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 (line 84-109). In practice, the physics-based corrections applied to FY-4A LMI are still imperfect for the following reasons:

(1) The 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, leading to a parallax correction error on the order of kilometers. Temporally, the satellite CTH product has a resolution of 15 minutes, which is too coarse for rapidly developing convective cells; time interpolation introduces additional errors. Even with satellite CTH, the residual error after parallax correction should not exceed 10 km. However, our matching results show that only 58.2% of LMI lightning locations have a residual error within 20 km, and these residuals exhibit clear periodic variations with time and geographic location, suggesting that they are caused by thermal deformation or other periodic errors. This 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.

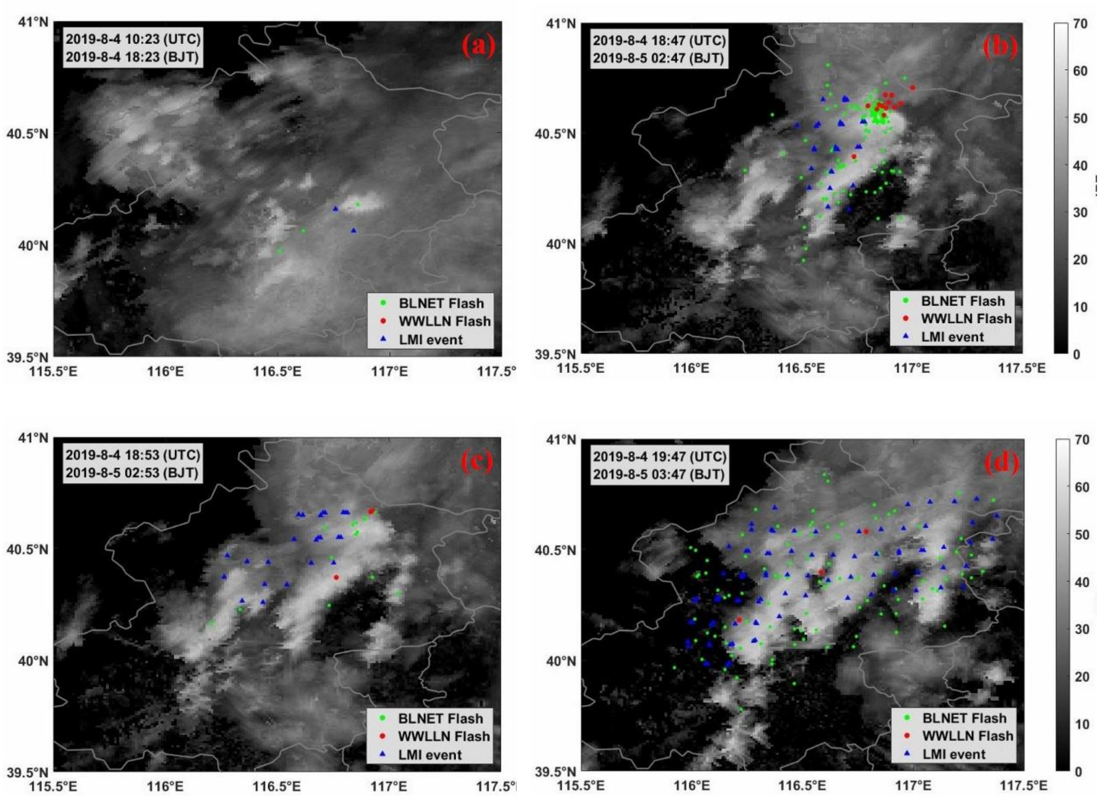

Figure 1: LMI event deviation at different stages of a severe convective event over Beijing on 4 August 2019.

(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.

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).

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.

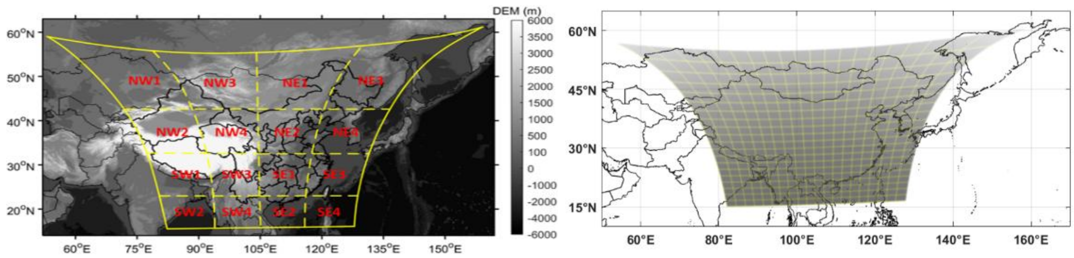

Figure 2. Subregion division in the operational algorithm, from Hui et al. (2020), and the subregion division used in this study.

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 (line 208-225).

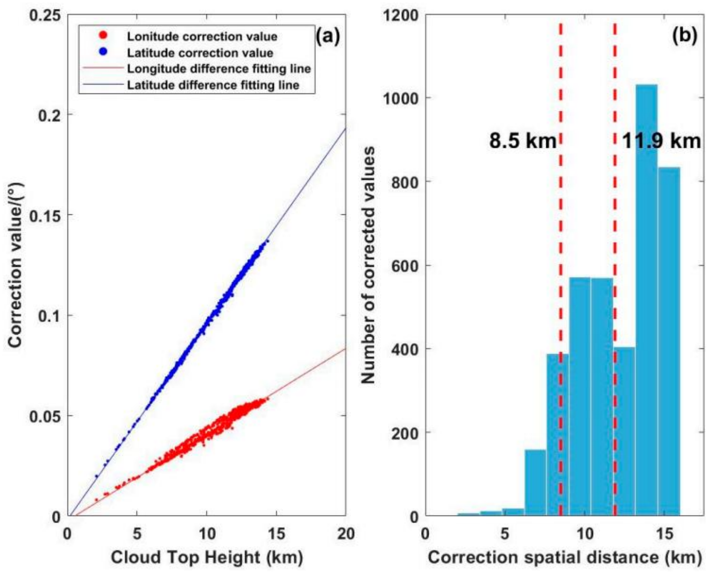

Figure 3 Correction values of latitude–longitude differences and spatial distance of LMI event from Zhang et.al., (2023). https://doi.org/10.3390/rs15194856

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 (line 358-366), 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.

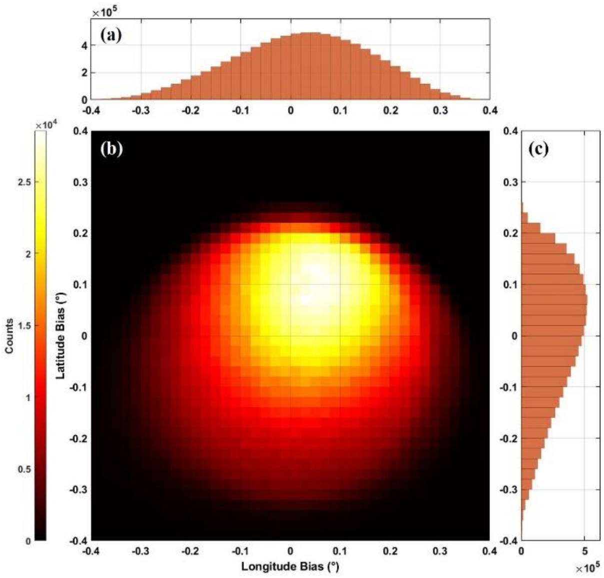

Figure 4: Distributions of coordinate differences between LMI events and WWLLN flashes before correction. The statistics are based on data collected during the boreal warm season (March–September) from 2019 to 2023.

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 4, line 358-373).

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.

This section of the description has been added to the revised manuscript (line 163-169)

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 the above descriptions in the revised manuscript (line 123-128).

