2014 23rd International Lightning Detection Conference
18 - 19 March • Tucson, Arizona, USA
5th International Lightning Meteorology Conference
20 - 21 • Tucson, Arizona, USA

Validation of a ground strike point identification algorithm based on ground truth data

Stéphane Pédeboy
Météorage
Pau, France
sp@meteorage.com

Wolfgang Schulz
Aldis
Vienna, Austria
w.schulz@ove.at

Abstract*—Lightning locating systems (LLS) can provide flash data derived from individual return stroke based on a flash grouping algorithm. However the latter considers negative cloud-to-ground (CG) flashes striking the ground in a unique point represented by the location of the first return stroke. According to video observations flashes have often different ground strike points. This can be a limitation in some engineering applications like the lightning risk assessment where the actual number of ground contacts is an important parameter. To get around this limitation Météorage has developed an algorithm allowing the identification of the location of the ground strike points (GSP) based on a statistical clustering ('k-means') method. The effectiveness of this algorithm, using operational LLS data, is tested on a total of 227 negative CG flashes observed with high speed video cameras in Austria and in France, in 2012 and 2013 respectively. The comparison between GSP computation and video observations reveals a GSP detection efficiency (DE) of about 95%. In addition the algorithm is able to discriminate between strokes creating a new ground contact (NGC) or using a pre-existing channel (PEC) in 83% out of the 767 observed strokes. The analysis shows that the limitation of the model is highly depending on the DE and location accuracy (LA) of the LLS collecting the data. Nevertheless, the fairly good results obtained with the GSP identification algorithm permits to build from existing VLF/LF LLS lightning data a hierarchical interlocked data structure composed of chronological events starting with the flash as the root event which is composed of GSPs being containing themselves strokes. This new dataset describes in a more complete way some lightning parameters related to a flash (e.g. flash multiplicity and number of ground strike points per flash) and their individual relationship, giving room to the improvement of engineering and research applications.

Keywords— lightning locating systems; ground strike point or terminations; video measurements; detection efficiency; flash data;

I. INTRODUCTION

Lightning Locating Systems (LLS) operating in the VLF/LF range locate the individual return strokes occurring in a cloud-to-ground (CG) lightning flash. Using a spatial and temporal clustering algorithm it is possible to group these strokes in a consistent chronological sequence containing one or several individuals assigned to the initiating root event so called 'flash', the luminous phenomenon that can be observed during thunderstorms [Cummins et al, 1998]. This method is simple and easy to implement but it produces data suffering from the assumption that all the strokes in a flash use the same ionized channel created by the first return stroke. Obviously this is not the case in nature where about half of the downward negative CG flashes exhibit multiple ground strike points (GSP), resulting in a average number of GSP per flash ranging from 1.5 to 1.7 GSP/flash with a mean separation distance in the order of 1.8 km [Rakov and Uman, 2003; Tottappillil et al, 1992; Fleenor et al, 2008; Saba et al, 2006; Stall et al, 2009]. The terrain and the local meteorological conditions seem also to highly influence this parameter [Cummins, 2012] making it difficult to correct with an averaged factor derived from high speed video records analysis.

Due to the growing use of lightning data in engineering applications there is great interest in providing the most accurate lightning information. The lightning protection community uses the lightning density derived from LLS flash data for the lightning risk assessment because it can provide high spatial resolution data. Unfortunately with the flash data being the first return stroke position it is likely that the computed risk might be actually underestimated by a factor of about two, which is roughly the mean number of GSP per negative CG flash in nature. Because of the recent improvements of performance in the lightning detection technology, Météorage has developed a GSP identification algorithm capable of detecting and locating individual ground contact points from flash and stroke data. It is based on a clustering data analysis technique, so called the "k-means method" which uses a local search approach to partition the stroke locations into clusters representing GSP locations.

