

Response to the Editor and Reviewers

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 Editor,

We thank you and the two referees for the careful and constructive evaluation of our manuscript. In response to the comments raised during the discussion, we have substantially revised the manuscript and Supplement.

The main revisions include an expanded independent AERONET evaluation covering three periods and 104 station–date matchups, additional assessment of the predefined aerosol assumptions using AERONET observations, clarification of the role of MCD19A2 AOD in the residual-learning framework, and further methodological details and sensitivity analyses for the 6S/LUT retrieval. We have also broadened the regional comparison and revised the Discussion and Conclusions to state the applicability and remaining limitations of the method more explicitly.

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

Reviewer #1

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 (p. 12, lines 302–305) 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 (p. 15, lines 383–395) 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.”

Section 2.3 (p. 5, lines 128–130) 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 (p. 25, lines 589–593) 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 (p. 6, lines 151–157) 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 (p. 2, lines 12–14).

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 (p. 8, lines 208–210) 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 (p. 13, lines 335–337) 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 (p. 20, lines 480–483) 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 (p. 9, lines 90–95) provides the corresponding sensitivity results.

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 (p. 5, lines 121–125) 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 (p. 10, lines 263–267) 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 (p. 10, lines 256–260) 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 (p. 9, lines 243–247) 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 (p. 8, lines 211–216) 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 of the Supplement (p. 8, lines 78–89).

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 (p. 12, lines 306–312) 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 (p. 12, lines 324–328) 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 (pp. 2–3, lines 15–22).

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 (pp. 13–14, lines 351–358) 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. Because correlated predictors may share or redistribute MDI importance, these values are interpreted as descriptive properties of the fitted model rather than as causal effects (Roscher et al., 2020). The effect of the surface-term correction on the intermediate surface term and final AOD retrieval is examined explicitly in Sects. 4.2 and 4.3.”

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 (p. 20, lines 483–485) 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 (p. 22, lines 509–515) 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 (p. 14, lines 371–374) 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 (Fig. 4: p. 14, lines 363–365; Fig. 5: p. 15, lines 380–382; Fig. 6: p. 16, lines 398–400).

Reviewer #2

Dear Reviewer,

We thank the reviewer for the careful and constructive assessment. In response to the concerns about validation scope, aerosol-model representativeness, and methodological reproducibility, we substantially revised the manuscript and Supplement.

Most importantly, the independent AERONET evaluation now covers three periods and 104 station–date matchups in total: 27 in spring 2025, 49 in autumn–winter 2025, and 28 in spring 2026. The autumn–winter analysis used a season-specific static prior together with a seasonally adjusted LUT and RF model, whereas the spring-2026 analysis used the spring-2025 configuration without retraining or recalibration. We also evaluated the prescribed aerosol spectral dependence using 678 AERONET records from four sites spanning 34 observation dates and added methodological details on the 6S/LUT implementation, gas-transmittance treatment, angular interpolation, matching cost, regional comparison, and data availability. The revised manuscript also states the scope and remaining limitations more explicitly.

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 describes a method to obtain an a priori surface reflectance for aerosol optical depth (AOD) retrieval from the WSI satellite instrument over the Beijing–Tianjin–Hebei (BTH) region in China. It is a two-step approach, first creating a static reference data base (beforehand) and then applying a correction based on mismatch at 443 nm (on a scene-by-scene basis). The static data base is obtained using atmospheric correction of WSI data from low-AOD days using MODIS MAIAC AOD retrievals as an input. The correction step improves performance of the AOD retrieval, based on comparisons of results using the static vs. corrected surface reflectance against ground-based AERONET observations at one location. The idea seems promising and to my knowledge is novel. Generally a prior surface constraint is necessary because surface reflectance modeling errors are a large contributor to AOD retrieval uncertainty, particularly when AOD is low, and a single-view multispectral imager generally does not have sufficient information content to constrain both spectral aerosol and surface information from a single overpass. As such the manuscript falls within AMT's scope. I appreciate the authors having incorporated suggestions made during the Quick Review process as this has made the manuscript easier to follow.

