Validation and Sensitivity Testing of a Lightning Grouping Algorithm on Thunderstorm Climatologies

Tobias Gerken[1], Hayden Abbot[1], Chelsea Lang[1], Dudley Bonsal[1], Zhoujun Duan[2], Henry Way[1], Mace Bentley[1]

(1) School of Integrated Sciences, James Madison University, Harrisonburg, VA, USA
(2) Department of Computer Science, James Madison University, Harrisonburg, VA, USA

Background:

Air pollution has the potential to modify the lifespan and intensity of thunderstorms and properties of lightning. By modulating cloud and rainfall processes, low concentrations of air pollutants can promote shallow convection with limited lightning activity. Higher air pollutant concentrations up to a threshold, can promote more vertical transport of droplets and increases in lightning production. Using cloud-to-ground (CG) flash data from the U.S. National Lightning Detection Network (NLDN; Vaisala Inc) we apply a thunderstorm tracking algorithm[1] using spatio-temporal clustering to group lightning into events.

Methods:

1. We extracted warm season (May - Sep) lightning flashes in a 250 km radius surrounding Washington, DC. flash data for this analyses were obtained from the U.S. National Lightning Detection Network (NLDN; Vaisala, Inc.) for the years 2006–2020.
2. Lightning grouping algorithm[1] was applied using standard parameters yielding a total of 74328 events with ≥10 flashes considered thunderstorm initiations (Table 1).
3. The algorithm has 3 tunable parameters for the event detection:
   1. Maximum flash distance: Δr default value: 15 km
   2. Maximum flash delay: Δt default value: 15 min
   3. Time distance tradeoff (TLD) $\Delta r \times \Delta t$ default value: 150 km min
4. Two sensitivity experiments were conducted:

TLD-Experiment:	Δr, Δt = default;	TLD = [100,125,150,175]
Distance/Delay Experiment:	Δr = 15 km, Δt = 20 min Δr = 17.5 km, Δt = 15 min Δr = 20 km, Δt = 15 min Δr = 20 km, Δt = 20 min	TLD = 200

Research Questions:

- What is the sensitivity of the grouping algorithm to changes in tunable maximum flash distance (Δr), delay (Δt), and time distance parameter ($\Delta r \times \Delta t$)?
- What meteorological conditions impact the sensitivity response?
- Can we find optimal values for these parameters?
- Do detected flash groupings correspond to observed convective systems?

