2018 25th International Lightning Detection Conference & 7th International Lightning Meteorology Conference
March 12 - 15 | Ft. Lauderdale, Florida, USA

Comparison of two cloud to ground stroke datasets- Vaisala and WSI

Kathy Gray
Department of Math and Statistics,
California State University-Chico
Chico, CA USA

Don Latham
Retired-USFS
Six Mile Systems LLC
Missoula, MT USA

Paul Sopko
Missoula Fire Sciences Laboratory
US Forest Service, Rocky Mountain Research Station
Missoula, MT USA
psopko@fs.fed.us

***Abstract*— Two lightning stroke data sets for the same time period from a 7854 sq.km. area in the Black Hills, South Dakota USA , one from Vaisala and the other from WSI (Weather Services International now The Weather Company) were compared with respect to number of events, number of positive and negative strokes, amplitude and number of subevents. There was also an attempt to match strokes between the two data sets. The first stroke recorded for each data set took place on June 15 2012 at 3:49 PM MDT and the last stroke for each data set occurred on June 26 2012 at 2:58 PM MDT. The WSI data set has 3974 strokes while the Vaisala data set has 3342 strokes. The data from each set was split into negative and positive stroke sets and were compared with respect to count and match. Daily time plots for negative and positive strokes show more negative strokes for Vaisala with Vaisala showing a much larger negative to positive ratio than WSI. The spatial distribution of strokes that matched up with respect to a specified time criteria are discussed. This comparison suggests that WSI over-reports cloud strokes as ground strokes and records higher peak amplitudes with respect to the Vaisala data.**

I. INTRODUCTION

Locations for the termini of cloud-to-ground lightning events are used by land management agencies, power companies, municipalities, and providers of weather information to public and private entities. In the case of federal land management agencies, a request for proposal is usually issued, and evaluated for the criteria stated therein. Two sources, Vaisala Inc., (https://www.vaisala.com/en/products/data/data-sets/nldn) and WSI (Weather Services International) were chosen in the past. At the time of this study, WSI data was provided by TOA systems (http://toasystems.com/) and was used in the United States Precision Lightning Network (USPLN). Changes to the TOA data made by WSI may be present in the supplied dataset. Note that WSI is now the Weather Company, an IBM business, and currently uses lightning data from Vaisala INC.

Each of these sources has to some degree provided information as to location accuracy (ground truth) for its products, as stated in their website information. This small study provides a comparison of the outputs of these two products over common spatial and temporal fields.

A direct comparison of the Vaisala (NLDN) and WSI (USPLN) products does not seem to be available although Jacques et.al., (2011) did a comparison of the WSI (USPLN) network to NASA's Cloud-to-Ground Lightning Surveillance System (CGLSS-II).

A very broad classification of lightning events separates those with ground termini, or CG, from in-cloud or cloud-air events, or IC. CG events are further separated into positive and negative according to the sign of charge lowered to ground by the discharge(s) in the event. A discussion of nomenclature may be found in Cooray (2015). For this paper, we simply note that a lightning event to ground is termed a flash, and that a flash may be composed of one or more discharges called strokes (Rakov, 2013). Most users are interested in the timestamped location of a ground terminus, whether the discharge is positive or negative, the value of the electric current flow in the discharge, and a measure of the location accuracy. For most users, discharges in IC events are not germane, and are not included in the data stream. If only flashes are reported, the timestamp and location will be that of the first (or only) stroke, and the number of strokes may be included in the data.

It might seem that a direct event-for-event comparison between systems is not possible because ground truth data of the actual event location(s) are not known. However, each of the systems uses GPS timestamping. Basically, each station in each system assigns a local GPS timestamp to a detected discharge. Timestamped data are used together with selection criteria at a central processor to determine the type of discharge, calculate its location, and form the event. Events can thus be directly compared through selection of time windows and discharge details. Also, area wide statistics for common time windows may be calculated.

This study uses data as summarized below, and applies summary statistics as indicated, to compare and contrast the data from two sources. The questions to be answered include: Would a user see a difference between the two products? If so, could the difference be sufficient enough that decision-making would be altered?

