

Monte Carlo Simulations for Evaluating the Accuracy of Geostationary Lightning Mapper Detection Efficiency and False Alarm Rate Retrievals

KATRINA S. VIRTS^a AND WILLIAM J. KOSHAK^b

^a *University of Alabama in Huntsville, Huntsville, Alabama*

^b *NASA Marshall Space Flight Center, Huntsville, Alabama*

(Manuscript received 18 May 2022, in final form 3 October 2022)

ABSTRACT: Performance assessments of the Geostationary Lightning Mapper (GLM) are conducted via comparisons with independent observations from both satellite-based sensors and ground-based lightning detection (reference) networks. A key limitation of this evaluation is that the performance of the reference networks is both imperfect and imperfectly known, such that the true performance of GLM can only be estimated. Key GLM performance metrics such as detection efficiency (DE) and false alarm rate (FAR) retrieved through comparison with reference networks are affected by those networks' own DE, FAR, and spatiotemporal accuracy, as well as the flash matching criteria applied in the analysis. This study presents a Monte Carlo simulation-based inversion technique that is used to quantify how accurately the reference networks can assess GLM performance, as well as suggest the optimal matching criteria for estimating GLM performance. This is accomplished by running simulations that clarify the specific effect of reference network quality (i.e., DE, FAR, spatiotemporal accuracy, and the geographical patterns of these attributes) on the retrieved GLM performance metrics. Baseline reference network statistics are derived from the Earth Networks Global Lightning Network (ENGLN) and the Global Lightning Dataset (GLD360). Geographic simulations indicate that the retrieved GLM DE is underestimated, with absolute errors ranging from 11% to 32%, while the retrieved GLM FAR is overestimated, with absolute errors of approximately 16% to 44%. GLM performance is most severely underestimated in the South Pacific. These results help quantify and bound the actual performance of GLM and the attendant uncertainties when comparing GLM to imperfect reference networks.

KEYWORDS: Lightning; Satellite observations; Sensitivity studies


1. Introduction

Building on two decades of lightning observations from low Earth orbit, Geostationary Operational Environmental Satellites-R series (GOES-R) satellites carried the first lightning sensors into geostationary orbit. Geostationary Lightning Mappers (GLM) aboard *GOES-16*, operational in GOES-East position at 75.2°W since 18 December 2017, and *GOES-17*, operational in GOES-West position at 137.2°W since 12 February 2019, continuously observe total lightning occurrence throughout the tropics and midlatitudes, from the western tip of Africa westward to New Zealand and the Aleutian Islands (Rudlosky and Virts 2021). *GOES-T* launched in March 2022 and became *GOES-18* upon attaining geostationary orbit; it carries a third GLM instrument. *GOES-18* is now in its final operational location at 136.8°W longitude. GLM data have several useful operational applications including weather monitoring and prediction (e.g., Bruning et al. 2019; Kong et al. 2020; Wang et al. 2021), wildfire detection (Bitzer 2017; Rudlosky et al. 2020), lightning chemistry/climate and air quality studies (Koshak 2017), and bolide detection (Jenniskens et al. 2018; Rumpf et al. 2019; Brown et al. 2019). These varied applications have proceeded in parallel with calibration

and validation activities and investigative studies into GLM performance.

There are two particularly important metrics employed to assess GLM performance. The first is its flash detection efficiency (DE), i.e., the percentage of lightning flashes within the GLM hemispheric-scale field of view that GLM typically detects. The second is the percentage of its recorded flash detections that are in fact nonlightning phenomena or noise; this percentage is referred to as the flash false alarm rate (FAR). Knowledge of GLM's DE and FAR is necessary for many purposes, including climatological comparisons and integration with other lightning sensors. The instrument design specifications require GLM performance of at least 70% DE and less than 5% FAR averaged over the entire day and field of view (Goodman et al. 2013). Bateman and Mach (2020) evaluated GLM DE and FAR against a "virtual" network synthesized from ground-based lightning detection systems operated by Earth Networks and Vaisala and reported an aggregate GLM DE of >70%, while the FAR specification was met only over the well-sampled continental United States (CONUS). A follow-on study by Bateman et al. (2021) explored the effect of broadening the temporal matching criteria between GLM and reference network flashes (i.e., not requiring strict flash-to-flash matching). They reported, for temporal matching windows of ± 10 min, aggregate DE values of >90% and FAR of about 5% (GLM-16) up to 20% (GLM-17).

In addition to these aggregate studies, analysis of GLM performance has also focused on the impact of time of day (Bateman and Mach 2020; Zhang and Cummins 2020; Bateman et al. 2021),

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Katrina S. Virts, katrina.virts@uah.edu

storm type and cloud characteristics (Murphy and Said 2020; Rutledge et al. 2020), and flash characteristics including size and duration (Lang et al. 2020; Zhang and Cummins 2020). However, other factors can also affect the calculated performance of GLM, including the performance and geographic coverage variations of the reference networks themselves as well as the location and timing accuracy of both GLM and the reference networks. Little investigation into these factors has been performed to date, in part because the performance of the reference networks is both imperfect and imperfectly known.

Assessing GLM's detection performance is complicated by the type and availability of reference sensors and networks. Lightning mapping arrays (LMAs) provide high DE, but only for limited areas [e.g., Chmielewski and Bruning (2016) reported DE > 95% within 100 km of the network sensors]. The Lightning Imaging Sensor aboard the International Space Station (ISS-LIS) has a flash DE around 60% (Blakeslee et al. 2020), but samples a given location for only ~90 s every 1.5 days. Ground-based, long-range detection networks such as the Earth Networks Total Lightning Network (ENTLN; Earth Networks 2014; Zhu et al. 2022) and Vaisala's Global Lightning Dataset (GLD360; Vaisala 2014) detect global lightning continuously and are best suited for evaluating GLM performance throughout its field of view, although that comparison is complicated because these sensors detect lightning differently (GLM via optical signals, reference networks via monitoring radio frequencies; see Zhang and Cummins 2020). Rudlosky et al. (2017) reported GLD360 flash DE over the Western Hemisphere of 44% with respect to the LIS on the Tropical Rainfall Measuring Mission (TRMM) satellite. ENTLN exhibits larger regional variations; comparison with TRMM-LIS revealed flash DE ranging from 10% to 72% (average: 27%) over the Western Hemisphere (Rudlosky 2015). Bitzer and Burchfield (2016) employed a Bayesian technique to compare ground network detections of strokes/pulses (i.e., the components of a lightning flash) and estimated the upper limit of the global absolute stroke detection efficiency to be 57% for ENTLN and 60% for the Vaisala networks. The DE of both ENTLN and GLD360 have increased in recent years with additional sensors and improved detection techniques, although no more recently published hemispheric-scale DE statistics are available. Published studies of reference network FAR over the full Western Hemisphere domain are lacking. However, day-to-day examination of the lightning flashes reported by ENTLN and GLD360 suggests that their FARs are quite small.

Given these unknowns, it is desirable to assess both GLM performance metrics and the associated error bars due to unknown reference network performance. The purpose of this study is to quantify via Monte Carlo simulations how GLM performance metrics are affected by a range of realistic reference network performance characteristics, including their geographic variations. Results from these simulations give context, unique quantitative insight and provide error bars to the GLM performance metrics that have been analyzed and published in recent years.

In short, this study clarifies what impact reference network performance imperfections have on assessments of GLM

performance, and therefore provides a more robust/accurate estimate of the true GLM performance. For example, if a reference network false alarms, it gives the false impression that GLM missed a detection (i.e., that GLM has lower DE). Similarly, if a reference network misses a legitimate flash that GLM rightly detects, it gives the false impression that GLM is producing false alarms (i.e., that GLM has a higher FAR). The unique benefit of our simulations is that these kinds of nuances are tracked and flagged within the simulated retrievals so that we can quantify the *specific deleterious impact the reference network imperfections have on the assessment of GLM performance*. Apart from deploying additional independent lightning measurements at added cost/time, it is worth emphasizing here that the authors know of no other means to attain the valuable insight offered by our simulations.

2. Data and methods

a. Data sources

GLM detects the optical signals produced by lightning by monitoring a ~1 nm spectral band at 777.4 nm in the near-infrared (Rudlosky et al. 2019). GLM reports the illumination of individual pixels at ~500 frames per second above the dynamically varying background. Illuminated pixels, with footprint sizes ranging from ~8 km at nadir to ~14 km near the limb, are referred to as *events*. In the GLM ground system, contiguous events within the same frame are clustered into *groups*, which are further clustered into *flashes* using spatio-temporal windows of 330 ms and 16.5 km (Mach 2020; GOES-R Algorithm Working Group and GOES-R Series Program Office 2018). Tuning of the ground system software has improved GLM data quality over time, with the greatest improvements taking place in 2019 with adjustments to the second-level thresholds and implementation of a blooming filter (Rudlosky and Virts 2021). This study, with an analysis period of 2020/21, exclusively uses GLM data gathered after these ground system improvements.

As indicated in section 1, the two global lightning networks best suited for comparison with GLM are those operated by Earth Networks and Vaisala. Earth Networks sensors monitor wideband frequencies (1 Hz to 12 MHz; Liu and Heckman 2012), and a time-of-arrival technique is employed for geolocation of the detected waveforms. In the Western Hemisphere, the sensor network is most dense over southern Canada and CONUS, as well as portions of Brazil. Rudlosky (2015) reported an aggregate flash DE of 27% over the Western Hemisphere for ENTLN with respect to TRMM LIS. An upgrade in December 2021 resulted in an overall 149% increase in pulse detections for ENTLN (Zhu et al. 2022); updated aggregate DE statistics have not yet been published but are expected to increase in the future. For this study we employ the ENGLN dataset (Earth Networks 2014), in which lightning strokes detected by Earth Networks sensors are supplemented with strokes detected by the World Wide Lightning Location Network (WWLLN; Abarca et al. 2010; Rudlosky and Shea 2013), providing improved global

