7.3 PRELIMINARY RESULTS FROM THE INCLUSION OF LIGHTNING TYPE AND POLARITY IN THE IDENTIFICATION OF SEVERE STORMS

Martin J. Murphy
Vaisala Inc., Louisville, CO, U.S.

ABSTRACT:
To date, the literature on the use of lightning information to distinguish severe thunderstorms from ordinary storms has focused primarily on the "lightning jump", the rapid increase in total flash rate that is generally associated with a strengthening of the storm. The lightning jump, however, ignores other characteristics such as the proportions of cloud and cloud-to-ground (CG) flashes within the total flash rate and the proportions of positive- and negative-polarity flashes, CG strokes, and individual cloud pulses within the flashes. In this work, we present results from an analysis of more than 1000 thunderstorm cells sampled from most of the continental U.S. by the U.S. National Lightning Detection Network (NLDN) on a number of days in 2015 and 2016 when both severe and non-severe thunderstorms occurred. The sample was chosen to represent the real proportion of severe thunderstorms, generally considered to be around 10% or less of all thunderstorms in the U.S. All of these storms were observed by the NLDN in its post-2013 upgraded state, when its cloud flash detection efficiency was between 40-60% and its CG flash detection efficiency was more than 95%, and thus, the NLDN served as a continental-scale total lightning network. In this study, the severe storms were identified on the basis of proximity in space and time to severe storm reports compiled by the NOAA Storm Prediction Center. The objective of the study is a preliminary assessment of whether the inclusion of parameters such as cloud flash fraction and separate rates of positive and negative cloud and CG flashes, strokes, and pulses provide any additional value on top of the total flash rate in terms of distinguishing between severe and non-severe storms in a representative sample of all thunderstorms from around the U.S.

1. Introduction
In November 2016, the U.S. launched the first geostationary satellite ever to bear an optical lightning imaging instrument, known as the Geostationary Lightning Mapper (GLM). As described in Goodman et al. (2013), the eventual GLMs on both GOES-E and GOES-W satellites will cover most of the western hemisphere between approximately 50° S latitude and 50° N latitude, with a pixel size (and thus, spatial resolution) of 8 km at the equator and 14 km at the edges. These instruments are expected to provide a spatially uniform total lightning flash detection efficiency (DE) between 70-90%. In this paper, the term "total lightning" refers to the combination of cloud and cloud-to-ground flashes. The combination of high DE of total lightning and wide area of coverage is the primary benefit of space-based optical monitoring of lightning. Although methods have been developed to estimate the fraction of cloud-to-ground (CG) flashes and flash types from a large sample of optically-detected flashes (e.g. Koshak and Solakiewicz, 2015), the GLM cannot uniquely identify lightning flash type. Optical systems also cannot identify the polarity, or direction of vertical current flow, of the detected discharges.

The high total lightning DE of space-borne optical sensors, coupled with prior literature indicating that CG lightning data alone is poorly related to storm intensity and severity (see literature review provided by Schultz et al. 2011), have led to an emphasis on rapid increases in total flash rate as indicators of storm severity. These rapid increases in total flash rate were originally called "lightning jumps" by Williams et al. (1999), and subsequently, a substantial body of literature has developed around how effectively these jumps can be used to differentiate between ordinary thunderstorms and those storms that produce damaging winds, large hail, or tornadoes, collectively defined, at least in the U.S., as "severe" thunderstorms.

Most quantitative lightning jump algorithms are related in some way to that of Gatlin and Goodman (2010). When 2-minute intervals were used, as in this study, Gatlin and Goodman defined the "lightning jump" as any 2-minute interval in which the rate of change of total flash rate was at least two standard deviations above a weighted, running average of the rate of change during the preceding 10 minutes. It should be noted that many variations on the lightning jump algorithm exist: Gatlin and Goodman described

*Corresponding author address: Martin J. Murphy, Vaisala Inc., 194 S. Taylor Ave., Louisville, CO 80027 U.S.; e-mail: martin.murphy@vaisala.com

the configurable options in their appendix, with their preferred configuration (described in their section 4, figure 7) involving a weighted smoothing process. Schultz et al. (2011) adopted a different variation, without smoothing but with the addition of a 10 flash/minute minimum total flash rate requirement.

1.1 Prior studies of lightning jumps

In the absence of continuous observations of total lightning from space-based sensors, prior studies have investigated the efficacy of the lightning jump using data from ground-based total lightning detection systems, primarily Lightning Mapping Arrays (LMAs; Thomas et al. 2004). The following literature review summarizes the results from these prior studies.

Gatlin and Goodman examined a sample of 20 spring thunderstorms in the vicinity of Huntsville, Alabama, of which 19 were severe storms and only one was non-severe. Gatlin and Goodman matched Severe Storm Reports (hereafter *SSRs*) from the database maintained by the NOAA Storm Prediction Center with thunderstorm cells identified from radar and lightning data. They then applied the lightning jump algorithm to the LMA observations from these storms. They used a standard contingency table event-based verification methodology, counting a “hit” as any SSR having a lightning jump during the 30-minute period preceding the SSR. They found that 90% of the SSRs had a lightning jump during their respective 30 minute periods, with a false alarm ratio (FAR) of 40% (26 out of 65 jumps). False alarms in their study were defined as lightning jumps that did not have a corresponding SSR during a 30-minute time window beginning at the time of the lightning jump.

