Large-scale, automated detection of gigantic jets using machine learning and sensor fusion

Georgia Tech Research Institute
Severe Storms Research Center

Georgia Tech Research Institute

Levi Boggs[1], Morris Cohen[2], Douglas M Mach[3], Steve Cummer[4], John M Trostel[1], Jeffery Burke[1], Jeffrey Claiborne Smith[5], and Michael Peterson[1]

1: GTRI, Smyrna, GA USA | 2: Georgia Tech, Atlanta, GA USA | 3: USRA, Huntsville, AL USA | 4: Duke, Durham, NC USA | 5: SETI, Mountain View, CA USA

Abstract

Gigantic jets (GJs) are a type of transient luminous event (TLE) which also includes sprites, elves, halos, and blue jets [Pasko 2010, doi: 10.1029/2009JA014860]. However, GJs are unique in that they directly couple electric charge reservoirs in the troposphere (i.e. thunderclouds) with the lower ionosphere, allowing significant amounts of charge (100s of C) to flow between these regions. We do not understand how this affects the ionosphere and global electric circuit. Past observations are very limited, resulting from ground-based cameras getting lucky enough to capture an event while looking over a distant thunderstorm [Liu et al. 2015, doi: 10.1038/ncomms6995]. Additionally, GJ-producing storms are normally accompanied by substantial areas of stratiform clouds obscuring the view, and they tend to occur more often over the ocean. To solve this problem of limited detection capability, we have developed a pipeline that utilizes machine learning and sensor fusion of multiple sensing modalities (optical, VLF, ELF). Our pipeline can detect GJs over nearly a hemisphere and operate 24/7, potentially revolutionizing how GJs are detected and paving the way for other TLE and unique lightning event detection.

Our pipeline begins by performing detection with data from the Geostationary Lightning Mapper (GLM), which is a staring optical imager in geostationary orbit that detects the 777.4 nm (OI) triplet from lightning leaders [Goodman et al. 2013, doi: 10.1016/j.atmosres.2013.01.006]. Gigantic jets have unique signatures in the GLM data from past studies [Boggs et al. 2019, doi: 10.1029/2019GL082278]. We have developed a supervised, ensemble machine learning classifier that detects potential gigantic jets in the GLM data. Considering we have an imbalanced dataset, we use data imbalance techniques such as Synthetic Minority Oversampling Technique (SMOTE) when training the classifier. Next, we combine data from multiple sensing modalities to vet the candidate GJs from the classifier in multiple stages. The first stage filters the candidate GJs to the stereo GLM region [Mach and Virts, 2021, doi: 10.1175/JTECH-D-21-0078.1], and calculates the stereo altitudes for all the events. GJs have stereo altitude sources consistently between 15-25 km altitude from the leader escaping the cloud top [Boggs et al. 2022, doi: 10.1126/sciadv.abl8731]. Next, we match the events spatiotemporally to GLD360 data to remove cloud-to-ground events. Subsequently, we use a statistical GOES ABI model (developed at GTRI) to filter out events that have differing parent storms from our truth database. Finally, we use a multi-parameter extremely low frequency (ELF) vetting model (developed by Duke) to filter out the remaining non-GJ events. After a few complete detection and vetting cycles, we have found tens of new events with a high degree of confidence. With further development of our pipeline and deployment to the entire GLM field-of-view (not limited to stereo region), we anticipate hundreds of new detections per year, significantly more than ground-based cameras, allowing for comprehensive studies relating gigantic jets to the other atmospheric phenomena.

