acp-2020-1009: Seasonal analysis of submicron aerosol in Old Delhi using high resolution aerosol mass spectrometry: Chemical characterisation, source apportionment and new marker identification

Response to the reviewers' comments

We thank the reviewer for providing detailed and helpful comments on our manuscript. Below we respond to each reviewer comment in turn (reviewers' comments are shown in italics), indicating the changes we have made.

Response to Anonymous Referee #1

The article presents data and findings that are of interest to the scientific community and in a location where air pollution is very high and affects a large number of people. The methods used are sound. The interpretation of the data and the conclusions are well rooted in the data with minimal speculation. The presentation of the data and the results are good, although it could use some more clarity especially in the figures as mentioned below in more detail. The location where the measurements were carried out presents a number of challenges such as high temperatures and high relative humidity that can be really tricky for instruments. I think the authors have generated a great dataset in those challenging conditions. The authors also did an excellent job at interpreting such a complex mixture of sources. The manuscript is of high quality and within the scope of the journal. I recommend the publication after minor revisions.

We thank the reviewer for their very supportive comments on the merits of our study and its presentation.

Detailed comments:

1- The manuscript similarly to many manuscripts based on AMS data and PMF, makes extensive use of acronyms. These acronyms are probably very familiar to the authors and to experienced AMS users, however, they tend to be hard to follow for readers less involved with AMS and PMF analysis. I recommend making a list of acronyms to help the reader follow the text.

We agree and thank the reviewer for this suggestion. A list of acronyms has been added in Table S1.

Table S1 – List of abbreviations

Abbreviations

IGDTUW	***Indira Gandhi Delhi Technical University for Women***
HR-ToF-AMS	***High-resolution time-of-flight aerosol mass spectrometer***
PTR-QiTOF	***High-resolution proton transfer reaction mass spectrometer***
PM_1	***Sub-micron particulate matter***
SOA	***Secondary organic aerosol***
VOC	***Volatile organic compound***
OrgNO	***Organic nitrogen oxide species***
BC	***Black carbon***
LWC	***Liquid water content***
PCDDs	***Polychlorinated dibenzodioxins***
PCDFs	***Polychlorinated dibenzofurans***
$\overline{OS_C}$	***Carbon oxidation state***
PMF	***Positive Matrix Factorisation***
COA	***Cooking organic aerosol***
NHOA	***Nitrogen-containing hydrocarbon-like organic aerosol***
SFOA	***Solid fuel organic aerosol***
HOA	***Hydrocarbon-like organic aerosol***
SVBBOA	***Semi-volatility biomass burning organic aerosol***
LVOOA	***Low-volatility oxygenated organic aerosol***
SVOOA	***Semi-volatility oxygenated organic aerosol***
PAH	***Polyaromatic hydrocarbons***
UnSubPAH	***Unsubstituted PAH***
MPAH	***Methyl-substituted PAH***
OPAH	***Oxidised PAH***
NOPAH	***Nitrogen-oxygen substituted PAH***
APAH	***Amino PAH***
VK	***Van Krevelen***

2- Figure 1. I suggest adding the time series of the standard AMS species (NO3, SO4, NH4, Chl, and Org) and BC. This will give the reader a good bird's eye view of the dataset, maybe merging it with some version of Figure 5.

Figures 1, 2 and 5 have been combined to produce a new summary figure (Figure 1). Please note that the collection efficiency during the Diwali Period (05 to 14/11/18) has been changed to a lower CE of 0.8 in order to obtain a PM_1 vs. $PM_{2.5}$ gradient (~0.9) that is similar to the other measurement periods. This change resulted from a suggestion made by the second reviewer. The concentration results have therefore slightly changed in many averages and statistics, and all the relevant figures and tables have been updated accordingly.

