Toward lightning mapping with long-baseline LF networks

Martin J. Murphy
Vaisala Inc.
Louisville, CO, U.S.
martin.murphy@vaisala.com

Ryan K. Said
Vaisala Inc.
Louisville, CO, U.S.
ryan.said@vaisala.com

Abstract— Recent experiences with the comparison between cloud lightning pulse (IC pulse) data from the U.S. National Lightning Detection Network (NLDN) and Lightning Mapping Arrays (LMAs) has shown that, even though the NLDN is a network with roughly 350-km sensor baselines and was originally designed to optimize sensor baselines based on the coverage of each sensor, the IC pulse data from the NLDN nevertheless exhibits some of the characteristics of a lightning mapping system. Several cases have now been observed in which the NLDN captured between 25-40% of the total horizontal extents of long flashes and geo-located between 10-25 points within those flashes. This capability is brought about in part by a pulse train processing capability that was first described by Murphy et al. [2004] in conjunction with the 2013 upgrade of the NLDN sensors to LS7002s, as described by Nag et al. [2016] at the previous ILDC. Several examples of this capability are highlighted in this paper. Potential future enhancements to this capability, including altitude estimation, are outlined, with examples as available.

Keywords—LF; NLDN; mapping

I. INTRODUCTION

Lightning mapping from ground-based networks is traditionally done at VHF frequencies using systems such as the Lightning Mapping Array (LMA) [Thomas et al. 2004]. More recently, however, mapping capabilities have also been introduced from networks operating at lower frequencies (generally from VLF into MF, but generically referred to as “LF” in this paper).

Lyu et al. [2014] presented a network of five sensors with 15-20 km baselines and an operating bandwidth of 1-400 kHz. The LF signals detected by these sensors, at the sub-millisecond level, contain trains of short pulses of the type discussed in section II of this paper, but in the method of Lyu et al., a cross-correlation is used to identify a time of arrival associated with each train (or portion of a train) within time windows of 350 µs.

The time-of-arrival (TOA) calculation of the 3-D positions of all sources is based on a look-up table of pre-calculated arrival times. This approach permits a five-station network to perform full 3-D location calculations, and it also has the advantage of handling large uncertainty in the altitude dimension with sources close to the ground. Very detailed 3-D maps of several flashes are demonstrated through the use of this technique.

Bitzer et al. [2013] described HAMMA, an array of 7 sensors with 10-15 km baselines and a 3-dB frequency response of 1 Hz to 400 kHz. This response goes significantly closer to DC than the system of Lyu et al. [2014] but has the same upper frequency limit. The HAMMA E-field recordings are continuous or nearly continuous and are parsed into “events” that can be located by standard LMA-style 3-D TOA. The event extraction is done manually, and usually the peaks of signals are taken as the key time points. This is doable with a short-baseline network because inter-pulse time separations are very often greater than the inter-sensor time differences (that is, sensor baselines / c), such that the confusion of different pulses from different source locations is usually not an issue, or at least it is relatively minimal compared with a longer-baseline network.

The Broadband Observation network for Lightning and Thunderstorm, “BOLT,” is described by Yoshida et al. [2014]. The typical sensor baselines in this network are similar to those of both Lyu et al. [2014] and Bitzer et al. [2013], 15-25 km. The full sensor bandwidth is quoted as 800 Hz to 500 kHz, although a slightly band-passed data set (1-500 kHz) is used to identify pulse peaks. The pulse peak identifications are done in 80-µs windows, basically identical to LMAs in their normal operational configurations. Standard LMA-like 3-D TOA is used to do geo-location, which is possible given the larger number of stations, 11, in this network relative to Lyu et al. [2014]. As in Lyu et al. [2014], very detailed 3-D maps of several flashes are given by Yoshida et al.
Karunaratne et al. [2013] deployed a network in which 6 sensors in the interior are on approximately 15-20 km baselines and then 4 additional sensors around the periphery are on roughly 75-100 km baselines. The normal trigger channel (ch. 1) has a bandwidth of 1.6-630 kHz, but in their paper, the authors used another channel (ch. 3) with a bandwidth from 1 Hz to 2.6 MHz. The signals were digitized at 5 MHz (except at one station, where it was 1 MHz), and then peaks were identified with 0.2-µs resolution. Cross-correlation was used to determine the time differences of peaks between pairs of stations. Then, an essentially typical 3-D TOA method was applied to compute positions, except that their method involved a 2-stage estimation in which stations within 6 km (horizontal) distance of the initial position estimate were eliminated in order to get rid of the static E-field change component at nearby stations, given the greatly enhanced low-frequency response of the channel used in the study.

