INTERNATIONAL CONFERENCE ON LIGHTNING PROTECTION
33rd International Conference on Lightning Protection
25 - 30 September 2016
Estoril, Portugal

Lightning Location Systems and Interstroke Intervals: Effects of Imperfect Detection Efficiency

Mirjana Stojilović[1], Mohammad Azadifar[2], Marcos Rubinstein[1] and Farhad Rachidi[2]

1 University of Applied Sciences Western Switzerland, Yverdon-les-Bains, Switzerland
2 Swiss Federal Institute of Technology (EPFL), EMC Laboratory, Lausanne, Switzerland
{mirjana.stojilovic, marcos.rubinstein}@heig-vd.ch, {mohammad.azadifar, farhad.rachidi}@epfl.ch

***Abstract*—Lightning location systems (LLSs) are often used to obtain parameters of lightning flashes. However, the imperfect detection efficiency (DE) of these systems affects the accuracy of the results. In this paper, we focus on the effects of the imperfect DE on the probability distribution of interstroke intervals. Two DE models are considered here: one dependent on the peak amplitude of return stroke currents and the other independent of any stroke parameter. For these two models, the results show that the distribution of interstroke intervals obtained from strokes detected by an LLS depends essentially on the average DE in %. In addition, we demonstrate that the average DE has a noticeable influence on the geometric mean (GM) of the measured interstroke intervals. Finally, we provide a mathematical expression for estimating the DE of an LLS; this expression uses the GM of the interstroke intervals obtained from the LLS data and assumes that the distribution of interstroke intervals for the geographical location of concern is known.**

Keywords-detection efficiency; interstroke interval; lightning location system; flash

I. Introduction

Lightning location systems provide global information on occurrence characteristics of lightning discharges. Inferred lightning parameters by LLSs can be used for early warning, weather forecasting, risk assessment, geophysical research, power utilities, and other applications [1].

LLSs can report various parameters of the detected lightning flashes, such as the individual stroke peak amplitudes, stroke locations, stroke polarities, flash multiplicity, interstroke intervals, and flash type (cloud-to-ground vs. cloud lightning) [2]. Imperfect detection efficiency (DE), location accuracy (LA), and peak estimation along with inhomogeneity of LLS networks might cause some inaccuracies in lightning parameter distributions inferred from LLS data [3]. Numerous studies have been conducted to evaluate and improve the performance of LLS networks either by hardware development or algorithm improvement (see [1] for a recent review). These studies have led to improved performance and, hence, to more accurate lightning parameter distributions derived by these networks.

Statistical approaches have also been applied to investigate the effect of the imperfect LLS performance on the inferred lightning parameter distributions and to obtain estimates of the actual distributions. The effect of the detection efficiency on the return stroke peak current distribution, as well as corrections for the distribution, were presented in [4]. Additionally, a sensitivity analysis was performed on the Brazilian network and the influence of the DE on the peak current estimation, and confidence ellipse was reported in [5]. The distributions of interstroke intervals and flash multiplicity reported by LLS networks were compared with those obtained from a separate field measurement system in [6].

An approach to analyze the effects of the imperfect stroke-detection efficiency of LLS on the statistical distribution of interstroke intervals was proposed in [7]. In this paper, we implement it and analyze the results. Two types of DE are considered here: uniform DE and return stroke current dependent DE. We also present a novel methodology for estimating the DE of LLS, based on the geometric mean of the detected interstroke intervals.

The paper is organized as follows. Section II presents the analysis method used in [7]. The effects of the DE on the interstroke intervals distribution are presented and discussed in Section III. Some concluding remarks are given in Section IV.

II. Methodology

For the analyses presented in this paper, we created a program in Matlab to simulate the effects of the imperfect detection efficiency of lightning location systems. This program accepts as inputs the total number of flashes, the probability distribution of the interstroke intervals and of the stroke currents, the probability distribution of the flash multiplicities, and the detection efficiency of the LLS. Based on the inputs, the program creates a dataset of lightning strokes [7] satisfying the input probability distributions. Then, it eliminates some of the strokes, making sure the detection efficiency specification is fulfilled. The DE can be specified either as a percentage (ratio of the number of detected to the total number of strokes) or in the form of a probability distribution, which might depend, as in this work, on the stroke

