Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Aerosol Inversion Based on Vector Radiative Transfer Model and Machine Learning Method

He Huang
Yangzhou Meteorological Bureau
Bochen Zhang
ZBC@yzu.edu.cn

Yangzhou University
Yong Qian
Yangzhou Meteorological Bureau
Le Liu
Yangzhou Meteorological Bureau
Yuanfu Niu
Yangzhou Meteorological Bureau

Article

Keywords: Vector radiative transfer model, Satellite remote sensing, AOD retrieval, Machine learning method

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11710252/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

While geostationary imagers can track diurnal aerosol optical depth (AOD), separating atmospheric and surface signals remains challenging. Here, we combine a vector radiative transfer model (VRTM) with multi-spectral machine learning to enhance AOD retrieval from FengYun-4A (FY-4A) AGRI observations and evaluate the algorithm across different operational scenarios. In VRTM-based synthetic inversions, AOD recovery reached numerical precision with a known surface, whereas jointly fitting AOD and surface reflectance increased the root mean square error (RMSE) to 0.121, highlighting the need for multi-channel constraints. For independent FY-4A AGRI pairs at 550 nm, our five-channel random forest algorithm improved retrieval accuracy to a predictive coefficient of determination ($R^2_{\mathbf{pred}}$) of 0.601 and an RMSE of 0.212, compared with 0.519 and 0.232 for the single blue-channel model. Scenario-specific validation further showed that the five-channel algorithm achieved high accuracy for temporal retrieval at monitored pixels ($R^2_{\mathbf{pred}}$ = 0.704 in date-grouped validation) and an $R^2_{\mathbf{pred}}$ of 0.497 when applied to unmonitored locations. Overall, this framework improves geostationary AOD retrieval accuracy and provides clear guidance for deploying machine learning retrievals across temporal monitoring and spatial transfer applications.

Introduction

Atmospheric aerosols influence climate by scattering and absorbing sunlight and by interacting with clouds. Monitoring their distribution also supports air-quality assessment, particularly when emissions, transport and removal change aerosol loading within a day. Aerosol optical depth (AOD) measures the extinction of light by aerosols through the atmospheric column and provides a common observable for satellite and ground-based monitoring. Retrieving AOD from reflected sunlight requires separating aerosol scattering from reflection by the underlying surface[1–3]. Algorithms such as MODIS Dark Target, Deep Blue and MAIAC use different spectral and surface constraints to make that separation[4–6].

Geostationary imagers extend AOD monitoring to subdaily timescales. Repeated column observations can track the timing and extent of smoke or dust transport, providing context for surface air-quality measurements. Specifically, the Advanced Geostationary Radiation Imager (AGRI) aboard Fengyun-4A repeatedly samples the same locations and supports multispectral AOD retrieval over land[7,8]. Whereas the sensor view direction remains nearly fixed at a given pixel, solar zenith and relative azimuth change through the day. Land surfaces reflect light anisotropically, so their bidirectional reflectance varies with solar angle even when the surface and aerosol column remain stable. Solar zenith also changes the optical path through the atmosphere, which affects the sensitivity of the signal to aerosol. Consequently, an error in representing the surface can introduce a diurnal signal into retrieved AOD[9–11].

The Unified Linearized Vector Radiative Transfer Model (UNL-VRTM) provides a controlled setting for examining how aerosols, surface reflectance and geometry shape the signal measured by a satellite[12]. Statistical emulators can approximate these forward calculations, and related retrieval studies have

combined simulated scenes with learned models[13–15]. Accurate predictions of a signal can still leave its origin ambiguous when changes in AOD and surface reflectance produce similar signals. In contrast to forward validation, synthetic inversion with controlled surface constraints probes whether the signal contains enough information to recover model AOD.

For retrieval from observations, AERONET provides ground-based spectral AOD for assessing performance on new dates and at new locations[16,17]. Testing new dates at familiar pixels retains the local surface and viewing environments represented in training, whereas withholding pixels changes those environments[18–20]. We compare forward emulation at withheld geometries, examine how a known surface constrains model-space inversion and assess the transfer of FY-4A AOD retrieval at 550 nm across dates and pixels. Figure 1 links the three tracks and distinguishes the simulation experiments from the satellite retrieval.

Methods

FY-4A AGRI observations and ground collocation

The 2018–2021 satellite archive extracts calibrated reflective solar-channel top-of-atmosphere (TOA) apparent reflectances and solar–satellite viewing angles from FY-4A AGRI Level 1 full-disk radiometric and geolocation products (FDI/GEO) paired with Level 2 cloud-mask products (CLM on a nominal 4 km grid) at hourly UTC intervals. Annual extraction footprints comprise 1,183, 1,254, 1,306 and 1,306 fixed 4 km land pixel locations across 2018–2021, aligned by physical pixel coordinates across years to yield 4,756,842 confident-clear pixel-time records and 14,731 ground-collocated pairs.

The nearest retained pixel within 10 km and the nearest eligible ground observation within ± 30 min of the nominal hour form one pair per site, pixel and satellite time; ties in ground time use the earlier observation. Duplicate spectra and satellite records are removed. Confirmed AERONET Version 3 Level 2.0 direct-Sun measurements[17] take priority, yielding 7,175 pairs. The remaining 7,556 pairs derive from pre-screened sun-photometer spectral records at the same monitoring sites (processed under identical collocation and cloud-screening protocols prior to final Level 2.0 post-deployment archival) and are retained to expand multi-year temporal coverage. For simultaneous positive spectral AOD, 500 and 675 nm take priority, followed by other available wavelength pairs in a fixed priority order using reported filter wavelengths. Spectral AOD is interpolated to 550 nm via a local Ångström power law with λ_1 and λ_2 in nanometers[21]:

$$\ln \tau_{550} = \ln \tau_{\lambda_1} + \frac{\ln\left(\tau_{\lambda_2} / \tau_{\lambda_1}\right)}{\ln\left(\lambda_2 / \lambda_1\right)} \ln\left(550 / \lambda_1\right).$$

1

Clear-sky screening requires the Level 2 confident-clear mask (CLM = 3) and restricts solar and viewing zenith angles (SZA and VZA) to [0°, 80°) with solar and satellite azimuths in [0°, 360°]. Each of the five non-cirrus solar channels must be finite and within [0, 1.5], with at least one positive value, while viewing zenith and azimuth angles must agree with the nominal geostationary orbital geometry at 104.7°E within 0.5° and 1.0°, respectively. Because the Level 2 archive lacks an independent snow or inland-water mask, surface screening relies on the fixed land-pixel selection and radiometric bounds. Relative azimuth angle (RAA) is defined as the folded azimuthal difference between satellite azimuth φ_v and solar azimuth φ_s (Eq. (2)):

$$\Delta\varphi = |\varphi_v - \varphi_s|, \quad \mathrm{RAA} = \min(\Delta\varphi, 360^\circ - \Delta\varphi).$$

2

UNL-VRTM forward database and geometry-grouped emulation

The UNL-VRTM lookup table (LUT) contains 803,088 states spanning 13 solar zenith angles (SZA), 13 viewing zenith angles (VZA), 11 relative azimuth angles (RAA), eight AOD values, nine Lambertian surface reflectance values, and six aerosol IDs[12,22]. For each state, the LUT stores monochromatic Stokes I outputs in normalized radiance units at 470, 650, 825, 1375, 1610, and 2250 nm.