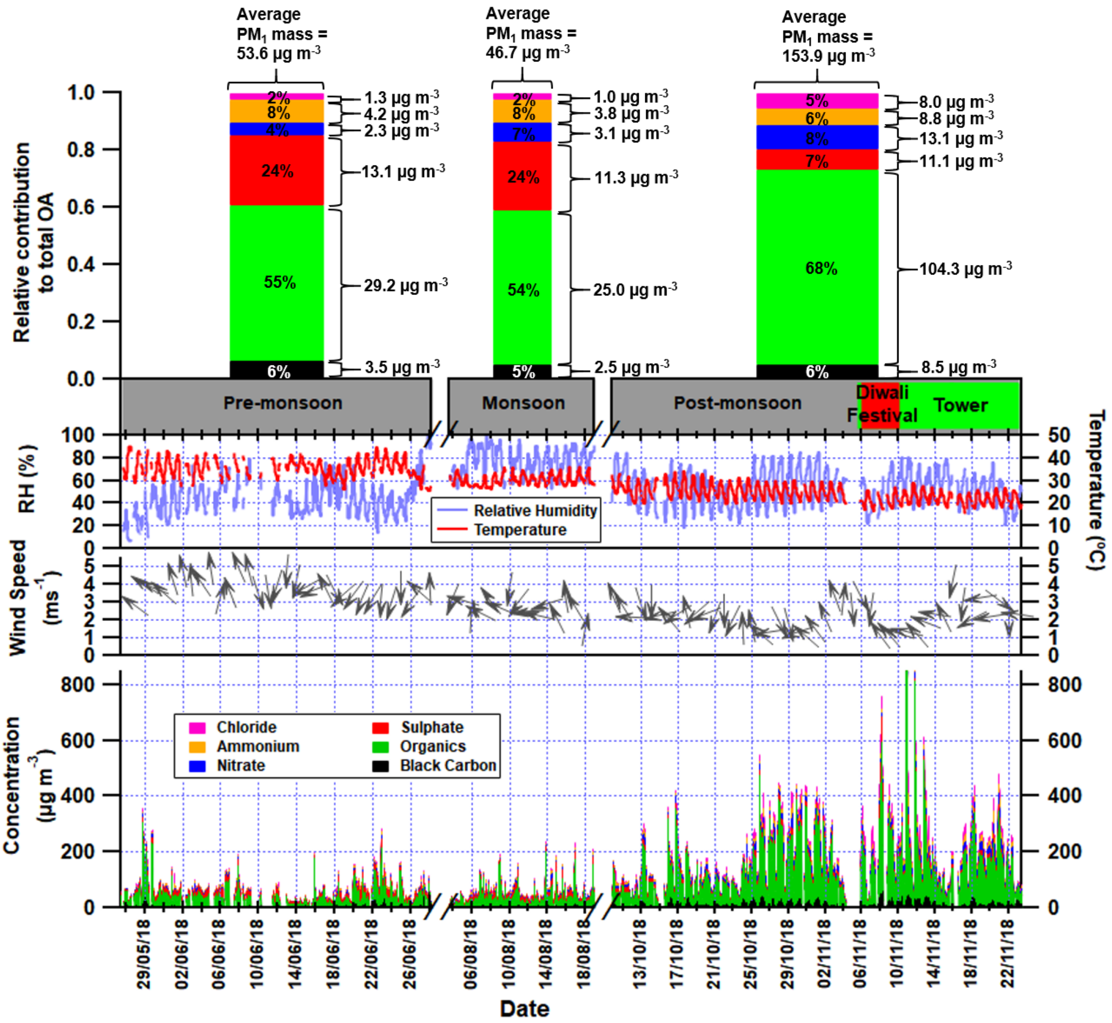

Figure 1 – First Panel: Average relative contributions of chloride, ammonium, nitrate, sulphate, organic aerosol and black carbon to the total PM$_1$ mass loadings in the pre-monsoon, monsoon and post-monsoon periods. The average concentrations of each species are shown to the right of each bar (see Table S5 for values and statistics). Second panel: Gant chart showing the measurement periods where the red region shows the Diwali festival and the green region shows when the inlet was moved to a 30 m tower. Third panel: time series of the relative humidity and the temperature for the three measurement periods. Fourth panel: time series of the wind speed with arrows showing wind direction. Fifth panel: time series of stacked concentrations of aerosol species showing total PM$_1$.

3- Figure 3. I suggest adding dark and light hours with a shaded area for the transition/changing light conditions over the measurement period to help the reader form a picture of the data presented.

Shading for dark and light hours has been added to figures 2 and 8.

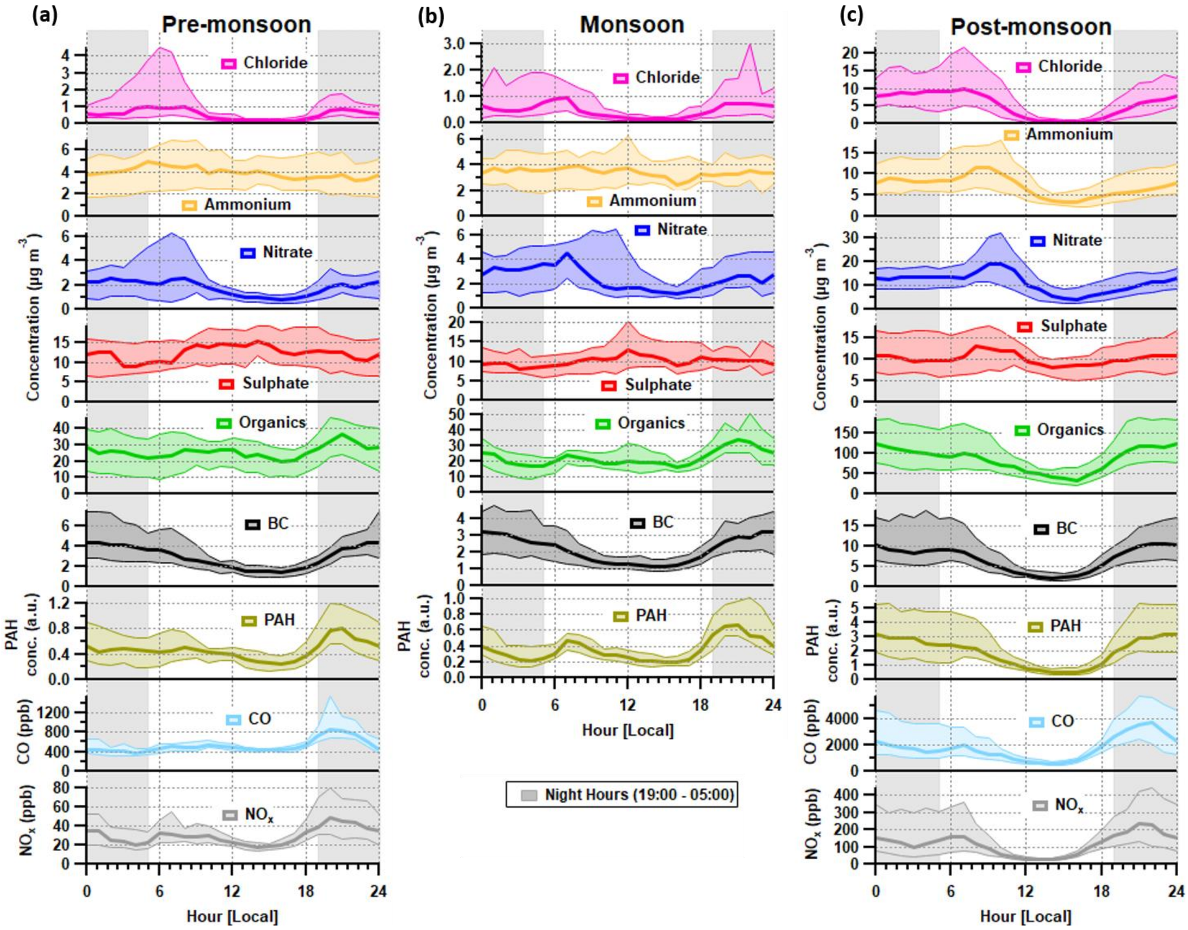

Figure 2 – Median diurnal cycles for aerosol chemical species and for BC, CO and NO_x during the (a) pre-monsoon, (b) monsoon and (c) post-monsoon periods. The median concentration is represented by the thick line and the interquartile range is represented by the shading. Regions shaded in grey are night hours. Data for CO and NO_x are not available for the monsoon period.

4- A lot of figures have tiny labels that are really hard to read especially once the manuscript is printed. In Figure 4 the percentage numbers in each panel are very small I suggest using fewer vales and a larger font. Also, "mean" and "calm" (panels ae), as well as the Wind Speed values in panel (f) are almost not readable in the printed version.

The wind and pollution roses (now presented in Figures 3 and 9) have been increased in size and altered to show wind vectors that are sized based on the contribution to the mean (rather than sized based on frequency of counts). We hope this creates a clearer picture for observing the wind directional relationships within the data. The font size of all labels have also been increased.

