AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD024663

Key Points:
- LMA flash-clustering algorithm able to process years of data and provide insight into algorithm performance and LMA detection
- LMA flash density differences with satellite values largest in Colorado and may be dependent on flash altitude and environment
- Variability in flash characteristics highlights performance of algorithm and guides future improvements

Correspondence to:
B. R. Fuchs,
brfuchs@atmos.colostate.edu

Citation:
Fuchs, B. R., E. C. Bruning, S. A. Rutledge, L. D. Carey, P. R. Krehbiel, and W. Rison (2016), Climatological analyses of LMA data with an open-source lightning flash-clustering algorithm, *J. Geophys. Res. Atmos.*, *121*, doi:10.1002/2015JD024663.

Received 15 DEC 2015
Accepted 12 JUL 2016
Accepted article online 15 JUL 2016

Climatological analyses of LMA data with an open-source lightning flash-clustering algorithm

Brody R. Fuchs[1], Eric C. Bruning[2], Steven A. Rutledge[1], Lawrence D. Carey[3], Paul R. Krehbiel[4], and William Rison[4]

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA, [2]Department of Geosciences, Texas Tech University, Lubbock, Texas, USA, [3]Department of Atmospheric Science, University of Alabama in Huntsville, Huntsville, Alabama, USA, [4]Langmuir Laboratory for Atmospheric Research, Geophysical Research Center, New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA

Abstract Approximately 63 million lightning flashes have been identified and analyzed from multiple years of Washington, D. C., northern Alabama, and northeast Colorado lightning mapping array (LMA) data using an open-source flash-clustering algorithm. LMA networks detect radiation produced by lightning breakdown processes, allowing for high-resolution mapping of lightning flashes. Similar to other existing clustering algorithms, the algorithm described herein groups lightning-produced radiation sources by space and time to estimate total flash counts and information about each detected flash. Various flash characteristics and their sensitivity to detection efficiency are investigated to elucidate biases in the algorithm, detail detection efficiencies of various LMAs, and guide future improvements. Furthermore, flash density values in each region are compared to corresponding satellite estimates. While total flash density values produced by the algorithm in Washington, D. C. (~20 flashes $km^{-2}\,yr^{-1}$), and Alabama (~35 flashes $km^{-2}\,yr^{-1}$) are within 50% of satellite estimates, LMA-based estimates are approximately a factor of 3 larger (50 flashes $km^{-2}\,yr^{-1}$) than satellite estimates in northeast Colorado. Accordingly, estimates of the ratio of in-cloud to cloud-to-ground flashes near the LMA network (~20) are approximately a factor of 3 larger than satellite estimates in Colorado. These large differences between estimates may be related to the distinct environment conducive to intense convection, low-altitude flashes, and unique charge structures in northeast Colorado.

1. Introduction

1.1. Lightning Processes

A lightning flash is a continuous plasma channel that develops bidirectionally away from a region containing a strong electric field [*Maggio et al.*, 2005] and propagates into potential energy wells [*Coleman et al.*, 2003]. Along this plasma channel, breakdown processes and currents produce electromagnetic radiation across a wide spectrum of frequencies. In focusing on the plasma channel and the processes along it, we are consistent with the flash definition that is used by ground- and space-based instruments [*Christian et al.*, 1999, 2003; *Cummins and Murphy*, 2009]. Regardless of the detection method, the flash is the highest-level grouping of detected processes to emanate from a connected set of local breakdown processes. The large electric fields present in thunderstorms are hypothesized to be produced by charge-separating collisions between precipitation ice (graupel and hail) and ice crystals in the presence of supercooled liquid water [e.g., *Takahashi*, 1978; *Saunders et al.*, 1991]. Accordingly, lightning is tightly coupled to the microphysical and dynamical processes within a storm [e.g., *Williams*, 1985; *Carey and Buffalo*, 2007; *Fuchs et al.*, 2015; *Schultz et al.*, 2015]. Indeed, considerable evidence has been shown that lightning flash rates are strongly tied to updraft speeds, for example [*Deierling and Petersen*, 2008], and ice water contents [*Petersen et al.*, 2005] in the mixed-phase region (0 to −40°C). The connection between lightning flash characteristics and storm physics can be used to study processes not observable by other methods. The lightning mapping array (LMA)-based flash rate algorithm described in this study provides characteristics of flashes, such as size and location, in addition to total flash counts to facilitate investigation of more storm processes.

1.2. Lightning Mapping Array Detection of Lightning

The advent of LMA networks has provided unprecedented detail of lightning channels on the substorm scale. LMAs use time-of-arrival (TOA) techniques from multiple stations to detect the time and location of

very high frequency (VHF) radiation bursts (or "sources") produced by the discontinuous propagation during breakdown that forms lightning channels [*Rison et al*., 1999]. Radiation is detected in an unused local television channel frequency (usually 60–66 MHz or ~5 m wavelength) to avoid contamination from local noise. There may be a few to a few thousand VHF sources reported by the LMA for a single lightning flash, depending on its spatial extent, its proximity to the network, and network detection efficiency [*Fuchs et al*., 2015]. Detection of VHF sources along lightning channels results in highly accurate three-dimensional mapping of the in-cloud portions of intracloud (IC) and cloud-to-ground (CG) flashes. See Figure 1 from *Thomas et al*. [2004] for an example of the detail that can be detected by an LMA. Subflash processes, such as the bidirectional nature of lightning breakdown, can also be observed with an LMA [*Behnke et al*., 2005; *Maggio et al*., 2005; *van der Velde and Montanyà*, 2013]. Such detail has facilitated investigations of storm processes, such as cloud microphysics and turbulence [*Bruning et al*., 2007; *Bruning and MacGorman*, 2013], in addition to macroscale electrical characteristics of storms. Examples of such investigations include the vertical charge structure and total lightning activity in thunderstorms occurring in geographically diverse regions in the United States [e.g., *Lang and Rutledge*, 2011; *Fuchs et al*., 2015].

1.3. LMA Source Clustering

Since LMA networks only detect VHF sources produced by flash propagation, processing of those sources is required to get higher-level information about the overall flash. These algorithms are known as flash-counting or flash-clustering algorithms. Multiple flash-clustering algorithms currently exist, and all perform the same basic function of clustering VHF sources by space and time into their parent flashes [e.g., *Thomas et al*., 2003; *MacGorman et al*., 2008; *McCaul et al*., 2009]. The grouping of sources is possible because the spatial and temporal separation between successive sources within a single flash is much smaller than the separation of sources in different flashes in all but the most extreme flash rate storms [see *Fuchs et al*., 2015, Figure 3]. This paper utilizes the open-source flash-clustering algorithm first discussed in *Fuchs et al*. [2015]. The algorithm not only produces total flash rates but also calculates characteristics about each flash such as location (initiation and centroid) and the plan position area of each flash. The plan position area of a flash can be thought of as the area enclosed by a rubber band wrapped around the flash viewed from above [*Bruning and MacGorman*, 2013].

This algorithm currently performs the similar basic functionality as other flash-clustering algorithms. In fact, flash rate differences between the XLMA algorithm [*Thomas et al*., 2003] and this algorithm have been shown to be consistently within 10–15% of each other [*Fuchs et al*., 2015]. However, this algorithm is able to efficiently process large amounts of LMA data to produce climatological-scale results, such as those shown in this paper, and is conducive to real-time use. The results in the paper elucidate the strengths and weaknesses of this algorithm, provide some insight about the ways the algorithm may be improved, and raise some points about lightning physics. Since the algorithm is open source, anyone may contribute to the improvement of the algorithm.

1.4. LMA-Detected Lightning in the Context of Other Data Sets

Lightning flashes produce radiation across a large range of frequencies, each with distinct propagation characteristics [*Shumpert et al*., 1982; *Budden*, 1988; *Cummins and Murphy*, 2009; *Nag et al*., 2015]. Since various lightning detection systems detect radiation from different portions of the spectrum, they are subject to their own strengths and weaknesses. Lower frequencies (from the very low frequency to high-frequency bands), subject to relatively little attenuation, are utilized by long-range detection systems such as Global Lightning Detection 360 network (GLD360) [*Said et al*., 2013] and Earth Networks Total Lightning Network (ENTLN) [*Rudlosky*, 2015]. These systems are able to continuously monitor lightning activity across the globe but are only able to detect strong flashes (typically CGs), since radiation from those processes is at a lower frequency and typically travels long distances before reaching detectors. The National Lightning Detection Network (NLDN) [*Cummins and Murphy*, 2009] is able to continuously detect weaker lightning flashes at the cost of spatial coverage and can discriminate between IC and CG flashes, although the uncertainty in the classification depends on several factors [*Nag et al*., 2013]. Optical lightning detectors aboard satellites, such as the Lightning Imaging Sensor (LIS) [*Christian et al*., 1999] and the Optical Transient Detector (OTD) [*Christian et al*., 2003], are able to provide information about global lightning distributions via snapshots of storms but are unable to provide temporal evolution of individual storms and are best suited for climatological studies. It is important to note that this limitation will be remedied with the launch of the

Geostationary Lightning Mapper (GLM) aboard the Geostationary Earth Observing System R series satellite (GOES-R). GLM will continuously detect optical emission from lightning on a hemispheric scale from approximately 55°N to 55°S [*Goodman et al*., 2013].

1.5. Differences Between LMA and Satellite Flash Data

Lightning flash climatologies from the LMA provide some valuable information when compared to corresponding climatologies constructed from satellite-based LIS and OTD optical sensors. Some differences are inevitable because of the differences between detection methods and data collection characteristics of the two systems. These differences must be kept in mind when comparing these data sets and may actually provide useful information about the flashes themselves when assessing any differences between flash climatologies.

LMA networks detect lightning fundamentally different from satellites. Satellite sensors use a charge-coupled device (CCD) to detect near-optical (777.4 nm) wavelength emissions from channel heating associated with oxygen excitation and relaxation [*Christian et al*., 1999, 2003]. Conversely, the LMA detects VHF (~5 m wavelength) radiation produced by discontinuous lightning channel propagation [*Rison et al*., 1999]. Multiple sensors detect each source of radiation, and a TOA method is used to estimate the location and time of each radiation source. Optical radiation is subject to larger rates of attenuation when propagating through a thunderstorm, which may make it more difficult to detect flashes with certain characteristics such as those that do not produce much light.

In addition to the detection frequencies, the data collection characteristics are also quite different between the two systems. LMAs are stationary and always on, so they continuously map the channels that comprise each lightning flash, permitting temporal analysis of storm evolution and climatological analysis. Since LMAs continuously detect lightning, they sample a distribution of storm intensities and life cycle phases as each storm initiates, matures, and then eventually decays. Given a long enough record, LMAs should sample the entire spectrum of storm types and life cycle phases to produce a representative measure of lightning characteristics near the network. Conversely, satellites capture short snapshots of lightning in a storm and may capture multiple storms during an overpass of a particular region. All of the satellite overpasses from multiple years are aggregated to produce an estimate of total flash density in a grid box. Presumably all the storms sampled during an overpass will not have identical intensities and will be in different phases of their life cycle. Therefore, we argue that it is reasonable to assume that the long record of overpasses should sample the full spectrum of storm intensities, from weak to the very intense. The end result is similar to the continuous observations of storms by an LMA network.

Another factor that warrants consideration is that the operational period of the LMA networks is not concurrent with the operational period of the satellites. As a result, the climatologies constructed by both instruments were not sampling the same population of storms. This is of particular concern in the Washington, D. C., and Colorado regions, which were observed by the OTD (north of ~38° latitude), which flew from 1995 to 2000. It is possible that a small sample of storms could bias the results in these regions. The Alabama region is at lower latitude than the other regions and was observed by the LIS instrument (~ south of 38° latitude), which was operational for much longer than the OTD (1997–2014). All of these caveats must be taken into account when attempting to compare climatologies constructed from LMA and satellite systems.

1.6. Overview of Paper

This study describes the development of the open-source flash-clustering algorithm and how the algorithm groups LMA sources into flashes. We will then demonstrate use of the algorithm in the form of lightning flash climatologies for the Washington, D.C., northern Alabama, and northeast Colorado LMA networks. These networks are located in environmentally distinct locations [*Williams et al*., 2005; *Fuchs et al*., 2015], which provide a wide range of storm intensities. In addition to flash densities, we will examine flash characteristics such as flash area, initiation height, duration and average power, and the sensitivity of these quantities to network detection characteristics. This will allow us to fully investigate the flash-clustering algorithm and the detection performance of the LMA networks included in this study. Flash density climatologies will be compared to climatologies from satellite-based optical detectors (LIS and OTD), and implications of the similarities and differences will be explored.

2. Algorithm Description

2.1. Clustering Method

This study uses the flash algorithm introduced (but only briefly described) by *Fuchs et al*. [2015]. This algorithm is similar to other published flash algorithms [*McCaul et al*., 2009; *MacGorman et al*., 2008; *Wiens et al*., 2005; *Thomas et al*., 2003] in that it uses spatial and temporal thresholds to cluster individual VHF sources into flashes. The major differences with this algorithm are the ability to process large amounts of data and that it is open source. Here we further describe how the algorithm groups sources into flashes and some of the implementation details.

LMA data are reported as tuples of time, latitude, longitude, and altitude coordinates. Each source has an associated chi-square metric, which describes the goodness of fit of the overdetermined solution for source location and time in addition to the number of stations that contributed to the location retrieval [*Thomas et al*., 2004]. Initial filtering is performed to discard solutions (LMA sources) detected by fewer than six stations (by default) or a chi-square value greater than 1.0. This is in an attempt to filter out noise (i.e., VHF sources not produced by lightning). Sensitivity studies (not shown) for a few select storms indicate that the flash counts and characteristics are not very sensitive to the choice of the chi-square threshold.

