Reply on RC1

Major Comments

Comment 1:

Physics-based corrections (e.g., payload misalignment, cloud-top height parallax, thermal deformation), followed by an additional empirical correction based on the WWLN were applied. This combination is reasonable, but it remains unclear why a substantial residual empirical correction is still required after applying physically based models. Whether these arise from CTH uncertainties, instrument/navigation limitations, or incomplete physical representation.

Response:

We agree that this is a crucial question, and the description in our manuscript was not sufficiently detailed. In the following, we provide a detailed answer and will add the corresponding content in the revised manuscript. In practice, the physics-based corrections applied to FY-4A LMI are still imperfect for the following reasons:

(1) The cloud-top height (CTH) parallax correction depends on the accuracy of CTH retrievals from FY-4A satellite data, which suffers from spatiotemporal uncertainties.
Spatially: Comparison between satellite CTH products and radar observations shows a vertical error of about 1 km. Assuming all other conditions remain unchanged, the resulting parallax correction error is also on the order of kilometers.
Temporally: The current satellite CTH product has a temporal resolution of 15 minutes. For multilayer clouds and rapidly developing convective cells, the CTH value matched to a lightning event is not the CTH at the exact moment of the discharge. Time interpolation can only approximate the true value, inevitably introducing errors.
Ideally, radar data with higher spatiotemporal resolution and accuracy could further reduce parallax errors, but our goal is full-coverage, all-timescale correction. Radar data are difficult to obtain over the entire domain, so we are limited to satellite CTH products. Nevertheless, even with satellite CTH, the residual error after parallax correction should not exceed 10 km. However, when we match LMI events with ground-based lightning location networks, we find residual errors of 15 km or more, and these errors exhibit clear periodic variations with time and geographic location. This suggests that the residual is likely due to thermal deformation or other periodic errors, which leads to the second point.

(2) According to our experimental results (as shown in Fig. 1), the uncorrected LMI Level-2 products still exhibit large residual errors, indicating that the corrections applied in the operational algorithm are incomplete. As illustrated by the severe convective system case over Beijing on 4 August 2019, the locations of LMI events show a systematic southwestward

deviation relative to the lightning strokes detected by BLNET and WWLLN. These deviations are not a fixed overall offset but rather systematic biases with diurnal variation and spatial heterogeneity (e.g., thermal deformation signatures), indicating that the physical corrections in the operational algorithm are incomplete. However, the physical corrections applied in the operational algorithm are not fully documented in the published literature. It is difficult and could lead to redundant corrections to pinpoint exactly which correction step(s) the residuals originate from. Therefore, the only feasible approach is to match LMI events with ground-based lightning location network data, statistically derive the spatiotemporal characteristics of these residuals, and thereby improve the accuracy of LMI products. This is precisely the main methodology of the present study: an empirical spline-fitting correction based on WWLLN data.

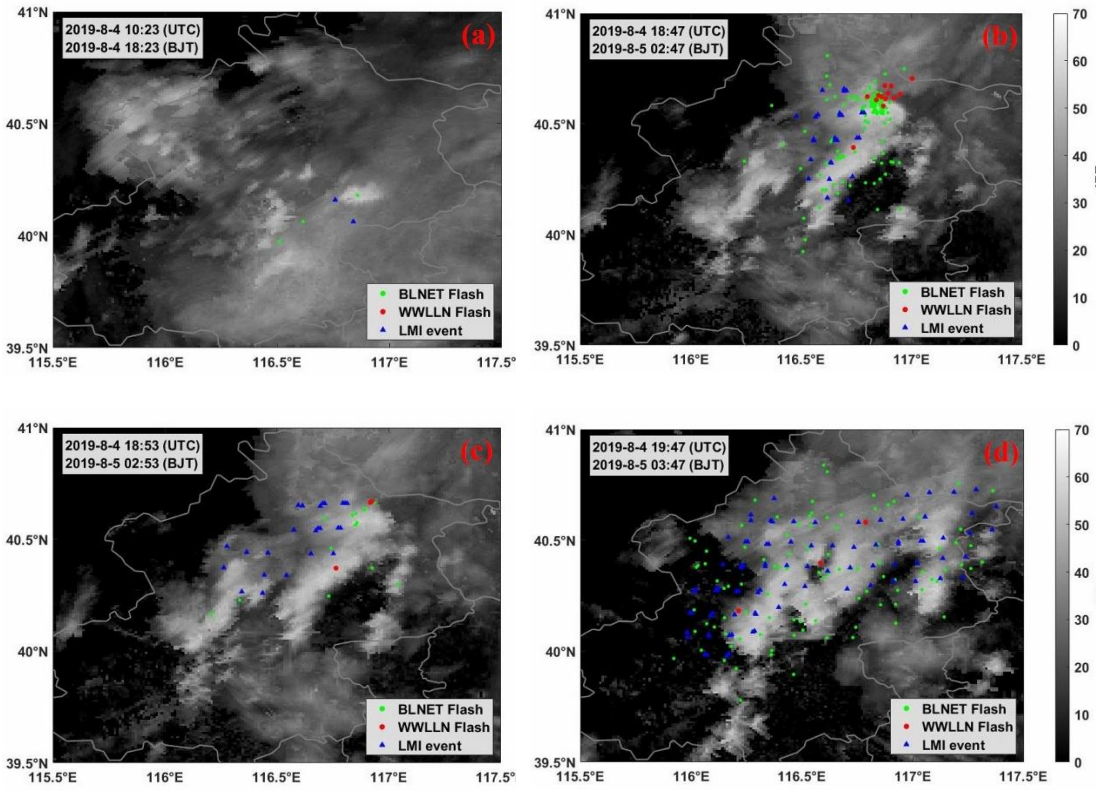

Figure 1: LMI event deviation at different stages of a severe convective event over Beijing on 4 August 2019.

(3) Additionally, random or semi-random errors such as satellite platform jitter, attitude control inaccuracies, and inter-detector response inconsistencies cannot be fully resolved by existing physical models. These errors are not yet systematically addressed in the literature. They contribute to the observed scatter in lightning locations and cannot be eliminated by deterministic physical corrections. Our empirical spline-fitting approach partially absorbs these random-like errors by leveraging statistical smoothing over time, reducing their impact on the final corrected dataset. Of course, the explanation of this statistical smoothing will be

detailed in our responses to other comments.