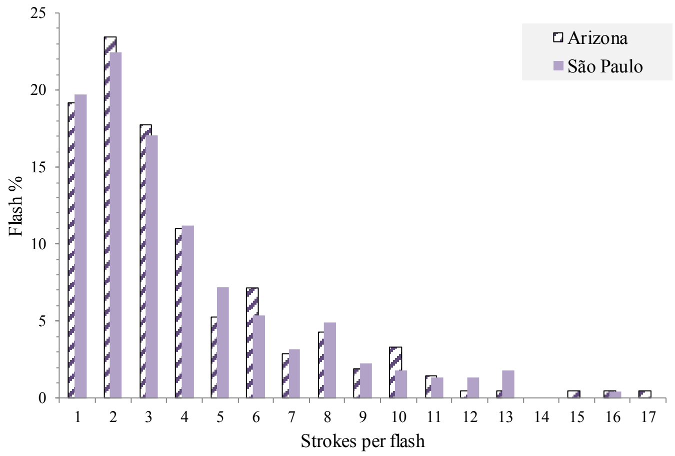

Figure 1. Percentage of flashes with different number of strokes per flash (multiplicity) observed in Arizona and São Paulo. Adopted from [8].

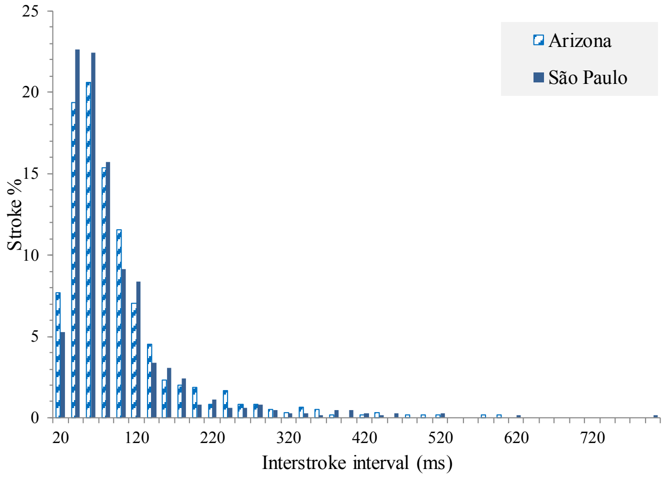

Figure 2. Distribution of interstroke intervals in Arizona and São Paulo. Adopted from [8].

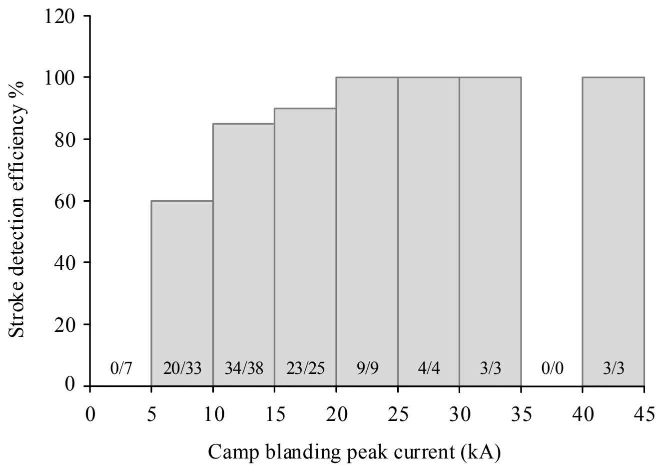

Figure 3. Stroke detection efficiency as a function of peak current. The ratio given inside each column represents the number of strokes detected by the lightning detection system over the number of strokes recorded directly. Adopted from [9].

current peaks. The program outputs are a plot of the cumulative probability distributions of the initial and the final set of interstroke intervals, as well as the geometric mean and standard deviation derived from the interstroke interval distributions. Although we use it in this paper exclusively to study the interstroke interval distribution, the program can be readily adapted to study the effect of the detection efficiency on other parameters, such as the peak current.

A. Flash Multiplicities and Interstroke Intervals

The program allows users to specify the probability distributions of flash multiplicity and interstroke intervals, and fills the dataset accordingly. In this paper, we use the distributions corresponding to the data measured using high-speed video observations in the city of São Paulo (Brazil) and in Arizona (United States) [8]. These two sets of experimental results exhibit similar distributions of flash multiplicities and interstroke intervals (Figs. 1 and 2), despite being captured at sufficiently different locations. Nevertheless, they were used here as inputs essentially because they had not been affected by LLS imperfections.