The perspective offered by this method is to introduce a new lightning flash parameter, the GSP being today the missing link between flash and stroke data. Indeed the whole lightning flash entity measured by VLF/LF LLS can be define as a hierarchical interlocked data structure composed of flash, GSP and strokes data organized in a tree like structure. The root of the structure is the flash itself encompassing all lower layers. The second layer is the GSP that is a branch of the lightning flash made of one or several subsequent return strokes. Finally at the third layer are the return strokes possibly creating a new separated channel or using a pre-existing one. This categorization can be done by the GSP algorithm introducing a new stroke classification in the overall lightning flash dataset.

Promising first results were obtained during a validation of the algorithm based on a small dataset of high speed video records collected around Tucson and in Austria [Pédeboy, 2012a]. Those results motivated us to run further analysis on a larger video dataset with LLS data based on technology which includes the latest improvements in location accuracy (LA) to confirm the first results and validate the reliability of such a technique. This report presents in details the methodology and the results based on a cross comparison of GSP computed data and video observations. Several parameters are computed like the GSP detection efficiency (DE) highly depending on the stroke DE, determining the capability to count the correct number of ground terminations, derived from the latter result the number of GSP per flash. Further the efficiency of the discrimination between strokes creating a new ground contact (NGC) or using a pre-existing channel (PEC) is investigated. The results are then analyzed and discussed in perspective with some applications of the GSP data in the near future.

II. PRESENTATION OF THE VALIDATION METHOD

The general idea of this work consists of cross comparing the GSP data computed from operational LLS data with GSP data derived from high speed video camera records which are assumed to be a ground truth reference dataset. A detailed analysis of the results permits to focus on crucial parameters which are the GSP detection efficiency, GSP multiplicity and the NGC/PEC discrimination effectiveness.

The first step of the method consists of building the reference ground truth dataset from flash video records. Video are visualized to precisely identify the individual return strokes occurring in a flash, measure their time of occurrence down to the millisecond and determine whether it produces a NGC or it uses a PEC created by a previous stroke. After analysis a stroke may then be either categorized as 'NGC' or 'PEC'. In addition to this categorization, the order of ground contacts in a flash is determined according to the chronological time of occurrence. So it is possible to know for any observed stroke to which GSP in the flash it belongs to. Therefore the video data consists of a list of accurately time stamped and categorized strokes serving as ground truth reference. A practical example is presented in table 1 with a flash composed of four strokes occurring on the 18th of June at 23:42:29. The strokes are time sorted according to their order, the first stroke being at the top. Looking at the column labelled 'Video data' one can see the first three strokes are termed as NGC1 to NGC3 meaning they created each a new ground termination. The last stroke is flagged as PEC3 meaning it used the pre-existing channel created by stroke #3.

Table 1- Example of GSP algorithm output

Date	Time	Nano second	Longitude (deg. Dec.)	Latitude (deg. Dec.)	Peak Current (kA)	Video data	GSP algo results
18/06/2013	23:42:29	880078199	3.687	47.905	11.62	NGC1	NGC1
18/06/2013	23:42:29	887845715	3.661	47.892	-49.91	NGC2	NGC2
18/06/2013	23:42:29	905166174	3.660	47.909	-14.19	NGC3	NGC3
18/06/2013	23:42:29	909970874	3.664	47.909	-8.1	PEC3	PEC3

The second step of the method consists of requesting the operational LLS database to retrieve the flash and stroke data corresponding to the reference dataset. Then the GSP algorithm processes the LLS data to determine the ground terminations in every flash of the list and assigns each stroke with a label following the same rule as the one presented here above in the video analysis (see table 1 column GSP algo results).

Finally the comparison consists in a simple row by row check determining whether the computed GSP data match with the video observations or not. From this analysis it is possible to compute some statistics, e.g. the GSP detection efficiency, multiplicity and NGC/PEC discrimination effectiveness describing the real capability of the algorithm to produce reliable data.

III. DATA USED IN THE VALIDATION METHOD

This work relies on data coming from Austria and France. ALDIS and Météorage, respectively the Austrian and French national LLS operators, operate compatible systems manufactured by Vaisala which are updated with the latest technology improvements resulting in high quality lightning data.

