Reply on RC2:

This manuscript by Wang et al. investigates the relative importance of aerosol chemical composition and particle size in determining the Absorption Ångström Exponent (AAE), using both ground-based observations and column-integrated AERONET data within an interpretable machine learning framework (SHAP). This study presents a novel and insightful application of machine learning to a challenging problem in aerosol science. The quantification of the relative roles of composition and size on AAE is a significant contribution. However, the manuscript requires major revisions to address the critical issue of causality in the radiative forcing analysis, clarify key methodological steps (especially the temporal matching), and provide a more integrated and critical discussion of the multi-platform datasets. With these revisions, the paper will be well-suited for publication in Atmospheric Chemistry and Physics.

Response：Thank you for your thorough review and constructive suggestions. We appreciate your recognition that our work provides a novel and insightful use of interpretable machine learning (SHAP) to quantify the relative roles of aerosol chemical composition and particle size in predicting AAE. We also appreciate your assessment that, with appropriate revisions, the manuscript can be well-suited for publication in *Atmospheric Chemistry and Physics*. We have carefully considered your major concerns and revised the manuscript accordingly.:

In the revised manuscript, we explicitly reframe the radiative forcing results as a

sensitivity analysis conditional on the adopted radiative transfer assumptions and retrieval constraints, rather than as a strict causal claim. We have expanded the Methods to provide a step-by-step description of the temporal alignment procedure between in situ observations and AERONET products (including time windows, averaging strategy, and the rationale for each choice). Discussion to more explicitly compare what each platform represents (near-surface and column-integrated), the implications for AAE interpretation. We also highlight where the two datasets converge and where they diverge, and we discuss plausible physical explanations for both. Our responses to the comments are marked in blue font and the changes to the text are marked in red in this document. Unless otherwise specified, the line numbers in this document refer to the line numbers in the revised manuscript.

Major Comments

Causality and the Role of AAE in Radiative Forcing (Lines 224-242, 424-459, and Section 4): The most significant conceptual issue lies in the framing of AAE's role in radiative forcing. The authors state that they aim to "elucidate the critical role of AAE in radiative effects" and later use SHAP to quantify AAE as a "driver" of ADRF and ARFE variations. This implies a causal relationship where AAE is an independent variable controlling radiative forcing. However, as the authors themselves expertly demonstrate in the first half of the paper, AAE is not a fundamental physical property; it is a diagnostic metric that is itself driven by the same factors that control radiative forcing: composition (BC, BrC, dust) and size. To then turn around and treat AAE as an independent "driver" of radiative effects is circular. The fundamental drivers are the

microphysical and chemical properties (e.g., BC concentration, dust loading, fine-mode radius). AAE is a valuable observational constraint precisely because it integrates these properties, but it is a consequence, not a cause. The analysis of AAE's relationship with DRF is still scientifically valuable, but its framing must be corrected. The study should not claim to quantify AAE's "driving" role. Instead, it should be framed as an investigation into the diagnostic power of AAE. The goal should be to understand how well this convenient, measurable optical parameter can explain or predict variability in radiative forcing. For example, showing that AAE is a strong predictor of TOA forcing efficiency (as in Figure 7) is a useful result: it suggests that if you can measure or constrain AAE, you have a powerful tool for estimating the aerosol's radiative effect. This is a correlative/ diagnostic relationship, not a causal one. The language throughout Sections 3.4 and 4 must be revised to reflect this, replacing terms like "driver," "regulator," and "governs" with phrases like "is associated with," "can help predict," or "serves as a key diagnostic for."

Response：Thank you for the constructive comment. We fully agree that AAE is not a fundamental causal control variable for radiative forcing. Instead, AAE should be viewed as a concise indicator of the spectral dependence of aerosol absorption, arising from underlying microphysical and chemical characteristics such as the relative contributions of BC, BrC, and dust, particle size distribution, and mixing state. These same properties, together with aerosol loading, also govern radiative transfer and therefore influence ADRF and ARFE. We acknowledge that our previous wording could be read as implying causality, and we have revised the manuscript to remove this

ambiguity.

In the revised version, we explicitly frame the radiative part of the study as an assessment of the diagnostic value of AAE_{col} for AERONET-derived radiative metrics, rather than an attempt to quantify a driving role of AAE. Specifically, we clarify that the interpretable machine-learning models quantify predictor importance. We now avoid statements that could be interpreted as "changing AAE causes a radiative response," and instead describe that AAE_{col} is or helps predict radiative variability, consistent with the fact that AAE_{col} co-varies with aerosol composition and size regime. Terms such as "driver," "regulator," and "governs" have been replaced with wording that reflects a diagnostic/predictive relationship. The key scientific value of our radiative analysis is that AAE emerges as a strong predictor of TOA forcing efficiency; thus, constraining AAE provides a practical and observation-accessible way to better estimate aerosol radiative impacts. Our revisions are as follows:

Lines 322-328 (in the revised manuscript):

Similarly, to evaluate aerosol radiative impacts, XGBoost, RF, and CatBoost models also were trained using distinct predictor sets for different radiative metrics. The AERONET ADRF and ARFE products are generated by a radiative-transfer calculation (Section 2.4); therefore, our goal is not to replace radiative transfer. Here machine-learning model is used to quantify the relative importance of AAE_{col} as a predictor of ADRF and ARFE variability, rather than implying a causal pathway where AAE_{col} independently drives ADRF and ARFE.