In summary, the WWLLN-based empirical correction serves as a "data-driven closure" that accounts for residual systematic biases not captured by existing physical models (including CTH interpolation errors, incomplete thermal deformation correction, and uncorrected platform jitter effects). We will add the above detailed discussion in the Introduction of the revised manuscript.

Comment 2:

Authors produced a "unified correction dataset" using subregional (20 × 20) curve fitting. But that depends on lightning occurrence. This questions about the spatial consistency, especially in data-sparse regions.

Response:

You raise a valid concern regarding the spatial consistency of our subregional correction, especially in data-sparse regions. We will clarify the rationale behind the 20×20 subdivision as follows.

The 20×20 grid is not arbitrary. In the operational processing chain of LMI, the full field of view is divided into a 4×4 grid (16 operational subregions, as shown in Fig. 2), within which detection parameters (background threshold, clustering criteria, etc.) are independently tuned (Hui et al., 2020). Consequently, each of these 16 operational subregions has its own homogeneous detection efficiency and noise characteristics.

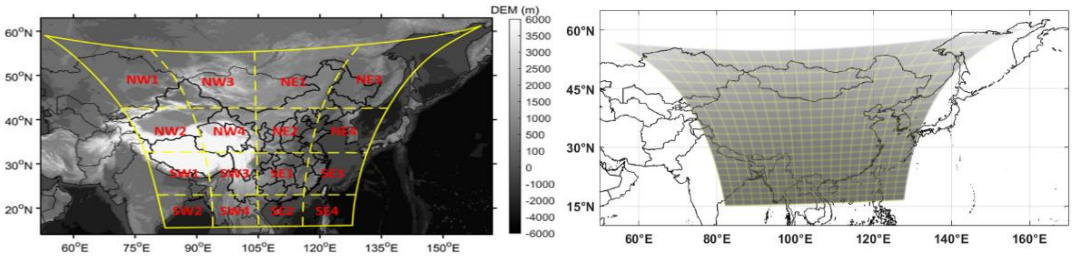

Figure 2. Subregion division in the operational algorithm, from Hui et al. (2020), and the subregion division used in this study.

To avoid mixing data from different operational subregions within a single analysis subregion—which would cause performance inconsistencies due to different detection parameters and algorithmic behaviors—we require that all lightning events within each of our analysis subregions belong to the same operational subregion. Therefore, the number of analysis subregions along each dimension must be a multiple of 4, such as 8×8, 20×20, or 40×40. We tested multiples in the range of 16×16, 20×20, and 24×24 to determine the optimal subdivision. The 16×16 subdivision resulted in subregions that were too coarse

(each covering ~2° in latitude/longitude), smoothing over the systematic deviation patterns we aimed to resolve. Moreover, although we expected that larger subregions would provide more samples, in some marginal areas with sparse lightning occurrence, larger subregions actually expanded the spatial extent of data scarcity, further reducing the effective sample density. The 24×24 subdivision produced subregions that were too small, leading to insufficient sample sizes in many subregions, especially in marginal or low-lightning areas. The 20×20 subdivision strikes an optimal balance, providing sufficient spatial resolution while maintaining robust sample sizes (median >500 events per day in active regions). Supporting evidence comes from Zhang et al. (2023), who applied CTH parallax correction to the Beijing region (approximately 2° × 2° grid). They found that within such a grid, the longitude and latitude correction values for all lightning events were very close under the same conditions, essentially following a straight line (as shown in Fig. 2). This demonstrates that within a grid of about 2° size, the deviation characteristics are relatively homogeneous. Our 20×20 subdivision results in subregions of comparable scale, which justifies the assumption that a single fitted correction curve is applicable to all events within a subregion. We will add the above detailed explanation in Section 3.1 of the revised manuscript.

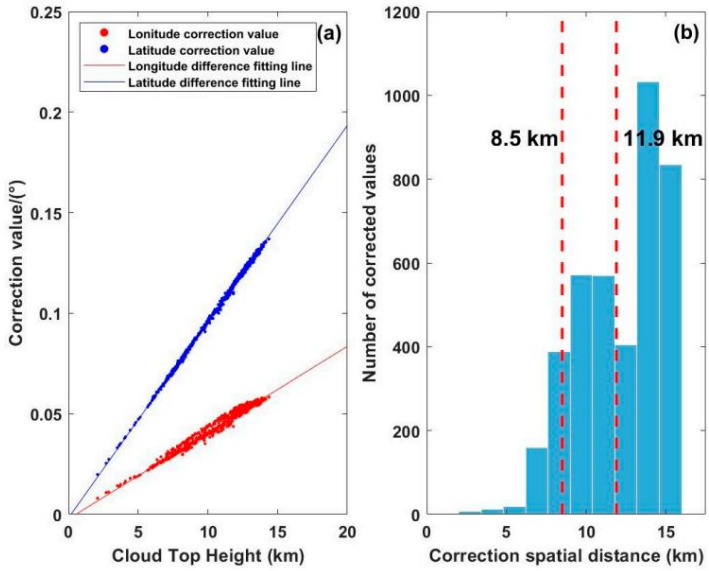

Figure 3 Correction values of latitude–longitude differences and spatial distance of LMI event from Zhang et.al., (2023). https://doi.org/10.3390/rs15194856

Comment 3:

The manuscript relies on WWLLN. It has biases due to its low and spatially variable detection efficiency, bias toward strong CG strokes, and location uncertainty (several km). The authors should therefore: Quantify residual error after physics-based correction alone, Justify the need for the empirical step, and Discuss how WWLLN limitations affect the robustness of the final dataset. Authors partially justified the use of the WWLN (by its global coverage); however,

this should be balanced against its known limitations in detection efficiency and accuracy.

Response:

We thank the reviewer for raising this important issue. In the revised manuscript, we have addressed all three requested aspects as follows.

