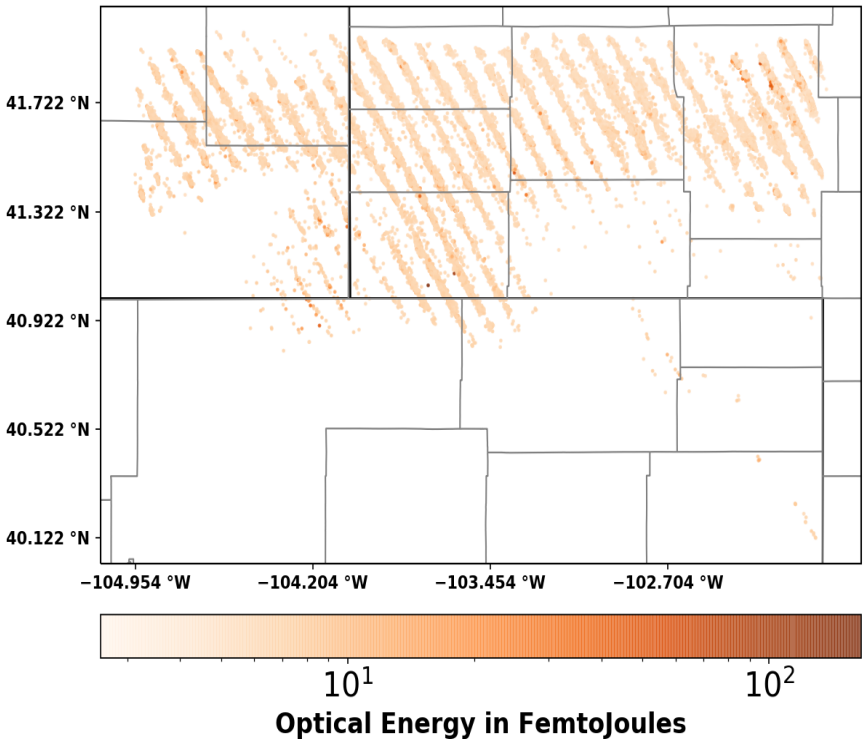
The VARMA model applied to this dataset of events utilizes an AR order of one and an MA order of two. The model results in an AIC of 225,751.72 and a maximum value of the likelihood function of -112,842.86. The plot of generated lightning events from Scenario 2 is shown in Figure 24. The pattern of Figure 24 compared to Figure 23 is comparable; however, as is observed for the same comparison for Scenario 1, the created events form in lines, which are oriented northwest-southeast.

The map of the lightning data created using VARMA from Scenario 2 with statistical noise added is found in Figure 25. The noise is added to both the latitude and longitude for each event, and limits the events to stay within 11 km of the beginning location, the approximate resolution of the GLM data from Scenario 2. The pattern still resembles that which was found in Figure 24; however, the linear pattern is no longer apparent. The clusters of events throughout the four hours now appears as a continuous field of data as it did in Figure 23. A handful of generated events stray away from the original area of where lightning was found. They mostly form to the south of the area.

Figure 26 shows the entire fabricated lightning scene divided into four different time frames of one hour each. At a time of 0200Z-0259Z, two distinct regions of lightning events are clustered together, one in southeastern Wyoming and the other in western Nebraska. From 0300Z-0359Z, only one cluster is discovered. This is likely due to both clusters tracking east, and the one that originated in Nebraska no longer being within the FOV. From 0400Z-0459Z, the cluster that was first seen in Wyoming has tracked further east and split into two separate bunches. Finally, from 0500Z-0559Z, the cluster has pushed to the northeast into Nebraska and reformed into a singular collection of events.



(a) Plotted is a VARMA model's best fit predicated by data from Scenario 2. The area with the white backdrop is an enlarged picture of the data.



(b) The linear patterns discussed in the analysis section are the most apparent in this plot. 41,435 synthetic events are plotted. The AIC is the lowest for Scenario 2, as found in the figure caption below.

Figure 24.