To isolate numerical operator conditioning from empirical sensor cross-calibration uncertainties, we analyzed the lookup table directly in native monochromatic model coordinates (τ_{model} and normalized Stokes *I*) across the six discrete aerosol optical regimes (IDs 1–6). This controlled formulation benchmarks forward emulation and surface–aerosol equifinality near the nominal reflective solar channel centers of FY-4A AGRI (with 1375 nm representing the 1.37 μm cirrus absorption band) prior to introducing sensor spectral response convolution or observational noise.

Retaining only finite, Stokes *I* entries lookup-table nodes with complete six-channel support. Within each discrete aerosol regime, the forward operator applies bounded five-dimensional multilinear interpolation across all continuous state dimensions whenever all contributing hypercube vertices are valid across the required channels.

The forward emulators take the cosines of SZA, VZA and RAA, τ_{model}, ρ_s and six one-hot aerosol regime indicators as predictors. A fixed random partition assigns the 1,859 solar–viewing geometry triples as whole groups into 1,301 initial training, 186 inner validation and 372 independent test groups, ensuring that each channel's neural network (NN) and random forest (RF) are evaluated on identical held-out states. Each NN comprises two hidden layers of 48 rectified linear units trained using standardized inputs and targets, the Adam optimizer (initial learning rate 0.001, batch size 2,048, L_2 weight decay 0.0001) and early stopping within 50 epochs (patience 7), before refitting at the selected epoch count across all 1,487 non-test geometry groups. Each forward RF comprises 64 regression trees grown to a maximum depth of 20 with at least three records per terminal leaf and 80% bootstrap subsampling[23].

Synthetic inversion and surface constraints

Evaluating seven AOD interval midpoints, eight surface reflectance midpoints, six aerosol regimes and three observed SZA quartile geometries generates 1,008 off-node synthetic test cases. Forward synthesis and inverse retrieval share the bounded multilinear interpolator across channels 1, 2, 3, 5 and 6 under noiseless conditions, excluding state combinations that encounter invalid interpolation vertices. When surface reflectance and aerosol regime are prescribed, τ_{model} is recovered by minimizing the unweighted spectral residual across all valid piecewise-linear AOD segments.

The joint fit estimates $\tau_{\mathbf{model}}$ and a common Lambertian $\rho_{\mathbf{s}}$ with known geometry Ω and aerosol ID m. The unweighted objective compares the interpolated signal $\hat{I}_{\mathbf{c}}$ with the synthetic signal $I_{\mathbf{c,syn}}$:

$$J\left(\tau_{\mathrm{model}},\ \rho_{s}\right)\ =\ \sum_{c\ \in\ \{1,\,2,\,3,\,5,\,6\}}\left[\hat{\mathrm{I}}_{c}\left(\Omega\ ,\ \tau_{\mathrm{model}},\ \rho_{s},\ m\right)\ -\ I_{c,\mathrm{syn}}\right]^{2}.$$

3

In the joint inversion, we constrain τ_{model} to [0.05, 2.0] and ρ_{s} to [0.02, 0.8], initializing nonlinear least squares from the valid grid node with minimal objective residual (up to 100 function evaluations) and recording convergence status, boundary clamping, residual norms and local Jacobian conditioning. Because a single scalar channel provides at most a rank-one Jacobian with respect to the two continuous unknowns (τ_{model}, ρ_{s}), multispectral observations across channels with contrasting surface-to-aerosol sensitivity ratios are mathematically required to regularize the joint inverse problem.

Observation-trained random forests and blocked validation

Both observational FY-4A RF regressors predict τ_{550} from the retained AGRI reflective solar signals and viewing–illumination geometry cosines. The single-channel blue RF uses channel 1 (470 nm), whereas the five-channel VSWIR RF incorporates channels 1, 2, 3, 5 and 6 (470, 650, 825, 1610 and 2250 nm), excluding the 1375 nm cirrus-screening band. Each ensemble comprises 400 bootstrap-sampled regression trees with at least five records per terminal leaf and all candidate predictors evaluated at each split. Station identifiers, geographic coordinates and ground-derived surface reflectance priors are excluded from the predictor set, and each training record is weighted by the reciprocal of the number of collocated stations sharing the same satellite pixel-time within a training fold.

Interannual temporal transfer evaluations train on complete preceding years (2018–2020 → 2021, and 2020 → 2021 for single-year comparison), requiring all training reference timestamps to precede the test-year boundary by at least 30 min. Date-grouped five-fold cross-validation withholds complete UTC dates while grouping any dates linked by a shared ground observation. Leave-one-pixel-out validation withholds each physical 4 km AGRI pixel together with all collocated ground stations and years, removing any ground observations shared with the held-out set from the training pool.

For N references y_i with mean $\bar{y}$ and predictions $\hat{y}_i$, we report Pearson R, its square $r^2 = R \times R$, the predictive coefficient of determination R^2_{pred}, RMSE, MAE and signed bias:

$$R^2_{\text{pred}} = 1 - \frac{\sum_{i=1}^{N} \left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{N} \left(y_i - \bar{y}\right)^2},$$

4

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\hat{y}_i - y_i\right)^2},$$

5

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left|\hat{y}_i - y_i\right|, \quad \text{Bias} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{y}_i - y_i\right).$$

6

Positive bias indicates overestimation. Percentile intervals use 500 resamples of complete UTC dates with predictions held fixed. Station intervals use within-site UTC dates and paired draws for both RF configurations. Figure 6 uses the MODIS Dark Target land envelope as a reference guide, independent of the screening criteria[4].

Results

Observation coverage and forward emulation

From 39,103,863 dated FY-4A pixel-time records during 2018–2021, 4,756,842 passed confident-clear and geometry quality screening. The ground-collocated evaluation dataset comprises 14,731 matched pairs across 1,185 UTC dates and 12,256 distinct satellite pixel-times at seven monitoring sites (six 4 km AGRI pixels across the Beijing cluster, Xuzhou and Hong Kong; Fig. 2 and Table 1). Annual sample sizes were 3,499, 4,200, 2,837 and 4,195 during 2018–2021, respectively. The 2021 independent temporal holdout spans five sites (four pixels), whereas the 2018–2019 training archive additionally includes 1,014 pairs at Xuzhou and 53 pairs across 11 UTC dates at Yanqihu.

Table 1
FY-4A collocated observations by monitoring site and year. *N* denotes total collocated pairs and Days counts distinct UTC dates at each site. Level 2.0 denotes confirmed AERONET Version 3 Level 2.0 direct-Sun pairs; the remaining records derive from pre-screened sun-photometer spectral archives processed under identical collocation and cloud-screening protocols.

