Project JetNet: Hemispheric-scale gigantic jet detection network

Levi Boggs, Jeffrey C. Smith, Douglas Mach, Steve Cummer, John Trostel, Jeffrey Burke, Jessica Losego

1: GTRI, Smryna, GA USA;
2: SETI, Mountain View, CA USA;
3: USRA, Huntsville, AL USA;
4: Duke, Durham, NC USA

Introduction

In this presentation we will provide an overview and present preliminary results from a multi-institutional collaborative project, which seeks to detect gigantic jets over hemispheric scales using a combination orbital and ground-based sensors and machine learning. Gigantic jets are a type of transient luminous event (TLE, Pasko 2010, doi: 10.1029/2009JA014860) that escape the cloud top of a thunderstorm and propagate up to the lower ionosphere (80-100 km altitude), transferring tens to hundreds of Coulombs of charge. Our detection methodology primarily uses the Geostationary Lightning Mapper (GLM), which is a staring optical imager in geostationary orbit that detects the 777.4 nm (OI) triplet commonly emitted by lightning (Goodman et al. 2013, doi: 10.1016/j.atmosres.2013.01.006). Gigantic jets have been shown to have unique signatures in the GLM data from past studies (Boggs et al. 2019, doi: 10.1029/2019GL082278; Boggs et al. 2022, doi: 10.1126/sciadv.abl8731). Thus far, we have built a preliminary, supervised machine learning model that detects potential gigantic jets using GLM, and begun development on a series of vetting techniques to confirm the detections as real gigantic jets. The vetting techniques use a combination of low frequency (LF) and extremely low frequency (ELF) sferic data, in combination with stereo GLM measurements. When our detection methodology grows in maturity, we will deploy it to all past GLM data (2018-present), with the potential to detect thousands of events each year, allowing correlation with other meteorological and atmospheric measurements. We will share the database of gigantic jet detections publicly during and at project conclusion (2025), allowing other researchers to use this data for their own research.

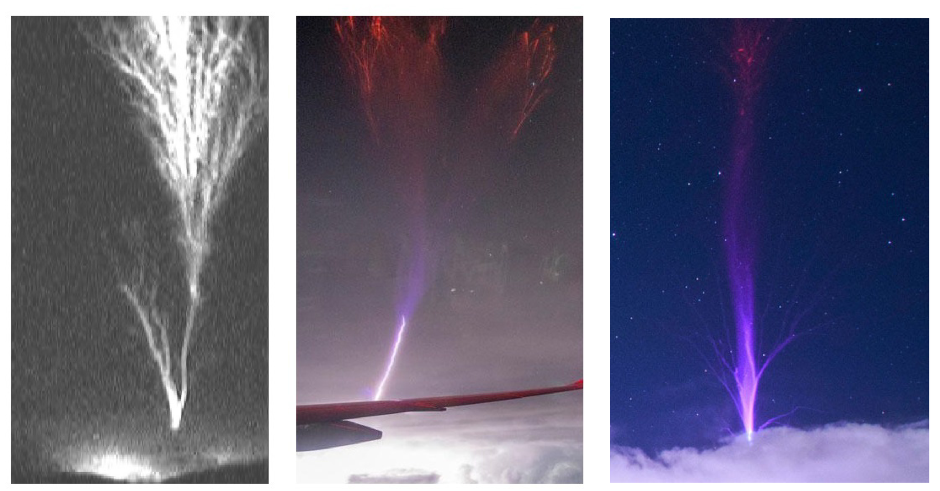

Figure 1. Gigantic Jets captured from ground-based cameras. (left) From *Boggs et al. 2019* (middle) from passenger on an airliner (right) from *Yang et al. 2020.*

