2008
20th International Lightning Detection Conference
21-23 April • Tucson, Arizona, USA
2nd International Lightning Meteorology Conference
24-25 April • Tucson, Arizona, USA

A method to determine relative stroke detection efficiencies from multiplicity distributions

Schulz W.[1] and Cummins K.[2]
1. Austrian Lightning Detection and Information System (ALDIS), Kahlenberger Str.2A, 1190 Vienna, Austria, w.schulz@ove.at
2. Institute of Atmospheric Physics, University of Arizona, Tucson, USA.

1. Introduction

The performance of a lightning location system (LLS) can vary with time as a result of changes in the location and performance of the sensors that comprise the network. The most common changes are (1) the addition or removal of sensors and (2) updated sensor technology. Changes in LLS performance can lead to significant changes in estimated lightning parameters (e.g. peak current and flash multiplicity statistics - Cummins and Bardo, 2004). Given these facts, it is important to be able to quantify these network changes in terms of the relative detection efficiency (DE) between various network configurations.

A method to determine the overall relative stroke DE using peak current distributions has been presented by Cummins and Bardo [2004], and is described in detail in an upcoming CIGRE report by Task Force C4.404A. This approach has the limitation that any changes in network configuration that alter the individual stroke peak current estimates will introduce errors in the DE calculation. Further, it does not provide relative flash DE without additional calculations involving multiplicity measurements.

The method described in this paper is not subject to the limitations noted above, although we show that it is sensitive to other LLS-derived parameters. The basis for this detection efficiency (DE) correction using multiplicity distributions is the work by Rubinstein [1995], where he shows the relation between flash and stroke DE.

2. Theory

2.1 Relation between Flash and Stroke DE

Rubinstein [1995] shows that the detected flash multiplicity distribution N_f can be calculated from the "actual" flash multiplicity distribution N_f^a using

$$N_f(m) = \sum_{n=m}^{\infty} F(n,m)\, N_f^a(n) \qquad m=1,2,...,\infty \tag{1}$$

where F(n,m) is the probability to detect an n-stroke flash as m-stroke flash. With the assumption that individual stroke DE (defined as P) is independent of stroke order, F(n,m) can be calculated according to Eq. (2).

$$F(n,m) = \binom{n}{m} P^m \, (1-P)^{n-m} \qquad (2)$$

In this example F(n,m) is completely specified by the value of P. In our work we use two somewhat more general models for determining F(n,m):

a) Different but constant stroke DEs for different stroke orders are allowed.
b) In addition to condition a), first-stroke DE is allowed to depend on multiplicity.

For both cases we calculate F(n,m) using a recursive function (see Eq. 3) where p(i) is the DE of the ith stroke order and q(i)=1-p(i). If the real multiplicity (n) is increased, then F(n+1,m) may be calculated as

$$F(n+1,m) = q(n+1) * F(n,m) + p(n+1) * F(n,m-1) \qquad (3)$$

The first product term in Eq. 3 handles the case where the additional stroke is not detected but ”m” other strokes are, and the second term handles the case where the additional stroke is detected and one of the other “m” strokes is not detected. The two extreme terms of the recursive function F(n,0) and F(m,m) are defined in Eq. 4 and Eq. 5.

$$F(n,0) = \prod_{i=1}^{n} q(i) \qquad (4)$$

$$F(m,m) = \prod_{i=1}^{m} p(i) \qquad (5)$$

Each column in the matrix F is related to a different multiplicity value. The mechanism to implement different first stroke DEs is to calculate all elements of matrix F related to a specific multiplicity using the related first stroke DE.

2.2 Relative DE Correction

Our new method to determine stroke DE corrections relies on describing the problem in terms of relative DEs. If we are interested in the DE of a LLS for a certain period with bad DE (N_{flow}) relative to a period with good DE (N_{fhigh}) we simply have to rewrite Eq.(1) as

$$N_{flow}(m) = \sum_{n=m}^{\infty} F(n,m)\, N_{fhigh}(n) \qquad m=1,2,\ldots,\infty \qquad (6)$$

In the case where one assumes only two different stroke DEs, one for first and one for subsequent strokes (p_{first} and p_{sub}), then each F(n,m) term in Eq. (6) contains those two stroke DEs as unknowns. With a nonlinear least square algorithm (see Appendix A) it is possible to determine those two unknowns (starting with an initial “guess”), and therefore determine estimates of the relative stroke DEs for the network.

