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

This supplement provides additional methodological details for the unified QA mask, the blue-band surface-residual learning model, the AERONET matchup validation, the scene usage for residual learning and cross-product consistency analysis, and the predefined aerosol models used in the 6S-based LUT simulations. In Table S1, blue, green, red, NIR, and SWIR2 denote the DQ-1 WSI TOA reflectance bands used for QA masking. The variables vis_mean and vis_std denote the mean and standard deviation of the blue, green, and red reflectance, respectively. BT10.8 denotes the 10.8 μm brightness temperature. The variable grad denotes the Sobel gradient magnitude of smoothed vis_mean, and texture denotes the absolute residual between vis_mean and its Gaussian-smoothed field.

Table S1. Thresholds and morphological operations used in the unified QA mask.

Target	Feature or rule	Threshold or operation	Role / output class
Thick cloud	Visible mean reflectance	vis_mean > 0.18	cloud
Thick cloud	Visible whiteness	vis_std < 0.04	cloud
Thin cloud / haze	Blue reflectance	0.18 < blue < 0.55	cloud / haze
Thin cloud / haze	Visible mean reflectance	vis_mean > 0.15	cloud / haze
Vegetation suppression	NDVI	NDVI < 0.35	constraint
Snow-like target suppression	NDSI	NDSI < 0.30	constraint
Cold-cloud branch	10.8 μm brightness temperature	BT10.8 < 292 K	cloud
Warm-haze branch	10.8 μm brightness temperature and blue reflectance	BT10.8 < 300 K and blue > 0.18	haze
Cloud edge / texture	Visible gradient	grad > 0.015	cloud edge
Cloud edge / texture	Visible texture	texture > 0.020	cloud edge
Cloud morphology	Binary dilation, hole filling, small-object removal	2 iterations; min_size = 80; additional 3 × 3 dilation	cloud
Snow	NDSI	NDSI > 0.40	snow
Clear water	MNDWI	MNDWI > 0.40	water
Complex water	NDVI and spectral constraints	NDVI < 0.30, blue > red, NIR < 0.15, MNDWI > −0.25	water
Water morphology	Hole filling and small-object removal	min_size = 10	water

Snow morphology	Small-object removal	min_size = 10	snow
QA encoding	Class values	0 = clear, 1 = cloud/haze, 2 = water, 3 = snow, 255 = no data	QA mask

Table S2. Configuration of the blue-band residual-learning model.

Item	Setting
Target variable	Blue-band surface residual
Residual definition	LUT-derived reference blue-band surface term minus static blue-band surface prior
Training period	Spring 2025
Training scenes	22 March, 22 April, and 19 May 2025
Internal validation scene	28 April 2025
AOD range for residual-sample construction	0.01–0.50
Maximum samples per training scene	80,000 pixels
Maximum samples for internal validation	60,000 pixels
Input features	Static blue-band surface prior, blue-band TOA reflectance, NDVI, solar zenith angle, viewing zenith angle, relative azimuth angle, and elevation
External AOD as RF input feature	No
Number of trees	500
Maximum tree depth	22
Minimum samples per leaf	40
Minimum samples for split	80
Maximum feature fraction	0.7
Reference blue-band surface term range	0–0.8
Residual outlier filter	Absolute residual ≤ 0.12
Predicted residual constraint	−0.10 to 0.10
Corrected blue-band surface term constraint	0–0.8
Online inference requirement	No external AOD product is required

The residual threshold of 0.12 is used only as an outlier-screening criterion during residual-sample construction and does not directly enter the online AOD inversion. It excludes extreme discrepancies between the LUT-derived reference blue-band surface term and the static prior. During online application, the RF-predicted residual is constrained to −0.10–0.10 as a safeguard against unusually large model corrections, while the corrected blue-band surface term is restricted to 0–0.8,

consistent with the admissible range used for the LUT-derived reference surface term. These constraints are intended to prevent anomalous surface corrections from propagating into the subsequent LUT-based retrieval.