In addition, they have developed a mobile Video and E-Field Recording System (VFRS) to collect ground truth data for their LLS performance validation. This system was introduced in Austria in 2004 whereas it is new in France since 2012. The same type of system was used in Brazil and presented by Saba in [Saba et al, 2006]. It consists of a GPS time synchronized high speed video camera coupled with a flat capacitive antenna measuring the local E-Field changes continuously. Although both systems are based on the same principles and use the same E-field antenna, the cameras are different. The camera operated in France is a Phantom MIRO M310 running at a rate of 7000 fps with a resolution of

768x480 pixels automatically triggered by a lightning detector. The recorded video sequences are transferred from the camera memory to a connected laptop. The camera used in Austria runs at 200 fps with a resolution of 640x480 pixels [Schulz and Saba, 2009]. Both companies have hired a storm chaser in charge of tracking and observing thunderstorms with the VFRS in their countries. Furthermore each LLS operated in these countries shows a validated good performances in terms of DE and LA which are key features for the effectiveness of the GSP identification using a clustering method relying on object separation distances like the 'k-means'.

The reason to use observation data coming from different areas is to limit regional effects, being related either to the meteorology, the terrain or the LLS performance. All those regional effects could bias the results. It must be noted that the lightning data used in this study are operational data used on an as-is basis meaning the results of the GSP identification are obtained under normal operations (no reprocessing).

A. Video records dataset

A total of 227 negative CG flashes exhibiting 767 strokes were observed during 12 thunderstorms, between June to September, in several parts of Austria in 2012 and France in 2013. Approximately one third of the flashes were observed in Austria. After analysis 397 GSP could be clearly identified, again one third in Austria. It must be noted that about 22% of the total flashes are single stroke flashes ranging from 17% in France up to 29% in Austria. This latter result is significantly higher than the percentage usually reported in literature [Rakov, 2007] but could be explained because of the specific local conditions in Austria. The average number of observed GSP per flash is 1.75 with a maximum of 2.02 reached in the mountainous region of the Mont Ventoux (France) on July.

The average distance between observers and lightning flashes was 25 km, ranging from 2 km up to 45 km. The observers focused on CG flashes only and therefore they often time managed to record the area between the cloud base and the ground resulting in full scale stroke images.

However, it is sometimes difficult to get a clear field view of the ground when hills, buildings or any other object are placed between the observer and the lightning flash. In this case, the sudden flash of light produced by the return stroke tells if there is an attachment to the ground or any other assets. In addition the geometry of the channels provides information whether or not a GSP is created because it is assumed that a subsequent stroke which is using exactly the same path to ground than the previous stroke is likely to have the same ground strike point. Unfortunately, when a video is not clear enough because of heavy rain shadowing the lightning flash or a bad position of the observer, the data is discarded since it is considered as not enough reliable and subject to errors during the comparison with the lightning data.

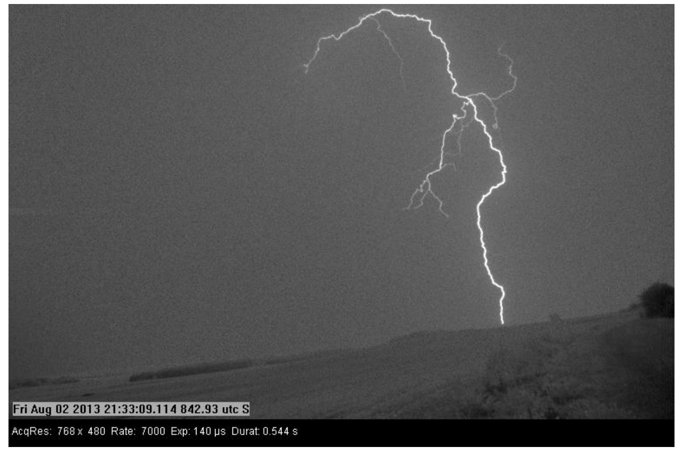

Figure 1 – Example of return stroke in France

Fig. 1 shows an image taken from a video made in France illustrating the brightness and the quality of the records. The operational settings of the camera are displayed in the lower left corner with the date and time of the image used to determine the stroke time occurrence. Indeed, with observation distances smaller than fifty kilometers the timing error committed due to propagation delays is less than 170 µs making the time recorded by the camera directly useable for strokes time stamping without any correction for propagation time of the field due to the distance.

