Response to reviewer comments on the manuscript "*Distinct aerosol populations and their vertical gradients in central Amazonia revealed by optical properties and cluster analysis*", submitted for publication at ACP.

Dear Editor, we would like to thank you and both reviewers for their valuable comments and useful suggestions for improving our manuscript. Below, you can find answers and actions for each individual comment from the reviewers. In order to make it easier to identify the individual answers and actions, we used the following color code strategy:

- **In black are the reviewer's comments.**
- **In blue are the author's responses.**
- ***In orange are the text modifications we made in the manuscript.***

Responses to Reviewer #1 of the manuscript "Distinct aerosol populations and their vertical gradients in central Amazonia revealed by optical properties and cluster analysis" by Valiati et al., submitted for publication in Atmospheric Chemistry and Physics

The manuscript titled "Distinct aerosol populations and their vertical gradients in central Amazonia revealed by optical properties and cluster analysis" by Valiati et al. presents multi-instrument datasets (optical, chemical, and size-resolved measurements) at two different heights and the application of clustering algorithm to characterize of Amazonian aerosol dynamics and sources. The study leverages five years of vertically resolved in-situ data (2018–2023) from the Amazon Tall Tower Observatory (ATTO), applying unsupervised machine learning (k-means clustering) to optical intensive parameters and black carbon concentrations. By stratifying observations seasonally and vertically (60 m and 325 m), the authors identify different aerosol populations associated with background conditions, long-range transport (LRT) events (e.g., Saharan dust and African biomass burning), and regional biomass-burning episodes. The methodology is well based in the literature, it extends the field by linking aerosol intensive properties to source and transformation processes and this study provide a valuable optical characterization of different aerosol populations at the analyzed site. The manuscript deserves publication since it is methodologically sound, comprehensive, and clearly written. However, I recommend minor revisions before acceptance.

Dear editor, we thank Reviewer #1 for the positive feedback and for recognizing the value of this manuscript, which utilizes an extensive dataset of atmospheric measurements from the ATTO site and is based on a sound methodology for aerosol characterization. The general and specific comments are addressed in detail over the next pages of this response letter.

General comments:

- The manuscript explains the limitations of the clustering method using intensive optical parameters, such as increased uncertainty in Ångström exponents at low scattering/absorption, sensitivity to instrument calibration, and the subjective weighting applied in k-means. Please motivate a bit more why you chose this input aerosol parameter for the k-means clustering and why do you prefer to perform clustering with optical parameters instead of PNSD and ACMS data. It has been demonstrated the utility and effectiveness of clustering methods to characterize different aerosol population according to PNSD data. In this sense, other parameter (like intensive optical parameters) can be used as ancillary information for the clustering interpretation. Explain why SAE, AAE, and eBC were chosen as clustering inputs instead of, or alongside, PNSD metrics. What specific insights do these optical parameters provide about aerosol composition and source attribution that PNSD and chemical speciation do not capture?

  We thank the reviewer for this important comment regarding the motivation of our choices. The optical properties being used as the seed variables are based on a few reasons: Firstly, the Ångström matrix is a well-known method of separating possible particle populations based on intensive optical properties, at least in a qualitative way. Optical measurements are also more widely available, since the instruments are usually cheaper than SMPSs and ACSMs, which can be a factor for further applications of our method in other sites worldwide. Lastly, aerosol populations at ATTO have been previously characterized regarding their chemical properties and size distributions by a few authors, especially Holanda et al., 2023. Building on this, we seek a novel approach by connecting the optical properties to what's already known from previous work.

  To clarify these points, the following text was added to Section 2.4:

  "*This variable choice was motivated by the Ångström matrix, which is a well-known method of separating possible particle populations based on intensive properties, as well as the availability of measurements of the aerosol particles' optical properties.*"

- Throughout the whole manuscript (where applicable) verify the statistical significance between the averaged parameters. Several sections report differences between clusters or vertical levels (e.g., in SSA, refractive index, BrC fraction, MSE), but the manuscript does not consistently demonstrate that these differences are statistically robust. I recommend the authors to select appropriate statistical tests. For example, use non-parametric tests (e.g., Wilcoxon rank-sum or Kruskal–Wallis) for distributions that are not Gaussian, or t-tests/ANOVA when normality holds.

