Reply on CC1:

This paper conducts a detailed study on the influencing factors of ground and atmospheric column AAE, revealing distinct patterns of influence. However, it is evident that there is a lack of correlation and comparison between the studies on the ground and atmospheric column. Furthermore, a significant portion of the text is devoted to discussing the SHAP of TOA, ATM, and BOA with respect to atmospheric column optical properties. However, are TOA, ATM, and BOA calculated based on atmospheric column optical properties and radiative transfer models? Therefore, SHAP is likely just a data-driven decomposition and description of the traditional radiative transfer simulation process. In summary, I suggest a major revision.

Response: Thanks to your constructive comments. We have improved our manuscript in the following two aspects: (1) Strengthening the explicit linkage between ground and column AAE, rather than presenting them as two parallel but weakly connected analyses; (2) Clarifying the computational chain for TOA, ATM, and BOA radiative effects. We have also implemented a major revision focusing on a new cross-scale comparison section and a clearer methodology and interpretation for the SHAP analysis of radiative effects in the revised manuscript. 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:

1. It is necessary to enhance the comparison between ground and atmospheric column AAE, especially to explain why carbonaceous aerosols and dust are not the main driving factors on the ground.

Response: Thank you for the suggestion. We have strengthened the comparison between surface and columnar AAE by adding a dedicated synthesis section (Section 3.4) that clarifies their different definitions, representativeness, and composition metrics, and by expanding the discussion of surface factors. In addition to the cross-scale synthesis, we have strengthened the discussion of the controlling factors for near-surface AAE in Section 3.2, providing a clearer interpretation of why certain predictors (e.g., dust-related fraction and size metrics) emerge as important while carbonaceous fractions show limited explanatory power during the campaign. Our revisions are as follows:

Lines 405-467 (in the revised manuscript):

The EC mass fraction shows no correlation with AAE_{sfc} ($r = 0.09$, $p = 0.49$; Fig. 2a). This is plausible because EC is an operational thermal fraction and does not directly represent the optically effective BC absorption, which can be substantially modified by mixing state and coating (Petzold et al., 2013). Similarly, the OM mass fraction is not significantly correlated with AAE_{sfc} ($r = -0.11$, $p = 0.40$; Fig. 2a). In contrast to study dominated by biomass burning, where light-absorbing organic carbon can account for > 50% of the mass fraction and strongly enhance AAE (Wang et al., 2021). During the Beijing campaign, however, OM contributes only ~19% of total $PM_{2.5}$ mass and BrC fractions therefore are relatively low. Although BrC exhibits intrinsically high AAE values (Laskin et al., 2015; Moosmüller et al., 2011), its impact is diminished in the mixed aerosol matrix due to the influence of other dominant compositions.

We observed a statistically significant negative correlation between AAE_{sfc} and carbonaceous aerosol AAE (AAE_{CA}) (Fig. 2b), indicating that the non-carbonaceous aerosol had a significantly stronger role in shaping the absorption spectral dependence under complex pollution conditions. Due to nitrogen dioxide (NO_2) concentrations were elevated at night (Fig. S8), which can interfere with PAX instruments, particularly at shorter wavelengths (Arnott et al., 2000; Gyawali et al., 2012). Therefore, we restrict the analysis here to daytime data (Fig S9). This pattern therefore cannot be ascribed simply to inter-instrument discrepancies.

AAE_{sfc} exhibites a significant positive correlation with the mass fraction of FMD ($r = 0.79$, $p < 0.01$) and a negative correlation with nd-WSII ($r = –0.78$, $p < 0.01$) (Fig. 2a). The AAE_{sfc} enhancement associated with FMD can be attributed to metal oxides such as hematite and goethite, which strongly absorb in the UV wavelengths and steepen the spectral dependence (Bi et al., 2016). By contrast, nd-WSII (mostly sulfates, nitrates, and ammonium particles) primarily behaves as a weakly absorbing (nearly scattering-only) component in the visible–near-infrared (Seinfeld & Pandis, 2016), and an increase in its mass fraction therefore tends to dilute the contribution of absorbing species to total $PM_{2.5}$ absorption. In our surface dataset, this dilution effect is expected to reduce the relative importance of short-wavelength absorbers and, in turn, weaken the apparent wavelength dependence of bulk absorption, leading to lower AAE_{sfc} when nd-WSII dominates. We note that a “lensing effect” associated with non-absorbing coatings has been reported to enhance AAE_{CA} (Cappa et al., 2012; Zhang et al., 2025). However, the carbonaceous components contributed only a small fraction of $PM_{2.5}$ mass during our campaign. Consequently, any potential lensing-related enhancement was likely too small relative to the total aerosol and variability to yield a detectable positive correlation between AAE_{sfc} and the nd-WSII mass fraction. In this regime, nd-WSII is better interpreted as a marker of secondary inorganic aerosol loading that mainly increases scattering and dilutes absorber fractions.

