Author response to referee comments

We thank the referee for the valuable comments which have greatly helped us improve the manuscript. Please find below our responses (in blue) after the referee comments (in black). The changes in the revised manuscript are written in italic.

Anonymous Referee #1

This article reports the first time the binPMF algorithm has been applied to VOCUS PTR-MS data, as applied to forested environments. There would be a strong interest in this type of broad-base work to try to generalize biogenic emissions, as these can have a profound effect on atmospheric chemistry. While I would say this certainly fits thematically within ACP's scope, right now the paper currently feels unfinished as a research article because while it demonstrates the instrument and algorithm 'working', it currently fails to identify what new understanding this confers to atmospheric science, beyond a running commentary of the authors' interpretations of the factors. I therefore recommend that this paper be published after major revisions. This could take the form of either a research article that is more focused on the atmospheric science arising from the work, or a technical note that explores the technicalities in more detail (I have queries regarding the methodology, see below). While it could in theory present a 'measurement report' based on this work, I feel that this may not be in the spirit of what the authors intended.

Response: We thank Anonymous Referee #1 for the careful review and inputs which helped improving the overall quality of our work. We agree that as a research article, new understanding towards atmospheric science from this work should be highlighted. Therefore, for the "4 Results and discussion" part, we revised the structure of the section to highlight our major scientific findings and add further analysis to gain insights into the atmospheric processes of monoterpenes and sesquiterpenes.

Compared to preexisting studies, this study performed binPMF analysis on Vocus PTR-TOF data and identified both primary emission sources and secondary oxidation processes of atmospheric organic vapors in two forested environments. For the first time, organic precursors, the lightly oxidized products, and the more oxidized products were separated as individual PMF factors. The relative abundances of these factors can be utilized by modelers to evaluate simulation output, improve model performance, and provide new perspectives to understand gas-phase physicochemical processes. Based on the interpretation of the results relating to oxidation processes, further insights were gained regarding monoterpene and sesquiterpene reactions. For example, a strong relative humidity (RH)-dependence was found for the behavior of sesquiterpene lightly oxidized compounds. High concentrations of these compounds only occur at high RH, yet similar behavior was not observed for monoterpene oxidation products. These findings highlight the need for further studies to delve into gas-phase atmospheric processes of monoterpenes and sesquiterpenes.

More details can be found below as response to the referee's first comment.

Comments:

The manuscript currently presents the results very systematically, but it is difficult to see what the reader is supposed to get from these. The authors provide a commentary on their interpretation of the factors, but I am not sure I learned anything new or significant about atmospheric chemistry on reading these. If this is to be presented as a research article, the paper needs to be refocused towards the new scientific insight or a testable hypothesis.

Response: We agree with the reviewer that as a research article, the new scientific insight or hypothesis should be highlighted to make it clear to the readers. Therefore, the structure of the section “4 Results and discussion” was revised as follows:

“4 Results and discussion

- 4.1 Choice of PMF solution and factor interpretation
- 4.2 Source identification in the Landes forest
- 4.3 Source identification in the southern Finnish boreal forest
- 4.4 Comparison among different factors
- 4.5 Comparison between the two forests
- 4.6 Insights into terpene oxidation processes
  - 4.6.1 Monoterpene oxidations
  - 4.6.2 Sesquiterpene oxidations”

In Section 4.4., the identified factors were compared with each other. Based on the similar temporal behavior of Factor L3 (C_6 and C_7 lightly oxidized products) and Factor L7 (C_{13} lightly oxidized products) and our current knowledge of the corresponding compounds, the C_{13} oxidized compounds are speculated to be produced through the dimer formation mechanisms of C_6 and C_7 species. The time series of monoterpene lightly oxidized products and sesquiterpene lightly oxidized products do not follow very well with each other, suggesting probably different atmospheric processes. This is further investigated in Section 4.6.

In Section 4.5, spatial comparison between the two forests were discussed regarding the relative abundances of different identified factors. For the common sources identified in both forests, they show similar mass profiles, indicating that the sources and processes are indeed similar despite the quite different regions the forests are in.

In Section 4.6, based on the separation of terpene oxidation processes with varying oxidation degrees, further insights were gained regarding monoterpene and sesquiterpene oxidations (Figure 14, Figure 15, Figure S16, and Figure S17). A strong relative humidity (RH)-dependence was found for the sesquiterpene lightly oxidized compounds, as well as the correlation between them and the products of [OH] × [sesquiterpenes] or $[O_3]$ × [sesquiterpenes]. However, these RH dependences were not observed for monoterpene lightly oxidized compounds.

Overall, for the first time, the source identification of atmospheric organic vapors measured by Vocus PTR-TOF separated both primary emission sources and secondary oxidation processes with varying oxidation degrees. The relative abundances of organic precursors, the lightly oxidized products, and the more oxidized products can be utilized by modelers to evaluate simulation output, improve model performance, and provide new perspectives to understand gas-phase physicochemical processes. Based on further investigation of monoterpene and sesquiterpene reactions in the atmosphere, a strong RH-dependence was found for the behaviour of sesquiterpene lightly oxidized products but not for that of monoterpene lightly oxidized products.

The corresponding changes can be seen in the revised manuscript as follows:

4.4 Comparison among different factors

The monoterpene factor and sesquiterpene factor correlate very well with each other at both sites (Fig. 11; r^2 *= 0.69 in the Landes forest and* r^2 *= 0.59 at the SMEAR II station). The emissions of monoterpenes and sesquiterpenes are both strongly influenced by temperature. Their signals peak at night with the effect of the shallow boundary layer. In the daytime, the low signals of the monoterpene and sesquiterpene factors are likely a combination of enhanced atmospheric mixing after sunrise and the rapid photochemical consumption of monoterpenes and sesquiterpenes. The signal of monoterpene factor is around 15 times higher than that of sesquiterpene factor at the SMEAR II station while it is around 60 times in the Landes forest. Previous studies found that sesquiterpene emissions from pines, spruces, and birches under normal conditions were 5-15% of total monoterpene emissions by mass (Rinne et al., 2009 and references therein), in line with our observations.*

