Authors' Responses to Referee #2 Comments

We gratefully acknowledge Referee #2 for their valuable and constructive comments, which helped improve the clarity and impact of this manuscript. We have reproduced their comments below in black text and we have number-listed Referee #2'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.

1. L10. "During these LRT events, the BB fraction of PM2.5 dominates by frequency and amount, averaging 11.14 µg/m3 (38%). On average, dust and volcanic degassing contribute 6.77 µg/m3 (34%) and 6.46 µg/m3 (30%) of the concentrations." This phrase might be confusing, specially the second part. What the authors really mean is that averaging over LRT events, dust and volcanic degassing contribute 34% and 30% of PM2.5? Is that the total? Something should be said in the abstract to at least provide the reader with an idea of the observed frequency of LRT events, or its typical duration (so those other numbers could be better contextualized).

Thank you for your comment. Please see our response to Referee #2's Comment #2 below.

2. L12. "Of the three, dust events showed fewer affected days." I would consider rewriting.

To answer this comment and the previous one, we include the annual average frequency of the LRT events identified for the study period. This new information was added in L7 as:

"Annually, we found that on LRT of aerosols from BB, dust and volcanic degassing influence approximately 13, 8 and 13% of days, respectively."

3. L63. "In Colombia, the Aburrá Valley (AV) has made substantial progress in monitoring and identifying agents of the state of air quality in the territory, managing to report significant affectation driven by external sources (SIATA, 2021)." Consider re-writing or removing it altogether.

Following the comment, the sentence is removed and rewritten according to the idea of the paragraph. Now, this paragraph starts as:

"In the last few years, different studies in Colombia have made substantial contributions in monitoring and identifying impacts caused by open fire emissions on air quality (see e.g., Hernández et al., 2019; Mendez-Espinosa et al., 2019; Ballesteros-González et al., 2020; Rincón-Riveros et al, 2020; Henao et al., 2021). Nonetheless…"

4. L64 "In the territory, as on the national scale" To which territory do the authors refer to? Please, consider removing or rewriting.

We removed the text in L63 and changed it to consider Referee #2's Comment #5 below.

5. L65. There are at least 2 relevant studies in the region that could possibly enrich the discussion in the introduction:

- https://doi.org/10.1016/j.atmosenv.2019.01.051 which demonstrate the high concentration of PM2.5 and ozone in the Orinoco river basin during high BB seasons.
- https://doi.org/10.5194/acp-20-7459-2020 which shows the correlation of BB tracers with regional biomass burning activity.

We appreciate the provided references. As suggested, the papers have been included in different phrases in the introduction:

L20: "Long-range transport (LRT) of aerosols influences the chemical composition of air over hundreds to thousands of kilometers (Kaneyasu et al., 2014; Wang et al., 2015; **Rincón-Riveros et al., 2020**)"

L63: "In the last few years, different studies in Colombia have made substantial contributions in monitoring and identifying impacts caused by open fire emissions on air quality (see e.g., **Hernández et al., 2019;** Mendez-Espinosa et al., 2019; Ballesteros-González et al., 2020; **Rincón-Riveros et al, 2020;** Henao et al., 2021)."

L46: "Therefore, transboundary emissions from open fires in the Orinoco basin and the Caribbean are significant drivers of intra-annual periods of hazardous air quality for Colombian cities such as Bogotá, Medellín, Arauca, Yopal, Bucaramanga and Villavicencio (Mendez-Espinosa et al., 2019; **Hernandez et al., 2019; Rincón-Riveros et al., 2020;** Henao et al., 2021; Rodríguez-Gómez et al., 2022)...".

6. L71. "To the AV, obtaining the …" Consider removing "To the AV".

The recommendation was accepted, and L71 changed.

7. L90. It would be useful to qualify this statement with data. For example, how many daily exceedances were observed in a given year? Or what is the annual mean PM2.5 concentration in the AV?

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 (following Referee #1's Comment #2).

"... 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 [RP2] in the most urbanized zone of the Aburrá Valley. The MED-BEME site corresponds to the location of the chemical sampling campaign (shadowed row).