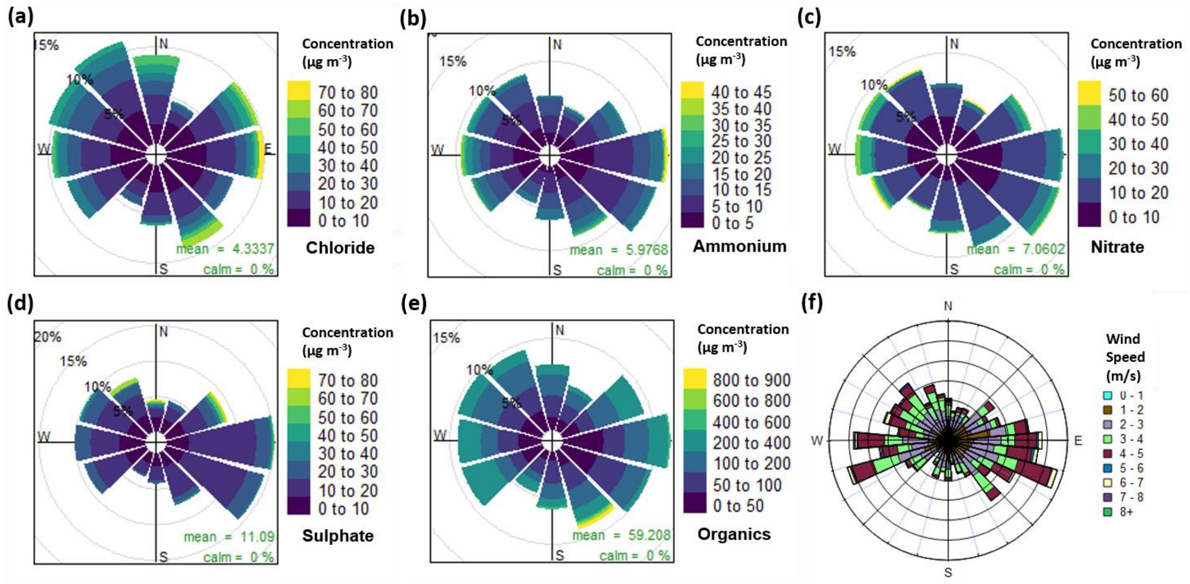

Figure 3 – Pollution roses for (a) chloride, (b) ammonium, (c) nitrate, (d) sulphate and (e) organic aerosol, along with (f) a wind rose plot for all measurement periods combined. The pollution roses show 30° wind vectors and their size is proportional to the percentage contribution to the mean concentration. The vectors are divided into concentration bins based on the colour scale in the legend.

5- Figure 5. I suggest merging it with figure 1 as mentioned in comment 2

This has been done – see previous response.

6- Figure 6. Dates are very small, please increase the font.

The font size has been increased on this figure (now Figure 4).

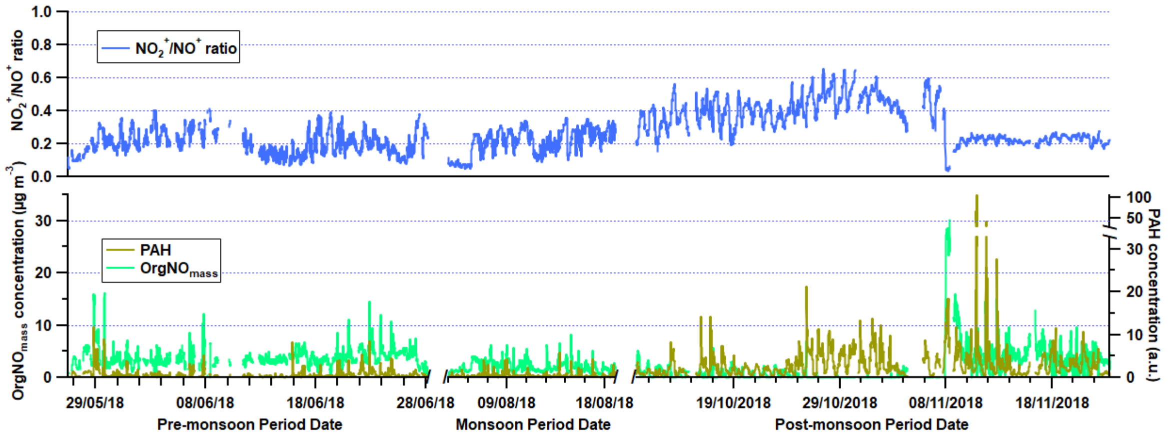

Figure 4 – Upper panel: time series of NO_2^+/NO^+ ratio in the three measurement periods. Lower panel: time series of polyaromatic hydrocarbon (PAH) uncalibrated concentrations and organic nitrogen oxide species ($OrgNO_{mass}$) concentrations.

7- Figure 7 the y axis could be harmonized. It's ok to keep a different scale but I suggest keeping the same number of ticks.

The y-axis has been harmonized on this figure (now Figure 5).

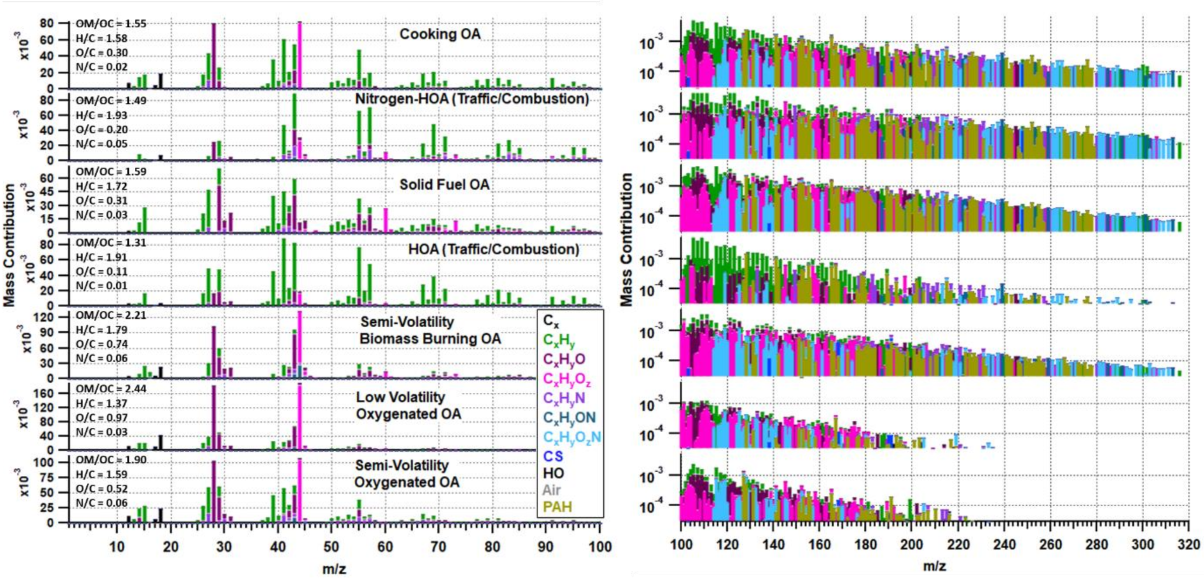

Figure 5 – Organic-only PMF solution with elemental ratios shown for each factor in the left-hand corner of each spectrum. The mass spectra on the left show *m/z* 12-100 on a linear scale, while the spectra to the right show *m/z* 100-320 on a logarithmic scale. The peaks are coloured based on the chemical families shown in the legend.

8- Figure 10. the y axis labels for the O:C, H:C and N:C ratios panel are too small. Please reduce the number of ticks, decide how many to put there (3?) and maybe increase the font.