In the Landes forest, a factor of C_6 *and* C_7 *lightly oxidized products (Factor L3) was resolved in the low mass range and a factor representative of* C_{13} *lightly oxidized products (Factor L7) was identified in the high mass range. Interestingly, these two factors show a close correlation with each other (*r^2 *= 0.64). The* C_6 *oxygenated compounds have been observed during the oxidation processes of benzene and* C_7 *oxygenated compounds from toluene oxidations (Sato et al., 2012; Zaytsev et al., 2019). These compounds can also be directly emitted from biogenic or anthropogenic sources (Conley et al., 2005; Pandya et al., 2006; Rantala et al., 2015). The temporal behaviour of Factor L7 is similar to that of Factor L3, indicating potentially similar formation pathways of these lightly oxygenated compounds. Therefore, the* C_{13} *oxidized compounds are speculated to be produced through the dimer formation mechanisms of* C_6 *and* C_7 *species (Valiev et al., 2019). In addition,* $C_{13}H_{20}O_3$ *can be direct emissions of methyl jasmonate (Meja), which is a typical green leaf volatile used in plant-plant communications for defensive purposes (Cheong and Choi, 2003). But considering the close correlation between Factor L3 and Factor L7, we conclude that these C13 lightly oxidized compounds are formed from atmospheric oxidation processes, not direct plant emissions.*

Monoterpene lightly oxidized products and sesquiterpene lightly oxidized products were resolved as individual factors at both sites (Factor L7 vs. Factor L10 in the Landes forest and Factor S5 vs. Factor S6 at the SMEAR II station). While the diurnal variations of monoterpene lightly oxidized products are similar to those of sesquiterpene lightly oxidized products, their time series do not follow very well with each other, suggesting the probably different formation pathways or different factors influencing the atmospheric processes of monoterpenes and sesquiterpenes. More discussions can be found in Sect. 4.6.

In this study, the source apportionment analysis was performed separately on two subranges of the mass spectra. It can happen that the same factor is identified in both subranges. For example, both Factor L2 and Factor L9 are defined as the plume event during the measurements. The time series of Factor L2 and Factor L9 show a high correlation coefficient of 0.93 and correlate tightly with aromatic compounds, indicating the major influence of anthropogenic sources. As mentioned above, the air masses in the Landes forest were relatively stable during our observations with wind speed below canopy < 1 m s^{-1}*. Therefore, the influence of long-range regional transport on the atmosphere in the forest is expected to be minor. We speculate that the plume event is a result of local anthropogenic disturbances favored by the lower boundary layer height at night.*

4.5 Comparison between the two forests

To give an overview of the source distributions in the two forest ecosystems, we calculated the mass fraction of each factor based on their average signal intensities. We acknowledge that it is not a perfect method to quantify the contributions of various sources and formation processes. The sensitivities of different VOCs measured by the PTR instruments may vary by a factor of 2-3 (Sekimoto et al., 2017; Yuan et al., 2017). The uncertainties can come from the challenge to convert the signal intensity to atmospheric concentrations because of problematic calibrations, especially given that many unknown molecules exist in the mass spectra. The major bins at m/z 81 Th and 137 Th, which were initially excluded to perform PMF analysis, were counted into their corresponding *factors. For example, the signals of the discarded bins at m/z 81 Th and 137 Th were estimated by multiplying their isotope signals by the corresponding scale number and added to the factor representing monoterpenes. The average mass fractions of various PMF factors in total measured organic vapors are shown in Fig. 12.*

While the atmospheric environment and ecosystem processes differ markedly in the Landes forest and the southern Finnish boreal forest, the results of this study reveal similar biogenic sources and oxidation processes in these forest environments. For instance, the biogenic VOCs at the two sites are both dominated by monoterpenes, with the average fractions of 29% in the Landes forest and at the SMEAR II station. These two forests are both characterized by pine trees, with dominant emissions of α-pinene and β-pinene (Riba et al., 1987; Simon et al., 1994; Hellén et al., 2018). According to the PMF results, isoprene and its major oxidation products in these environments (mainly C_4H_6O) contribute 14% and 21% in the two ecosystems, respectively. Factors indicative of sesquiterpenes are identified in the high mass range at both sites. The average contribution of sesquiterpenes (0.5% in the Landes forest and 1.7% at the SMEAR II station) is much smaller than that of monoterpenes and isoprene. Factors of the lightly oxidized products, more oxidized products, and organic nitrates of monoterpenes/sesquiterpenes in total contribute 8% and 12% of the measured organic vapors in the Landes forest and at the SMEAR II station, respectively.

The factor related to $C_4H_8H^+$ ion was resolved at both sites and contributes 10% in the Landes forest and 16% at the SMEAR II station. According to the discussions by Li et al. (2020), the observation of $C_4H_8H^+$ in the Landes forest can be attributed to several sources. For instance, the protonated butene may contribute to the $C_4H_8H^+$ signal, which is emitted by biogenic or anthropogenic sources (Hellén et al., 2006; Zhu et al., 2017). Another possible explanation is that the $C_4H_8H^+$ ion is produced during the fragmentation of many VOCs in the PTR instruments (Pagonis et al., 2019). The green leaf volatiles (GLV) have been found to fragment at m/z 57 Th inside the PTR instruments, which are a group of six-carbon aldehyde, alcohols and their esters released by plants. Furthermore, butanol can easily lose an OH during the PTR source ionization and produce prominent $C_4H_8H^+$ peaks (Spanel and Smith, 1997). Therefore, the condensation particle counters (CPCs) using butanol for aerosol measurements at the site could also be an important source of $C_4H_8H^+$ ions, although the exhaust air from these instruments has been filtered using charcoal denuder. At the SMEAR II station, the bivariate polar plot where the concentrations of air pollutants are shown as a function of WS and WD indicates that high signals of $C_4H_8H^+$ generally occur when the wind comes from the north (Fig. S15). Located in the north of the measurement container is a particle measurement cottage with several CPCs inside using butanol. A previous study at this station also found that $C_4H_8H^+$ signals detected by PTR-TOF mainly come from butanol used by aerosol instruments (Schallhart et al., 2018). Therefore, it is expected that Factor S1 at the SMEAR II station is mainly contributed by butanol fragmentation inside the instrument where butanol comes from nearby aerosol instruments.

Figure 13 presents the comparison of the mass spectra of the common sources identified at both sites, with the x and y axis showing the mass fraction of different bins in the factor profile. The scattering in the plots is mainly caused by mass bins with much lower mass fractions. However, the dominant bins with high mass contributions in the factor profiles generally correlate well and are located close to the 1:1 line. It shows that the mass spectra of the common sources match well in these two forests and the sources and processes are indeed similar despite the quite different regions the forests are in.