My main concern is about the extent of the data processed and analysed.

It is very restricted: to the BTH region only and for three months only (March to May 2025). It is not clear how generalisable to other parts of the world or times the results would be (despite the authors' claims that it is, it is not supported by the analysis). This short scope also makes it hard to judge the quantitative comparisons with AERONET. The data volume is not enough to make robust conclusions, and it is common for new algorithm papers to describe higher data volumes than this. Since the satellite launched in April 2022 it should be easy to process more data, even if only in the same study region. If nothing else this will also shed light on issues such as whether the initial 3-month data base is valid for other years, or times of year. Additionally, the aerosol optical models seem to have been selected for convenience and are not well justified by the analysis shown in the manuscript. These concerns are touched on by the authors in section 5.2 but I think more should be done on them in the current work. As a result, I recommend major revisions and encourage the authors to expand the analysis to make the results more robust. My specific comments are below:

Response: *We thank the reviewer for the positive assessment of the motivation and novelty of the study and for identifying the limited validation scope and aerosol-model justification as the main concerns. We agree that the original spring-2025 analysis was too limited to support broad conclusions regarding temporal robustness or general applicability. We have therefore substantially expanded the analysis in the revised manuscript.*

The independent AERONET evaluation now covers three periods and 104 station–

date matchups in total: 27 in spring 2025, 49 in autumn–winter 2025, and 28 in spring 2026. For autumn–winter 2025, a season-specific static prior was constructed, the LUT was adjusted for the seasonal viewing and atmospheric conditions, and the RF model was retrained using representative scenes from that period. In contrast, the spring-2026 evaluation used the spring-2025 retrieval configuration without retraining or recalibration, providing an interannual evaluation under a similar seasonal regime. We also broadened the regional spatial analysis by replacing the earlier municipal cutouts with a BTH-scale comparison. To address the concern regarding the predefined aerosol models, we added an observational comparison using 678 AERONET records from four BTH sites spanning 34 observation dates to examine whether the prescribed spectral AOD dependence is representative of the study region. The revised manuscript also states more explicitly that the present results do not establish year-round or cross-regional transferability, and that dense summer vegetation, snow-dominated surfaces, broader aerosol variability, and regions outside BTH remain to be evaluated. These revisions directly address the two main issues raised in the general comment—the limited data volume and the justification of the aerosol assumptions. The corresponding changes and additional analyses are described in detail in our responses to Comments 6, 7, and 12 below.

Specific comments

1. Line 122: What is the satellite overpass time window? This should be stated here.

Response: *We have added the DQ-1 WSI overpass time in both the sensor description and the AERONET matchup section. The nominal overpass time over the study region is approximately 13:30 Beijing time, corresponding to approximately 05:30 UTC. We have also clarified the temporal matching windows used for the different validation periods.*

Changes in the manuscript:

Sect. 2.2 (p. 4, lines 99–101):

“The DQ-1 atmospheric environment monitoring satellite was launched on 16 April 2022 and operates in a Sun-synchronous orbit at an altitude of approximately 700 km, with an overpass time of about 13:30 Beijing time over the study region (Wang et al., 2024b).”

Sect. 2.3 (p. 5, lines 121–126):

“The nominal DQ-1 WSI overpass time over the study region is approximately 13:30 Beijing time (approximately 05:30 UTC).”

For spring 2025 and spring 2026, AERONET observations between 02:00 and 06:00 UTC were used, while the autumn–winter 2025 evaluation used observations within ± 60 min of the nominal overpass time.

2. Line 126: A citation is needed for the MAIAC aerosol data set here.

Probably: Lyapustin, A., Wang, Y., Korkin, S., and Huang, D.: MODIS Collection 6 MAIAC algorithm, Atmos. Meas. Tech., 11, 5741–5765, <https://doi.org/10.5194/amt-11-5741-2018>, 2018.