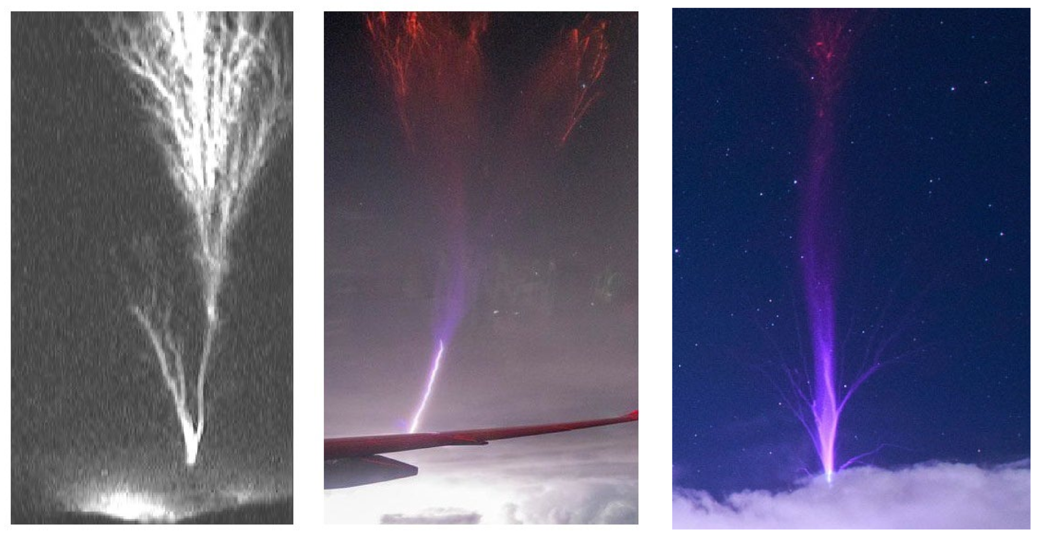

Figure 1. Gigantic Jets captured from ground-based cameras. (left) From *Boggs et al. 2019* (middle) from passenger on an airliner (right) from *Yang et al. 2020*.

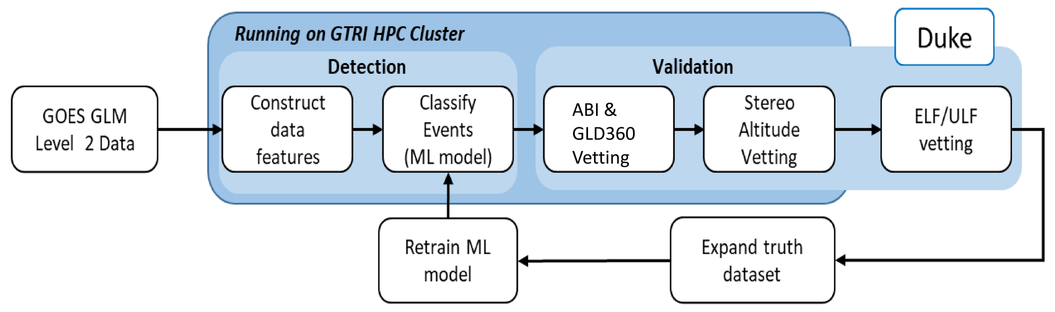

Figure 2. Detection pipeline for this project, consisting of GOES GLM data, GLD360 data, ABI data, and ELF data.

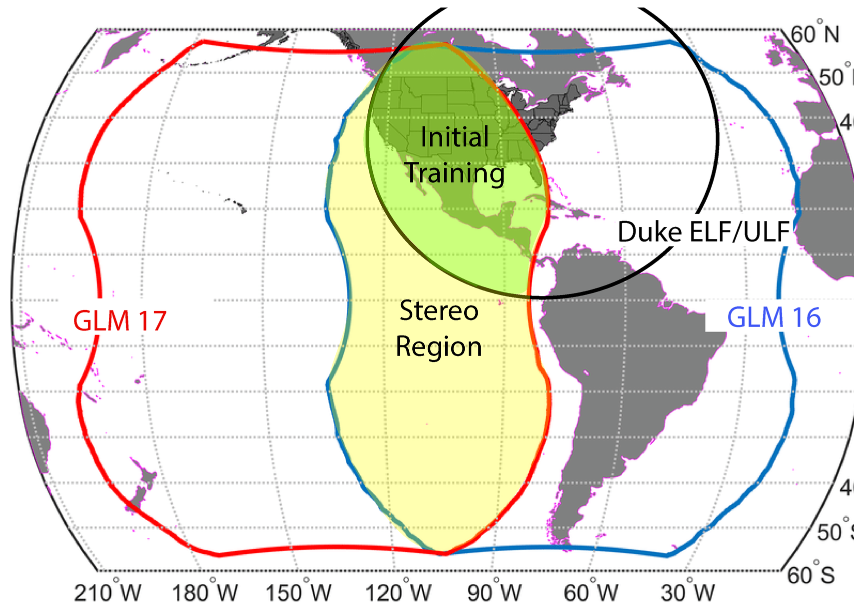

Figure 3. Detection system field-of-view, with regions for initial training and the GLM stereo region.

