
Response to Reviewer #1

Manuscript ID: egusphere-2026-2510

Title: A blue-band surface residual correction method for land aerosol optical depth retrieval from DQ-1 WSI

Journal: Atmospheric Measurement Techniques

Dear Reviewer,

We thank the reviewer for the careful evaluation and constructive comments. In response, we revised the manuscript and Supplement to clarify the role of MCD19A2 AOD in the residual-learning framework, expand the independent AERONET evaluation, and provide additional details on the retrieval, quality-control, validation, and post-processing procedures. The revised AERONET evaluation now comprises 104 station–date matchups across spring 2025, autumn–winter 2025, and spring 2026.

Our detailed responses are provided below. For clarity, the reviewer’s comments are shown in **black** font, our responses in “***ITALIC BOLD BLUE***” font, and quoted revisions to the manuscript in “***Red***” font.

General comment

This manuscript presents a blue-band surface residual correction method for land aerosol optical depth (AOD) retrieval from DQ-1 WSI, combining machine-learning-based correction of the surface reflectance prior with a physically based LUT retrieval framework. The topic is relevant to AMT, and the reported results show a clear improvement over the static-prior retrieval. The manuscript is generally well structured, but several aspects of the methodology and validation require further clarification, particularly the role of MCD19A2 AOD in constructing the RF training target, the limited independent AERONET validation, and some details of the retrieval and post-processing procedures. These issues are important for assessing the robustness and reproducibility of the method, but do not appear to undermine the overall retrieval framework. I believe the manuscript can be considered for publication after the issues below are adequately addressed. My specific comments are listed below.

Response: *We thank the reviewer for the positive assessment of the manuscript and for the constructive comments. We have revised the manuscript and Supplement to clarify the role of MCD19A2 AOD in the residual-learning framework, expand the independent AERONET evaluation, and provide additional details on the retrieval, quality-control, and post-processing procedures. The specific revisions are described in our point-by-point responses below.*

Specific comments

1. The RF model does not directly predict AOD, but the residual target used for RF training is constructed using MCD19A2 AOD. The

manuscript should therefore distinguish more carefully between “not directly predicting AOD” and “being independent of external AOD information”. Some statements regarding the independence of the proposed approach from external AOD constraints appear somewhat strong and should be revised accordingly.

Response: *Thank you for pointing this out. We have revised the manuscript to distinguish more clearly between the offline construction of the RF training target and the online AOD retrieval. MCD19A2 AOD is used during the offline LUT inversion to derive the reference surface term and therefore contributes indirectly to the RF training target. However, it is not included among the RF input features and is not required during online retrieval. We have also revised the relevant statements in the manuscript to avoid implying that the proposed approach is completely independent of external AOD information.*

Changes in the manuscript:

The corresponding text in Sect. 3.4 now reads:

“MCD19A2 AOD is used in the offline derivation of the LUT-derived reference surface term and therefore contributes indirectly to the RF training target. It is not included among the RF input features and is not required during online retrieval.”

2. The independent AERONET validation contains only 27 station–date matchups. Although the reported improvement relative to the static-prior retrieval is substantial, this sample size remains relatively limited for assessing the overall robustness and general applicability of the method. The authors should state this limitation more explicitly and interpret the validation results with appropriate caution. If expanding the validation dataset is not feasible, the manuscript should also clarify the practical

reasons for the limited number of matchups and more clearly define the conditions over which the current validation supports the retrieval performance.

Response: *We agree that the original 27 spring-2025 matchups provided limited temporal coverage. We therefore expanded the independent AERONET evaluation to 104 station–date matchups across three periods: 27 in spring 2025, 49 from two sites spanning 26 independent observation dates in autumn–winter 2025, and 28 from three sites spanning 17 independent dates in spring 2026. The autumn–winter evaluation used a season-specific static prior together with a seasonally adjusted LUT and RF model, whereas the spring-2026 evaluation used the spring-2025 configuration without retraining or recalibration. The limited number of spring-2025 matchups results partly from the requirement that both DQ-1 WSI retrieval schemes be valid for the same station–date sample, which ensures a direct comparison but reduces the available sample size. We have also revised the Discussion and Conclusions to limit the interpretation to the tested BTH conditions and to avoid implying year-round or cross-regional transferability.*

Changes in the manuscript:

Section 4.1 now includes the additional validation results:

“Figure 5 shows the AERONET validation results for autumn–winter 2025. After excluding scenes used for RF training and held-out internal validation, 49 station–date matchups from two AERONET sites were retained, covering 26 independent observation dates.”