8. L107. “However, the sampling in this later period was typically between 3 to 14 days. Therefore, while the temporal sampling resolution did decrease with time, we still have periods of intense sampling and measurements across the majority of the period.” Please rephrase as it is confusing.

Following the suggestion, we rephrase the paragraph to provide clarity and additional information about the sampling frequency. The changes were implemented in L104

“While April 2019 to July 2020 represented an intense sampling campaign with samples every three days, the frequency of the surface site observations became less intense after July 23, 2020 (i.e. up to a maximum of 2 weeks during periods of routine sampling). However, there were two extended gaps in the campaign from November 2020 to mid-March 2021 and mid-September 2021 to March 2022. Despite the decrease in sampling frequency, the measurements still provide sufficient temporal coverage to get robust seasonal and annual information on aerosol concentration level and composition in this study.”

9. L123. “Official campaign concentrations of PM2.5 were measured by a Low Volume PM2.5 ambient air sampler”. Could the authors clarify this statement? What are “official” concentrations? How do the Low-vol concentrations differ from the High-volume sampler derived concentrations? Where the latter concentrations not determined at all? Please clarify and correct the manuscript accordingly.

Both measurements were part of the campaign but with different objectives. The Low-Vol equipment was used to measure the $PM_{2.5}$ concentrations during the campaign. This instrument met the Colombian national requirements (following the CFR 40 Appendix L to

Part 50 - US-EPA, 2017), so was classified as the "official" $PM_{2.5}$ sampler of the campaign. The Hi-Vol equipment was mainly used for the chemical characterisation of measured aerosol, which had the capacity to measure higher mass required for some of the chemical characterisation methods. The samples from this instrument were collected relative to the Australian/New Zealand Standard AS/NZS 3580.9.14:2013 Method 9.14 (Standards New Zealand, 2013).

The clarification was added to the paragraph in L124:

"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 PM2.5 concentrations for the Low-volume are used for the positive matrix factorization model."

10. L125. "In addition to the carbonaceous matter, species measured included secondary organic carbon". Please re-write for clarity. SOC was not measured, but it was inferred from the measurements.

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)"

11. L150. "mean absolute percentage error of 21.5%". Is this 21.5 percent overestimation relative to the MED-BEME station? Or 21% underestimation?

Thank you for the observation; more information was added in L149 to interpret the equipment measurement differences better.

We found a good agreement between the campaign (low-vol sampler) instrument and the official MED-BEME station. For the study period, $PM_{2.5}$ concentrations from the automatic instrument had a minor overestimation against the reference method with a mean bias error of -0.76 µg/m3. The corresponding mean absolute error (MAE) was 21.5%. For $PM_{2.5}$ measurements, the low-volume as a reference method provides better precision and accuracy than the MED-BEME sensor (Tasić et al., 2012), which follows equivalent methods. Despite this, the official sensor provides continuous measurements that are used in this study for more robust comparisons. Regarding temporal variability, the Pearson correlation coefficient was 0.84, highlighting good consistency between them.

12. L187. "Here, if less than four days with values greater than the specified threshold were detected, then they were classed as outliers and removed (i.e., we are focusing on LRT events, which we define as lasting more than half a week". This is a key point in the manuscript and one that should be subject to a more specific description. Why focus on 4-day events? Dust events from LRT can impact a given location for a single day but contribute over 90% of PM2.5

to that given location on that day. If the decision is due to the sparsity of PM2.5 samples, then it should be clearly stated.

Thank you for the suggestion. The text of the paragraph was rewritten, and more details have been included to explain the selected filters more clearly from L187:

"A 7-day rolling window was used to accurately identify prolonged and intense periods of LRT events. Within this window, at least 4 days had to have values above the respective thresholds to be classified as a LRT event. We subjectively chose 4 days of elevated values due to the sampling frequency of the campaign. Here, campaign temporal sampling was ≥ 3 days, so these criteria were required to get representative samples of the aerosol characterization for the PMF analysis."

