

Response to reviewer #2

Comments from reviewer #2:

Major concerns

1. The study determines a series of organic markers in PM_{2.5} with two-hour resolution in Hong Kong for one month. However, the data management is inadequate. The mean and median values for most species are very different (at least for those shown in Figure 4), probably due to variability in compound concentrations. Therefore, it is required to know the distribution of the variables, to confirm the assumption of the normal distribution done by the authors.
2. The authors must show the distribution of the variables, thereby justifying the type of statistics to be used, whether parametric or non-parametric, or some other mathematical proposal for data management. They should double-check any statements or suggestions they made.
3. Throughout the document, no statistical evidence is shown to support the conclusions of the values comparisons between the different periods. Nor the use of Pearson correlations. Subjectivity should be avoided in the discussion. Appropriate statistical tests will verify the hypotheses proposed by the authors. The article cannot be published in its current state and must be resubmitted for consideration.

We sincerely appreciate the professional comments on data management. Since the three major concerns are related to each other, we provide a combined response to them. In brief, we have examined the distributions of the observational data with Shapiro-Wilk tests and found that in most cases they did not follow a normal distribution, which is actually a common phenomenon for urban air quality measurement. As a response, we show more metrics of the data, including mean \pm 95% confidence interval, median, 25th percentile, and 75th percentile. Moreover, non-parametric Mann-Whitney U tests have been performed to compare any two sets of data that does not conform to the normal distribution. We've also changed all the coefficients of determination to Pearson correlation coefficients throughout the manuscript. The methods are described in a new section 2.4.

2.4 Statistical analysis

Due to uneven emission patterns, non-linear chemistry and changing meteorological impacts, the distribution of urban air pollutants concentration is often uncertain. This determines how we describe the data. In this study, we examined the data distribution with the Shapiro-Wilk tests. It was found that the concentrations of most species studied did not follow a normal distribution, neither over the entire observation period nor within sub-periods defined by air mass categories, with only a few exceptions (e.g., palmitic acid in Case I and 2-MGA in Case II). Therefore, it is insufficient to analyze and discuss the data based only on the mean values. As a supplement, we also show the 25th percentile, median

and 75th percentile where necessary. However, to avoid redundancy, not all the metrics are referenced in all discussions. Moreover, given the non-normal distribution of the data, non-parametric Mann-Whitney U tests were performed to compare any two sets of data. The significance levels are expressed as p values. Both the Shapiro-Wilk test and the Mann-Whitney U test were implemented using the `scipy.stats` package in Python. All the correlation coefficients shown in this paper are derived from Pearson correlation analysis and are significant ($p < 0.05$), unless otherwise specified.

For details, please refer to lines 178-189.

Correspondingly, the discussions have been updated with the addition of median where necessary. It is our pleasure to invite the reviewer to review the revisions highlighted in blue throughout the manuscript and supplement. In this file, our point-by-point responses are formatted in italics.

Minor comments

1. Lines 100-104. Place the mass of the compound isotopically labeled with carbon 13, added to the TAG.

The volume of internal standards that were added on top of the samples was 5.22 μL and the concentrations for 41 species were in the range of 0.0625–0.5 $\text{ng}/\mu\text{L}$. So, the added mass varied between 0.33 ng and 2.61 ng for different species. They were either deuterated (^2H -containing) or ^{13}C -containing compounds, which had little impact on the target compounds in the air. More information of the internal standards is provided.

A 5.22 μL of mixture of deuterated (^2H -containing) or ^{13}C -containing compounds at the concentration of 0.0625-0.5 $\text{ng}/\mu\text{L}$ was used as internal standards (IS), which was injected on top of every ambient sample and subject to the same treatment and analysis procedures. Therefore, the ISs could track and correct for the changes in instrument sensitivity. More details about the AMS and TAG operations, including the IS information, can be found in our previous studies (Lyu et al., 2020; Huo et al., 2022).

For details, please refer to lines 109-113.

2. Lines 105-109. Change below to "throughout the document".

Accepted with thanks.

3. Lines 110-114. This way of associating the variables with their respective concentrations is confusing. Each value must be associated with the corresponding contaminant. Apply to the entire document.

Thanks for the suggestion. In view of the other comments on data management, we are committed to showing more metrics to describe the data. Thus, the values are removed here and presented in a supplementary table.

Table S2 shows the various metrics (mean, median and percentiles) of the trace gases, meteorological parameters and LWC.

For details, please refer to lines 121-122 and Table S2.