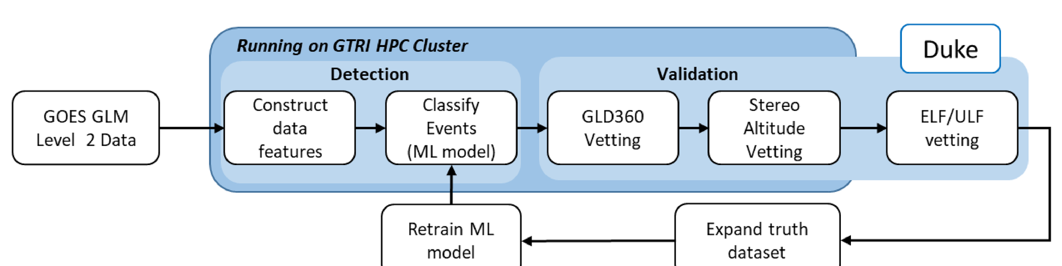

Figure 2. Detection pipeline for this project, consisting of GOES GLM data, GLD360 data, and ELF data.

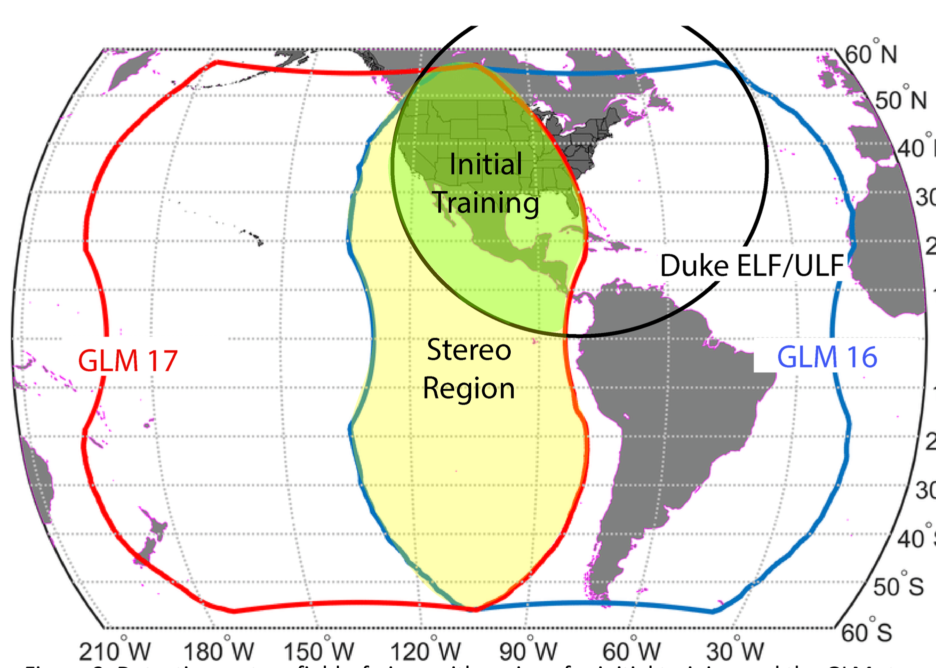

Figure 3. Detection system field-of-view, with regions for initial training and the GLM stereo region.

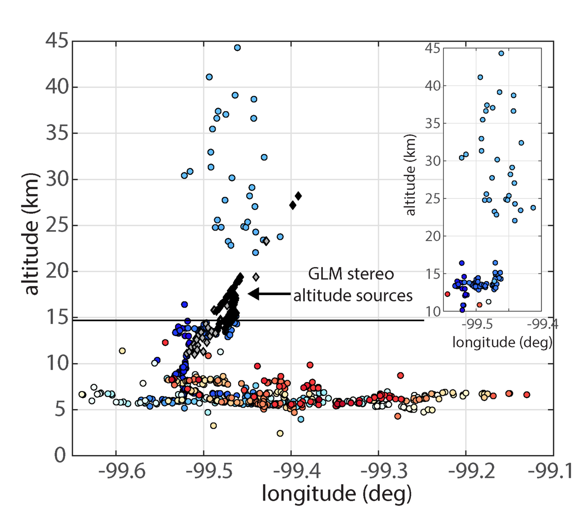

Figure 4. VHF LMA sources (colored dots) and GLM stereo sources (diamonds) for a GJ. The panel inset shows the VHF LMA data alone. Adapted from Boggs et al. 2022.