II. DATA

Data were obtained from two lightning detection networks commonly used in North America: Weather Services International (WSI, currently The Weather Company) derived from their United States Precision Lightning Network (USPLN) and Vaisala obtained from the familiar National Lightning Detection Network (NLDN). These lightning stroke datasets came from an area in the Black Hills of South Dakota and Wyoming USA and were constructed from a circle centered at 43.91478N, -10428469W with a 100 mile radius (Figure 1). A chart comparing some basic attributes of both networks is shown in Table 1. Aside from the slight difference in sensor baselines and lack of WSI error ellipse data, the two networks are similar in most respect, making this comparison simpler.

Both the Vaisala and WSI data sets were reduced so that they cover the exact same time frame. The time frame selected for comparison was chosen for similarities in both data sets for the first and last stroke event. The first stroke recorded for each data set took place on June 15 2012 at 3:49PM MDT and the last stroke for each data set occurred on June 26 at 2:58PM MDT. The Vaisala (NLDN) dataset was obtained for another study and used with permission. The WSI (USPLN) data was ordered to match the Vaisala data both spatially and temporally.

Uman (2001) demonstrated that all lightning strokes have an interval time less than 160 milliseconds (ms) therefore for this study, we considered a stroke a subevent if it was less than 160 ms after the previous stroke. The Vaisala dataset contained 3342 strokes recorded during this time period while the WSI contained 3974 stroke events. A stroke histogram for both data sets used for this comparison is shown in Figure 2.

TABLE 1. Base attributes of NLDN and USPLN networks

Attribute	NLDN	USPLN
Scale	North America	North America
Sensor Baseline	~300km	~250km
Detection technique	Time of Arrival	Time of Arrival
Flash/Stroke	Stroke	Stroke
Error Ellipse Data	Yes	No

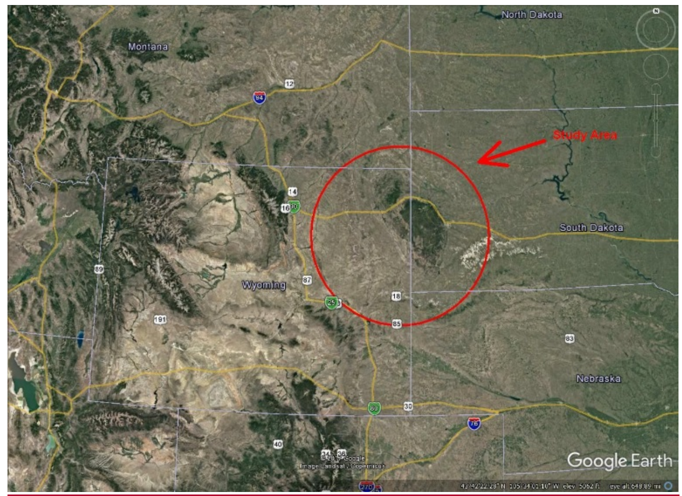

Fig. 1. Study Area

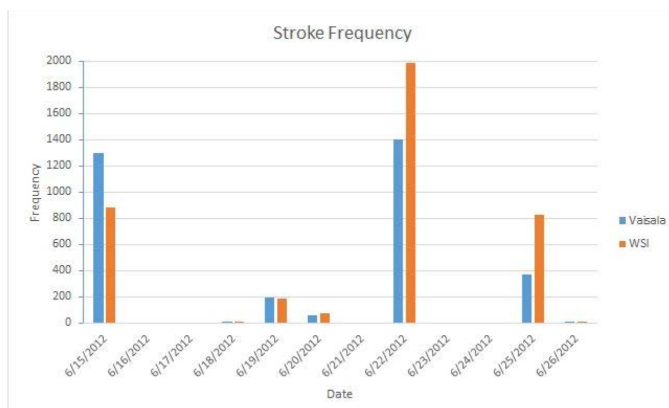

Fig. 2. Stroke histogram for WSI and Vaisala data used for comparison

For each event, the Vaisala dataset contained the following parameters: date/time (MDT), latitude, longitude (decimal degrees), amplitude (kA with polarity), chi-squared value, 99% semi-major error ellipse axis, 99% semi minor error ellipse axis, angle semi-major error ellipse axis and number of sensors participating in the stroke detection. The WSI parameters were: date/time, milliseconds, latitude, longitude (decimal degrees) and amplitude (kA with polarity).