Mezentsev and Fullekrug [2013] presented one very short baseline array of 10 sensors in southwestern England, where baselines were 0.2 to 1.3 km, and another 7-sensor array with 2-10 km baselines in southeastern France. Note that the term “baseline”, as applied to these arrays, includes not only the typical nearest-neighbor sensor spacings but also the maximum dimension of the array as well. These arrays are essentially like single-station interferometers in the LF from the point of view described in Cummins and Murphy [2009] where baselines are shrunk to the point that TOA differences between sensors turn into measurements of angle from the array as a whole. Thus, these two arrays determine azimuth and elevation angles to, rather than the positions of, LF sources.

The network described by Liu et al. [2017] includes both LF and VHF components. The LF channel is digitized at 5 MHz, but the frequency response of the LF channel is not specifically stated. The authors report an “extracting window” to get pulse features, with a time duration of 1 msec in the LF channel. However, the paper also says that both LF and VHF signals are digitized “continuously”, so presumably the sensors provide a continuous stream of those 1-ms windows, such that as many as 1000 LF events per second are possible. Baselines are something like 10-20 km in the interior, but there are also 2-3 sensors around the periphery that are on longer baselines of approximately 50 km or so. A total of 14 sensors is reported.

The above-mentioned networks that do lightning mapping at LF all share the common feature of operating at short baselines ranging from approximately 10 to 50 km (not counting the very short baselines of the arrays in Mezentsev and Fullekrug). The U.S. National Lightning Detection Network™ (NLDN), by contrast, has typical baselines of 300-350 km. With sensor separations of this magnitude, the NLDN was designed primarily to geo-locate individual cloud pulses (IC pulses) within both cloud flashes (IC flashes) and cloud-to-ground (CG) flashes. As described by Nag et al. [2016], in August, 2015, a new algorithm that processes multiple pulses within pulse trains became part of the operational algorithm suite in the NLDN. At that same time, the standard geo-location algorithm that deals with single pulses (both IC pulses and CG return strokes) was also upgraded. By comparing flashes observed by the combined NLDN algorithm package with data from Lightning Mapping Arrays (LMAs) in Oklahoma [MacGorman et al. 2008] and Colorado [Lang et al. 2014], we find that the NLDN is now capable of capturing a substantial amount of the horizontal extents of some large flashes. In this paper, we document several examples of the quasi-mapping capability of the NLDN and a practical example of the pulse-train processing that contributes to this capability. We also describe several cases where the NLDN did not provide much spatial extent, and we make a preliminary effort to interpret the sometimes mapping-like behavior of the NLDN in the context of previous work on the expected sensitivity of LF systems in comparison with VHF mapping observations.

II. Data

Data from the LMAs in Oklahoma and Colorado were being delivered in quasi-real time to the Vaisala facility in Colorado from 2014 through early 2017. The original purpose of the data feed was to permit the validation of cloud lightning flash detection efficiency of the NLDN (as in Murphy and Nag [2015]). Thus, we were not originally concerned with small data drop-outs, nor with the question of whether more and/or better LMA sources might be obtained via reprocessing to provide greater detail at the sub-flash level. In this paper, we simply use the near-real time data that we had available, even though it turns out to be valuable to examine the details of the flashes that were and were not well mapped by the NLDN. To date, we have not (yet) requested any reprocessed LMA data sets that may be available from any of the selected cases presented in this paper.

The NLDN data were from the operational algorithm suite described by Nag et al. [2016]. As described by those authors, the updated algorithm that classifies CG strokes and IC pulses is known to classify a noticeable population of low-current positive events as CG strokes, even though they are likely to be IC pulses. This mis-classification is not relevant in the context of this paper, where the topic of interest is spatial mapping by the NLDN. Thus, all NLDN events are included in this paper, regardless of the accuracy of their classification.

In all figures and discussion in the following sections, all times are given in UTC. The LMA and NLDN data are displayed in Cartesian coordinates in units of km, using origin latitude, longitude coordinates of (34.9, -97.6) in the Oklahoma cases and (40.7, -105.0) in the Colorado cases.

III. Pulse Train Processing and a Practical Example

Murphy et al. [2004] described a process by which correct time-of-arrival differences could be determined from a collection of pulses whose inter-pulse times were significantly shorter than the propagation times of the signals between neighboring sensors in a long-baseline network such as the NLDN. Without such a process, if individual sensors simply report a single time-of-arrival pertaining to one pulse in the train, or individual times-of-arrival from whatever subset of pulses happens to be detected, then it is very likely that the central processor incorrectly associates the times of arrival reported by different sensors. The result is either badly located lightning positions of poor quality, or lightning positions that are ultimately rejected by the chi-square quality control parameter in the central processor.