At its core, the clustering algorithm uses the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [*Ester et al*., 1996] algorithm, which is a general machine-learning clustering algorithm that assumes no a priori cluster shape. DBSCAN emulates the space and time thresholding behavior of other flash-sorting algorithms but is distinct because it searches for clusters of high-density LMA sources in a four-dimensional space-time matrix [*Pedregosa et al*., 2011], rather than searching for the mere presence of LMA sources that are in close proximity to each other [*Fuchs et al*., 2015]. In order to use the DBSCAN algorithm implemented in the Python scikit-learn package [*Pedregosa et al*., 2011], it is necessary to form a unified space-time coordinate vector that is searched for high-density clusters.

DBSCAN begins by picking random sources to start searching for clusters (or flashes). Each flash begins with a core LMA source surrounded by at least a specified number of points, each at a distance less than or equal to a specified maximum distance threshold. In the limit of the lowest possible density, a flash is a set of points at a distance exactly equal to the specified distance from the core source (which is also the flash centroid) and distributed uniformly in all directions. A linear segment of points belonging to a flash would have an average spacing somewhat less than the specified distance. Connecting sets of points that meet the above criteria forms the flash. Therefore, the DBSCAN algorithm operates much like a simple space-time thresholding algorithm for randomly distributed points but places a more stringent requirement on peripheral points. Compared to a simple dot-to-dot algorithm, which can connect a long string of points together as long as one other point is under the threshold distance, DBSCAN requires an outer point to be near another point which itself must be within the threshold distance of at least the minimum specified number of other sources.

Since the separation in both distance and time between sources in a flash must be considered simultaneously, a transformation must be made to make the units compatible for use in the DBSCAN algorithm. The space-time coordinate data is normalized with the weighted Euclidean distance used by *Mach et al*. [2007]. After the spatial coordinates are converted to an Earth-centered Cartesian coordinate system, the location of the LMA coordinate center is subtracted. Dividing the spatial dimensions by a prescribed distance value and the time dimension by a specified time value normalizes the coordinates. The maximum distance value must be large enough to include the nominal distance between VHF sources produced in a flash while not being too large to group nearby flashes together. In *Fuchs et al*. [2015], the normalization distance value was chosen to be 3 km for the Colorado network and 6 km for the less sensitive Alabama and D. C. networks, in accordance with values from other algorithms [*MacGorman et al*., 2008; *McCaul et al*., 2009]. The time coordinate is normalized by 150 ms. Unlike *McCaul et al*. [2009] the algorithm does not use a distance-dependent normalization value that loosens the clustering criteria for sources far from the network, where source location errors are a significant concern [*Thomas et al*., 2004]. Doing so attempts to compensate for missed sources with low powers that may result in missed flashes. We decided not to employ this adaptive thresholding for this paper because we want to provide a base performance for the algorithm before introducing any uncertainties associated with adaptive thresholding. Furthermore, investigating the impacts of detection efficiency on flash characteristics also allows us to characterize the performance of the LMAs included in this study.

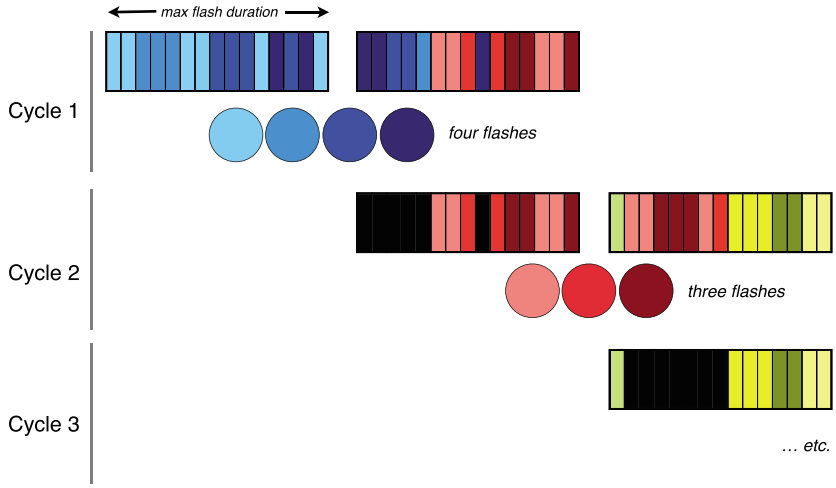

Figure 1. Graphical illustration of the streamed clustering approach that utilizes a rolling buffer of VHF sources. In each cycle, the two horizontal bars represent the first and second halves of the source buffer. The vertical subdivisions of each large horizontal bar represent VHF sources. Colors encode clusters of sources identified as flashes. At the end of the first cycle, all sources (in blue) in the first half of the buffer are assigned to flashes and cleared. Some sources (in blue) from the second half of the buffer are removed because they were also clustered as part of flashes during the first cycle. The remaining sources in the second half of the buffer (in red) are joined by a new half buffer of sources (in yellow and red), and the process repeats.

LMA data rates may be on the order of 10^5 sources min^{-1}. The DBSCAN implementation in scikit-learn uses a pairwise distance matrix to determine clustering, which is an $O(N^2)$ operation, making it prohibitive to process an entire thunderstorm at once. To work around this limitation, we take advantage of the fact that individual lightning flashes are time limited and impose (by default) an arbitrary 3 s maximum duration on the lightning flashes. Impacts of this threshold will be explored in sections 3 and 4. Processing is then accomplished in a streaming mode, where a VHF source buffer whose time span is twice the maximum possible flash duration ensures that flashes less than or equal to the maximum duration are not split (Figure 1). After DBSCAN identifies clusters within the buffer, those clusters with a source in the first half of the buffer are pruned out and further processed as flashes. Some sources will remain in the second half of the buffer, and these are retained for processing with a new half buffer of sources. This streaming process results in significant speed increase. As an example, the algorithm is fast enough to run in real time in the Colorado region, where LMA source rates from approximately 1000 flashes can be on the order of 1 million sources in a matter of 5 min.

2.2. Clustering Parameters

DBSCAN cluster formation is controlled by two parameters: a Euclidean distance threshold ε and the minimum number of points N_{min} required for a cluster to not be considered noise. Because the coordinate vector has been normalized to the specified distance and time values, we set ε equal to 1.0, since ε is essentially a multiplicative constant for the space and time thresholds. In previous studies, N_{min} has ranged from 2 to 10, depending on the sensitivity of the particular LMA network [*McCaul et al.*, 2009; *Fuchs*, 2014; *Fuchs et al.*, 2015]. For this study, N_{min} is set to 2 for Alabama and D. C. and set to 10 for Colorado. While singleton points and small clusters of less than N_{min} points (noise in DBSCAN's terminology) are not valid DBSCAN clusters, they are still removed from the first half of the processing buffer and retained in the final output data as small flashes. Later, when calculating flash rates, it is customary to filter these small flashes out by once again thresholding valid flashes on their number of points.

Similar to other existing algorithms, this algorithm is sensitive to the choices of clustering parameters. Larger space and time thresholds used in normalization may result in more sources being included in flashes or incorrectly grouping two nearby flashes together into one larger flash. Conversely, smaller space and time thresholds may incorrectly result in breaking a large flash into multiple smaller flashes. N_{min} may affect the number of flashes identified in addition to the characteristics of flashes, such as size and duration. Smaller

values of N_{min} will permit smaller clusters that may not be physical flashes, while a larger N_{min} will be more restrictive but may miss smaller flashes that have fewer sources, particularly for flashes that are far from the network. Differences in clustering parameters are mandated by differences in network sensitivities. The Alabama and D. C. networks are less sensitive than the Colorado network and therefore detect fewer low-power sources, resulting in fewer total detected sources and potentially missed flashes. To compensate for the lower detection efficiency of the Alabama and D. C. networks, a larger spatial threshold is employed. This attempts to compensate for sources produced by lightning channels that are missed and may result in a breakup of a flash because source distances are too far apart.

2.3. Differences With Existing Algorithms

The algorithm and instructions for use are currently available at https://github.com/deeplycloudy/lmatools. Since it is open source, anyone may download the package and contribute to it. The package is modular and permits other flash-clustering algorithms to be added and used. Additionally, it has been incorporated into an automated framework by *Fuchs et al.* [2015] and can process many LMA files at a time to produce large-scale results, such as those shown in this study. However, one of the goals of this paper is to show the strengths and weaknesses of this algorithm and outline potential ways that the algorithm may be improved.

By adopting an open-source transdisciplinary machine-learning package, we gain the benefit of rapid dissemination of bug fixes identified by a much larger community. This reduces the software maintenance burden on the lightning community, whose fundamental software task is reduced to lightning-specific preparation of the data for use in a standardized algorithm.

3. Results

3.1. Washington, D.C., Region

3.1.1. Spatial Flash Variability

The algorithm was applied to 8 years (2007–2014) of LMA VHF source data for the Washington, D. C., region. The network was composed of 8 stations before 2009 and 10 stations after 2009. The addition of stations likely increased the detection of efficiency of the LMA. However, quantifying this impact is difficult because the storms were not identical before and after the addition. Unlike the other LMA networks included in this study, the D. C. LMA is configured to detect radiation of slightly higher frequency (local Channel 10 versus Channel 4). This is important to note because lightning produces lower intensity radiation at higher frequencies, which are closer to the inherent noise floors of the stations, making source detection slightly more difficult by this network. Approximately 10 million flashes were gridded on a 0.15° latitude × 0.2° longitude grid (to make them approximately square). Conclusions reached in this study were not dependent on grid box size. We are assuming that the LMA detection efficiency is 100% over the network for all flashes, including CGs. This is because there is no currently accepted detection efficiency behavior for LMA networks. The implications of this assumption should be kept in mind when analyzing the results.

We wanted to investigate not only raw flash counts (and densities), which can be compared to other measurements such as satellite observations, but also characteristics of the flashes themselves. Figure 2 shows maps of VHF LMA source density, total lightning flash density, plan position flash area, and vertical distribution of flash initiation heights along with an azimuthally integrated relationship as a function of distance from the center of the network. The VHF source density in Figure 2a has a maximum near the center of the network with generally decreasing values farther from the network center. LMA source detection is line of sight, so sources that are close to the network are less likely to be blocked by the curvature of the Earth. Additionally, VHF sources are assumed to radiate isotropically [*Rison et al.*, 1999]. Consequently, low-power sources are less likely to be detected if they are far from the network because the power falls off according to the inverse-square law.

Figure 2b shows a map of flash density processed by the algorithm. A pattern similar to the LMA source densities consisting of a maximum near the center of the LMA with decreasing values with increasing distance from the network is evident. The average flash density values within 50 km of the network are approximately 18 flashes $km^{-2}\,yr^{-1}$ (Figures 2b and 2e). The decrease in flash density with increasing distance from the network is slower than the LMA source density, as shown in Figures 2b and 2e. The LMA source density falls off to half of its maximum value at approximately 50 km from the center of the network, while the flash density falls off to half its maximum value at approximately 125 km.

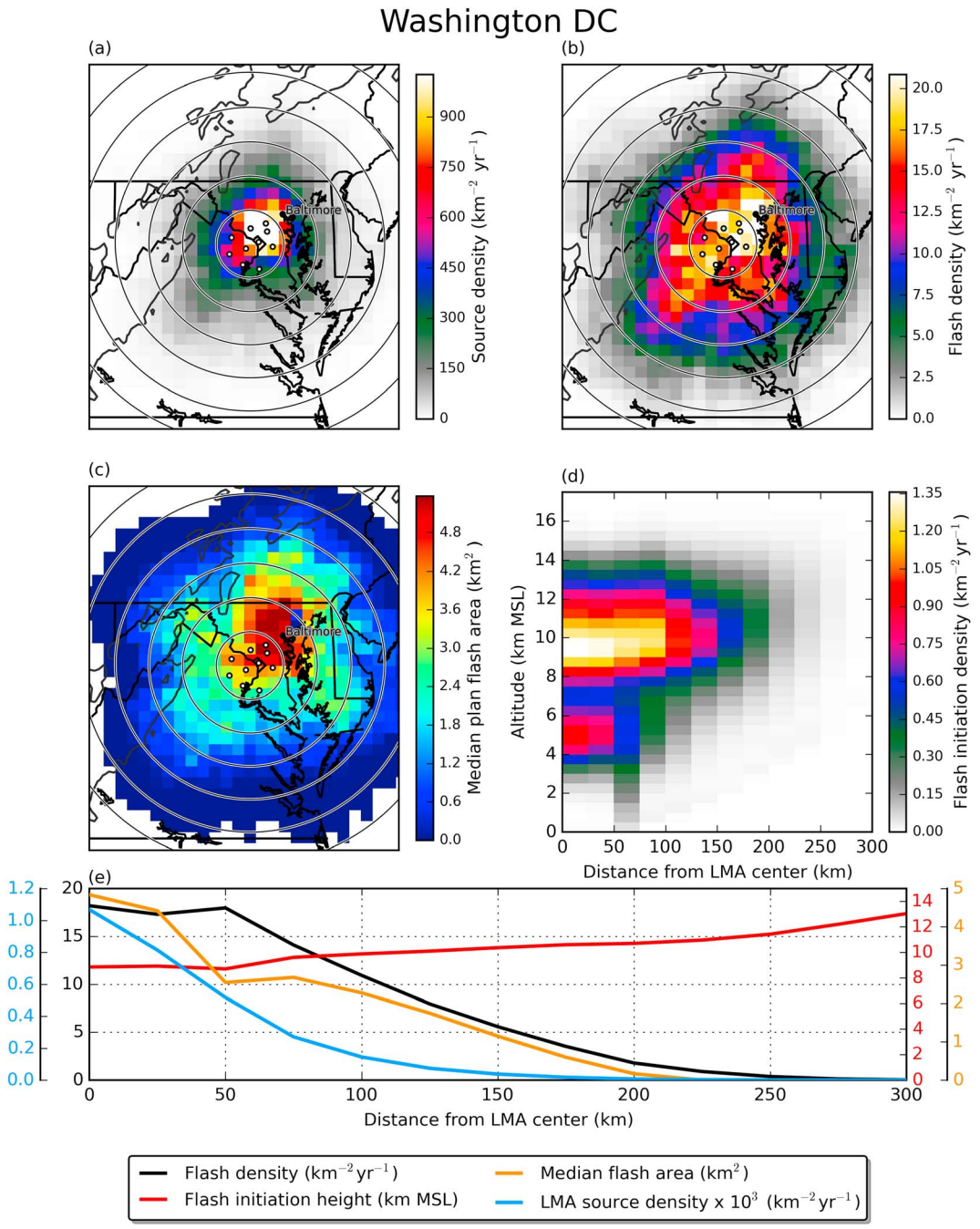

Figure 2. (a) LMA source density for the Washington, D. C., network on a 0.15° × 0.2° grid. (b) Total lightning flash density from the clustering algorithm on the same grid. (c) Median plan flash area for all flashes within a grid box. Locations with less than 1 flash km^{-2} yr^{-1} are excluded to minimize outlier effects. Dark gray contours in Figures 2a–2c are 300, 1000, and 2000 m MSL ground elevation. (d) Flash initiation height density as a function of distance from the center of the network. (e) Flash quantities azimuthally integrated to give behavior as a function of distance from the center of the network. White dots indicate the locations of LMA stations present during the analysis period. Range rings are increments of 50 km.