4. Lines 120-129. The description is very confusing. Explain more clearly why organic contaminants were not quantified with the TAG. Explain what the authors refer to as “internal standard scaled peak areas”. Check if it is an accepted term in the scientific community. Enter the corresponding equation for calculation.

As stated above, the fixed amount of internal standards was injected on top of every ambient sample and subject to the same subsequent processes, e.g., derivatization, transfer and analysis. Therefore, the variations of their signals over time were caused by changes in instrument sensitivity only. For example, the sensitivity decreases when the ion source of the mass spectrum gets dirty. The unstable sensitivity hinders understanding of the true variations of target compounds in the air. It is why we need to use internal standards to correct for the measured variations in target compounds. Specifically, we calculated the ratio of the peak area of target compounds to the peak area of the corresponding internal standards, which was defined as IS-scaled peak area. This method is widely used in chemical analysis (Williams et al., 2007; Kreisberg et al., 2009; Williams et al., 2010). However, inspired by this comment, we think it is more appropriate to define the ratio as IS-corrected response. The corrected responses would be converted to concentrations of target compounds if we had external standards to establish the concentration-response relationships. Unfortunately, we ran out of external standards in this sampling campaign, thus the target compounds were not quantified for their concentrations. The description has been revised.

Given the consistent amount of ISs injected on top of all individual samples, the variations in peak areas of ISs over time could only be attributed to changes in instrument sensitivity (including recovery rate). To understand the true variations of target compounds in the air, we calculated the ratio of the peak areas of target compounds to the peak areas of corresponding ISs with similar structures and close retention time, and the ratio is deemed as IS-corrected response. The pairing of target compounds and ISs are the same as that adopted in Lyu et al. (2020). However, we ran out of the external standards (multiple concentrations of authentic standards that are exactly the same as target compounds or surrogate standards) in this sampling campaign. Therefore, the relationship between IS-corrected response and concentration was not determined, and the concentrations of the detected species were not quantified. We use the IS-corrected responses of target compounds for analysis instead.

For details, please refer to lines 130-138.

5. Lines 151-155. Explain in detail and in the supplementary material the calculation of the method detection limit (MDL). Explain why it is only one value. In general, it should be one MDL for each compound. What sense does the MDL in $\mu\text{g}/\text{m}^3$ if the units of the compounds are shown in relative areas?

The MDL of $0.201 \mu\text{g}/\text{m}^3$ was for $\text{PM}_{1\text{-OM}}$, an individual component quantitatively measured by AMS. The calculation method is described in a supplementary text. It was the molecular markers detected by TAG that were not quantified. Their MDLs were not available, and we adopted those determined in a previous study.

The MDL for $\text{PM}_{1\text{-OM}}$ measured by AMS was determined to be $0.201 \mu\text{g}/\text{m}^3$, as described in Text S1. Since the molecular markers measured by TAG were not quantified in this sampling campaign, the MDLs determined for the same instrument ~8 months earlier were adopted (Lyu et al., 2020). For the newly detected species, e.g., phthalic acid, 2-MGA, cis-2-MBT, and 9-oxononanoic acid, we assume 10% of the average IS-corrected response as the dimensionless MDL.

For details, please refer to lines 164-168 and Text S1.

6. Lines 170-174. Avoid subjectivity throughout the document. It is not appropriate to simply say: slightly higher, slightly lower, etc., because nothing conclusive is got it. The level of significance associated with the statistical test must be set. In this case the averages are compared. As mentioned above, the use of the average must first be justified, and subsequently, the result of the comparison test and the significance level must be given. This applies to the rest of the comparisons made throughout the document.

We appreciate the reviewer for the scientific rigor. In response, we have examined the distribution of the data with Shapiro-Wilk test, used more metrics (mean, median and percentiles) to describe the data, and conducted non-parametric Mann-Whitney U tests to compare two sets of data that does not follow normal distribution.

Table S4 lists the statistics of the $\text{PM}_{1\text{-OM}}$, sulfate, nitrate, and ammonium in $\text{PM}_{1\text{-OM}}$. $\text{PM}_{1\text{-OM}}$ was found to be the most abundant component, followed by sulfate, ammonium and nitrate, whether based on the mean or median. The fraction of $\text{PM}_{1\text{-OM}}$ ($47.1 \pm 2.2\%$) was comparable to that at an urban background site in HK (43.8%), but significantly lower ($p < 0.05$) than that at a roadside site ~350 m away (57.7%) (Yao et al., 2021,2022). From the perspective of air mass categories, $\text{PM}_{1\text{-OM}}$ was the lowest in the most common marine air, followed by that in Case II with coastal air and Case I with continental air, same for CO and O_3 (Tables S5-S7).