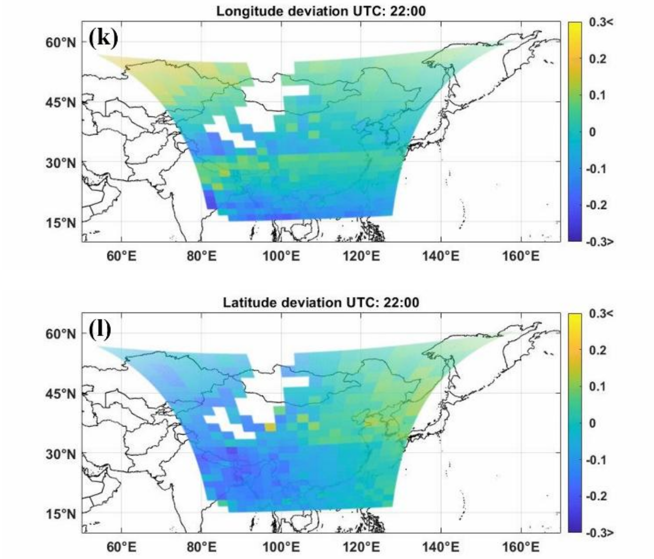

Figure 4: Spatiotemporal distribution of lightning geolocation deviations for LMI events over the Northern Hemisphere observational domain. The left column shows the longitudinal deviations, and the right column shows the latitudinal deviations.

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 (line 410-445).

We selected a severe convective weather system that occurred over Beijing from 4 to 5 August 2019. Figure 5 compares the LMI event locations before and after correction with BLNET flashes, WWLLN flashes, and radar composite reflectivity. The time window 18:00–20:00 UTC is examined because the thermal deformation effect was most significant during this period, consistent with the diurnal variation noted earlier. As shown in Fig. 5, 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. 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.

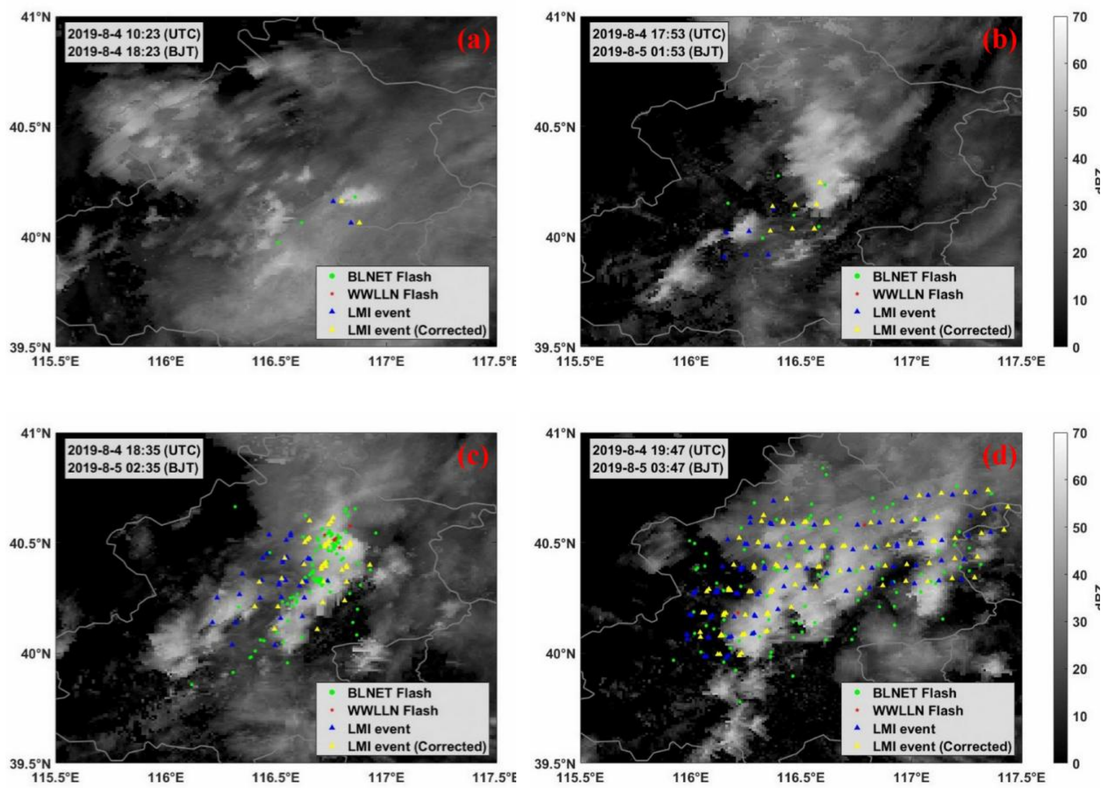

Figure 5: Comparison of LMI correction results before and after with BLNET and WWLLN flash. The red dots represent the matched WWLLN flash data, the green dots indicate the matched BLNET flash data, the yellow dots show the LMI events data after correction, and the blue dots show the LMI events data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

In addition to the visual comparison, we also conducted a quantitative evaluation of the LMI geolocation deviation 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. However, it is worth noting that the relatively broad thresholds (3 s, 30 km) inevitably introduce mismatches from non-corresponding convective cells. As shown in Fig. 5c, the BLNET flashes within the matching window cover almost the entire high-reflectivity convective region, which dilutes the statistical improvement and masks the true correction effect. To better quantify the actual correction performance, we conducted a sensitivity experiment specifically for the BLNET validation thresholds (see Appendix B). Under the stricter thresholds (2 s, 20 km), which effectively removed much of the background noise, the statistical results from the coordinate differences between LMI events and BLNET flashes (Figs. 6a and 6b) show that the distributions of both zonal and meridional deviations become more convergent after correction and exhibit a Gaussian-like distribution centered at zero. Specifically, 70% of the longitude deviations are within ±0.069° with a mean of 0.016°, and 70% of the latitude deviations are within ±0.046° with a mean of −0.016°. These accuracy levels are comparable to those achieved by post-processing corrections using regional high-precision lightning location networks (Zhang et al., 2026), confirming the effectiveness of the proposed method. Overall, the spatial alignment between the corrected LMI events and the BLNET flashes, as well as with radar reflectivity cores, is visibly improved in the spatial plots, confirming the overall effectiveness of the correction. This demonstrates that the WWLLN-based empirical spline-fitting method, when combined with existing physical corrections, effectively reduces systematic geolocation biases in LMI observations, as further validated by independent BLNET and radar data.

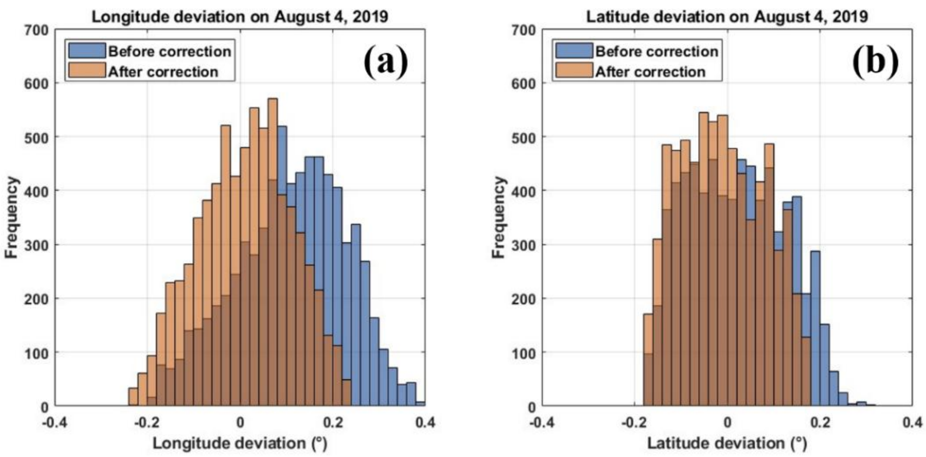

Figure 6: Distributions of coordinate differences between LMI events and BLNET flashes before and after correction.

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. The above descriptions have been added to the revised manuscript (line 463-474).

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. (line 390-403)

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.

