2017 International Symposium on Lightning Protection (XIV SIPDA), Natal, Brazil, 2nd – 6th October 2017.

Evaluation of EUCLID IC/CG Classification Performance Based on Ground-Truth Data

H. Kohlmann, W. Schulz
OVE-ALDIS
Vienna, Austria
h.kohlmann@ove.at, w.schulz@ove.at

S. Pedeboy
Meteorage
Pau, France
sp@meteorage.com

Abstract—**This work compares the classification accuracy (CA) of two algorithms applied to data from the EUCLID lightning location system (LLS). As CA we call the accuracy of a LLS to correctly distinguish between cloud-to-ground (CG) and intra cloud (IC) discharges. The ground-truth data, used for this evaluation, was taken from optical and electric field data measured in various regions in Austria (2012 and 2015) and France (2014). The data set contains CG and IC discharges of positive and negative polarity. The data set was split up into further sub-categories as long as the number of data was still sufficient to give reasonable results. For a coarse overview of the algorithm performances, the total CA was first calculated for each year and country for both polarities. Furthermore, for the class of CG discharges, the CA of first return strokes, the CA of subsequent return strokes (with and without respect to the polarity) as well as the CA of IC events with respect to their polarity was evaluated. Specifically the subdivision into classes of amplitudes of the peak currents for different events can give further insight to the performance of the algorithms. For that reason the total CA, the CA of negative and positive events and CG and IC was analyzed.**

The evaluation shows that in combination with the new sensor data format LS the new algorithm exhibits an improvement of 2% at the CA. In combination with the old data format IMPACT, the new algorithm performs worse than the old one. In total, IC events have been classified much better by the new algorithm, irrespective of the sensor data format. CG discharges on the other hand show a worse CA throughout all years for the new algorithm.

Keywords—lightning location systems; performance of LLS

I. Introduction

The performance of LLS is becoming more and more important. Different evaluations for performance parameters, among which for example detection efficiency, classification accuracy location accuracy and outliers are important ones, are available. Different methods to determine several performance parameters are described by Nag et al. [1]. Due to improvements during the last years lightning location systems (LLS) detect more and more cloud pulses. Therefore the accuracy of a network to correctly distinguish between cloud-to-ground (CG) strokes and intra cloud (IC) discharges is becoming more important. We call this performance parameter “classification accuracy” (CA). There are different methods to classify events. Some LLS are classifying based on the peak-to-zero (PTZ) time, some on the altitude of the discharge and some on a combination of different waveform parameters using a linear discriminant analysis.

Small positive events have been analyzed several times in literature for their occurrence. Campaigns in Southern Arizona and Texas-Oklahoma suggest that most (~90%) of the positive small events (<10 kA) are actually cloud discharges [2]. Further Biagi et al. [3] show in their paper that between 1% and 7% of the positive events between 0 kA and 10 kA and 1% to 30% of positive events between 10 kA and 20 kA were cloud to ground discharges. Because of the results from those papers show that low peak current positive CG strokes are not frequent, classification algorithms often have a threshold below which detected positive discharges are classified as IC events in order to improve the CA. In [3] an amplitude level of 15 kA is mentioned for which the number of false classifications of positive events is equal to correct reports. Further in this work, a classification accuracy between 50% and 87% for negative CG strokes smaller than 10 kA is reported.

Only few papers dealing with the CA performance of LLS for the complete peak current range are available in the literature. In a paper by Mallick et al. [4] the CA of rocket triggered return strokes is analyzed. Those return strokes are comparable to natural subsequent strokes regarding their characteristics. Mallick et al. reported a CA for rocket triggered return strokes of 96%. Fleenor et al. [5] reported that between 34% and 54% of IC events are falsely classified as CG strokes. Further Zhang et al. [6] reported that 78% of negative and 100% of positive cloud pulses were classified correctly and the total CA of IC pulses is 92%. In Zhu et al. [7], the NLDN (National Lightning Detection Network) with the configuration of summer 2014 classified 86% of IC events correctly. It is also stated in that study that from 24 preliminary breakdown pulses 46% were detected and 82% were classified correctly. The CG stroke CA for first and subsequent return strokes (RS) taken together was 91%. For first RS the CA was higher than for subsequent RS. Negative subsequent strokes were reported to have a CA of 90%. Further Zhu et al. [7] provides the CA of various types of events such as positive and negative discharges, CG and IC or first return strokes (RS) and subsequent RS. In Zhu et al. [7] the following CAs have been given: For all CG events 92% CA were reported. Negative strokes have 92% CA among which negative first RS have 96% CA and neg. subsequent RS have 90% CA. For positive strokes the CA is 96% (among which positive first RS have 95% CA and positive subsequent RS have 100% CA). For first RS and subsequent RS altogether the CAs are 96% and 90% respectively. Further, for all IC events in that study, the CA is 86%.