Once the stroke DEs and related F(n,m) are computed, the relative flash DE for the “low” condition can be calculated. This is done with Eq. 7 where the total number of flashes for the “low” condition (Numerator) is related to the total number of flashes for the “high” condition (Denominator).

$$DE_{flash} = \frac{\sum_{n=1}^{\infty}\left(\sum_{n=m}^{\infty} F(n,m)\, N_{fhigh}(n)\right)}{\sum_{n=1}^{\infty} N_{fhigh}(n)} \quad m = 1,2,\ldots,\infty \tag{7}$$

3. Test of the algorithm

In order to verify the calculation of the relative DE, we have tested the algorithm in several ways:

- We assumed a first-stroke DE, a subsequent-stroke DE and a multiplicity distribution N_{fhigh}, and used those values to calculate the multiplicity distribution N_{flow} using Eq. (6). We then used the two multiplicity distributions in the new algorithm to estimate the stroke DEs. This calculation procedure resulted in stroke DEs identical to the assumed ones.
- We assumed a “real” multiplicity distribution and calculated distributions for two conditions of the network -- N_{fhigh} (with p_{first}=0.9 and p_{sub}=0.7) and N_{flow} (with p_{first}=0.75 and p_{sub}=0.5). We then used those two multiplicity distributions to calculate the relative DEs. Since we can directly compute the relative flash DE between both distributions N_{fhigh} and N_{flow} we were able to compare with the resulting flash DE obtained using the correction algorithm. Also in this case the algorithm converged to the correct value.
- We tested the algorithm with real Austrian data to evaluate the effect of different maximum multiplicities for the distributions (up to a maximum of 20). We found that there is no significant difference in the result if we ignore flashes with multiplicity greater than 14. Only the calculation time increases significantly.

4. Results with real data

We used data from the NLDN for this analysis because DE corrections (using the correction algorithm described in the CIGRE report from Task Force C4.404A) were already available. These existing corrections provide an overall stroke DE for the individual 2x2 degree regions shown in Figure 4.1. For our evaluation we have chosen a region close to Tucson (region 169) and a region in Florida (region 120). In those two regions we compare the performance of the NLDN between 1999 and 2004 (after the 2002-2003 upgrade of the network).

