# JGR Atmospheres

**Key Points:**
- An algorithm restores the shapes of Geostationary Lightning Mapper event detections and resamples them on a target grid to make imagery
- Imagery capitalizes on GLM's improved sensitivity, resulting in improved display of meteorological processes compared to point displays
- Statistics from imagery support prior work distinguishing a second class of extensive, propagating lightning

**Correspondence to:**
E. C. Bruning,
eric.bruning@ttu.edu

# Meteorological Imagery for the Geostationary Lightning Mapper

**Eric C. Bruning[1], Clemens E. Tillier[2], Samantha F. Edgington[2], Scott D. Rudlosky[3], Joe Zajic[4], Chad Gravelle[5], Matt Foster[6], Kristin M. Calhoun[7,8], P. Adrian Campbell[7,8], Geoffrey T. Stano[9,10], Christopher J. Schultz[10], and Tiffany C. Meyer[7,8]**

[1]Department of Geosciences, Texas Tech University, Lubbock, TX, USA, [2]Lockheed Martin, Palo Alto, CA, USA, [3]Center for Satellite Applications and Research, National Oceanic and Atmospheric Administration, College Park, MD, USA, [4]Integrity Applications, Chantilly, VA, USA, [5]National Weather Service Southern Region Headquarters, National Oceanic and Atmospheric Administration, Fort Worth, TX, USA, [6]National Weather Service Operations Proving Ground, National Oceanic and Atmospheric Administration, Kansas City, MO, USA, [7]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK, USA, [8]National Severe Storms Laboratory, Office of Atmospheric Research, National Oceanic and Atmospheric Administration, Norman, OK, USA, [9]ENSCO, Inc., Huntsville, AL, USA, [10]SPoRT, Marshall Space Flight Center, National Aeronautic and Space Administration, Huntsville, AL, USA

**Abstract** The Geostationary Lightning Mapper (GLM) on the Geostationary Operational Environmental Satellite-R series of weather satellites provides point geolocations of lightning flashes that are further comprised of a hierarchy of geolocated groups and events. This study describes an open-source method for reconstruction of imagery from those point detections that retains the quantitative physical measurements made by GLM, restores the spatial footprint of the events, and connects that spatial footprint to the groups and flashes. Meteorological signals are demonstrated to be more apparent in the gridded imagery than in the point detections, leading to their adoption by the United States National Weather Service as the first GLM product available in their real-time displays. Analysis of a mesoscale convective system over Argentina confirms that there is a class of propagating lightning observed by GLM (distinct from that in storm cores) that can be visualized and quantified using our imagery-based approach.

**Plain Language Summary** This paper describes a method for creating imagery from the Geostationary Lightning Mapper instrument on the new geostationary weather satellites launched by the United States. The imagery overlays directly on and can be animated like other weather satellite images, making it more suitable for diagnosing thunderstorm behavior than the simple lightning location plots easily made from the publicly available data. The imagery can be summed to create analyses of lightning on climate time scales. The imagery illustrates that extensive lightning discharges exist in some storm systems and that distant ground strike points are joined by a single extensive lightning channel in the cloud.

## 1. Introduction

The Geostationary Lightning Mapper (GLM) on the Geostationary Operational Environmental Satellite (GOES)-R series of spacecraft is a new Earth-observing capability (Goodman et al., 2013; Rudlosky et al., 2019, 2019). It is, along with the Advanced Baseline Imager (ABI; Schmit et al., 2005, 2017), one of the two Earth-pointing instruments on GOES-R. GLM continuously monitors optical emission of light from cloud tops, with the design requirement of detecting more than 70% of lightning flashes on a 24 hr basis over a field of view covering a subset of the full disc spanning ±54° latitude. The 24-hr design requirement translates to an expected nighttime flash detection efficiency of about 90%. Technologically, GLM is a 500 frame per second camera, though the data volume and challenges inherent in detecting small-amplitude transient pulses of light against a daytime bright cloud background means the data are not processed and reported as a simple high-speed video. Instead, the GLM data are distributed as pixel-level triggers above the bright cloud background, clustered through a three-level hierarchy into lightning flashes. Ground processing algorithms use the spatial and temporal adjacency of triggered pixels to identify lightning flashes. The triggers are reported as geolocated points, so the only information available to the end user about the spatial extent over which the trigger is valid is through their implicit spacing when adjacent pixels are triggered in the same frame.

Furthermore, measurement of the spatial extent of each lightning flash was not one of the original design requirements for the GLM (Goodman et al., 2010). Rather, the goal was minimally sufficient spatiotemporal coherency at the event level so that lightning flashes could be separated from nonlightning triggers. However, the sensitivity of GLM as built is in fact sufficient to allow for mapping of the spatial extent of discharges and so it is valuable to have a GLM data set that exploits this capability and represents the aggregate spatial extent of each flash. The feasibility, characteristics, and utility of one method of constructing such imagery is demonstrated by this study.

GLM also continues the heritage of prior space-based optical measurements of lightning and so matches the data reporting format of those sensors. In practice, that format can lead to the GLM data being treated by end users as point data, when in fact it arises from a pixel-based detector array. The Lightning Imaging Sensor (LIS) and the Optical Transient Detector (Christian et al., 2000; Goodman et al., 2010; Mach et al., 2007) utilized a similar detection principle, though with a limited sampling duration of $O(10^2 \text{ s})$ for any one location. In the literature, one may find storm cells visualized as an accumulation of flash centroids (e.g., Buechler et al., 2000) or as disjoint imager pixel polygons, points, or ellipses showing the extent of individual flashes (e.g., Cecil et al., 2005; Goodman et al., 2013; Peterson & Liu, 2013; Peterson et al., 2017). Climatological analyses (Albrecht et al., 2016; Beirle et al., 2014; Cecil et al., 2014; Christian et al., 2003) of LIS and Optical Transient Detector data simply accumulated flash properties at the flash centroid. Evidently, there is no standard approach for accumulating the flash data that preserves the mapped extent of the flashes while also permitting long-term aggregation into a climatology and as discussed below, the necessary data and metadata for doing so are not obviously available from the routinely produced GLM data as available to the public.

The use of GLM data in climatological analyses is expected to grow because the World Meteorological Organization and the Global Climate Observing System have declared lightning an essential climate variable (Aich et al., 2018) for studying climate and climate change. A global constellation of geostationary optical sensors will soon be in place, with a lightning imager already present on the Chinese FY-4A satellite (Yang et al., 2017) and expected in the next few years on Meteosat Third Generation over Europe and Africa (Stuhlmann et al., 2005). The preliminary World Meteorological Organization concept for a lightning climate data set fuses long-range radio-frequency lightning networks with space-based observations on a 0.1 grid (World Meteorological Organization, 2019), so a well-defined gridded lightning product from GLM takes on added scientific relevance as a foundation for future work. Furthermore, Meteosat Third Generation will include accumulated lightning products as part of its baseline product suite and for which the GLM imagery described in this study will serve as a reference (the first author of this study is advising on that effort).

The space-based optical lightning mapping imagery described in this study is applicable on space and time scales from the storm scale to the climatological. GLM's sensitivity as built means that we can rely on GLM's sensor to create quantitative photographs of the spatial extent of lightning flashes and their flickering. After being rectified as an image, the data set is trivially accumulated so as to quantify lightning and its spatial extent over arbitrary time intervals, from 2 ms to years. The method is shown to support qualitative visualization of thunderstorm structure, and analysis of the imagery over long time intervals reveals that patterns in flash properties correspond to two distinct classes of lightning: discharges on the scale of individual storms and larger mesoscale discharges. This study will focus on a case from South America, where, in the absence of a dense radar network, we are able to infer details of individual, extensive discharges while also illustrating mesoscale thunderstorm structure and evolution with the same data set.

One application of the method described in this study is time-series GLM lightning imagery like that typically produced by weather satellites. After a rapid development effort beginning in January 2018, our approach has been implemented as the first GLM product available within the U.S. National Weather Service (NWS) and is also available as an open-source software package (Bruning, 2019). A real-time feed for academic and research users is available through Unidata (Fulker et al., 1997) alongside those ABI data sets already available (Goebbert et al., 2019).

## 2. Background

The GLM imagery is based on a heritage of visualization techniques prototyped with other total lightning mapping systems, especially very high frequency (VHF) systems such as the Lightning Detection and Ranging System (LDAR, Boccippio et al., 2000; Poehler & Lennon, 1979) and Lightning Mapping Array (LMA)

(Rison et al., 1999). In VHF data, it is common to observe thousands of points that overlap in any 2-D cross-section of the 4-D point cloud. Recent work in the GOES-R Proving Ground (Calhoun et al., 2013, 2014; Calhoun, 2015; Goodman et al., 2012) applied visualization techniques that summarized the VHF point detections as time series imagery. These included simple methods such as gridded counts of the points and more elaborate methods that clustered the points into flashes using space–time thresholds (Fuchs et al., 2016; Murphy, 2006; McCaul et al., 2009; Thomas et al., 2003) prior to counting flash-level properties on a grid. Such methods have value in distinguishing high density clusters of points produced by a single lightning channel from those high densities produced by multiple discharges in the same location. Detection efficiency of the VHF source points also decreases much more rapidly with range than flash detection efficiency, so working at the flash level serves to normalize lightning counts as a function of range.

One of the earliest efforts to formally specify these visualization methods was by Lojou and Cummins (2004), who defined flash extent density (FED) as the count of all flashes that passed through a grid box during some period of time. After identifying clusters of points that belonged to each flash, neighboring points were connected, and each flash was said to touch a box if any edge of the connectivity graph passed through a box. (Depending on the study, "density" may actually refer to a simple count, or a quantity with density units resulting from division of the count by the size of a grid box and the time interval of accumulation. This study uses counts.) Bruning and MacGorman (2013) used the point locations instead of the graph edges in their definition of FED, causing the activated grid boxes for some flashes to be noncontiguous. While visually undesirable, it was easier to implement computationally. However, in both approaches, the extent density calculation identified the number of unique flash IDs at each grid cell.

Such visualizations of flash spatial extent allowed for visual identification of storm cells and the associated local flash rate at its center. In a time series, FED made apparent fluctuations within individual thunderstorm cells and in an animation, it was possible to identify storm tracks over a longer period of time by following the cell centroids. This contrasted with a grid created from the sum of the flash centroids or flash initiation locations (first VHF sources), which were quite sparse (Figure 4 in Bruning & MacGorman, 2013).

The other quantity introduced by Bruning and MacGorman (2013) was average flash area (AFA). While counting flashes answers the elementary question of "how many," the meteorological value of depicting flash areas was based on the expected contrast between less spatially extensive flashes in storm cores (dominated by closely coupled and highly variable updrafts and downdrafts) and more extensive flashes in stratiform or anvil regions of thunderstorms (dominated by transport of hydrometeors in relatively stratified mesoscale or synoptic-scale flow) (Bruning & MacGorman, 2013; Bruning & Thomas, 2015; Schultz et al., 2015).

The extensive channels mapped by LMAs have proven useful in connecting the location of widely separated ground strike locations to a single parent discharge (e.g., Lang et al., 2017; Lu et al., 2013). GLM imagery showing the flash extent is expected to have similar interpretive utility, allowing for previously unknown relationships between ground strikes to be analyzed. For instance, Lyons et al. (2019) reported an extensive, propagating flash with numerous ground strike locations where the GLM data were useful in understanding the connectivity of the discharge beyond the LMA line of sight horizon. The greatest separation between a pair of events attached to the area of contiguous illumination of the flash was 547 km. Peterson (2019) reported the largest discharge to date of 674 km in an examination of all GLM data from 2018. Recently, Peterson et al. (2017), Peterson et al. (2017), and Peterson et al. (2018) utilized LIS data to visualize the spatially extensive illumination of clouds by lightning and showed how the connectivity of successive cloud illumination centroids within the same flash could be used to visualize the lightning channel. Additional flash metrics and visualization techniques from those studies complement and extend the methods described in this study. The aim of this study is more foundational, aiming to quantify only those properties reported in the basic GLM data without performing further analysis of the flash properties.

### 2.1. The GLM Data Set

The GLM detector is a charge-coupled device (CCD) pixel array and has a variable pixel pitch that keeps the ground sample distance of the pixels from exceeding 14 km over the continental US (CONUS; GOES-R Series Program Office, 2019). Most of the field of view has a ground sample distance of less than 10 km and at nadir, 8 km. The GLM electronics monitor the background scene brightness detected by the CCD and on a frame-by-frame basis check for pixels that exceed a threshold that depends on the background. After filtering for shot noise, glint, radiation tracks, blooming due to stray light or glint, and other artifacts (GOES-R

Series Program Office, 2019; Goodman et al., 2010; Rudlosky et al., 2019), a detection is reported at each pixel exceeding that background in a single 2-ms frame. In the GLM data set, the triggered pixels are known as events, and the adjacent, triggered pixels in a single frame form groups. In the case of lightning, a triggered detection usually corresponds to the pool of light emitted from the top of a cloud surface illuminated after the light from the lightning channel has scattered through the cloud. Groups that have any events within 16.5 km and 330 ms (Goodman et al., 2010) in weighted Euclidean distance (Mach et al., 2007) are clustered into GLM flashes, encompassing the flickering light output expected to be produced by a connected light-ning channel discharge tree. If processing delays accumulate, the clustering algorithm permits flashes to be truncated after 3 s (Peterson, 2019). The three-level parent–child hierarchy resulting from a clustering algo-rithm applied to an image is an application of connected component labeling theory (Shapiro, 1996). These data are the baseline measurements made publicly available in the GLM level 2 (L2) data files (GOES-R, Series Program Office, 2017) and are similar to the definitions used in the heritage LIS and Optical Tran-sient Detector instrument processing (Christian et al., 2000; Goodman et al., 2010; Mach et al., 2007). The net effect of this processing is to produce a data set that is largely light pulses produced by lightning. This study is not concerned with the effectiveness of the filters and clustering (which at the time of publication continued to be refined), so it makes no modification to the flash, group, and event data as produced by the operational L2 processor.

The variable optical energy across each group and its spatial extent are the primary physical measurement quantities and are referred to as a lightning detection data set. As with many geophysical measurements, signals produced by one phenomenon (here, lightning) can be produced by other physical phenomena. For example, GLM has also detected meteorite bolides (Jenniskens et al., 2018; Rumpf et al., 2019), and these are preserved by the processing method in this study.