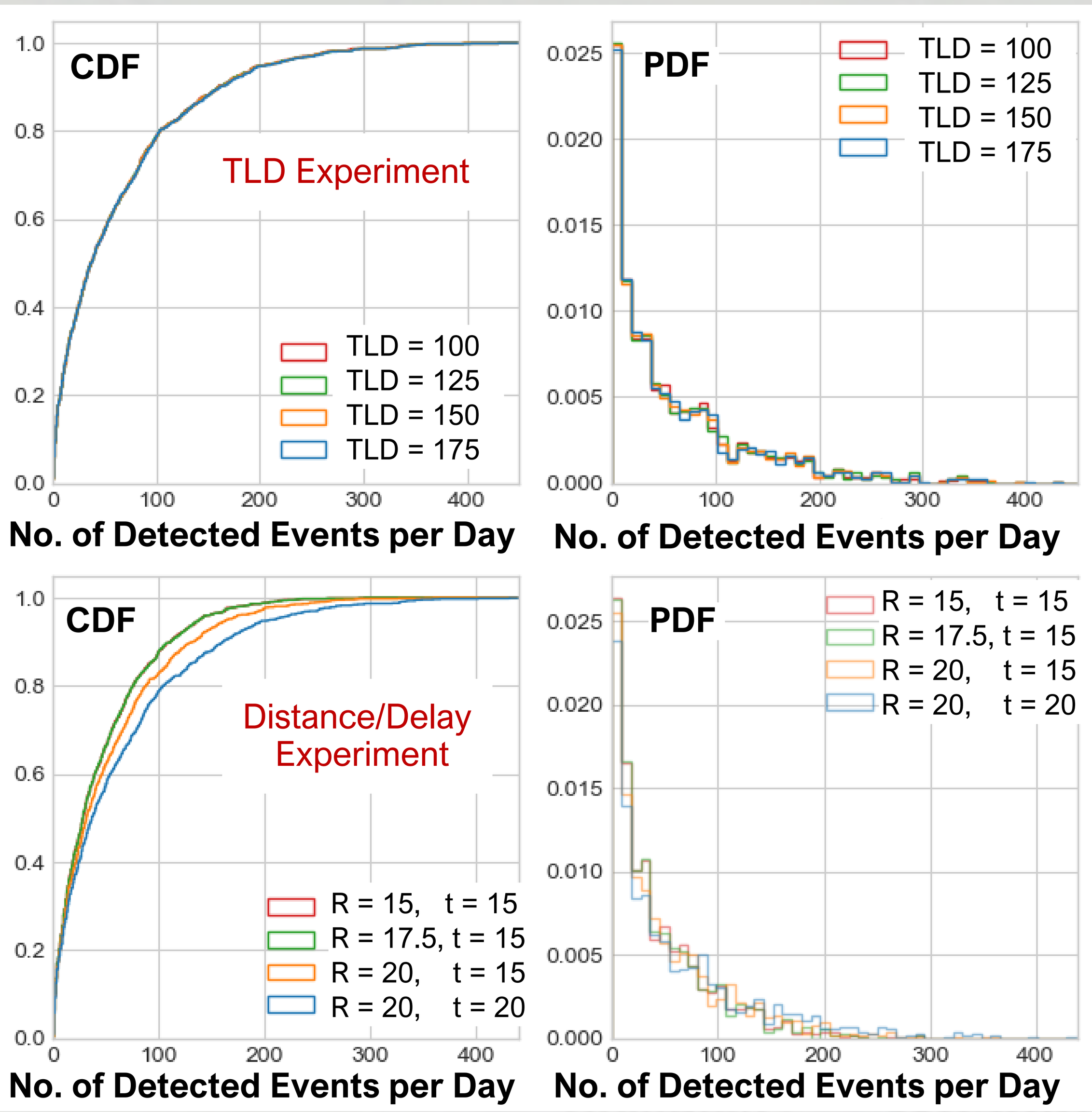

Figure 1: Change in number of detected events per day for the TLD experiment (upper row) and Distance/ Delay Experiment (lower row).

Results reveal that the lightning detection algorithm shows little sensitivity to the TLD parameter.

