Thunderstorm characteristics from cluster analysis of lightning

Michael L. Hutchins[1]* and Robert H. Holzworth[1]

[1]Department of Earth and Space Sciences, University of Washington, Seattle Washington, U.S.A.

ABSTRACT: Application of clustering algorithms to ground based lightning detection networks expands the real time global observations of lightning from strokes to flashes and strokes to thunderstorms. Lightning detection networks, such as the World Wide Lightning Location Network (WWLLN) or the Earth Networks Total Lightning Network (ENTLN), are then able to identify, locate, and analyze nearly every active thunderstorms within their operational range. Global thunderstorm information allows for research into the climatological structures of thunderstorm behavior on large spatial and temporal scales. Flash clustering allows for new network diagnostics, such as flash multiplicity and thunderstorm detection efficiency.

An algorithm is used to cluster strokes to flashes and thunderstorms for both the WWLLN and ENTLN. Cross validation of the networks is performed with the located thunderstorms and comparisons of their inferred areas and duration. Overall WWLLN detects 61% of all ENTLN thunderstorm clusters and 80% of thunderstorms larger than 10^3 km^2. In ther reverse analysis, ENTLN detection of WWLLN thunderstorms, ENTLN detects 86% of all WWLLN thunderstorms over North America. On average WWLLN observes thunderstorm clusters lasting 10 minutes and spanning 66 km^2, ENTLN observes average durations of 10 minutes and 60 km^2. Within thunderstorms the average time between flashes is 21 seconds as seen by WWLLN and 10 seconds by ENTLN, with a strong regional dependence on season. Clustering algorithms applied to lightning detection networks allow for a new range of analysis of thunderstorm effects, network performances, and the links between lightning and thunderstorms.

DISCUSSION

Clustering the located strokes of both WWLLN and ENTLN networks provides thunderstorm and flash data. The clustering is performed with the Density-Based Spatial Clustering of Application with Noise (DBSCAN) algorithm [*Ester et al.*, 1996; *Kriegel et al.*, 2011] following the application metholodgy discussed in *Hutchins et al.* [2014]. DBSCAN is used over other clustering methods as it clusters based on the spatial and temporal distance between strokes with robust handling of noise (e.g. nearby thunderstorms). The flash clustering uses the same algorithm as the thunderstorm clustering, with adjustment of the spatial and temporal clustering parameters.

Within a thunderstorm cluster lightning detection networks are able to measure two properties: the interstroke and interflash timing. The full WWLLN interstroke time, Figure 1a, naturally shows two distinct peaks: one at 40 ms and another at 100 seconds with a natural inflection at 1 second. With ENTLN the interstroke distribution does not show the same distinct peaks (Figure 1c) because it is able to locate more strokes in each flash, the tail of the interstroke time distribution overlaps the interflash time distribution. In the WWLLN interevent time distribution there is a spike of events near 10 μs caused by the same event recorded twice by the network, this can also occur with ENTLN but less frequently.

After the network events are clustered into flashes the time distributions can be split into the time between strokes in the same flash (black) and time between flashes (blue) in Figure 1b and 1d. The stroke and flash distribution can be fit as lognormal distributions for each day of the year, an example set of fits is shown with the dashed lines in Figure 1b and 1d. This fitting allows for the daily tracking of the interstroke (Figure 2a) and interflash (Figure 2b) times for both networks. With WWLLN the global distributions

*Corresponding author, email: mlhutch@uw.edu, Postal address: University of Washington, Box 351310, Seattle WA 98195-1310 U.S.A.