II. DATA & METHODOLOGY

The data used in this analysis are from the EUCLID (European Cooperation for Lightning Detection) network. Figure 1 shows the sensor layout of the EUCLID network. During the years from 2012 to 2015, which is the period where data are used for this paper, the network itself has not changed essentially. What on the other hand has changed is the provided sensor data format. The former is called IMPACT data format and contains only little information about the measured waveforms. The new format, the LS data format, provides more parameters of the measured waveforms. Hence, for the data sets used for the analysis, the 2012 (Austria) and 2014 (France) EUCLID data are based on the IMPACT data format and 2015 (Austria) on the LS data format. Additionally, both classification algorithms use the information of the closest two not over-ranged sensors for classification.

The old classification method is based on the peak-to-zero (PTZ) time solely, and the new classification is based on a linear discriminant analysis (LDA) of different waveform parameters.

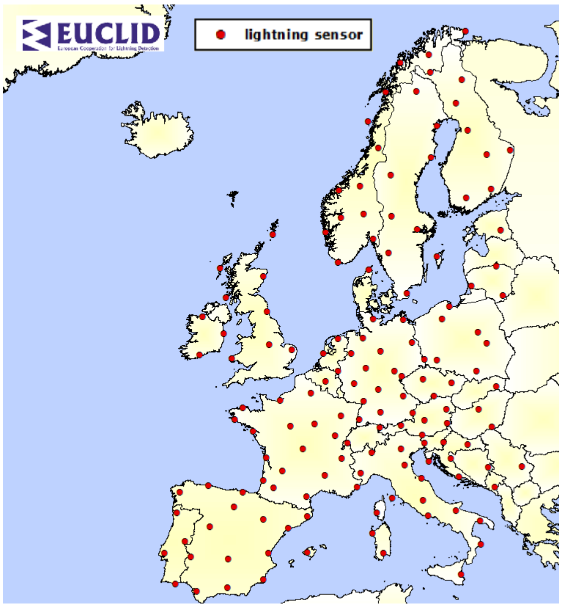


Figure 1: EUCLID Network

The evaluation was performed for the years 2012, 2014 and 2015 because ground-truth data from Austria and France are available. The data were recorded with a video and E-field recording system described in [8], [9] and [10]. The locations of the events are shown in Figure 2.

For the analysis, data from optical measurements and corresponding E-field measurements from Austria were chosen in a way that a certain number of flashes with multiple events are available (such as first strokes, subsequent strokes and IC pulses) summing up to a total of about 150 to 200 discharges in the years 2012 and 2015. The number of ground truth events for the different years and countries is listed in Table 1.

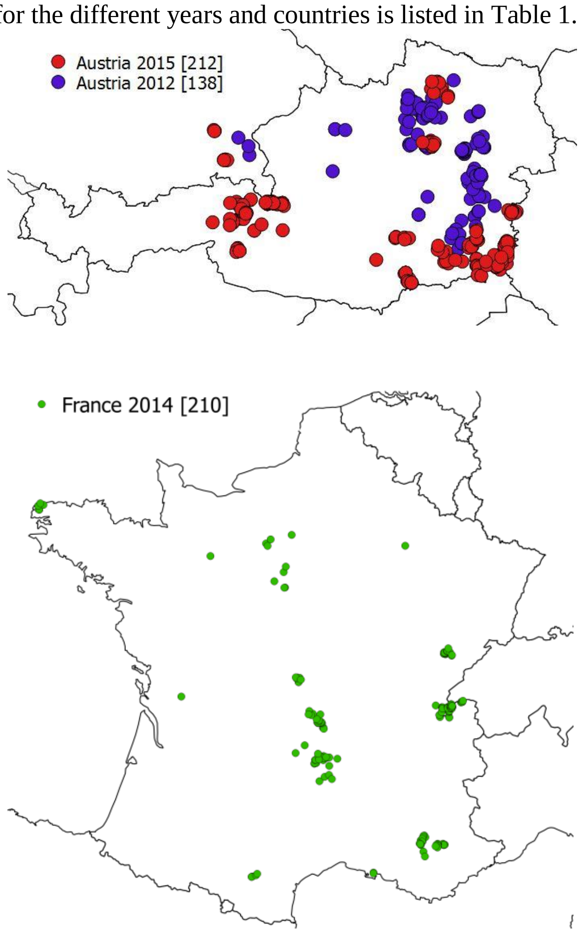


Figure 2: Evaluated ground truth events in Austria and France

The ground truth data from France were selected to suite the chosen data set from Austria 2012, 2015 regarding the number of discharges and class of events as well as possible. It can be seen in Table 1, that the data set of France 2014 contains only a small number of IC pulses. Among the 11 IC events, 9 were negative discharges. Because of the small number of IC events in France only CG data are compared to the Austrian CG data.