B. Return Stroke Currents

The cumulative statistical distribution of peak stroke currents is lognormal. According to [10], negative first stroke currents can be obtained assuming that the median and the logarithmic standard deviation equal 30 kA and 0.265, respectively. In the same way, negative subsequent stroke currents can be calculated assuming that the median and the logarithmic standard deviation equal 12 kA and 0.265, respectively. These distribution parameters are based on Berger's direct lightning current measurements in Switzerland from 1963 to 1971. Given the total number of flashes, the flash multiplicities and the current distributions, our program determines the stroke current peaks.

C. Detection Efficiency

Prior to running the program, one should specify a DE either as uniform (in percent, independent of the characteristics of the stroke the system is trying to detect) or as a function, for example, of the stroke current. We implement one such function based on the data published in [9] and shown in Fig. 3. The detection efficiency in [9] was obtained using triggered lightning, for which the distance from the sensors to the strike point is known and unchanged for all strokes. Although that is not the case for the data from São Paulo and Arizona used in the paper, we included the results of [9] because we are interested in the effect of an amplitude dependent detection efficiency vs. a constant DE and not in the actual detection efficiency.

If the detection efficiency is given as a percentage of all the strokes, independent of the stroke characteristics, then a random set of strokes is removed, making sure that the number of detected strokes over the initial total number of strokes equals the detection efficiency.

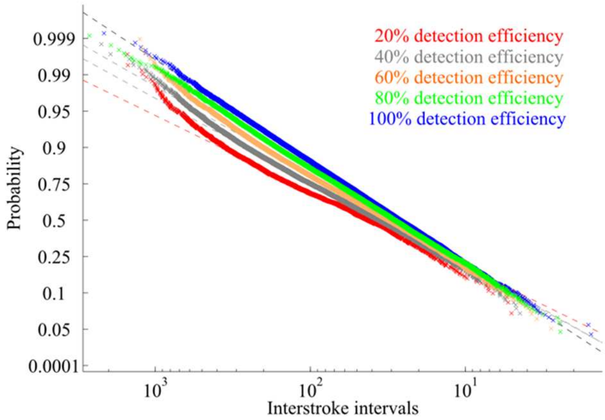

Figure 4. Cumulative probability distribution of interstroke intervals for São Paulo and detection efficiency ranging from 20% to 100%.

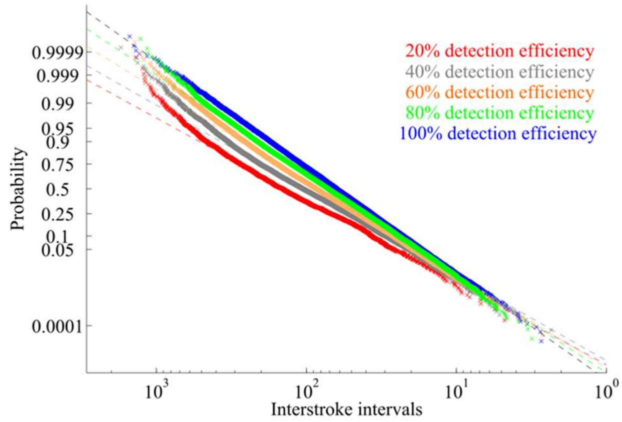

Figure 5. Cumulative probability distribution of interstroke intervals for Arizona and detection efficiency ranging from 20% to 100%.

III. Results

A. Effects of Non-Ideal Uniform Detection Efficiency

To evaluate the effect of an imperfect detection efficiency on the probability distribution of interstroke intervals, we ran several tests. The first two tests were performed on the data measured using high-speed video observations in São Paulo [8] and Arizona [8], respectively. A data vector containing 1e5 flashes was created for both cases. Interstroke intervals and flash multiplicities were assigned in accordance with Figs. 1 and 2. The detection efficiency was varied from 20% to 100%, in steps of 20%. The resulting cumulative probability distributions are plotted in Figs. 4 and 5. As expected, both figures demonstrate the following trends:

1) When the DE decreases, the number of detected strokes decreases as well. Except for single-stroke flashes, every missed stroke leads to a lost data point in the interstroke intervals dataset. Thus, the lower the DE, the lower the number of points in the corresponding interstroke intervals plot.

To understand this tendency, let us take an n-stroke flash and assume a stroke detection efficiency $deff$. Here, we assume that the probability of detecting a stroke is the same for all strokes and that it has no effect on the probability of detecting other strokes. Each of these flashes has $n-1$ interstroke intervals and thus contributes to $n-1$ points on the graph. The probability that all n strokes will be detected equals

$$p_n = deff^n . \tag{1}$$

For example, if $deff$ is 20% and n is three, p_3 will be 0.8% only. If $deff$ is 80%, p_3 will be 51.2%. Therefore, the chances that no data point will be lost from the plot are considerably lower than the detection efficiency itself.