Lines 605-701 (in the revised manuscript):

3.4 The Diagnostic Power of Columnar AAE for Radiative Forcing and Efficiency in Beijing

Joint analysis of the boxplots and SHAP diagnostics revealed a robust, layer-dependent correlation between the AAE_{col} and ADRF. As AAE_{col} increases from 0–1 to 2–4.5, cooling at the TOA intensifies, atmospheric heating weakens, and cooling at the BOA is alleviated (Fig. 6a-6c). This pattern is consistent with a shift from more BC-like absorption toward regimes with stronger short-wavelength absorption signatures and higher scattering fractions, commonly associated with mixtures involving BrC and mineral dust. SHAP method confirm that AAE_{col} is the third strongest predictor (~16%) after AOD (~56%), and comparably to SSA (~18%) at TOA and consistently shifts ADRF toward more negative values (Fig. 6d). At BOA, AAE_{col} explains only ~4% of the model importance. BOA cooling is primarily explained by AOD (~65.0%) and SSA (~16 %) (Fig. 6e). In the ATM, AOD and SSA remain the leading predictors, while AAE_{col} still shows importance comparable to surface albedo (SA) (both ~12%) (Fig. 6f). Mechanistically, higher AAE_{col} is commonly associated with BrC and dust, which exhibit higher SSA but lower mass absorption efficiencies (MAE), thereby enhancing backscattering and solar escape (more negative TOA forcing), reducing absorption (weaker atmospheric heating), and producing a net transmission effect that mitigates BOA cooling.

To better show columnar AAE's impact on ADRF, we introduce the ARFE, which removes the scaling by aerosol loading and highlights intrinsic optical controls. At TOA, AAE_{col} serves as a key diagnostic of cooling efficiency in the model, with mean |SHAP|

reaching ~40.0%, exceeding the asymmetry factor (g), SSA, and SA even when AOD was conditioned at 25th (Fig. S11), 50th (Fig. 7), 75th percentiles (Fig. S12), or mean (Fig. S13). Larger AAE_{col} is associated with more negative TOA ARFE (Fig. 7d), indicating that, for comparable loading, regimes with steeper absorption spectra tend to exhibit stronger TOA cooling efficiency. At BOA, ARFE is predicted primarily by SSA (~50%), followed by g and SA, with AAE_{col} predicting more modestly (~8%) (Fig. 7e). In this layer, higher SSA and larger g tend to make ARFE less negative, consistent with reduced absorption and more forward-directed scattering leading to greater transmittance for a fixed AOD. In the ATM, SSA is the dominant predictor of the heating-efficiency (>50%), with AAE_{col} and SA providing secondary information (both ~17%), while **g** plays a minor role (Fig. 7f). Higher AAE_{col} is linked to lower atmospheric heating efficiency, reflecting a shift toward aerosol types with weaker mass absorption than BC, and higher SSA further suppresses in-column absorption. Overall, these results do not imply that AAE_{col} is a causal driver of radiative forcing and radiative forcing efficiency; rather, AAE_{col} acts as a compact descriptor of absorption spectral shape that co-varies with underlying composition and size regimes. The strong association between radiative forcing and ARFE therefore suggests that constraining AAE can meaningfully improve estimates of forcing efficiency in radiative assessments.

4 Conclusions

LAAs exert a strong influence on the Earth's radiation budget, yet the spectral dependence of their absorption, commonly summarized by the AAE, remains poorly

constrained in urban regions. Here we combined a winter in situ observation in Beijing with a long-term AERONET column data (2001–2019) and an interpretable machine-learning framework to quantify how composition and particle size influence AAE and to evaluate what AAE implies for radiative effects.

Near the surface in wintertime Beijing, AAE variability co-varied primarily with enhanced fractions of fine mineral dust and water-soluble inorganic ions, underscoring that non-carbonaceous species can substantially modulate local absorption spectra in addition to BC and BrC. At the column level, SHAP diagnostics identified CAI is the most informative predictor of columnar AAE, followed by BrC and BC. Among particle size metrics, the fine-mode effective radius is the leading size-related predictor and accounts for about 29% of the cumulative importance of all size parameters, whereas non-absorbing composition (coarse and fine non-absorbing dust and non-absorbing carbonaceous aerosols) played only a minor role. Because this study is based on Beijing observations, the identified predictor importance reflects a polluted urban environment influenced by both anthropogenic aerosol and episodic dust. In cleaner regions the relationships may weaken due to lower absorption signal, whereas in more dust-dominant regions the role of dust-related predictors would likely become even stronger.

For radiative impacts, our results highlight the diagnostic value of columnar AAE rather than implying a causal control. In the model trained on AERONET radiative products, columnar AAE is among the most informative predictors for TOA ADRF (~16%, comparable to SSA) and becomes the leading predictor for TOA ARFE (~40%), with higher columnar AAE associated with more efficient TOA cooling under loading-

controlled conditions. By contrast, columnar AAE contributes much less to the prediction of ATM and BOA ADRF and ARFE, where AOD and SSA remain the primary predictors.

Overall, the findings of our study demonstrate the multifactorial influences of AAE by composition and size and highlight its strong correlation with the vertical partitioning of radiative forcing, especially at the TOA. Consequently, accurately constraining AAE is essential for a realistic representation of aerosol radiation interactions in regional and global models.

