Response to Reviewer #1 Comments

Thank you very much for your valuable comments on our manuscript. These comments are highly appreciated and have been very helpful in improving the quality of our manuscript and strengthening the overall research.

Comment 1: Line 103: It is not clear what the authors mean by "stable WWLLN observations" .

Reply: We agree that the phrase "stable WWLLN observations" should be defined more explicitly at its first use. In the Introduction, we had noted that the initial years of the WGLC record were influenced by WWLLN network build-out and that reprocessed subsets are therefore needed for applications requiring a temporally stable baseline. Here, by "stable WWLLN observations", we refer specifically to the 2013–2024 WWLLN/WGLC period used for model training and evaluation, excluding the initial 2010–2012 years that are more strongly affected by network build-out. We have revised the sentence to make this definition explicit at Lines 104-107: "*Leveraging ERA5 reanalysis data, we employ an ensemble approach—integrating XGBoost, RF, LightGBM, and DNN—trained against the temporally stable WWLLN/WGLC observations from 2013 to 2024, following the exclusion of the initial 2010–2012 network build-out period.*".

Comment 2: Line 126: Please clarify whether the stroke density is accumulated for every month or averaged.

Reply: We thank the reviewer for pointing out this ambiguity. We used the WGLC monthly 5 arcmin lightning-density time-series product. In the WGLC processing chain, georeferenced WWLLN stroke detections are gridded to hourly rasters, converted to stroke density, corrected for detection efficiency, and then aggregated into monthly totals. Therefore, the monthly density fields used in this study represent monthly aggregated stroke-density fields, rather than multi-year monthly mean climatological values. The variable is reported as lightning stroke density in units of strokes km^{-2} day^{-1}. We have revised Sect. 2.2 to clarify this definition at Lines 134-137: "*We utilized the WGLC monthly 5 arcmin stroke-density time-series product, in which the density fields are aggregated at monthly time steps and reported in units of strokes km^{-2} day^{-1}, rather than as raw monthly stroke counts. The data were conservatively aggregated to a 0.25° grid to align with the ERA5 predictors.*".

Comment 3: Lines 188-192: It is not clear what additional advantage the inclusion of tree-based models provides compared to using only the DNN, or vice versa. The authors should explain, from a theoretical perspective, how combining these different models is expected to improve the robustness and overall performance of the final ensemble.

Reply: We thank the reviewer for this helpful suggestion. We have revised Sect. 3.2 to clarify the rationale for combining tree-based models with the DNN. Specifically, we now explain that DNNs provide a different inductive bias by learning flexible nonlinear representations from standardized predictors, thereby complementing the partition-based representations of tree models. Relevant references have also been added at Lines 195-199: "*To complement these partition-based approaches, we incorporated a DNN, which excels at feature learning and generalization(Zhang et al., 2016). DNNs provide a different inductive bias by learning flexible nonlinear representations from standardized predictors, which can complement the partition-based*

representations of tree models (Zhang et al., 2016).".

Comment 4: Lines 247-248: Please refer to Table 1 here for the statistical metrics related to the ridge regression model.

Reply: We thank the reviewer for this suggestion. We have added a reference to Table 1 at this point to direct readers to the statistical metrics of the ridge regression ensemble.

Comment 5: Lines 258-261: What is the performance of the random forest (RF) model in this context? Figure 3f suggests that the RF model performs better than the other individual models, with a larger fraction of grid cells exhibiting high R^2 values, and in some cases even outperforming the ridge regression ensemble. Please discuss this result.

Reply: We thank the reviewer for this insightful comment. We have revised Sect. 4.1 to discuss the strong RF performance shown in Fig. 3f at Lines 311-317: "*RF also shows strong grid-cell-level performance, with a relatively large fraction of grid cells falling into high-R^2 intervals. This reflects the complementary information provided by grid-cell-level diagnostics, which emphasize spatially distributed local skill, whereas the pooled test-set metrics in Table 1 summarize overall predictive performance across all samples. However, RF also exhibits a larger train–test discrepancy than the ridge ensemble, suggesting a greater tendency toward overfitting and supporting the use of the ridge ensemble for the final reconstruction.*". The revised text clarifies that RF performs well in grid-cell-level diagnostics and can locally exceed the ridge ensemble, whereas the pooled metrics in Table 1 summarize overall sample-level performance. We also note that RF shows a larger train–test discrepancy than the ridge ensemble, indicating a greater tendency toward overfitting. This distinction explains why RF can show strong local skill while the ridge ensemble remains more suitable for the final reconstruction.