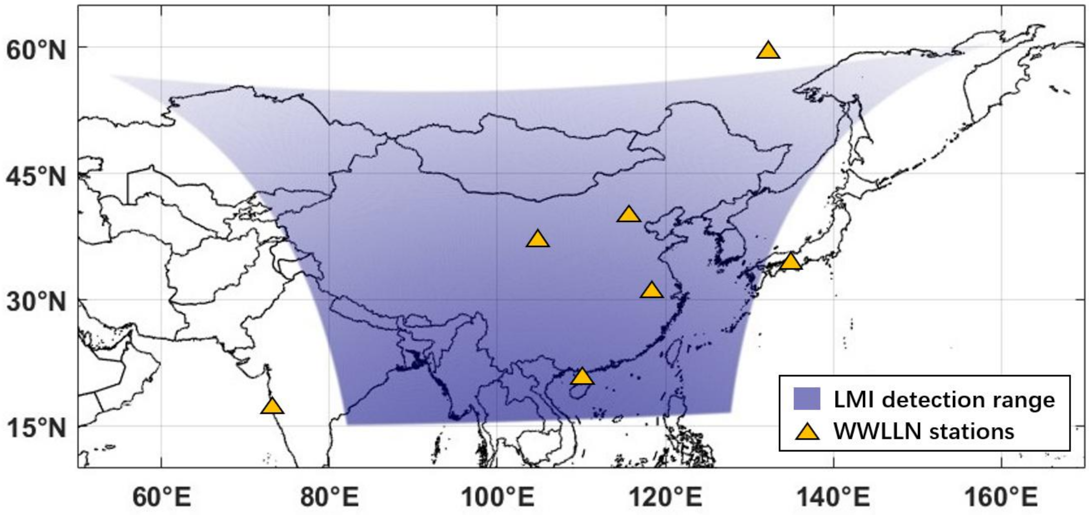

Figure 7. The LMI field of view changes from a rectangle to an inverted trapezoid due to projection.

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 ⩽ 0.4 s and the spatial separation was ⩽ 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. 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 explanations and the corresponding references in the revised manuscript (Section 3.1, line 187-201).

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 (line 208-224).

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.

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 (line 315-326). 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.

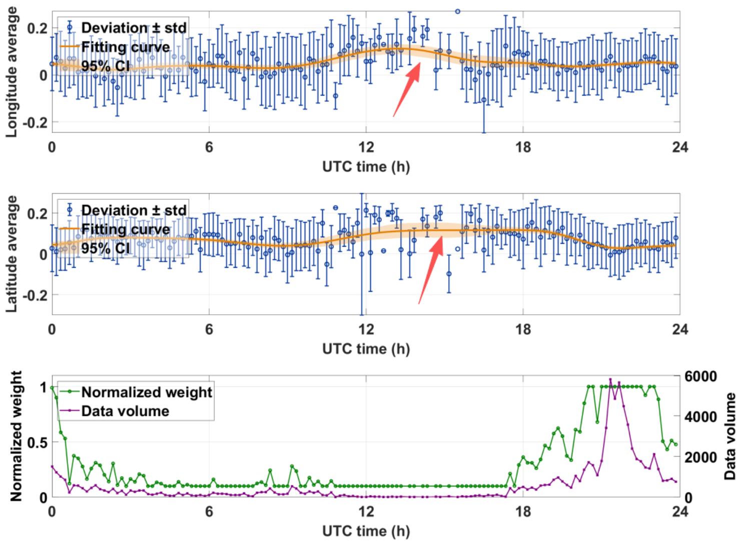

Figure 8. Shaded bands of the 95% confidence intervals for the fitted curves.

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 add this description in the conclusion (line 463-484).

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 have added a brief comparison with GOES-R GLM in the Conclusions section (line 463-474).
The corrected dataset achieves a coarse-estimate mean geolocation error within approximately 15 km (about 1.5 pixels) across most of the LMI field of view, with the actual accuracy being better. To provide a more rigorous assessment, we conducted an independent validation using the high-precision Beijing Broadband Lightning Network (BLNET) for a severe convective case on 4 August 2019. The results show that for 70% of the events, longitude deviations are within 0.069° (mean 0.016°) and latitude deviations are within ±0.046° (mean −0.016°), corresponding to a comprehensive geolocation error within approximately one pixel at the Beijing latitude (~40°N). This independent validation demonstrates the effectiveness of the proposed method and the reliability of the corrected LMI data.
This kilometer-scale accuracy is approaching the geolocation performance of the GOES-R GLM (approximately 4 km at nadir and ~14 km near the edge), despite the remaining gap due to differences in instrument design, calibration maturity, and the availability of high-quality ground reference data. This comparison has been incorporated into the revised Conclusions to place our results in a broader context.

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 (line 134-138) 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.

Reply on RC2

Comment 1:

Line 49: It will be possible to apply the authors' correction method to FY-4C LMI data, but not in real-time since it relies on other lightning data. Is such a post-processed dataset planned? More generally, will lessons learned from FY-4A mitigate the need for such a method due to improved operational algorithms for FY-4C?

Response:

We thank the reviewer for this forward-looking question.

Applicability of the correction method to FY-4C and planned post-processed dataset:

Yes, our correction method can in principle be applied to FY-4C LMI data, because the fundamental sensor design and the geolocation error sources (e.g., CTH parallax, thermal deformation, and platform jitter) are similar between FY-4A and FY-4C, although the latter includes hardware upgrades. However, as the reviewer correctly notes, our method relies on ground-based reference data (WWLLN) and therefore cannot be applied in real time. We are planning to produce a post-processed corrected dataset for FY-4C once a sufficient amount of coincident LMI and WWLLN data becomes available (e.g., after the first boreal summer of FY-4C operation). This dataset would be released in a similar format to the FY-4A product presented in this study.

However, it should be noted that FY-4A is positioned at 104.7°E, whereas FY-4C is positioned at 133°E. This means that the Lightning Mapping Imager (LMI) onboard FY-4C primarily observes ocean areas rather than land. Consequently, for lightning correction over coastal and near-ocean regions, the proposed method (using WWLLN as reference) remains applicable, similar to the approach used for FY-4A. For deep ocean areas where WWLLN has limited station coverage and lower detection reliability, the correction strategy may need to be adjusted. In such areas, alternative reference datasets such as the Lightning Imaging Sensor

(LIS) onboard the Tropical Rainfall Measuring Mission (TRMM) could be considered as a replacement for WWLLN, given that LIS offers high-quality lightning detection over ocean regions. A multi-source data fusion approach that integrates information from different satellite platforms (e.g., LIS, WWLLN) may provide a more comprehensive correction capability for FY-4C LMI observations over the vast ocean. This will be an important direction for our future work.

Lessons learned and potential mitigation for FY-4C operational algorithms:
We fully agree that the experience gained from FY-4A should help improve the operational calibration and navigation algorithms for FY-4C. The systematic biases we identified (e.g., the diurnal thermal deformation signature and the CCD-array-dependent longitudinal offset) provide specific targets for hardware and software improvements. It is likely that FY-4C will benefit from better thermal control, more accurate onboard calibration, and refined geometric correction models, which may reduce the magnitude of residual errors compared to FY-4A. Nevertheless, we believe that certain residual errors (e.g., random platform jitter and inter-detector inconsistencies) are difficult to completely eliminate by operational algorithms alone. Therefore, our empirical post-processing method may still be valuable for FY-4C to achieve the highest possible geolocation accuracy for scientific applications, especially for climate-quality datasets. We will add a brief discussion of this point in the Conclusions or a future outlook section of the revised manuscript (line 475-484).

Comment 2:
Line 75: How are thermal distortions (forced by the diurnal cycle) related to launch-induced payload displacement? Presumably those would occur even in the absence of any launch effects.

Response:
We thank the reviewer for this clarifying question. We would like to distinguish between two different types of displacement:
Launch-induced payload displacement: a permanent, static mechanical offset caused by the large acceleration during rocket launch. On-orbit thermal distortion: a periodic, dynamic deformation caused by diurnal variations in solar radiation flux after the satellite enters its operational orbit. Even in the absence of any launch effects, thermal distortion would still occur. FY-4A is in a geostationary orbit at an altitude of approximately 35,800 km. Due to Earth's shadow, the satellite is only blocked from solar radiation for about 2 hours per day; for the rest of the time it is exposed to sunlight. As the Earth rotates, the angle between the LMI viewing direction and the solar direction changes continuously, causing a periodic variation in the solar radiation flux entering the instrument. This in turn induces a diurnal pattern of thermal deformation on the CCD array. The geometric relationship is schematically illustrated

in Fig. 8.