Particle size also plays a critical role. AAE_{sfc} is negatively associated with the fine-mode mean diameter from SMPS (D_{SMPS}, r = –0.58; Fig. 2c) and positively associated with the coarse-mode mean diameter from APS (D_{APS}, r = 0.58; Fig. 2d). Here, the correlation analysis is used as an exploratory step to describe these first-order relationships, whereas the standardized multiple linear regression (MLR) estimates the multivariate associations after accounting for predictor covariation. Consistent with the bivariate results, the MLR yields a negative standardized coefficient for D_{SMPS} (–0.02) and a positive coefficient for D_{APS} (0.44), confirming that the coarse-mode size metric provides the stronger size-related contribution in the multivariate setting.

The composition terms show a similarly coherent pattern across the two analyses. FMD is positively associated with AAE_{sfc} in the bivariate correlations and remains positive in the MLR (0.35), whereas nd-WSII shows a negative association and remains negative in the MLR (–0.16). Importantly, particle size and composition are not independent in this winter dataset. Periods with larger coarse-mode diameters tend to coincide with enhanced fine mineral dust fraction (r=0.64; Fig. S10), consistent with stronger dust influence. Conversely, periods characterized by smaller fine-mode diameters are associated with elevated nd-WSII fraction (r=0.89; Fig. S10), consistent with secondary inorganic build-up and hygroscopic growth that increase scattering and dilute the relative contribution of absorbing components. Together, these results indicate that higher AAE_{sfc} is associated with a regime of larger particles and stronger dust contribution, whereas lower AAE_{sfc} occurs when secondary inorganic matter is more influential and dust contributions are reduced. Overall, AAE_{sfc} is influenced not only by carbonaceous aerosols, but also strongly by other chemical components, particularly mineral dust-related particles, non-dust water-soluble inorganic ions, and particle-size distributions.

Lines 569-604 (in the revised manuscript):

3.4 The comparison between surface and columnar AAE.

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.

2. The article initially employs many correlations, followed by regression analysis using Multiple Linear Regression (MLR) for ground-based AAE, while the atmospheric column is studied using seven interpretable machine learning models for AAE and six radiative factors. There is considerable overlap in data analysis functions among correlations, MLR, and machine learning. Firstly, why isn't interpretable machine learning used for ground-based AAE? The same method makes it easier for readers to compare driving patterns between ground and column AAE.

Response: Thank you for the suggestion that applying the same interpretable machine-learning framework to both surface and column datasets could facilitate comparison. In our study, however, the near-surface chemical predictors are constrained by offline $PM_{2.5}$ filter sampling, resulting in 58 valid surface samples for the multivariate analysis. With this limited sample size, machine-learning models can produce unstable feature rankings and may arbitrarily split importance across correlated variables, making SHAP attribution less robust for the surface case.

Therefore, we adopted a standardized multiple linear regression (MLR) for AAE_{sfc} to quantify the relative strength of the main factors, while using bivariate correlations primarily as an exploratory step to illustrate first-order relationships and guide predictor selection. Overall, correlation describes first-order associations; MLR estimates multivariate associations under covariation. To improve transparency and address uncertainty in the regression inference, we additionally performed bootstrap resampling (1000 replicates) of the 58 windows and now report the bootstrap mean coefficients and 95% confidence intervals. In contrast, the AERONET 2001–2019 record provides a much larger sample size and broader regime coverage, for which machine-learning is appropriate to capture nonlinearities and interactions. Our specific revisions can be found in response to your first (Lines 406-468) and third comment (Lines 202-237).

