Ground strike point properties derived from observations of the European Lightning Location System EUCLID

Dieter Roel Poelman
Royal Meteorological Institute of Belgium
Brussels, Belgium
dieter.poelman@meteo.be

Hannes Kohlmann
Austrian Lightning Detection and Information System (ALDIS)
Vienna, Austria
h.kohlmann@ove.at

Wolfgang Schulz
Austrian Lightning Detection and Information System (ALDIS)
Vienna, Austria
w.schulz@ove.at

Stephane Pedeboy
Météorage
Pau, France
sp@meteorage.com

Lukas Schwalt
Institute of High Voltage Engineering and System Performance
Graz University of Technology
Graz, Austria
lukas.schwalt@tugraz.at

Abstract*—In order to evaluate the lightning risk to a particular structure, it is common practice to follow the guidelines set out in IEC 62305-2, i.e., the reference standard for lightning risk calculation. Amongst the various components that influence the total risk, the flash density is a key parameter. However, flashes have on average more than one ground termination point. This study seeks to ascertain whether existing ground strike point (GSP) algorithms estimate correctly the actual observed number of GSPs per flash based on observations made by high-speed cameras. In addition, lightning data as observed by the European Cooperation for Lightning Detection (EUCLID) network are used in combination with a particular GSP algorithm to retrieve the temporal behavior of GSPs in two topographically different regions in Europe, i.e., Austria and Belgium, over a ten-year period from 2012 to 2021. We find that although most GSP algorithms over- or underestimate to some extent the number of GSPs per flash, this number is fairly close to the observed value as derived from the ground-truth observations. Furthermore, it is found that the average number of GSPs per flash is highest during the summer months. Finally, a diurnal trend is visible where the number of GSPs per flash is lowest between 12 and 18 UTC (Universal Time Coordinated).

***Keywords*—ground strike point, negative cloud-to-ground flash**

I. Introduction

In order to evaluate the lightning risk to a particular structure, it is common practice to follow the guidelines set out in IEC 62305-2 [1], i.e., the reference standard for lightning risk calculation. Amongst the various components that influence the total risk estimation, the flash density N_G is one of the key parameters. The latter is expressed as the number of lightning flashes per km^2 per year. Present-day lightning location systems (LLSs) are capable of locating with high accuracy, i.e., of the order of up to a few hundreds of meters, most (if not all) of the individual strokes that make up a lightning flash. Yet, the location of a flash is per definition determined by the location of the first cloud-to-ground (CG) stroke within the flash. High speed camera observations have been employed in many studies and provided proof of on average more than one ground strike point (GSP) within multiple-stroke flashes [2, 3, 4, 5]. On the other hand, a recent study [6] investigated the performance of so-called GSP algorithms to correctly determine the ground strike points by means of high-speed camera observations. It was found that the GSP algorithms perform well, with success rates up to about 90% to retrieve the correct type of the strokes in the flash, i.e., whether the stroke creates a new termination point or follows a pre-existing channel. However, an essential aspect that was not brought to light in this latter study was the ratio of the number of GSPs retrieved by the algorithms to the number of GSPs as observed in the different ground-truth data sets. This sheds light on whether a particular GSP algorithm over- or underestimates the number of ground strike points.

A description of the ground-truth and LLS data used is provided in Section II. Section III briefly explains the various GSP algorithms examined, while in Section IV the algorithm's ability to correctly retrieve the number of observed GSPs is further examined. Additionally, a GSP algorithm is applied to group the individual strokes as observed by the European Cooperation for Lightning Detection (EUCLID) over a ten-year period from 2012 to 2021, in order to retrieve GSP characteristics over a larger spatial and temporal scale in Austria (AT) and in Belgium (BE). We conclude and summarize in Section V.