Schultz et al. (2011) analyzed 711 thunderstorms in four areas of the U.S. where LMA observations were available in order to have a more geographically diverse sample. Of their 711 storms, 255 were severe (35.9%) and the other 456 were non-severe (64.1%). Different from Gatlin and Goodman, Schultz et al. did not apply a weighted smoothing to the rates of change of flash rate but they did add the requirement that the total flash rate be at least 10 flashes/minute in order to remove spurious lightning jumps due to statistical noise at low flash rates. They also extended the time window relating lightning jumps and SSRs to 45 minutes, rather than 30 minutes as used by Gatlin and Goodman. With these modifications, Schultz et al. observed an FAR of 36%. A subset of the false alarms was found to have occurred shortly after another lightning jump that was correlated with an SSR. When Schultz et al. removed such false alarms, the effective net FAR was reduced to 22%.

Rudlosky and Fuelberg (2013) had an even larger sample, 1252 storms, of which 384 (30.7%) were severe and 868 (69.3%) were non-severe. All of these storms were within 150 km of a single LMA near Washington, DC. They applied the lightning jump algorithm in the same way as Schultz et al. (2011) (although they also tested it without the 10 flash/min minimum rate threshold). The contingency table values (POD, FAR) were not presented in their study. However, with the 10 flash/min minimum, they found that 53.7% of the non-severe thunderstorms contained lightning jumps, and when the minimum flash rate requirement was switched off, that percentage increased to 76.4% of the non-severe thunderstorms.

Metzger (2010) opted not to use the Gatlin-like algorithm and instead defined a lightning jump as any sustained increase in the total flash rate of at least 5 flashes/minute over a one-minute period and required that this rate of change be sustained over at least 3 minutes. He analyzed 34 storms in Arizona and Texas, 9 of which were non-severe, such that the proportions of severe and non-severe storms in the sample were 73.5% and 26.5%. A total of 73 lightning jumps were found. All nine of the non-severe storms had at least one lightning jump, and 16 of the 73 jumps occurred somewhere in the non-severe storm sample. Metzger also noted that a number of lightning jumps in the severe storms were not correlated with the actual production of severe weather, and he noted that this led to a “high” FAR. The FAR value was not given outright, but based on the number of reported hits (39) and the number of reported lightning jumps (73), we can infer a FAR value of 46.6%.

The preceding studies used radar data to define thunderstorm cells, and both the lightning data and the SSRs were assigned to the radar-defined storms. Miller et al. (2015), by contrast, used a lightning-based clustering method rather than radar to examine 470 storm clusters, some of which were multicellular. Importantly, only 53 of these storms (11.3%) were severe, while the remaining 417 storms (88.7%) were non-severe. They applied the Gatlin/Schultz-like lightning jump algorithm to the lightning-defined clusters.

They found that the lightning jumps yielded a POD of severe weather events that was comparable to that of Schultz et al. (2011) but that the FAR was significantly higher: Over all storms (the single-cell and multi/super-cell categories given by Miller et al.), the FAR was 87.9% rather than 36%.

Similar to Miller et al. (2015), lightning data alone were also used by Farnell et al. (2016) to identify storm cells. They applied the 2σ lightning jump algorithm of Schultz et al. (2009), which is the same as that applied by Schultz et al (2011) and the same as that used in the present study. However, all of their storms were taken from days with severe weather in Catalonia (NE Spain), and given the small size of Catalonia relative to the continental U.S., that choice yielded a large proportion of severe storms: Of the 179 storms in their sample, 108 (60%) produced severe weather, primarily large hail, while the remaining 71 (40%) were non-severe. The FAR was only 10%, but it is noteworthy that radar data were used to augment ground-based severe storm reports.

1.2 Objectives of the present study

Space-borne optical sensors and ground-based lightning detection networks are expected to provide complementary information. The former provide high total lightning DE over very large areas of the world. The latter provide high CG flash DE, with lower total lightning DE, but they also contribute polarity, lightning type classification at the individual event level (CG stroke or cloud pulse), and peak currents of CG strokes (and a rough equivalent in the case of cloud pulses). In the context of lightning jumps, which can produce high FARs as just shown, the question arises regarding the extent to which information derived from ground-based lightning observations, particularly polarity and lightning type classification, might be able to enhance the lightning jump observations derived from total flash rate observations.

In addition to that primary question, we also seek in this study to include a representative sample of severe and non-severe thunderstorms with geographically diverse sampling. According to the U.S. National Weather Service, only about 10% of all thunderstorms in the U.S. produce severe weather (see http://www.nws.noaa.gov/om/severeweather/resources/ttl6-10.pdf). The current study includes 3350 storms, of which 6.5% are severe, as described in more detail in the Methods section. These 3350 storms were taken from eight different storm days in 2015 and 2016, and they occurred all the way from eastern Washington and Oregon through the inter-mountain west, Great Plains, Great Lakes, as well as all of the southern U.S. and up the east coast from north-central Florida to approximately New York City.

It is important to note that the present study is limited just to lightning information. In practice, multiple sources of information are used operationally, including satellite, radar, and numerical weather prediction data. The combined effects of these, with or without lightning data, on the ability to identify storms that are likely to become severe, are discussed by Cintineo et al. (2014) and Bedka et al. (2015). This also includes the possibility of extending the lead times of severe weather warnings via the use of combined data sets. Such improvements fall outside the scope of this paper, given its exclusive use of lightning data and comparisons with prior literature where lightning was the sole, or at least major, focus.

