Table S1. Sensitivity analysis of σ_0 and τ

σ_0	τ	**Pseudo-label count (×10⁶)**	**MAE**	**RMSE**	**R**
0.03	0.10	8.1	0.067	0.108	0.86
0.05	**0.10**	**9.4**	**0.061**	**0.097**	**0.89**
0.08	0.10	10.2	0.063	0.101	0.88
0.05	0.05	8.7	0.064	0.102	0.87
0.05	0.15	10.1	0.062	0.099	0.88
0.05	0.20	10.8	0.065	0.105	0.86

Table S2. Comparison of loss functions for high-AOD conditions (AOD > 1.0)

Loss function	**MAE**	**RMSE**	**Bias**	**R**
Huber loss	0.182	0.295	0.021	0.78
Weighted MSE	**0.176**	0.312	−0.046	0.75
Quantile loss (τ = 0.7)	0.189	0.301	**0.008**	**0.73**

Table S3. Detailed information about the sensitivity analysis of key parameters.

δ	N_{window}	$NDVI_{thresh}$	*MAE*	*RMSE*	*R*	*CCC*	*MAPD*
0.05	3x3	0.15	0.071	0.112	0.86	0.85	13.2
0.1	3x3	0.15	0.067	0.107	0.87	0.86	12.5
0.2	3x3	0.15	0.061	0.097	0.89	0.88	11.1
0.3	3x3	0.15	0.058	0.092	0.9	0.9	10.4
0.4	3x3	0.15	0.059	0.094	0.9	0.89	10.6
0.5	3x3	0.15	0.062	0.098	0.89	0.88	11.2
0.3	1x1	0.15	0.062	0.099	0.89	0.88	11.5
0.3	3x3	0.15	0.058	0.092	0.9	0.9	10.4
0.3	5x5	0.15	0.06	0.095	0.9	0.89	10.9
0.3	3x3	0.1	0.06	0.086	0.89	0.88	11
0.3	**3x3**	**0.15**	**0.058**	**0.086**	**0.96**	**0.90**	**10.4**
0.3	3x3	0.2	0.059	0.094	0.9	0.89	10.7
0.3	3x3	0.25	0.061	0.097	0.89	0.88	11.2
0.3	3x3	0.3	0.064	0.101	0.88	0.87	11.9

To quantify the contribution of each architectural and training design, we per-form controlled ablations experiments (Table 5), keeping all other settings fixed. Starting from a frozen-backbone baseline (RMSE = 0.176), introducing a short freeze warm-up phase modestly stabilizes optimization (RMSE = 0.172). Enabling progressive unfreezing of the backbone yields a larger improvement (RMSE = 0.163), con-

firming the importance of end-to-end adaptation to the AERONET. Building on this schedule, adding adapters improves to RMSE = 0.159, while enabling MMD-based distribution alignment alone (without adapters) achieves RMSE = 0.157, indicating that increased adaptation capacity and explicit feature alignment each provide independent benefits. The full configuration (progressive unfreezing + adapters + MMD) attains RMSE = 0.086, R=0.962, R2=0.925, within EE = 87.2%, the best across all settings, demonstrating the synergy between the schedule and the alignment objective.

Table S4. Comprehensive summary of AOD Retrieval Ablation experiment.

Variant	RMSE	R	R^2	EE%	N
(a) Backbone Frozen (no adapt)	0.176	0.879	0.773	71.3	761
(b) + Freeze Warm-up (no unfreeze)	0.172	0.887	0.786	72.2	761
(c) +Pro-Unfreeze (last 1 block)	0.163	0.902	0.813	74.7	761
(d) (c) + Adapters only	0.159	0.910	0.828	76.1	761
(e) (c) + MMD only	0.157	0.915	0.837	77.0	761
(f)Full：Prog-Unfreeze +Adapters +MMD	**0.086**	**0.962**	**0.925**	**87.2**	**761**

