LIGHTNING FLASH MULTIPLICITY MEASUREMENTS BY THE U. S. NATIONAL LIGHTNING DETECTION NETWORK

Craig J. Rodger[(1)], **Noah A. Russell**[(2)]

[(1)] *Department of Physics, University of Otago, PO Box 56, Dunedin, New Zealand; crodger@physics.otago.ac.nz*
[(2)] *Department of Psychology and Neuroscience Research Centre, PO Box 56, University of Otago, Dunedin, New Zealand; noah@psy.otago.ac.nz*

ABSTRACT

The National Lightning Detection Network (NLDN) detects the majority of cloud-to-ground lightning occurring over the United States. As part of a 1994 upgrade, a new algorithm was employed to group strokes into a flash. We find that the algorithm successfully classifies the majority of subsequent strokes into their appropriate flash. However, a small percentage of subsequent strokes are misclassified as independent flashes. An improved algorithm for grouping cloud to ground (CG) strokes into flashes is suggested, based on a statistical flash definition. Our study indicates that previously recognised under-reporting of flush multiplicity is almost certainly due to NLDN detection efficiency.

INTRODUCTION

Today, commercial lightning location networks operate in many regions of the world, using multiple stations to locate the source of lightning discharges through their electromagnetic radiation. An example is the United States National Lightning Detection Network (NLDN), which in 1996 used 106 sensors located over the continental United States to achieve a typical location accuracy of 0.5 km [1]. The NLDN network detects lightning discharges through the electromagnetic radiation from high-current pulses in the discharges, termed strokes. It is conventional to refer to the total lightning discharge as a flash (lasting ~1 s), made up of various discharge components, among which are the ~1 ms strokes [2]. The multiplicity refers to the number of strokes in a flash. In 1994 NLDN was upgraded to improve location accuracy and detection efficiency, involving changes to detection sensors and new processing techniques [1]. We believe this algorithm is still currently in use. As part of the 1994 NLDN upgrade, a modification was made to the algorithm which groups individual cloud to ground (CG) strokes into a CG flash and computes the flash multiplicity. A new algorithm was applied, grouping strokes into flashes using spatial and temporal clustering. The post upgrade average flash multiplicity ranged from 1.9-2.1 over 2 years [1], reasonably similar to the average flash multiplicity of 2.6 observed in south-eastern Brazil [3].

However, it is generally thought that the actual average flash multiplicity in CG lightning is ~3.5-4.5 [4]. It has been suggested by [1] that the lower average flash multiplicity reported by NLDN is due to subsequent strokes having peak currents which are on average somewhat less than half of the first stroke values [2], and hence too weak to be detected by NLDN.

The change in flash multiplicity algorithm from angle-based towards a spatial and temporal clustering method lead to a significant decrease in the average flash multiplicity, moving further away from the generally accepted value of ~4.5 strokes per flash. After the NLDN upgrade in 1994, the flash count increased from 17.4 million yr^{-1} over 1989-1993 to 22.7 million yr^{-1} in 1995 [5], raising the possibility that some of the events classified as flashes by the current flash multiplicity algorithm might be better classified as subsequent strokes inside a flash. In this paper we consider the appropriateness of the post-upgrade algorithm used to determine flash multiplicity and its implications on flash detection statistics after the upgrade. Based on this examination an improved flash algorithm is suggested.

DATA SET

The lightning data used in this study was collected by the NLDN continuously over the months of June and July, 1996, and covers all NLDN-observed CG flashes for the entire continental United States. During the months of June and July 1996, 11.4×10^6 flashes were observed, with an average flash multiplicity of 2.1. In 1994, June and July were the most active months of the year, in which ~50% of the year's flashes were detected [5]. This data was made available in "extended ASCII output" format to researchers taking part in the Colorado centred SPRITES'96 Campaign [6]. Note that this data has been pre-processed by NLDN into flashes using the spatial and temporal clustering algorithm described below. A higher accuracy study of the type presented here, should be undertaken using unprocessed data. However sufficient funding was not available to purchase such a large dataset.