- GJs have hot leader that pokes above the cloud top → GLM stereo measurements
- Stereo sources map to vertical channel segment
- GLM sources are localized in lat, lon space
- GLM optical energy typically bright – no scattering/absportion from cloud

Results

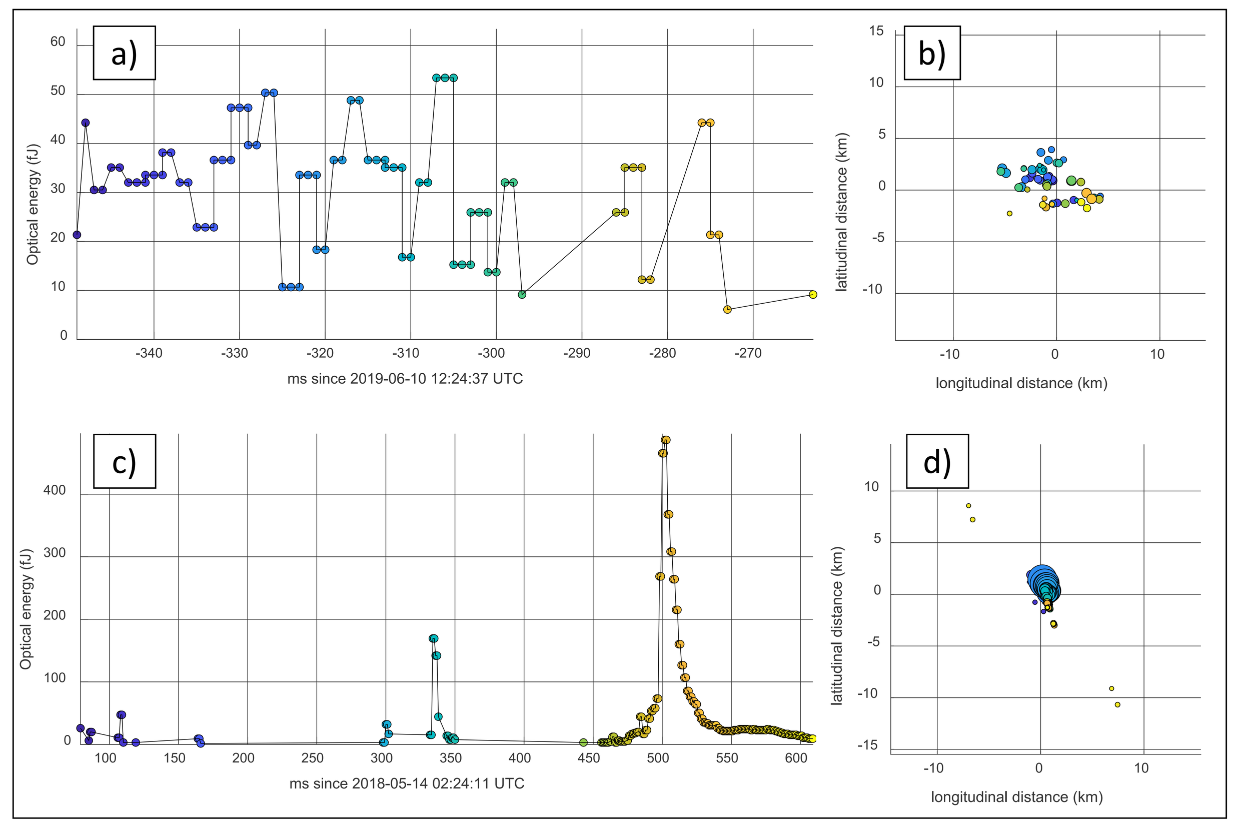

Figure 5. a) lightcurve (GLM energy vs. time) for a non-GJ flash. b) spatial (lat,lon) distribution for the sources in a). Panels c) and d) show similar plots for a confirmed GJ. The sources in b) and d) are sized according to energy.