2. Methods

2.1 General

All data in this study are taken from the U.S. National Lightning Detection Network (NLDN) and include all CG strokes and cloud pulses. All case studies are from 2015 and 2016, so that they include the roughly 50% cloud flash DE described by Murphy and Nag (2015). The individual CG strokes and cloud pulses are grouped into flashes using the algorithm described in Murphy and Nag (2015). Data from both the flash level and the individual strokes / cloud pulses are used. Any flash that contains at least one CG stroke is defined as a CG flash, regardless of how many cloud pulses it also contains. Any flash that includes only cloud pulses is defined here as a *pure IC* flash.

In contrast to space-borne optical sensors, ground-based networks such as the NLDN have lower total lightning DE due principally to their more limited cloud lightning DE. After a full upgrade in 2013, the cloud flash DE of the NLDN is on the order of 50% (Murphy and Nag, 2015). The cloud-to-ground (CG) flash DE of the NLDN is at least 95% over the continental U.S. following the 2013 upgrade. The location accuracy of CG strokes is also very good, with a median value of 150-250 m (Nag et al. 2014). Importantly, the NLDN and other ground-based lightning locating systems also provide the polarity and estimated peak current of every CG stroke and cloud pulse, and the NLDN is able to distinguish between CG strokes and cloud pulses with an overall classification accuracy of 90% or better.

2.2 Thunderstorm tracking

Thunderstorms were identified using an algorithm described in Murphy (2016). This algorithm assigns lightning flashes to storm objects on a quasi-interrupt-driven basis with short, configurable update intervals, nominally set to 60 seconds. At each update interval, new lightning flashes are "attracted" to nearby thunderstorm objects on the basis of the number of flashes already in those objects and the distance of the object centroids to each new lightning flash. That distance is subject to a maximum limit, such that any new flashes that are not sufficiently close to any existing storm object become the initial flashes in new storm objects. Older lightning flashes gradually age out of thunderstorm objects via a smooth, age-dependent weighting function. The same age-dependent weighting function is also used to determine the influence of each flash on the centroid position of the storm object itself, in order to produce storm tracks that are as smooth as possible. Ultimately, in this study, the positions of storm object centroids and the various information about their lightning contents are made available every two minutes, to be consistent with the use of the 2σ lightning jump algorithm in Schultz et al. (2009, 2011) and Gatlin and Goodman (2010). The lightning information that is provided at these two-minute update intervals includes the total flash count, the numbers of positive and negative CG flashes, the numbers of positive and negative pure IC flashes, the numbers of positive and negative CG strokes, the numbers of positive and negative cloud pulses, and the numbers of positive and negative cloud pulses that were produced exclusively by pure IC flashes. Various ratios of these quantities are also calculated as needed.

The above-mentioned storm tracking algorithm also picks up small thunderstorms that produce very little lightning and have no realistic probability of association with SSRs, and occasionally, it also picks up a small subdivision of a larger storm. Thus, we initially filter the output from the storm tracking algorithm to require that the storm have a total lifetime of at least 20 minutes and that it produce at least 20 flashes over its lifetime.

2.3 Matching severe storm reports to thunderstorms

Severe Storm Reports (SSRs) are taken from the daily filtered CSV files that are compiled and maintained by the NOAA Storm Prediction Center (SPC). In the U.S., "severe storms" are defined as those that produce winds of 93 km/hr or greater, hail of diameter 2.54 cm or greater, or a tornado, or any combination of those three. Much has been written about the use and limitations of SSRs (e.g. Carey and Rutledge 2003 and references therein; Witt et al., 1998; Trapp et al. 2006). The National Weather Service's principal interest in SSRs is to verify severe weather warnings, and to this end, a single observation of a severe weather event is sufficient to verify a single warning. If multiple thunderstorms pass nearly simultaneously over an area that has a severe weather warning, there may or may not be an SSR that matches each storm, even though more than one of the storms may have generated a severe event. Multiple SSRs per warning are not precluded in the SSR database, but they are not necessarily actively sought out either; some National Weather Service forecast offices have been shown by Weiss et al. (2002) to be more active than others at recording SSRs, leading to geographic inhomogeneity in the SSR database. Despite the limitations, the SPC database of SSRs is the most comprehensive source of information about severe weather over the U.S., and it is used in this study.

To associate each SSR with a thunderstorm track, if possible, we start by looking at all thunderstorm track points that are within 50 km of the SSR and that occurred at or before the time of the SSR. The motivation behind the "at or before" time matching is two-fold: (1) An observer may report a severe event late. Prior studies of SSRs mentioned by Witt et al. (1998) noted a late bias in tornado reports, and one may reasonably assume that this extends to hail and wind SSRs as well, given that any event necessarily has to occur before it can be reported. (2) Developments within the thunderstorm that lead to severe weather occur primarily aloft and usually happen before the severe weather event manifests itself on the ground. To name just one example, a hailstone that just reaches the "severe" limit of 2.54 cm diameter requires approximately 14 minutes to fall from an altitude of 10 km to sea level based simply on its near-sea level terminal velocity. The electrification process and its evolution likewise occur aloft, and thus it is reasonable to assume some lead time between changes in lightning activity and severe weather at ground level.

