*Supplement of*

# Reconstruction of Global 0.25° Land Lightning Density from 1979 to 2025 based on an ensemble machine learning


## Hao Zheng et al.

*Correspondence to:* Jun Wang (wangjun@nju.edu.cn)

**Table S1.** Description of input variables.

| Feature name | Description | Unit |
|---|---|---|
| cape | CAPE | J $kg^{-1}$ |
| cape_tp | CAPE * total precipitation | / |
| kx | K-index | K |
| tp | total precipitation | m |
| cp | Convective precipitation | m |
| lsp | Large scale precipitation | m |
| lcc | Low cloud cover | / |
| mcc | Middle cloud cover | / |
| hcc | High cloud cover | / |
| t2m | 2m temperature | K |
| msl | mean sea level pressure | Pa |
| u100 | 100 metre U wind component | m $s^{-1}$ |
| v100 | 100 metre V wind component | m $s^{-1}$ |
| tciw | Total column cloud ice water | kg $m^{-2}$ |
| tclw | Total column cloud liquid water | kg $m^{-2}$ |
| viiwd | Total column vertically-integrated divergence of cloud frozen water flux | kg $m^{-2}$ $s^{-1}$ |
| vilwd | Total column vertically-integrated divergence of cloud liquid water flux | kg $m^{-2}$ $s^{-1}$ |
| viiwe | Total column vertically-integrated eastward cloud frozen water flux | kg $m^{-1}$ $s^{-1}$ |
| vilwe | Total column vertically-integrated eastward cloud liquid water flux | kg $m^{-1}$ $s^{-1}$ |
| viiwn | Total column vertically-integrated northward cloud frozen water flux | kg $m^{-1}$ $s^{-1}$ |
| vilwn | Total column vertically-integrated northward cloud liquid water flux | kg $m^{-1}$ $s^{-1}$ |
| totalx | Total totals index | K |
| lai_lv | Leaf area index, low vegetation | / |
| lai_hv | Leaf area index, high vegetation | / |
| q850 | Specific humidity (850 hPa) | kg $kg^{-1}$ |
| w | Vertical velocity (500 hPa) | Pa $s^{-1}$ |
| mrh600_800 | Mean relative humidity between 600 and 800 hPa | % |
| abs_lat | Absolute latitude | / |
| month_sin | $\sin(2\pi m/12)$, m being the month of the year | / |
| month_cos | $\cos(2\pi m/12)$, m being the month of the year | / |
| lightning_density | Lightning stroke density | strokes $km^{-2}$ $day^{-1}$ |

**Table S2.** Five-fold cross-validation evaluation metrics for the four base models (XGBoost, RF, LightGBM, and DNN) during 2013–2024. Values are reported as the mean and standard deviation across the five validation folds.

| Model | $R^2$ | | RMSE | | MAE | |
|---|---|---|---|---|---|---|
| | mean | std | mean | std | mean | std |
| XGBoost | 0.6498 | 0.0010 | 0.0114 | 0.0000 | 0.0034 | 0.0000 |
| RF | 0.6519 | 0.0018 | 0.0114 | 0.0000 | 0.0030 | 0.0000 |
| LightGBM | 0.6872 | 0.0009 | 0.0108 | 0.0000 | 0.0032 | 0.0000 |
| DNN | 0.6344 | 0.0022 | 0.0117 | 0.0000 | 0.0036 | 0.0000 |