Scenario 2 - 41,435 Observations	
AIC : 225,751.72	**L : -112,842.86**

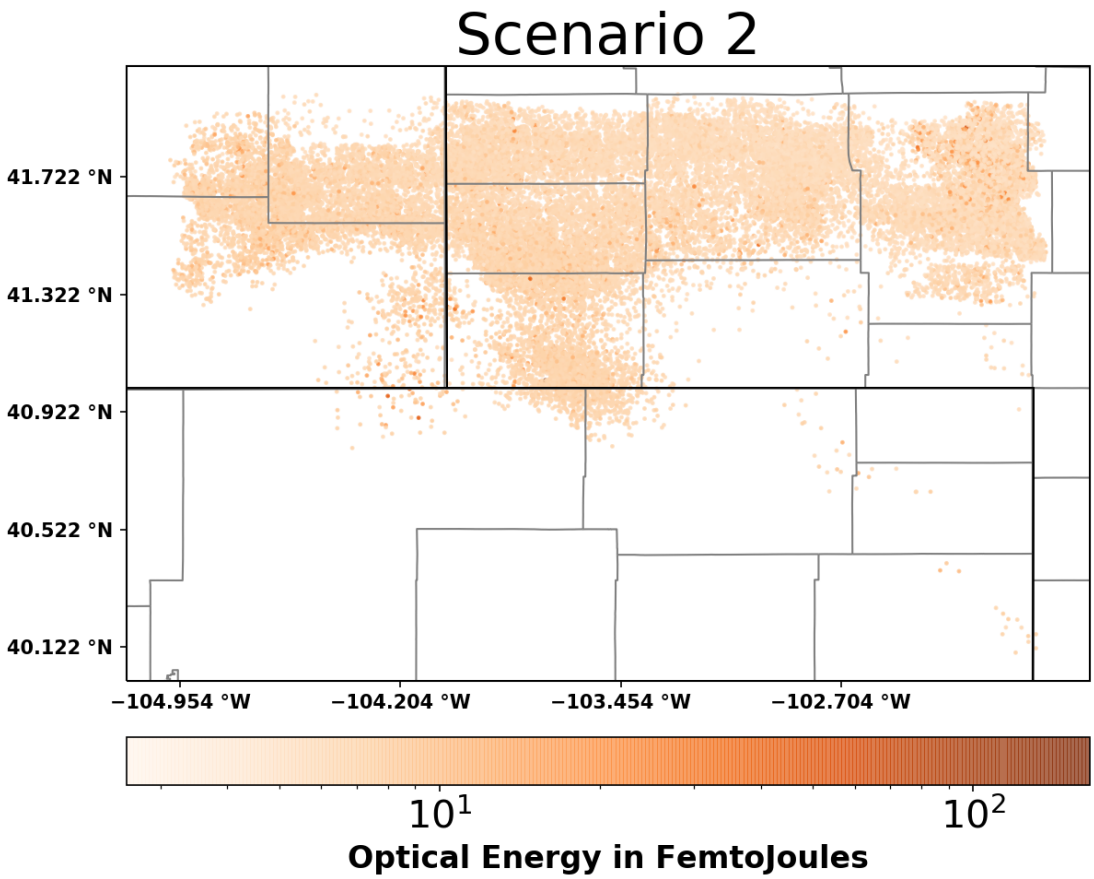
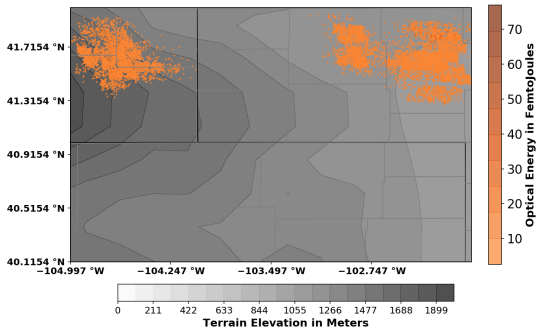
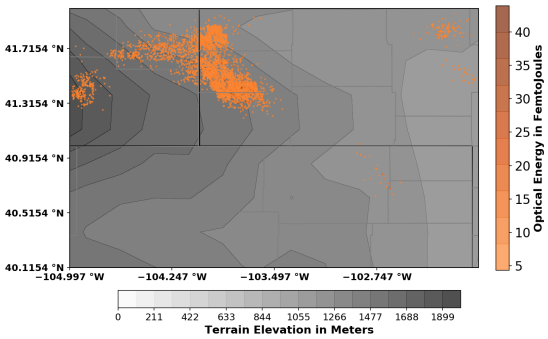


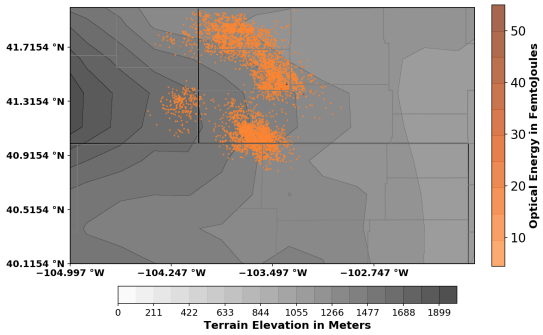
Figure 25. This figure contains the data from Scenario 2 with a VARMA model applied and random uniformly distributed spatial noise added. The noise allows each pixel to occupy a random spot in a 5.5 by 5.5 km box around the original spot, agreeing with the GLM resolution in this region. Even this minimal amount of noise removes the striations observed in Figure 24.



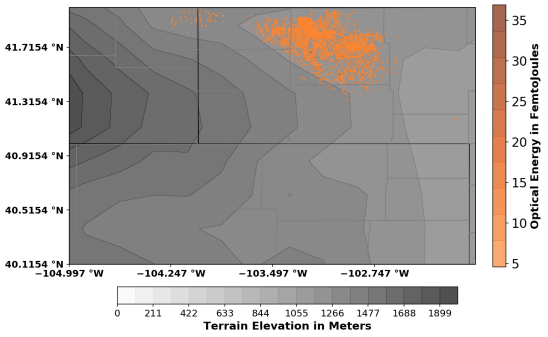
(a) **June 26 2020 0200Z-0300Z**



(b) **June 26 2020 0300Z-0400Z**



(c) **June 26 2020 0400Z-0500Z**



(d) **June 26 2020 0500Z-0600Z**

Figure 26. The data from Figure 25 plotted into four equal, consecutive time frames is plotted in this figure. Unlike the other two scenarios, a west-east flow is evident. There is a large discrepancy in terrain for the background image for this data; it is displayed on each subplot, though no inferences regarding its effective on the lightning is made.

Scenario 3 originates from a collection of lightning events that occurred on July 5 2020 from 2000Z-2030Z on the southeast coast of Florida (Table 3, Figure 8). The FOV includes all latitudes from 25.73° North to 27.52° North, and all longitudes from -95.43° West to -93.43° West. The strikes form a pattern of three separate regions oriented from north to south when plotted out for the entire 30 minute block. 85,087 events were detected during the 30 minute period, and their locations are plotted in Figure 27.

The VARMA model applied to this dataset of events utilizes an AR order of one and an MA order of two. The model results in an AIC of 493,472.38 and a maximum value of the likelihood function of -246,703.19. The lightning events produced via the VARMA model tailored to Scenario 3 are plotted in Figure 28. The outline and locations of the lightning clusters are identical to the original data (Figure 28). However, striations are observed in the created data. Lines are apparent running west-east throughout each different area of lightning. Moreover, a small number of events stray away from the original locations of all events.

The map of the lightning data created using VARMA from Scenario 3 with statistical noise added is found in Figure 29. The uniformly distributed noise is added to both the latitude and longitude for each event, and limits the events to stay within 11 km of the beginning location, the approximate resolution of the GLM data from Scenario 3. The linear pattern found in Figure 15 is not observed when noise is added to the data, and it resembles that of the raw GLM data, as it did for both Scenario 1 and Scenario 2. The data with noise still contains the handful of synthetic events that deviate away from the main formations of lightning.