Once the foregoing matching step is complete, if multiple storm tracks are found within the spatial constraint of 50 km just mentioned, then we prefer to take the closest track point, provided that the SSR is in the down-shear direction with respect to the storm track. That down-shear condition is also consistent with the "at or before" time matching condition: It attempts to take into account the natural "lead distance" associated with the lead time between the evolution of the thunderstorm and electrification process aloft and the manifestation of severe weather on the ground. The down-shear condition is measured by taking the dot product between two unit vectors: (a) one connecting a candidate matching time point on the storm track to the SSR, and (b) the storm's velocity vector at the same time point. If that dot product is 0.5 or greater, then the SSR is down-shear of the candidate storm track time point to within 60 degrees, and this is regarded as a preferable match. The 60-degree restriction is waived if the SSR is within 10 km of its potential associated track point from the storm. Figure 1 shows an example of several matches, some including the 60-degree requirement and some in which it was waived, in a long-track storm in eastern Iowa (the Mississippi River is seen toward the right-hand side of the image). The storm track is given by the circle symbols (our internal storm ID number is also plotted alongside but is difficult to read and not relevant here). The SSRs are shown by the triangle symbols and are connected to the best-match upstream storm track point by thin lines. The color coding indicates the progression of time along the storm track and is not relevant here. Although there are track points at various points in the storm's life cycle that are physically closer to some of the SSRs, particularly the first, that SSR is assigned to an up-shear track point that is considered to be more representative of the evolution of the storm aloft that ultimately led to the severe weather event itself.

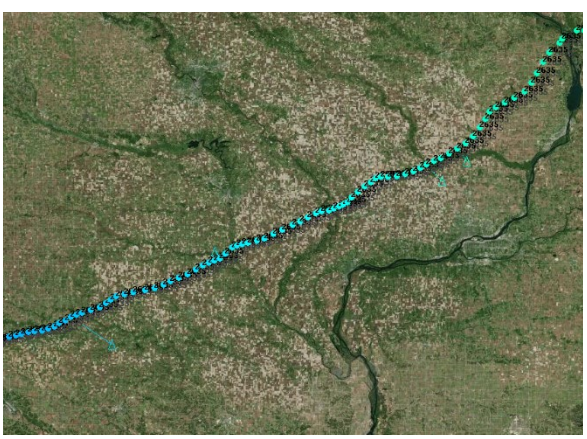

Figure 1. Association of SSRs (triangles) with a storm track based on lightning data (circles).

Given the previously mentioned limitations of SSRs, there are times when only a single SSR is matched to a particular storm. In order to focus on storms that are truly severe, we require at least two SSR matches to consider a storm to be in the "severe" population. Any storm with 0-1 SSR matches is taken as part of the non-severe storm population. The result is that 217 of our 3350 storms, or about 6.5%, are considered "severe", and the remaining 93.5% are regarded as non-severe. In section 3 (Lightning Jump Results) below, we also present the results when 1 or more SSR is allowed to define a storm as "severe".

2.4 Selected case studies
Each of the case studies crosses 00:00 UTC, as expected given the optimal time of thunderstorm activity in North America. The specific case study dates and the populations of non-severe and severe storms are given in Table 1.

Table 1. Summary of case studies

Dates	Total storms	Severe	Non-severe
2015-03-31/04-01	180	33	147
2015-04-08/09	415	44	371
2015-04-09/10	477	60	417
2015-04-24/25	386	22	364
2015-05-31/06-01	362	13	349
2015-06-01/02	540	18	522
2015-06-02/03	402	17	385
2016-03-08/09	588	10	578
Total	**3350**	**217**	**3133**

2.5 Lightning jumps and verification
The lightning jump analysis in this study uses the same configuration of the 2σ lightning jump algorithm as Schultz et al. (2011). This includes

the minimum flash rate of 10 flashes/minute and the convention of consolidating those lightning jumps that are separated by 6 minutes or less. Although the NLDN has approximately 50% cloud flash DE, we do not alter the lightning jump configuration on the basis of DE. If anything, the use of the original lightning jump definition on a data set that has lower total lightning DE ought to be overly restrictive and result in fewer lightning jumps than might be expected if the data set were from an LMA or other total lightning mapping system.

Our verification is somewhat stricter than that of Schultz et al. (2011) in that we start a 30-minute warning window, rather than 45 minutes, when a lightning jump starts. However, we have the capability of changing that time window to 45 minutes or any other value (see section 3, Lightning Jump Results). We likewise adopt the convention of Schultz et al. (2011) of counting as a "hit" only the earlier lightning jump if more than one jump occurs within 30 minutes prior to a severe weather event. Likewise, in the false alarm count, we allow the option of either including or excluding those jumps that are "false alarms" simply by virtue of the fact that a prior jump initiated a warning at an earlier time.

We use the standard contingency table to verify, but rather than computing the critical success index (CSI), however, we prefer to compute the equitable threat score (ETS) instead. The latter statistic is generally considered to be more robust in verifying rare events, such as severe weather, than CSI, but it also requires an estimate of the correct negative count in the contingency table. Our method of estimating this is described next.

The estimation of the correct negative count, to round out the contingency table and thus enable the calculation of ETS, first involves defining an "event time window". In the case of a lightning jump-initiated warning, the event time window is the 30-minute time window that begins with the jump itself. In the case of an SSR, the "event" only has one time stamp as given in the SPC database. However, in reality, almost all severe weather occurrences have some time duration, and the receipt of an actual report constitutes only one point in time in the real lifetime of the severe weather occurrence. In the absence of actual data on this, we open a 30-minute window to correspond to each SSR as well. However, to be consistent with the uncertainty in SSR times and to be consistent with trying to focus on what occurs aloft prior to the manifestation of severe weather on the ground, we define the "SSR time window" to start 20 minutes prior to the SSR time stamp and end 10 minutes after. Having thus defined what constitutes "events", both jumps and SSRs, we then simply count the total number of events and the total time spent in events. In the case where multiple events overlap in time, e.g. multiple lightning jump time windows overlap and/or at least one lightning jump window overlaps an observed SSR window, we do not over-count the time. That is, each two-minute time interval that contains one or more lightning jump and/or SSR events is simply counted once as a time interval that belongs to the "event" population. With the total number of events and the total time spent in events thus defined, we can determine the average time per event. Next, the total number of two-minute time points that are not associated with any "event" is counted as the "non-event" time. (Of course, in doing this, we take into account whether or not the non-severe storm population includes storms that have only one SSR match – hence, some SSR matches can end up in the "non-event" count.) Lastly, the total "non-event" time is divided by the average time per event to arrive at a count of "non-events", which becomes the correct negative count in the contingency table.