Response: *Thank you for pointing this out. We have added the suggested MAIAC reference when introducing the MCD19A2 aerosol product.*

Changes in the manuscript:

Sect. 2.3 (p. 5, lines 131–133):

“The MODIS MAIAC MCD19A2 AOD product (Lyapustin et al., 2018) was also used to support AOD-screened surface-prior construction and regional cross-product comparison.”

3. Line 136-140: I think everything from “To ensure” to “navigation data” should be deleted because it is to be expected that the algorithm is working in calibrated physical units and that geometry is needed. If this text is kept in some form then a link or reference to the CRESDA calibration data should be given (if this calibration is not something

which is included by standard in the WSI level 1 data).

Response: *We agree that the original wording was unnecessarily general. We have therefore revised this part to describe only the preprocessing steps actually applied to the WSI data. Because the visible-to-shortwave-infrared channels were converted from digital numbers to TOA reflectance using the official CRESDA calibration coefficients, we retained this information for reproducibility and added the corresponding CRESDA calibration-data source.*

We have also added the CRESDA data portal for the DQ-1 WSI observations in the Code and data availability section.

Changes in the manuscript:

Sect. 2.4 (p. 5, lines 141–145):

“The WSI visible-to-shortwave-infrared channels were converted from digital numbers to top-of-atmosphere (TOA) reflectance using the official radiometric calibration coefficients released by the China Centre for Resources Satellite Data and Application (CRESDA). The 10.8 μm thermal infrared channel was converted to brightness temperature, while solar zenith angle, viewing zenith angle, and relative azimuth angle were obtained from the accompanying navigation data.”

Code and data availability (p. 25, lines 594–603):

“DQ-1 WSI data used in this study were obtained from the China Centre for Resources Satellite Data and Application (CRESDA) data portal and remain subject to its data-access policy. The official DQ-1 radiometric calibration coefficients and spectral response information are available from the CRESDA calibration-data webpage. The processed matchup data, validation statistics, and figure data generated in this study are available from the corresponding author upon reasonable request.”

4. Lines 144, 150: Acronyms such as MNDWI, NDVI, NDSI should be defined at first use.

Response: *We have defined these indices at their first use in the revised manuscript.*

Changes in the manuscript:

Sect. 2.4 (p. 6, lines 148–151):

“Water pixels were identified mainly using the modified normalized difference water index (MNDWI; Xu, 2006), the normalized difference vegetation index (NDVI), and related spectral constraints, while snow pixels were detected using the normalized difference snow index (NDSI).”

5. Sections 3.1, 3.1 and in general: the manuscript is not always clear about what it means by “reflectance”, e.g. line 170 talking about “surface reflectance terms” (what quantities?). I recommend adopting nomenclature from Schaepman-Strub et al (2006) where possible:

<https://www.sciencedirect.com/science/article/pii/S0034425706001167>

Reading sections 3.1 and 3.2 talking about the use of MAIAC AODs and WSI reflectance, what exactly is being calculated for the static data base?

The paper mentions Lambertian assumption using 6S for atmospheric correction but then talks about geometry being used for the random forest correction later. What I infer is being done is that the static data base is created using Lambertian atmospheric correction of WSI images using MAIAC AOD (at which wavelength, this needs to be stated?) but that aerosol optical properties (i.e. spectral and directional dependence of scattering and absorption) come from OPAC which is built into 6S (not from MAIAC – which creates an inconsistency). The multiple scenes used to generate the data base are combined to give the final data base value using rules based on brightness and NDVI (are these the “matching criterion” used to pick aerosol model mentioned in line 208? This is also unclear). Thus there is no direct accounting for BRDF effects in here.

Then the random forest (RF) method used in the correction approach takes in geometry so I guess is making the first-order correction for BRDF effects (since if I read correctly it is trained on 3 of the individual scenes going into the data base), but there is no inherent physical BRDF model. Is all this interpretation correct? This all feels very empirical but I suppose if it works it works, at least within this limited domain. Although the authors state that the retrieval is not circular because the MAIAC AOD is not input into the process here, I am not fully convinced, because the training target of this RF method is a discrepancy between the data base and the surface reflectance obtained using MAIAC AOD so it is still indirectly in there. However, since the training was done based on low-AOD conditions it seems likely that the bulk of the RF correction targets e.g. BRDF effects and not indirectly consistency with MAIAC aerosols.