“To examine interannual performance, the spring-2025 retrieval configuration was applied to spring 2026 without using 2026 observations for retraining or recalibration. The evaluation included 28 station–date matchups from three AERONET sites spanning 17 independent observation dates.”

The manuscript also clarifies the common-sample requirement:

“A station–date pair was retained only when both DQ-1 WSI retrievals were valid. This strategy ensures consistent sampling between the two retrievals, although it also

constrains the final number of matchups.”

The Conclusions now state:

“These results do not establish year-round or cross-regional transferability. Further evaluation should therefore target additional surface and aerosol regimes, BRDF effects, and longer independent observational records outside BTH.”

3. Please clarify what is meant by “haze” in the cloud–haze screening procedure. Since aerosol haze is itself the target of AOD retrieval, it is important to explain how aerosol-loaded but otherwise valid observations are distinguished from contamination that should be masked. The authors should also clarify whether this screening could preferentially affect the retention of high-AOD cases.

Response: *We have clarified the meaning of “haze” in the QA terminology. In this study, “haze” refers to semi-transparent cloud-like contamination identified from combined spectral, thermal, and spatial characteristics rather than ordinary aerosol-loaded clear-sky conditions. AOD itself is not used as a screening criterion, and high-AOD observations are therefore not explicitly excluded. We nevertheless note that aerosol-loaded pixels showing cloud-like spectral, thermal, or spatial characteristics may still be removed by the QA procedure.*

Changes in the manuscript:

Section 2.4 now states:

“For QA purposes, ‘haze’ refers to pixels showing cloud-like spectral, thermal, or spatial characteristics rather than aerosol-loaded clear-sky conditions. Because AOD itself is not used as a screening criterion, high-AOD observations are not explicitly excluded. However, aerosol-loaded pixels showing cloud-like spectral, thermal, or spatial characteristics may still be removed by the QA procedure.”

A corresponding clarification has also been added below Table S1 in the Supplement.

4. Section 3.2: The manuscript considers both continental and urban aerosol models. Please indicate which aerosol model was selected for the main retrieval cases and clarify how the scene-level model selection was implemented. The selection criterion should be stated explicitly so that this part of the retrieval procedure can be reproduced.

Response: *We have clarified the use and selection of the two predefined aerosol models. The continental model was used for the offline construction of the static surface prior and LUT-derived reference surface terms. During the final online retrieval, the continental and urban models were both retained as candidates. Retrieval was performed separately under the two models, and the model with the lower median matching cost over all valid pixels was selected at the scene scale. We also added an AERONET-based comparison of the normalized spectral AOD dependence of the two predefined aerosol models to provide regional context for their representativeness.*

Changes in the manuscript:

Section 3.2 now states:

“The continental model was used for offline construction of the static surface prior and LUT-derived reference surface terms, whereas both models were retained for the final online retrieval.”

Section 3.4 further specifies the scene-level selection criterion:

“Retrieval was performed separately under the continental and urban aerosol models, and the model with the lower median $J_{i,m}$ over all valid pixels was selected at the scene scale.”

5. In the comparison with MCD19A2, collocated pixels are excluded when $AOD < 0.05$ in either product. Please justify this threshold and report approximately how many samples are removed by this criterion. Since excluding low-AOD cases may influence the regression slope, bias,

and other reported statistics, the possible effect of this filtering on the comparison should be clarified.

Response: *The $AOD < 0.05$ criterion was used to exclude very low- AOD collocations from the cross-product comparison. We have now quantified the number of samples affected by this criterion and added a sensitivity comparison in the Supplement. The screen directly removed 1,951 collocated pixels, corresponding to 0.96 % of the samples entering this step.*

We also recalculated the cross-product statistics without the $AOD < 0.05$ exclusion. The resulting changes in R , $RMSE$, MAE , MBE , slope, and intercept were very small, indicating that the reported cross-product agreement is not sensitive to this screening criterion.

Changes in the manuscript:

Section 4.4 now states:

“Pixels with AOD values below 0.05 in either product were excluded; this screen removed 0.96 % of the collocated samples, and omitting it produced only minor changes in the comparison statistics (Table S8).”

Table S8 in the Supplement provides the corresponding sensitivity results. With the current screening, $R=0.8629$, $RMSE = 0.1311$, $MAE = 0.0926$, $MBE = 0.0120$, slope = 0.9939, and intercept = 0.0141; without the $AOD < 0.05$ exclusion, the corresponding values are 0.8633, 0.1309, 0.0925, 0.0126, 0.9904, and 0.0158.