The time alignment process of Murphy et al. [2004] begins with a relatively coarse alignment of the times of arrival of key
pulses detected by multiple sensors. Normally, this key pulse is the one with the largest absolute amplitude within a set of pulses spanning a total time period of a couple of milliseconds, although the choice of the key pulse turns out not to be absolutely critical. Following the initial time alignment, the method then identifies the sensor that detected the greatest number of pulses and uses that pulse sequence as a reference. The pulse sequences detected by the remaining sensors are then compared with the reference sequence with the objective of matching as many inter-pulse time intervals as possible. As long as at least some minimum number of time intervals can be matched, from some minimum number of sensors, then the time alignment is considered successful, and the central processor proceeds to calculate the positions of the successfully aligned pulses using the relevant times of arrival as well as an angle measurement from each sensor that is representative of the pulse train.

Fig. 1 shows a set of pulses detected by four NLDN sensors in the central U.S. following just the coarse time alignment stage of the pulse train processing. The pulses were detected over a period of 2 msec ($2 \times 10^6$ nanoseconds as shown on the time axis of the graph). The measured amplitudes of all pulses are represented by the vertical axes in Fig. 1, which all have the same scale. It is not readily obvious which pulses align in time until the fine-scale interval alignment stage is complete. The result of that stage is shown in Fig. 2, together with some visual guides to several clusters of pulses that match and can therefore be geo-located by the NLDN central processor. Altogether, 17 pulses from just these four sensors were successfully time-aligned via this process, and 14 of those produced successful quality-controlled positions. Figure 3 shows the net result of all

![Figure 1](image-url)
Figure 2. Same pulse train from the same four sensors as in Fig. 1, except after the interval-based time alignment stage, with boxes to draw attention to four sub-groups of pulses within the train that are successfully aligned.
NLDN detections in the flash in question, a large flash that extended into the stratiform region of a small mesoscale convective system (MCS) to the west of Oklahoma City on 2017-04-13. In Fig. 3, data from the Oklahoma LMA are shown by small dots, and data from the NLDN are shown by large squares (IC pulses) or triangles (CG strokes).

IV. RESULTS

In general, MCSs appear to be favored storm systems from the point of view of getting the NLDN to depict the spatial extents of large flashes well. Figure 4 shows a three-minute segment of another MCS over the Oklahoma LMA, this time on 2016-05-17 from 07:29-07:32 UTC. Several very large flashes can be seen in this example, with the NLDN depicting almost the complete spatial extents of several of them.

Broadly similar, but not identical, behavior is observed in the area of the Colorado LMA. Figure 5 shows three minutes of data from a line of storms on 2016-06-13 with an incipient stratiform region just to the east of Cheyenne, WY. The system evolved into a better-defined MCS a couple hours later as it propagated out of the LMA coverage area. The developing stratiform region in this storm is to the northwest of the main line of storms, roughly between 40 and 70 km on the Y axis in Fig. 5. The LMA shows a couple of horizontally extensive flashes farther south along the western edge of the line, as well as a couple of extensive flashes that stick out to the east of the line, presumably into forward anvils. The NLDN picks up some of the spatial extents of the extensive flashes farther south and east, but not as much as is shown in the incipient stratiform region on the northwest end of the line.
Figure 4. Data from 2016-05-17 07:29 – 07:32 UTC, over southwestern Oklahoma. As in Fig. 3, color denotes time, starting from dark blue and ending in dark red. Small dots are data from the Oklahoma LMA, large squares are IC pulses from the NLDN, and triangles are CG strokes from the NLDN.
The following paragraphs and accompanying figures describe several cases of flashes in which the NLDN was only able to provide a small amount of detail and some where the NLDN provided roughly complete spatial mapping of the flashes.