Nevertheless despite the very good resolution of the images it may happen that a GSP cannot be detected on the video. This can be the case when the observation distance is so big that simultaneous GSP which are closely located to each other are drowned in the brighter channel, if the strokes are aligned with respect to the observer so the closer one hides the one behind. These cases are expected to be relatively rare and do not produce a significant bias in the video data reference.

B. LLS lightning data

ALDIS and Météorage operate both the latest technology in lightning location based on LS700X sensors, manufactured by Vaisala [Cummins et al, 2011; Vergeiner et al, 2013; Pédeboy, 2012b]. Arrival times and directions of the signals measured by the sensors are combined by the location algorithm to locate the corresponding strokes which are afterwards grouped with the flash grouping algorithm [Honna et al, 2011a]. Since 2011, both systems have been using the sensor based onset time correction and time propagation corrections leading to a median LA error estimated by several studies to range from about 100 m in Austria to 200 m in the US NLDN depending on sensors baselines [Schulz, 2009; Pédeboy, 2012; Honna et al, 2011b].

Out of the 227 negative flashes and 767 strokes of the reference dataset, a subset of 221 flashes and 691 strokes could be correlated with lightning data resulting in an overall flash and stroke DE of 97% and 90% respectively. The Austrian network performs very well exhibiting a 100% flash DE and 91% stroke DE. The French LLS performance is a bit lower because of bigger sensor baselines but is also very good, since the flash DE is 97% and stroke DE is 90%. This demonstrates the good performance of the LLS and thus the quality of LLS data.

From the LLS data subset mentioned above, the GPS algorithm computed a total of 379 GSP.

IV. RESULTS

The results obtained from the cross comparison between computed and observed ground contacts were at first analyzed with a quantitative approach focusing on the total number of GSP being detected at the flash scale. From this analysis it is possible to derive the GSP DE as the total number of computed GSP related to the total number of observed GSP. From this global count it is possible to determine the computed GSP multiplicity as the total number of computed ground contacts divided by the total number of flashes. Both parameters determine the capability of the algorithm to produce reliable data for statistical analysis. In a second step, an analysis at the stroke scale with a qualitative approach permits to check the capability of the algorithm to discriminate strokes between NGC and PEC. The algorithm is considered to fail when the stroke is not exactly consistent with the video observation implying that a perfect match on the type of the stroke (NGC or PEC) and on the GSP order is expected. This is a very strict criterion because the important thing for the algorithm is the capability to discriminate between NGC or PEC regardless the stroke position in the sequence of a flash. The results are presented here below in the three following sections

a) *Determination of the GSP DE*

The results obtained after comparison between computed and observed GSP are presented in table 1. It summarizes the number of flashes, GSP and strokes observed on videos and obtained from the GSP algorithm computation. The ratio of the figures in the second and first row gives the corresponding DE in the third row. The following row refers to the GSP algorithm and gives the count of flashes with correct, overestimated or underestimated computed number of GSP. The last row presents the number of observed or computed GSP per flash. These results are presented in columns for Austria, France and the whole study region.

According to table 1, the absolute GSP DE obtained from the total dataset is a bit higher than 95% which is a surprisingly good result meaning the error committed on the real number of ground termination count is less than 5%. It is important to note, that this result includes not only the failure of the algorithm to identify the GSP but also the undetected strokes by the LLS as well, describing so the absolute GSP DE. The real performance of the algorithm can be assessed when the undetected strokes are removed from the math, leading to a 97% GSP DE.

