Los Alamos NATIONAL LABORATORY

A 30-Minute Lightning Flash? Clustering and Counting Flashes in Very High Flash Rate Thunderstorms

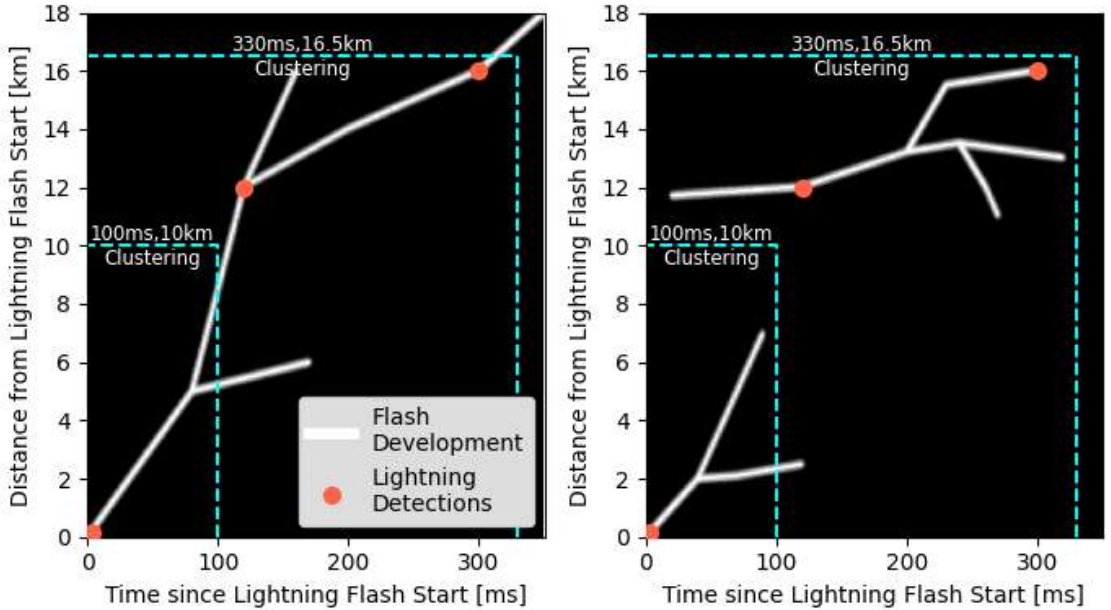
Michael J. Peterson

11/13/2023

LA-UR-23-31998

Flash Clustering is Highly Sensitive to the Chosen Thresholds in High Flash Rate Thunderstorms

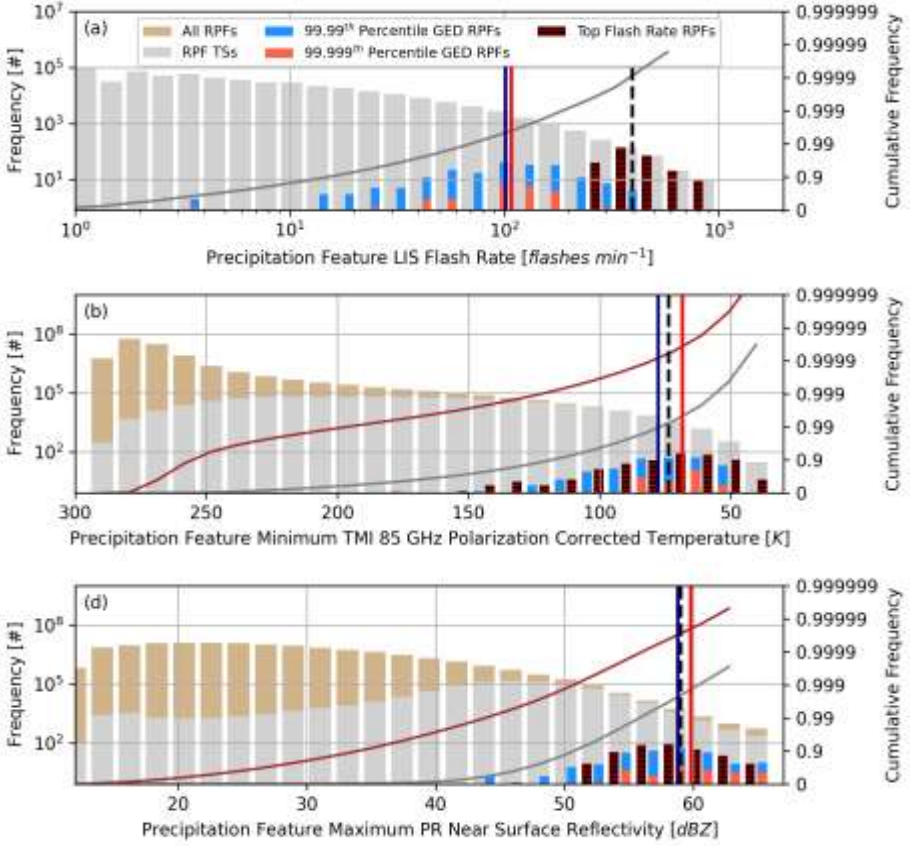
- Long-range and satellite lightning sensors detect sparse events (red dots) on portions the evolving lightning channels (white lines)
- Lightning detectors try to identify distinct flashes by assuming that close events are in the same flash
  - This assumption is usually valid, but it breaks in high flash rate storms and/or with low-sensitivity detectors
  - Flash counts and properties also depend on the space / time thresholds used in flash clustering