Associated with each event, group, and flash are the following data. Each includes a centroid latitude and longitude, time (both start and end for flashes), and ID number for that entity. As child levels in the hier-archy, groups and events also included a parent ID for the next level up the hierarchy. Therefore, there were navigational coordinates in space, time, as well as hierarchy. Physical properties for each entity are the radiant energy and the area of each group and flash.

The GLM L2 data (GOES-R Algorithm Working Group and GOES-R Series Program, 2018) comply with the Climate and Forecast (Eaton et al., 2011) and National Aeronautics and Space Administration meta-data conventions and are stored in NetCDF 4 (Unidata, 2016) files. The flash, group, and event variables are each organized along their own dimension, giving the data a tabular structure with each dimension corre-sponding to a separate table. Such formal specification of the structure of the data enables implementation of aspects of the GLM data processing as a generic traversal operation, as described below.

### 2.2. Data Display Methods and Their Utility

Figure 1 shows a display easily made from the GLM L2 data, where different symbols are used to represent the centroid locations of a flash, its constituent groups, and the constituent events that make up each group.

While it is programmatically convenient to create a point display, it is a poor visualization of the pools of light detected by GLM. The events overlap one another, so it is not possible to infer the number of events triggered at each location. Furthermore, while the event spacing implies the arrangement of the underlying GLM CCD, the visualization does not represent how light was detected across the span of the pixels.

Therefore, the goal is to utilize the available information at L2 to produce imagery that depicts the spatial extent of each GLM event while also quantifying counts and other physical quantities (such as area and energy) in the data set. The quasi-regular spacing of GLM events complicates the production of such imagery.

If GLM events were evenly spaced in some coordinate system, the events could be accumulated as points (as with LMA data) on a grid that matches the CCD configuration. However, GLM's variable CCD pitch does not meet this requirement. Other possible target grids, such as a regular latitude–longitude grid, exacerbate the problem further. Consider such a grid at 0.1° (about 10 km) resolution. The convergence of regular longitude lines toward the Earth's poles means that two events spaced evenly from the geostationary point of view contribute to a single grid cell on the equatorward side of the grid, while artificial gaps are introduced toward the poles.

The problem also exists on the storm scale. Figure 1 depicts a regular 0.1° grid. On this grid, which is larger than the typical GLM event spacing, there are occasions where two events are within the same grid cell.
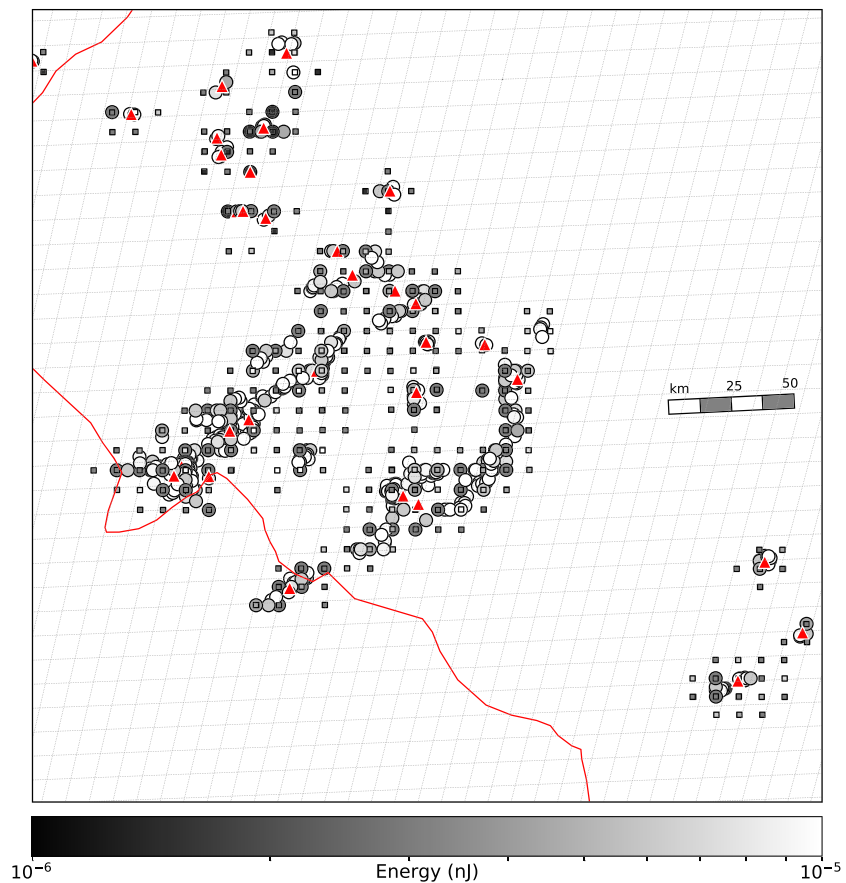
**Figure 1.** Geostationary Operational Environmental Satellites-16 Geostationary Lightning Mapper events (squares), groups (circles), and flashes (red triangles) from 1348–1349 UTC on 13 November 2018 near the border of Uruguay and Brazil. Numerous events overlap one another, with the most recent event plotted on top. Brighter events and groups are in lighter colors, logarithmically scaled between 1 and 10 fJ total energy at the Geostationary Lightning Mapper aperture. Spacing between events is about 10 km. A regular latitude–longitude grid with spacing of 0.1° is included for reference.

Such double counting leads to artifacts when resampling the GLM data. For instance, if we wish to represent on the grid the number of times lightning flickered at that location (in the next section, this is termed as group extent density), it is reasonable to assume that we might do so by counting the number of events that fell within any given grid cell. In the case where a single group has illuminated some large area, the group would show up in some locations as a count of 2, which is a false representation of how many times light flickered at that location.

At smaller grid sizes, gaps may be introduced, and both problems were observed in a prototype gridded product developed for operations in the U.S. NWS. Figure 2 shows an example from an early prototype NWS display and illustrates how counting events on a regular latitude–longitude grid (at 8- or 9-km equivalent grid spacing at this location) leads to gaps and holes where there is actually an unbroken field of events. A larger grid spacing can eliminate gaps (at 10 km and mostly at 9 km) but leads to double counting of events. Each of these choices has a different pattern of maxima and minima, which suggests that an analyst might infer a different interpretation of the cellular structure of the underlying thunderstorm as a function of the analysis approach. In our correspondence with a range of expert and nonexpert users of GLM beyond the NWS, we have observed that it is all too easy to introduce such artifacts. The GLM grids defined in this study do not exhibit such artifacts, so adoption of this study's methods will prevent future users from unknowingly introducing them. As such, this study provides a universal solution to the gap and double-counting problems while maintaining a direct connection to the events, groups, and flashes as observed, counted, and reported by GLM.
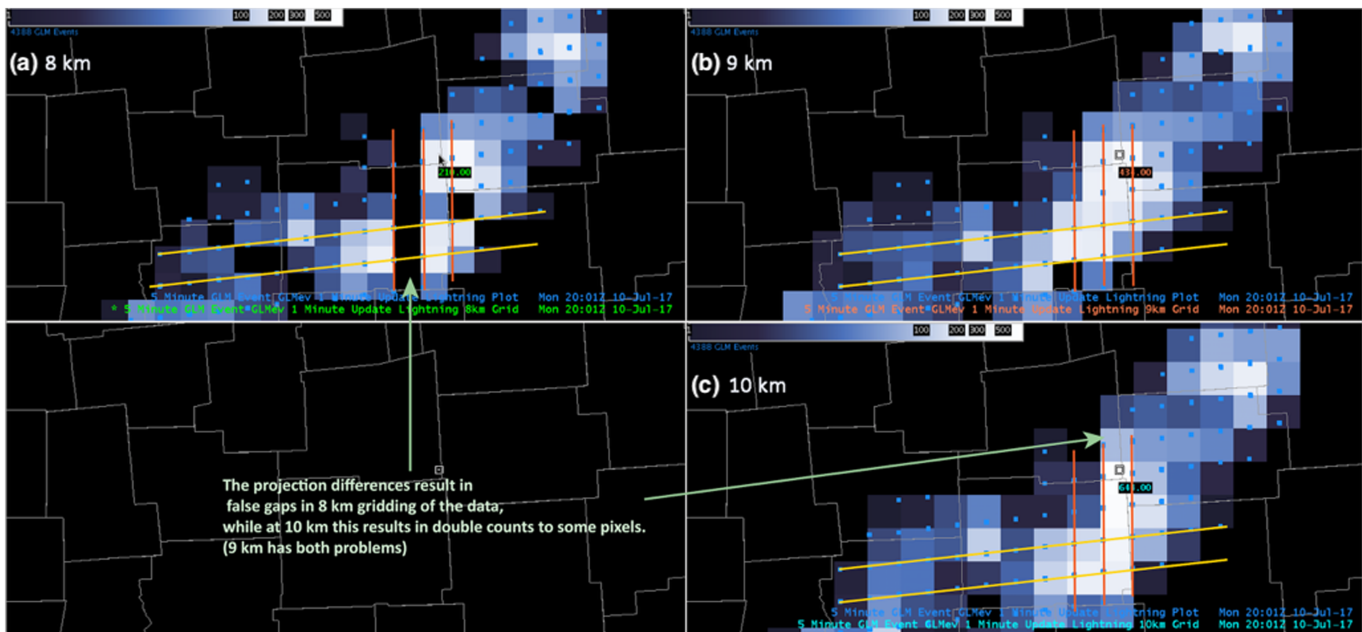
**Figure 2.** Five-minute accumulation of Geostationary Lightning Mapper (GLM) events on a regular latitude and longitude grid in the National Weather Service display software, with equivalent local ground sample distance of (a) 8, (b) 9, and (c) 10 km. Data are from 20:01 UTC on 10 July 2017 over central Ohio. Yellow and orange lines connect GLM event centroids (blue dots). The target grid and implicit event grid (tied to GLM's sensor) are not the same (and cannot be made to match over the whole field of view), resulting in spatial aliasing. The result is gaps at 8 km, double counting at 10 km, and both problems at 9 km.

In summary, a gridded display requires careful remapping of the actual spatial footprint of each GLM event onto a target grid to account for the variable pitch GLM CCD. The approach described next results in a gridded display that is, effectively, an oversampled (nominal 2-km pixels), antialiased remapping of the GLM CCD (nominal 8-km pixels) to a target grid. While many approaches exist for image remapping and interpolation, the GLM data are reported as points without reference to the underlying CCD, preventing the use of such approaches. Instead, each event's spatial footprint is accumulated on the target grid directly.

## 3. Definition of the GLM Imagery Products

Figure 3 shows six products that were created by accumulating the centroid, extent, area, and energy of the GLM flashes, groups, and events for 1 min. The centroid and extent products are counts, with the group counts being greater than or equal to the flash counts at every location. Also shown are all strokes detected by the Vaisala Global Lightning Dataset (GLD)360 network (Said & Murphy, 2016; Said et al., 2010) in the same minute. The GLD360 strokes will be discussed later as part of an analysis of the structure of the thunderstorm complex from which this example snapshot was extracted.

Group (and flash, not shown) centroid density (GCD and FCD) is shown in Figure 3f and is a count of the circles in Figure 1. When accumulated on the 2-km target grid, the centroids are quite sparse, making them ill-suited for visualization by comparison to the gridded products in the other panels of the figure. Furthermore, one- or two-event groups (or flashes) have radiance-weighted centroids that align with the GLM CCD (for instance, just below the "2" in the color bar), leading to a less-than-random appearance to the centroid distribution. While less desirable for direct human inspection, the centroid products have the advantage that the total flash or group count can be determined for any geographic subdomain by simple summation. Therefore, the centroid grids are expected to have utility when constructing automated machine queries of flash rate in, for instance, a storm cell. We chose not to implement the event centroid density product. All event centroids were aligned to the 2-D grid of GLM CCD pixels, leading to a product that looked quite similar to the events in Figure 1. The accumulation process for the flash and group centroids on a target grid requires a straightforward identification of the single target grid cell to which that centroid belonged and will not be discussed in detail in this study.

In contrast to the centroid densities, the flash and group extent products (FED and GED) illustrate how lightning discharges illuminated the parent thunderstorm. Quantitatively, they count the number of flashes
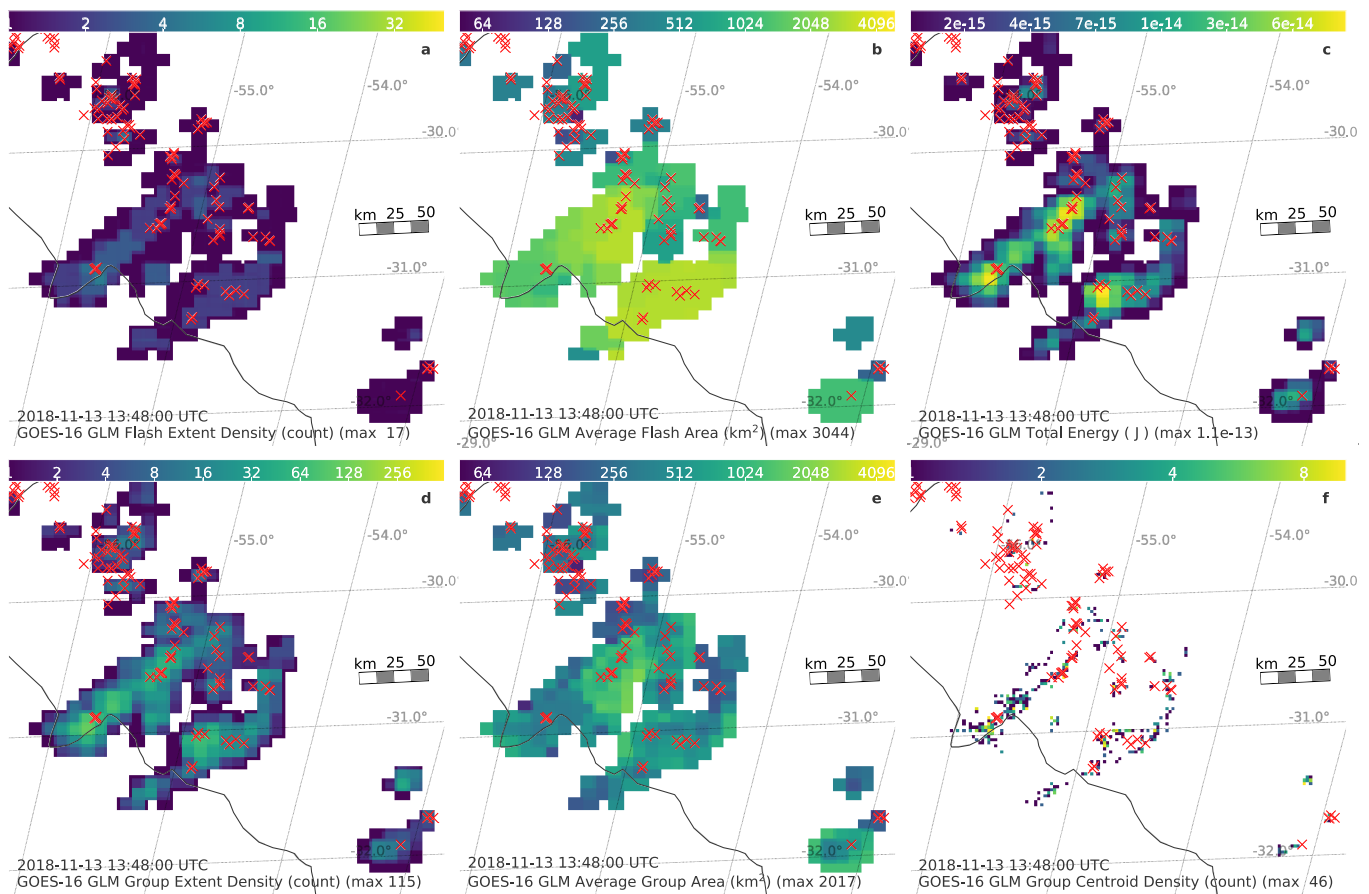
**Figure 3.** Geostationary Operational Environmental Satellites-16 Geostationary Lightning Mapper gridded imagery from 1348–1349 UTC on 13 November 2018 near the border of Uruguay and Brazil, corresponding to the points in Figure 1. Six accumulations of the point data are shown: (a) flash extent density (count), (b) average flash area (square kilometer), (c) total energy (joule), (d) group extent density (count), (e) average group area (square kilometer), and (f) group centroid density (count). Reference latitude–longitude lines are indicated with gray dotted lines. Each red "X" marks a Global Lightning Dataset 360 stroke detected during the same minute. The strokes were assumed to be emitted from the optimized lightning ellipsoid so they would align with the Geostationary Lightning Mapper observations.

