2014 23rd International Lightning Detection Conference
18 - 19 March • Tucson, Arizona, USA
5th International Lightning Meteorology Conference
20 - 21 March • Tucson, Arizona, USA

Method to Improve Location Accuracy of the GLD360

Ryan Said
Vaisala, Inc.
Boulder Operations
194 South Taylor Avenue, Louisville, CO, USA
ryan.said@vaisala.com

Amitabh Nag
Vaisala, Inc.
Boulder Operations
194 South Taylor Avenue, Louisville, CO, USA
amitabh.nag@vaisala.com

***Abstract*—An algorithm change to the central processor of the network producing the GLD360 dataset is proposed to reduce the population of events with large (>5 km) location errors. In this study, we compare the data reprocessed using the proposed algorithm change to that from the realtime GLD360 dataset. The relative location accuracy using the National Lightning Detection Network™ as a reference is evaluated using data from July 20, 2013. With respect to the realtime dataset, the median location error of the reprocessed dataset decreased from 3.0 km to 1.7 km, and the 90th percentile decreased from 15 km to 6.5 km. The variation of the relative location error of the reprocessed dataset with local time is also investigated. The median location error is found to double near the local sunrise hours. This degradation in location accuracy is correlated with a drop in lightning event polarity estimation accuracy.**

Keywords—lightning location systems; VLF propagation; GLD360; performance improvements; location accuracy

I. INTRODUCTION

The GLD360 dataset is generated by a long-range lightning detection network. The network is owned and operated by Vaisala, Inc., and has been operational since 2009. Recent validation studies against reference precision networks, such as the comparison against data from the National Lightning Detection Network (NLDN) in Said et al [2013], show a time-averaged median location accuracy of ~2—5 km. However, the location error distribution shown in Fig. 2a in Said et al [2013] is bimodal, with a significant population of events clustered at ~10--15 km error. Similar results are reported by Mallick et al. [2014, ILDC/ILMC] who use rocket-triggered lightning data as ground-truth to examine the performance characteristics of GLD360.

This paper is concerned with a method to improve the overall location accuracy of the GLD360 dataset. As discussed in Section II, the location accuracy is dependent on the correct identification of the event polarity, and on each sensor's ability to reliably identify a given feature in the measured waveform. The primary origin of the population of large error (>5 km) events reported in Said et al [2013] is feature identification errors at one or more sensors; the algorithm change proposed in Section III addresses these errors. Location accuracy errors caused by incorrect polarity estimation, however, are not addressed by the proposed algorithm change. In this paper, we discuss the underlying cause for these remaining polarity errors, seek to establish a pattern for their occurrence, and comment on future work that may help to reduce their numbers.

II. SUMMARY OF GLD360'S ARRIVAL TIME ALGORITHM

The proposed algorithm change discussed in Section III involves special treatment of certain arrival time measurements reported by individual sensors. To place the proposed change in context, this section begins with an overview of the arrival time measurement scheme currently used in the GLD360 dataset network. The algorithm used by the network to compensate for long-distance propagation is also detailed in Said et al [2010].

A. Description of Arrival Time Correction Algorithm

Using ground-based sensors sensitive to the Very Low Frequency (VLF; 3—30 kHz) band [Cohen et al., 2010], the network uses a waveform matching algorithm to identify distinct features of individual waveforms sampled from radio atmospherics (colloquially, sferics) emitted from cloud-to-ground (CG) strokes and large cloud pulses. The arrival time of one of these features, which may be the rising portion of the ground wave or a zero-crossing point located on the rising or falling edge of an ionospheric reflection, is transmitted back to the central processor (CP). Once the CP achieves an initial position estimate, it uses the approximate propagation distance to each sensor to apply a propagation correction factor based on the specific timing feature.

The waveform matching algorithm begins with a cross-correlation between the measured sferic and a locally stored waveform bank. Currently, the waveform bank is comprised of sferic waveforms from negative CG strokes indexed at distances ranging from 100 km to 6000 km. There is a separate waveform bank for daytime and nighttime propagation conditions. Each detected sferic is cross-correlated with the appropriate waveform bank using both the sferic and its negative. The peak cross-correlation of each polarity determines an estimated distance for that polarity. Once an initial location fix is determined at the CP, the true propagation distance to each sensor is compared with the estimated distances for each polarity. The polarity that results in the smallest estimated propagation distance error gives the estimated polarity from the respective sensor. The overall polarity for the event is determined from a weighted sum of the polarity estimates from each contributing sensor.

