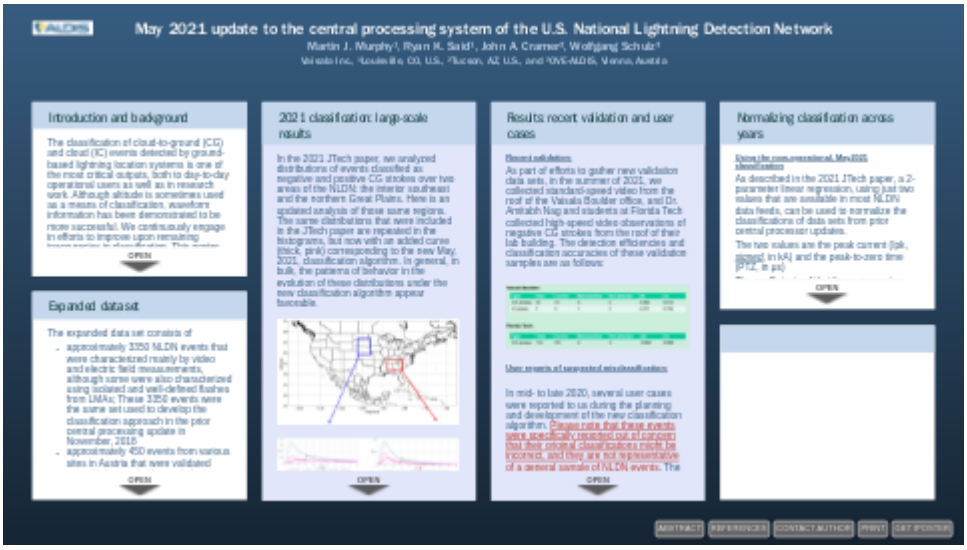
May 2021 update to the central processing system of the U.S. National Lightning Detection Network



Martin J. Murphy[1], Ryan K. Said[1], John A. Cramer[2], Wolfgang Schulz[3]

Vaisala Inc., [1]Louisville, CO, U.S., [2]Tucson, AZ, U.S., and [3]OVE-ALDIS, Vienna, Austria

PRESENTED AT:

INTRODUCTION AND BACKGROUND

The classification of cloud-to-ground (CG) and cloud (IC) events detected by ground-based lightning location systems is one of the most critical outputs, both to day-to-day operational users as well as in research work. Although altitude is sometimes used as a means of classification, waveform information has been demonstrated to be more successful. We continuously engage in efforts to improve upon remaining inaccuracies in classification. This poster provides an update on our latest improvements as applied specifically in the U.S. National Lightning Detection Network™ (NLDN™).

In March, 2021, we published a *JTech* paper on the state of the NLDN up to and including an upgrade to the central processing system on 2018-11-07 (https://journals.ametsoc.org/view/journals/atot/38/3/JTECH-D-19-0215.1.xml (https://journals.ametsoc.org/view/journals/atot/38/3/JTECH-D-19-0215.1.xml))

The 2018-11-07 central processor update included an updated classification algorithm but with some suspected remaining limitations.

On the basis of those remaining limitations, in late 2020 (while the 2021 paper was under review), we undertook another round of work on classification after expanding the data set with a sample of about 450 video-validated events from Austria

Objectives with the expanded data set

- increase sample size
- increase sample diversity, with an eye toward filling in more detail about the full “sample space” of CG and IC events

Outcome of the 2020 work:

Central processor update in the NLDN that went into operations on 2021-05-27

Objectives of this presentation

- new classification results
- analysis of new 2021 validation data set results and some pre-2021 user reports of suspected misclassifications
- how to normalize to the current operational classification using the same method presented in the 2021 *JTech* paper