Response: *We agree that the original description did not sufficiently distinguish the different reflectance quantities and the roles of MCD19A2, the 6S aerosol models, the reference-date selection, and the RF correction. We have therefore reorganized and expanded Sects. 3.1–3.4.*

In the revised manuscript, “surface reflectance” in the LUT framework is explicitly defined as a Lambertian-equivalent quantity under the Lambertian surface assumption. During offline construction of the static surface prior, MCD19A2 AOD at 550 nm is used only to constrain aerosol loading in the LUT inversion. The spectral and angular aerosol optical properties are specified by the predefined 6S aerosol model rather than taken from MAIAC. The brightness–NDVI score is used only to select the reference date for each pixel and is not the criterion used for aerosol-model selection.

We have also clarified the role of viewing geometry in the RF model. Geometry-related variables allow the RF to represent geometry-dependent variability in the surface residual empirically, but the current framework does not contain an explicit physical BRDF model. We therefore do not interpret the RF correction as a physical BRDF correction.

Finally, we agree with the reviewer that MCD19A2 is not completely absent from the residual-learning process. Because MCD19A2 AOD is used in the offline LUT inversion that generates the reference surface term, it contributes indirectly to the RF training target. We now state this explicitly. MCD19A2 AOD is not an RF input feature and is not required in the online retrieval.

Changes in the manuscript:

Sect. 3.1 (pp. 6–7, lines 170–184):

“The proposed framework consists of an offline preparation stage and an online AOD retrieval stage. The offline stage is used to construct the static surface prior and train the blue-band residual model, whereas the online stage performs AOD retrieval without requiring any external AOD product.”

“For these scenes, collocated MCD19A2 AOD is used only as an external atmospheric constraint in the LUT inversion, together with WSI TOA reflectance, observation geometry, and elevation, to derive LUT-based reference surface terms.”

Sect. 3.2 (p. 7, lines 192–195):

“Under the Lambertian surface assumption, the 6S radiative transfer model (Kotchenova et al., 2006; Vermote et al., 1997) was used to simulate top-of-atmosphere (TOA) reflectance, with surface reflectance represented as a Lambertian-equivalent quantity in the LUT retrieval (Schaeppman-Strub et al., 2006).”

Sect. 3.3 (p. 9, lines 239–247):

“After invalid pixels affected by clouds, water, and snow were removed, the collocated MCD19A2 AOD at 550 nm was used to constrain aerosol loading in the LUT inversion, while the spectral and angular aerosol optical properties were specified by the predefined 6S continental aerosol model. Together with DQ-1 WSI TOA reflectance, observation geometry, and elevation, these inputs were used to

interpolate the LUT and inversely estimate the Lambertian-equivalent surface reflectance for the blue, green, red, and near-infrared bands on a pixel-by-pixel basis.”

Sect. 3.4 (pp. 11–12, lines 298–305):

“The inclusion of solar and viewing geometry allows the RF to represent geometry-dependent variability in the surface residual empirically. No explicit physical BRDF model is included in the current framework. 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.”

6. This does lead in to a general concern (touched on in section 5.2 but not enough in my view): the study presents results only for a small geographic area and for three months (March-May 2025). It is not clear how generalisable this will be. For example, I would expect changing vegetation phenology and changing solar angles throughout the year to influence apparent surface reflectivity which would likely mean that additional data bases are needed for other times of year (this is what e.g. the NASA Deep Blue approach does). It’s also not clear that a data base created for one year would be suitable for use in other years.