13. Figure 4. Caption and legend could be improved. No mention is made of the PM2.5 variable there. Is it monthly PM2.5 for the site? Or is ir PM2.5 attributable to LRT events? Similarly, the "All events" bar, which is black, it is not clear if there were any LRT events in which the three sources were impacting the site simultaneously.

Thank you very much for noticing the problems in the caption. In accordance with the comments, we improved the caption to clarify:

"Figure 4. Monthly frequency of days with BB-LRT (green bars), Dust-LRT (wheat bars), and Volcanic-LRT (blue bars) events as identified from the CAMS reanalysis. Overlapped events are depicted in dark blue (BB and Volcanic), orange (BB and Dust), and red (Volcanic and Dust) bars since different LRT events can happen at the same time. White bars represent the frequency of days without LRT events, while the black line shows the monthly average $PM_{2.5}$ concentration ($\mu g/m^3$) for the MED-BEME station."

Regarding black bars, we removed this category from the figure's legend after verifying there is no overlapping among the three events.

14. L315. Seasonality?

Change accepted on L315. Please see our response to Referee #2's Comment #15 below.

15. L315. "some non-event days in the different months occur" ?? Please, consider re-writing for clarity.

The beginning of the paragraph was rewritten as follows (L315):

"Although LRT events display a marked seasonality, a significant percentage of days in each month have a negligible impact from LRT events (see white bars in Fig. 4), suggesting that intraseasonal variations are also relevant in explaining the occurrence of these events."

16. L353. "On the other hand, the concentration of PM2.5 right after Volcanic-LRT significantly decreases" …. This assertion is hard to see from Figure 7c.

The sentence was rewritten to explain more clearly what we wanted to highlight. The Mann-Whitney comparison supports a lower average concentration from 1 to 7 days after the peak of Volcanic degassing events. The following text will be added to L353:

"On the other hand, the concentration of PM2.5 immediately following the peak of Volcanic-LRT showed significantly lower levels (p-value ≤ 0.1), contrasting with the subsequent days (Dte_8 to Dte_{15})."

17. L442. "The lower TCSO2 threshold derived in this study is likely linked to the CAMS product we used". It is also possible that using SO2 observations (if available from the monitoring network) for the Volcanic-LRT events could help.

The station does not have an SO_2 analyser, so we tried to use data from other stations in the city. However, the SO_2 record only had 38.53% valid data for the study period. Therefore, we decided not to include the pollutant.

Added references

AS/NZS: Methods for sampling and analysis of ambient air - Method 9.14: Determination of suspended particulate matter - PM2.5 high volume sampler with size selective inlet - Gravimetric method, https://www.standards.govt.nz/shop/asnzs-3580-9-142013/, 2013

Hernandez, A. J., Morales-Rincon, L. A., Wu, D., Mallia, D., Lin, J. C., & Jimenez, R. (2019). Transboundary transport of biomass burning aerosols and photochemical pollution in the Orinoco River Basin. *Atmospheric Environment*, 205, 1-8. https://doi.org/10.1016/j.atmosenv.2019.01.051

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

Rincón-Riveros, J. M., Rincón-Caro, M. A., Sullivan, A. P., Mendez-Espinosa, J. F., Belalcazar, L. C., Quirama Aguilar, M., and Morales Betancourt, R.: Long-term brown carbon and smoke tracer observations in Bogotá, Colombia: association with medium-range transport of biomass burning plumes, *Atmos. Chem. Phys.*, 20, 7459–7472, https://doi.org/10.5194/acp-20-7459-2020, 2020.

Tasić, V., Jovašević-Stojanović, M., Vardoulakis, S., Milošević, N., Kovacević, R., and Petroví c, J.: Comparative assessment of a real-time particle monitor against the reference gravimetric method for PM10 and PM2.5 in indoor air, Atmospheric Environment, 54, 358–364, https://doi.org/https://doi.org/10.1016/j.atmosenv.2012.02.030, 2012

US-EPA: 40 CFR Appendix L to Part 50 - Reference Method for the Determination of Fine Particulate Matter as PM2.5 in the Atmosphere, https://www.govinfo.gov/app/details/CFR-2017-title40-vol2/CFR-2017-title40-vol2-part50-appL/context, 2011.