4.6 Insights into terpene oxidation processes

Terpenes undergo varying degrees of oxidations in the atmosphere and produce a large variety of organic compounds with different volatilities (Donahue et al., 2012; Ehn et al., 2014). With the sub-range PMF analysis performed in this study, terpene reaction products with varying oxidation degrees are successfully separated. The sources of monoterpene lightly oxidized products, sesquiterpene lightly oxidized products, monoterpene more oxidized compounds, and monoterpene-derived organic nitrates are identified in both forests with distinct characteristics. These factors account for 8-12% of the measured organic vapors in the two forests. It provides a *great opportunity to gain insights into terpene oxidation processes. Because some environmental parameters, for example, measurements of UVB to estimate OH concentration, are not available in the Landes forest, the results from SMEAR II station are presented as follows.*

4.6.1 Monoterpene oxidations

The oxidation processes of monoterpenes at the SMEAR II station have been investigated by several previous studies, mostly based on the highly oxidized compounds. Utilizing non-negative matrix factorization analysis on iodide-adduct CIMS data at the SMEAR II station, Lee et al. (2018) found that the gas-phase organic species subgroup of $C_{6\text{-}10}H_yO_{\geq 7}$ *showed distinct daytime diel trends. Yan et al. (2016) conducted source apportionment of HOMs at the SMEAR II station and separated various HOM formation pathways, such as monoterpene ozonolysis and monoterpene oxidation initiated by* NO_3 *radical. In this study, three types of monoterpene reaction products were detected: monoterpene lightly oxidized compounds, monoterpene more oxidized compounds, and monoterpene-derived organic nitrates. The latter two were not clearly separated into different factors at the SMEAR II station due to the similarities in their overall time trends. For example, the time series of* $C_{10}H_{15}NO_5H^+$ *correlate well with those of* $C_{10}H_{16}O_4H^+$ *and* $C_{10}H_{16}O_5H^+$ *(*$r^2 > 0.61$*).*

Consistent with previous observations, monoterpene more oxidized products (i.e., $C_{10}H_{16}O_4$ *and* $C_{10}H_{14}O_5$*) have a broad high distribution throughout the day due to the active photochemical processes during daytime. Monoterpene-derived organic nitrates (i.e.,* $C_{10}H_{17}NO_4$*,* $C_{10}H_{15}NO_5$*, and* $C_9H_{13}NO_6$*) are mainly characterized by a distinct morning peak at around 8:00, approximately 2 h after the NO peak. But their intensities are also elevated at night. PMF analysis of* NO_3^- *CIMS dataset observed similar diurnal variations of terpene organic nitrates factor at a forest site in the southeastern US (Massoli et al., 2018). Compared with β-pinene and most other monoterpenes, the overall organic nitrate yield from α-pinene +* NO_3 *is rather low (Fry et al., 2014; Kurtén et al., 2017). Laboratory studies found that using iodide-adduct FIGAERO-HR-ToF CIMS,* $C_{10}H_{15}NO_6$ *is the most abundant organic nitrate in both gas- and particle-phase measurements of α-pinene +* NO_3 *reactions (Nah et al., 2016). Boyd et al. (2015) mainly detected* $C_{10}H_{17}NO_4$*,* $C_{10}H_{15}NO_5$*,* $C_{10}H_{17}NO_5$*, and* $C_{10}H_{15}NO_6$ *with iodide-adduct CIMS from the α-pinene +* NO_3 *system. Using* $C_{10}H_{17}NO_5$ *and* $C_{10}H_{15}NO_6$ *as the examples, we checked their correlations with the products of [OH] × [monoterpenes], [*O_3*] × [monoterpenes], and [*NO_3*] × [monoterpenes] in different periods of the day (Fig. 14; Fig. S16). Comparatively,* $C_{10}H_{17}NO_5$ *and* $C_{10}H_{15}NO_6$ *correlate better with the products of [OH] × [monoterpenes] and [*O_3*] × [monoterpenes] during daytime (9:00~18:00). However, for the product of [*NO_3*] × [monoterpenes], its correlation coefficients with* $C_{10}H_{17}NO_5$ *and* $C_{10}H_{15}NO_6$ *are a bit higher at night (20:00 to 4:00 of the next day). These results indicate that monoterpene-derived organic nitrates can be mainly formed by the* NO_3*-initiated oxidations at night, but in daytime by the OH and* O_3*-initiated oxidations followed by NO termination of the* RO_2*.*

4.6.2 Sesquiterpene oxidations

The lightly oxygenated compounds from sesquiterpene reactions present a big morning peak and elevated signal intensities at night, similar to the diurnal variations of monoterpene lightly oxidized products. Hellén et al. (2018) showed that at the SMEAR II station, O_3 *oxidation dominated the first step of sesquiterpene reactions for the whole year. It has also been observed in central Amazonia that sesquiterpenes contributed the highest to total* O_3 *reactivity although sesquiterpene concentrations were much lower than those of monoterpenes and isoprene (Yee et al., 2018). At the SMEAR II station, emissions of sesquiterpenes are dominated by β-caryophyllene (Hellén et al., 2018). Photooxidation of β-caryophyllene in the chamber experiments resulted in high aerosol yield and is expected to strongly influence SOA formation (Jaoui et al., 2013). Using the mass spectrometric techniques, Jokinen et al. (2016) observed the production of highly oxidized organic compounds from β-caryophyllene ozonolysis, i.e., monomers* $C_{15}H_{24}O_{7,9,11}$ *and* $C_{15}H_{22}O_{9,11}$*, and dimers* $C_{29}H_{46}O_{12,14,16}$ *and* $C_{30}H_{46}O_{12,14,16}$*. However,* *due to the instrumental limitation, only the lightly oxidized products from sesquiterpene reactions were identified in this study.*

Interestingly, a strong RH-dependence was observed for the correlations between sesquiterpene lightly oxidized compounds and the product of [OH] × [sesquiterpenes] or [O_3] × [sesquiterpenes]. These products represent the oxidation rates of sesquiterpenes with OH radical and O_3. As shown in Fig. 15, the corresponding correlation coefficients vary significantly with RH. In addition, the signal intensities of sesquiterpene lightly oxidized products also show high dependence on RH. At lower RH (RH<40%), the signal intensities of sesquiterpene lightly oxidized products are relatively low and correlate closely with the product of [OH] × [sesquiterpenes] and [O_3] × [sesquiterpenes]. The high signal intensities of sesquiterpene lightly oxidized products occur when RH>70% but the correlation between sesquiterpene lightly oxidized compounds and the product of [OH] × [sesquiterpenes] or [O_3] × [sesquiterpenes] is more scattered. Such high RH-dependence was not observed for monoterpene lightly oxidized compounds (Fig. S17). These findings have not been observed by previous studies and the reasons behind remain unclear. High-RH conditions typically occur during nights with temperature inversion (Zha et al., 2018), while RH below 40% generally only occurs at the station during sunny days. Future studies are needed to dig deep into the atmospheric processes of sesquiterpenes and monoterpenes.