2.6 Down-sampling of false alarms

In the literature review in section 1.1 and our description of our own storm sample, we have deliberately made note of the portion of non-severe storms among the total storm sample. We regard this as a critical piece of information and one that must be normalized properly between our study and the prior literature in order to make proper sense of the results. In addition to presenting results from our full sample of 3350 storms, we have also down-sampled both the non-severe storm count and a best estimate of the number of false alarm jumps associated with non-severe storms to match each of the studies cited in section 1.1 except Rudlosky and Fuelberg (2013) because they did not provide contingency table statistics.

The first step in the down-sampling process, reducing the number of non-severe storms to match the proportions of non-severe storms from the prior studies cited above, is straightforward: To do this, we simply fix the number of severe storms at 217, the number in the current study, and compute a number of non-severe storms, *N*, to match any desired proportion, *p*, of non-severe storms, according to the following:

$$N = S\frac{p}{1-p}$$

where *S* is the number of severe storms (217).

The second step, scaling the number of jumps associated with non-severe storms, is admittedly subjective because that information is rarely if ever actually reported in the literature. To estimate it, however, we note that our sample of 3133 non-severe storms included 1159 storms that had one or more lightning jumps, with a total of 1858 jumps. In other words, 37% of our non-severe storms had at least one jump, but an average of 1.6 jumps occurred per non-severe storm that had jumps (that is, 1858 jumps / 1159 storms). With this information, it is possible to estimate the number of non-severe storms that would have had jumps after we down-sample the population of non-severe storms, and to estimate the number of jumps that would have been associated with those. That information, at last, permits us to normalize our results to the particular proportions of non-severe storms reported in each of the prior studies cited above.

2.7 Best use of ground-based lightning network data

Lastly, we have the question of using additional information derivable from ground-based network data to improve upon the performance of the lightning jump method alone. We have examined a large number of such enhancements, and in addition, we have also examined how each of those would perform on its own if lightning jumps were not taken into consideration at all. In the end, we found that the highest ETS value was obtained by filtering lightning jumps with a set of three non-jump criteria:

(a) total flash rate is at least 30 per 2-min interval and the pure IC flash fraction is at least 0.5 and greater than or equal to a threshold that decreases linearly with total flash rate, or
(b) the rate of negative-polarity IC pulses due to pure IC flashes is greater than or equal to a second threshold that also decreases linearly with increasing total flash rate, or
(c) the total flash rate is at least 55 per 2-minute interval

Note that the latter criterion is based on total flash rate itself, not the rate of change of total flash rate, which is the basis of the lightning jump.

The motives behind the choice of these three criteria are as follows

(a) in severe storms, cloud flashes are expected to dominate over CG flashes more so than in ordinary storms, and therefore, the fraction of pure IC flashes among total flashes should be high around the time when storms become severe
(b) due to turbulent disruption of the normal, relatively orderly, layered charge structure of ordinary storms (Bruning and MacGorman, 2013), and possibly due to the rapid rearrangement of charge by the high flash rate itself, severe storms might be expected to have a disorderly arrangement of charge, with small pockets of positive and negative charge rather than more orderly layers of charge. In ordinary convection, with a mid-level negative and upper-level positive charge layer, the majority of cloud pulses in pure IC flashes should be of positive polarity. However, in a disorderly charge structure, a larger proportion of cloud pulses in pure IC flashes is expected to be negative. Note that this does not apply to the cloud pulses associated with CG flashes; many, if not most, of those are due to the preliminary breakdown process and have the same polarity as the return strokes, which may not be relevant to the question of storm severity
(c) overall, the total flash rate is generally high in severe storms

3. Lightning Jump Results

The contingency table from our sample of 3350 storms, with the Schultz et al. (2011) lightning jump configuration applied to them, is presented in Table 2A, along with the probability of detection (POD), false alarm ratio (FAR), and equitable threat score (ETS) derived from it. As noted in the Methods section, we opened a 30-minute warning time window, as in Gatlin and Goodman (2010), when a lightning jump occurred. A total of 743 SSRs occurred in the 217 storms in our severe storm population. Almost half of those had a lightning jump in the 30 minutes leading up to the SSR time, giving a POD of 0.499. There were 2144 false alarm lightning jumps. Of these, 1858 (86.7%) occurred in the set of non-severe storms, and 286 (13.3%) occurred in the severe storm population. Of the 2144 false alarms, 444 were "false alarms" due to the fact that a warning was already in effect due to a prior jump (see Schultz et al. (2011) verification method), and therefore, our FAR remains roughly the same even if we take those out, as Schultz et al. did: 0.821 vs 0.852.

Table 2A. Contingency table, standard 2σ lightning jump algorithm, all 3350 storms, using a 30-minute warning time window.

