2006
19th International Lightning Detection Conference
24-25 April • Tucson, Arizona, USA
1st International Lightning Meteorology Conference
26-27 April • Tucson, Arizona, USA

WHEN FLASH ALGORITHMS GO BAD

Martin J. Murphy*
Vaisala Inc.
Tucson, Arizona

1. INTRODUCTION

Flash algorithms are an important component of data reduction for VHF total lightning mapping systems that detect large quantities of VHF radiation sources (or “events”) for each lightning flash. A flash algorithm attempts to determine which events belong to the same flash and then provide some reduced set of information about the VHF data. An example of a reduced information set is a flash rate trend over the life of a cell or storm.

A variety of flash algorithms exists for VHF time-of-arrival (TOA) lightning mapping systems. All of them basically involve simple distance and time criteria (Thomas et al. 2003, Wiens et al. 2002; Murphy and Cummins, 2000; Williams et al. 1999). That is, if events are closer together than X km in space and Y seconds in time, they are considered part of the same flash. Some algorithms have additional criteria, such as a cap on the total duration of a flash. Algorithms may also vary in terms of how the criteria are applied (for example, whether the distances are in two dimensions or three). The first operational algorithm, developed for use in the original Lightning Detection and Ranging (LDAR) system at the NASA Kennedy Space Center (KSC), had a two-part method that aggregated events into branches (lightning channel segments) and then grouped branches into flashes. The distance and time criteria used in these flash algorithms, together with the specific details of other constraints, determine how many flashes are generated.

There is general, anecdotal agreement that a wide range of parameter values and criteria will produce the correct flash count in situations with a low rate of discrete flashes. However, in complicated storm situations with high lightning rate, it eventually becomes difficult to identify discrete flashes, and different algorithms with different parameter values will produce very different flash counts. Little quantitative analysis of this “uncertainty” in flash algorithms has been performed to date. To our knowledge, only McCormick (2003) has attempted to quantify the variability among algorithms in high lightning-rate storms. She compared several combinations of time, distance, and duration constraints with each other and with a modified version of the NASA-KSC algorithm for two large MCSs observed near Dallas-Fort Worth, Texas. The worst-case differences in flash count were a factor of 4. The modified KSC algorithm always produced the lowest flash count, and a set of parameters generally consistent with Wiens et al. (2002) produced the highest. It is helpful to keep McCormick’s results in mind when interpreting, for example, the flash rates quoted by Wiens et al. (2005), who used the same algorithm as Wiens et al. (2002) to calculate peak flash rates above 300 min^{-1} in a severe storm in the high plains during STEPS 2000. Had the modified KSC algorithm been used on the same storm, the peak flash rate might have been significantly lower (even without the additional constraint used by Wiens et al., discussed later).

McCormick (2003) only had the two MCS cases to investigate, and both produced very high lightning rates. Therefore, it was not possible to probe lower-rate storms to find the flash rate at which all algorithms converged on a single discrete flash count. This study examines a larger number of storm cases with a variety of flash rates. Our objective is to identify exactly when significant uncertainty becomes an issue for a flash algorithm by starting with low-rate storms and gradually moving to storms with higher flash rates and more complex geometry.

2. METHOD

The flash algorithm used in this study was first described by Lojou and Cummins (2005). This algorithm attempts to reconstruct the channel structure as illustrated in Figures 1a-1b, which show the individual sources associated with a large flash and the reconstructed channel segments, respectively. Like other algorithms, this one has two variable parameters, the maximum distance and maximum time between two events that may be joined together. In the event that multiple sources satisfy the distance and time criteria with respect to a new event, the new event is joined to the closest source in distance. The algorithm has no other constraints (e.g., maximum flash duration). This algorithm was designed for use with 2-D or 3-D VHF interferometer network data, but it applies equally well to VHF TOA data. All data used in this analysis are from the Vaisala LDAR II (TOA) system in the Dallas-Fort Worth area of Texas (Demetriades et al. 2002).