All axes labels have been increased in font size and the number of ticks have been reduced on this figure (now Figure 8).

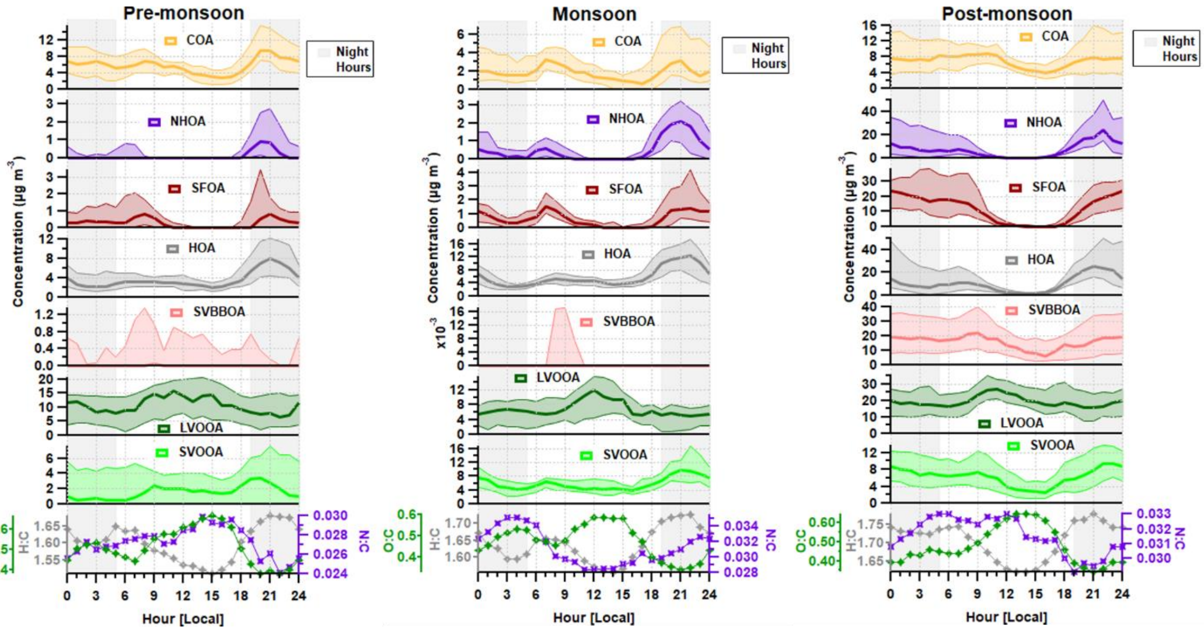

Figure 8 – Median diurnal cycles of the factor solutions for the three measurement periods (interquartile range indicated by the shading) along with the elemental ratios. Regions shaded in grey are night hours.

9- Figure 11. “mean” and “calm” almost not readable

The font size has been increased on this figure (now Figure 9).

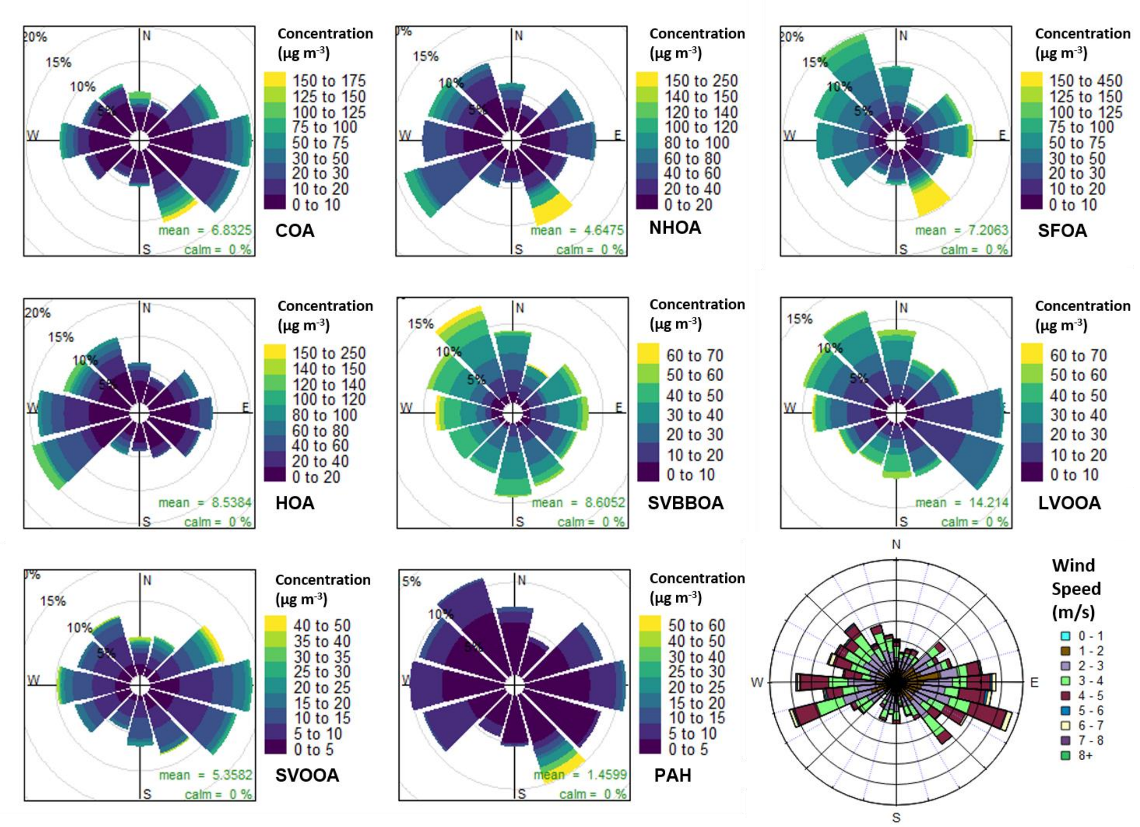

Figure 9 - Pollution roses for each factor and uncalibrated PAH concentrations along with a wind rose. The pollution roses show 30º wind vectors and their size is proportional to the percentage contribution to the mean concentration. The vectors are divided into concentration bins based on the colour scale in the legend.

10- Figure 12. Panel (a) the numbers of ticks could be harmonized by making it the same (4?).

Ticks are now harmonized. This figure has also been moved to the SI as, on further consideration, we feel it isn’t crucial to the main storyline (now Figure S13).