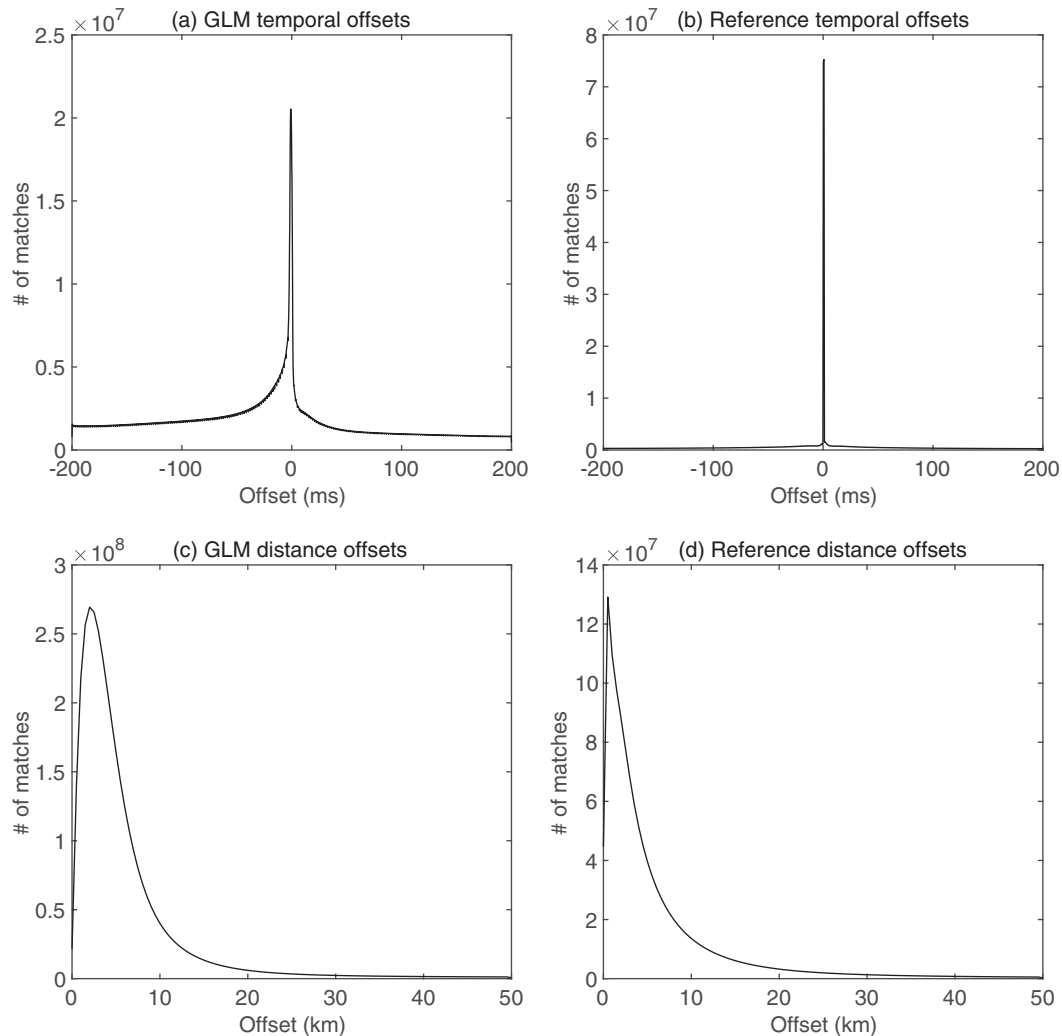


FIG. 1. (top) Temporal and (bottom) distance offset distributions for (left) GLM compared with the reference networks and (right) the reference networks compared with each other. Statistics are calculated for January–December 2021 and are used in the Monte Carlo simulations.

coverage. Vaisala sensors have a large detection range through monitoring very low frequency (VLF) radio wavelengths (~ 500 to ~ 50 kHz; Said et al. 2013; Mallick et al. 2014; Vaisala 2014). Geolocation of the detected waveforms is performed using both magnetic direction finding and time-of-arrival methods. The most recent GLD360 location algorithm upgrade was implemented in 2015 (Said and Murphy 2016), and Rudlosky et al. (2017) reported a flash DE of 44% aggregated over the Western Hemisphere.

b. Spatiotemporal offset distributions

The actual spatial and temporal accuracy of GLM and the reference networks cannot be known perfectly but can be estimated through intercomparison. For this study, GLM and reference network spatiotemporal characteristics are baselined using 1 year of matched group/stroke observations from 2021. For this analysis only, we use group-level data to obtain more

precise offset statistics. GLM groups and reference network strokes are statistically matched if the observations are within ± 200 ms and 50 km of the group centroid. GLM distance offsets from the ground networks peak around 3 km, and temporal offsets exhibit a prominent peak (Fig. 1, left column), with GLM slightly lagging the reference networks for reasons discussed in Virts and Koshak (2020). The two ground-based reference networks exhibit close overall agreement with each other (Fig. 1, right column), with offset distributions peaking sharply near 0 ms and between 0 and 1 km. Greater agreement is observed between the two reference networks than between GLM and the reference networks, due in part to the coarser spatial and temporal resolution of GLM. In addition, these sensors detect different physical processes of the lightning flash via optical signals (GLM) and radio frequencies (the reference networks) and will not always align.

c. Monte Carlo simulations

The lightning simulations follow these steps:

- 1) Define a set of N “true” lightning flashes within the GLM domain with realistic timing and geographic distribution.
- 2) Define the true DE such that $DE_{\text{true}} = n_T/N$, where n_T is the number of detections that are associated with a real flash. In this step, $100 - DE_{\text{true}}\%$ of the true flashes are randomly discarded. DE_{true} is defined separately for GLM and the reference network.
- 3) Define the true FAR such that $FAR_{\text{true}} = n_F/n$, where n_F is the number of detections that are not associated with any real flash (i.e., the number of false alarms) and n is the total number of detections ($n = n_T + n_F$). The FAR is defined by adding flashes that are randomly distributed spatially throughout the GLM domain and temporally throughout the analysis period (for our simulations, one day). FAR_{true} is defined separately for GLM and the reference network.

The outcome of steps 1–3 is two sets of lightning flashes, one representing GLM and the other representing a generalized reference network (R).

- 4) Define the location and timing errors by applying time and location offsets characteristic of GLM and R. For each simulated GLM flash, apply a temporal offset drawn at random from the distribution in Fig. 1a and a location offset in a random direction with magnitude drawn at random from the distribution in Fig. 1c. Similarly, each simulated reference network flash receives spatiotemporal offsets randomly drawn from Figs. 1b and 1d, such that the simulated reference network flashes incorporate characteristics of both ENGLN and GLD360. To examine the impact of spatiotemporal uncertainty, a multiplicative factor may be applied to the offset distributions prior to their application in the simulation.
- 5) Match the simulated GLM and reference network flashes. There are currently no standard flash matching criteria for GLM. The initial criteria adopted for this study are ± 200 ms (i.e., 200 ms before the flash starts to 200 ms after the flash ends) and 50 km between flash centroids. To examine the impact of the selected matching criteria, a multiplicative factor may be applied to the initial criteria. Note that in order to align with the standard procedures employed in evaluating GLM performance (e.g., Bateman and Mach 2020; Bateman et al. 2021), a single GLM flash may be matched to multiple reference network flashes, or vice versa.
- 6) Calculate the retrieved GLM DE and FAR. GLM DE is calculated as the percent of R flashes that can be successfully matched with GLM flashes using the selected matching criteria, i.e., the percent of R flashes that GLM detects. GLM FAR is calculated as the percent of GLM flashes not successfully matched with an R flash.

The simulations thus have eight tunable parameters: GLM DE, GLM FAR, the magnitude of the GLM spatiotemporal errors, reference network DE, reference network

FAR, the magnitude of the reference network spatiotemporal errors, and the distance and temporal matching criteria applied. It is possible to covary all these simultaneously, and we have performed simulations throughout a large parameter space. However, to reduce computation time and simplify interpretation of the results, we focus on five simulation categories as summarized in Table 1, examining the retrieved GLM DE and FAR while varying the following parameters:

- True GLM DE (40%–100%) and FAR (0%–40%) for three different true reference network DE scenarios (40%, 70%, and 90%)
- True reference network DE (40%–100%) and FAR (0%–10%)
- The magnitude of the reference network spatiotemporal offsets (from 50% to 800% of the observed values in Figs. 1b,d)
- Distance matching criteria (from 50% to 800% of the initial criteria, i.e., from 25 to 400 km)
- Temporal matching criteria (from 30% to 10⁶% of the initial criteria, i.e., from 60 ms to 200 s)

A full set of simulations thus includes 158 variations in the parameter combinations. The simulation technique could start with any set of lightning flashes with a realistic geographic and temporal distribution. For this study, we define the “true” lightning flashes N to be the daily observed GLM-16 flashes every 5 days during 2020, i.e., all flashes from 0000 to 2359 UTC on 1, 6, 11, ..., January, etc. This means that 74 full sets of simulations are performed, permitting evaluation of the standard deviation of the retrieved GLM performance metrics. Note also that for these initial simulations, the tunable parameters are applied universally. Geographic variations due to varying DE across the ground networks are considered in a second phase of simulations, described in section 5.

3. Simulations varying GLM and reference network performance

The first scenario tests the simulation methods by calculating retrieved GLM performance as a function of a suite of true GLM DE and FAR values (Table 1, scenario 1). That is, we are comparing the GLM performance metrics defined and applied in step 2 of the simulation (the “true” values) with the same performance metrics calculated in step 6 of the simulation via matching the simulated GLM and reference network flashes to each other (the “retrieved” values). Retrieved GLM DE varies with the true GLM DE but is independent of the true reference network DE (Fig. 2a; note that the three colored lines representing reference network DE are nearly identical) and the true GLM FAR (not shown). There is not an exact 1:1 correspondence between the true and retrieved GLM DE. Rather, GLM DE is underestimated by $\sim 9\%$ – 10% in the case of true DE = 100%, due in part to the true reference network FAR of 5%, since only a fraction of these false flashes are likely to be successfully matched with a true GLM flash. Errors in the retrieved GLM DE decrease as the true GLM DE decreases, which may initially seem

TABLE 1. Overview of Monte Carlo simulation types, indicating the true (applied) value of each tunable parameter. See text for further details. Boldface font indicates the parameters being varied in the simulation (i.e., parameters not at their nominal values).

Simulation type	1) Vary GLM DE and FAR	2) Vary reference network DE and FAR	3) Vary reference network offsets	4) Vary distance matching criteria	5) Vary temporal matching criteria
Simulation specifics	GLM DE: 40%, 50%, ..., 100% GLM FAR: 0%, 10%, ..., 40% Reference network DE: 40%, 70%, 90% Reference network FAR: 5%	GLM DE: 70% GLM FAR: 5% Reference network DE: 40%, 50%, ..., 100% Reference network FAR: 0%, 2.5%, ..., 10% GLM offsets: Observed Reference network offsets: Observed Distance matching criteria: Initial (50 km) Temporal matching criteria: Initial (± 200 ms)	GLM DE: 70% GLM FAR: 5% Reference network DE: 70% Reference network FAR: 5% GLM offsets: Observed Reference network offsets: 0.5, 1, 2, 4, 8 \times observed Distance matching criteria: Initial (50 km) Temporal matching criteria: Initial (± 200 ms)	GLM DE: 70% GLM FAR: 5% Reference network DE: 70% Reference network FAR: 5% GLM offsets: Observed Reference network offsets: Observed Distance matching criteria: 0.5, 1, 2, 4, 8 \times initial (50 km) Temporal matching criteria: Initial (± 200 ms)	GLM DE: 70% GLM FAR: 5% Reference network DE: 70% Reference network FAR: 5% GLM offsets: Observed Reference network offsets: Observed Distance matching criteria: Initial (50 km) Temporal matching criteria: 0.3, 1, 3, 10, ..., 1000 \times initial (± 200 ms)

counterintuitive. However, lightning-producing storms frequently produce many flashes, so this decrease in retrieval errors is likely because other nearby reference network flashes were successfully matched with the remaining GLM flashes.