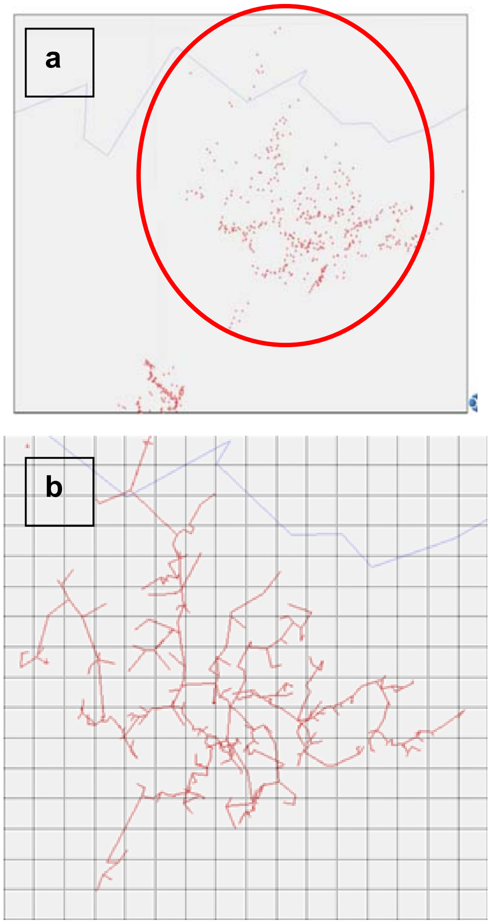

Fig. 1(a) source positions in flash (circled), (b) close-up of branch and flash reconstruction.

The strategy for the main part of this analysis is to identify cases with low rates of discrete flashes first. Two samples, from 12 July 2005, are analyzed here. One is an isolated cell, and one is a broken line of three cells. In both cases, flash rates are low enough that flashes can be counted manually. We then process the same data through the flash algorithm with 400 different combinations of the parameters. The maximum time separation varies between 0.05 and 1 second (inclusive) in steps of 0.05 sec, and the maximum distance parameter varies between 1 and 20 km (inclusive) in steps of 1 km. From this analysis, we are able to identify a subset of the parameter space over which the algorithm produced the correct number of flashes for these two low-rate situations. Finally, we apply the same subset of the parameter space to analyze more complex situations in order to quantify the flash algorithm uncertainty.

In addition to the main analysis, we also look into the effects of detection efficiency loss as a function of distance from the network center. For this task, we use the same low-rate data from 12 July 2005, and we gradually apply a threshold to cut off lower-amplitude (source power) events. Finally, we also use a simple visual examination of the raw event data in a complex storm to try to narrow the range of likely flash counts and thereby constrain the possible range of parameter values. For that task, we examine data from a complex MCS visually, 1 sec at a time, in an attempt to do approximate manual flash counts per second. These are then compared to the flash algorithm results.

3. ANALYSIS OF LOW-RATE STORMS ON 12 JULY 2005

Figures 2a and 2b show color contour plots of the flash counts generated by the algorithm as a function of the two free parameters for the isolated cell and broken line cases, respectively. The legend shows the flash count ranges applicable to each color. The correct flash

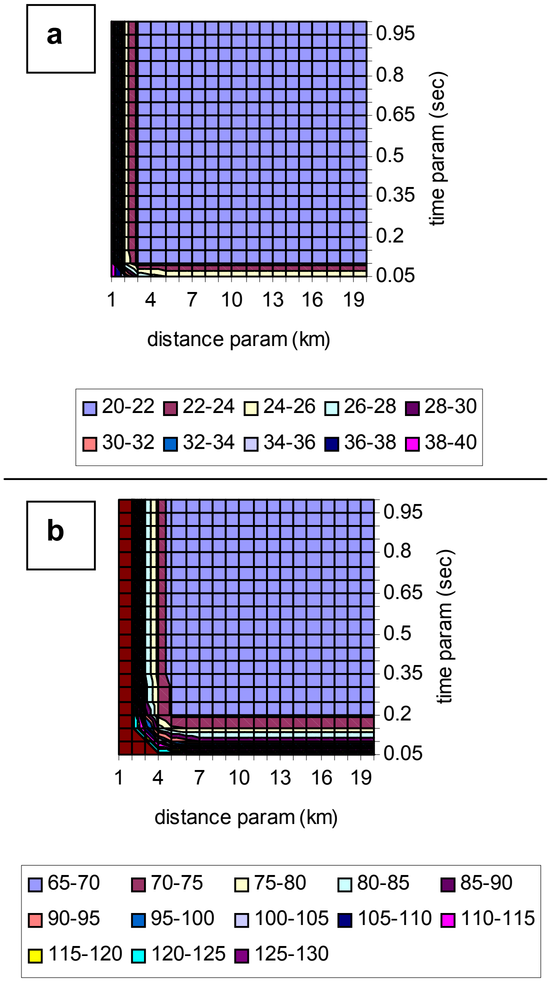

Fig. 2. Contour plot of flash counts for (a) isolated cell and (b) broken line storms on 12 July 2005.

counts in each case are represented by the broad area of uniform color seen at all but the lowest parameter values. When either the time or distance parameter gets too small, flashes are broken up into small pieces. The converse is not true when the parameters are big, at least within the range of parameter values that we use in this analysis. Note also in Figs. 2a and 2b that the broken line of storms starts to have its flashes broken apart at higher parameter values than the isolated cell. For this reason, the parameter values that correspond to correct flash counts in the broken line case set the parameter limits for the rest of the study. These limits are 0.2-1 sec for the time and 5-20 km for the distance. This subset of the original parameter space reduces the number of parameter combinations from 400 to 272 in the remainder of the analysis.

