2019 12–14 JUN HK APL
The 11th Asia–Pacific International Conference on Lightning
June 12–14 Hong Kong,China
APL 2019

# Performance validation of a ground strike point algorithm

D. R. Poelman
Royal Meteorological Institute of Belgium
Brussels, Belgium
dieter.poelman@meteo.be

S. Pédeboy
Météorage
Pau, France

W. Schulz
OVE-ALDIS
Vienna, Austria

***Abstract*— An important input parameter in lightning protection studies is the lightning flash density. Lightning Location Systems (LLS) do provide flash data, with a single location allocated to each flash. However, cloud-to-ground (CG) flashes are known to exhibit one or more ground strike points (GSP). Therefore, having a tool that is able to determine the different GSP within a single flash is of great importance to correctly investigate the potential risk of lightning damage. In this study a GSP identification algorithm, developed by Météorage, is tested against high-speed video measurements in order to validate the ability to reproduce the observed GSP in the field. The ground truth data were taken in Austria (2012, 2015), Brazil (2008), France (2013-2016), Spain (2017-2018), and USA (2015) and are correlated to operational LLS data in order to extract the location, peak current estimate and other parameters serving as input for the GSP algorithm. As a result, the validation of the GSP algorithm is based on 824 flashes with a total of 2413 strokes. Averaged over all the datasets the GSP algorithm is able to identify correctly new ground contacts (NGC) in 93% of the cases, whereas 82% of the strokes following previously existing channels (PEC) were captured accurately by the algorithm. It becomes clear that the actual performance of the algorithm depends on 1) the so-called distance parameter within the algorithm itself, i.e. the distance criterion to group individual strokes within a single GSP, and on 2) the location accuracy (LA) of the LLS.**

***Keywords*— ground strike point, video measurements**

## I. Introduction

The primary input parameter in lightning risk assessment applications is the lightning flash density ($N_g$), defined as the number of cloud-to-ground (CG) flashes per square kilometer per year. In the past $N_g$ was extracted based on the amount of observed thunderstorm days $T_d$. Various relationships have been published ever since [1] correlating $N_g$ and $T_d$ depending on e.g. geographical regions, and are mostly of the following form:

$$N_g = a\, T_d^{\,b}\ [\mathrm{km}^{-2}\mathrm{yr}^{-1}] \tag{1}$$

With *a* and *b* variables. Nevertheless, not two thunderstorms are the same. Hence, the above $N_g$-$T_d$ relation only provides an initial guess about the true amount of occurred CG flashes.

Moreover, a CG flash is the combination of CG strokes that match in time and space. Present-day lightning location systems (LLS), operating in the VLF/LF frequency range, are capable to locate a large percentage (if not all) of those individual return strokes that comprise a single CG lightning flash. Furthermore, more often than not, the CG flash exhibit multiple ground strike points (GSP), with an average number of GSP per flash varying in between 1.5 and 1.7 [2-6]. However, the flash location is usually defined by the location of the first return stroke in the flash; thus not taking into account the different GSP. Hence, using $N_g$ may underestimate the computed risk to structures due to lightning flashes to Earth. Therefore, the GSP value linked to each individual flash is an important parameter to take into account for lightning risk purposes.

In this work, a GSP algorithm, developed by Météorage, is tested against distinct sets of video measurements. Note that this work is an extension of a study presented in [7, 8]. In Section II the different ground-truth datasets are examined and the quality determined. Section III briefly reports on the GSP algorithm and is followed by the results in Section IV.

## II. Ground-truth Datasets

Ground-truth (GT) data were gathered over several years and different geographical areas. Those data were obtained by high-speed video measurements taken in Austria (AT) in 2012 and 2015, Brazil (BR) in 2008, France (FR) in between 2013-2016, Spain (ES) during 2017-2018 and USA (US) in 2015. Subsequently, all the strokes in the different flashes are cross-correlated in time to observations made by the local LLS in order to assign the location, peak current estimate, semi-major axis (SMA) of the 50% confidence ellipse and other lightning parameters when provided by the LLS. In addition, the video imagery enables to distinguish whether each individual stroke creates a new ground contact point (NGC) or follows a pre-existing channel (PEC). Besides the type being an NGC or PEC, the sequence is stored as well, i.e. whether for example a particular NGC is the first or $i^{th}$ NGC in the flash or to which NGC a certain PEC belongs to. Those datasets are then further fine-tuned by excluding flashes for which the LLS did not detect all the individual strokes. The latter is done since it influences directly the GSP algorithm. In addition, positive flashes are excluded as well.