3. Why does the ground-based MLR directly state in the methods section that black carbon, brown carbon, and dust are not included or as the "intercept term" (Lines 179-180)? The apparent reason for exclusion may stem from the correlation analysis in the results section (e.g., Section 3.2). I am not denying your viewpoint, on the contrary, I find it very interesting. Unfortunately, the author did not focus on analyzing possible mechanisms and instead conducted extensive repetitive analysis and modeling.

Response: Thanks to your comment, we have revised the Methods to better justify the predictor selection for the ground-based MLR. There seems to be a misunderstanding in this comment: black carbon and brown carbon are not included but dust is included in our original manuscript. Specifically, the $PM_{2.5}$-resolved fine mineral dust (FMD) fraction showed the strongest positive association with AAE_{sfc} (e.g., Fig. 2a) and is retained as a key predictor in the MLR.

We have tested an extended MLR formulation that included EC fraction and OM fraction in addition to the predictors used in the reduced model. In practice, however, the extended specification produced poor and non-interpretable coefficient estimates, suggesting that the regression became ill-conditioned when these additional fraction-type predictors were introduced.

To address your concern and to ensure that our inference is robust, we conducted a nonparametric bootstrap with 1000 resamples for both model specifications (reduced model without EC/OM and extended model with EC/OM), refitting the regression for each resample and summarizing the coefficient distributions using percentile-based 95% intervals. The bootstrap analysis confirms that the reduced model yields stable and physically interpretable estimates for key predictors. In contrast, the extended model remains highly unstable under bootstrap resampling: the coefficients associated with EC and OM (and several other compositional predictors) exhibit orders-of-magnitude inflation ($\sim 10^{12}$) and frequent sign reversals, with 95% bootstrap intervals spanning both negative and positive values, indicating that component-wise attribution is not robust in the extended specification. These results are now provided in the Supplementary Material (Table S2, S3).

In addition, our correlation analysis shows that EC and OM fractions are not significantly associated with AAE in this dataset ($p > 0.05$; Section 3.2), implying limited explanatory power for AAE in the ground-based MLR. Therefore, combining (i) the lack of significant correlation and (ii) the bootstrap evidence of coefficient instability when included, we retain the MLR without EC/OM. In this final formulation, the intercept is used in the standard regression sense to represent the baseline and residual influences not explicitly parameterized, rather than implying that EC/OM are physically equivalent to the intercept. Our revisions are as follows:

Lines 202-209 (in the revised manuscript):

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

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

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.

Lines 225-236 (in the revised manuscript):

Notably, to further evaluate the robustness of the regression coefficients, we conducted a nonparametric bootstrap analysis with 1000 resamples. We also tested an extended model 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.

4. The interpretable machine learning for the six radiative quantities seems (e.g., Figs. 6-7) to merely describe the traditional radiative transfer simulation process via a data-driven machine learning method. It is suggested to move it to the Supplementary Information (SI), with only the main conclusions appearing in the main text. Is this comment correct? Perhaps you need to clarify how radiation variables are obtained, rather than just relying on literature, why machine learning is applicable, and what new discoveries machine learning has made.

Response: Thank you for the comment. The radiative quantities (ADRF and ARFE at TOA, ATM, and BOA) are ultimately produced by a radiative-transfer model. Our goal is to use an interpretable machine learning to quantify the relative importance predicting ADRF and ARFE over a long-term dataset rather than to replace the radiative transfer model. In the revised manuscript, we explicitly frame the radiative part of the study as an assessment of the diagnostic value of AAE_{col} for radiative metrics, rather than an attempt to quantify a driving role of AAE. We have revised the Methods to explicitly describe how ADRF and ARFE are obtained in AERONET (radiative-transfer-based broadband shortwave flux differences between aerosol-free and aerosol-laden conditions, following García et al., 2008, and we added the standard forcing definitions in the manuscript).

We also clarified why ML is applicable here: The dependence of ADRF/ARFE on the predictors is nonlinear and involves interactions (e.g., between AOD and SSA, or between SSA and surface albedo), which cannot be fully summarized by bivariate correlations or simple linear sensitivities over a long-term dataset. Interpretable ML provides (i) a flexible surrogate mapping that captures nonlinearities and interactions, and (ii) a unified attribution metric (mean｜SHAP｜) to rank predictors consistently.

