NWRA Lightning Analysis

April Gao, Janice Kim, Elaine Zhang
UW MSDS
March 17, 2025

Introduction

Tropical cyclones (TCs) are among the most powerful and destructive weather systems on Earth, posing significant threats to life, property, and infrastructure. To mitigate these impacts, researchers continuously monitor various storm characteristics such as intensity, wind speed, convection patterns, and lightning activity. Of particular interest is the role of lightning as a potential indicator of changes in storm dynamics. Previous studies suggest that lightning activity within the inner core of a tropical cyclone, typically defined as the region within 100 kilometers of the storm center, may be closely linked to fluctuations in storm intensity. Research presented by Molinari, Moore, & Idone (1999) suggests that lightning outbreaks in the core of weakening or steady hurricanes may signal rapid intensification, while in deepening hurricanes, lightning outbreaks in the core may indicate weakening or reversal.

Understanding these patterns offers valuable insights into TC intensification patterns and could enhance TC intensification forecasting accuracy. Currently, predictive models for TC intensification do not perform very well. By analyzing lightning occurrences in relation to tropical cyclone development and evolution, this project aims to uncover potential correlations that could inform predictive models. These findings could significantly contribute to efforts to improve early warning systems and disaster preparedness strategies, benefiting both research and operational forecasting communities.

Problem Statement

This study aims to analyze tropical cyclone and lightning data to investigate and explore the relationship between lightning patterns and the intensification cycle of tropical cyclones. The intensity of tropical cyclones is measured by wind and pressure. We access global data on tropical cyclones, including metrics on wind, convection, lightning occurrences, and tropical cyclone storm center locations from worldwide stations for this project. The tropical cyclone and lightning data are available from two sources: World Wide Lightning Location Network (WWLLN) and Geostationary Lightning Mapper (GLM). We perform our analysis on the WWLLN dataset for this project.

The first part of this analysis involves defining a WWLLN lightning burst. We conducted statistical analysis on lightning stroke data to identify bursts based on lightning density, defining a burst as a 30-minute time bin with a spike in lightning activity relative to the storm's overall lightning trends. As of now, spikes in lightning are visually identified, Solorzano et al. 2018., rather than through structured, mathematical methods. The second part of the analysis explores the relationship between lightning bursts and tropical cyclone intensity. Previous research indicates that intense, frequent lightning may be followed by a lull period, often occurring near the tropical cyclone's eye. However, the exact relationship between lightning activity and the stages of tropical cyclone intensification remains unclear. To explore this relationship, we applied statistical analyses, such as examining distributions and identifying significant thresholds, to determine lightning activity associated with different intensification stages. These stages, defined by changes in wind speed over a 24-hour period, include Neutral, Intensifying, and Weakening.

Our primary aim is to understand these statistical relationships and build a formula. We initially focused on inner core lightning density, defined as within 100 km of the storm center, due to its relative simplicity as a smaller, more contained area with fewer extraneous variables. After performing analysis on the inner core, we performed similar analyses on the rainband, defined as 200-400 km from the storm center. Rainband lightning is further categorized by shear quadrants defined by vertical wind shear angle at the time of the lightning. We then investigated the relationship between the detected lightning bursts and tropical cyclone intensification stages for both inner core and rainband lightning.

Data Pipeline

The datasets provided by NorthWest Research Associates (NWRA) include 984 tropical cyclones (TC) over 6 basins from 2010 to 2020. Each tropical cyclone (TC) is associated with three data files: (1) a World Wide Lightning Location Network (WWLLN) file containing lightning stroke timestamps and north/east distances from the storm center, (2) a storm track file with wind speed, pressure, and storm center coordinates recorded at regular intervals, and (3) a MATLAB file with vertical wind shear vector magnitude and angle data for each WWLLN lightning stroke. Each of these files were combined across TCs to get one combined WWLLN lightning file, one combined TC track file, and one combined WWLLN lightning and vertical wind shear file for rainband lightning. We then calculated the maximum wind speed per TC to get the TC's category using the Saffir-Simpson Hurricane Wind Scale (Figure 1). The data was then filtered down to TCs categorized as Category 1 or higher, leaving us with 472 TCs for use in subsequent burst and intensification analysis.