Figure 30 shows the entire fabricated lightning scene split into four different time frames of 7.5 minutes each. The plots for all of the time spans are similar, as multiple areas of lightning events are observed. Most of these areas show sustained output

throughout the entire 30 minute duration. An area of lightning on the southern tip of Florida can be seen dissipating from 2007:30Z to 2030:00Z.

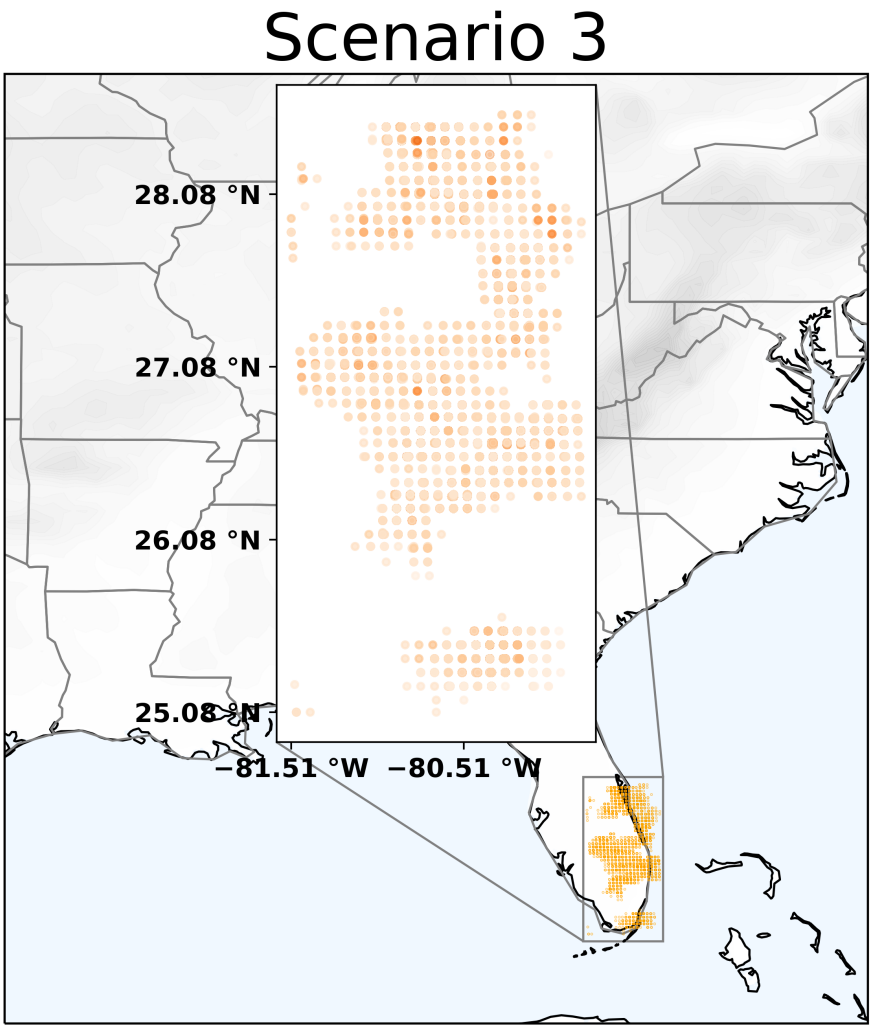
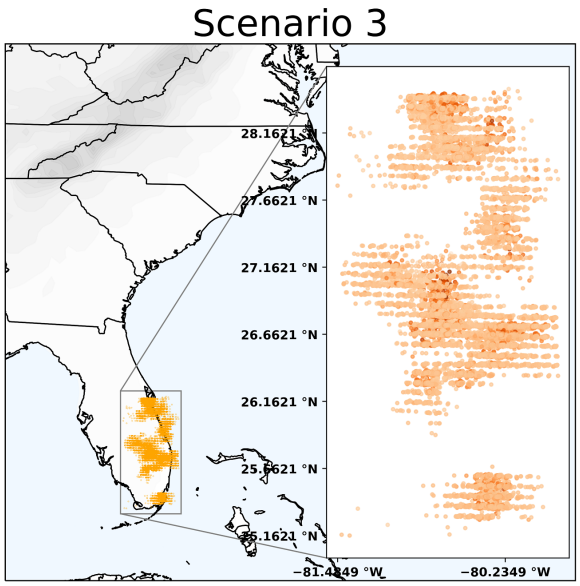
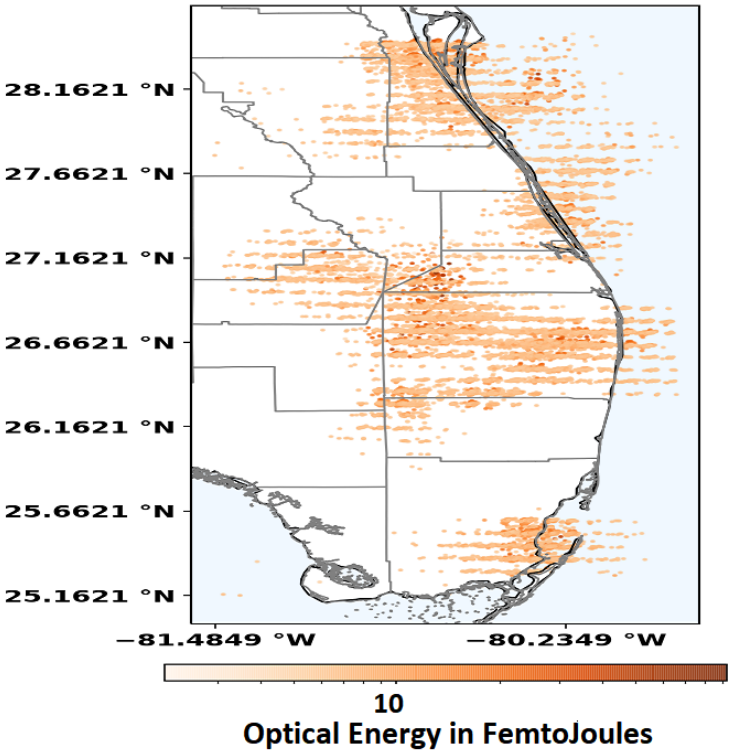


Figure 27. The location of every GLM event collected during Scenario 3, a 30 minute period in January of 2020. Multiple clusters of lightning are observed for this period as opposed to the one general area detected in the other two scenarios. The GLM resolution of approximately 11 km pixels is clear in this plot.