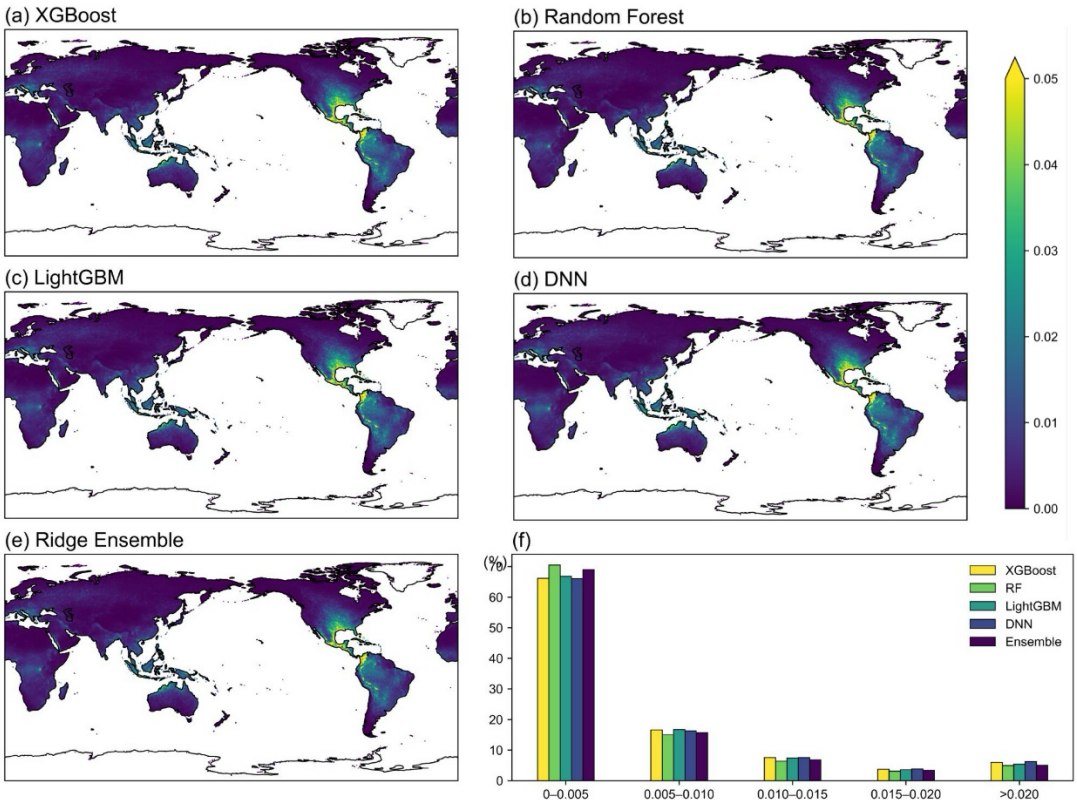


**Figure S1.** Spatial distribution of the RMSE for four base models **(a)** XGBoost, **(b)** RF, **(c)** LightGBM, **(d)** DNN and **(e)** ridge regression ensemble. The bar chart **(f)** shows the percentage of global grid cells falling within different RMSE intervals for each model.

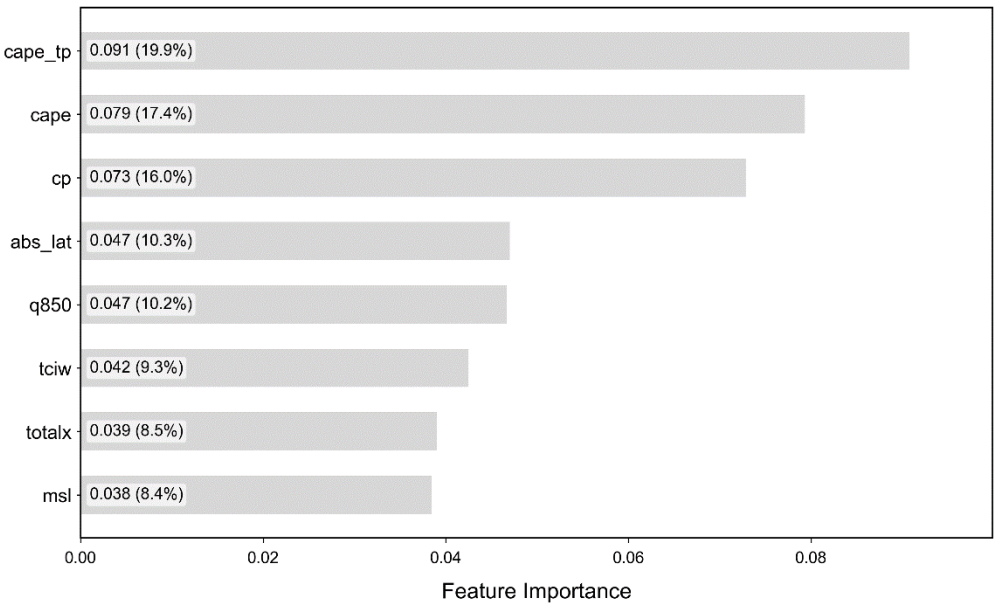


**Figure S2.** RF feature importance ranking for lightning-density prediction.

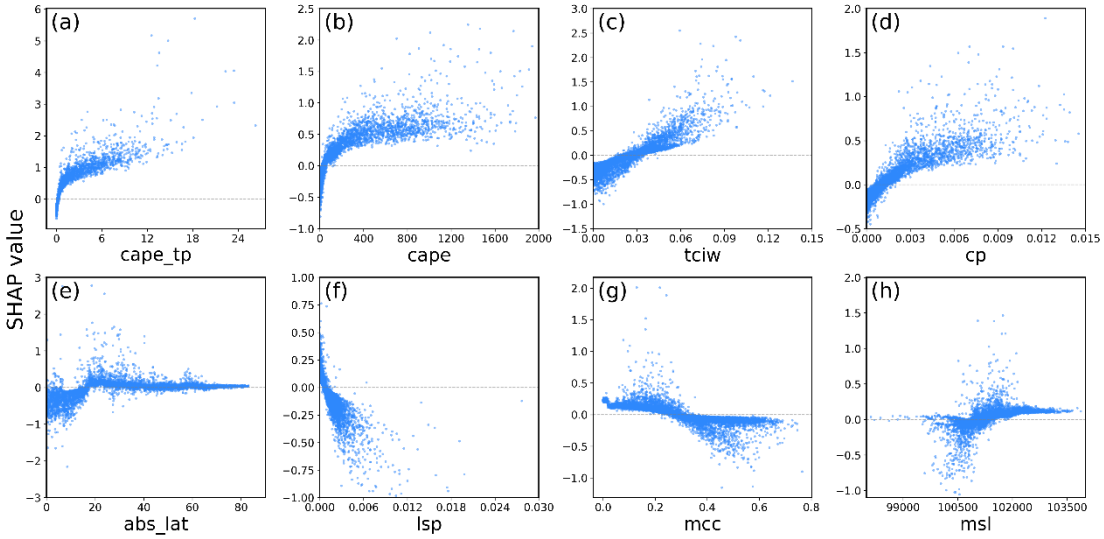


**Figure S3.** SHAP dependence plots for the top eight influencing factors identified by the LightGBM model.