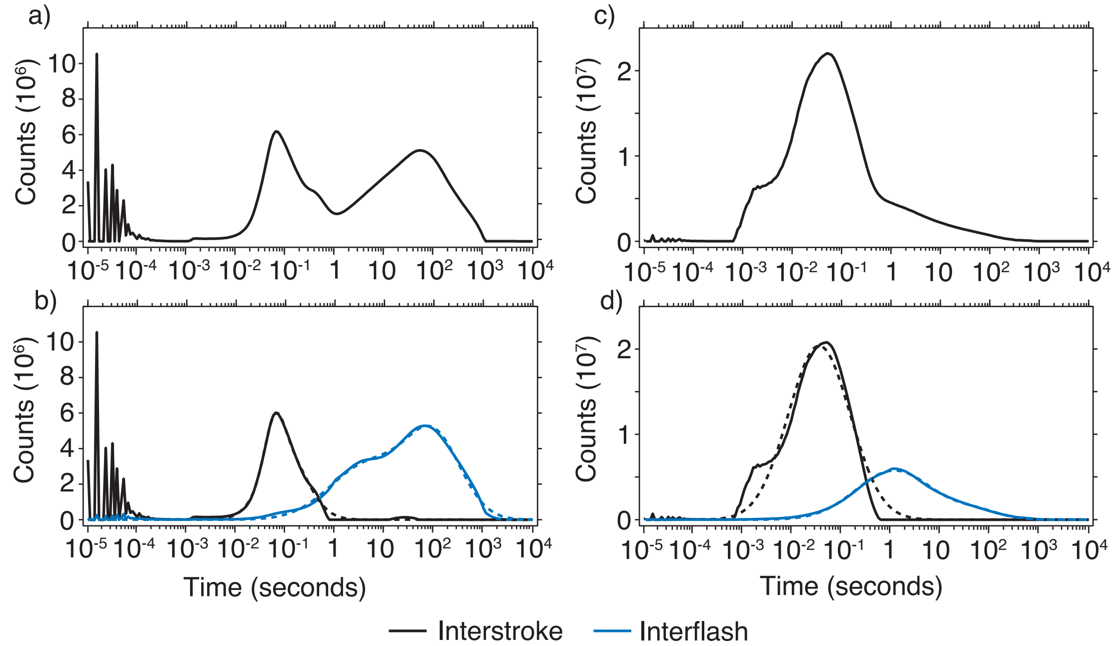

Figure 1: The interevent times for WWLLN (a) and ENTLN (c) and the interstroke (black) and interflash (blue) time distributions for WWLLN (b) and ENTLN (d). The dashed lines correspond to the best lognormal fits of the distributions.

(black) remain relatively centered at 71 ms and 100 seconds; over North America WWLLN averages (blue) are 60 ms for interstroke and 39 seconds for interflash times. ENTLN (red) has lower daily averages of 53 ms and 17 seconds due to the higher detection efficiency. The North American WWLLN distribution more closely matches the seasonal behavior present in the ENTLN distribution with a small offset in timing.

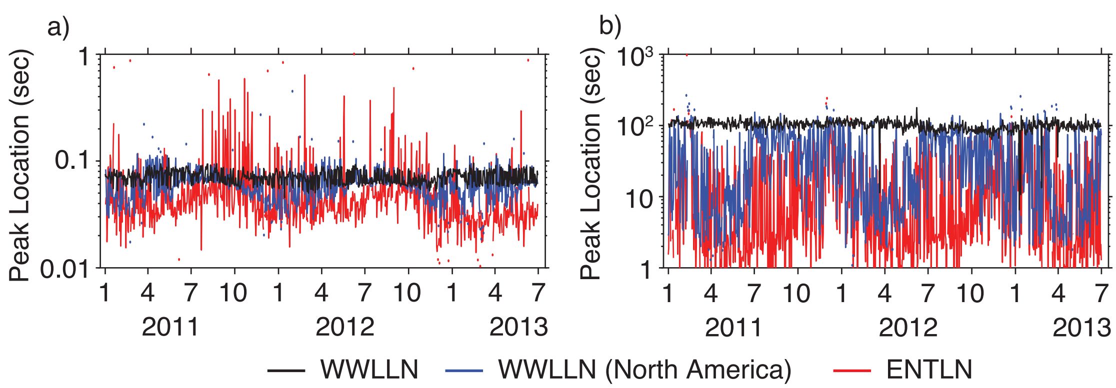

Figure 2: Peak of interstroke (a) and interflash (b) times for WWLLN (black), WWLLN over North America (blue), and ENTLN (red). Points beyond 2^{th} and 98^{th} percentiles shown as dots for WWLLN North America and ENTLN.

Compared to ENTLN, WWLLN performs better when detecting large and active thunderstorm regions. The lower performance for smaller thunderstorm clusters may be due to the performance of WWLLN, the performance of ENTLN, or the cutoff in the clustering that requires at least 3 strokes for a cluster, removing 39% of strokes from the analysis. Including these strokes would lead to less robust matches between the networks and conflate a stroke to thunderstorm detection efficiency and the thunderstorm to thunderstorm detection efficiency. For the thunderstorms WWLLN does detect, the characteristics of the thunderstorm are on par with those of ENTLN.

ACKNOWLEDGMENTS: The authors wish to thank the World Wide Lightning Location Network (http://wwlln.net), a collaboration among over 50 universities and institutions, and the Earth Networks Total Lightning Network for providing lightning location data used in this paper.

References

Ester, M., H. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise., in *2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996.

Hutchins, M. L., R. H. Holzworth, and J. B. Brundell, Diurnal variation of the global electric circuit from clustered thunderstorms, *Journal of Geophysical Research - Space Physics*, (119), 620–629, doi: 10.1002/2013JA019593.Received, 2014.

Kriegel, H.-P., P. Kröger, J. Sander, and A. Zimek, Density-based clustering, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(3), 231–240, doi:10.1002/widm.30, 2011.