Table 1: Number of ground truth events

	negative events	positive events	CG events	IC events	Total
Austria 2012	85	53	88	50	138
Austria 2015	136	76	95	117	212
France 2014	138	72	199	11	210

For evaluation of the CA, the available raw sensor data of the events were reprocessed (recalculated) with both, the old and the new classification algorithm. For positive events we used for the old and for the new algorithm an IC threshold of 5 kA meaning that all positive events below 5 kA were classified as an IC pulse. The reprocessing result was then compared to the ground-truth data. In case that the type of event was clearly identified by means of either the optical measurement or the E-field measurement or both and the result of the classification algorithm coincided with or differed from the observation, the classification was taken as correct or incorrect respectively. Special attention was paid to the case, where the result of the two classification algorithms differed. Those cases where different classifications were observed were double checked for the Austrian events with the video and E-field observations. In case of the French data, the raw video and E-field data was not available. Any doubt in the ground-truth data regarding IC or CG led to the decision to not include the event into the evaluation.

III. Analysis

In this part of the work the results of the evaluation of all data sets are presented. This is done by means of tables where relevant numbers of event occurrences are given and histograms showing the CA for those events. That intends to make reasoning as fast as possible.

To start with an overview of the performance of the old versus the new classification algorithm, Table 2 shows the CA of positive, negative and total events for each year and region.

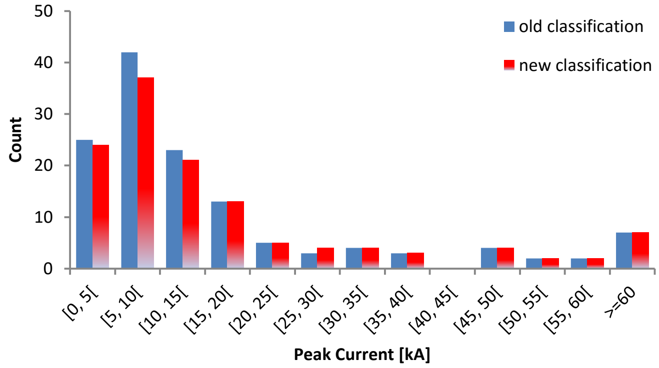
Table 2: CA for negative and positive events for all three campaigns.

		CA pos. events	CA neg. events	Total
2012 Austria	old classification	96% (51/53)	97% (82/85)	97%
	new classification	96% (51/53)	88% (75/85)	91%
2015 Austria	old classification	88% (67/76)	86% (117/136)	87%
	new classification	95% (72/76)	85% (116/136)	89%
2014 France	old classification	79% (57/72)	93% (129/138)	89%
	new classification	90% (65/72)	82% (113/138)	85%

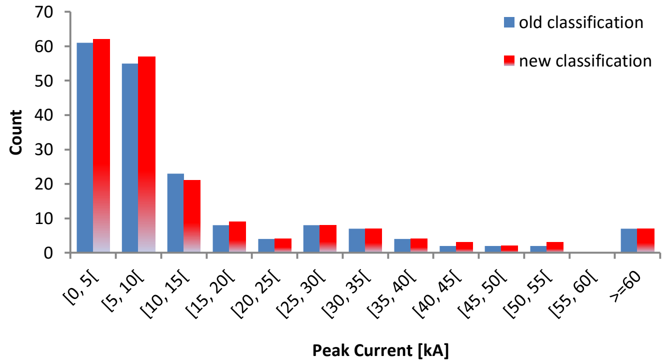
An improvement of the CA in comparison to the old classification is observed for positive discharges in 2015 (Austria) where the CA increased from 88% to 95% and for positive discharges in 2014 (France) where the CA increased from 79% to 90%. The CA for positive events did not change in 2012 (Austria). Table 2 shows that, negative discharges have been classified worse in all years. The worst classification is observed for 2014 (France) with a degradation from 93% to 82%. The rightmost column is the total CA per year and classification algorithm. It can be concluded that for the old data format (IMPACT), which was used in Austria 2012 and in France 2014, the new classification algorithm exhibits worse classification results. The LS data format, used in Austria in 2015, together with the new classification shows a small total classification improvement (from 87% to 89%) and negative events have only been classified worse by 1%.