III. Anaylsis

The strokes were split between positive and negative strokes based on the value for amplitude. Strokes in the WSI data sets were matched to individual Vaisala strokes by two criteria: (1) a time difference of 10 milliseconds or less and (2) polarity. If more than one Vaisala stroke was within 10 ms of a WSI stroke then the closest stroke in distance was considered a match. The difference between time and distance were calculated for the stroke matches and analyzed with summary statistics and graphs. Days were selected that had an abundance of lightning strokes (June 15,22 and 25) (Fig.2) and graphs were created to demonstrate the number of positive and negative strokes throughout the day for both Vaisala and WSI. The negative and positive ratio was calculated and graphs were used to compare the ratio across time between Vaisala and WSI. To avoid having a ratio with a denominator of zero a constant of one was added to each stroke number. Finally, stroke amplitude for matched pairs were compared to see if there were any differences in value. Some dates were excluded because of a complete lack of strokes on those days.

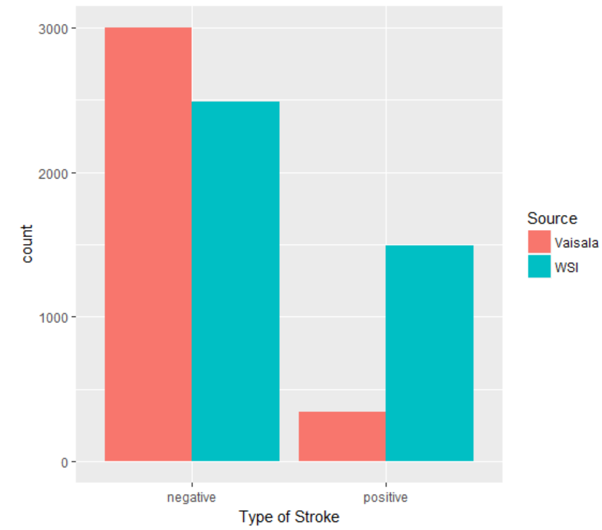

Fig. 3. Total number of positive and negative strokes

IV. Results

For the three selected days, the Vaisala data contained 2999 negative strokes and 343 positive strokes and the WSI data had 2484 negative strokes and 1490 positive strokes (Fig. 3). Forty-six percent (1381) of the 2999 negative Vaisala strokes were within 10 ms of a negative WSI strike. Thus, 54% (1618) of the 2999 Vaisala negative strokes were not matched while 44.4% (1103) of the 2484 negative WSI strokes were not matched. The average time difference between the matched negative strokes was 0.00057 seconds with a standard deviation of 0.0012 seconds (Table 2). Seventy-eight percent (268) of the 343 Vaisala positive strokes were matched to a WSI positive strike. Twenty-two percent (75) of the Vaisala positive did not match a WSI positive strike. Additionally, 82% (1222) of the WSI positive strokes were not matched. This was a result of having many more WSI positive strokes than Vaisala positive strokes (Fig.3). The average difference between times for WSI and Vaisala matched positive strokes was 0.0005 seconds with a standard deviation of 0.0009 seconds (Table 2). The histogram reveals that most matches were within 0.001 seconds of each other (Fig.4). Two peaks 1 ms apart implies truncation error in the reported timing data. The mean difference in distance between matched strokes was 3.072 km (sd=4.32 km) for negative strokes and 1.500 km (sd=1.16 km) for positive strokes (Table 2 and Fig. 5).

Seventy-eight of the 1618 unmatched negative Vaisala strokes were within 10 ms of an unmatched positive WSI stroke. Additionally, 54 of the 75 unmatched positive Vaisala strokes were within 10 ms of an unmatched negative WSI stroke.

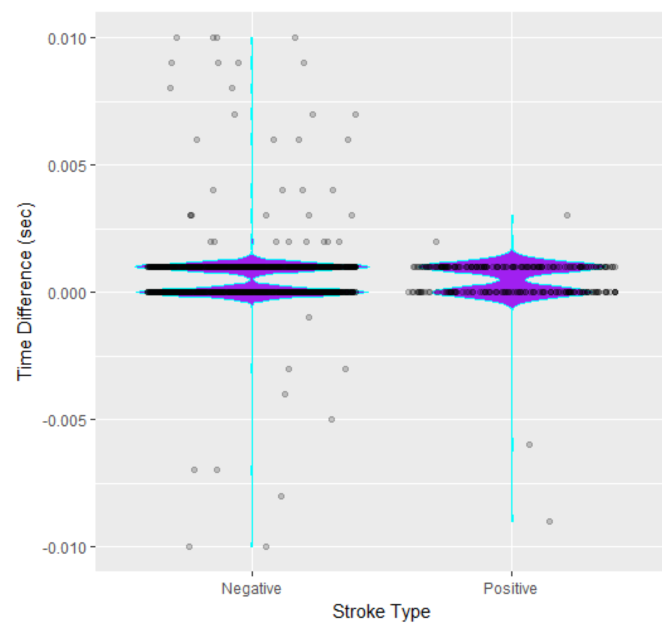

Fig. 4. The difference in time between Vaisala and WSI matched positive and negative strokes.