Station	2018	2019	2020	2021	*N*	Days	Level 2.0
Beijing	0	182	610	880	1,672	368	182
Beijing-CAMS	1,173	1,289	1,020	1,049	4,531	950	4,526
Beijing-PKU	1,023	1,164	0	464	2,651	548	0
Beijing-RADI	0	821	708	797	2,326	560	0
Hong Kong Sheung Shui	516	464	499	1,005	2,484	524	2,467
Xuzhou-CUMT	787	227	0	0	1,014	239	0
Yanqihu	0	53	0	0	53	11	0

At 470 nm, the slope of the top-of-atmosphere radiance response to AOD varies substantially with surface reflectance, viewing geometry and aerosol regime, becoming weak or slightly negative over brighter surfaces and under absorbing aerosol configurations (Fig. 3a–c). The 1375 nm signal remains near zero and exhibits minimal sensitivity across the evaluated AOD range (Fig. 3d), as strong mid-tropospheric water vapor absorption near 1.38 μm attenuates both boundary-layer aerosol scattering and surface reflection[24,25]. By contrast, the 1610 and 2250 nm shortwave infrared window channels included in the five-channel RF retain strong sensitivity to underlying surface reflectance variations across seasons and viewing geometries[8,10], although translating these noiseless sensitivity curves into operational retrieval skill depends on sensor signal-to-noise characteristics.

With entire solar–viewing geometries withheld from training, predictive $R^2_{\mathbf{pred}}$ ranged from 0.902 to 0.999 for the neural networks (NNs) and from 0.932 to 0.994 for RF (Fig. 4). The number of valid LUT states ranged from 693,456 to 803,088 across channels, with both models evaluated on identical test states within each channel. RF achieved higher predictive $R^2_{\mathbf{pred}}$ and lower RMSE at all five non-cirrus wavelengths. The largest improvement occurred at 470 nm, where $R^2_{\mathbf{pred}}$ increased from 0.909 to 0.956 and RMSE decreased from 0.0141 to 0.00986, a reduction of approximately 30% (Fig. 4a, b). At 650 and 825 nm, RF reduced RMSE from 0.0142 to 0.0117 and 0.0123, respectively (Fig. 4c–f). The NN regression lines fell below the 1:1 line toward the upper end of the signal range, whereas the RF fits followed the diagonal more closely.

At 1610 and 2250 nm, both models achieved higher predictive scores than at the visible and near-infrared wavelengths (Fig. 4i–l). RF attained $R^2_{\mathbf{pred}}$ values of 0.959 and 0.971, compared with 0.944 and 0.959 for the NNs. The corresponding RMSE values decreased from 0.0105 to 0.00897 at 1610 nm and from 0.00976 to 0.00825 at 2250 nm. Across the five non-cirrus channels, the point clouds showed

broader dispersion at low to intermediate intensities and generally narrowed at higher intensities. Off-diagonal branches remained visible in the RF predictions, particularly at 650 and 825 nm, despite their lower aggregate errors.

The model ranking reversed at 1375 nm (Fig. 4g, h). The NN achieved $R^2_{\mathbf{pred}}$ =0.999 compared with 0.994 for RF. Both point clouds closely followed the 1:1 line, with a narrower spread for the NN. This weak signal is consistent with strong water-vapor absorption near 1375 nm, which attenuates radiation from the surface and lower atmosphere[24,25].

Inverse conditioning under surface constraints

The interpolator exactly reproduced the valid values at all 803,088 serialized nodes. Of 1,008 off-node cases, 718 supported all five required signals and 290 encountered invalid corners. With surface reflectance and aerosol ID supplied, recovery reached numerical precision. Joint AOD and surface estimation yielded RMSE of 0.121 and mean absolute error (MAE) of 0.018 (Fig. 5). The joint estimates included 672 interior closures, 44 fits with nonzero residual and two boundary solutions. A few large errors increased RMSE despite close recovery in most cases.

Temporal and geographic generalization in FY-4A

The five-channel RF had lower RMSE than the blue-channel RF in each of the four principal evaluations (Table 2). With training in 2018–2020, the five-channel model achieved R = 0.776, $R^2_{\mathbf{pred}}$ = 0.601, RMSE = 0.212 and bias = + 0.012 on 4,195 pairs in 2021. Figure 6a,b shows underprediction in parts of the high-AOD tail despite the small pooled bias. Restricting training to 2020 raised the five-channel RMSE to 0.240 on the same test pairs.

Table 2

Principal FY-4A independent evaluations. *N* is the number of pairs. Date and pixel tests pool out-of-fold predictions. Brackets give 95% percentile intervals from 500 complete UTC-date resamples of fixed predictions. *R* is Pearson correlation and r^2 its square; R^2_{pred} compares unadjusted predictions with the test-set mean. Errors are in dimensionless AOD at 550 nm; bias is prediction minus reference.

Evaluation	RF	*N* / dates	*R*	r^2	R^2_{pred}	RMSE	MAE	Bias
2018–2020 → 2021	Blue	4,195 / 294	0.737	0.543 [0.425, 0.637]	0.519 [0.406, 0.591]	0.232 [0.202, 0.266]	0.148	−0.046
2018–2020 → 2021	Five	4,195 / 294	0.776	0.602 [0.505, 0.681]	0.601 [0.495, 0.668]	0.212 [0.185, 0.242]	0.139	+ 0.012
2020 → 2021	Blue	4,195 / 294	0.708	0.501 [0.396, 0.588]	0.428 [0.332, 0.489]	0.253 [0.222, 0.291]	0.160	−0.073
2020 → 2021	Five	4,195 / 294	0.737	0.543 [0.449, 0.616]	0.486 [0.401, 0.540]	0.240 [0.208, 0.277]	0.149	−0.060
Date-grouped five-fold	Blue	14,731 / 1,185	0.757	0.573 [0.533, 0.613]	0.572 [0.532, 0.610]	0.238 [0.222, 0.252]	0.154	−0.006
Date-grouped five-fold	Five	14,731 / 1,185	0.844	0.712 [0.689, 0.736]	0.704 [0.683, 0.725]	0.198 [0.184, 0.211]	0.125	−0.007
Leave-one-pixel-out	Blue	14,731 / 1,185	0.656	0.431 [0.383, 0.474]	0.413 [0.349, 0.464]	0.278 [0.263, 0.293]	0.171	−0.019
Leave-one-pixel-out	Five	14,731 / 1,185	0.708	0.502 [0.458, 0.541]	0.497 [0.446, 0.537]	0.258 [0.244, 0.271]	0.159	−0.008

In five-fold cross-validation grouped by date, the five-channel RF achieved R = 0.844, R^2_{pred} =0.704, and RMSE = 0.198 at monitored pixels. When entire satellite pixels were withheld, R decreased to 0.708 and R^2_{pred} to 0.497, while RMSE increased to 0.258 (Fig. 6c, d). Prediction errors were therefore larger at withheld pixels than on new dates at pixels represented in training.