Once a polarity is chosen for a given stroke or cloud pulse, a propagation delay correction is applied to the waveform feature (such as a ground wave or subsequent zero-crossing) corresponding to the chosen polarity from each sensor. As an example, consider the 2D histogram of zero-crossing delays under daytime propagation conditions, plotted in microseconds after speed-of-light propagation, shown in In Fig. 1a, all misidentified zero-crossing delays are separated by ~60 microseconds from the primary cluster at ~80 microseconds. We call this separation a "two-cycle" difference: the difference in offset delays corresponds to the time separation between two zero-crossings in the sferic waveform. This behavior is a result of the relationship between arrival time accuracy and peak current polarity determination in the GLD360 network. Since each zero-crossing has an associated polarity, determined by the slope of the waveform at that point, the zero-crossing delay is associated with the polarity of the causative stroke. Hence, if the polarity is known, the arrival time error is either limited by the spread in a given zero-crossing feature (typically <10 microseconds), or the arrival time error is shifted by ~60 microseconds (corresponding to an incorrectly identified zero-crossing).

a. These sensor measurements all correlated with a waveform bank entry that is known to have a low time variance zero-crossing feature occurring two zero-crossing marks after the initial ground wave. For distances ranging from ~2000 km to ~5000 km, this feature arrives ~80 microseconds after the speed-of-light line.

The peak cross-correlation delay of the waveform bank entry with the measured sferic ideally identifies the same feature in the sferic. The absolute time of this feature is sent back to the CP, with an identifier that uniquely labels it in the evolution of the sferic waveform. With this identifier, the CP can apply a fitted propagation correction to refer this arrival time to the speed-of-light propagation time. For the case of the second zero-crossing, the CP subtracts the appropriate distance index of an empirically-determined quadratic fit curve of the cluster centered near 80 microseconds in Fig. 1a.

Fig. 1. (a) 2D histogram of the zero-crossing delays of the second zero crossing after the ground wave, under a daytime ionosphere. Delay-distance bins with fewer than 100 sensor events are omitted. (b) Fraction of events which reported the correct polarity based on the peak cross-correlation-derived propagation distance estimate (discussed in Section III).

As seen in In Fig. 1a, all misidentified zero-crossing delays are separated by ~60 microseconds from the primary cluster at ~80 microseconds. We call this separation a "two-cycle" difference: the difference in offset delays corresponds to the time separation between two zero-crossings in the sferic waveform. This behavior is a result of the relationship between arrival time accuracy and peak current polarity determination in the GLD360 network. Since each zero-crossing has an associated polarity, determined by the slope of the waveform at that point, the zero-crossing delay is associated with the polarity of the causative stroke. Hence, if the polarity is known, the arrival time error is either limited by the spread in a given zero-crossing feature (typically <10 microseconds), or the arrival time error is shifted by ~60 microseconds (corresponding to an incorrectly identified zero-crossing).

a, the majority of zero-crossing arrival times occur between 60 and 100 microseconds after the speed-of-light propagation line. The clusters of delays after 140 microseconds and before 30 microseconds represent events where the peak cross-correlation delay with the waveform bank entry is offset by two zero-crossings. The cluster centered at ~10 microseconds after the speed-of-light line corresponds to the zero-crossing just before the rise of the ground wave. The cluster centered at ~140 microseconds corresponds to two zero-crossing after the expected feature.

In Fig. 1a, all misidentified zero-crossing delays are separated by ~60 microseconds from the primary cluster at ~80 microseconds. We call this separation a "two-cycle" difference: the difference in offset delays corresponds to the time separation between two zero-crossings in the sferic waveform. This behavior is a result of the relationship between arrival time accuracy and peak current polarity determination in the GLD360 network. Since each zero-crossing has an associated polarity, determined by the slope of the waveform at that point, the zero-crossing delay is associated with the polarity of the causative stroke. Hence, if the polarity is known, the arrival time error is either limited by the spread in a given zero-crossing feature (typically <10 microseconds), or the arrival time error is shifted by ~60 microseconds (corresponding to an incorrectly identified zero-crossing).