Matching Offline and Online Measurements for Model Training (Section 2.2, Lines 175-186): A critical methodological detail is insufficiently explained. The multiple linear regression model in Section 2.2 uses offline chemical composition data (FMD fraction, nd-WSII fraction) matched with online AAE and size distribution data. The authors state that data were "temporally matched to the corresponding online measurements based on sampling periods" (daytime 09:00-20:30; nighttime 21:00-08:30). This averaging over ~11.5-hour and ~11.5-hour periods is a significant source of uncertainty and potential bias. Within a single daytime or nighttime filter sample, the aerosol composition, size distribution, and AAE are likely highly variable due to changes in emissions (e.g., rush hour), boundary layer dynamics, and chemistry. Assigning a single, averaged composition value to the highly temporally resolved online data within that period assumes a static relationship that may not hold. The authors must: （1）Explicitly justify why this temporal resolution is sufficient to capture the relationships they are investigating. （2）Discuss the potential for "ecological

fallacy" or averaging bias—where the relationship between variables at an aggregated level differs from the relationship at a high-resolution level. （3）Ideally, provide an uncertainty estimate for how this temporal mismatch might affect the regression coefficients and conclusions.

Response: Thank you for the comment. We have added a critical methodological detail insufficiently. First, our $PM_{2.5}$ chemical composition data are derived from integrated daytime and nighttime filter samples, representing mean composition over the sampling periods of 09:00–20:30 and 21:00–08:30, respectively. Therefore, AAE_{sfc} and the size-related parameters were averaged over the same sampling intervals as the filter samples, so that both chemistry and optical variables represent the same time-integrated period. Using hourly online values together with one integrated chemical sample would repeatedly assign the same chemistry to multiple higher-frequency optical observations, which would artificially inflate the sample size and lead to pseudo replication. For this reason, our analysis is intentionally formulated at the filter timescale, and the relationships investigated by the MLR are interpreted as 11.5 h mean composition–optics relationships, rather than hour-by-hour relationships. Besides, we note that this scale-consistent alignment is common in studies combining integrated filter chemistry with continuous optical measurements, where online optical data are averaged to the filter sampling intervals before comparison or regression analysis, such as Bernardoni et al. (2021) and Wang et al. (2021). We also acknowledge this limitation and have added a sentence in the manuscript noting that future studies would benefit from online, high-time-resolution measurements of aerosol chemical composition,

which would enable a more direct linkage between short-timescale composition variability and AAE

Second, we now explicitly acknowledge the possibility of aggregation bias (or ecological fallacy), in the sense that associations at the filter-averaged timescale may differ from those at finer temporal resolution. However, this does not invalidate the present framework, because our offline chemistry data do not resolve within-period variability and therefore do not support inference at hourly resolution. In other words, the MLR in Eq. (5) is not intended to infer high-frequency causal relationships, but to identify the dominant composition–size associations at the same temporal support as the filter chemistry. We have revised the Methods to make this scale dependence explicit.

Third, to quantify the uncertainty introduced by averaging, we added a representativeness uncertainty analysis based on the within-period variability of AAE_{sfc}. Specifically, for each day and night sampling period, we calculated the standard deviation of the hourly AAE_{sfc} values and summarized its frequency and cumulative distributions in Fig. S2. This result shows that within-period variability is generally moderate: approximately 90% of the sampling periods have a standard deviation no greater than 0.35, while larger fluctuations are limited to a relatively small tail of windows. We therefore interpret this dispersion as a representativeness uncertainty associated with window averaging, while retaining the window-mean AAE_{sfc} as the only physically consistent quantity for matching with integrated filter chemistry.

In addition, to evaluate how limited sample size and temporal averaging may affect

the regression inference, we performed a nonparametric bootstrap analysis (1000 resamples) for the surface MLR. The bootstrap results indicate that the reduced model yields stable coefficient signs and relative magnitudes, with good predictive skill for AAE_{sfc} (the coefficient of determination (R^2) = 0.75, root mean square error (RMSE) = 0.13, mean absolute error (MAE) = 0.10; Table S2). We also tested an extended model including EC and OM fractions, but its coefficients were highly unstable under bootstrap resampling (Table S1), and EC and OM were not significantly associated with AAEsfc in this campaign. We therefore retained the parsimonious formulation in Eq. (5). Our revisions are as follows:

Lines 202-236 (in the revised manuscript):

The influence of particle size and chemical composition on AAE_{sfc} was assessed using a standardized multiple linear regression:

$$\widehat{AAE_{sfc}} = a + \mathrm{b} \times \widehat{\mathrm{FMD}} + \mathrm{c} \times \mathrm{nd}\widehat{-\mathrm{WSII}} + \mathrm{d} \times \widehat{\mathrm{D}_{SMPS}} + \mathrm{e} \times \widehat{\mathrm{D}_{APS}} \qquad (5)$$

where $\widehat{AAE_{sfc}}$ denotes the standardized AAE_{sfc}; a represents the intercept term, any remaining influence not parameterized by the selected predictors is captured by the intercept term; b, c, d, and e are regression coefficients; $\widehat{\mathrm{FMD}}$, $\mathrm{nd}\widehat{-\mathrm{WSII}}$, $\widehat{\mathrm{D}_{SMPS}}$, and $\widehat{\mathrm{D}_{APS}}$ are standardized variables of FMD fraction, nd-WSII fraction, and mean diameters from SMPS and APS, respectively. To ensure consistent temporal support between offline chemistry and online optical measurements, we aggregate AAE_{sfc} (and size-related parameters) over the same sampling windows and use these window-mean values. We note that AAE_{sfc} can vary within a given sampling period; however, such within-period variability is not resolvable by the integrated filter chemistry and