- Flash clustering will cause some flashes to be artificially divided (as in the restrictive clustering in the left panel) and certain clusters of multiple flashes to be artificially merged (as in the loose clustering in the right panel). The chosen approach determines the frequency of either issue

The Most Intense LIS / GLM Thunderstorms have Low Flash Rates from Artificial Flash Merging

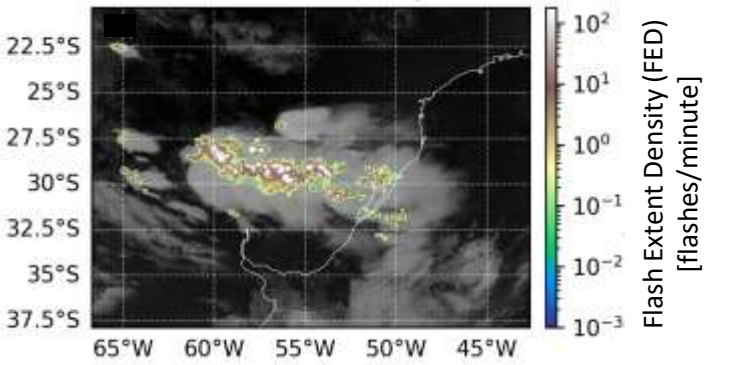
- At low flash rates, artificial merging occurs infrequently in random flashes
- As local thunderstorm flash rates increase to near 1/(clustering time threshold), artificial merging becomes assured
- Group Extent Density (GED) is not affected by artificial merging
  - The top GED thunderstorms in the TRMM data (red, blue at right) have low LIS flash rates (top)
  - Yet, they include the most intense storms in the TRMM Microwave Imager (center) and Precipitation Radar (bottom) data that rival the top LIS flash rate storms (black)
  - The flash rates in top GED storms are being artificially suppressed by flash merging