Category	Sustained Winds (Knots)
1	64-82 kt
2	83-95 kt
3	96-112 kt
4	113-136 kt
5	137 kt or higher

Fig. 1: Saffir-Simpson Hurricane Wind Scale

The inner core and rainband lightning datasets were then separated by calculating the direct distance of each lightning stroke to the storm center. Using the hypotenuse method, which applies the Pythagorean theorem, the direct distance of each lightning stroke to the storm center was calculated based on its eastward and northward distances from the center. Lightning within 100 km of the storm center was classified as inner core lightning, while lightning between 200 km and 400 km was classified as rainband lightning.

For the rainband lightning data, we assigned each lightning stroke to a shear quadrant. Using the vertical wind shear vector angle data, we calculated the shear angle Cloud-to-Ground (CG) by subtracting the shear vector angle from the geographic angle of the lightning stroke relative to North. We used the inverse tangent function (arctan) applied to each lightning stroke's distance North and East of the storm center to calculate the geographic angle relative to North, where a lightning stroke due east of the storm center would yield a geographic angle of 90 degrees. After getting the shear angle CG, we classified the lightning stroke's shear quadrant using the shear angle CG as follows: 0–90° to Downshear Left (DL), 90–180° to Downshear Right (DR), 180–270° to Upshear Right (UR), and 270–360° to Upshear Left (UL) (Figure 2).

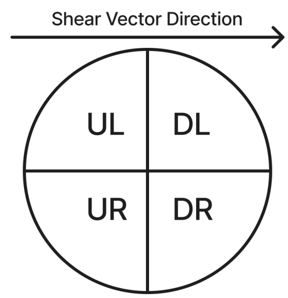

Fig. 2: Shear quadrant diagram, with shear vector to the right

To define a burst of lightning, we need the number of lightning strokes associated with a time period. As such, we binned the lightning data for each TC into 30-minute time bins and calculated the number of lightning strokes in each. For the rainband lightning, we calculated these counts by shear quadrant, yielding four counts per time bin. To analyze the relationship between lightning and TC intensification stages, binned lightning data was joined with the nearest wind speed and pressure observation. Due to the differences between track file observation intervals and lightning stroke observation intervals, this join is a potential source of error. Each TC has time bins from the first to the last track file observation, regardless of lightning count, to ensure complete representation of the TC.

To analyze lightning burst patterns with intensification stages at a granular level, we assigned intensification stages and current categories for each time bin. The intensification stages are defined below (Figure 3) , where the changes in wind speeds are forward-looking 24-hour differences. The current category of each time bin is defined using the time bin's wind speeds and the Saffir-Simpson Hurricane Wind Scale (Figure 1). For example, a time bin from 6 PM yesterday will be classified as Weakening if the difference between yesterday's wind speed and today's 6 PM wind speed is -15 knots. If wind speed today is 70 knots, we assign the time bin to Category 1. The current category should not be confused with the TC's overall category - unlike the overall TC category, the current category is evaluated for each time bin and not the maximum wind speed overall. Due to the nature of the intensification stage calculations, we also include an Unidentified stage, denoting the last 24 hours of the storm where there is no data to calculate the change in winds needed to attribute the time period to an intensification stage.

Intensification Stage	Change in Winds (Knots) in 24 Hours (Jiang and Ramirez, 2013)
Weakening	<-30 to -10
Neutral	-10 to 10
Intensifying	10 to >30

Fig. 3: Intensification stage definitions, Jiang and Ramirez, 2013

Exploratory Analysis

To better understand the lightning and storm tracking data, exploratory data analysis was performed on the dataset as a whole and by basin. For the overall exploratory analysis, we compared the unfiltered and filtered datasets for all TCs, where the filtered dataset only includes TCs with a maximum sustained wind speed of ≥ 64 knots. For the basin exploratory analysis, we focused on the filtered dataset and compared the 6 basins.

The overall lightning activity analysis explored (1) the total lightning count for each tropical cyclone, (2) the number of storms per year, (3) the average lightning frequency by hour, and (4) the duration of tropical cyclones. Each of these distributions were plotted for both filtered and unfiltered datasets and do not differentiate between basins or years.