  The reviewer raised a very important point, as the statistical significance of all claims based on probability distributions and mean/median values should be tested. This was done for all analyzed differences using the Z-test of the

means and the Wilcoxon rank-sum test. Sections of the text that didn't mention it explicitly were changed appropriately to further clarify this point:

On Section 3.3:

"*These differences are statistically significant, and indicate that the aerosols arriving in central Amazonia during LRT events from Africa are highly efficient in absorbing radiation and much more refringent than those of local/regional biogenic origin.*"

On Section 3.4:

"*The real part has a more pronounced vertical gradient during the wet season, with higher median values closer to the canopy of 1.47 and 1.58 during near-pristine and LRT events, respectively. In contrast, medians of 1.29 and 1.41 were obtained at 325 m, highlighting the effectiveness of coating in reducing the refringence of aerosols. Once again, the statistical significance of these claims was verified with the Wilcoxon rank-sum test, which resulted in virtually null p-values.*"

On Section 3.5:

"*The MSE at 325 m was nearly identical during the wet season clusters, reaching (3.96 ± 0.07) m2 g−1 at near-pristine conditions and (4.09 ± 0.17) m2 g−1 for LRT events. In fact, when performing a Z-test of these values, a statistical difference is not achieved within a 95% confidence interval (Z = 0.7).*"

"*Similarly, low-impact and dry-season LRT conditions also showed nearly the same mass scattering efficiencies at 325 m, at (3.60 ± 0.05) m2 g−1 and (3.70 ± 0.20) m2 g−1, respectively, compatible within the 95% confidence interval, and around 9% lower than the wet season average.*"

Specific comments:

- Line 135: PNSDs measurements were performed using a TSI SMPS in the range between 10-400 nm. Why did you choose this range without considering a wider range which would allow you to characterize a more realistic accumulation mode (up to 500 – 800 nm, for example)? Considering a wider range in the PNSD data will provide very useful information for the aerosol population characterization.

  We thank the reviewer for this observation. The selected range (10 - 400 nm) is a limitation of the instrument that has been used at the ATTO site since the beginning of the measurements, in 2014. Franco et al., 2022, DOI: 10.5194/acp-22-3469-2022) discuss these issues in more detail.

- Lines 166-170: Why did you consider only two inorganic species? What about the contribution of H2SO4 or NH4HSO4? Probably, you expect a negligible effect of this species, I recommend supporting this statement with previous publication in the same area.

We thank the reviewer for this comment. In our study, the mass concentrations of sulfate were determined using the AMS default fragmentation table method (Allan et al., 2004; Canagaratna et al., 2015). This standard approach reports bulk sulfate and does not resolve individual species such as H2SO4, NH4HSO4, or (NH4)2SO$_4$. Despite that, at the ATTO site and in central Amazonia more broadly, sulfate represents only a minor fraction of submicron aerosol mass: typically <5–10% during near-pristine conditions, and up to ~15% during long-range transport events (Chen et al., 2009; Artaxo et al., 2022). Thus, any internal partitioning among different sulfate species would have only a negligible influence on the bulk density estimates compared to the dominant role of organics. We have clarified this in the revised manuscript by explicitly stating that the sulfate mass reported here refers to the bulk signal derived from the fragmentation table and therefore includes all sulfate species.

The following text was added to Section 2.2.3:

"In this study, sulfate mass concentrations were derived using the default fragmentation table method (Allan et al., 2004; Canagaratna et al., 2015), which reports bulk sulfate and does not resolve individual species (e.g., H_2SO_4, NH_4HSO_4, $(NH_4)_2SO_4$). Given that sulfate typically accounts for <5–15% of PM_1 mass at ATTO (Chen et al., 2009; Artaxo et al., 2022), this simplification has a negligible impact on the bulk density estimates."

- Lines 169-171: The plots and regression coefficients of the comparison of ACSM and SMPS-derived mass concentration should be included in the manuscript (maybe in the supplement) to support the statement mentioned in these lines.