For a more detailed analysis, Figure 3 shows the histograms of the total number of negative and positive events for different peak current intervals. The intervals of the peak current classes are $\Delta I_{peak} = 5$ kA and are represented by square brackets. E.g. [0, 5[means that the interval includes 0 kA but excludes 5 kA.

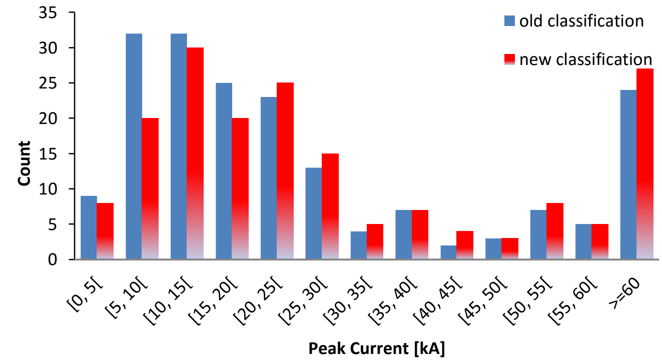
Comparing IMPACT data format (Austria 2012, France 2014) and LS data format (Austria 2015) Figure 3b) shows an improvement of the CA for amplitudes in the range of 0 to 20 kA for the LS data format together with the new algorithm. In that peak current range, the new classification algorithm is better for all intervals than the old algorithm except for [10, 15[kA, where the CA of the new algorithm for the LS data format is worse.



a) Austria 2012



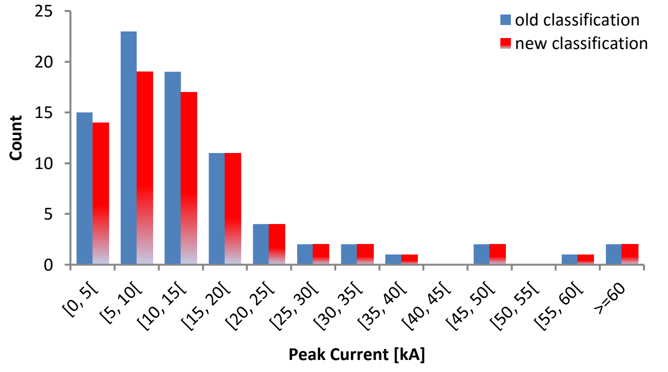
b) Austria 2015



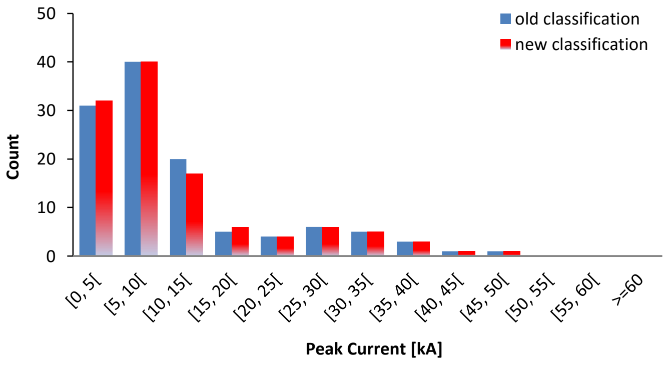
c) France 2014

Figure 3: Histogram of correct classifications of all events versus peak current

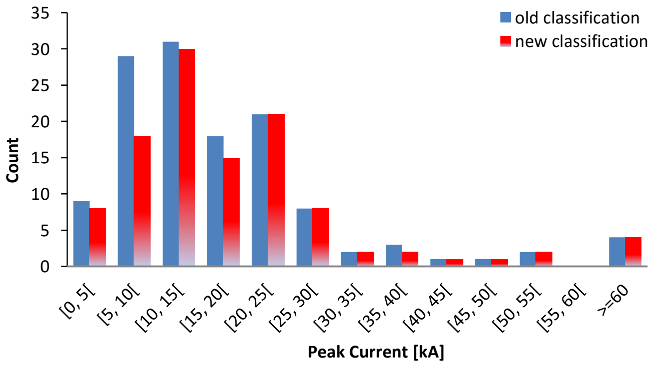
In Austria classification differences can basically be seen only for small peak currents what is not the case for France. In France even for peak currents greater than 20 kA results are different. The evaluation of the French data exhibits a different CA performance for certain peak current ranges. For peak currents less than 20 kA the old classification is better and for peak currents greater than 20 kA the new algorithm classifies better. Although in the ≥60 kA bin in Figure 3c a lot of events are accumulated, the CA for high peak currents is still better on average for the new classification. By looking at Figure 4 one can see that for large peak currents of negative discharges there is practically no difference in the CA for all years. By further looking at Figure 5c) one can see that the positive events of France, 2014 are responsible for the better classification performance for peak currents greater than 20 kA in France. For the worse classification performance below 20 kA in France dominantly negative discharges are responsible but also some positive events below 20 kA are classified worse by the new algorithm. Practically no difference in the classification of positive events in Austria 2012 except for the 5-10 kA range where the old algorithm classified slightly more events correctly and in the 25-30 kA interval, where the new algorithm classified better.