We then analyzed wind speed and pressure data for tropical cyclones by: (1) comparing peak and minimum intensity, (2) examining the distribution of each TC's percent change between its most and least intense periods, and (3) investigating the relationship between pressure and wind speed by plotting peak intensity wind speeds against corresponding pressure measurements.

We sanity checked our datasets with the WWLLN website by plotting storm center coordinates and daily lightning density for a few specific storms, Seymour (EPAC_16_20), Hagibis (WPAC_19_20), and Jose (ATL_17_12). For each of the lightning density plots, we also separated out the inner core and rainband lightning to better understand the inner core and rainband lightning regions.

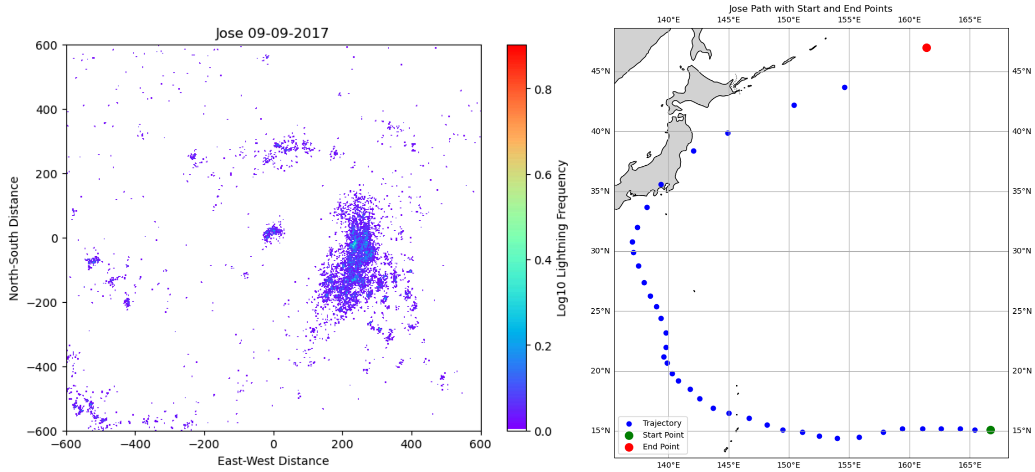

Fig. 4: Lightning density plot for Jose 9/9/2017 (left), storm path for Jose

The plot on the left demonstrates coordinates of lighting strokes with North-South distance from the storm center on the y-axis and East-West distance from the storm center on the x-axis. Each dot represents the log-transformed lightning frequency, where red represents the most frequent and purple represents the least frequent (Figure 4). The plot on the right demonstrates the storm center path with a green dot as a start point, blue dots as the trajectory, and a red dot as the end point (Figure 4). Storm centers were plotted using longitude and latitude as opposed to the relative distances in the lightning density plot.

Moving to the basin-level exploratory analysis, we explored the inner core and rainband lightning separately, and further separated the rainband data by shear quadrant. The dataset includes a total of six basins: the Atlantic Ocean basin (ATL), Central Pacific basin (CPAC), Eastern Pacific basin (EPAC), Indian Ocean basin (IO), Southern Hemisphere basin (SHEM), and Western North Pacific basin (WPAC).

For the inner core basin-level exploratory analysis, we created a boxplot to compare the distribution of storm durations across basins. We looked at the number of lightning strokes associated with each overall TC category, current category, and intensification category for each of the six basins. Since each basin has a different number of storms, the distribution of lightning counts across basins was calculated, weighted by the number of storms, to determine the average number of lightning strokes per storm, where we found the ATL basin to have the highest number of average lightning strokes. We also created a stacked bar chart that displays the distribution of log-transformed lightning counts across the different basins, colored by intensification stage and current category. The analysis was replicated on the rainband data, with adjustments for shear quadrants.

From this basin-level exploratory analysis, we found that storms in the ATL basin have the largest variation in storm duration, while storms in the EPAC, SHEM, and WPAC basins have similar durations. Upon looking at the distribution of the rainband lightning across intensification stages, the ATL basin has most of its lightning in the neutral stage, while the WPAC basin has most of its lightning in the intensifying stage. This aligns with the distribution of time bins by intensification stage - the ATL basin has more time bins in the neutral stage than other stages, and the WPAC basin has more time bins in the intensifying stage. The exploratory analyses performed in this section helped with our understanding of the data and overall storm behavior through exploring overall duration, characteristics of basins, descriptive statistics of lightning stroke counts, wind speed, and pressure. It also provided a comprehensive view of relationships among the intensification stages, lightning count, basins, and shear quadrants before moving on to the lightning burst identification and intensification statistical analysis.