(a) Shown is the locations of VARMA model generated lightning strikes with a subplot to enhance the data points.



(b) An enlarged picture of the events found in Figure 28a. The striations continue to materialize in the VARMA data, oriented west-east in this case. The shape of the synthetic data mirrors the various clusters of convection the GOES-16 GLM detected. The metrics for the model performance and event count are found in the figure caption.

Figure 28.

Scenario 3 - 85,087 Observations	
AIC : 493,472.38	L : -246,703.19

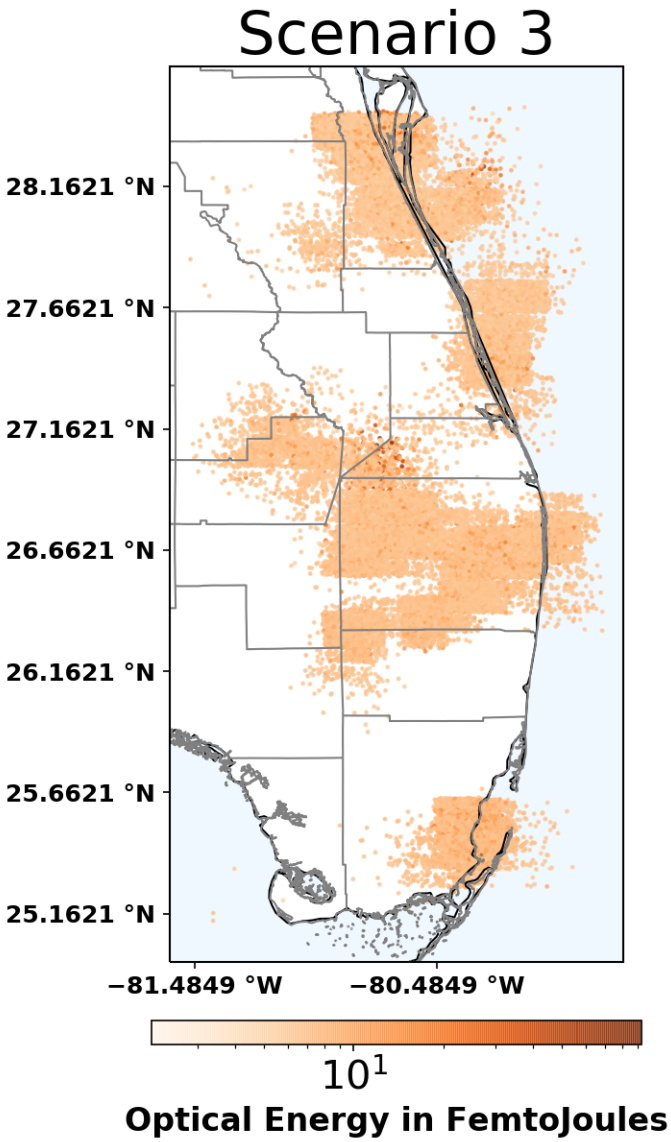
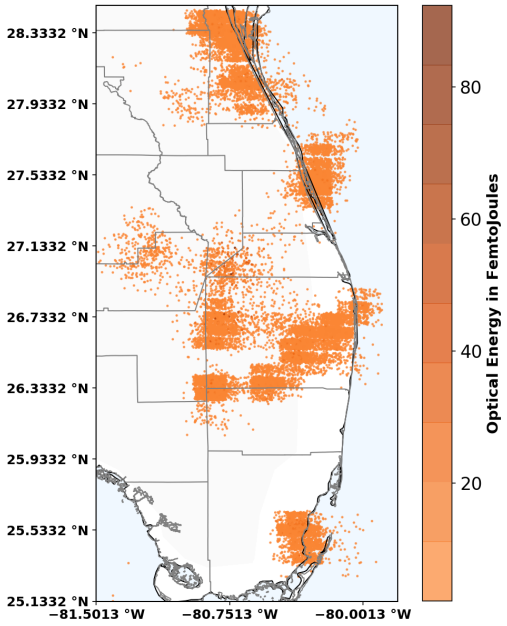
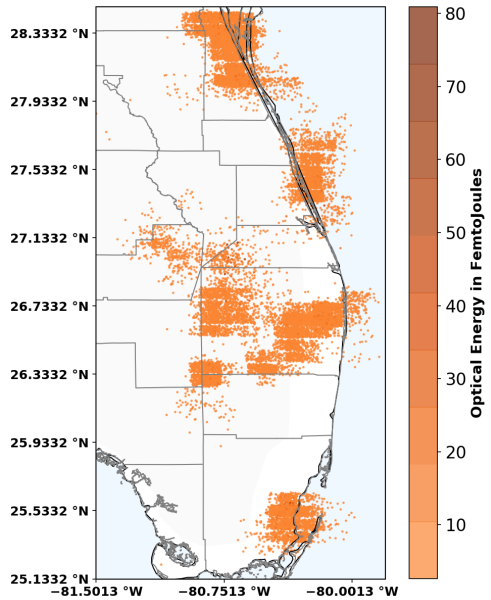