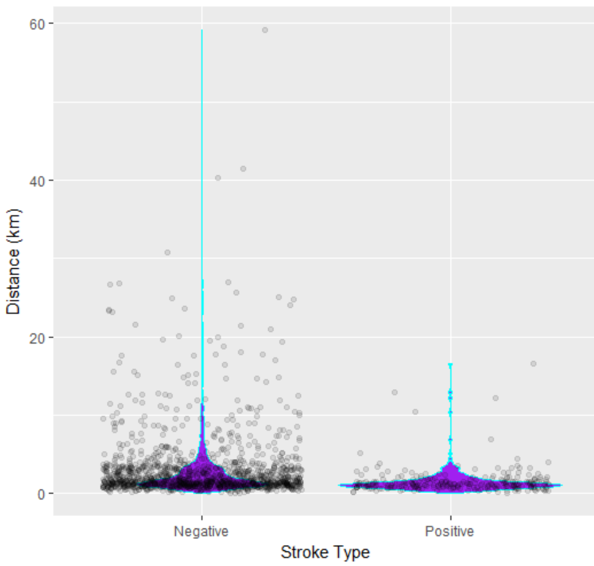

Fig. 5. Distance between Vaisala and WSI matched positive and negative strokes.

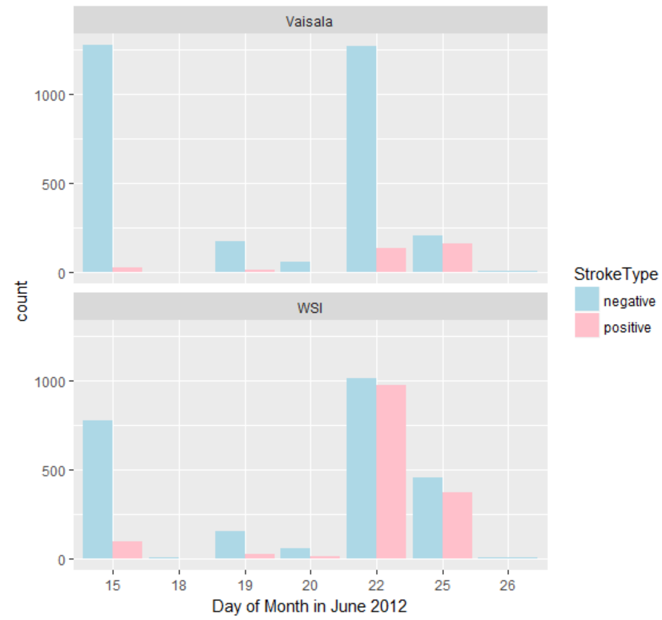

Fig.6. Positive and negative stroke numbers for Vaisala and WSI.

The number of strokes per day were summarized over the period of the study. On June 15, 2012, Vaisala reported 1302 total strokes while WSI reported only 880 total strokes (Fig. 6). Vaisala reported many more negative strokes (1275) than WSI (780) (Fig. 6). On the two other days, 6/22/2012 and 6/25/2012, WSI reported more total strokes. However, on 6/22/2012, WSI had more positive strokes than Vaisala and less negative strokes (Fig. 6). On 6/25/2012, WSI had more negative and positive strokes than Vaisala (Fig. 6).

The number of strokes across time per day was examined for the days 6/15/2012, 6/22/2012, and 6/25/2012. There were not enough strokes on the other days to consider additional analysis. In order to create a smoother graph, data were aggregated every ten minutes and the points represent number of strokes for each ten minute period. Fig. 7 shows the number of strokes and Fig. 8 displays the negative to positive ratio for the three days considered. On 6/15/2012 and 6/22/2012, Vaisala had a much larger negative to positive ratio (Fig. 8) throughout the day than WSI. Both Vaisala and WSI reported a small number of positive strokes for 6/15/2012 while on 6/22/2012, WSI reported a large number of positive strokes while Vaisala reported a low number of positive strokes throughout the day. On 6/25/2012, Vaisala and WSI had a similar negative to positive ratio of strokes throughout the day (Fig. 8). WSI reported both more positive and more negative strokes on 6/25/2012 (Fig. 7).