Lightning Burst Identification

We define a lightning burst as a spike in activity relative to the storm's overall lightning, identified by comparing a 30-minute lightning stroke count to other time bins within the same storm. Despite being included in the exploratory analysis, this section only includes lightning data associated with wind speeds greater than 40 knots and time bins with lightning counts greater than 0 in the calculation of lightning burst thresholds. The excluded data points skew the lightning burst thresholds to detect any lightning count greater than zero as a burst. Lightning counts are log-transformed to help reduce skewness and improve visibility before calculating lightning burst threshold values. As such, all lightning count-related statistics used in calculating the lightning burst threshold are on the log-transformed scale.

We use six different lightning burst threshold calculation methods to identify lightning bursts - three different techniques with two thresholds each (Figure 5). Refer to the appendix for more details on each method's definition. The distance measurements were based on the standard norm used to determine values that are sufficiently far from the mean or median.

Method Name	**Definition**	**Threshold Equation**	**Threshold Name**
Median Absolute Deviation	$MAD = median(\|X_i - median(X)\|)$	$median + 4 * MAD$	MAD1
		$median + 5 * MAD$	MAD2
Interquartile Range	$IQR = Q_3 - Q_1$	$Q_3 + 1 * IQR$	IQR1
		$Q_3 + 1.5 * IQR$	IQR2
Log-normal Standard Deviations	$SD = e\^(\mu + (\sigma^2 / 2)) * \sqrt{}(e\^(\sigma^2) - 1)$	$\mu + 2 * \sigma$	LOGN1
		$\mu + 3 * \sigma$	LOGN2

Fig. 5: Table displaying the 6 threshold calculation methods

Before calculating burst thresholds, we log-transform the lightning counts, replace invalid pressure values with None, and create a column for the 3-level intensification stages. Initial analysis included time bins with zero lightning events in the threshold calculations, but it was found that their inclusion skews the threshold to zero, causing any lightning count greater than zero to be flagged as a burst. As this is not realistic, we chose to exclude time bins with zero lightning events from subsequent analysis. We then applied the 6 threshold calculation methods to each TC's inner core and rainband lightning data separately, excluding time bins with wind speeds less than 40 knots and time bins with zero lightning count. We plotted a few specific TCs to view the detected bursts (Figure 6, 7) and export the data for the Power BI Dashboard and intensification analysis. For rainband datasets, we also plotted the lightning activity and detected bursts by shear quadrant. Each visualization's background is color-coded by intensification stage or current category to provide context for each detected lightning burst.

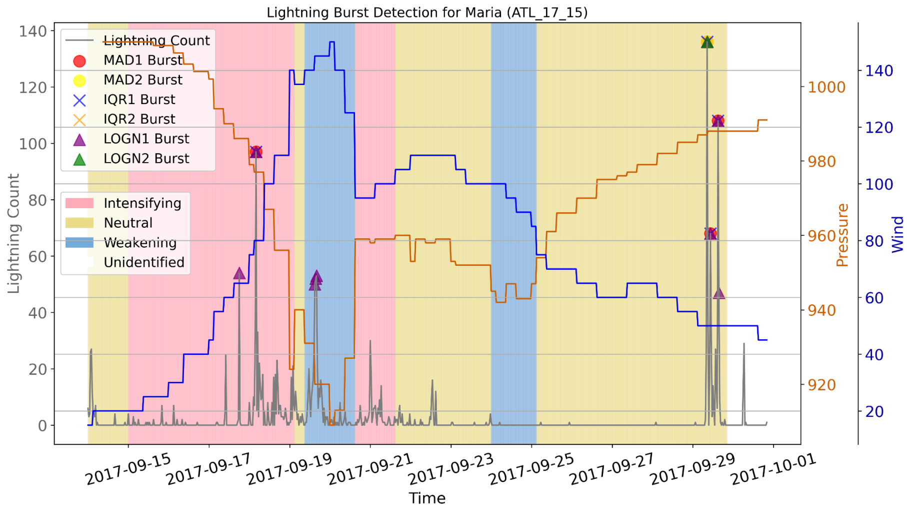

Fig. 6: Maria's inner core lightning and detected bursts, background colored by intensification stage