B. Sensor Polarity Estimation Fidelity

The arrival time estimation algorithm described in Section IIA assumes the correct polarity is known. If the incorrect polarity is assumed, then the peak cross-correlation offset is typically set to one of the adjacent opposite-polarity zero-crossing times in the waveform, resulting in an arrival time estimation error of ~30 microseconds. Since each sensor measurement is equally likely to exhibit this bias, there is usually not a consistent error offset bias associated with incorrectly-identified polarity events.

Fig. 2 shows the fraction of sensor measurements that give the correct polarity based on the propagation distance estimate from the waveform bank cross-correlation, binned according to propagation distance and the elevation of the sun at the midpoint between the lightning discharge and the receiver. 1,181,805 sensor measurements were matched to NLDN-reported negative CG strokes detected on July 20, 2013 with peak current magnitudes between 20 and 60 kA. Of these matched sensor events, 1,057,783 (89.5%) sensor records reported a lower propagation distance error for the correct (in this case, negative) polarity. For propagation distances greater than ~800 km, where the peak in the waveform transitions from the ground wave to the first ionopsheric reflection, there is a noticeable dip in the fraction of correct sensor polarity estimation near the day-night terminator. This dip is likely due to the current lack of intermediate day-night waveform bank entries. Sensor measurements at distances >5000 km also exhibit a large degree of polarity estimation error, due to the more dispersed nature of the waveform at these distances. There is also a dip in polarity estimation fidelity near 2500 km for all times of day. This distance corresponds to the point at which the maximum zero-crossing slope transitions from the first to the second zero-crossing after the ground wave. In this

Fig. 2. (a) 2D histogram of the zero-crossing delays of the second zero crossing after the ground wave, under a daytime ionosphere. Delay-distance bins with fewer than 100 sensor events are omitted. (b) Fraction of events which reported the correct polarity based on the peak cross-correlation-derived propagation distance estimate (discussed in Section III).

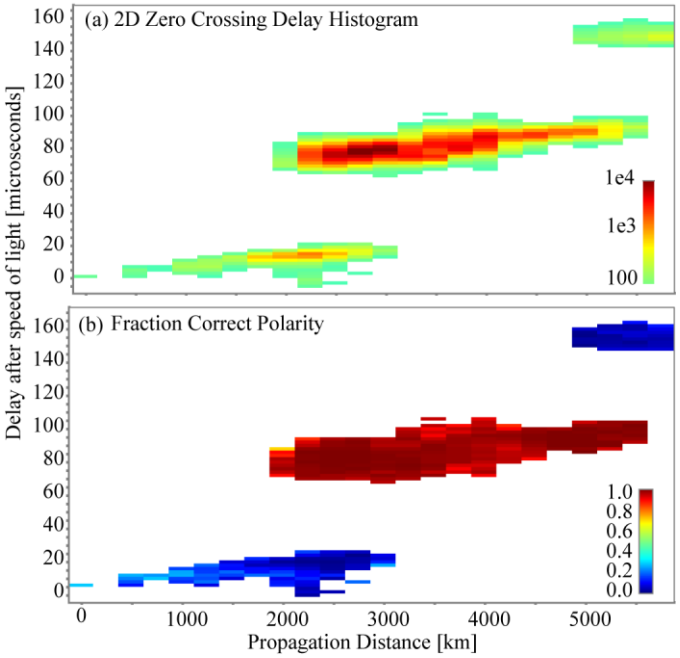

Fig. 3. (a) 2D histogram of the zero-crossing delays of the second zero crossing after the ground wave, under a daytime ionosphere. Delay-distance bins with fewer than 100 sensor events are omitted. (b) Fraction of events which reported the correct polarity based on the peak cross-correlation-derived propagation distance estimate (discussed in Section III).