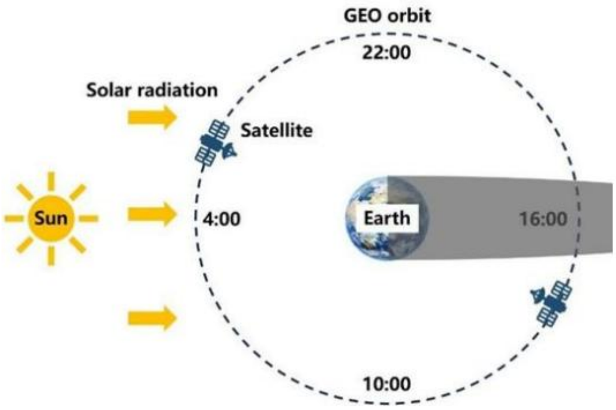

Figure 8 The schematic diagram illustrating the impact of solar radiation on FY-4A LMI thermal deformation. The time is in UTC. Figure from Zhang et.al., (2026). https://doi.org/10.3390/rs18010163

Therefore, in our study we have considered both the static, launch-induced displacement and the periodic residual biases caused by on-orbit thermal deformation, which are the primary focus of our empirical correction. These two error sources are not contradictory; together they constitute the complete picture of LMI geolocation errors.

Comment 3:

Line 129: what were the other “ground-based lightning observations” used by Fan et al. (2018) for comparison with WWLLN?

Response:

Thank you for the question. According to Fan et al. (2018), the “other ground-based lightning observations” used for comparison with WWLLN were primarily the Cloud-to-Ground Lightning Location System (CGLLS) developed by the State Grid Corporation of China, for the period 2013–2015. CGLLS is designed to detect cloud-to-ground strokes, and its data were used to evaluate the detection efficiency and location accuracy of WWLLN over the mid-southern Tibetan Plateau. The quantitative results show that the average spatial separation between matched WWLLN and CGLLS strokes in the mid-southern Tibetan Plateau was 9.97 km, and that between matched WWLLN and LIS flashes over the entire Tibetan Plateau was 10.93 km. Furthermore, the detection efficiency of WWLLN increased markedly with stroke peak current; over the mid-southern Tibetan Plateau, the DE of WWLLN for CG lightning was 9.37% and for total lightning was 2.58%. We will add this description in the revised manuscript (line 152-153).

Comment 4:

Line 131: “of [WWLLN is] around 10 km”

Response:

Thank you for pointing this out. We have corrected Line 131 from "of around 10 km" to "of WWLLN is around 10 km" in the revised manuscript.

Comment 5:

Line 141: please provide a reference for "marginal effect analysis"

Response:

Thank you for pointing this out. In the revised manuscript (Section 3.1), we will add the following statement with appropriate references:

"Currently, the selection of spatiotemporal matching thresholds between different lightning detection systems mainly relies on sensitivity analysis combined with the marginal effect principle (Zhu, 2018; Thompson et al., 2014)."

Comment 6:

Line 179: "the weight assigned to the k-th data point" is defined as omega sub k with an overbar, but the formula also has a tilde above the overbar. In the definition of Eq. 2, there is an additional notation that adds a superscript "raw" to omega. Please clarify how each of these symbols are related to one another, and their possibly varied role in the spline fitting process. For example, is Eq. 2 only a starting estimate for omega?

Response:

We thank the reviewer for the careful reading and for pointing out the ambiguity in the notation. In the revised manuscript, we will clarify the relationship between the different weight symbols as follows:

$\tilde{\varpi}_k$(appearing in Eq. 1, with a tilde above the overbar) is the final weight used in the fidelity term of the spline fitting. It is obtained by normalizing ϖ_k^{raw} to a reasonable range (0.1–1.0) and applying additional constraints, such as downweighting or excluding intervals with extremely small standard deviations, and those manually flagged as outliers.

ϖ_k^{raw}(defined in Eq. 2) is the raw weight computed solely from the number of samples N_k and the standard deviation σ_k. It represents an initial reliability estimate for each 10-minute interval.

We will add the above detailed explanation in the equation section of the revised manuscript (line 270-272).

Comment 7:

Section 3.1: matching of WWLLN strokes to LMI events is described here. I would have expected matching to group centroids. In the case of the a bright optical pulse comprised of many pixels, does the authors' fitting procedure only include those events that are in the 3s and 30 km space/time window? Are the events outside this window simply discarded in the fitting analysis?

Response:

We thank the reviewer for raising this important question. Initially, we also considered using LMI groups for matching, because by definition a group is the centroid of all events within a single frame that are spatially adjacent (<16.5 km), representing the center of a lightning optical footprint. However, through experimental comparisons, we found that using a single centroid (group) to represent lightning location is not always accurate. As shown in Fig. 9, (line 109) when compared with ground-based lightning location network data (e.g., BLNET) and radar composite reflectivity, the full spatial distribution of events actually better represents the lightning optical footprint than the group centroid, and the overall geolocation deviation of LMI lightning is more clearly visible. Furthermore, using groups would introduce additional residual errors associated with the clustering algorithm itself. Therefore, we chose events (the most fundamental detection unit, a single pixel in a single frame) as the matching basis.

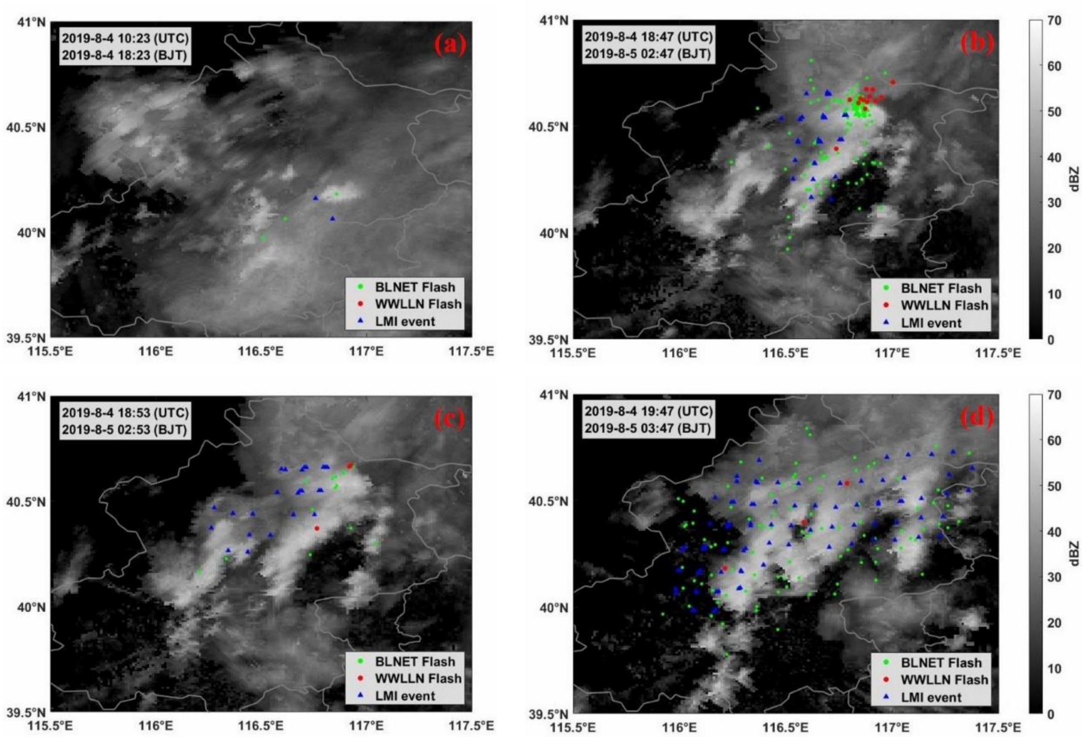

Figure 9: LMI event deviation at different stages of a severe convective event over Beijing on 4 August 2019.