We appreciate the reviewer's attention to this detail. This statement is now supported by Figure S1, which shows the evolution of the collection efficiency over time and the linear relationship between ACSM and SMPS-derived concentrations.

- Lines 225-228: In the manuscript, it's pointed out that the main driver of the aerosol population at this site is the dry/wet condition. Since it's mentioned that the k-means clustering is applied to two different datasets: the dry and wet condition seasons. As a quality check for the input variables choice, have you tried to perform the clustering with k=2 for the whole dataset? According to this analysis, we expect that each cluster represents one of the main seasons. If this quality check is not satisfactory, I am a bit skeptical in the decision of the input variables of the algorithm. I'm curious to see a figure like Fig. S2b with the cluster frequency of each two cluster over it.

We thank the reviewer for raising this important point. The seasons at ATTO, as we defined them in this manuscript, are based on the aerosol number concentration, which is not a seed variable for our clustering method. Despite this, the BC concentration is highly correlated with the aerosol concentration, but some variations of aerosol composition give rise to some distinct

features. In the wet season, LRT events bring high BC and mineral dust concentrations, and during the dry season, there are occurrences of relatively clean conditions after rainfall.

Furthermore, the “transition months” of June and July usually register higher aerosol concentrations without the increase in BC, as the vegetation in the Amazon is not dry enough to burn yet. Due to these discrepancies, and the variability of BC from one season to the other being the most important difference in the aerosol properties, the k-means algorithm with 2 clusters for the entire year mainly separated periods with higher BC concentration in one cluster and cleaner conditions in the other.

In our tests, shown in Figure 1, when we run the k-means algorithm for 2 clusters on the entire dataset, 65% of the points align with their respective seasons. Most of the disagreements arise during the early dry season, when rainfall is minimal, yet BC concentrations remain relatively low compared to the dry season peak. Additionally, LRT events contribute to higher BC concentrations, even though they occur during the wet season. These periods mostly agree with the “low-impact” and “LRT-wet” clusters of our analysis, respectively, so we consider this quality check satisfactory.

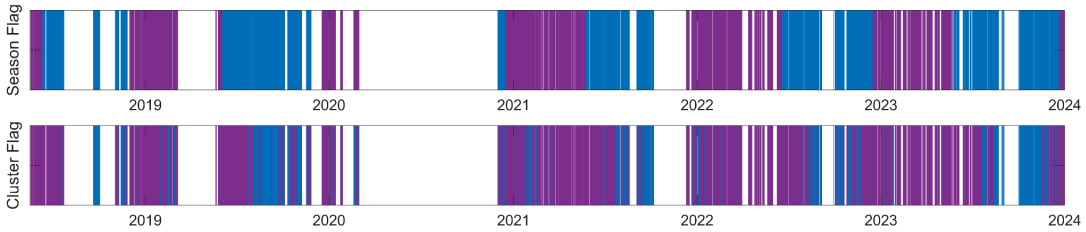

***Figure 1**: Comparison of the season definitions following the methodology described in the manuscript (wet season in purple and dry season in blue) with the cluster assignment test proposed by the reviewer.*

- Line 460: “Section 3.5 Aerosol mass scattering efficiency”. In my opinion, I found that the last section (on aerosol mass scattering efficiency) reads somewhat independently of the earlier parts of the study. In the introduction of the manuscript all the optical parameters (SSA, refractive index, MSE, etc…) are presented at the same time showing their usefulness for characterizing aerosol population properties. If this section were more clearly related to the objectives of the document, the cohesion of the entire document would be improved. I recommend, for example, to clarify at the begging of this section the purpose of this analysis within the manuscript main goals.

  We thank reviewer #1 for the suggestion. Indeed, the mathematical and theoretical definitions at the beginning of the section may confuse the reader from the main message of our study. In this case, the point is the determination of another important intensive aerosol property, which relates mass concentration to visibility and radiation scattering, during the conditions outlined by each cluster, which show differences throughout the year and the two height levels of the tower.