ML Model

GLM Data

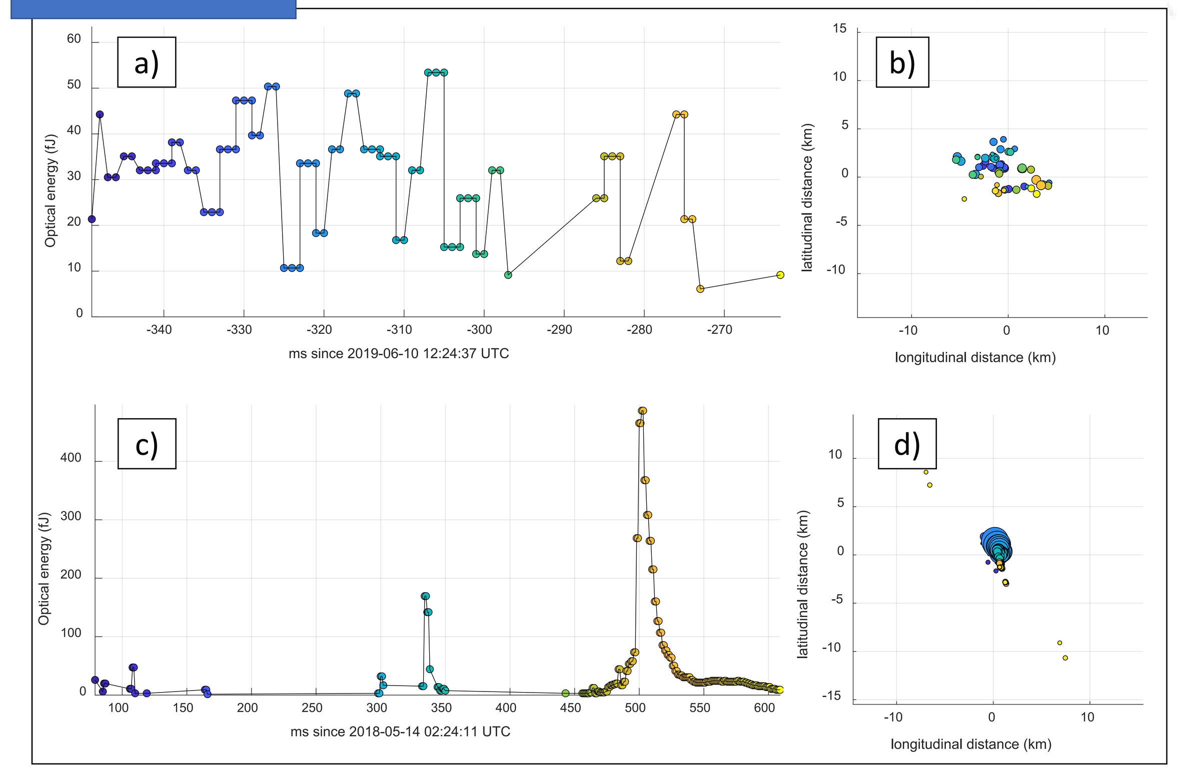

Figure 4. a) lightcurve (GLM energy vs. time) for a non-GJ flash. b) spatial (lat,lon) distribution for the sources in a). Panels c) and d) show similar plots for a **confirmed GJ.** The sources in b) and d) are sized according to energy.