The first simulation scenario further demonstrates that retrieved GLM FAR varies with true GLM FAR and true reference network DE (Fig. 2c), but not with true reference network FAR (not shown). A perfectly performing GLM (FAR = 0%) is nevertheless evaluated as having retrieved FAR values ranging from 15% to 60% depending on reference network performance, as GLM flashes not detected by the reference networks are recorded as false. As true GLM FAR increases, the retrieved FAR also increases, although more slowly, as some subset of the large quantities of false GLM flashes is close enough to a reference network flash to be successfully matched. The impacts of imperfect reference network performance are further explored in the following simulations.

The second simulation scenario tests the impact of varying reference network performance metrics (DE: 40%–100%; FAR: 0%–10%) upon a GLM performing exactly at its specifications (DE: 70%; FAR: 5%; Table 1; Goodman et al. 2013). Retrieved GLM DE varies with true reference network FAR (Figs. 3a,b). For perfect reference network FAR of 0%, one might expect the retrieved GLM DE to not statistically differ from the true GLM DE. However, the simulations indicate that the true and retrieved GLM DEs do significantly differ, with an absolute error of $\sim 2\%$. This difference can be explained by the spatiotemporal offsets applied to both

datasets that cause a small percentage of simulated GLM and reference network flashes to not successfully match their counterparts, thus lowering the retrieved GLM DE. Retrieved GLM DE further decreases as the true reference network FAR increases, and for an extreme case of reference network FAR = 10%, true GLM DE is underestimated by almost 9%. However, as mentioned in section 1, while hemispheric-scale reference network FAR has not been investigated in the published literature, it is likely at the low end of the range considered in this simulation scenario.

Retrieved GLM FAR varies with true reference network DE (Figs. 3c,d). For perfect reference network DE of 100%, the retrieved GLM FAR is $\sim 10\%$, representing an absolute error of 5%. As the reference network DE deteriorates, overestimation of the GLM FAR increases, with absolute errors exceeding 55% when the reference network DE = 40%. This is noteworthy in the context of reported reference network flash DE values of 30%–40% averaged over the Western Hemisphere (Rudlosky 2015; Rudlosky et al. 2017). While improvements in recent years have likely increased the global DE of both reference networks (e.g., Zhu et al. 2022), the results in Figs. 3c and 3d indicate that for locations with comparatively low reference network DE, GLM FAR will be significantly overestimated.

The third simulation scenario examines the effect of reference network spatiotemporal accuracy. In the aggregate, the reference networks exhibit close spatiotemporal agreement (Fig. 1), and submillisecond timing agreement is observed throughout the GLM-16 domain (Fig. 4a). In contrast, the

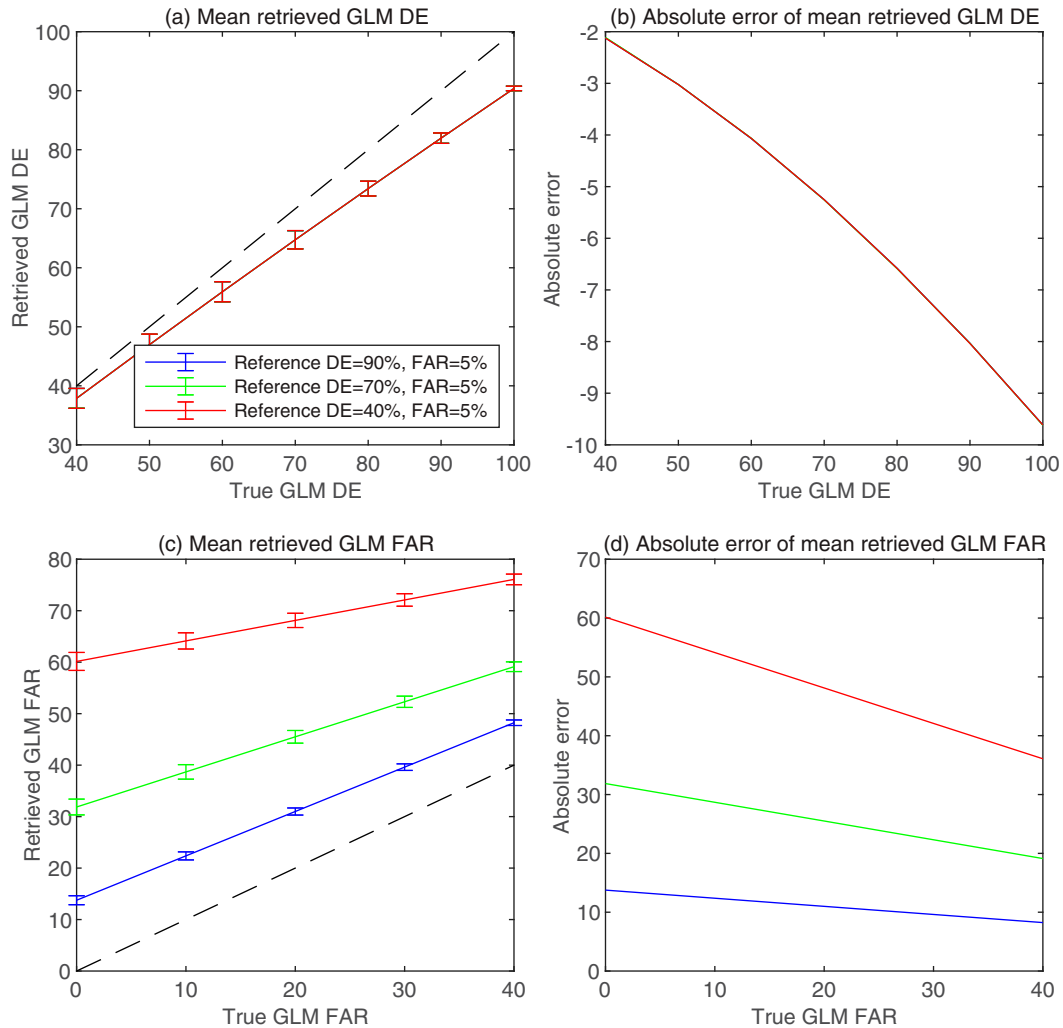


FIG. 2. Simulation results varying true GLM DE and FAR (for full specifics, see scenario 1 in Table 1). (a) Mean and (b) absolute error of mean retrieved GLM DE as a function of true GLM DE. (c) Mean and (d) absolute error of mean retrieved GLM FAR as a function of true GLM FAR. Colored lines represent three reference network performance scenarios and are nearly identical in (a) and (b). Error bars in (a) and (c) represent two standard deviations, and dashed lines indicate a 1:1 relationship.

peak distance offsets vary regionally (Fig. 4b). Close agreement is observed over CONUS and central South America, where both networks have dense sensor coverage. Away from these locations, distance offsets increase to $\sim 3\text{--}4$ km over much of Central America, the western Atlantic, and the northern Pacific. Even larger offsets up to 10 km or more are observed over the South Pacific. These results motivate examination of the effect of these spatial uncertainties on retrieved GLM performance.

In the third simulation scenario, both GLM and reference network performance are defined as identical (DE: 70%; FAR: 5%) while the magnitude of the reference network spatiotemporal offsets varies from 50% to 800% of the observed aggregate values (Fig. 1; Table 1). For example, the observed reference network offset distributions peak at 0.8 ms and 0.5 km, so offset distributions at 8 times the observed values

would peak at 6.4 ms and 4 km with correspondingly longer tails. While these simulations vary both spatial and temporal offsets, the impact of varying the temporal offsets is negligible since the peak offsets are two orders of magnitude smaller than the temporal matching window. For reference network offsets from 50% to 100% of the observed values, the error of the retrieved GLM DE is 5% or less, while retrieved GLM FAR errors of $\sim 30\%$ reflect the defined reference network DE (Fig. 5). As reference network offsets increase, so too does the magnitude of the retrieval errors. For reference network offsets that are 8 times the observed values (i.e., still smaller than the distance offset values observed in parts of the South Pacific), absolute errors around 40% and 65% are observed for the retrieved GLM DE and FAR, respectively.

Uncertainties in reference network location accuracy (Fig. 4) and DE (Rudlosky 2015; Rudlosky et al. 2017) in remote areas

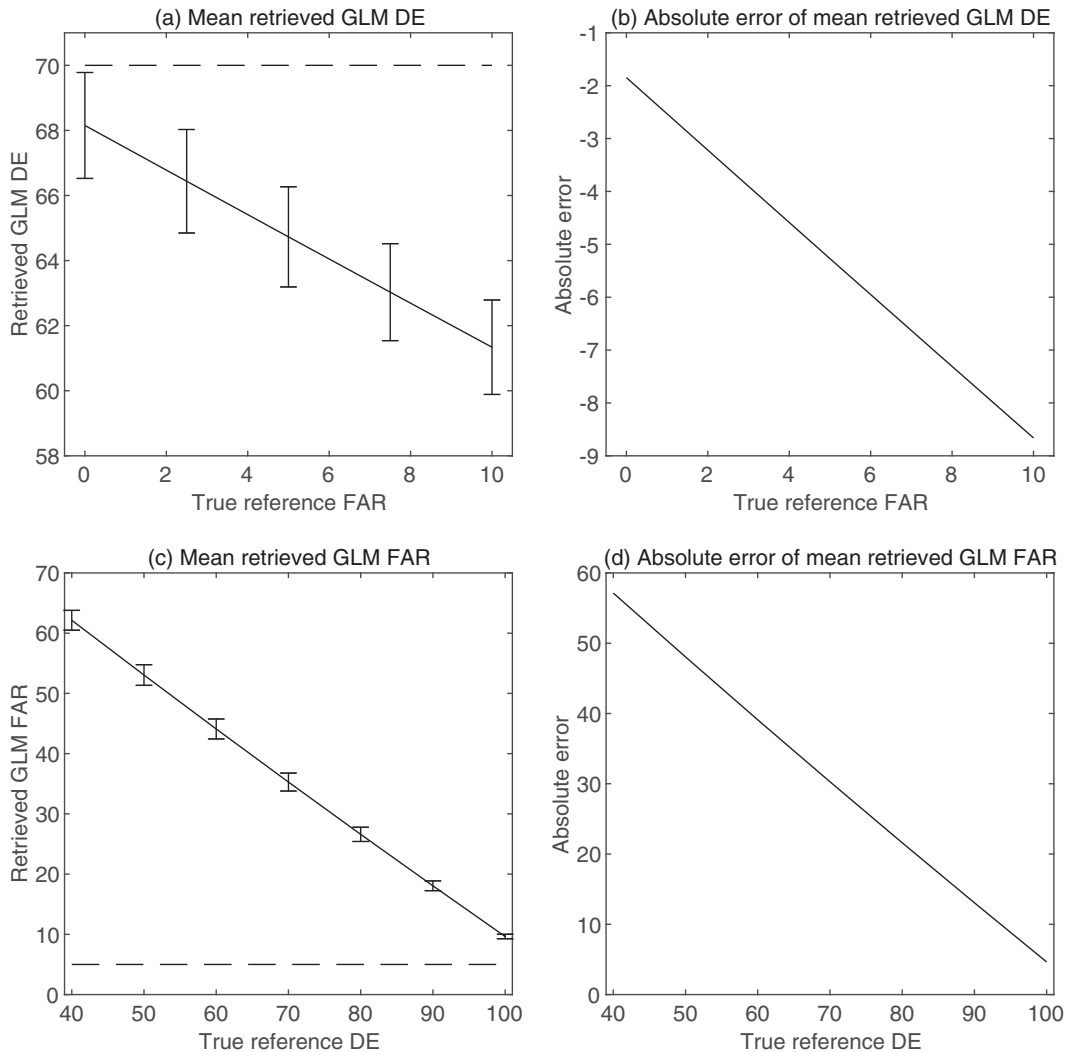


FIG. 3. Simulation results varying true reference network DE and FAR (for full specifics, see scenario 2 in Table 1). (a) Mean and (b) absolute error of mean retrieved GLM DE as a function of true reference network FAR. (c) Mean and (d) absolute error of mean retrieved GLM FAR as a function of true reference network DE. Error bars in (a) and (c) represent two standard deviations, and dashed lines indicate the true GLM metrics.

of the GLM-16 domain illustrate why a recent study by Bateman et al. (2021) experimented with broadened flash matching criteria in the examination of GLM-16 performance metrics. Broadened matching windows sacrifice strict flash-to-flash matching, allowing other nearby flashes to “match” a GLM lightning observation for the purpose of arriving at more accurate aggregate statistics. In the following section, we investigate this same issue via Monte Carlo simulation.