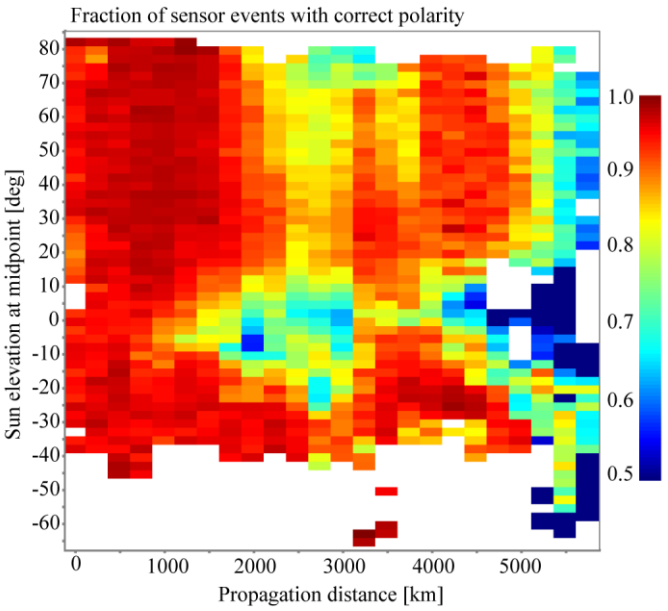

Fig. 4. Fraction of sensor events with the correct polarity versus distance and sun elevation.

transition region, the peak cross-correlation difference between the two polarities is not as large. The CP algorithm down-weights the polarity estimation contribution from more distant (>2000 km) events, but it does not down-weight the contribution from mixed day/night propagation paths.

III. Proposed CP Algorithm Change

In Fig. 1a, all misidentified zero-crossing delays are separated by ~60 microseconds from the primary cluster at ~80 microseconds. We call this separation a "two-cycle" difference: the difference in offset delays corresponds to the time separation between two zero-crossings in the sferic waveform. This behavior is a result of the relationship between arrival time accuracy and peak current polarity determination in the GLD360 network. Since each zero-crossing has an associated polarity, determined by the slope of the waveform at that point, the zero-crossing delay is associated with the polarity of the causative stroke. Hence, if the polarity is known, the arrival time error is either limited by the spread in a given zero-crossing feature (typically <10 microseconds), or the arrival time error is shifted by ~60 microseconds (corresponding to an incorrectly identified zero-crossing).

b shows the fraction of events in each delay-distance bin that had a correct sensor-reported polarity estimate, that is, the distance estimation error corresponding to the matched polarity was lower than that for the opposite polarity. The majority of sensor measurements with the correctly (incorrectly) identified zero-crossing feature report the correct (incorrect) polarity based on the propagation distance estimate. Thus, if the correct polarity is known, the reported estimated distance from the sensor can be used to flag an arrival time measurement as unreliable.

This nearly complete correlation between a correctly identified zero-crossing feature and the sensor-estimated polarity is expected based on the cross-correlation scheme. Since each cross-correlation result is tied to a specific polarity, each zero-crossing is flanked by two cross-correlation offsets corresponding to the opposite polarity. Thus, a "two-cycle" error that jumps to the next zero-crossing of the same polarity is likely to be flanked by at least one cross-correlation result which is a better match to the waveform.

Fig. 5. (a) 2D histogram of the zero-crossing delays of the second zero crossing after the ground wave, under a daytime ionosphere. Delay-distance bins with fewer than 100 sensor events are omitted. (b) Fraction of events which reported the correct polarity based on the peak cross-correlation-derived propagation distance estimate (discussed in Section III).

In the realtime GLD360 system, these two-cycle error events are addressed by a zero-level reassignment scheme [Said et al 2010]. For the second zero level, for example, events closer than a given distance are assumed to correspond to the zero-crossing preceding the ground wave. To recover a useful arrival time estimate, the CP reassigns the zero-crossing marker based on the propagation distance.

This paper proposes a change to this algorithm: After the CP determines the polarity based on all available sensor measurements, it revisits the distance estimation error for each sensor for the given polarity. Instead of reassigning the zero-crossing marker based on distance only, if the reported distance of the chosen polarity gives a larger distance estimation error than that of the opposite polarity, then the CP assumes the zero-crossing identifier to be incorrect. In addition, in these cases, the CP does not attempt to recover the zero-crossing time by assuming a different identifier. Instead, the zero-crossing time is discarded, and a simple threshold crossing time is used with an appropriately large assumed arrival time estimation error. This threshold measurement represents an incoherent time measurement that loosely tracks the arrival time of the peak energy in the waveform packet.

Fig. 3 shows the histogram and Cumulative Distribution Function (CDF) of the corrected arrival time error of the 1,181,805 matched sensor events shown in Fig 2. The results are separated by sensor events which report the correct (top panel) and incorrect (bottom panel) polarities based on the peak waveform cross-correlation result. Events with the correct polarity are more tightly clustered, since they correspond to a coherent measurement with a known zero-crossing delay. The 84^{th} – 16^{th} percentile error difference is 10 microseconds, corresponding to ~5 microsecond root-mean-squared (RMS) error assuming a Gaussian distribution.