To investigate the characteristics of flashes themselves, Figure 2c shows a map of the median plan flash size for all flashes in a spatial bin. Similar to the VHF source density map, median flash size decreases with range from the maximum value. However, the location of maximum flash area is offset northeast of the network center toward the Baltimore area.

The vertical distribution of flash initiation height locations is shown in Figure 2d with respect to the distance from the network center. The flash initiation density maximum around 10 km above mean sea level (MSL) is consistent with IC flashes that initiate between midlevel negative charge and upper level positive charge in normal polarity storms common in this region [*Zajac and Rutledge*, 2001; *Fuchs et al*., 2015]. The average flash initiation heights of these flashes, shown in Figure 2e, increase from 9 km MSL over the network to 13.5 km MSL at 300 km from the center of the network. This follows from the expected R^2 source height location errors outside of an LMA network, where R is the distance from the network center to the source [*Thomas et al*., 2004]. The relative minimum around 7 km MSL is consistent with the main midlevel negative charge region, which is not a likely initiation point for lightning flashes [*MacGorman et al*., 2001], because it is a potential well, regardless of charge polarity [*Maggio et al*., 2005]. The relative maximum around 5 km MSL is consistent with the large electric field between midlevel negative charge and lower level positive charge in normal polarity storms (Figure 2d). The lack of flashes at low altitudes and longer range is due in part to Earth's curvature blocking line of sight for those particular sources.

3.1.2. Range Sensitivities of Flashes

Decreasing source detection efficiency impacts the algorithm-derived flash characteristics. If sources produced by a flash are missed, an incomplete picture of the flash results. Implications of this are explored in Figure 3 for the D. C. region. First, Figure 3a shows distributions of flash durations, partitioned by distance from the LMA center. Flash duration is defined here as the time difference between the first and last sources attributed to a flash. A consistent decrease in flash duration is apparent as distance increases from the LMA center. This is likely due to undetected sources that are occurring at the beginning or the end of the flash, thus shortening the duration of the flash. It is important to note that for flashes within 50 km of the network where values are most likely to be representative of the true flashes, the distribution of durations is nearly lognormal (note the log scale on the vertical axis) with a median around 200 ms. Additionally, less than 3% of flashes had more than a 1 s duration, suggesting that the 3 s cutoff in flash processes is more than sufficient.

Distributions of plan position flash areas calculated by the algorithm are shown in Figure 3b. Similar to the behavior of flash durations, the algorithm-calculated flash areas decrease with range from the network. Just as undetected sources near the beginning or end of the flash can shorten the duration, undetected sources near the spatial boundary of the flash will result in a calculated area that is smaller than the physical flash. It should be noted that the algorithm assigns an area of 0 if there are only two sources in a detected flash.

The distributions of the average source power in a flash shown in Figure 3c exhibit the opposite trend of previous panels. Average flash power increases monotonically with distance from the network. This is consistent with the inverse-square law, as weaker sources are more likely to go undetected if they are farther from the network, resulting in detection of only the strongest source powers in flashes far from the network. Conversely, the network is able to detect weaker sources inside of the network, which is why the median of the average source powers is approximately 0 dBW.

Figure 3d shows the distributions of the number of detected source points per flash. The median values decrease rapidly with increasing distance from the network, from approximately 20 sources per flash within the network to 2 at 250 km from the network. The lines in Figure 3d indicate the two thresholds at 2 and 10 points. It is clear that the flash count values are highly sensitive to the choice of N_{min}, particularly for flashes far from the network. There appears to be no best choice for this threshold. The choice of two minimum points may erroneously classify flashes simply because two VHF sources may be coincidentally adjacent; however, a strict threshold of 10 would result in missing nearly all of the flashes at 250 km from the center of the network. Because of this, flash rates and quantities should be met with a great deal of skepticism at these long ranges from the network center. It should be noted, however, that the strict chi-square source detection parameter should filter out much of the noise sources and mitigate identification of erroneous flashes.

Figure 3e shows a map of the number of lightning hours for the D. C. region. If one flash occurs in a grid box within a certain hour, it is counted as one lightning hour. This can be used as a measure of how common thunderstorms and lightnings are in the D. C. region and can be compared to other regions in the study. Over the center of the network, there are approximately 80 lightning hours, with decreasing values with distance from the network. The spatial variation in the lightning hour field is relatively smooth, much more so than the flash density map.

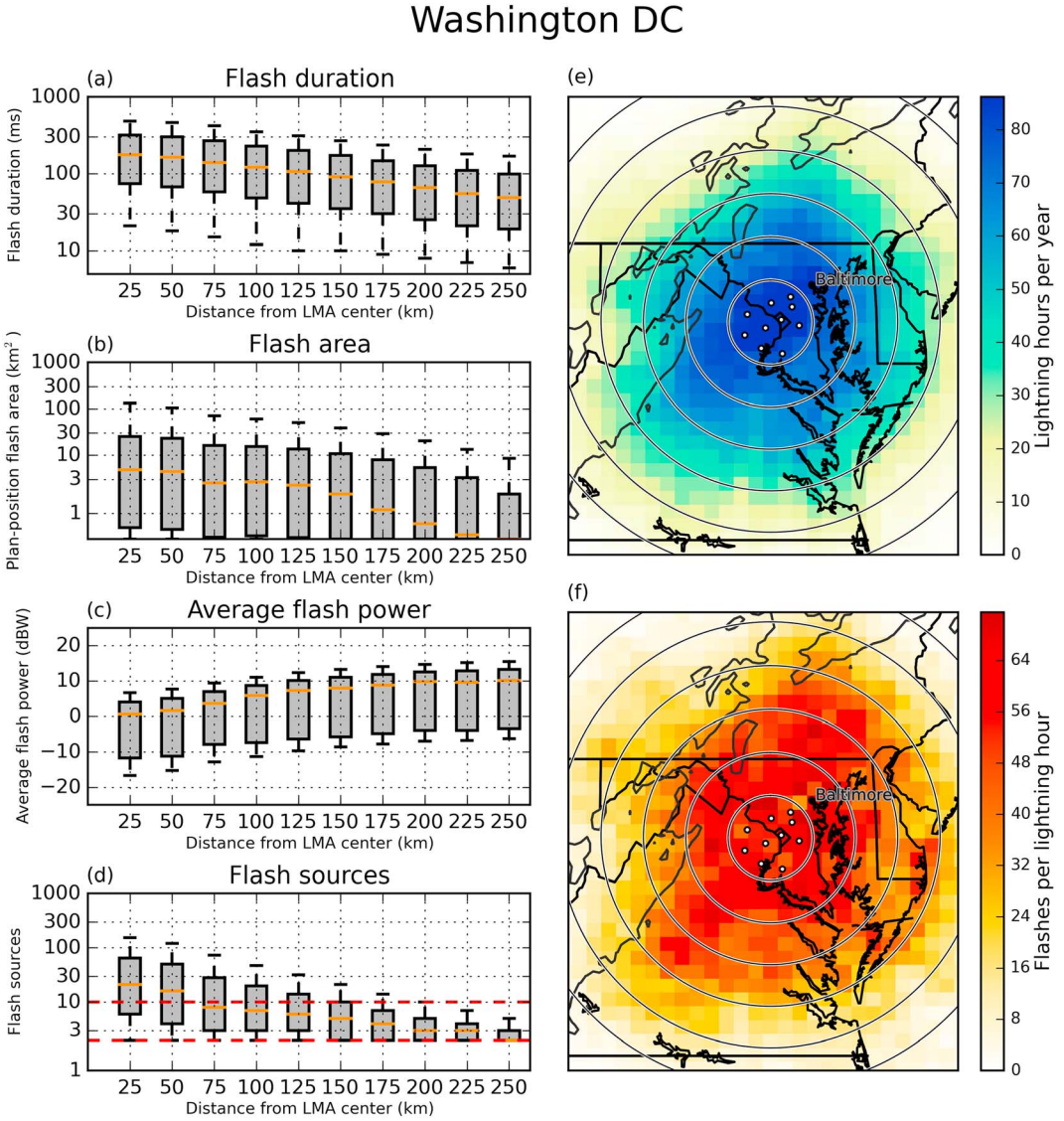

Figure 3. (a) Distributions of flash duration for flashes in each range bin from the D. C. network center. Bars indicate the median value. Top and bottom of the box indicate the 25th and 75th percentiles, respectively. Whiskers indicate the 5th and 95th percentiles, respectively. (b) Similar to Figure 3a for plan position flash area. (c) Similar to Figure 3a for the average source power within a flash. (d) Similar to Figure 3a for the total number of points associated with each flash. The red dashed lines indicate the typical thresholds of 2 and 10 points for reference. (e) Map of lightning hours for each grid box. Indicative of how many hours during a year at least one detected flash occurred in a particular grid box. (f) Map of the average number of flashes that occurred during an hour when at least one lightning flash occurred. This can be thought of as an average grid total flash rate.

The combination of flash density and lightning hour maps provides another useful piece of information. Figure 3f shows a map of the average number of lightning flashes that occur in a grid box during an hour that at least one flash occurs. This metric can be thought of as an average grid box total flash rate, as the units are flashes h^{-1}, but is not strictly a cell flash rate because multiple storms may occur in a grid box during a given hour. Regardless, this metric gives an indication of the intensity of the lightning-producing storms in this region. The values near the network are relatively uniform, ranging from approximately 50 to 70 flashes h^{-1}. The slower trend of decreasing values with range is not surprising when compared to flash densities, considering that only one flash in a grid box is needed to trigger a lightning hour.

3.1.3. NLDN CG Lightning

NLDN CG flash densities are subtracted from the total flash densities provided by the algorithm to estimate IC flash densities in each grid box. Bulk total and CG flash density quantities are subtracted in each grid box,

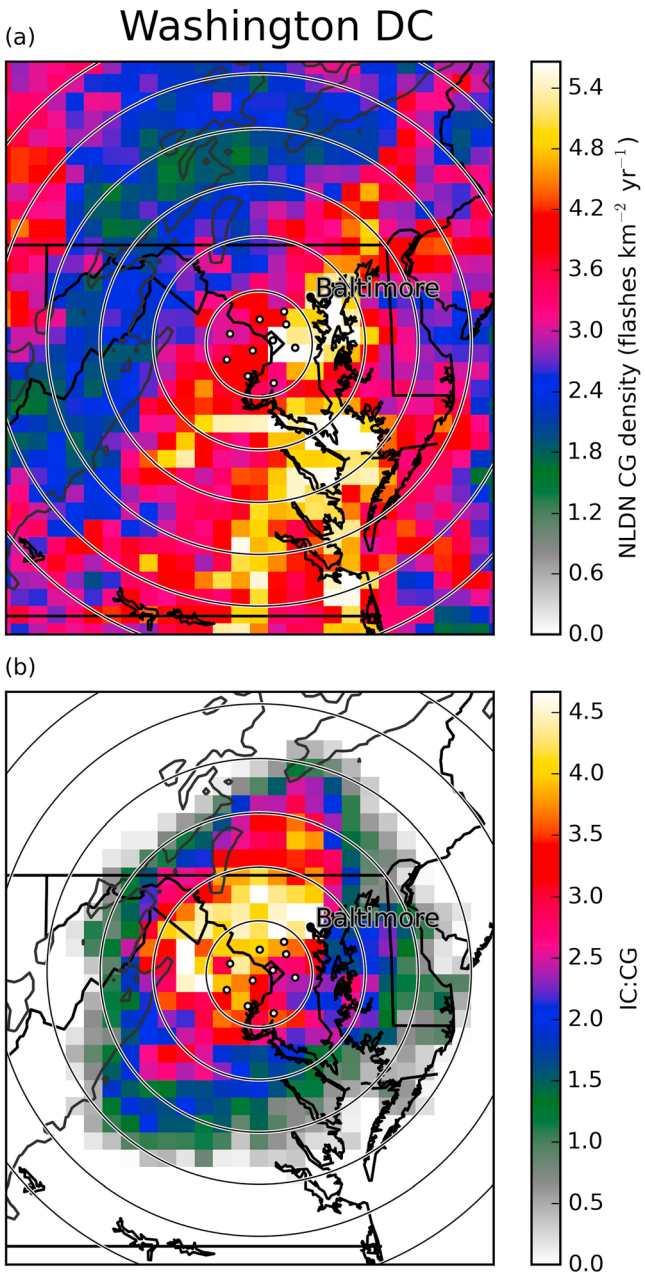

Figure 4. (a) Peak-current-filtered CG flash density map from the NLDN from the same time frame as the available LMA data in D. C. (b) Derived IC:CG values from the NLDN CG rate and the LMA-calculated total flash rate assuming 100% detection efficiency. White dots indicate the locations of LMA stations present during the analysis period. Range rings are increments of 50 km.

following the methodology of *Boccippio et al.* [2001]. Recall that we are making the assumption of 100% flash detection efficiency with the LMA. Only NLDN CG flashes, subject to peak-current filtering [*Cummins and Murphy*, 2009] that occurred during the time period of 2007–2014, are included in this study to ensure that the same storms are being included in this analysis. The detection efficiency of the NLDN network is above 95% throughout the D. C. domain. The detection characteristics of the NLDN may have been altered after its upgrade in 2013 [*Nag et al.*, 2013]; however, we assumed 100% detection efficiency for the NLDN for the entire observation period, since it will not change any of the conclusions in this paper.