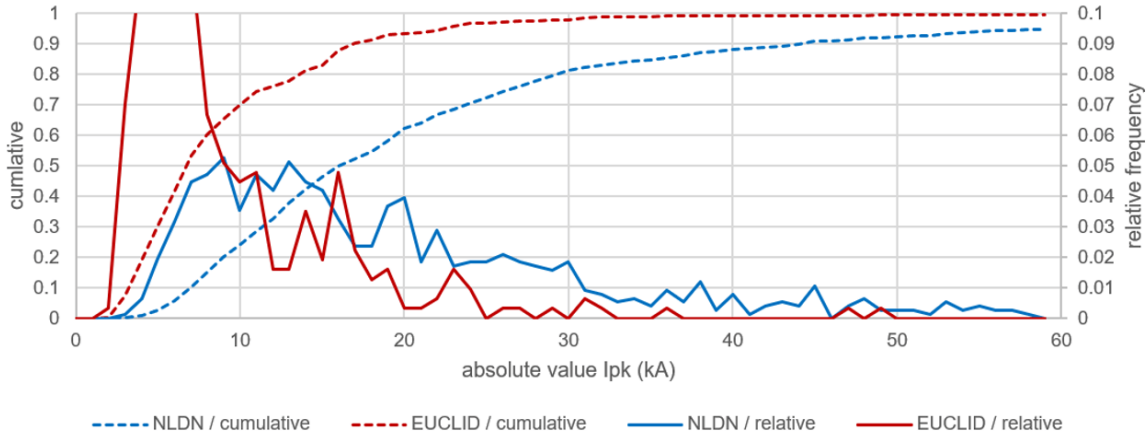
EXPANDED DATA SET

The expanded data set consists of

- approximately 3350 NLDN events that were characterized mainly by video and electric field measurements, although some were also characterized using isolated and well-defined flashes from LMAs; These 3350 events were the same set used to develop the classification approach in the prior central processing update (2018-11-07)
- approximately 450 events from various sites in Austria that were validated using high-speed video and electric field measurements and detected by the EUCLID network (see Schulz et al. 2015 and Schwalt et al. 2020)

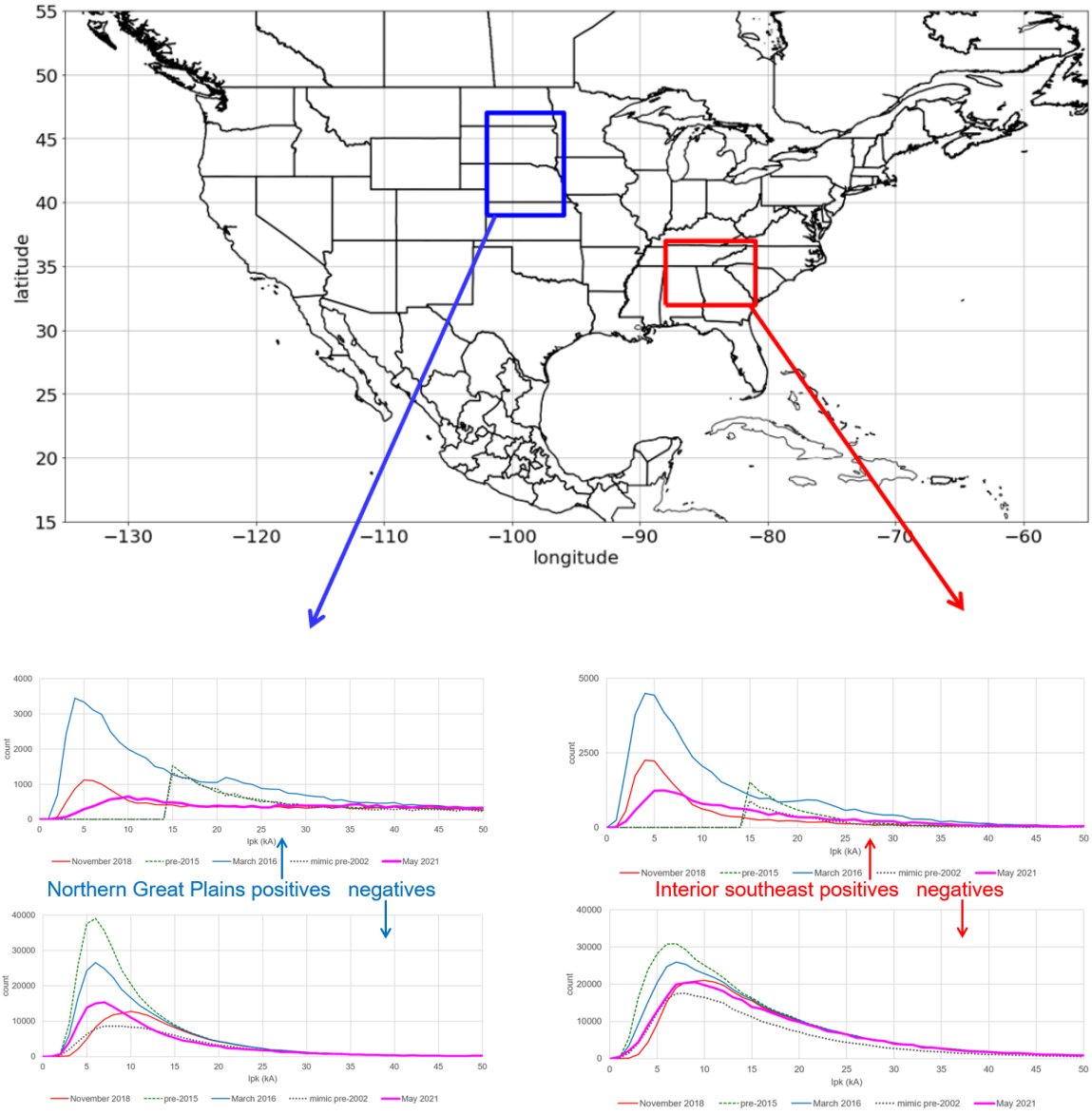
Although the EUCLID data set included a number of IC pulses as well as CG strokes, most of the CG strokes in Austrian summer thunderstorms are negative; thus the positive data set was almost entirely IC pulses.