For details, please refer to lines 193-198.

As stated above, With the improvement of data management, many discussions are updated, especially in section 3.2. However, we still retain the discussions based on averages, because the small number of high values are true values and reflect the short-duration but intensive emissions or chemical formation. Please review the detailed revisions in the revised manuscript.

7. In a separate section, the statistical tests must be described.

As stated above, we have improved the data management based on the insightful comments. For details, please refer to our response to the major concerns and the new section 2.4.

8. Lines 190-194. This assertion is not clear. Figure S4a shows that levoglucosan maximum indeed comes from the east, but between 7 and 18 km away from the site in the hour before its emission. From figure 1, it is not clear how the authors assume that the biomass combustion is located in the first kilometer east of the site. Clarify.

The original Figure S4a (Figure S5a in this revised supplement) shows higher levels of levoglucosan under easterly winds. However, it does not indicate the distance between the emission source and the sampling site, because we do not know the time between levoglucosan emission and detection. Besides, the TAG did not analyze instant samples but collected the sample for 90 minutes and analyzed it afterwards in each 2h cycle. We extrapolated the local origin of levoglucosan based on the following facts: i) there were several funeral parlors and temples east of the sampling site within 1km, where burning incense, candles, banknotes, paper products was common; and ii) there was no known source of biomass burning further east. The funeral parlors and temples were marked in the original Figure 1, which are replaced with the distribution of restaurants – a much larger source of PM₁-OM – in the updated Figure 1. For clarity, the discussions are revised as follows.

Further, we found that high levels of levoglucosan were associated with easterly winds (Figure S5). As stated in section 2.1, the burning of incense, candles, banknotes, paper products was common within 1 km east of the site. There was no known source of biomass burning further east, and east of HK are waters. Therefore, the observed variations of levoglucosan were mainly attributed to the combined effect of local emissions and wind patterns, different from the regional transport of biomass burning plumes from mainland to HK in cool seasons (Sang et al., 2011; Chow et al., 2015).

For details, please refer to lines 223-227.

9. Lines 215-219. What is the evidence to associate these alkanes with sacrificial activities? Explain what the authors refer to as sacrificial activities.

With ‘sacrificial activities’, we meant burning of incenses, candles, joss papers and paper artifacts nearby. Studies identified alkane emissions from burning some of these biomass products, and the observed alkanes exhibited similar diurnal patterns to biomass burning tracers (e.g., levoglucosan). That’s why we attributed the alkanes partially to ‘sacrificial activities’. For clarity, the following revisions are made.

Besides, both the bihourly mean and median C₂₆ – C₃₁ n-alkanes were observed with obvious peaks at 8:00 and 20:00 and might also be related to the burning of incenses, candles, joss papers and paper artifacts nearby, which was confirmed in other scenes (Lyu et al., 2021; Song et al., 2023). Other sources, e.g., vehicle emissions, might also contribute to the peak at 8:00 that was more pronounced than levoglucosan.

For details, please refer to lines 266-270.

10. Lines 220-224. Azelaic acid is not shown in Figure S5.

Azelaic acid is shown in the updated Figure S7. In response to the other reviewer, more discussions on azelaic acid are added. Please see lines 270-273 and lines 283-285.

11. Lines 220-229. The pyrene is a clear example that the average is not the best metric to compare periods. In this case, the median of the hour 20:00 is like the rest of the hourly medians. This change the discussion and conclusions.

Indeed, mean-based discussion may not be enough for the data that does not follow normal distribution. As stated above, we adopt more metrics in the revised manuscript, and the discussion is revised as follows.

Pyrene, as a representative of PAHs, was observed with higher median and mean values during 6:00-22:00. The median values showed peaks in the morning and early evening, while the mean value was the highest at 20:00. Interestingly, the high levels of pyrene correlated well with levoglucosan ($r = 0.94$) at 20:00 (Figure S9) and exhibited no correlation with oleic acid which also had the highest levels in the early evening. Therefore, the burning of biomass products nearby was responsible for the occasionally elevated levels of PAHs in the early evening. At the other times, the variation of pyrene was consistent with that of NO_x ($r = 0.80$), indicating their common source of vehicle emissions. This also explained the bimodal distribution of the bihourly median values of pyrene.

For details, please refer to lines 273-279.

12. Line 227. It is well known that pyrene and NO_x are emitted by vehicles that burn diesel. Could this correlation be associated with this source?

Yes. Pyrene at times other than 20:00 was likely to be emitted by vehicles. Please see our response to comment #11.