GLM Stereo Data

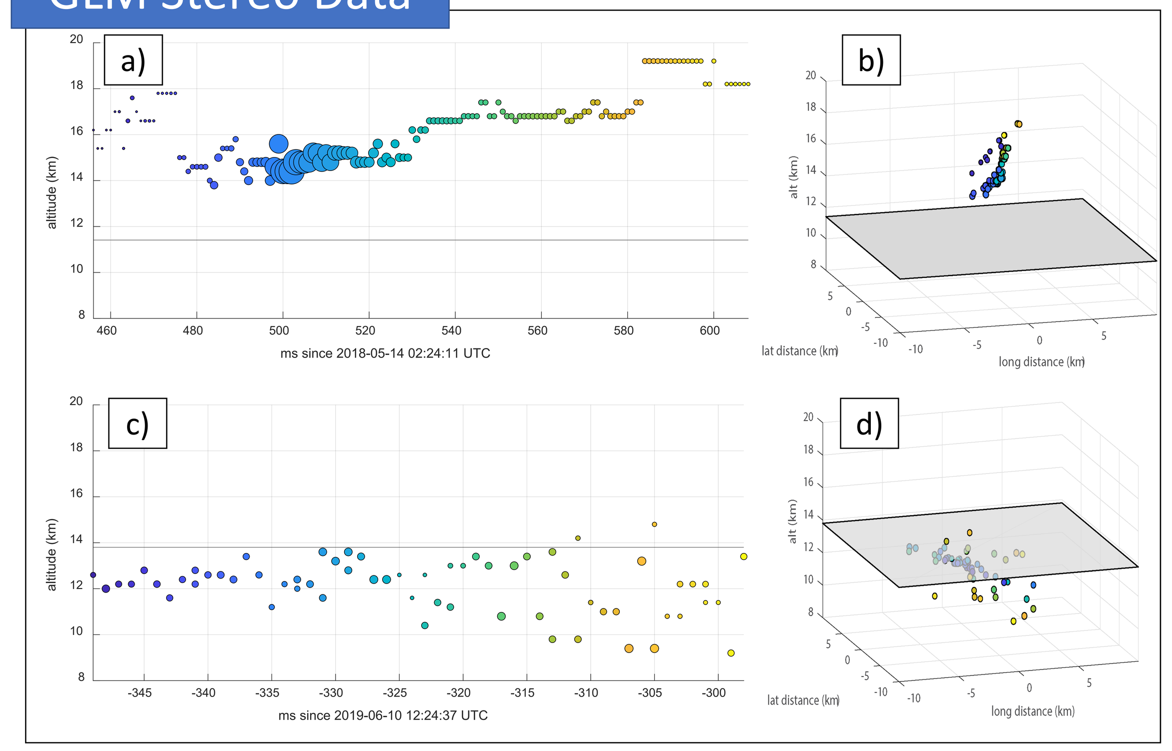

Figure 5. a) Stereo altitude vs. time for a confirmed GJ. Sized according to optical energy. b) spatial distribution (lat, lon, alt) of the sources in a). Colored according to time. The grey plane represents the cloud top. Panels c) and d) show similar plots for a non-GJ flash.

Machine Learning Feature Creation

- Use group-level GLM data, parsed from each GLM flash
- Feature categories:
  1. propagation: summarizes lateral extent / lateral coverage of GLM groups
  2. gridding: re-grid group centroid data to 4 km grid, captures energy distribution over the grid
  3. temporal: captures temporal evolution and associated temporal waveform properties

Features	Category
Max Distance	Propagation
Cumulative propagation	Propagation
Area	Propagation
Max Count Pixel	Gridding
Gradient N-S	Gridding
Gradient E-W	Gridding
Summed Energy Pixel	Gridding
Max Continuous Duration	Temporal
Integrated Energy	Temporal
Lightcurve Smoothness	Temporal

Table 1. Truncated feature list for the machine learning model.