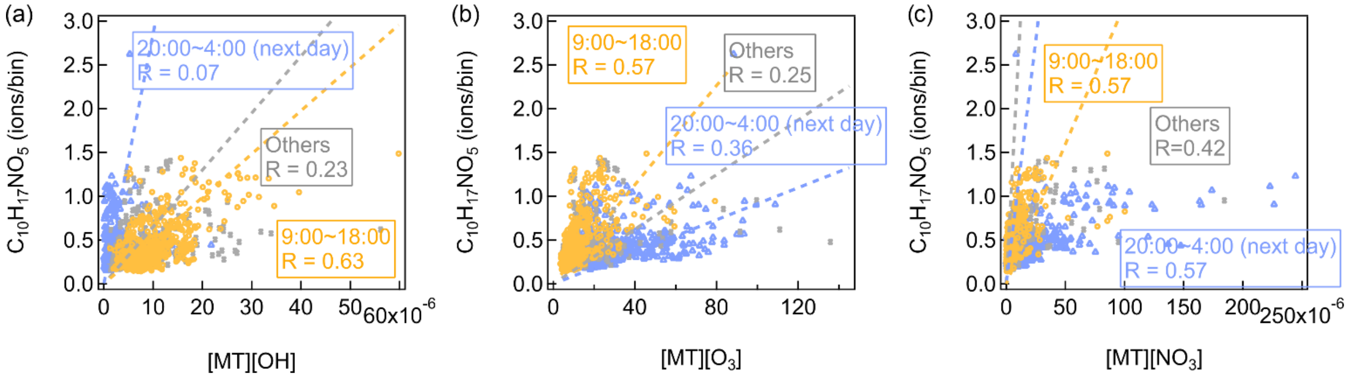

Figure 14. Scatter plots of $C_{10}H_{17}NO_5$ versus the product of (a) [OH] × [monoterpenes], (b) [O_3] × [monoterpenes], and (c) [NO_3] × [monoterpenes]. Different colours represent different periods of the day.

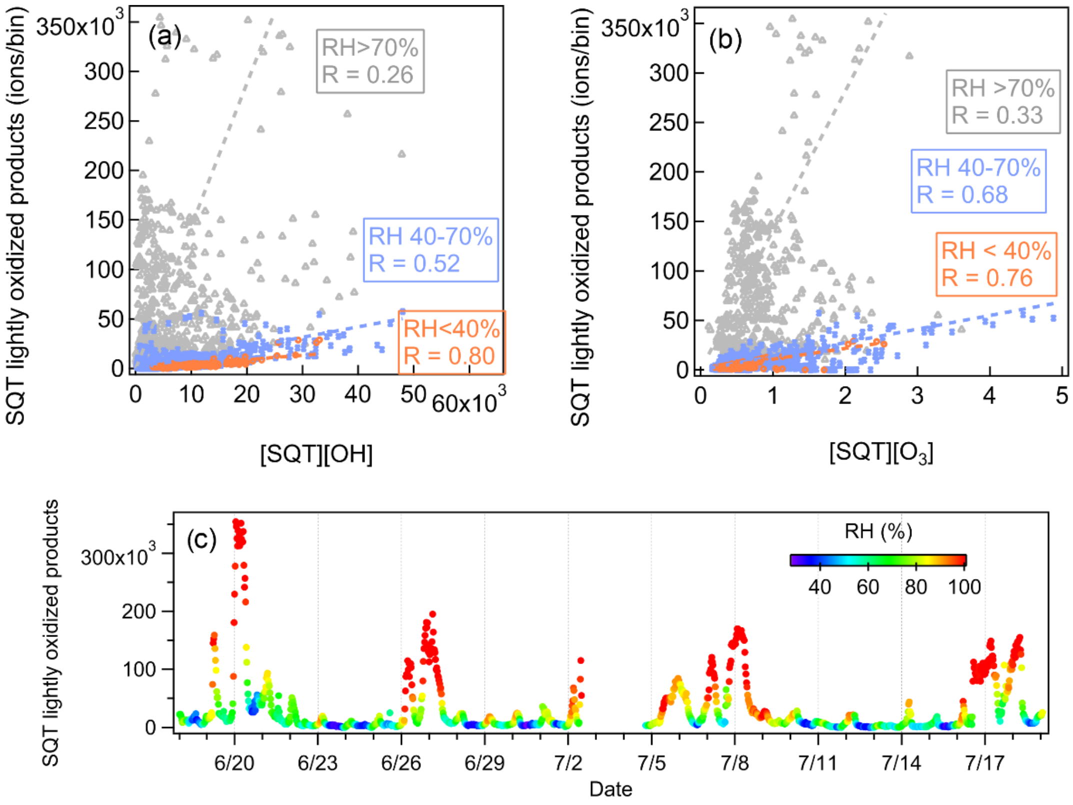

Figure 15. Scatter plots of sesquiterpene lightly oxidized products versus the product of (a) [OH] × [sesquiterpenes], and (b) [O_3] × [sesquiterpenes]. Different colours indicate different ranges of RH. (c) Time series of sesquiterpene lightly oxidized products colored by RH.

A certain amount of work in this paper goes into arriving at factorisations that aren't simply dominated by the big signals. This is probably to be expected because the gas phase VOC ensemble is likely to have many more degrees of freedom than can be accurately represented by the PMF and furthermore, many peaks will have isomers that won't be resolved using PTR. This is done by removing the main monoterpene signal and separating the mass spectrum into different regions. However, this comes across as a little subjective and prevents a direct association between the peaks in the two regions. Did the authors attempt a more conventional approach, such as applying a 'model error' parameter to downweight the larger peaks? More attention should be paid to demonstrating what the effects of not following these procedures in either case, perhaps shown in the supplement.

Response: As the reviewer points out, there are multiple ways in which data can be scaled before factorization, each one giving more or less weight to certain signals in the mass spectra. Earlier studies from our group have explored in detail e.g. scaling according to intensity or mass-to-charge ratio (Fig. 3 and Fig. S7 in Äijälä et al., 2017). The ultimate added value of such labor-intensive approaches is largest when the factorization results are ambiguous or hard to verify. In the current work, we tried a simple approach (removing the main peaks of the largest signals), which will be easy also for others to replicate. This produced factors that made sense both chemically and through their temporal behavior, which lends confidence in the results. The sub-range analysis, which we earlier have shown to be very powerful in separating out less abundant factors (Zhang et al., 2020), also provides a type of "internal verification" when factors with similar temporal and chemical features are resolved from the two different mass ranges. In the end, there is no single “correct” way to factorize atmospheric data, and the validity of the approach should be referenced to the results, and the conclusion that can be drawn from them.