Performance also varied among the seven monitoring sites. In date-grouped validation, Pearson R ranged from 0.645 at Hong Kong Sheung Shui to 0.924 at Yanqihu, and site-specific RMSE ranged from 0.172 to 0.250. The Yanqihu estimate was based on observations from 11 UTC dates. Under leave-one-pixel-out validation, performance declined more strongly at the geographically isolated Hong Kong Sheung Shui and Xuzhou-CUMT sites than at the clustered Beijing stations.

The confirmed Level 2.0 and local archive subsets differed in site coverage, observation period, and reference provenance. Retrieval errors also varied with aerosol loading, season, and illumination geometry. Underestimation became more pronounced toward the upper end of the reference AOD distribution. Seasonal RMSE ranged from 0.164 in winter to 0.215 in summer, while RMSE across solar zenith angle strata ranged from 0.179 to 0.233.

Subdaily performance was examined using two Beijing-CAMS episodes from the screened archive of 4,756,842 pixel-time records. For the five-channel RF trained on pre-2021 data, the low-AOD autumn episode on 17 October 2021 comprised nine hourly matched pairs, with a mean reference AOD of 0.092 and an RMSE of 0.045. The moderate summer haze episode on 10 August 2021 comprised eight matched pairs, with a mean reference AOD of 0.543 and an RMSE of 0.180. The larger error during the summer episode was consistent with the increased retrieval errors at higher aerosol loading.

Discussion

The high forward emulation scores reflect accurate reproduction of stored lookup-table intensities, including channels that are weak in magnitude and minimally sensitive to column aerosol loading. At 1375 nm, strong water vapor absorption suppresses radiative contributions from lower-tropospheric aerosols and the underlying surface, so high statistical emulation fidelity in that band does not imply constraint on column AOD[24,25]. Prescribing surface reflectance in the noiseless off-node inversion removed a continuous degree of freedom and enabled numerical closure to machine precision, whereas jointly estimating surface reflectance and AOD permitted compensating parameter combinations to fit nearly identical multispectral signals. Because both experiments assumed a known aerosol regime and shared the bounded multilinear forward operator, they isolate intrinsic inverse conditioning in model space.

The 1610 and 2250 nm shortwave infrared channels can help constrain the surface contribution as vegetation, soil and illumination vary across seasons and viewing geometries[8,10,11]. The five-channel RF also adds two shorter wavelengths, so the observed improvement over the blue-channel model does not isolate a shortwave infrared effect. Without an explicit surface predictor, the RF learns local associations among radiometry, geometry and AOD. Such associations are less likely to hold at unmonitored pixels, where differences in local surface reflectance, aerosol optical properties, cloud screening and reference sampling degrade spatial transfer.

For broader regional mapping, the multivariate Area of Applicability (AoA) framework quantifies predictor-space dissimilarity relative to training samples and calibrates a dissimilarity threshold from cross-validation to flag pixels where empirical predictions extrapolate beyond training support[26]. Defining an operational AoA across the FY-4A disk requires spatially blocked validation across monitoring sites that span the target domain's full joint distribution of surface reflectance, aerosol regime and viewing geometry.

Because the confirmed Level 2.0 and pre-screened local archive strata sample different combinations of sites and multi-year periods, their metric differences reflect spatiotemporal regime differences alongside reference processing. Pairing hourly nominal satellite scan times with point sun-photometer observations also introduces spatiotemporal representativeness uncertainty[27,28]. Furthermore, the date-grouped bootstrap intervals quantify sampling variability conditional on fixed out-of-fold predictions rather than full ensemble refitting or radiometric calibration uncertainty.

Together, these three experimental tracks demonstrate that high forward emulation accuracy can coexist with ill-conditioned inverse AOD recovery when surface reflectance is unconstrained, and that temporal skill at monitored pixels exceeds spatial transferability to unmonitored pixels. Operational deployment of physical-statistical hybrid inversions across the full FY-4A AGRI disk will therefore require coupling radiative transfer emulators with sensor spectral response convolutions, explicit bidirectional surface reflectance priors and multivariate Area of Applicability screening.

Declarations

Data availability

AERONET Version 3 direct-Sun observations are publicly available from the NASA Goddard Space Flight Center (https://aeronet.gsfc.nasa.gov/), and FY-4A AGRI Level 1 and Level 2 products are publicly accessible from the National Satellite Meteorological Center (https://data.nsmc.org.cn/). Summary evaluation tables are included in this article, and the processed sun-photometer collocation records and LUT simulation datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Code availability

The analysis and plotting scripts used to reproduce all validation metrics, bootstrap confidence intervals, figures and tables are archived with the evaluation package.

Funding

This work was supported by the Huaihe River Basin Meteorological Center (Grant HRM202601), the Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites (Grant 2006NSMCQ03), the Jiangsu Provincial Meteorological Bureau (Grant KQ202618), Yangzhou

Meteorological Bureau (Grant Z202501) and Yangzhou Lvyang Jinfeng Program (YZLYJFJH2023YXBS100).

Acknowledgments

We thank the AERONET site investigators and teams for maintaining the observations, the National Satellite Meteorological Center for the FY-4A products and the UNL-VRTM developers.

Author contributions

H.H. and B.Z. conceived and designed the study, implemented the radiative transfer emulation and satellite retrieval experiments, and drafted the manuscript. Y.Q., L.L. and Y.N. processed the FY-4A AGRI satellite archives, performed the sun-photometer collocation and quality control, and contributed to the validation analysis. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Kaufman, Y. J., Tanré, D. & Boucher, O. A satellite view of aerosols in the climate system. *Nature* **419**, 215–223 (2002).
2. Kaufman, Y. J. et al. Passive remote sensing of tropospheric aerosol and atmospheric correction for the aerosol effect. *J. Geophys. Res. Atmos.* **102**, 16815–16830 (1997).
3. Li, Z. et al. Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective. *Ann. Geophys.* **27**, 2755–2770 (2009).
4. Levy, R. C. et al. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* **6**, 2989–3034 (2013).
5. Hsu, N. C. et al. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *J. Geophys. Res. Atmos.* **118**, 9296–9315 (2013).
6. Lyapustin, A., Wang, Y., Korkin, S. & Huang, D. MODIS Collection 6 MAIAC algorithm. *Atmos. Meas. Tech.* **11**, 5741–5765 (2018).
7. Yang, J., Zhang, Z., Wei, C., Lu, F. & Guo, Q. Introducing the New Generation of Chinese Geostationary Weather Satellites, Fengyun-4. *Bull. Am. Meteorol. Soc.* **98**, 1637–1658 (2017).
8. Xie, Y. et al. Aerosol Optical Depth Retrieval Over South Asia Using FY-4A/AGRI Data. *IEEE Trans. Geosci. Remote Sens.* **60**, 4104814 (2022).
9. Gupta, P. et al. Applying the Dark Target aerosol algorithm with Advanced Himawari Imager observations during the KORUS-AQ field campaign. *Atmos. Meas. Tech.* **12**, 6557–6577 (2019).
10. Kim, M., Levy, R. C., Remer, L. A., Mattoo, S. & Gupta, P. Parameterizing spectral surface reflectance relationships for the Dark Target aerosol algorithm applied to a geostationary imager. *Atmos. Meas.*

Tech. **17**, 1913–1939 (2024).