Regarding the matching window: a bright optical pulse composed of many pixels will produce multiple LMI events within a single frame. In our procedure, each LMI event is matched independently to WWLLN strokes within the 3 s temporal window and 30 km spatial window. A single WWLLN stroke may be matched to multiple LMI events if they fall within the thresholds. We do not enforce one-to-one correspondence; instead, our approach aligns the spatial “footprints” of the two datasets.

Events outside the [3 s, 30 km] window are excluded from the fitting analysis. However, based on the sensitivity analysis, the marginal gain beyond these thresholds is very small, and including distant events would increase noise and false associations without significantly improving the curve fitting. As long as a sufficient number of valid matches are retained within the window, the discarded events do not adversely affect the robustness of the fitted correction curves. We will clarify these points in the revised manuscript (line 188-200).

Comment 8:

Please mark the four subregions shown in Fig. 4 on Fig. 3

Response:

We thank the reviewer for this constructive suggestion. In the revised manuscript (line 327) , we will add a reference schematic diagram (as shown in the Fig. 10) that indicates the locations of the four subregions presented on the full subregion division map. This will help readers quickly identify the spatial positions of the selected subregions. We will update the figure caption and the relevant text accordingly.

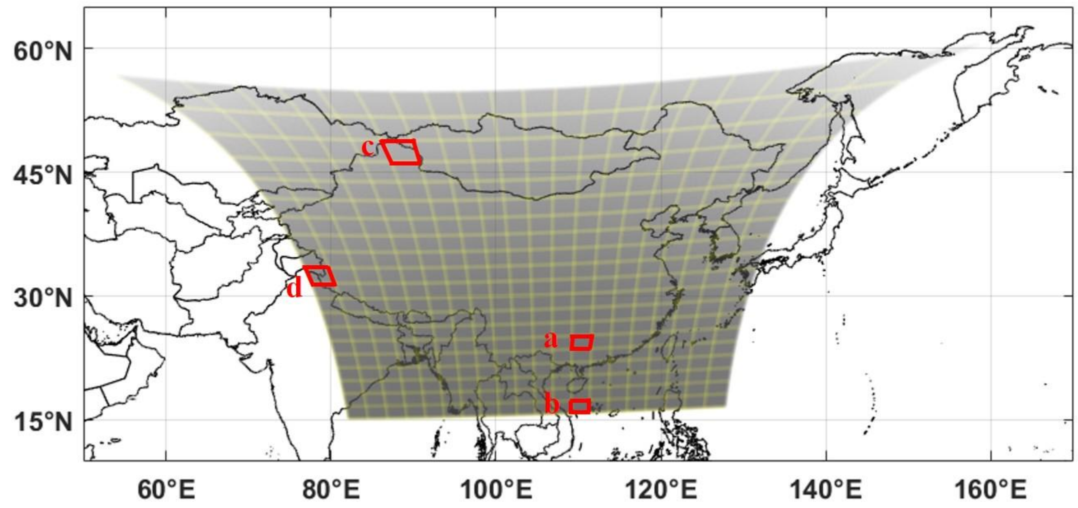

Figure 10 Schematic diagram of the locations of the four study case subregions.

Comment 9:

Line 243: I don't see an east-to-west propagation of the error pattern. To me it looks more like a steady amplification of regions that always have a relative maximum of errors.

Response:

We thank the reviewer for the keen observation. The reviewer's point is very accurate. The wording "east-to-west propagation" in our original manuscript is indeed not precise and may be misleading. In fact, the thermal deformation influence on the LMI CCD array exhibits a pattern of gradual expansion or amplification from the southeast toward the entire domain, rather than horizontal propagation from east to west. Therefore, we will replace "propagation" with “expansion” to more accurately describe the spatial evolution of the error pattern. We will correct this wording in the revised manuscript (line 336-337) and adjust the related text accordingly. We appreciate the reviewer’s valuable comment.

Comment 10:

Line 289: Typically, more samples should lead to improved location, but the conclusion here is that a low number of samples leads to better fits. Why? Does this suggest a shortcoming in some part of the fitting approach?

Response:

We thank the reviewer for this insightful question. The observed phenomenon is indeed interesting and deserves clarification.

In lightning-dense regions (e.g., areas dominated by large mesoscale convective systems), lightning activity is widespread over hundreds of kilometers, and different convective cells may overlap in space and time. Under our relatively broad matching thresholds (3 s, 30 km), such widespread activity inevitably introduces mismatches from non-corresponding convective cells into the matching pool, leading to increased noise and larger standard deviations in the deviation estimates. Consequently, the fitted curves may appear less smooth.

In contrast, in lightning-sparse regions (e.g., Xinjiang, Mongolia), lightning events are typically isolated, associated with weak, localized convection, and spatially concentrated. Under the same broad matching thresholds, the risk of mismatching across different convective cells is much lower. The matched events are therefore highly consistent in space, resulting in smaller standard deviations and visually “better” fits (i.e., smoother curves).

However, this does not mean that the fit is more reliable. The sample size is very small, and the statistical representativeness and extrapolation capability of the fitted curve are actually poor.

Therefore, the statement in our original manuscript that “regions with fewer lightning events yield better fits” needs to be revised, as it is not accurate. A smoother fitted curve in low-sample regions merely indicates local smoothness, not greater overall reliability. True reliability still depends on sample size. We will clarify this distinction in the revised manuscript (line 390-397) to avoid misunderstanding. Meanwhile, for regions with extremely limited sample sizes (e.g., Xinjiang and Mongolia), we ultimately choose not to apply the empirical correction to avoid the risk of insufficient data.

Comment 11:

Line 291: I’m not convinced that the large errors in the NE are due to the large ground footprint of the detector elements. If that were the case, the NW part of the domain should exhibit the same pattern.

Response:

We thank the reviewer for pointing out this logical inconsistency. Our original statement attributing the large errors in the northeastern part solely to the large ground footprint of the detector elements is indeed insufficient.

First, it should be noted that the left half of the LMI CCD array experienced a permanent mechanical offset due to the large acceleration during rocket launch, breaking the original geometric symmetry. Consequently, the northwest and northeast regions are no longer mirror images of each other, and a simple pixel-footprint comparison between the two regions is not valid.

Second, regarding lightning activity characteristics: in the northwest (e.g., Xinjiang), lightning activity is sparse, convection is weak, and lightning occurrences are relatively concentrated, resulting in fewer matched LMI–WWLLN events but lower matching noise. In contrast, the northeast exhibits relatively frequent lightning activity with active convection, and lightning is widely distributed and scattered. Under our broad matching thresholds, this widespread activity tends to introduce more mismatches, leading to higher matching noise. It is important to emphasize that this does not mean the geolocation performance in the northeast is necessarily better than that in the northwest. On the contrary, the larger sample size in the northeast makes its statistical results more stable and reliable, whereas the northwest suffers from high statistical uncertainty due to insufficient samples. Furthermore, the standard deviation of the residuals in the northeast is indeed consistent with the increasing ground footprint of LMI pixels from ~7.8 km at nadir to >20 km toward the edge of the field of view, indicating that pixel projection geometry does contribute to the larger errors.
We will add the above more detailed analysis in the revised manuscript (line 398-403). We appreciate the reviewer's careful reading and constructive criticism.

Comment 12:

Line 300: While I don't object to distribution of a CSV file format, which is easily understood and machine readable (if lacking in terms of metadata), it is not true that NetCDF files will be larger. Naively reading the 20190406.csv dataset (49.2 MB) with pandas and writing it back to NetCDF with xarray was a bit smaller (48.7 MB), even naively using 64 bit ints and floats for all columns. However, the file can easily be made much smaller. Year, month, day, hour, minute and second can all be stored in much fewer than 64 bits, and packing the floats as either 16 or 32 bit floats, or as ints using scale_factor and add_offset, will save further space. Finally, internal zlib compression can also be used on NetCDF files and will be transparently decoded by the library. Finally, the authors might also consider using the CF metadata standard to encode time as a single variable.