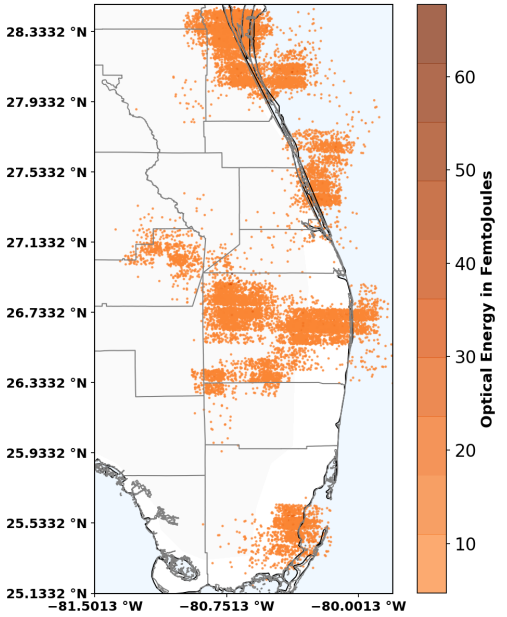
Figure 29. Random uniformly normal statistical noise is applied to the data from Figure 28. The events no longer form lines oriented west-east. The distinct areas of lightning signatures remain separated despite the spatial noise being added.



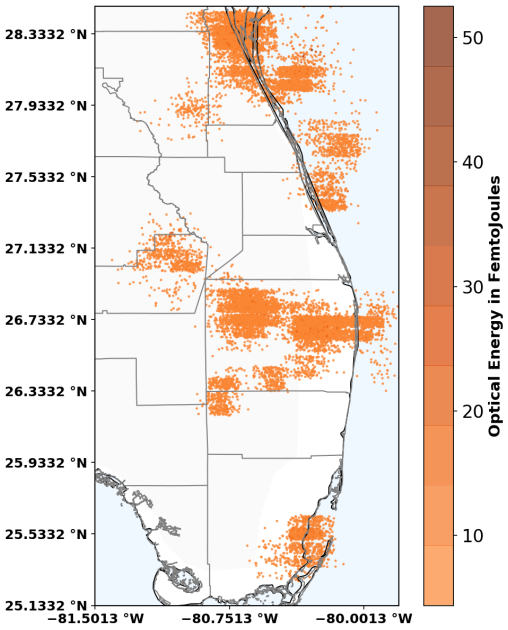
(a) **July 5 2020 2000:00-2007:30Z**



(b) **July 5 2020 2007:30-2015:00Z**



(c) **July 5 2020 2015:00-2022:30Z**



(d) **July 5 2020 2022:30-2030:00Z**

Figure 30. Scenario 3 contains the shortest timetable of the three scenarios with a duration of 30 minutes. Despite this fact, 85,087 GLM events were still collected, double the number of Scenario 2. This timetable is a probable reason for why little variation is discovered between the four iterations. The optical energies vary throughout the sequence, as was the case for both Scenario 1 and Scenario 2.

V. Discussion and Conclusions

5.1 Discussion

5.1.1 High Energy Event and Flash Comparison

The location of GLM events that are within the top 99.68% and 99.98% are calculated for this research (see Chapter 4.1). These results are calculated to compare with the findings of Rudlosky et al. (2019) to evaluate if the location of the two phenomena are complementary. At a first glance, this research suggests that the locations and densities of high energy events are not congruous with those of high energy flashes. Similar characteristics are found, such as the higher superbolt density values and higher average energies found towards the northern edge of the GOES-16 FOV. Otherwise, exact resemblances cannot be observed. Some explanations for the differences can be deciphered; for instance, the greater concentration of high energy events found in the Southeastern CONUS is likely skewed by the increased flash density that occurs in that region during the summer months (Figure 8). The larger amount of bright events over the Sierra Madres compared to the Rocky Mountains is presumably due to the greater amounts of convection the Sierra Madres gets. The region has a large amount of lightning activity compared to the rest of the Americas (Rudlosky et al., 2019).

The clear division between flash energies over the ocean and over the land is not seen for event energies (Figure 5, Figure 9). It is probable that this discrepancy is found due to the way in which flash energies are calculated: by taking the sum of all events (Koshak et al., 2018). Rudlosky et al. (2019) calculated that the average event has an optical energy of 7.1 fJ over the ocean and 6.0 fJ over land (see Chapter 2.3). The average event optical energy is higher over the ocean, but only by a margin of 18% compared to 83% for flashes (Rudlosky et al., 2019). The events per flash

count is also higher over the ocean compared to over the land, as the count is 39.3 for over land and 57.0 for over the ocean (Rudlosky et al., 2019). This further supports the proposition that the locations of high energy flashes and high energy events are not analogous.

5.1.2 VARMA Model Performance

The statistical method of VARMA is applied to both FHEEs and scenarios of GOES-16 GLM data (see Chapter 3.3). The goal of this approach is to replicate the presentation of the collection of GLM events for any given timeframe in a manner in which the resolution is not constrained as it is for the GLM (Goodman et al., 2012). FHEEs are analyzed to discover how well VARMA models work on small scales and also how well they reproduce optical energies. Scenarios are analyzed because they represent a complete FOV, akin to the ASSET.

The VARMA models used in this research perform adequately when generating latitudes and longitudes for FHEE 1 and 3 (Figure 11, Figure 17). The shapes and locations of those synthetic flashes match with those of their original counterparts, and no abnormal patterns are detected (Figure 11, Figure 17). In regards to FHEE 2, the VARMA model applied is able to match the location of the original for most of the events, but it contains a handful of events that stray away from the primary location (Figure 14). This generates an output that would not be realistically created from one individual lightning flash. Thus, the model for FHEE 2 did not perform adequately as it did for FHEE 1 and 3.

More research would need to be completed to find out why the method broke down for FHEE 2; however, three observations are noted. Firstly, FHEE 2 has the highest event count of the three, as it contains 351 events compared to the 335 for FHEE 1 and 164 for FHEE 3 (Table 4, Table 5, Table 6). It could be the case where

that number is too large to adequately formulate a best-fit equation for that particular AR and MA order for the model. Secondly, FHEE 3 has the longest flash duration of the three; the time from first event to last event is 2.149 seconds (Table 5). This long time span could cause issues for the model. Thirdly, FHEE 2 has a lower maximum value of the likelihood function than the other two. This signifies that the model fits the original data less optimally than the others (see Chapter 2.5).