13. Line 228. Does the term funerary refer to crematoriums for humans and animals? since there are funeral homes that do not have this service. If so, the authors should cite studies showing the type of PAHs emitted by these sources found in this study. This information will be useful to strengthen their hypothesis.

Crematoriums for humans or animals were not the service of the nearby funeral parlors. The high levels of PAHs at 20:00 were likely related to the burning of biomass products (incenses, candles, joss papers and paper artifacts). In Chinese culture, burning these things to gods and souls is a kind of memorial and offering. There have been many studies showing PAH emissions from biomass burning.

However, we do not want to challenge the culture with this little piece of evidence. And for clarity, we rephrase this sentence to highlight the burning of biomass products.

Therefore, the [burning of biomass products](#) nearby was responsible for the [occasionally elevated levels](#) of PAHs in the early evening.

For details, please refer to lines 276-277.

14. Line 240. It's confusing. Is it sum or division?

The sentence has been rephrased for clarity.

According to Noziere et al. (2011), the ratio of 2-methylerythritol to the sum of [2-methylerythritol and 2-methylthreitol, referred to as 2-MT1/2-MTs hereinafter](#), indicates the sources of 2-MTs: primary emissions (0.35), NO_x-lean photooxidation (0.61), NO_x-rich photooxidation (0.76), and aqueous phase oxidation (0.90).

For details, please refer to lines 297-299.

15. Figure 3. Why does the time scale go from 6 to 26? What are the units? Eliminate the $\mu\text{g}/\text{m}^3$ on the scale on the right. Change color between ammonium, nitrate and coastal air bar. It's confusing. Change the color between the UV and continental air bar. It's confusing.

The number on the x-axis represented date in June 2019, which is updated with a standard date format. The colors are also changed to avoid confusion. Other changes can be found in the updated Figure 3. The unit $\mu\text{g}/\text{m}^3$ is for the species measured by AMS, which were quantified. We have added a note into this panel highlighting they were measured by AMS.

16. Figure 7a. Why does the time scale go from 6 to 26? What are the units? Improve sharpness.

Date format changed. The figure has been updated as suggested.

17. Figure S2. Why does the time scale go from 6 to 26? What are the units? Place the color bar for marine, continental and coastal air, as in figure 3.

Date format changed. The figure has been updated as suggested.

18. Figure S5. Explain the calculation to normalize the variables.

The normalization method is described in a supplementary text, which is referenced in the figure caption.

Figure S7. Diurnal patterns of selected OA markers represented by the normalized mean values of IS-corrected response (see Text S2 for normalization method). The species are grouped based on the similarity in patterns.

Figure S8. Same as Figure S7, but showing normalized median values (see Text S2 for normalization method).

Text S2. Normalization of diurnal profiles

The analysis of diurnal profiles focuses on the variations rather than magnitudes of the species. To accommodate the species with same patterns but different orders of magnitude in a same figure, we adopt a linear normalization approach to process the data, according to the formula:

$$\text{Normalized value}_t = \frac{x_t - \min}{\max - \min}$$

where x_t is the mean (or median) value of the species in hour t ; and \min and \max are the minimum and maximum of all the mean (or median) values throughout the day, respectively.

For details, please refer to Text S2 and the updated Figures S7-S8.

19. Figure S6. There is inconsistency between the graphs and the legend.

Thanks for pointing out this mistake. The figure is updated.

20. Table S2. Verify the names of the compounds.

Verified. Thanks.

References:

Kreisberg, N. M, Hering, S. V., Williams, B. J., Worton, D. R., and Goldstein, A. H.: Quantitation of hourly speciated organic compounds in atmospheric aerosols, Measured by in-situ thermal desorption aerosol gas chromatography (TAG), Aerosol Sci. Technol., 43(1), 38–52, 2009.

Williams, B. J., Goldstein, A. H., Millet, D. B., Holzinger, R., Kreisberg, N. M., Hering, S. V., White, A. B., Worsnop, D. R., Allan, J. D., and Jimenez, J. L.: Chemical speciation of organic aerosol during the International Consortium for Atmospheric Research on Transport and Transformation 2004: Results from in situ measurements, J. Geophys. Res., 112, D10S26, doi:10.1029/2006JD007601, 2007.

Williams, B. J., Goldstein, A. H., Kreisberg, N. M., Hering, S. V., Worsnop, D. R., Ulbrich, I. M., Docherty, K. S., and Jimenez, J. L.: Major components of atmospheric organic aerosol in southern California as determined by hourly measurements of source marker compounds, Atmos. Chem. Phys., 10, 11577–11603, <https://doi.org/10.5194/acp-10-11577-2010>, 2010.