Events with an incorrect polarity have an arrival time based on the incoherently measured threshold delay. Averaged over all arrival time estimates, the RMS error estimate from the CDF plot is ~12 microseconds.

IV. Results

This section presents location accuracy results for events reprocessed using the algorithm change proposed in Section III. Sensor data used to generate the production GLD360 dataset on July 20, 2013 is used as the input to a geo-location process equipped with the new algorithm. Reprocessed and production data are compared to all NLDN-reported ground strokes in the latitude/longitude bound [25,50]/[-150,-50] degrees. The stroke matching algorithm uses a time-space coincident window of 150 microseconds and 50 km, respectively. Since the median location accuracy of NLDN is better than 500 m [Nag et al., 2011], the NLDN-reported CG strokes are taken as ground truth in this analysis.

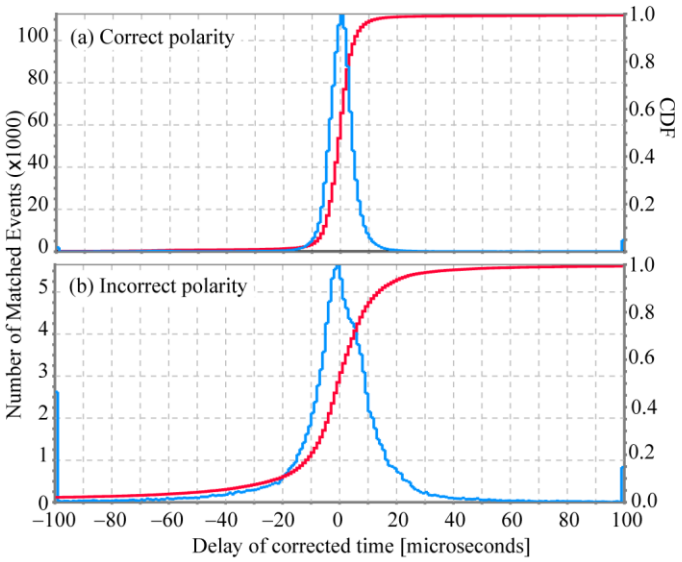

Fig. 6. Arrival time error histogram and CDF for sensor events with the (a) correct and (b) incorrect estimated polarity.

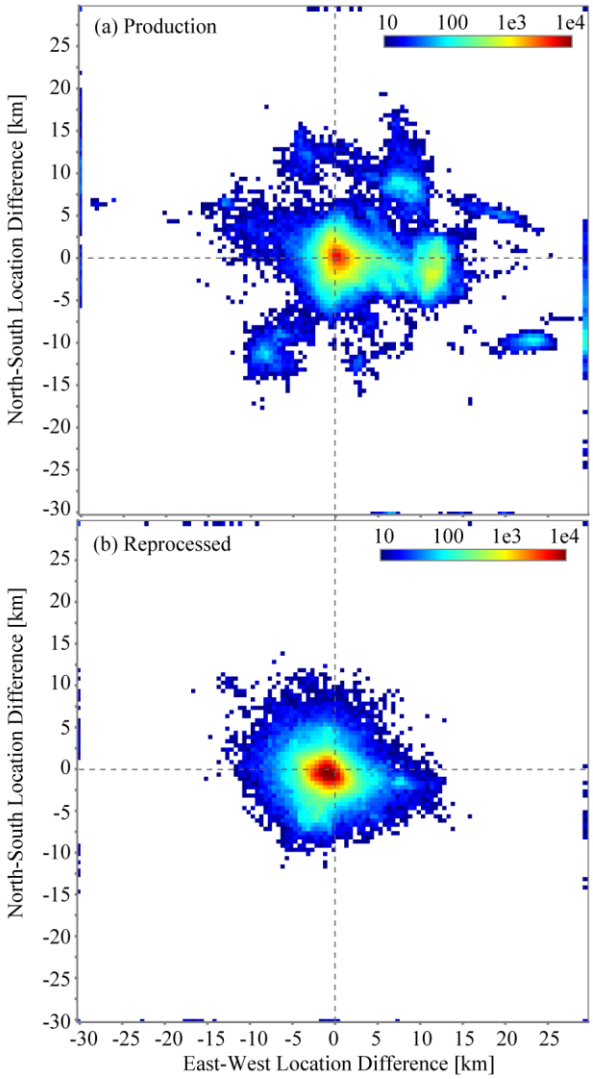

Fig. 7. Histogram of location differences between (a) realtime (production) GLD360 events and (b) reprocessed data, using NLDN ground strokes as a reference.