II. Data

A. Ground-truth

Ground-truth data used in this study are the same as described in [5, 6]. For all the details about the data sets used, the interested reader is referred to those latter papers. Only a concise summary is provided in the following. First, it is important to point out that the flash grouping is based on the observations made by the high-speed video cameras. Second, since focus in this particular study is on the GSP characteristics on the European continent, the data sets are limited to the ones gathered in Austria (AT) in 2012, 2015, 2017 and 2018, France (FR) during 2013-2016 and Spain (ES) in 2017-2018. Eventually, a total of 904 flashes, comprising 2450 strokes, are used in this analysis of which the LLSs have detected all the strokes within the flash. Of that group of 2450 strokes, 1514 create a new GSP; leading to an average of 1.67 GSPs per flash in the combined data sets.

In addition, for each stroke, the location and, e.g., estimated peak current, is extracted from the observations made by a local LLS, being the Austrian Lightning Detection and Information System (ALDIS) network in case of the Austrian data set and the French national LLS operated by Météorage (MTRG) in case of the data sets of France and Spain. Notice that ALDIS and MTRG are similar networks, whereby *i*) sensors in both networks detect the electromagnetic radiation of lightning discharges in the low-frequency (LF) band, *ii*) the sensors are from the same manufacturer, i.e., Vaisala, *iii*) the networks have similar baselines, and *iv*) hence perform at the same level in terms of detection efficiency (DE) and location accuracy (LA). The values of DE and LA are comparable to the ones stated in Section II.B.

B. EUCLID

The European Cooperation for lightning Detection (EUCLID) consists of more than 150 sensors spread across Europe and locates cloud-to-ground (CG) strokes and intracloud (IC) pulses after applying a combined time-of-arrival (TOA) and direction finding (DF) method (https://www.euclid.org). The performance of EUCLID in terms of DE, LA and peak current estimation is tested in a continuous manner by exploiting on the one hand direct lightning measurements from instrumented towers, and video and E-field records collected in different places on the other hand. Based on the instrumented tower data, it is derived that the LA is of the order of 100 m, while the DE for negative CG strokes and flashes reaches 70% and 96%, respectively. Using video and E-field records, the DE of negative CG strokes and flashes is determined to be 84% and 98%, respectively [7, 8, 9].

III. GSP ALGORITHMS

A GSP algorithm can be used to group the observed CG strokes of a flash in one or more GSPs. In [6] the success rate of different GSP algorithms was investigated to retrieve the correct type of the strokes in the flash, i.e., whether the stroke creates a new termination point or follows a pre-existing channel. In here, the algorithms' ability to retrieve the correct number of GSPs, as observed in the high-speed video images, is examined. Before jumping to the results, the main characteristics of the different algorithms are briefly touched upon. For in-depth information, the reader may consult [6].

A. Algorithm 1 (A1)

This algorithm implements an iterative K-means method of looping chronologically through the various strokes of a flash. The location of the 1st GSP is given by the location of the 1st stroke. Successively, the distance between the following strokes and the existing GSPs is tested against a threshold value. If it is lower than the distance-threshold, the stroke belongs to the closest GSP, otherwise it creates a new GSP in case the distances are all larger than the threshold. The GSP positions are updated by the end of an iteration, whereby a weight is given to each stroke that is inversely proportional to the respective semi-major axis (SMA) of the error ellipse. In addition, a stroke is assigned to the previous GSP regardless its position when the absolute peak current $|I_p|$ is below 6 kA and/or the SMA is larger than 2 km.

B. Algorithm 2 (A2)

The initial step of this iterative K-means method is to separate strokes in a flash into two main groups, i.e., those with small and those with large SMA values, based on a user-defined threshold. Then the group with low-SMA values are