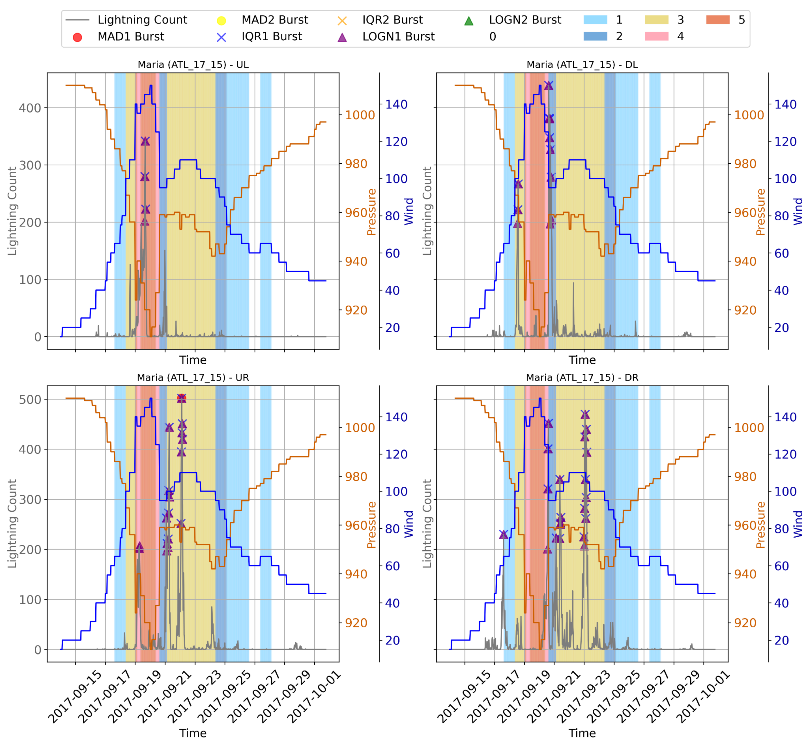

Fig. 7: Maria's rainband lightning and detected bursts by shear quadrant, background colored by current category

We calculate a basin-level lightning burst threshold for the inner core lightning, where we apply the same basin-level threshold value to all storms in the basin. Each TC's time bins were split by current category into three category groups (0-2, 1-2, 3-5) to calculate the basin's mean, median, minimum and maximum threshold values, as well as the standard deviation, number of time bins evaluated, and number of time bins identified as a burst for each threshold method. We then looked at "effective" thresholds for each basin - thresholds that identify at least one burst. These effective thresholds are used to calculate a basin-level threshold for each category group (0-2, 1-2, 3-5) and applied to the basin to identify bursts. The CPAC basin is not included in basin-level analysis due to data sparsity.

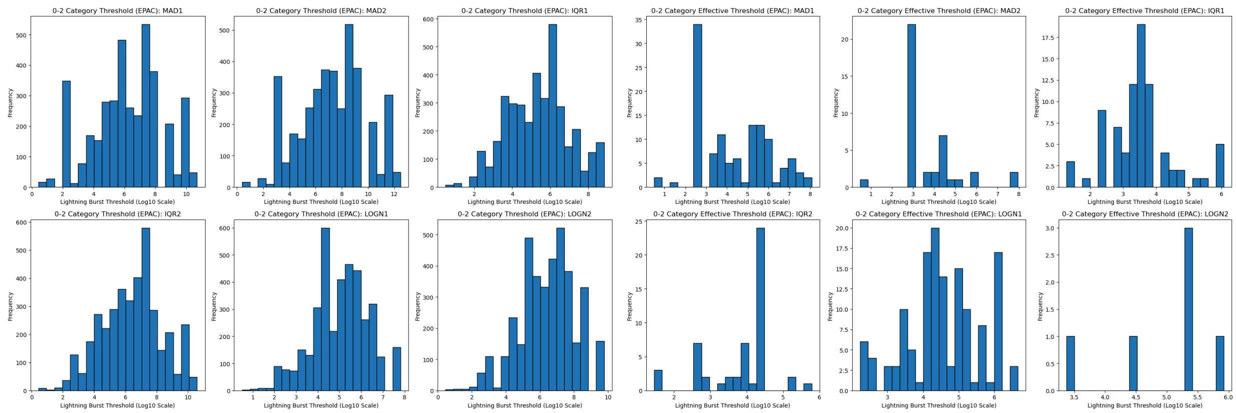

Fig. 8: EPAC Categories 0-2 threshold distributions (left), effective threshold distributions (right)