First, Figure 4a shows the NLDN CG density in units of flashes km^{-2} yr^{-1}. The significant spatial variation in CG activity is evident. Specifically, the highest values lie along the Atlantic coast and near Baltimore. Values decrease both inland toward the higher terrain and offshore to the east. Figure 4b shows a map of the ratio of the estimated IC flash rate (based on LMA total flash rate and NLDN CG flash rate) to the NLDN CG flash rate. The maximum values are near 5–6 in the Baltimore area, with generally decreasing values farther from the network. This decrease is not surprising, since flashes are being missed by the LMA and clustering algorithm at long range but are still being detected by the NLDN, which has little variability in detection efficiency over the range of the LMA network [*Cummins and Murphy*, 2009]. The IC:CG values are most trustworthy near and inside the network (~50 km), where LMA detection efficiency is the highest. Accordingly, the differences between IC:CG values shown here are 50–100% higher than in *Boccippio et al.* [2001] and more recently *Medici et al.* [2015], which used satellite-based optical sensors for total flash rate estimates. Implications for these differences are explored in section 4.

3.2. Northern Alabama

3.2.1. Spatial Flash Variability

The northern Alabama LMA network covers portions of Alabama, Mississippi, Georgia, and Tennessee. The network consists of 12 stations in northern Alabama centered near Huntsville and 2 stations near Atlanta,

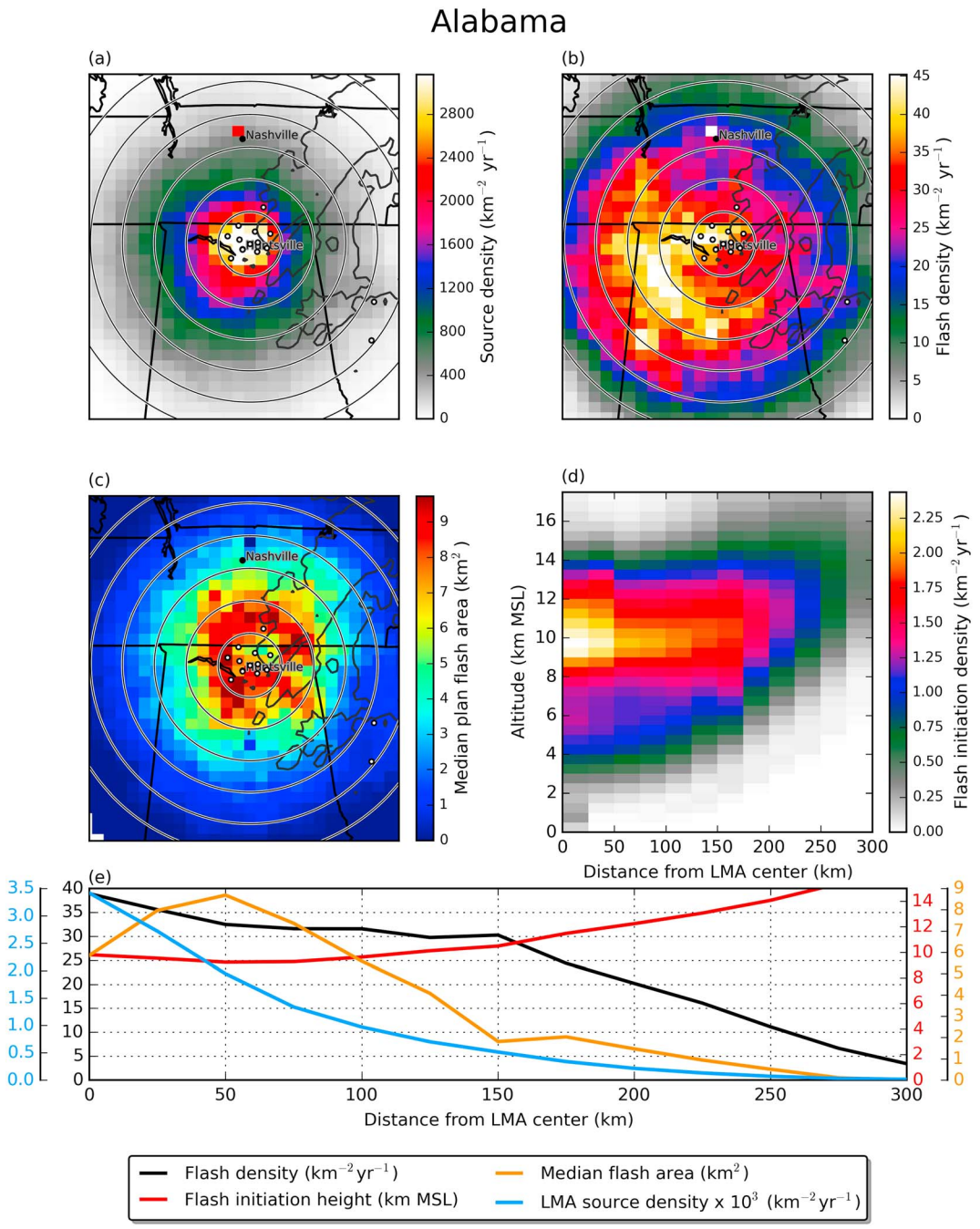

Figure 5. (a) LMA source density for the northern Alabama network on a 0.15° × 0.2° grid. (b) Total lightning flash density from the clustering algorithm on the same grid. (c) Median plan flash area for all flashes within a grid box. Locations with less than 1 flash km^{-2} yr^{-1} are excluded to minimize outlier effects. Dark gray contours in Figures 5a–5c are 300, 1000, and 2000 m MSL ground elevation. (d) Flash initiation height density as a function of distance from the center of the network. (e) Flash quantities azimuthally integrated to give behavior as a function of distance from the center of the network. White dots indicate the locations of LMA stations present during the analysis period. Range rings are increments of 50 km.

Georgia. Approximately 43 million flashes from 7 years of LMA data (2008–2014) are shown here. The LMA source density map in Figure 5a shows that the maximum is located near the center of the network with a rapid decrease outside of the network, similar to the D. C. LMA. However, the magnitudes of the maximum LMA source density values are approximately a factor of 3 larger than in the D. C. region. The preference for sources on the south and west side of the network is apparent. Detection artifacts due to asymmetries

in station locations are likely not the cause [*Thomas et al*., 2004; *Koshak et al*., 2004], since station locations are approximately uniform.

The maximum in source density near Nashville, Tennessee, is not physical, however. Further analysis (not shown) indicates that a VHF emitter in the LMA frequency window is responsible for the additional sources there. An attempt to remove these sources is difficult, if not impossible, without a large radar data set (to determine the presence of storms), and is beyond the scope of the present study. Additionally, the characteristics of the sources (such as altitude and power) were not substantially different from sources produced by lightning (not shown).

A map of total flash density is shown in Figure 5b. The average value within the LMA network is around 35 flashes $km^{-2} yr^{-1}$, much higher than the values in the D. C. region. The bias of lightning activity to the south and west is also evident in the total flash density map in Figure 5b, providing some evidence that the source density variability is indeed a physical signal rather than a detection artifact of the LMA. Lower flash densities over the elevated terrain in the eastern portion of the network suggest that the terrain may have an impact on storms in the area. Since the terrain only rises to approximately 300 m above the altitude of Huntsville, it is unlikely that any line-of-sight blockage preventing LMA detection of VHF radiation is occurring. The decrease in flash density with range in the Alabama network is much more gradual than the D. C. network, as the azimuthally integrated flash density falls to half of its maximum value around 225 km, much longer than the 125 km in the D. C. region. This is likely due in part to the greater number of stations in the Alabama network (12) compared to D. C. in addition to the greater sensitivity as a result of lower noise floors. Note that the same point with high source densities near Nashville results in high flash densities in that same area. This implies that the production of sources by the anomalous emitter is close enough in space (~1 km) and time (~0.1 s) to be erroneously clustered into a flash by the algorithm.

The median plan flash areas map in Figure 5c and the azimuthally integrated values in Figure 5e show an unexpected result. The median flash area peaks around 50 km from the center of the network, which is located just outside of the network radius. Median flash areas in that region are approximately 50% larger than over the center of the network. Unlike flash density, no significant flash area variability with azimuth is observed. Possible explanations for this will be explored in section 4.

The vertical distribution of flash initiation altitudes in the Alabama region shows a similar trend to the D. C. region. Most flashes initiate around 9–10 km MSL, likely between the upper level positive charge region and the midlevel negative charge region of a normal polarity storm. Indeed, *Fuchs et al*. [2015] found that the overwhelming majority of storms in this region had inferred positive charge at temperatures colder than −40°C, consistent with normal polarity storms containing positively charged ice crystals in the upper portions of the cloud [*Williams*, 1985]. The detected initiation altitude increases with distance from the network, as the mode height increases from 10 km over the network to 14 km MSL at 300 km from the network due to source height location errors [*Thomas et al*., 2004]. Detected flashes at lower altitudes (<5 km MSL) are relatively rare, particularly at ranges greater than ~200 km because those are below the horizon due to Earth's curvature.

3.2.2. Range Sensitivities of Flashes

Similar to the D. C. region, flashes are composited and analyzed by their distance from the center of the network to investigate the impacts of detection characteristics on the flashes themselves. The flash durations in Figure 6a show that the algorithm's representation of flashes tends to become shorter for flashes far from the LMA, similar to the D. C. region. Note that the distribution of flash durations inside of 50 km is quite similar to corresponding flashes in the D. C. region. Undetected sources near the beginning or the end of the physical flash will effectively shorten the detected flash. Similar reasoning is consistent with the general decrease in flash area with increasing distance from the network in Figure 6b. The plan position area estimate is sensitive to sources near the periphery of a flash. If those sources are not detected, the area estimate is erroneously low. This effect is particularly apparent when comparing the 95th percentiles of flash area for each distribution. It is important to note that the largest median area corresponds to flashes 50–75 km from the center of the network. Inspecting the 75th and 95th percentiles of flash area around 50–75 km show that fewer flashes have very large area, suggesting that some detail of the flashes may be lost at that range, assuming no substantial difference in physical flash properties. However, inspecting the 25th percentile of flash sizes at that range shows that there are fewer small flashes, suggesting that

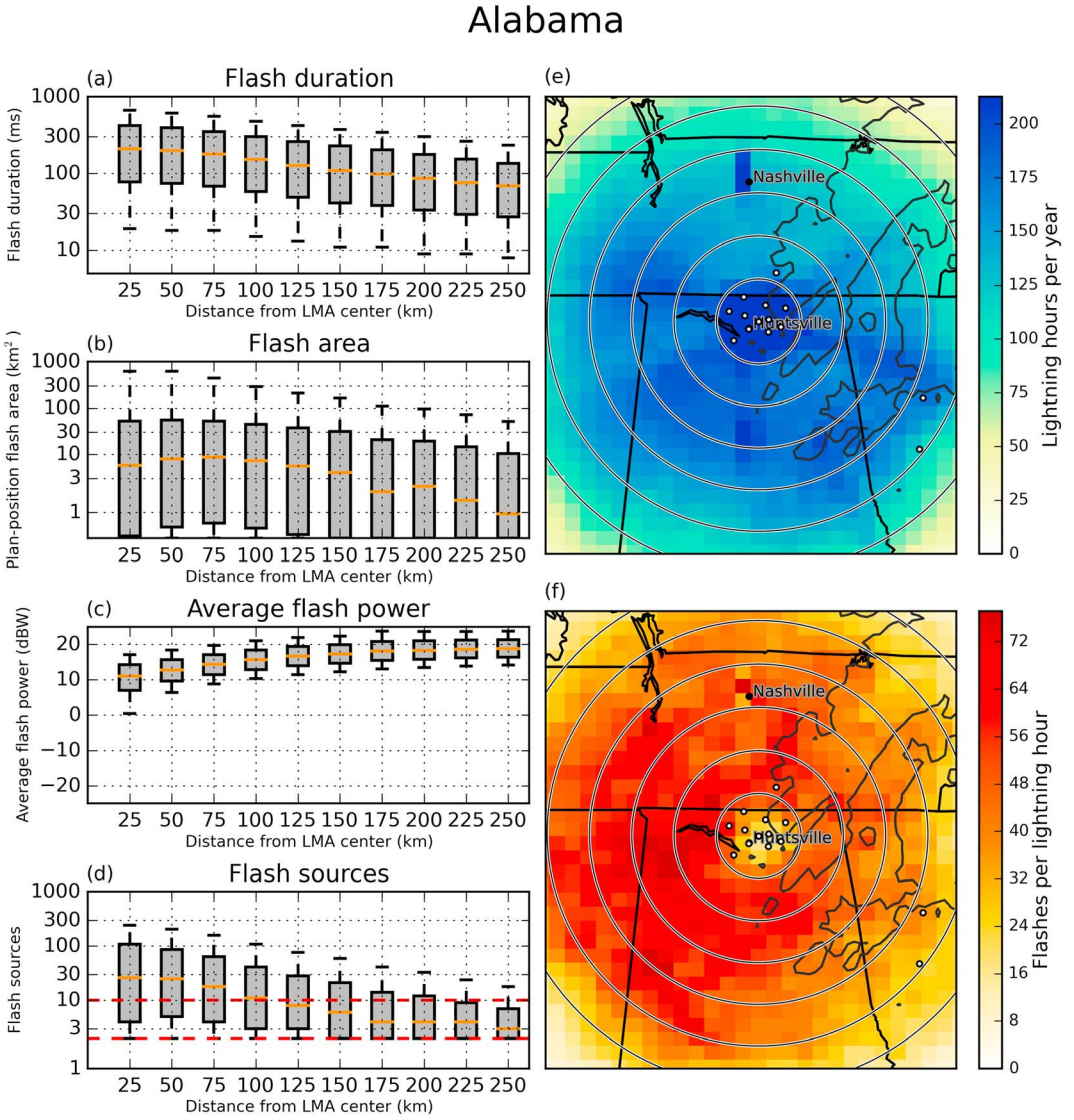

Figure 6. (a) Distributions of flash duration for flashes in each range bin from the northern Alabama network center. Bars indicate the median value. Top and bottom of the box indicate the 25th and 75th percentiles, respectively. Whiskers indicate the 5th and 95th percentiles, respectively. (b) Similar to Figure 6a for plan position flash area. (c) Similar to Figure 6a for the average source power within a flash. (d) Similar to Figure 6a for the total number of points associated with each flash. The red dashed lines indicate the typical thresholds of 2 and 10 points for reference. (e) Map of lightning hours for each grid box. Indicative of how many hours during a year at least one detected flash occurred in a particular grid box. (f) Map of the average number of flashes that occurred during an hour when at least one lightning flash occurred. This can be thought of as an average grid total flash rate.

the smallest flashes are undetected at that range, thereby increasing the median flash area. The decrease in the 25th percentile of flash sizes at longer ranges is likely owed to undetected sources in a flash, artificially decreasing the calculated flash size.