Response:

We thank the reviewer for the very professional and detailed suggestions. The following improvements to the dataset:

Time format: The original separate columns for year, month, day, hour, minute, and second have been consolidated into a single ISO 8601 string column (e.g., 2019-03-31 00:00:01.512), with millisecond precision.

Numeric precision: Longitude, latitude, and radiance values are now stored with three decimal places as 32-bit floats. Cloud top height is stored as an integer (16-bit, in meters), which significantly reduces file size.

File formats: Both CSV and NetCDF formats are provided. The NetCDF files use internal zlib compression (level 9) and appropriate data type reduction (e.g., 32-bit integers with scale/offset for time, 16-bit floats or 32-bit floats for geolocation). The resulting NetCDF files are approximately 60% smaller than the original CSV files, while preserving complete metadata and readability. CSV remains available as a lightweight, human-readable option for quick inspection and simple processing.

Metadata: The NetCDF files comply with the Climate and Forecast (CF) metadata conventions, including variable long names, units, and coordinate reference information, ensuring discoverability and interoperability.

We have updated the data availability section and the attribute descriptions in Appendix A (Table A1) accordingly. We greatly appreciate the reviewer's constructive and expert suggestions, which have substantially improved the usability and reusability of the dataset..

Reply on CC1

Specific Comments

Comment 1:

The detection efficiency of WWLLN is spatially non-uniform, and is relatively low in regions with sparse station coverage (e.g., the Tibetan Plateau, Xinjiang, and Mongolia). Although the authors mention in Section 2.3 that the mean location accuracy of WWLLN is approximately 10 km, the following issues are insufficiently discussed: How does the intrinsic location error of WWLLN propagate into the final correction results? Does it significantly affect the geolocation accuracy assessment (approximately 15 km)? The manuscript notes that sparse lightning occurrence in Xinjiang and Mongolia leads to fitting difficulties, but makes no attempt to disentangle the two. Given that this distinction has direct implications for the reliability of the correction in these regions, a brief discussion would be warranted.

Response:

Thanks the reviewer for raising this important issue.

1. Propagation of WWLLN location error into the correction results

The intrinsic location uncertainty of WWLLN (about 10 km) is largely random in nature. Our fitting procedure aggregates many matched LMI–WWLLN event pairs within each subregion and 10-minute time bin, and calculates the mean deviation over these samples. The random errors from WWLLN tend to cancel out during this averaging process. Consequently, the fitted correction curves primarily reflect the systematic biases of LMI rather than the random noise of WWLLN. After correction, 74.8% of events fall within 20 km, and most residual errors are around 15 km, which is comparable to WWLLN's own uncertainty. This indicates that further improvement would require higher-quality reference data (e.g., regional networks). This point will be clarified in the revised manuscript (line 367-373).

2. Distinguishing WWLLN limitations from sparse lightning occurrence

In regions such as Xinjiang and Mongolia, both WWLLN detection efficiency and natural lightning occurrence are low. It is difficult to quantitatively separate the two factors. Instead, we adopt a practical criterion: where the number of matched LMI－WWLLN events is insufficient for robust curve fitting (e.g., below a minimum threshold), we do not apply the empirical correction. In such data-sparse regions, we retain the original physics-based corrections and flag the data accordingly. Thus, regardless of whether the low match count is due to WWLLN inefficiency or true lightning scarcity, the outcome is the same – no empirical correction is applied. A brief discussion of this rationale will be added in the revised manuscript (line 390-397), and users will be advised to exercise caution when using data from these regions.

Comment 2:

A temporal threshold of 3 seconds is selected for matching LMI events with WWLLN flashes. Since individual lightning flashes typically occur within a fraction of a second, a 3-second window is physically quite broad. Please clarify the rationale for this choice. Specifically, is this intended to account for clock synchronization uncertainties between the satellite platform and the ground-based network, or to capture a statistically sufficient number of regional events?

Response:

Thanks the reviewer for this question. The selection of the 3 s temporal window serves both purposes, but the primary driver is to obtain a statistically sufficient number of matched events.

First, clock synchronization uncertainties between the FY-4A satellite and WWLLN stations can introduce timing errors on the order of a second. A 3 s window provides a safe margin to account for such asynchrony.

Second, and more importantly, both LMI and WWLLN have relatively low detection efficiencies. Using a very narrow temporal window (e.g., <1 s) would result in an insufficient number of matched events for robust curve fitting, especially in regions with moderate lightning activity. As shown in our sensitivity analysis (Fig. 2a), the number of matched events increases rapidly up to ~3 s and then plateaus, indicating that a 3 s window captures most true coincidences without adding excessive false matches. This marginal effect principle has been used in previous studies (e.g., Thompson et al., 2014; Zhu, 2018).

This study does not pursue a strict one-to-one match between individual strokes. Instead, the matching strategy aims to align the overall spatial "footprint" of LMI events and WWLLN flashes. The 3 s window is a practical compromise that balances physical plausibility and statistical robustness, especially given the low detection efficiencies of both systems.
This rationale will be clarified in the revised manuscript (Section 3.1, line 187-201).

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: 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 3:

The smoothing parameter λ in Equation (1) is a key parameter controlling the trade-off between data fidelity and curve smoothness (line 175). However, the manuscript does not specify the value of λ or the rationale behind its selection (e.g., cross-validation, the L-curve method). Nor does it discuss the sensitivity of the correction results across different subregions to the choice of λ. It is also unclear whether a single uniform value of λ is applied to all 400 subregions or whether it is determined adaptively for each subregion. The authors should either justify this choice in the text or present a sensitivity test in a supplementary appendix.

Response:

Thanks the reviewer for pointing out this omission. In the experiment, a fixed smoothing parameter λ was used, corresponding to a smoothing parameter of 0.5 in the 'SmoothingSpline' method of MATLAB's fit function. λ ranges from 0 to 1, where 0 gives a completely smooth (over-fitted) curve and 1 forces the curve to pass through all data points (no smoothing). λ = 0.5 provides the best balance between removing high-frequency noise and preserving the diurnal trend of the deviation signal. For consistency and to avoid overfitting, the same λ value was applied uniformly to all 400 subregions

Furthermore, to verify the stability of the fitting and assess sensitivity to λ, we calculated the 95% confidence intervals of the fitted curves for each subregion using bootstrap resampling (500 replicates). The results show that at λ = 0.5, the confidence bands are narrow (typically <2 km in most regions), and the fitted curves are not highly sensitive to λ variations within the range of 0.4–0.6.

The above explanation and sensitivity test results will be incorporated into the equation description in Section 3.2 (line 256-262) and the result discussion in Section 4 (line 315-326) of the revised manuscript.

Comment 4:

The assessment of correction performance (Figures 6 and 7) relies exclusively on WWLLN data, which is the same reference used to construct the correction curves. This is, in effect, a circular validation, as the same reference data are used both to derive the correction and to assess its accuracy. We strongly recommend including at least one independent dataset for cross-validation, such as BLNET or ADTD. If access to such data is restricted, this limitation should be stated plainly, along with a candid discussion of how it affects confidence in the reported results.

Response:

We thank the reviewer for this constructive suggestion. We will add a case study in the revised manuscript (line 410-445). The specific content is as follows:

For the severe convective case over Beijing on 4 August 2019, which was previously used to illustrate the systematic southwestward deviations (see Introduction), Figure 11 compares the LMI event locations before and after correction with BLNET flashes, WWLLN flashes, and radar composite reflectivity. The time window 18:00–20:00 UTC is examined because the thermal deformation effect was most significant during this period, consistent with the diurnal variation noted earlier. As shown in Fig. 11, 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. 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.