Probably the best case of spatial mapping by the NLDN that we have observed thus far occurred in the MCS over western Oklahoma on 2016-05-17. A very extensive flash propagated into the stratiform region from 07:29:11 to 07:29:18. Figure 6 shows both a plan view and an altitude-time view of the flash. Unfortunately, the near-real time LMA data from this case exhibited data gaps, which are clearly visible in the altitude-time view in Fig. 6. We tentatively interpret this event as a single flash in part because the NLDN data are approximately continuous during the LMA data gaps (roughly 0.5 second apiece) and in part because of a rough inference of the channel development. Namely, the total horizontal channel structure is spread out over an area of about 17000 km$^2$ and the total linear propagation length of all channel segments within that area is probably at least 300 km or so, based on what is visible from the LMA data. Thus, assuming that the flash initiated with a normal-polarity intracloud form in the convective region of the MCS, the average propagation speed of negative leaders in this flash would be around $5 \times 10^4$ m/s, which is within a factor of 2 of the typical negative leader propagation speed shown by Van der Velde and Montanya [2013], also based on LMA data. The altitude-time part of Fig. 6 suggests that there was, indeed, some sort of upward development between 7-10 km near the beginning of the flash, suggestive of normal-polarity IC development in the convective region. Between seconds 1-3 of the flash, the altitude-time view indicates a gradual lowering of the LMA sources, most of which went to the north of the flash initiation (although some also developed southward). During the final 2 sec. of the flash, the LMA indicates sources at higher altitudes. These occurred at the far back end of the stratiform region. They could be consistent with a distinct, higher-altitude stratiform-only area of VHF emissions observed near the back of another MCS stratiform region by Carey et al. [2005], and thus, these later sources perhaps should be considered part of a separate flash.

As the MCS in Wyoming on 2016-06-13 propagated farther east after 02:30 UTC, the NLDN again showed some skill at
mapping flashes that most likely propagated into the stratiform region. One such flash occurred at 02:39:18 and is shown in Fig. 7. The difficulty with this time period, when the MCS was better developed than around 00:51 (Fig. 5), is that it had propagated well outside the area of best coverage of the Colorado LMA by 02:39. Thus, Fig. 7 does not show a lot of detail in the LMA data. Based on the temporal continuity, of the LMA data in this case rather than the NLDN data, it appears that this is also a single flash, although the plan view might suggest otherwise. The altitude-time view indicates initial VHF activity at 5-6 km altitude, followed roughly 100-200 msec later by what may be normal-polarity bi-level intracloud development. The upper-altitude VHF sources appear to start descending toward the end of the flash, suggesting propagation into the stratiform region of the MCS. Interestingly, the final LMA source and the final NLDN IC pulse, shown in dark red in the altitude-time projection, occurred in almost the same plan-view position and at almost the same time, suggesting that this is not an outlier in either network’s data, but rather, a genuine component of the flash.

Late in the afternoon on 2016-07-19 (early 2016-07-20 UTC), a broken line of storms propagated from the west-southwest over an area from south of Denver to north and northeast of Fort Collins, CO. These storms produced several horizontally extensive flashes, as viewed from the Colorado LMA, but relatively little in the way of spatial mapping by the NLDN. Figures 8-9 show two flashes, at 01:22:30 and 01:42:01, respectively. Both flashes appear to begin with presumably normal-polarity intracloud development, with the upper-altitude portion descending slightly with time. The presumably negative leaders that trace out the full extents of both flashes are completely unaccompanied by any NLDN detections. Instead, in both cases, the NLDN only detected pulses near the flash origins. However, these NLDN detections, again in both cases, did not occur temporally at the start of the flashes. Instead, they occurred after the full horizontal extent of the flashes was already mapped out by the upper-altitude negative leader, and once some much sparser, lower-altitude LMA sources began to appear, presumably from positive leader development in the presumably negative charge region in the middle altitudes of the storm. The LMA clearly showed a small number of sources that retraced the spatial extents mapped out by the initial negative leaders, so we infer that both flashes generated recoil processes that propagated through the upward channel developed during the initial breakdown and then all the way back out along the initial horizontal channels. These recoil processes appear to be associated with the NLDN detections, which were spatially, but not temporally, close to the flash initiation locations. Recoil processes moving away from the flash origin in the channels mapped by negative leaders are consistent with observations by Van der Velde and Montanya [2013].

Figure 6. Plan view (left) and altitude-time view (right) of flash at 07:29:11 on 2016-05-17 over the Oklahoma LMA. Color in both panels corresponds to time from the start of the flash. LMA VHF sources are shown by small dots. NLDN IC pulses are shown by large, filled squares. Events that were classified as CG strokes by the operational NLDN classification method are shown by triangles pointing up (positive) or down (negative). The same format is used in subsequent figures.
Figure 7. Flash at 02:39:18.4 UTC on 2016-06-13, near the northeastern extent of the Colorado LMA coverage area, in the same format as Fig. 6.

