Response to the reviewers' comments on "Interpretable Machine Learning Quantifies Composition and Size Controls on Aerosol Spectral Absorption"

We would like to sincerely thank the editor for giving us the opportunity to revise and resubmit our manuscript titled "Interpretable Machine Learning Quantifies Composition and Size Controls on Aerosol Spectral Absorption". We are also grateful to thank the two referees and Dr. Xiyao Chen for carefully reading and reviewing our manuscript. The comments and suggestions have significantly contributed to improving the quality of the manuscript. We have carefully addressed all the points raised by the reviewers, and we believe that these revisions have significantly enhanced the quality of the manuscript. Our responses to the comments are marked in blue font in this document. The changes to the text are marked in red in a separate marked-up version of the manuscript. Please find below a detailed point-by-point response to the reviewers' comments.

Reply on RC1:

This manuscript by Wang et al. investigates the complex drivers of the Absorption Ångström Exponent (AAE) using a combination of high-resolution field observations in Beijing and multi-year AERONET columnar data. The study addresses a persistent

challenge in atmospheric science: quantitatively disentangling the relative importance of chemical composition versus particle size in determining aerosol spectral absorption. The authors employ Shapley Additive exPlanations (SHAP) to provide a ranked attribution of AAE drivers. The study quantifies how AAE variations directly influence aerosol radiative forcing efficiency (ARFE), finding that AAE is a primary driver of cooling efficiency at the Top of the Atmosphere (TOA). The manuscript is well-structured and the methodology is fairly robust, with the caveats detailed below.

Response: Thank you for your overall assessment and for recognizing the significance of our effort to quantitatively disentangle the roles of chemical composition and particle size in influencing aerosol spectral absorption (AAE). We also appreciate your positive evaluation of the manuscript structure and the general robustness of the methodology, as well as your recognition of the value of using SHAP for ranked attribution and of linking AAE variability to aerosol radiative forcing efficiency (ARFE).

We have carefully considered all of your specific comments. In the revised manuscript, we have strengthened the methodological description and justification where needed, expanded the discussion of key mechanisms and interpretation (including how composition- and size-related factors jointly influence AAE), and added supporting analyses and text to improve transparency and robustness, such as Mie-theory sensitivity test. Detailed point-by-point responses to each comment are provided below, and all revisions are marked in the updated manuscript for ease of review. 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 issues

The study relies heavily on SHAP to attribute drivers of AAE which identifies associations, not necessarily causal links. The authors acknowledge nonlinearity and collinearity among predictors, however, if two variables are highly correlated (e.g.,D_{APS} and FMD mass fraction, r=0.64), SHAP can sometimes "split" the importance between them in ways that don't reflect physical reality. The manuscript would benefit from a more explicit discussion on whether the "importance" found by the model is supported by Mie theory or other physical optical models to bridge the gap between statistical importance and physical causation.

Response: Thank you for the comment. We have revised the manuscript to clarify this issue and to avoid causal wording when discussing SHAP results. We also note that the example of collinearity between D_{APS} and FMD (r = 0.64) pertains to our surface dataset, whereas the SHAP analysis is applied to the AERONET column dataset using AERONET/GRASP-derived composition proxies and size parameters; D_{APS} and FMD are not included as SHAP predictors in the column models. Nevertheless, collinearity is still present among some AERONET predictors (e.g., among size parameters and component fractions).

Besides, to bridge the gap between statistical attribution and physical plausibility, we added a simple Mie-theory sensitivity test that isolates two primary physical factors influencing AAE: (i) the spectral absorption strength via the imaginary refractive index

k, and (ii) the particle-size structure via systematic shifts of the fine- and coarse-mode characteristic radii, while keeping the remaining factors fixed. Specifically, using the AERONET volume size distribution as the baseline, we computed an AAOD-weighted effective AAE (440–870 nm) as a function of radius. We then performed three perturbation experiments (Text S3 and Fig. S11):

(1) Absorption-spectral sensitivity: imaginary refractive index at 440 nm (k_{440}) was scaled by factors of 0.6, 0.8,1.0, 1.2, and 1.5 with the size distribution held fixed. The resulting curves show a clear and monotonic response, with larger k_{440} producing higher AAE across the relevant size range. This confirms that enhanced short-wavelength absorption physically increases AAE, consistent with the interpretation of absorption-related predictors in the SHAP analysis.