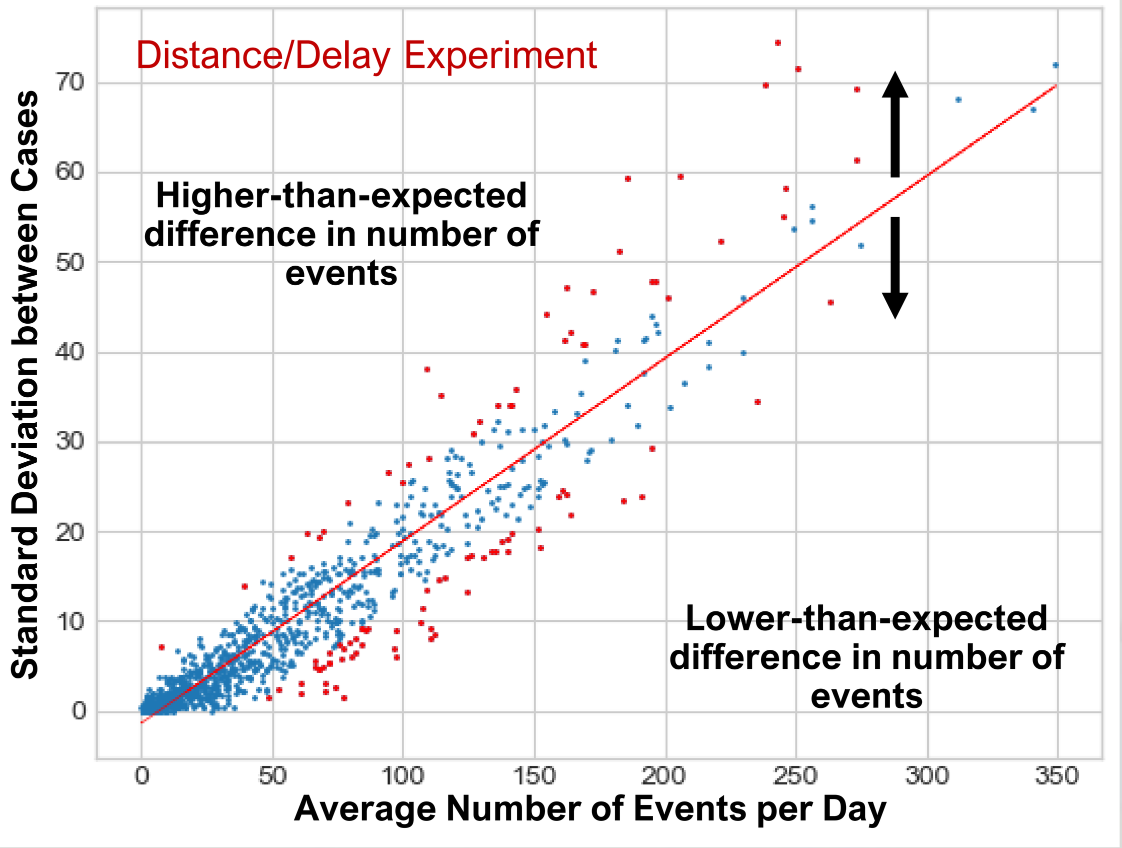

***Figure 2 (left):** Focusing on the Distance/ Delay Experiment we find a linear relationship between the total number of detected lightning events per day (x-axis) and the standard deviation of detected events per day between cases (y-axis).*

Interestingly, we find both large positive and negative deviations from the relationship (highlighted in red). These days may help explain the reason for the event detection algorithm sensitivity.