6. Section 4.1 / Table S3: The AERONET temporal matching window is relatively broad compared with the DQ-1 overpass time. Please justify the choice of this window and briefly discuss the possible influence of temporal aerosol variability on the validation statistics. Given the limited number of station–date matchups, it would also be useful to comment on the potential sensitivity of the reported performance to the temporal collocation criterion.

Response: *We have added the nominal DQ-1 WSI overpass time and clarified the reason for using the 02:00–06:00 UTC window for the spring validation. The relatively broad window was used to retain sufficient AERONET observations for the available DQ-1 WSI overpass dates. We also explicitly acknowledge that temporal aerosol variability within this interval may introduce additional uncertainty into the validation statistics.*

Because the number of independent spring matchups is already limited, further narrowing the temporal window would reduce the available validation sample. We therefore retained the original spring collocation window and have stated this limitation explicitly in the revised manuscript.

Changes in the manuscript:

Section 2.3 now states:

“The nominal DQ-1 WSI overpass time over the study region is approximately 13:30 Beijing time (approximately 05:30 UTC). The 02:00–06:00 UTC window was used to retain sufficient AERONET observations for the available DQ-1 WSI overpass dates. Temporal aerosol variability within this window may introduce additional uncertainty into the validation statistics.”

7. The reference-date switch ratio is defined relative to the default value $\alpha = 0.06$. Therefore, the minimum value at $\alpha = 0.06$ follows, at least in part, from the definition itself and does not by itself provide independent evidence supporting the choice of this parameter. The corresponding interpretation should be revised, and the authors should clarify what additional criterion was used to justify $\alpha = 0.06$.

Response: *We agree with this point and have revised the interpretation accordingly. The reference-date switch ratio is now used only as a descriptive sensitivity metric and is no longer interpreted as independent evidence supporting $\alpha = 0.06$.*

No independent optimization criterion was used to determine $\alpha = 0.06$. The value was adopted empirically to balance the brightness and vegetation-state consistency

terms, and the sensitivity analysis is used only to characterize the behavior of the reference-date selection around this setting.

Changes in the manuscript:

Section 3.3 now states:

“Here, $\alpha = 0.06$ was adopted empirically to balance the brightness and vegetation-state consistency terms.”

and:

“The reference-date switch ratio is defined relative to $\alpha = 0.06$ and is therefore used only as a descriptive measure of changes in reference-date selection.”

8. Please provide a short physical explanation for the brightness term in the reference-scene selection score. It is not immediately clear why a relatively brighter candidate surface state should provide a more reliable static surface reference. Since this term directly affects reference-scene selection, its physical or empirical basis should be made more explicit.

Response: *We have clarified that the brightness term is an empirical component of the reference-date selection score rather than a strict physical constraint. Within the quality-controlled candidate scenes, this term favors relatively high retrieved surface-reflectance states and reduces the selection of anomalously low surface solutions. The multi-band mean is used to reduce the influence of band-specific noise and outliers.*

Changes in the manuscript:

Section 3.3 now states:

“Brightness represents the overall retrieved surface-reflectance level of a pixel on a candidate date and was used as an empirical term in reference-date selection. Within the quality-controlled candidate scenes, this term empirically favors relatively high retrieved surface-reflectance states and reduces the selection of anomalously low surface solutions. The multi-band mean reduces the influence of band-specific noise and outliers.”

9. Section 3.3: $\text{AOD} \leq 0.5$ is treated as a low-AOD condition for construction of the surface reference. This threshold appears relatively high, particularly for a retrieval that relies strongly on the blue band, where aerosol effects can remain substantial at such AOD levels. Please provide a clearer justification for this choice and discuss whether residual aerosol effects under these conditions could influence the constructed surface reference.

Response: *We have clarified the role of the AOD threshold. The 0.01–0.50 range is used to limit the influence of heavier aerosol loading rather than to assume that aerosol effects below 0.50 are negligible. For all retained candidate pixels, the collocated MCD19A2 AOD remains an atmospheric constraint in the LUT inversion, so aerosol effects within this range are explicitly accounted for when deriving the reference surface term.*

Changes in the manuscript:

Section 3.3 now states:

“Candidate pixels were restricted to MCD19A2 AOD values of 0.01–0.50 to limit the influence of heavier aerosol loading on surface-term estimation. The collocated MCD19A2 AOD was retained as an atmospheric constraint in the LUT inversion to account for aerosol effects within this range. Thus, aerosol effects at $\text{AOD} \leq 0.50$ were not treated as negligible.”