4. Simulations varying matching criteria

In the fourth simulation scenario, GLM and the reference network are defined to perform identically (DE: 70%; FAR: 5%), while the distance matching criteria range from narrow to very broad (i.e., 50%–800% of the initial matching criteria; Table 1). For matching criteria stricter than the initial criteria,

the retrieved GLM performance underestimates the true performance, with DE errors up to 11% and FAR errors over 35% (Fig. 6). As the criteria are relaxed, there is shift toward overestimating GLM DE up to ~5%, while GLM FAR errors decrease to ~20%. The standard deviation of the retrieved GLM metrics also increases strongly as the criteria are broadened, but even for the most permissive criteria, the retrieved FAR still significantly differs from the true FAR. Further broadening the distance matching criteria increases the probability that the matched flashes may be produced by two different storms.

Broadening the distance matching criteria only moderately reduces errors in the retrieved GLM FAR, so the fifth simulation scenario instead explores the effect of varying the temporal matching criteria from 30% to 10⁶% of the initial values (± 60 ms to ± 200 s; Table 1). The curves are steep for criteria

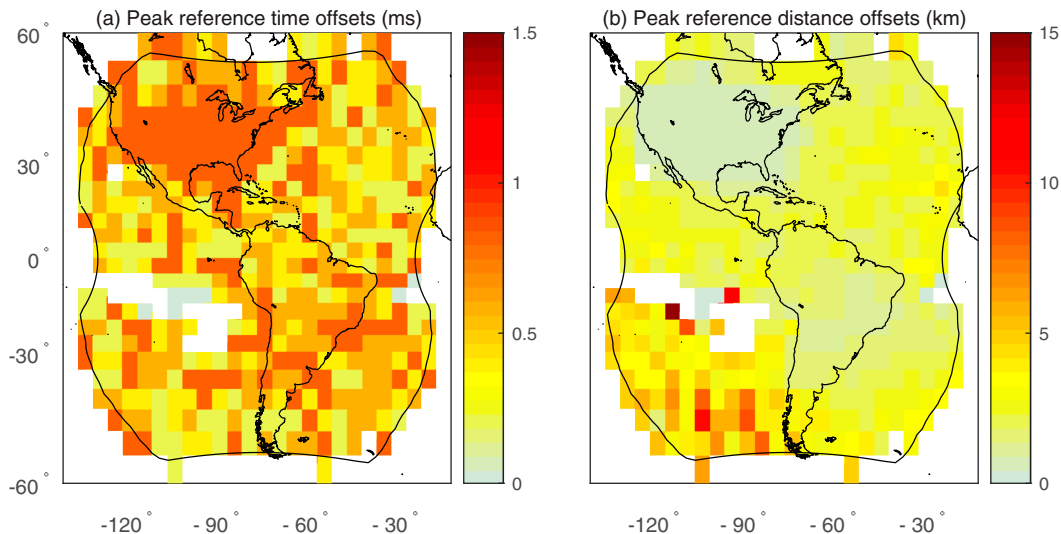


FIG. 4. Geographic distributions of peak reference network (a) temporal (ms) and (b) spatial offsets (km). Nominal GLM-16 field of view is shown by the black line.

near the initial values (Fig. 7), prompting the adoption of a logarithmic plotting scale. This suggests the sensitivity of the retrieved performance metrics when strict flash-to-flash matching is required, in agreement with Bateman et al. (2021, their Figs. 5 and 9). As the criteria are broadened to permit matching of different flashes produced by the same storm, the retrieved GLM DE is increasingly overestimated, approaching an asymptote at 95% which reflects the true reference network FAR of 5% (Table 1). The retrieved GLM FAR also exhibits asymptotic behavior, converging toward the true value of 5%. These results agree with and illustrate the purpose of Bateman et al. (2021) in using broadened temporal matching criteria, especially when evaluating GLM FAR.

A question that naturally follows is which temporal matching criteria should be applied to obtain the most reliable estimate of GLM performance. The answer depends on the use case; e.g., when analyzing the characteristics of reference network flashes that GLM is unable to detect, strict flash-to-flash matching would likely be needed. However, our simulation procedure provides a mechanism for investigating the optimal matching criteria for evaluating aggregate GLM DE and FAR. To accomplish this, we run two additional sets of simulations with varying temporal matching criteria, one with varying GLM performance and the other with varying reference network performance (Table 2).

The sixth simulation scenario demonstrates that as the temporal matching window is broadened, the retrieved GLM DE converges to 95% regardless of the true GLM DE, again reflecting the true reference network FAR of 5% (Fig. 8; Table 2). The crossover point with absolute errors of 0% occurs for matching criteria ranging from ± 0.3 s (for low true GLM DE) up to ± 3 s (for high true GLM DE). In contrast, absolute errors of 0% for retrieved GLM FAR are observed only for temporal matching windows broader than $\sim \pm 100$ s.

The seventh simulation scenario explores the inverse situation where GLM performance is fixed and reference network

performance is allowed to vary (Fig. 9; Table 2). As the temporal matching criteria broaden, the retrieved GLM DE converges to a value equivalent to 100% minus the true reference network FAR, with absolute errors of 0% occurring for matching criteria ranging from ± 0.3 s (for low true reference network FAR) up to ± 0.6 s (for high true reference network FAR). The results in Fig. 9 further demonstrate that the retrieved GLM FAR converges to the true GLM FAR regardless of the true reference network DE. However, for all scenarios except that of unrealistically perfect reference network performance, broad matching windows on the order of minutes are needed before the absolute errors converge to 0%.

In answer to the question of what matching criteria produce the most accurate estimates of aggregate GLM performance, the results in Figs. 8 and 9 indicate that the optimal criteria are different for DE and FAR. Broad matching criteria significantly overestimate the true GLM DE but are necessary to correctly estimate GLM FAR, especially in areas of poor reference network detection performance, where narrow flash-to-flash matching criteria unfairly punish GLM for detecting flashes below the local reference network detection thresholds. Our results suggest that matching windows on the order of ± 1 s produce the closest estimate of GLM DE, while matching windows on the order of minutes provide the most reliable estimate of GLM FAR (although the exact windows selected are less important when evaluating FAR).

5. Simulations varying reference network geographical performance

a. Geographic simulation methods

The simulations described in the foregoing sections have defined the reference network performance universally. In reality, as discussed in section 1, the performance of ground reference networks varies geographically in ways that cannot be

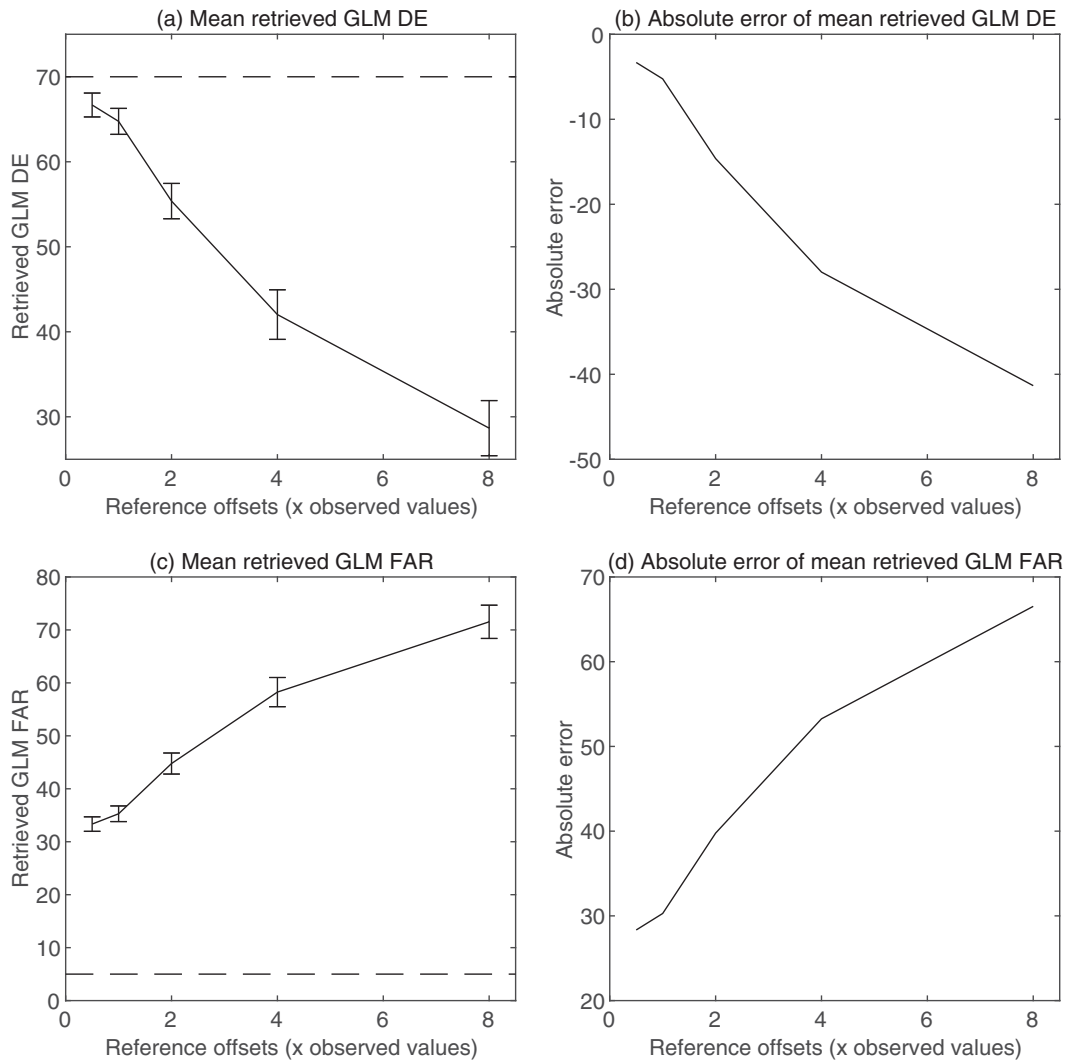


FIG. 5. Simulation results varying the magnitude of reference network spatiotemporal offsets (for full specifics, see scenario 3 in Table 1). (a) Mean and (b) absolute error of mean retrieved GLM DE as a function of reference network offsets ranging from 50% to 800% of the observed offsets. (c),(d) As in (a) and (b), but for retrieved GLM FAR. Error bars in (a) and (c) represent two standard deviations, and dashed lines indicate the true GLM metrics.

fully known in the absence of an additional, perfectly performing reference source. To estimate the effect of these variations on retrieved GLM performance via Monte Carlo simulation, it is necessary that the simulated reference network data performance varies in a geographically realistic manner.