TABLE 2. Summary statistics for time (sec) and distance (km) between matched strokes.

Metric	Type	n	Mean	Median	Std. Dev	Min	Max
Time (sec)	Negative	1381	0.0006	0.001	0.0012	-0.01	0.010
	Positive	268	0.0005	0.001	0.0009	-0.009	0.003
Distance (km)	Negative	1381	3.072	1.664	4.32	0.0186	59.17
	Positive	268	1.500	1.155	1.16	0.0873	16.5

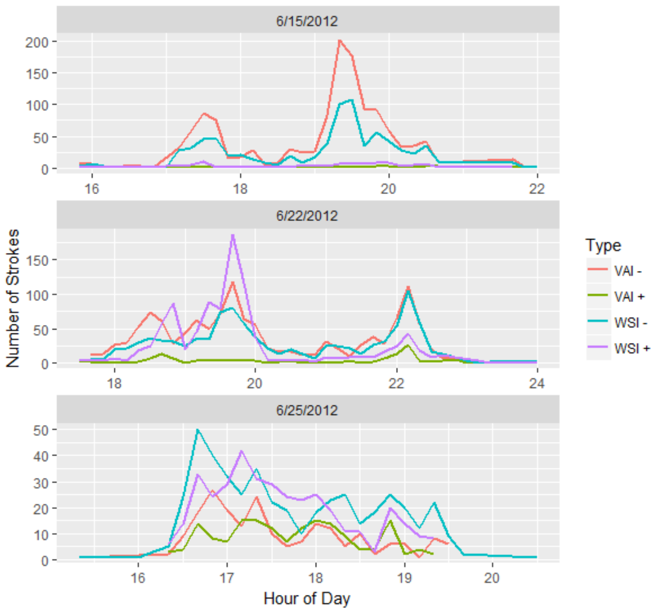

Fig. 7. Total number of strokes for selected days.

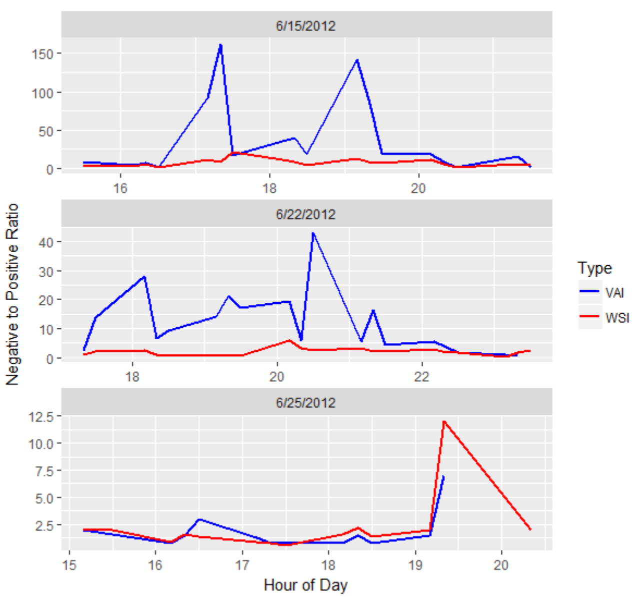

Fig. 8 Negative to positive ratio for selected days for Vaisala and WSI.

Amplitude for matched strokes was compared by subtracting the WSI recorded amplitude from the Vaisala amplitude for positive matched strokes and the reverse for negative matched strokes. Thus, positive differences meant that Vaisala had a larger (in magnitude) recorded amplitude. The distribution of differences appears to be bimodal for both positive and negative strokes (Fig. 9). The bulk of the differences are centered around zero; however, for negative strokes there is another group of strokes with differences of about -50 kA. Similarly, for positive strokes there is a cluster of strokes that have a difference with a mean of about -75 kA. For both negative and positive matched strokes, there is a tendency for WSI to have a larger recorded amplitude (Fig. 9). Fig. 10 shows that there were many more negative matched strokes for 6/15/2012 and