Table 1 - Detailed results at the flash scale

Parameter	Austria	France	Total
Number of observed flashes/GSP/strokes	82/126/254	145/271/513	227/397/767
Number of computed flashes/GSP/strokes	82/123/232	139/256/459	221/379/691
DE (%) Flashes/GSP/strokes	100/97/91	96/94/89	97/95/90
Flashes with correct/ over/underestimated GSP	71/7/4	107/22/10	178/29/14
Video/computed GSP per flash	1.54/1.50	1.87/1.84	1.75/1.71

A regional analysis shows a better result in Austria compared to France, with a GSP DE of 97 and 94% respectively. Looking at the results obtained at the different observation sites in France it is possible to determine the LLS DE and the GSP counting errors on these specific areas. The size of the reduced datasets collected in these small regions range from 10 to more than 50 flashes. Of course the smaller the data set the less relevant the result is, but this rough analysis determines that one can expect a 5% error in the GSP counting when LLS operates at least with a 96% flash and 83% stroke DE. The error is increasing up to 10% when the performance drops down to 90%/75% flash/stroke DE. Having in mind the general performance of modern LLS [Nag et al, 2013; Vergeiner et al, 2013], one can expect an error smaller than 10% on the total number of GSP.

b) *Number of GSP per flash*

The analysis shows (see table 1) that the mean computed number of GSP per flash is 1.71 compared to the 1.75 in the video observations. At a regional scale the observed and detected mean number of GSP per flash are respectively 1.54 and 1.50 in Austria and 1.87 and 1.84 in France. This discrepancy is currently unclear but could be related to the high single stroke flash rate in Austria.

The error committed on the average GSP per flash is about 2% on the total dataset ranging from 2.4% in Austria and 1.5% in France. With such small errors one can consider the computed GSP data match very nicely with the ground truth data. This is of a great interest since it supports the hypothesis that the GSP values provided by LLS can be useful in lightning statistics.

Looking into more detail, it is noticeable that the algorithm manages to compute the exact number of GSP per flash for 81% of the detected flashes whereas this parameter is overestimated in 13% and underestimated in 6% of the remaining flashes. The main cause of underestimation is not only due to undetected strokes but also to a wrong stroke assignment as PEC instead of NGC because of stroke location errors. The overestimation is mainly related to stroke location accuracy issues. Indeed, strokes that are poorly located tend to bias the GSP algorithm which computes a NGC instead of a PEC. It is interesting to note that most of the misclassified strokes do no exhibit big semi major axis of confidence (SMA) or sensor measurement discrepancy (CHI^2) making it difficult to fix this classification issue.

A regional analysis shows that the percentage of flashes with the correct number of GSP per flash is 87% in Austria compared to 77% in France. The better overall performance of the ALDIS LLS and the high proportion of single stroke flashes explain the better results obtained in Austria. However it is interesting to note the percentage of overestimated flashes is of the same order in France (6%) and in Austria (5%). Both findings are related to the fact that the DE is obviously different but the relative LA very similar on both systems.

c) NGC/PEC discrimination

Table 2 also summarizes the regional and global results of this work into detail for several parameters. A total of 636 strokes were perfectly classified as either NGC or PEC resulting in a stroke type discrimination effectiveness of 83% relative to the total observed strokes. The criteria used to determine the discrimination effectiveness is conservative as it takes into account not only the fact the stroke creates a channel or not, but also if it is assigned with the correct GSP order. For instance a NGC assigned with a wrong order will be considered as a misclassified. It must be noted that practically most of the GSP algorithm problems are related to NGC/PEC misclassification.

The previous already good result is increasing up to 92% when the undetected strokes are removed from the calculation.

Table 2- Detailed results at the stroke scale

Parameter	Austria	France	Total
Number of correct stroke type	212	424	636
Stroke type effectiveness	83% (91%[a])	83% (92%[a])	83% (92%[a])
Number of fully consistent flashes	55 (24[b])	71 (24[b])	126 (48[b])
Fully consistent flash DE	67% (67%[a])	49% (51%[a])	56% (57%[a])

[a] In this result the undetected flashes and their corresponding GSP were removed from the calculation

[b] Number of unique stroke flashes

It is interesting to note that 56% (126/227) of the total flashes including 48 single stroke flashes were completely consistent with the observation meaning the GSP algorithm succeeded to be in total agreement with the video data.