a) Negative events 2012 Austria



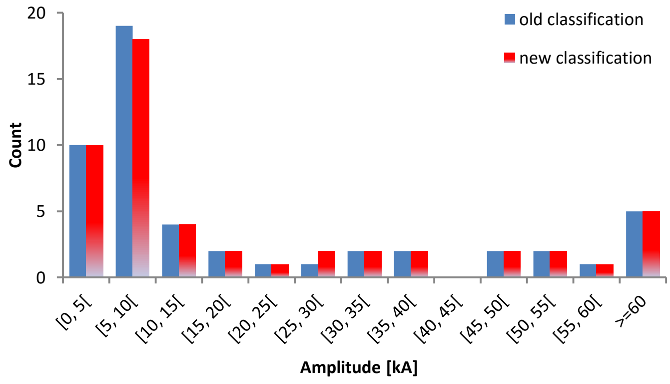
b) Negative events 2015 Austria



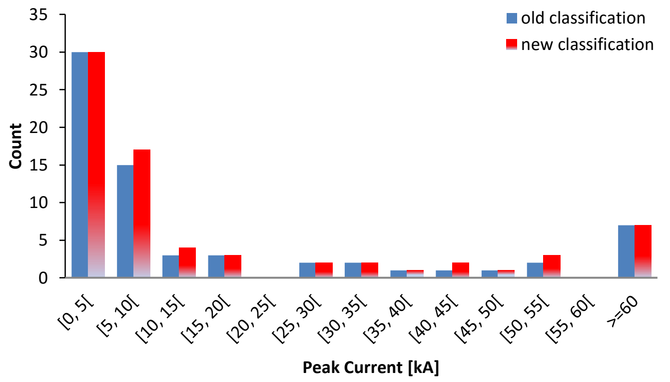
c) Negative events 2014 France

Figure 4: Correct classification of negative discharges versus peak current

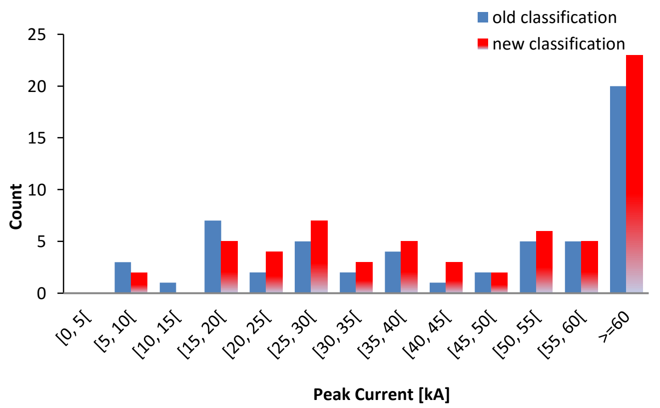
Due to the configured 5 kA limit – see section II - all positive events below 5 kA in Figure 5 in Austria are cloud pulses.