Table S3. Configuration of the AERONET matchup validation.

Item	Setting
Ground-based reference	AERONET Version 3 Level 1.5 instantaneous observations
Study region	Beijing-Tianjin-Hebei region
Temporal matching window	02:00-06:00 UTC, corresponding to 10:00-14:00 Beijing time
Satellite retrievals compared	Static-prior LUT retrieval and residual-corrected retrieval
Spatial matching window	3 × 3 DQ-1 WSI pixels centered on each AERONET site
Minimum valid satellite pixels	At least 1 valid pixel within the 3 × 3 window
Satellite valid AOD range	$0 < AOD < 3.0$
Ground-based AOD wavelength conversion	AOD at 550 nm interpolated from 500 nm AOD and the 440–675 nm Ångström exponent
AERONET temporal aggregation	Mean AERONET AOD within the temporal matching window for each site-date pair
Satellite spatial aggregation	Mean satellite AOD within the valid pixels of the 3 × 3 window
Common-sample requirement	A site-date pair was retained only when both the static-prior LUT retrieval and the residual-corrected retrieval were valid
Expected error envelope	$\pm (0.05 + 0.15 \times$ AERONET AOD$)$
Validation unit	Station-date matchup
Final number of matchups	27

Note. The final number of matchups is limited because a station-date pair was retained only when both retrieval schemes were valid under the same temporal and spatial matching criteria.

Table S4. Scene usage and data-split configuration for residual learning and cross-product consistency analysis.

Scene date	Usage in this study	Role in leakage control

22 March 2025	RF residual-model training	Excluded from the MCD19A2 hold-out-date cross-product consistency analysis
22 April 2025	RF residual-model training	Excluded from the MCD19A2 hold-out-date cross-product consistency analysis
19 May 2025	RF residual-model training	Excluded from the MCD19A2 hold-out-date cross-product consistency analysis
28 April 2025	Internal validation of residual module	Excluded from the MCD19A2 hold-out-date cross-product consistency analysis
20 March 2025	Excluded from cross-product consistency analysis	Excluded because of abnormal weather or large-area mask failure
29 April 2025	Excluded from cross-product consistency analysis	Excluded because of abnormal weather or large-area mask failure
10 May 2025	Excluded from cross-product consistency analysis	Excluded because of abnormal weather or large-area mask failure
Other matched dates not listed above	Hold-out-date cross-product consistency analysis	Dates used for RF training and internal validation were excluded

Note. Dates used for RF residual-model training and internal validation were explicitly excluded from the MCD19A2 cross-product consistency analysis to reduce data-leakage concerns.

The LUT simulations used the predefined continental and urban aerosol models implemented in 6S. Table S5 summarizes their representative spectral optical properties, including aerosol single-scattering albedo (SSA) and normalized spectral aerosol optical depth, $\tau a(\lambda)/\tau a(550)$.

Table S5. Representative spectral optical properties of the predefined continental and urban aerosol models used in the 6S simulations.

Wavelength/μm	Continental SSA	Urban SSA	Continental $\tau a(\lambda)/\tau a(550)$	Urban $\tau a(\lambda)/\tau a(550)$
0.443	0.9004	0.6955	1.2346	1.3032
0.490	0.8993	0.6937	1.1221	1.1569
0.550	0.8932	0.6888	1.0000	1.0000
0.681	0.8839	0.6743	0.7959	0.7532
0.865	0.8570	0.6300	0.5974	0.5280
2.130	0.7104	0.2877	0.2364	0.1430

Note. SSA denotes aerosol single-scattering albedo, and τa(λ)/τa(550) denotes aerosol optical depth normalized to 550 nm. The values were obtained directly from the predefined 6S aerosol models with AOD at 550 nm set to 1.0. The 0.550 μm row is included only as the normalization reference and is not a DQ-1 WSI retrieval band.