4. ANALYSIS OF HIGHER-RATE STORMS

We now apply the aforementioned 272 parameter value combinations to four storm situations that involved gradually more rapid and complex lightning activity. These cases, in order of complexity, are: 6 August 2005 (widespread air-mass convection), 14 June 2005 (several multicell clusters), 16 June 2002 (MCS) and 6 April 2003 (supercells). For the more complicated situations, such as the MCS on 16 June 2002, we find that the 272 flash algorithm runs result in a distribution of flash counts with a long tail. This is shown in Figure 3 for a particular two-minute period of the MCS case. Because this general distribution shape occurs frequently, we represent the uncertainty in the flash count as the spread between the median and 95th percentile flash counts. Each storm case is broken up into two-minute intervals, and one pair of values (median, 95th percentile) is produced for each interval.

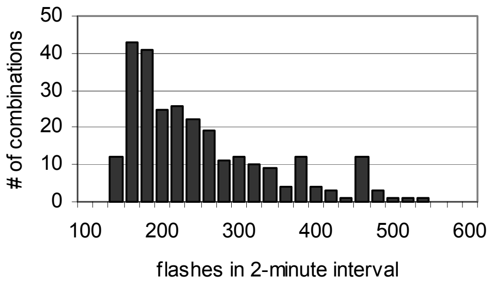

Fig. 3. Histogram of flash counts for a 2-minute interval of a large MCS. Each “count” represents one run of the flash algorithm with a particular combination of parameter values.

In Figure 4, we show the spread between the median and 95th percentile flash counts as a function of the median flash count for the collection of two-minute intervals. Each case has a different symbol and color. In Fig. 4, we find that, once the flash rate reaches 50-60 flashes (2 min)$^{-1}$, we notice a steady increase in the spread of the distribution. There is a significant jump in the uncertainty in going from the multicell cluster case on 14 June 2005 to the large MCS case of 16 June 2002.

Figure 5 shows the same information but this time expressed as a relative uncertainty, that is, (95th% - median) / median. For flash rates up to about 60 (2 min)$^{-1}$, the relative uncertainty is generally between 5-20%, but occasionally exceeds 100% for the complex MCS and supercell cases.

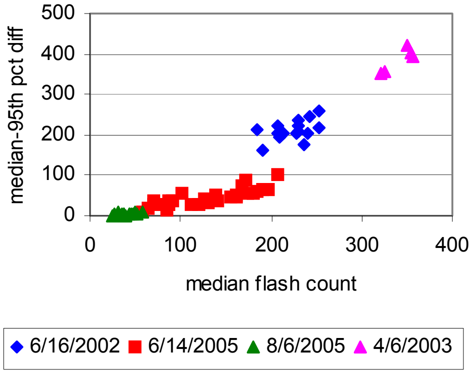

Fig. 4. The difference between the 95th percentile and median flash counts (see Fig 3) vs. the median flash count for four storm cases.

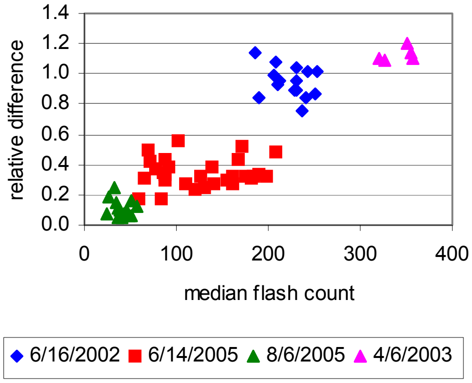

Fig. 5. Difference between 95th percentile and median flash counts (see Fig. 3) *relative* to the median flash count, plotted as a function of the median flash count.