1. Quantification of residual errors after physics-based correction alone

As originally shown in Fig. 4, the residual errors after applying only the physics-based corrections were presented, but a detailed quantitative description was missing. In the revised manuscript, we will add the following quantitative analysis. Based on over nine million matched LMI–WWLLN pairs from 2019 to 2023, after applying only the existing physics-based corrections (CTH parallax, navigation calibration, and partial thermal deformation correction), the residual geolocation errors are substantial: only 58.2% of LMI events have a deviation within 20 km. For the zonal and meridional components, 82.9% and 83.5% of differences lie within 20 km of the WWLLN reference, respectively. These statistics confirm that systematic residuals of approximately 20–30 km persist after the existing physical corrections.

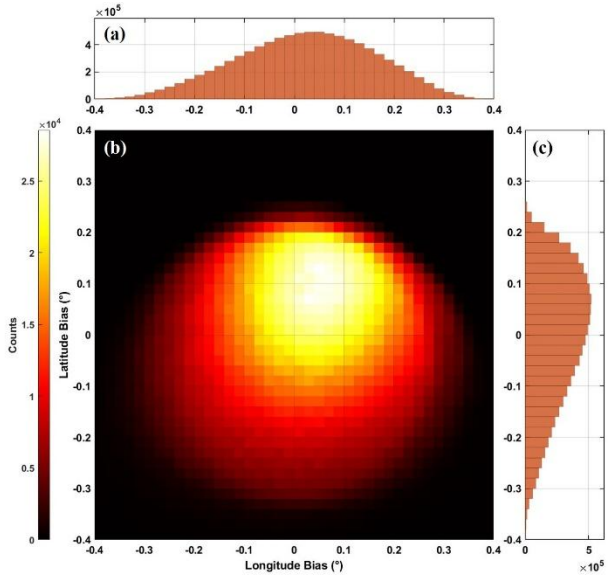

Figure 4: Distributions of coordinate differences between LMI events and WWLLN flashes before correction. The statistics are based on data collected during the boreal warm season (March–September) from 2019 to 2023.

2. Justification for the empirical correction step

As explained in Comment 1, the remaining systematic residuals exhibit clear diurnal and spatial patterns (e.g., thermal deformation signatures and a persistent longitudinal bias), indicating that the existing physical corrections are incomplete. However, the operational LMI Level-2 products have undergone multiple correction steps, not all of which are fully

documented in the literature. Isolating the exact source of each residual is difficult and could lead to redundant corrections. Therefore, we adopt an inverse approach: using ground-based lightning location data (WWLLN) as a reference for lightning-dense regions, we empirically fit the spatiotemporal pattern of the residuals via a spline-based method. This empirical correction serves as a data-driven closure to absorb residual systematic biases not captured by the existing physical models (including CTH interpolation errors, incomplete thermal deformation correction, and uncorrected platform jitter effects). The necessity of this step is thus well justified.

3. Impact of WWLLN limitations on the robustness of the final dataset
We acknowledge that WWLLN has inherent limitations: detection efficiency over China is approximately 30–50% (with regional variations; e.g., ~72% over the Tibetan Plateau, Fan et al., 2018), location uncertainty of about 10 km, and a bias toward high-current CG strokes. It is important to note that the 10 km uncertainty does not imply a 10 km displacement for every stroke; rather, WWLLN tends to locate high-current events within a region, and its locations generally fall within the strongest radar reflectivity cores. Thus, WWLLN should be treated not as absolute ground truth but as a reference for the spatial distribution of lightning-dense areas. Our empirical correction aims to reduce the relative systematic offset between LMI and this reference, not to correct detection efficiency.

After applying our correction, the residual errors are significantly reduced: the proportion of LMI events with a deviation ≤15 km increases from 36.0% to 51.7%, and the proportion with a deviation ≤20 km increases from 58.2% to 74.8%. For the zonal component, 90.3% of longitude differences are within 20 km (compared to 82.9% before correction); for the meridional component, 90.5% of latitude differences are within 20 km (compared to 83.5% before correction). After correction, the average deviation for most LMI events is approximately 15 km, and these residual errors are comparable to WWLLN's own location uncertainty (~10 km). This indicates that further improvement would require a higher-quality reference (e.g., regional networks). To mitigate the impact of WWLLN's limitations, we have added an independent cross-validation using BLNET (a high-precision regional network) over Beijing to confirm that the correction improves LMI geolocation without being biased by WWLLN's deficiencies. For data-sparse regions where WWLLN performance is poor (e.g., Xinjiang and Mongolia), we do not apply the empirical correction. These discussions have been added to the revised manuscript (Sections 3, 4, and the Conclusions).

Comment 4:
Recent studies (e.g., Wu et al., 2024, https://doi.org/10.1016/j.jastp.2024.106194) using higher-quality regional networks, such as the National Lightning Monitoring Network, have

shown significant discrepancies in FY-4A LMI performance. The authors should:

1. Discuss the trade-off between coverage and data quality,
2. Evaluate how WWLLN limitations affect correction accuracy,
3. and Consider (or discuss) the role of regional high-DE networks for validation or hybrid approaches. The exclusion of oceanic regions further indicates that the correction is not uniformly applicable, which should be explicitly acknowledged.

Response:

We thank the reviewer for raising this important point regarding the trade-off between coverage and data quality.

1. Trade-off between coverage and data quality

We fully acknowledge that high-quality regional lightning location networks (e.g., National Lightning Monitoring Network (NLMN); Beijing Broadband Lightning Network, BLNET) have much higher detection efficiency (e.g., NLMN >90%) and location accuracy (e.g., NLMN better than 500 m) compared to WWLLN. However, their coverage is limited to specific areas of China (primarily the Chinese mainland), whereas the goal of our study is to perform a full-field-of-view correction of FY-4A LMI observations over the entire Northern Hemisphere, including China, Mongolia, Kazakhstan, Southeast Asian countries, and considerable ocean areas. This requires a reference dataset with global or near-global coverage, which currently only WWLLN can provide.

2. How WWLLN limitations affect correction accuracy

As detailed in our response to Comment 3, WWLLN has inherent limitations: detection efficiency over China is approximately 30–50% (with regional variations, e.g., ~72% over the Tibetan Plateau), location uncertainty of about 10 km, and a bias toward high-current CG strokes. Our empirical correction aims to reduce the relative systematic geolocation shift between LMI and this ground-based reference, not to perform a one-to-one absolute match and correction. After applying our correction, the residual errors are significantly reduced (e.g., the proportion of LMI events with a deviation ≤20 km increases from 58.2% to 74.8%). In most regions, the average deviation after correction is approximately 15 km. These residual errors are comparable to WWLLN's own location uncertainty, indicating that further improvement would indeed require higher-quality references such as regional networks.