Additionally, while this is a proof of concept for the BTH region, it’s not clear how the availability of suitable scenes to construct the static data base would be elsewhere in the world (due to e.g. availability of low-AOD, cloud-free, snow-free days and sufficient stability of surface characteristics). Presumably the WSI satellite is intended to eventually lead to AOD retrievals outside the BTH region (indeed, outside the cities

themselves – the only maps shown are described as urban areas, we don’t actually see any broader maps) so I would have appreciated some more analysis (or at least discussion) of that. It is not unusual for new algorithm papers to have a somewhat broader spatial/temporal scope.

Response: *We agree that the original spring-2025 analysis was too limited to support broader claims of temporal generalizability. We have therefore substantially expanded the temporal evaluation.*

First, we added an autumn–winter 2025 evaluation. For this period, a season-specific static surface prior was constructed from four autumn–winter reference scenes, the LUT was extended to larger solar zenith angles using the Midlatitude Winter atmospheric profile, and the RF model was retrained using the corresponding seasonal scenes. The resulting AERONET evaluation contains 49 station–date matchups from two sites spanning 26 independent observation dates. Second, we added a spring-2026 interannual evaluation. In this case, the spring-2025 residual model and retrieval configuration were applied without retraining or recalibration. The evaluation contains 28 station–date matchups from three AERONET sites spanning 17 independent observation dates. This evaluates the spring-2025 configuration in a different year under a similar seasonal regime. We also replaced the earlier municipal cutouts with a broader BTH regional comparison and added a true-color image to provide spatial context. At the same time, we have revised the manuscript to avoid claiming broader transferability than is supported by the data. The current study still does not systematically evaluate dense summer vegetation, snow-dominated conditions, or regions outside BTH. We now state these limitations explicitly and identify extension to additional seasons and regions as future work.

Changes in the manuscript:

Sect. 4.1 (p. 15, lines 383–395):

“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.”

Sect. 5.2 (p. 24, lines 547–561):

“Across the three evaluation periods, the AERONET analysis includes 104 station–date matchups: 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 was performed using 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. These additional evaluations substantially extend the temporal coverage of the original analysis, although validation remains confined to the BTH region and does not constitute a systematic year-round or cross-regional assessment.”

“Dense summer vegetation, snow-dominated surfaces, and regions outside the BTH domain were not systematically evaluated.”

Sect. 5.3 (p. 24, lines 573–575):

“Broader aerosol-type constraints and more systematic validation over additional regions, seasons, and independent ground-based sites will also be needed before the framework can be evaluated for broader operational application.”

7. Line 195: the continental and urban OPAC models in 6S seem to have been picked because of their names and because they are convenient to use. There is no scientific justification provided. There should be some discussion of why this is a reasonable choice, or if it was because it was

convenient, at least say that. These OPAC models are very old and I'm not sure how reasonable the urban one in particular is. Looking at the Supplement they have fairly similar spectral dependence (and there's no evidence provided that this spectral dependence is common for the BTH region) and are quite absorbing (especially the urban model, with single scattering albedo, SSA, below 0.7 at all wavelengths). As one simple step the authors could look at e.g. AERONET retrievals in this region and see how closely these models fall to observed spectral dependence of AOD and SSA. If they are similar, this is evidence that it is a reasonable choice. If not, then more justification and/or a better model is needed.

Response: *We agree that the original manuscript did not provide sufficient justification for the predefined aerosol models. We have therefore added both their representative optical properties and an observational comparison with AERONET. The revised Supplement now reports the single-scattering albedo and normalized spectral AOD of the predefined continental and urban 6S models. Their regional spectral-AOD representativeness was then examined using 678 AERONET records from four BTH sites spanning 34 observation dates. Both predefined models fall within the observed 5th–95th percentile range of normalized spectral AOD. The continental model is closer to the AERONET median spectral dependence, whereas the urban model has a steeper spectral dependence.*

The continental model is therefore used for offline construction of the static surface prior and LUT-derived reference surface terms, while both continental and urban models are retained as candidates during online retrieval and are selected at the scene scale according to the matching cost.

The added AERONET analysis evaluates the normalized spectral dependence of AOD only; it does not validate the prescribed SSA. This limitation is now stated

explicitly in Sect. 3.2.