		FORECASTS	
		YES	NO
OBS	YES	371	372
	NO	2144	15329
		POD	**0.499**
		FAR	**0.852**
		ETS	**0.096**

Our POD of 0.499 should theoretically be closest to that of Gatlin and Goodman (2010) because we both opened 30-minute warning time windows with the lightning jumps, but Gatlin and Goodman actually had a substantially higher POD of 0.74. The fact that our POD is much lower might be due simply to the much greater geographic diversity of storms and variety of severe storm conditions that we sampled in this study. The remaining studies cited in section 1 are expected to have rather different POD values because they either used different warning time windows or no time window that was explicitly called out. Specifically, Schultz et al. (2011) and Miller et al. (2015) both applied 45-minute windows, whereas Metzger (2010) and Farnell et al. (2016) did not specifically mention any time window. Tables 2B and 2C below show the results that we get from our storm sample if we change the warning time window to 45 minutes (Tab. 2B) or effectively remove it completely by raising it to 120 minutes (Tab. 2C).

Table 2B. Contingency table, standard 2σ lightning jump algorithm, all 3350 storms, using a 45-minute warning time window.

		FORECASTS	
		YES	NO
OBS	YES	445	298
	NO	2125	10770
		POD	**0.599**
		FAR	**0.827**
		ETS	**0.112**

Table 2C. Contingency table, standard 2σ lightning jump algorithm, all 3350 storms, using a 120-minute warning time window.

		FORECASTS	
		YES	NO
OBS	YES	531	212
	NO	2158	5804
		POD	**0.715**
		FAR	**0.803**
		ETS	**0.113**

Tables 2A-2C show that, as the warning time window is gradually raised, the POD progressively increases. The number of correct negatives progressively decreases because the total time per event increases (see subsection 2.5 of Methods). The number of false alarms, however, remains relatively constant, decreasing slightly at the 45-minute time window and then increasing again slightly with the 120-minute window. Mostly, this is because the 1858 false alarms that occurred in non-severe storms continue to be false alarms regardless of the warning time window. The remaining, and much smaller, portion of false alarms in the severe storm population redistribute themselves somewhat as the warning time window changes. The initial change of that window from 30 to 45 minutes causes some false alarms to become hits. The latter increase of that window, however, results in a bigger group of lightning jumps that occur while a warning window due to an earlier jump is already in effect. By default, jumps that occur while a warning is already in effect are counted as false alarms, and thus, the count of false alarms increases again, unless the repeat warning cases are removed. Even when repeat warnings are removed, the net effect is still that the FAR is dominated by the false alarms that occurred in non-severe storms, and the FAR decreases relatively little as the warning time window is increased: When repeat-warning “false alarms” are removed, the FARs are 0.774 with the 45-minute warning window and 0.710 with the 120-minute window.

If we redefine a “severe” storm as having any SSR, rather than a minimum of two, then the number of false alarms due to non-severe storms drops from 1858 to 1496, while the number of hits increases from 371 (Table 2A) to 482. The FAR, therefore, becomes 0.808. However, of the false alarms, only 364 are due to repeat warnings. Thus, the FAR is still dominated by jumps occurring primarily in the non-severe storm population, and the exclusion of repeat-warning “false alarms” changes the FAR relatively little, to 0.776.

Thus, the preceding results indicate that *the representation of non-severe storms in the sample is critical*. All of the prior literature cited in section 1, with the exception of Miller et al. (2015), had a substantially lower proportion of non-severe storms than the 93.5% in our sample of 3350 storms or the 90% proportion indicated by the National Weather Service. Therefore, as described in section 2.6, we have attempted to down-sample the non-severe storm population, and the corresponding number of jumps due to non-severe storms, from our sample to match the proportions reported in the prior literature. In this analysis, we ignore the small number of repeat warnings (41 when the warning time window is 30 minutes) because, as already shown above, they have little effect on the FAR value. Table 3 shows the results of this down-sampling, and the discussion of this starts in the next paragraph.

Table 3. False alarm ratio (FAR) down-sampled from our data set and compared with original FARs

Reference	**Warning time window (min)**	**Hits based on our sample**	**Proportion non-svr storms**	**Non-svr stms based on our sample**	**Non-svr FA jumps based on our sample**	**FAR as reported**	**FAR down-sampled, our data**
Gatlin	30	371	0.050	11	6	**0.40**	**0.44**
Metzger	not reported	531	0.265	78	46	**0.47**	**0.40**
Farnell	not reported	531	0.400	145	87	**0.10**	**0.42**
Schultz	45	445	0.641	387	229	**0.36**	**0.53**
Miller	45	445	0.887	1703	1010	**0.88**	**0.74**

As discussed in section 2.6, the FAR down-sampling effort is based on down-sampling the number of non-severe storms from our sample, and the corresponding number of lightning jumps due to non-severe storms. We keep our original population of severe storms and its associated lightning jumps the same. The distribution of those jumps between hits and false alarms depends on the warning time window, which is why that value, and the corresponding number of hits, are given in the second and third columns of Table 3. The middle three columns of Table 3 show (a) the proportion of non-severe storms in the original literature citation, (b) the number of non-severe storms that our sample would have in order to match that proportion, and (c) the number of false alarm lightning jumps that our sample would have, given the new number of non-severe storms. The final two columns of Table 3 show the FAR as originally reported in the literature and the FAR of our storm sample following the down-sampling.

The down-sampled FARs from our data sample are within about ±0.15 of the originally reported FARs in all cases except Farnell et al. (2016). Those authors reported that they used radar information to supplement ground-based severe storm reports in storms that occurred over sparsely-populated and/or mountainous areas. A more significant discrepancy might be expected with respect to the Metzger (2010) study, given that his definition of a lightning jump was rather different from that of Schultz et al. (2011), which is also used here. Aside from that, discrepancies are also to be expected because (a) we require at least 2 SSRs to define a “severe” storm, although this turns out to have a minor effect on FAR as described above, (b) the NLDN has lower cloud flash DE than LMAs, also as discussed above, and/or (c) we have a totally different sample of storms, with very different diversity in terms of both geographic and storm type considerations. Given all that, however, the fact that the down-sampled and original FAR values are almost all within about ±0.15 is remarkable and indicative of the critical importance of the representation of non-severe storms in the data sample.