To improve clarity and further connect the sections with the goals of this manuscript, the following text was added to the beginning of Section 3.5:

"*The MSE can also be used to characterize the particle populations, as it indicates the scattering power of an aerosol type and relates mass concentration to visibility and aerosol optical depth.*"

Responses to Reviewer #2 of the manuscript "Distinct aerosol populations and their vertical gradients in central Amazonia revealed by optical properties and cluster analysis" by Valiati et al., submitted for publication in Atmospheric Chemistry and Physics

Dear Editor, we would like to thank Reviewer #2 for the valuable comments and useful suggestions to improve our manuscript. Below, you can find answers and actions for each individual comment.

Manuscript by Valiati et al. presents long-term in situ data (5 years) from atmospheric aerosol measurements at tall tower using state-of-the-art aerosol instrumentation in the Amazon region. The study provides a unique data set investigating changes in aerosols optical properties in combination with their chemical composition, vertical gradient and size distribution as well as aerosol origin.

From this perspective, I find this study suitable for publication in the ACP. I have only a few minor comments that would be worth to address.

Dear editor, we thank Reviewer #2 for the positive feedback and for recognizing the value of this manuscript, which utilizes a large dataset based on in situ instrumentation and arrives at a characterization of the aerosol properties at two heights over the Amazon. The specific comments will be addressed over the next pages of this response.

- Fig. 2, matrix: Could you add to this Ångström matrix a classification scheme presented e.g. by Cappa et al. (2016) (in their Fig. 8d) for better orientation of possible aerosol composition? And can you provide a same matrix for 60 m data? Considering the different year-variations of SAE (Fig. 1a), certain changes could also be reflected in this image. It can be included in the supplement.

  We thank reviewer #2 for the suggestion. The classification from Cappa et al. (2016) was included in the plots, as shown in Figure 2. The 60 m Ångström matrix indeed shows different features, as the common occurrence of larger particles prevents the SAE from reaching values higher than 2.0 most of the time.

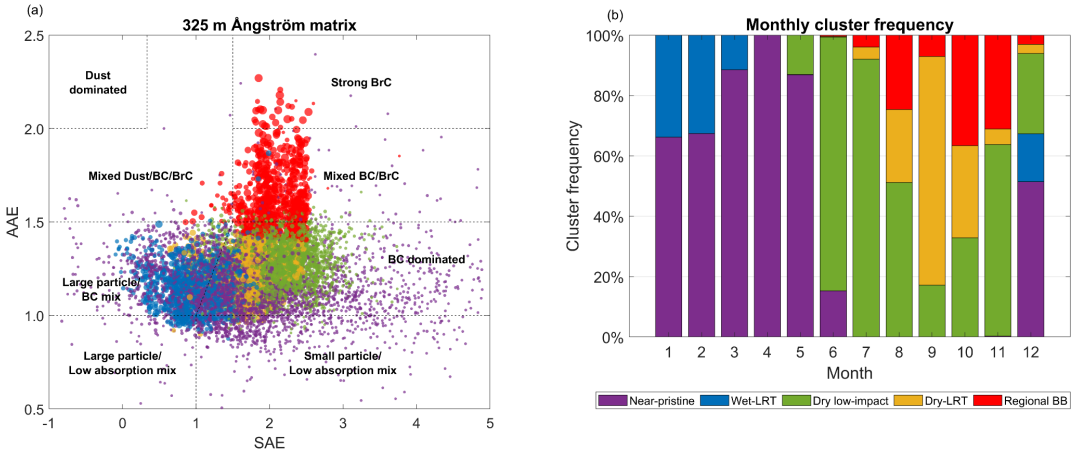

***Figure 2**: Updated version of Figure 2 from the manuscript, with the classification from Cappa et al., 2016.*

Figure S7 now shows the Ångström matrix for the 60 m measurements.