Changes in the manuscript:

Sect. 3.2 (p. 8, lines 203–210):

“The spectral AOD dependence prescribed by the two models was compared with AERONET observations from the BTH region. The comparison included 678 records from four sites spanning 34 observation dates and was used to characterize the regional spectral AOD range. Both models fell within the observed 5th–95th percentile range, with the continental model closer to the median spectral dependence and the urban model showing a steeper spectral dependence (Fig. S1). The AERONET comparison evaluates the prescribed spectral AOD dependence only; the predefined SSA values were not independently evaluated in this study.”

Table S5 (pp. 5–6, lines 44–51); Fig. S1 (pp. 6–7, lines 52–60).

8. Lines 201–203: More information is needed on the gas absorption cross sections and how the correction was done. Spectral responses are not shown, absorption cross sections are not shown, source of gas concentrations is not given. I’m assuming either the assumption of scattering-absorption decoupling was made or the cross-sections and spectral responses were built into 6S and it was handled there. But none of this is stated.

Response: *We have expanded the description of the spectral and gaseous-absorption treatment. The 6S simulations were performed at the nominal WSI central wavelengths and did not explicitly convolve the measured WSI spectral response functions. Gaseous absorption was handled within 6S using the gaseous contents specified by the selected atmospheric profile; no additional external gas-concentration data or absorption cross sections were introduced.*

We have also clarified how the directional global-gas transmittances were combined with the total-scattering transmittances used in the LUT calculation. In addition, we

added a sensitivity test in the Supplement to quantify the influence of this gas-transmittance treatment on both simulated TOA reflectance and retrieved AOD.

Changes in the manuscript:

Sect. 3.2 (p. 8, lines 211–216):

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

“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.”

Table S7 (pp. 8–9, lines 78–90).

9. Table 2: the angular steps here feel really coarse, is this truly sufficient? Most other approaches use finer angular grids. Do the authors have an estimate of uncertainty introduced by this angular sampling? I would expect some uncertainty from the interpolation step because top of atmosphere (TOA) reflectance often does not vary linearly with angular changes.

Response: *We agree that the relatively coarse angular grid requires an explicit interpolation-error assessment. We therefore added a dedicated sensitivity test for the baseline spring 0.443 μm LUT.*

A total of 160 geometries located between the LUT angular nodes were evaluated by comparing direct 6S simulations with LUT-interpolated TOA reflectance. The overall interpolation bias was 0.0052, the RMSE was 0.0078, and the 95th percentile absolute error was 0.0185. The error increased toward large viewing zenith angles. These results and the angular dependence of the interpolation error

are now reported in Fig. S2 and Table S6.

Changes in the manuscript:

Sect. 3.2 (p. 9, lines 218–223):

“For the baseline spring LUT, the effect of angular discretization was evaluated by comparing direct 6S simulations with LUT-interpolated 0.443 μm TOA reflectance for 160 geometries located between the LUT nodes. The resulting bias and RMSE were 0.0052 and 0.0078, respectively, and the 95th percentile of the absolute difference was 0.0185. The interpolation error increased toward large viewing zenith angles (Fig. S2 and Table S6).”

10. Section 3.4: Just commenting to note that while I understand the intent of the correction approach I am not an expert on the RF method so would defer to other reviewers on whether the chosen method and level of analysis are sufficient for this part.

Response: *Thank you for noting this. Although no specific methodological change was requested, we expanded Sect. 3.4 and Table S2 so that the RF component can be assessed more directly. The revision now reports the training/validation split, hyperparameters, input variables, feature-importance calculation, residual constraints, and the role of MCD19A2 in constructing the training target.*

Changes in the manuscript:

Sect. 3.4 (pp. 11–12, lines 298–312):

“A random forest (RF) model was used to learn the nonlinear mapping between the blue-band surface reflectance residual and the input features (Breiman, 2001; Wei et al., 2019a). The input features included blue-band surface reflectance from the static library, blue-band TOA reflectance, NDVI derived from red and near-infrared TOA reflectance, solar zenith angle, viewing zenith angle, relative azimuth angle, and elevation. The inclusion of solar and viewing geometry allows the RF to represent geometry-dependent variability in the surface residual empirically. No explicit physical BRDF model is included in the current framework. 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.”