a) Positive events 2012 Austria



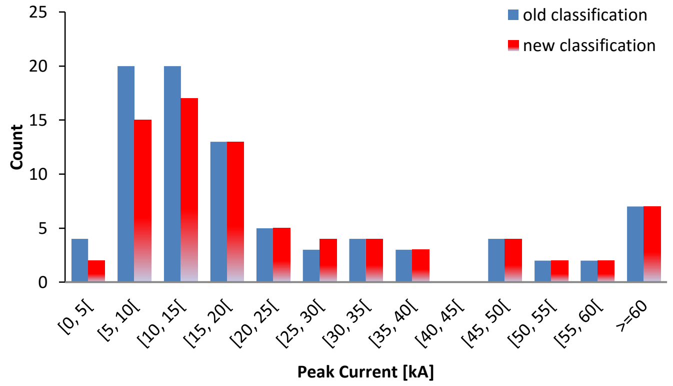
b) Positive events 2015 Austria



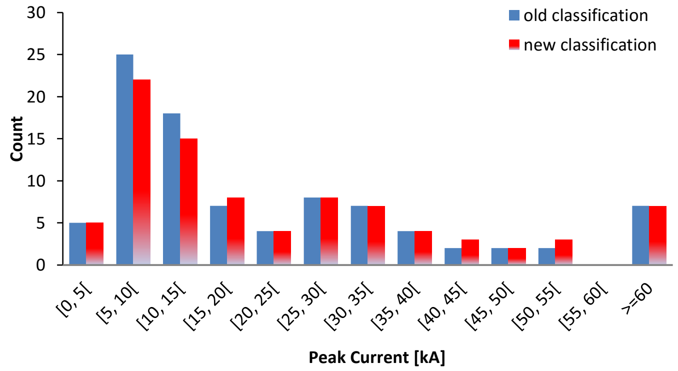
c) Positive events 2014 France

Figure 5: Correct classification of positive discharges versus peak current

When separating the data into CG (Figure 6) and IC strokes (Figure 7), the following can be seen: Below 15 kA CG strokes have been classified worse with the new algorithm for both the IMPACT (Austria 2012) and the LS data format (Austria 2015). Since in the French data set there were only 11 IC discharges (9 negative IC events and 2 positive IC events) which all lie below 15 kA, Figure 3c, Figure 4c and Figure 5c all represent CG events above 15 kA. Above 20 kA the new classification algorithm performs better than the old algorithm for the data in France (see Figure 3c). Regarding IC events, the results of this study show that the performance of the new classification was better or equal throughout all peak current intervals of both Austrian data sets.



a) CG events Austria 2012



b) CG events Austria 2015

Figure 6: Correct classification of CG discharges versus peak current

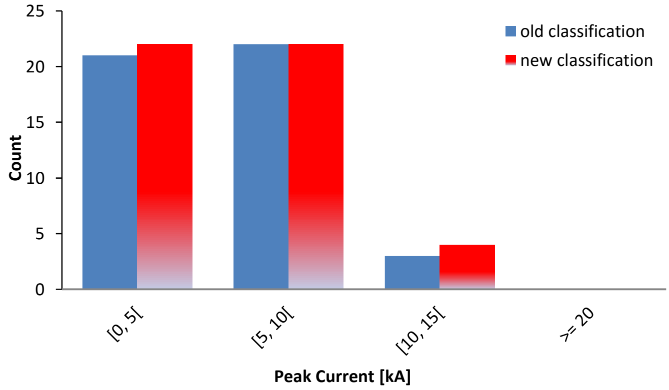
In Table 3 and Table 4 the IMPACT data (2012 Austria and 2014 France) and LS (2015 Austria) sensor data formats can be compared.

Table 3 shows the CA of IC events in Austria 2012 and 2015. For the IMPACT data, the new algorithm shows an improvement from 92% to 96% in total. Both positive and negative events were classified better with the new algorithm. An improvement was also observable for French data, which is not included in Table 3 as the small amount of events makes the result not very representative. Though, the total CA of IC events has improved from 73% to 91%. Among the 11 French IC discharges there were 9 negative events, all below 15 kA, and 2 positive events between 5 and 10 kA.

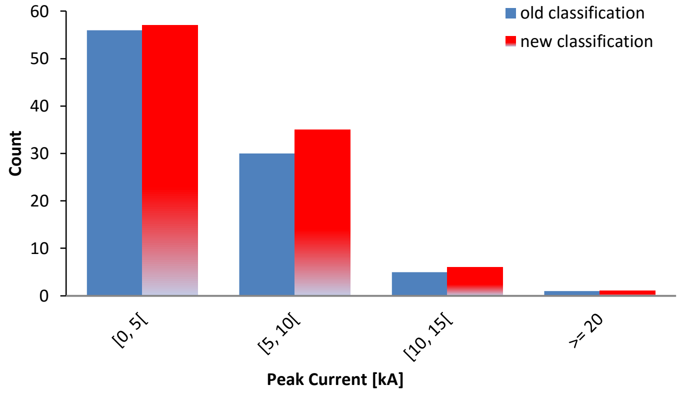
For the LS sensors the CA has improved from 86% to 93% in total. Also here, all positive and negative events have been classified better.