(2) Fine-mode size sensitivity: the fine-mode characteristic radius was shifted by factors of 0.8, 0.9, 1.0, 1.1, and 1.2 while keeping the fine-mode integrated volume unchanged. This produces a systematic change in the AAOD-weighted AAE, demonstrating that plausible variations in fine-mode size alone can modify AAE even when k is fixed.

(3) Coarse-mode size sensitivity: the coarse-mode characteristic radius was shifted by factors of 0.8, 0.9, 1.0, 1.1, and 1.2 under the same volume-conserving constraint. The AAE response is again systematic, and the sensitivity mainly emerges once coarse particles contribute substantially to the AAOD-weighted integral. Our revisions are as follows:

Lines 511-554 (in the revised manuscript):

Machine learning analysis further quantified relative contributions, as illustrated in Fig. 5a. It is found that showed that CAI had the strongest explanatory power, accounting for ~19% of the model's predictive power, confirming the dominant role of dust in amplifying spectral absorption. 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). Among the size-related predictors, R_{fine} alone accounted for about one quarter (~29%) of the cumulative importance of all size metrics, making it the most influential size parameter. In addition, its importance was also clearly higher than CNAI and FNAI (Fig. 5a). During the prediction process, it is observed that higher values of BrC, CAI, and volume concentrations of coare-mode (vol_{coarse}) corresponded to higher SHAP values and higher values of other predictors corresponded to smaller SHAP values (Fig. 5a). These responses are fully consistent with the correlations between AAE_{col} and these parameters (Fig. 3b–d; Fig. 4c-d).

To connect these statistical attributions to physically plausible behavior, we performed a simple Mie-theory sensitivity analysis (Text S3; Fig. S12). First, we varied the imaginary part of the refractive index at 440 nm (k_{440}) while keeping the size distribution fixed. Second, keeping refractive index fixed and shifting the fine- and coarse-mode radii to 80%, 90%, 100%, 110%, and 120% of their baseline values (with mode-integrated volume conserved).

Varying k_{440} produces a substantially larger change in the modeled AAE_{col} than the size-shift experiments (Fig. S12), indicating that changes in short-wavelength absorptivity exert stronger leverage on AAE than variations in modal radii. Because we

altered k_{440} while keeping k_{870} unchanged, this experiment specifically isolates enhanced absorption in the short wavelengths which is consistent with increased contributions from aerosols that preferentially absorb at shorter wavelengths, such as absorbing mineral dust and brown carbon. Spectral refractive indices retrieved by AERONET are known to vary systematically across aerosol types and can be used to infer absorbing components, supporting the interpretation of k as a proxy for composition-related absorption variability (Dubovik et al., 2002; Wang et al., 2013). Fine-mode shifts produce a clearer change in the AAE_{col} than coarse-mode shifts (Fig. S12b, S12c), consistent with R_{fine} being the leading size predictor in Fig. 5a. Taken together, the SHAP results and the Mie sensitivity tests support a consistent interpretation that AAE_{col} is not influenced by BC or BrC alone; it is primarily associated with dust and secondarily by particle-size structure (size metrics together ~35%) (Fig. 5a), underscoring the need to account for both composition and size when evaluating spectral absorption.

Lines 62-105 in the supplementary:

Text S3. Mie theory sensitivity experiments.