therefore cannot be explicitly attributed at finer temporal resolution. To transparently characterize the associated representativeness uncertainty, we quantify the within-window dispersion of AAE_{sfc} using the standard deviation across all sampling windows and provide its frequency and cumulative distributions (Fig. S2). In particular, ~ 90% of the sampling periods show a standard deviation no greater than 0.35. This result indicates that window-mean AAE_{sfc} provides a reasonable representative value at the filter timescale. Due to power outage on 27 December 2023 and 3 January 2024, daytime data for 27 December and both daytime and nighttime data for 3 January were unavailable. In future studies, higher-time-resolution measurements of aerosol chemical composition would be valuable for more directly linking short-timescale composition variability with AAE.

Notably, to further evaluate the robustness of the regression coefficients, we conducted a nonparametric bootstrap analysis with 1000 resamples. We also tested an extended specification including EC and OM fractions as additional predictors. However, the extended model yielded highly unstable coefficient estimates under bootstrap resampling, with strong dispersion and frequent sign changes (Table S1). In contrast, the reduced model provides stable and physically interpretable coefficients for the key predictors and demonstrates good predictive skill for AAE_{sfc} (the coefficient of determination (R^2) = 0.75, root mean square error (RMSE) = 0.13, mean absolute error (MAE) = 0.10; Table S2). Consistent with these robustness results, our correlation analysis further indicates that EC and OM fractions are not significantly associated with AAE_{sfc} during this campaign (Section 3.2). Therefore, we retained the parsimonious

formulation without EC and OM fractions for subsequent analyses (Equation (5)).

Lines 109-122 in the supplementary:

Table S2. Summary statistics of standardized MLR coefficients and model performance from 1000 bootstrap resamples for the extended model specifications.

coef	mean	std	p2.5	p97.5
a	0.006633	0.069351	-0.12142	0.148516
EC	-7.5E+11	9.37E+11	-2.7E+12	9.93E+11
OM	-2.1E+12	2.57E+12	-7.3E+12	2.73E+12
FMD	-7.3E+12	9.12E+12	-2.6E+13	9.67E+12
nd-WSII	-7.1E+12	8.9E+12	-2.5E+13	9.43E+12
D_{SMPS}	0.032792	0.125148	-0.18773	0.299769
D_{APS}	0.446719	0.086028	0.300897	0.630525
R^2	0.74	0.03	0.67	0.77
RMSE	0.13	0.01	0.13	0.15
MAE	0.10	0.01	0.09	0.16

The extended model refers to the standardized MLR specification including EC and OM fractions. coef denotes the regression coefficient (including the intercept term, “a”). mean and std are the bootstrap mean and standard deviation of each coefficient across 1000 resamples. p2.5 and p97.5 are the 2.5th and 97.5th percentiles of the bootstrap distribution, respectively, forming the percentile-based 95% bootstrap confidence interval.

Table S3. Summary statistics of standardized MLR coefficients and model performance from 1000 bootstrap resamples for the reduced model specifications.

coef	mean	std	p2.5	p97.5
a	0.00	0.07	-0.13	0.14
FMD	0.35	0.17	0.04	0.71
nd-WSII	-0.16	0.17	-0.50	0.17
D_{SMPS}	-0.02	0.12	-0.24	0.26
D_{APS}	0.44	0.09	0.30	0.64
R^2	0.74	0.02	0.68	0.76
RMSE	0.13	0.01	0.13	0.15
MAE	0.10	0.01	0.09	0.11

The reduced model refers to the standardized MLR specification excluding EC and OM fractions.

Lines 128-130 in the supplementary:

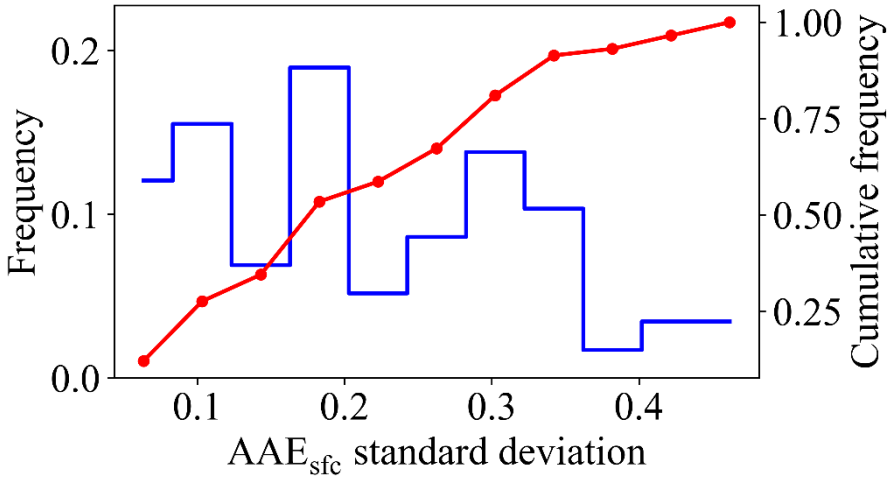

Figure S2. Frequency (blue) and cumulative frequency (red) distributions of within-window AAE_{sfc} standard deviation (11.5 h).