In the 1994 NLDN upgrade, a multiplicity algorithm was employed that grouped strokes into flashes using spatial and temporal clustering. To be included in a flash, strokes must lie within 10 km of the first stroke, occur within 1 s of the first stroke and also no longer than 0.5 s after the immediate previous stroke in that flash. Strokes that qualify for inclusion in more than one flash are placed into the closest flash. Strokes located between 10 km and 50 km from the first stroke are included in the flash if the strokes are not clearly separated from the flash because the estimates of their location confidence regions overlap. The maximum flash multiplicity limit is taken as 15 strokes. The reported flash location and peak current estimate reported are those of the first stroke; subsequent strokes included in the flash are not required to have the same polarity as the first stroke.

EXAMINATION OF FLASH DATA

Performance of the NLDN spatial and temporal clustering algorithm

In order to examine the effect of the post-upgrade flash multiplicity algorithm on NLDN observations we examine the time difference between observed CG flashes. We expect a subsequent stroke that has been misclassified as a flash will be spatially and temporally close to the flash which it is in fact part of, while any two correctly classified flashes, which we will term a "CG pair", should have essentially random timings relative to one another. The timing differences between NLDN detected CG pairs are shown in Fig. 1, where a 3 ms running average has been used to clarify features in the CG pair timing differences. Here all flashes reported by NLDN have been included in the calculation of CG pair timing differences as long as they occur within 3 s of one another.

Clearly the majority of CG pairs included in Fig. 1 are independent of one another, leading to an approximately constant number of CG pair timing differences (~39500 pairs/ms). This indicates that the post-upgrade NLDN multiplicity algorithm correctly identifies the majority of subsequent strokes detected by the network. However, there are also distinct deviations in Fig. 2 from the anticipated uniform distribution. The first of these is the smaller than expected number of CG pairs for low timing differences (1-10 ms), as much as 8200 pairs/ms (~21%) less than the mean. Subsequent examination (not shown) indicates this decreases with separation distance between the CG pair, becoming insignificant >~2000 km. It appears that the NLDN is not detecting a small number of CG flashes that occur ≤10 ms after another flash. This is consistent with the known performance of the sensors which make up the NLDN, which have a dead time of approximately 5 ms [Ken Cummins, Global Atmospherics, personal communication, 2000], leading to a slight decrease in NLDN detection efficiency for CGs occurring ≤10 ms, even though these discharges may be separated by >500 km.

In addition to the deviation for CG pairs with very small (1-10 ms) timing differences, Fig. 1 shows 3 other significant deviations from the uniform background: an excess of CG pair time differences from ~10 to ~250 ms, a sharp increase beginning at 500 ms that decays away over ~700 ms, and finally, a smaller increase beginning at 1 s. These deviations are potentially caused by the misclassification of strokes by the current NLDN algorithm, which are then listed in our data-set as distinct flashes. The increases in CG Pairs in Fig. 2 at 0.5 s and 1 s are particularly suggestive, as these are coincident with temporal thresholds in the NLDN algorithm.

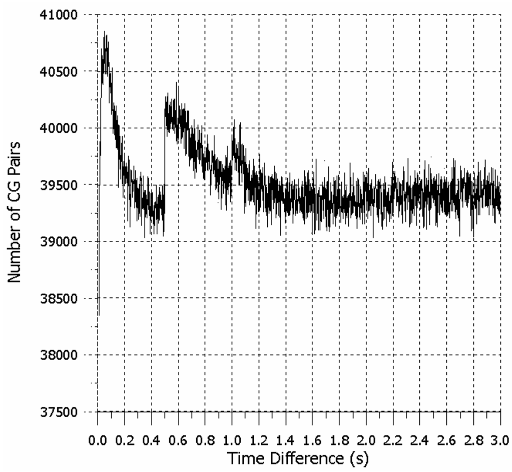

Fig. 1. The number of NLDN detected CG flash pairs versus the timing difference between the beginnings of each flash. A 3 ms running average has been used to clarify features in the CG pairs.