Comment 6: Section 4.1: Although RMSE and MAE are useful performance metrics, they are reported in absolute units. To better assess the magnitude of the errors, the mean value of the target variable should also be reported alongside these statistics. Alternatively, and preferably, I recommend reporting normalized RMSE and normalized MAE so that the errors are expressed as percentages rather than absolute values.

Reply: We thank the reviewer for this helpful suggestion. We agree that RMSE and MAE reported only in absolute units do not fully convey the relative magnitude of model errors. Following the reviewer's recommendation, we added normalized RMSE (nRMSE) and normalized MAE (nMAE), which express the errors as percentages relative to the mean observed lightning density of the corresponding training or test set.

We added a dedicated Supplementary Information section describing the calculation of nRMSE and nMAE, including the equations and the full train/test normalized error metrics (Sect. S1 and Table S3). We also revised Sect. 4.1 to briefly summarize these normalized-error results at Lines 294-297 and to refer readers to the Supplementary Information: "*The normalized error metrics provide an additional scale-aware evaluation of model errors and indicate that the ridge regression ensemble achieves the lowest test nRMSE and a test nMAE comparable to the lowest value among the tested models (Sect. S1 and Table S3).*".

Text S1. *Normalized error metrics*

To facilitate a scale-aware interpretation of model errors, normalized RMSE (nRMSE) and normalized MAE (nMAE) were calculated in addition to the absolute RMSE and MAE values. The normalized metrics were calculated as follows:

$$nRMSE_s = \frac{RMSE_s}{\bar{y}_{obs,s}} \times 100\% \quad (1)$$

$$nMAE_s = \frac{MAE_s}{\bar{y}_{obs,s}} \times 100\% \quad (2)$$

where s denotes the data split, either the training set or the test set, and $\bar{y}_{obs,s}$ *denotes the mean observed lightning density of the corresponding split.*
The mean observed lightning density was 0.0050267982 strokes km^{-2} day^{-1} for the training set and 0.0050160284 strokes km^{-2} day^{-1} for the test set. The resulting normalized error metrics are shown in Table S3. Given the sparse and right-skewed distribution of lightning density, these mean-normalized metrics were interpreted primarily as scale-aware comparisons among the tested models.

Table S3. *Normalized error metrics for the four base models and the ridge regression ensemble.*

Model	***nRMSE train***	***nRMSE test***	***nMAE train***	***nMAE test***
XGBoost	*210.87 %*	*231.26 %*	*63.66 %*	*65.79 %*
RF	*165.12 %*	*227.27 %*	*39.79 %*	*59.81 %*
LightGBM	*185.01 %*	*219.30 %*	*57.69 %*	*63.80 %*
DNN	*214.85 %*	*231.26 %*	*67.64 %*	*69.78 %*
Ridge regression	*167.10 %*	*215.31 %*	*47.74 %*	*59.81 %*

Comment 7: Figure 4: In line with the previous comment, please report the bias in panel (e) in percentage. Also, is the reported correlation coefficient calculated from the zonal mean comparison? If so, please state this explicitly in the figure caption.

Reply: We thank the reviewer for this helpful suggestion. We retained the zonal-mean absolute bias and added a new panel showing the zonal-mean percentage bias to provide a complementary relative-error measure. We also clarified in the Fig. 4 caption that the Pearson correlation coefficient in panel (d) was calculated from the zonal-mean values across latitude bands:

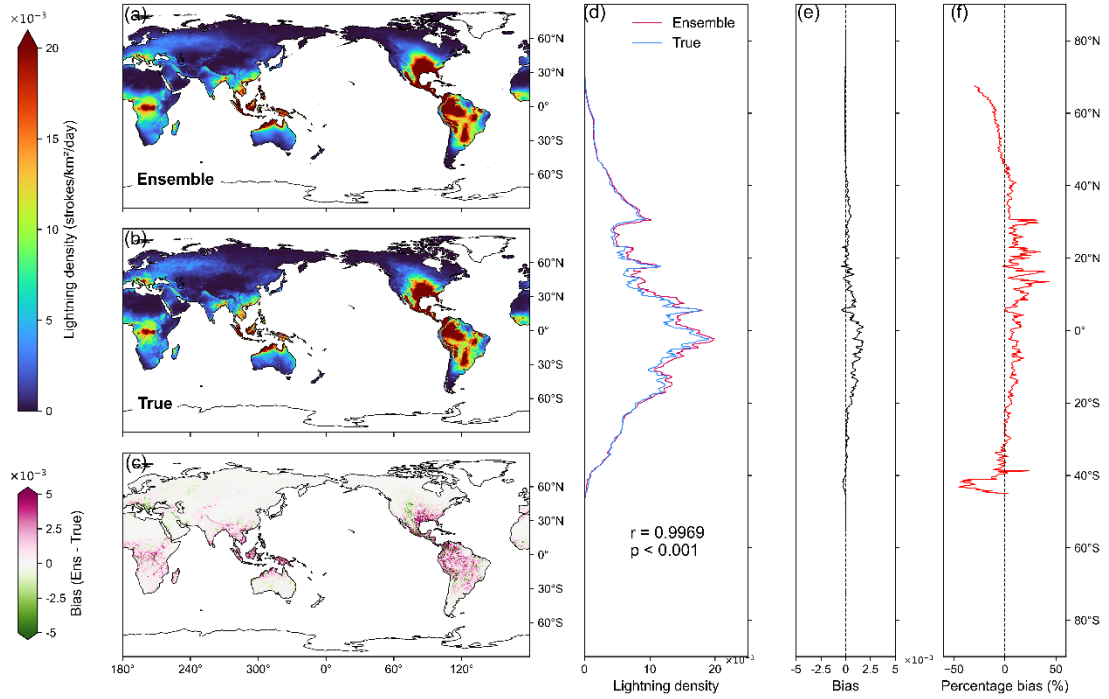

Figure 4. *Spatial distributions of lightning density for* ***(a)*** *the ridge regression ensemble and* ***(b)*** *observations, alongside* ***(c)*** *the associated bias during 2013–2024. The corresponding zonal mean lightning densities for the ensemble and observations are shown in* ***(d)****. The Pearson correlation coefficient and p-value indicated within panel* ***(d)*** *were calculated from the zonal mean values across latitude bands. The zonal mean absolute bias and percentage bias are shown in* ***(e)*** *and* ***(f)****, respectively. Lightning density is expressed in strokes km^{-2} day^{-1}.*

And we have updated the corresponding text in Sect. 4.2 at Lines 340-347:"*The reconstructed and observed zonal-mean profiles showed a high correlation (r = 0.9969, p < 0.001; Fig. 4d), indicating that the reconstruction reproduced the broad tropical-to-polar gradient in lightning density.*" and "*The absolute and percentage differences further characterized latitude-dependent discrepancies in magnitude, with the dominant error being a tropical positive bias within approximately ±20° latitude (Figs. 4e–f).*".

Comment 8: More generally, I recommend reporting bias and other comparison statistics as percentages as well throughout the manuscript (for example, Fig. 5d) rather than in absolute units only.

Reply: We thank the reviewer for this helpful recommendation. We have added percentage-based measures where appropriate while retaining the corresponding absolute quantities as complementary information. As described in our responses to Comments 6 and 7, mean-normalized RMSE and MAE have been added in Sect. S1 and Table S3 to complement the error metrics reported in absolute units.

And we added percentage-bias and percentage-difference panels to 5:

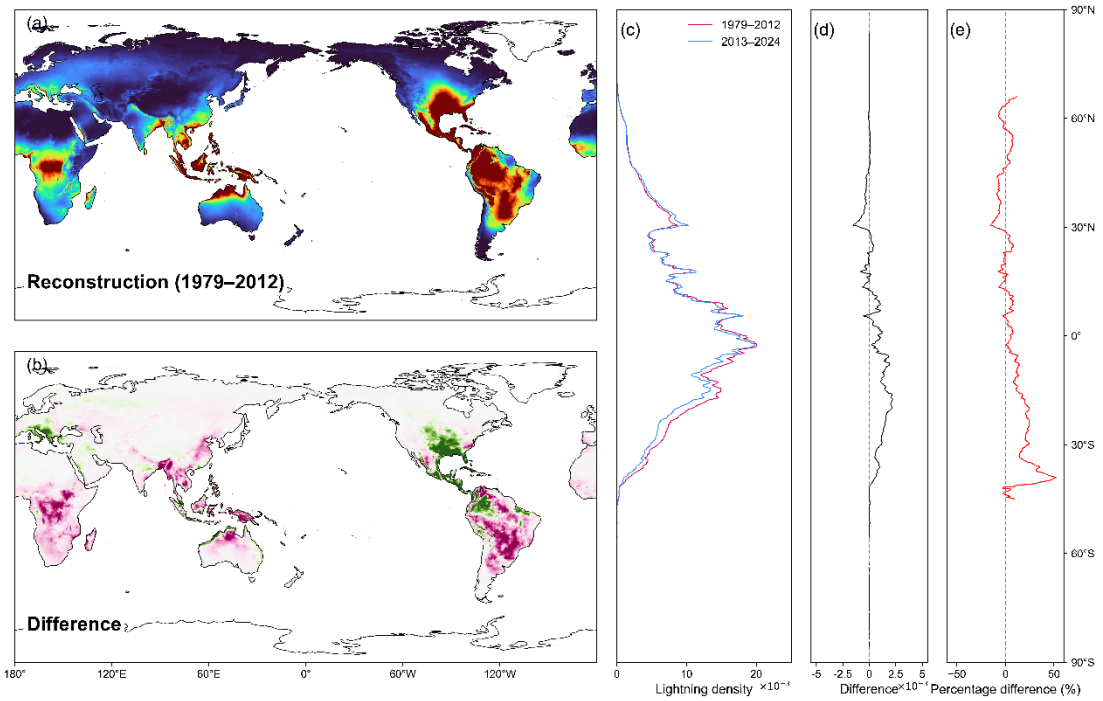

Figure 5. *Global distribution and zonal-mean structure of the reconstructed lightning density.* ***(a)*** *Multi-year mean global lightning density from the ensemble reconstruction during 1979–2012.* ***(b)*** *Difference in multi-year mean lightning density between 1979–2012 and 2013–2024, calculated as 1979–2012 minus 2013–2024.* ***(c)*** *Zonal-mean lightning density for the two periods. Panels* ***(d)*** *and* ***(e)*** *show the corresponding zonal-mean absolute and percentage differences, respectively, with the percentage difference calculated relative to the 2013–2024 climatology. Lightning density is expressed in strokes* km^{-2} day^{-1}.

The relevant figure captions and Sect. 4.2 have been revised accordingly at Lines 369-374: "*The zonal-mean absolute difference further indicated that positive differences were concentrated between the equator and approximately 45°S, whereas negative differences were most evident near 30°N (Fig. 5d). The corresponding percentage-difference profile provided a complementary relative measure, showing positive relative differences mainly between the equator and approximately 40°S and negative relative differences near 30°N and at higher northern latitudes (Fig. 5e).*".

Comment 9: Figure 6: The y-axis limits in panels (b) and (d) could be adjusted to better highlight the variations in the annual time series.

Reply: We thank the reviewer for this helpful suggestion. We agree that the original y-axis ranges in Figs. 6b and 6d compressed the apparent interannual variability. We have therefore narrowed the y-axis limits in both panels to more closely encompass the full ranges of the annual time series, while retaining a small margin around all data points. This adjustment improves the visibility of the year-to-year variations without excluding any values or altering the underlying data and analyses. The revised Fig. 6 is shown below:

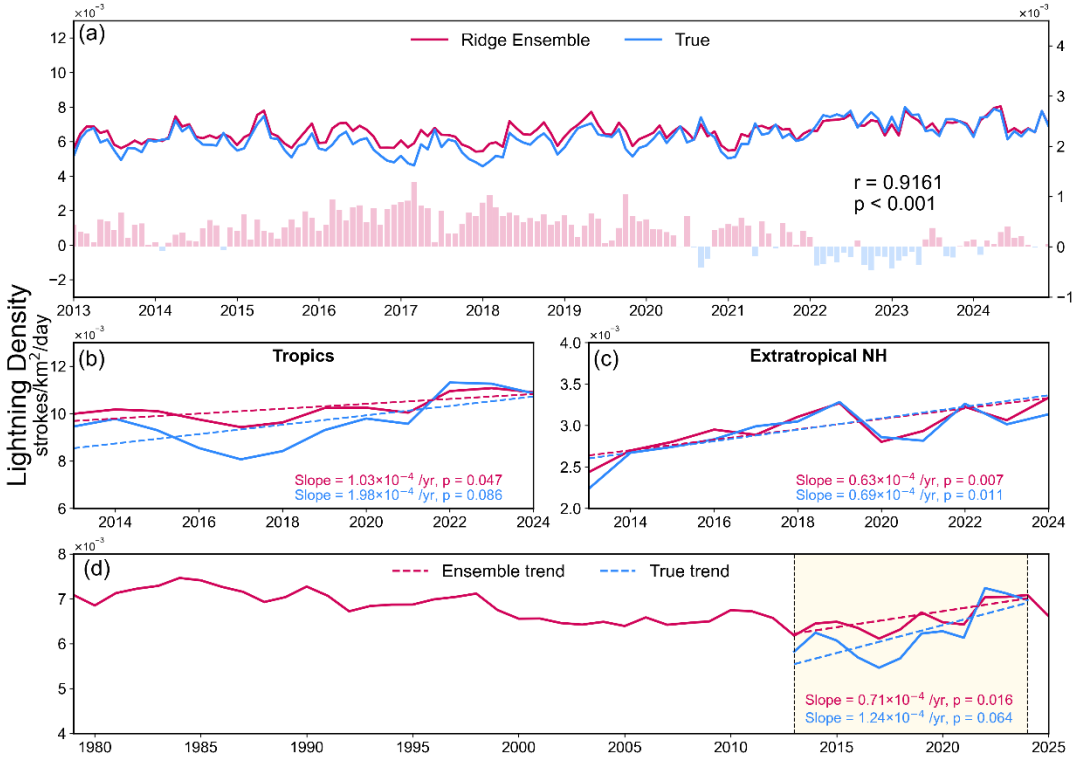

Figure 6. ***(a)*** *Monthly global mean lightning density during 2013–2024 predicted by the ridge regression ensemble and observations, with residuals shown as bars.* ***(b)*** *and* ***(c)*** *present annual mean lightning density during 2013–2024 over the tropics (|lat| ≤ 30°) and the extratropical Northern Hemisphere (30° < lat ≤ 90°), respectively.* ***(d)*** *Annual global mean lightning density from the ridge regression ensemble (1979–2025), with the observational period (2013–2024) highlighted. For panels* ***(b)–(d)****, Theil–Sen trend lines and the corresponding Mann–Kendall p-values were calculated exclusively for 2013–2024 and are shown for both the ensemble and the observations.*

Comment 10: Lines 366-367: It would be helpful to introduce the distinction between flash-rate density and stroke density at this point, rather than at line 378. This would allow readers to better understand why a qualitative comparison is more appropriate than a direct quantitative assessment.

Reply: We thank the reviewer for this helpful suggestion. We agree that the distinction between flash-rate density and stroke density should be explained when the comparison with LIS/OTD is first introduced. We therefore moved the definition of the two lightning metrics to the beginning of the relevant paragraph at Lines 456-459 and Lines 471-474: "*Because LIS/OTD reported flash-rate density, whereas WWLLN and the ridge regression ensemble trained on it represented stroke density, the two products did not quantify exactly the same lightning metric. A flash represents a complete lightning discharge and may comprise multiple individual strokes. Therefore, direct agreement in absolute magnitude was not expected.*" and "*These discrepancies were consistent with this difference in lightning metrics, as well as the distinct sampling and detection characteristics of spaceborne optical sensors and ground-based very-low-frequency (VLF) networks (Kaplan and Lau, 2021, 2022; Cecil et al., 2014; Rudlosky and Shea, 2013).* "

Comment 11: Figure 8: Since the ridge regression ensemble is trained using WWLLN data, a direct comparison between the ensemble output and LIS/OTD observations should account for the inherent differences between the WWLLN and LIS/OTD datasets. I therefore suggest including an additional panel, similar to panel (d), showing the correlation between WWLLN and LIS/OTD over their common period of availability. The corresponding discussion in the text should also be revised accordingly.