or groups that passes through a given location. Flash and group accumulation on the grid shows the GLM event footprint. The jagged shape of the storm cells is due to the shape of the events and so represents visually the method of detection by the GLM CCD. (Weather radar displays utilize a similar approach: the azimuthal and range resolution of the data are represented with varying, spatially extensive pixel sizes.) By definition, each location touched by a group has only one event, so event extent density and group extent density are identical.

The AFA and average group area (AGA) products exhibit the same spatial footprint as the extent products but instead of counting the number of flashes or groups, they depict the average area of flashes or groups that illuminated a given location. AFA and AGA are equivalent to an area-weighted FED and GED, respectively, and as shown below, are calculated by accumulating the total area before dividing by the count. An average event area product was not produced because it would have contained no meteorological signal, as event areas are a fixed characteristic of the GLM CCD.

Finally, total optical energy (TOE) is constructed from the energy of each event and spread uniformly across each event footprint. It has the intuitive interpretation of the illumination of the storm as observed at the GLM instrument—strictly, it is the total energy within the 777.4-nm oxygen triplet received at the GLM entrance pupil. Note that the minimum detectable TOE is expected to vary as a function of the triggering threshold, which varies with position across the CCD and as a function of background illumination, as with LIS that had an observed diurnal dependence (Peterson & Liu, 2013).
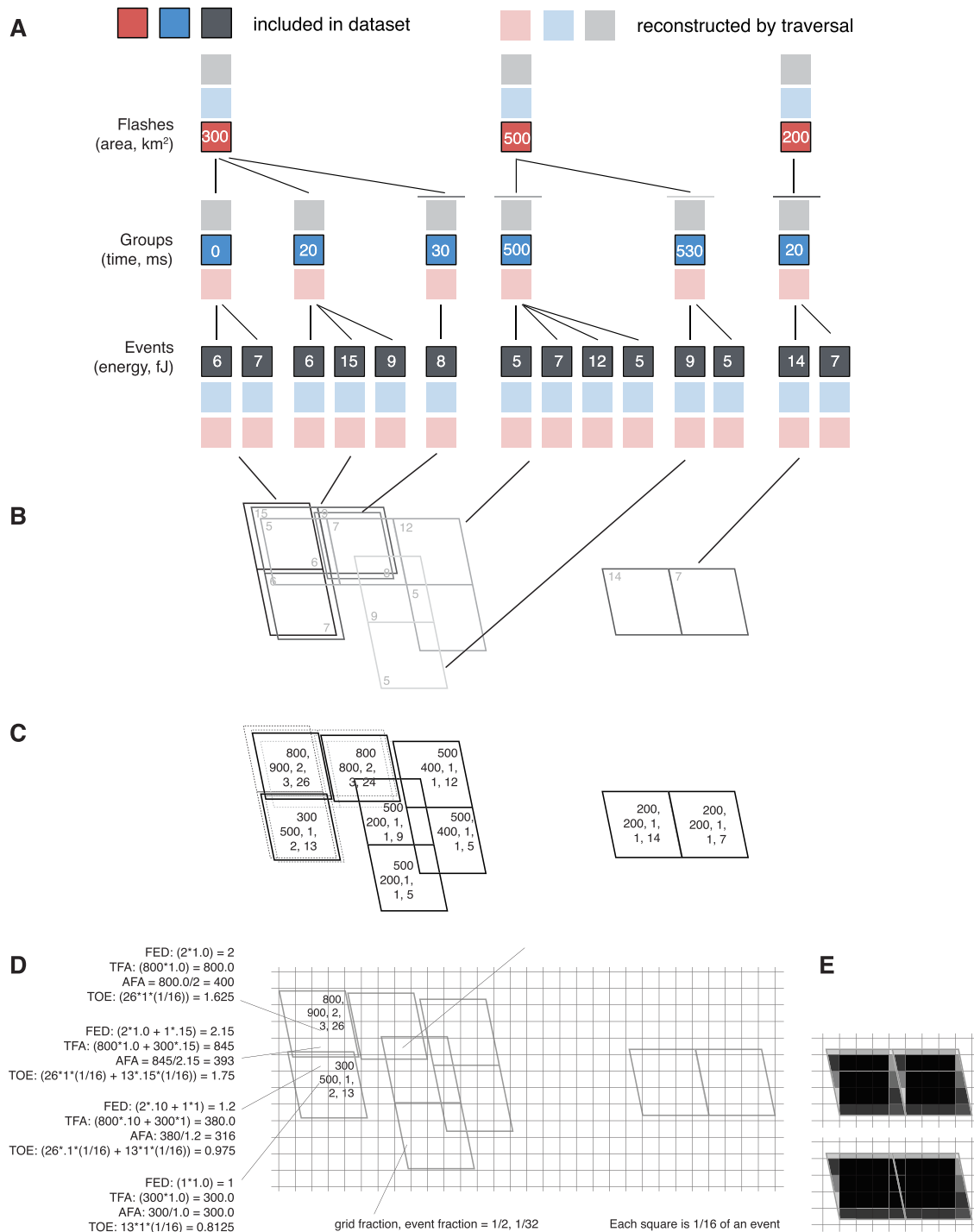
**Figure 4.** (a) The flash/group/event hierarchy. Saturated shading at each level represents data pertaining to that level, while desaturated shades indicate data from parents that have been replicated or from children that have been summarized. Sample flash areas, group times, and event energies are indicated. (b) Spatial representation of event centroids (circles) and footprint quadrilaterals (artificially tilted to exaggerate the misalignment with the target fixed grid), including (exaggerated) jitter, and the fifth group at 530-ms illustrates a large, nonjitter navigation excursion. Lines link each group in the hierarchy to its spatial representation. (c) Spatially aggregated events (solid lines) and original event positions (dotted lines). Numbers indicate aggregated total flash area (square kilometre), total group area(square kilometre), flash count, group count, and energy (fJ). (d) Slicing of events by the target grid. Sample calculations for flash extent density, total flash area, average flash area, and total optical energy are shown for four grid cells. (e) Color-shaded visualization of (top) $f_t$ for the subevent polygons and (bottom) sum of $f_t$ for all subevent polygons intersecting the target grid.

Spreading the total flash or group energy across their constituent events would only result in a smoothed version of the event total energy. There may be some meteorological signal in the average flash or group energy but with no clear rationale from prior work, study of such products is left for future work. At the same time, it may also be of interest to examine the standard deviation of flash sizes and energies at a given location. Bruning and Thomas (2015) examined moments of the flash size (square root of area) distribution and noted a relationship between them and a simple model of electrostatic energy and to the fractal length of the lightning channel. It is an open question whether their area-based energy model can be reconciled with optical flash energies and areas, though a similar statistical approach may prove fruitful.

## 4. Procedure for Constructing GLM Imagery

The GLM imagery is constructed as though the observed optical pulses are accumulated at the CCD in its native coordinates and then resampled to a final image with regular pixel spacing. The final image mimics the coordinate system and resolution of the ABI L1b and L2 imagery for ease of overlay on those data. The processing method described in this study is implemented in the Python programming language in a package we call `glmtools` (Bruning, 2019). As an open-source package, it is available to any and all researchers. Additional implementation details are provided in Appendix A, which discusses how we handled the quasi-standard scaled integer encoding used in several GOES data sets and references the open source tools upon which we built our method and to which we contributed.

Figure 4 illustrates the steps in the gridding process. Each of the panels a–e will be described throughout the rest of this section.

### 4.1. Traversal of the Flash/Group/Event Hierarchy

The first step in accumulating the GLM data requires reconstruction of the event–flash linkages from the minimally sufficient data provided at GLM L2 (Figure 4a).

The structure of the GLM data as a hierarchical point data set (stored as three flat data tables linked by common identification integers) is a one-to-many mapping with a tree-like network structure and is also that of a foreign key schema as used in a relational database (Codd, 1970). Recognition of the formal structure of the data set allowed us to extend standard tooling built around such concepts to facilitate easy traversal of the tree. Three fundamental operations that operated on the data are as follows: (1) replication of properties from parents to (grand)children, (2) summarization of child properties at a (grand)parent level, and (3) pruning of the data set. Appendix A describes the implementation of the traversal method.

We adopt the following notation: an array of values $V$ with length corresponding to dimension $d$ and related to an optional foreign key dimension $k$ is denoted $_kV_d$. For GLM, the entity keys at each level in the hierarchy are $K_f$, $K_g$, and $K_e$, the flash, group, and event IDs, respectively. The corresponding parent ID variables for event and group dimensions are, respectively, $_gK_e$ (event_parent_group_id) and $_fK_g$ (group_parent_flash_id). Implicit in the data set is the $_fK_e$ (event_parent_flash_id; conventionally, this is a grandparent but parent is used for brevity). $_fK_e$ was calculated as part of the processing as it was needed to calculate FED. For each event, $K_f$ was replicated from the flash level to the event level by ascending the hierarchy from event to group (using $_gK_e$ to find the corresponding index in $K_g$) and then group to flash (using $_fK_g$ to find $K_f$.)

The generalized traversal operation was used to build an enhanced GLM reader that automatically calculated variables needed for further processing, yielding an augmented GLM data set that included additional variables as though they were produced by the ground system in the first place. The replication $_fK_e$ and the summarizations of $_gN_f$ (flash_child_group_count), $_eN_f$ (flash_child_event_count), and $_eN_g$ (group_child_event_count) were calculated by default as soon as the GLM L2 data were loaded. Such preprocessing facilitated gridding as well as queries of the L2 data hierarchy for other applications, such as geographical subsetting of a population of flashes (by pruning) and deep-dive analyses of a single flash and its groups and events.

Work on the GLM data set traversal problem led us to recognize that many lightning data sets, as well as other data sets from across atmospheric science (such as storm cell identification and tracking) shared a similar structure. Formal specification of parent–child hierarchy metadata naming conventions in the Climate and Forecast Standards (Eaton et al., 2011) would enable broad, library-level support for traversal operations on any data set with such structure. The GLM reader described above is a proof of concept of such an effort.

### 4.2. Lightning Ellipsoid Removal and Navigation to Fixed Grid

The second GLM preprocessing step is calculation of the fixed grid coordinates corresponding to the geolocated GLM L2 data. The GLM events are geolocated by assuming that light is emitted at a standard cloud top height corresponding to a virtual "lightning ellipsoid"(GOES-R Series Program Office, 2019) whose equatorial and polar heights are modified from the Geodetic Reference System 1980 ellipsoid that defines the Earth's surface in the GOES coordinate reference system (GOES-R, Series Program Office, 2017). The latitude and longitude reported at L2 is determined by identifying the corresponding point on the ground that is below the lightning ellipsoid (Figure 5). At launch, the assumed cloud top height varied from 16 km at the equator to 6 km at the poles and was lowered to 14 km on 15 October 2018 to minimize time-averaged location errors (K. Virts, personal communication; article expected in this Journal of Geophysical Research special issue). After the adjustment, the ground-referenced locations using the revised lightning ellipsoid align better with other ground-referenced data like weather radar or ground strike detections (Cummins & Murphy, 2009).

On the O(1) min time scales appropriate for meteorological observations of thunderstorms, variable cloud top heights are observed and result in some residual error in the ground-referenced geolocation (Figure 5). In real-time multisensor displays, which might include ABI imagery, GLM, and ground-based data and maps, use of the lightning ellipsoid results in offsets of the GLM data from each of the other data sets. In such a situation, where users are already used to dealing with satellite parallax, it is desirable to reduce the number of possible offsets back to 1 by navigating the GLM data to align with the ABI data on the fixed grid.

The fixed grid coordinates are the east and north angular position of each GLM detection with respect to the subsatellite point. The GOES fixed grid coordinates ($x$, $y$) do not require any assumption about ellipsoid height because they are simply the direction in which the instrument observed an event. The GLM CCD is also aligned along ($x$, $y$), making the fixed grid a natural coordinate system for subsequent calculations. The GLM L1b processing computes and utilizes the fixed grid coordinates then discards them during L2 processing. Therefore, the fixed grid coordinates of the flash, group, and event centroids are calculated from the L2 data by inverting the equations of Bezooijen et al. (2016).