11. Su, T., Laszlo, I., Li, Z., Wei, J. & Kalluri, S. Refining aerosol optical depth retrievals over land by constructing the relationship of spectral surface reflectances through deep learning: Application to Himawari-8. *Remote Sens. Environ.* **251**, 112093 (2020).
12. Wang, J. et al. A numerical testbed for remote sensing of aerosols, and its demonstration for evaluating retrieval synergy from a geostationary satellite constellation of GEO-CAPE and GOES-R. *J. Quant. Spectrosc. Radiat. Transf.* **146**, 510–528 (2014).
13. Vicent, J. et al. Emulation as an Accurate Alternative to Interpolation in Sampling Radiative Transfer Codes. *IEEE J. Sel. Top. Appl. Earth Obs Remote Sens.* **11**, 4918–4931 (2018).
14. Jia, C. et al. Satellite Aerosol Retrieval Using Scene Simulation and Deep Belief Network. *IEEE Trans. Geosci. Remote Sens.* **60**, 4104516 (2022).
15. Tao, M. et al. A robust and flexible satellite aerosol retrieval algorithm for multi-angle polarimetric measurements with physics-informed deep learning method. *Remote Sens. Environ.* **297**, 113763 (2023).
16. Holben, B. N. et al. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.* **66**, 1–16 (1998).
17. Giles, D. M. et al. Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmos. Meas. Tech.* **12**, 169–209 (2019).
18. Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).
19. Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. & Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **101**, 1–9 (2018).
20. Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 4540 (2020).
21. Eck, T. F. et al. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *J. Geophys. Res. Atmos.* **104**, 31333–31349 (1999).
22. Xu, X., Wang, J. & Ding, S. *User's Guide for UNL–VRTM: A Unified Linearized Vector Radiative Transfer Model, Version 1.3* (University of Nebraska–Lincoln, 2014).
23. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
24. Gao, B. C., Kaufman, Y. J., Tanré, D. & Li, R. R. Distinguishing tropospheric aerosols from thin cirrus clouds for improved aerosol retrievals using the ratio of 1.38-μm and 1.24-μm channels. *Geophys. Res. Lett.* **29**, 1890 (2002).
25. Gao, B. C. & Kaufman, Y. J. Selection of the 1.375-μm MODIS channel for remote sensing of cirrus clouds and stratospheric aerosols from space. *J. Atmos. Sci.* **52**, 4231–4237 (1995).

26. Meyer, H. & Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **12**, 1620–1633 (2021).
27. Ichoku, C. et al. A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophys. Res. Lett.* **29**, 8006 (2002).
28. Schutgens, N. et al. On the spatio-temporal representativeness of observations. *Atmos. Chem. Phys.* **17**, 9761–9780 (2017).

Figures

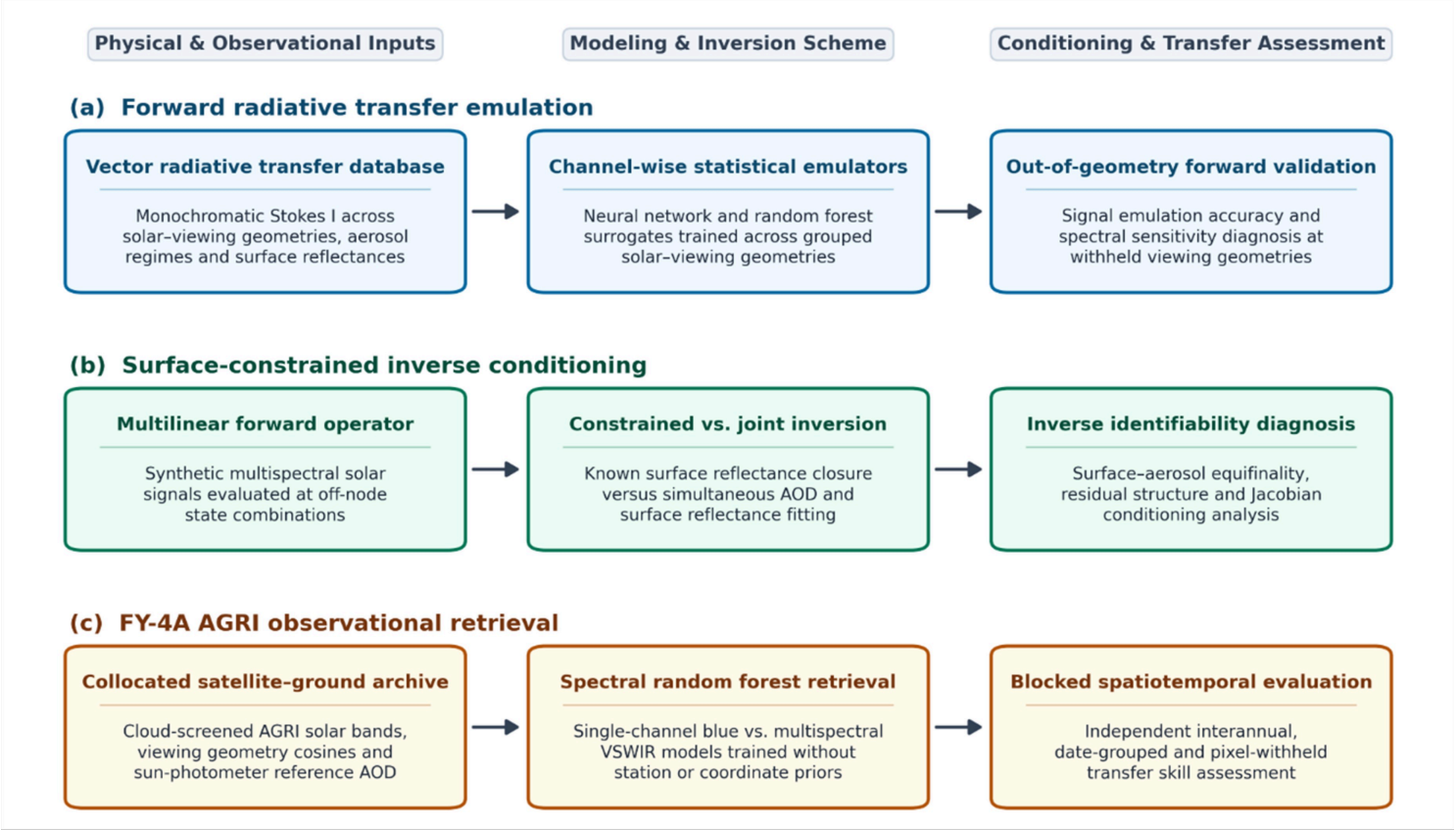

Figure 1

Methodological framework linking radiative transfer modeling, inverse conditioning and geostationary satellite retrieval evaluation. (a) Forward emulation evaluates channel-wise statistical surrogates of monochromatic Stokes *I* across withheld solar–viewing geometries in model space. (b) Synthetic off-node inversion compares aerosol optical depth (τ_{model}) recovery under known surface reflectance against simultaneous aerosol–surface fitting to diagnose parameter equifinality. (c) Observational retrieval evaluates single-channel blue and multispectral visible–shortwave infrared (VSWIR) random forest models on collocated FY-4A AGRI and ground-based sun-photometer records across independent interannual, date-grouped and pixel-withheld validation designs.