The distribution of average flash powers (Figure 6c) follows a similar trend to the D. C. region. Lower average flash powers inside and near the network result from the ability of the LMA to detect weaker sources at close range. The average flash powers near the network in Alabama are much higher (~10 dBW) than in the D. C. region (~0 dBW). Given the lower noise floors and greater number of stations, the lower source powers in D. C. seem counterintuitive, but the difference can likely be explained by the different frequencies used by the networks. Recall that lightning produces lower power radiation at higher frequencies. The number of points per flash in Figure 6d supports the notion of the more sensitive Alabama network. While the points

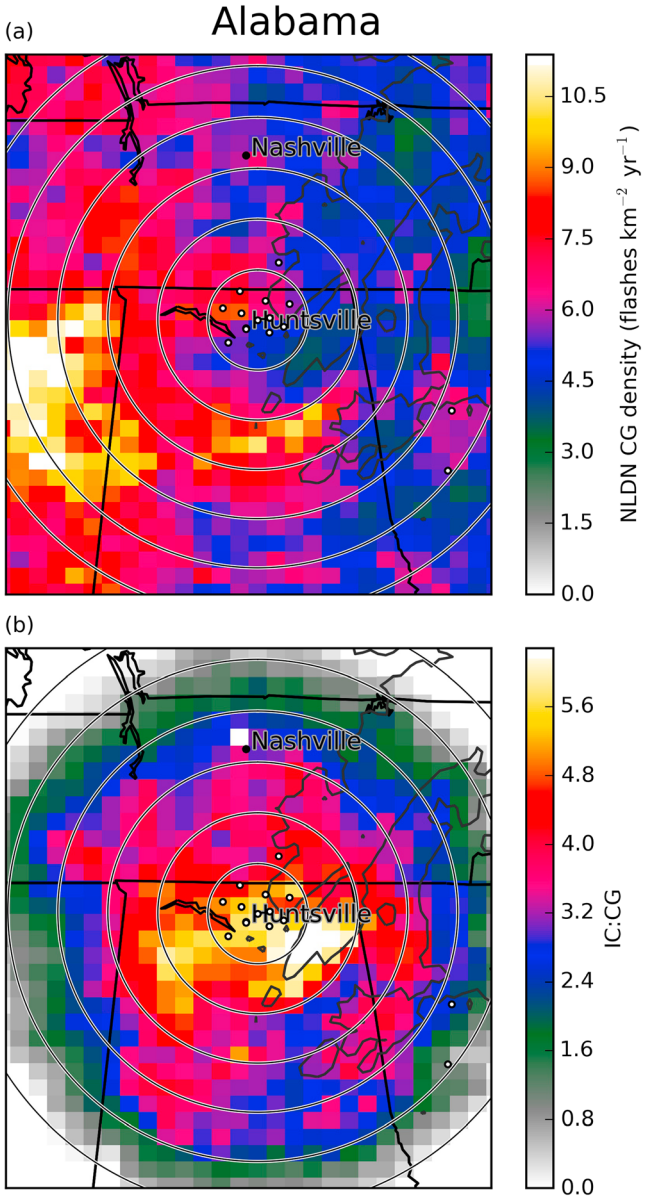

Figure 7. (a) Peak-current-filtered CG flash density map from the NLDN from the same time frame as the available LMA data in northern Alabama. (b) Derived IC:CG values from the NLDN CG rate and the LMA-calculated total flash rate assuming 100% detection efficiency. White dots indicate the locations of LMA stations present during the analysis period. Range rings are increments of 50 km.

per flash follows the same trend as the D. C. region, the median values are much higher than those in the D. C. region. For flashes between 225 and 250 km from the Alabama network, approximately 20% of the flashes are made up of at least 10 points, compared to approximately 1% of similar flashes in the D. C. region. Similar to the D. C. network, flash rates and quantities at this long range should be met with skepticism and should be analyzed to ensure their quality.

The number of lightning hours in each grid box is shown in Figure 6e and indicates a different thunderstorm environment than in the D. C. region. With values above 200 lightning hours per year inside 50 km, this region is an active region for thunderstorm development and persistence. Figure 6f shows that the average grid box flash rates during a lightning hour in Alabama are not substantially higher than in the D. C. region. This suggests that there are more numerous lightning-producing storms in Alabama than in D. C., rather than more intense thunderstorms producing more lightning per storm. These results agree with the cell-based flash rates from *Fuchs et al.* [2015], which showed similar cell flash rates between the Alabama and D. C. regions. The maximum in lightning hours (Figure 6e) and relative minimum in grid total average flash rate (Figure 6f) over the network is a surprising feature which suggests that spurious clusters of points are being considered a flash by the algorithm. However, this effect appears to be limited to the interior of the network, a small portion of the whole domain, where network sensitivity is the highest.

3.2.3. NLDN CG Lightning

Figure 7a shows the distribution of peak-current-filtered NLDN CG flash density for the Alabama region. The westward preference for lightning production is also evident here, as northeastern Mississippi and northwestern Alabama have substantially higher CG flash densities than the eastern side of the network where the high terrain is located. Figure 7b shows that the average value of IC:CG within the network is around 5, nearly twice as high as previous estimates from *Boccippio et al.* [2001]. The IC:CG should be most trustworthy inside of the network (~50 km), as the highest flash detection efficiency is located there. Note the highest IC:CG values to the southeast of Huntsville over a region of high terrain known as Sand Mountain. This location corresponds to a relative minimum in NLDN CG density, while no comparable drop is observed in total flash density.

3.3. Northeast Colorado

3.3.1. Spatial Climatologies

The northeast Colorado LMA network was installed during the winter of 2011–2012 for the DC3 field campaign [*Barth et al*., 2015]. The network consists of 15 stations ranging from locations just northeast of Denver up to the Wyoming and Nebraska borders. Currently, it is the newest of the LMA networks in the United States and has the lowest noise floor of any network, due to the updated electronics and remote locations of most stations. The diameter of the network is approximately 90 km, also the largest of any network to date. This is expected to further increase the detection efficiency [*Rison et al*., 1999; *Thomas et al*., 2004]. However, data during most of 2013 were not used since the network only had five to six active stations operating then due to various technical difficulties. For this reason, results in this section include approximately 10 million flashes from 2012 and 2014.

The LMA source density map in Figure 8a shows much larger magnitudes and spatial variability than either of the other networks. Complex topography in the region and a small sample size of 2 years are likely significant contributors to the variability of the LMA source density map. Unlike the other networks, the maximum VHF source density is not located within the network but rather northeast of the network near Fort Morgan, about 110 km northeast of Denver. A sharp gradient in LMA source density is present in the foothills of the Rocky Mountains as well. Rapid decrease in LMA source density is observed toward the east around 150 km from the center of the network, farther than the other networks.

The flash density map in Figure 8b exhibits some different behaviors than the source density map. The magnitudes of flash density are striking, with values as high as 80 flashes $km^{-2}\,yr^{-1}$ in some areas. The average flash density inside the network (~50–55 flashes $km^{-2}\,yr^{-1}$) is approximately 3 times larger than in the D. C. network and almost twice as large as the Alabama network.

Concerning the spatial distribution of flash densities, the values are much more homogenous east of the Rocky Mountains, while lightning activity falls off rapidly toward the west in the Rocky Mountains. This is also shown in Figure 8e, as the flashes in the eastern portion of the domain are indicated separately. The large gradient is likely due to weaker storm intensities in the foothills where thermodynamic environments are not conducive to intense convection, relative to the adjacent foothills where instability and cloud base heights are higher [*Williams et al*., 2005; *Jirak and Cotton*, 2006; *Fuchs et al*., 2015]. A relative maximum in lightning density is present east of Denver on the north side of the Palmer Divide. This region is hypothesized to be a favorable location for storm initiation due to the Denver Convergence and Vorticity Zone (DCVZ) [*Crook et al*., 1990]. More years of data should help verify the spatial patterns in flash density, as spatial noise from a low number of storm days will be averaged out. Note that with the strict clustering thresholds employed in this region, we suspect that these flash density quantities may actually be a lower bound on the estimate of flash densities, particularly far from the network.

A drastically different signal in median plan flash area is observed in the Colorado region compared to the other regions, as the larger flash areas over the Rocky Mountains dominate the signal. The sharp longitudinal gradient in flash area is coincident with both the flash density gradient and the elevation gradient. Put another way, there are fewer but larger flashes over the Rocky Mountains compared to the adjacent plains of Eastern Colorado. This signal is consistent with the claims made by *Bruning and MacGorman* [2013] that weaker storms with less turbulence produce fewer but spatially larger flashes. They hypothesize that the flash size spectrum is strongly related to the amount of turbulence in a storm, because larger charge reservoirs build up in weaker storms that result in large flashes when breakdown is initiated. This is in contrast to strong storms with intense vertical motions that produce numerous pockets of charge that result in frequent but smaller flashes [*Bruning and MacGorman*, 2013; *Basarab et al*., 2015].

Perhaps a less obvious, but nonetheless surprising signal, is apparent in the data if only flashes that occurred in the eastern half of the domain are considered (dotted lines in Figure 8e). The median flash size increases from the center of the network to 125 km, after which the median flash decreases monotonically, a trend similar to the Alabama network. The flashes in the eastern half of the domain are treated separately in this case to remove the terrain and meteorological effects of the Rocky Mountains to the west.

The vertical distribution of flash initiation points is also substantially different than the other regions. The maximum in flash initiation density is located around 7 km MSL, coincident with the location of the relative

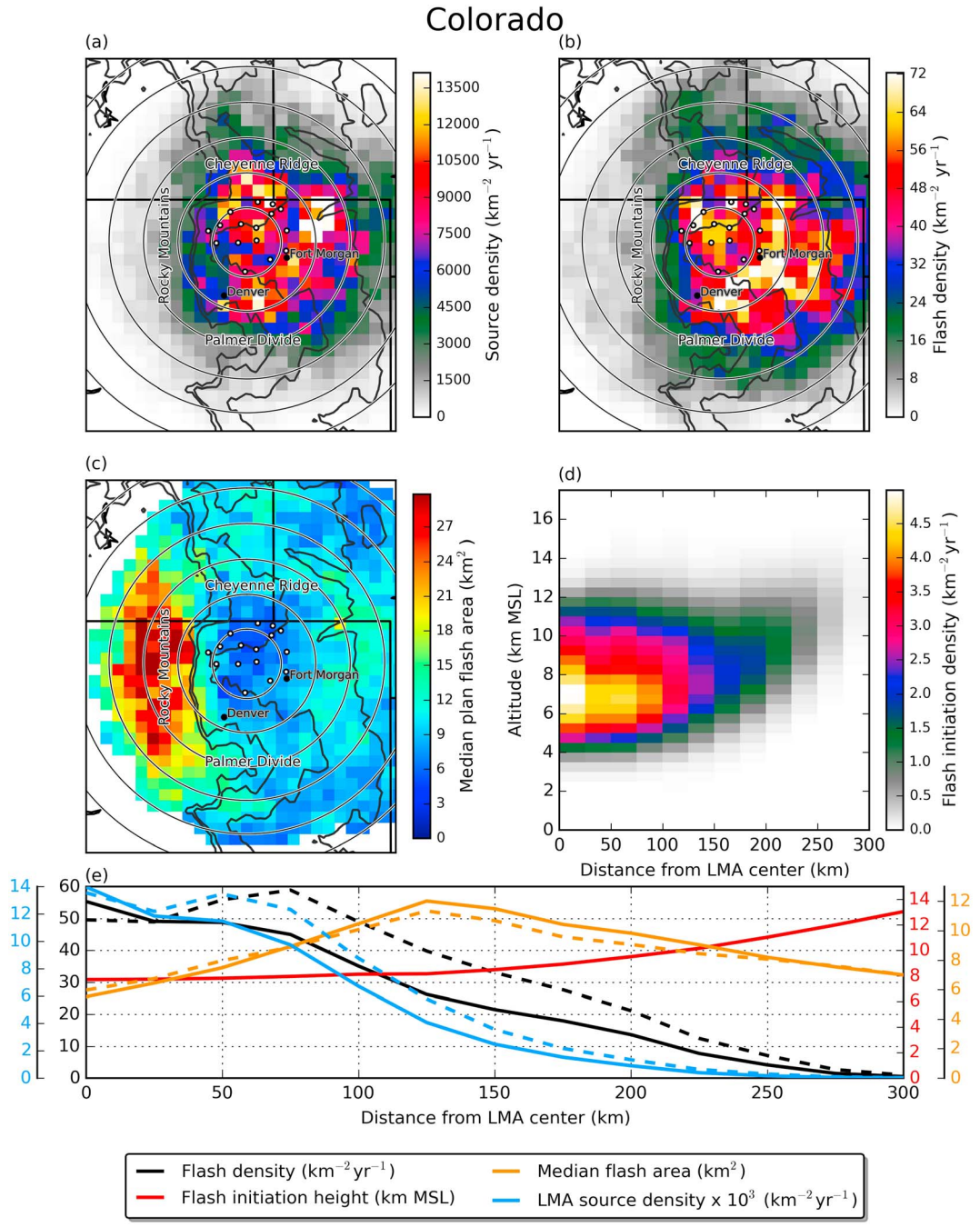

Figure 8. (a) LMA source density for the Colorado network on a 0.15° × 0.2° grid. (b) Total lightning flash density from the clustering algorithm on the same grid. (c) Median plan flash area for all flashes within a grid box. Locations with less than 1 flash km^{-2}yr^{-1} are excluded to minimize outlier effects. Dark gray contours in Figures 8a–8c are 1200, 1500, 1800, 2000, and 4000 m MSL ground elevation. (d) Flash initiation height density as a function of distance from the center of the network. (e) Flash quantities azimuthally integrated to give behavior as a function of distance from the center of the network. Dotted lines indicate values when only considering sources and flashes in the eastern portion of the network. White dots indicate the locations of LMA stations present during the analysis period. Range rings are increments of 50 km.