GLM image domains matching the ABI mesoscale or CONUS sectors have flashes that cross the edges of the image. To account for this possibility, the event centroid ($x_e$, $y_e$) is used to find all event centroids within the field of view. The $_fK_e$ of these events are reduced to their set of unique $K_f$, which are then used to prune the data set to only those flashes overlapping the domain of interest. This feature allows for the stitching of individual image tiles if necessary and shows how the availability of $_fK_e$ combined with pruning functionality facilitated correct processing of the data. By contrast, pruning on a flash centroid ($x$, $y$) falling to the left of a tile edge would leave out events to the right of the tile edge when creating the left tile; creation of the right tile would not find the missing events because their corresponding flash centroid was not within the right tile.

### 4.3. Event Spatial Precision and Aggregation

The removal of the lightning ellipsoid and navigation to fixed grid coordinates remove some geolocation uncertainties, but spacecraft jitter, thermal distortion throughout the diurnal solar cycle, and navigation uncertainty related to coastline identification mean that the GLM CCD position and orientation vary with time with respect to the fixed grid. The geolocation data available at L2 incorporate spacecraft attitude information and the best available coastline navigation and thermal modeling, applied prior to L1b, and accounts for such motions.

The net effect is that each group (and its events) have a slightly different location in fixed grid coordinates, and processing each of these positions separately was too slow. However, at most times, the event motions within 1 min were much less than the 56-microradian target (nominally 2 km) fixed grid size, permitting aggregation of flash, group, and event properties at a representative event location. Specifically, median motion from typical jitter is 8–9 microradians, except when spacecraft station-keeping activities (typically less than 50 microradians) or motion of the solar arrays needed to calibrate the GOES-16 space environment sensors (Solar Ultraviolet Imager and X-ray Irradiance Sensors; typically less than 250 microradians, about the size of a GLM pixel) are performed. Figure 4b depicts an exaggerated version of variability of the event centroids and the shape of each event as eventually reconstructed during a later step.
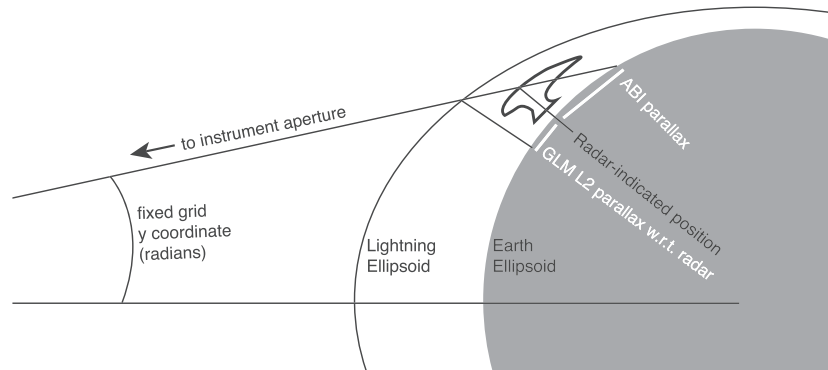
**Figure 5.** Illustration of the navigation error that results from differences between the assumed lightning ellipsoid height (exaggerated) and the actual height of light emission at cloud top. The cross-section is through the Earth (shaded), with North Pole up; the satellite position is above the equator to the left. Light from cloud top travels toward the satellite at some fixed grid coordinate angle and intersects both the lightning ellipsoid (used to navigate the Geostationary Lightning Mapper L2 position) and the earth ellipsoid (as is used for creating Advanced Baseline Imagery). The correct ground-relative position of the cloud top as indicated by a weather radar is also indicated.

The space and time scales described above are exploited to discretize the event locations to the nearest $s = 28$ microradians — much larger than the typical jitter but smaller than the larger station-keeping excursions. These new, discretized locations are associated with an aggregated event ID $K_a$ along a new dimension in the data set and calculated as

$$x_s = \text{int}[(x_e - x_0)/s], \tag{1}$$

$$y_s = \text{int}[(y_e - y_0)/s], \tag{2}$$

$$x_{s1} = \text{int}[(x_1 - x_0)/s], \tag{3}$$

$$K_a = x_s + y_s * x_{s1}, \tag{4}$$

where $(x_e, y_e)$ is the fixed grid coordinate of the event; $x_0, x_1, y_0$, and $y_1$ give the minimum and maximum possible $x$ and $y$ values; $x_s$ and $y_s$ are discrete fixed grid coordinates; and $x_{s1}$ is the maximum possible discretized value of $x$.

A new variable along the event dimension $_aK_e$ (event_parent_lut_id) links the original event to its aggregate. $x_s$ and $y_s$ are only needed to calculate $K_a$. For each $K_a$, a centerpoint location $(x_a, y_a)$ is calculated from the mean of the constituent events' fixed grid coordinates $(x_e, y_e)$.

As illustrated in Figure 4c, the result is that most events consolidate together locally, while more sporadic larger excursions, when they happen, result in a few separate events displaced substantially from the other aggregations.

An alternate approach would have been to aggregate the events based on their source pixel on the CCD (though such information was not available at L2). That would have been sufficient to account for typical jitter, but larger excursions would have been located incorrectly unless the accumulation were performed at a sufficiently small time slice. Accumulation by CCD pixel would have then required the use of image interpolation routines to map the CCD image to the fixed grid, but remapping a dense image at a possibly high frame rate would have presented other performance challenges. The approach described here is instead able to exploit the sparse nature of the GLM observations.

### 4.4. Accumulation of Flash and Group Properties
The aggregated event dimension of the data set is used to summarize the flash, group, and event properties at the aggregated event location, as shown by the numbers for each event in Figure 4c.

$_fK_e$ and $_gK_e$ for each event are reconstructed in the initial hierarchical traversal and when combined with $_aK_e$, permit the aggregation of flash and group level properties. Spatially extensive flash and group rates
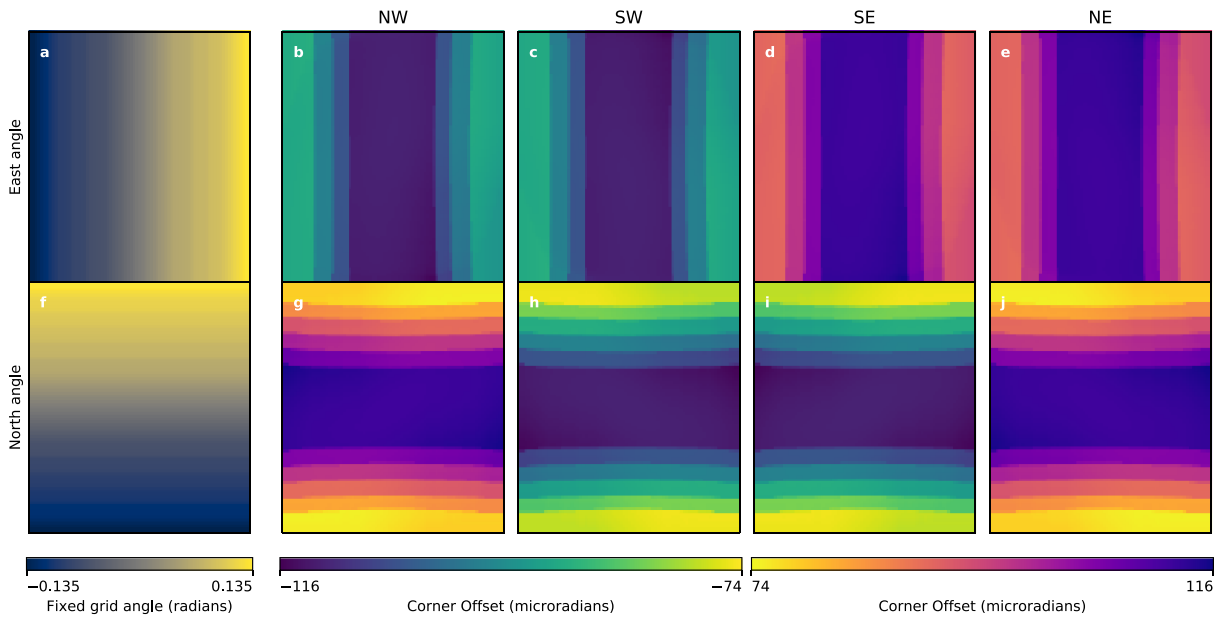
**Figure 6.** The corner point lookup table used to reconstruct the event quadrilateral. Each panel is an 87 × 77 array of values corresponding to the typical characteristics of the Geostationary Lightning Mapper charge-coupled device at that position. Positions are indicated by the (a) fixed grid east coordinate and (b–e) the corresponding corner point offsets for the fixed grid east coordinate in the northwest, southwest, southeast, and northeast directions; (f) fixed grid north coordinate and (g–j) the corresponding corner point offsets for the fixed grid north coordinate in the northwest, southwest, southeast, and northeast directions.

$(_aN_f, _aN_g)$ are calculated by counting the number of unique $K_f, K_g$, respectively, at each discretized event location. At each discretized event location, the total flash and group areas $_aA_f, _aA_g$ and event energy $_aJ_e$ are found by simple summation. The total area is used later to calculate the average flash and group areas on the eventual target grid.

Each of these totals are added to the internal data set structure along the aggregated event dimension. At this stage, preprocessing ends and the event accumulation on the target grid begins.

### 4.5. Slicing

Accumulation requires remapping the geometry of the discretized events onto the target grid. The first step is to determine the four corner points of each event and the overlap of that event with the target grid, as depicted in Figures 4D and 4E. The geometry of the overlap region is determined using a polygon clipping approach.

Because they are not available at L2, the event shapes are reconstructed using corner points provided as an 87 × 77 lookup table defined in fixed grid coordinates every 1,000 microradians. For any given $(x_a, y_a)$ in fixed grid coordinates, the corner points are linearly interpolated from this table. A lookup table is used instead of a direct, analytic reconstruction because the pixel shapes included the less-than-straightforward combined effect of lens barrel distortion and the variable pitch configuration of the GLM CCD.

Figure 6 shows $(x, y)$ for each entry in the lookup table and the corresponding offsets for each of the corners that together define the quadrilateral for each GLM event. The interpolation approach introduces a small amount of error that is much less than the size of an individual GLM pixel. Median errors are less than 50 m, and the 99.99th percentile error was less than 500 m.

Each nominal 8 × 8-km GLM event corresponds to about 16 2 × 2-km target grid cells (25 for a 10-km event), plus any additional partially overlapped target grid cells. A contiguous span of target grid cell indices $(g_{xi}, g_{yi})$ (and their associated corner points) are identified by converting the corner points of each aggregated event to $(g_{xi}, g_{yi})$ and then padding the index by two grid cells in each direction to ensure full coverage of each event by the target grid. For each aggregated event, the result is a set of target grid cell quadrilateral corner points.

The next step is to find the intersection of each target grid quadrilateral with its aggregated event's quadrilateral. For each aggregated event, a list of zero or more subevent polygons (one for each overlapping target grid quadrilateral) is generated. The subevent polygons are the small polygons illustrated in Figure 4d, and determining their variable shape requires the use of a polygon clipping algorithm. Each subevent polygon is tagged with the ID of its parent aggregated event $K_a$, that is, $_aK_p$ is added for each subevent polygon.

For each subevent polygon, a variety of additional properties are calculated (Figures 4D and 4E). The $(g_{xi}, g_{yi})$ associated with each subevent polygon is preserved. The centroid position $(x_p, y_p)$ (mean of the polygon coordinate vertices) and area $A_p$ of each subevent polygon are calculated. The fractional coverages $f_e = A_p/A_e$ (of the aggregated event) and $f_t = A_p/A_t$ (of the target grid cell) are determined. Note that $f_e = f_t A_t/A_e$, where the ratio of the grid cell and the event areas are locally approximately constant, though we did not make use of this fact. $A_t$ could be calculated once for all grid cells, but $A_p$, $f_t$, $f_e$, and $A_e$ can change on a per-event basis depending on how the event shifts with respect to the grid and where the event is located in the field of view.

### 4.6. Accumulation on Target Grid

Accumulation on the target grid makes use of the preaccumulated properties for each discretized event polygon and weighs them by either $f_e$ (in the case of total energy) or $f_t$ (in the case of spatially extensive properties such as count or total area). The relevant distinction is whether a quantity is a conserved physical measurable (as with energy) or is a representation of a spatially broadcast property of an entity at some other level in the hierarchy.

The total value for any target grid cell $i$ for any field $VVV$ is the sum of zero or more contributing subevent polygons $j$. For conserved quantities,

$$VVV_i = \sum_j f_{ej}(_aV_d)_j, \tag{5}$$

while for extensive properties,

$$VVV_i = \sum_j f_{tj}(_aV_d)_j, \tag{6}$$

where the $_aV_d$ for each field is as listed in Section 4.4. $V$ is the quantity to be accumulated and comes from level $d$ of the hierarchy. The precursor $_a$ denotes that $V_d$ has been preaggregated at a representative event location.

As shown in Figure 4d, the accumulations of $_aJ_e$ for $VVV =$ TOE, $_aN_f$ for $VVV =$ FED, and $_aN_g$ for $VVV =$ GED are used as the final grids. The accumulations of $_aA_f$ and $_aA_g$ result in total flash area ($VVV =$ TFA) and total group area ($VVV =$ TGA), respectively, and are divided by FED and GED to create AFA and AGA. Also accumulated are the flash and group centroids (FCD and GCD), which lack spatial extent and do not require the polygon clipping information.

### 4.7. Time Accumulation

Production of imagery at 1-min intervals (combining three 20-s GLM L2 files) matches the routine ABI mesoscale sector sampling rate. To obtain other time accumulations, it is not necessary to reprocess the L2 data. Instead, the 1-min grids can be manipulated algebraically to give the desired accumulation interval. For all of the products except AFA and AGA, the accumulation is a simple sum. AFA and AGA are multiplied by FED and GED to retrieve TFA and TGA for each minute, which can then be summed, and finally divided by the summed FED and GED (respectively) for the new accumulation interval.

## 5. Example: A Mesoscale Convective System in Argentina

As of 00 UTC on 13 November 2018, thunderstorms were ongoing over northeastern and eastern Argentina, where upscale organization of convective complexes to the east of the Sierras de Córdoba is common (Mulholland et al., 2018). By 13 UTC, a convective complex had developed and persisted over extreme eastern Argentina, southwestern Brazil, Uruguay, and Paraguay.