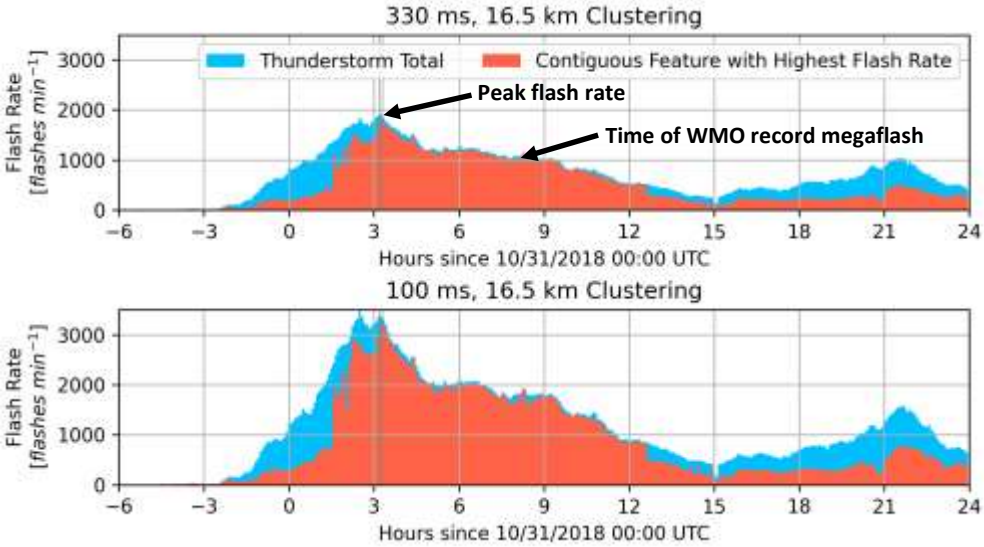


Los Alamos NATIONAL LABORATORY

Clustering Threshold Sensitivity in the Highest Flash Rate GLM Thunderstorm

- To directly compare clustering performance in high flash rate thunderstorms, we created “virtual network” data by merging GLM and ENGLN lightning detections
- Two sets of clustering thresholds are considered:
  - 330 ms, 16.5 km
  - 100 ms, 16.5 km
- The top flash rate storm was a Mesoscale Convective System (MCS) over Argentina on October 31st 2018
  - Convective flash rates were suppressed under GLM clustering, yielding a peak of 1902 flashes/min
  - Applying the shorter time threshold increases the peak flash rate to 3470 flashes/min

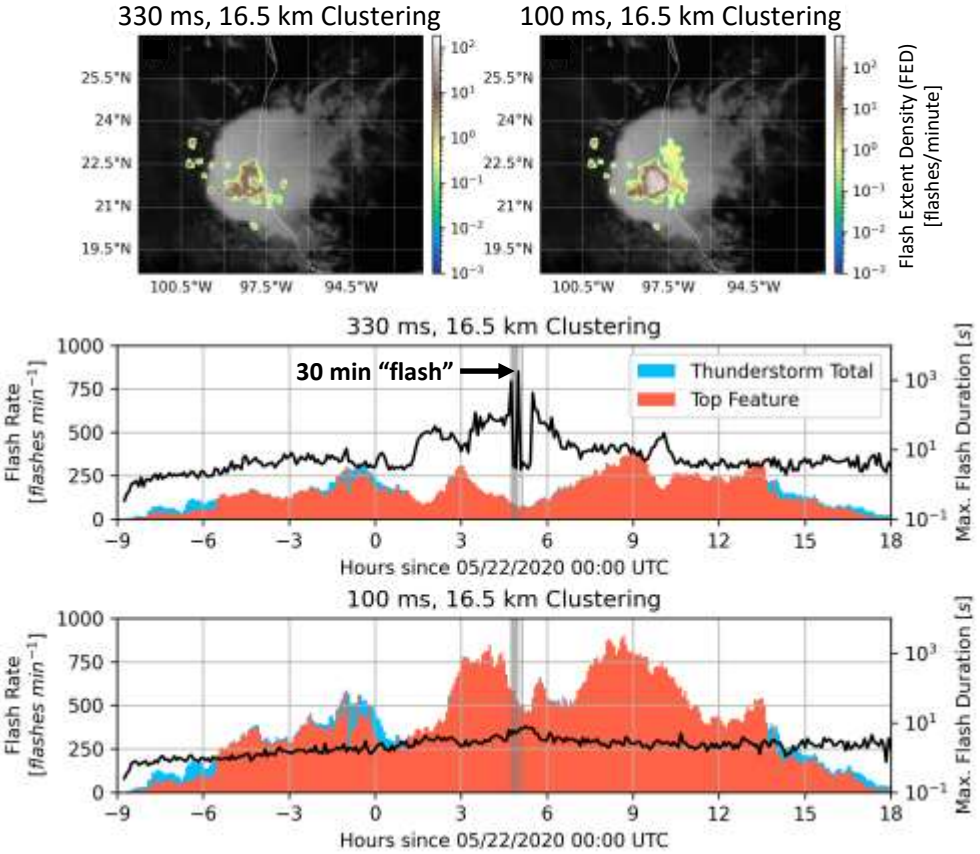




Los Alamos NATIONAL LABORATORY

Severe Clustering Failures Reduce a Thunderstorm to a Single Flash Spanning a 30 Minute Period