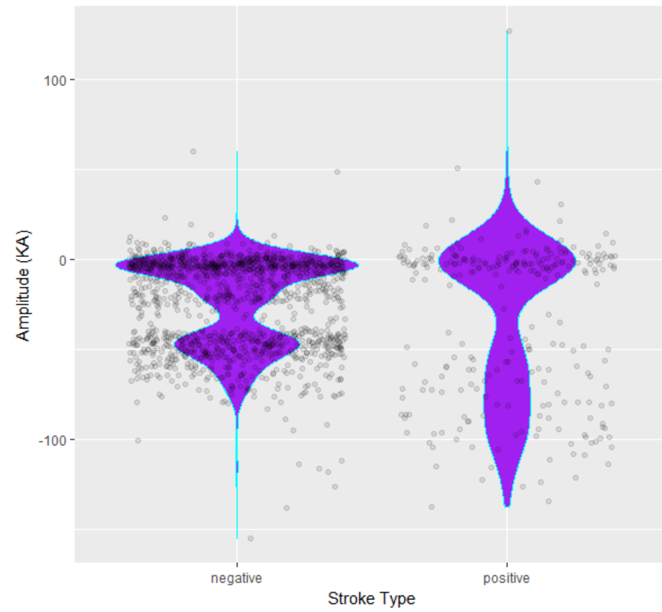

Fig. 9 Difference in amplitude between the matched strokes. For both positive and negative strokes, a positive difference represents a stroke where Vaisala had a stroke that was larger in magnitude.

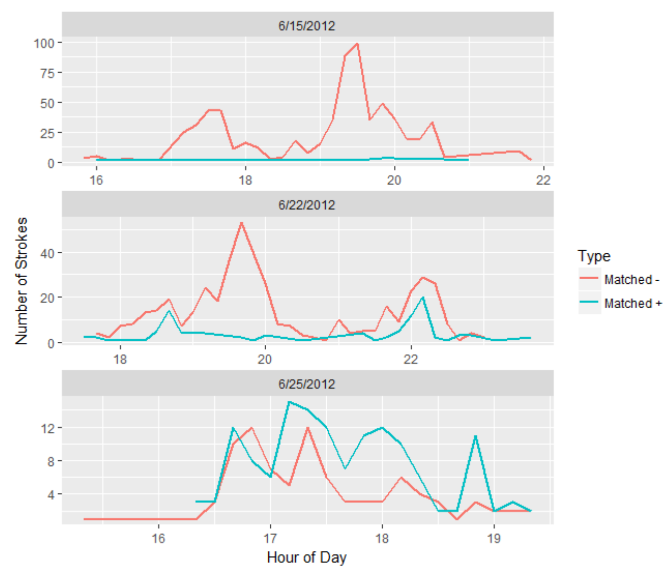

Fig. 10 Number of matched and unmatched strokes on selected days

6/22/2012 while on 6/25/2012 the number of matched strokes for positive and negative were about equal.

V. Conclusions and Discussion

The questions that were asked in this study were: would a user see a difference between the two products, Vaisala and WSI (USPLN)? If so, could the difference be sufficient so that decision-making would be altered?

If users were provided with comparison data from the two lightning detection systems for their area of interest, they would definitely see a difference in the number, timing, and location of reported lightning strokes. This study covers a small area in South Dakota and Wyoming. It does not include additional data, such as radar and satellite information about the thunderstorms on the dates studied. In addition, the study did not include the timelines and locations in a tool such as Google Earth which would provide a better, dynamic view of the data.

This study did discover some points of interest using the provided data. The WSI data had considerably more positive strokes than the Vaisala data (Figure 3), especially on the 15th and 22nd of June 2012. The larger number of positive strokes in the WSI data compared to the Vaisala data is similar to the findings in Jacques et.al. (2011) for WSI (USPLN) vs. the Cape Canaveral lightning array (CGLSS-II). The data in Figure 7, displaying number of strokes, and Figure 8, displaying ratios of negative to positive strokes, shows this quite clearly. WSI negative to positive ratios remain near 2.5 for all periods except the end of the day on 6/25. Note that on that day, after about 18:20 MDT, Vaisala had no strokes of either polarity, and WSI showed negative strokes for the rest of the day - a puzzle. Because the data were aggregated for 10 minute periods, any subtle timing with respect to polarity is washed out. Note that in general, WSI reports more positive strokes along with negative strokes than Vaisala. Could WSI be categorizing cloud strokes as positive ground strokes? Figure 10 appears to support this hypothesis because many positive strokes in the WSI data could not be matched to positive strokes in the Vaisala data.