Vetting

Duke Sferic Data

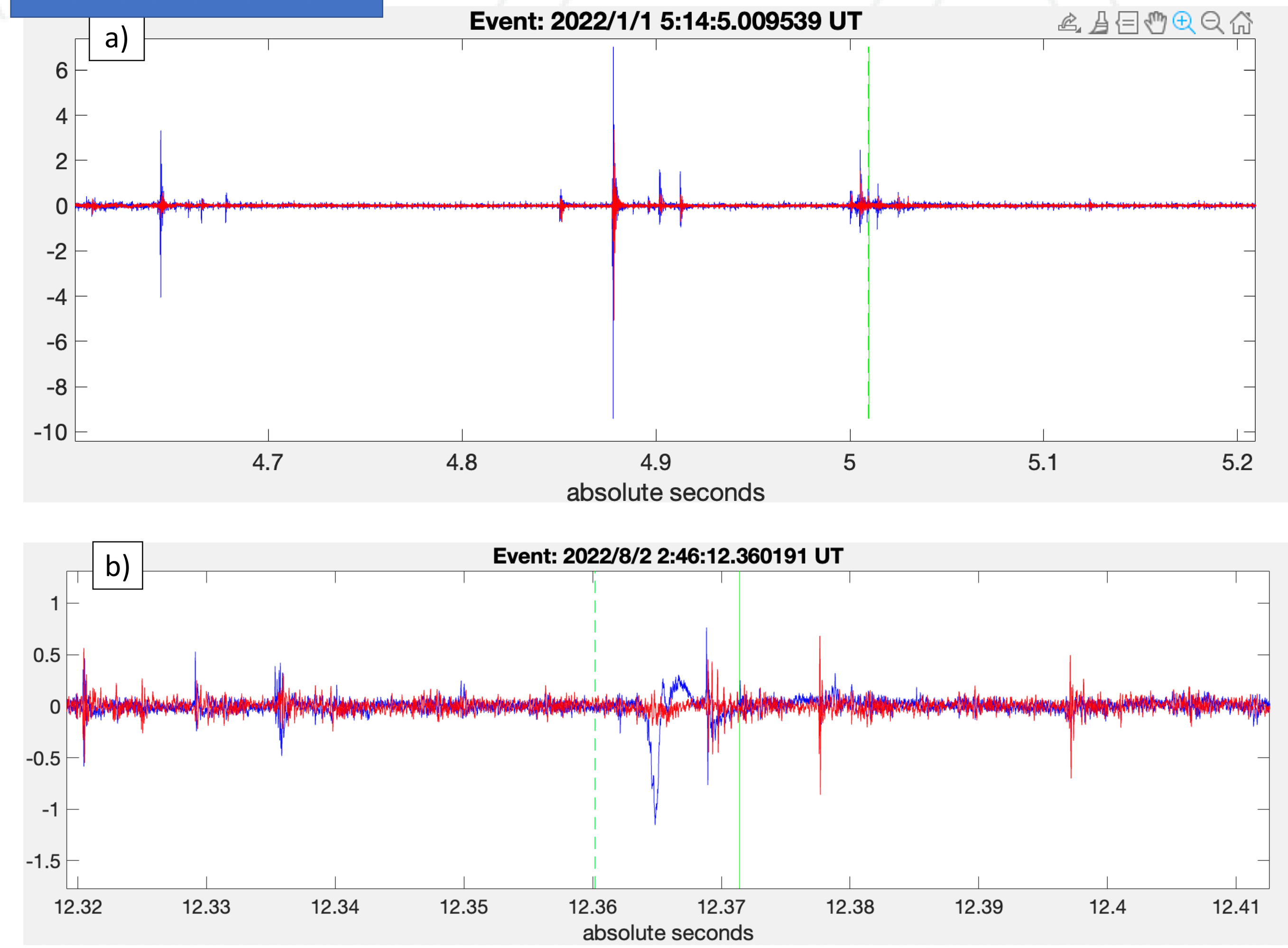

Figure 6. Duke VLF/ELF data for a) non GJ event (large cloud-to-ground followed by intracloud activity and b) a GJ (denoted as slow blue pulse). Green lines denote times and uncertainties from GLM detection candidate jets.