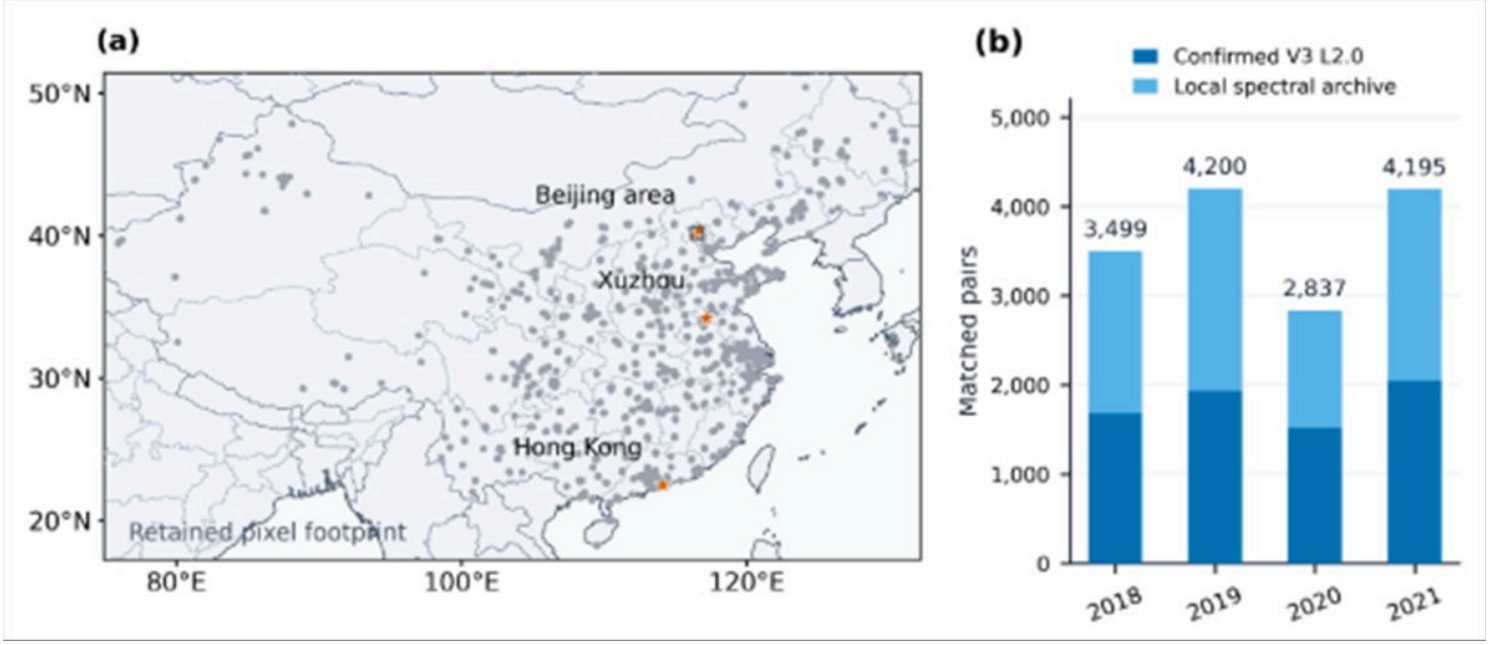

Figure 2

FY-4A observation coverage and ground collocation during 2018–2021. (a) Spatial distribution of retained 4 km pixels with at least one valid confident-clear observation (gray circles) and the three ground-collocated station regions in the Beijing area, Xuzhou and Hong Kong (orange stars); (b) annual collocated pair counts partitioned into confirmed AERONET Version 3 Level 2.0 direct-Sun measurements and pre-screened local sun-photometer spectral archive records.

