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 (Russell et al., 2010). 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

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

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,

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

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

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

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:

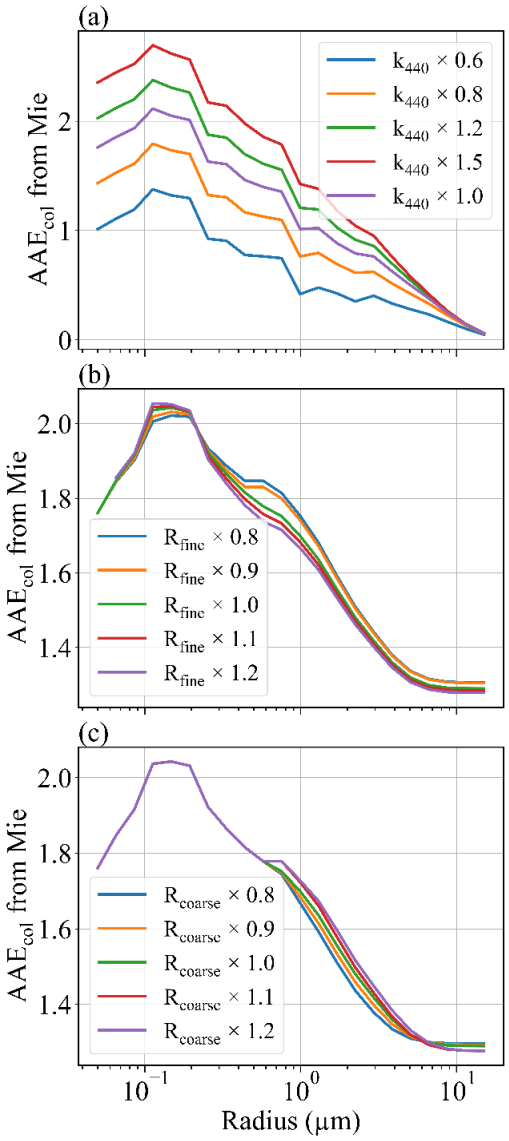

Figure S11. Mie-theory sensitivity experiments linking composition-related absorptivity and size structure to AAE. (a) Sensitivity to shortwave absorptivity: the imaginary refractive index at 440 nm (k_{440}) is scaled by factors of 0.6, 0.8, 1.0, 1.2, and 1.5 while the size distribution and k_{870} 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:

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

where $\widehat{AAE_{sfc}}$ denotes the standardized AAE_{sfc}; a represents the intercept term, any remaining influence not parameterized by the selected predictors is captured by the intercept term; b, c, d, and e are regression coefficients; $\widehat{\mathrm{FMD}}$, $\mathrm{nd}\widehat{-\mathrm{WSII}}$, $\widehat{\mathrm{D}_{SMPS}}$, and $\widehat{\mathrm{D}_{APS}}$ are standardized variables of FMD fraction, nd-WSII fraction, and mean diameters from SMPS and APS, respectively. To ensure consistent temporal support between offline chemistry and online optical measurements, we aggregate AAE_{sfc} (and size-related parameters) over the same sampling windows and use these window-mean values. We note that AAE_{sfc} can vary within a given sampling period; however, such within-period variability is not resolvable by the integrated filter chemistry and therefore cannot be explicitly attributed at finer temporal resolution. To transparently characterize the associated representativeness uncertainty, we quantify the within-window dispersion of AAE_{sfc} using the standard deviation across all sampling windows and provide its frequency and cumulative distributions (Fig. S2). In particular, ~ 90% of the sampling periods show a standard deviation no greater than 0.35. This result

indicates that window-mean AAE_{sfc} provides a reasonable representative value at the filter timescale. Due to power outage on 27 December 2023 and 3 January 2024, daytime data for 27 December and both daytime and nighttime data for 3 January were unavailable. In future studies, higher-time-resolution measurements of aerosol chemical composition would be valuable for more directly linking short-timescale composition variability with AAE.

Notably, to further evaluate the robustness of the regression coefficients, we conducted a nonparametric bootstrap analysis with 1000 resamples. We also tested an extended specification including EC and OM fractions as additional predictors. However, the extended model yielded highly unstable coefficient estimates under bootstrap resampling, with strong dispersion and frequent sign changes (Table S1). In contrast, the reduced model provides stable and physically interpretable coefficients for the key predictors and demonstrates good predictive skill for AAE_{sfc} (the coefficient of determination (R^2) = 0.75, root mean square error (RMSE) = 0.13, mean absolute error (MAE) = 0.10; Table S2). Consistent with these robustness results, our correlation analysis further indicates that EC and OM fractions are not significantly associated with AAE_{sfc} during this campaign (Section 3.2). Therefore, we retained the parsimonious formulation without EC and OM fractions for subsequent analyses (Equation (5)).

Lines 109-122 in the supplementary:

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

coef	mean	std	p2.5	p97.5
a	0.006633	0.069351	-0.12142	0.148516
EC	-7.5E+11	9.37E+11	-2.7E+12	9.93E+11

OM	-2.1E+12	2.57E+12	-7.3E+12	2.73E+12
FMD	-7.3E+12	9.12E+12	-2.6E+13	9.67E+12
nd-WSII	-7.1E+12	8.9E+12	-2.5E+13	9.43E+12
D_{SMPS}	0.032792	0.125148	-0.18773	0.299769
D_{APS}	0.446719	0.086028	0.300897	0.630525
R^2	0.74	0.03	0.67	0.77
RMSE	0.13	0.01	0.13	0.15
MAE	0.10	0.01	0.09	0.16

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

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

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

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

Lines 128-130 in the supplementary:

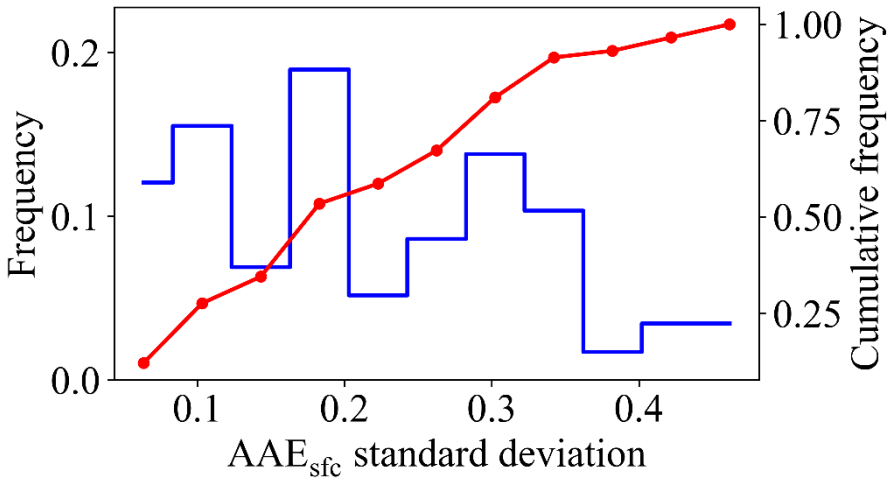

Figure S2. Frequency (blue) and cumulative frequency (red) distributions of within-window AAE_{sfc} standard deviation (11.5 h).

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".

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.

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.