To provide a physical consistency check for the SHAP-based attribution, we conducted a set of Mie-theory sensitivity experiments to isolate how (i) short-wavelength absorptivity and (ii) particle-size structure can influence the absorption Ångström exponent between 440 and 870 nm. The experiments use the AERONET-retrieved column volume size distribution (VSD), expressed as $\frac{dV}{dlnr}$ on a discrete radius grid r spanning approximately 0.05–15μm. For each radius bin, the number size

distribution is obtained by

$$\frac{\mathrm{dN}}{dlnr} = \frac{1}{\frac{4}{3}\pi r^3} \times \frac{dV}{dlnr} \quad (1)$$

Given the complex refractive index $m = n - ik$, where n is the real part and k is the imaginary part, we compute the absorption efficiency $Q_{abs}(r, \lambda)$ using Mie theory. The AAOD at wavelength λ is then evaluated by integrating absorption cross section (up to a proportionality constant) weighted by the number distribution,

$$AAOD_\lambda \propto \int Q_{abs}(r,\lambda)\, \pi r^2 \frac{\mathrm{dN}}{dlnr} dlnr \quad (2)$$

Because our goal is sensitivity and mechanistic consistency rather than an exact retrieval of absolute AAOD, the proportionality constant is not required; all comparisons are made within the same computational framework. The AAE between 440 and 870 nm is computed as

$$AAE = -\frac{ln(\frac{AAOD_{440}}{AAOD_{870}})}{ln(\frac{440}{870})} \quad (3)$$

To isolate the influence of composition-related changes, we varied only the imaginary refractive index at 440 nm (k_{440}) while holding the size distribution fixed. Specifically, k_{440} was scaled by multiplicative factors of 0.6, 0.8, 1.0, 1.2, and 1.5, while k_{870} was kept unchanged. The real part n was held constant in these tests (1.5, the mean AERONET-retrieved value). This design emphasizes changes in absorption that preferentially affect shorter wavelengths, consistent with increased contributions from aerosol components that absorb more strongly in the near-UV (e.g., absorbing mineral dust and brown carbon), and avoids conflating this effect with broadband absorbers

whose spectral dependence is weak.

To isolate the role of particle-size structure, we held the refractive index fixed and perturbed the size distribution by shifting the characteristic radii of the fine and coarse modes. We separated the VSD into a fine subset and a coarse subset using a radius threshold of r = 0.6 μm. For each subset, we performed a log-radius shift,

$$\left(\frac{dV}{dlnr}\right)_{new}(\mathrm{r}) = \left(\frac{dV}{dlnr}\right)_{old}\left(\frac{\mathrm{r}}{s}\right) \quad (4)$$

where s is the shift factor. We used s = 0.8,0.9,1.0,1.1, and 1.2, corresponding to moving the characteristic radius to 80%, 90%, 100%, 110%, and 120% of the baseline value. The shift is implemented by linear interpolation in lnr, with values outside the original subset radius range set to zero. To ensure that this perturbation represents a change in size location rather than loading, we renormalize the shifted subset so that its mode-integrated volume is conserved,

$$\int \left(\frac{dV}{dlnr}\right)_{new} dlnr = \int \left(\frac{dV}{dlnr}\right)_{old} dlnr \quad (5)$$

The unperturbed subset is kept unchanged, so that “fine-mode shift” modifies only the fine subset while the coarse subset remains fixed, and vice versa for “coarse-mode shift”. The resulting sensitivity curves are summarized in Fig. S11.

Lines 162-173 in the supplementary:

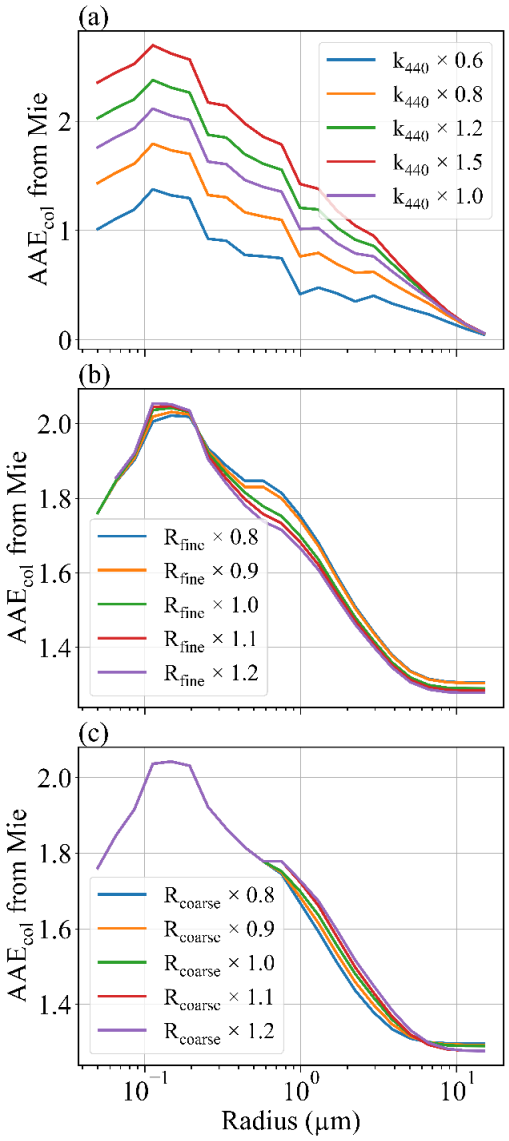

Figure S11. Mie-theory sensitivity experiments linking composition-related absorptivity and size structure to AAE. (a) Sensitivity to shortwave absorptivity: the imaginary refractive index at 440 nm (k_{440}) is scaled by factors of 0.6, 0.8, 1.0, 1.2, and 1.5 while the size distribution and k_{870} are held fixed (the real part is fixed at n = 1.5, the AERONET mean). (b) Sensitivity to fine-mode size: the fine-mode characteristic radius is shifted to 80%, 90%, 100%, 110%, and 120% of its baseline value while conserving the fine-mode integrated volume; the coarse mode is unchanged. (c) Sensitivity to coarse-mode size: the coarse-mode characteristic radius is shifted to the

same factors while conserving the coarse-mode integrated volume; the fine mode is unchanged. Curves are shown as a function of particle radius; changes in AAE reflect the differing impacts of absorptivity and size perturbations under otherwise fixed conditions.

Near-surface chemical data was collected via offline filter sampling (day/night blocks), while optical data was online (hourly). While the authors matched these temporally, the coarse resolution of the chemical data (12-hour averages) likely masks the fine-scale diurnal variability seen in the optical AAE_{sfc}.

Response: Thank you for the comment. We acknowledge this limitation and have added a sentence in the revised 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. We also 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, this scale-consistent alignment is a standard approach in field studies that combine offline filter chemistry with continuous absorption measurements, where online optical data are routinely averaged to the filter sampling intervals (e.g., 12 h or 24 h) before comparison or regression analysis. For example, Bernardoni et al. (2021) analyzed 12-h resolved $PM_{2.5}$ samples together with Aethalometer absorption measurements on matching time windows, and Wang et al. (2021) similarly combined 24-h filter-based analyses with high-frequency Aethalometer absorption data to investigate the impacts of chemical composition on AAE.

To transparently address your concern about AAE_{sfc} variability, we quantified the within-window dispersion of AAE_{sfc}. Specifically, we computed the standard deviation of AAE_{sfc} within each 11.5 h window and summarized its frequency distributions with cumulative frequency curves (new Fig. S2). The cumulative distributions show that within-window variability is typically moderate: approximately 90% of the windows have a within-window standard deviation ≤ 0.35, while larger fluctuations occur only in a small tail of rare windows. This result indicates that window-mean AAE_{sfc} provides a reasonable representative value at the filter timescale. We have added a brief statement in the manuscript to clarify the timescale consistency and to point readers to Fig. S2, and we keep the main analysis based on window-mean values.