When all thresholds were included in calculating the basin-level threshold value, the value was too high to detect any bursts. As such, using effective thresholds to calculate the basin-level burst threshold helps bring the basin-level value down and detect bursts across various storms. Further investigation is needed to validate these findings and apply the methods to rainband data.

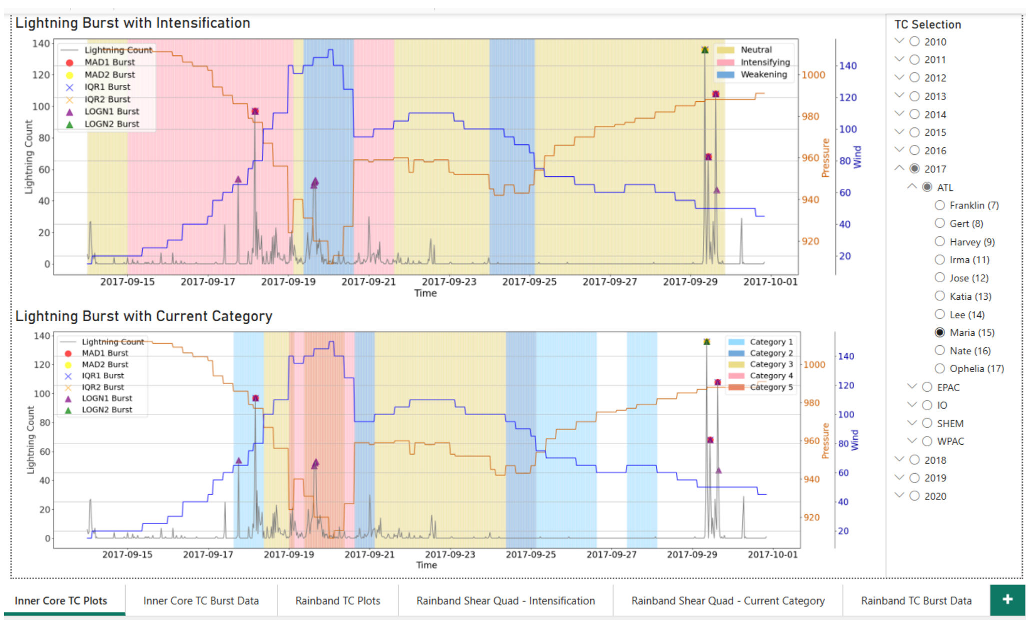

Fig. 9: Power BI dashboard

To facilitate easy viewing of each TC's lightning activity, detected bursts, and intensification data, we created a Power BI dashboard (Figure 9). The dashboard allows users to select a tropical cyclone and view plots for both inner core and rainband datasets. It also includes a tabular view to explore detailed lightning counts associated with each time bin and whether that time bin was detected as a burst by any of the six detection methods.

We identified lightning bursts using six different lightning burst threshold methods and created a Power BI dashboard showing lightning activity, identified bursts, with background coloring based on the time bin's intensification stage or current category. All data is plotted regardless of whether or not the point was included in the lightning burst threshold calculation. Each threshold calculation method is marked with a different color and shape. The left y-axis indicates lightning counts, the right y-axes represent wind and pressure, and the x-axis plots timestamps in year-month-day format.

Lightning Burst Intensification Analysis

The following section presents statistical analyses of lightning bursts in relation to tropical cyclone intensification stages. We investigate the relationship between lightning activity and tropical cyclone intensification stages using various lightning burst detection methods to examine distribution patterns. The data used in this analysis consist of inner core and rainband lightning bursts categorized by intensification stage: Intensifying, Neutral, Weakening, and Unidentified. The Unidentified stage represents the last 24 hours of a tropical cyclone, where we no longer have the data to calculate the forward-looking 24-hour change in wind speed used to classify storm intensification stage. This section will only include detection methods IQR1, IQR2, and LOGN1. We choose to exclude both MAD methods from this section due to its high variability, as well as the LOGN2 method due to the sparsity in detected bursts.

