https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Open Access Earth System Science Data Discussions

Corrected event dataset of FY-4A LMI, 2019-2023

Yuansheng Zhang[1,2], Xiushu Qie[1,2], Rubin Jiang[1], Dongjie Cao[3], Jing Yang[1], Dongfang Wang[1], Mingyuan Liu[1], Dongxia Liu[1], Zhuling Sun[1], Hongbo Zhang[1]and Shanfeng Yuan[1]

[1] State Key Laboratory of Atmospheric Environment and Extreme Meteorology, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[2] College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[3] National Satellite Meteorological Center, Chinese Meteorological Administration, Beijing 100081, China

Correspondence to: Xiushu Qie (qiex@mail.iap.ac.cn) and Rubin Jiang (jiangrubin@mail.iap.ac.cn)

Abstract. The Lightning Mapping Imager (LMI) aboard the Fengyun-4A (FY-4A) satellite, once one of the only two geostationary lightning detection payloads operating in space, has accumulated a substantial volume of observational data. Extensive efforts have been made to correct lightning geolocation deviations, including payload misalignment correction, cloud-top height parallax correction, and thermal deformation correction. These measures have substantially improved the geolocation accuracy of LMI observations. However, individual correction schemes are not necessarily applicable across the entire LMI field of view; furthermore, a comprehensive, unified correction dataset has yet to be established, which has limited the wider utilization of LMI data. To address the remaining systematic geolocation deviations in current LMI lightning products, we propose a new correction method based on reference data from the World Wide Lightning Location Network (WWLLN). Using ground-based lightning observations as a benchmark, the LMI field of view is subdivided into 400 subregions arranged in a 20 × 20 grid. Within each subregion, sensitivity experiments are conducted to match spaceborne LMI detections with ground-based lightning events, thereby quantifying the systematic deviation for each subregion. A weighted curve-fitting approach is then applied to the coordinate deviations derived from the matched events to obtain a correction curve for each subregion. These subregional correction curves are subsequently mapped back to the image coordinates, enabling an analysis of the spatiotemporal variability of lightning geolocation deviations across the full LMI coverage. Finally, the fitted curve values are applied as correction terms to the original data, resulting in the construction of a new, refined correction dataset. Building upon the existing Level-2 lightning products, this method significantly enhances the geolocation accuracy of LMI observations. The coordinate deviations between LMI detections and ground-based lightning network observations exhibit pronounced convergence in both the zonal and meridional components, indicating a substantial improvement in geolocation performance. In addition, a domain-wide assessment of geolocation accuracy reveals that, except for regions such as Xinjiang and Mongolia where lightning occurrence is too sparse to support robust curve fitting, the geolocation accuracy across most of the LMI field of view is relatively stable, with an average error of approximately 15 km (about 1.5 pixels), achieving high practical accuracy. The corrected dataset is publicly available at https://doi.org/10.11888/Atmos.tpdc.303312 (Zhang et al., 2026).

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Open Access Earth System Science Data Discussions

1 Introduction

Lightning is a hazardous atmospheric discharge phenomenon associated with severe convective weather and represents a major contributor to casualties from meteorological disasters(Qie et al., 2023). As a long-distance discharge phenomenon in nature, lightning serves as the largest natural source of nitrogen oxides (NOx) in the troposphere, which not only regulates the atmospheric oxidation capacity in the upper troposphere but also profoundly shapes chemical cycling processes within the troposphere(Brune et al., 2021). Lightning also triggers wildfires, which in turn influence the evolutionary trajectories of species and ecosystems. Accurate lightning detection is therefore of paramount importance for weather monitoring and disaster mitigation. With advances in space-based lightning imaging technology, large-scale lightning observation systems have developed rapidly over the past decade. So far, three series of lightning imagers based on geostationary satellites have been operating in space, providing a probable global distribution of lightning activity. The Geostationary Lightning Mapper (GLM) onboard the Geostationary Operational Environmental Satellite (GOES) R-series—GOES-16, -17, -18 and -19—was launched in 2017, 2018, 2023 and 2024, respectively(Goodman et al., 2013; Loto Aniu et al., 2023; Peterson et al., 2022a; Rudlosky et al., 2019). This establishes a multi-satellite collaborative observation system. In parallel, China's next-generation geostationary meteorological satellite, Fengyun-4A (FY-4A), carrying the Lightning Mapping Imager (LMI), was launched on 11 December 2016. During boreal summer, FY-4A primarily monitored lightning activity over China and surrounding regions in the Northern Hemisphere, whereas in boreal winter it focused on the eastern Indian Ocean and western Australia in the Southern Hemisphere(Cao, 2016; Yang et al., 2017). After several years of continuous operation, FY-4A was decommissioned in early 2024. Subsequently, on 27 December 2025, the next-generation geostationary satellite FY-4C was successfully launched, carrying an upgraded version of the LMI. In addition, the Lightning Imager (LI) onboard the Meteosat Third Generation (MTG) satellite was launched on 13 December 2022, providing coverage of Europe, Africa and the surrounding seas (Holmlund et al., 2021). This further expanded the global capability for continuous geostationary lightning observations.

For the Geostationary Lightning Mapper (GLM), Carr et al. (2020) addressed geolocation errors by using coastline identification and registration as a spatial reference. They further incorporated data from distributed temperature sensors into compensation algorithms to mitigate deviations caused by thermal deformation. Consequently, the GLM achieved a lightning location accuracy of approximately 4 km at the nadir. Building on this work, (Peterson and Mach et al. (2022)conducted a comprehensive evaluation of GLM lightning detection performance, including the characteristics of lightning optical sources, detection range and height, detection thresholds, and clustering algorithms. In addition, GLM implemented a cloud-top height (CTH) parallax correction model to mitigate parallax-related geolocation deviations(Buechler et al., 2018). Overall, GLM has demonstrated stable performance with relatively high lightning geolocation accuracy. However, as the CTH parameter was derived through comparisons between satellite observations and ground-based lightning location networks, the correction accuracy was influenced by the latitude of lightning occurrences to a certain extent.

As an experimental payload, the LMI onboard FY-4A has been the subject of extensive validation efforts. Hui et al. (2020) examined the preliminary observational data and radiative characteristics of LMI by comparing lightning optical signals

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

detected by LMI with those from other lightning observation systems, such as the Lightning Imaging Sensor (LIS) and the World Wide Lightning Location Network (WWLLN). The consistency among these datasets provided initial evidence of the reliability of LMI observations. Nevertheless, Cao et al. (2021)compared LMI and LIS lightning observations from 2018 to 2020 and found that the nighttime detection efficiency of LMI was significantly higher than its daytime counterpart, indicating that the majority of LMI lightning detections occur at night. Chen et al. (2021)compared LMI observations with high-precision data from the Beijing Broadband Lightning Network (BLNET) and found that LMI preferentially detects lightning in shallow clouds, whereas lightning within deep convective clouds is more difficult to capture, indicating spatially and physically uneven detection efficiency. To address these limitations, a series of studies focusing on data error correction have been conducted. Drawing on the navigation registration strategy used for GLM, Cheng et al. (2021) and Wang et al. (2021) proposed a double-edge algorithm that employs coastlines as reference features for image registration during daytime, while nighttime geolocation deviations were estimated from thermal plate measurements to infer payload displacement induced by launch acceleration. This approach partially reduced thermal deformation–related geolocation deviations in LMI, achieving a positioning accuracy of approximately one pixel during daytime and within three pixels at night. Zhang et al. (2023) proposed an ellipsoidal cloud-top height correction model that utilizes post-processed CTH data, effectively mitigating parallax-induced geolocation deviations in LMI lightning observations and further improving detection accuracy. Building on this work, Zhang et al. (2026) demonstrated the feasibility of using regional ground-based lightning location data, following CTH parallax correction, to correct thermal deformation–induced geolocation deviations of satellite payloads. This approach further enhanced the lightning geolocation accuracy of LMI.