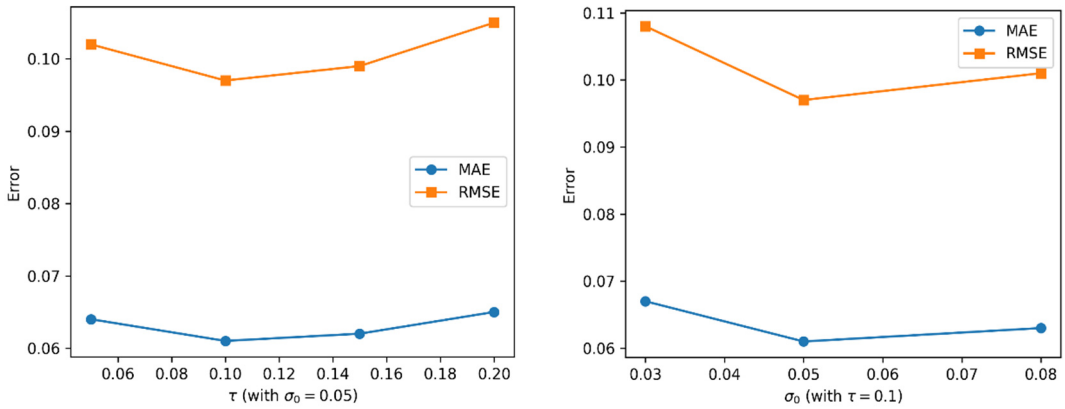


Figure S1. Sensitivity analysis of pseudo-label selection parameters. (a) Variation of MAE and RMSE with while fixing $\tau = 0.1$; (b) Variation of MAE and RMSE with while fixing $\sigma_0 = 0.05$.

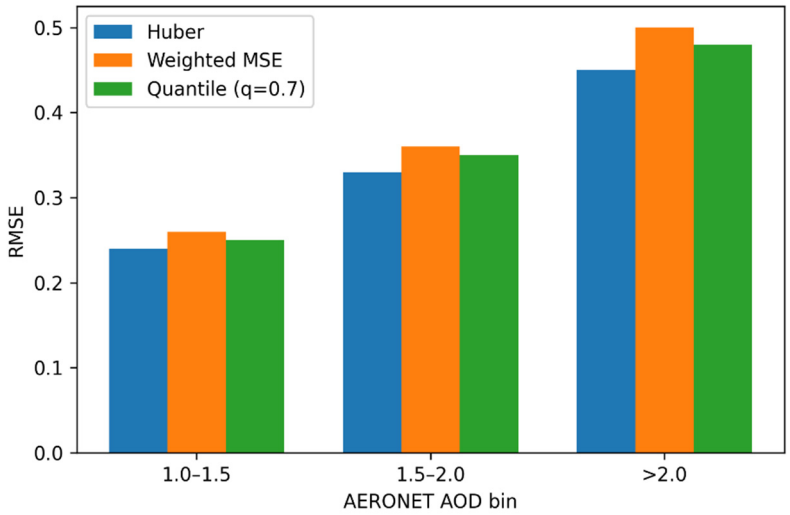


Figure S2. Comparison of loss functions for high-AOD conditions (AOD > 1.0). (b) RMSE computed within different AERONET AOD bins (1.0–1.5, 1.5–2.0, and >2.0) for models trained with different loss

functions. The binned evaluation highlights the relative fitting stability of each loss function under increasingly high-AOD regions.