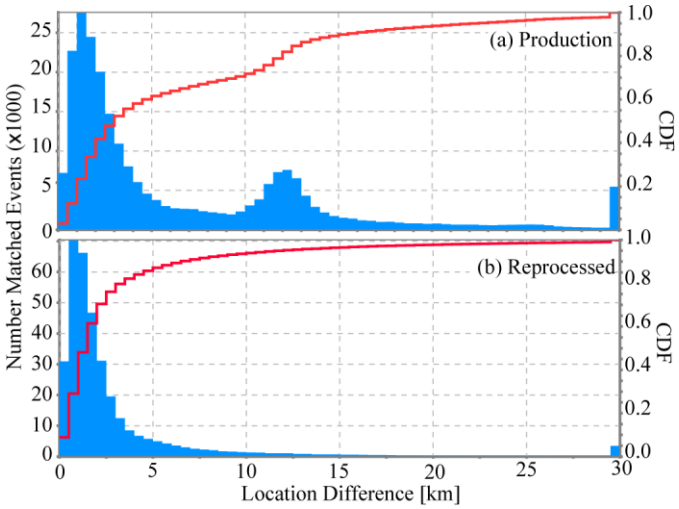

Fig. 8. Histogram and CDF of absolute location differences between matched (a) production and (b) reprocessed events compared to NLDN strokes.

There are multiple metrics that are important to quantify the practical performance of a lightning location system (LLS), including stroke location accuracy, stroke and flash detection efficiency, and the peak current estimation accuracy. We did verify that the flash detection efficiency and average peak current estimation error for July 20, 2013 with respect to NLDN data did not degrade as a result of the algorithm change. However, this paper is focused only on the location accuracy metric, and so the detailed influence of the proposed algorithm change on the detection efficiency and peak current estimation is left to a future discussion.

A. Comparison with Production Dataset

Fig. 4 plots the 2D histogram of location differences between the GLD360 realtime (top panel) and reprocessed (bottom panel) dataset and NLDN ground strokes. The logarithmically spaced color scale indicates the number of matched events falling into a given 0.5 km x 0.5 km location error bin. Bins with fewer than 10 events over the 24-hour period are omitted from the plot. The edges of each plot accumulate all events between the 30 km plot boundary and the maximum location difference of 50 km used by the matching algorithm.

The large cluster in the matched realtime dataset centered near the origin of Fig. 4a corresponds to events where the arrival time was taken from a properly identified waveform feature. Several clusters of location errors ~12 km from the origin are also seen in this distribution, with the largest such cluster exhibiting a consistent ~12 km eastward bias. As discussed above, these large location error events typically result from one or more sensors contributing a "two-cycle" arrival time error to the optimization algorithm.

The benefit to the location accuracy of the proposed algorithm change is clearly seen in the location error distribution in 4b. Visually, most of the outlier clusters are absent, and the large cluster of events with +12 km eastward bias has been greatly reduced.

To better quantify the overall location accuracy performance, Fig. 5 plots histograms of the absolute value of the location errors shown in Fig. 4. The cluster near 12-km error due to two-cycle errors is clearly seen in the production dataset. In the reprocessed location error histogram, this feature is absorbed in the background error distribution. The median location error decreases from 3.0 km to 1.7 km, and the 90th percentile decreases from 15 km to 6.5 km.

B. Location Accuracy versus Local Time

Fig. 6b plots three absolute location accuracy distribution metrics – the 16th, 50th, and 84th percentiles--versus local time. For most local times, the median location accuracy is <2 km. However, near 5 LT (local sunrise), the median location error jumps to ~4 km. This spike can be understood from Fig 6a, which plots the fraction of reprocessed events with the same polarity as the matched NLDN-reported ground stroke. During the sunrise hours, over 10% of the matched events have an incorrect polarity. As mentioned above, the propagation correction scheme is dependent on correct polarity estimation. During the sunrise hours, the relative proportion of incorrect polarity estimation events rises, leading to a larger distribution of ~30 mircrosecond sensor error measurements. Another factor which likely plays a role in the increased location uncertainty is a more complicated ionospheric profile at the dawn terminator, which may not be sufficiently compensated by a linear interpolation between the daytime and nighttime propagation correction factors.

