Summary

This manuscript presents a 30 m annual forest litterfall production dataset for China covering 2000–2024. The authors compiled ground-based litterfall observations from the Chinese Ecosystem Research Network (CERN) and published literature and developed a spatial matching approach to account for coordinate uncertainty. Landsat-derived spectral variables, TerraClimate climatic variables, and topographic variables were integrated using a Random Forest model. After quality control, 384 annual litterfall records were retained for model development. The model achieved a testing R^2 of 0.72 based on ten-fold cross-validation. The resulting dataset was further used to investigate the spatial patterns, temporal trends, and climatic associations of forest litterfall production across China.

The study is relevant to *Earth System Science Data*, and the long-term, high-resolution dataset has potential value for forest carbon and nutrient cycling studies and ecosystem modeling. However, several methodological issues require further clarification or analysis, as detailed below.

Major issues

1. Insufficient assessment of the representativeness of ground observations for nationwide spatiotemporal prediction. After all filtering and quality-control procedures, only 384 annual litterfall records were retained for model development. Given that the resulting product provides annual estimates for all forested areas of China at 30 m resolution over a 25-year period (2000–2024), the representativeness of the ground observations is critical for assessing the reliability and applicability of the dataset. However, the manuscript currently provides only limited information on the composition and coverage of these 384 records. The authors should provide a more detailed summary of the final dataset, including the number of unique sites and the distribution of observations across years, forest types, climate zones, and major geographic regions. This information is necessary to assess the extent to which the training data represent the spatial and temporal domain of the final product.

2. **Lines 137–141:** Two aspects of the treatment of incomplete monthly observations require further justification. First, why are records with more than six months of observations considered sufficient for estimating annual litterfall? Given the strong seasonality of litterfall, the resulting annual estimate may depend strongly on which months are missing. Second, is it appropriate to fill missing months using observations from the same month in the nearest available year? Litterfall can vary substantially among years, and such cross-year substitution may alter the actual interannual variability. The authors should justify these choices and report how many annual records and monthly values were affected by this procedure.

3. The manuscript reports R^2, RMSE, and MAE from ten-fold cross-validation, but these metrics do not characterize the spatial uncertainty of the final 30 m litterfall product. Given the limited field observations and nationwide extrapolation over 25 years, prediction uncertainty may vary substantially across space. I suggest that the authors provide spatially explicit uncertainty estimates, such as prediction intervals, bootstrap-based uncertainty, or Random Forest ensemble variability. Uncertainty should also be considered when reporting national estimates and long-term trends. At minimum, the manuscript should discuss where predictions are likely to be less reliable and should be interpreted with caution.

4. The choice of Random Forest could be better justified. Random Forest is used as the sole modeling algorithm, mainly because of its ability to capture complex and nonlinear relationships. However, the rationale for selecting RF over other commonly used machine-learning approaches is not sufficiently discussed. The authors should provide a clearer justification for this choice. If feasible, a comparison with alternative algorithms would further demonstrate the robustness of the model selection.

Minor issues

1. **Figure 2:** The training set is reported as n=3456, whereas only 384 annual records were retained for model development. This value appears to result from pooling the training samples across the ten cross-validation folds 384×9=3456, meaning that the same observations are included multiple times across the pooled training folds rather than representing 3456 independent samples. Please clarify this in the figure caption. In addition, the test results show an apparent compression of the predicted range, particularly an underestimation of relatively high litterfall values. This pattern should be briefly discussed.

2. **Lines 135–140:** Please report the number of observations removed or retained at each major QC step. This would make the filtering process leading to the final 384 records more transparent.

3. **Lines 170–179:** Please provide a table listing all predictor variables, their original spatial/temporal resolutions, temporal aggregation methods, and final units. This would substantially improve reproducibility.

4. **Figure 4:** The caption refers to the percentage contribution of each "litter component," whereas the plotted categories appear to be forest types (DNF, ENF, DBF, and EBF). "Forest type" may therefore be more appropriate than "litter component."

5. Several typographical and formatting errors are present throughout the manuscript, for example "In this study,we" (**Lines 110–120**), "1.4T g yr^{-1}" (**Lines 225–245**), and "litterfall prodcution dataset" (**Section 5.1**). Please carefully proofread the manuscript and correct these and other similar formatting, spacing, and spelling errors.