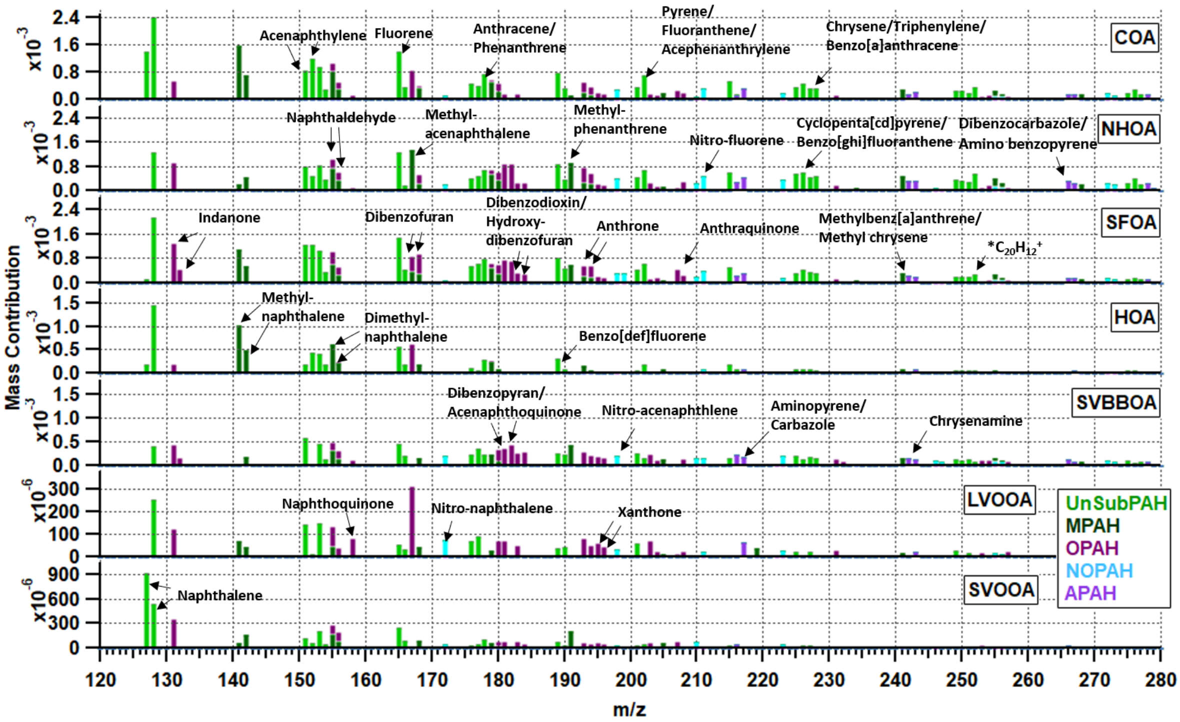

Figure S13 – PAH factor mass spectra showing The PAH families: UnSubPAH, MPAH, OPAH, NOPAH and APAH. *The peak at m/z 252 relating to the ion $[C_{20}H_{12}]^+$ is a list of PAHs overlapping in mass and includes benzo[b]-, benzo[j]- and benzo[k]flouranthene, along with benzo[a]- and benzo[e]pyrene.

11- Figure 14. I recommend increasing the resolution for the top panel (VK diagrams) the dots are lost even in the electronic version if zoomed in.

Attempts to increase the resolution weren’t successful so we increased the marker size instead to show the data more clearly in this figure (now Figure 11).

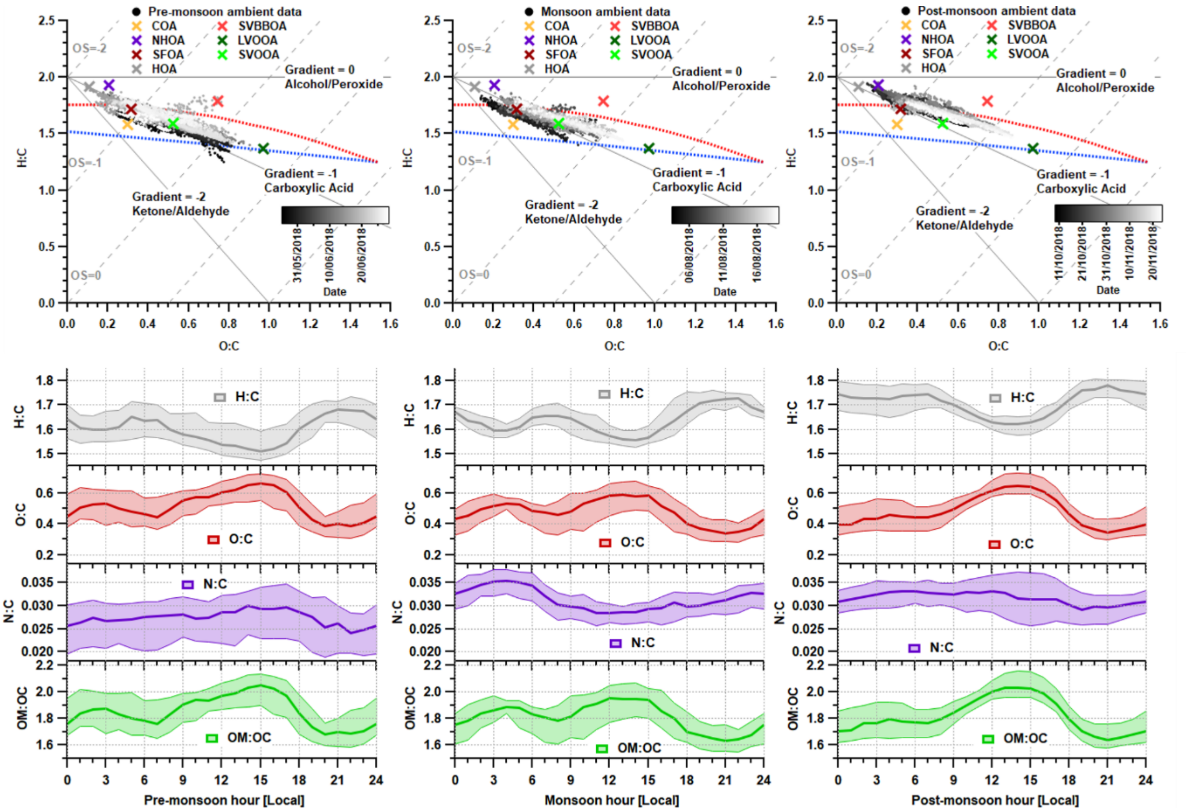

Figure 11 – Van Krevelen (VK) diagrams and median diurnal cycles for elemental ratios during each measurement period. Each VK diagram contains the H:C vs O:C data for the PMF solutions and the raw AMS measurements which are coloured based on the time of the measurement. The carbon oxidation states ($\overline{\mathrm{OS}}_C \approx 2O/C - H/C$) are shown using grey dashed lines and the functional group gradients are shown using solid grey lines. The blue and red dashed lines demarcate the region where published ambient OOA measurements are commonly found (Ng et al., 2011).

12- In the abstract, I recommend adding a mention that sulfate is the largest mass fraction for the pre-monsoon and monsoon periods.