Although substantial efforts have been devoted to the correction and validation of lightning observations from the LMI, resulting in notable progress, several limitations persist in existing correction approaches. First, most previous studies have focused on localized regions or specific time periods. Although the proposed correction strategies perform well under these conditions, they have not been systematically evaluated across the full observational domain, limiting confidence in their applicability. Second, several issues remain unresolved, including systematic geolocation deviations arising from the use of different post-processing strategies across subregions, deviations induced by satellite platform jitter, and the identification and removal of interference signals such as intense solar flares. Currently, the correction efforts are largely independent and fragmented, lacking effective integration and coordination. Consequently, a comprehensive, consistently formatted, and fully corrected lightning dataset has yet to be established, which substantially constrains the broader application of lightning observations in meteorological forecasting, hazard monitoring, and climate analysis. In view of this, the present study builds upon previous work by synthesizing and integrating existing correction methodologies and explicitly addressing the multiple sources of geolocation deviation that persist in current LMI lightning products. We perform a systematic, full–field-of-view correction of LMI lightning observations and construct a complete, coherent, and reliable corrected lightning dataset. This dataset provides a robust foundation for accurate lightning monitoring and advances research in lightning meteorology and climatology, while also offering higher-quality observational constraints for data-driven applications such as artificial intelligence–based analysis and the validation and evaluation of numerical weather and climate models.

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

2 Data

The LMI onboard FY-4A was the first spaceborne lightning detector capable of observing both cloud-to-ground and intra-cloud lightning in China. Correcting LMI observations across the full field of view requires ground-based lightning location data with broad spatial coverage and long-term temporal continuity. Data from the WWLLN provide an appropriate reference for this purpose, as demonstrated by Zhang et al. (2026). Nevertheless, ground-based validation is less effective over oceans due to the lower natural occurrence of lightning and the reduced sensitivity of detection resulting from the sparse station networks. Consequently, this study focuses on correcting LMI observations over the Northern Hemisphere and does not consider data over the Indian Ocean in the Southern Hemisphere.

2.1 FY-4A LMI event

The LMI onboard FY-4A employed a 400 × 300 × 2 CCD focal plane array, operated at a wavelength of 777.4 nm, and sampled at a frame rate of 2 ms (Cao, 2016; Yang et al., 2017). The LMI field of view covered China and its adjacent seas (dark-blue region in Fig. 1), with a nadir spatial resolution of 7.8 km. LMI used a real-time event processor (RTEP) to dynamically estimate the mean background optical radiance, which was subsequently used as a threshold for background discrimination. Pixels within each frame whose radiance exceeded this threshold were extracted and defined as event data. Spatially adjacent events occurring simultaneously within the same frame were grouped to form group data, which were further processed using a dedicated clustering algorithm to classify and aggregate detections into flash data products. In this study, we employed the fundamental LMI event-level products provided by the National Satellite Meteorological Center of the China Meteorological Administration and applied a cloud-top height (CTH) parallax correction as a preprocessing step, following the model proposed by Zhang et al. (2023).

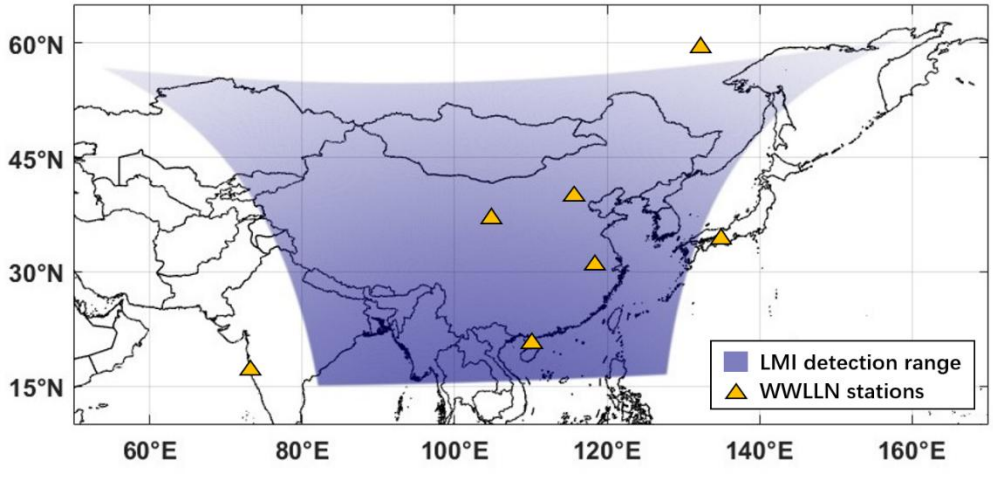


Figure 1: LMI observational domain in the Northern Hemisphere and the distribution of WWLLN stations

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

2.3 WWLLN flash

The ground-based lightning observations from the WWLLN is used as a reference to investigate whether systematic geolocation deviations associated with thermal deformation are present. This provides a basis for assessing the potential of WWLLN data to correct thermal deformation–induced geolocation deviations in LMI observations across the full field of view. WWLLN detects global lightning activity in near real time by capturing electromagnetic radiation in the very low frequency (VLF) band (3–30 kHz) and using GPS timing to determine the precise arrival times of lightning pulses at individual stations. The network comprises more than 70 stations worldwide, five of which are located within or near the LMI observational domain in East Asia (orange triangles in Fig. 1). WWLLN primarily detects high–peak-current lightning events, including both intra-cloud and cloud-to-ground discharges, with detection efficiency increasing for stronger return strokes(Abarca et al., 2010; Rodger et al., 2006). Fan et al. (2018) compared WWLLN data with ground-based lightning observations and found that, during 2013–2015, approximately 72% of WWLLN-detected lightning events over the central and southern Tibetan Plateau were intra-cloud discharges. The mean location accuracy of around 10 km. In this study, the WWLLN dataset spans the boreal summer of 2019. The WWLLN data used here are provided at the flash level. A flash is defined as a cluster of lightning discharges occurring within 0.5 s and separated by no more than 30 km.

3 Methods

3.1 Matching between spaceborne and ground-based lightning observations

Before quantifying the geolocation deviations of LMI lightning observations, it is necessary to match spaceborne lightning detections with ground-based lightning events. To determine appropriate matching thresholds, we conducted a series of sensitivity experiments, as shown in Fig. 2. As illustrated in Fig. 2a, regardless of the spatial window applied, the rate of increase in the number of matched lightning events becomes stable once the temporal threshold exceeds 3 s. We therefore adopt 3 s as a robust temporal matching threshold. Under this condition, the spatial matching window was further refined through, a marginal effect analysis, as shown in Fig. 2b. The results indicate that spatial thresholds of 25 km and 30 km yield optimal marginal effects. To ensure a sufficiently large number of matched lightning events, a 30-km spatial threshold was ultimately adopted for the matching procedure.