More specifically concerning this study, the measured signals at *m/z* 81Th and *m/z* 137 Th were much higher than the others. In the Vocus PTR-TOF, *m/z* 81Th mainly comes from the fragmentation of *m/z* 137 Th (monoterpenes) and therefore follows the characteristics of *m/z* 137. With the inclusion of these super high peaks (Figure 1a), the mass profiles of three factors were quite similar and dominated by monoterpenes at *m/z* 137 Th and the major fragment at *m/z* 81 Th. After exclusion of these high peaks, the mass profiles were more distinct and representative of different factors and at the same time, their temporal behaviors were not interfered (Figure 1b). While the parent ions at *m/z* 137 Th and *m/z* 81 were excluded, their corresponding isotopes were retained, effectively downweighting their contributions to the PMF results. The time series of the resolved factors with and without the inclusion of these super high peaks are almost identical. As suggested by the reviewer, the time series and mass profiles of the resolved factors with the inclusion of monoterpene peaks are added in the supplement as Figure S1.

After the exclusion of monoterpene high peaks, if the entire mass spectrum was used for PMF analysis without subranges, factors identified in the high mass range in this study cannot be resolved. As shown in Figure 2, with the entire mass spectrum as PMF input, most identified factors in the low mass range were resolved although there were some mixing of different factors. For example, the factors of C_6 and C_7 lightly oxidized products, a plume event, monoterpenes, unknown source, monoterpene lightly oxidized products, and isoprene and its oxidation products, were clearly seen. However, the PMF analysis cannot separate the factors of sesquiterpenes, sesquiterpene lightly oxidized products, monoterpene more oxidized products, monoterpene-derived organic nitrates, and C_{13} lightly oxidized products. Increasing the number of factors for PMF run did not help.

In this study, with the factorization on subranges of the mass spectra, different factors representing primary emission sources and secondary oxidation processes were identified in both mass ranges. The association between these two ranges were further explored by comparison of their time series, diurnal variations, and correlation analysis (Figure 11 in the manuscript). For example, the factors of a plume event were resolved in both mass ranges and their time series correlated closely with each other. The monoterpene factor in the low mass range showed a good correlation with the sesquiterpene factor in the high mass range. Interestingly, the factor of C_6 and C_7 lightly oxidized products in the low mass range correlated very well with the factor of C_{13} lightly oxidized products in the high mass range, which lead to the speculation that the C_{13} oxygenated compounds are produced through the dimer formation mechanisms of C_6 and C_7 species. In addition, the factor of monoterpene lightly oxidized products showed a poor correlation with the factor of sesquiterpene lightly oxidized products. Without the PMF analysis on subranges of mass spectra, these factors and different processes cannot be separated. Zhang et al. (2020) performed factor analysis on subranges of mass spectra measured by NO_3^- CIMS, and found that the formation of daytime dimer and the monoterpene dimers from the combined products of NO_3 and O_3 oxidations cannot be resolved without the subrange approach.

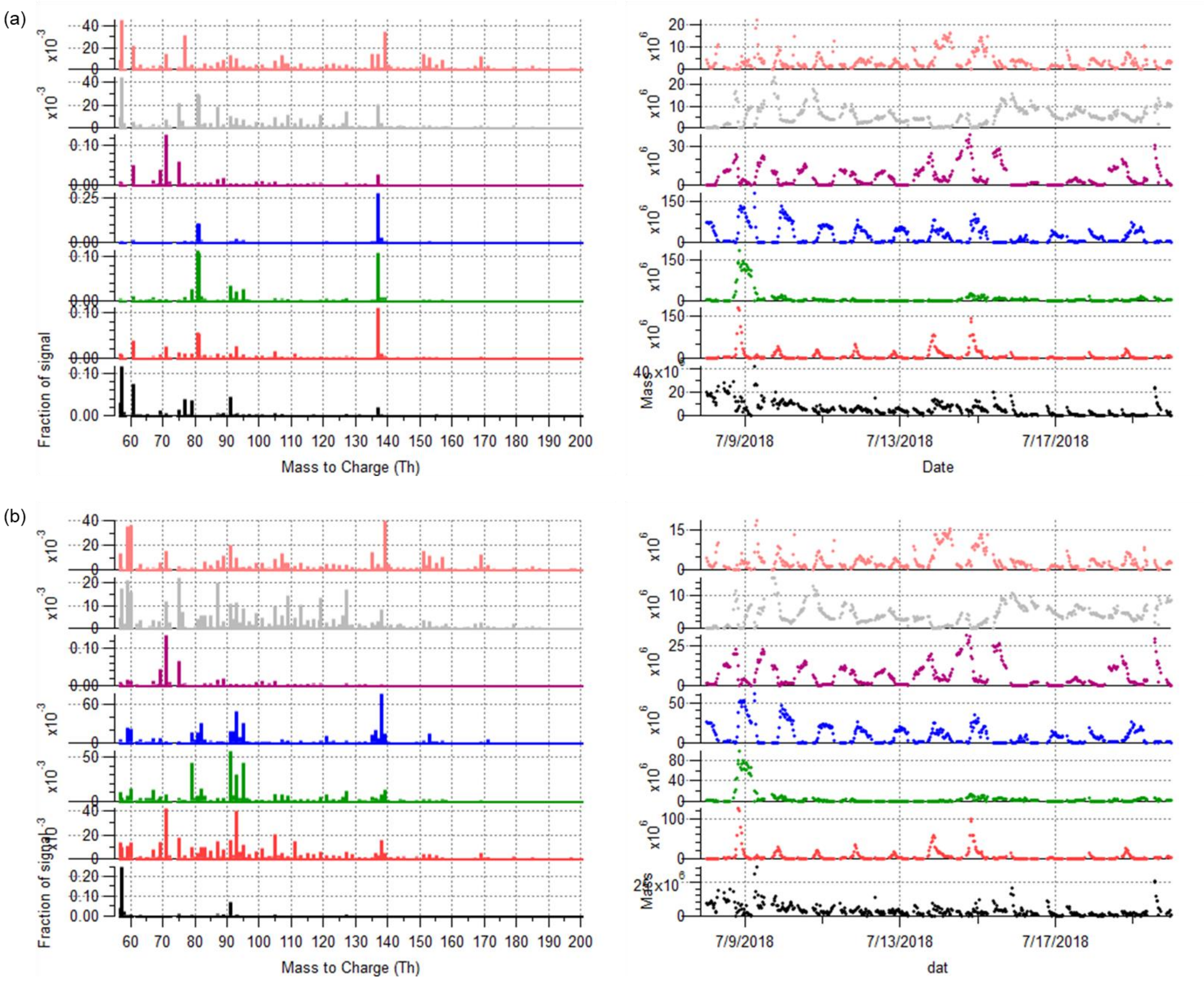

Figure 1. The mass profiles and time series of the seven-factor solution for the low mass range in the Landes forest (a) with and (b) without the inclusion of the signals at *m/z* 81 Th and *m/z* 137 Th.