3. Role of regional high-DE networks for validation and the applicability of correction over ocean areas

To independently evaluate the performance of our correction, we have added a cross-validation study over the Beijing area using BLNET, a high-precision total-lightning

network (see our response to Comment 5 for details). This provides a robust assessment of the residual errors and demonstrates the added value of using regional networks for validation.

Regarding ocean areas: Unlike regional networks, WWLLN's VLF-based detection capability can extend far into ocean areas. As shown in Fig. 4, the southeastern part of the LMI field of view (oceanic region) exhibits frequent lightning activity, and the derived geolocation deviations transition smoothly from land to ocean. This indicates that WWLLN is sufficiently reliable for lightning detection over ocean areas to a certain depth, and the correction can be applied there without distinguishing between "land lightning" and "ocean lightning". Therefore, we do not exclude ocean areas from the empirical correction; rather, we apply the same correction procedure wherever sufficient matched LMI – WWLLN pairs are available. The only exceptions are regions where lightning occurrence is extremely sparse where curve fitting is not feasible. This is already noted in the original manuscript (e.g., Xinjiang, Mongolia). We will explicitly clarify this point in the revised manuscript.
We will add all the above descriptions in the revised manuscript.

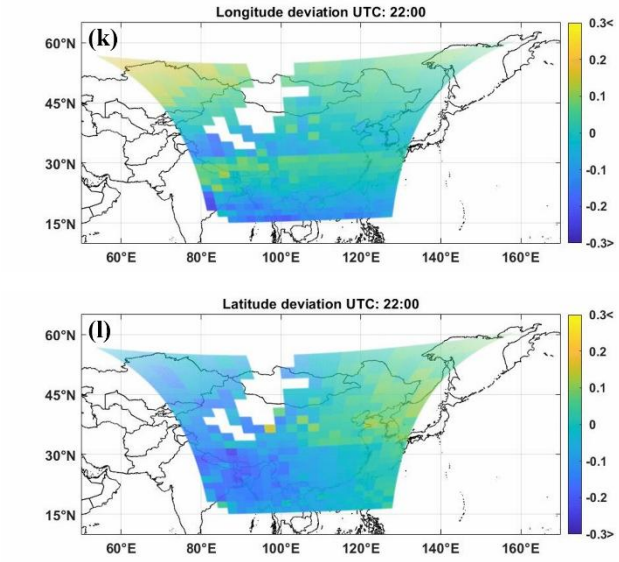

Figure 4: Spatiotemporal distribution of lightning geolocation deviations for LMI events over the Northern Hemisphere observational domain. The left column shows the longitudinal deviations, and the right column shows the latitudinal deviations.

Comment 5:

The authors should include case studies comparing lightning locations before and after correction with CTT/CTH and/or radar reflectivity, to demonstrate improved alignment with convective structures. Where possible, comparison with regional networks such as the China Lightning Detection Network or National Lightning Monitoring Network would strengthen validation.

Response:

We thank the reviewer for this constructive suggestion. As mentioned in our response to Comment 4, we will add a case study in the revised manuscript.

We selected a severe convective weather system that occurred over Beijing from 4 to 5 August 2019. Figure 5 presents a comparison of LMI event locations before and after correction, along with BLNET flashes, WWLLN flashes, and radar composite reflectivity. The time period from 18:00 to 20:00 UTC was chosen because the impact of thermal deformation was most significant during this interval. As shown in the figure, after correction, the LMI events align more closely with the areas of strongest radar reflectivity and show better agreement with both BLNET and WWLLN flash locations, indicating a notable improvement in correction performance. It should be noted that due to differences in detection principles, the detection efficiencies of LMI and WWLLN are relatively low. Consequently, the number of LMI events and WWLLN flashes displayed in the figure is significantly lower than that of BLNET flashes, which has a much higher detection efficiency. Another important point to note is that in Fig. 5b, some WWLLN flashes appear outside the strongest radar reflectivity cores. This is because WWLLN flash locations more promptly reflect the regions of most intense convective activity, whereas the radar reflectivity product has a lower temporal resolution (approximately 5 minutes). In the figure, all lightning strokes occurring between two consecutive radar frames are overlaid on the preceding radar frame, which can create an apparent "misalignment" effect.