ABI & GLD360 Data

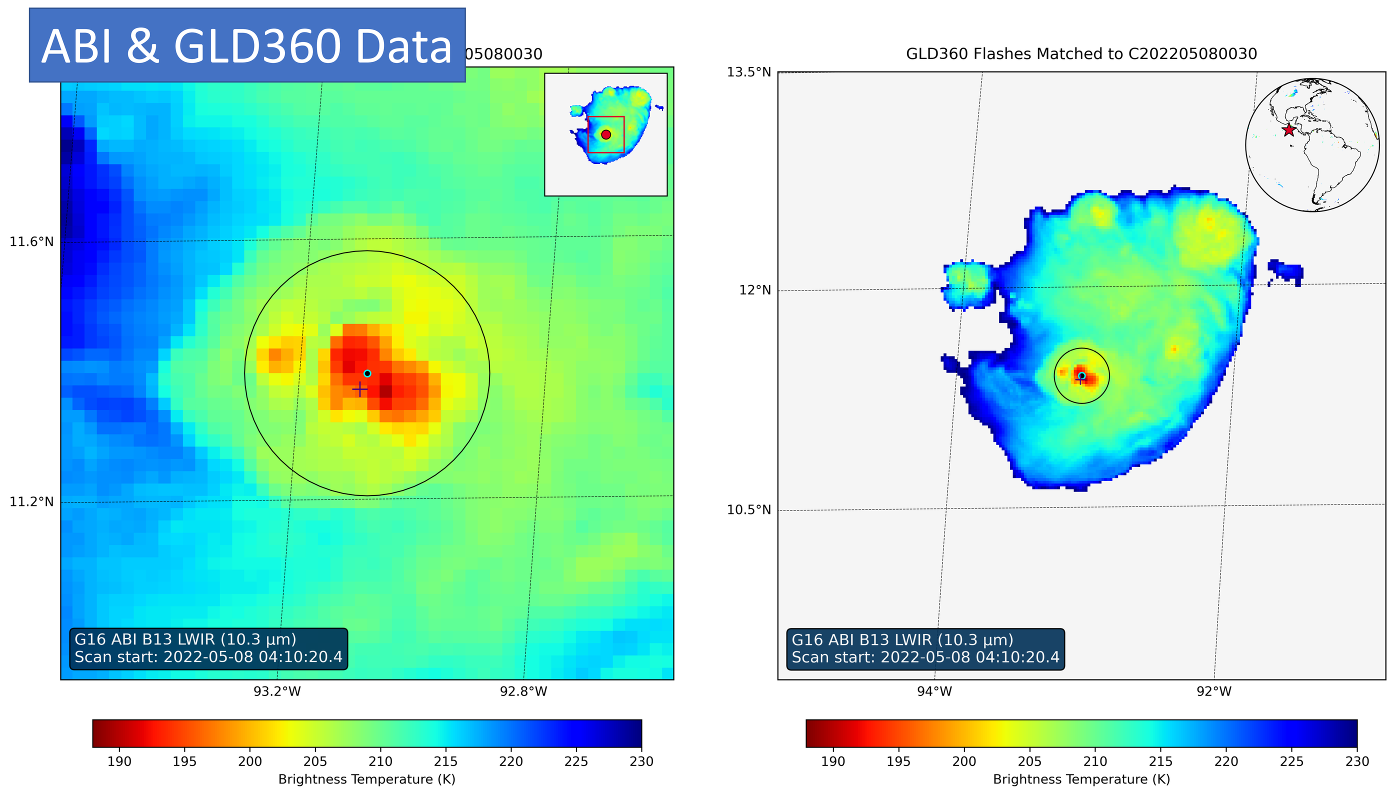

Figure 7. ABI data of GJ. (Left) zoomed region showing the cold cloud top, a 40 km radius buffer around the GLM detection, and a GLD 360 positive event (+). (Right) zoomed out region of panel to the left.