Figures 2a and S7 now have the classification scheme presented by Cappa et al. (2016), with the figure legends altered accordingly with the text: "*the different highlighted regions relate to possible particle composition as introduced by Cappa et al. (2016).*"

In the text, we added short discussions about the new figure in Section 3.2:
"*Furthermore, Figure S7 shows the Ångström matrix for the 60 m measurements highlighted with the colors that represent each cluster.*"
"*Figure S7 also shows that during near-pristine periods the SAE rarely surpasses 2.0 at 60 m, highlighting the role of larger particles found at this height.*"

- The same eBC trend at both heights (60 vs. 325 m, Fig. 1c) is an interesting result. Some shorter time measurements on towers show that eBC concentrations usually decrease with height - e.g. Sun et al. (2020). The vertical distribution of other components in individual clusters is also similar (60 vs. 325 m, Fig. 4e), suggesting that atmospheric mixing at both levels is similar. Could this indicate that the mixed planetary boundary layer (PBL) is much higher than 325 m? Can you compare this with other studies where the concentration gradient changes with elevation?

  Indeed, the convective mixing that starts in the morning reaches the top of the ATTO tower in only a few hours, and grows to over 1 km in height during most days. This explains the diurnal similarities between the aerosol concentrations at both heights. Another feature that can be discussed when comparing our result with the findings of Sun et al. (2020) is that in our case the tower is in a remote forest area, and in this 2020 paper's case the tower is located in a large city, where BC is locally emitted. The absence of a BC height gradient during the entire year is indicative of BC particles from faraway sources that reach both heights at ATTO somewhat evenly. Franco

et al., 2024 (DOI: 10.5194/acp-24-8751-2024) also observed this distinct pattern of similar BC concentrations, even higher at 325 m, at ATTO during the wet season, which could also be related to more turbulence closer to the canopy mixing the air more efficiently.

In the revised manuscript, this discussion was introduced by the addition of the following text in Section 3.1:
"*such as long-range transport of absorbing particles, unlike the findings of Sun et al. (2020), obtained for a large urban area, where the observed height gradient of BC concentration is indicative of nearby ground sources of pollution.*"

- 167: „…and ρOrg calculated based on Kuwata et al." Please, provide average organic matter density + standard deviation you calculated for the measurement period.

  We appreciate the reviewer's suggestion. In this study, the organic matter density (ρOrg) was estimated following Kuwata et al. (2012). We calculated daily values and reported the median ± standard deviation over the entire measurement period: 1.53 ± 0.11 g cm^{-3} at 60 m and 1.65 ± 0.07 g cm^{-3} at 325 m. These values have been added to the manuscript in Section 2.2.3.

  "The SMPS-derived mass concentration was calculated by multiplying the volume concentration, assuming spherical particles, by the average aerosol density at the site. The ACSM aerosol composition was used to estimate aerosol density, with the following densities for the species: $\rho(NH_4)_2SO_4$ = 1.77 g cm^{-3}, ρNH_4NO_3 = 1.72 g cm^{-3}, ρBC = 1.8 g cm^{-3}, and ρOrg calculated based on Kuwata et al. (2012). The resulting median ± standard deviation values of ρOrg, based on daily calculations, were 1.53 ± 0.11 g cm^{-3} at 60 m and 1.65 ± 0.07 g cm^{-3} at 325 m."

- 170: you write: „Lastly, the scatter and linear regression between ACSM daily concentrations and SMPS-derived daily concentrations minus eBC indicate that the methodology could accurately represent ACSM aerosol concentrations at both heights." Can you prove this statement, for example with a graph in the supplement?

  We appreciate the reviewer's attention to this detail. This statement is now supported by Figure S1, which shows the evolution of the collection efficiency over time and the linear relationship between ACSM and SMPS-derived concentrations.

- Some parts of the manuscript are too long and should be shortened or moved to the supplement. An example is section 2.4. (Clustering procedure), but in general, I think the entire methodology could be shortened appropriately.

  We thank reviewer #2 for the suggestion. Some parts of Sections 2.2.1 and

2.4 were shortened to improve readability, which are highlighted below.

"*A validation analysis for the absorption coefficients measured at ATTO can be found in Franco et al. (2024), where it has been found that, after corrections, the MAAP and AE33 measurements at the same height are within their respective uncertainty when interpolated to the same wavelength. Due to this agreement, it was chosen that the default eBC concentration is estimated using the MAAP measurements, while the AE33 is used when there are gaps in the MAAP data.*" was replaced by "*A validation analysis for the absorption coefficients measured at ATTO can be found in Franco et al. (2024), where it has been found that, after corrections, the MAAP and AE33 measurements at the same height are within their respective uncertainty when interpolated to the same wavelength.*" on Section 2.2.1.