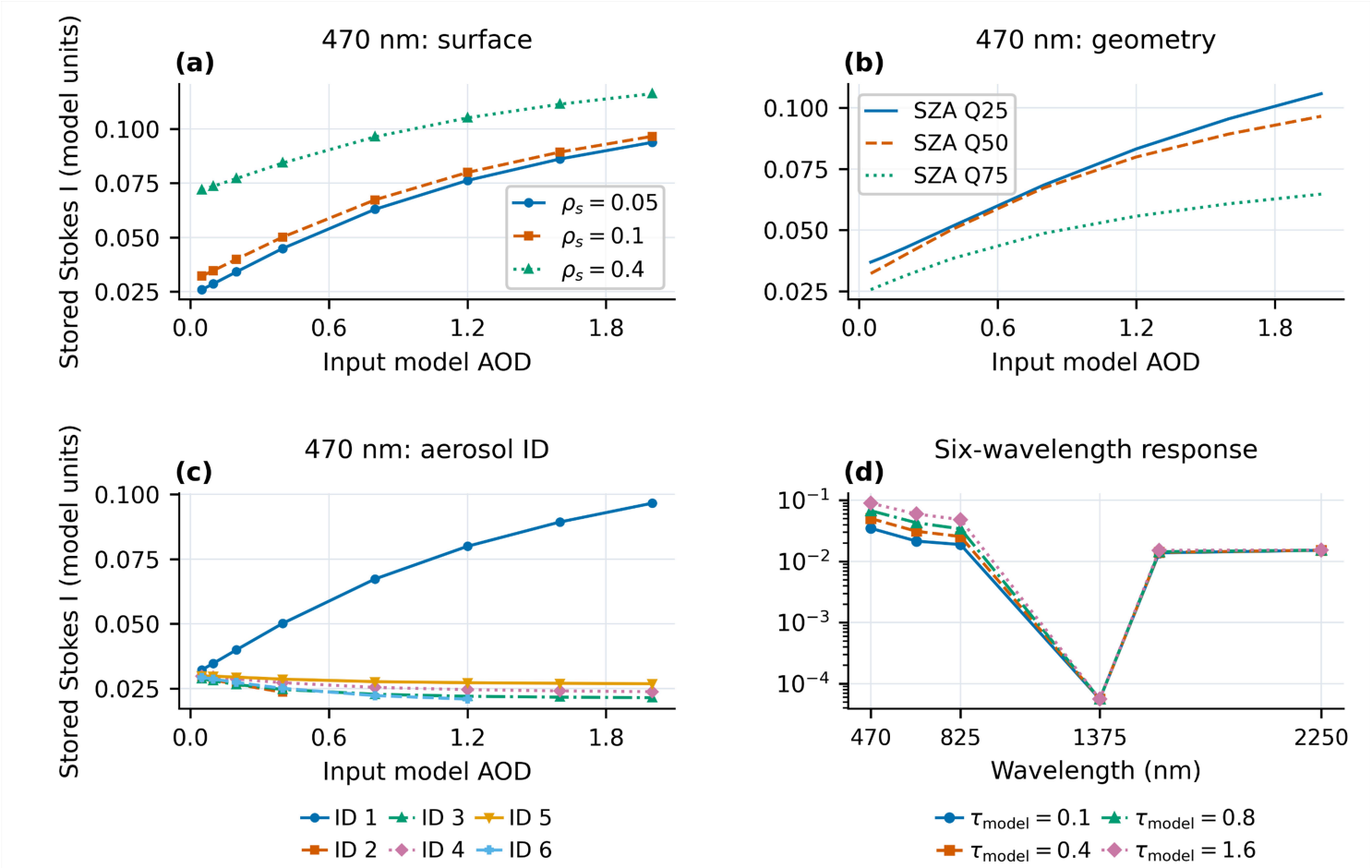

Figure 3

Conditional forward radiative response to model AOD in the UNL-VRTM lookup table. Panels (a–c) show 470 nm Stokes *I* sensitivity to Lambertian surface reflectance (ρ_s), viewing–illumination geometry and categorical aerosol optical regime (IDs 1–6), respectively. Panel (d) shows the six-channel spectral response across 470–2250 nm on a logarithmic intensity axis. Geometry tuples (SZA, VZA, RAA) correspond to the 25th, 50th and 75th solar zenith angle quartiles: Q25 (42.22°, 47.65°, 2.83°), Q50 (57.30°, 47.65°, 18.58°) and Q75 (68.69°, 47.67°, 54.21°). Baseline slices fix Q50 in (a,c,d), ρ_s = 0.1 in (b–d) and aerosol ID 1 in (a,b,d).

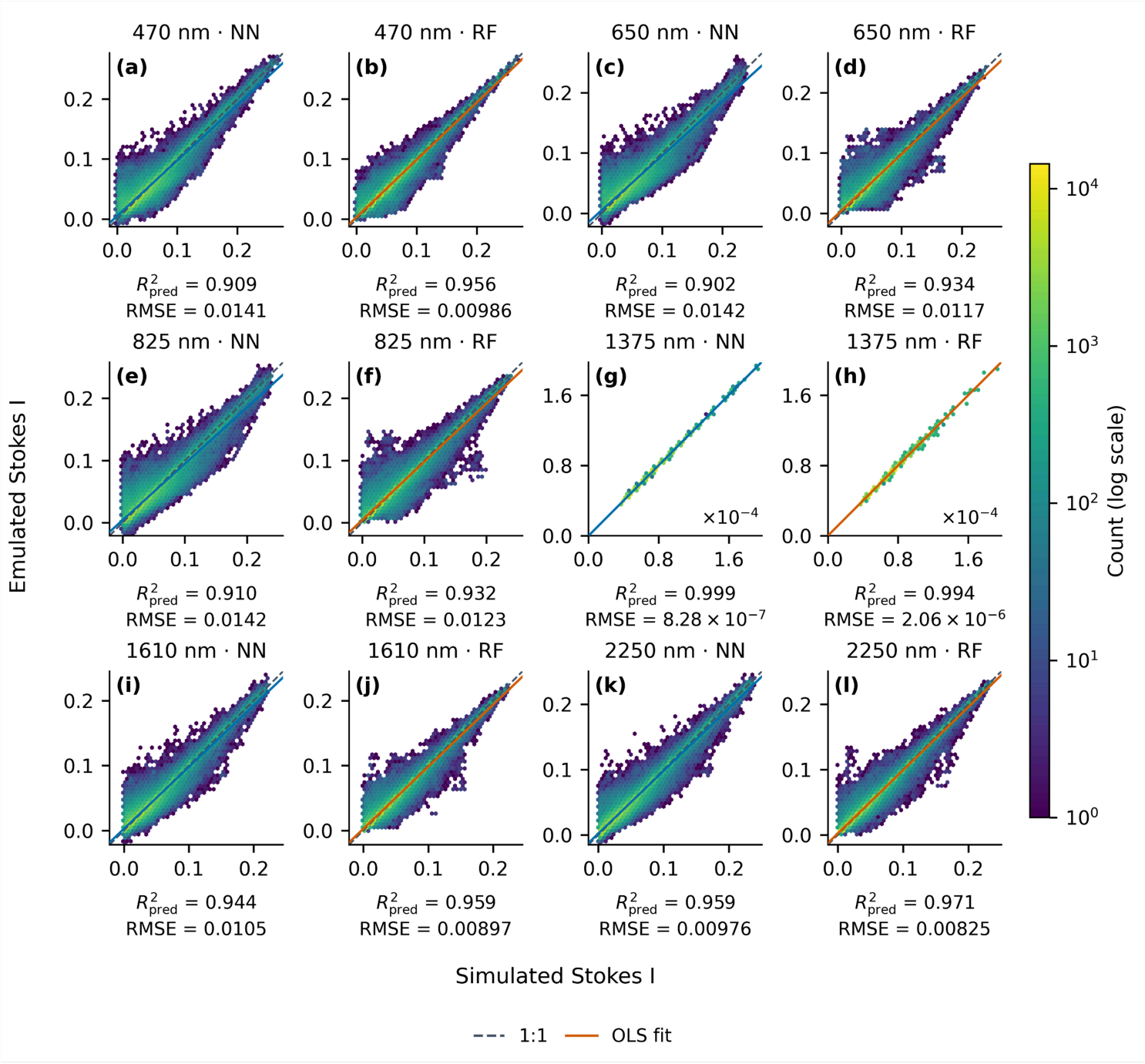

Figure 4

Geometry-held-out emulation at the six stored wavelengths. Panels (a–l) follow channel order, with NN then RF for each channel. Paired panels share test states and axis limits; count normalization is shared across panels. Dashed lines denote 1:1 agreement and colored lines are descriptive ordinary least-squares fits. RMSE is in stored Stokes-*I* units, including at 1375 nm where both axes use a × 10^{-4} multiplier.