Contrary to the GSP DE, the stroke type discrimination does not vary with the country and so the LLS. Indeed the effectiveness is the practically the same in Austria and in France as shown in table 2. This result might be explained by similar LA errors in both systems operated by ALDIS and Météorage.

V. DISCUSSION

The results obtained from the GSP algorithm can be considered as very good with an error of 2%, on the number of GSP per flash computation and less than 10% error on the NGC/PEC discrimination. Since these errors are mostly due to a lack of performance in the LLS it is crucial to know operational parameters like LA and flash/stroke DE including the IC/CG discrimination before using the GSP algorithm. It must be noted that flash, GSP and strokes are phenomenon connected to the same event. A flash may produce one or several GSP that may be composed of one or several strokes. As a result, GSP DE is bounded between flash and stroke DE. Therefore it is possible to assess the GSP DE knowing the flash and stroke DE of a given LLS.

a) Effect of DE issues on the GSP data

Due to the limited stroke DE one or more strokes may be absent in the chronological strokes sequence of a flash leading to either an undetected GSP or a NGC/PEC discrimination error. Therefore the IC/CG discrimination errors by the LLS cause also GSP DE issue since the entire stroke sequence can be broken because the flash grouping algorithm excludes any IC from a CG flash. Finally IC/CG discrimination and DE issues produce in our analysis a similar effect because a misclassified stroke is treated like an undetected stroke. Of course flashes with a low stroke multiplicity are very sensitive to this effect and therefore the statistics derived from a LLS dataset containing numerous single stroke flashes can exhibit big errors. Another effect of such holes in the stroke sequence results in an NGC/PEC discrimination error, e.g. a real PEC becomes NGC when the stroke creating the channel is undetected.

However DE issues affect the GSP identification algorithm to a lesser extent when using modern LLS, compared to stroke location errors.

b) Effect of LA issues on the GSP data

Location errors generally depend on the number of sensors reporting which is normally related to the stroke peak current: the weaker the current the bigger the location error. The drawback is that strokes with small peak currents usually produce in PEC whereas location errors are likely to force the GSP algorithm to separate strokes from their original cluster materializing the GSP resulting in the creation of a fake new channel resulting in an overestimation of the GSP count. To get around this issue the GSP algorithm uses the SMA to weight the influence of a less accurate stroke location in the clustering. In addition, strokes exhibiting peak currents lower than 6 kA are forced to be classified as PEC. It is known from video studies that strokes with a stroke order equal or greater than 5 are likely to use a pre-existing channel [Fleenor et al, 2008]. However due to the limited stroke DE of LLS the order of a stroke may not be as reliable as its peak current measurement reason why this latter parameter is preferably used in the algorithm. However this solution is not always efficient to handle correctly strokes with a high order or a low peak current and may produce sometimes mismatches. Further analysis might be necessary to improve this point using both parameters.

In addition, stroke location errors make it more difficult to detect closely spaced GSPs from LLS data than the widely spaced ones. When the separation distance between two GSP is in the range of the LLS LA they are likely to be merged together. In the GSP algorithm there is a parameter which limits the resolution of the clustering. It is set to 500 m corresponding to two times the mean location errors expected in modern LLS. This latter value is a bit higher according to the latest performances validation in some very performing system studies [Schulz et al, 2014]. However, we recommended keeping a kind of security factor related to this distance to limit the GSP overestimation due to clustering errors. We think that the number of closely branched flashes is not very large and therefore we do not need to reduce the GSP algorithm resolution.

