This study uses GOES-16 and GOES-17 ABI observations and MLP- and Transformer-based models to post-process NOAA ABI and NASA Dark Target AOD products. AERONET AOD is used as the reference within a spatio-temporal cross-validation framework. The fusion of dual-satellite observations and the use of intra-day geostationary time series are potentially useful, and the data-partitioning strategy is relatively elaborate.

However, the manuscript does not sufficiently establish the necessity of residual correction relative to direct machine-learning retrieval. The selection of the Transformer and its comparison with the MLP are also not adequately justified. In addition, the key conclusions regarding spatial generalization, the physical meaning of correction under cloud contamination, spatial-resolution enhancement, and the general applicability of the method require more targeted experiments and discussion.

Major comments

1. A table should be added to present the model inputs and outputs more clearly

Although Table A1 lists some input variables, a summary table should be included in the main text. It should specify the inputs, input dimensions, sequence length, quality-flag treatment, and outputs for the single- and dual-satellite models. The authors should clearly define whether the model predicts an AOD residual, an AOD correction, or the final corrected AOD, and provide the corresponding equations.

2. The necessity of correcting an already retrieved product has not been established

Is correction required because of inherent limitations in the original retrieval algorithm, or is it simply used to improve retrieval accuracy? If the sole objective is to improve accuracy, why not directly train a machine-learning model to predict the final AOD rather than learning residual errors to compensate for the original algorithm?

A direct AOD model using the same ABI reflectances, geometrical variables, and auxiliary inputs should be included and compared with the residual-correction approach. Otherwise, the manuscript does not demonstrate that using the original AOD product as an input is necessary.

3. The choice of the Transformer and the fairness of the comparison are insufficiently justified

Given that time-series data are used, why was a Transformer selected instead of commonly used sequence models such as LSTM or GRU? The manuscript does not sufficiently explain this choice.

More importantly, the Transformer receives a complete time series, whereas the MLP receives only a single time step. The current comparison therefore cannot distinguish whether the improvement comes from temporal information or from the Transformer architecture. An LSTM or GRU using the same sequence input should be added, together with ablation experiments separating the contributions of temporal context and model architecture.

4. Comparisons with a broader range of machine-learning models are needed

An MLP alone is not a sufficient baseline for demonstrating the advantage of the Transformer. Representative models such as random forest, XGBoost, or LightGBM should be considered, together with at least one sequence model such as LSTM or GRU. Comparable input information and hyperparameter-search budgets should be used.

5. The current data partitioning cannot fully eliminate spatial similarity between training and test data

Clustering AERONET stations within 25 km reduces direct leakage between nearby stations, but different folds within CONUS may still contain similar climate, surface, and aerosol conditions.

The authors should report the distance from each test station to the nearest training station and conduct a stricter continuous-region holdout experiment, for example based on eastern and western CONUS, climate zones, or larger spatial blocks. Otherwise, the claims regarding spatial generalization should be moderated.

6. Overall bias should be reported, and high-AOD underestimation requires further discussion

The main results include correlation, RMSE, and MAE, but not overall bias. Bias should be added to the principal evaluation.

Figure 6 shows that the correction reduces overestimation at low AOD but also causes substantial underestimation at high AOD. This behaviour should not be attributed only to the limited number of high-AOD samples. Sample size, bias, and RMSE should be reported for separate AOD intervals, and the possible effects of MSE loss and residual-learning range compression should be discussed.

7. The anomalous fifth CV split in Figure 4 is not explained

The original NOAA ABI AOD shows substantially higher RMSE and MAE and lower correlation in the fifth spatial split. The authors should describe the corresponding region, stations, sample size, AOD distribution, quality flags, and land-cover composition, and determine whether the overall conclusions are influenced by a small number of anomalous stations or regions.

8. Spatial differences before and after correction should be presented

Figure 7 only shows station-level RMSE and does not illustrate how the complete satellite AOD field is modified. Representative cases should show the original AOD, corrected AOD, and correction magnitude.

The authors should also examine whether the corrected fields contain nonphysical discontinuities, stripes, or block-like structures, and provide seasonal or multi-year mean correction maps where appropriate.

9. The physical meaning of correction under cloud contamination requires further clarification

When non-negligible cloud contamination is detected, does the original AOD retrieval still satisfy the basic assumptions required for a physically meaningful retrieval? If the original AOD is invalid, the model output may represent a statistical estimate based on the training distribution rather than a correction of valid satellite AOD.

The physical meaning of Aerosol_Cloud_Fraction_Land should be clearly defined. Near-cloud aerosol enhancement, thin-cloud contamination, and invalid retrievals should be distinguished. Performance should be evaluated separately according to cloud fraction, cloud distance, and quality flag.

10. The general applicability and main contribution of the method are overstated

The method is evaluated only for GOES-16 and GOES-17, which carry highly similar ABI instruments, and only over CONUS. Therefore, the conclusion that the method can generally correct satellite AOD retrievals is too broad.

The authors should clarify whether the main contribution is dual-satellite fusion, temporal modelling, or residual correction. Residual-based correction itself is not a new general methodology. If additional sensors or regions cannot be tested, the claims in the abstract and conclusions should be substantially narrowed.

11. The claimed spatial-resolution enhancement or downscaling has not been demonstrated

Reprojecting a 2 km or 10 km AOD product onto a 500 m grid using nearest-neighbour interpolation does not demonstrate that genuine 500 m AOD information has been retrieved. The effective spatial resolution of the output should be clarified and evaluated using cases with clear spatial gradients.

Otherwise, the terms “spatial-resolution enhancement” and “downscaling” should be replaced by a more accurate description, such as generating post-processed estimates on a 500 m grid.

12. Reprojection and spatial collocation may introduce pseudo-replication

Nearest-neighbour resampling may replicate one native 2 km or 10 km AOD pixel into multiple nominal 500 m samples. The authors should clarify whether multiple training samples originate from the same native satellite pixel.

In addition, matching AERONET to a single 500 m grid cell may not represent the native footprint of the 2 km or 10 km product. The sensitivity to nearest-pixel matching, spatial averaging, and native-pixel matching should be evaluated.

13. Performance should be evaluated separately for each product quality category

The main results include all quality flags, and the large improvement may be dominated by medium- and low-quality retrievals. Bias, RMSE, MAE, and correlation should be reported separately for high-, medium-, and low-quality data. The authors should also determine whether the correction degrades any originally high-quality AOD retrievals.

Minor comments

1. The phrase “in the 10 years 2020–2022” is incorrect. The period covers three years and should be corrected.
2. The statement “Atmospheric aerosols, also referred as Particulate Matter” is imprecise, because atmospheric aerosols and particulate matter in the regulatory air-quality context are not fully equivalent concepts.
3. The terms “post-process correction,” “bias correction,” “enhancement,” and “downscaling” should be clearly defined and used consistently throughout the manuscript.
4. The numbers of NOAA and NASA DT samples should be reported by year, spatial fold, quality category, and AOD interval to clarify the representativeness of the experiments.
5. The expression “MLP transformer” in the captions of Figures 8 and 9 should be checked and likely replaced by “Transformer.”
6. The limitations should explicitly state that the combined model uses only samples for which valid data from both satellites are available, that the evaluation is restricted to CONUS, and that operation when one satellite is unavailable has not been validated.