The VARMA models utilized in this research perform well analytically when they generate the average optical energies. The average synthetic optical energy has a divergence from the GLM mean of .05 fJ for FHEE 1, .85 fJ for FHEE 2, and 4.49 fJ for FHEE 3 (Table 4, Table 5, Table 6). These values are all within one standard deviation relative to the original data. However, the ranges and standard deviations of the synthetic optical energies are not representative of the real ones. For example, the physical observation from FHEE 1 has an optical energy range of 398.67 fJ, which is almost five times that of the synthetic observation (Table 4). These are numbers that should be as close to each other as possible. The smallest disparity in the standard deviations is 8.5 fJ, found in FHEE 2 (Table 5). This is still a value that is half from the expected number.

The analyses of the numerical comparisons of the FHEEs suggests that the VARMA models utilized for this research are not able to account for the pattern of an FHEE. As highlighted in Chapter 3.3.1, these flashes are specifically chosen because of their large range of optical energies. The standard deviations are large compared to most other flashes and contain extremely bright events relative to the others (see Chapter 3.3.1). More research would need to be performed to discover if VARMA models are able to accurately recreate optical energies for flashes which do not contain abnormally bright events.

The VARMA models applied to an entire FOV performed equivalently for all

three specified databases of GLM event times, optical energies, latitudes, and longitudes, or scenarios as they are called for the purpose of this research. Overall, the models work well regarding how well the generated areas of events mirror the real ones. The synthetic events are predominantly located in the areas where the authentic data was found when analyzing each scenario as the entire time block and also split into even iterations. This is particularly apparent in Scenario 3, where four different regions of convection are depicted accurately (Figure 28).

While the VARMA process is adequate in representing the regions of real lightning events, the arrangement of the data is inadequate in all three scenarios. Lines of events form in all three generated plots. This development is especially apparent in Scenario 2, where noticeable empty space is found between the aforementioned lines (Figure 24). There is no particular standard these formed lines follow, as they north-south for Scenario 1 (Figure 20), northwest-southeast for Scenario 2 (Figure 24), and west-east for Scenario 3 (Figure 25). Noise in the form of random uniformly distributed latitude and longitude values has to be added to each scenario to create synthetic scenes that do not have abnormal patterns (Figure 21, Figure 25, Figure 29). The noise does not seem to corrupt the data, as no irregularities are detected when parsing the data into even time blocks, one as brief as 7.5 minutes for Scenario 3 (Figure 30).

An important distinction between the comparison of using FHEE and scenarios to create lightning is the number of events inputted. Though FHEEs with a high events per flash count are singled out in this study, the number of data points for them is minuscule compared to the number of lightning signatures that are seen over a set time frame across a given environment. The event count for each FHEE is on the order of hundreds in this study, while the scenarios have event counts of either hundreds of thousands (Scenario 1) or tens of thousands (Scenarios 2 and 3).

This high event count limited both the AR and moving average (MA) orders for the scenarios, as higher values are unable to run due to computational limitations in the model's current iteration. Essentially, applying the VARMA process to FHEEs allows "greediness" to be utilized. A "greedy" algorithm is one that will always take an optimal solution when solving a problem, no matter how much time and computational power is sacrificed in the process (Black, 2005).

The effects of this greedy approach on the synthetically generated lightning observations are best quantified by the AIC. As highlighted in Chapter 2.5, lower AIC numbers signify a model that is of a higher quality, or one that is more representative of the input data. The lesser number of parameters (signified by k) decreases the value of the AIC (Equation 5), and the ability to have much higher orders for the AR and MA terms increases the maximum value of the likelihood function, which ultimately results in dramatically lesser AIC totals. There is a stark contrast between FHEE and scenario AIC values. The AIC values for FHEE 1, 2, and 3 are on the order of thousands (Table 4, Table 5, Table 6). The AIC values for the scenarios are 1,080,661.40 for Scenario 1, 225,751.72 for Scenario 2, and 93,472.38 for Scenario 3 (see Chapter 4.3). These values are all many orders higher than the FHEEs calculations. This discrepancy in AIC is likely the cause for why FHEEs do not require statistical noise to look realistic but scenarios do. The exact threshold the AIC needs to be to avoid the data being skewed is not known.

5.2 Future Work

The processes used to produce both synthetic lightning strikes and scenes in this research is intended to work for any time and day the GLM collected data and location within the GOES-16 FOV. Therefore, more data can be processed to see if there are discrepancies in the season, time of day, and location of GLM events. This

would involve analyzing the location of high energy events for more than one season so that better comparisons can be made to the GLM flash data from Rudlosky et al. (2019).

As highlighted in Chapter 2.4, cloud properties are an important component of ASSET and also a crucial ingredient for lightning. Future research could be accomplished to implement algorithms that restrict lightning event generation in areas that do not contain clouds. Additionally, variables such as CTH and CTP could be actualized from the available ASSET data and be used as part of the scene formation process to tailor lightning output to cloud properties.

Finally, the VARMA method can be improved to better recreate lightning scenes and FHEEs. This would involve tweaking the input parameters to cause the synthetic data to fit the original data more effectively. Improvements to this aspect of the models could negate the need for statistical noise in the scenarios and cause the event energy values to be more similar for created and real data. Also, high-performance computing could be used to allow for more statistical greediness to be enabled in the model runs by increasing the AR and MA orders. Lastly, more variety could be observed in the output if time is included as a dependent variable instead of an independent variable.

5.3 Conclusions

The objective of this thesis is to explore the legitimacy of applying VARMA models to GLM data in order to create synthetic lightning events which are comparable to the original events yet not restricted by resolution. A review of lightning flashes and events is presented, as well as study of the statistical method of VARMA. No previous research is found that attempts to generate lightning data using VARMA models or other statistical methods.