A Chi-Squared test was conducted to assess whether there is a significant difference in lightning burst counts across different intensification stages. The results indicate a statistically significant difference ($p<0.05$) for both the inner core and rainband regions, suggesting that lightning bursts vary by cyclone intensity stage. However, when the dataset was stratified by basin, some basins exhibited no significant differences for certain detection methods. For example, in the Central Pacific (CPAC) basin, inner core lightning did not show a statistically significant difference. To further investigate, we analyzed the data by current categories and found that lightning burst distributions are strongly influenced by the current category. Specifically, in the inner core region, time bins with current Category 1 and 2 exhibited fewer statistically significant differences. To refine our analysis, we grouped the categories into two broader groups: Categories 0–2 and 3–5. The statistical results for both the inner core and rainband regions for the two category groups demonstrated a significant difference ($p<0.01$) in lightning burst distribution across intensification stages.

We used heatmaps to visualize the distribution of burst counts across categories, reinforcing the trends identified in the statistical analysis. The heatmap indicates that most lightning activity occurs during the intensification stage when considering all lightning data for both the inner core and rainband regions. A similar trend is observed at the basin level, with the exception of the CPAC and EPAC basins in the inner core region and the ATL basin in the rainband region. The TC category-level heatmap reveals that Categories 0, 1, and 2 experience more lightning bursts during intensification than during other stages for both inner core and rainband regions. Category 2 in the inner core shows a more even distribution of lightning bursts between the intensification and weakening stages. Categories 3, 4, and 5 exhibit a similar pattern, with more bursts occurring during the weakening stages for the inner core and rainband. To further validate this trend, we looked at the grouped Categories 0–2 and 3–5.

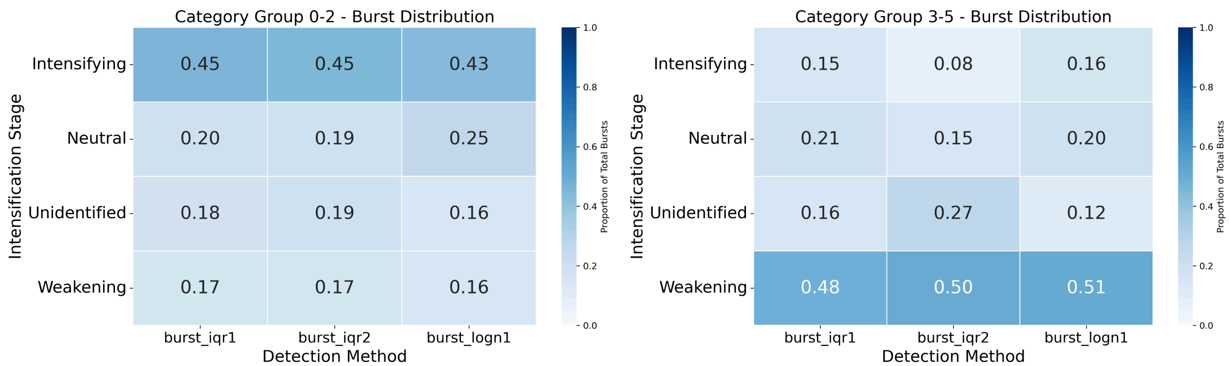

Fig. 10: Heatmaps for inner core lightning analysis, Categories 0-2 (left), Categories 3-5 (right)

The heatmaps in Figure 10 clearly show that lightning bursts are more frequent during the intensification stage for Categories 0–2 and during the weakening stage for Categories 3–5 for the inner core and rainband.

In conclusion, the result confirms that lower-category time bins exhibit the highest lightning burst activity primarily during the Intensifying stage. Higher-category time bins experience more bursts during the Weakening stage. The log-normal 2 sigma method (LOGN1) demonstrates exceptional sensitivity in comparison with other methods

Limitations and Future Work

This study has several limitations. Joining the six-hour interval track file data with higher-resolution WWLLN measurements (≤1 hour) introduces potential sources of error. Burst identification may be less effective for extreme, unpredictable, or small storms, as well as storms that make landfall. The WWLLN dataset itself contains inherent spatial and temporal detection biases due to reliance on ground-based sensors, potentially skewing lightning activity quantification in remote oceanic regions. Collectively, these limitations may propagate uncertainties in establishing robust relationships between lightning variability and cyclone dynamics.