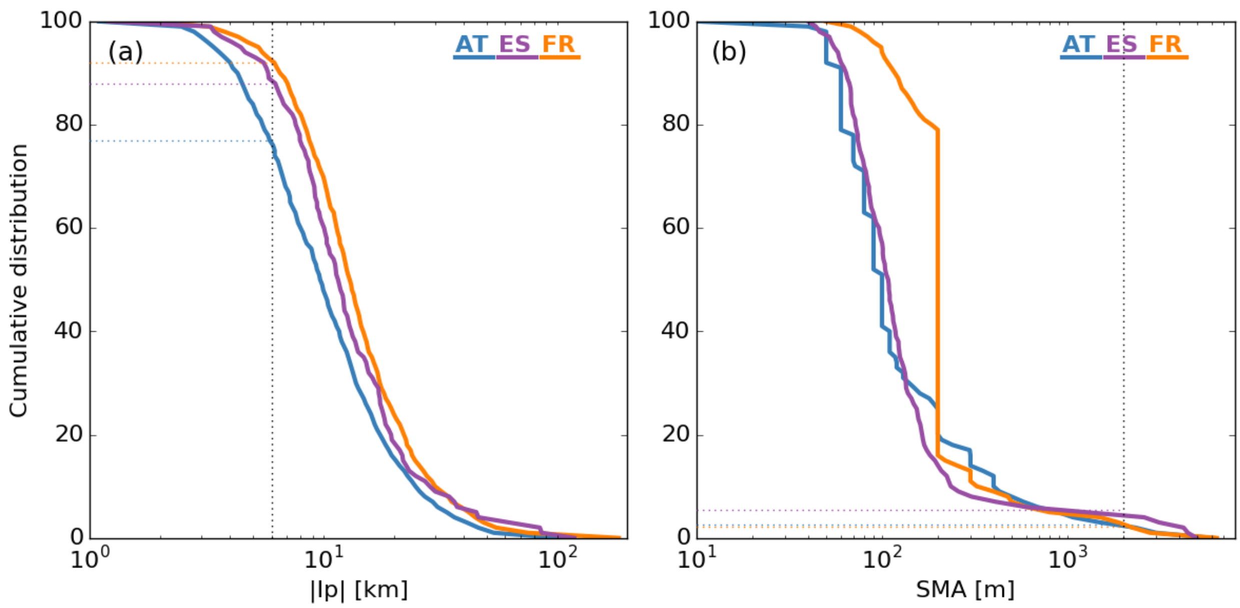

Fig. 1: Cumulative distribution for (a) absolute peak current $|I_p|$, and (b) the length of the semi-major axis (SMA) of the 50% probability ellipse. In (a) the 6 kA, and in (b) the 2 km value is indicated by the vertical dotted line.