Besides, to quantify the uncertainty introduced by limited sample size and window-averaging, we additionally performed nonparametric bootstrap resampling

(1000 replicates). We now report the bootstrap mean coefficients and 95% confidence intervals for the standardized MLR (Table S2, S3), which show that the signs and relative magnitudes of the key coefficients are stable at the 11.5-h timescale. 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:

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

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:

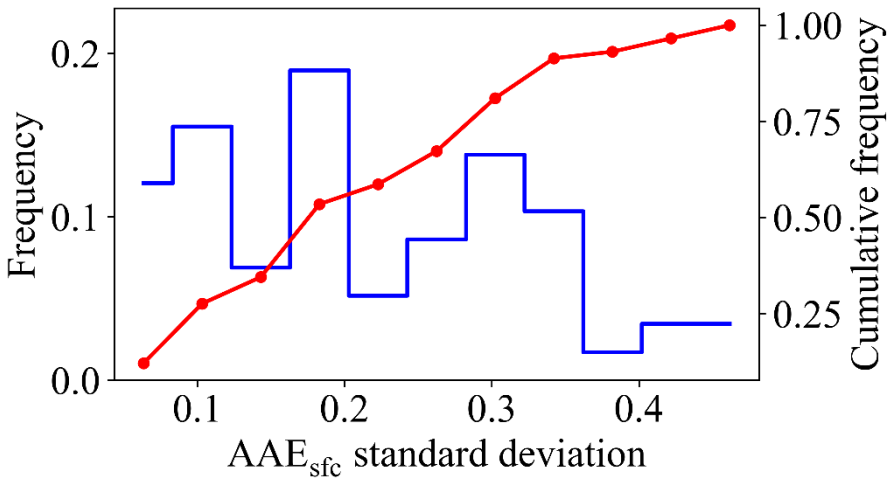

Figure S2. Frequency (blue) and cumulative frequency (red) distributions of within-window AAE_{sfc} standard deviation (11.5 h).

The authors describe using a 'consistent split' for model training, testing and validation but never explicitly state what that is. Please describe the method for splitting and why that's appropriate for this dataset.

Response: Thank you for the suggestion. By "consistent split" we mean that we use one fixed and reproducible partition of the dataset into training and testing subsets, and apply the same partition across all machine-learning models so that their performance can be compared on an identical test set. Specifically, we randomly split the dataset into 80% training and 20% testing samples using a fixed random seed (for reproducibility). The split was performed once and then kept unchanged for all models and all target variables. This approach is appropriate here because the long-term AERONET record provides a sufficiently large sample size for model development, while the use of an independent held-out test set enables an objective evaluation of model performance on unseen data. Using the same split across models avoids

differences in performance that could arise purely from different test samples. We have clarified this procedure in the Methods section. Our revisions are as follows:

Lines 312-318 (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 on the test set. 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.

Minor issues

The authors focus on a specific site in Beijing, which given the observational constraints seems reasonable, but it would be beneficial to include a sentence or two in the conclusion explicitly stating how these results might change (and therefore their relevance) in cleaner or more dust-dominant global regions.

Response: Thank you for the suggestion. In the revised Conclusion, we added a short paragraph clarifying that the relative importance of the identified predictors is expected to depend on regional aerosol regime. In cleaner environments with lower aerosol loading, the statistical relationships may weaken because the signal-to-noise ratio in absorption properties is smaller, whereas in more dust-influenced regions the role of dust-related predictors would likely become even stronger and may outweigh the contribution of anthropogenic fine-mode components. This addition helps place our results in a broader global context while retaining the Beijing site as the primary observational basis. Our revisions are as follows:

Lines 684-688 (in the revised manuscript): 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-influenced regions the role of dust-related predictors would likely become even stronger.

Please mention the ranges quoted in the results section. Are these 1 standard deviation, or something else?

Response: Thank you for the comment. Thank you for pointing this out. We have clarified the statistical meaning of the quoted ranges throughout the Results section. Our revisions are as follows:

Lines 387-390 (in the revised manuscript): Aerosol absorption coefficients exhibit 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).

L217, "coare-mode" should be corrected to "coarse-mode".

Response: Thank you for the suggestion. We have corrected "coare-mode" should be corrected to "coarse-mode".