Future work will focus on enhancing lightning burst identification by excluding landfall data, as landfall is often associated with increased lightning activity in tropical cyclones. The threshold analysis will need to exclude lightning strikes associated with storm centers within 100 km of land. Basin-level threshold calculations will also be performed on the rainband data, as this project only calculated basin-level thresholds for inner core data. We expect more cohesive trends for basin-level thresholds in the rainband data. Additionally, conducting the same analysis using Geostationary Lightning Mapper (GLM) image data may provide deeper insights into lightning burst behavior. Furthermore, the integration of machine learning techniques will be explored to systematically detect lightning bursts, enabling their application in future predictive models.

Acknowledgements

Thank you to our sponsors, Dr. Jeremy Thomas and Dr. Natalia Solorzano at NWRA, for your guidance and help in making this project successful.

WWLLN Data Acknowledgment - The authors are using data from the World Wide Lightning Location Network, a collaborative consortium of over 70 worldwide collaborators, managed at the University of Washington.

References

Anon (n.d.), Saffir-Simpson Hurricane Wind Scale. Available from: https://www.nhc.noaa.gov/aboutsshws.php (Accessed 15 March 2025)

Cecil, D.J. and E.J. Zipser 2002b: Reflectivity, ice scattering, and lightning characteristics of hurricane eyewalls and rainbands. Part II: Intercomparison of observations. Mon. Wea. Rev., 130, 785–801.

Corbosiero, K. L., and J. Molinari, 2002: The effects of vertical wind shear on the distribution of convection in tropical cyclones. Mon. Wea. Rev., 130, 2110–2123, https://doi.org/10.1175/1520-0493(2002)130<2110:TEOVWS>2.0.CO;2.

DeMaria, M., and J. Kaplan, 1994: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic Basin. Wea. Forecasting, 9 , 209–220.

Fierro, A. O., S. N. Stevenson, and R. M. Rabin, 2018: Evolution of GLM-Observed Total Lightning in Hurricane Maria (2017) during the Period of Maximum Intensity. Mon. Wea. Rev., 146, 1641–1666, https://doi.org/10.1175/MWR-D-18-0066.1.

Jiang, H., and E. M. Ramirez, 2013: Necessary Conditions for Tropical Cyclone Rapid Intensification as Derived from 11 Years of TRMM Data. J. Climate, 26, 6459–6470, https://doi.org/10.1175/JCLI-D-12-00432.1.

Solorzano, N. N., J. N. Thomas, and C. Bracy (2018), Monitoring tropical cyclones with lightning and satellite data, Eos, 99, https://doi.org/10.1029/2018EO092439.

Appendix

Definitions

ATL - Atlantic basin
CPAC - Central Pacific basin
Current category - category defined by time bin wind speeds
Effective thresholds - thresholds that detect at least one time bin as a burst
Inner core - within 100 km of storm center
IO - Indian Ocean basin
IQR - Interquartile Range
LOGN - Log-normal
MAD - Median Absolute Deviation
Overall TC category - a TC's category defined by its maximum wind speed
Rainband - 200-400 km of storm center
SHEM - Southern Hemisphere basin
TC - tropical cyclone
WWLLN - World Wide Lightning Location Network
WPAC - Western North Pacific basin

Lightning Burst Threshold Calculation Methods:

Method Name	**Definition**	**Variables**
Median Absolute Deviation	$MAD = median(\|X_i - median(X)\|)$	X_i - the data points $median(X)$ - median of the dataset $\|X_i - median(X)\|$ - absolute deviations from the median
Interquartile Range	$IQR = Q_3 - Q_1$	Q_1 - first quartile (25th percentile) Q_3 - third quartile (75th percentile)
Log-normal Standard Deviations	$SD = e\hat{}(\mu + (\sigma^2 / 2)) * \sqrt{(e\hat{}(\sigma^2) - 1)}$	μ = mean of the associated normal distribution σ = standard deviation of the associated normal distribution e = Euler's number (≈ 2.718)