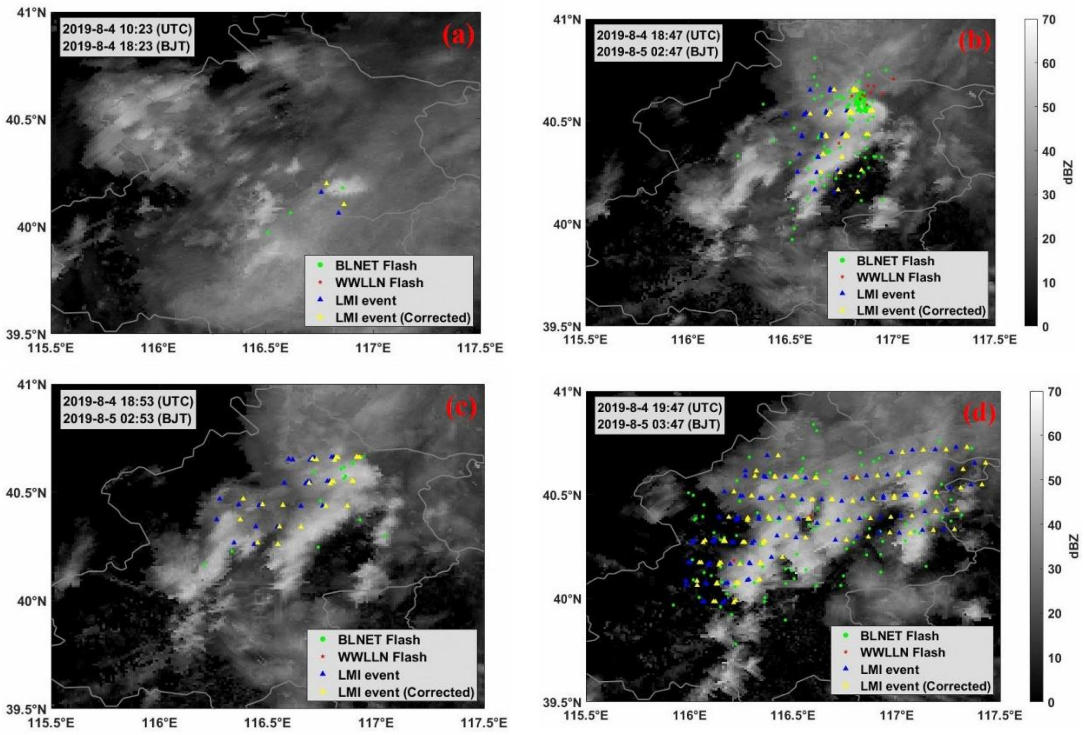

Figure 5: Comparison of LMI correction results before and after with BLNET and WWLLN flash. The red dots represent the matched WWLLN flash data, the green dots indicate the matched BLNET flash data, the yellow dots show the LMI events data after correction, and

the blue dots show the LMI events data before correction. The background indicates radar composite reflectivity.

In addition to the visual comparison, we also conducted a quantitative evaluation of the LMI geolocation for this severe convective system using BLNET as an independent reference. Figure 6 shows the distributions of coordinate deviations between LMI events (before and after correction) and the matched BLNET flashes, based on 20,710 matched pairs. Overall, before correction, LMI events exhibited a systematic southwestward deviation relative to BLNET flashes, WWLLN flashes, and the high-radar-reflectivity cores. After correction, the LMI events align more closely with the ground-based lightning location results and generally cover the high-radar-reflectivity areas. Specifically, the median distance error decreases from 20.11 km to 17.63 km, and the longitude error shows a notable improvement, with the median decreasing from 0.1449° to 0.1022°. The overall correction performance is satisfactory, demonstrating that the WWLLN-based empirical spline-fitting method effectively reduces systematic geolocation biases in LMI observations, as further validated by independent BLNET and radar data.

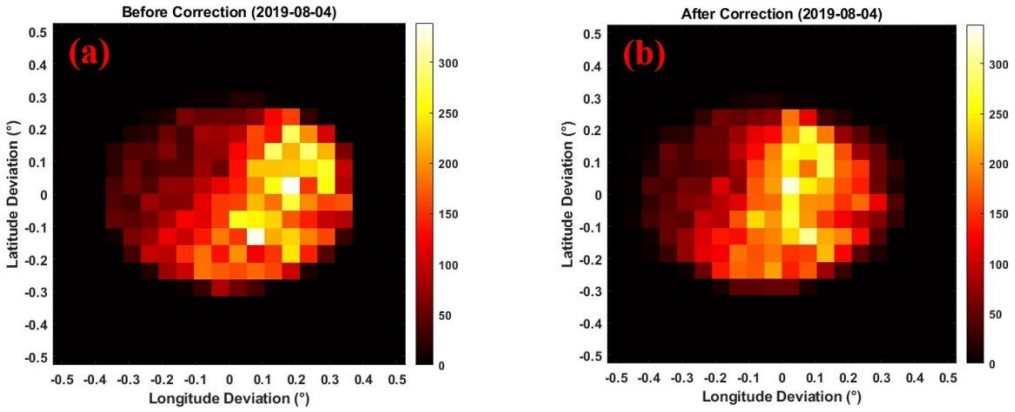

Figure 6: Distributions of coordinate differences between LMI events and BLNET flashes before and after correction.

Comment 6:

Please add a discussion of remaining limitations after correction is needed, including residual errors (~15 km), dependence on WWLLN, and implications for different applications (storm-scale vs climatology). This is particularly important given reduced reliability in low-lightning regions, as acknowledged by the authors.

Response:

Thanks to the reviewer for this important suggestion. In the revised manuscript, we have added a dedicated “Limitations” subsection in the Conclusions (as well as relevant discussions in Section 4 and the Discussion). The following points are now explicitly addressed.

1. Residual errors after correction

As detailed in our response to Comment 1, the existing physics-based corrections are imperfect due to multiple factors (CTH parallax uncertainties, incomplete thermal deformation correction, and uncorrected platform jitter). After applying our empirical correction, based on over nine million matched LMI–WWLLN pairs from 2019 to 2023, the proportion of LMI events with a geolocation deviation ≤15 km increases from 36.0% to 51.7%, and the proportion with a deviation ≤20 km increases from 58.2% to 74.8%. For the zonal component, 90.3% of longitude differences are within 20 km (compared to 82.9% before correction); for the meridional component, 90.5% of latitude differences are within 20 km (compared to 83.5% before correction). The median deviation after correction is approximately 14–18 km depending on the region. These residual errors are comparable to or slightly better than WWLLN's own location uncertainty (~10 km). Given LMI's native pixel resolution (~7.8 km at nadir, >20 km at the edge) and the inherent limitations of the reference data, further reduction of residual errors would require higher-quality reference data (e.g., regional networks with sub-kilometer accuracy) or fusion with other observations.

2. Dependence on WWLLN

As discussed in our response to Comment 3, WWLLN has inherent limitations: detection efficiency over China is approximately 30–50% (with regional variations, e.g., ~72% over the Tibetan Plateau), location uncertainty of about 10 km, and a bias toward high-current CG strokes. Our correction reduces the relative systematic offset between LMI and WWLLN but does not correct detection efficiency or random errors beyond WWLLN's own uncertainty. Users should be aware that the corrected dataset inherits the spatial and temporal sampling characteristics of WWLLN to some extent, particularly in regions where WWLLN station coverage is sparse.

3. Implications for different applications

Storm-scale studies (e.g., lightning data assimilation into convective-scale numerical models with typical grid spacings of 3–15 km, convective initiation, lightning jump, and storm tracking): The residual errors of 15–20 km for the majority of events are acceptable. As demonstrated in our added case study (Beijing, 4 August 2019), the corrected LMI events are generally collocated with the highest radar reflectivity cores. The remaining uncertainty should be considered when performing point-to-point comparisons with high-resolution observations. Climatological studies (e.g., gridded lightning climatology at 0.25° or coarser, trend analysis, regional lightning frequency mapping): The residual error is negligible because random components average out over long time periods and coarse spatial aggregation. The corrected dataset is well suited for climate-scale analyses without introducing significant bias.