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. S12), 50th (Fig. 7), 75th percentiles (Fig. S13), or mean (Fig. S14). 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:

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

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:

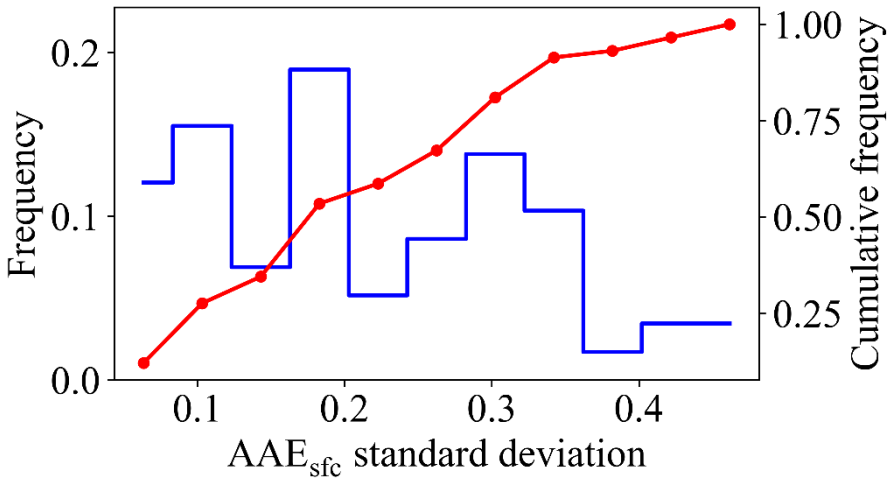

Figure S2. Frequency (blue) and cumulative frequency (red) distributions of within-window AAE_{sfc} standard deviation (11.5 h).

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):

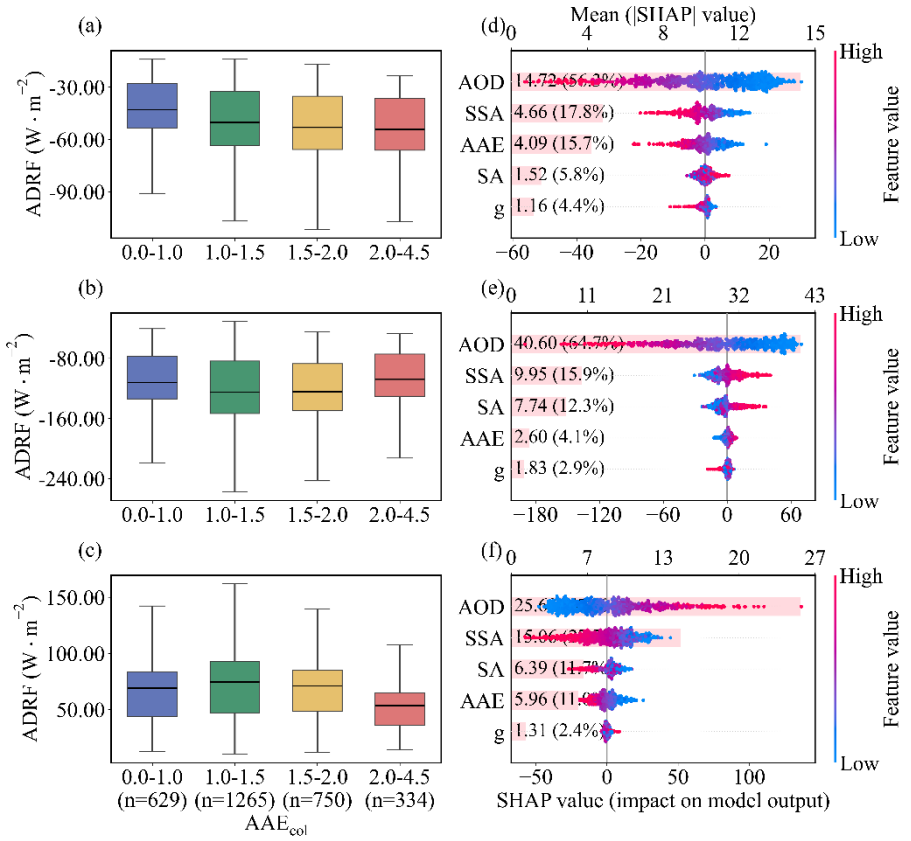

Figure 6. The relationship between columnar AAE (AAE_{col}) and aerosol direct radiative forcing (ADRF). (a–c) Box plots of ADRF at the top of the atmosphere (a), bottom (b), and in the atmosphere (c) as a function of AAE_{col}. The sample sizes for the AAE_{col} bins 0.0–1.0, 1.0–1.5, 1.5–2.0, and 2.0–4.5 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):