To facilitate comparison with GLM flash data, flash clustering is applied separately to both ENGLN and GLD360 stroke data using spatiotemporal windows of 330 ms and 5.5 km (i.e., the clustering criteria employed for LIS; Mach et al. 2007). Next, a 5° latitude–longitude grid is created within the GLM-16 field of view. ENGLN and GLD360 flashes within each grid box are matched to each other, producing two DE values and two FAR values at each grid box (one for each network with respect to the other). From these values, “best” and “worst” reference network scenarios are defined for each grid box, where the “best” scenario is the higher of the two DE values and the lower of the two

FAR values, and vice versa for the “worst” scenario. In this way, the analysis is reference network agnostic, in that it does not matter which real-world network performs better at a given location. The important outcome is a realistic upper and lower bound on reference network performance as a function of geography.

Other specifics of the geographic simulations are given in Table 3. GLM performance is defined such that it exactly meets its specifications of 70% DE and 5% FAR in each grid box. The observed aggregate GLM offsets (i.e., Figs. 1a,c) are applied throughout the domain, while the reference network offsets in each grid box are those observed for the reference networks with respect to each other.

b. Results

Geographic variations in reference network performance produce geographic error patterns in the retrieved GLM

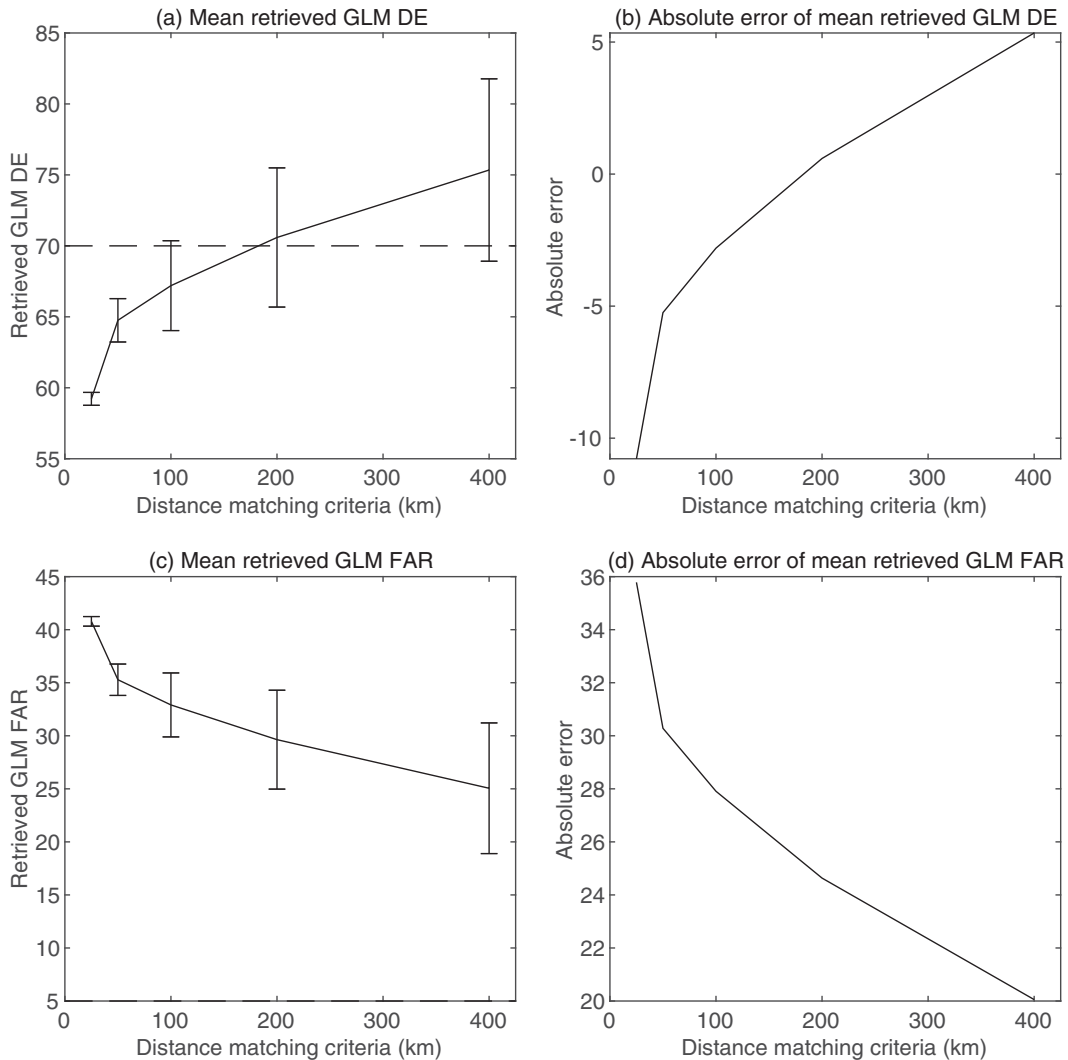


FIG. 6. Simulation results varying the distance matching criteria (for full specifics, see scenario 4 in Table 1). (a) Mean and (b) absolute error of mean retrieved GLM DE as a function of distance matching criteria. (c),(d) As in (a) and (b), but for retrieved GLM FAR. Error bars in (a) and (c) represent two standard deviations, and dashed lines indicate the true GLM metrics.

performance metrics (Fig. 10). The best agreement is observed over the eastern United States, northwestern Atlantic, and central South America, where absolute errors for the best reference network scenario are as low as 0%–10%. Larger deviations are observed in geographic areas removed from the dense sensor networks, with GLM performance most severely underestimated in the Southern Hemisphere Pacific. In this region, absolute errors for the *best* reference network scenario range up to 20%–40% for DE and 30%–50% for FAR.

Overall, the retrieved GLM DE aggregated over all flashes ranges from 38% for the worst reference network scenario to 59% for the best reference network scenario, implying absolute errors of 11%–32% from the true GLM DE of 70% throughout the domain. Errors in the retrieved aggregate GLM FAR are even larger, ranging from 16% for the best reference network scenario to 44% for the worst reference

network scenario. Aggregating the GLM performance metrics over all grid boxes rather than all flashes yields still larger errors (Fig. 10, lower-right corner of each panel) because the areas of comparatively poor reference network coverage tend to experience less frequent lightning than the convectively active and comparatively well sampled areas over the eastern United States and central South America. Geographic simulations using GLM-17 flash data exhibit retrieval errors of similar magnitude to those for GLM-16 (Fig. 11).

6. Conclusions

Ongoing efforts to evaluate GLM flash detection performance are complicated by the fact that the flash detection performance of the long-range ground-based detection networks against which it is evaluated is both imperfect and imperfectly

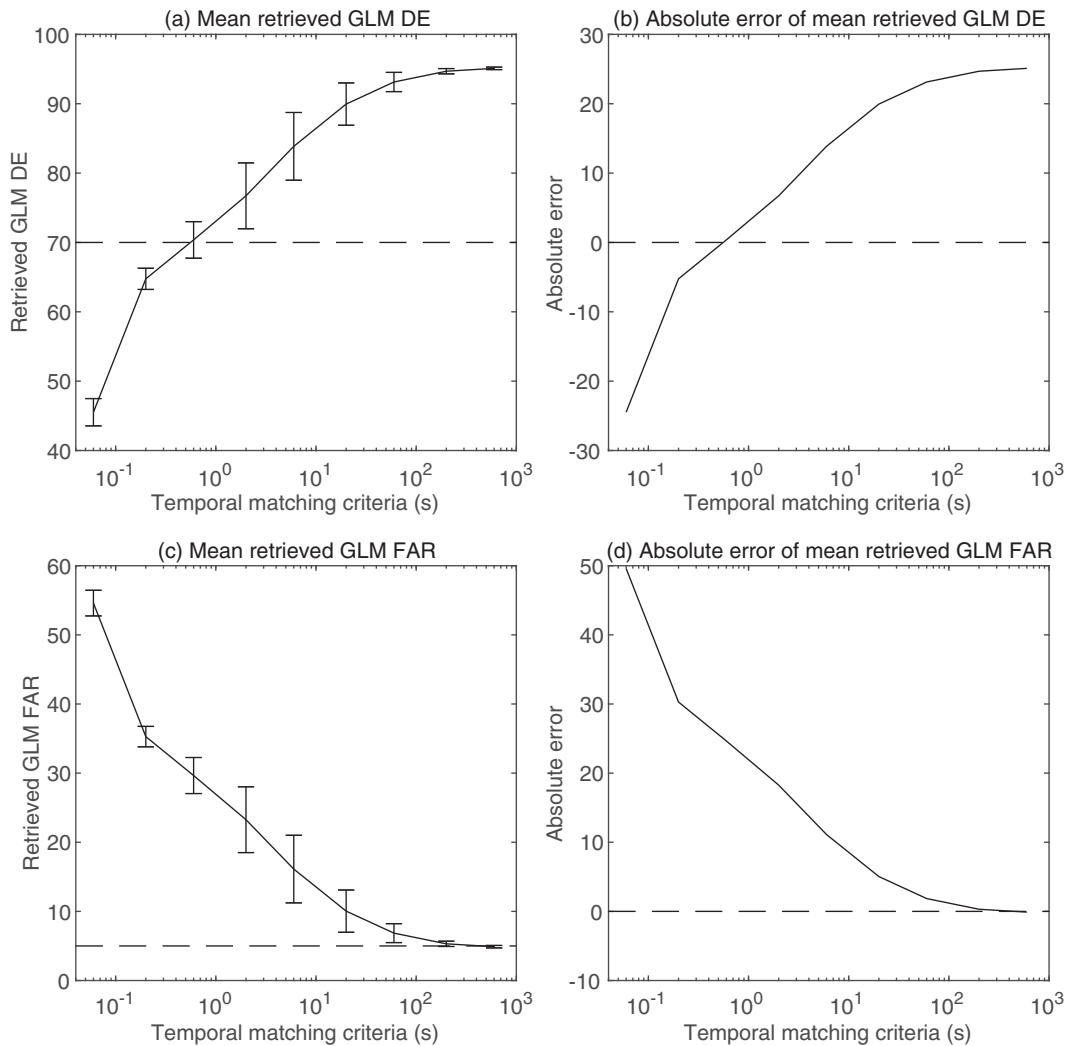


FIG. 7. Simulation results varying the temporal matching criteria (for full specifics, see scenario 5 in Table 1). (a) Mean and (b) absolute error of mean retrieved GLM DE as a function of temporal matching criteria. (c),(d) As in (a) and (b), but for retrieved GLM FAR. Error bars in (a) and (c) represent two standard deviations, and dashed lines indicate the true GLM metrics. Note the logarithmic scale of the abscissa.

known. This study explored the impact of imperfect reference network performance on two primary GLM performance metrics, the detection efficiency (DE) and false alarm rate (FAR). Exploration was performed via Monte Carlo simulations with eight tunable parameters: true GLM DE, true GLM FAR, the magnitude of the GLM spatiotemporal errors, true reference network DE, true reference network FAR, the magnitude of the reference network spatiotemporal errors, and the distance and temporal matching criteria applied. A variety of simulation scenarios were explored with various combinations of realistic GLM and reference network characteristics (Table 1).