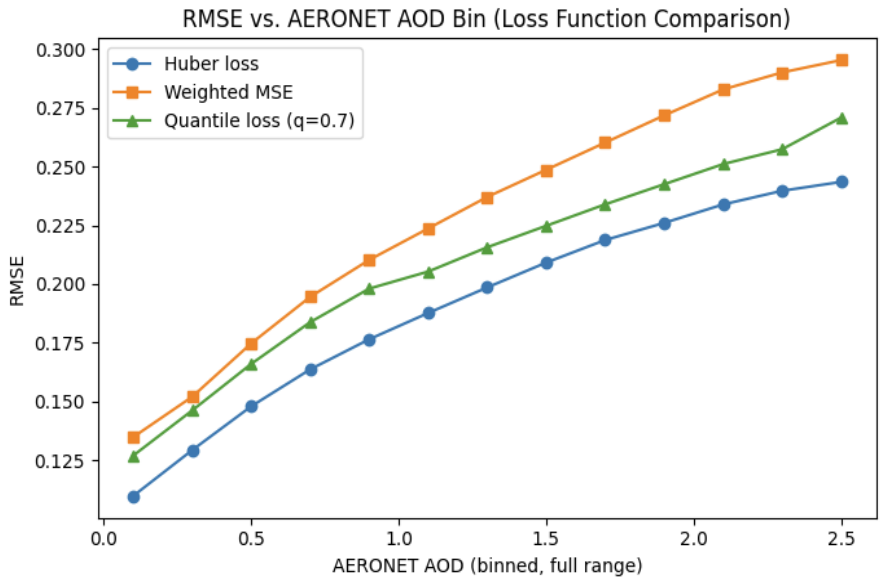
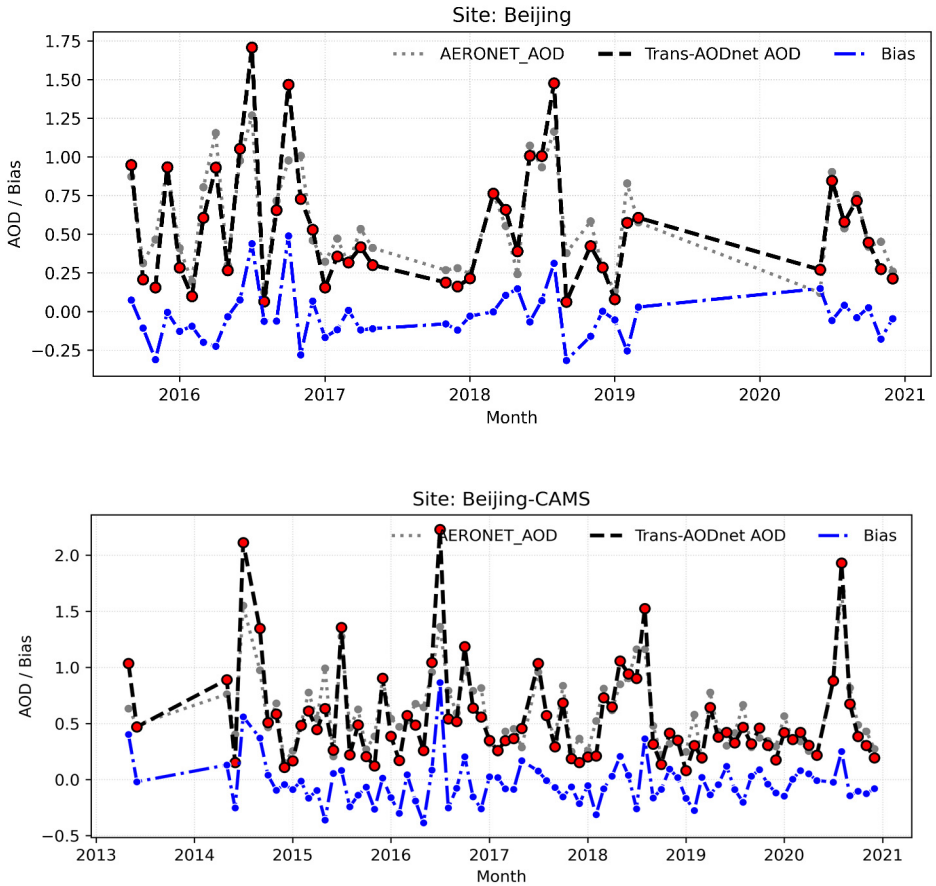