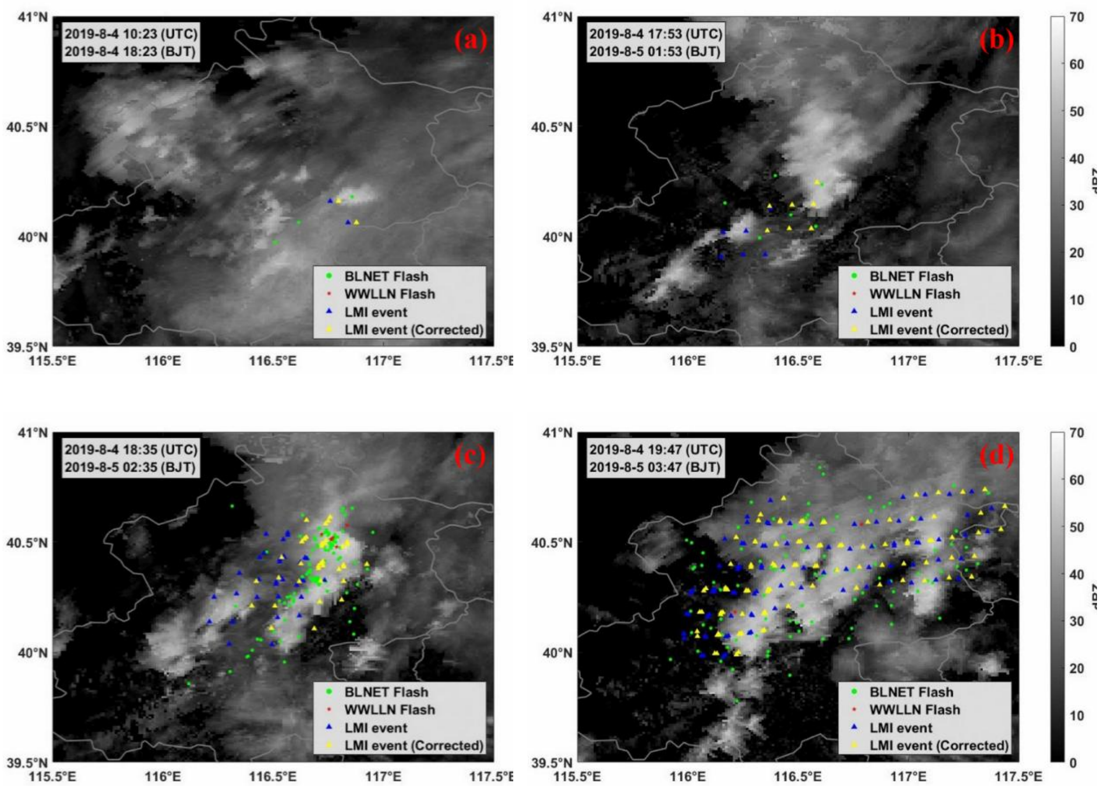

Figure 11: Comparison of LMI correction results before and after with BLNET and WWLLN flash. The red dots represent the matched WWLLN flash data, the green dots indicate the matched BLNET flash data, the yellow dots show the LMI events data after correction, and the blue dots show the LMI events data before correction. The background indicates radar composite reflectivity.

In addition to the visual comparison, we also conducted a quantitative evaluation of the LMI geolocation deviation for this severe convective system using BLNET as an independent reference. Figure 12 shows the distributions of coordinate deviations between LMI events (before and after correction) and the matched BLNET flashes. However, it is worth noting that the relatively broad thresholds (3 s, 30 km) inevitably introduce mismatches from non-corresponding convective cells. As shown in Fig. 11c, the BLNET flashes within the matching window cover almost the entire high-reflectivity convective region, which dilutes the statistical improvement and masks the true correction effect. To better quantify the actual correction performance, we conducted a sensitivity experiment specifically for the BLNET validation thresholds (see Appendix B). Under the stricter thresholds (2 s, 20 km), which effectively removed much of the background noise, the statistical results from the coordinate differences between LMI events and BLNET flashes (Figs. 12a and 12b) show that the distributions of both zonal and meridional deviations become more convergent after correction and exhibit a Gaussian-like distribution centered at zero. Specifically, 70% of the longitude deviations are within ±0.069° with a mean of 0.016°, and 70% of the latitude deviations are within ±0.046° with a mean of −0.016°. These accuracy levels are comparable to those achieved by post-processing corrections using regional high-precision lightning location networks (Zhang et al., 2026), confirming the effectiveness of the proposed method. Overall, the spatial alignment between the corrected LMI events and the BLNET flashes, as well as with radar reflectivity cores, is visibly improved in the spatial plots, confirming the overall effectiveness of the correction. This demonstrates that the WWLLN-based empirical spline-fitting method, when combined with existing physical corrections, effectively reduces systematic geolocation biases in LMI observations, as further validated by independent BLNET and radar data.

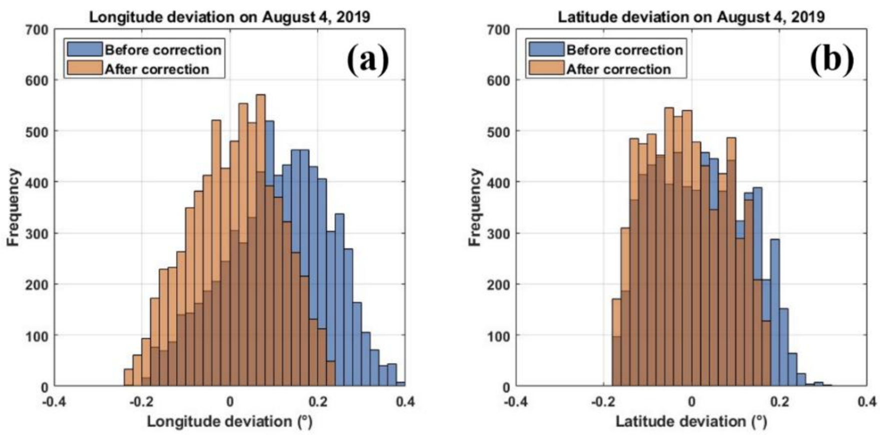

Figure 12: Distributions of coordinate differences between LMI events and BLNET flashes before and after correction.

Comment 5:

The manuscript mentions that additional constraints are applied to downweight time intervals with extremely small standard deviations (lines 191–194) in order to prevent unrealistically large weights. However, neither the specific form of these constraints nor the threshold values are provided. This information is essential for reproducibility and should be included in the main text or in an appendix.

Response:

Thanks the reviewer for raising this reproducibility concern. The specific constraints and thresholds applied to downweight time intervals with extremely small standard deviations are detailed below.

1. Automatic screening (standard deviation threshold)

For each 10-minute interval, if the standard deviation σ_k<0.02 or σ_k is NaN, the raw weight ϖ_k^{raw} is set to NaN (i.e., excluded from the fitting). The threshold of 0.02° is chosen because, given the intrinsic uncertainties of LMI and WWLLN, a standard deviation smaller than this is physically implausible and typically results from a very small sample size (e.g., only one matched event).

2. Manual inspection (outlier identification)

After initial fitting for all subregions, the time series of deviations for each subregion are visually inspected. As shown in the example (subregion [1,16]), an obvious outlier appears around UTC 5:00. Statistics reveal that only one matched LMI - WWLLN event exists in this subregion for that specific time bin over the five-year period (2019–2023), leading to an extremely small standard deviation. If left untreated, this outlier would receive a disproportionately large weight and severely distort the fitted curve. Therefore, we manually identify all such obvious outliers across all subregions and set their weights to NaN. All manually flagged intervals are documented in the code annotations (see Appendix).

3. Final weight

After the above automatic and manual steps, the raw weights of the remaining valid intervals are normalized to the range [0.1, 1.0] to obtain the final weight $\tilde{\varpi}_k$.

These details will be added to Section 3.2 (line256-261, 270-272), of the revised manuscript.