- The worst case of artificial flash merging occurred in May 22nd 2020 along the Mexican gulf coast
- All of the virtual network detections in the central updraft over a ~30-minute period satisfied the GLM clustering thresholds, yielding a single flash (top left)
  - The GLM clustering model works well for general convective lightning and megaflashes
  - 330 ms is too broad for very high flash rate storms
  - The shorter 100-ms clustering (top right, bottom) produced a more reasonable peak flash rate of 900 flashes/minute
    - **It still overclustered the lightning detections during the 30-minute "flash," artificially suppressing the thunderstorm flash rate**

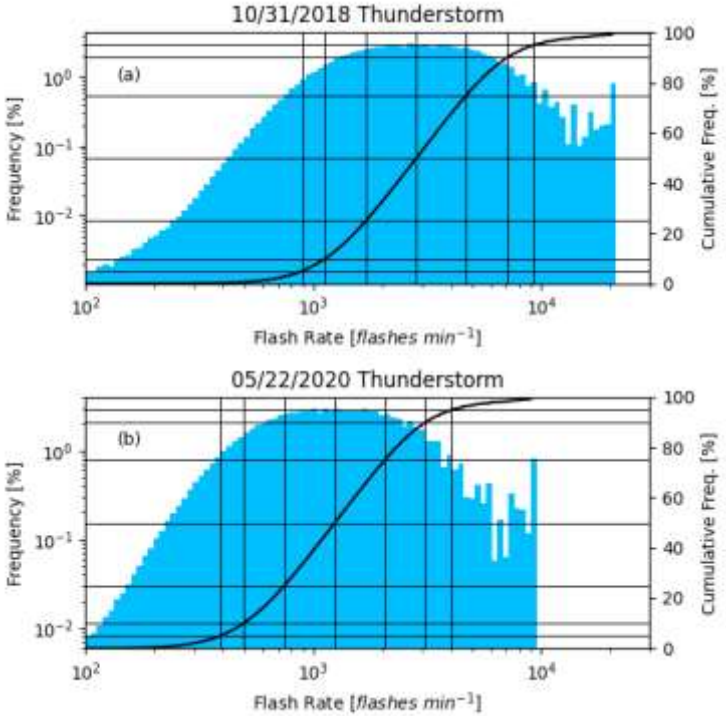


Los Alamos NATIONAL LABORATORY

Alternate Approaches for Counting Flashes in Very High Flash Density Thunderstorms

- Three alternate approaches have been proposed to provide realistic flash rates / flash rate trends for very high flash rate storms
  - Completely redefine how we cluster lightning data (i.e., using Machine Learning, etc.)
    - Pros: may lead to a one-size-fits-all solution
    - Cons: highly experimental, needs robust validation
  - Apply a dynamic clustering time threshold that varies by flash rate (proposed by Doug Mach)
    - Pros: tackles the problem directly
    - Cons: need to know flash rate to set threshold, dynamic threshold model still needs validation to ensure the flash rates are physical
  - Use a statistics-based approach to estimate flash rates / trends from thunderstorm group rates and GLM energy thresholds (right)
    - Pros: leverages existing work, variability primarily arises from electrified cloud type rather than storm intensity
    - Cons: currently only applies to GLM, not ground or virtual networks

Flash Rate Probabilities from GLM Group Rates and Energy Thresholds



Los Alamos NATIONAL LABORATORY

Conclusion

- **Very high flash rate thunderstorms break LIS / GLM flash clustering**
  - Detections become “overclustered” into amalgamated flashes that merge multiple distinct lightning flashes into a single “flash” feature in the data
  - GLM additionally chops these amalgamated flashes into an arbitrary number of 101-group pieces, obscuring the problem in the operational data
- **When clustering is broken by these rare cases of extreme thunderstorms, flash characteristics, flash rates, and flash rate trends become unreliable**
  - This degrades the quality of GLM products that are used both in operations and scientific research
- **Mitigating these issues requires handling the data in a consistent manner**
  - Effort is needed to develop a clustering strategy that can handle both low flash rate stratiform clouds producing megaflashes and high flash rate convective clouds generating thousands of localized discharges
    - A number of promising clustering strategies have been proposed, but they require validation
    - Alternatively, avoiding clustering entirely and deriving thunderstorm flash rates / trends from group rate statistics is also proving useful

Questions?