Disconnect Between Surface and Columnar Models (Sections 3.2 and 3.3): The manuscript essentially presents two independent modeling efforts: one for surface AAE (using MLR with 4 predictors) and one for columnar AAE (using ML models with 9 predictors). The connection between these two parts is weak. The surface analysis uses direct physical measurements, while the columnar analysis uses retrieved optical properties and a chemically inverted dataset. The manuscript would be strengthened by a more explicit discussion of how these two perspectives complement each other. For instance, does the dominant role of dust at the surface (FMD) align with the importance of CAI (coarse-mode absorbing dust) in the column? How do the limitations of one dataset inform the interpretation of the other? A dedicated paragraph synthesizing these

findings and acknowledging their different physical meanings would greatly improve the manuscript's coherence.

Response: Thank you for this helpful suggestion sincerely. In the revised version, we added a dedicated synthesis paragraph (Section 3.4) to explicitly explain how the two perspectives complement each other and why a one-to-one correspondence is not expected. Specifically, we clarify that AAE_{sfc} is derived from the wavelength dependence of the near-surface absorption coefficient $b_{abs,\lambda}$ and is interpreted together with chemically measured $PM_{2.5}$ composition (e.g., EC, OM, nd-WSII, and $PM_{2.5}$-resolved dust/FMD), whereas AAE_{col} is derived from the spectral dependence of column AAOD and its associated components (BC, BrC, CAI, CNAI, and FNAI) are optically retrieved fractions that depend on prescribed optics and mixing assumptions.

We also highlight the main cross-scale consistency: the surface predictor FMD (fine mineral dust within $PM_{2.5}$) and the column predictor CAI (absorbing dust including substantial coarse-mode contributions, ~0.6–15 μm) both indicate that dust-related short-wavelength absorption is an important contributor to elevated AAE. Finally, we discuss how the limitations of each dataset inform interpretation of the other. For example, $PM_{2.5}$ measurements under-represent coarse-mode and elevated-layer contributions captured in column retrievals, while AERONET component categories are retrieval constructs and therefore not directly comparable to measured chemical fractions (e.g., nd-WSII vs FNAI, OM vs BrC, EC vs optically defined BC). This revision strengthens the coherence of the manuscript and clarifies the complementary value of the two approaches. Our revisions are as follows:

Lines 570-604 (in the revised manuscript):

Sections 3.2 and 3.3 provide two complementary perspectives on AAE. The near-surface campaign (December 2023–January 2024) represents a specific winter pollution regime, whereas the AERONET analysis provides a longer-term perspective (2001–2019). Despite these differences, the two analyses converge on a consistent mechanistic interpretation. AAE increases when short-wavelength absorption becomes relatively stronger, and dust-related absorption plays a central role in influencing this spectral dependence. In the surface analysis, the fine mineral dust fraction within $PM_{2.5}$ is significantly associated with elevated AAE_{sfc} (Fig. 2a). In the column analysis, the absorbing dust component (CAI), which includes substantial coarse-mode contributions (radius about 0.6–15 μm), likewise ranks among the most informative predictors for AAE_{col} (Fig. 5a). Despite the different size ranges and vertical weighting, both indicators consistently support the interpretation that dust-related enhancement of short-wavelength absorption, and is linked to higher AAE.

It is also worth noting that the AAE_{col} (1.47 ± 0.56) was found to be lower than that derived from the surface field campaign (Fig. 1), but this difference should not be interpreted as a comparison between column and surface values. The two quantities differ in both temporal representativeness (multi-year climatology versus a one-month winter campaign) and measurement definition (AAOD-based column integration versus near-surface absorption coefficients), so their absolute magnitudes are expected to vary with aerosol regime, meteorology, and the contribution of elevated layers. Therefore, our emphasis is on the consistency of predicting factors and mechanisms,

rather than a direct comparison of mean values.

Finally, the two datasets complement each other in terms of strengths and limitations. The surface measurements provide chemically explicit constraints but are restricted to $PM_{2.5}$, thereby under-representing coarse-mode dust and any elevated-layer contributions. The AERONET analysis offers direct links to radiative quantities, but its component variables are retrieval-based optical constructs that depend on prescribed optics and mixing assumptions (Dubovik et al., 2000; Sinyuk et al., 2020; Li et al., 2019). As a result, several categories are not directly interchangeable (e.g., surface nd-WSII versus retrieved non-absorbing components, surface OM versus optically defined BrC, and thermal EC versus optically defined BC). Taken together, the surface campaign provides process-level chemical context for short-term variability, while the AERONET record generalizes the interpretation across regimes and links AAE to column radiative effects with dust-related absorption emerging as the clearest cross-scale consistency.

Uncertainty in AERONET-Inverted Chemical Composition (Section 2.3): The analysis of columnar AAE relies heavily on the AERONET chemical composition product (BC, BrC, CAI, etc.). It is crucial to remind readers that these are not directly measured but are retrieved from inversions of spectral sun photometer measurements, which come with their own assumptions and uncertainties. The manuscript briefly cites Zhang et al. (2024), but a more critical discussion is warranted here, especially given the central role of these data in Figure 5. What are the primary assumptions in this retrieval? (e.g., regarding refractive indices, mixing state, particle shape). What is the

estimated uncertainty for each component (BrC, BC, dust) as provided by the retrieval algorithm or the literature? A short statement acknowledging these limitations and citing key references on the uncertainties of AERONET inversions (e.g., Dubovik et al., 2000; Sinyuk et al., 2020) would provide necessary context for the robustness of the SHAP results.