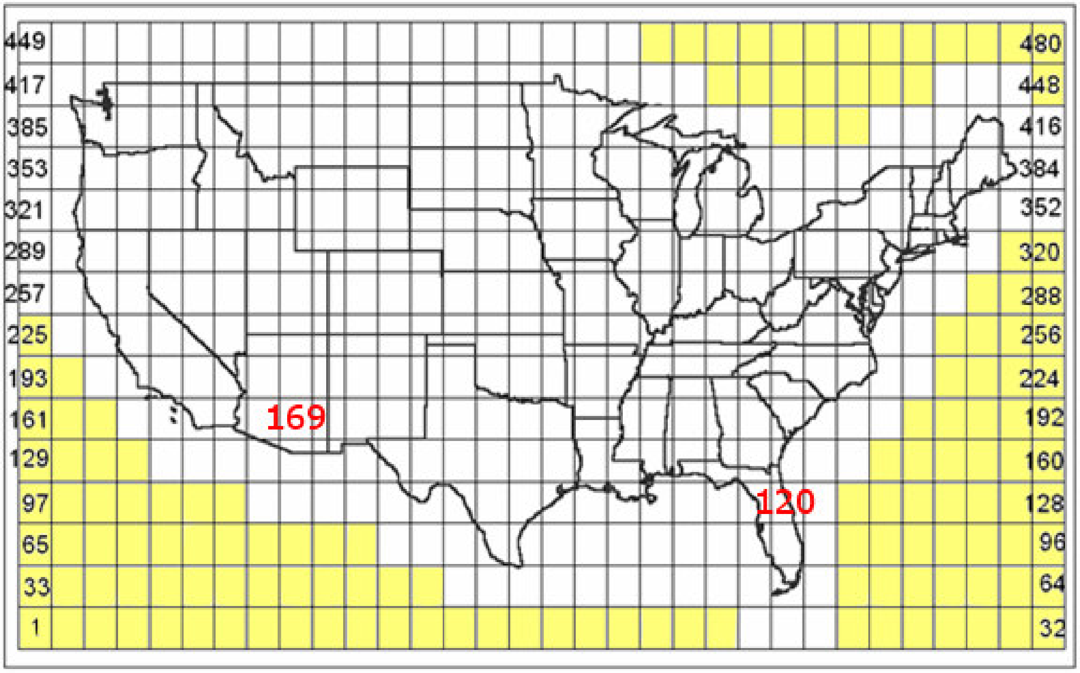

Fig. 4.1: Red numbered regions will be used for comparison

In the initial analysis we assumed that all stroke DEs are independent of multiplicity, and that first and subsequent strokes can have different (but constant) DE values. The results for the two regions and for different data sets are compared to the original DE correction made with the algorithm described in the CIGRE report and are given in Tables 1 and 2. Estimated DE values were obtained for three conditions, as described below. The overall stroke DE (first column) for the new method is computed using a variation of Eqs. (1) and (3) in Cummins and Bardo [2004]. The derivation is given in Appendix B (see Eq. B6).

It can be seen that if all data are used ("All Data" condition in Tables 1-2) an unrealistically small first stroke DE (Region 169: 13%; Region 120: 10%) is obtained. This small first-stroke DE is unlikely, given that first strokes normally exhibit peak currents greater than subsequent strokes, which mean that first-stroke DE should not be lower than subsequent stroke DE.

Table 1: Region 169 - first stroke DE is constant

	Stroke DE [%]	1st Stroke DE [%]	Sub. Stroke DE [%]	Flash DE [%]
DE correction according to CIGRE	49	-	-	58
All data	43	13	63	50
Flashes with first stroke peak current >10kA	50	31	62	63
Flashes with first stroke peak current >15kA	53	40	60	71

Table 2: Region 120 - first stroke DE is constant

	Stroke DE [%]	1st Stroke DE [%]	Sub. Stroke DE [%]	Flash DE [%]
DE correction according to CIGRE	74	-	-	88
All data	53	10	74	57
Flashes with first stroke peak current >10kA	72	90	66	96
Flashes with first stroke peak current >15kA	-	-	-	-

The cause of this small first-stroke DE could be a combination of several factors, e.g. misclassified cloud strokes, strokes with a bad position (outlier) that are not grouped with other strokes in the flash, flashes with multiplicity higher than 15 which are split in a flash with 15 strokes and a flash with the remaining strokes, and the assumption that first stroke DE is independent of multiplicity (low multiplicity is associated with low-current first strokes – Orville et al., 2002; Schulz et al., 2005). All of these possible causes will tend to produce excess low-peak-current single-stroke flashes which lead to low first stroke DEs.

To test the influence of an excess number of single-stroke flashes, we explored the sensitivity of the estimated DE to small increases in the fraction of single-stroke flashes. We employed the second simulation test condition described in Section 3 to create “true” multiplicity distributions, and then the fraction of single-stroke flashes for the “high” DE condition was artificially increased. When an additional 1% of the flashes were forced to be single-stroke, the estimated first-stroke DE decreased from 83.8% to 65.4% and the flash DE decreased from and 89.4% to 77.1%. When an additional 10% of the flashes were forced to be single-stroke, the first-stroke and flash DEs decreased further to 21.7% and 46.1%, respectively. Clearly first-stroke and flash DE estimates are very sensitive to errors in the fraction of single-stroke flashes. We note that the estimated subsequent stroke DE varied less than 1% from the “true” value for both simulated conditions.

To test the influence of low-peak-current flashes on DE estimates derived from measured LLS data, we calculated the DEs for conditions where we excluded flashes with first-stroke peak current less than 10 kA and less than 15 kA. As can be seen from Tables 1 and 2, both the first-stroke and flash DE values increase when flashes with low-peak-current first strokes are excluded. However, for region 169 the first-stroke DE is still lower than the subsequent-stroke DE. Note also that the overall stroke DE values for flashes with first-stroke peak current greater than 10 kA are very similar to the overall stroke DE according to the method published at CIGRE. A correction for region 120 and first-stroke peak currents greater than 15 kA was not possible because the algorithm did not converge to a reasonable result.