Reply: We thank the reviewer for this helpful suggestion. We agree that the ensemble–LIS/OTD comparison should be interpreted relative to the inherent agreement and differences between WWLLN and LIS/OTD. We therefore added Supplementary Fig. S5:

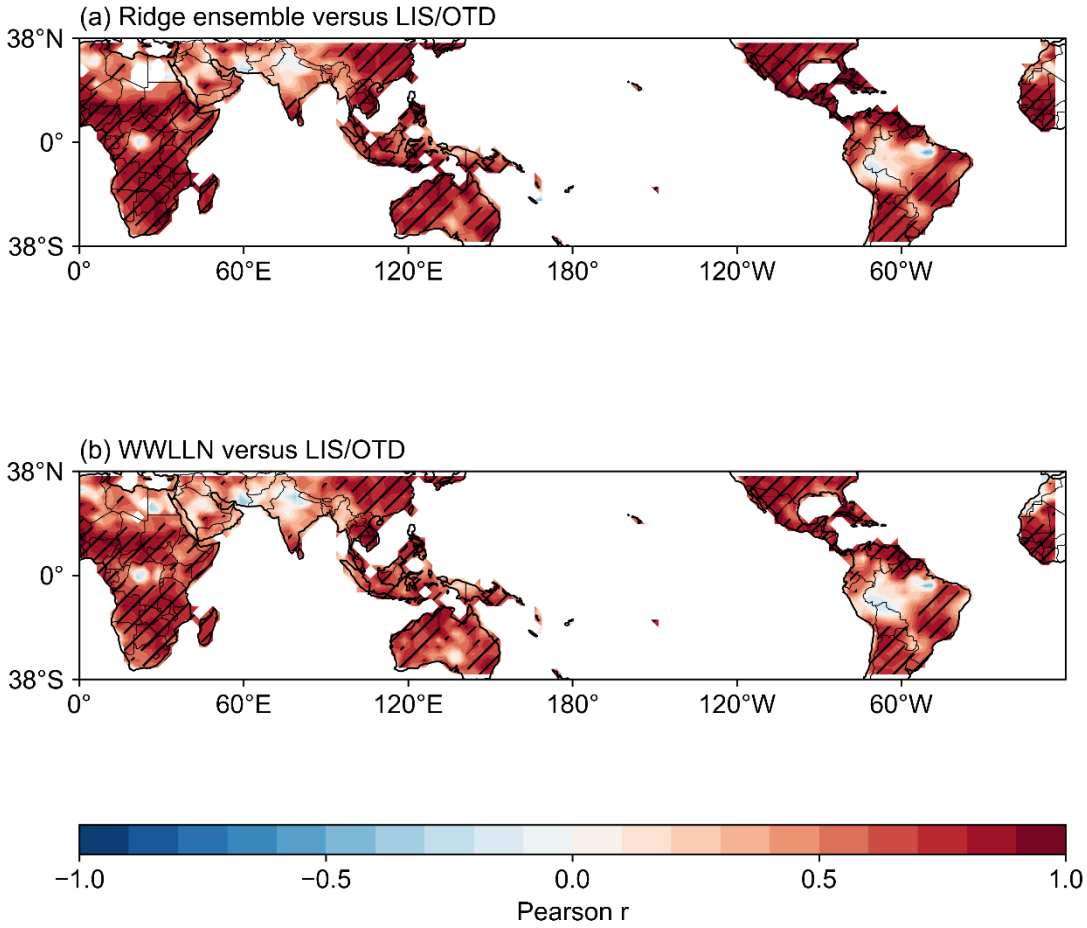

Figure S5. *Spatial distributions of grid-cell Pearson correlation coefficients calculated from monthly lightning density time series within 38°S–38°N during the common 2013–2014 period:* ***(a)*** *the ridge regression ensemble versus LIS/OTD and* ***(b)*** *WWLLN versus LIS/OTD. Hatched areas indicate correlations significant at $p<0.05$.*

We also revised Sect. 4.3 accordingly at Lines 463-466 and Lines 476-478:"*To account for the inherent differences between the two observational products, we additionally calculated grid-cell monthly correlations between the ridge regression ensemble and LIS/OTD and between WWLLN and LIS/OTD over their common 2013–2014 period (Fig. S5).*" and "*The monthly correlation maps over the common 2013–2014 period showed broadly similar regional patterns for the ridge ensemble–LIS/OTD and WWLLN–LIS/OTD comparisons (Fig. S5).*".