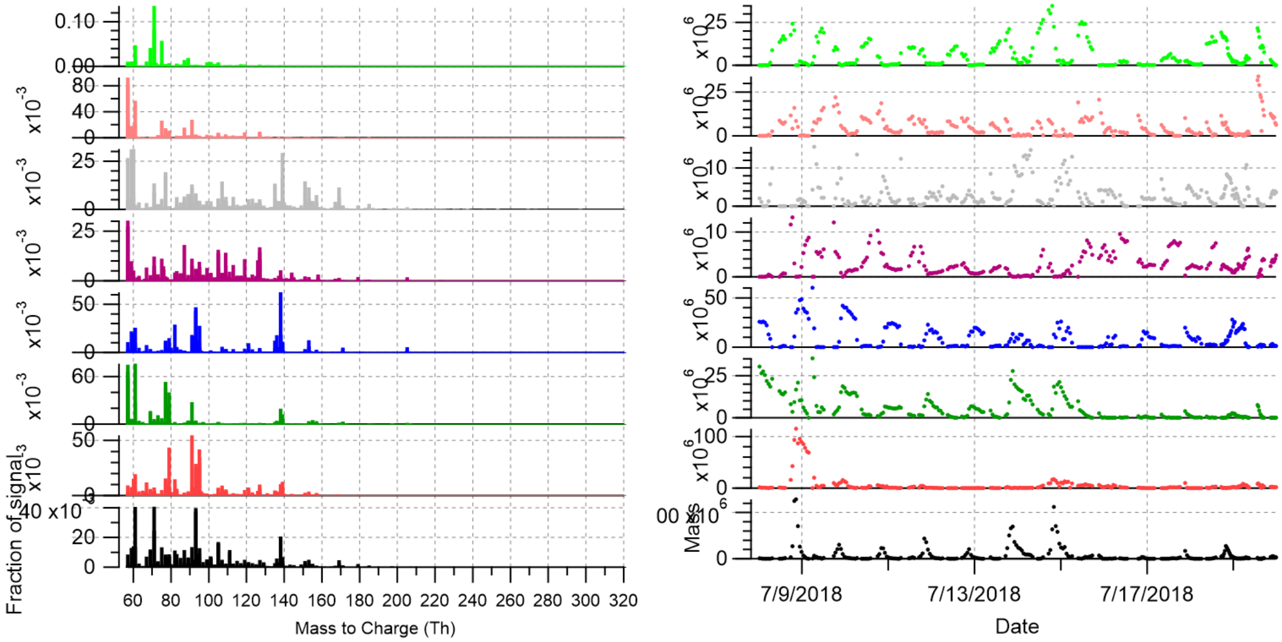

Figure 2. The mass profiles and time series of the eight-factor solution in the Landes forest with the entire mass spectrum as input of PMF analysis. We varied the FPEAK value between -1 and +1 with the step of 0.2. Taking the high mass range of 201-320 Th at the SMEAR IIstation as the example,

One might expect that given the number of degrees of freedom available, there will be a level of rotational ambiguity in the solution sets. This certainly would appear to be the case in figure 11, where all of the factors appear to contain traces of siloxane. Was the amount of rotational freedom available explored?

Response: The rotational freedom of the PMF solutions in this study was explored through use of the FPEAK parameters. For the optimal solutions, we varied the FPEAK value between -1 and +1 with the step of 0.2. For the low mass range of 51-200 Th of the Landes and SMEAR II dataset, the variations in FPEAK value did not influence the mass profile and time series much. For the high mass range of 201-320 Th, we saw the changes especially in the factor profiles by varying FPEAK values. For the Landes measurements, Figure 3 shows the factor profiles of the eight-factor solution with FPEAK = 0, +0.6, and -0.6. The time series of different factors for these FPEAK values are similar. After a detailed evaluation, we found no evidence that solutions with FPEAK value away from zero are preferable. However, for the high mass range of the SMEAR II measurements, as expected by the reviewer, the solutions with positive values of FPEAK work better than that with FPEAK = 0 in terms of the factor profiles. As shown in Fig. 4, by varying FPEAK with positive values, the factor profile of monoterpene more oxidized products (including organic nitrates) contained less traces of siloxanes and showed elevated fractions of the fingerprint peaks. After evaluation, we decided to choose the solution with FPEAK = +0.6 for the high mass range of the SMEAR II dataset.

The corresponding information of rotational ambiguity has been added in the revised manuscript (Lines 230-239): "*The rotational freedom of the PMF solutions was explored through the use of the FPEAK parameters. For each of the optimal solutions, we varied the FPEAK values between -1 and +1 with the step of 0.2. For the low mass ranges of the Landes and SMEAR II dataset, the varying FPEAK values did not change the factor profiles and time series much. For the high mass range of the Landes measurements, we saw variations especially in the factor profiles by varying FPEAK values. But after a detailed evaluation, we found no evidence that solutions with*

FPEAK values away from zero were preferable. However, for the high mass range of the SMEAR II measurements, the solutions with positive values of FPEAK worked better than that with FPEAK = 0 in terms of factor profiles. The factor time series were similar when FPEAK values varied. But for the factor profiles with positive FPEAK values, the factor of monoterpene more oxidized products including organic nitrates contained less traces of siloxanes and showed elevated fractions of the corresponding fingerprint peaks (Fig. S12). After evaluation, we chose the solution with FPEAK = +0.6 for the high mass range of the SMEAR II dataset.”

Figure 9, Figure 10, Figure 12, and Figure 13 have been updated accordingly.