V. Discussion

The arrival time procedure reviewed in Section II that enables accurate arrival time estimates from sensor measurements depends on both an accurate polarity estimate at the CP and the correct identification of a specific feature in the waveform. Section III introduced an algorithm change that appropriately down-weights the contribution from misidentified waveform features at the sensor. As shown in Figures 4 and 5, this algorithm change significantly reduced the population of large error events caused by these so-called two-cycle errors.

If the polarity estimate based at the CP is incorrect, then the arrival time estimate errors increase due to the opposite polarity cross-correlation at the sensor. This effect can be seen during the sunrise hours in Fig. 6, where the dip in polarity estimation accuracy is correlated with a decrease in location accuracy. Since the polarity estimation is tied to the fidelity of the cross-correlation procedure (and the resulting propagation distance estimate) at the sensor, this degradation in polarity estimation accuracy is likely tied to poor cross-correlation performance during this day/night transition time. Thus, further improvement in location accuracy for this time period will likely come from the introduction of mixed day/night waveform bank entries at the sensor.

We note that the degradation in location accuracy in Fig. 6b at the dusk terminator is much less severe, with only a slight rise in the 84th percentile and a nearly constant median location error line. It is unclear if this lack of degradation is a result of the network geometry with respect to the terminator line or a more gradual day-night waveform transition across the dusk terminator.

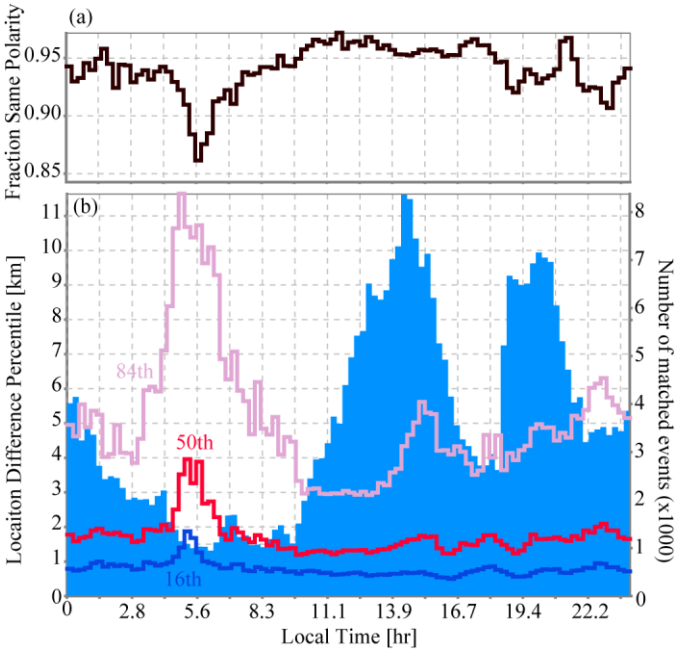

Fig. 9. (a) Fraction same polarity between the reprocessed dataset and NLDN strokes. (b) 16th, 50th, and 84th percentile of the location error distribution versus local time for the reprocessed dataset.

ACKNOWLEDGMENT

The authors wish to thank John Cramer, Tommy Turner, Joe Nowak, and Dave Fincher for their ongoing work on running and maintaining the GLD360 network.

REFERENCES

Cohen, M. B., U. S. Inan, and E. W. Paschal (2010), Sensitive broadband ELF/VLF radio reception with the AWESOME instrument, IEEE. Trans. Geosci. Remote Sens., 48, 3-7.

Mallick et al. (2014); 23rd International Lightning Detection Conference & 5th International Lightning Meteorology Conference, 18-21 March, Tucson, Arizona.

Nag., A. et al. (2011), Evaluation of U.S. National Lightning Detection Network performance characteristics using rocket-triggered lightning data acquired in 2004-2009, J. Geophys. Res., 116.

Said, R. K., U. S. Inan, K. L. Cummins (2010), Long-range lightning geolocation using a VLF radio atmospheric waveform bank, J. Geophys. Res., 115.

Said, R. K., M. B. Cohen, and U. S. Inan (2013), Highly intense lightning over the oceans: Estimated peak currents from global GLD360 observatoins, J. Geophys. Res. 118.