The probability that all n strokes will be missed equals

$$p_0 = (1-deff)^n . \tag{2}$$

2) In Figs. 4 and 5, the lefthand side of the distributions converges towards one and the righthand side towards zero. This is due to the very nature of the cumulative probability distribution.

3) The lower the DE, the farther the cumulative probability plot from the perfect one. The first factor causing the difference in the plobability plots is the loss of interstroke interval data points due to the missed strokes. The second is the appearance of new interstroke intervals. Let us analyze this using, as an example, a three-stroke flash. If only the first or the last stroke is missed, one interstroke interval will be lost. If, however, only the middle stroke is missed, two interstroke intervals will be lost. Yet, they will be replaced by one new interstroke interval equal to their sum. The probability that only one stroke is missed equals

$$p_2 = deff^2 \,(1-deff) . \tag{3}$$

The probability that two strokes are missed is

$$p_1 = deff \,(1-deff)^2 . \tag{4}$$

Using (1) to (4) and assuming $deff$ is 80%, it follows that, from the interstroke intervals associated with 3-stroke flashes, 36% will be deleted, 64% will be retained, and 12.8 % new values will be added. The ratio between the number of retained points and the number of new points is five. In addition, since short interstroke intervals are more likely to occur than large ones (see Fig. 2), most of the new interstroke intervals will not exceed the maximum interstroke interval in the original dataset, and thus the final distribution will not be considerably influenced by the new values. However, if $deff$ is only 20%, 96% of interstroke intervals will be lost, 4% will be retained, and 3.2% will be replaced by new values. The ratio between the number of retained points and the number of new points becomes 1.25. As a consequence, new points will have more effect on the probability distribution. Since the new interstroke intervals are higher in value than their predecessors, the mean value of the interstroke interval set will exhibit an increase, and the distribution will be shifted towards higher values of interstroke intervals. A similar analysis can be carried out for

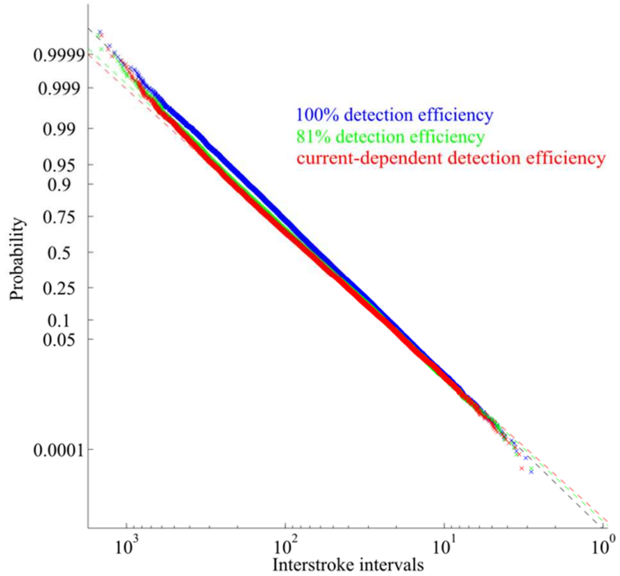

Figure 6. Cumulative probability distribution of interstroke intervals for Arizona and current-dependent detection efficiency (in red). Since this detection efficiency results in the removal of 19% of samples, it is compared with the cumulative probability for 81% uniform detection efficiency. The results show that there is no significant difference in the final cumulative probability distribution. The dataset size was 1e5 flashes.

other flash multiplicities. It is these effects, accumulated from all flashes of various multiplicities that cause the shifts in probability distributions, as shown in Figs. 4 and 5.

B. *Effects of Current-Dependent Detection Efficiency*

The LLS stroke detection efficiency can be a function of the peak current. Fig. 3 shows one such DE [9].

When applied to the initial dataset, this DE results in 19% of the strokes being missed. The obtained cumulative probability distribution of interstroke intervals is plotted in Fig. 6 (in red). For comparison purposes, we also plotted the distribution obtained using a uniform detection efficiency of 81% (in green). We can see from the figure that these two distributions almost completely overlap. This suggests that the distribution depends essentially on the equivalent DE in %. The slight difference observed here is due to the fact that the current-dependent DE and the stroke order are correlated; indeed, since the first strokes are statistically larger in amplitude than subsequent strokes, the current-dependent DE used in this paper implies that, statistically, more first strokes than subsequent strokes will be detected.