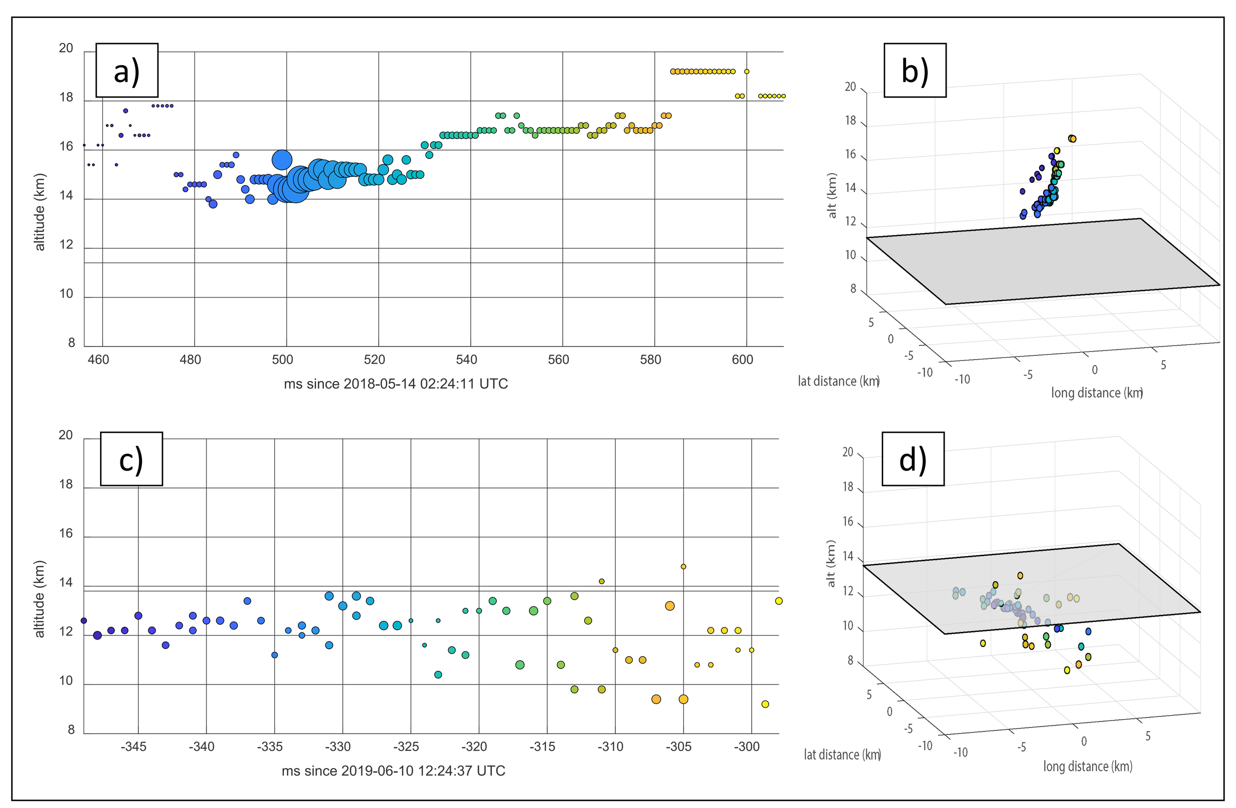

Figure 6. a) Stereo altitude vs. time for a confirmed GJ. Sized according to optical energy. b) spatial distribution (lat, lon, alt) of the sources in a). Colored according to time. The grey plane represents the cloud top. Panels c) and d) show similar plots for a non-GJ flash.