4. Reduced reliability in low-lightning regions

We explicitly acknowledge that in regions where lightning occurrence is sparse (e.g., Xinjiang, Mongolia, and some remote ocean areas), the number of matched LMI – WWLLN events is insufficient for robust curve fitting. Therefore, we do not apply the empirical correction in such regions; instead, we retain the original physics-based corrections and flag the data accordingly. Users are advised to exercise caution when using data from these low-lightning-occurrence areas. This limitation is clearly stated in the revised manuscript (Section 4 and the Conclusions).

All the above descriptions have been added to the revised manuscript.

Comment 7:

It is not clear whether the correction varies with viewing geometry (e.g., satellite zenith angle) or latitude; this should be discussed.

Response:

We thank the reviewer for raising this question regarding the dependence of the correction on viewing geometry.

The LMI onboard FY-4A has a northward viewing inclination. Its focal plane array is a 400 × 300 × 2 CCD matrix (as described in Section 2.1), and when projected onto the Earth's surface, it forms an inverted trapezoidal field of view as shown in Fig. 7. Theoretically, if we assume that the influence of thermal deformation is uniformly distributed across the entire CCD array, then due to the geometric projection effect (northward inclination causing the inverted trapezoidal footprint), this uniform CCD-level deformation will project onto the Earth's surface as a geolocation bias that gradually increases from south to north. In other words, detector elements with larger northward viewing inclination (i.e., those closer to the northern edge of the field of view) will experience a larger projected displacement than those near the center or southern edge.

Fortunately, FY-4A is a geostationary satellite, meaning its position relative to the Earth is fixed, and the viewing geometry of the payload does not change over time. Therefore, we do not need to account for the complex, time-varying geometric transformations typical of polar-orbiting satellites. This stability allows us to apply the same correction scheme consistently to all time periods across the full dataset. In our subregional approach, the effect of viewing geometry is implicitly captured because each subregion corresponds to a fixed set of detector elements with constant viewing angles. As a result, the fitted correction curves derived for each subregion are directly applicable to all lightning events observed within that subregion, regardless of time.

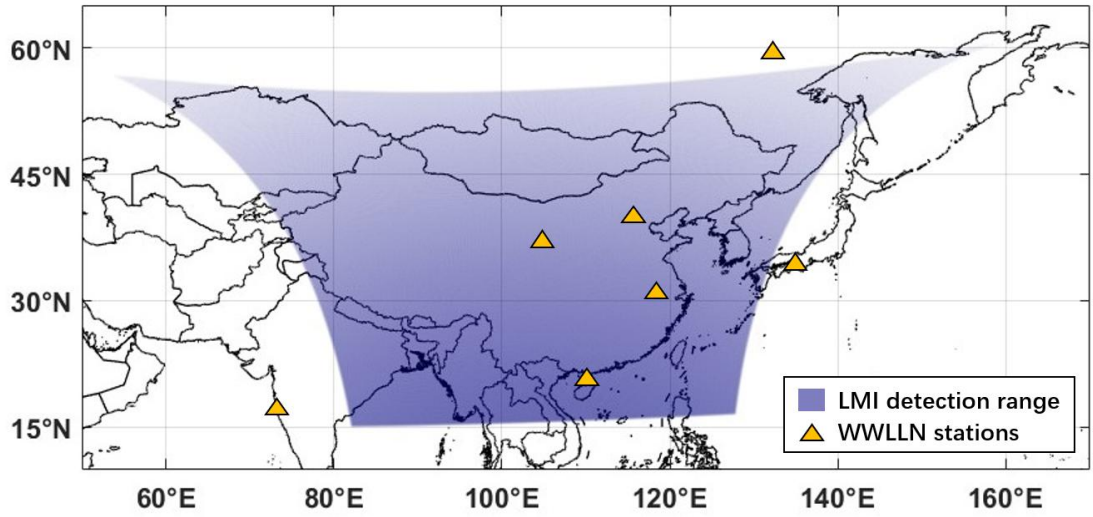

Figure 7. The LMI field of view changes from a rectangle to an inverted trapezoid due to projection.

Minor Comments

Comment 1:

The description of the matching criteria between FY-4A LMI events and the WWLN should be clarified, specifically the choice of spatial and temporal thresholds.

Response:

Thank you for this comment. We will refine the description of our matching criteria in Section 3.1 to clarify the basis for choosing the spatial and temporal thresholds.

Currently, the selection of spatiotemporal matching thresholds between different lightning detection systems mainly relies on sensitivity analysis combined with the marginal effect principle. For example, Thompson et al. (2014) compared WWLLN and ENTLN (Earth Networks Total Lightning Network) with LIS data, considering a WWLLN/ENTLN stroke as a match to an LIS group if the temporal difference was $\leqslant$ 0.4 s and the spatial separation was $\leqslant$ 0.15° . Zhu (2018) used sensitivity tests to determine an optimal matching window of 1 s and 0.5° (≈50 km) between LIS and the ADTD network.

The temporal threshold of 3 s was chosen because a single lightning discharge typically lasts less than 1 second, but LMI and ground-based lightning detection systems observe the same discharge with different temporal sampling and signal propagation delays. As shown in the sensitivity analysis (Fig. 2a), the number of matched events increases rapidly as the temporal window expands from 0 to ~3 s, and beyond 3 s the growth rate levels off. This indicates that a 3 s window captures the vast majority of true coincidences without introducing excessive false matches. We will cite appropriate references in the revised manuscript.