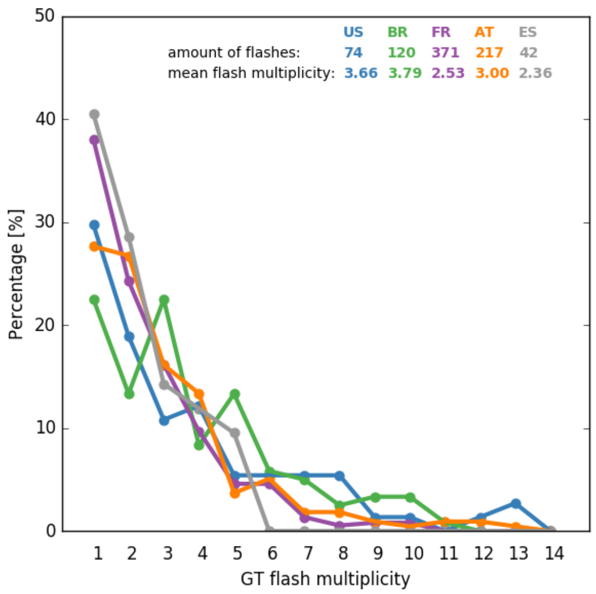


Figure 1. Flash multiplicity distribution as recorded by high-speed camera observations in different regions. The amount of flashes and the mean flash multiplicity, i.e. number of strokes per flash, for each ground truth dataset is indicated at the top of the plot.

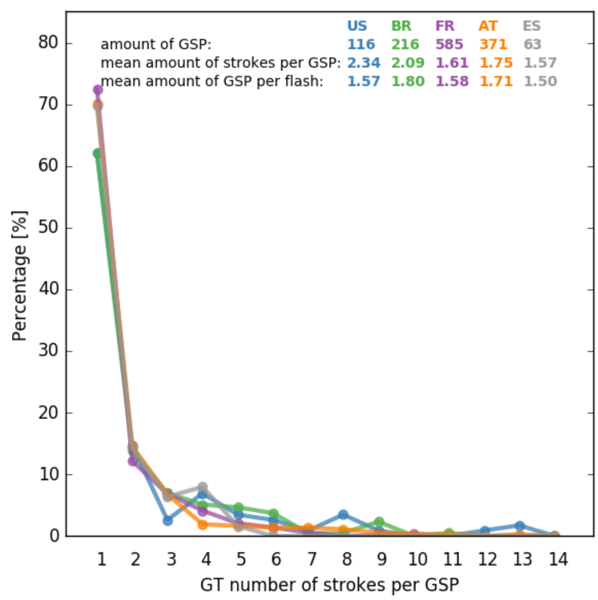


Figure 2. Distribution of the number of strokes per GSP. The amount of GSP per dataset, as well as the mean amount of strokes per GSP and the mean amount of GSP per flash for each dataset is indicated at the top of the plot.

In this way, a total of 824 negative CG flashes, containing 2413 strokes, are used to ingest in the GSP algorithm.

The individual datasets in terms of flash multiplicity and amount of flashes are depicted in Figure 1. The percentage of single stroke flashes varies in between 25% (BR) and 40% (ES). The largest dataset by far is the one of France, obtained over four consecutive years (2013-2016), and includes 45% of the total amount of flashes altogether used in this study. Mean flash multiplicities vary in between 2.36 (ES) and 3.79 (BR).

The location accuracy (LA) of the different LLS can be determined as well from the video measurements by using those strokes that follow the same channel as determined from the consecutive high-speed images. As such, these strokes are assumed to strike ground at the same point. Following the procedure by [9], the differences between the stroke positions within a flash are then computed from the position distances in the LLS data and are downscaled by $\sqrt{2}$. This scaling is necessary because both positions are subject to random errors [9, 10]. Since there is the possibility that the channel geometry and/or the actual ground contact varied slightly from stroke to stroke and was not resolved by the camera, the differences determined by this method should be regarded as upper bounds of the actual position differences. The LA for the different LLS retrieved by the above method

TABLE I. LOCATION ACCURACY [KM]

| | LLS | | | | |
|---|---|---|---|---|---|
| | ***US*** | ***BR*** | ***FR*** | ***AT*** | ***ES*** |
| mean | 0.53 | 2.0 | 0.51 | 0.6 | 0.33 |
| median | 0.13 | 0.84 | 0.15 | 0.14 | 0.09 |
| $95^{th}$ percentile | 2.76 | 8.3 | 2.24 | 3.42 | 1.54 |

is listed in Table 1. From this, it is obvious that the LA of the Brazilian LLS is not as good as compared to the other LLS. This will have an obvious effect on the output of the GSP algorithm.