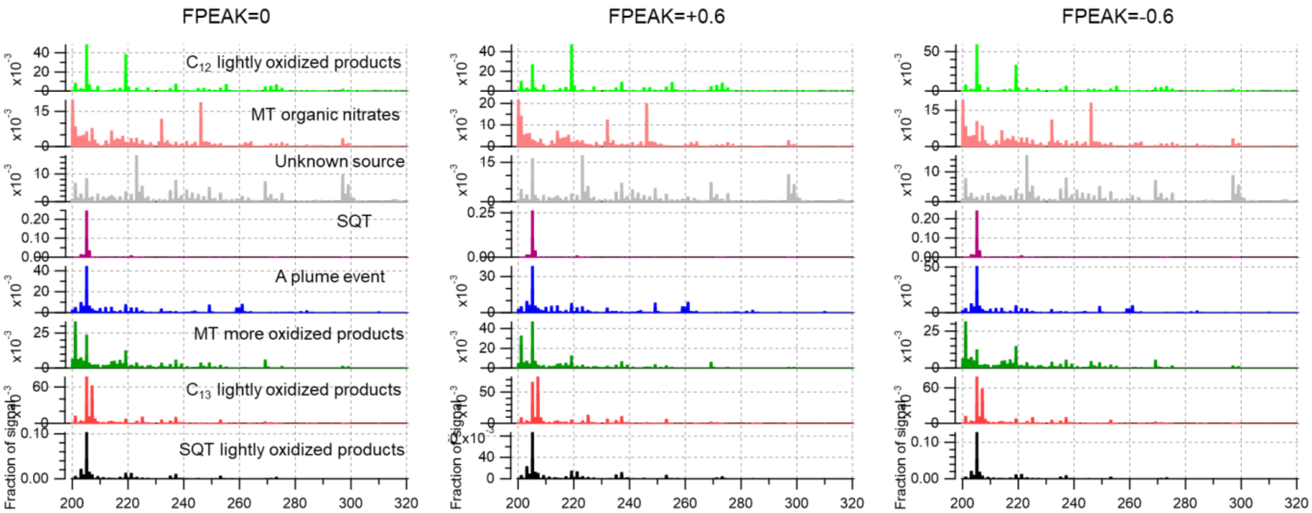

Figure 3. The factor profiles of the eight-factor solution for the high mass range of the Landes measurements with FPEAK = 0, +0.6, and -0.6.

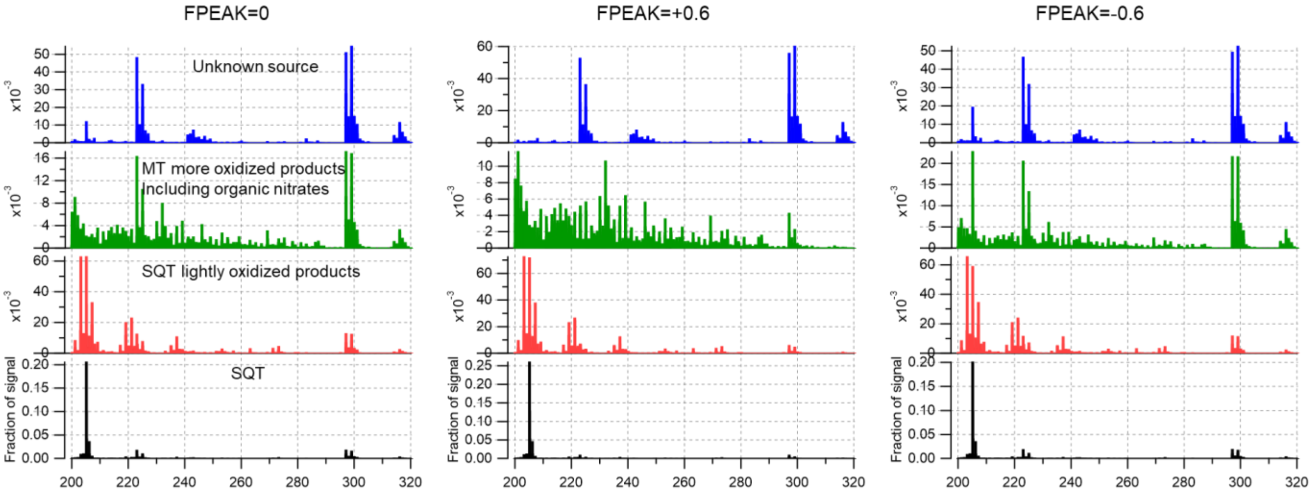

Figure 4. The factor profiles of the four-factor solution for the high mass range of the SMEAR II measurements with FPEAK = 0, +0.6, and -0.6.

The observation that reaction products did not contribute as much to the mass budget is perhaps expected because of their chemical lifetime. However, can the authors be sure the these (presumably more polar) molecules were being detected with equal efficiency? Have the authors tried comparing with a mechanistic model like the MCM or GECKO-A?

Response: The sensitivities of different VOCs in the PTR instrument are not equal and are linearly related to their proton-transfer reaction rate constants when ion transmission efficiency and fragmentation ions are considered (Sekimoto et al., 2017; Krechmer et al., 2018). According to Sekimoto et al. (2017), the reaction rate constants of different molecules significantly depend on their molecular mass, elemental composition, and functionality. In this study, we acknowledge that it is not a perfect method to quantify the mass fraction of different factors based on their average signal intensities as shown in the pie charts of Figure 12. The related uncertainties are discussed in the manuscript (Lines 419-423): “We acknowledge that it is not a perfect method to quantify the contributions of various sources and formation processes. The sensitivities of different VOCs measured by the PTR instruments may vary by a factor of 2-3 (Sekimoto et al., 2017; Yuan et al., 2017). The uncertainties can come from the challenge to convert the signal intensity to atmospheric concentrations because of problematic calibrations, especially given that many unknown molecules exist in the mass spectra.”

In this study, a large mass fraction of the gas-phase organic species were measured and classified including the precursors, the lightly oxidized products, and the more oxidized products, which was not achieved by previous studies. Although it is out of the scope of the current study to perform model simulations, our results provide good data base for potential model study in the future to compare model simulations with our ambient observations, improve model performance, and help scientists better understand the complex atmospheric chemistry. Still, the lack of speciation of e.g. the monoterpenes with the PTR approach remains a challenge for mechanistic modeling, as the oxidation product distributions will vary tremendously depending on the exact VOC distributions in the forests.