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

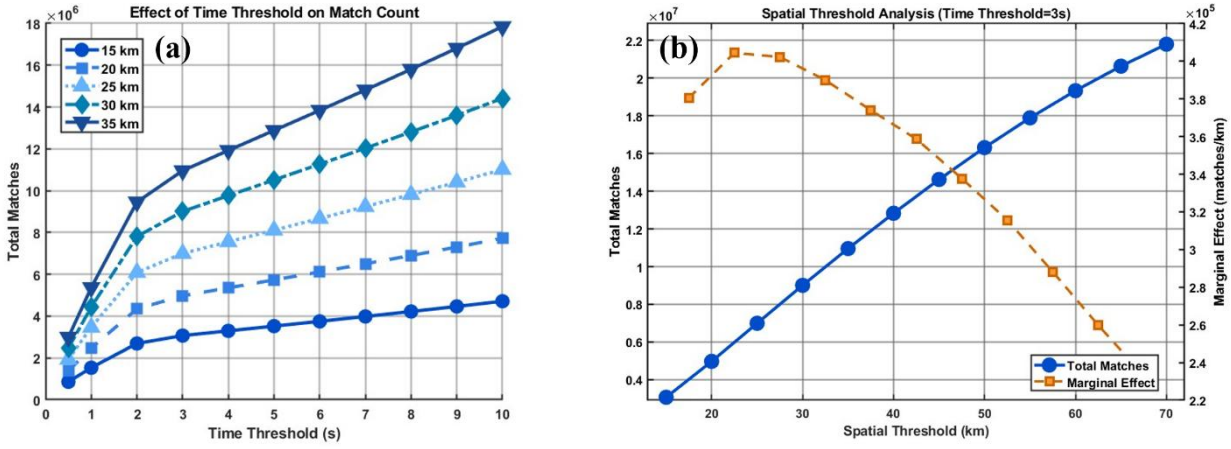


Figure 2: Sensitivity experiments on spatiotemporal matching thresholds for matching LMI events with WWLLN flashes. (a) Statistics of the number of matched events under different spatiotemporal threshold settings; (b) marginal effect analysis under a fixed temporal threshold of 3 s.

As the number of detected lightning events is finite, performing geolocation deviation analysis at the level of individual detection units is impractical. Instead, detection units within a certain spatial extent must be treated collectively and analyzed as an ensemble. This requires a regionalized analysis of the LMI observational domain. Notably, the operational LMI processing algorithm itself adopts a regionalization strategy(Hui et al., 2020), which is therefore taken into account in the design of our partitioning scheme. Considering both the availability of lightning observations and computational efficiency, we divide the LMI field of view into 400 subregions arranged in a 20 × 20 grid for subsequent analysis.

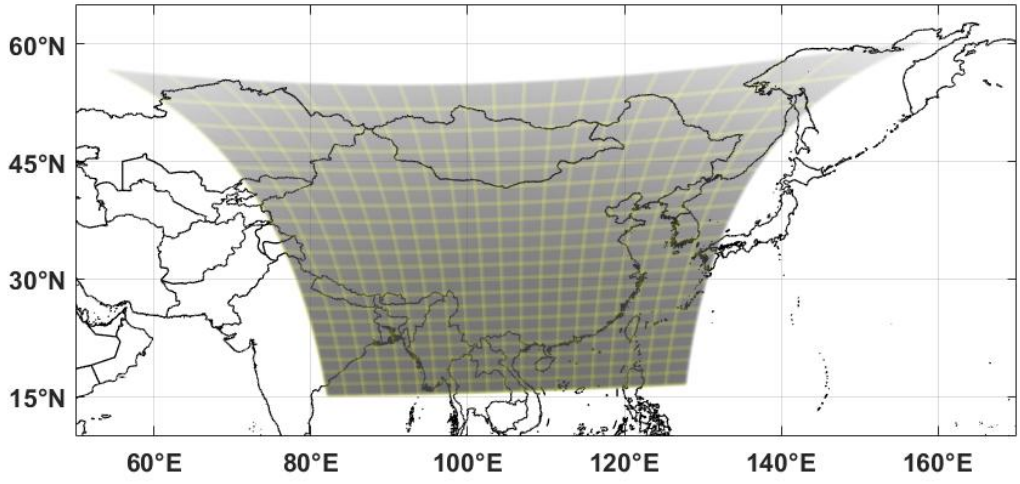


Figure 3: Subdivision of the LMI observational domain in the Northern Hemisphere

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Open Access Earth System Science Data Discussions

3.2 Curve-fitting–based correction method

As noted in the Introduction, the operational lightning products have already undergone multiple layers of correction, some of which are redundant, while others fail to address certain sources of bias. Consequently, the experiments presented here seek to address the geolocation deviation problem directly from a numerical perspective, using the final operational LMI products as the starting point. Building on the subdivision of the LMI observational domain into 400 subregions described, we aggregate the lightning location deviation values within each subregion along a 24-h time axis, using 10-min intervals. For each interval, the mean and standard deviation of the deviation vectors are calculated. These time series are then corrected using a curve-fitting approach, a methodology that has been validated in several previous studies on satellite payload geolocation bias correction(Wang et al., 2021; Zhang et al., 2021; Zhang et al., 2026). In recent thermal deformation correction experiments, Zhang et al. (2026) employed a composite model consisting of multiple superimposed Gaussian functions, which proved effective for correcting thermal deformation over the eastern portion of the LMI field of view. However, when extending this approach to full-domain correction, the systematic offsets associated with the left-hand CCD array of the LMI result in longitudinal and latitudinal deviation components. They deviate substantially from the ideal bivariate normal distribution. This necessitates the adoption of a more flexible fitting strategy that aligns more closely with the empirical structure of the deviation data, thereby ensures correction quality.

Accordingly, we adopt a spline-based fitting approach. Spline fitting is inherently robust to sporadic strong outliers occurring at individual time intervals and, while maintaining overall smoothness, can adaptively preserve the genuine diurnal variation of the deviation signal. This avoids overfitting induced by high-frequency noise and provides a stable representation of the underlying systematic bias. The fitted function $f(t)$ is obtained by minimizing the following functional:

$$min_f[\sum_{k\in v}\widetilde{\varpi}_k(\mu_k - f(t_k))^2 + \lambda\int[f''(t)]^2\,dt]\,, \quad (1)$$

Where $\sum_{k\in v}\widetilde{\varpi}_k(\mu_k - f(t_k))^2$ represents the weighted data fidelity term, Here, k denotes the k-th 10-min interval, and $v$denotes the set of valid data points (excluding NaNs and outliers).$t_k$is the time corresponding to the k-th interval (in hours), $\mu_k$is the mean observed location deviation within that interval, and $f(t_k)$is the value of the fitted function evaluated at t_k, The difference $\mu_k - f(t_k)$ corresponds to the fitting residual, and ϖ_k denotes the weight assigned to the k-th data point. The term $\lambda\int[f''(t)]^2\,dt$is the smoothness constraint, where $f''(t)$ is the second derivative of the fitted function with respect to time, representing its curvature, and λ is the smoothing parameter that controls the trade-off between fidelity to the data and overall smoothness of the fitted curve.

Because the number of LMI observations varies across time intervals and the data quality is not uniform, different intervals should contribute unequally to the fitting process. Intervals characterized by a larger sample size and lower dispersion are expected to be more reliable and therefore should be assigned higher weights. Accordingly, the weight is defined as a function that jointly accounts for the number of samples and the variability within each time interval, such that time bins with more observations and smaller spread exert a stronger influence on the fitted curve. Eq. (2):

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Open Access Earth System Science Data Discussions

$$\varpi_k^{raw} = \frac{N_k}{\sigma_k + \varepsilon}, \tag{2}$$

Where N_k denotes the number of samples within the k-th time interval, σ_k is the corresponding standard deviation, and $\varepsilon = 10^{-6}$ is a small constant introduced to avoid division by zero.