For the spatial threshold, with the temporal window fixed at 3 s, we examined the marginal gain in matched events as the spatial radius increased from 5 km to 50 km (Fig. 2b). The curve flattens beyond 25 km, suggesting that further enlarging the window does not

significantly increase true matches but may increase the risk of false associations. Considering that WWLLN has an intrinsic location uncertainty of about 10 km and LMI pixel size ranges from ~7.8 km at nadir to >20 km near the edge, a spatial threshold of 30 km represents a reasonable compromise between statistical robustness and physical plausibility. Admittedly, the [3 s, 30 km] window appears relatively broad. Because both LMI and WWLLN have lower detection efficiency compared to regional lightning location networks, applying a very narrow spatiotemporal window would result in too few matched events to obtain sufficient data for robust curve fitting. As shown in the radar reflectivity map in our Introduction, WWLLN locations generally fall within the highest reflectivity areas, and the region covered by LMI events also corresponds to high radar reflectivity. Therefore, our matching strategy is to match all WWLLN strokes within the threshold for a single LMI event, aiming to align the "footprints" of the two datasets. Similar approaches using relatively broad spatiotemporal windows for satellite-to-ground lightning matching have been used in previous studies. For example, in evaluating FY-4A lightning detection performance, some studies employed a statistics-based matching method, selecting a relatively large window and then determining optimal thresholds by analyzing the probability distributions of temporal and spatial differences. In the Beijing-Tianjin-Hebei region, a study comparing FY-4A LMI with ground-based lightning data used matching thresholds of 2 s and 50 km. In Zhejiang province, a study on lightning activity under different underlying surfaces adopted the same temporal threshold but found that the optimal spatial threshold varied with surface condition. These studies demonstrate that choosing a relatively broad matching window is an effective strategy under different research objectives and data conditions, allowing the matching results to better reflect the overall distribution characteristics of satellite-based and ground-based lightning data. Based on the above considerations, our chosen [3 s, 30 km] matching window is reasonable and sufficient for the matching requirements of this study.

We will incorporate the above explanations and the corresponding references in the revised manuscript.

References mentioned in this response:

Thompson, K. B., et al. (2014). A comparison of two ground-based lightning detection networks against the Lightning Imaging Sensor (LIS). Journal of Atmospheric and Oceanic Technology, 31(10), 2191–2205. DOI: https://doi.org/10.1175/JTECH-D-13-00186.1

Zhu, J. (2018). Comparison of the satellite-based Lightning Imaging Sensor (LIS) against the ground-based national lightning monitoring network. Progress in Geophysics, 33(2), 541–546. DOI: 10.6038/pg2018AA0631

Comment 2:

Please clearly explain how the 20 × 20 subregional division was selected. Are the results sensitive to grid size?

Response:

As we have detailed in our response to Comment 2, the selection of the 20×20 subdivision is not arbitrary but is directly constrained by the operational 4×4 regionalization of the LMI official algorithm (Hui et al., 2020). To avoid mixing data from different operational subregions—which would cause performance inconsistencies due to different detection parameters and algorithmic behaviors—each of our analysis subregions must be fully contained within a single operational subregion. Therefore, the number of analysis subregions along each dimension must be a multiple of 4.

Within this constraint, we tested multiples of 4: 16×16, 20×20, and 24×24. The 16×16 subdivision resulted in subregions that were too coarse (each covering ~2° in latitude/longitude), smoothing over the systematic deviation patterns we aimed to resolve. The 24×24 subdivision produced subregions that were too small, leading to insufficient sample sizes in many subregions, especially in marginal or low-lightning areas. The 20×20 subdivision struck an optimal balance, providing sufficient spatial resolution while maintaining robust sample sizes (median >500 events per day in active regions). Supporting evidence comes from Zhang et al. (2023), who demonstrated that within a grid of about 2° size, the deviation characteristics are relatively homogeneous.

Consequently, the results are not sensitive to small perturbations of the grid size that remain multiples of 4, as we have validated by comparing 16×16 and 20×20. However, grids that are not multiples of 4 (e.g., 10×10, 30×30) would cut across operational region boundaries and are physically inconsistent, so they were not considered. We will add this justification clearly in Section 3.1 of the revised manuscript.

Comment 3:

The authors should explicitly state whether day–night differences in LMI detection efficiency were considered in the correction process.

Response:

This is an important aspect. First, we acknowledge that LMI detection efficiency exhibits day-night differences: during daytime, solar radiation reflected by cloud tops raises the background optical radiance, which increases the detection threshold, resulting in lower detection efficiency compared to nighttime. The number of detected lightning events directly affects the quality of our fitted correction curves—more events in a given time bin generally lead to more reliable deviation estimates.

In theory, if lightning occurred uniformly throughout the day, we would expect more events at night and fewer during the day. However, based on our experimental data, the actual number of lightning events in each time bin is primarily governed by the occurrence of severe

convective weather, rather than by the day-night cycle. In many subregions, we observed that nighttime event counts are not consistently higher than daytime counts; instead, the counts vary largely with storm activity. Therefore, the influence of day-night detection efficiency differences on our correction is not straightforward.

To address this, we introduced a weighting scheme in our curve-fitting procedure (Equation 2 in the manuscript). The weight for each time bin is defined as a function of both the number of samples and the standard deviation intervals with more events and smaller spread are assigned higher weights, exerting a stronger influence on the fitted curve. This weighting mechanism implicitly accounts for differences in data reliability across different time bins, including those caused by day-night variations in detection efficiency. Bins with very few events (whether due to low detection efficiency or lack of lightning activity) are naturally downweighted.

In summary, our correction process does not explicitly separate or correct for day-night detection efficiency differences. However, the weighting scheme effectively mitigates their impact by emphasizing time intervals with larger sample sizes and more consistent deviation patterns. This approach ensures that the fitted correction curves are primarily driven by periods with sufficient lightning activity, regardless of whether they occur during day or night. We will add a clear statement to this effect in the revised manuscript (Section 3.2 or 4).

Comment 4:

Uncertainty estimates associated with the fitted correction functions are not clearly presented and should be included.

Response:

To quantify the uncertainty of the fitted correction functions, we apply a bootstrap resampling method to estimate the 95% confidence intervals for the longitude and latitude deviation curves of each subregion. Considering the balance between computational efficiency and statistical robustness, the number of bootstrap resamples is set to 500. The time axis covers 24 hours with a 10-minute interval, resulting in 144 time points.

In each resampling iteration, we draw with replacement from the set of valid 10-minute time bins (i.e., those with sufficient lightning events and not flagged as outliers) to create a new sample of the same size as the original valid set. The weight of each time bin (determined by the number of lightning events and the standard deviation; see Eq. 2) is preserved. We then apply the same spline fitting procedure (including the same smoothing parameter and periodic boundary handling) to the resampled data. After 500 iterations, the 2.5th and 97.5th percentiles of the fitted values at each time point are taken as the 95% confidence interval for that time point.