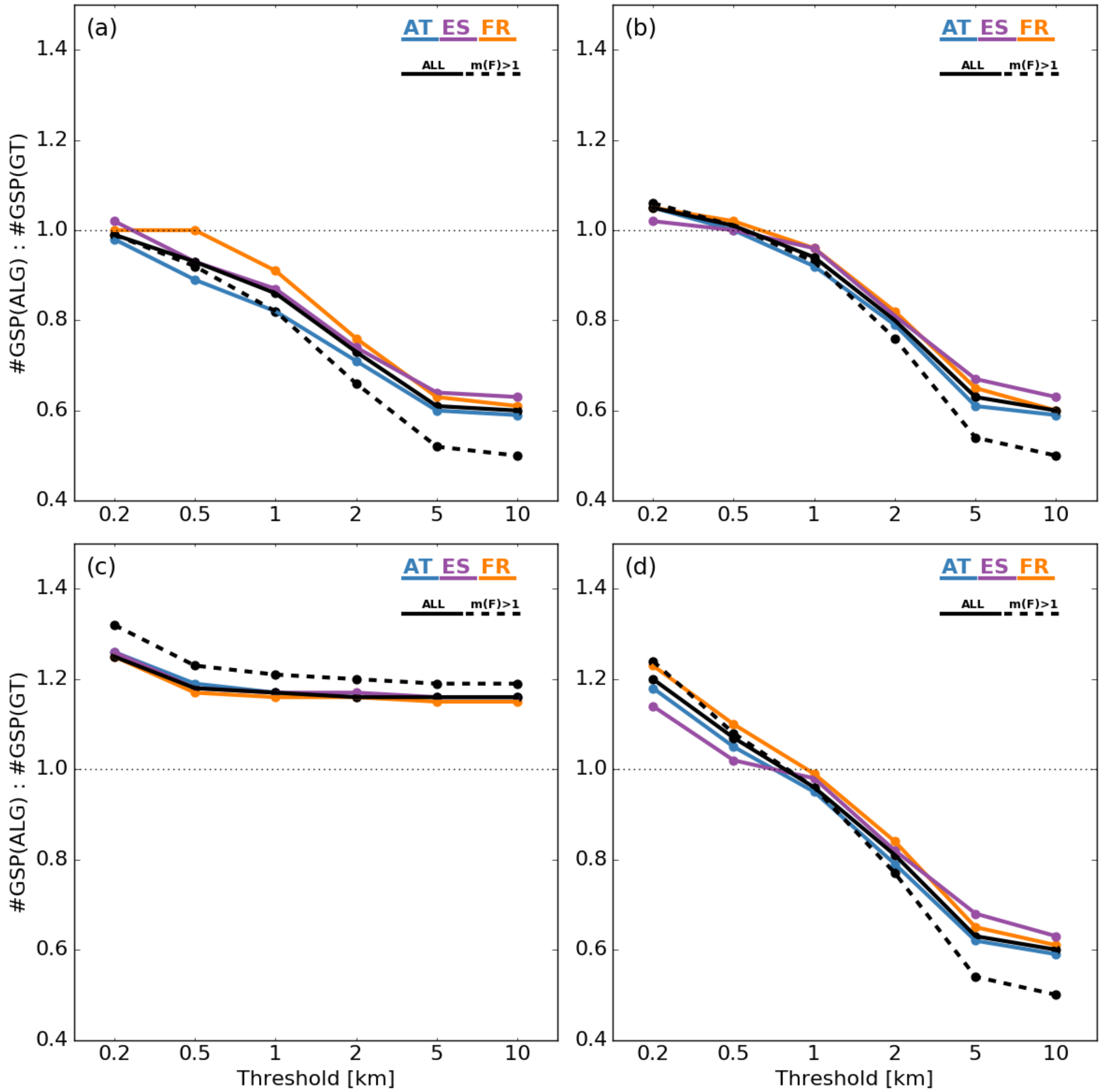

Fig. 2: The ratio of the number of GSPs as retrieved by a particular algorithm (ALG) to the number of GSPs based on the ground-truth (GT) data. The results for A1, A2, A3 and A4 are plotted in (a), (b), (c), and (d), respectively. The black solid line is the result for the combined datasets, whereas the dashed black line excludes single stroke flashes from the data sets. The latter therefore includes only flashes with multiplicity greater than one, i.e., m(F) > 1.

first grouped into GSPs based on a K-means method in a similar manner as A1. Subsequently, the algorithm attempts to group the high-SMA strokes in the flash to the existing GSPs based on elliptical scaling. If not possible, the high-SMA stroke creates a new GSP.

C. Algorithm 3 (A3)

Non-iterative method whereby a stroke is only assigned to an existing GSP when the distance falls below a certain threshold and when the 50% probability error ellipse overlaps with one or more of the other error ellipses already assigned to that particular GSP. If not, the stroke creates a new GSP.

D. Algorithm 4 (A4)

This algorithm – which was not described in [6] – consists of two steps. First, the location of the strokes is used as input in DBSCAN (Density Based Spatial Clustering of Applications with Noise, see [10]). The 'minimum number of points'-criterion of DBSCAN is set to one, meaning that a single point will create a cluster in itself with a distinct label. Any other point that lies within the distance criterion to a cluster will be assigned to that particular cluster. Second, to compute the final clusters, a K-means algorithm is used. The strokes are given as input to the K-means clustering algorithm, while the size of the error ellipses is used as the