10. Section 3.2: Please provide some additional information on the implementation of the 6S simulations and LUT. In particular, please clarify whether the actual DQ-1 WSI spectral response functions were used rather than only the nominal central wavelengths. The manuscript also mentions an additional gaseous-absorption correction applied to the

LUT; the implementation of this correction should be described in sufficient detail to allow the retrieval procedure to be reproduced.

Response: *We have expanded the description of the 6S/LUT implementation. The simulations were performed at the nominal WSI central wavelengths and did not explicitly convolve the measured WSI spectral response functions. This is now stated directly in the manuscript.*

We have also clarified the atmospheric profile and gaseous-absorption treatment. Gaseous absorption was treated within 6S using the gaseous contents specified by the selected atmospheric profile. The downward and upward scattering transmittances stored in the LUT were obtained from the 6S total-scattering transmittances, and the corresponding directional global-gas transmittances were applied in the TOA reflectance calculation. A sensitivity test was additionally included in the Supplement to quantify the influence of this treatment.

Changes in the manuscript:

Section 3.2 now states:

“The 6S simulations were performed at the nominal WSI central wavelengths listed in Table 2 without explicit convolution with the measured spectral response functions.”
and:

“Gaseous absorption was treated within 6S using the gaseous contents specified by this profile, without additional external gas-concentration data or absorption cross sections. The downward and upward scattering transmittances stored in the LUT were obtained from the total-scattering transmittances provided by 6S, and the corresponding directional global-gas transmittances were applied to obtain the transmittances used in the TOA reflectance calculation.”

The corresponding sensitivity analysis is provided in Table S7. The median absolute change in simulated 0.443 μm TOA reflectance was 6.52×10^{-5} , the median absolute AOD change was 0.002, and 97.1 % of the test cases showed $|\Delta\text{AOD}| \leq 0.02$.

11. Section 4.2: Please clarify whether the 28 April 2025 scene was used

at any stage of hyperparameter tuning, feature selection, model selection, or other optimization. If it was used for any of these purposes, it would be more appropriate to describe it consistently as an internal validation scene rather than as a fully independent test case. This distinction is important when interpreting the reported generalization performance.

Response: *We have clarified the role of the 28 April 2025 scene. It was reserved as a held-out internal validation scene and was not used for static-prior construction, RF training, feature selection, or hyperparameter tuning. The spring RF training scenes were 22 March, 22 April, and 19 May 2025. We now describe the 28 April scene consistently as held-out internal validation rather than as a fully independent external test case.*

Changes in the manuscript:

Section 3.4 now states:

“For the spring configuration, the three reference scenes used for static-prior construction in Sect. 3.3 were also used to generate RF residual-training samples. The scene acquired on 28 April 2025 was reserved for held-out internal validation and was not used for static-prior construction, RF training, feature selection, or hyperparameter tuning.”

12. Several empirical thresholds are applied to the training residuals, predicted residuals, and corrected surface reflectance. Please explain the rationale for these thresholds more clearly. It would also be useful to report the fraction of retrievals affected by the clipping or bounding operations. If a non-negligible fraction of pixels is constrained, the possible influence of these empirical limits on the final AOD retrieval should be discussed.

Response: *We have clarified the roles of the empirical thresholds in the revised*

manuscript and Supplement. The absolute residual threshold of 0.12 is used only during offline residual-sample construction to remove extreme discrepancies between the LUT-derived reference surface term and the static prior. During online retrieval, the predicted residual is constrained to $[-0.10, 0.10]$ as a safeguard against extreme RF extrapolation, and the corrected blue-band surface term is constrained to $[0, 0.8]$, consistent with the admissible range used for the LUT-derived reference surface term.

We also examined the fractions of valid retrieval pixels affected by the two online constraints. Neither constraint was activated in the evaluated retrievals (0 % for both), indicating that these bounds acted only as safeguards and did not affect the reported retrieval results. This information has now been added below Table S2 in the Supplement.

Changes in the manuscript:

Section 3.4 now explains the rationale for the online constraints:

“Specifically, the RF-predicted residual was limited to $[-0.10, 0.10]$ as a safeguard against extreme model extrapolation, and the corrected blue-band surface term was constrained to $[0, 0.8]$, consistent with the admissible range used for the LUT-derived reference surface term.”

The threshold definitions and their roles are also summarized in Table S2 of the Supplement.

13. The manuscript discusses the relative importance of the RF input variables. Please clarify how the feature importance shown in Fig. 3 was calculated. Since some of the predictors may be correlated, a brief note on how the importance values should be interpreted would also be helpful.