The added value is not a new RT mechanism, but a quantitative attribution of predictor importance across TOA, ATM, and BOA and across loading regimes. In particular, our SHAP analysis shows that AAE_{col} is among the most informative predictors for TOA ADRF (comparable to SSA) and becomes the leading diagnostic predictor for TOA ARFE when aerosol loading is controlled (AOD-conditioned analysis). This result is practically important because it indicates that constraining AAE_{col} can substantially improve estimates of forcing efficiency, even though AAE itself is not treated as a causal driver. Our revisions are as follows:

Lines 290-304 (in the revised manuscript):

These radiative quantities are computed within the AERONET inversion radiative-transfer module under cloud-free conditions, using AERONET-retrieved aerosol optical properties and surface albedo as inputs. ADRF is defined as the difference in broadband shortwave radiative fluxes between aerosol-free and aerosol-laden conditions (García et al., 2008):

$$ADRF_{TOA} = F_{0,TOA}^{\uparrow} - F_{TOA}^{\uparrow} \tag{8}$$

$$ADRF_{BOA} = F_{BOA}^{\downarrow} - F_{0,BOA}^{\downarrow} \tag{9}$$

$$ADRF_{ATM} = ADRF_{TOA} - ADRF_{BOA} \tag{10}$$

where F and F_0 denote radiative fluxes with and without aerosols, and arrows indicate upward or downward fluxes. ARFE is defined as radiative forcing per unit aerosol optical depth:

$$ARFE = \frac{ADRF}{AOD_{550}} \tag{11}$$

where AOD_{550} is the AOD at 550 nm. Defined this way, negative ADRF and ARFE indicate shortwave cooling.

Lines 322-347 (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.

For ADRF, five optical properties (AOD, single scattering albedo (SSA), asymmetry parameter (g), surface albedo (SA), and AAE_{col}) were used as inputs. For ARFE, the target definition (ARFE = ADRF/AOD) was kept unchanged; however, AOD was included during model fitting together with SSA, g, SA, and AAE_{col} so that the models could learn any residual nonlinearity and interactions involving AOD. Performance was again evaluated using a consistent training–testing split, with 80% of the dataset used for training and the remaining 20% for testing. The evaluation was quantified by R^2, RMSE, and MAE. The performance metrics for the three models are summarized in Fig. S8-S9. CatBoost in our case was retained as the best-performing model across TOA, BOA, and ATM. R^2, RMSE, and MAE are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{14}$$

where n represents the number of input samples. y_i and $\hat{y}_i$ are the observed and predicted values, respectively; $\bar{y}$ refers to the mean of the target values predicted by the model. In this study, y corresponds to the target variable, including AAE_{sfc} (Section 2.2), AAE_{col}, ADFR, and ARFE in this Section.

Lines 605-701 (in the revised manuscript):

3.5 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.

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.

Minor comments:

Lines 44-46: Provide AAE for dust and brown carbon separately.

Response: Thank you for the suggestion. We have revised the Introduction to provide typical AAE ranges for brown carbon and mineral dust separately. Our revisions are as follows:

Lines 47-50 (in the revised manuscript): For example, BrC AAE is frequently reported to be ~2–6 depending on source and aging, whereas dust AAE is typically ~2–4 owing to shortwave absorption by iron oxides (Bergstrom et al., 2007).

In the Introduction, it is necessary to separately review the ground and column AAE, with fewer or unclear reviews of column AAE. Why is it necessary to study both ground and column AAE simultaneously?

Response: Thank you for the comment. We have revised the Introduction. We now clarify that AAE can be observed by in situ multi-wavelength absorption measurements and by surface-based remote sensing. In situ observations provide high-precision process constraints and are widely used as benchmarks for evaluating retrievals and models, whereas AERONET provides column-integrated aerosol properties from the surface to the top of the atmosphere. And we clarify why studying both ground and column AAE is necessary: they constrain different spatial/vertical volumes and therefore reveal scale-dependent drivers; combining them helps interpret the representativeness gap and better constrain aerosol vertical structure and radiative impacts. Our revisions are as follows:

Lines 54-84 (in the revised manuscript): AAE has been characterized using multiple observational approaches, including in situ multi-wavelength absorption measurements and surface-based remote sensing retrievals (Li et al., 2022). In situ observations provide high-precision, process-resolving constraints on aerosol absorption spectra near the surface and therefore serve as an important benchmark for evaluating remote sensing products and model simulations (Gliß et al., 2021). In contrast, surface-based remote sensing can retrieve aerosol properties integrated over the entire atmospheric column, such as Aerosol Robotic Network (AERONET), enabling a broader view of aerosol spectral absorption and its radiative properties

(Dubovik et al., 2000). Combining in situ and column retrievals is particularly valuable because they constrain complementary aspects of aerosol spectral absorption. In situ measurements are sensitive to near-surface processes (emissions, hygroscopic growth and aging) but have limited spatial and vertical representativeness, whereas AERONET provides column-integrated constraints that are directly connected to radiative impacts but can be influenced by vertical layering and retrieval assumptions (Li et al., 2022). Therefore, integrating near-surface with column AAE enables us to provide improved observational guidance for models, and better constrain column characteristics relevant to radiative forcing.

Both near-surface and column AAE vary with particle size distribution, chemical composition, and mixing state (Russell et al., 2010; Scarnato et al., 2013; Li et al., 2016; Schuster et al., 2016a; Sotiropoulou et al., 2025). For instance, near-surface BC AAE may decrease as BC cores grow or as aggregates become more compact during aging processes (Liu et al., 2018). Recent numerical simulation further indicates that secondary organic coatings can increase near-surface AAE, with sensitivity to coating thickness (Zhang et al., 2025). In contrast, photochemical bleaching lowers BrC ultraviolet absorption and near-surface AAE (Wang et al., 2019). Russell et al., (2010) showed that column AAE values are strongly correlated with aerosol composition or type. Heterogeneous aging of long-range-transported dust may enhance absorption, also affecting column AAE (Tian et al., 2018). The magnitudes and signs of these effects depend on location, season, and processing history, complicating both measurements and modeling and propagating to radiative forcing uncertainty (Sand et al., 2021; Li et al., 2022; Ponczek et al., 2022).

Line 180, 304-307, 323-325, and so on: These are key points, but too simple. Need more discussions.

Response: Thank you for the suggestion. we have expanded the explanation and discussion for Line 180 and the related points. Our specific revisions can be found in response to your third major comment (Lines 202-209 and Lines 225-236 in the revised manuscript). In addition, we have revised the discussions around Lines 304–307 and 323–325 to provide clearer physical interpretation and stronger linkage between the correlation analysis and the multivariate regression results. Our specific revisions can be found in response to your first major comment (Lines 405-467 in the revised manuscript).

Lines 251-254, Fig. 1: It is not clear or direct for reads to understand! Perhaps Figure 1 contains a lot of information, but lacks visualization skills, making it difficult to highlight key points.

Response: Thank you for the suggestion. 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. Our revisions are as follows:

Lines 358-396 (in the revised manuscript):

Figure 1 provides the near-surface AAE (AAE_{sfc}) variability and its co-variation with $PM_{2.5}$ composition and particle size during the Beijing campaign. 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. Notably, periods with elevated FMD fractions generally coincide with higher AAE_{sfc}, whereas intervals dominated by nd-WSII tend to correspond to lower AAE_{sfc}, consistent with dust-related enhancement of short-wavelength absorption. These co-variations motivate the quantitative attribution in Section 3.2, where we assess how the fractions of FMD and nd-WSII relate to the observed spectral absorption dependence.

The overall distribution of AAE_{sfc} is summarized in Fig. 1b. AAE_{sfc} ranges from 0.90 to 3.0 and occurs most frequently between1.10–2.0, with a mean value of 1.64 ± 0.32. A pronounced high-AAE_{sfc} tail (values above 2.0) occurs episodically (Fig. 1b), suggesting intermittent enhancement of short-wavelength absorption. Such elevated values likely resulted from winter heating emissions (Tian et al., 2019; Yan et al., 2017) and mineral dust contributions (Fig. 1a), both known to raise AAE (Liu et al., 2018).