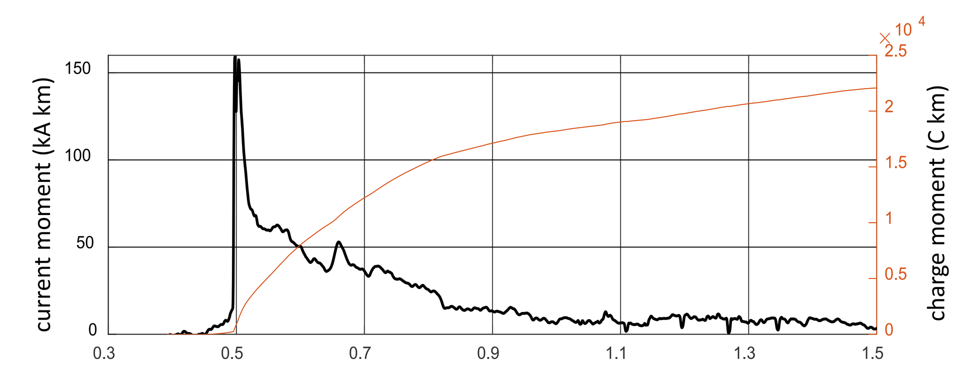

Figure 7. ELF data for the gigantic jet in Figure 5c captured by Duke University.

- Duke ELF data: current moment and charge moment change (CMC).
- GJs: average CMC of 5,800 C km for truth events
- Large positive CG events: average of 794 C km (Lyons et al. 2003)

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, summarizes energy distribution over the grid
  3. temporal: summarizes time 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. Feature list for the machine learning model.

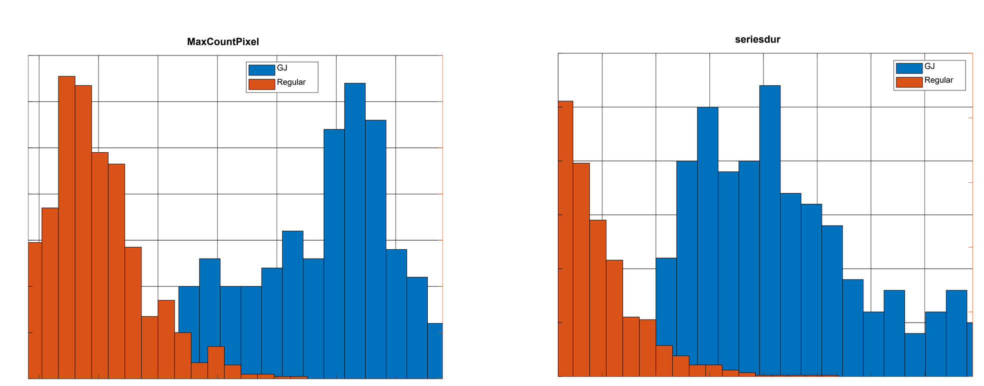

Figure 8. Distributions for the feature (left) Max Count Pixel and (right) Max Continuous Duration. Note: the GJ feature data includes synthetic SMOTE data.

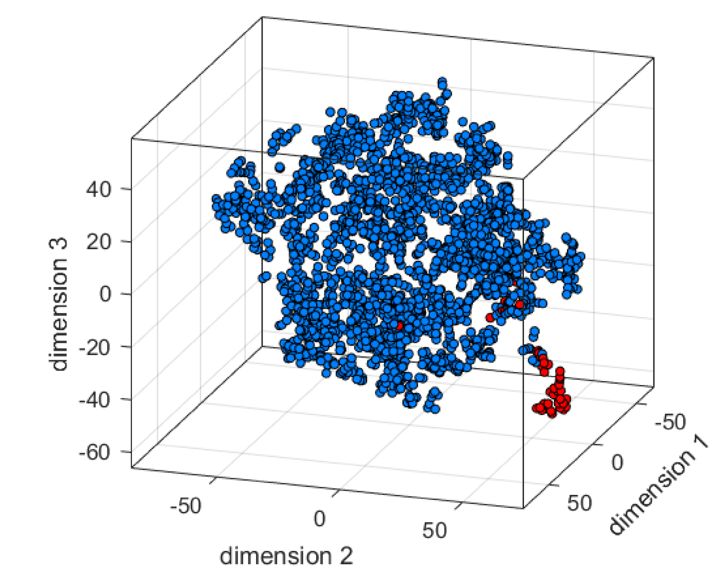

Figure 9. t-distributed stochastic neighbor embedding (t-SNE) plot showing the GJs (red) and non-GJs (blue). t-SNE plots distill a complex feature space into a 3D representation.