In practice, certain time intervals may contain only a very small number of lightning detections. In such cases, the estimated standard deviation can be artificially small, which would otherwise lead to an unrealistically large weight. To address this issue, additional constraints are imposed to downweight intervals characterized by extremely small standard deviations. Furthermore, all time-series data from the 400 subregions were manually inspected, and time intervals containing extreme outliers were explicitly identified and excluded from the analysis.

During the fitting procedure, we also found that some data points, while not classified as extreme outliers, still deviated substantially from the overall diurnal trend and could noticeably distort the fitted curve. To further suppress the influence of such points, an additional smoothing step was applied. Specifically, the spline-fitting results were subsequently processed using a three-point moving average, which further stabilizes the correction curve while preserving the dominant temporal structure.:

$$\hat{f}_{final}(t_k) = \frac{1}{3}\left[\hat{f}(t_{k-1}) + \hat{f}(t_k) + \hat{f}(t_{k+1})\right], \tag{3}$$

Where $\hat{f}_{final}(t_k)$ denotes the final smoothed curve used for correction, while $\hat{f}(t_k)$ represents the spline-fitted value obtained from Eq. (1). The terms t_{k-1}, t_k, t_{k+1} correspond to three consecutive 10-min time intervals. To ensure continuity of the diurnal cycle, the curve is constructed using a cyclic replication approach, in which the end of the 24-h period is smoothly connected to the beginning of the next day, thereby preserving temporal continuity across the 24-h boundary.

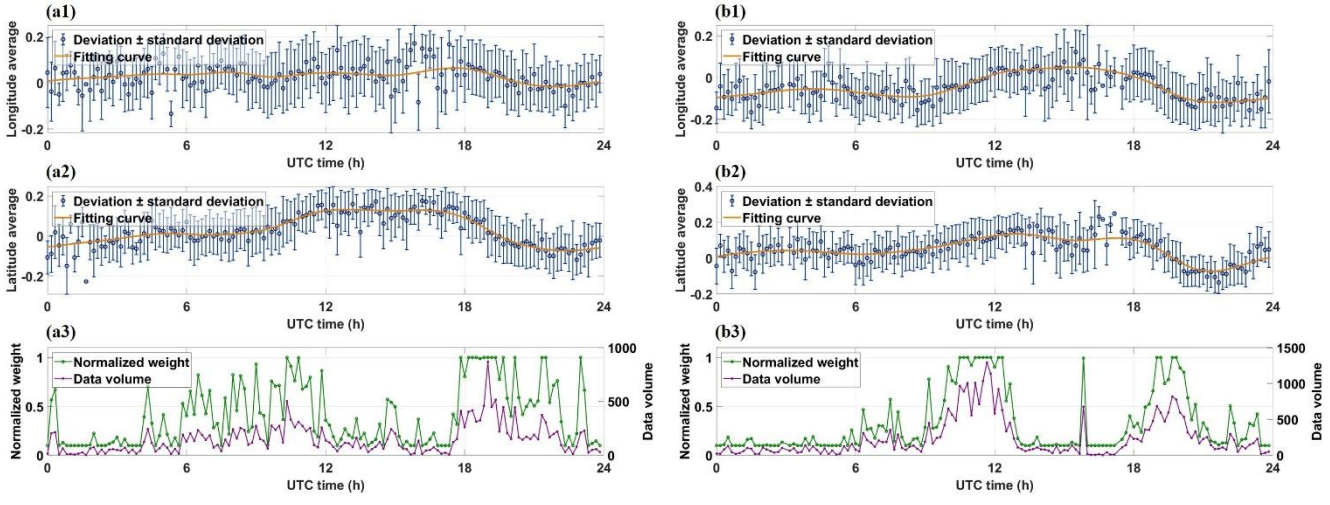
4 Results and discussion

Following the methodology outlined in Section 3, we compute the mean and standard deviation of the geolocation deviations for all observations within each corresponding time step. Figure 4 presents several representative subregions, illustrating the longitudinal and latitudinal deviations together with the associated weight calculations. Overall, in contrast to previous studies(Cao et al., 2021; Chen et al., 2021; Zhang et al., 2023), we do not observe a pronounced imbalance characterized by substantially more nighttime than daytime detections across the full domain. Instead, the data volume exhibits a more irregular temporal distribution, with lightning counts in a given subregion and time interval often being strongly influenced by episodic extreme weather events, such as typhoons(Zhang et al., 2015). Figure 4a shows the fitted deviation curves for a subregion in the eastern part of the LMI field of view. In the latitudinal component, a pronounced deviation emerges from approximately 10:00 UTC and weakens around 20:00 UTC, a pattern consistent with the thermal deformation correction results reported by Wang et al. (2021) and Zhang et al. (2026). This characteristic clearly indicates a thermally induced deformation driven by the diurnal cycle of solar radiation. In contrast, the longitudinal deviations are less distinct. This is attributable to the superposition

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Open Access Earth System Science Data Discussions

of thermally induced deformation and meridional distortion arising from the projection of the LMI CCD array onto the Earth's ellipsoidal surface. Although these effects do not fully coincide, a discernible displacement can still be identified near 20:00 UTC. In Fig. 4b, we examine the southernmost subregion of the LMI observational domain. This region lies closest to the satellite nadir and experiences minimal projection-induced distortion; consequently, its spatiotemporal deviation characteristics provide a particularly clear representation of systematic and periodic error sources. Here, a similar thermally driven periodic deviation is evident. Figure 4c shows a subregion near Xinjiang and Inner Mongolia, where lightning occurrence is extremely sparse. As a result, the available data volume is very limited, with many time intervals entirely missing, substantially increasing the difficulty of curve fitting and leading to a higher probability of extreme outliers. Nevertheless, certain deviation characteristics remain discernible. In particular, the longitudinal deviations exhibit a clear systematic offset, despite the small sample size. This is attributed to a westward displacement of the left half of the LMI detector array caused by acceleration during rocket launch, an effect that is further amplified by the proximity of this subregion to the northwestern edge of the LMI field of view. Figure 4d presents a subregion located near the boundary of the observational domain. Its longitudinal and latitudinal deviation patterns are broadly similar to those shown in Fig. 4a; however, extended periods of missing data are evident. Such prolonged data gaps can induce substantial biases in the fitted curves. To mitigate this issue, we adjust the weighting scheme and apply additional smoothing during the fitting process, allowing the curves to exhibit a predictive characteristic over data-sparse intervals and ensuring smooth temporal transitions. The underlying cause of these extended data gaps is likely related to the partitioning strategy employed in the operational processing algorithm, an issue that is examined in greater detail in the following analysis.。