Instead of using the 'k-means' method which cannot identify GSP in closely spaced ground contact flash cases it might be interesting to use an identification method based on the physical stroke parameters analysis. Such a method, based on a linear discriminant analysis, has been developed using the US NLDN data with good results [Cummins, 2012]. One of the advantages of this method in the closely spaced ground terminations is the stroke classification. This work intended to test the integration of the discriminant analysis in the 'k-means' GSP identification algorithm to determine its possible benefits when the GSP separation distance is in the order of the algorithm resolution. Unfortunately, the linear parameters being computed with the US NLDN data are not fully applicable to data used in this study because of signal rise time inconsistencies due to the different ground conductivity between the US, France and Austria. As a result it was not possible to run this test before new linear discriminant parameters are determined according to the regional lightning datasets. This work could not be carried out on time before releasing this paper but it is still interesting to test and eventually integrate this method in the Météorage's GSP algorithm for the specific cases of short GSP separation distances.

Of course, flashes exhibiting simultaneous terminations to the ground, whatever the GSP separation distance is, are systematically underestimated by both methods since the LLS cannot discriminate simultaneous strokes and will locate one stroke, in the best case.

Nevertheless, the results from this work tends to prove the relevancy of using a spatial clustering method to identify ground terminations based on LLS data collected with modern and performing LLS. That is the flash data could benefit from the GSP algorithm enabling to deliver a complete and organized lightning flash structure made of flash, GSP and stroke data to engineering or research applications.

VI. CONCLUSION

The flash grouping algorithm used by most of the LLS world wide implements a simple flash model that groups strokes based on spatial and temporal criteria. It assumes that all the strokes use the same channel created by the first stroke. As a result, the flash location is assigned to the location of the first return stroke of the sequence. This could be a limitation in some engineering applications like lightning risk analysis. To get around this issue, Météorage has developed a GSP identification algorithm using a spatial clustering method based on LLS lightning data. The latter introduces a new lightning flash parameter, the GSP being the missing link between flash and stroke in the current LLS data making it possible to completely describe a lightning flash measured by a VLF/LF LLS as a hierarchical interlocked data structure. A lightning flash consists of a number of GSP which are themselves composed of one or several strokes.

This work investigates the capability of the algorithm to correctly identify and detect GSP in flashes based on high speed video camera data sets collected in Austria and France. Several parameters such as the GSP count, number of GSP per flash and the NGC/PEC discrimination effectiveness are analyzed. A quantitative approach permits to assess the committed errors on the computed data. Those errors are 5% on the total GSP count and 2% on the number of GSP per flash. This result demonstrates the capability of the algorithm to produce relevant data on ground terminations resulting in reliable statistics. Further a qualitative approach at the stroke scale determined the capability of the algorithm to discriminate the type of stroke, meaning producing NGC or using PEC. In 83% of the cases the type of the stroke was correctly assigned, reaching up to 91% when the undetected strokes were removed from the comparison.

The good results obtained in this study are linked to the high quality of the LLS used that exhibit a mean flash DE of 97%,a stroke DE of 90% and a LA better than 250m. This demonstrates the importance of the LLS data quality in getting reliable results when using a spatial clustering algorithm for GSP identification. However, a rough analysis shows that good results can be obtained with most of the modern LLS in operation worldwide.

Errors in GSP computed data are essentially due to the limited performance of the LLS collecting the lightning data. When the stroke location error is too big or for flashes exhibiting ground contacts with short separation distances, the use of a discriminant analysis on the stroke parameters might help improving the GSP identification. This could not be tested and validated in this work but we are still interested to carry this study out in the future.

Finally, the Météorage's GSP identification algorithm can be considered as efficient and capable to produce a new LLS derived lightning parameter which can complete the already existing LLS data and built a hierarchical data structure defining more accurately what a lightning flash is according to VLF/LF LLS measurements opening doors for improvement of several engineering and research applications. E.g. the lightning density used in risk assessment, currently based on flash data, could become more accurate in the future when GSP data for longer time periods will be available.