Response：Thank you for the suggestion sincerely. We have revised Section 2.3 to clarify that these component fractions are derived from Sun–sky photometer inversions and a component mixing model (GRASP/Component), which assumes prescribed dry-component refractive indices and an internal-mixing rule (e.g., Maxwell–Garnett effective medium approximation), and treats fine and coarse modes separately. We also note that particle non-sphericity (relevant for dust) is handled through spheroid-based treatments in AERONET inversion schemes. Regarding uncertainty, we emphasized that absorption-related retrieval sensitivity depends strongly on aerosol loading. Our analysis already restricted the dataset to $AOD_{440} > 0.4$, which corresponds to the Level 2 AERONET V3. Moreover, we have cited Li et al. (2019) to provide a literature-based assessment of the uncertainties associated with the retrieved aerosol components. Our revisions are as follows:

Lines 256-285 (in the revised manuscript):

Notably, the column chemical components (BC, BrC, CAI, CNAI, and FNAI) used here are retrieval-based and should not be interpreted as directly measured chemical mas. They are inferred from spectral Sun–sky photometer observations through the AERONET inversion (which retrieves column-integrated size distribution and complex

refractive index from AOD and sky radiances) and a subsequent component-mixing framework (GRASP/Component) that maps the retrieved optical constraints to optically equivalent component fractions (Dubovik et al., 2000; Sinyuk et al., 2020; Li et al., 2019). In doing so, the component retrieval necessarily relies on prescribed assumptions, notably fixed complex refractive indices for the dry components, an internal-mixing rule (commonly Maxwell–Garnett effective medium approximation) to compute effective optical properties, and constraints on how absorbing components are partitioned between fine and coarse modes (Li et al., 2019). For dust, non-sphericity is treated using spheroid-based scattering models rather than purely spherical Mie theory (Dubovik et al., 2006).

These assumptions introduce additional uncertainty beyond the base AERONET inversion. As background, absorption-related AERONET inversion products (e.g., SSA/AAOD) are substantially less stable at low aerosol loading; under favorable loading conditions, SSA uncertainty is typically on the order of ~0.03, while it increases rapidly as AOD decreases (Dubovik et al., 2000; Sinyuk et al., 2020). Component volume fractions inherit this sensitivity and, in addition, respond to uncertainties in prescribed component optics and mixing rules. Sensitivity tests in the GRASP/Component literature indicate that, for $AOD_{440} \geq 0.4$ and sufficiently non-negligible component fractions, the uncertainty in retrieved BC, CAI, FNAI, and CNAI volume fractions is commonly within ~50%, whereas BrC generally remains more uncertain at low BrC fractions but can approach the ~50% level when BrC becomes a substantial contributor (Li et al., 2019). Nevertheless, this approach has been applied

by Zhang et al. (2022) and Zhang et al. (2024), who obtained reliable aerosol chemical-component information from remote-sensing measurements. To reduce the low-loading regime where absorption and component retrievals are most uncertain, we restricted our analysis to $AOD_{440} > 0.4$.

Minor Comments

1. Line 89-90: The phrasing "an ensemble of models was initially trained, after which the optimal model was selected" is slightly ambiguous. Clarify that you trained multiple model types and selected the best-performing one (CatBoost) for the final interpretation, as described later in Section 2.4. This is good practice, but the wording could be more precise.

2. Line 245-247: The acronyms SMPS and APS are used but were introduced in Section 2.1.1. Since this is the start of the Results section, it might be helpful to briefly re-introduce them as "fine-mode (SMPS) and coarse-mode (APS) particle sizers" for readers who may not have the methods section fresh in mind.

3. Line 340-341: The sentence "The AAEcol (1.47±0.56) was also suggested to be greater than that derived from the surface field campaign" is a bit awkward. Replacing "was also suggested to be" with "was found to be" or "was also higher than" would be clearer.

4. Line 384: "explaining ~50% of model performance." It would be more precise to say "explaining ~50% of the model's predictive power (as measured by mean absolute SHAP value)" or something similar, as SHAP importance sums to the total model output, not necessarily a performance metric like R^2.

5. Figures 6 and 7: The box plots in (a-c) are very effective. The SHAP summaries in (d-f) are informative. Consider adding the sample size (n) for each AAE bin in the box plots to give the reader a sense of the statistical robustness of each category.

6. Line 225: Change "in determining" to "in predicting" or "in explaining the model's estimation of."

7. Line 424: Change "revealed a robust, layer-dependent coupling" to "revealed a robust, layer-dependent correlation."

8. Line 477: Change "is a key regulator of" to "is a strong predictor of" or "contains valuable information for estimating."

9. Line 484-486: Change "demonstrate the multifactorial control of AAE" to "demonstrate the multifactorial influences on AAE" and "highlight its pivotal role in partitioning radiative forcing" to "highlight its strong correlation with the vertical partitioning of radiative forcing."

