Authors' Responses to Referee #1 Comments

We thank Referee #1 for their valuable and constructive comments, which were helpful in improving the clarity and impact of this manuscript. Here, we have provided more information/discussion on our methods, particularly the PMF approach, to make it clearer to the reader the research we have done in this study. We have reproduced their comments below in black text and we have number-listed Referee #1's comments for clarification when addressing comments relevant to both referees. Our responses are in blue text and any additions to the manuscript are in red text. Our reference to line numbers is based on the initially submitted manuscript.

Specific comments:

1. L12. Are the three dust events enough for the conclusions of the work. Consider re-writing.

In this L12, we are talking about the three kinds of LRT events assessed, not about the number of events. Nonetheless, following this and Comment #2 from Referee #2, we clarified the statement, giving information about the annual frequency of the identified events for the study period.

2. L 90. A table with the days exceeding the PM2.5 should be included for each year and clarify if you are using the average the authors mean.

We re-wrote the sentence in L91 to improve clarity and added a table in the supplemental material (Table S1) showing the average $PM_{2.5}$ concentration and daily $PM_{2.5}$ exceedances for each year of the study period as suggested.

"... the average daily concentrations of $PM_{2.5}$ frequently exceed national and international standards (WHO, 2021, 15 ug/m^3) in the valley, with more than 60% of days exceeding this limit at most stations located in the urban areas of the AV (see Supplementary Table S1)."

Table added in supplementary material:

Table S1. $PM_{2.5}$ average concentration (ug/m^3) and percentage of exceedance of daily WHO [RP1] (15 µg/m^3) and national (37 µg/m^3) standards for urban stations from the Aburrá Valley's official air quality network during the study period.

Official AQ Station	2019			2020			2021			2022		
	Average	% WHO	% National	Average	% WHO	% National	Average	% WHO	% National	Average	% WHO	% National
CEN-TRAF	28.6	99.2	14.2	25.5	87.2	10.9	25.0	97.0	1.9	26.7	98.6	5.5
ENV-HOSP	17.5	54.0	3.0	16.8	41.0	6.8	14.3	39.2	0.3	15.3	46.0	0.5
EST-HOSP	18.8	63.3	4.4	17.2	45.1	5.7	14.8	43.8	0.3	15.7	50.4	0.5
ITA-CJUS	23.4	91.5	6.8	22.7	79.0	10.7	20.8	87.7	1.4	23.7	90.7	4.9
ITA-CONC	18.5	66.3	2.7	19.6	64.2	6.8	16.6	61.4	0.3	16.7	47.1	0.3
MED-ALTA	22.7	92.6	5.2	19.9	69.7	6.0	18.6	79.7	0.5	19.4	81.6	1.1
MED-ARAN	22.2	86.0	7.4	20.1	62.8	9.6	17.8	68.2	0.5	19.3	77.8	0.5
MED-BEME	21.7	83.0	4.7	19.5	64.2	7.1	19.9	78.1	0.5	20.2	82.5	0.5
MED-LAYE	19.2	71.5	2.2	17.7	50.3	5.5	17.1	65.2	0.3	17.7	61.9	1.4
MED-TESO	17.2	54.0	3.3	16.8	42.3	6.8	14.9	45.8	0.3	15.7	50.1	1.1
MED-VILL	18.9	66.8	3.8	18.4	53.8	6.8	16.2	57.0	0.8	17.4	63.6	0.3
SAB-RAME	18.5	63.0	3.3	17.8	50.8	6.8	15.6	50.4	0.3	17.4	63.8	0.8

Note: this table includes the official air quality stations in the most urbanized zone of the Aburrá Valley. The MED-BEME site corresponds to the location of the chemical sampling campaign (shadowed row).