ACKNOWLEDGMENT

This work could not have been carried out without the video data. The authors cordially thank the two passionate storm chasers, Xavier Delorme (http://www.temps-orages.pro/) and Christian Vergeiner, for sleepless nights and their professionalism in gathering so useful data. We do not forget about Paul Barnéoud, a French colleague at Météorage, for his patience and precision in analyzing the huge French video dataset. We are also very grateful to Ken Cummins for his unfailing availability and his precious comments.

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(D8), 9035–9044, doi:10.1029/98JD00153.

Cummins K. L., M. J. Murphy, J. A. Cramer, W. Scheftic, N. Demitriades, A. Nag, 2010: Location accuracy improvements using propagation corrections: A case study of the US NLDN, 21st ILDC.

Cummins, K.L., N. Honma, A.E. Pifer, T. Rogers, M. Tatsumi (2011), Improved detection of winter lightning in the Tohoku region of Japan using Vaisala's LS700x Technology, 3rd Intl. Symp. on Winter Lightning, Sapporo, Japan, June15-16, 2011.

Cummins K. L., "Analysis of multiple ground contacts in cloud-to-ground flashes using LLS data: the impact of complex terrain", International Lightning Detection Conference, April 2012.

Fleenor S.A., C. J. Biagi, K.L. Cummins, E.P. Krider,"Characteristics of cloud-to-ground lightning in warm-season thunderstorms in the great plains", ILDC2008.

Honna H., K. L. Cummins, M. J. Murphy, A.E. Pifer, T. Rogers, 2011a: Improved lightning locations in the Tokohu region of Japan using propagation and waveform onset corrections, 3rd International Symposium on Winter Lightning (ISLW)

Honna H., K. L. Cummins, M. J. Murphy, A.E. Pifer, T. Rogers, 2011b: Improved lightning locations in the Tokohu region of Japan using propagation and waveform onset corrections, 3rd International Symposium on Winter Lightning (ISLW)

Nag A., M.J. Murphy, K.L. Cummins, A.E. Pifer, J.A. Cramer, "Performances characteristics improvements of the US National Lioghtning Detection Network", Asia-Pacific International Conference on Lightning, Seoul 2013.

Pédeboy S., "Identification of the multiple ground contacts flashes with lightning location systems", ILDC 2012a Boulder.

Pédeboy S., " Using 20 years of lightning data in ground flash density statistics in France", International Conference on Lightning Protection, 2012, Vienna.

Rakov V., Uman M., 3Some properties of negative cloud-to-ground lightning flashes versus stroke order", J. Geophys. Res., 95(D5),5447-5453.

Rakov V. A.,"Lightning phenomenology and parameters important for lightning protection", IX SIPDA 2007, Iguaçu – Brasil

Rakov V. A.,"Lightning phenomenology and parameters important for lightning protection", IX SIPDA 2007, Iguaçu – Brasil

Saba M. F., M. G. Ballarotti and O. Pinto Jr, 2006: Negative cloud-to-ground lightning properties from high speed video observations, JGR, vol 111, D03101.

Schulz W., M. Saba, "First Results of Correlated Lightning Video Images and Electric Field Measurements in Austria", SIPDA 2009

Schulz W., D. Poelman, S. Pédeboy, C. Vergeiner, H. Pichler, G. Diendorfer, S. Pack, 2014: Performance validation of the European Lightning Location System EUCLID, International Colloquium on Lightning and Power Systems, CIGRE Lyon 2014.

Stall C, K.L. Cummins, E.P. Krider, J. Cramer (2009), Detecting Multiple Ground Contacts in Cloud-to-Ground Lightning Flashes. Journal of Atmospheric & Oceanic Technology, vol. 26 (11), pp. 2392-2402, November 2009.

Thottappillil R., V. A. Rakov, M. Uman, W. Beasley, M. Master and D. Shelukhin, 1992: Lightning subsequent stroke electric field peak greater than the first stroke peak and multiple ground terminations, JGR, 97(D7), 7503-7509.

Vergeiner C., W. Schulz, S. Pack (2013), " On the Performance of the Austrian Lightning Detection and Information System (ALDIS)", 11 Höfler´s Days, 2013.