“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. Up to 80,000 pixels were randomly sampled from each training scene, and up to 60,000 pixels were sampled from the validation scene. Following previous studies on random-forest hyperparameters and tuning strategies (Probst et al., 2019), the main RF settings were as follows: 500 trees, maximum depth of 22, minimum leaf size of 40, minimum split size of 80, and maximum feature fraction of 0.7, while other parameters were kept at their default values.”

Table S2 (pp. 2–3, lines 15–28).

11. Line 302: the cost function used needs to be defined. I am guessing it is either absolute residual TOA reflectance, square residual TOA reflectance, or one of these normalised by some assumed measurement uncertainty? Since the thresholds mentioned later in this paragraph are small numbers my guess is probably not normalised?

Response: *We have now defined the matching cost explicitly. It is the squared residual between observed and LUT-simulated 0.443 μm TOA reflectance, minimized over the discrete AOD nodes. The cost is not normalized by measurement uncertainty. We have also clarified how this cost is used for the scene-level aerosol-model selection and added a sensitivity check for the matching-cost threshold.*

Changes in the manuscript:

Sect. 3.4 (pp. 12–13, lines 330–339):

$$(5) J_{i,m} = \min_{\tau \in A} [\rho_{i,0.443}^{\text{obs}} - \rho_{i,0.443,m}^{\text{sim}}(\tau)]^2,$$

and states:

“The matching cost was not normalized by measurement uncertainty. 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.”

12. Section 4.1: the data volume is really not adequate to make a meaningful assessment of performance. It is probably safe to say that the residual correction approach works better than the static LUT, but not much else. This is only 27 points from a single location, and not telling how representative that is of the broader region. The authors acknowledge this is limited but in my view agreeing it is limited is not enough, more data should be processed to get towards a more robust data volume. I would also suggest the authors do the AOD comparison at multiple wavelengths; since the spectral dependence of AOD is fixed in their retrieval approach to one of two models (continental or urban) it would be interesting to see if any one wavelength seems more stable than the others. Scatter plots could also be coloured by the AERONET Ångström exponent to see whether that seems to covary with the error of individual points. That would provide a useful indication of whether the remaining main error source is aerosol spectral dependence as opposed to something else (e.g. surface reflectance, SSA).

Response: *We agree that the original 27 spring-2025 matchups were too few for a robust assessment of retrieval performance. We therefore expanded the independent AERONET evaluation to three periods, giving 104 station–date matchups in total: 27 in spring 2025, 49 in autumn–winter 2025, and 28 in spring 2026. For autumn–winter 2025, after excluding RF training and internal-validation*

scenes, 49 matchups from two AERONET sites spanning 26 independent observation dates were retained. After residual correction, the RMSE decreased from 0.086 to 0.071, while the fraction within the expected-error envelope increased from 67.3 % to 85.7 %.

For spring 2026, the spring-2025 retrieval configuration was applied without retraining or recalibration. The evaluation included 28 matchups from three AERONET sites spanning 17 independent observation dates. The RMSE decreased from 0.330 to 0.153 and the MBE from 0.269 to 0.115 after residual correction. The reviewer also suggested examining the spectral dependence of the retrieval. In the present framework, AOD is parameterized at 550 nm, while the final matching objective uses the 0.443 μm TOA reflectance with spectral dependence prescribed by the selected 6S aerosol model. Separate wavelength-by-wavelength AOD scatterplots would therefore not constitute independent retrieval tests. Rather than further subdividing the original small matchup set by Ångström exponent, we evaluated the prescribed spectral dependence directly using 678 AERONET records from four BTH sites spanning 34 observation dates. This provides a more direct observational assessment of whether the assumed aerosol spectral dependence is representative of the study region.