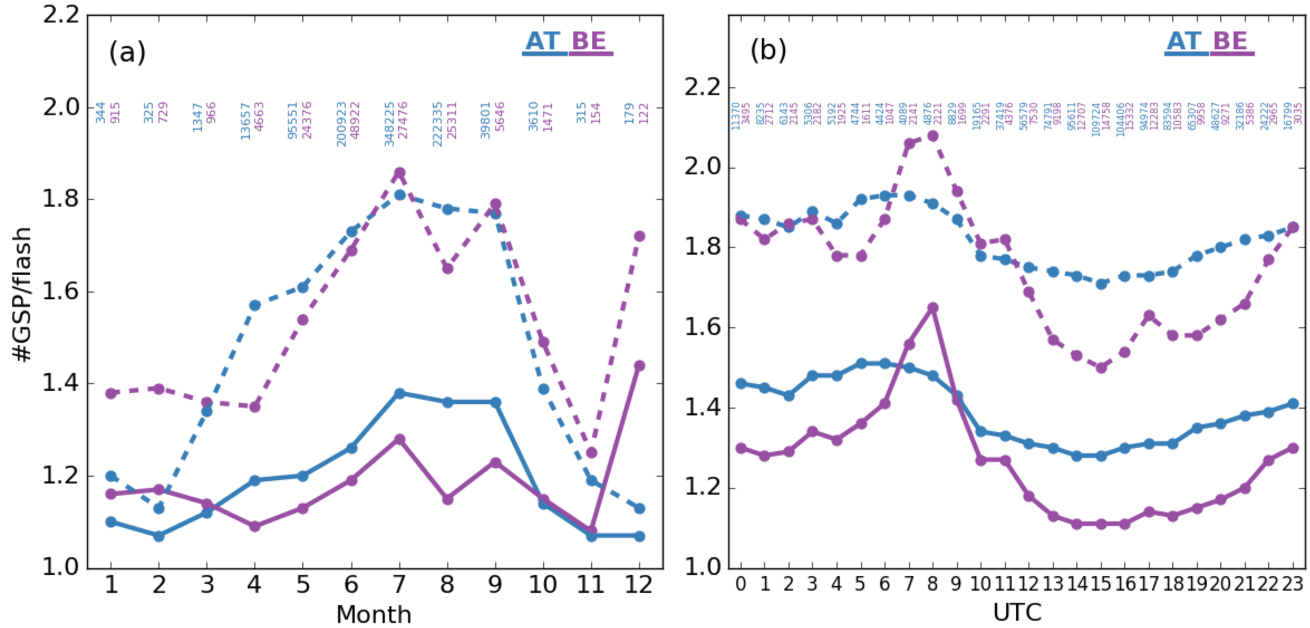

Fig. 3: (a) Average monthly number of GSPs per flash in Austria (AT) and Belgium (BE) over the period 2012-2021, based on A1. Dashed lines are the values excluding single stroke flashes from the EUCLID observations. (b) Similar to (a), but now the diurnal distribution. The actual number of all flashes per bin is listed at the top of the figure.

inversely proportional weight for the computation of the cluster centroid. At initialization, the number of distinct cluster labels, as determined in the previous step, is used as the required K-value.

Figure 1 depicts the cumulative distribution for the absolute peak current and the SMA for the different data sets. Note that the SMA distribution found for FR is caused by the data itself, i.e., for the time period 2013-2015 the SMA km-scale was provided with just one single digit after the comma, whereas in 2016 three digits were provided by the LLS operator. It can be seen that 77%, 88% and 92% of the strokes have an absolute peak current larger than 6 kA for AT, ES and FR, respectively. In case of SMA, all three data sets find only a limited number of strokes, i.e., 2-5%, with an SMA larger than 2 km. As such, since A1 and A4 are somewhat similar in nature, it is mainly the peak current of some of the strokes that will result in a difference between A1 and A4.

IV. RESULTS

The ratio of the number of GSPs retrieved by the different algorithms to the number of GSPs as observed in the high-speed video recordings is plotted in Fig. 2. A ratio greater/smaller than one indicates an over/underestimation by the respective algorithm. Since the distance threshold is a fundamental criterion common to all algorithms, the algorithm's performance (in terms of the calculated number of GSPs) is evaluated for different distance thresholds ranging from 200 m up to 10 km.