4. Filtered Lightning Jump Results

The foregoing results indicate that the false alarm ratio is the primary issue in the use of lightning jumps to indicate which storms are likely to become severe. Our next objective, then, is to see if there is a way to augment the lightning jumps with other information available from ground-based lightning networks in such a way as to reduce the FAR while, ideally, not affecting POD and overall skill too severely. After examining relationships between many non-jump lightning characteristics and our storm sample, in a multitude of 1-D and 2-D histograms, we settled on a combination of the three values described in section 2.7. Namely, those characteristics are (1) the pure IC flash fraction, (2) the rate of negative IC pulses due to pure IC flashes, and (3) the total flash rate, with the first two criteria having thresholds that are inversely dependent upon the total flash rate.

We apply this augmented set of criteria as a filter on the basic lightning jumps. Specifically, we filter out any lightning jump that does not also pass the threshold of one or both of the first two above-named criteria, subject to thresholding on total flash rate. The augmented criteria threshold(s) must be passed within ±15 minutes of the time of the lightning jump. The time of the warning becomes the later of the two times, either the lightning jump or the threshold(s) based on the augmented criteria. This way, we require that both the lightning jump and the filter criteria threshold be satisfied before we initiate the warning time window. Then, we open a 30-minute warning time window and verify in the same way as described in section 2.5 above.

With that, the contingency table and corresponding performance statistics from the filtered lightning jump analysis are presented in Table 4. This table is comparable to Table 2A, where the warning time window was also set to 30 minutes. We find that the raw number of false alarms is lower, by a factor of 2.6, when we filter the lightning jumps vs. when we use all lightning jumps. Predictably, however, the number of hits is also lower, but only by a factor of 1.4. Thus, the net decrease in false alarm events provides a small, but noticeable, drop in FAR of around 0.09. The equitable threat score is therefore substantially higher under the filtered lightning jump method (0.146) than it was with the unfiltered lightning jumps (0.096). Obviously, however, neither value of ETS is truly impressive, indicating that, ultimately, it appears to be difficult to use lightning data alone to discern storms that are likely to become severe from all other storms in an overall population of storms that is representative of the real-world thunderstorm situations faced by operational forecasters.

Table 4. Contingency table after filtering lightning jumps using the augmented set of criteria described in section 2.7.

		FORECASTS	
		YES	**NO**
OBS	**YES**	265	478
	NO	831	18583
		POD	**0.357**
		FAR	**0.758**
		ETS	**0.146**

5. Conclusions and Future Work

In this study, we have taken a large and geographically diverse sample of 3350 thunderstorms, in which 93.5% were non-severe storms, and attempted to reproduce the results of the lightning jump analysis from prior literature and then augment that with additional information derivable from a ground-based lightning location system. We find that the skill (that is, ETS) of a lightning jump algorithm can be improved by filtering the jumps using other, non-jump lightning characteristics. However, the ETS in both the filtered and unfiltered cases is quite low and probably not operationally useful.

Crucially, we find that we are more or less able to reproduce the FAR values from prior studies by down-sampling the number of non-severe storms from this study, and the number of lightning jumps occurring in those non-severe storms. This suggests the criticality of a representative sample of non-severe storms to the overall conclusions. This is one factor that was not raised by Miller et al. (2015).

By contrast, Miller et al. (2015) did raise two other considerations that we have not yet taken into account in this study: (1) the use of lightning data to identify and track thunderstorms, and (2) the non-LMA source of the total lightning information. The former issue affects how lightning flashes are assigned to thunderstorm cells, and thus might impact the determination of lightning jumps. The latter factor could also affect the determination of lightning jumps via the total lightning detection efficiency, the manner in which IC pulses and CG strokes are grouped into flashes, or some combination of the two. In the present study, we suspect that the use of NLDN, with its approximately 50% DE of pure IC flashes, should tend to produce fewer lightning jumps than LMA data, but it would be worth confirming that suspicion by comparing LMA and NLDN lightning jumps derived from a single set of thunderstorms that are tracked by the same algorithm.

Theoretically, it might be argued that the present study is too diverse in its sampling of geography and storm type. The Miller et al. (2015) study, however, provides some counter-argument to that, insofar as they studied just a single NWS county warning area and they also separated the data sample by convective mode, and the FAR was still 90% even in cases identified as part of the multi-cell / supercell convective mode. It is well worth considering whether some kind of *a priori* knowledge should be taken into account in attempts to filter lightning jumps. However, the implication of the results of Miller et al. and the current study, in the context of combined data sets as used in Bedka et al. (2015) and Cintineo et al. (2014), is that the *a priori* knowledge should be from non-lightning sources. One possibility is sounding information, specifically MUCAPE, CIN, wet-bulb zero altitude, shear, and helicity. We have made preliminary efforts in the direction of including key sounding parameters in the classification or filtering of lightning jumps but have no concrete results to present yet.