Machine Learning Initial Testing

- 21 confirmed GJs from ground-based video
- Randomly sampled several hundred flashes as non-GJ
- Used Synthetic Minority Oversampling Technique (SMOTE, Fernandez et al. 2018) to boost the features of minority (GJ) class to several hundred
- Built random forest model, with 150 maximum splits (i.e. branches), and 50 learners (i.e. trees)
- Testing:
  - Reserved 30% of initial data (confirmed GJs and non-GJs) for testing
  - With other 70%, boosted GJ class with SMOTE
  - Trained random forest with SMOTE data
  - Tested with the original 30% (confirmed GJs and non-GJs)
  - Performed 5 iterations of this
  - Precision: 83% (5/6) → limited truth GJ events
  - Recall: 83% (5/6) → limited truth GJ events
  - False positive rate: 0.435% (1/229)

Test confusion matrix

Truth		Predicted: GJ	Predicted: Non-GJ
	GJ	5	1
	Non-GJ	1	228

Machine Learning Deployment

- We have deployed to 1 month of data (10-12 million flashes)
- Model predicted ~13,000 possible GJs
  - false positive rate: 0.1%
- Currently working on vetting the events
  - Preliminary precision: ~1% → low as expected from first deployment. Similar results as GLM bolide detection (Smith et al. 2021).
  - Stereo altitude vetting
  - Duke ELF vetting
  - Slowly building larger truth database to re-train ML algorithm

Conclusions

- Initial work has started to build an automated pipeline to detect GJs
- The pipeline uses GLM data primarily, with GLM stereo data and ELF data for vetting
- Features to discriminate GJs from non-GJs use different GLM data characteristics such as group propagation, group energy distribution, and group time evolution
- Initial testing/training of a machine learning model shows high performance on limited test datasets
- Large-scale deployment shows similar false positive rates, but diminished precision as expected from first model iteration
- Vetting will produce more truth GJ events and improved model performance
- Once model is mature (2024), GJ detection list will be hosted by **NASA's Global Hydrometeorology Resource Center Distributed Active Archive Center (GHRC DAAC) and freely available in an easy-to-read format**

References

Pasko, Victor P. "Recent advances in theory of transient luminous events." Journal of Geophysical Research: Space Physics 115.A6 (2010).

Goodman, Steven J., et al. "The GOES-R geostationary lightning mapper (GLM)." Atmospheric research 125 (2013): 34-49.

Boggs, Levi D., et al. "First observations of gigantic jets from geostationary orbit." Geophysical Research Letters 46.7 (2019): 3999-4006.

Yang, Jing, et al. "Analysis of a gigantic jet in southern china: Morphology, meteorology, storm evolution, lightning, and narrow bipolar events." Journal of Geophysical Research: Atmospheres 125.15 (2020): e2019JD031538.

Boggs, Levi D., et al. "Upward propagation of gigantic jets revealed by 3D radio and optical mapping." Science advances 8.31 (2022): eabl8731.

Lyons, Walter A., et al. "Characteristics of sprite-producing positive cloud-to-ground lightning during the 19 July 2000 STEPS mesoscale convective systems." *Monthly Weather Review* 131.10 (2003): 2417-2427.

Mach, Douglas, and Katrina Virts. "A technique for determining three-dimensional storm cloud-top locations using stereo optical lightning pulses observed from orbit." Journal of Atmospheric and Oceanic Technology 38.11 (2021): 1993-2001.

Smith, Jeffrey C., et al. "An automated bolide detection pipeline for GOES GLM." Icarus 368 (2021): 114576.

Fernández, Alberto, et al. "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." Journal of artificial intelligence research 61 (2018): 863-905.

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