Table 3: CA of IC events

2012 Austria	old classification	new classification
Pos. IC	97% (30/31)	100% (31/31)
Neg. IC	84% (16/19)	89% (17/19)
Total IC	92% (46/50)	96% (48/50)
2015 Austria	old classification	new classification
Pos. IC	90% (47/52)	100% (52/52)
Neg. IC	82% (46/56)	86% (48/56)
Total IC	86% (93/108)	93% (100/108)



a) IC Austria 2012



b) IC Austria 2015

Figure 7: Correct classification of IC discharges versus peak current

For IMPACT data events, we see for positive CG events in the French data set that the new algorithm exhibits an improvement from 79% to 90% (see also the improvement for peak currents greater than 20 kA in Fig. 5c) and for negative CG a degradation from 95% to 81% because of the peak currents below 20 kA (Fig. 4c). While for negative CG the worse CA with the new algorithm can also be observed in the Austrian data of 2012, which exhibits a degradation from 100% to 88% the positive CG classification shows a degradation from 95% to 91% instead of an improvement.

For LS data events, positive CG events (2015 Austria) are classified with the same accuracy for both classification algorithms and a degradation from 89% to 85% was observed for negative CG events. In total the CA has decreased from 88% to 85%.

Table 4: CA of CG events by polarity

2012 Austria	old classification	new classification
Pos. CG	95% (21/22)	91% (20/22)
Neg. CG	100% (66/66)	88% (58/66)
Total CG	99% (87/88)	89% (78/88)
2015 Austria	old classification	new classification
Pos. CG	83% (20/24)	83% (20/24)
Neg. CG	89% (71/80)	85% (68/80)
Total CG	88% (91/104)	85% (88/104)
2014 France	old classification	new classification
Pos. CG	79% (55/70)	90% (63/70)
Neg. CG	95% (123/129)	81% (104/129)
Total CG	89% (178/199)	84% (167/199)

We have further analyzed special types of IC and CG events. For initial breakdown (IB) events we determined the detection efficiency (DE_{IB}) and classification accuracy (CA). E-field data that contained CG events were analyzed for typical IB waveforms before the first return stroke. We disregard any data where no significant IB waveform showed up in the E-field data. When at least one pulse of the IB event was detected, irrespective of classification, the IB was taken as detected. Altogether we observed in Austria for 34 flashes an IB in 2012 and 36 in 2015 respectively. The DE_{IB} in 2015 (33 %) is significantly larger compared to the DE_{IB} in 2012 (18 %). This increase can be attributed to a slightly more sensitive new sensor (LS7002) and an improved location algorithm which groups sensor information in a better way to the same event. For the LS sensor data (2015 Austria) the new classification algorithm shows an increased CA of 94% (see Table 5).

Table 5: DE_{IB} of IB events and CA of IB pulses

	DE_{IB} in %	Old CA in %	New CA in %
2012 Austria	18% (6/34)	94% (15/16)	94% (15/16)
2015 Austria	33% (12/36)	83% (15/18)	94% (17/18)

For data in Austria 2012 the mean number of detected IC events per IB is 2.5 with a sample variance of 4.3 which tells that this mean value is not very trustworthy because of the small sample size. Among the 6 detected IB events in 2012, one was detected with 4 and one with 6 IC pulses. The remaining 4 events were detected with only one or two IC pulse per IB. For 2015 Austria the mean number of detected IC events per IB is 1.58 with a sample variance of 1.17 which is a better value compared to 4.3 in the previous case.

A further CA analysis for first RS and subsequent RS of CG events was made. The results are given in Table 6. It shows that positive first RS are classified better with the new algorithm in Austria 2015 and France 2014. Negative first and subsequent RSs are classified equally or worse in all cases. For the LS sensor data (2015, Austria), positive first RS were classified better with an improvement from 86% to 91% for the new algorithm. Negative first RS in that year were classified with the same CA by both algorithms.

Table 6: CA of first RS / subsequent RS by polarity

Austria 2012	old classification	new classification
Pos. first RS	95% (20/21)	95% (20/21)
Neg. first RS	100% (34/34)	91% (31/34)
Neg. subs. RS	100% (32/32)	84% (27/32)
Austria 2015	old classification	new classification
Pos. first RS	86% (19/22)	91% (20/22)
Neg. first RS	83% (34/41)	83% (34/41)
Neg. subs. RS	95% (37/39)	85% (33/39)
France 2014	old classification	new classification
Pos. first RS	80% (53/66)	92% (61/66)
Neg. first RS	97% (72/74)	81% (60/74)
Neg. subs. RS	93% (51/55)	80% (44/55)

The CA of positive subsequent strokes was not analyzed because of the small sample size – one event in 2012, two events in 2015 and four events in 2014.