Anonymous Referee #2

This paper reports on data collected by a VOCUS PTR-ToF-MS at two forest sites. The VOCUS PTR-ToF is a powerful tool to characterize both biogenic and anthropogenic emissions due to it's high sensitivity and broad range of detectable organic compounds. For the first time, the binned positive matrix factorization (binPMF) algorithm has been applied to VOCUS data. Applying PMF to binned data with subsequent high-resolution peak fitting and identification of peaks found to be relevant is a clever way of data reduction in rich datasets as obtained by modern non-selective CIMS techniques.

The paper is technically sound; the authors describe individual PMF factors in great detail, but, unfortunately, the paper does not go beyond a description of observations, and I agree with Referee #1 that it feels unfinished at this stage. I therefore recommend that this paper be published only after major revisions.

Response: We thank Anonymous Referee #2 for the careful review and inputs which helped improving the overall quality of our work. We agree that as a research article, the paper should go beyond a description of PMF source apportionment and highlight new understanding towards atmospheric science from this work. Therefore, in the revised manuscript, our major findings are highlighted and more insights are gained into monoterpene and sesquiterpene oxidations. Please see more details in our responses to Referee #1.

Comments:

I think that the paper does not identify oxidation processes as stated in the abstract, nor does it provide a more comprehensive understanding of gas-phase organic chemistry.

Response: In this study, a large mass fraction of the gas-phase organic species were measured and classified. In addition to the precursors, their lightly oxidized products and more oxidized products were separated as individual factors. Based on the interpretation of these factors related to oxidation processes, further insights were gained regarding monoterpene and sesquiterpene reactions. In addition, the relative abundances of organic precursors, the lightly oxidized products, and the more oxidized products can be utilized by modelers to evaluate simulation output, improve model performance, and provide new perspectives to understand gas-phase physicochemical processes.

We revised the abstract as follows:

"Atmospheric organic vapors play essential roles in the formation of secondary organic aerosol. Source identification of these vapors is thus fundamental to understand their emission sources and chemical evolution in the atmosphere and their further impact on air quality and climate change. In this study, a Vocus proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF) was deployed in two forested environments, the Landes forest in southern France and the boreal forest in southern Finland, to measure atmospheric organic vapors, including both volatile organic compounds (VOCs) and their oxidation products. For the first time, we performed binned positive matrix factorization (binPMF) analysis on the complex mass spectra acquired with the Vocus PTR-TOF and identified various emission sources as well as oxidation processes in the atmosphere. Based on separate analysis of low- and high-mass ranges, fifteen PMF factors in the Landes forest and nine PMF factors in the Finnish boreal forest were resolved, showing a high similarity between the two sites. Particularly, terpenes and various terpene reaction products were separated into individual PMF factors with varying oxidation degrees, such as lightly oxidized compounds from both monoterpene and sesquiterpene oxidations, monoterpene-derived organic nitrates, and monoterpene more oxidized compounds. Factors representing monoterpenes dominated the biogenic VOCs in both forests, with less contributions from the isoprene factors and sesquiterpene factors. Factors of the lightly oxidized products, more oxidized products, and organic nitrates of monoterpenes/sesquiterpenes accounted for 8-12% of the measured gas-phase organic vapors in the two forests. Based on the interpretation of the results relating to oxidation processes, further insights were gained regarding monoterpene and sesquiterpene reactions. For example, a strong relative humidity (RH)-dependence was found for the behavior of sesquiterpene

lightly oxidized compounds. High concentrations of these compounds only occur at high RH, yet similar behavior was not observed for monoterpene oxidation products. These findings highlight the need for further studies to delve into gas-phase atmospheric processes of monoterpenes and sesquiterpenes."

The authors divided the mass spectra into two regions: 51 to 200 Th and 201 to 320 Th. Furthermore, they excluded two masses with high signal intensities (m81 and m137) from the PMF analysis, since these peaks were dominating the mass profiles. As far as I understand, both actions are necessary due to the fact that ambient concentrations of organic species and oxidaion products vary by many orders of magnitude, and the PMF method cannot resolve small signals. Maybe it's worth coming up with either a peak-by-peak normalization method prior to PMF analysis or feed the algorithm with logarithmized signal intensities. Please see this comment being made out of curiosity rather than critical.

Response: As the reviewer points out, there are multiple ways in which data can be scaled before factorization, each one giving more or less weight to certain signals in the mass spectra. Earlier studies from our group have explored in detail e.g. scaling according to intensity or mass-to-charge ratio (Fig. 3 and Fig. S7 in Äijälä et al., 2017). The ultimate added value of such labor-intensive approaches is largest when the factorization results are ambiguous or hard to verify. In the current work, we tried a simple approach (removing the main peaks of the largest signals), which will be easy also for others to replicate. This produced factors that made sense both chemically and through their temporal behavior, which lends confidence in the results. The sub-range analysis, which we earlier have shown to be very powerful in separating out less abundant factors (Zhang et al., 2020), also provides a type of "internal verification" when factors with similar temporal and chemical features are resolved from the two different mass ranges. In the end, there is no single "correct" way to factorize atmospheric data, and the validity of the approach should be referenced to the results, and the conclusion that can be drawn from them.

Specific comments: Figures 4,7,9 and 12: the y-axis' unit is 'ions/bin' - I think that should be changed into something like 'ions/factor'.

Response: These figures show the time series of different factors. The unit corresponds to the binned signal intensities measured by the mass spectrometer and should be "ions/bin".

line 62 and 307: replace "complicated" with "complex"

Response: Replaced.

line 182: please specifiy what 'high' means.

Response: As shown in Figure S2, for some bins, the scaled residual can as high as ±200. In the revised manuscript (Line 196), it is specified as "*For some bins the residuals are still high (the scaled residuals as high as ±200).*"

line 293: "much higher intensities" - please rephrase

Response: Done.

References

Äijälä, M., Heikkinen, L., Fröhlich, R., Canonaco, F., Prévôt, A. S. H., Junninen, H., Petäjä, T., Kulmala, M., Worsnop, D., and Ehn, M.: Resolving anthropogenic aerosol pollution types – deconvolution and exploratory classification of pollution events, Atmos. Chem. Phys., 17, 3165–3197, https://doi.org/10.5194/acp-17-3165-2017, 2017.

Zhang, Y., Peräkylä, O., Yan, C., Heikkinen, L., Äijälä, M., Daellenbach, K. R., Zha, Q., Riva, M., Garmash, O., Junninen, H., Paatero, P., Worsnop, D., and Ehn, M.: Insights into atmospheric oxidation processes by performing factor analyses on subranges of mass spectra, Atmos. Chem. Phys., 20, 5945–5961, https://doi.org/10.5194/acp-20-5945-2020, 2020.