The finding that non-jump lightning characteristics, including the pure IC flash fraction, still do not raise the ETS to a level that might be considered “skillful”, is broadly consistent with a recent climatological study of the IC flash fraction over the U.S. by Medici et al. (2017, submitted). Their study finds areas of the U.S. that have similar climatological values of the pure IC flash fraction yet very different thunderstorm characteristics and tendencies toward severe weather, most notably, the interior of the northwestern U.S. vs. the high plains of the central U.S. They suggest, as a result, that high pure IC flash fraction is not uniquely driven by storm intensity or severity, nor by overall thunderstorm and lightning occurrence in general. Thus, there appears to be a clear need to try to separate those properties of lightning that are essentially climatological from those that may be specifically related to storm severity.

One potentially serious limitation in the present study is the use of SSRs, and specifically, the assignment of SSRs to thunderstorm tracks. A unique and different approach to identifying hazardous thunderstorms in Finland is given by Rossi et al. (2013). Their method links emergency call information with storm tracking information and takes population density into account to correct the anticipated lack of emergency calls in sparsely populated areas. Conceptually, this idea is appealing in the context of the present study due to the wide geographic diversity of storms used here. An initial effort to apply the general concept of the Rossi et al. study to this study has thus far produced inconclusive results, but additional effort in that direction is warranted.

6. References

Bedka, K.M., C. Wang, R. Rogers, L.D. Carey, W. Feltz, and J. Kanak, 2015: Examining deep convective cloud evolution using total lightning, WSR-88D, and GOES-14 super rapid scan datasets, *Wea. and Fcstg.*, **30**, 571-590.

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

Carey, L.D. and S.A. Rutledge, 2003: Characteristics of cloud-to-ground lightning in severe and nonsevere storms over the central United States from 1989-1998, *J. Geophys. Res.*, **108**, doi:10.1029/2002JD002951.

Cintineo, J.L., M.J. Pavlonis, J.M. Sieglaff, and D.T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection, *Wea. and Fcstg.*, **29**, 639-652.

Farnell, C., T. Rigo, and N. Pineda, 2016: Lightning jump as a nowcast predictor: Application to severe events in Catalonia, *Atmos. Res.*, **183**, 130-141.

Gatlin, P.N. and S.J. Goodman, 2010: A total lightning trending algorithm to identify severe thunderstorms, *J. Atmos. Ocean. Tech.*, **27**, 3-22.

Goodman, S.J. et al., 2013: The GOES-R Geostationary Lightning Mapper (GLM), *Atmos. Res.*, **125-126**, 34-49.

Koshak, W.J. and R.J. Solakiewicz, 2015: A method for retrieving the ground flash fraction and flash type from satellite lightning mapper observations, *J. Atmos. Ocean. Tech.*, **32**, 79-96.

Medici, G., K.L. Cummins, D.J. Cecil, W.J. Koshak, and S.D. Rudlosky, 2017: The intra-cloud lightning fraction in the contiguous United States, submitted to *Mon. Wea. Rev.*

Metzger, E.L., 2010: *The relationship between total cloud lightning behavior and radar-derived thunderstorm structure*, M.S. Thesis, Naval Postgraduate School, Monterey, CA, U.S.

Miller, P.W., A.W. Ellis, and S.J. Keighton, 2015: The utility of total lightning trends in diagnosing single-cell thunderstorm severity: Examples from the central Appalachians region, *J. Oper. Met.*, **3**, 92-98.

Murphy, M.J. and A. Nag, 2015: Cloud lightning performance and climatology of the U.S. based on the upgraded U.S. National Lightning Detection Network, Seventh Conf. on Meteorological Applications of Lightning Data, Phoenix AZ, U.S., Amer. Meteorol. Soc.

Murphy, M.J., 2016: Short-term thunderstorm forecast and severe weather alert system and method, U.S. Provisional Application 62/368,625

Nag, A., M.J. Murphy, K.L. Cummins, A.E. Pifer, and J.A. Cramer, 2014: Recent evolution of the U.S. National Lightning Detection Network, 23rd Intl. Lightning Detection Conf., Tucson, AZ, U.S., Vaisala Inc.

Rossi, P.J., V. Hasu, K. Halmevaara, A. Mäkelä, J. Koistinen, and H. Pohjola, 2013: Real-time hazard approximation of long-lasting convective storms using emergency data, *J. Atmos. Ocean. Tech.*, **30**, 538-555.

Rudlosky, S.D. and H.E. Fuelberg, 2013: Documenting storm severity in the mid-Atlantic region using lightning and radar information, *Mon. Wea. Rev.*, **141**, 3186-3202.

Schultz, C.J., W.A. Petersen, and L.D. Carey, 2011: Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends, *Wea. and Fcstg.*, **26**, 744-755.

Schultz, C.J., W.A. Petersen, and L.D. Carey, 2009: Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather, *J. Appl. Met. and Clim.*, **48**, 2543-2563.

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**, doi:10.1029/2004JD004549.

Trapp, R.J., D.M. Wheatly, N.T. Atkins and R.W. Przybylinski, 2006: Buyer beware: Some words of caution on the use of severe wind reports in post-event assessment and research, *Wea. and Fcstg.*, **21**, 408-415.

Weiss, S.J., J.A. Hart and P.R. Janish, 2002: An examination of severe thunderstorm wind report climatology: 1970-1999, Preprints, 21st Conf. on Severe Local Storms, San Antonio, TX, U.S., Amer. Meteorol. Soc., 446-449.

Williams, E. et al., 1999, The behavior of total lightning activity in severe Florida thunderstorms, *Atmos. Res.*, **51**, 245-265.

Witt, A., M.D. Eilts, G.J. Stumpf, E.D. Mitchell, J.T. Johnson and K.W. Thomas, 1998: Evaluating the performance of WSR-88D severe storm detection algorithms, *Wea. and Fcstg.*, **13**, 513-518.