The negative CG events are more interesting. The graph below shows the histograms and cumulative distributions of absolute value of peak current in validated, labeled negative CG strokes in the NLDN (blue) and EUCLID (red) data sets. Our goals in expanding the data set were to increase sample size and also increase sample diversity in order to make sure that the full feature spaces of both CG strokes and IC pulses were captured more thoroughly. However, the significant difference between the NLDN and EUCLID peak current distributions appears to have created an unanticipated challenge to the classification of negative events.



2021 CLASSIFICATION: LARGE-SCALE RESULTS

In the 2021 *JTech* paper, we analyzed distributions of events classified as negative and positive CG strokes over two areas of the NLDN: the interior southeast and the northern Great Plains. Below is an updated analysis of those same regions, but now with an added curve (thick, pink) in each distribution plot corresponding to the now-operational 2021-05-27 classification algorithm. In general, the patterns of behavior in the evolution of these distributions under the new classification algorithm appear favorable.



With positive events in particular, the 2018-11-07 classification algorithm left behind modes in the "+CG" distributions at around 4-6 kA. Capturing independent validation information about such low-amplitude positive events remains elusive, but they have long been suspected (e.g. Cummins et al. 1998) of being misclassified IC pulses. The 2021-05-27 classification algorithm now classifies a large portion of those low-current positive events as IC pulses. It also preserves as +CG strokes a large number of events with peak current > +15 kA that the 2018-11-07 algorithm classified as IC pulses.

RESULTS: RECENT VALIDATION AND USER CASES

Recent validation:

As part of efforts to gather new validation data sets, in the summer of 2021, we collected standard-speed video from the roof of the Vaisala Boulder office, and Dr. Amitabh Nag and students at Florida Tech collected high-speed video observations of negative CG strokes from the roof of their lab building. The detection efficiencies and classification accuracies of these validation samples are as follows:

Vaisala Boulder:

Type	Total	Correct	Misclassified	Not detected	DE	CA
CG strokes	18	13	3	2	0.889	0.812
IC pulses	7	3	1	3	0.571	0.750

Florida Tech:

Type	Total	Correct	Misclassified	Not detected	DE	CA
CG strokes	114	110	2	2	0.982	0.982

User reports of suspected misclassification:

In mid- to late 2020, several user cases were reported to us during the planning and development of the new classification algorithm. Please note that these events were specifically reported out of concern that their original classifications might be incorrect, and they are not representative of a general sample of NLDN events. The following table is a summary of these 13 user-reported events. Of the 13 events, only 3 (23%) were correct by their original classifications, but after the new classification was applied, 11 of the 13 (85%) are correctly classified.

date	2020-04-19	2020-04-19	2020-04-19	2013-05-19	2013-05-19	2013-05-19	2013-05-19	2013-05-19	2013-05-19	2020-05-26	2020-05-26	2020-07-11	2020-07-11
polarity	–ve	–ve	–ve	–ve	–ve	–ve	–ve	–ve	–ve	+ve	+ve	+ve	+ve
valida-tion	video	video	video	LMA	LMA	LMA	LMA	LMA	LMA	transmission-line faults on electric power grids			
correct type	CG	CG	CG	IC	IC	IC	IC	IC	IC	CG	CG	CG	CG
original type	CG	CG	CG	CG	CG	CG	CG	CG	CG	IC	IC	IC	IC
new type	CG	CG	CG	IC	IC	CG	CG	IC	IC	CG	CG	CG	CG

NORMALIZING CLASSIFICATION ACROSS YEARS

Using the now-operational, May2021 classification

As described in the 2021 *JTech* paper, a 2-parameter linear regression, using just two values that are available in most NLDN data feeds, can be used to normalize the classifications of data sets from prior central processor updates.