The heat map in Fig. 1a further illustrates the time-of-day evolution of AAE_{sfc} across the campaign, and the accompanying diurnal profile highlights a clear nighttime enhancement relative to daytime. AAE_{sfc} showed a clear night-high and day-low pattern (Fig. 1a), consistent with the evolution of particle size distributions. Fine-mode number concentrations derived from SMPS increased during the morning rush hours and nighttime residential activity (Fig. 1c). By contrast, coarse-mode diameters from APS were larger in the early morning and decreased during the day (Fig. 1d). These results demonstrate that AAE_{sfc} was co-regulated by both composition and size, providing the observational evidence for the subsequent machine-learning analysis to quantify their relative contributions and radiative implications.

Figure S6 further shows the multi-wavelength absorption coefficients and their diurnal behavior. Aerosol absorption coefficients exhibited a clear spectral decrease from the near-UV to the near-IR, with mean values of 13.19 ± 9.91, 6.80 ± 6.15, and 3.77 ± 3.27 Mm^{-1} at 375, 532, and 870 nm, respectively (mean ± one standard deviation). (Fig. S6). The corresponding mass absorption efficiencies are relatively low (0.49 ± 0.24, 0.21 ± 0.08, and 0.12 ± 0.04 $m^2 \cdot g^{-1}$), reflecting the dominance of nd-WSII, which accounted for 42.9% of $PM_{2.5}$ mass (Fig. 1a). Absorption coefficients at three wavelengths are consistently higher absorption at night and a peak around 23:00 (Fig. S6), driven by reduced tropospheric boundary layer height, lower afternoon temperatures and wind speeds (Fig. S7), and enhanced emissions from nighttime traffic and heating (Guo et al., 2016; Zhao et al., 2019).

MLR and machine learning first need to report R^2 and RMSE, such as in Lines 313-314. These are the most direct indicators to win the trust of readers. Please check the whole paper.

Response: Thank you for the suggestion. We have added MLR and machine learning performance metrics (R^2, RMSE, and MAE, cross-validated) to improve transparency. Our revisions are as follows:

Lines 229-233 (in the revised manuscript): 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).

Lines 312-317 (in the revised manuscript): Model performance was evaluated using a consistent training–testing split (80% of dataset were used for the training set, and 20% were used for the test set) and quantified by R^2, RMSE, and MAE. The RF model achieved an R^2 of 0.58, an RMSE of 0.43, and an MAE of 0.30. In comparison, the CatBoost model yielded an R^2 of 0.64, an RMSE of 0.40, and an MAE of 0.29, while the XGBoost model showed an R^2 of 0.64, an RMSE of 0.40, and an MAE of 0.30 (Fig. S8).

Lines 334-347 (in the revised manuscript): Performance was again evaluated using a consistent training–testing split, with 80% of the dataset used for training and the remaining 20% for testing. The evaluation was quantified by R^2, RMSE, and MAE. The performance metrics for the three models are summarized in Fig. S8-S9. CatBoost in our case was retained as the best-performing model across TOA, BOA, and ATM, as it showed the highest or near-highest R^2 together with the lowest or near-lowest RMSE and MAE among the tested models. R^2, RMSE, and MAE are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{14}$$

where n represents the number of input samples. y_i and $\hat{y}_i$ are the observed and predicted values, respectively; $\bar{y}$ refers to the mean of the target values predicted by the model. In this study, y corresponds to the target variable, including AAE_{sfc} (Section 2.2), AAE_{col}, ADFR, and ARFE in this Section.

Lines 336-338: Organize sections around this point.

Response: Thank you for the suggestion. We have addressed this by adding a dedicated synthesis paragraph that explicitly links the surface and column perspectives and highlights the cross-scale consistency of dust-related absorption. Our specific revisions can be found in response to your first major comment (Lines 570-604).

References:

García, O. E., Díaz, A. M., Expósito, F. J., Díaz, J. P., Dubovik, O., Dubuisson, P., Roger, J. -C., Eck, T. F., Sinyuk, A., Derimian, Y., Dutton, E. G., Schafer, J. S., Holben, B. N., and García, C. A.: Validation of AERONET estimates of atmospheric solar fluxes and aerosol radiative forcing by ground-based broadband measurements, J. Geophys. Res., 113, 2008JD010211, https://doi.org/10.1029/2008JD010211, 2008.