3. Authors should clarify what do they mean by official concentrations?

The $PM_{2.5}$ samples from the Low-Vol equipment were defined as the reference method following Colombian national requirements aligned with the CFR 40 Appendix L to Part 50 – (US-EPA, 2011), the reason why, for this study, this was agreed as the "official" $PM_{2.5}$ measurement for the campaign. But we agreed it was confusing, so clarify that in L124 as follows:

"The PM25 was additionally sampled by a Low-volume sampler (Reference: Wilbur TE-WILBUR - Tish). Since these measurements followed the reference method described by the 40 CFR Part 50 standards suggested by the US-EPA (2011) and adopted by Colombian regulations (MinAmbiente-Colombia, 2010), the calculated average 24-hour $PM_{2.5}$ concentrations for the Low-volume are used for the positive matrix factorization model."

4. L 125: Secondary organic carbon cannot be measured, as author stated, so this should be clarified.

The sentence in L126 was changed to clarify that the SOC was calculated in the study.

"To complement the characterization of carbonaceous matter, the secondary organic carbon (SOC) was calculated using the elemental carbon trace methodology (Huntzicker et al., 1986)"

5. Are the results statistically significant?

Following the body of the comment, we understand that this question targets the PMF results with concern about the size of the datasets. We recognize that the sample size is a critical

factor for the PMF results' reliability and we did previously consider the factors/metrics for the uncertainties for this method. As a result, we have added the text below to improve the manuscript. Two additional rows were also added to Table 2.

For Methodology section in L253

For the PMF we anticipate small datasets since the target events may not significantly impact the entire $PM_{2.5}$ chemical campaign. Nonetheless, multiple studies with PMF samples ranging from 14 to 30 have reported useful and meaningful results (Yu et al., 2015; Haghnazar et al., 2022; Via et al., 2022). As the sample dataset decreases, rotational ambiguity caused by infinite valid solutions strongly affects the results and increases overall uncertainty (Manousakas et al., 2017). To mitigate the error, the software EPA PMF v 5.0 allows for estimating the effect of random errors and rotational ambiguity in the dataset using bootstrapping (BS) and Displacement (DISP) tools. While BS evaluates random errors by performing 100 runs with randomly relocated blocks of observation of the original dataset, DISP focuses on indicating rotational ambiguity by adjusting up and down all values in the factors profile restricted by 4 allowed changes in the calculated Q (dQmax) and monitoring major factors swaps (Noris and Duvall, 2014).

Additionally, constraining the base run can improve the solution when data is limited by reducing the rotational space (Dai et al., 2020). The PMF software has the functions to "pull down maximally," "pull up maximally," "set to zero," and manually set the profile concentrations. While the first two options are soft constraints, the third and fourth are hard constraints and require a high level of confidence in the magnitude of the profile contributions. For this study, only soft contains are contemplated. The constraint increases in the final Q value, which should be less than 5% (i.e. %dQ < 5%), the recommended maximum change (Noris and Duvall, 2014). For this study, the %dQ was set by default <0.5%.

For the results section in L364:

In evaluating errors of the PMF, none of the three simulated scenarios showed significant rotational ambiguity, nor were substantial random errors in the dataset after running the DISP and BS methods. Results for the DISP method showed no factor swaps for all dQmax values. In the BS analysis, outputs were considered stable, yet not all base factors were mapped to the boot factors. On average, the percentage of factors correctly mapped was 84%, which is in line with (Noris and Duvall, 2014), where a minimum of 80% mapped factors are suggested for interpretability and to support the number of factors selected. The target BB-LRT, Dust-LRT, and Volcanic-LRT profiles for the models were mapped in 88%, 80%, and 98% of the runs, respectively. The quantiles 25th and 75th of the $PM_{2.5}$ contribution rates calculated for the target profiles in the BS runs are presented in Table 2. The contribution rate variability represented by the BS runs is particularly important for small database models, where contribution rate results might be unstable compared to the more stable profile (Feng et al., 2023).