Additional text has been added to the abstract which now reads as follows:

"…Old Delhi is one of the most polluted locations in the world, and PM_1 concentrations reached ~750 µg m^{-3} during the most polluted period, the post-monsoon, where PM_1 increased by 178% over the pre-monsoon period. Sulphate contributes the largest inorganic PM_1 mass fraction during the pre-monsoon (24%) and monsoon (24%) periods with nitrate contributing most during the post-monsoon (9%). The organics dominate the mass fraction (54-68%) throughout the three periods and using positive matrix factorisation (PMF) to perform source apportionment analysis of organic mass, two burning-related factors were found to contribute the most (35%) to the post-monsoon increase…"

In the methods section

13- at lines 129-131 the Authors mention that they calibrated the AMS "throughout the campaign". I recommend adding a sentence explaining how many times and when (e.g., before, middle, and after?).

Details on the timings of the calibrations have been added: "The HR-TOF-AMS was calibrated fortnightly over the three campaigns (11 calibrations in total) for its ionisation efficiency of nitrate (IE) and the relative ionisation efficiency (RIE) of other inorganic compounds using nebulised 300 nm ammonium nitrate, sulphate and chloride." A summary table has also been added containing the calibration results (Table S2).

Table S2 – Relative ionisation efficiencies (RIE), ionisation efficiencies (IE) and collection efficiencies (CE)

Season	IE	RIE NH_4^+	RIE SO_4^{2-}	RIE Cl^-	CE
Pre-monsoon	2.92E-07	4	1.45	2.07	0.5
Monsoon	2.92E-07	4	1.45	2.07	0.5
Post-monsoon preflux period (11/10/18 - 03/11/18)	2.89E-07	4	1.45	2.07	0.5
Post-monsoon Diwali period (05/11/18 - 14/11/18)	3.14E-07	4	1.45	1.05	0.8
Post-monsoon post-Diwali (14/11/18 - 23/11/18)	3.14E-07	4	1.45	1.05	0.5

14- At lines130 to 135 the Authors mention that in their analysis they had to use different CEs to match the "PM2.5 filter measurements". I recommend expanding this sentence explaining which measurements they are referring to, carried out by which group, with which instrument, and at what time resolution.

Details have been added to the measurements we refer to: "A collection efficiency (CE) of 0.5 was confirmed through comparisons with gravimetric $PM_{2.5}$ filter measurements taken throughout the pre- and post-monsoon campaigns by Birmingham University. The AMS measurements were added to the BC measurements to give total PM_1 which was averaged to match the sample intervals of the filters (6 and 12 h)." More details on this is being published in the companion paper to this study which includes filter measurement comparison graphs (Reyes-Villegas et al., 2020).

15- At lines 146 -148 the sentence "Therefore, only peaks which significantly improved the open and closed signal residuals were fitted regardless of the residuals in the difference (diff = open – closed) signal." is unclear and leaves the reader wonder which peaks were not included. I understand that the fitting at higher m/z is tricky, but I am wondering if the authors can modify or expand on the sentence to clarify the process to the reader, maybe explaining which peaks were not fitted and why.

We have expanded the sentence in the text to inform the reader of the fitting procedure and to ensure the explanation is more understandable to the reader: "The PIKA software allows the user to fit peaks based on a reduction in the residuals between the measured signal and the fit. There are measured signals for open (when the chopper is open), closed (when the chopper is closed) and diff (= open – closed), meaning there are also three sets of residuals. Neighbouring peaks may overlap and cause the diff residuals to improve whilst not improving the open and closed residuals. This becomes more relevant when moving to higher *m/z* ions as peaks become broader. Therefore, only peaks which significantly improved the open and closed signal residuals were fitted."

16- Lines 163-165: ". . . black carbon (BC) measurements which were taken using an Aethalometer AE-31 and corrected for by a Single Particle Soot Photometer (SP-2; Droplet Measurement Technology, Boulder, CO) (Reyes-Villegas et al., 2020).". This sentence is quite vague. I understand that there is a reference to look up, however, I recommend adding a short sentence giving a few more details, e.g., explaining briefly 1) how the Aethalometer data were corrected 2) if/when and for how long the SP2 was co-located with the Aethalometer.

The correction method is now included along with the number of days the SP2 was co-located with the Aethalometer: "The AMS measurements were compared to a number of co-located instruments including black carbon (BC) measurements which were taken using an Aethalometer AE-31. The BC measurements were corrected using the Weingartner method (Weingartner et al., 2003) and by using reference measurements from a Single Particle Soot Photometer (SP-2; Droplet Measurement Technology, Boulder, CO) which was co-located for ~7-days (Reyes-Villegas et al., 2020).” More details are also presented in the cited reference (Reyes-Villegas et al., 2020).

Results

17- Lines 235-238: here and in a few other parts, the authors cite “personal communication with Ben Langford”. In all those cases I think that this information should be removed as it doesn’t seem critical to the point of the sentences unless a paper has been published in the meantime and can be properly referenced.

We have updated the citation which refers to the abstract of a talk given at the European Aerosol Conference which summarised the results (Di Marco et al., 2019).

18- Line 458: “UnSubPAHs” acronym not defined

A definition has been added: “The PAH composition of COA is mainly unsubstituted PAHs (UnSubPAHs) and also contributes....”

Conclusions

19- Lines 894-895: “These high post-monsoon concentrations have been linked to an increase in burning emissions mainly from crop residue and solid fuel.” Are higher concentrations only due to an increase in burning emissions or the boundary layer height affect these concentrations as well? If that’s the case, I recommend adding it here.

We agree and have added this statement within the quoted sentence: “These high post-monsoon concentrations have been linked to an increase in the boundary layer height affect and an increase in burning emissions, mainly from crop residue and solid fuel.”

Supplementary Information

20- Figure S2 y-axis label too small

All figures relating to multilinear regression analysis have been increased in font size.