High energy event climatology is compared to that of flash energy. These observations suggest that the locations of the two phenomena are not similar; thus more research needs to be performed to better understand high energy event climatology. The VARMA process is applied to three FHEEs found in various locations at different times. This procedure is generally able to recreate a spatial representation of the authentic flashes without being restricted by GLM resolution. However, the numerical values produced by the models are not precise. The VARMA process is then tested on locations and time frames that contain many flashes, called scenarios. Statistical noise is required to produce synthetic lightning scenes that do not have abnormal patterns in the data.

Appendix A

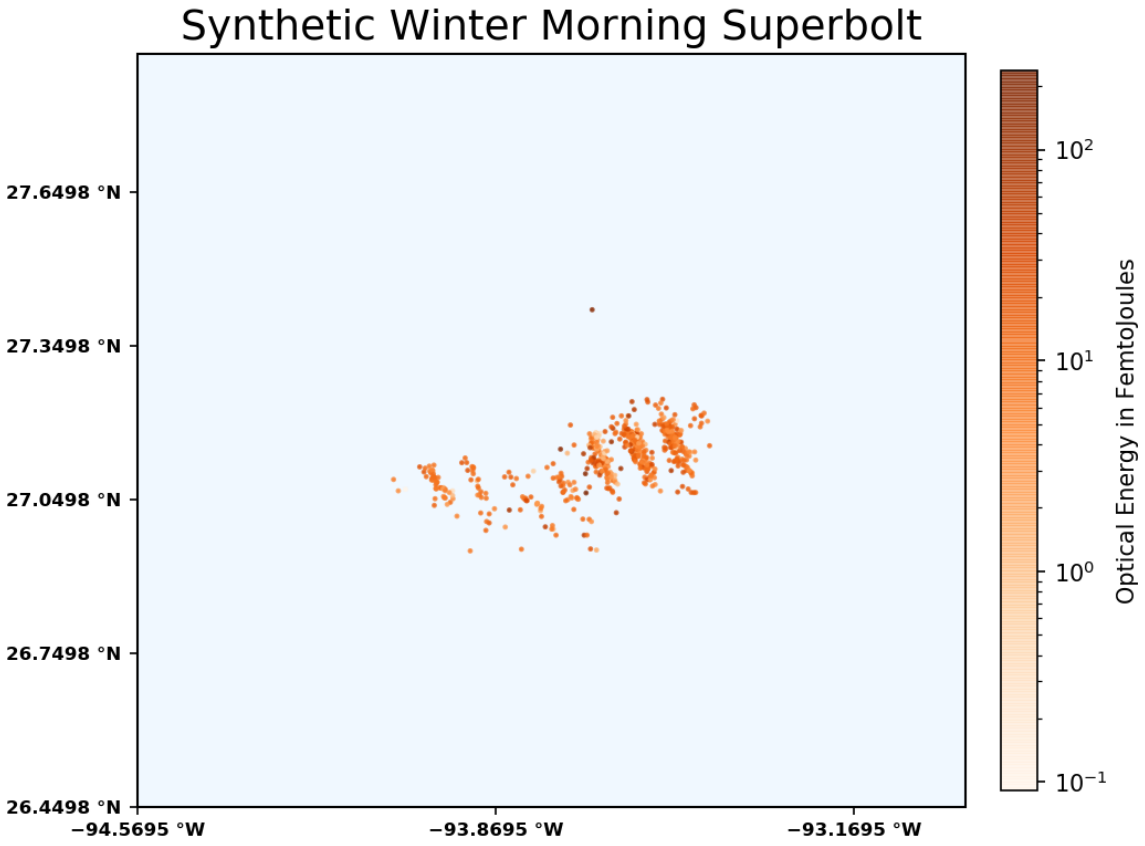


Figure 31. This figure shows a synthetic lightning flash generated from a VARMA model. The linear sequence of the synthetic events (oriented North-South in this case) is reviewed to assure that it is not present when evaluating how the model performed. The original GLM flash contained 607 events, which could be a factor for the poor model performance.

Bibliography

Box, G., Jenkins, G., & Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*.

Bui, V., Chang, L.-C., & Heckman, S. (2015). A Performance Study of Earth Networks Total Lighting Network (ENTLN) and Worldwide Lightning Location Network (WWLLN). https://doi.org/10.1109/CSCI.2015.120

Cecil, D. J., Buechler, D. E., & Blakeslee, R. J. (2014). Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmospheric Research*, *135-136*, 404–414. https://doi.org/10.1016/j.atmosres.2012.06.028

Date, S. (2019). The Akaike Information Criterion. https://towardsdatascience.com/the-akaike-information-criterion-c20c8fd832f2

Earth Networks. (2020). Earth Networks - World Weather Data Experts. https://www.earthnetworks.com/

Fierro, A. O., Mansell, E. R., Ziegler, C. L., & Macgorman, D. R. (2013). Application of a Lightning Data Assimilation Technique in the WRF-ARW Model at Cloud-Resolving Scales for the Tornado Outbreak of 24 May 2011. https://doi.org/10.1175/MWR-D-11-00299.1

GHRC. (2020). Lightning Imaging Sensor (LIS) - Sensor - Observations — GHRC Lightning. https://ghrc.nsstc.nasa.gov/lightning/overview_lis_instrument.html

Goodman, S., Nesdis, N., & Mach, D. (2012). *NOAA NESDIS CENTER for SATELLITE APPLICATIONS and RESEARCH ALGORITHM THEORETICAL BASIS DOCUMENT GLM Lightning Cluster-Filter Algorithm* (tech. rep.).