In order to evaluate the importance of the dependence of first-stroke peak current on multiplicity, we have augmented the model to include this relationship. Prior studies indicate that there is a factor of about 2 difference between the amplitude of a single-stroke flash and the amplitude of a first stroke in a flash with multiplicity 10, with a monotonic increase in peak current as a function of multiplicity [Orville et al., 2002; Schulz et al., 2005]. We assume a linear relationship between the first-stroke DE and the multiplicity. To explore this model, we also assumed a variety of specific values for the slope of this relationship, and then we estimated both the intercept of this first-stroke-DE:multiplicity relationship and the subsequent-stroke DE. We have tested the assumption with data from region 120 for different slopes and found the lowest squared-error value of the optimization for a slope of 0.05, yielding a factor of about 6 between a single stroke flash and a flash with multiplicity 10, as shown in Fig. 4.2. The squared-error value in this case is even lower than the squared-error value for the calculation without dependence of first-stroke DE on multiplicity.

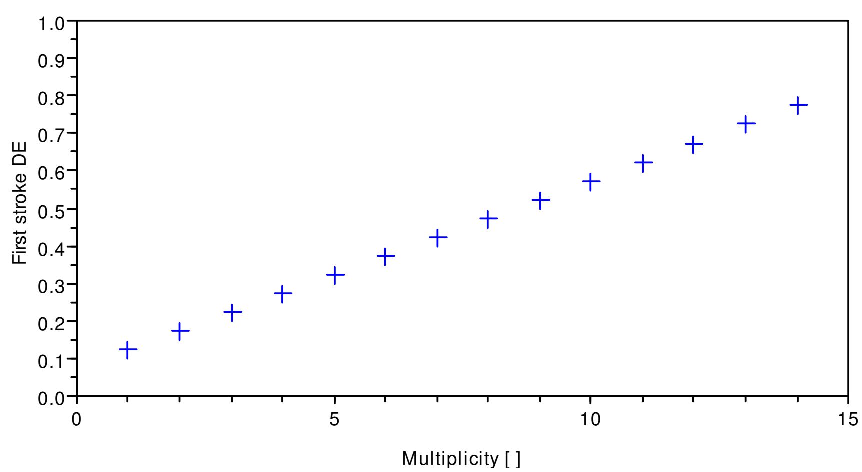

Fig. 4.2: First stroke DE as a function of multiplicity (slope = 0.05).

For this slope we computed a subsequent-stroke DE of 66% and a flash DE of 56%. It is interesting to note that the flash DE did not increase significantly. We also performed this calculation for the multiplicity distribution of flashes excluding first-stroke peak currents less than 10 kA. This condition had the smallest squared-error when we employed a smaller slope, which is reasonable considering that the relation between first-stroke peak currents and multiplicity will be weaker when flashes with low-current first strokes are excluded. The results were similar to the values given in Table 2 for the >10 kA condition.

5. Summary/Discussion

In this paper we introduced a new method to determine relative changes in detection efficiency of a LLS based on multiplicity distributions. Contrary to earlier methods that employ peak current distributions, this new method provides direct estimates of first stroke DE, subsequent stroke DE and flash DE. We have shown that the model with the lowest squared-error (of those tested) is a model that allows for different first- and subsequent-stroke DEs, and also allows first stroke DE to depend on flash multiplicity.

By applying this new method we implicitly assume that the underlying "true" multiplicity distribution is the same for both periods, and that no other factors contributed to the measured multiplicity values. It is further important to note that the stroke to flash grouping algorithm and its configuration should be the same for both periods.

The new method seems to be quite sensitive to "additional" (erroneous) single-strokes flashes. These may occur as a result of misclassified cloud strokes, outliers, limits in the stroke-to-flash grouping algorithm, and limits to the model that relates the two multiplicity distribution to each other. It is clear from this work, as well as other work related to DE models, that modelling allows us to identify and evaluate anomalies in real-world lightning location data.