Figure S3. Line plot of RMSE as a function of binned AERONET AOD for different loss functions



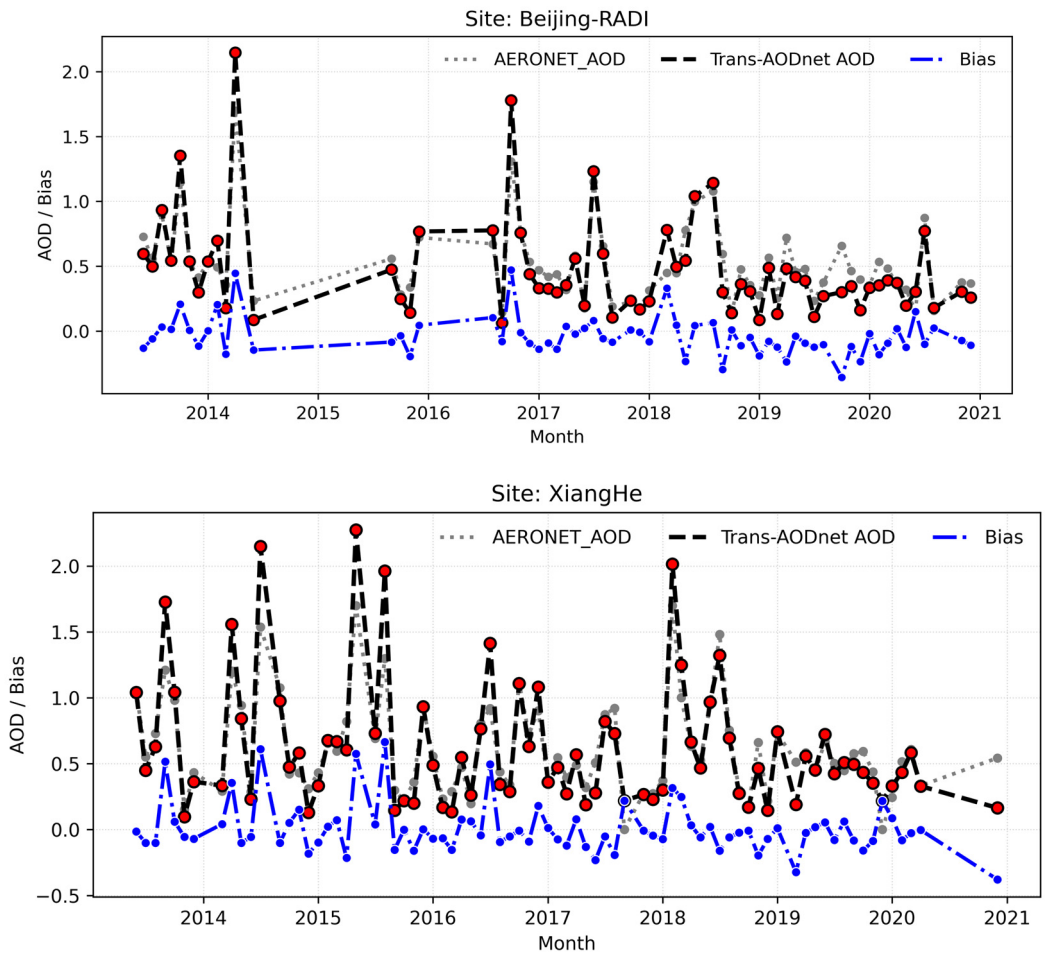


Figure S4. Site-specific monthly AOD time series comparisons between Trans-AODnet predictions and AERONET observations at Beijing, Beijing-CAMS, Beijing-RADI, and XiangHe under a time extrapolation setting (training: 2013–2018; testing: 2019–2021). Gray dotted lines denote AERONET AOD, black dashed lines with red markers indicate Trans-AODnet AOD, and blue dash-dotted lines represent the bias (predicted minus observed)