minimum in the other regions. This implies that the vertical distribution of electric fields is drastically different in Colorado storms compared to Alabama or D. C. storms. Indeed, *Fuchs et al*. [2015] found that storms in Colorado exhibited different charge structures than storms in Alabama or D. C. Storms in Colorado were much more likely to possess an anomalous charge structure, characterized by strong low-level or midlevel

positive charge when compared to a normal charge structure, which is characterized by strong midlevel negative and upper level positive charge. It is then possible that the locations of strongest electric fields are located at different heights, as Figure 8d suggests. The rapid decrease in flash densities past 100 km is due in large part to the presence of the Rocky Mountains and the lack of flashes there. Inspecting Figure 8 e shows that flash densities do not drop off until roughly 200 km from the center of the network for flashes in the eastern portion of the domain. The effect of height location errors can be observed in the increasing flash heights between 200 and 300 km from the network (Figure 8d).

3.3.2. Range Sensitivity of Flashes

In an effort to remove the effects of terrain on the lightning characteristics, only flashes within the eastern half of the Colorado LMA domain were grouped based on their distance from the LMA center and analyzed to investigate the impacts that detection variations have on calculated flash characteristics. In Figure 9a, it is immediately apparent that the sensitivity of flash characteristics to network proximity is markedly different from the other networks. The median duration holds steady around 200 ms, regardless of distance from the LMA, out to 250 km. The consistency of flash duration with range is evidence of the network's high detection efficiency. It should be noted, however, that the distributions of flash durations within the network are remarkably similar to corresponding flashes in D. C. and Alabama.

A similar signature is present in Figure 9b, which shows the plan position flash areas with respect to range from the LMA. The median flash areas increase from 5 km^2 inside of 25 km up to 12 km^2 at 150 km and decreases to 9 km^2 at 250 km. It is important to note here that the median flash area is actually larger at 250 km than inside of 25 km. Since this analysis is restricted to the flashes in the eastern portion of the domain for reasons discussed earlier, this provides strong evidence for the high sensitivity of the Colorado network. There is a maximum in median flash area at moderate ranges from the LMA center, similar to Alabama. However, looking at various percentiles of the distribution reveals a different behavior than the Alabama network. All major percentiles of flash area increase from the center of the network out to approximately 125 km, suggesting that detected flashes may be larger coincident with smaller flashes going undetected. Given the complex terrain in this region, it is reasonable to suspect that physical flash characteristics may vary with terrain, and that may be influencing these relationships. However, it is extremely difficult to decouple those potential impacts from detection artifacts.

The distribution of average flash powers in Figure 9c reveals a consistent behavior for all ranges from the LMA center. The consistency provides some evidence of the high detection efficiency of the CO LMA. The 5th percentile of all distributions is above 5 dBW, indicating that few flashes have low average power.

The number of points per flash in the Colorado region (Figure 9d) is much larger than either the Alabama or D. C. regions. For flashes within the network, the median number of points per flash is over 100 or approximately a factor of 3 more than in D. C. and a factor of 2 than in Alabama. Nonetheless, the decrease in sensitivity at long range is evident as the median number of points monotonically decreases to approximately 30 points per flash at 250 km, still greater than any distribution of flashes in Alabama or D. C. It is possible that physical flashes are being missed at distances around 250 km due to the strict 10-point threshold in this region, implying that these flash density values may be a lower bound on the real flash density value.

The number of thunderstorm hours in the Colorado region, shown in Figure 9e, reveals some undeniable spatial patterns. Immediately evident is the maximum in the sharp elevation gradient in the foothills of the Rocky Mountains, with an eastward extension on the Palmer Lake Divide, south of Denver. Somewhat surprising are both the values over the network and the location of maximum flash density north and east of Denver. The number of lightning hours in these areas is approximately 90–100 per year, similar values to those in the D. C. region and about a factor of 2 lower than in Alabama. Since the values of flash density are much larger in Colorado than in the other regions, it follows that more lightning is produced per storm. This is borne out in Figure 9f, which shows average grid total flash rates as high as 180, roughly a factor of 3 larger than either Alabama or D. C. This is in accordance with the *Fuchs et al.* [2015] hypothesis that the unique thermodynamic ingredients in this region conspire to produce intense, highly electrified storms in Colorado. Note that the foothills of the Rocky Mountains, location of the lightning hours maximum, have a very low average grid total flash rate.

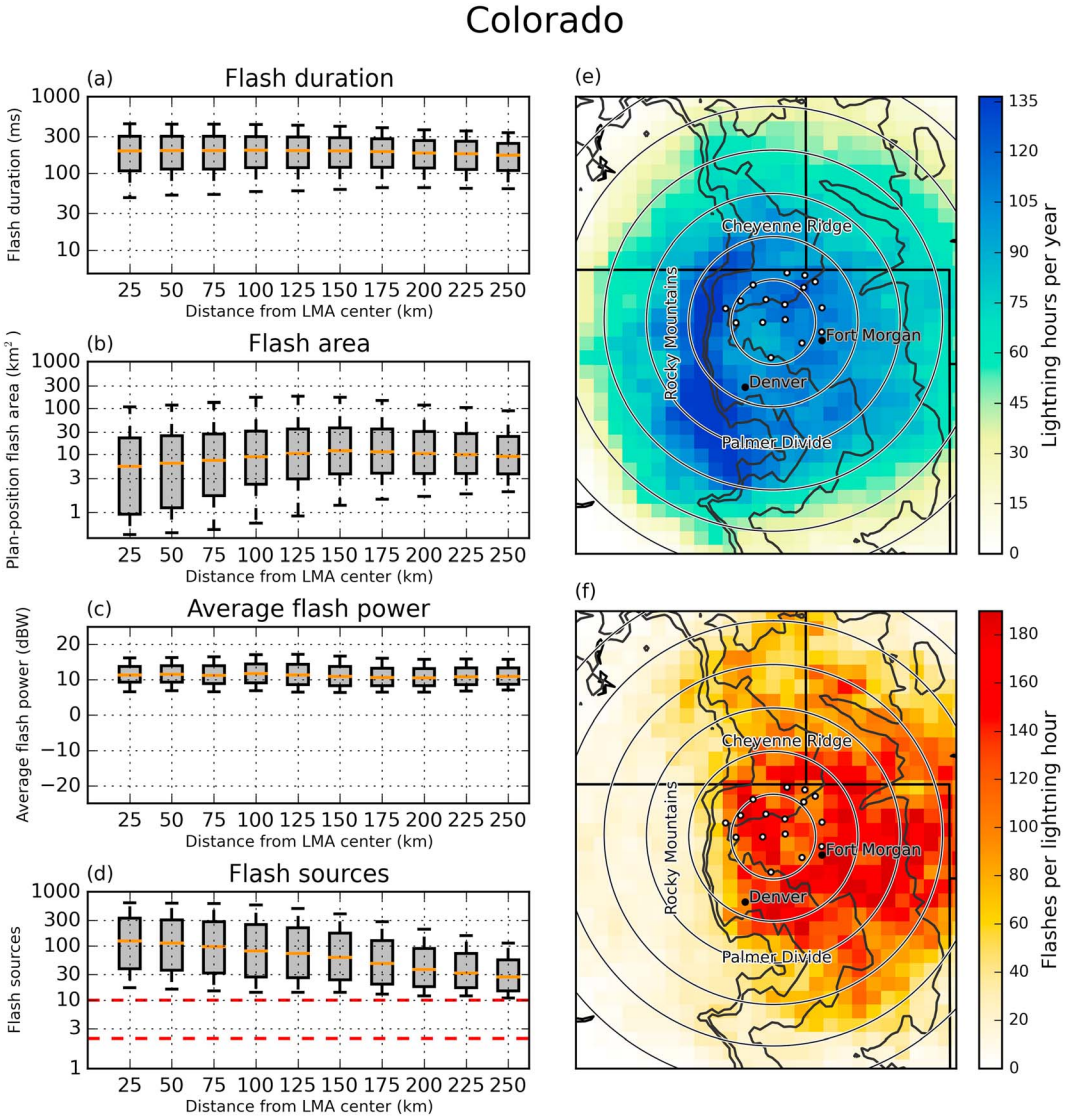

Figure 9. (a) Distributions of flash duration for flashes in each range bin from the Colorado network center. Bars indicate the median value. Top and bottom of the box indicate the 25th and 75th percentiles, respectively. Whiskers indicate the 5th and 95th percentiles, respectively. (b) Similar to Figure 9a for plan position flash area. (c) Similar to Figure 9a for the average source power within a flash. (d) Similar to Figure 9a for the total number of points associated with each flash. The red dashed lines indicate the typical thresholds of 2 and 10 points for reference. (e) Map of lightning hours for each grid box. Indicative of how many hours during a year at least one detected flash occurred in a particular grid box. (f) Map of the average number of flashes that occurred during an hour when at least one lightning flash occurred. This can be thought of as an average grid total flash rate.

3.3.3. NLDN CG Lightning

The local terrain undoubtedly influences the production of CG flashes in the Colorado region. Studies with longer data sets have shown a longitudinal gradient of CG flash density along the foothills of the Rocky Mountains in addition to a latitudinal gradient in the adjacent plains [*Zajac and Rutledge*, 2001; *Orville and Huffines*, 2001; *Vogt and Hodanish*, 2014]. Specifically, CG density maxima are present on the relatively higher terrain of the Palmer Divide and the Cheyenne Ridge (~1.8 km MSL) and a minimum in the Platte River valley (~1.5 km MSL). The limited temporal extent of this data set shows similar features, with a maximum of CG activity to the east of Denver on the north side of the Palmer Divide in Figure 10a.

The large total flash density values in this region result in much higher IC:CG values than previous studies [e.g., *Boccippio et al.*, 2001]. Maximum IC:CG values in Figure 10b approach 25 near the network and in the

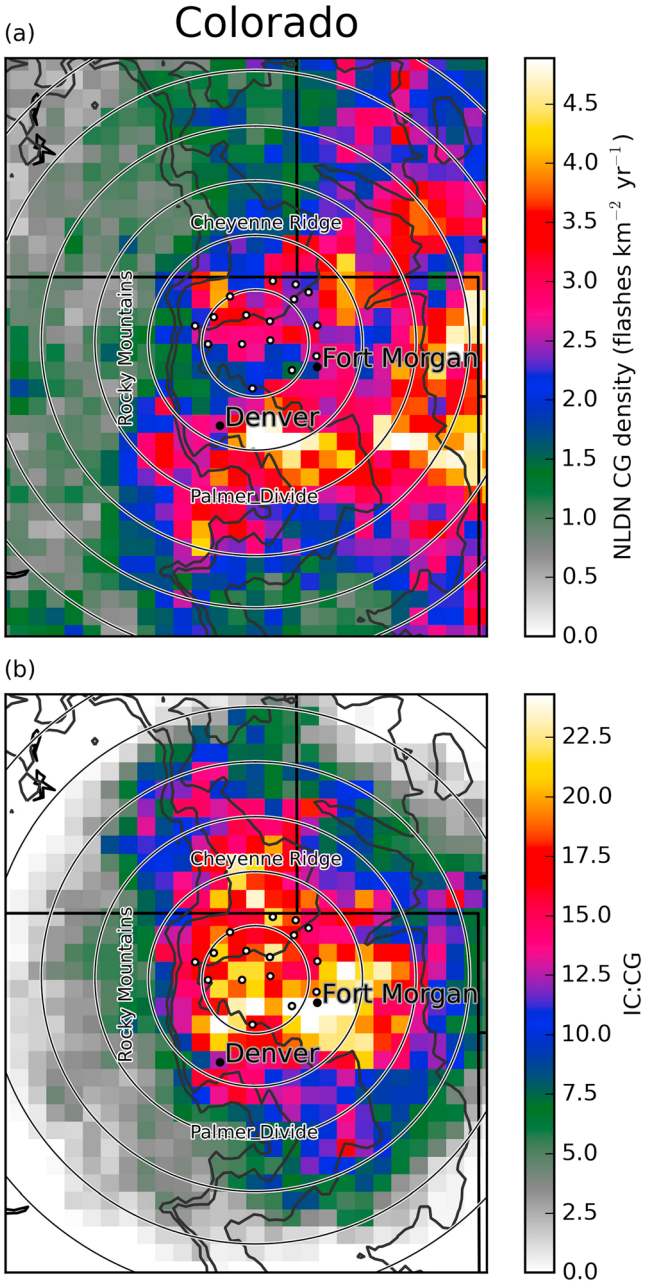

Figure 10. (a) Peak-current-filtered CG flash density map from the NLDN from the same time frame as the available LMA data in Colorado. (b) Derived IC:CG values from the NLDN CG rate and the LMA-calculated total flash rate assuming 100% detection efficiency. White dots indicate the locations of LMA stations present during the analysis period. Range rings are increments of 50 km.