Response: *We have clarified both the calculation and interpretation of the RF feature importance. The values shown in Fig. 3 are the impurity-based feature importance, i.e. mean decrease in impurity (MDI), from the trained RF model. We have also added a caution that correlated predictors may share or redistribute MDI*

importance. The reported values are therefore interpreted as descriptive measures of the fitted residual model rather than as causal effects.

Changes in the manuscript:

Section 3.4 now states:

“To further illustrate how the residual-learning model uses the input information, the relative feature importance was calculated using the impurity-based feature importance (mean decrease in impurity, MDI) of the trained random forest model, as shown in Fig. 3. The static blue-band surface prior, solar zenith angle, and blue-band TOA reflectance had the largest relative importance values, accounting for 28.3 %, 24.4 %, and 19.5 %, respectively. Viewing zenith angle and NDVI also contributed to the prediction, whereas relative azimuth angle and elevation had lower importance values. Previous studies have shown that feature-importance analysis can help characterize how machine-learning models use input variables and improve model interpretability in Earth-science applications (Roscher et al., 2020). Because correlated predictors may share or redistribute MDI importance, the values reported here are interpreted as descriptive measures of the fitted residual model rather than as causal effects.”

14. In the MCD19A2 comparison, please provide a quantitative and reproducible definition of the “residual cloud-contamination outliers” removed before calculating the cross-product statistics. The current description is somewhat subjective. The number or fraction of samples excluded by this procedure should also be reported, since outlier removal may affect the reported correlation, regression, and error statistics.

Response: *We have replaced the qualitative description with an explicit and reproducible criterion. Probable residual-cloud outliers were defined as pixels satisfying both $DQ-1\ AOD > 1.0$ and $DQ-1\ AOD - MCD19A2\ AOD > 0.8$. This criterion removed 1,621 samples, corresponding to 0.80 % of the collocated samples.*

Changes in the manuscript:

Section 4.4 now states:

“Probable residual-cloud outliers were defined as pixels with DQ-1 AOD > 1.0 and DQ-1 AOD – MCD19A2 AOD > 0.8; this criterion removed 1,621 samples (0.80 %).”

15. The very large number of collocated pixels in the MCD19A2 comparison should not be interpreted as the same number of statistically independent observations, because neighboring pixels within the same scene are spatially correlated. Please report the number of independent observation dates included in this analysis and make the distinction between pixel count and independent sampling more explicit.

Date-by-date summary statistics in the Supplement would also be useful for assessing whether the reported agreement is consistent across scenes rather than being dominated by a subset of dates.

Response: *We agree that the collocated pixel count should not be interpreted as the number of statistically independent observations. We have revised the manuscript to report explicitly that the cross-product comparison contains 200,507 spatial samples from 31 independent observation dates and to note that neighboring pixels within individual scenes are spatially correlated.*

The purpose of this analysis is to assess aggregate regional cross-product consistency on dates excluded from training and internal validation, rather than to provide date-level independent accuracy estimates; independent retrieval accuracy is evaluated against AERONET. We therefore report the number of independent observation dates and explicitly distinguish the spatial sample count from independent sampling. A date-by-date statistical table was not included in the present revision, because we retain this analysis as a regional cross-product consistency assessment rather than an independent validation dataset.

Changes in the manuscript:

Section 4.4 now states:

“A total of 200,507 collocated spatial samples from 31 independent observation dates were retained. Because neighboring pixels within the same scene are spatially correlated, the pixel count should not be interpreted as the number of statistically independent observations.”

and:

“This comparison is used to assess cross-product spatial consistency, whereas independent retrieval accuracy is evaluated against AERONET.”

16. Section 4.1 / Fig. 4: Please provide the expected-error expression directly in the main text or figure caption. This would make the validation figure easier to interpret without requiring the reader to refer to the Supplement. It should also be made clear that this expected-error envelope is used as a reference benchmark rather than as a DQ-1-specific uncertainty specification.

Response: *We have added the expected-error expression directly to the main text and the AERONET validation figure captions and clarified its role. The EE envelope is defined as $\pm(0.05 + 0.15 \times AOD_{AERONET})$, following previous satellite AOD validation studies, and is used as a common validation benchmark rather than as a DQ-1-specific uncertainty estimate.*

Changes in the manuscript:

Section 4.1 now states:

“Here, the expected-error (EE) envelope was defined as $\pm(0.05 + 0.15 \times AOD_{AERONET})$, following previous satellite AOD validation studies (Sayer et al., 2013), and was used as a common validation benchmark rather than as a DQ-1-specific uncertainty estimate.”

The same expression is also provided directly in the captions of Figs. 4–6.