https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

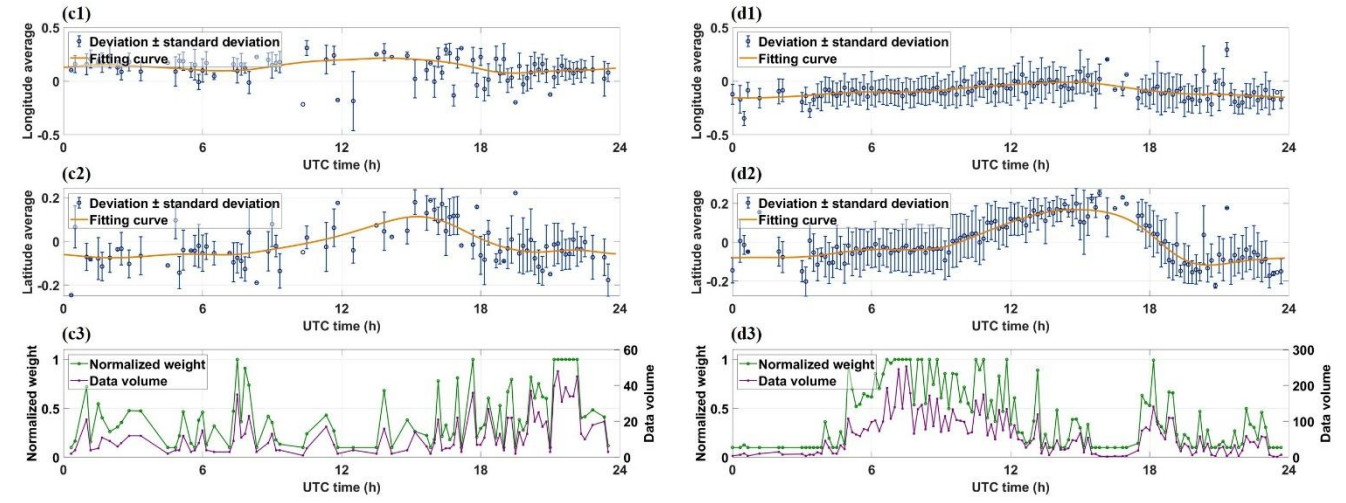
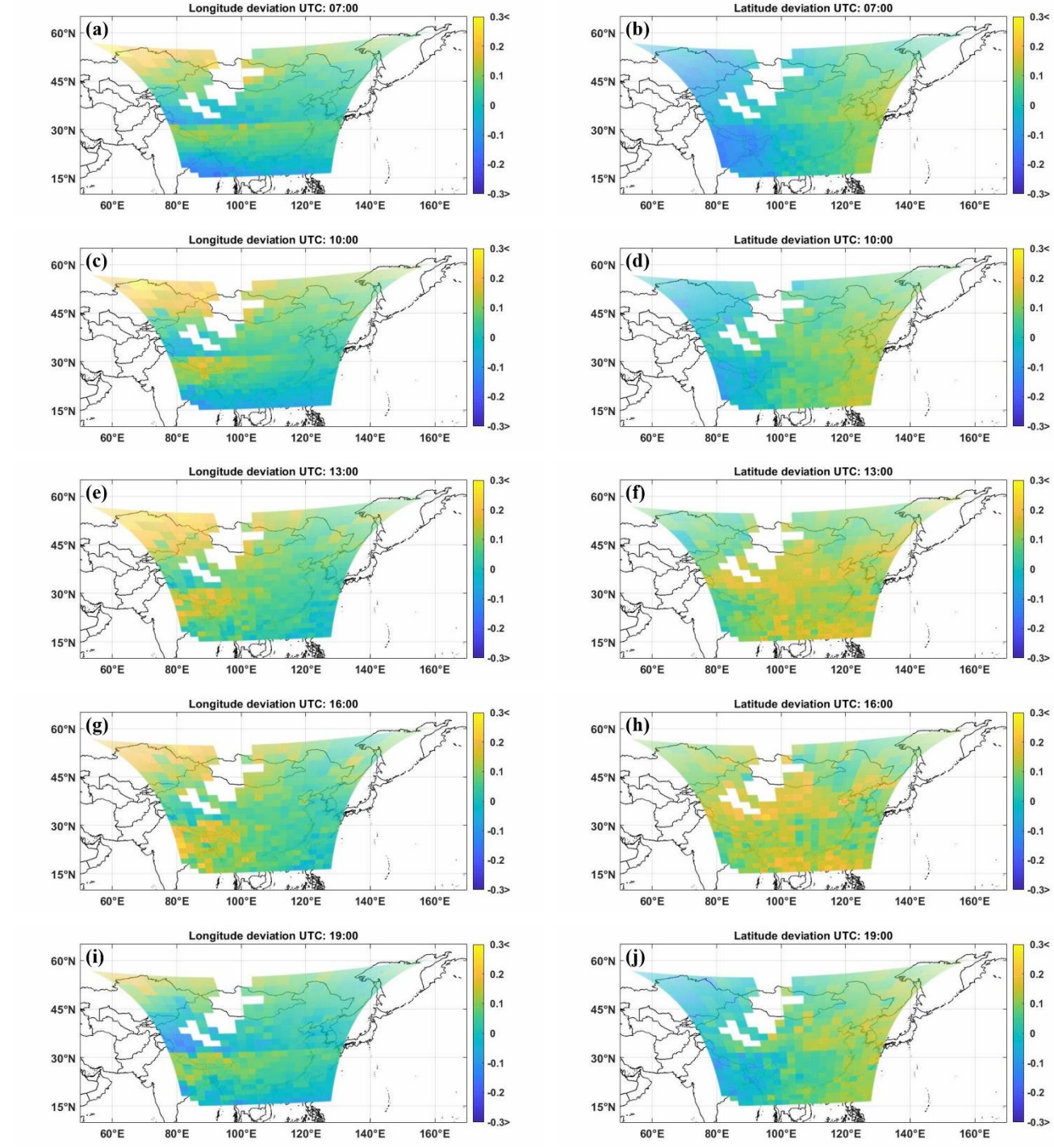


Figure 4: Fitted longitude and latitude deviation curves and corresponding weight distributions for selected LMI event subregions. (a1, a2, a3) Longitude deviation curve fitting, latitude deviation curve fitting, and weight distribution for subregion "15, 12"; (b1, b2, b3) Corresponding statistics for subregion "20, 13"; (c1, c2, c3) Corresponding statistics for subregion "2, 4"; (d1, d2, d3) Corresponding statistics for subregion "10, 1".

240

By aggregating the fitted results from all subregions across the full LMI field of view, we obtain the spatiotemporal distribution of lightning geolocation deviations over the entire observational domain (Fig. 5). From the latitude deviation patterns, a pronounced thermal deformation signal emerges, initiating around 10:00 UTC and propagating from east to west. This signal intensifies until approximately 16:00 UTC, after which it gradually weakens, stabilizing near 22:00 UTC before entering the 245 next diurnal cycle. This spatiotemporal evolution is fully consistent with the thermal deformation characteristics reported by Zhang et al. (2026) and other previous studies. In contrast, the longitude deviations exhibit a different characteristic. A marked systematic offset is evident on the western half of the LMI CCD array, attributable to intrinsic displacement of the detector plane. This offset exerts a stronger influence in the zonal (longitude) direction than in the meridional (latitude) direction, thereby partially masking the canonical eastward intensification and subsequent westward attenuation associated with thermal 250 deformation. Nevertheless, a clear temporal modulation of the longitude deviation remains discernible. Another noteworthy feature is the presence of distinct "boundaries" within the LMI observational domain, which arise from the regionalized processing strategy adopted in the operational LMI algorithm. Specifically, the algorithm partitions the field of view into a 4 × 4 grid comprising 16 regions, within which lightning characteristics are processed independently ((Hui et al., 2020). This regionalized treatment introduces discontinuities and anomalous geolocation characteristic near the boundaries, giving rise to 255 the systematic, boundary-like deviation patterns observed around 30°N in Fig. 5, as well as the extended temporal gaps in lightning detections evident for certain edge subregions (for example, Fig. 4d). These artifacts underscore the influence of algorithmic partitioning on the spatial coherence of LMI lightning products and motivate further investigation, which we address in subsequent sections.