The mean uncertainty (in km) for a subregion is defined as the average half-width of the confidence band over the 24-hour period. For data-sparse subregions (e.g., Xinjiang,

Mongolia) or those where fitting failed, uncertainty is not calculated and such subregions are flagged in the dataset.

In the result presentation, the 95% confidence intervals are shown as shaded bands around the fitted curves (e.g., in Fig. 8), allowing readers to visually assess the diurnal reliability of the correction.

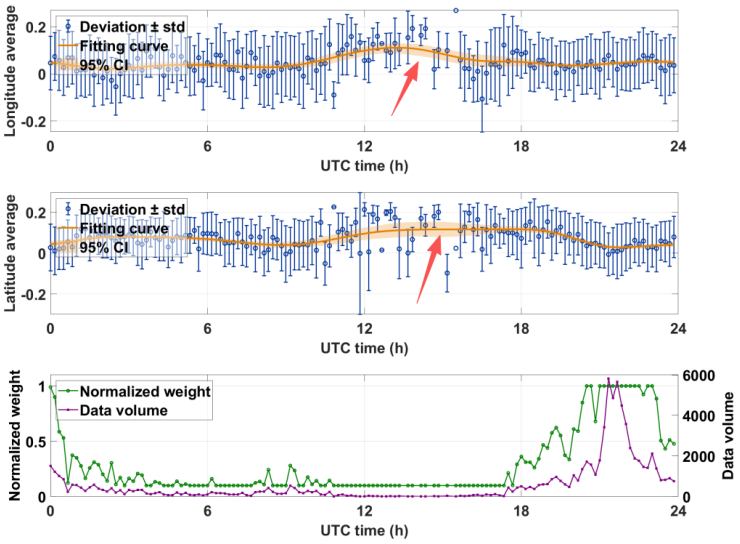

Figure 8. Shaded bands of the 95% confidence intervals for the fitted curves.

Comment 5:

The reported ~15 km residual error should be discussed in terms of its implications for different applications (e.g., storm-scale vs climatological studies).

Response:

Thanks to the reviewer for this suggestion. In the revised manuscript, we will add a dedicated discussion on the implications of the ~15 km residual geolocation error for different scientific applications.

For storm-scale studies (e.g., lightning data assimilation, convective initiation, lightning jump, and storm tracking):

The ~15 km residual error is comparable to the typical size of a convective core (~10–20 km) and approaches the theoretical limit given LMI's native pixel resolution (~7.8 km at nadir, >20 km at the edge) and the ~10 km location uncertainty of WWLLN. As demonstrated in the new case study added in the revised manuscript (e.g., the 4 August 2019 convective system over Beijing), the corrected LMI events are generally collocated with the highest radar reflectivity cores. For applications such as lightning data assimilation into numerical weather prediction models (which typically have grid spacings of 3–15 km for convective-scale models), a 15 km error is acceptable and still provides useful constraints on convection.

For climatological studies (e.g., gridded lightning climatology at 0.25° or coarser, trend analysis, or regional lightning frequency mapping):
The ~15 km error is negligible when aggregated onto coarse grids (e.g., 0.25° corresponds to ~25–28 km in mid-latitudes). Even at finer resolutions (e.g., 0.1° ≈ 10–11 km), the random component of the residual error tends to cancel out when averaging over long time periods (e.g., seasonal or annual scales). Therefore, the corrected dataset is well suited for climate-scale analyses without significant bias.
For applications requiring higher precision (e.g., lightning-induced NOx estimation with point-by-point alignment, or validation of high-resolution satellite products):
Users should be aware that the 15 km residual remains, and further improvements would require higher-quality reference data (e.g., regional lightning location networks with sub-kilometer accuracy) or fusion with other observations. We will clearly state this limitation in the revised manuscript.
We will add this description in the results analysis.

Comment 6:

The manuscript would benefit from a brief comparison with other space-based lightning sensors, such as the GOES-R Geostationary Lightning Mapper or TRMM Lightning Imaging Sensor, to place the results in context.

Response:

Thanks to the reviewer for this suggestion. In the revised manuscript, we will add a brief comparison with other space-based lightning sensors in both the Introduction and Conclusions sections to place our results in a broader context.
The Geostationary Lightning Mapper (GLM) onboard the GOES-R series achieves a nadir geolocation accuracy of approximately 4 km after rigorous navigation and thermal deformation corrections, benefiting from advanced onboard calibration and well-established ground reference data (e.g., Carr et al., 2020). The Lightning Imaging Sensor (LIS) on the TRMM satellite, operating in low Earth orbit, attained a location accuracy of about 3 – 6 km. In contrast, the FY-4A LMI is Chinese first experimental geostationary lightning mapper and initially suffered from larger geolocation errors. After applying our proposed correction, the LMI dataset achieves a substantial improvement: approximately 90% of corrected lightning events have a geolocation error within 15 km, and 70% within 10 km, over most land areas of the Northern Hemisphere. This represents a significant step forward but still lags behind the performance of GLM. The remaining gap is primarily attributable to several factors: (1) the occasional (random) displacement of the detector array; (2) differences in calibration algorithms between the two instruments; and (3) the limited availability of high-quality ground reference data over the full LMI field of view (in contrast to GLM, which benefits from dense ground networks over the Americas). This comparison helps users understand the

relative quality and appropriate application scope of our corrected dataset.
We will add this comparison in the revised Introduction and Conclusions, with appropriate citations.

Comment 7:
The terminology distinguishing event, group, and flash should be defined clearly at the beginning for consistency.
Response:
Thanks to the reviewer for this suggestion. In the revised manuscript, we will add clear definitions of event, group, and flash in Section 2.1 (or in a separate terminology box) to ensure consistency throughout the paper. The definitions are as follows:
Event: A single pixel in a single frame (2 ms) whose radiance exceeds the background threshold.
Group: A cluster of spatially adjacent (<16.5 km) events within the same frame. The location of the group is the centroid of the polygon formed by all qualifying events.
Flash: A cluster of groups within a certain temporal window (<330 ms). The location of the flash is the centroid of the polygon formed by all qualifying groups.