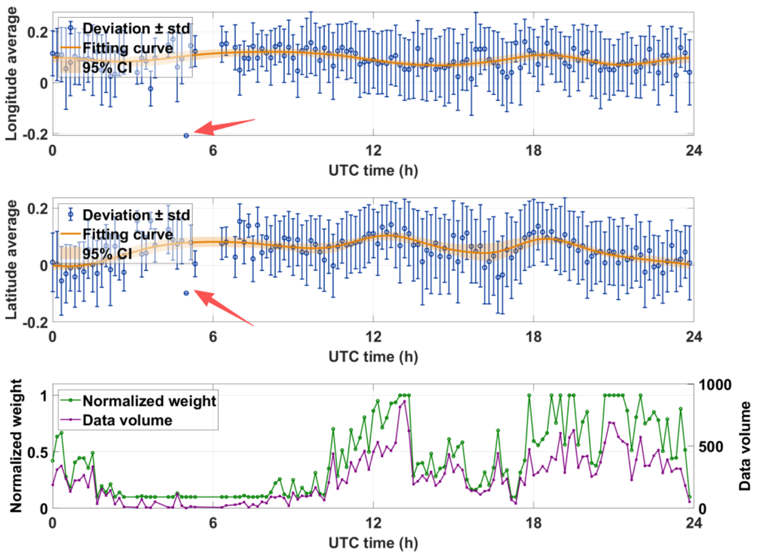

Figure 13 An outlier case in the fitting of subregion [1, 16].

Comment 6:

The color scale in Figure 5 is set to [−0.3, 0.3], with values exceeding the range annotated as “0.3<” and “−0.3>”. However, no information is provided regarding the proportion of pixels falling outside this range or their geographic distribution. It would be helpful to report this in the caption or the main text.

Response:

We thank the reviewer for pointing out the insufficient description of the color scale in Fig. 5. In the revised manuscript (line332-335), we will add the following clarification:

“It should be noted that the color scale in Fig. 7 is capped at ±0.3°; values exceeding this range (mainly occurring in a few northwestern subregions due to large observation angles and the launch-induced meridional deviation of the left-half CCD array) are displayed as “0.3<” or “−0.3>” to avoid stretching the color scale and obscuring the spatiotemporal variations in the rest of the domain.”

Technical Corrections

Comment 1:

Line 56: The citation “(Peterson and Mach et al. (2022)” contains a formatting error and should be corrected.

Response:

The citation in line 56 has been corrected to the proper format, for example ‘Peterson et al. (2022)’. All references have been reviewed and similar formatting issues have been corrected throughout the manuscript.

Comment 2:

Line 120: The section number “2.3” should be “2.2”.

Response:

The section number has been corrected from 2.3 to 2.2.

Comment 3:

There are two typographical issues worth noting: a spurious colon at the end of line 200, and a Chinese period ("。") at the end of line 235. Both should be removed. Please also verify whether the captions of Figures 1, 3, and 7 require a terminal period “.” in accordance with journal style.

Response:

Thanks to the reviewer for the careful check. The spurious colon at line 200 and the Chinese period at line 235 have been removed. Periods have been added to the captions of Figures 1, 3, and 7, and other similar issues have been corrected throughout the manuscript.”

Comment 4:

The abstract runs longer than necessary. Tightening the abstract to focus on three elements — the principle of the correction method, the quantified improvement in accuracy, and the key value of the dataset — would improve its impact.

Response (English):

We thank the reviewer for the very pertinent comment. We have shortened the abstract and concentrated it on three main aspects: the scientific problem and methodology, the quantitative improvement in model accuracy, and the value of the dataset. The revised abstract (line 9-24) is as follows:

The Lightning Mapping Imager (LMI) aboard FY-4A has accumulated substantial observational data. To address remaining systematic geolocation deviations, we propose a correction method using World Wide Lightning Location Network (WWLLN) as a reference. The LMI field of view is divided into 400 subregions (20 × 20 grid). Within each subregion, sensitivity experiments match LMI events with ground-based lightning to quantify systematic deviations, followed by a weighted curve-fitting approach to derive subregional correction curves. The fitted curves are applied to the original Level-2 products to construct a refined correction dataset. Based on over nine million matched LMI-WWLLN pairs (2019–2023), after correction the proportion of events with deviation ≤15 km increases from 36.0% to 51.7%, and ≤20 km increases from 58.2% to 74.8%. The coordinate deviations converge substantially, indicating improved geolocation performance. Across most of the LMI field of view, the coarse-estimate average error is within approximately 15 km (about 1.5 pixels), with the actual accuracy being better, excluding those undetermined regions with sparse lightning (e.g., Xinjiang, Mongolia) where robust fitting is not feasible. Independent validation using the high-precision Beijing Broadband Lightning Network (BLNET) for a severe convective case on 4 August 2019 shows that for 70% of the events, longitude deviations are within ±0.069° (mean 0.016°) and latitude deviations are within ±0.046° (mean −0.016°), with the comprehensive geolocation error within approximately one pixel, achieving accuracy levels comparable to those obtained using regional high-precision lightning network post-processing corrections. The corrected dataset is publicly available at https://doi.org/10.11888/Atmos.tpdc.303312 (Zhang et al., 2026).

Comment 5:

The Conclusion (Section 6) repeats much of what is already in the abstract. Rather than recapping the methodology, the authors could use this space to look ahead. For instance, a brief discussion of how this dataset might support AI-based lightning analysis or serve as a benchmark for numerical model evaluation would connect naturally with the remarks in the

Introduction (lines 97–98).

Response:

The reviewer's comment is very pertinent. We agree and will rewrite the Conclusions section to avoid repetition with the abstract. Instead of recapping the methodology, here is the revised conclusion (line 456-484):

Spaceborne lightning observations are inevitably affected by solar radiation, satellite platform jitter, and systematic offsets of the detector array, which together introduce periodic and random geolocation deviations. To address these issues for the FY-4A LMI, this study develops an empirical correction framework based on a robust spline-fitting model using WWLLN as a reference for lightning-dense regions. The method quantifies the spatiotemporal patterns of residual deviations and applies fitted correction curves to the original Level-2 products, resulting in a publicly available corrected dataset (https://doi.org/10.11888/Atmos.tpdc.303312). Independent validation using the high-precision BLNET over Beijing confirms the effectiveness of the correction.

The coarse-estimate mean geolocation error of the corrected dataset is within 15 km (about 1.5 pixels) across most of the LMI field of view, with the actual accuracy being better. Independent validation using BLNET over Beijing shows that for 70% of the events, longitude deviations are within 0.069° and latitude deviations are within ±0.046°, corresponding to a comprehensive geolocation error within approximately one pixel, approaching the accuracy level of the GOES-R GLM. This residual error has different implications depending on the application. For storm-scale studies, such as lightning data assimilation, convective initiation, lightning jump, and storm tracking, the kilometer-scale error is comparable to the typical size of a convective core (~10–20 km) and is acceptable for convective-scale numerical models (typical grid spacings of 3–15 km), as demonstrated by the improved spatial alignment with radar reflectivity cores in the Beijing case study. For climatological studies, such as gridded lightning climatology at 0.25° or coarser, trend analysis, or regional lightning frequency mapping, the residual error is negligible, as random components tend to cancel out over long-term averaging. For applications requiring higher precision, for example lightning-induced NOx estimation with point-by-point alignment or validation of high-resolution satellite products, users should be aware that the kilometer-scale residual remains.

The corrected dataset opens new avenues for advancing lightning research and its applications. It can serve as a benchmark for training artificial intelligence models in lightning nowcasting, for example using deep learning to predict convective initiation or lightning jumps, and for identifying lightning-structure relationships from satellite data. The dataset also provides an independent observational constraint for evaluating numerical weather and climate models, particularly the performance of convective parameterizations and lightning parameterization schemes in models such as WRF and GRAPES. Furthermore, it enables long-term climatological analyses of lightning activity over the Northern Hemisphere with improved spatial consistency. Looking forward, the methodology can be extended to FY-4C LMI data, though adjustments will be needed for its different orbital position (133°E) and predominantly oceanic view. Future work will also explore multi-source data fusion, for example combining WWLLN with satellite-borne lightning imagers such as LIS, to further enhance geolocation accuracy, especially over oceans.

Comment 6:

There are multiple instances where a space is missing between a word and a citation bracket (e.g., Lines 34, 37, 43, 48, and 59).

Response:

We thank the reviewer for pointing out these issues. We have corrected the missing spaces at the indicated locations and have carefully checked the entire manuscript. All necessary spaces between words and citation brackets have been inserted where needed. A thorough proofreading has been performed before resubmission.