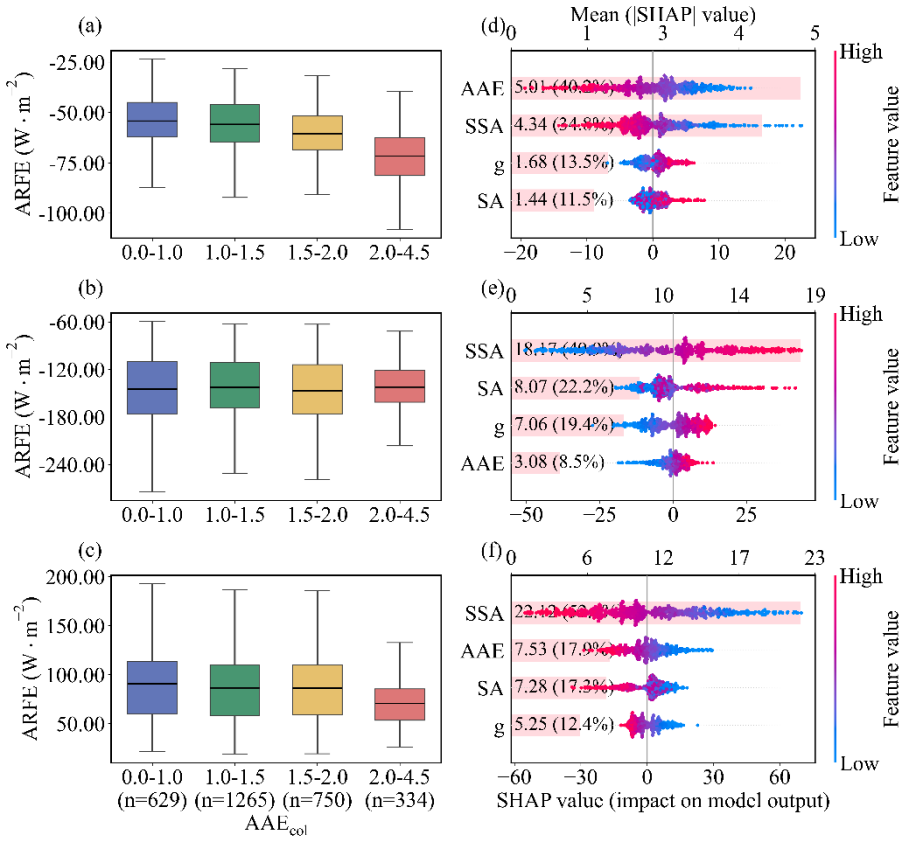

Figure 7. The relationship between columnar AAE (AAE_{col}) and aerosol radiative forcing efficiency (ARFE). (a–c) Box plots of ARFE at the top of the atmosphere (a), bottom (b), and in the atmosphere (c) as a function of AAE_{col}. The sample sizes for the AAE_{col} bins 0.0–1.0, 1.0–1.5, 1.5–2.0, and 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.

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:

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

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; $\widehat{FMD}$, $\widehat{nd-WSII}$, $\widehat{D_{SMPS}}$, and $\widehat{D_{APS}}$ 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):

$$ADRF_{TOA} = F^{\uparrow}_{0,TOA} - F^{\uparrow}_{TOA} \quad (8)$$

$$ADRF_{BOA} = F^{\downarrow}_{BOA} - F^{\downarrow}_{0,BOA} \quad (9)$$

$$ADRF_{ATM} = ADRF_{TOA} - ADRF_{BOA} \quad (10)$$

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:

$$ARFE = \frac{ADRF}{AOD_{550}} \quad (11)$$

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. S12), 50th (Fig. 7), 75th percentiles (Fig. S13), or mean

(Fig. S14). 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} rangs 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:

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

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

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

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.