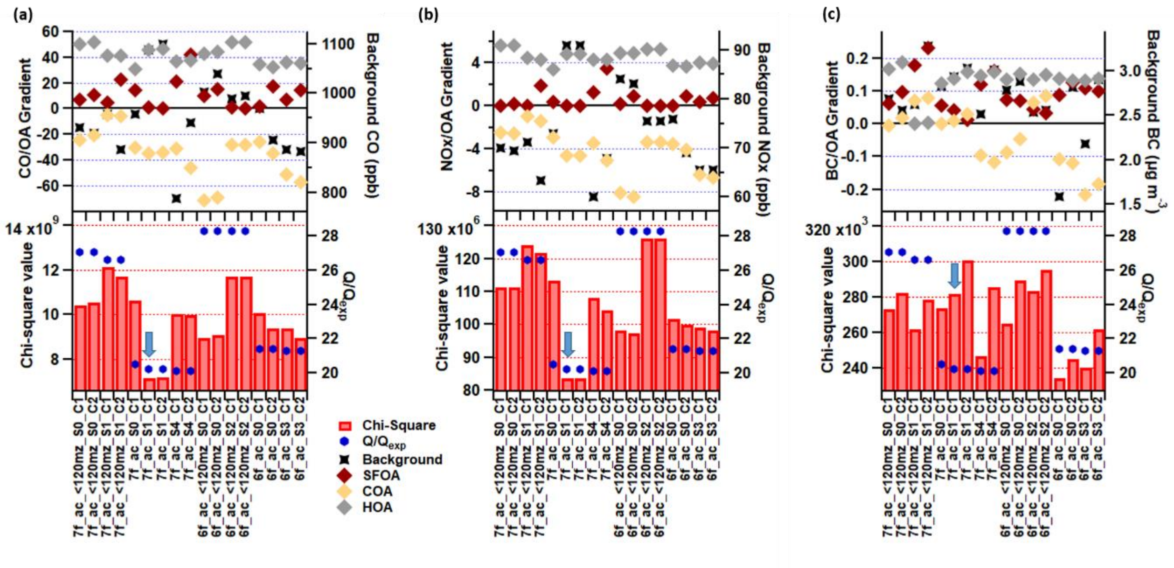

Figure S1 – Trilinear regression analysis results for the PMF solutions taken from the all-periods-combined case. Results are shown for the fit using (a) CO, (b) NO$_x$ and (c) BC as external tracers. Gradient contributions for factors SFOA, COA and HOA are shown alongside the background concentration of the tracer (black) which is estimated using the intercept of the linear regression. The chi-square value (red markers), the Q/Q$_{exp}$ (blue markers) and the chosen final solution (labelled with a blue arrow) are also shown below.

21- Page 12: "Error! Not a valid bookmark self-reference." Should be "Table S2"

This has been corrected.

22- Figure S8: "Polar graphs showing the concentrations . . ." add units of concentrations.

Units of concentration have been added to this caption: "**Polar graphs showing the concentrations (in µg m^{-3}) by wind direction for…**"

23- Figure S9 "Mean" and "calm" not legible.

The Figure S9 has been altered to show wind vectors that are sized based on the contribution to the mean. The font size has also been increased.

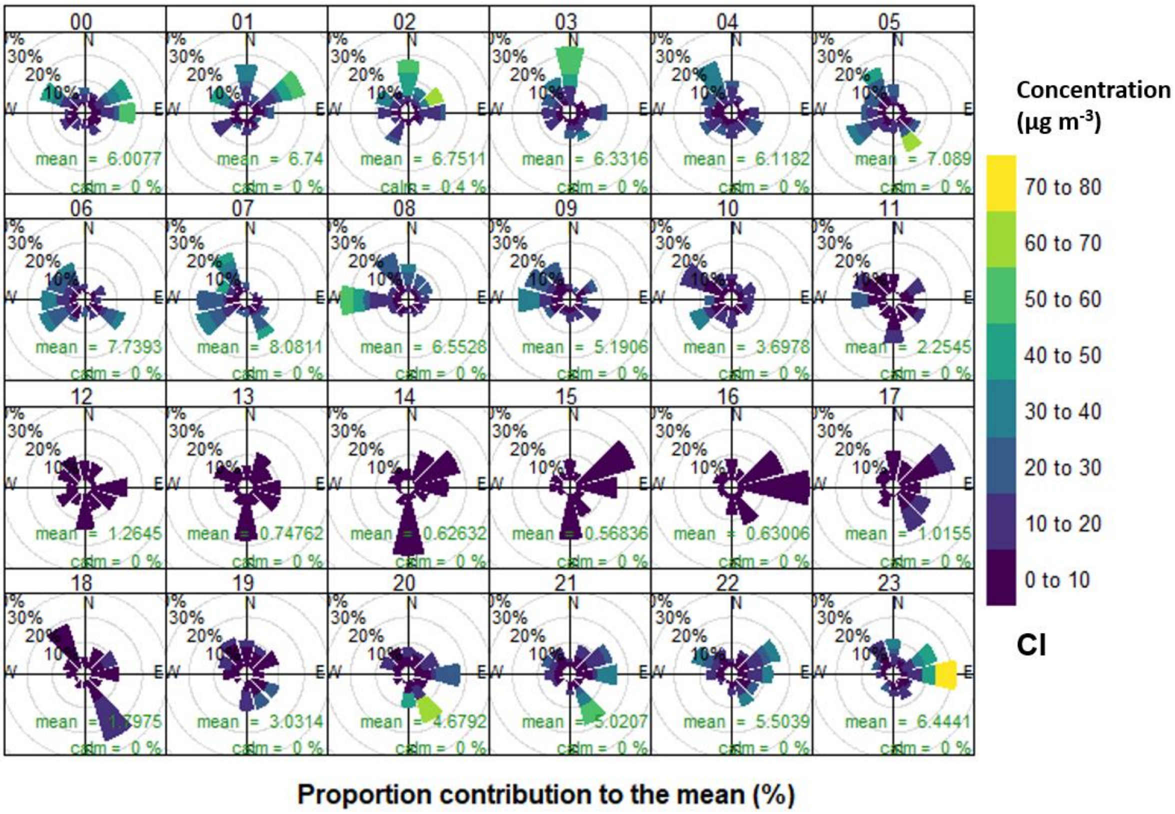

Figure S9 – Chloride pollution roses for each diurnal hour for all measurement periods combined, where the data is binned into 30º wind vectors and the size of each bin is proportional to its contribution to the mean concentration. The counts are divided into concentration bins based on the colour scale in the legend. Units are μg m^{-3}.

24- Figure S13: add that the points not labeled neither "Delhi" nor "Chack2018" come from Table S3.

The caption has been updated to make it clearer that all points included in the Van Krevelen diagram are from Table S6: "**Van Krevelen (VK) diagram for the mass spectra of the organic aerosol factors listed in Table S6. The data labelled with Chak2018 are from the study by Chakraborty et al. (2018) and those labelled with Delhi are values from this study.**" (now Figure S15).

25- Page 20: "The factor mass profiles and their diurnal cycles during each measurement period are summarized in Figure S14". I think it should be "Figure S15"

This has been corrected and Figure S15 is now Figure S17: "The factor mass profiles and their diurnal cycles during each measurement period are summarised in Figure S17"

26- Figure S15 and S16: y-axis labels are too small

All axis labels have been increased in font size (now Figures S17 and S18).

Finally, reading the manuscript I have been wondering why the Authors decided not to run the PMF in bootstrap mode for the 7 solutions combined periods.

We did conduct a bootstrapping analysis. However, the different factors show a high degree of correlation because they share common controls through boundary layer dynamics and temperature. This is a situation in which bootstrapping analysis can often be incorrect or ambiguous (Ulbrich, 2011).