We therefore focused the revision on the reviewer’s primary concern—the limited validation volume—while also directly examining the prescribed aerosol spectral dependence. A fully multi-band retrieval analysis would require band-dependent aerosol sensitivity and retrieval errors to be treated independently and is left for future work.

Changes in the manuscript:

Sect. 4.1 (p. 15, lines 383–395):

“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.”

Sect. 3.2 (p. 8, lines 203–210)); Fig. S1 (pp. 6–7, lines 52–60); Sect. 5.3 (p. 24, lines 571–572):

“The residual-correction concept could also be extended from the 0.443 μm band to additional visible bands if band-dependent aerosol sensitivity and cross-band error characteristics are explicitly considered (Litvinov et al., 2025).”

13. Section 4.4: For me, presenting Figure 9 as a trio of cutouts of municipal boundaries is not so useful without the broader regional context. My suggestion is just to show the whole broader BTH area these three locations all sit in (i.e. including the spaces between them). and add a panel showing a true colour image from this day so the reader had broader context. So this would be 5 larger panels instead of 12. About all I glean from the current image is that the residual-corrected method matches MAIAC better than the static one, but it’s hard to gauge much more.

Response: *We agree that the earlier municipal cutouts did not provide sufficient regional context. We have therefore replaced this presentation with a broader BTH-scale comparison and added a DQ-1 WSI true-color panel.*

The revised figure now presents the true-color image, static blue-band surface prior, static-prior LUT AOD, residual-corrected AOD, and MCD19A2 AOD over the same broader BTH extent. This allows the spatial structure of the retrievals and the locations of cloud-affected or masked areas to be interpreted within a common regional context.

Changes in the manuscript:

Sect. 4.4 / Fig. 11 (p. 21, lines 486–495):

“Figure 11. Spatial comparison over the broader Beijing–Tianjin–Hebei (BTH) region

on 28 May 2025. (a) DQ-1 WSI true-color image; (b) static blue-band surface reflectance used as the surface prior; (c) static-prior LUT AOD; (d) residual-corrected AOD; and (e) MCD19A2 AOD.”

The corresponding text now states:

“Figure 11 presents the DQ-1 WSI true-color image, the static blue-band surface prior, the static-prior LUT AOD, the residual-corrected AOD, and MCD19A2 AOD over the same spatial extent.”

14. Line 620: links and/or data DOIs are needed for this section.

Response: *We have revised the Code and data availability section to provide direct access information for the principal datasets used in this study. The revised text now includes the NASA AERONET data portal, the DOI for the MODIS MAIAC MCD19A2 Version 6.1 product, the CRESDA data portal for DQ-1 WSI observations, and the official CRESDA webpage for DQ-1 radiometric calibration coefficients and spectral response information.*

Changes in the manuscript:

Code and data availability (p. 25, lines 595–603):

“The code used for QA-mask generation, blue-band surface-residual learning, and LUT-based AOD retrieval is available from the corresponding author upon reasonable request. AERONET Version 3 observations are publicly available from the [NASA AERONET data portal](#). The MODIS MAIAC MCD19A2 Version 6.1 aerosol product is publicly available from NASA Earthdata under DOI 10.5067/MODIS/MCD19A2.061. DQ-1 WSI data used in this study were obtained from the [China Centre for Resources Satellite Data and Application \(CRESDA\) data portal](#) and remain subject to its data-access policy. The official DQ-1 radiometric calibration coefficients and spectral response information are available from the [CRESDA calibration-data webpage](#). The processed matchup data, validation statistics, and figure data generated in this study are available from the corresponding author upon reasonable request.”

Closing statement

We again thank the Handling Editor and both referees for their careful evaluation and constructive comments. We believe that the revisions have substantially improved the clarity, transparency, reproducibility, and scientific context of the manuscript, while more clearly defining the scope and limitations of the present study. We hope that the revised manuscript is now suitable for further consideration for publication in *Atmospheric Measurement Techniques*.

Thank you for your consideration.

Yours sincerely,

Jingxuan Zhao, on behalf of all authors

29 September 2026