In addition, Figure 9 indicates that the algorithm WSI uses to determine the stroke current provides in general a higher current for both positive and negative strokes than Vaisala. A bimodal structure is clear for of both positive and negative strokes. This figure also displays one of the difficulties with system comparisons.

Both systems presently use hyperbolic location algorithms to locate the position of the stroke terminus on the ground. The available system information does not, for understandable proprietary reasons, include either the location of the detection stations or the algorithms used for data processing, such as stroke type at the local station and/or at the central processor. GPS timing is definitely used at each station, and one or more timestamps and other information are passed to the central processor. Available datasets do not contain this information, so the two systems could not be compared to determine why these differences exist

For example, a good analysis of relative position on the ground can go no further than that given in Table 2. Although the data supplied by Vaisala includes error estimates for position, WSI data includes no such information. So, for matched strokes, only the statistics of 3 km mean difference and 4.32 standard deviation for negative strokes and 1.5 km mean and 1.16 standard deviation for positives are useful. Comparison of matched strokes using error statistics would be more useful. Vaisala error statistics, given in the data as major and minor ellipse axes, were almost all within 1 km by 1 km for both positive and negative strokes. So it would appear that the WSI data mat have higher location errors.

Even with a generous time matching window of 10 ms, approximately half of the negative strokes and one quarter of positive Vaisala strokes were not matched. Also, seventy-eight of the 1618 unmatched negative Vaisala strokes were within 10ms of an unmatched positive WSI stroke. Additionally, 54 of the 75 unmatched positive Vaisala strokes were within 10ms of an unmatched negative WSI stroke. There is no way to determine from the data why this is so, especially since local GPS timestamps can, with little effort, be easily determined in the neighborhood of 100 ns (Allan, 1997). Again, proprietary restrictions prevent analysis: there are no stroke structure details.

The data as received have implied accuracy of 1ms, clear from Table 2 and Figure 4, thus a matching window of 10ms seems appropriate. Two peaks 1ms apart in Figure 4 indicates round off error in the timing data. The window cannot be set too broad because interstroke intervals can be just somewhat larger (Cooray, 2015).

Using only the data as available to a typical user, then, it appears that the answer to the questions “is there a difference between the two products” and “is the difference sufficient so decision making would be changed” has to be “yes” and “yes”. Analysis of the data clearly shows the difference between the data sets. Clearly, the two systems “see” the same incoming electromagnetic signals and act upon them. But the outputs seem to be very different.

Decisions based on these data could be considerably altered, depending of course on the intended use. Electric utilities require as accurate a ground position as possible and may depend upon stroke amplitude as well. Land management agencies may, for purposes of wildland fire location and investigation, require accurate position. In addition, some of these agencies depend on differentiating between positive and negative stroke locations for fire search purposes. Other users may have quite different requirements. This small study will, hopefully, assist those responsible in obtaining the best data available for their use.

Acknowledgment

The authors would like to thank Vaisala and WSI for providing the data used in this study. We would also like to specifically thank Bob Keane and Bret Butler of the Missoula Fire Sciences Laboratory for providing funding for the statistical analysis and presentation of this paper.

REFERENCES

Allan, D.W., N. Ashby, and C.C. Hodge. (1997) The Science of Timekeeping, Hewlett Packard Application Note 1289.

Cooray V. (2015) An Introduction to Lightning. Pg. 114 ISBN 978-94-017-8937-0 doi 10.1007/978-94-017-8938-7

Jacques, A.A., J.P. Koermer, and T.R. Boucher, (2011) Comparison of the United States Precision Lightning Network™ (USPLN™) and the Cloud-to-Ground Lightning Surveillance System (CGLSS), Fifth Conference on the Meteorological Applications of Lightning Data, Amer. Meteorol. Soc., Seattle, WA.

Rakov V.A. (2013) The Physics of Lightning. Surv Geophys 34:701-729 doi 10.1007/s10712-013-9230-6

Uman M.A. (2001) The Lightning Discharge. Pg. 19 ISBN 978-80-486-4146-3