Jensen. (2015). MIT School of Engineering — Is there a way to harness electricity from lightning? https://engineering.mit.edu/engage/ask-an-engineer/is-there-a-way-to-harness-electricity-from-lightning/

Karagiannidis, A., Lagouvardos, K., Lykoudis, S., Kotroni, V., Giannaros, T., & Betz, H. D. (2019). Modeling lightning density using cloud top parameters. *Atmospheric Research*, *222*, 163–171. https://doi.org/10.1016/j.atmosres.2019.02.013

Koshak, W., Bitzer, P., & Goodman, S. (2018). *Flash Optical Energy from the Geostationary Lightning Mapper* (tech. rep.).

MacGorman, D. R., Straka, J. M., & Ziegler, C. L. (2001). A lightning parameterization for numerical cloud models. *Journal of Applied Meteorology*, *40*(3), 459–478. https://doi.org/10.1175/1520-0450(2001)040⟨0459:ALPFNC⟩2.0.CO;2

Marshall, T. C., McCarthy, M. P., & Rust, W. D. (1995). Electric field magnitudes and lightning initiation in thunderstorms. *Journal of Geophysical Research*, *100*(D4), 7097–7103. https://doi.org/10.1029/95JD00020

NASA. (2020). Vaisala US NLDN Lightning Flash Data. https://ghrc.nsstc.nasa.gov/uso/ds_docs/vaiconus/vaiconus_dataset.html

NLSI. (1995). Lightning's Social and Economic Costs - National Lightning Safety Institute. http://lightningsafety.com/nlsi_lls/sec.html

NOAA. (2017). *Geostationary Lightning Mapper (GLM)* (tech. rep.). http://www.goes-r.gov/spacesegment/glm.html

NOAA. (2019). GOES Satellite Network. https://www.nasa.gov/content/goes

NSSL. (2020). Severe Weather 101: Lightning. https://www.nssl.noaa.gov/education/svrwx101/lightning/

Petersen, D., Bailey, M., Beasley, W. H., & Hallett, J. (2008). A brief review of the problem of lightning initiation and a hypothesis of initial lightning leader formation. https://doi.org/10.1029/2007JD009036

Peterson, M., & Lay, E. (2020). GLM Observations of the Brightest Lightning in the Americas. *Journal of Geophysical Research: Atmospheres*, *125*(23). https://doi.org/10.1029/2020JD033378

Price, C., & Rind, D. (1992). A simple lightning parameterization for calculating global lightning distributions. *Journal of Geophysical Research*, *97*(D9), 9919–9933. https://doi.org/10.1029/92JD00719

Rakov, V., Uman, M., & Rambo, K. (1998). New insights into lightning processes gained from triggered-lightning experiments in Florida and Alabama. *Wiley Online Library*. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97JD02149

Rudlosky, S. D., Goodman, S. J., Virts, K. S., & Bruning, E. C. (2019). Initial Geostationary Lightning Mapper Observations. *Geophysical Research Letters*, *46*(2), 1097–1104. https://doi.org/10.1029/2018GL081052

Shafer, P. E., & Fuelberg, H. E. (2006). A statistical procedure to forecast warm season lightning over portions of the Florida peninsula. *Weather and Forecasting*, *21*(5), 851–868. https://doi.org/10.1175/WAF954.1

Turman, B. N. (1977). Detection of lightning superbolts. *Journal of Geophysical Research*, *82*(18), 2566–2568. https://doi.org/10.1029/jc082i018p02566

Williams, E. R. (1985). LARGE-SCALE CHARGE SEPARATION IN THUNDERCLOUDS. *Journal of Geophysical Research*, *90*(D4), 6013–6025. https://doi.org/10.1029/jd090id04p06013

Wilms, I., Basu, S., Bien, J., Matteson, D. S., Basu, /., & Basu, S. (2019). *Sparse Identification and Estimation of Large-Scale Vector AutoRegressive Moving Averages* (tech. rep.). http://faculty.bscb.cornell.edu/~bien/

Yair, Y., Lynn, B., Price, C., Kotroni, V., Lagouvardos, K., Morin, E., Mugnai, A., & Llasat, M. d. C. (2010). Predicting the potential for lightning activity in Mediterranean storms based on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields. *Journal of Geophysical Research*, *115*(D4), D04205. https://doi.org/10.1029/2008JD010868

Young, S., Steward, B., & Gross, K. (2017). Development and validation of the AFIT scene and sensor emulator for testing (ASSET).

Zhu, Y., Lyu, W., Cramer, J., Rakov, V., Bitzer, P., & Ding, Z. (2020). Analysis of Location Errors of the U.S. National Lightning Detection Network Using Lightning Strikes to Towers. *Journal of Geophysical Research: Atmospheres*, *125*(9). https://doi.org/10.1029/2020JD032530

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	Jun 2019 — Mar 2021

4. TITLE AND SUBTITLE

SYNTHETIC LIGHTNING GENERATION EMPLOYING AUTOREGRESSIVE-MOVING-AVERAGE (ARMA) MODELS

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Powers, Seth R., 1st Lt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

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

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENP-MS-21-M-131

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

Undisclosed

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION / AVAILABILITY STATEMENT

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Lightning is a meteorological phenomenon that occurs frequently in numerous locations every single day on Earth. Comprehending the way it materializes spatially and chronologically is imperative to develop a realistic environmental scene. Live lightning data can be fed into a scene, but that process is costly. Therefore, this work explores if lightning data can be generated synthetically using vector autoregressive-moving-average (VARMA) models. Geostationary Lightning Mapper (GLM) data is used as the basis for the study. Lightning climatology is examined and compared to previous research to gain insight into the targeted areas. Individual lightning flashes are analyzed to inspect how well the process works on a smaller scale. Then, entire regions are evaluated to simulate lightning creation in a larger setting. The results of each step suggest that VARMA models are sufficient at lightning generation up to a certain degree. However, the techniques used in this study have the potential to be improved and allow the models to mirror expansive scenes containing many lightning strikes.

15. SUBJECT TERMS

Geostationary Lightning Mapper, Vector Autoregressive-Moving-Average

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Maj Peter Saunders, AFIT/ENP
U	U	U	U	70	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4505; Peter.Saunders@afit.edu