Final base models were constrained to improve the correspondence between the chemical profiles found by the PMF and the profiles expected based on the identified emission sources. Specific constraints were defined in the different modeling scenarios to refine the factor profiles. The model's factors are presented in Figure S3. For the BB-LRT model, three soft

constraints were applied to the "coal boiler" factor to pull up the EC1, Se, and As concentrations. Conversely, the Ag, Se, and EC1 concentrations were pulled down in Factors "ceramic industry", "gasoline" and "Diesel". For the Dust-LRT model, four soft constraints were established: for "vehicular emissions", the concentrations of EC and Ni were maximally adjusted downward and upward, respectively; for "resuspended material," Ni was reduced to improve the fit; and for "biomass burning," OC was maximally increased. In the volcanic-LRT event scenario, two soft constraints were implemented. For volcanic-LRT, we pulled down the concentration of Cu and Mg, while Se was pulled up for "secondaries" and Cu was pulled down for "vehicular". For the models the %dQ (i.e. the Q change because of the constrain) is < 1%

In addition, this study includes different independent methodologies aiming to add robustness to the appreciated small dataset PMF assessment. For instance, a partial answer to one comment not included in the "specific comment", the delivery described as the monthly frequency used the daily average CAMS's products instead of the samples; the back trajectories and a meteorological assessment was made as complements. The $PM_{2.5}$ comparison in Figure 7 includes the daily average $PM_{2.5}$ concentration, and after presenting the PMF result, the comparison of all measured compounds was made in Figure 9.

6. Figure 8 shows the profiles identified, but they are not clear as the one identified as a Dust and volcanic have similar composition, how can the authors make sure of the name of the profile. More details should be given.

Targeting Referee #1's main comment, we wanted to clarify first that all measured elements were used for dust and volcanic aerosols. The campaign did not measure some key trace elements from April 2021, including K, Mg, and Na concentrations. Because these elements are trace elements for these two factors, the decision was to restrict the samples to the period where all elements were measured instead of removing them from the model, as is explained in L253. The carbonaceous species were removed, but the total OC and EC were still included in the models, considering the elements, not carbon, as the target compounds for these sources.

Given the diverse chemical structures of the sources and the mixing between them due to the different phenomena associated with atmospheric dispersion and dynamics, identifying physically significant profiles is based on groups of chemical components called pseudo components according to the relevant species or sources. That is why a mixture of sources is often a constraint. To accurately label the factors, it is crucial to utilize tracers supported by literature and thoroughly evaluate the sources at the site. The site sources assessment is backed by Gómez et al. (2021) (as L257 says) and the first part of the study for LRT events. Dust and volcanic aerosol share key tracers identified as crustal tracers, and those have marked contributions in the majority of compounds. In addition to Figure 8, Figure S3 is a good complement for appreciating what is described for the profile identification and then recognizing key differentiation compounds better. As described in L394, in addition to the crustal tracers, the contribution of the anions led by F-, NO3, and SO4, together with Na and K, are crucial trace compounds for the Volcanic-LRT factor. Meanwhile, for the Dust-LRT factor, Ca and Ti were key minerals (described from L384).

It is important to highlight that both factor profiles belong to independent models, so differentiating one from the other was not a direct necessity. Besides, other analyses supported both events, helping us back up the results.

An additional reference was added to back up the volcanic factor selection:

"Cu and Zn are other tracers observed here and identified before for Colima Volcano in the southeast of the ring of fire (Miranda et al., 2004)."

7. What is the purpose of the results in Figure 9?

Figure 9 has two main objectives. The first objective is similar to Figure 7 for $PM_{2.5}$, comparing the compound concentrations for days affected by a LRT event and the closest and most meteorologically similar days (i.e., days before and after the event), as described in L404. The second is present cation concentration and some index magnitudes (e.g., OC/EC) not included in the PMF, as described by L409. Since the campaign had a big discontinuity in the cations measurement, these were not used in the PMF for the profile characterization. This is part of the methodologies that, as was explained at the beginning of the document, aim to increase robustness in the analysis, in this case, for the selected profile.

We noticed that the connection between L409 and L424 and Figure 9 might not be evident, so we have rewritten those sections, including clearly referencing the figure.