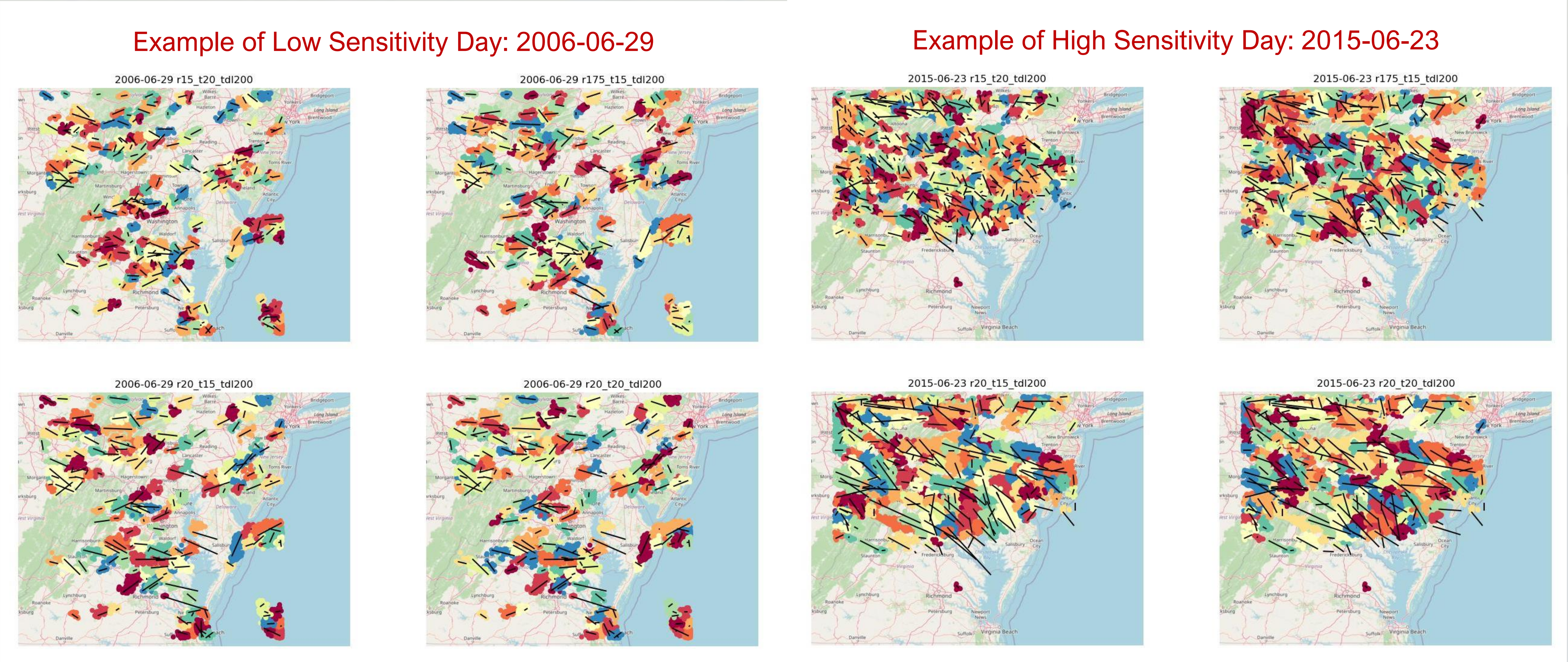

***Figure 3 (top):** Using the results of Figure 2, we show an example of a low sensitivity day (left panels) with lower-than-expected difference in detected thunderstorm event numbers and a high sensitivity day (right panels) with higher-than-expected changes in event numbers. Each dot represents a lightning flash. Grouped flashes of the same event are given the same color. Black solid lines show the distance between first and last flash location for each event.*

A main difference between the two days is the spatio-temporal clustering of thunderstorm events with the low sensitivity day exhibiting much clearer event separation and smaller scale thunderstorms.