GLD360 and GLM Data

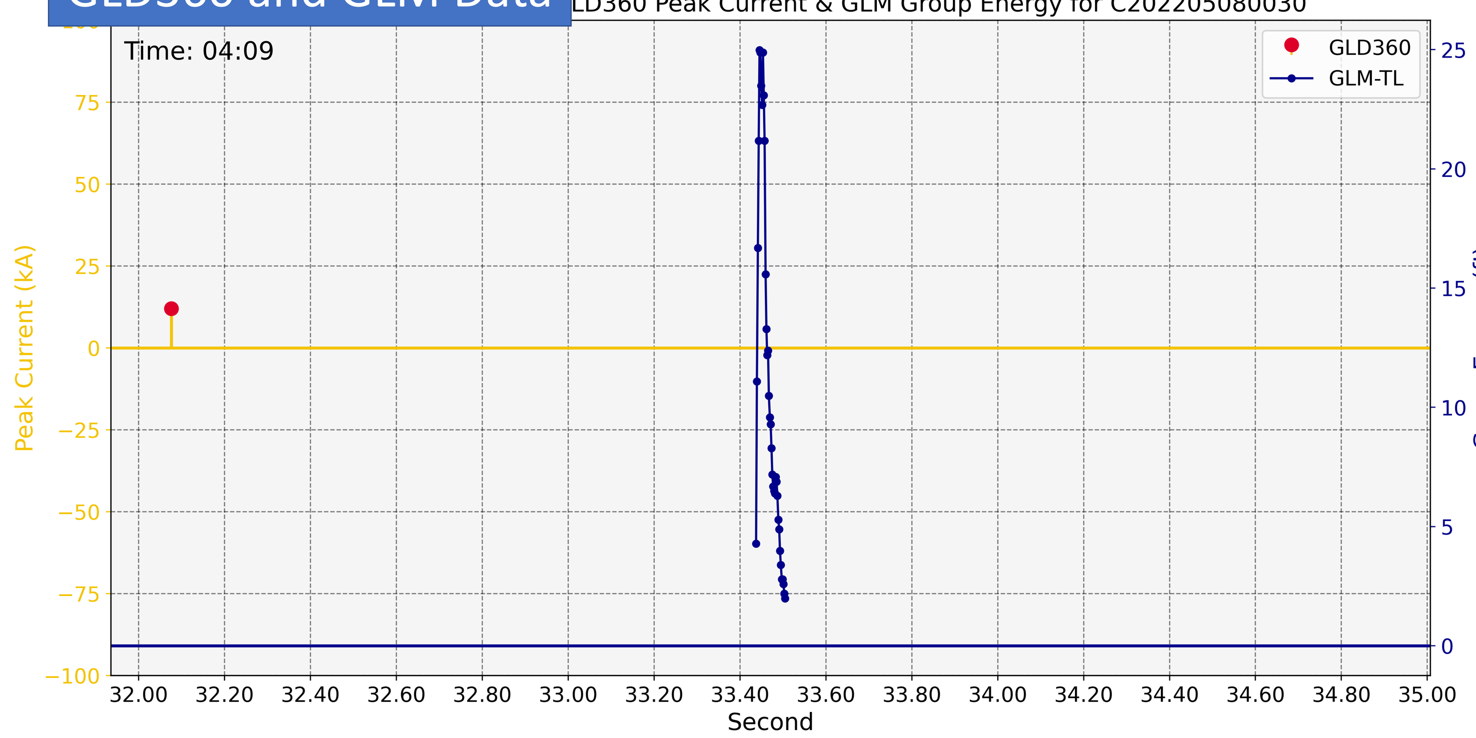

Figure 8. GLD360 (red circle) overlaid on a timeseries of GLM energy (lightcurve). GJ is denoted by the blue curve. The GLD360 event is far displaced in time from the GJ (not concurrent).

ABI & GLD360 Data

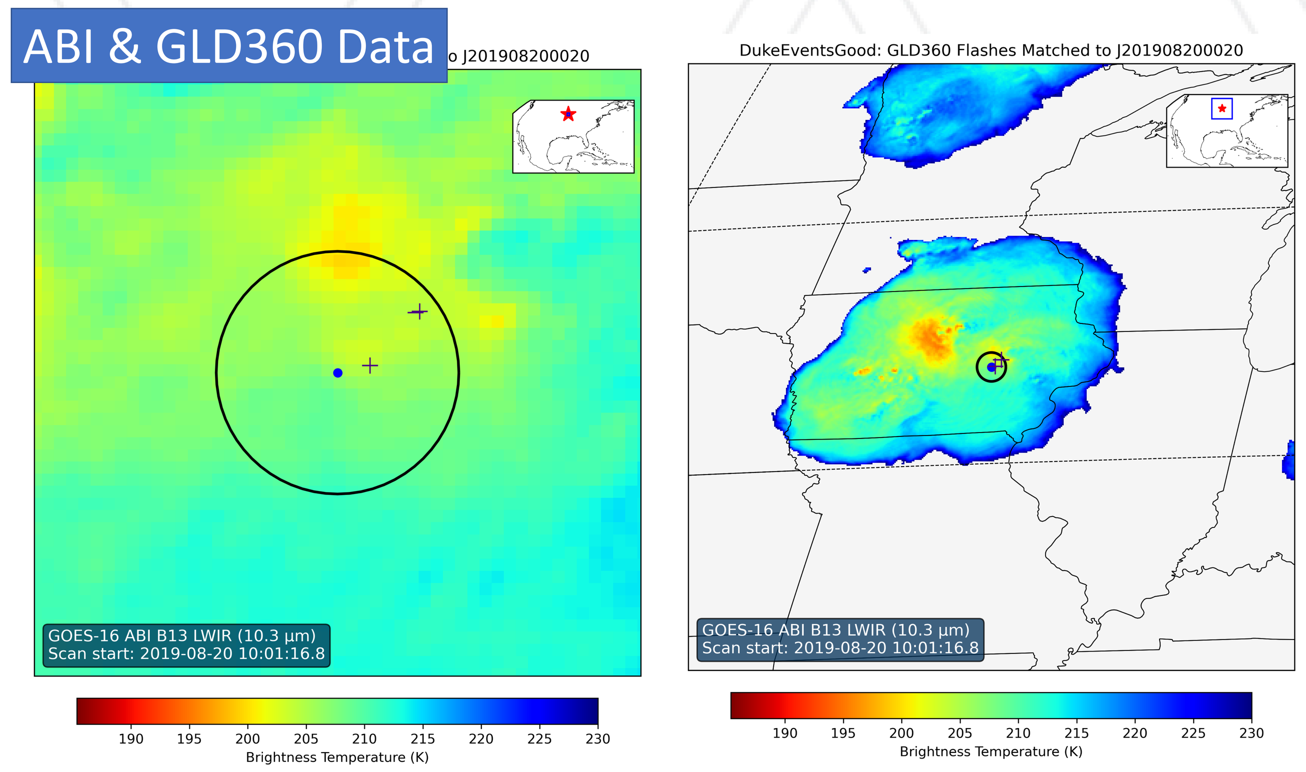

Figure 9. ABI data of CG discharge. (Left) zoomed region showing the cold cloud top, a 40 km radius buffer around the GLM detection, and some GLD 360 events. (Right) zoomed out region of panel to the left.