Figure 3 already showed 1 min of gridded data from this case. Study of the accumulation of the 1-min grids over varying intervals serves to illustrate the meteorological signals that were detected by GLM through its combination of continuous monitoring and spatial mapping of lightning activity. This example demonstrates
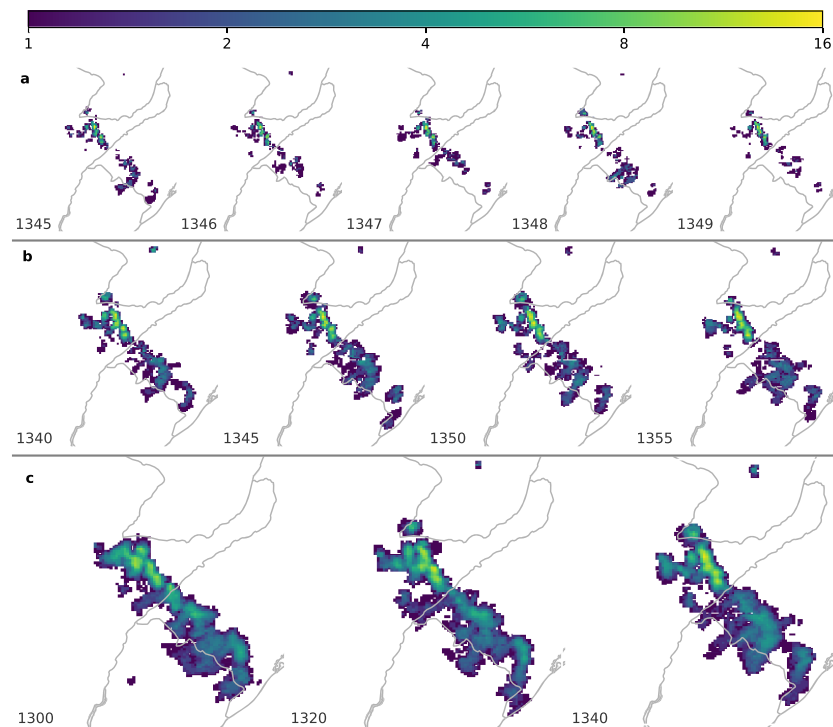
**Figure 7.** Geostationary Lightning Mapper Flash Extent Density on 13 November 2018 during the 1300 UTC hour over southern Paraguay, eastern Argentina, southwestern Brazil, and northern Uruguay. UTC times are indicated in each panel. Panels (a), (b), and (c) show 1-, 5-, and 20-min accumulations of lightning activity, respectively. One panel in each row is an accumulation of the row above it. The color bar indicates flash counts for the 1-min images and is scaled up by 5 and 20 for the other two rows.

lightning-only inferences that could be improved upon through fusion with other satellite data, as well as radar, storm environment, and ground-based lightning detection data where available (e.g., Cintineo et al., 2018; Wang et al., 2012). Such analyses are beyond the scope of this study.

Figures 7–9 show accumulations of FED, AFA, and TOE at 20-, 5-, and 1-min intervals. A portion of the data from 1348 UTC in panel (a) of each figure corresponds to Figures 1 and 3. Consistent with findings in preoperational trials in the United States, the 1-min accumulations of lightning products were patchy, while the 5- and 20-min accumulations more clearly showed a northwest-southeast line of storms and the system-scale motion to the east. For reference, Figure 10 shows all strokes detected by the Vaisala GLD360 network at corresponding times, as also shown in Figure 3.

### 5.1. Qualitative Meteorological Interpretation

Complementary signals in each product combined to give confidence in the meteorological interpretation of the imagery. The linear structure, long life, and large extent of the convective system suggests that it was a mesoscale convective system of the leading-line, trailing stratiform type (Biggerstaff & Houze, 1991; Parker & Johnson, 2000). In such a system, deep convective drafts with the largest vertical velocities would be expected in the leading line, where relatively large flash rates and small flash sizes (Carey et al., 2005; Bruning & MacGorman, 2013; Steiger et al., 2007) are found. To the rear, less frequent but larger flashes would be expected in trailing stratiform precipitation.

The general pattern of FED and AFA (Figures 7 and 8) was consistent with expectations from storm mode. Furthermore, the northwestern part of the line had much higher flash rates and smaller average flash sizes and relatively infrequent large, trailing flashes. The southeastern part of the line had consistently larger flashes along the convective line and more frequent and larger average flashes into the trailing stratiform region. The duration of these stratiform flashes in some cases exceeded the 3 s clustering algorithm limit, leading to artificial splitting of larger flashes and larger FED values than a human analyst would infer. This structure suggests that the mixed-phase updraft volume was greater in the northwestern part of the line (esp. Figures 24c and 24d of Schultz et al., 2015), while hydrometeor sedimentation, front-to-rear transport, and related stratiform region precipitation processes were more dominant in the southeastern part of the line.
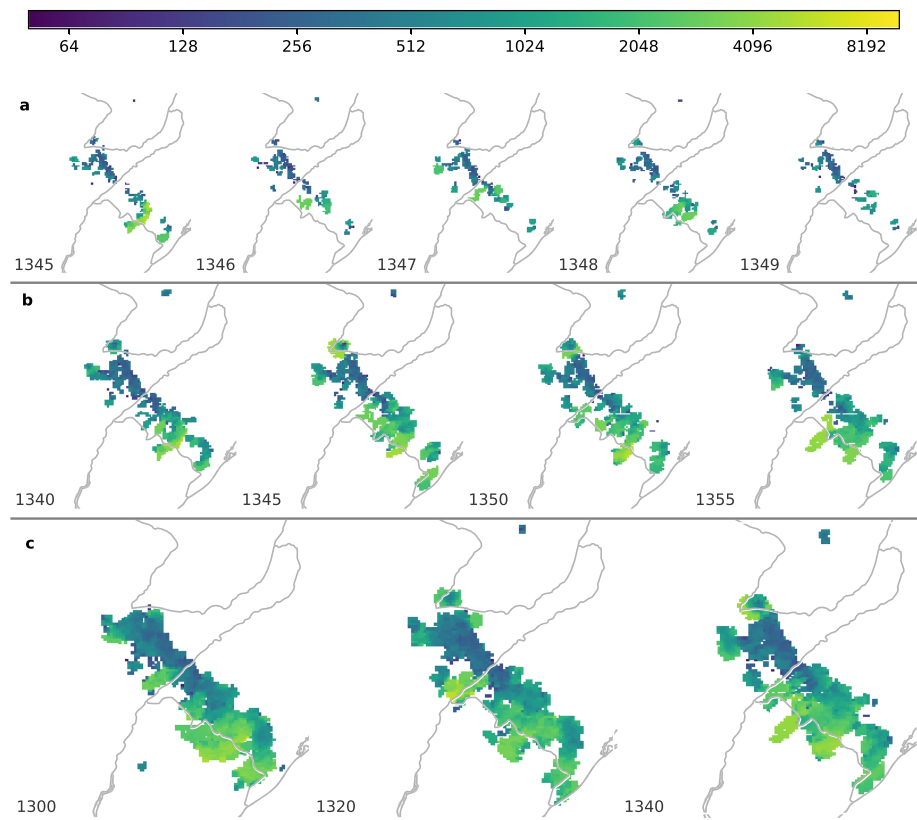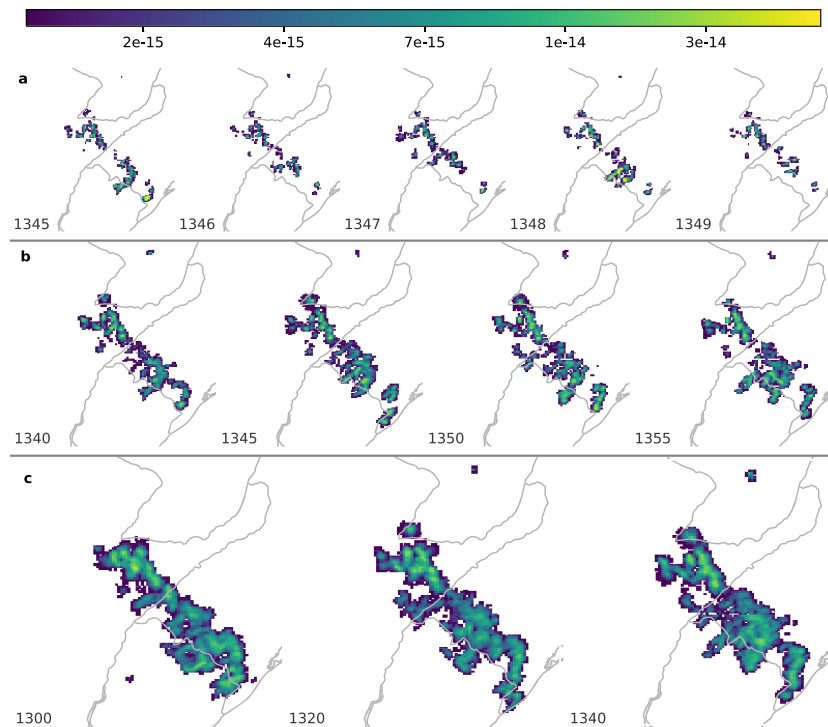
**Figure 8.** Geostationary Lightning Mapper Average Flash Area on 13 November 2018 during the 1300 UTC hour over southern Paraguay, eastern Argentina, southwestern Brazil, and northern Uruguay. UTC times are indicated in each panel. Panels (a), (b), and (c) show 1-, 5-, and 20-min accumulations of lightning activity, respectively. One panel in each row is an accumulation of the row above it. The color bar indicates the average area (square kilometer) with the same scale for all panels.

On long time scales, TOE (Figure 9) was brightest in the northwestern part of the line, and consolidated in relatively small cells. However, compared to the other two products, the relative difference between the convective line and the trailing stratiform region was not as visually obvious in total energy. The flashes at 1348 and 1349 UTC illustrated how this pattern occurred: at that time, a few bright flashes with large areal extents to the southeast emitted more light than the dozens of flashes in the convective line to the northwest. Bruning and Thomas (2015) noted that large average flash size would be associated with more energetic flashes, though the large flashes at 1345 and 1348, with similar area, also had very different optical output.

A comparison of the lightning stroke data from the GLD360 network (Figure 10) to the GLM data (Figures 7 and 8) shows that overall pattern of lightning and the location of relatively high flash rates is shared by the two lightning locating systems. A close examination illustrates the benefit of the GLM data in connecting the paths of extensive lightning channels to their corresponding stroke locations.

First, consider the zoomed portion of the 1348 UTC minute in Figure 3, where two extensive chains of groups extend rearward from the convective line. The FED is low, but the flash area, total energy, and group counts are large, suggesting that only a few flashes generated a large number of very bright groups in quick succession. The GLM group centroids confirm this impression, nearly continuously tracing the apparent path of lightning channels that extend for over 200 km. Along these channels, GLD360 strokes confirm the ground connection points of these extensive discharges, which in this case also tended to be colocated with local maxima in total energy.

Consider also the 20-min aggregations beginning at 1320 UTC and 1340 UTC in Figures 7c – 10c. Immediately south of the highest flash rates, an extensive set of GLM channels with large AFA protrudes from the trailing edge of the convective line and crosses the intersection of Argentina, Uruguay, and Brazil. The GLD strokes associated with those channels remain entirely along the border of Argentina and Brazil

**Figure 9.** Geostationary Lightning Mapper Total Optical Energy on 13 November 2018 during the 1300 UTC hour over southern Paraguay, eastern Argentina, southwestern Brazil, and northern Uruguay. UTC times are indicated in each panel. Panels (a), (b), and (c) show 1-, 5-, and 20-min accumulations of lightning activity, respectively. One panel in each row is an accumulation of the row above it. The color bar indicates total energy (joule) for the 1-min images and is scaled up by 5 and 20 for the other two rows.

and then bend back north. The GLM illumination shows that this set of trailing channels also branched further south into Uruguay but did not connect to the ground per GLD360. More flashes were detected by LIS than GLD360 in South America in a study by Said and Murphy (2016), and since none of these systems has perfect flash detection efficiency (Bitzer & Burchfield, 2016) and they observe somewhat different parts of the lightning process, the differences observed here are not surprising. Regardless, such patterns show how GLM can connect ground strike location data to the total extent of the electrified cloud. Note also that the dense cluster of strokes on the western end of the line-trailing channels corresponds to approximately equivalent brightness, somewhat higher flash rates (in both data sets), but much smaller average flash sizes, distinguishing the extensive channels (likely in stratiform cloud) to the east from what might be an embedded, deeper convective cell trailing the line.

### 5.2. Quantification

The qualitative meteorological interpretation suggested that lightning in the trailing, more stratified clouds to the rear of the convective line was larger. Utilizing LIS data, Peterson et al. (2017) and Peterson et al. (2018) introduced the idea of propagating flashes and suggested that they were a separate class of discharge where group centroids moved in a preferential direction away from the first group so as to elongate the flash footprint. They also diagnosed another class of very bright flash with large illuminated area, so-called "superbolts," in which the group centroids did not propagate. They are probably related to the very large, single-stroke discharges in Holzworth et al. (2019) and references therein. In this section, we evaluate whether a quantitative assessment of the imagery can distinguish at least two separate classes of lightning in this case from Argentina or if the flashes steadily get larger in the tail of a distribution of a single population of lightning discharges.