The retrieved GLM DE varies as a function of true GLM DE and true reference network FAR (Figs. 2, 3) but is independent of the true reference network DE. Similarly, the retrieved GLM FAR varies as a function of true GLM FAR

and true reference network DE but is independent of the true reference network FAR. Reference network flash DE aggregated over the Western Hemisphere has been reported in the literature as ~30%–40% (Rudlosky 2015; Rudlosky et al. 2017), while the upper limit of reference network stroke DE has been estimated at 55%–60% (Bitzer and Burchfield 2016). Low reference network DE can increase errors in the retrieved GLM FAR up to 50%–60% (Fig. 3).

The two reference networks employed in this study exhibit strong temporal agreement throughout the GLM domain, but their spatial agreement deteriorates away from the well-sampled CONUS and central South America regions (Fig. 4). As uncertainties in reference network lightning geolocation increase, so do the uncertainties in the retrieved GLM performance metrics. For locations where reference network distance offsets are 400%–800% of the observed values (i.e., peaking near the

TABLE 2. Overview of Monte Carlo simulations for investigating the optimal matching criteria, indicating the true (applied) value of each tunable parameter. See text for further details. Boldface font indicates the parameters being varied in the simulation (i.e., parameters not at their nominal values).

Simulation type	6) Vary temporal matching criteria and GLM performance	7) Vary temporal matching criteria and reference network performance
Simulation specifics	GLM DE: 30%, 40%, ..., 100% GLM FAR: 0%, 10%, 20% Reference network DE: 70% Reference network FAR: 5% GLM offsets: Observed Reference network offsets: Observed Distance matching criteria: Initial (50 km) Temporal matching criteria: 0.3, 1, 3, 10, ..., 3000 × initial (±200 ms)	GLM DE: 70% GLM FAR: 5% Reference network DE: 30%, 40%, ..., 100% Reference network FAR: 0%, 2.5%, 5% GLM offsets: Observed Reference network offsets: Observed Distance matching criteria: Initial (50 km) Temporal matching criteria: 0.3, 1, 3, 10, ..., 3000 × initial (±200 ms)

observed values over the South Pacific), retrieved GLM DE can be underestimated up to 30%–40% and retrieved FAR can be overestimated up to 55%–65%.

Uncertainties in reference network DE and location accuracy prompted the authors of a recent study to experiment

with broadened matching criteria in their evaluation of GLM performance (Bateman et al. 2021). Our Monte Carlo simulations illustrate the impact of varying matching criteria on retrieved GLM performance. Varying the distance matching criteria from 50% to 800% of the initial values (i.e., 25–400 km)

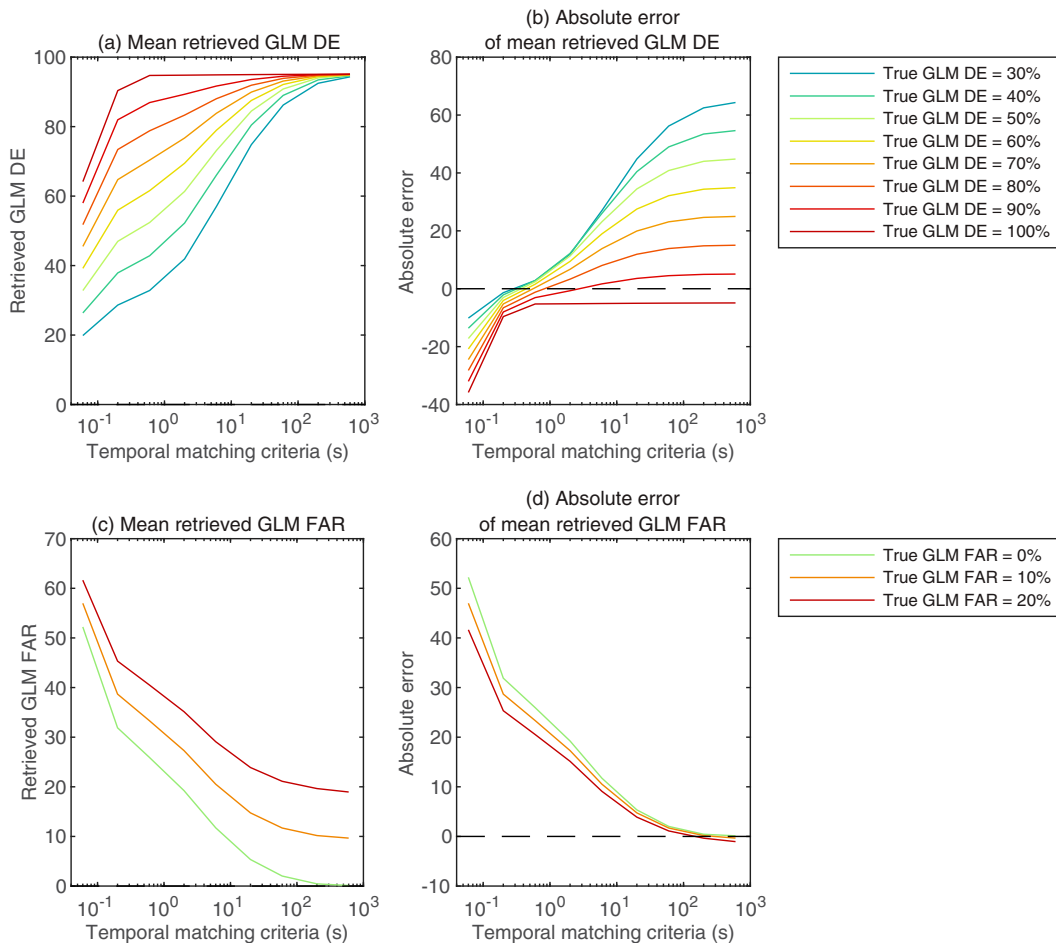


FIG. 8. Simulation results varying the temporal matching criteria and true GLM DE and FAR (for full specifics, see scenario 6 in Table 2). (a) Mean and (b) absolute error of mean retrieved GLM DE as a function of temporal matching criteria, for varying true GLM DE (colored lines). (c),(d) As in (a) and (b), but for retrieved GLM FAR, for varying true GLM FAR (colored lines). Dashed lines in (b) and (d) indicate absolute errors of 0. Note the logarithmic scale of the abscissa.

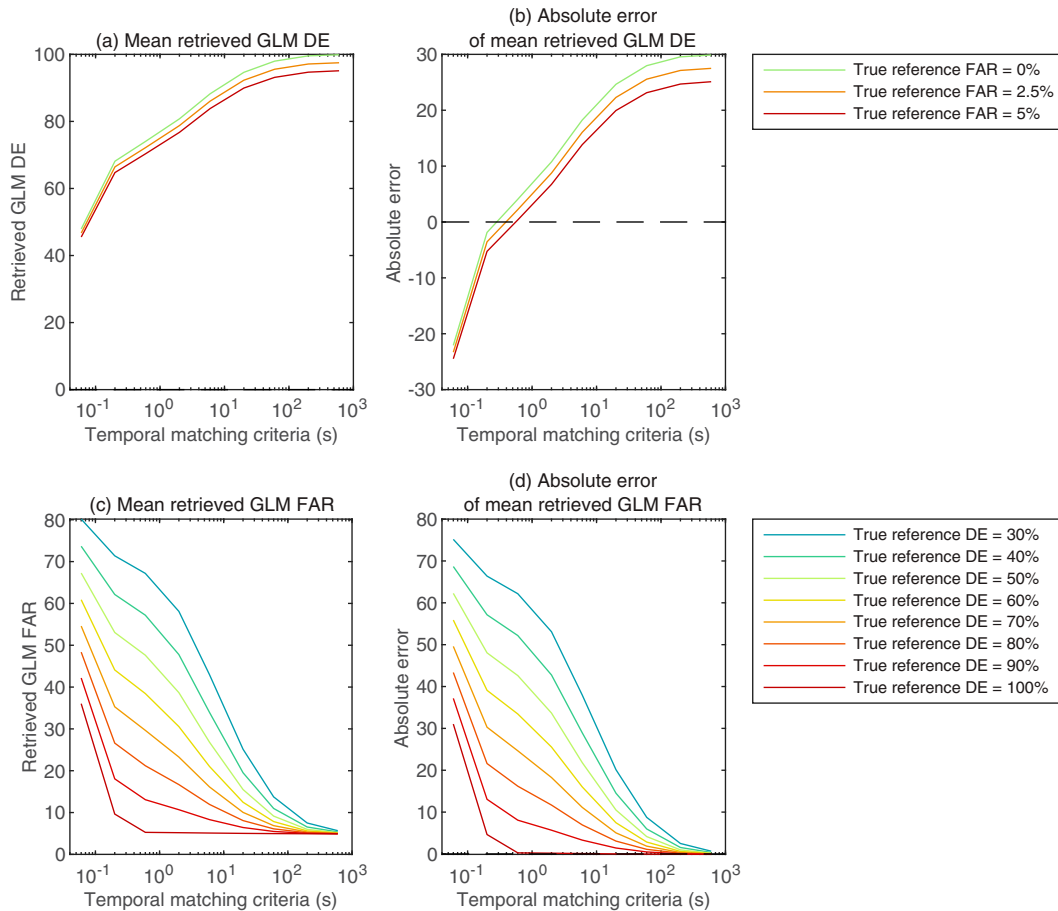


FIG. 9. Simulation results varying the temporal matching criteria and true reference network DE and FAR (for full specifics, see scenario 7 in Table 2). (a) Mean and (b) absolute error of mean retrieved GLM DE as a function of temporal matching criteria, for varying true reference network FAR (colored lines). (c),(d) As in (a) and (b), but for retrieved GLM FAR, for varying true reference network DE (colored lines). Dashed lines in (b) and (d) indicate absolute errors of 0. Note the logarithmic scale of the abscissa.

produces variations up to 10% in the retrieved GLM DE, while the retrieved FAR errors decrease by ~15% (Fig. 6). The standard deviation of the retrievals increases substantially as the distance criteria are broadened, as the ability to identify a “match”

TABLE 3. Overview of Monte Carlo geographic simulations, indicating the true (applied) value of each tunable parameter. See text for further details. Boldface font indicates the parameters being varied in the simulation (i.e., parameters not at their nominal values).