C. *Estimating LLS detection efficiency*

Let us analyze the behavior of the geometric mean (GM) and the geometric standard deviation (GSD) of the detected interstroke intervals as a function of the LLS detection efficiency. To that end, two datasets are created: one for São Paulo and one for Arizona, based on information from the histograms in Figs. 1 and 2.

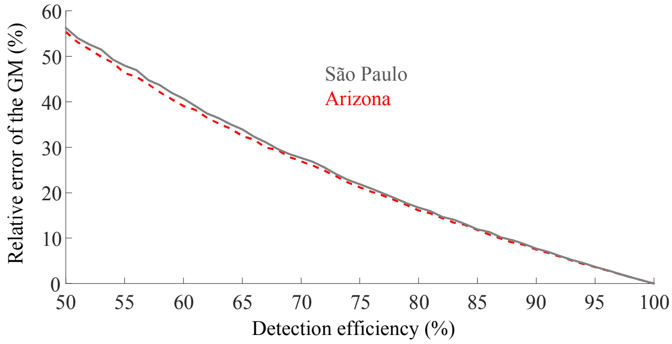

Figure 7. Relative error of the geometric mean of the interstroke interval as a function of DE. The dataset of 1e5 flashes was created based on histograms in Figs. 2 and 3.

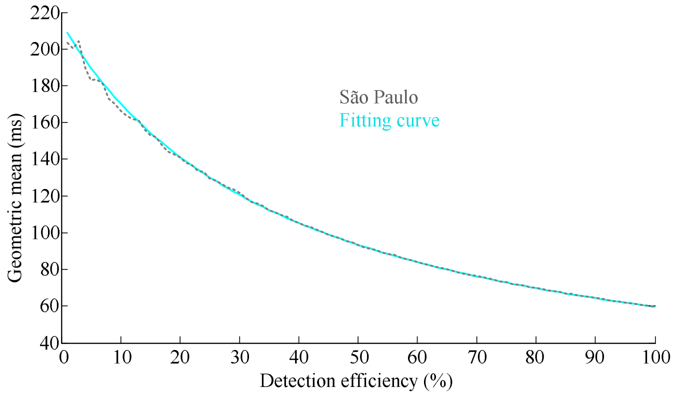

Figure 8. Geometric mean of the interstroke interval as a function of DE for interstroke intervals observed in São Paulo (in gray). A fitting curve is plotted in blue.

Let us assume that the DE is uniform and independent of any stroke parameter. The relative error of the GM, when the DE ranges from 50% to 100%, for data recorded in both locations, is shown in Fig. 7. As expected, for both locations, the relative error approaches zero as the DE improves. Additionally, it is in fact affected primarily by the DE and not by the geographical location for the Arizona and São Paulo data considered in this paper.

A plot of the GM of the interstroke intervals as a function of the detection efficiency is shown in Fig. 8. By inspection, we found that a function of the form

$$gmean = \frac{a}{deff + b}, \tag{5}$$

where the variable $deff$ denotes the value of the DE in % and $gmean$ denotes the value of the GM, provides a very good fit for the values of the GM obtained from our dataset. In (5), the parameters a and b depend on the distribution of the interstroke intervals. The following function is found to fit well the São Paulo data:

$$gmean = \frac{8274.9}{deff + 38.6044}. \tag{6}$$

The corresponding fitting curve is plotted in Fig. 8 in blue. The function for the Arizona data is

$$gmean = \frac{8698.4}{deff + 39.63}. \qquad (7)$$

Interestingly, the coefficients *a* and *b* in (6) have values similar to those of *a* and *b* in (7), although the interstroke intervals data originates from two very distant geographical locations.

If reliable interstroke interval data exist for a particular geographical location, i.e., if the coefficients *a* and *b* are known, one can estimate the DE (in %) of an LLS based on (5) and on the GM of the interstroke intervals it detects:

$$deff = \frac{a}{gmean} - b. \qquad (8)$$

The relative error of the GSD for both sets of data is below 10% in the whole range of DE and decreases with the increase in DE. Moreover, it is below 0.5% for DEs higher than 90%. Hence, one can assume that the GSD does not vary significantly with the DE.