Figures 11a, 11b, 11d, and 11e show the joint distribution of FCD, FED, GCD, and TOE with average flash area for 13–14 UTC on 13 November 2018. The joint distribution was calculated by counting the number of 2-km pixels having that pair of values during each 5-min accumulation interval in the hour. In the density versus AFA plots, the logarithmic scaling visually compresses what are actually very long tails of low-rate,
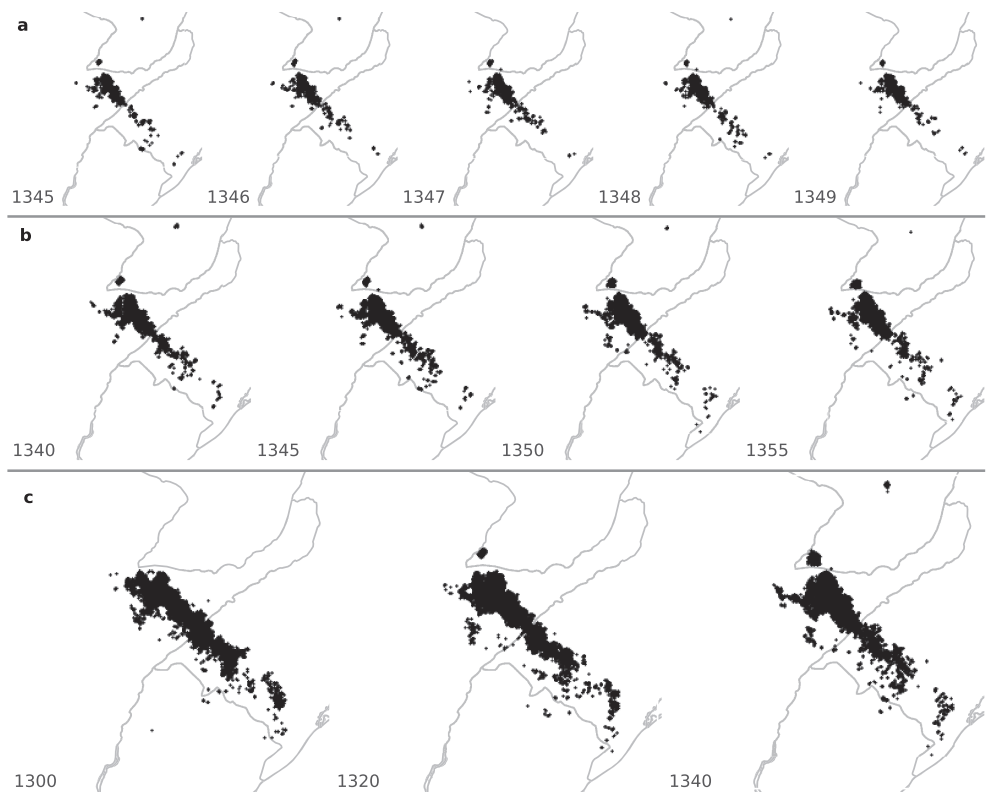
**Figure 10.** Global Lightning Dataset 360 strokes detected on 13 November 2018 during the 1300 UTC hour over southern Paraguay, eastern Argentina, southwestern Brazil, and northern Uruguay. UTC times are indicated in each panel. Panels (a), (b), and (c) show 1-, 5-, and 20-min accumulations of lightning activity, respectively. One panel in each row is an accumulation of the row above it.

large area and high-rate, small area pixels. This distribution explains why a logarithmic color mapping tends to work well for display of GLM imagery, as was empirically determined in operational trials of these products. TOE also spans a few orders of magnitude, though the range of values observed is similar regardless of AFA.

A reference line is drawn in each panel of Figure 11 at 272.25 km$^2$, which is the square of the 16.5-km separation threshold used in the GLM flash clustering algorithm (Goodman et al., 2010). The other reference line is at 1,000 km$^2$ and separates two peaks in the GCD and TOE distributions (Figures 11d and 11e). Above 1,000 km$^2$, the 5-min flash and group rates are $\leq 10$ and almost always only a few. The peak in the large flash counts occurred at $\approx 2,500$ km$^2$ and $2 \times 10^{-14}$ J, and the most common GCD value at that size was one group centroid per 5 min. There was no secondary peak in FCD (Figure 11a) among the large flash pixels, but a few flash centroids were observed at those sizes. The FED–AFA pair with the largest counts also occurred at 2,500 km$^2$ at values of one flash per 5 min (Figure 11b). As flash rates increased, the large flashes in FED became less common, and the most common flash area value at large flash rates converged on the 272.25-km$^2$ line.

Large flashes are necessarily overrepresented in the count of all *FED* and pixels (Figure 11b) because they by definition cover more pixels. (There is also a uniform, artificial $\approx 25\times$ increase in the total pixel counts for a 10 km pixel because the distributions are drawn from the individual 2 km pixels that make up the images, and the $\approx 10$-km GLM pixels are oversampled to a final 2-km resolution.) Therefore, Figure 11c shows counts for only those pixels having flash centroids so as to have only one sample of the distribution per flash. After removing the artificial overcounting of large flash pixels, Figure 11c shows that the FED–AFA pair with the largest counts has moved to $\approx 25 - 50$ per 5 min at 272.25 km$^2$. At lower flash rates, the counts decreased smoothly from this peak toward a wide range of flash areas.

Figure 11f is analogous to Figure 11c but for those pixels having group centroids. The FED distribution once again has a maximum within the high flash rate tail, but a secondary maximum is restored at $\approx 2,500$ km$^2$
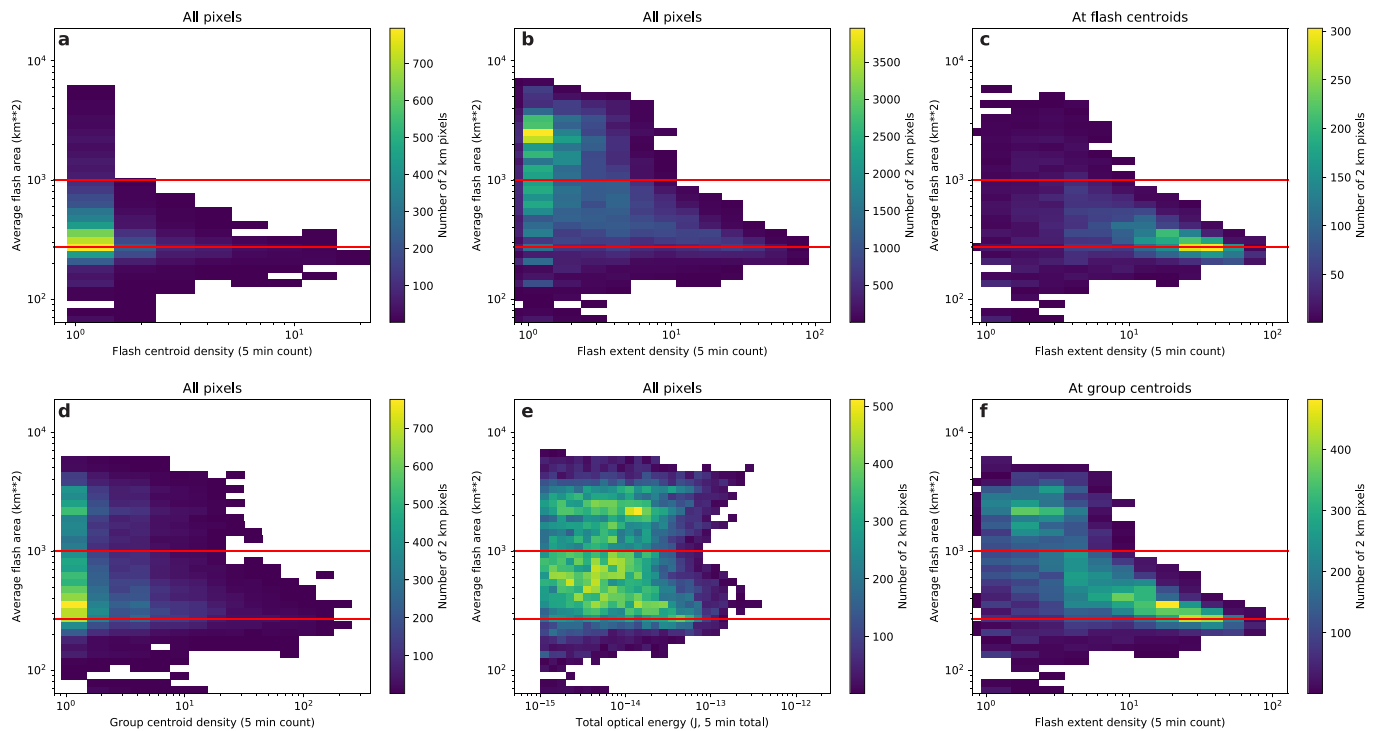
**Figure 11.** Joint distribution of (a) flash centroid density, (b, c, f) flash extent density, (d) group centroid density, and (e) total optical energy with average flash area showing the number of 2-km pixels in each 5-min accumulation having that pair of values. Panels (c) and (f) use only those 2-km pixels where there is at least one flash or group centroid, respectively. The number of 2-km pixels is ≈ 25× the number of Geostationary Lightning Mapper charge-coupled device pixels for a 10-km Geostationary Lightning Mapper pixel.

and at rates of a few flashes per 5 min. While large area flashes are once again overrepresented on a per-flash basis, they are overrepresented in proportion to the number of groups that occurred.

The question posed at the beginning of this section was whether large flashes are a distinct population from others in the storm, and the answer depends on whether one examines the data in Figure 11c alone or in concert with Figures 11d–11f. Joint utilization of the group centroid data alongside flash area and FED helps establish that there is a secondary population of flashes that in some locations have frequent groups. Since groups count the number of times current flowed along a lightning channel (in a way that emitted light in excess of GLM's detection threshold), there is a physical basis for making the analytic choice to favor those low flash rate regions that flicker more. Furthermore, the findings herein are consistent with the definitions of Peterson et al. (2017) and Peterson et al. (2018), who defined propagating flashes in terms of the behavior of group centroids. Finally, these flashes are probably not superbolts. The enhancement of the secondary maximum in the FED versus AFA parameter space relies on the presence of groups in spatially separated pixels, while superbolts illuminate a very large area due to a single very bright group that is likely to coincide with the location of other group centroids. Future work might study the parameter space combinations (involving TOE) that are associated with known superbolts and the underlying cloud structure that gives rise to them.

The statistics presented herein demonstrate a purely image-based approach that might be suitable for bulk detection and classification of the parts of storms that contain propagating lightning flashes. An image-based approach has advantages in the limited computational environment of an operational setting since explicit reconstruction of the group connectivity graph is not required.

### 5.3. Discussion
The analysis in this section was focused solely on 1 hr of multiparameter flash imagery from a large mesoscale convective system in Argentina. A few other comments are in order as future work is considered.

First, it is not clear whether the convergence of high flash rate pixels within the convective line on a characteristic flash size is a physical characteristic of the lightning or if it is an artifact of the instrument resolution and/or the group-to-flash clustering threshold. Is the clustering algorithm configured optimally so as to reveal the most physically relevant flash size or would we see a decrease to smaller flash sizes if smaller optical pulses could be observed? Repeating the work of Mach et al. (2007) with GLM is the necessary first step in evaluating the effect of algorithm choices and that work is underway (D. Mach, personal communication). Furthermore, the scene-background-dependent detection threshold and/or cloud structures that systematically attenuate and scatter light may contribute to the lower envelope of flash size at high flash rates, as well as to the final illuminated area of larger flashes.

Second, the exact area, group, and rate thresholds that separated the two lightning classes may be unique to this storm day or storm mode; and for some modes, a second class of lightning may not be present at all. How did the separation of the two classes of lightning emerge with time and are there other patterns intrinsic in the space and time evolution of the flash parameters that characterize other storm modes? Such questions are beyond the scope of this study, but there appears to be a rich vein of meteorologically driven inquiries that could be asked of the GLM data.

Third, affirmation of two classes of lightning in this study, in combination with recent studies of an apparently separate class of superbolts, is a helpful reminder that the lightning discharge (and its detection by any given instrument) is a varied process. New measurements of lightning are revealing that it might be ill-posed to discuss a single population of lightning flashes with some statistical variability around a typical discharge. That perspective implies shared flash dynamics among the whole population and further ignores how the lightning discharge is convolved with the underlying cloud structure of each thunderstorm. The meteorology drives the creation of electrical energy and its distribution within the cloud, on time scales from 1 s to hours and on spatial scales from single ice crystals to 100 s of kilometers. Electrical energy is then discharged by processes that similarly span several orders of magnitude of space and time (e.g., Mazur et al., 1998). In the case of GLM, multiple scattering through varied cloud optical depths further modifies the signal directed at the satellite. In the mesoscale convective system from Argentina, similar TOE was produced by very different FED and AFA optical energy as a function of position along the line, and two large flashes had similar area but different optical output, illustrating the practical expectation that flash-to-flash variability appears on meteorological scales. Simply counting flashes glosses over the varied information available from flash to flash and suggests there is opportunity to link that variability to the cloud structure driven by storm processes.

## 6. Data Dissemination and Operational Applications

The GLM processing method described here was adapted for use in the U.S. NWS, through ISatSS, a satellite preprocessing system that prepared data produced by the GOES Ground System for dissemination and use in the second generation NWS Advanced Weather Information Processing System (AWIPS-II), the primary display software used nationwide by local weather forecast offices. Like `glmtools`, ISatSS was implemented in the Python language, so it was possible to drop `glmtools` into the existing processing infrastructure with relatively minor changes to the front-end scripts that configured the GLM processing.

Initial NWS integration work in February to March 2018 was followed by a preoperational demonstration of the products. As those efforts matured, products were made available to forecasters in the National Severe Storms Laboratory Hazardous Weather Testbed Spring Experiment's Experimental Warning Program in May (Calhoun, 2018; Calhoun et al., 2014). Throughout the summer, the NWS Operations Proving Ground led further field tests with 20 forecast offices having subject matter experts and/or prior experience with lightning data sets such as LMA and LDAR.

During the preoperational trials, a variety of optimization and training activities took place. For example, during particularly active storms or periods of sun glint, GLM event rates of $O(10^4 s^{-1})$ exceeded what could be processed in real time, leading to the implementation of the aggregated event method described above. Other data artifacts included persistent nonlightning triggers by bright cloud tops combining with instrumental noise over the Bahamas (Rudlosky et al., 2019) and secondary scattering of light by low clouds. Examples were identified and incorporated into training materials.

As the 13 November 2018 example demonstrated, accumulations over longer periods of time resulted in greater spatiotemporal coherence to the imagery that was useful in inferring meteorological structure. Forecaster experimentation with 1-min versus 5-min products that updated each minute showed that the 5-min product/1-min update was preferred by forecasters for severe storm nowcasting. Initial feedback from NWS National Centers with large domains of responsibility was that additional, longer, accumulation times may be more appropriate for their needs. Other AWIPS optimizations, such as rounding of the FED product to the nearest integer value, allowed for dissemination of a single-byte product that was easier for forecasters to interpret and consumed much less memory in AWIPS than the 8-byte floating point grids used internally in `glmtools`.

The availability of highly tailored training materials was essential to the acceptance of products by forecaster end users. Foundational training included the basics of product definitions and an explanation of expected artifacts. Equally important were sound physical models paired with application narratives in the form of one- to two-page quick guides or short training vignettes showing the application of GLM to specific storm scenarios.