As mentioned before, the video measurements permit to classify each stroke as NGC or PEC. Using this information, the amount of GSP as seen in the GT data can be determined. This is shown in Figure 2. The mean amount of GSP per flash varies in between 1.50 (ES) and 1.80 (BR). The difference amongst the datasets is related to the unequal portion of single stroke flashes within the datasets. In addition, the mean amount of strokes per GSP ranges from 1.57 (ES) to 2.34 (US).

## III. GSP ALGORITHM

The GSP algorithm (ALG), as developed by Météorage, has been described in detail in [8]. In short, it is based on a k-means method to cluster individual CG strokes belonging to a particular flash into one or more GSPs. In order to run the algorithm, the latitude, longitude and the semi-major axis (SMA) of the 50% confidence ellipse is required for each individual CG stroke within the flash. This information is retrieved from the observations of a local operational LLS. Key parameters in the GSP algorithm are 1) the maximum distance limit (DIST), i.e. the distance above which a new termination point is created, and 2) the maximum SMA. The algorithm determines the GSP in an iterative process, whereby strokes are assigned to the closest ground contact. Subsequently, the location of a GSP is calculated as the mean of the locations of the strokes that belong to it. Note that the location accuracy of each individual stroke, as indicated by the SMA, is taken into account using a weighting factor inversely proportional to the SMA to calculate the GSP position. As such, strokes that are poorly located have a limited effect on the GSP location. For more information, the interested reader is referred to [8].

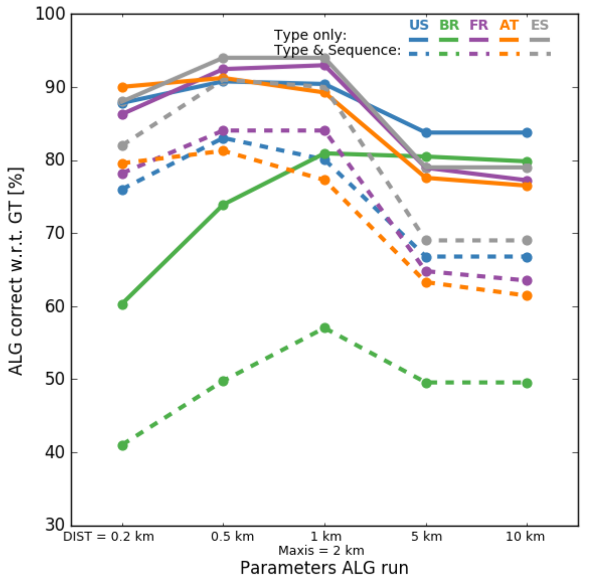


Figure 3. Ability of the GSP algorithm (ALG) to correctly predict the type, i.e. NGC versus PEC, of the CG stroke in the flash (solid lines), as well as when the sequence is taken into account (dashed lines) with respect to the different GT datasets. The input parameters of ALG are indicated at the bottom of the plot.

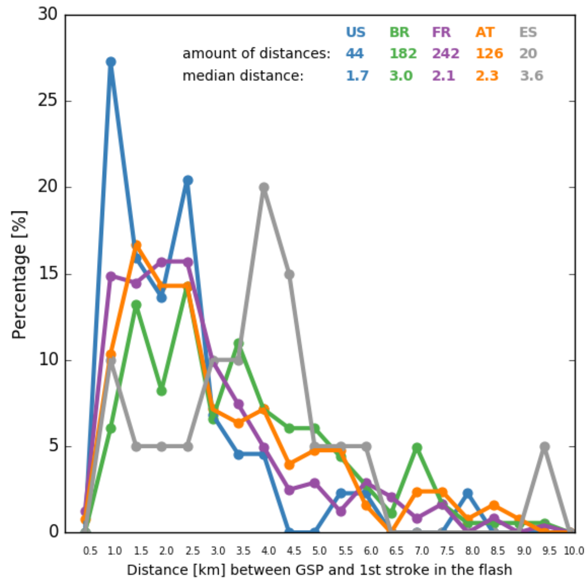


Figure 4. Distribution of the distances retrieved by ALG in between the GSP and the first stroke within the flash. The amount of distances used and the median distance retrieved for the different datasets are indicated at the top of the plot.