In [6] it became evident that for all algorithms the best success rate was found adopting a distance threshold of 500 m. Hence, the outcome of the different GSP algorithms are compared in the following for this threshold. From Fig. 2, it follows that for the combined data sets A1 underestimates the number of GSPs. However, applying A1 only to FR results in a ratio equal to one. This is not surprising, since A1 was designed to perform the best over France. On the other hand, A1 underestimates the number of GSPs per flash for AT for a distance threshold of 500 m; a consequence of the lower average peak current observed in Austria (see Fig. 1). Of all the algorithms, A2 performs best with a ratio equal to one at 500 m both for the combined data set as well as for the individual data sets. In general, A3 overestimates the number of GSPs per flash by almost 20%. A4 also overestimates somewhat at 500 m, though to a lesser degree than A3. Keep in mind that the stroke DE of LLSs is less than 100%. Hence, some strokes can be missed that potentially create a new GSP. The fact that A4 overestimates slightly the number of GSPs compared to what is observed in the ground-truth data sets, would compensate the latter. Hence applying A4 to LLS data in turn leads to a number of GSPs close to what actually occurs in nature.

In the remainder of the study, lightning data as observed by EUCLID are ingested into A1 to retrieve the temporal behavior of GSPs in two topographically different regions in Europe, i.e., Austria and Belgium, over a ten-year period from 2012 to 2021. A1 is used in this case, simply because at the time of writing the results for A1 are the only ones available with respect to calculating the GSPs based on EUCLID data. It is envisioned that a future follow-up study will investigate in more detail the differences between some of the GSP algorithms. In any case, we believe that applying either one of the GSP algorithms will show the same patterns as described hereafter.

Figure 3a depicts the average monthly number of GSPs per flash for AT and BE. Solid lines are the results taking into account all flashes, while the dashed lines exclude single stroke flashes from the EUCLID observations. It is found that the average number of GSPs per flash is highest during the summer months. This becomes especially visible when single stroke flashes are omitted from the data. On average, the number of GSPs per flash is somewhat higher in AT compared to the average value found in BE. The reason for this discrepancy is food for thought and under further investigation. But a difference in topography and/or thunderstorm type between AT and BE can impact the

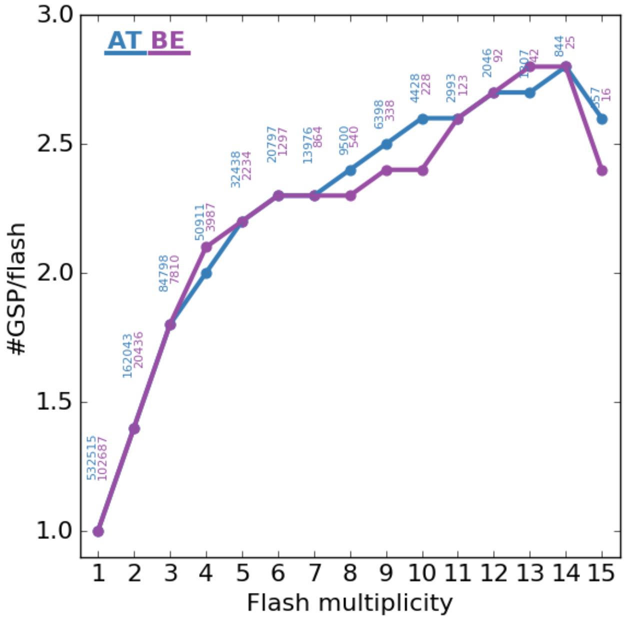

Fig. 4: Distribution of the average number of GSPs per flash as a function of flash multiplicity. The actual number of flashes per multiplicity is listed as well.

outcome. Figure 3b is similar to Fig. 3a, but in this case the diurnal behavior of the number of GSPs per flash is illustrated. The results show a trend that agrees for both regions, whereby a minimum is observed in the afternoon hours between about 12h-18h UTC (Universal Time Coordinated). These findings are novel and to the best of the authors' knowledge, this observation has not been published in the literature yet. Since the reasons for these monthly and diurnal trends/patterns are unclear, any attempt to explain them would be purely speculative at the time of writing and out of the scope of the paper. Surely, this will be further investigated in the future.