Fort Morgan area. These values are striking, considering that they are approximately a factor of 3 higher than in *Boccippio et al*. [2001]. Even though CG production is substantially lower in the Rocky Mountains compared to the adjacent plains, IC:CG values are also much lower. This is in accordance with the much lower total flash rates and larger flash sizes, hypothesized to result from different storm environments and weaker vertical motions in storms over the Rocky Mountains.

The numbers here likely represent a lower bound on the actual IC:CG values, since we are assuming that all CG flashes are detected by the LMA. CG flashes typically produce very little VHF radiation near the ground; however, there is usually plenty of in-cloud activity associated with any CG flash. For this study, we have made no adjustments to the total flash counts and assume that the detection efficiency of the LMA is 100% because no published detection efficiencies currently exist. This is obviously not correct but should be close for flashes within the network (~50 km from the network center).

3.4. Flash Merging

In an effort to test and improve the performance of the algorithm, a simple flash-merging model was implemented. The goal of this was to mitigate the errors produced by the algorithm separating branches of one larger physical flash into multiple smaller flashes. Separation of branches may result in erroneously high values of flash rates and densities while producing erroneous flash characteristics as well. The simple model implemented here tests for nearby flashes that are sufficiently close in time. More specifically, if the initial point of flash is within 150 ms and 3 km from any point in any other flash, that flash is considered as a branch to the earlier nearby flash by the model. This model was tested on isolated cells with maximum reflectivity above 55 dBZ (to select the strongest storms with the most lightning) from the storm database produced by *Fuchs et al*. [2015]. Figures 11a, 11c, and 11e show the total number of flashes in each cell and the number of merged branches in Alabama, D. C., and Colorado, respectively. The merged branches can be thought of as the decreased total flash number for each cell. A linear least squares best fit slope is also indicated in each panel. The slope represents the percentage decrease of total flashes as a result of the leader merging. The average decrease in flash

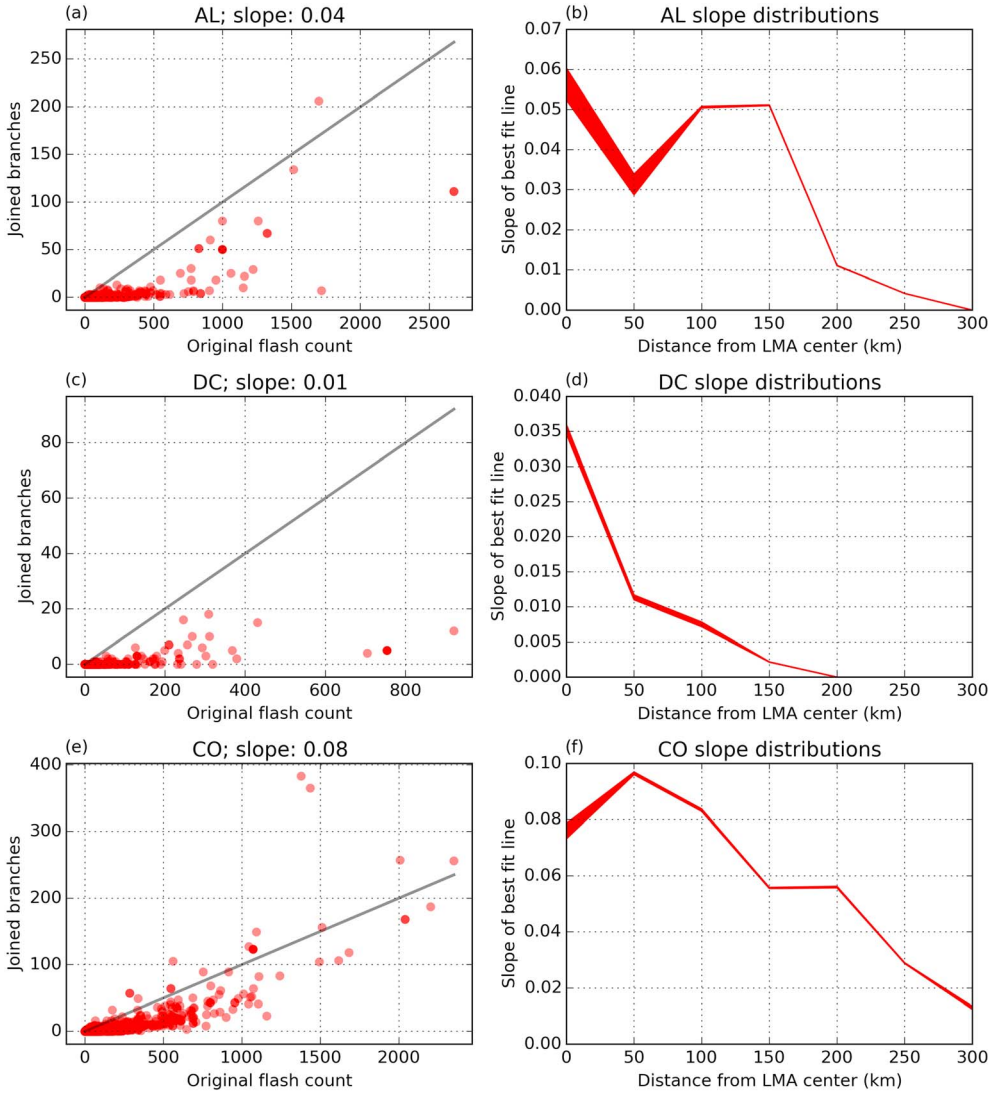

Figure 11. (a) Scatterplot of original number of flashes in a cell and the number of flashes joined together in Alabama. The black line indicates that 10% of flashes have been joined together. The slope of the least squares best fit line is indicated in the title. (b) The slope of the best fit line if only considering cells in a binned distance from the Alabama network center. The width of the line indicates the 5th and 95th percentile of slopes using the jackknife method. (c) Same as Figure 11a for the D. C. region. (d) Same as Figure 11b for the D. C. region. (e) Same as Figure 11a for the Colorado region. (f) Same as Figure 11b for the Colorado region.

number is 4% for Alabama storms, 1% for D. C. storms, and 8% for Colorado storms. That the effect is largest in Colorado is somewhat surprising, since it is the most sensitive network in this study. A possible explanation is that the greater network sensitivity in Colorado resolves lightning channels more fully, making them easier to meet the proximity criteria to other parts of the flash. Conversely, it would be more difficult to meet the proximity criteria if networks with lower sensitivity are not resolving parts of the lightning channels.

To assess the dependence of the simple flash-merging model on the detection efficiency, storms were partitioned by distance from the network center and a similar least squares fit was performed for each population of cells. Figures 11b, 11d, and 11f show the slope of the linear fit as a function of distance from the network center storms in Alabama, D. C., and Colorado, respectively. There is a general decreasing trend with increasing distance from the center of the network, particularly in D. C. and Colorado, meaning that fewer leaders are merged to parent flashes at longer ranges. This may be the result of decreasing detection sensitivity resulting in branches not meeting proximity criteria at longer ranges, similar to the larger percentage of leaders merged in Colorado, compared to the other regions. Slope sensitivities to particular cells are indicated by

the thickness of the lines using the jackknife resampling method [*Efron*, 1982]. There is more sensitivity to individual storms at short range because the cell sample size is smaller, in large part because the area of an annulus depends on its radii. Assuming a constant density of storms, there will be fewer storms near the network center, compared to farther from the network.

4. Summary and Discussion

4.1. Flash Algorithm

This paper has described a new open-source flash-clustering algorithm designed to perform the same basic function of spatial and temporal VHF source grouping as other flash-clustering algorithms. One of the key differences with this open-source algorithm is that it facilitates collaboration and continuous improvements by the larger lightning community. By virtue of the algorithm's ability to automatically process large amounts of LMA data, this study is the first of its kind to provide climatological-scale analyses on millions of LMA-detected lightning flashes in multiple regions of the United States with distinct environmental characteristics. These results highlight some of the differences between the LMA networks in Alabama, D. C., and Colorado. Some of the strengths and weaknesses of the algorithm were elucidated, which may steer future improvements to the algorithm.

4.2. Flash and Detection Characteristics

Comparisons between flash distributions partitioned by distance from the LMA showed some differences between the different networks. Flash characteristics in D. C. exhibited strong sensitivity to proximity to the network center, suggesting that an appreciable number of sources go undetected. The proximal dependence of flash characteristics is weaker in Alabama, but sensitivities still exist and may bias estimates of flash area integrated over a storm. Flash characteristics in the Colorado region were shown to be much less sensitive to network proximity. This is expected because the Colorado network has the most stations and lowest noise floors of any permanent network currently in operation.

It is these differences in network proximity sensitivity between networks that highlight some of the shortcomings of the algorithm and differences in network detection efficiencies. There are a couple of factors that contribute to network sensitivity. First, since VHF radiation sources radiate isotropically, the power flux decreases rapidly with distance, resulting in undetected sources (particularly weak ones) at long distances. This is an unavoidable problem determined by the laws of physics and the noise floor of a particular network.

Errors of source location estimation increase rapidly with distance outside of the network due to the geometry of intersecting hyperbolas [*Thomas et al*., 2004]. This may result in VHF sources being scattered in space and time, to the point where sources are no longer close enough to meet the clustering thresholds. A potential remedy to this problem is to loosen the clustering thresholds with increasing distance, known as "adaptive thresholding" in *McCaul et al*. [2009]. This would likely result in more detected flashes by the algorithm. The effect on flash characteristics is not immediately obvious and would require a comparison of flash characteristics before and after the adaptive thresholding is applied. Additionally, there are actually competing effects on the flash area calculation: the source errors and the undetected sources, the former increasing the calculated area and the latter decreasing the calculated area.

Incorrectly classifying a branch of a lightning flash as a separate flash has been a known problem in flash-clustering algorithms [*Thomas et al*., 2003; *MacGorman et al*., 2008]. If a branch is delayed by a time interval after the start of the parent flash, it may be identified as a separate flash, thereby incorrectly inflating the number of flashes in storm. This phenomenon needs to be understood in a quantitative manner to ensure that the flash counts and characteristics are representative of the physical flashes that are being detected. A simple model for merging branches to their parent flashes was attempted in this study. Figure 11 shows that the effects separating leaders was minimal in most cases. Storms in Colorado were most sensitive to the branch-merging model; however, flash counts were decreased by more than 10% in an extremely small number of cases.

For all the observed differences in network sensitivities and, consequently, flash quantities between the regions, some similarities also arise. The distributions of flash durations are remarkably similar between regions, particularly if only considering flashes within each network. In D. C., the 5th, 50th, and 95th percentiles of flash duration are approximately 25, 200, and 500 ms, respectively. In Alabama, the 5th, 50th, and 95th

percentiles of flash duration are approximately 25, 200, and 700 ms, respectively. In the plains of Colorado (the eastern half of the network), the 5th, 50th, and 95th percentiles of flash duration are approximately 50, 200, and 500 ms, respectively. Recall that we are only considering the eastern half of the Colorado domain to mitigate terrain effects. The distributions of plan position flash area are likewise similar in all three regions. The similarities of calculated lightning flash quantities provide some confidence that the algorithm is performing as expected. The differences in flash power and number of flash points rely on the detection of LMA sources and are consequently going to have differences based on the ability of each network to detect VHF radiation produced by lightning. It is encouraging to see that variations in flash characteristics far from the network are smooth, which suggests that those flashes with few points are indeed flashes and not random noise being clustered into a flash.

4.3. Comparisons With Satellite Values

One must bear in mind the detection and data collection differences between LMA and satellite when making comparisons of corresponding climatologies. Recall that satellites detect optical radiation produced by lightning, while LMA networks detect VHF radiation produced by discontinuous lightning breakdown. LMA networks continuously detect the same storms, while satellites get a snapshot of lightning in a storm.

Average flash densities within the D. C. network (where LMA detection efficiency is highest) were 18 flashes $km^{-2}\,yr^{-1}$ compared to 10–15 flashes $km^{-2}\,yr^{-1}$ for the combined LIS/OTD climatology (HRFC_COM product from http://thunder.nsstc.nasa.gov; *Cecil et al*. [2014]). LMA flash densities over the northern Alabama network are roughly 33% larger than comparable satellite observations (~40 versus ~30 flashes $km^{-2}\,yr^{-1}$) by *Christian et al*. [1999, 2003] and *Cecil et al*. [2014]. It should be noted that these values did not significantly change if gridded to the $0.5° \times 0.5°$ resolution of *Cecil et al*. [2014]. The comparisons between LMA- and satellite-constructed climatologies in D. C. and Alabama are relatively close to each other. In both of these regions, the large majority of storms are generally weak with low flash rates [*Fuchs et al*., 2015]. Weaker storms are hypothesized to have less turbulence and correspondingly larger flashes [*Bruning and MacGorman*, 2013].