## IV. RESULTS

As a first step, ALG is validated while changing maximum distance allowed between a given stroke and an existing GSP. Different runs are performed while altering DIST from 0.2 km to 10 km. The effect on the ability of ALG to correctly predict the type, i.e. either NGC or PEC, and type and sequence of the stroke is plotted in Figure 3. First, the results match quite well amongst the different datasets, except for BR. This is not surprising since the LA of the LLS used in BR is lower by a factor of about four in comparison to the other LLS. As such, ALG will create more frequent NGC, whereas those strokes are most probably of type PEC. Second, choosing DIST in between 0.5-1 km results in the best performance of ALG. Combining all the datasets together leads to an overall classification success by ALG of 88.4%, with 93.3% for NGC and 82.3% for PEC. This performance drops somewhat further by 10% when in addition the sequence is taken into account. Other values of DIST lower in general the performance of ALG w.r.t. GT. In the following, results are described by setting DIST at 0.5 km, with a maximum SMA of 2 km.

Figure 4 plots the distribution of the distances between the locations of the different GSP as retrieved by ALG and the location of the first stroke within a flash. The median distance varies in between 1.7 km (US) and 3.6 km (ES). Note that the lower limit depends on the maximum stroke grouping distance DIST used in ALG, being 0.5 km in this case. The somewhat different distribution for US and ES, compared to FR and AT, can be attributed to the limited size of the datasets. On the other hand, one should keep in mind the lower accuracy with which the strokes in BR are located by the LLS. FR and AT exhibit similar behavior in the separation distance, with a median distance of 2.1 - 2.3 km.

Finally, the outcome of ALG is split into four different possibilities w.r.t. GT. The percentage of occurrence of those four possibilities is plotted in Figure 5 as a function of the median peak current Ip. Solid symbols take into account all strokes, whereas open symbols are the result when first strokes in each flash are excluded. The latter is done since first strokes are per definition correctly assigned by ALG as NGC. Let's first focus on the correctly assigned strokes by ALG w.r.t. GT, i.e. triangles (blue) and squares (green). One notices that the median Ip for strokes that create an NGC is larger than those following a PEC. Except for BR, ALG assigns NGC and PEC in approximately 90% of the cases correctly. As mentioned before, the quality of the LLS influences the results of the GSP classification which can be seen in the case of BR. In general, excluding the first strokes (open symbols) lowers the classification success by about 10 to 20%. Finally, the median Ip for strokes that are incorrectly classified by ALG is most of the time lower compared to the correctly classified strokes. This is clearly visible in the dataset of US, where for instance correctly classified NGC have higher Ip than incorrect classified NGC. Since strokes with lower peak current are in general detected by a lower amount of sensors, the accuracy to locate those strokes decreases as well. Thus the peak current can partly explain why those strokes are incorrectly classified by ALG.

## V. CONCLUSIONS

Negative CG flash data retrieved from several ground-truth campaigns in different geographical regions in the world are correlated to LLS data to deduct the corresponding discharge parameters. This information is ingested into a GSP algorithm to validate its ability to retrieve the different termination points within each flash. It becomes clear that the actual performance of the algorithm depends on 1) the so-called distance configuration parameter of the algorithm itself, i.e. the distance criterion to group individual strokes to a certain GSP, and on 2) the location accuracy of the LLS. Overall, the algorithm is able to correctly assign the type of stroke in 88% of the cases. More specifically, 93% of the NGC are determined correctly, whereas 82% of the PEC are captured accurately by the algorithm.

