Pipeline Development and Statistical Analysis of Bolides Detected by the GOES Geostationary Lightning Mappers

J. C. Smith[1], R. Morris[1], R. Longenbaugh[2], J. Dotson[3], N. McCurdy[3],
[1]The SETI Institute, Mountain View, CA (jsmith@seti.org), [2]Sandia National Labs., [3]NASA Ames Research Center

Introduction: The Geostationary Lightning Mapper (GLM) instrument [1] onboard the GOES 16 and 17 satellites has been shown to be capable of detecting bolides in the atmosphere [2]. Due to its large, continuous field of view and immediate public data availability, GLM provides a unique opportunity to detect a large variety of bolides, including those in the 0.1 to 3 m diameter range that complements current ground-based meteor detection systems, which are typically sensitive to smaller objects. We have deployed a machine learning (ML) based bolide detection and light curve generation pipeline [3] on the NASA Advanced Supercomputer Facility, with detections being promptly published on a NASA hosted publicly available website, https://neo-bolide.ndc.nasa.gov. The goal is to generate a large catalog of calibrated bolide light curves to provide an unprecedented data set for three main purposes: 1) to inform meteor entry models on how incoming bodies interact with the atmosphere, 2) to infer the pre-entry properties of the impacting bodies and 3) to statistically analyze bolide impact populations across the globe. The pipeline has now been operational for over 3 years and we have amassed a catalogue of over 3500 bolides.

Pipeline Architecture and ML training: As is common practice in machine learning, an iterative approach was taken to the pipeline development. Work began by utilizing classic filter design where human-derived heuristics were considered to create a set of detection filters tuned to what a human would expect for a bolide event in GLM data [4]. This manually tuned detector had a detection precision of about 1.5%. The resulting detection set was therefore 98.5% false positives. Manual human vetting then labeled the small number of legitimate bolides. Once the initial training set was obtained machine learning training could begin. After training each new classifier, the GLM data was reprocessed, and a subset was again manually labeled. After several iterations, progressively more competent ML classifiers were developed. Our current best classifier utilizes a random forest [5] and has a detection precision of 88%, meaning almost nine out of ten automated detections are legitimate. Final human vetting is still performed before publication on the website. By utilizing the NASA Advanced Supercomputing Facility, we can capitalize on massive parallelization and reprocess the full historical dataset in a matter of hours.

Statistical Analysis of Detected Bolides: To date, we have posted 3549 bolides on our public website, all being manually vetted by a human before publishing. Effort has been made to identify and eliminate all systematic biases in the detection pipeline. A principal source of false positives is solar glints, where sun glare reflects off the Earth producing a signal in GLM that appears bolide-like. Solar power generation farms, in particular, can cause large numbers of concentrated false positives at specific locations on the globe. After removing these biases in the data, we can begin to statistically analyze the bolide distributions. We wish to measure the distribution of bolides as a function of latitude to assess various theories of bolide impact distributions.

A marked increase in detection is very apparent during meteor showers. We are particularly sensitive to the Leonid meteor shower and found the volume of bolides in 2020 to be considerably higher than in 2019 or 2021. After experimentation, we have not found a systematic detection bias in our pipeline and so the increase appears to be legitimately in the data. We have also found a strong bias for detection during the night hours for each satellite, as would be expected.

Outlook: Our ultimate goal is near 100% precision and automatic posting of detections with no human intervention. We will then be able to potentially post bolides within a minute of impact. Critical to this goal is automation of the manual human validation step. The human vetting utilizes the GOES Advanced Baseline Imager (ABI) [6] data to identify and remove false positives due to lightning and LEO satellites. We intend to develop a deep learning-based classifier of ABI images, along with other external data streams, to automate the human validation step.

References:

[1] Goodman, S.J., et al. The GOES-R Geostationary Lightning Mapper (GLM). Atmos. Res. 125, 34–49. (2013)
[2] Jenniskens, P., Albers, J., Tillier, C.E., et al., 2018. Detection of meteoroid impacts by the Geostationary Lightning Mapper on the GOES-16 satellite. Meteorit. Planet. Sci. 53 (12), 2445–2469. (2018)
[3] Smith, J.C., et al., "An Automated Bolide Detection Pipeline for GOES GLM", Icarus, Vol. 368, 1 (2021)
[4] Rumpf, C.M., et al., An algorithmic approach for detecting bolides with the geostationary lightning mapper. Sensors 19 (5), 1008. (2019)
[5] Breiman, L., Random forests. Mach. Learn. 45 (1), 5–32. (2001)
[6] Schmit, T.J., Gunshor, M.M., Menzel, W.P., Li, J., Bachmeier, S., Gurka, J.J., 2005. Introducing the next-generation advanced baseline imager (ABI) on GOES-R. Bull. Amer. Meteorol. Soc. 8, 1079–1096.