In order to generalize these observations, the methodology presented in this section should be tested using other models of the DE that consider possible correlation between the DE, the flash multiplicity, the stroke order, and the interstroke intervals. However, to the best of our knowledge, such correlations have not yet been studied.

IV. CONCLUSION

We investigated the effect of a lightning location system's imperfect detection efficiency on the measured statistical distribution of interstroke intervals. To that end, we generated a dataset that corresponds to the observed lognormal distributions of interstroke intervals, current amplitudes, and stroke multiplicities presented in the literature. We assumed that dataset to be ground-truth. Then, we eliminated some of its strokes to simulate the imperfect detection efficiency.

Two different models for the detection efficiency were used: one independent of any stroke parameter and another that assumes that the detection efficiency depends on the peak amplitude of the return strokes [9]. The results show that the effect of an imperfect detection efficiency on the statistical distribution of interstroke intervals is essentially independent of the detection efficiency model.

A lightning location system's average detection efficiency has a noticeable effect on the geometrical mean of the interstroke interval distribution. Its effect on the geometric standard deviation of the same distribution is negligible.

Based on our analyses, we propose Equation (8) to estimate the detection efficiency, given the geometric mean of the interstroke intervals obtained from LLS data. Since the parameters in this expression may be a function of the geographical location, this formula should be particularized to the region where the interstroke intervals are being measured.

ACKNOWLEDGMENT

Financial support from the Swiss National Science Foundation (Project No. 200021_147058) is acknowledged. The authors would like to thank Dr. Marcelo Saba and Dr. Antonio Varela Saraiva for providing the data plotted in Figs. 2 and 3.

REFERENCES

[1] A., Nag, M. J. Murphy, W. Schulz, and K. L. Cummins, "Lightning locating systems: Insights on characteristics and validation techniques," *Earth and Space Science*, vol. 2, no. 4, pp. 65–93, Apr. 2015, doi:10.1002/ 2014EA000051.

[2] W. Schulz, K. Cummins, G. Diendorfer, and M. Dorninger, "Cloud-to-ground lightning in Austria: A 10-year study using data from a lightning location system," *J. Geophys. Res.*, vol. 110, no. D9, May 2005, doi:10.1029/2004JD005332.

[3] Cummins K.L. and E.A. Bardo, "On the relationship between lightning detection network performance and measured lightning parameters," paper presented at the *1st International Conference on Lightning Physics and Effects*, Belo Horizonte, Brazil, Nov. 7–11, 2004.

[4] W. Schulz, G. Diendorfer, R. Iorio, F. Hofbauer and A. Stimmer, "Correction of lightning density and lightning current distributions for detection efficiency," Paper presented at the *23th Int. Conference on Lightning Protection*, Florence, Sep. 1996.

[5] C. R. Mesquita, R. N. Dias, S. Visacro. "Analysing the Effects of LLS Detection Efficiency on Lightning Parameter Data", *IX International Symposium on Lightning Protection*; Nov. 2007, Foz do Iguaçu, Brazil.

[6] W. Schulz, and G. Diendorfer, "Flash multiplicity and interstroke intervals in Austria," Proceedings of the 28th International Conference on Lightning Protection, Inst. of Electr. Installation Eng. of Jpn., Kanazawa, 2006, pp. 402 – 404.

[7] M. Azadifar, M. Stojilovic, M. Rubinstein, and F. Rachidi, "Influence of LLS Detection Efficiency on the Measured Distribution of Interstroke Intervals," Presented at American Electromagnetics International Symposium (AMEREM), Albuquerque, New Mexico, USA, Jul. 27–31, 2014.

[8] A. C. V. Saraiva, M. M. F. Saba, O. Pinto Jr., K. L. Cummins, E. P. Krider, and L. Z. S. Campos, "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.*, vol. 115, no. D11, Jun. 2010. doi: 10.1029/2009JD012604.

[9] A. Nag, S. Mallick, V. A. Rakov, J. S. Howard, C. J. Biagi, J. D. Hill, M. A. Uman, D. M. Jornad, K. J. Rambo, J. E. Jerauld, B. A. DeCarlo, K. L. Cummins, and J. A. Cramer, "Evaluation of U.S. National Lightning Detection Network performance characteristics using rocket-triggered lightning data acquired in 2004–2009", *J. Geophys. Res.*, vol. 116, no. D2, Jan. 2011, doi: 10.1029/2010JD014929.

[10] V. A. Rakov et al., "Lightning parameters for engineering applications," *CIGRE* WG C4.407, 2013.