The two monthly correlation maps were presented together in the Supplement, while Fig. 8d retained the longer-term 1996–2014 annual comparison.

Comment 12: Figures 9 and 10: These figures are difficult to interpret because they use abbreviations for the control parameters without defining them. I recommend providing the expanded forms of these abbreviations in at least one of the figures/caption to improve readability.

Reply: We thank the reviewer for this helpful suggestion. We agree that the predictor abbreviations used in Figs. 9 and 10 should be explicitly defined. To avoid overcrowding the axes of the multi-panel SHAP summary and dependence plots, we retained the concise variable labels within the figures and added the expanded forms of all abbreviations to the captions of both Figs. 9 and 10. We also refer readers to Table S1 for the complete definitions and units of all predictors. The revised captions are provided below:

Figure 9. *Key predictors of lightning density. SHAP summary plots illustrating the top eight variables for* ***(a)*** *XGBoost and* ***(b)*** *LightGBM: cape_tp (CAPE×total precipitation), cape (convective available potential energy), tciw (total column ice water), abs_lat (absolute latitude), cp (convective precipitation), mcc (medium cloud cover), lsp (large-scale precipitation), and msl (mean sea level pressure). The colors represent the feature values (relative magnitude), while the horizontal bars denote the mean absolute SHAP values, indicating global feature importance.*

Figure 10. *SHAP dependence plots for the top eight influencing factors identified by the XGBoost model. The y-axis represented the SHAP value for a specific feature, reflecting its marginal contribution to the predicted lightning density, while the x-axis denoted the feature value. Predictor abbreviations are defined in the caption of Fig. 9. These dependence plots characterize the XGBoost base model rather than the final ridge regression ensemble.*

Comment 13: Section 5: The discussion should also address the current challenges that prevent the development of lightning density datasets at daily resolution. A monthly temporal resolution is too coarse for many studies of lightning, which is closely associated with short-lived and extreme meteorological processes. It would also be valuable to discuss whether a pathway exists for developing a daily lightning climatology using currently available satellite observations or by combining multiple satellite products.

Reply: We thank the reviewer for this valuable suggestion. We revised Sect. 5.3 at Lines 679-686 to clarify the limitation of the monthly resolution and to briefly discuss the main requirements and potential pathway for developing future daily lightning products, including higher-frequency predictors, sparse-target modeling, and harmonized satellite observations.

"*A further limitation of GLLDR v1 was its monthly temporal resolution, which could not resolve individual thunderstorms or short-lived extreme events. Developing a daily product would additionally require daily lightning observations, higher-frequency meteorological predictors, and modeling strategies designed for sparse and zero-inflated targets, such as a two-stage occurrence–intensity framework. Harmonized polar-orbiting and geostationary satellite products could support regional or satellite-era daily reconstructions as a practical first step, although producing a globally consistent, multi-decadal daily lightning climatology would remain challenging.*"

References:

Cecil, D. J., Buechler, D. E., and Blakeslee, R. J.: Gridded lightning climatology from TRMM-LIS

and OTD: Dataset description, Atmos Res, 135, 404–414, 10.1016/j.atmosres.2012.06.028, 2014.
Kaplan, J. O. and Lau, K. H. K.: The WGLC global gridded lightning climatology and time series, Earth Syst Sci Data, 13, 3219–3237, 10.5194/essd-13-3219-2021, 2021.
Kaplan, J. O. and Lau, K. H. K.: World Wide Lightning Location Network (WWLLN) Global Lightning Climatology (WGLC) and time series, 2022 update, Earth Syst Sci Data, 14, 5665–5670, 10.5194/essd-14-5665-2022, 2022.
Rudlosky, S. D. and Shea, D. T.: Evaluating WWLLN performance relative to TRMM/LIS, Geophys Res Lett, 40, 2344–2348, 10.1002/grl.50428, 2013.
Zhang, J. B., Zheng, Y., Qi, D. K., Li, R. Y., and Yi, X. W.: DNN-Based Prediction Model for Spatio-Temporal Data, 24th Acm Sigspatial International Conference on Advances in Geographic Information Systems (Acm Sigspatial Gis 2016), Artn 92
10.1145/2996913.2997016, 2016.