Response: Thank you for these helpful editorial suggestions. We have revised the manuscript accordingly to improve clarity, precision of wording, and consistency. The main changes include: (i) clarifying that multiple model types were trained and CatBoost was selected as the best-performing model for interpretation; (ii) briefly re-introducing SMPS and APS at the start of the Results section; (iii) improving several sentences to avoid ambiguous or causal wording; (iv) refining SHAP-related phrasing to accurately describe predictive attribution rather than “model performance”; and (v) adding sample sizes to the AAE-bin box plots in Figs. 6–7. Besides, during revision, we rechecked the statistics shown in Fig. 1b and found that the previously reported

mean value had been inadvertently recorded as the mean AAE of carbonaceous aerosol rather than the mean value of the observed AAE_{sfc} distribution. As a result, the value originally given as 1.28 ± 0.39 was incorrect. This has now been corrected to 1.64 ± 0.32. Specific revisions are as follows:

Lines 109-111 (in the revised manuscript): Subsequently, we train an ensemble of machine-learning models to predict columnar AAE. The best-performing model (CatBoost) is selected for the final prediction, as described later in Section 2.4.

Lines 359-363 (in the revised manuscript): In Fig. 1a, the stacked bars show the window-resolved $PM_{2.5}$ mass fractions of non-dust water-soluble ions (nd-WSII), fine mineral dust (FMD), organic matter (OM), and elemental carbon (EC), overlaid with AAE_{sfc} and the mean particle diameters derived from the fine-mode (SMPS) and coarse-mode (APS) measurements.

Lines 583-585 (in the revised manuscript): It is also worth noting that the AAE_{col} (1.47 ± 0.56) was found to be lower than that derived from the surface field campaign (Fig. 1), but this difference should not be interpreted as a comparison between column and surface values.

Lines 514-516 (in the revised manuscript): BrC was second (18.5%) and BC was third (13.9%), together with CAI explaining ~50% of model's predictive power (as measured by mean absolute SHAP value)

Lines 624-634 (in the revised manuscript):

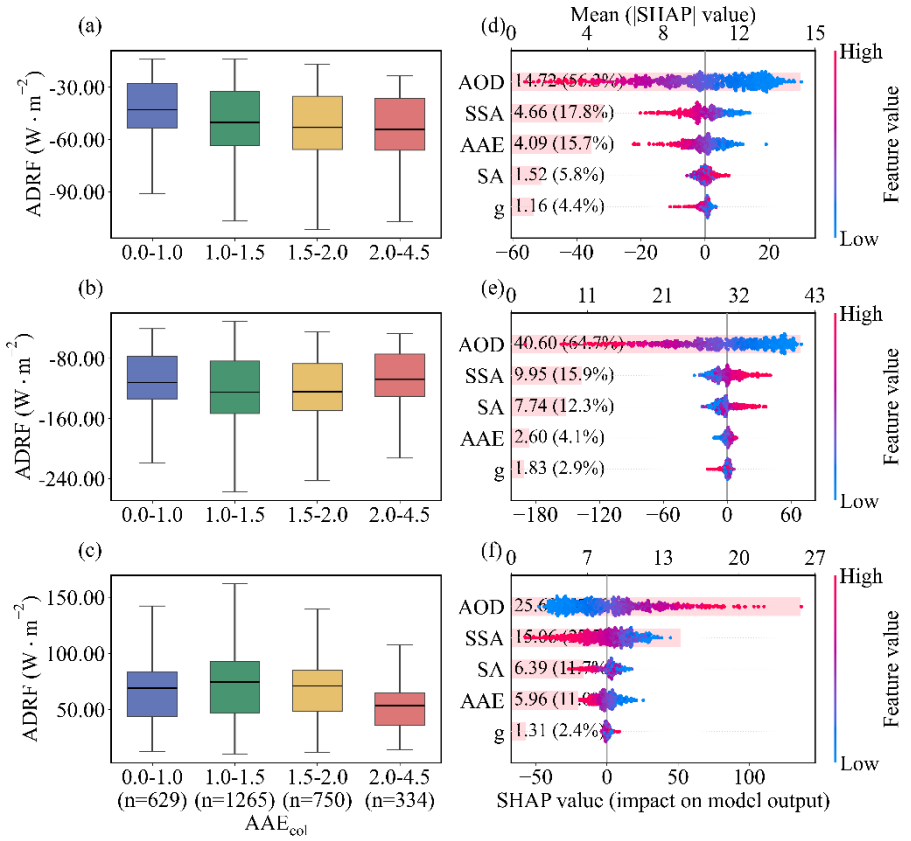

Figure 6. The relationship between columnar AAE (AAE_{col}) and aerosol direct radiative forcing (ADRF). (a–c) Box plots of ADRF at the top of the atmosphere (a), bottom (b), and in the atmosphere (c) as a function of AAE_{col}. The sample sizes for the AAE_{col} bins 0.0–1.0, 1.0–1.5, 1.5–2.0, and 2.0–4.5 are n = 629, 1265, 750, and 334, respectively, and are identical for panels (a)–(c). (d–f) SHAP analysis quantifies the relative contributions of aerosol optical depth (AOD), single scattering albedo (SSA), asymmetry parameter (g), surface albedo (SA) and AAE_{col} in predicting ADRF variations at the top of the atmosphere (d), at the bottom of the atmosphere (e), and in the atmosphere (f). The mean absolute SHAP values (numbers in parentheses) indicate the relative contribution of each predictor to the model output.