One of the most significant surprises of preoperational testing was that GLM showed very little lightning in some storms in which other lightning detection systems showed very high flash rates. Initial hypotheses are that the light from some flashes simply cannot exit the cloud when the optical depth of the cloud is too large or lightning flashes occur too far from cloud top (Thomas et al., 2000). It is also possible that some flashes produce large radio frequency pulses without making light or that aggressive filtering to reduce GLM false alarms is eliminating some single-light-pulse flashes. A full explanation awaits the results of detailed comparison studies underway in 2019, but it is important for users of the GLM data in any form to be aware of the phenomenon.

During summer 2018, 93 forecast offices from the central, eastern, and southern regions of the United States obtained access to the FED product over the CONUS domain (matching the ABI CONUS sector) as their first routinely available GLM data. All forecast offices, including those in the western United States, Alaska, and Hawaii, received access to products from the GOES-17 GLM during summer 2019 (matching the ABI sector covering the western CONUS and adjacent Pacific Ocean to Hawaii), completing the initial rollout of experimental GLM products to nearly all NWS local forecast offices and convective weather service units. As of publication, NWS National Centers were also receiving prototype 4-km equivalent, full-disk GLM imagery from GOES-16 and GOES-17 (e.g., the Aviation Weather Center and National Hurricane Center). Over 50 volunteer test offices continued to evaluate AFA and TOE as of the time of publication. Future application of these and the full suite of seven or more GLM products described in this study will rely on the development and availability of operational logic and training methodologies.

Users without access to NWS data feeds could process the GLM L2 data on their own using the open-source software that implemented the algorithms described in this study. To support users without such processing infrastructure, a real-time feed of GLM imagery has been made available to academic users through the Unidata Internet Data Distribution infrastructure alongside the routinely distributed ABI feed (Fulker et al., 1997; Goebbert et al., 2019). For example, one access route for the GOES-East CONUS GLM imagery products with an easily browsable catalog is http://thredds.unidata.ucar.edu/thredds/catalog/satellite/goes/east/products/GeostationaryLightningMapper/CONUS/current/catalog.html.

## 7. Concluding Remarks

This study described an algorithm suitable for gridding the GLM point detections reported by the GLM ground system while restoring their spatial footprint. The imagery produced by this process served as an alternate version of the GLM Level 2 data that retained the available quantitative information about lightning counts, energy, and size while being more suitable for display and analysis in meteorological settings.

The analysis of a mesoscale convective system in Argentina demonstrated that the imagery generated by our method can be used not only for visualization but also for quantitative evaluation of hypotheses drawn from theory and to test the robustness of patterns thought to be observed in the data. Looking at the spatially extensive flash properties allows examination of the local flash characteristics and places those characteristics in a storm-relative context. The prevalence of multiple peaks in the joint distribution of multiple physical

**Table B1**
*Table of Variables*

| Symbol | Variable Name | L2 |
|---|---|---|
| $x$ | Fixed grid projection east coordinate | |
| $y$ | Fixed grid projection north coordinate | |
| $s$ | Event discretization scale | |
| $(x_a, y_a)$ | Fixed grid projection coordinate aggregate event centroid | |
| $(x_e, y_e)$ | Fixed grid projection coordinate event centroid | |
| $(x_p, y_p)$ | Fixed grid projection coordinate of sub-event polygon centroid | |
| $g_{xi}$ | Target fixed grid east index | |
| $g_{yi}$ | Target fixed grid north index | |
| $K_f$ | flash_id | x |
| $K_g$ | group_id | x |
| $K_e$ | event_id | x |
| $K_a$ | event_lut_id | |
| $J_f$ | flash_energy | x |
| $J_g$ | group_energy | x |
| $J_e$ | event_energy | x |
| $A_f$ | flash_area | x |
| $A_g$ | group_area | x |
| $A_e$ | event_area | |
| $A_p$ | Subevent polygon area | |
| $A_t$ | Target grid cell area | |
| $f_e$ | Fraction of event covered by subevent polygon | |
| $f_t$ | Fraction of target grid cell covered by subevent polygon | |
| $_g K_e$ | event_parent_group_id | x |
| $_f K_e$ | event_parent_flash_id | |
| $_f K_g$ | group_parent_flash_id | x |
| $_a K_e$ | event_parent_lut_id | |
| $_a K_p$ | Subevent polygon parent aggregate event ID | |
| $_g N_f$ | flash_child_group_count | |
| $_e N_f$ | flash_child_event_count | |
| $_e N_g$ | group_child_event_count | |
| $_a N_f$ | flash_aggregate_count | |
| $_a N_g$ | group_aggregate_count | |
| $_a A_f$ | flash_aggregate_area | |

parameters drawn from the spatially extensive flash imagery contrasts with the limited detail available in considering the point-centroid distributions as a function of flash area. This study supports earlier work that found a second class of propagating lightning in space-based optical lightning data sets.

FED has been adopted by the U.S. NWS, but the other products require further study before operational adoption. Such studies should focus on the physical basis for their application; much of this work has been completed for AFA. The seven GLM imagery products documented in this study are available (by processing with an open-source package) for research use in carrying out such studies and for basic research made possible by the first continuous, hemispheric-scale lightning mapping observations.

## Appendix A: `glmtools` **implementation details**

The GLM L2 data were produced by the GOES Ground System in the NetCDF binary format (Unidata, 2016). Data were stored as integers with scale and offset factors to be used in conversion to floating point

**Table B1**
*Continued*

| Symbol | Variable Name | L2 |
|---|---|---|
| $_aA_g$ | group_aggregate_area | |
| $_aJ_e$ | event_aggregate_energy | |
| $VVV$ | Accumulated variable on the target grid | |
| TOE | Total optical energy | |
| FED | Flash extent density | |
| GED | Group extent density | |
| TFA | Total flash area | |
| TGA | Total group area | |
| AFA | Average flash area | |
| AGA | Average group area | |
| FCD | Flash centroid density | |
| GCD | Group centroid density | |

physical values. Time variables were given in seconds or milliseconds since some date and time, as indicated in metadata. These metadata were part of the Climate and Forecast (CF) Metadata Standards (Eaton et al., 2011).

We read the data with the `xarray` Python library (Hoyer & Hamman, 2017), which adopted the CF data model for its internal data structure, and automatically processed all of the above metadata, giving an in-memory data structure comprised of floating point arrays in physical units or, for time, a `datetime` array format that provided a human-friendly interface for querying the data by time. Such automated application of machine-readable metadata had the practical benefit of removing the tedium experienced in some other programming languages of making the data set ready for scientific calculations while preserving access to the fast array processing functionality provided by Python's scientific software ecosystem (Oliphant, 2007; van der Walt et al., 2011).

Some GLM variables, including those for event latitude and longitude, included the quasistandard `_Unsigned='true'` attribute. It indicated that the signed integer variable to which it was attached should be cast to an unsigned type before application of the scale and offset. Because it was not part of the CF Standards, it was not automatically applied by `xarray`, leading to floating point arrays with incorrect values. We contributed a patch to the `xarray` library (see release notes for v. 0.10.0; Hoyer et al., 2017) that automated application of the `_Unsigned` attribute, removing the need for special-case processing of certain GLM variables.

The polygon clipping implementation described here used the general-purpose algorithm of Vatti (1992) as implemented in the Clipper library (Johnson, 2014) and wrapped into the Python language (Johnson et al., 2015). It made no assumptions about polygon shape. Speed improvements might be possible by taking advantage of the convex quadrilateral geometry, by implementing a faster clipping function, and avoiding the round trip of data in and out of the data structures in the underlying C++ library. However, those tradeoffs were deemed acceptable in comparison to the development time trade-off in writing and vetting a new clipping library from scratch.

Our implementation utilized the indexing features of `xarray` (Hoyer & Hamman, 2017) and its database-like `groupby` operation to build a generic hierarchical traversal that could operate on any data set having a foreign key structure. Specification of the names of the ID variables $K_d$ and the parent ID variables $_kK_d$ in hierarchical order was sufficient to allow for machine traversal of the data structure at arbitrary levels of depth. Replications were constructed by providing the NetCDF name of any parent variable $V$ and the target child dimension $d$. Summarization was requested similarly but with parent $d$ and child $V$. Because the traversal operator was aware of the hierarchy, ascent or descent of as many levels as necessary was possible. Pruning was requested by specifying the $K_d$ to keep, followed by traversal of both up and down the hierarchy to determine a self-consistent set. While specified in Python, the traversal process happened quickly because `xarray` vectorized the operation using low-level array operations and Python had a fast hash table (dictionary).

## Appendix B

## References

Aich, V., Holzworth, R., Goodman, S., Kuleshov, Y., Price, C., & Williams, E. (2018). Lightning: A new essential climate variable. *Eos*, *99*. https://eos.org/science-updates/lightning-a-new-essential-climate-variable

Albrecht, R. I., Goodman, S. J., Buechler, D. E., Blakeslee, R. J., & Christian, H. J. (2016). Where are the lightning hotspots on Earth? *Bulletin of the American Meteorological Society*, *97*, 2051–2068.

Beirle, S., Koshak, W., Blakeslee, R., & Wagner, T. (2014). Global patterns of lightning properties derived by OTD and LIS. *Natural Hazards and Earth System Sciences*, *14*(10), 2715–2726.

Bezooijen, R. W. H., Demroff, H., Burton, G., Chu, D., & Yang, S. (2016). Image navigation and registration for the Geostationary Lightning Mapper (GLM). *Proceedings of the SPIE*, *10004*, 100041N 27.

Biggerstaff, M. I., & Houze, R. A. (1991). Kinematic and precipitation structure of the 10–11 June 1985 squall line. *Monthly Weather Review*, *119*(12), 3034–3065.

Bitzer, P. M., & Burchfield, J. C. (2016). Bayesian techniques to analyze and merge lightning locating system data. *Geophysical Research Letters*, *43*, 12,605–12,613. https://doi.org/10.1002/2016GL071951

Boccippio, D. J., Heckman, S., & Goodman, S. J. (2000). A diagnostic analysis of the Kennedy Space Center LDAR network 1. Data characteristics. *Journal of Geophysical Research*, *106*(D5), 4769–4786. https://doi.org/10.1029/2000JD900687

Bruning, E. C. (2019). glmtools. https://github.com/deeplycloudy/glmtools/

Bruning, E. C., & MacGorman, D. R. (2013). Theory and observations of controls on lightning flash size spectra. *Journal of the Atmospheric Sciences*, *70*(12), 4012–4029.

Bruning, E. C., & Thomas, R. J. (2015). Lightning channel length and flash energy determined from moments of the flash area distribution. *Journal of Geophysical Research: Atmospheres*, *120*, 8925–8940. https://doi.org/10.1002/2015JD023766

Buechler, D. E., Driscoll, K. T., Goodman, S. J., & Christian, H. J. (2000). Lightning activity within a tornadic thunderstorm observed by the Optical Transient Detector (OTD). *Geophysical Research Letters*, *27*(15), 2253–2256.

Calhoun, K. M. (2015). Forecaster use of total lightning data for short-term forecasts and warnings in the Hazardous Weather Testbed. In *Seventh conference on the meteorological applications of lightning data, Paper 3.5, AMS Annual Meeting*, Phoenix, AZ, USA, invited.

Calhoun, K. M. (2018). Feedback and recommendations for the Geostationary Lightning Mapper (GLM) in severe and hazardous weather forecasting and warning operations: NOAA/NSSL Hazardous Weather Testbed and GOES-R Proving Ground. https://hwt.nssl.noaa.gov/ewp/projects/GLM-HWT-report_2018.pdf

Calhoun, K. M., Bruning, E. C., Kingfield, D. M., Rudlosky, S. D., Siewert, C. W., Smith, T., et al. (2013). Forecaster use and evaluation of pGLM data at the NOAA Hazardous Weather Testbed and GOES-R Proving Ground. In *Preprints, Sixth Conference on the Meteorological Applications of Lightning Data, AMS Annual Meeting*, Austin, TX, USA.

Calhoun, K. M., Smith, T. M., Kingfield, D. M., Gao, J., & Stensrud, D. J. (2014). Forecaster use and evaluation of real-time 3DVAR analyses during severe thunderstorm and tornado warning operations in the hazardous weather testbed. *Weather and Forecasting*, *29*(3), 601–613. https://doi.org/10.1175/WAF-D-13-00107.1

Calhoun, K. M., Stano, G. T., Bruning, E. C., Rudlosky, S. D., & Kingfield, D. M. (2014). The Hazardous Weather Testbed: Moving total lightning data from a research tool into forecast and warning operations. In *Intl. Conf. Atmos. Elec., Norman, OK, paper O–04–02*.

Carey, L. D., Murphy, M. J., McCormick, T. L., & Demetriades, N. W. (2005). Lightning location relative to storm structure in a leading-line trailing stratiform mesoscale convective system. *Journal of Geophysical Research*, *110*, D03105. https://doi.org/10.1029/2003JD004371

Cecil, D. J., Buechler, D. E., & Blakeslee, R. J. (2014). Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmospheric Research*, *135-136*, 404–414. https://doi.org/10.1016/j.atmosres.2012.06.028

Cecil, D. J., Goodman, S. J., Boccippio, D. J., Zipser, E. J., & Nesbitt, S. W. (2005). Three years of TRMM precipitation features. Part I: Radar, radiometric, and lightning characteristics. *Monthly Weather Review*, *133*, 543–566.

Christian, H. J., Blakeslee, R. J., Boccippio, D. J., Boeck, W. L., Buechler, D. E., Driscoll, K. T., et al. (2003). Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *Journal of Geophysical Research*, *108*, 4005.

Christian, H. J., Blakeslee, R. J., Goodman, S. J., & Mach, D. M. (2000). Algorithm Theoretical Basis Document (ATBD) for the Lightning Imaging Sensor (LIS). NASA/Marshall Space Flight Cent., Alabama, http://eospso.gsfc.nasa.gov/atbd/listables.html

Cintineo, J. L., Pavolonis, M. J., Sieglaff, J. M., Lindsey, D. T., Cronce, L., Gerth, J., et al. (2018). The NOAA/CIMSS ProbSevere model: Incorporation of total lightning and validation. *Weather and Forecasting*, *33*(1), 331–345. https://doi.org/10.1175/WAF-D-17-0099.1

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), 377–387. https://doi.org/10.1145/362384.362685

Cummins, K. L., & Murphy, M. J. (2009). An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN. *IEEE Trans. Electromagnetic Compatibility*, *51*(3), 499–518.