5. EFFECTS OF DETECTION EFFICIENCY LOSS (STORMS AT LONGER DISTANCE)

Figure 6 shows the source power distribution for all events within 75 km of the approximate center of the DFW network on 12 July 2005. The boldface vertical lines show where we have applied thresholds to the distribution (every 4 dB from 30 to 46 dBm) to mimic the effects of a gradual loss of detection efficiency with increasing distance from the network. Table 1 below shows the median, minimum and maximum flash counts as a function of source power threshold for an extended part of the broken-line storm used in section 3. The minimum flash count is also by far the most probable flash count, and therefore it also corresponds to the median. The maximum flash counts are again due to the break-up of flashes at the lowest values of the algorithm parameters. Note that we do not start to lose flashes until we cut off all source powers below 38 dBm, which corresponds to about the lowest 1/6 of the source power distribution. We lose increasing numbers of flashes as we cut off more events. For a threshold of 46 dBm, which corresponds to cutting off the lower 3/4 of the sources in Fig. 6, we lose about 1/5 of the original 73 flashes. These results have implications for detection efficiency modeling of VHF TOA networks. Specifically, although most flashes are retained even when we cut off 3/4 of the sources, it is not a safe assumption to say that all flashes sample randomly from the entire distribution of source powers. However, this sampling problem is relatively unimportant at distances closer than about 150 km, where most of the source power distribution is still sampled.

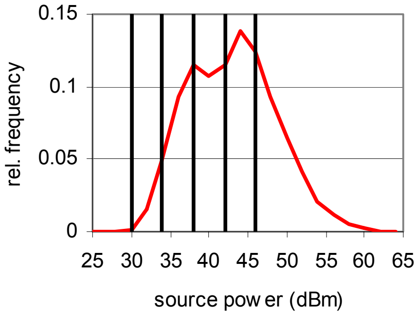

Fig. 6. Distribution of source powers for events within 75 km of DFW on 12 July 2005.

Table 1. Analysis of broken-line storm on 12 July 2005 using different thresholds on the source power of the events. Minimum, median, and maximum flash counts are given.

Threshold	min flash count	median flash ct	max flash count
NONE	73	73	79
30 dBm	73	73	79
34 dBm	73	73	78
38 dBm	71	71	75
42 dBm	67	67	76
46 dBm	58	58	65

6. NARROWING THE RANGE OF REALISTIC FLASH COUNTS IN COMPLEX STORMS

As we showed previously in Fig. 3, the algorithm produces a wide range of flash counts in complex storm situations as a function of its parameters. In an attempt to narrow the range of realistic flash counts (and therefore, realistic parameter value sets), we have analyzed a number of time periods from the 16 June 2002 MCS visually. We look at plots of all events in 1-second intervals, and the data stay on the plots for 5 seconds before disappearing. The data from the most recent second are shown in a different color from the older data. This is shown in Figures 7a-b, which show two consecutive 1-second intervals. We count flashes manually by considering any cluster of activity in the most recent second as a new flash as long as it is separated by at least a couple of km from any other activity, and as long as it is not clearly associated with a flash that was counted in the previous second. In this way, we build up a distribution of the rate of new flash activity per second over the whole MCS (convective and stratiform regions). We repeat this every second for a number of one-minute periods during the MCS. Though subjective, this analysis gives us an idea of the distribution of flashes per second produced by the MCS. This is shown in Figure 8. Note that the MCS most often exhibited 1-3 new flashes in any given second, and rarely did it produce more than 3. In terms of the 2-minute flash counts shown in Figures 3-5, most of these values would be in the range of 120-360, and rarely would the value exceed 360. Thus, the values at the lower end of the distribution shown in Fig. 3 are more likely to be representative. These values, in turn, can be produced when the time parameter is as low as 0.2 sec, but the distance parameter must be 7-8 km or greater. It appears that there is greater sensitivity to the distance parameter, at least in this case.

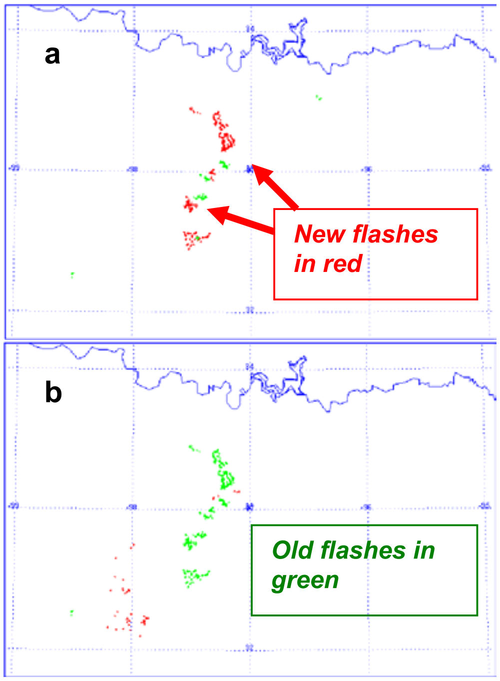

Fig. 7. Two consecutive seconds of lightning activity in a large MCS. The most recent 1 second of activity is shown in red. Examples are pointed out. Events from 1-5 seconds old are shown in yellow, and after 5 seconds, events are removed from the plot.