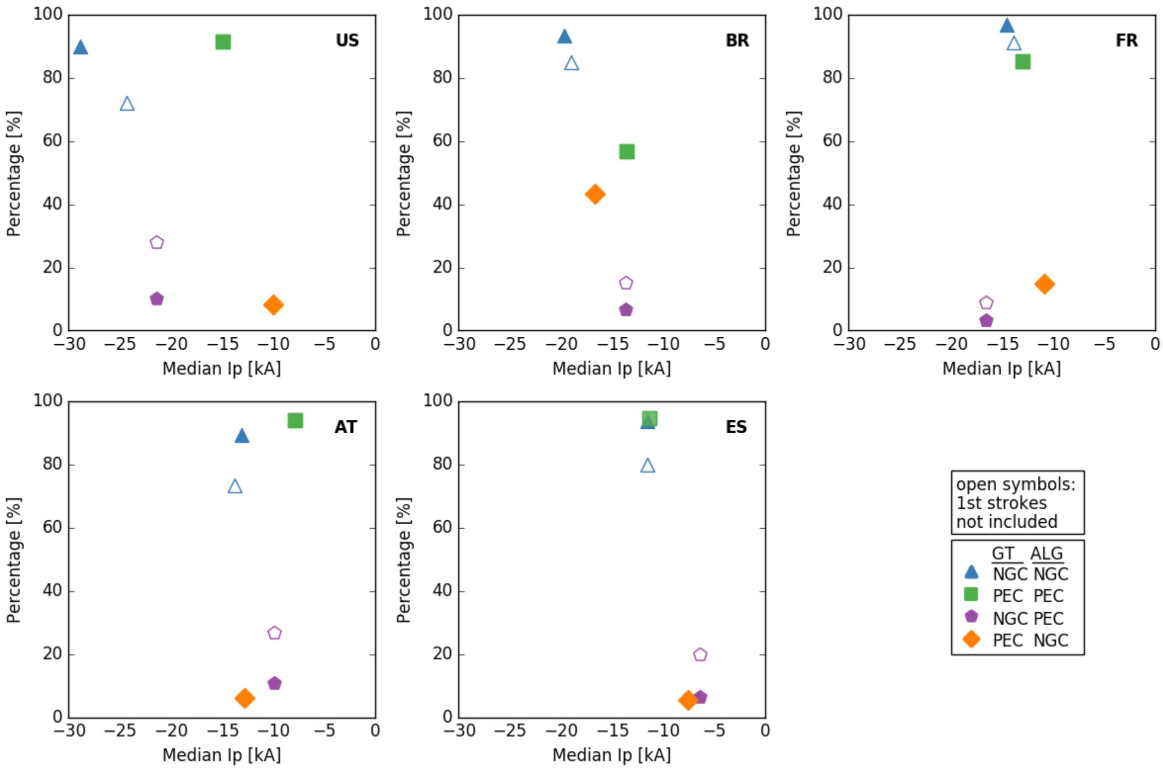


Figure 5. For each of the different GT datasets, the percentage of occurrence of the four different possibilities, indicated by the different symbols, is plotted as a function of their median peak current. Solid symbols make use of all the CG stroke data, whereas open symbols exclude the first strokes in each flash.

## ACKNOWLEDGMENT

The authors are grateful to Dustin Hill and Carlos Mata with Scientific Lightning Solutions, LLC (SLS) and Marcelo Saba from the National Institute for Space Research (INPE) for their effort to provide the ground-truth data sets of USA and Brazil used in this study.

## REFERENCES

[1] S. A. Prentice, “Frequency of lightning discharges”, in Lightning, Vol. I, Golde R. H., Academic Press, N. Y., 1997, pp. 465-496.

[2] V. Rakov, and M. Uman, “Some properties of negative cloud-to-ground lightning flashes versus stroke order”, J. Geophys. Res., 95(D5), 5447-5453.

[3] R. Thottappillil, V. A. Rakov, M. Uman, W. Beasley, M. Master, and D. Shelukhin, “Lightning subsequent stroke electric field peak greater than the first stroke peak and multiple ground terminations”, J. Geophys. Res., 97(D7), 7503-7509, 1992.

[4] S. A. Fleenor, C. J. Biagi, K. L. Cumins, E. P. Krider, “Characteristics of cloud-to-ground lightning in warm season thunderstorms in the great plains, Proceeding International Lightning Detection Conference, 2008.

[5] M. F. Saba, M. G. Ballarotti, and O. Pinto Jr, “Negative cloud-to-ground lightning properties from high speed video observations”, J. Geophys. Res., vol 111, D03101, 2006.

[6] C. Stall, K. L. Cummins, E. P. Krider, and J. Cramer, “Detecting multiple ground contacts in cloud-to-ground lightning flashes”, J. Atmos & Oceanic Technol., vol. 26 (11), pp.2392-2402, November 2009.

[7] S. Pedeboy, “Identification of the multiple ground contacts flashes with lightning location systems”, Proceeding in International Lightning Detection Conference, 2012.

[8] S. Pedeboy, and W. Schulz, “Validation of a ground strike point identification algorithm based on ground truth data”, Proceeding in International Lighting Detection Conference, 2014.

[9] 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., 112, D05208, doi:10.1029/2006JD007341, 2007.

[10] W. Schulz, C. Vergeiner, H. Pichler, G. Diendorfer, and K. Cummins, “Location accuracy evaluation of the Austrian Lightning Location System ALDIS”, Proceeding 22nd International Lightning Detection Conference, 2012.