GLD360 and GLM Data

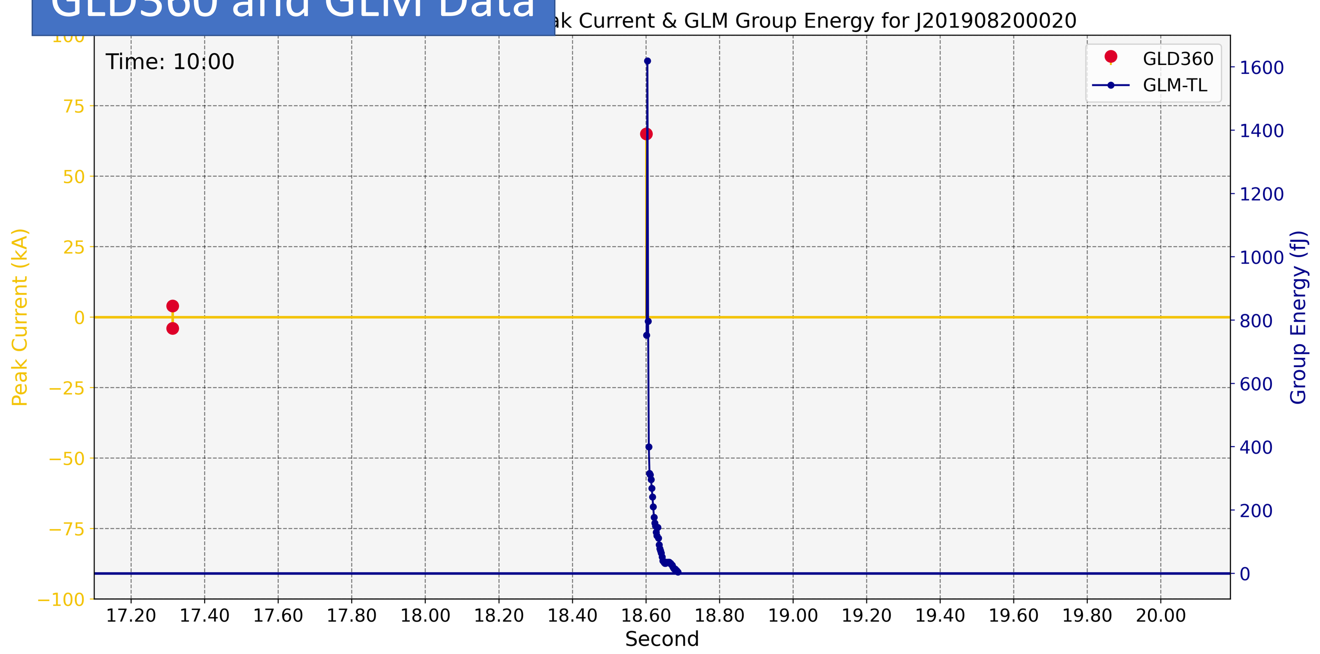

Figure 10. GLD360 (red circle) overlaid on a timeseries of GLM energy (lightcurve). The CG discharge is denoted by the blue curve. The GLD360 event is concurrent with the GLM lightcurve – most likely the return stroke generates the GLM optical energy and peak current (simultaneous emissions).

Conclusions

- **ML classifier has found ~20,0000 candidate GJs (low precision) so far (whole GLM GOV)**
- **Filtered down to several thousand detections in the initial training region**
- **Stereo vetting filters down to several hundred candidate GJs**
- **Duke, ABI, and GLD360 are final vetting tools. To date, have confidently identified approximately 25 GJs purely from our detection pipeline.**
- **Future work:**
  - **retrain ML model (create features to filter out CGs)**
  - **include newly found GJs in training data**
  - **increase throughput with the pipeline (more automation)**
  - **expand detections to entire GLM FOV (outside the stereo region)**

This work is supported by NSF awards:
AGS 2230383, AGS 2230384, AGS 2230385