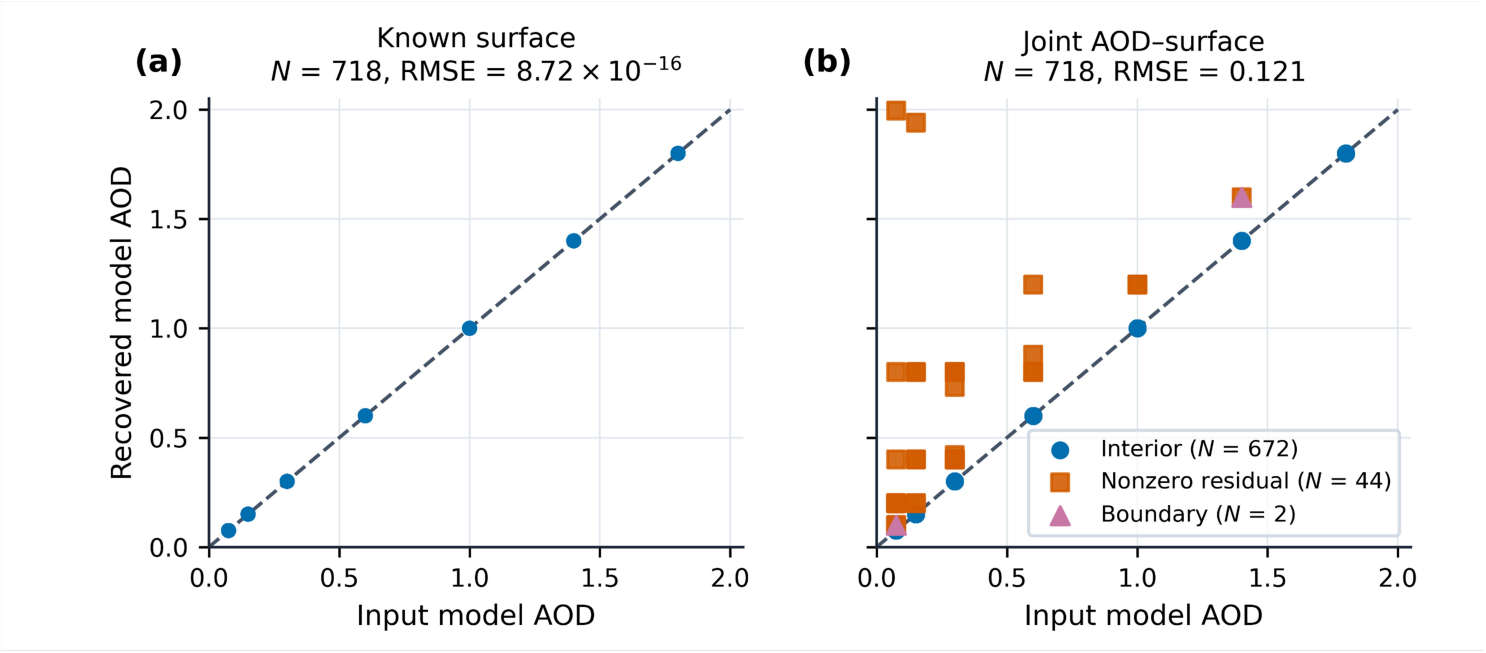

Figure 5

Noiseless off-node inverse behavior for (a) known surface reflectance and aerosol ID and (b) jointly fitted common Lambertian reflectance and AOD with known ID. All 718 supported cases are included: 672 interior closures, 44 nonzero-residual fits and two boundary solutions. A further 290 design cases have invalid interpolation corners. Generation and inversion share the bounded forward operator, so the experiment characterizes numerical conditioning in model space.

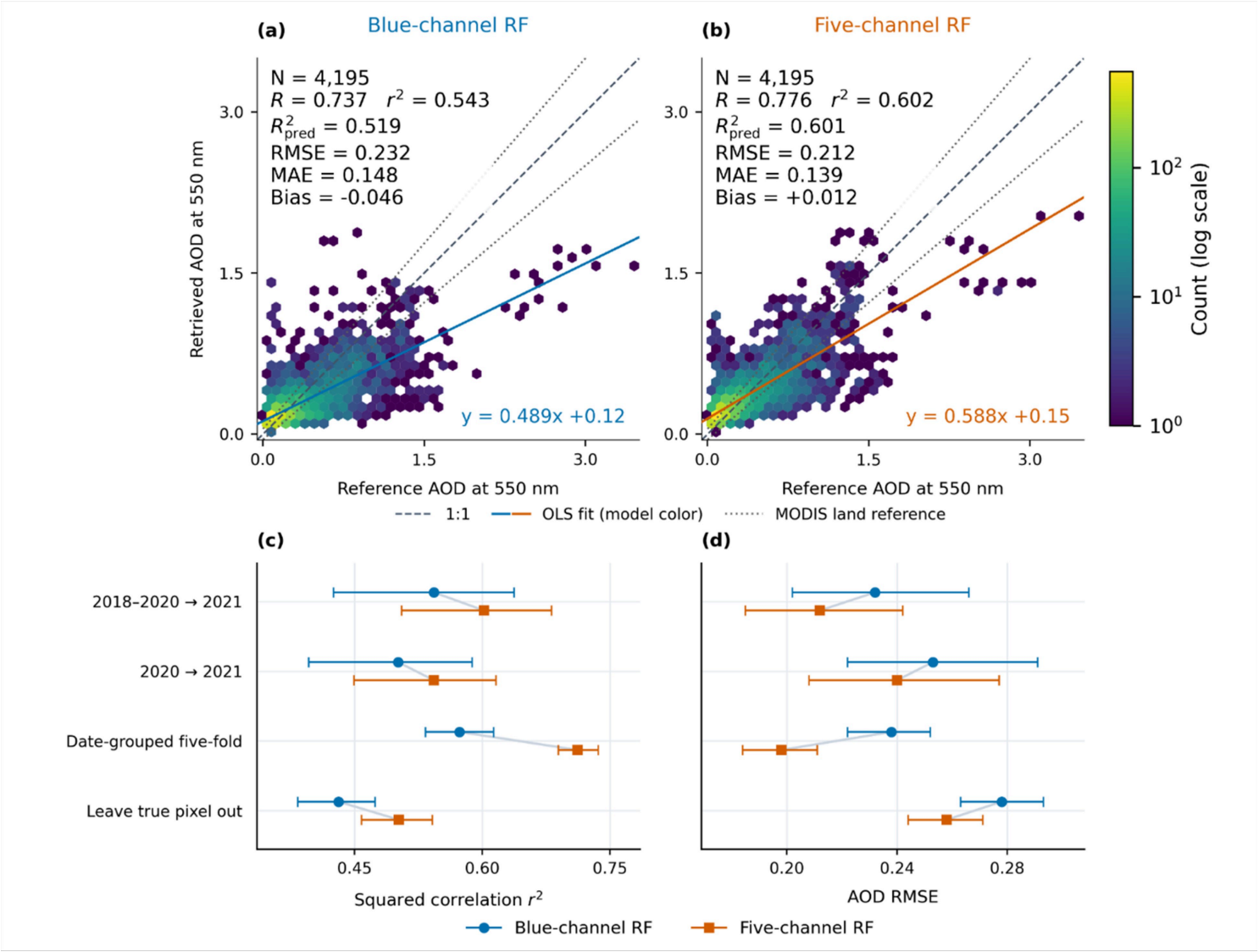

Figure 6

FY-4A temporal transfer and validation design. (a,b) The blue- and five-channel RFs trained in 2018–2020 predict the same 4,195 pairs in 2021. Dashed, colored and dotted lines show 1:1 agreement, descriptive ordinary least-squares (OLS) fits and the external MODIS Dark Target land expected-error envelope $y = x \pm (0.05 + 0.15x)$, respectively. (c,d) Squared correlation (r^2) and AOD RMSE with 95% confidence intervals from 500 UTC-date bootstrap resamples across the four evaluation designs (2018–2020 → 2021, 2020 → 2021, date-grouped five-fold and leave-one-pixel-out). The two 2021 evaluations each comprise 4,195 pairs, whereas the five date-grouped folds and six pixel holdouts each predict all 14,731 collocated pairs once.