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., et al. (2011). NetCDF Climate and Forecast (CF) Metadata Conventions. CF Conventions Governance Board, http://cf-pcmdi.llnl.gov/documents

Fuchs, B. R., Bruning, E. C., Rutledge, S. A., Carey, L. D., Krehbiel, P. R., & Rison, W. (2016). Climatological analyses of LMA data with an open-source lightning flash-clustering algorithm. *Journal of Geophysical Research: Atmospheres*, *121*, 8625–8648. https://doi.org/10.1002/2015JD024663

Fulker, D., Bates, S., & Jacobs, C. (1997). Unidata: A virtual community sharing resources via technological infrastructure. *Bulletin of the American Meteorological Society*, *78*(3), 457–468. https://doi.org/10.1175/1520-0477(1997)078<0457:UAVCSR>2.0.CO;2

GOES-R, Series Program Office (2017). Product definition and user's guide. Volume 3: Level 1b products. DCN 7035538, Revision F.1, available at https://www.goes-r.gov/users/docs/PUG-L1b-vol3.pdf

GOES-R, Series Program Office (2017). Product definition and user's guide. Volume 5: Level 1b products. DCN 7035538, Revision 1.1, available at https://www.goes-r.gov/users/docs/PUG-L2+-vol5.pdf

GOES-R Algorithm Working Group and GOES-R Series Program (2018). NOAA GOES-R Series Geostationary Lightning Mapper (GLM) Level 2 lightning detection: Events, groups, and flashes. NOAA National Centers for Environmental Information.

GOES-R Series Program Office (2019). GOES-R series data book. CDRL PM-14 Rev. A., May 2019, Available at https://www.goes-r.gov/downloads/resources/documents/GOES-RSeriesDataBook.pdf

Goebbert, K. H., Allen, J. T., Gensini, V. A., & Ramamurthy, M. (2019). Data-driven scientific workflows: A summary of new technologies and datasets explored at the Unidata 2018 workshop. *Bulletin of the American Meteorological Society*, *100*(2), ES97–ES99. https://doi.org/10.1175/BAMS-D-18-0265.1

Goodman, S. J., Blakeslee, R. J., Koshak, W. J., Mach, D., Bailey, J., Buechler, D., et al. (2013). The GOES-R Geostationary Lightning Mapper (GLM). *Atmospheric Research*, *125-126*, 34–49.

Goodman, S. J., Gurka, J., DeMaria, M., Schmit, T. J., Mostek, A., Jedlovec, G., et al. (2012). The GOES-R Proving Ground: Accelerating user readiness for the next-generation geostationary environmental satellite system. *Bulletin of the American Meteorological Society*, *93*(7), 1029–1040. https://doi.org/10.1175/BAMS-D-11-00175.1

Goodman, S. J., Mach, D. M., Koshak, W. J., & Blakeslee, R. J. (2010). Algorithm Theoretical Basis Document (ATBD) for the GLM lightning cluster-filter algorithm. NOAA/NESDIS Center for Satellite Applications and Research v2.0.

Holzworth, R. H., McCarthy, M. P., Brundell, J. B., Jacobson, A. R., & Rodger, C. J. (2019). Global distribution of superbolts. *Journal of Geophysical Research: Atmospheres*, *124*, 9996–10005. https://doi.org/10.1029/2019JD030975

Hoyer, S., Fitzgerald, C., & Hamman, J. (2017). xarray: v0.10.0.

Hoyer, S., & Hamman, J. (2017). xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1), 10. http://doi.org/10.5334/jors.148

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, *9*(3), 90–95.

Jenniskens, P., Albers, J., Tillier, C. E., Edgington, S. F., Longenbaugh, R. S., Goodman, S. J., et al. (2018). Detection of meteoroid impacts by the Geostationary Lightning Mapper on the GOES-16 satellite. *Meteoritics & Planetary Science*, *53*(12), 2445–2469. https://doi.org/10.1111/maps.13137

Johnson, A. (2014). Clipper software library. http://www.angusj.com/delphi/clipper.php

Johnson, A., Ratajc, G., Treyer, L., & Chalton, M. (2015). Pyclipper wrapper. https://pypi.org/project/pyclipper/

Lang, T. J., Pédeboy, S., Rison, W., Cerveny, R. S., Montanyà, J., Chauzy, S., et al. (2017). WMO world record lightning extremes: Longest reported flash distance and longest reported flash duration. *Bulletin of the American Meteorological Society*, *98*(6), 1153–1168. https://doi.org/10.1175/BAMS-D-16-0061.1

Lojou, J.-Y., & Cummins, K. L. (2004). On the representation of two- and three-dimensional total lightning information. In *Preprints, Conference on Meteorological Applications of Lightning Data, Paper 2.4, AMS Annual Meeting*, San Diego, CA, USA.

Lu, G., Cummer, S. A., Li, J., Zigoneanu, L., Lyons, W. A., Stanley, M. A., et al. (2013). Coordinated observations of sprites and in-cloud lightning flash structure. *Journal of Geophysical Research: Atmospheres*, *118*, 6607–6632. https://doi.org/10.1002/jgrd.50459

Lyons, W. A., C., B. E., Edgington, S. F., Tillier, C. E., MacGorman, D. R., Warner, T. A., et al. (2019). Documenting a 400+ km long mesoscale lightning flash in an MCS trailing stratiform region using GLM, NLDN, and OKLMA data. In *Preprints, Ninth Conference on the Meteorological Application of Lightning Data, Paper TJ20.1*, Phoenix, AZ.

Mach, D. M., Christian, H. J., Blakeslee, R. J., Boccippio, D. J., Goodman, S. J., & Boeck, W. L. (2007). Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor. *Journal of Geophysical Research*, *112*, D09210. https://doi.org/10.1029/2006JD007787

Mazur, V., Shao, X.-M., & Krehbiel, P. R. (1998). "Spider" lightning in intracloud and positive cloud-to-ground flashes. *Journal of Geophysical Research*, *103*(D16), 19811–19822. https://doi.org/10.1029/98JD02003

McCaul, E. W., Goodman, S. J., LaCasse, K. M., & Cecil, D. J. (2009). Forecasting lightning threat using cloud-resolving model simulations. *Weather and Forecasting*, *24*(3), 709–729.

Mulholland, J. P., Nesbitt, S. W., Trapp, R. J., Rasmussen, K. L., & Salio, P. V. (2018). Convective storm life cycle and environments near the Sierras de Córdoba, Argentina. *Monthly Weather Review*, *146*(8), 2541–2557. https://doi.org/10.1175/MWR-D-18-0081.1

Murphy, M. J. (2006). When flash algorithms go bad. In *1st international lightning meteorology conference*, Tuscon, AZ, USA, Vaisala, pp. 26–27. http://www.vaisala.com/en/events/ildcilmc/Documents/When%20Flash%20Algorithms%20Go%20Bad.pdf

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, *9*(3), 10–20. https://doi.org/10.1109/MCSE.2007.58

Parker, M. D., & Johnson, R. H. (2000). Organizational modes of midlatitude mesoscale convective systems. *Monthly Weather Review*, *128*(10), 3413–3436.

Peterson, M. (2019). Research applications for the Geostationary Lightning Mapper (GLM) operational lightning flash data product. *Journal of Geophysical Research: Atmospheres*, *124*, 10205–10231. https://doi.org/10.1029/2019JD031054

Peterson, M., Deierling, W., Liu, C., Mach, D., & Kalb, C. (2017). The properties of optical lightning flashes and the clouds they illuminate. *Journal of Geophysical Research: Atmospheres*, *122*, 423–442. https://doi.org/10.1002/2016JD025312

Peterson, M., & Liu, C. (2013). Characteristics of lightning flashes with exceptional illuminated areas, durations, and optical powers and surrounding storm properties in the tropics and inner subtropics. *Journal of Geophysical Research: Atmospheres*, *118*, 11,727–11,740. https://doi.org/10.1002/jgrd.50715

Peterson, M., Rudlosky, S., & Deierling, W. (2017). The evolution and structure of extreme optical lightning flashes. *Journal of Geophysical Research: Atmospheres*, *122*, 13,370–13,386. https://doi.org/10.1002/2017JD026855

Peterson, M., Rudlosky, S., & Deierling, W. (2018). Mapping the lateral development of lightning flashes from orbit. *Journal of Geophysical Research: Atmospheres*, *123*, 9674–9687. https://doi.org/10.1029/2018JD028583

Poehler, H., & Lennon, C. (1979). Lightning Detection and Ranging System (LDAR): System description and performance objectives. NASA, TM-74105.

Rison, W., Thomas, R. J., Krehbiel, P. R., Hamlin, T., & Harlin, J. (1999). A GPS-based three-dimensional lightning mapping system: Initial observations in central New Mexico. *Geophysical Research Letters*, *26*, 3573–3576.

Rudlosky, S., Goodman, S. J., & Virts, K. S. (2019). Lightning detection: Geostationary Lightning Mapper, *The GOES-R series: A new generation of geostationary environmental satellites* pp. ISBN–13: 978–0128143278, ISBN–10: 0128143274): Academic Press; Amsterdam, The Netherlands. http://www.worldcat.org/oclc/1127541502

Rudlosky, S. D., Goodman, S. J., Virts, K. S., & Bruning, E. C. (2019). Initial Geostationary Lightning Mapper observations. *Geophysical Research Letters*, *46*, 1097–1104. https://doi.org/10.1029/2018GL081052

Rumpf, C., Longenbaugh, R., Henze, C., Chavez, J., & Mathias, D. (2019). An algorithmic approach for detecting bolides with the Geostationary Lightning Mapper. *Sensors*, *19*(5), 1008. https://doi.org/10.3390/s19051008

Said, R. K., Inan, U. S., & Cummins, K. L. (2010). Long-range lightning geolocation using a VLF radio atmospheric waveform bank. *Journal of Geophysical Research*, *115*, D23108. https://doi.org/10.1029/2010JD013863

Said, R. K., & Murphy, M. J. (2016). GLD360 upgrade: Performance analysis and applications. In *Proc. 24th Int. Lightning Detection Conf. and Sixth Int. Lightning Meteorology Conf. 8 pp.*, San Diego, CA, Vaisala.

Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., & Lebair, W. J. (2017). A closer look at the ABI on the GOES-R series. *Bulletin of the American Meteorological Society*, *98*(4), 681–698. https://doi.org/10.1175/BAMS-D-15-00230.1

Schmit, T. J., Gunshor, M. M., Menzel, W. Paul, Gurka, J. J., Li, J., & Bachmeier, A. Scott (2005). Introducing the next-generation Advanced Baseline Imager on GOES-R. *Bulletin of the American Meteorological Society*, *86*(8), 1079–1096. https://doi.org/10.1175/BAMS-86-8-1079

Schultz, C. J., Carey, L. D., Schultz, E. V., & Blakeslee, R. J. (2015). Insight into the kinematic and microphysical processes that control lightning jumps. *Weather and Forecasting*, *1591–1621*, 30. https://doi.org/10.1175/WAF-D-14-00147.1

Shapiro, L. G. (1996). Connected component labeling and adjacency graph construction. *Machine Intelligence and Pattern Recognition*, *19*, 1–30. https://doi.org/10.1016/S0923-0459(96)80011-5

Steiger, S. M., Orville, R. E., & Carey, L. D. (2007). Total lightning signatures of thunderstorm intensity over north Texas. Part ii: Mesoscale convective systems. *Monthly Weather Review*, *135*(10), 3303–3324.

Stuhlmann, R., Rodriguez, A., Tjemkes, S., Grandell, J., Arriaga, A., Bézy, J.-L., et al. (2005). Plans for EUMETSAT's third generation Meteosat Geostationary Satellite Programme. *Advances in Space Research*, *36*(5), 975–981. https://doi.org/10.1016/j.asr.2005.03.091

Thomas, R. J., Krehbiel, P. R., Rison, W., Hamlin, T., Boccippio, D. J., Goodman, S. J., & Christian, H. J. (2000). Comparison of ground-based 3-dimensional lightning mapping observations with satellite-based LIS observations in Oklahoma. *Geophysical Research Letters*, *27*(12), 1703–1706.

Thomas, R. J., Krehbiel, P. R., Rison, W., Harlin, J., Hamlin, T., & Campbell, N. (2003). The LMA flash algorithm. In *Proc. 12th Int. Conf. on Atmospheric Electricity*, pp. 655–656.

Unidata (2016). Netcdf version 4.4.1 [software]. Boulder, CO: UCAR/Unidata Program Center, http://doi.org/10.5065/D6H70CW6

van der Walt, S., Colbert, SChris, & Varoquaux, G. (2011). The numpy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Vatti, B. R. (1992). A generic solution to polygon clipping. *Communications of the ACM*, *35*(7), 56–63. https://doi.org/10.1145/129902.129906

Wang, N.-Y., Gopalan, K., & Albrecht, R. I. (2012). Application of lightning to passive microwave convective and stratiform partitioning in passive microwave rainfall retrieval algorithm over land from trmm. *Journal of Geophysical Research: Atmospheres*, *117*, D23203. https://doi.org/10.1029/2012JD017812

World Meteorological Organization (2019). Lightning for climate: A study by the Task Team on Lightning Observation For Climate Applications (TT-LOCA) of the Atmospheric Observation Panel for Climate (AOPC). GCOS-227, https://library.wmo.int/index.php?lvl=notice_display&id=21405

Yang, J., Zhang, Z., Wei, C., Lu, F., & Guo, Q. (2017). Introducing the new generation of Chinese Geostationary Weather Satellites, Fengyun-4. *Bulletin of the American Meteorological Society*, *98*(8), 1637–1658. https://doi.org/10.1175/BAMS-D-16-0065.1

## Erratum

In the originally published version of this article, the figures published out of order. Figures 1, 2, and 9 published correctly, a duplicate of Figure 1 published as Figure 11, and Figure 8 was not published.

The following figures were also affected:

Figure 3 published as Figure 10

Figure 4 published as Figure 6

Figure 5 published as Figure 7

Figure 6 published as Figure 4

Figure 7 published as Figure 5

Figure 10 published as Figure 8

Figure 11 published as Figure 3

The figures have since been corrected, and this version may be considered the authoritative version of record.