IV. Discussion and Summary

This evaluation can be seen as a comparison of the performance of the algorithm that uses a linear discriminant analysis (LDA) for classification and the old algorithm which uses the peak-to-zero (PTZ) time to classify events with respect to two types of sensor data formats. These are IMPACT and LS data format where the latter offers more waveform parameters.

While considering that the results are depending on the choice and size of the data set and the location of ground truth recordings, the result of this study can be summarized as follows. The new data format together with new classification algorithm shows an improved CA. The overall CA improved from 87% to 89%, where positive events have been classified significantly better (increase from 88% to 95%) and negative events have been classified worse by 1% (from 86% to 85%). The overall improvement for the 2015 Austrian data accounts to the fact that the new classification algorithm performed much better on the total IC events and the data set contains 55% IC events (117/212) which gives these events a lot of weight in the result. The CG events were classified worse by the new algorithm throughout all evaluated years (see Table 4).

Irrespective of the data format, the new classification algorithm performs better on IC events for all years. The total CG classification of the new algorithm on the other hand has decreased for both data formats (see Table 4), but the best results are still achieved by the LS data format with a CA decrease from 88% to 85%.

The results for the new LS data format show that negative first return strokes (RS) are equally classified by both algorithms and positive first RS show a CA improvement from 86% to 91%. We conclude that the combination of the LS data format together with the new classification algorithm seem to have a positive effect on the CA of first return strokes. The CA of subsequent negative RS for the LS data format has suffered a degradation from 95% to 85% in this study.

Irrespective of the data format, negative subsequent RS have all been classified a lot worse by the new algorithm. The combination of IMPACT data and the new location algorithm has a big effect on negative events below 20 kA, which can be seen in Figure 4.

Further the detection efficiency of initial breakdowns (IB) has improved between 2012 and 2015 in Austria from 18% to 33%.

REFERENCES

[1] A. Nag, M. J. Murphy, W. Schulz, and K. L. Cummins, "Lightning Locating Systems: Insights on Characteristics and Validation Techniques," *Earth Sp. Sci.*, no. 2, pp. 65–93, Feb. 2015.

[2] K. L. Cummins *et al.*, "The U.S. National Lightning Detection Network: Post-Upgrade Status," *Am. Meteorol. Soc.*, 2006.

[3] C. J. Biagi, K. L. Cummins, K. E. Kehoe, and E. P. Krider, "National Lightning Detection Network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004," *J. Geophys. Res. Atmos.*, vol. 112, no. D5, p. D05208, 2007.

[4] S. Mallick *et al.*, "Performance characteristics of the NLDN for return strokes and pulses superimposed on steady currents, based on rocket-triggered lightning data acquired in Florida in 2004–2012," *J. Geophys. Res. Atmos.*, vol. 119, no. 7, p. 2013JD021401, Apr. 2014.

[5] S. A. Fleenor, C. J. Biagi, K. L. Cummins, E. P. Krider, and X.-M. Shao, "Characteristics of cloud-to-ground lightning in warm-season thunderstorms in the Central Great Plains," *Atmos. Res.*, vol. 91, no. 2–4, pp. 333–352, 2009.

[6] D. Zhang, K. L. Cummins, and A. Nag, "Assessment of Cloud Lightning Detection by the US National Lightning Detection Network using Video and Lightning," in *95th American Meteorological Society Annual Meeting*, 2015, vol. 2013, pp. 1–6.

[7] Y. Zhu, V. A. Rakov, M. D. Tran, and A. Nag, "A Study of National Lightning Detection Network Responses to Natural Lightning Based on Ground-Truth Data Acquired at LOG with Emphasis on Cloud Discharge Activity," *J. Geophys. Res. Atmos.*, p. 2016JD025574, Jan. 2016.

[8] W. Schulz, B. Lackenbauer, H. Pichler, and G. Diendorfer, "LLS data and correlated continous E-Field measurements," in *Lightning Protection (VIII SIPDA), 2005 International Symposium on*, 2005.

[9] W. Schulz and M. M. F. Saba, "First Results of Correlated Lightning Video Images and Electric Field Measurements in Austria," in *Lightning Protection (X SIPDA), 2009 International Symposium on*, 2009.

[10] W. Schulz and G. Diendorfer, "Flash Multiplicity and Interstroke Intervals in Austria," in *Lightning Protection (ICLP), 2006 International Conference on*, 2006, pp. 402–404.