"Unlike the PMF model, the comparisons in Fig. 9 contain analysis of the cations, the carbon matter species and the OC/EC and SOC/OC ratios for every type of event. Here, the major elements generally have a more significant increment in LRT events. Some elements supported the model's fingerprint (Figure 8), e.g., OC, OC1, OC2, SO4 2 for BB; Fe, Al, and Ti for dust; and Si, Al, Fe, Ca, Mg, and Na for volcanic aerosols (Figure 9). The OC was significantly higher for BB, presenting a median OC/EC ratio of 11.3 that surpasses common urban combustion ratios like from fossil fuel (~4), combustion, and Diesel exhausted (<1) (Pani et al., 2019). Although OC/EC is more commonly used to identify sources of urban combustion and BB, some studies have shown its potential for determining the influence of volcanic activity (Pongpiachan et al., 2019)."

"The elevated concentrations of ions in the days of events (see Fig. 9) also support the modeled profiles (Fig. 8) and align with the literature. In addition to the ions observed in the PMF profile for the BB, the cations K+ are representative ions (Rastogi et al., 2014; Moreno et al., 2023) that present significant increments for this type of event. Regarding the volcanic aerosol compositions (Fig. 9), the observed increment in Na+ and K+ also aligns with previous reports (Moreno et al., 2023; Mather et al., 2003; Roberts et al., 2018). On the other hand, although the PMF's fingerprint presented a high contribution of SO_2^{4-} and F^-, this was not enough for a significant rise in daily concentrations shown in Fig. 9."

Added references

Feng, J., Song, N., & Li, Y. (2023). An in-depth investigation of the influence of sample size on PCA-MLR, PMF, and FA-NNC source apportionment results. *Environmental Geochemistry and Health*, *45*(8), 5841–5855. https://doi.org/10.1007/s10653-023-01598-5

Haghnazar, H., Johannesson, K. H., González-Pinzón, R., Pourakbar, M., Aghayani, E., Rajabi, A., & Hashemi, A. A. (2022). Groundwater geochemistry, quality, and pollution of the largest lake basin in the Middle East: Comparison of PMF and PCA-MLR receptor models and application of the source-oriented HHRA approach. *Chemosphere*, *288*, 132489. https://doi.org/10.1016/j.chemosphere.2021.132489

Manousakas, M., Papaefthymiou, H., Diapouli, E., Migliori, A., Karydas, A. G., Bogdanovic-Radovic, I., & Eleftheriadis, K. (2017). Assessment of PM2.5 sources and their corresponding level of uncertainty in a coastal urban area using EPA PMF 5.0 enhanced diagnostics. *Science of The Total Environment*, *574*, 155–164. https://doi.org/10.1016/j.scitotenv.2016.09.047

MinAmbiente-Colombia: MANUAL DE DISEÑO DE SISTEMAS DE VIGILANCIA DE LA CALIDAD DEL AIRE, https://www.minambiente.gov.co/wp-content/uploads/2021/06/Protocolo_Calidad_del_Aire_-_Manual_Diseno.pdf, 2010

Noris, G. and Duvall, R.: EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, https://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf, 2014

Via, M., Chen, G., Canonaco, F., Daellenbach, K. R., Chazeau, B., Chebaicheb, H., Jiang, J., Keernik, H., Lin, C., Marchand, N., Marin, C., O'Dowd, C., Ovadnevaite, J., Petit, J.-E., Pikridas, M., Riffault, V., Sciare, J., Slowik, J. G., Simon, L., … Cruz Minguillón, M. (2022). Rolling vs. seasonal PMF: real-world multi-site and synthetic dataset comparison. *Atmospheric Measurement Techniques*, *15*(18), 5479–5495. https://doi.org/10.5194/amt-15-5479-2022

Yu, W., Liu, R., Wang, J., Xu, F., & Shen, Z. (2015). Source apportionment of PAHs in surface sediments using positive matrix factorization combined with GIS for the estuarine area of the Yangtze River, China. *Chemosphere*, *134*, 263–271. https://doi.org/10.1016/j.chemosphere.2015.04.049

Zhang, Y., Sheesley, R. J., Bae, M.-S., & Schauer, J. J. (2009). Sensitivity of a molecular marker based positive matrix factorization model to the number of receptor observations. *Atmospheric Environment*, *43*(32), 4951–4958. https://doi.org/10.1016/j.atmosenv.2009.07.009