Figure 8. Flash at 01:22:30.2 UTC on 2016-07-20 over the Colorado LMA, in the same format as Fig. 6.
It also appears that the long, horizontal channels of flashes that propagate into forward anvils do not produce signals that are detectable by the NLDN. Figure 10 shows a two-minute plan view of a multi-cellular cluster of storms in Oklahoma on 2017-04-21. These storms did not organize into either a line or an MCS while they were in the LMA domain, but westerly winds of 90-110 km/hr in the upper troposphere at the time did allow extensive forward anvils to develop from the cells in this cluster. These anvils were clearly visible in composite reflectivity from the Norman, OK, NEXRAD (not shown). The LMA observed a number of long horizontal flashes coming out of the east sides of the cells, into the anvils, but as shown in Fig. 10, these were not mapped spatially by the NLDN. The NLDN did capture large numbers of IC pulses in the cell interiors, just not out into the forward anvils.

Figure 11 shows one last example, evidently a bolt-from-the-blue CG flash about 10-20 km east of Rocky Mountain National Park at 02:26:20 UTC on 2016-07-20. The altitude-time panels of all flashes shown in Figs. 6-11 show all NLDN events at 0 km altitude. In the case of Fig. 11, however, it is important to note that the terrain in the area of this flash reaches to over 3 km altitude in some places. The lowest LMA sources are just above 4 km altitude, not far above local ground. Given the greater-than-usual amount of terrain blockage of VHF signals between this flash and the LMA, it is entirely possible that there may have been additional VHF sources closer to the ground in this flash. The NLDN detected the CG stroke at the western end of the flash and two IC pulses. The latter of those was located very close to the flash origin and at a time when the LMA also showed a small number of sources propagating back toward the flash origin from the return stroke location. The earlier of the two NLDN IC pulses occurred prior to the return stroke by about 20 ms, suggesting a leader pulse, but its location ended up about 5 km south of where the leader started its propagation to the west.
Figure 10. Plan view of data from 09:33 to 09:35 UTC on 2017-04-21 in a multi-cellular storm cluster over the Oklahoma LMA, in the same format as Fig. 5.

Figure 11. Bolt-from-the-blue CG flash on 2016-07-20 at 02:26:19.9 UTC, over the Colorado LMA, in the same format as Fig. 6.
V. PRELIMINARY CONCLUSIONS

The NLDN, with its typical sensor baselines of 300-350 km, was designed to have approximately 50\% cloud flash detection efficiency (DE) as a result of the 2013 sensor upgrade and 2015 algorithm upgrade. As a result of comparing NLDN data to LMA data in a number of storms, we have noticed that the enhanced cloud pulse detection capability of the NLDN actually permits the network to provide spatial mapping of some flashes. Our purpose in this paper has been to initiate a somewhat systematic study of horizontally extensive flashes in which the NLDN does and does not do well at mapping the spatial extent.

Of course, the question of spatial mapping should be kept separate from the question of cloud flash DE. Although our purpose in this paper is not an analysis of NLDN cloud flash DE, we should point out that all of the flashes presented in the preceding section have at least one NLDN-detected IC pulse. In the course of the present study, we also found some LMA-detected flashes that did not have any NLDN IC pulses. Broadly, the analysis in this paper is consistent with a cloud flash DE of at least 50\%, as originally expected of the network and as documented in Murphy and Nag [2015].

This study suggests that the NLDN has the best probability of mapping most or all of the spatial extent of flashes that propagate into the stratiform regions of MCSs. Such flashes migrate into large reservoirs of charge that are produced by a combination of advection of charge from the convective line and in situ charge generation [Carey et al. 2005 and references therein]. Horizontally extensive MCS stratiform flashes also often generate large positive CG strokes. Some of the NLDN events in the well-mapped Oklahoma flashes from 2016-05-17 were classified as positive CG strokes, but additional analysis is warranted to see whether these are truly CG or possibly misclassified IC pulses. Presumably, recoil processes associated with positive CG strokes, or perhaps any discharge in general that taps into large reservoirs of charge, might generate a wealth of LF signals that are detected and geo-located by the NLDN.

One question to be addressed in future studies is the degree to which the pulse train processing is responsible for the spatial detail observable in some NLDN flashes. As it turns out, in the flash in Fig. 3, the majority of the spatial extent description in the NLDN data is actually provided by the locations of pulses that were not part of the pulse trains. The pulse trains, in this case at least, tended to fill in details around several of the isolated pulses. But as of now, we have only looked in detail at a few flashes to see whether isolated IC pulses are actually the ones that essentially provide the boundaries of the NLDN’s observations of spatial extent. Additional questions that we defer to future work include an examination of the amplitudes and polarities of the IC pulses in flashes that are well-mapped by the NLDN, as well as whether the NLDN might be able to estimate the altitudes of these pulses.

REFERENCES