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026



https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

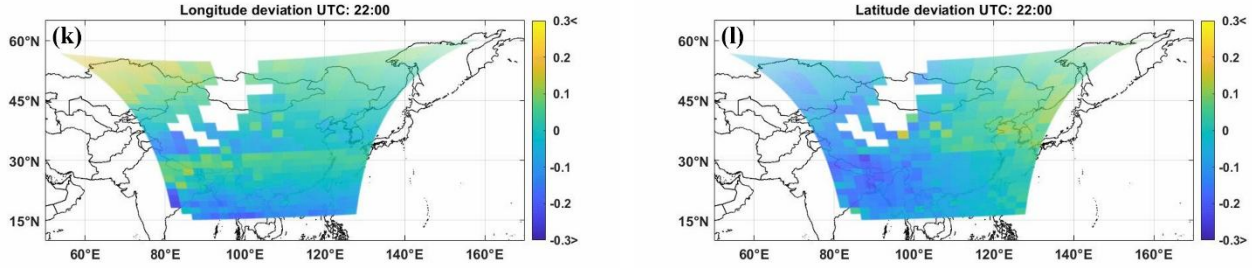


Figure 5: Spatiotemporal distribution of lightning geolocation deviations for LMI events over the Northern Hemisphere observational domain. The left column shows the longitudinal deviations, and the right column shows the latitudinal deviations. Panels (a, b) correspond to UTC 07:00; (c, d) UTC 10:00; (e, f) UTC 13:00; (g, h) UTC 16:00; (i, j) UTC 19:00; and (k, l) UTC 22:00. The statistics are compiled from LMI observations collected during the boreal warm season (March–September) from 2019 to 2023.

Figure 6 illustrates the distributions of coordinate deviations between LMI events and WWLLN flashes before and after correction, with WWLLN flashes used as the reference. As shown in Fig. 6b, even after preprocessing with the cloud-top height parallax correction, the majority of LMI lightning detections still exhibit a pronounced northward bias, accompanied by a moderate eastward bias. After applying the proposed correction (Fig. 6e), the distribution of coordinate differences exhibits a clear convergence. Most data points are more tightly clustered around zero and approximate a Gaussian distribution, with longitudinal deviations primarily confined to −0.1° to 0.15° and latitudinal deviations to −0.05° to 0.15°. The overall dispersion is substantially reduced, indicating a marked improvement in geolocation accuracy. It should be noted that relatively broad matching thresholds were adopted in this study, with the objective of associating LMI events with lightning-dense regions represented by ground-based observations, rather than enforcing a strict one-to-one correspondence between individual lightning discharges. As a result, some mismatches and “noise” are inevitable, leading to a small number of data points with comparatively large deviations. In addition, the intrinsic location uncertainty of WWLLN itself also contributes to the residual spread. Nevertheless, when sufficiently large sample sizes are available, these effects are effectively mitigated in a statistical sense, enabling robust correction of the systematic geolocation deviations.

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

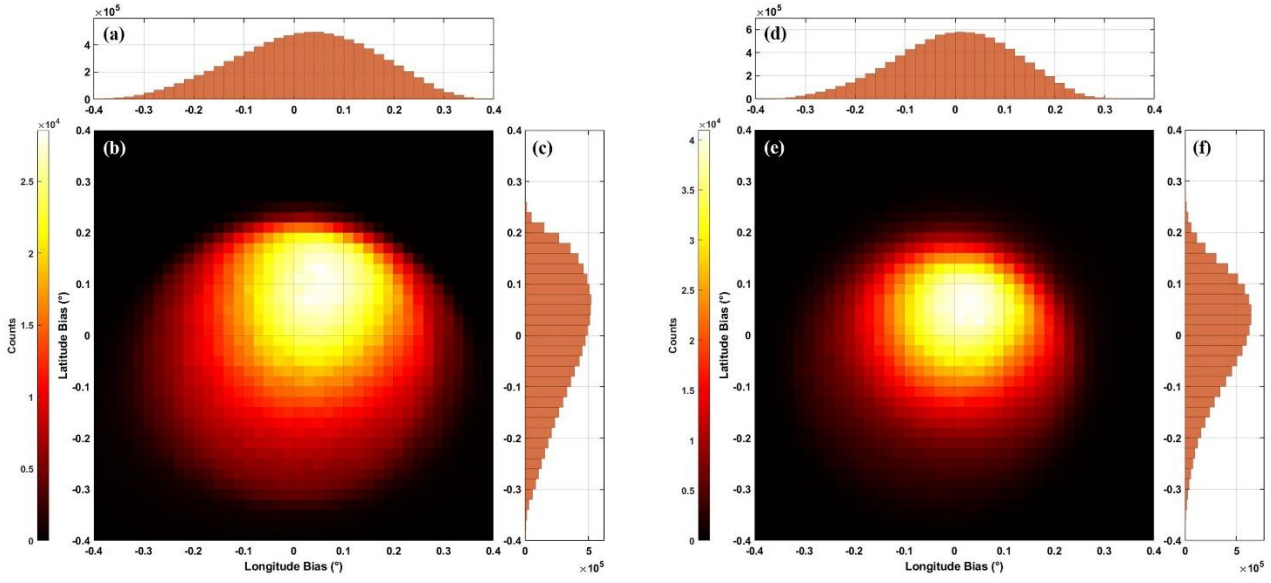


Figure 6: Distributions of coordinate differences between LMI events and WWLLN flashes before and after correction. (b) the density distribution of LMI event coordinate differences prior to correction, with panels (a) and (c) presenting the corresponding projections of the longitudinal and latitudinal deviations, respectively; (e) the density distribution after correction, while panels (d) and (f) display the corresponding longitudinal and latitudinal deviation projections. The statistics are based on data collected during the boreal warm season (March–September) from 2019 to 2023.

Finally, the fitted curve values are applied as correction terms to the lightning observations, and a corrected LMI event dataset is generated. Using this dataset, we evaluate the geolocation accuracy across the LMI observational domain (Fig. 7). The results show that, over most southern regions, the lightning geolocation accuracy is maintained at approximately 15 km (about 1.5 pixels), representing a relatively ideal level of performance. Note that parts of Xinjiang and Mongolia, as well as regions near India along the southwestern edge of the observational domain, experience sparse lightning activity. This leads to data gaps that preclude a robust accuracy assessment. However, subregions adjacent to these low–lightning-occurrence areas exhibit relatively higher geolocation accuracy. This can be attributed to the limited number of lightning events in these regions, which results in low matching noise and enables the fitted corrections to align more closely with ground-based lightning locations. A distinct degradation in geolocation accuracy is observed in the northeastern portion of the observational domain. This is primarily due to the intrinsically coarse spatial resolution of the detector elements in this region, where a single pixel can exceed 20 km in ground coverage, thereby degrading the overall data quality. Overall, after correction, the geolocation accuracy of LMI lightning observations is spatially stable across most of the observational domain.