Finally, Figure 4 displays the average number of GSPs per flash as a function of multiplicity. For this purpose, only multiplicities up to a value of 15 are taken into account since the sample size becomes too low at higher multiplicities. As expected, it is found that with increasing multiplicity, the average number of GSPs per flash increases accordingly.

V. SUMMARY

Distinct ground strike point algorithms have been tested in terms of their ability to estimate correctly the actual observed number of GSPs per flash based on observations made by high-speed cameras. In this regard, it is found that A2 performs best. Although the other GSP algorithms over- or underestimate to some extent the number of GSPs per flash, this number is anyhow fairly close to the observed value.

In order to study ground strike point characteristics on a larger temporal and spatial scale, observations from LLSs can be ingested into a GSP algorithm. However, in order to do so the particular LLS must have a high accuracy in terms of both detection efficiency and location accuracy. EUCLID meets this requirement, hence its observations from 2012 to 2021 are provided as input to one of the algorithms. A seasonal and diurnal trend is identified, in which the number of GSPs per flash is highest in the summer and lowest between 12 and 18 UTC, respectively. Finally, it is demonstrated that with increasing multiplicity, the average number of GSPs per flash increases accordingly. For example, a flash with four CG strokes creates on average two GSPs per flash, while this number increases to 2.5 GSPs per flash for a ten-stroke flash.

REFERENCES

[1] IEC 62305-2 (2010) Protection against lightning – Part 2: risk management.

[2] Rakov, V. A., Uman, M. A., and Thottappillil, R.: Review of lightning properties from electric field and TV observations, J. Geophys. Res., 99, 10745–10750, https://doi.org/10.1029/93JD01205, 1994.

[3] Valine, W. C. and Krider, E. P.: Statistics and characteristics of cloud-to-ground lightning with multiple ground contacts, J. Geophys. Res.-Atmos., 107, 4441, https://doi.org/10.1029/2001JD001360, 2002.

[4] Saraiva, A. C. V., Saba, M. M. F., Pinto, O., Cummins, K. L., Krider, E. P., and Campos, L. Z. S.: A comparative study of negative cloud-to-ground lightning characteristics in São Paulo (Brazil) and Arizona (United States) based on high-speed video observations, J. Geophys. Res., 115, D11102, https://doi.org/10.1029/2009JD012604, 2010.

[5] Poelman, D. R., Schulz, W., Pedeboy, S., Hill, D., Saba, M., Hunt, H., Schwalt, L., Vergeiner, C., Mata, C. T., Schumann, C., and Warner, T.: Global ground strike point characteristics in negative downward lightning flashes – Part 1: Observations, Nat. Hazards Earth Syst. Sci., 21, 1909–1919, https://doi.org/10.5194/nhess-21-1909-2021, 2021.

[6] Poelman, D. R., Schulz, W., Pedeboy, S., Campos, L. Z. S., Matsui, M., Hill, D., Saba, M., Hunt, H..: Global ground strike point characteristics in negative downward lightning flashes – Part 2: Algorithm validation, Nat. Hazards Earth Syst. Sci., 21, 1921–1933, https://doi.org/10.5194/nhess-21-1921-2021, 2021

[7] Schulz, W., Diendorfer, G., Pedeboy, S., and Poelman, D. R.: The European lightning location system EUCLID – Part 1: Performance analysis and validation, Nat. Hazards Earth Syst. Sci., 16, 595–605, https://doi.org/10.5194/nhess-16-595-2016, 2016.

[8] Poelman, D. R., Schulz, W., Diendorfer, G., and Bernardi, M.: The European lightning location system EUCLID – Part 2: Observations, Nat. Hazards Earth Syst. Sci., 16, 607–616, https://doi.org/10.5194/nhess-16-607-2016, 2016.

[9] Diendorfer, G.: A review of 25 years of lightning research in Austria from 1991–2015, in: World meeting on Lightning, 6–8 April 2016, Cartagena, Colombia, 2016.

[10] Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan: Fast Density-Based Clustering with R. Journal of Statistical Software, 91(1), 1–30. https://doi.org/10.18637/jss.v091.i01