The two values are the peak current (I_{pk}, *signed*, in kA) and the peak-to-zero time (PTZ, in μs)

The coefficients of that linear regression are now different due to the latest, 2021-05-27, central processor update:

coefficient	positive events	negative events
intercept (*a*)	0.900768	0.799103
signed Ipk (kA) (*b*)	–0.015322	0.022980
peak-to-zero time (μs) (*c*)	–0.041949	–0.029036

$$\hat{y} = a + b(I_{pk}) + c(PTZ)$$

$$if\ \hat{y} > 0, event\ is\ IC;\ \hat{y} \leq 0, event\ is\ CG$$

Effects of the above linear regression relative to the now-operational classification

Large samples of data from prior evolutions of central processor have been reprocessed through the currently operational classification and then again through the above linear regression.

The 3rd column of the table below shows the total number and percentage of all events (CG and IC) that are classified differently by the linear regression relative to the now-operational classification algorithm. That change is on the order of 5%; in other words, the above linear regression matches the now-operational classification algorithm on about 95% of events.

The last column of the table below shows the change in −CG strokes by using the linear regression vs. the now-operational classification. The linear regression tends to change the classification from CG to IC on about 8-10%; that is, use of the above linear regression causes a roughly 8-10% decrease in −CG strokes relative to the now-operational classification, and a corresponding increase in −IC pulses.

Central processor version	total events	net reclassified by linear rgr.	–CG strokes	–CG reclassified by linear rgr.
pre-2015	45918467	2514178 (5.5%)	12983030	–1272673 (–9.8%)
March 2016	38213559	1515645 (4.0%)	9504556	–856020 (–9.0%)
Nov. 2018	54257690	2037876 (3.8%)	11754667	–1005494 (–8.6%)

CONCLUSIONS

- The 2021-05-27 NLDN central processor update is part of ongoing efforts to make the classification of IC and CG events as accurate as possible
- Through future validation work, we will seek to understand any remaining classification inaccuracies and make any additional improvements
- In parallel, we are also investigating the determination of a classification confidence that can be made available in future NLDN data feeds

ABSTRACT

In March, 2021, we published a paper covering a range of upgrades made over the years to the U.S. National Lightning Detection Network (NLDN) – see https://journals.ametsoc.org/view/journals/atot/38/3/JTECH-D-19-0215.1.xml (https://journals.ametsoc.org/view/journals/atot/38/3/JTECH-D-19-0215.1.xml). In that paper and the work leading up to it, we noted that there was still room to improve upon the classification algorithm used to distinguish cloud-to-ground (CG) return strokes from cloud (IC) pulses. While the JTech paper was in press, we started working toward the objective of classification improvements using an expanded set of events that were independently characterized with a mix of video and electric field observations. A new classification algorithm, as well as some enhancements designed to improve upon the detection efficiency of IC pulses, were deployed in an update to the NLDN's central processing system on 27 May, 2021. In this presentation, we briefly describe the changes, and then we proceed to discuss our preliminary efforts to evaluate the performance of the new classification algorithm.

REFERENCES

Cummins, K.L., M.J. Murphy, E.A. Bardo, W.L. Hiscox, R.B. Pyle, and A.E. Pifer, 1998: A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network, *J. Geophys. Res.*, **103**, 9035-9044.

Rakov, V. A., et al., 2013: "CIGRE technical brochure on lightning parameters for engineering applications." *2013 International Symposium on Lightning Protection (XII SIPDA)*. IEEE, 2013.

Schulz, W., G. Diendorfer, S. Pedeboy, and D.R. Poelman, 2015: The European lightning location system EUCLID – part 1 Performance validation. *Nat. Hazards Earth Syst. Sci. Discuss.*, **3**, 5325-5355.

Schwalt, L., S. Pack and W. Schulz, 2020: Ground truth data of atmospheric discharges in correlation with LLS detections. *Elect. Power Sys. Research*, **180**, paper 106065.