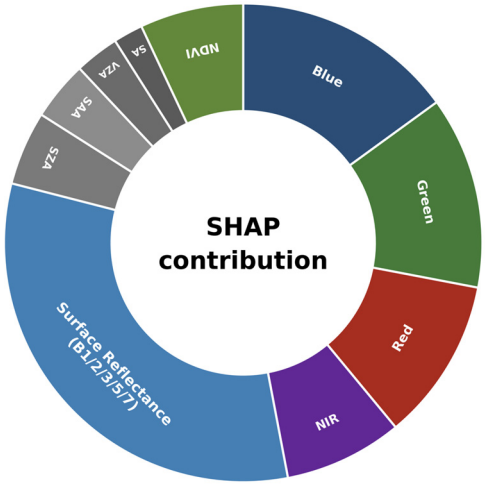


Figure S5. SHAP-based contribution of input variables in Trans-AODnet. Surface reflectance and visible–near-infrared spectral bands dominate the model response, while observation geometry parameters provide complementary constraints. Contributions are normalized by mean absolute SHAP values.

We further conducted pixel-level and spatial consistency assessments against Himawari-8 AHI AOD, as summarized in Figure 10. In total, 659 collocated samples with concurrent GF-1, MODIS, AERONET, and AHI observations were assembled. These collocations were used for pixel-to-pixel scatter–density analysis, where we show R, RMSE, and the fraction of points within EE, and for spatiotemporal trend comparisons. Over 2015–2023 (N = 659), Trans-AODnet exhibits the closest agreement with AERONET (EE = 89.9%, R = 0.966, RMSE = 0.087), outperforming both MAIAC (EE = 77.4%, R = 0.923, RMSE = 0.127) and Himawari-8 AHI (EE = 64.4%, R = 0.846, RMSE = 0.173). These results further substantiate the reliability and generalizability of the Trans-AODnet.