By a wide margin, the largest difference between LMA and satellite flash densities is in the northeast Colorado region. The average LMA flash densities within the network are approximately 50–55 flashes $km^{-2}\,yr^{-1}$, compared to 15–20 flashes $km^{-2}\,yr^{-1}$ for satellite detectors (OTD at latitudes greater than ~38°). This striking difference is roughly a factor 3 times larger for the LMA climatology than the OTD climatology. The flash sensitivity studies in this paper indicate the highest confidence in the flash density values in the Colorado region. Flash characteristics were least sensitive to network proximity, and the strictest clustering thresholds were imposed in this region (section 2). It is difficult to estimate error bars with the LMA flash density given the numerous unknowns that cannot be tested with millions of flashes. Given the comparisons with XLMA in *Fuchs et al*. [2015] and subjective analysis of several storms (similar to Figure 3 from *Fuchs et al*. [2015]), we estimate the errors to be approximately 10–20%. The differences between LMA and satellite climatologies are far outside of these bounds. It is worth noting that the short duration of both the LMA observation period of 2 years and the satellite observation of 5 years in Colorado may play a role in the large differences between the corresponding flash density estimates, especially because the observation periods are not concurrent. Additionally, the satellite climatologies are gridded more coarsely, which will smooth out small-scale variations in flash density, which may have otherwise had higher values. However, the coarse grid should not result in any missed flashes by the satellites and therefore should remain a valid comparison with azimuthally averaged values of flash density, as in Figures 2e, 5e, and 8e.

Perhaps not coincidentally, the flash density differences between satellite and LMA data sets are correlated with storm intensity. *Fuchs et al*. [2015] showed that isolated storms in Colorado most frequently produced large flash rates greater than 10 flashes min^{-1}. This was of particular importance for the bandwidth-limited OTD instrument, as storms with large flash rates sometimes saturated the instrument and resulted in an erroneously low flash rate. It is important to note that the bandwidth limit will not be a problem on the GLM instrument to be launched later this year. *Fuchs et al*. [2015] showed that storms in Colorado most frequently produced high flash rates and had the highest IC:CG values of any region in the study. It is reasonable to suppose that the flash area distribution in high flash rate storms is skewed toward smaller flashes, following *Bruning and MacGorman* [2013]. Those small flashes, more common in Colorado storms, may go undetected by satellites. Furthermore, comparisons of Figures 2d, 5d, and 8d reveal that a much larger fraction of flashes

initiate at lower altitudes in Colorado storms compared to Alabama or D. C. storms. The optical path for these flashes to reach satellite detectors is much longer than flashes in the upper portion of the cloud, which may result in undetected flashes by satellite instruments. This is corroborated by *Thomas et al*. [2000], which found that the LIS instrument detected a relatively small fraction of flashes that occurred in the lower portions of an Oklahoma storm. *Fuchs et al*. [2015] showed that a substantial fraction of storms in Colorado contained inferred midlevel positive charge, in contrast with storms in the Alabama and D. C. regions, suggesting that charge configurations in a storm may impact the number of lightning flashes detected by satellites during an overpass. By extension, regions with similar environments conducive to strong storms with a preference for low-altitude IC flashes (e.g., high cloud base height and instability) may have similarly undetected flashes.

Acknowledgments

The authors wish to thank Doug Stolz at CSU for relevant discussions and Paul Hein for technical help throughout this study. The authors would also like to thank the three anonymous reviewers for their suggestions and comments, all of which contributed to the improvement of the paper. Rich Blakeslee at NASA Marshall Space Flight Center provided the Alabama and Washington, D.C., LMA data. The LMA networks and data are maintained at New Mexico Tech. The UAH NSF DC3 award and National Science Foundation's Physical and Dynamical Meteorology (NDF PDM) Program (AGS-1063573) support the Alabama LMA and related research. Inquiries or supplemental information about the data used in this study should be directed to the corresponding author (brfuchs@atmos.colostate.edu). Funding for this study was provided by NSF grant AGS-1010657 and NOAA Proving Ground grant NA14OAR4320125.

References

Barth, M. C., et al. (2015), The deep convective clouds and chemistry (DC3) field campaign, *Bull. Am. Meteorol. Soc.*, *96*(8), 1281–1309.

Basarab, B., S. Rutledge, and B. Fuchs (2015), An improved lightning flash rate parameterization developed from Colorado DC3 thunderstorm data for use in cloud-resolving chemical transport models, *J. Geophys. Res. Atmos.*, *120*, 9481–9499, doi:10.1002/2015JD023470.

Behnke, S. A., R. J. Thomas, P. R. Krehbiel, and W. Rison (2005), Initial leader velocities during intracloud lightning: Possible evidence for a runaway breakdown effect, *J. Geophys. Res.*, *110*, D10207, doi:10.1029/2004JD005312.

Boccippio, D., K. L. Cummins, H. J. Christian, and S. J. Goodman (2001), Combined satellite and surface-based estimation of the intracloud-cloud-to-ground lightning ratio over the continental United States, *Mon. Weather Rev.*, *129*, 108–122.

Bruning, E. C., and D. R. MacGorman (2013), Theory and observations of controls on lightning flash size spectra, *J. Atmos. Sci.*, *70*, 4012–4029.

Bruning, E. C., W. D. Rust, T. J. Schuur, D. R. MacGorman, P. R. Krehbiel, and W. Rison (2007), Electrical and polarimetric radar observations of a multicell storm in TELEX, *Mon. Weather Rev.*, *135*(7), 2525–2544.

Budden, K. G. (1988), The propagation of radio waves, in *The Theory of Radio Waves of Low Power in the Ionosphere and Magnetosphere*, Cambridge Univ. Press, Cambridge, U. K.

Carey, L. D., and K. M. Buffalo (2007), Environmental control of cloud-to-ground lightning polarity in severe storms, *Mon. Weather Rev.*, *135*, 1327–1353.

Cecil, D. J., D. E. Buechler, and R. J. Blakeslee (2014), Gridded lightning climatology from TRMM-LIS and OTD: Dataset description, *Atmos. Res.*, *135*, 404–414.

Christian, H. J., et al. (1999), The Lightning Imaging Sensor, in *Proc. 11th Int. Conf. on Atmospheric Electricity*, vol. 1, pp. 746–749, ICAE, Huntsville, Ala.

Christian, H. J., et al. (2003), Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. Geophys. Res.*, *108*(D1), 4005, doi:10.1029/2002JD002347.

Coleman, L., T. Marshall, M. Stolzenburg, T. Hamlin, P. Krehbiel, W. Rison, and R. Thomas (2003), Effects of charge and electrostatic potential on lightning propagation, *J. Geophys. Res..*, *108*(D9), 4298, doi:10.1029/2002JD002718.

Crook, N. A., T. L. Clark, and M. W. Moncrieff (1990), The Denver cyclone. Part I: Generation in low Froude number flow, *J. Atmos. Sci.*, *47*, 2725–2742.

Cummins, K., and M. Murphy (2009), An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the US NLDN, *Electromagn. Compat.*, *51*, 499–518.

Deierling, W., and W. A. Petersen (2008), Total lightning activity as an indicator of updraft characteristics, *J. Geophys. Res.*, *106*, D16210, doi:10.1029/2007JD009598.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, vol. 38, Society for industrial and applied mathematics, Philadelphia.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996), A density-based algorithm for discovering clusters in large spatial databases with noise, *KDD*, *96*, 226–231.

Fuchs, B. R. (2014), Factors affecting lightning behavior in various regions of the United States, Master's thesis, Colo. State Univ., Fort Collins.

Fuchs, B. R., S. A. Rutledge, E. C. Bruning, J. R. Pierce, J. K. Kodros, T. J. Lang, D. R. MacGorman, P. R. Krehbiel, and W. Rison (2015), Environmental controls on storm intensity and charge structure in multiple regions of the continental United States, *J. Geophys. Res. Atmos.*, *120*, 6575–6596, doi:10.1002/2015JD023271.

Goodman, S. J., et al. (2013), The GOES-R geostationary lightning mapper (GLM), *Atmos. Res.*, *125*, 34–49.

Jirak, I. L., and W. R. Cotton (2006), Effect of air pollution on precipitation along the Front Range of the Rocky Mountains, *J. Appl. Meteorol. Climatol.*, *45*(1), 236–245.

Koshak, W., et al. (2004), North Alabama lightning mapping array (LMA): VHF source retrieval algorithm and error analyses, *J. Atmos. Oceanic Technol.*, *21*(4), 543–558.

Lang, T. J., and S. A. Rutledge (2011), A framework for the statistical analysis of large radar and lightning datasets: Results from STEPS 2000, *Mon. Weather Rev.*, *139*, 2536–2551.

MacGorman, D. R., J. M. Straka, and C. L. Ziegler (2001), A lightning parameterization for numerical cloud models, *J. Appl. Meteorol.*, *40*, 459–478.

MacGorman, D. R., et al. (2008), TELEX the thunderstorm electrification and lightning experiment, *Bull. Am. Meteorol. Soc.*, *89*, 997–1013.

Mach, D. M., H. J. Christian, R. J. Blakeslee, D. J. Boccippio, S. J. Goodman, and W. L. Boeck (2007), Performance assessment of the optical transient detector and lightning imaging sensor, *J. Geophys. Res.*, *112*, D09210, doi:10.1029/2006JD007787.

Maggio, C., L. Coleman, T. Marshall, M. Stolzenburg, M. Stanley, T. Hamlin, P. Krehbiel, W. Rison, and R. Thomas (2005), Lightning-initiation locations as a remote sensing tool of large thunderstorm electric field vectors, *J. Atmos. Oceanic Technol.*, *22*(7), 1059–1068.

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

Medici, G., K. L. Cummins, W. J. Koshak, S. D. Rudlosky, R. J. Blakeslee, S. J. Goodman, D. J. Cecil, and D. R. Bright (2015), The intra-cloud lightning fraction in the contiguous United States.

Nag, A., M. Murphy, A. Pifer, K. Cummins, and J. Cramer (2013), Upgrade of the US National Lightning Detection Network in 2013, in AGU Fall Meeting Abstracts, vol. 1, p. 0337.

Nag, A., M. J. Murphy, W. Schulz, and K. L. Cummins (2015), Lightning locating systems: Insights on characteristics and validation techniques, *Earth Space Sci.*, *2*(4), 65–93.
Orville, R. E., and G. R. Huffines (2001), Cloud-to-ground lightning in the United States: NLDN results in the first decade, *Mon. Weather Rev.*, *125*, 631–638.
Pedregosa, F., et al. (2011), Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, *12*, 2825–2830.
Petersen, W. A., H. J. Christian, and S. A. Rutledge (2005), TRMM observations of the global relationship between ice water content and lightning, *Geophys. Res. Lett.*, *32*, L14819, doi:10.1029/2005GL023236.
Rison, W., R. Thomas, P. Krehbiel, T. Hamlin, and J. Harlin (1999), A GPS-based three-dimensional lightning mapping system: Initial observations in central New Mexico, *Geophys. Res. Lett.*, *26*, 3573–3576, doi:10.1029/1999GL010856.
Rudlosky, S. D. (2015), Evaluating ENTLN performance relative to TRMM/LIS, *J. Oper. Meteorol.*, *3*(2).
Said, R., M. Cohen, and U. Inan (2013), Highly intense lightning over the oceans: Estimated peak currents from global GLD360 observations, *J. Geophys. Res. Atmos.*, *118*, 6905–6915, doi:10.1002/jgrd.50508.
Saunders, C. P. R., W. D. Keith, and R. P. Mitzeva (1991), The effect of liquid water on thunderstorm charging, *J. Geophys. Res.*, *96*, 11,007–11,017, doi:10.1029/91JD00970.
Schultz, C. J., L. D. Carey, E. V. Schultz, and R. J. Blakeslee (2015), Insight into the kinematic and microphysical processes that control lightning jumps, *Weather Forecasting*, doi:10.1175/WAF-D-14-00147.1.
Shumpert, T., M. Honnell, and G. Lott Jr. (1982), Measured spectral amplitude of lightning sferics in the HF, VHF, and UHF bands, *IEEE Trans. Electromagn. Compat.*, *3*, 368–369.
Takahashi, T. (1978), Riming electrification as a charge generation mechanism in thunderstorms, *Atmos. Sci.*, *35*, 1536–1548.
Thomas, R. J., P. R. Krehbiel, W. Rison, T. Hamlin, D. J. Boccippio, S. J. Goodman, and H. J. Christian (2000), Comparison of ground-based 3-dimensional lightning mapping observations with satellite-based LIS observations in Oklahoma, *Geophys. Res. Lett.*, *27*(12), 1703–1706, doi:10.1029/1999GL010845.
Thomas, R. J., P. R. Krehbiel, W. Rison, S. J. Hunyady, W. P. Winn, T. Hamlin, and J. Harlin (2004), Accuracy of the lightning mapping array, *J. Geophys. Res.*, *109*, D14207, doi:10.1029/2004JD004549.
Thomas, R., P. Krehbiel, W. Rison, J. Harlin, T. Hamlin, and N. Campbell (2003), The LMA flash algorithm, in *Proc. 12th Int. Conf. on Atmospheric Electricity*, pp. 655–656, International Commission on Atmospheric Electricity, Versailles, France.
van der Velde, O. A., and J. Montanyà (2013), Asymmetries in bidirectional leader development of lightning flashes, *J. Geophys. Res. Atmos.*, *118*, 13,504–13,519, doi:10.1002/2013JD020257.
Vogt, B. J., and S. J. Hodanish (2014), A high-resolution lightning map of the state of Colorado, *Mon. Weather Rev.*, *142*(7), 2353–2360.
Wiens, K. C., S. A. Rutledge, and S. A. Tessendorf (2005), The 29 June 2000 supercell observed during STEPS. Part II: Lightning and charge structure, *J. Atmos. Res.*, *62*, 4151–4177.
Williams, E. R. (1985), Large-scale separation in thunderclouds, *J. Geophys. Res.*, *90*, 6013–6025, doi:10.1029/JD090iD04p06013.
Williams, E. R., V. Mushtak, D. Rosenfeld, S. Goodman, and D. Boccippio (2005), Thermodynamic conditions that lead to superlative updrafts and mixed-phase microphysics, *Atmos. Res.*, *76*, 288–306.
Zajac, B., and S. A. Rutledge (2001), Cloud-to-ground lightning activity in the contiguous United States from 1995 to 1999, *Mon. Weather Rev.*, *129*, 999–1019.