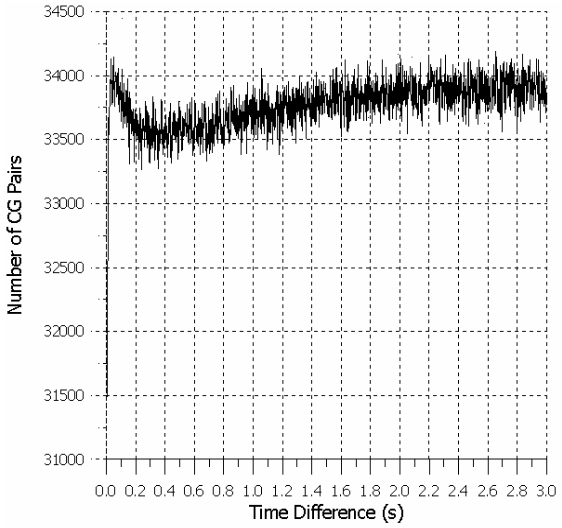

Fig. 2. The timing difference between CG flash pairs after the NLDN flash data has been reprocessed using the variant three modified multiplicity flash algorithm.

Modified flash multiplicity algorithms

Guided by Fig. 1, we have attempted to alter the definition of subsequent strokes to decrease the deviations shown in Fig. 1 and hence produce an improved algorithm for the classification of subsequent strokes into flashes. Our initial hypothesis was that strokes located at a distance >10 km from the first stroke might be misclassified as distinct flashes, and hence explain the greater numbers of CG pairs with timing differences <1 s. We therefore applied a new spatial window, defined by a flash spatial clustering radius, *r*=20 km, rather than *r*=10 km used by the current flash multiplicity algorithm. The strong peak in CG pair number occurring with separations of ~50 ms was reduced by the application of the variant one algorithm, so as to be approximately the same level as the constant level of CG pairs for separations >2 s (not shown) . This algorithm removed the sharp increase beginning at 500 ms, suggesting that this increase is due to subsequent strokes occurring more than 500 ms after the first stroke. Such strokes occur beyond the 0.5 s cutoff allowed for inter-stroke time differences, and as such are defined by the post upgrade algorithm as distinct flashes. However, the variant one algorithm leads to a significant decrease (an over-compensation) in the number of CG pair numbers with temporal separations beyond the ~50 ms peak, so that lightning discharges that are close spatially (≤20 km) and temporally (~100-1000 ms) from the initial stroke of an unrelated flash may be incorrectly grouped into that flash.

Having examined a number of possibilities, we propose a new algorithm variant where we relax the maximum temporal length of a flash (currently 1 s). In this algorithm the spatial clustering radius in variant three decreases linearly with time from (*t*=0 s, *r*=20 km) until the radius reaches zero (*t*=2 s, *r*=0 km). The timing difference between CG flash pairs after reprocessing using the new multiplicity algorithm (Fig. 2) shows considerably less variation than seen in the first modified algorithm (above), or for the multiplicity algorithm currently used by NLDN (Fig. 1), which we term the post-upgrade flash multiplicity algorithm. Most of the previously misclassified subsequent strokes are now successfully grouped into flashes, and the deviations from independent flash activity are now rather small.

With reference to the modified flash multiplicity algorithm variants, we can interpret the performance of the current multiplicity algorithm. The deviations in CG pair timing differences shown in Fig. 1 are strongly driven by the algorithm currently used to include strokes into a distinct flash. The excess of CG pair time differences from ~10 to ~250 ms is primarily due to subsequent strokes, physically associated with the first stroke, which are located more than 10 km from the first stroke. The sharp increase beginning at 500 ms is caused by subsequent strokes that are separated in time from the first stroke by more than the 0.5 s allowed for inter-stroke time differences. Finally, the smaller increase beginning at 1 s is due to subsequent strokes occurring outside the 1 s flash duration limit. The majority of misclassified subsequent strokes with timings <0.5 s after the first stroke are caused by the choice of spatial windowing, while those strokes which are >0.5 s after the first stroke are primarily misclassified through the choice of maximum inter-stroke time difference (the later events are almost entirely ≤10 km from the first stroke). All of these events represent a small number of subsequent strokes that are not recognized as such by the post upgrade NLDN algorithm, and are therefore listed as distinct flashes.

The post-upgrade flash multiplicity algorithm effectively defines a flash as a collection of strokes lasting no longer than 1 s, separated in time by no more than 0.5 s, and all located within 10 km from the first stroke. In the development of the modified multiplicity algorithms above we have taken a different approach to the definition of a flash. We treat a flash as a collection of strokes that are spatially and temporally close to one another (on spatial scales similar to the post-upgrade algorithm) but that are also statistically independent from other flashes. Such an approach may allow greater understanding of the discharge process, particularly when very large sets of "raw" stroke data are examined.