Simulation type	Vary reference network geographical performance
Simulation specifics	GLM DE: 70% at each grid box GLM FAR: 5% at each grid box Reference network DE/FAR: “Best” and “worst” scenario at each grid box GLM offsets: Observed aggregate values Reference network offsets: Observed at each grid box Distance matching criteria: Initial (50 km) Temporal matching criteria: Initial (± 200 ms)

increasingly depends on whether other convective storms are nearby.

A more stable approach is to broaden the temporal matching criteria (Figs. 7–9), which permits the matching of different flashes produced by the same thunderstorm. For matching criteria less than approximately ± 1 s, the retrieved performance metrics are quite sensitive to the exact matching window employed. As the criteria further broaden toward ± 200 s, the performance metrics exhibit asymptotic behavior. The retrieved GLM DE asymptotes to ~95% regardless of the true GLM DE, corresponding to the true reference network FAR of 5%, while the retrieved FAR asymptotes to the true GLM FAR of 5%. Our results give context to recent work by Bateman et al. (2021), who explored broadening the temporal matching criteria and reported aggregate DE values as high as >90% for both GLM-16 and GLM-17 and FAR values of ~5% (GLM-16) and <20% (GLM-17) when applying temporal matching criteria of ± 10 min. Our simulation results suggest that employing these broadest criteria overestimates the true GLM DE but is necessary in order to obtain a reasonable estimate of the true GLM

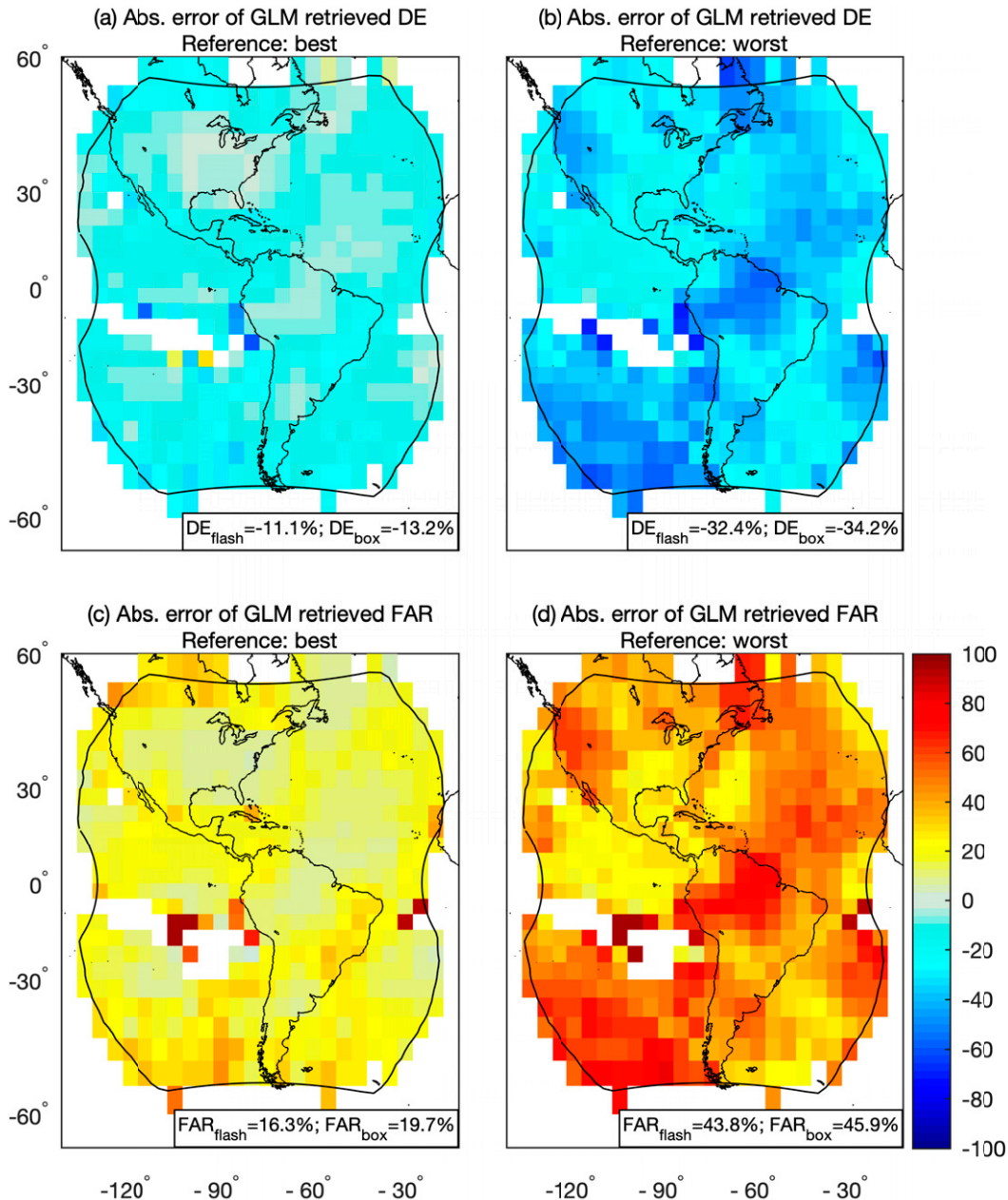


FIG. 10. Simulation results varying the reference network geographical performance. Absolute error of mean retrieved GLM DE when reference network performance (DE and FAR) is the (a) best and (b) worst scenario at each grid box. Retrieved DE errors aggregated over all flashes and all grid boxes are shown in the lower-right corner. (c),(d) As in (a) and (b), but for retrieved GLM FAR. Nominal GLM-16 field of view is shown by black line. See Table 3 and text for simulation details.

FAR. Sensitivity studies (Figs. 8, 9) suggest that matching windows on the order of ± 1 s and \sim minutes provide the most accurate estimates of aggregate GLM DE and FAR, respectively.

Published statistics on geographic variations in reference network coverage (Rudlosky 2015; Rudlosky et al. 2017; Murphy and Said 2020) prompted additional simulations to explore the impact of these variations on retrieved GLM

performance. We matched ENGLN and GLD360 flashes to each other to derive realistic upper and lower bounds on reference network performance throughout the GLM field of view. Geographic simulations indicate that retrieved GLM DE aggregated over all flashes is underestimated by $\sim 11\%$ – 32% , while the retrieved GLM FAR is overestimated by $\sim 16\%$ – 44% (Fig. 10). Errors of similar magnitude are

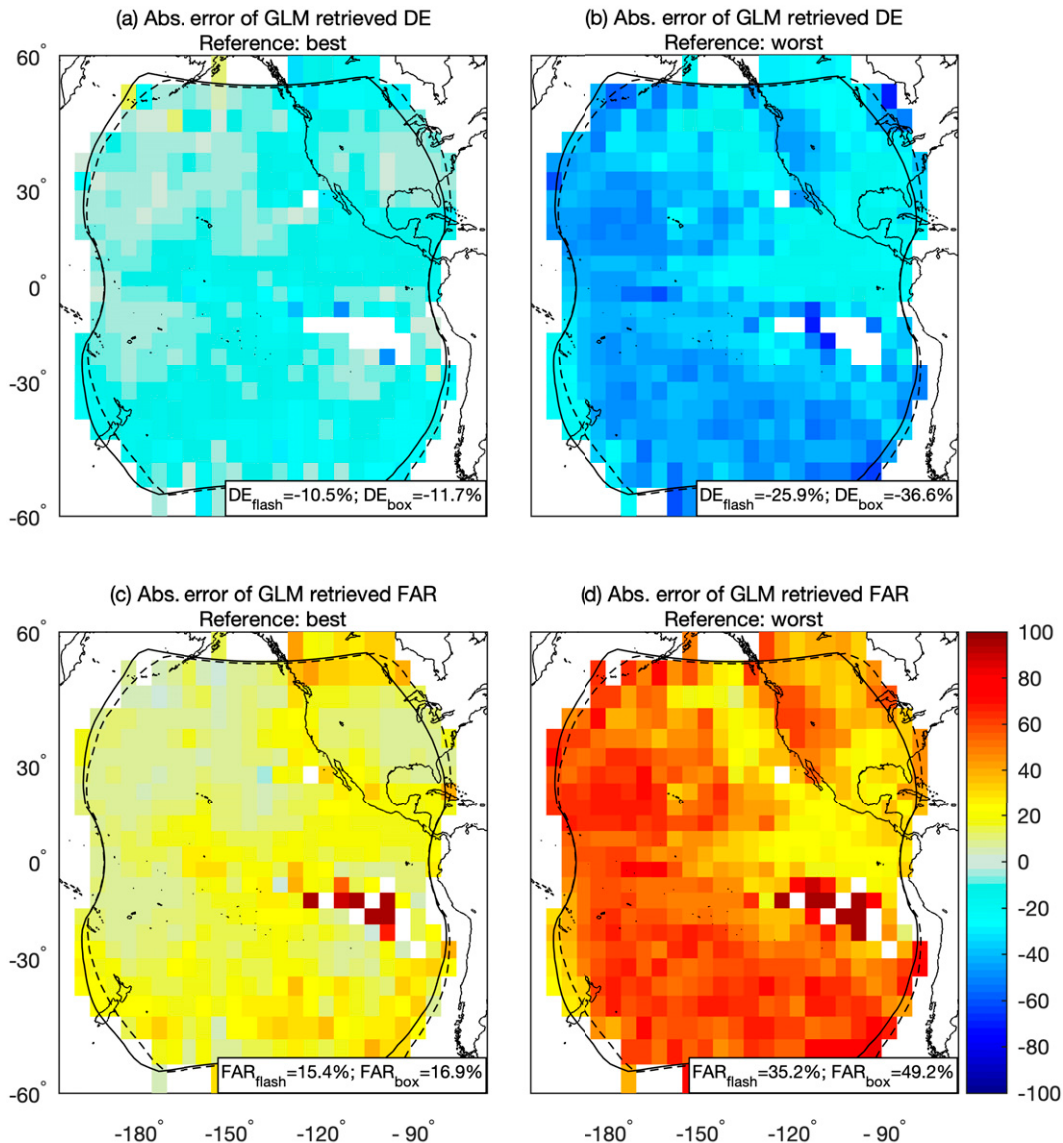


FIG. 11. As in Fig. 10, but for GLM-17. Nominal GLM-17 field of view for the 6 months each year when the satellite is in inverted yaw position is shown by dashed black line.

observed for GLM-17 (Fig. 11). GLM performance is most severely underestimated in the Southern Hemisphere Pacific, where retrieval errors for even the *best* reference network scenario range up to 20%–40% for DE and 30%–50% for FAR.