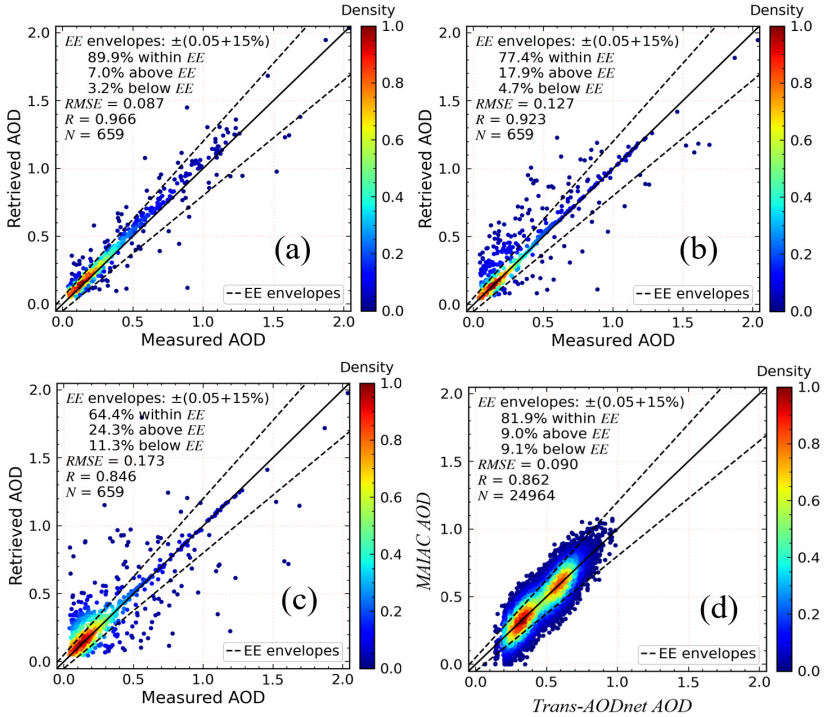


Figure S6. Spatiotemporal consistency and trend comparison among Trans-AODnet, Himawari-8 AHI, and MCD19A2 AOD. (a) Trans-AODnet AOD versus reference AERONET AOD, evaluating the GF-1-based retrievals against in situ measurements. (b) MAIAC AOD versus AERONET AOD for the same collocations, assessing the performance of the MODIS-based product relative to in situ observations. (c) Himawari-8 AHI AOD versus AERONET AOD, representing the evaluation of the geostationary AOD product against in situ measurements. (d) MAIAC AOD versus Trans-AODnet AOD, illustrating cross-dataset consistency between the two satellite-based AOD fields for all matched pixels.

To demonstrate improvements in pixel-level accuracy and cross-sensor consistency, Trans-AODnet is applied to Himawari-8 and evaluated against in-situ surface reflectance at four Beijing sites (Beijing, Beijing-CAMS, Beijing-RADI, Xianghe). Band-wise scatter plots compare Himawari-8 TOA and surface reflectance with in-situ measurements at 484, 560, 665, and 800 nm in Figure 16. Before AC, agreement is modest in the visible bands (R = 0.85; RMSE =0.06–0.08) and poor in the near-infrared (R= −0.34; RMSE = 47.39), indicating strong residual atmospheric effects. After AC, performance improved substantially across all bands (R=0.98; RMSE=0.02).

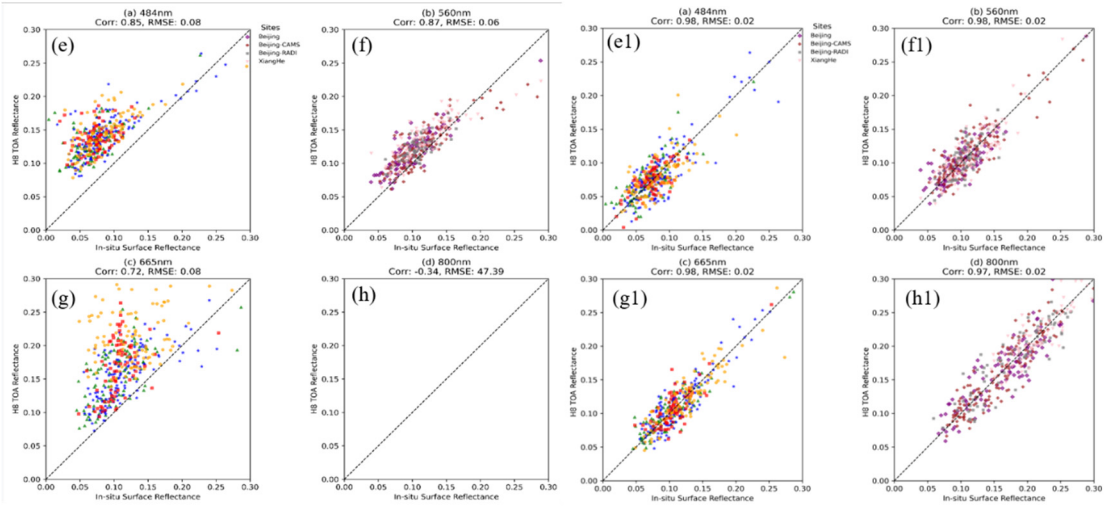


Figure S7. Scatterplot comparison of surface reflectance before and after atmospheric correction using Trans-AODnet applied to the Himawari-8 (H8) sensor. (e–h) show scatterplots comparing TOA reflectance from the Himawari-8 sensor (without atmospheric correction) with in situ surface reflectance for four spectral bands. (e1–h1) present the corresponding comparisons after atmospheric correction using Trans-AODnet derived AOD transferred to the H8 sensor. Each point represents a matched observation, with different colors indicating different sites. The dashed lines indicate the 1:1 reference.