NLDN detection of CG subsequent strokes

The NLDN flash data for June and July 1996 has been reprocessed using the second modified algorithm presented above to classify subsequent strokes into flashes. The reprocessed flash data contains 10.6×10^6 flashes, with an average flash multiplicity of 2.3. The modified algorithm for the classification of subsequent strokes into flashes put forward in this paper causes a small rise in the mean flash multiplicity (2.1 to 2.3). This change is not sufficient to explain the previously recognised under-detection by NLDN of flash multiplicity. Our study confirms that under-reporting of flash multiplicity is almost certainly due to NLDN detection efficiency, as the number of non-independent discharges present in the data set is low. From this we can conclude that the network is almost certainly not detecting the "missing" strokes, rather than misclassifying them.

DISCUSSION

After the NLDN upgrade in 1994, the flash count increased by ~30% to 22.7 million yr^{-1} in 1995 [5]. Our analysis suggests that a small proportion of the NLDN detected flashes would be better classified as subsequent strokes. Using the June and July 1996 data as a proxy for the yearly lightning activity and the modified algorithm for the classification of subsequent strokes into flashes put forward in this paper we estimate that the yearly NLDN flash count in 1995 should have been 21.1 million yr^{-1}. As is clear, the post-upgrade algorithm successfully classifies the majority of subsequent strokes into their appropriate flash, although our analysis suggests that ~7% of flashes currently NLDN reported are probably subsequent strokes rather than distinct flashes. The application of our approach to raw stroke data should enhance our understanding of the relationship between strokes and flashes. This would also enable an improvement in the modified algorithm presented here, which currently slightly over-compensates, grouping truly independent events into flashes. We expect that a non-linear spatial-temporal clustering function would be optimal to decrease all deviations seen in Fig. 2, based on joint probability distributions.

CONCLUSIONS

In this paper we consider the algorithm employed as part of the 1994 NLDN upgrade which groups individual CG strokes into a CG flash made up of multiple strokes. We find that the algorithm successfully classifies the majority of subsequent strokes into their appropriate flash. However, a small percentage of subsequent strokes appear to be misclassified as independent flashes. An improved algorithm for grouping CG strokes into flashes is suggested, based on a statistical flash definition and using a clustering radius around the first stroke that shrinks linearly with time at -10 km s^{-1}. The reprocessed flash data leads to an estimated 7% decrease in yearly flash numbers. It has been recognised that the mean flash multiplicity reported by NLDN is about half that which is generally accepted as typical (~2 rather than ~4). While the improved flash multiplicity algorithm leads to an increase in the average flash multiplicity (2.1 to 2.3), this change is not sufficient to explain the previously recognised under-detection by NLDN of flash multiplicity. We conclude that the network is almost certainly not detecting the "missing" strokes, rather than misclassifying them.

Acknowledgments. This work was supported by the New Zealand Marsden Research Fund contract LFE801. The authors would like to thank Global Atmospherics of Tucson for providing the lightning data for this study.

REFERENCES

[1] K. L. Cummins et al., A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network, *J. Geophys. Res*, 103, 9035-9044, 1998.
[2] M. A. Uman, *The Lightning Discharge, Int. Geophys. Ser.*, Vol. 39, Academic Press, San Diego, Calif., 1987.
[3] O. Pinto et al., Cloud-to-ground lightning flash characteristics in southeastern Brazil for the 1992-1993 summer season, *J. Geophys. Res.*, 101, 29627-29635, 1996.
[4] V. A. Rakov, M. A. Uman, and R. Thottappillil, Review of lightning properties from electric-field and TV observations, *J. Geophys. Res.*, 99, 10745-10750, 1994.
[5] R. S. Wacker, and R. E. Orville, Changes in measured lightning flash count and return stroke peak current after the 1994 U.S. National Lightning Detection Network upgrade 1. Observations, *J. Geophys. Res*, 104, 2151-2157, 1999.
[6] C. J. Rodger, Red sprites, upward lightning and VLF perturbations, *Rev. Geophys.*, 37, 317-336, 1999.