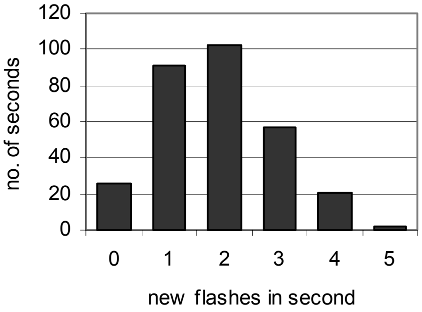

Fig. 8. Distribution of the number of new flashes in each second of a manual analysis of several minutes of lightning in an MCS. See text for further detail.

7. CONCLUSIONS

This analysis has shown that a flash algorithm begins to have significant uncertainty (that is, provide very different flash counts depending on its parameters) at flash rates that are quite low relative to some of the flash rates that have been quoted in the literature. Flash rates larger than about 30 flashes per minute are susceptible to significant uncertainty. When the algorithm produces a wide range of flash counts, we have found that the lower end of the distribution generally appears to be closer to reality based on visual inspection.

The algorithm used in this study tends to break flashes apart when its parameters are too small but does not show any indication of the opposite problem (combining flashes) when the parameter values are big. It is interesting to note that Wiens et al. (2005) found that their parameter values (0.15 sec for the time and 3 km for the distance) produced very large numbers of flashes with fewer than 10 sources, and they imposed a 10-source minimum on flashes in their analysis. The results presented in this paper, showing that even discrete flashes can be broken apart with such low parameter values, raise the question of whether the many small flashes found by Wiens et al. (2005) were actually small pieces of flashes. From a certain point of view, this issue might be regarded as academic, given that Wiens et al. also pointed out that the trends in total flash rate over the life of the storm were basically the same whether they looked at all flashes or only those with at least 10, 50, or 100 sources. However, to the degree that discrete flashes might have meaning beyond what they contribute to a flash rate time trend, this issue might be worth further analysis.

Our original intention in this paper was to find a flash rate at which the uncertainty jumps significantly and thus be able to label that point as a “saturation” flash rate beyond which flash counting is no longer meaningful. However, despite the large uncertainty involved in the high flash rates, Fig. 4 shows a clear trend toward larger flash rates for the complex and severe storms. The trend is evident even if we use the low-end flash counts from the distribution of possible values (refer to Fig. 3). These results therefore suggest that high flash rates (>30 min^{-1}) do have meaning. However, we should probably accompany high flash rates with some measure of the algorithm-related uncertainty.

REFERENCES

Demetriades, N.W.S., M.J. Murphy, and K.L. Cummins, 2002: Early results from the Global Atmospherics, Inc., Dallas-Fort Worth Lightning Detection and Ranging (LDAR II) research network. 6th Conf. on Integrated Observing Systems, Amer. Meteorol. Soc., Orlando, FL, 202-209.

Lojou, J-Y. and K.L. Cummins, 2005: On the representation of two- and three-dimensional total lightning information. Proceedings, Conf. on Meteorological

Applications of Lightning Data, Amer. Meteorol. Soc., San Diego, CA, paper 2.4.

McCormick, 2003: *Three-dimensional radar and total lightning characteristics of mesoscale convective systems.* M.S. Thesis, North Carolina State Univ.

Murphy, M.J., K.L. Cummins, and L.M. Maier, 2000: The analysis and interpretation of three-dimensional lightning flash information. Preprints, 16th Intl. Conf. on IIPS for Meteorology, Oceanography, and Hydrology, Long Beach, CA, Amer. Meteorol. Soc., 102-105.

Thomas, R., P. Krehbiel, W. Rison, J. Harlin, T. Hamlin, and N. Campbell, 2003: The LMA flash algorithm. Proceedings, 12th Intl. Conf. on Atmospheric Electricity, Intl. Commission on Atmos. Electricity, Versailles, France.

Wiens, K.C., S.A. Tessendorf, and S.A. Rutledge, 2002: June 29, 2000, STEPS supercell storm: Relationships between kinematics, microphysics, and lightning. Preprints, 21st Conf. Severe Local Storms, San Antonio, TX, Amer. Meteorol. Soc., 315-318.

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. Sci.*, **62**, 4151-4177.

Williams, E.R. *et al.*, 1999: The behavior of total lightning activity in severe Florida thunderstorms. *Atmos. Res.*, **51**, 245-264.