Lines 650-660 (in the revised manuscript):

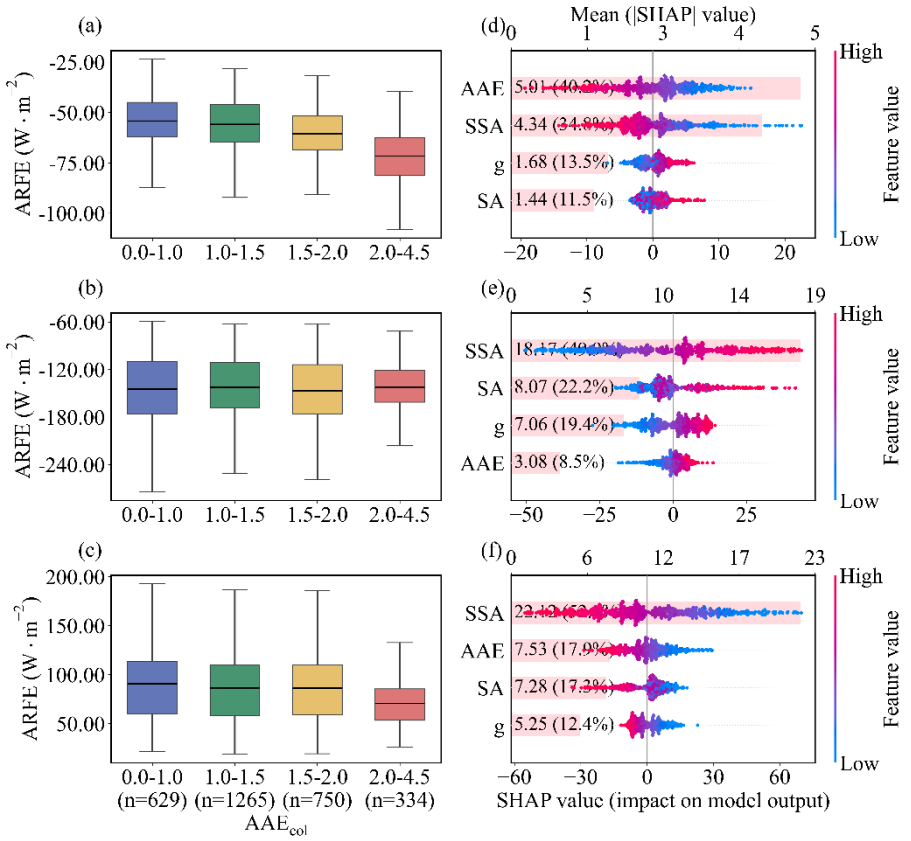

Figure 7. The relationship between columnar AAE (AAE_{col}) and aerosol radiative forcing efficiency (ARFE). (a–c) Box plots of ARFE at the top of the atmosphere (a), bottom (b), and in the atmosphere (c) as a function of AAE_{col}. The sample sizes for the AAE_{col} bins 0.0–1.0, 1.0–1.5, 1.5–2.0, and 2.0–4.5 are n = 629, 1265, 750, and 334, respectively, and are identical for panels (a)–(c). (d–f) SHAP analysis with AOD fixed at its median (50th percentile) quantifies the relative contributions of single scattering albedo (SSA), asymmetry parameter (g), surface albedo (SA) and AAE_{col} in predicting ARFE variations at the top of the atmosphere (d), bottom (e), and in the atmosphere (f). The mean absolute SHAP values (numbers in parentheses) indicate the relative contribution of each predictor to the model output.

Lines 318-321 (in the revised manuscript): Furthermore, the SHAP analysis was applied to decompose the model output into additive feature contributions, enabling

quantitative assessment of the relative contribution and sensitivity of individual aerosol composition and size parameters in predicting.

Lines 607-608 (in the revised manuscript): Joint analysis of the boxplots and SHAP diagnostics revealed a robust, layer-dependent correlation between the AAE_{col} and ADRF.

Lines 690-694 (in the revised manuscript): In the model trained on AERONET radiative products, columnar AAE is among the most informative predictors for TOA ADRF (~16%, comparable to SSA) and becomes the leading predictor for TOA ARFE (~40%), with higher columnar AAE associated with more efficient TOA cooling under loading-controlled conditions.

Lines 697-699 (in the revised manuscript): Overall, the findings of our study demonstrate the multifactorial influences of AAE by composition and size and highlight its strong correlation with the vertical partitioning of radiative forcing, especially at the TOA.

References:

Bernardoni, V., Ferrero, L., Bolzacchini, E., Forello, A. C., Gregorič, A., Massabò, D., Močnik, G., Prati, P., Rigler, M., Santagostini, L., Soldan, F., Valentini, S., Valli, G., and Vecchi, R.: Determination of Aethalometer multiple-scattering enhancement parameters and impact on source apportionment during the winter 2017/18 EMEP/ACTRIS/COLOSSAL campaign in Milan, Atmos. Meas. Tech., 14, 2919–2940, https://doi.org/10.5194/amt-14-2919-2021, 2021.

Li, L., Dubovik, O., Derimian, Y., Schuster, G. L., Lapyonok, T., Litvinov, P., Ducos,

F., Fuertes, D., Chen, C., Li, Z., Lopatin, A., Torres, B., and Che, H.: Retrieval of aerosol components directly from satellite and ground-based measurements, Atmos. Chem. Phys., 19, 13409–13443, https://doi.org/10.5194/acp-19-13409-2019, 2019.

Wang, Q., Liu, H., Ye, J., Tian, J., Zhang, T., Zhang, Y., Liu, S., and Cao, J.: Estimating Absorption Ångström Exponent of Black Carbon Aerosol by Coupling Multiwavelength Absorption with Chemical Composition, Environ. Sci. Technol. Lett., 8, 121–127, https://doi.org/10.1021/acs.estlett.0c00829, 2021.