In summary, our Monte Carlo simulation tool developed and applied in this work has demonstrated how a more accurate assessment of GLM performance can be obtained given the imperfections in the performance characteristics of the reference networks used for validating GLM. While deploying additional reference network systems would be required to fully clarify GLM performance, this new tool provides an avenue for understanding GLM bulk performance metrics without the added time and cost of installing new or expanded ground networks. The authors plan to continue using this tool

to assess GLM-18, which reached provisional-level maturity validation in October 2022.

Acknowledgments. The authors are thankful to have carried out this work under the GOES-R Series Science, Demonstration, and Cal/Val Program at Marshall Space Flight Center funded by NOAA/NESDIS Senior Scientific Adviser for GOES-R series, Dr. Daniel Lindsey, and the NOAA/NESDIS/STAR GOES-R series Algorithm Working Group Program Lead, Dr. Jaime Daniels.

Data availability statement. GLM data are accessible through the NOAA Comprehensive Large Array-data Stewardship System (CLASS; class.noaa.gov). Earth Networks and Vaisala Inc.

collected and provided the ENGLN and GLD360 data, respectively, to the Global Hydrology Resource Center (GHRC) as part of GLM cal/val activities.

REFERENCES

- Abarca, S. F., K. L. Corbosiero, and T. J. Galarneau Jr., 2010: An evaluation of the Worldwide Lightning Location Network (WWLLN) using the National Lightning Detection Network (NLDN) as ground truth. *J. Geophys. Res.*, **115**, D18206, <https://doi.org/10.1029/2009JD013411>.
- Bateman, M., and D. Mach, 2020: Preliminary detection efficiency and false alarm rate assessment of the Geostationary Lightning Mapper on the GOES-16 satellite. *J. Appl. Remote Sens.*, **14**, 032406, <https://doi.org/10.1117/1.JRS.14.032406>.
- , —, and M. Stock, 2021: Further investigation into detection efficiency and false alarm rate for the Geostationary Lightning Mappers aboard GOES-16 and GOES-17. *Earth Space Sci.*, **8**, e2020EA001237, <https://doi.org/10.1029/2020EA001237>.
- Bitzer, P. M., 2017: Global distribution and properties of continuing current in lightning. *J. Geophys. Res. Atmos.*, **122**, 1033–1041, <https://doi.org/10.1002/2016JD025532>.
- , and J. C. Burchfield, 2016: Bayesian techniques to analyze the merge lightning locating system data. *Geophys. Res. Lett.*, **43**, 12 605–12 613, <https://doi.org/10.1002/2016GL071951>.
- Blakeslee, R. J., and Coauthors, 2020: Three years of the lightning imaging sensor onboard the International Space Station: Expanded global coverage and enhanced applications. *J. Geophys. Res. Atmos.*, **125**, e2020JD032918, <https://doi.org/10.1029/2020JD032918>.
- Brown, P. G., and Coauthors, 2019: The Hamburg meteorite fall: Fireball trajectory, orbit, and dynamics. *Meteorit. Planet. Sci.*, **54**, 2027–2045, <https://doi.org/10.1111/maps.13368>.
- Bruning, E. C., and Coauthors, 2019: Meteorological imagery for the Geostationary Lightning Mapper. *J. Geophys. Res. Atmos.*, **124**, 14 285–14 309, <https://doi.org/10.1029/2019JD030874>.
- Chmielewski, V. C., and E. C. Bruning, 2016: Lightning mapping array flash detection performance with variable receiver thresholds. *J. Geophys. Res. Atmos.*, **121**, 8600–8614, <https://doi.org/10.1002/2016JD025159>.
- Earth Networks, 2014: Earth Networks Global Lightning Network. Global Hydrology Resource Center Distributed Active Archive Center, accessed 3 January 2022, <https://www.earthnetworks.com/why-us/networks/lightning/>.
- GOES-R Algorithm Working Group and GOES-R Series Program, 2018: NOAA GOES-R series Geostationary Lightning Mapper (GLM) level 2 lightning detection: Events, groups, and flashes. NOAA National Centers for Environmental Information, accessed 3 January 2022, <https://doi.org/10.7289/V5KH0KK6>.
- Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.*, **125–126**, 34–49, <https://doi.org/10.1016/j.atmosres.2013.01.006>.
- Jenniskens, P., and Coauthors, 2018: Detection of meteoroid impacts by the Geostationary Lightning Mapper on the GOES-16 satellite. *Meteorit. Planet. Sci.*, **53**, 2445–2469, <https://doi.org/10.1111/maps.13137>.
- Kong, R., M. Xue, A. O. Fierro, Y. Jung, C. Liu, E. R. Mansell, and D. R. MacGorman, 2020: Assimilation of GOES-R Geostationary Lightning Mapper flash extent density data in GSI EnKF for the analysis and short-term forecast of a mesoscale convective system. *Mon. Wea. Rev.*, **148**, 2111–2133, <https://doi.org/10.1175/MWR-D-19-0192.1>.
- Koshak, W. J., 2017: Lightning NO_x estimates from space-based lightning imagers. *16th Annual Community Modeling and Analysis System Conf.*, Chapel Hill, NC, CMAS, <https://core.ac.uk/reader/141519610>.
- Lang, T. J., and Coauthors, 2020: The RELAMPAGO Lightning Mapping Array: Overview and initial comparison with the Geostationary Lightning Mapper. *J. Atmos. Oceanic Technol.*, **37**, 1457–1475, <https://doi.org/10.1175/JTECH-D-20-0005.1>.
- Liu, C., and S. Heckman, 2012: Total lightning data and real-time severe storm prediction. *WMO Technical Conf. on Meteorological and Environmental Instruments and Methods of Observation*, Brussels, Belgium, WMO, P5(10), <https://community.wmo.int/teco-2012-programme>.
- Mach, D. M., 2020: Geostationary Lightning Mapper clustering algorithm stability. *J. Geophys. Res. Atmos.*, **125**, e2019JD031900, <https://doi.org/10.1029/2019JD031900>.
- , H. J. Christian, R. J. Blakeslee, D. J. Boccippio, S. J. Goodman, and W. L. Boeck, 2007: Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor. *J. Geophys. Res.*, **112**, D09210, <https://doi.org/10.1029/2006JD007787>.
- Mallick, S., and Coauthors, 2014: Evaluation of the GLD360 performance characteristics using rocket-and-wire triggered lightning data. *Geophys. Res. Lett.*, **41**, 3636–3642, <https://doi.org/10.1002/2014GL059920>.
- Murphy, M. J., and R. K. Said, 2020: Comparisons of lightning rates and properties from the U.S. National Lightning Detection Network (NLDN) and GLD360 with GOES-16 Geostationary Lightning Mapper and Advanced Baseline Imager data. *J. Geophys. Res. Atmos.*, **125**, e2019JD031172, <https://doi.org/10.1029/2019JD031172>.
- Rudlosky, S. D., 2015: Evaluating ENTLN performance relative to TRMM/LIS. *J. Oper. Meteor.*, **3**, 11–20, <https://doi.org/10.15191/nwajom.2015.0302>.
- , and D. T. Shea, 2013: Evaluating WWLLN performance relative to TRMM/LIS. *Geophys. Res. Lett.*, **40**, 2344–2348, <https://doi.org/10.1002/grl.50428>.
- , and K. S. Virts, 2021: Dual Geostationary Lightning Mapper observations. *Mon. Wea. Rev.*, **149**, 979–998, <https://doi.org/10.1175/MWR-D-20-0242.1>.
- , M. J. Peterson, and D. T. Kahn, 2017: GLD360 performance relative to TRMM LIS. *J. Atmos. Oceanic Technol.*, **34**, 1307–1322, <https://doi.org/10.1175/JTECH-D-16-0243.1>.
- , S. J. Goodman, K. S. Virts, and E. C. Bruning, 2019: Initial Geostationary Lightning Mapper observations. *Geophys. Res. Lett.*, **46**, 1097–1104, <https://doi.org/10.1029/2018GL081052>.
- , —, K. Calhoun, C. Schultz, A. Back, B. Kuligowski, S. Stevenson, and C. Gravelle, 2020: Geostationary Lightning Mapper value assessment. NOAA Tech. Rep. NESDIS 153, 46 pp.
- Rumpf, C. M., R. S. Longenbaugh, C. E. Henze, J. C. Chavez, and D. L. Mathias, 2019: An algorithmic approach for detecting bolides with the Geostationary Lightning Mapper. *Sensors*, **19**, 1008, <https://doi.org/10.3390/s19051008>.
- Rutledge, S. A., K. A. Hilburn, A. Clayton, B. Fuchs, and S. D. Miller, 2020: Evaluating Geostationary Lightning Mapper flash rates within intense convective storms. *J. Geophys. Res. Atmos.*, **125**, e2020JD032827, <https://doi.org/10.1029/2020JD032827>.
- Said, R., and M. Murphy, 2016: GLD360 upgrade: Performance analysis and applications. *24th Int. Lightning Detection Conf./Sixth Int. Lightning Meteorology Conf.*, San Diego, CA,

- Vaisala, <https://www.vaisala.com/sites/default/files/documents/Ryan%20Said%20and%20Martin%20Murphy.%20GLD360%20Upgrade%20Performance%20Analysis%20and%20Applications.pdf>.
- , M. B. Cohen, and U. S. Inan, 2013: Highly intense lightning over the oceans: Estimated peak currents from global GLD360 observations. *J. Geophys. Res. Atmos.*, **118**, 6905–6915, <https://doi.org/10.1002/jgrd.50508>.
- Vaisala, 2014: Global Lightning Dataset GLD360. Global Hydrology Resource Center Distributed Active Archive Center, accessed 3 January 2022, <https://ghrc.nsstc.nasa.gov/lightning/>.
- Virts, K. S., and W. J. Koshak, 2020: Mitigation of Geostationary Lightning Mapper geolocation errors. *J. Atmos. Oceanic Technol.*, **37**, 1725–1736, <https://doi.org/10.1175/JTECH-D-19-0100.1>.
- Wang, Y., N. Yussouf, E. R. Mansell, B. C. Matilla, R. Kong, M. Xue, and V. C. Chmielewski, 2021: Impact of assimilating GOES-R Geostationary Lightning Mapper flash extent density data on severe convection forecasts in a warn-on-forecast system. *Mon. Wea. Rev.*, **149**, 3217–3241, <https://doi.org/10.1175/MWR-D-20-0406.1>.
- Zhang, D., and K. L. Cummins, 2020: Time evolution of satellite-based optical properties in lightning flashes, and its impact on GLM flash detection. *J. Geophys. Res. Atmos.*, **125**, e2019JD032024, <https://doi.org/10.1029/2019JD032024>.
- Zhu, Y., M. Stock, J. Lapierre, and E. DiGangi, 2022: Upgrades of the Earth Networks Total Lightning Network in 2021. *Remote Sens.*, **14**, 2209, <https://doi.org/10.3390/rs14092209>.