Further investigations are necessary to fully understand the interaction between measurement errors and modelling errors. We also plan to explore the use of a weighting matrix in the optimization algorithm, allowing us to accommodate different errors associated with different multiplicity values.

Appendix A: Nonlinear Least Square Optimization

The nonlinear Least Square analysis used to get the two unknown uses the following procedure:

1) Estimate start values p_{first}^0, p_{sub}^0. Written as Matrix the

$$p^0 = \begin{vmatrix} p_{first}^0 \\ p_{sub}^0 \end{vmatrix} \tag{A1}$$

2) Calculate matrix A, which is the gradient of Eq. (6) with respect to the "p" values:

$$A = \begin{vmatrix} \frac{dN_{flow}(1)}{dp_{first}} & \frac{dN_{flow}(1)}{dp_{sub}} \\ \frac{dN_{flow}(2)}{dp_{first}} & \frac{dN_{flow}(2)}{dp_{sub}} \\ . & . \\ . & . \\ \frac{dN_{flow}(m)}{dp_{first}} & \frac{dN_{flow}(m)}{dp_{sub}} \end{vmatrix} \tag{A2}$$

3) Solve the following equation for the change in the "p" vector: L= $[\, N_{flow}^{n} - N_{flow}^{n-1} \,]$

$$A^T * A * p = A^T * L \tag{A3}$$

4) Correct previous "p" vector for the nth iteration using $p^n = p^{n-1} + p$.

5) Iterate until $| p^n – p^{n-1} | < \epsilon$, where ϵ is sufficiently small so that the gradient approaches zero.

Appendix B: The overall stroke DE

We define M as the "true" average flash multiplicity. This value is defined as the total number of strokes in a dataset (S_T), divided by the total number of flashes (F_T):

$$M = \frac{S_T}{F_T} \tag{B1}$$

Given an imperfect measurement system, not all strokes or flashes will be detected. We therefore define the measured average multiplicity (m) as the total number of *detected* strokes in a dataset divided by the total number of *detected* flashes. The detected quantities are simply the true quantities multiplied by their respected detection efficiency fractions (DE_s and DE_f), therefore:

$$m = \frac{S_T \cdot DE_s}{F_T \cdot DE_f} \tag{B2}$$

Given the definition of M in Eq. (B1), we can rewrite this equation as

$$m = \frac{M \cdot DE_s}{DE_f} \qquad \text{(B3)}$$

An alternate expression for the total number of detected strokes is the fraction of detected first strokes plus the fraction of subsequent strokes. This is accomplished by defining a first-stroke DE (DE_1) and a subsequent-stroke DE (DE_{su}), and recognizing that the number of subsequent strokes in a dataset is simply the number of flashes multiplied by (M-1). Therefore the total number of strokes can be expressed as

$$F_T \cdot DE_1 + F_T \cdot (M-1) \cdot DE_{su} = F_T \cdot (DE_1 + (M-1) \cdot DE_{su}) \qquad \text{(B4)}$$

Substituting Eq. (B4) into Eq. (B2) yields

$$m = \frac{DE_1 + (M-1)DE_{su}}{DE_f} \qquad \text{(B5)}$$

Note that Eqs. (B3) and (B5) are different forms of the same equation. Equating the right-hand-side numerators of these equations and re-arranging terms yields Eq. B6.

$$DE_s = \frac{DE_1 + (M-1)DE_{su}}{m} \qquad \text{(B6)}$$

References:

CIGRE Task Force C4.404A: Cloud-to-Ground Lightning Parameters Derived from Lightning Location Systems: The Effects of System Performance (in final review, November 2007).

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, Sao Paulo, Brazil, November 7-11, 2004.

Orville R.E., G.R. Huffines, W.R. Burrows, R.L. Holle, K.L.Cummins: The North American Lightning Detection Network (NALDN) — First Results: 1998–2000, Monthly Weather Review, Vol. 130, 2002.

Rubinstein M.: On the determination of the flash detection efficiency of lightning location systems given their stroke detection efficiency. EMC, 1995.

W. Schulz, K. Cummins, G. Diendorfer, and M. Dorninger: Cloud-to-Ground Lightning in Austria: A 10-years study using data from a Lightning Location System, Journal of Geophysical research, vol. 110, 2005