"*The absorption and scattering coefficients are extensive properties, as they scale with the aerosol concentration. These measurements also allow for the estimation of some intensive properties that relate to the chemical and physical properties of the particles that form the aerosol population. The single scattering albedo (SSA) was calculated with Equation 1, where σscat,λ and σabs,λ are the scattering and absorption coefficients measured at a particular wavelength.*" was replaced by "*The optical measurements also allow for the estimation of some intensive properties. The single scattering albedo (SSA) was calculated with Equation 1, where σscat,λ and σabs,λ are the scattering and absorption coefficients measured at a particular wavelength.*" on Section 2.2.1.

"*This study first employed four kinds of unsupervised clustering methods, feeding all data points to the algorithms, and compared the resulting clusters on several parameters in order to choose the one that fits the defined criteria for this analysis. The k-means algorithm performed the best on the silhouette and Davies-Bouldin index tests, commonly used tools to evaluate cluster assignments (Aliaga et al., 2025). Thus, the unsupervised k-means method was chosen to perform the cluster separation. The k-means algorithm is a well-known machine learning clustering method introduced by MacQueen, 1967, which consists of splitting the sample set into a previously defined number of clusters in an iterative process. More details can be found elsewhere (Li et al., 2022).*" was replaced by "*This study employed the k-means algorithm, which performed the best on commonly used tools to evaluate cluster assignments (Aliaga et al., 2025), and more details about the algorithm itself can be found elsewhere (MacQueen, 1967; Li et al., 2022).*" on Section 2.4.

- Fig. 6: Can you add correlations of individual fits to scatter plots?

The values of the R^2 correlation coefficients were added to the plots of Fig. 6 and the statistical significance of each fitted coefficient was verified by the p-value, which was less than 0.05 for all cases.

Figure 6 legend now says:

"*Scatter plots of non-refractive PM1 mass and scattering coefficient at 525 nm measured at 325 m (a-e) and 60 m (f-j) for each cluster, with linear fit and the R^2 determination coefficient. The shaded areas indicate the 95% confidence interval.*"

The second paragraph of Section 3.5 now says:
"Figure 6 shows the correlation plots between the aerosol mass and scattering coefficient measured at 60 and 325 m during the conditions flagged by each cluster, while Table 4 summarizes the values found by the linear regressions for both heights, which were statistically significant for all cases with the fitted coefficients showing $p < 0.05$."

- References, l.607, Andreae et al. (2015): Citation to preprint is provided, please update with link to final article.

Thanks for pointing this out. It was updated in the final revised version.

"The Amazon Tall Tower Observatory (ATTO) in the remote Amazon basin: overview of first results from ecosystem ecology, meteorology, trace gas, and aerosol measurements., Atmos. Chem. Phys. Discuss., 15, 11599–11726, https://doi.org/10.5194/acpd-15-11599-2015, 2015." was changed to *"The Amazon Tall Tower Observatory (ATTO): overview of pilot measurements on ecosystem ecology, meteorology, trace gases, and aerosols, Atmospheric Chemistry and Physics, 15, 10 723–10 776, https://doi.org/10.5194/acp-15-10723-2015, 2015."*

References:

Cappa, C. D., Kolesar, K. R., Zhang, X., Atkinson, D. B., Pekour, M. S., Zaveri, R. A., Zelenyuk, A. and Zhang, Q.: Understanding the optical properties of ambient sub-and supermicron particulate matter: Results from the CARES 2010 field study in northern California, Atmos. Chem. Phys., 16(10), 6511–6535, doi:10.5194/acp-16-6511-2016, 2016.

Sun, T., Wu, C., Wu, D., Liu, B., Sun, J. Y., Mao, X., Yang, H., Deng, T., Song, L., Li, M., Li, Y. J. and Zhou, Z.: Time-resolved black carbon aerosol vertical distribution measurements using a 356-m meteorological tower in Shenzhen, Theor. Appl. Climatol., 140(3–4), 1263–1276, doi:10.1007/s00704-020-03168-6, 2020.