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

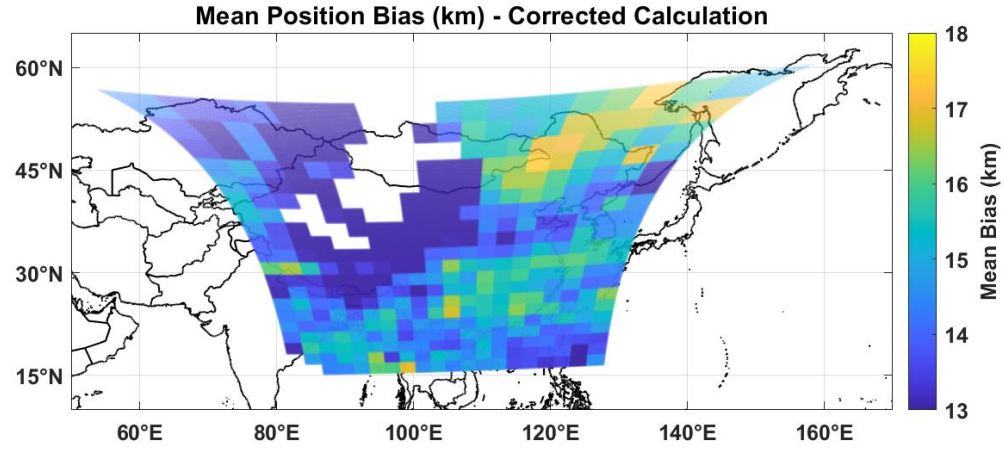


Figure 7: Spatial distribution of lightning geolocation accuracy over the LMI Northern Hemisphere observational domain

5 Data availability

The corrected FY-4A LMI event dataset produced in this study is publicly available at https://doi.org/10.11888/Atmos.tpdc.303312. Because lightning datasets are characterized by large data volumes and high temporal resolution, storing them in the conventional NetCDF (.nc) format would result in excessively large file sizes, making the datasets difficult to download and process. Therefore, the corrected lightning dataset generated in this study is stored in a tabular (.csv) format, with one file produced per day. The dataset covers corrected LMI event data from March to September for the period 2019–2023. Detailed descriptions of the table attributes are provided in Table A1 of Appendix A. Information on the acquisition of all other datasets used in this study is summarized in Table A2 of Appendix A.

6 Conclusion

Spaceborne lightning observations are inevitably affected by factors such as solar radiation, satellite platform jitter, and systematic offsets of the detector array, which together introduce both periodic and random geolocation deviations in lightning positioning. To address these issues for the FY-4A LMI, this study proposes a correction framework based on a robust spline-fitting model. Ground-based lightning location data are used as a reference to represent lightning-dense regions, and the positional deviations between spaceborne lightning detections and the reference observations are quantified. These deviations are then fitted using a robust spline model to derive correction curves, which are applied to mitigate both systematic and random errors in the spaceborne lightning observations. The proposed method is evaluated using combined spaceborne and ground-based lightning observations collected during the boreal warm seasons (March–September) from 2019 to 2023. The results demonstrate that this approach effectively improves the geolocation accuracy of FY-4A LMI lightning products. Owing to regional differences in lightning characteristics, detection efficiency, and payload displacement, distinct subregions exhibit

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

heterogeneous geolocation deviation patterns. The fitting framework adaptively captures these region-specific features and generates smooth correction curves tailored to the underlying deviation structures, thereby providing a robust correction strategy. In addition, this study identifies previously unreported, boundary-like systematic deviations arising from the operational partitioning strategy employed in the LMI processing algorithm. The proposed spline-based correction method is also shown to effectively mitigate these artifacts, further enhancing the spatial consistency and reliability of the corrected lightning dataset. The newly developed, highly efficient approach has been employed to construct a comprehensively corrected lightning dataset. After correction, the lightning geolocation deviations exhibit pronounced convergence, with most residual errors confined within ±0.15° (approximately 15 km, or about 1.5 pixels), approaching the theoretical limit of achievable accuracy.

Appendix A: Supplementary material

Table A1. Detailed descriptions of the FY-4A LMI Event Corrected Dataset

Column number	Attribute	Description
1	Year	Year derived from the FY-4A LMI event (LMIE) Level-2 lightning product
2	Month	Month derived from the Level-2 lightning product
3	Day	Day derived from the Level-2 lightning product
4	Hour	Hour derived from the Level-2 lightning product
5	Minute	Minute derived from the Level-2 lightning product
6	Second	Seconds derived from the seconds and milliseconds in the Level-2 lightning product files. Unit: s
7	Event Radiance	Radiance values of lightning locations in the Level-2 product provided by the China Meteorological Administration. Unit: μJ/m^2/ster
8	Longitude	Longitude of lightning locations after correction for cloud-top height parallax and systematic errors in this study. Unit: °E (Decimal degrees)
9	Latitude	Latitude of lightning locations after correction for cloud-top height parallax and systematic errors in this study. Unit: °N (Decimal degrees)
10	Cloud Top Height	Cloud-top height associated with individual lightning locations, derived from the cloud-top height product provided by the China Meteorological Administration. Unit: m

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Table A2. Overview of where to find the data and model used in the current work.

Data and Model Name	Variables	DOI or PID	Reference
FY-4A event (LMIE) product	LON, LAT	http://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx	Yang et al. (2017)
WWLLN flash	LON, LAT	http://wwlln.net	Abarca et al. (2010)
FY-4A AGRI Cloud Top Height (CTH) product	CTH	http://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx	Yang et al. (2017)
Ellipsoid CTH Parallax Correction model (ECPC)		https://doi.org/10.3390/rs15194856	Zhang et al. (2023)

Author contributions

YZ: data curation, investigation, validation, visualisation, writing (original draft), writing (review and editing), conceptualisation, formal analysis, methodology, soft ware. XQ: investigation, validation, methodology, writing (original draft), writing (review and editing), conceptualisation, formal analysis, funding acquisition, resources. RJ: data curation, formal analysis, methodology, writing (original draft), writing (review and editing). DC: data curation, methodology, writing (review and editing), resources. JY: data curation, methodology, writing (review and editing), resources. DW: data curation, methodology, writing (original draft), writing (review and editing). ML: writing (review and editing). DL: data curation, writing (review and editing). ZS: writing (review and editing). HZ: writing (review and editing). SY: methodology, writing (review and editing).

Competing interests

The authors declare no conflicts of interest.

Acknowledgements

The authors thank the National Satellite Meteorological Center for providing the FY-4A LMI data "http://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx (accessed on 30 December 2025)", and the WWLLN "http://wwlln.net (accessed on 30 December 2025)", a collaboration among over 40 universities and institutions, for providing the WWLLN data used in this paper.

Financial support

This research was funded by the National Natural Science Foundation of China (Nos. 42230609, 42027803) and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0760300).

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

References

Abarca, S. F., Corbosiero, K. L., and Galarneau, T. J.: An evaluation of the worldwide lightning location network (wwlln) using the national lightning detection network (nldn) as ground truth, Journal of Geophysical Research: Atmospheres, 115, 2009JD013411, https://doi.org/10.1029/2009JD013411, 2010.

Brune, W. H., Mcfarland, P. J., Bruning, E., Waugh, S., Macgorman, D., Miller, D. O., Jenkins, J. M., Ren, X., Mao, J., and Peischl, J.: Extreme oxidant amounts produced by lightning in storm clouds, Science, 372, 711-715, https://doi.org/10.1126/science.abg0492, 2021.

Buechler, D., Varghese, T., Armstrong, P., Bremer, J., Lamb, R., Fulbright, J., Goodman, S., Butler, J. J., Xiong, X. J., and Gu, X.: On-orbit validation of the geolocation accuracy of goes-16 geostationary lightning mapper (glm) flashes using ground-based laser beacons, in: Earth Observing Systems XXIII, SPIE, 2018/1/1, 107640J, 2018.