It may be possible to solve this issue using a different resampling method. The PET tool used in this study resamples the data matrix using subsets of original rows (mass spectra) and randomly replaces them with other rows from the original matrix (Ulbrich, 2011). Other software, such as the EPA PMF software, use block bootstrapping which is likely a better resampling method for this dataset. It attempts to account for the effects of serial correlation (or correlation between factors) by resampling via randomly selecting non-overlapping time periods or 'blocks' (i.e. of length 3-days) and creating a new dataset from these blocks (Paatero et al., 2014). This however does not account for large variations between the pre-monsoon and the post-monsoon (e.g. SVBBOA goes from ~0 to ~20 μg m^{-3}) because, for example, if a 'block' from the pre-monsoon is placed next to a block from the post-monsoon, this creates a large amount of variation that is not due to model error/solution instability. This has been noted before in previous studies that have implemented other resampling methods to counteract this (Hemann et al., 2009).

In order to account for seasonal variance, we could address this issue by doing bootstrapping on a seasonal basis, but such a detailed analysis is beyond the scope of this study. We would also be required to resample on a short time scale which is computationally demanding for a matrix with 2.39673×10^7 data points and is currently not supported by software such as EPA PMF. The additional issue is the time series dependence on the boundary layer which is likely not going to be accounted for when using current resampling methods or bootstrapping on a seasonal basis.

However, the variance can be explored by varying the SEED and these results show variance in the solutions. To deal with the problem of potential ambiguity, we applied multilinear regression analysis on multiple SEEDS as a more appropriate method for obtaining a solution in this scenario, which is an established method to deal with such situations (Allan et al., 2010; Young et al., 2015; Elser et al., 2016; Reyes-Villegas et al., 2016), as we were able to compare against multiple external tracers.

Nevertheless, in response to this comment, we have conducted a further bootstrapping analysis (100 iterations) on four selected possible all-combined solutions (6f_ac_S0_C1, 6f_ac_S3_C1, 7f_ac_S1_C1 and 7f_ac_S4_C1) using the PET tool and this gave favourable results in the multilinear regression analysis. We found that the chosen solution (7f_ac_S1_C1) gave a 5% decrease in the relative standard deviation of profiles and 7% decrease in relative standard deviation of the time series when compared to the nearest possible solution (7f_ac_S4_C1). However, it must be noted that these estimates could be incorrect or ambiguous for the reasons mentioned above.

References:

Allan, J. D., Williams, P. I., Morgan, W. T., Martin, C. L., Flynn, M. J., Lee, J., Nemitz, E., Phillips, G. J., Gallagher, M. W. and Coe, H.: Contributions from transport, solid fuel burning and cooking to primary organic aerosols in two UK cities, Atmos. Chem. Phys., 10(2), 647–668, doi:10.5194/acp-10-647-2010, 2010.

Chakraborty, A., Mandariya, A. K., Chakraborti, R., Gupta, T. and Tripathi, S. N.: Realtime chemical characterization of post monsoon organic aerosols in a polluted urban city: Sources, composition, and comparison with other seasons, Environ. Pollut., 232, 310–321, doi:10.1016/j.envpol.2017.09.079, 2018.

Elser, M., Huang, R. J., Wolf, R., Slowik, J. G., Wang, Q., Canonaco, F., Li, G., Bozzetti, C., Daellenbach, K. R., Huang, Y., Zhang, R., Li, Z., Cao, J., Baltensperger, U., El-Haddad, I. and André, P.: New insights into PM2.5 chemical composition and sources in two major cities in China during extreme haze events using aerosol mass spectrometry, Atmos. Chem. Phys., 16(5), 3207–3225, doi:10.5194/acp-16-3207-2016, 2016.

Hemann, J. G., Brinkman, G. L., Dutton, S. J., Hannigan, M. P., Milford, J. B. and Miller, S. L.: Assessing positive matrix factorization model fit: A new method to estimate uncertainty and bias in factor contributions at the measurement time scale, Atmos. Chem. Phys., 9(2), 497–513, doi:10.5194/acp-9-497-2009, 2009.

Di Marco, C. D., Langford, B., Cash, J. M., Mullinger, N., Helfter, C. and Nemitz, E.: Source apportionment analysis applied to aerosol eddy-covariance fluxes in Delhi, European Aerosol Conference. [online] Available from: https://www.costcolossal.eu/specialsessioneac2019/, 2019.

Ng, N. L., Canagaratna, M. R., Jimenez, J. L., Chhabra, P. S., Seinfeld, J. H. and Worsnop, D. R.: Changes in organic aerosol composition with aging inferred from aerosol mass spectra, Atmos. Chem. Phys., 11(13), 6465–6474, doi:10.5194/acp-11-6465-2011, 2011.

Paatero, P., Eberly, S., Brown, S. G. and Norris, G. A.: Methods for estimating uncertainty in factor analytic solutions, Atmos. Meas. Tech., 7(3), 781–797, doi:10.5194/amt-7-781-2014, 2014.

Reyes-Villegas, E., Green, D. C., Priestman, M., Canonaco, F., Coe, H., Prévôt, A. S. H. and Allan, J. D.: Organic aerosol source apportionment in London 2013 with ME-2: Exploring the solution space with annual and seasonal analysis, Atmos. Chem. Phys., 16(24), 15545–15559, doi:10.5194/acp-16-15545-2016, 2016.

Reyes-Villegas, E., Panda, U., Darbyshire, E., Cash, J. M., Joshi, R., Langford, B., Di Marco, C. F., Mullinger, N., Acton, W. J. F., Drysdale, W., Nemitz, E., Flynn, M., Voliotis, A., McFiggans, G., Coe, H., Lee, J., Hewitt, C. N., Heal, M. R., Gunthe, S. S., Shivani, Gadi, R., Singh, S., Soni, V. and Allan, J. D.: PM1 composition and source apportionment at two sites in Delhi, India across multiple seasons, Atmos. Chem. Phys. Discuss., 2020, 1–19, doi:10.5194/acp-2020-894, 2020.

Ulbrich, I. M.: Characterization of Positive Matrix Factorization Methods and Their Application to Ambient Aerosol Mass Spectra, Chem. Biochem. Grad. Theses Diss., Paper 35 [online] Available from: https://core.ac.uk/download/pdf/54848806.pdf, 2011.

Weingartner, E., Saathoff, H., Schnaiter, M., Streit, N., Bitnar, B. and Baltensperger, U.: Absorption of light by soot particles: determination of the absorption coefficient by means of aethalometers, J. Aerosol Sci., 34(10), 1445–1463, doi:10.1016/S0021-8502(03)00359-8, 2003.

Young, D. E., Allan, J. D., Williams, P. I., Green, D. C., Harrison, R. M., Yin, J., Flynn, M. J., Gallagher, M. W. and Coe, H.: Investigating a two-component model of solid fuel organic aerosol in London: Processes, PM1 contributions, and seasonality, Atmos. Chem. Phys., 15(5), 2429–2443, doi:10.5194/acp-15-2429-2015, 2015.