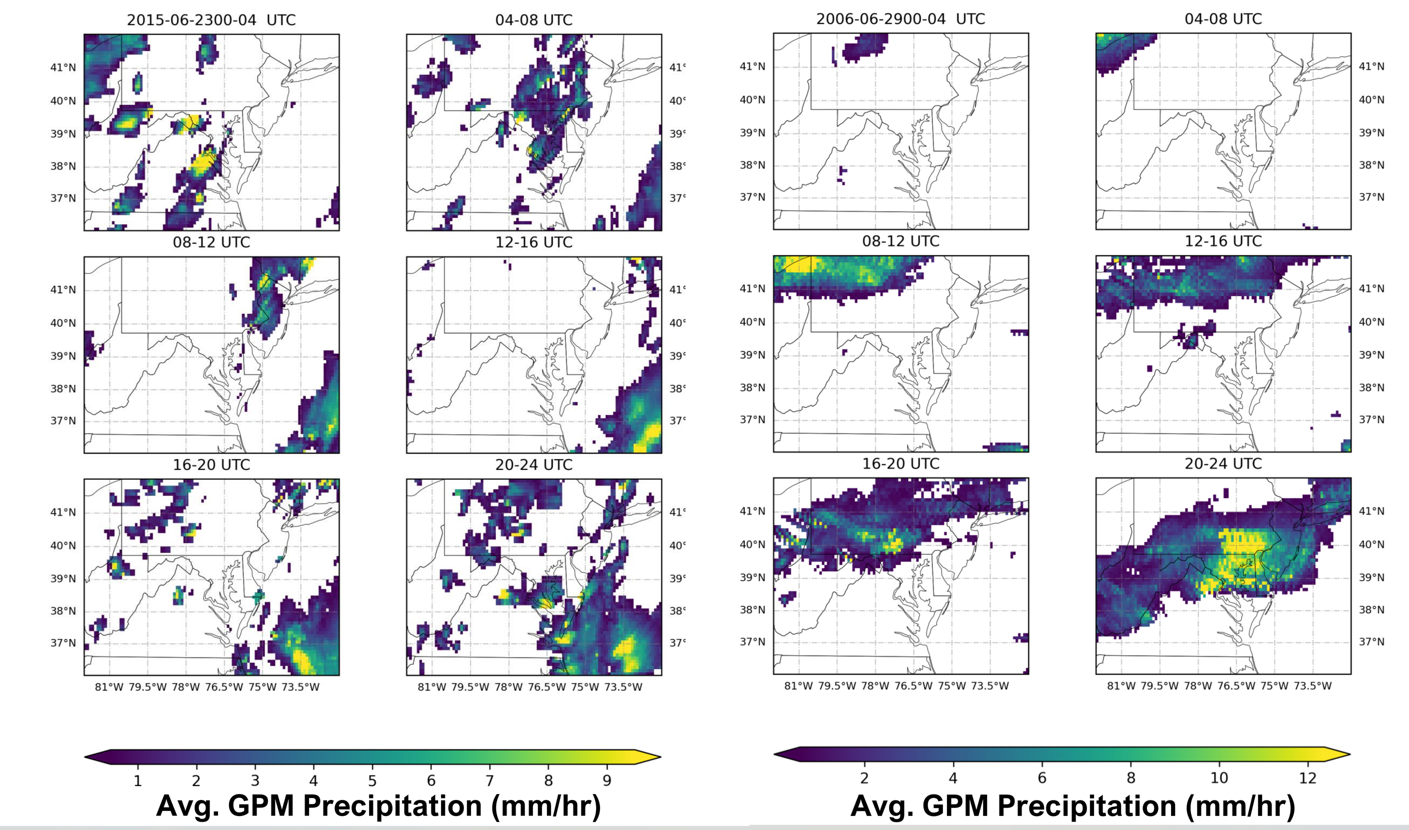

***Figure 4 (top):** Four hour averaged NASA Global Precipitation Measurement (GPM) calibrated precipitation rate (mm/hr) for low (left) and high (right) sensitivity days.*

In addition to spatio-temporal structure, there is also a mid-day precipitation pause for the low sensitivity day, highlighting the role of thunderstorm event timing for algorithm sensitivity.

Table 1: Lightning event climatology near Washington DC with 74328 detected thunderstorm events between (2006 and 2020) for baseline algorithm using suggested parameters for flash distance (Δr) and delay (Δt)

Month	Lightning Strokes (%)	Thunderstorm Events (%)	Avg. Flashes per Event	Max. Number Flashes	Avg. Duration (min)	Max Duration (min)
May	11.6	15.5	69	2474	43	265
Jun	25.3	28	83	3445	45	345
Jul	31	28.7	100	3877	46	335
Aug	24.2	20.6	108	7289	46	380
Sep	7.9	7.1	102	3840	45	270

Conclusions & Future Work:

- The lightning event grouping algorithm was found to be primarily sensitive to distance and delay cutoffs rather than the TLD parameter
- Algorithm sensitivity is dependent on synoptic conditions with high parameter sensitivity during days with persistent and tightly clustered lightning flashes.
- Additional research is needed to validate the algorithm against specific thunderstorm events and to determine optimal parameters for Δr and Δt.

[1] Tuomi, Tapio J., and Markku Larjavaara. "Identification and Analysis of Flash Cells in Thunderstorms." Quarterly Journal of the Royal Meteorological Society 131(607), (2005): 1191–1214. https://doi.org/10.1256/qj.04.64.

Acknowledgement: This work is funded through the National Science Foundation Award # 2104299 - *RUI: Characterization of Aerosol Effects within an Urban Lightning Climatology across a North American Climatic Gradient*