Cao, D.: The development of product algorithm of the fengyun-4 geostationary lightning mapping imager, Advances in Meteorological Science and Techology, 6, 94-98, https://doi.org/10.3969/j.issn.2095-1973.2016.01.014, 2016.

Cao, D., Lu, F., Zhang, X., and Yang, J.: Lightning activity observed by the fengyun-4a lightning mapping imager, Remote Sens., 13, 3013, https://doi.org/10.3390/rs13153013, 2021.

Carr, J. L., Tillier, C. E., and Shu, Y.: Validation of image navigation and registration for the geostationary lightning mapper, J. Appl. Remote Sens., 3, 1, https://doi.org/10.1117/1.JRS.14.032403, 2020.

Chen, Z., Qie, X., Sun, J., Xiao, X., Zhang, Y., Cao, D., and Yang, J.: Evaluation of fengyun-4a lightning mapping imager (lmi) performance during multiple convective episodes over beijing, Remote Sens., 13, 1746, https://doi.org/10.3390/rs13091746, 2021.

Cheng, K., Tong, X., Liu, S., Yan, X., and Li, H.: An on-orbit geometric calibration approach based on double tubes edge-binding feng yun-4a lightning mapping imager, Journal of Geomatics Science and Technology, 38, 477-484, 2021.

Fan, P., Zheng, D., Zhang, Y., Gu, S., Zhang, W., Yao, W., Yan, B., and Xu, Y.: A performance evaluation of the world wide lightning location network (wwlln) over the tibetan plateau, J. Atmos. Ocean. Technol., 35, 927-939, https://doi.org/10.1175/JTECH-D-17-0144.1, 2018.

Goodman, S. J., Blakeslee, R. J., Koshak, W. J., Mach, D., Bailey, J., Buechler, D., Carey, L., Schultz, C., Bateman, M., Mccaul, E., and Stano, G.: The goes-r geostationary lightning mapper (glm), Atmos. Res., 125-126, 34-49, https://doi.org/10.1016/j.atmosres.2013.01.006, 2013.

Holmlund, K., Grandell, J., Schmetz, J., Stuhlmann, R., Bojkov, B., Munro, R., Lekouara, M., Coppens, D., Viticchie, B., August, T., Theodore, B., Watts, P., Dobber, M., Fowler, G., Bojinski, S., Schmid, A., Salonen, K., Tjemkes, S., Aminou, D., and Blythe, P.: Meteosat third generation (mtg): continuation and innovation of observations from geostationary orbit, Bull. Amer. Meteorol. Soc., 5, https://doi.org/10.1175/BAMS-D-19-0304.1, 2021.

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Hui, W., Huang, F., and Liu, R.: Characteristics of lightning signals over the tibetan plateau and the capability of fy-4a lmi lightning detection in the plateau, Int. J. Remote Sens., 41, 4605-4625, https://doi.org/10.1080/01431161.2020.1723176, 2020.

Hui, W., Zhang, W., Lyu, W., and Li, P.: Preliminary observations from the china fengyun-4a lightning mapping imager and its optical radiation characteristics, Remote Sens., 12, 2622, https://doi.org/10.3390/rs12162622, 2020.

Loto Aniu, P. T. M., Davis, A., Jarvis, A., Grotenhuis, M., Rich, F. J., Califf, S., Inceoglu, F., Pacini, A., and Singer, H. J.: Initial on-orbit results from the goes-18 spacecraft science magnetometer, Space Sci. Rev., 219, 84, https://doi.org/10.1007/s11214-023-01032-3, 2023.

Peterson, M. and Mach, D.: The illumination of thunderclouds by lightning: 4. Volumetric thunderstorm imagery, Earth Space Sci., 9, https://doi.org/10.1029/2021EA001945, 2022.

Peterson, M., Light, T. E. L., and Mach, D.: The illumination of thunderclouds by lightning: 1. The extent and altitude of optical lightning sources, Journal of Geophysical Research: Atmospheres, 127, https://doi.org/10.1029/2021JD035579, 2022a.

Peterson, M., Light, T. E. L., and Mach, D.: The illumination of thunderclouds by lightning: 2. The effect of glm instrument threshold on detection and clustering, Earth Space Sci., 9, https://doi.org/10.1029/2021EA001943, 2022b.

Peterson, M., Light, T. E. L., and Mach, D.: The illumination of thunderclouds by lightning: 3. Retrieving optical source altitude, Earth Space Sci., 9, https://doi.org/10.1029/2021EA001944, 2022c.

Qie, X., Zhang, Y., Zhang, D., Yin, Y., Yu, Y., Lu, G., and Jiang, R. (Eds:): Principles and forecast of thunderbolt weather system, Science Press, Beijing, 2023.

Rodger, C. J., Werner, S., Brundell, J. B., Lay, E. H., Thomson, N. R., Holzworth, R. H., and Dowden, R. L.: Detection efficiency of the vlfworld-wide lightning location network (wwlln): initial case study, Ann. Geophys., 12, 3197-3214, https://doi.org/doi:10.5194/angeo- 24-3197-2006, 2006.

Rudlosky, S. D., Goodman, S. J., Virts, K. S., and Bruning, E. C.: Initial geostationary lightning mapper observations, Geophys. Res. Lett., 46, 1097-1104, https://doi.org/10.1029/2018GL081052, 2019.

Wang, J., Tong, X., Yang, L., Shang, J., Liu, C., Bao, S., Zhang, Z., and Yang, J.: Image navigation for fy-4a lightning mapping imager, Ieee J. Sel. Top. Appl. Earth Observ. Remote Sens., 14, 11450-11465, https://doi.org/10.1109/JSTARS.2021.3124955, 2021.

Yang, J., Zhang, Z., Wei, C., Lu, F., and Guo, Q.: Introducing the new generation of chinese geostationary weather satellites, fengyun-4, Bull. Amer. Meteorol. Soc., 98, 1637-1658, https://doi.org/10.1175/BAMS-D-16-0065.1, 2017.

Zhang, H., Zhao, X., Mei, Q., Wang, Y., Song, S., and Yu, F.: On-orbit thermal deformation prediction for a high-resolution satellite camera, Appl. Therm. Eng., 195, 117152, https://doi.org/10.1016/j.applthermaleng.2021.117152, 2021.

Zhang, W., Zhang, Y., Zheng, D., Wang, F., and Xu, L.: Relationship between lightning activity and tropical cyclone intensity over the northwest pacific, Journal of Geophysical Research: Atmospheres, 120, 4072-4089, https://doi.org/10.1002/2014JD022334, 2015.

https://doi.org/10.5194/essd-2026-70
Preprint. Discussion started: 20 March 2026

Zhang, Y., Cao, D., Yang, J., Lu, F., Wang, D., Liu, R., Zhang, H., Liu, D., Chen, Z., Lyu, H., Cai, W., Bao, S., and Qie, X.: A parallax shift effect correction based on cloud top height for fy-4a lightning mapping imager (lmi), Remote Sens., 15, 4856, https://doi.org/10.3390/rs15194856, 2023.

Zhang, Y., Qie, X., Cao, D., Yuan, S., Wang, D., Zhang, H., Liu, D., Sun, Z., Liu, M., Zhu, K., Jiang, R., and Yang, J.: Thermal deformation correction for the fy-4a lmi, Remote Sens., 18, 163, https://doi.org/10.3390/rs18010163, 2026.