Insights on Atmospheric Oxidation Processes by Performing Factor Analyses on Sub-ranges of Mass Spectra

Yanjun Zhang[1], Otso Peräkylä[1], Chao Yan[1], Liine Heikkinen[1], Mikko Äijälä[1], Kaspar R. Daellenbach[1], Qiaozhi Zha[1], Matthieu Riva[1,2], Olga Garmash[1], Heikki Junninen[1,3], Pentti Paatero[1], Douglas Worsnop[1,4], and Mikael Ehn[1]

[1] Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Helsinki, 00014, Finland

[2] Univ Lyon, Université Claude Bernard Lyon 1, CNRS, IRCELYON, F-69626, Villeurbanne, France

[3] Institute of Physics, University of Tartu, Tartu, 50090, Estonia

[4] Aerodyne Research, Inc., Billerica, MA 01821, USA

Corresponding author: yanjun.zhang@helsinki.fi

Abstract

Our understanding of atmospheric oxidation chemistry has improved significantly in recent years, greatly facilitated by developments in mass spectrometry. The generated mass spectra typically contain vast amounts of information on atmospheric sources and processes, but the identification and quantification of these is hampered by the wealth of data to analyze. The implementation of factor analysis techniques have greatly facilitated this analysis, yet many atmospheric processes still remain poorly understood. Here, we present new insights on highly oxygenated products from monoterpene oxidation, measured by chemical ionization mass spectrometry, at a boreal forest site in Finland in fall 2016. Our primary focus was on the formation of accretion products, i.e. “dimers”. We identified the formation of daytime dimers, with a diurnal peak at noon time, despite high nitric oxide (NO) concentrations typically expected to inhibit dimer formation. These dimers may play an important role in new particle formation events that are often observed in the forest. In addition, dimers identified as combined products of NO_3 and O_3 oxidation of monoterpenes were also found to be a large source of low-volatile vapors at night. This highlights the complexity of atmospheric oxidation chemistry, and the need for future laboratory studies on multi-oxidant systems. Neither of these two processes could have been separated without the new analysis approach deployed in our study, where we applied binned positive matrix factorization (binPMF) on sub-ranges of the mass spectra, rather than the traditional approach where the entire mass spectrum is included for PMF analysis. In addition to the main findings listed above, several other benefits compared to traditional methods were found.

1 Introduction

Huge amounts of volatile organic compounds (VOC) are emitted to the atmosphere every year (Guenther et al., 1995;Lamarque et al., 2010), which play a significant role in atmospheric chemistry and affect the oxidative ability of the atmosphere. The oxidation products of VOC can contribute to the formation and growth of secondary organic aerosols (Kulmala et al., 2013;Ehn et al., 2014;Kirkby et al., 2016;Troestl et al., 2016), affecting air quality, human health, and climate radiative forcing (Pope III et al., 2009;Stocker et al., 2013;Zhang et al., 2016;Shiraiwa et al., 2017). Thanks to the advancement in mass spectrometric applications, like the aerosol mass spectrometer (AMS) (Canagaratna et al., 2007) and chemical ionization mass spectrometry (CIMS) (Bertram et al., 2011;Jokinen et al., 2012;Lee et al., 2014), our capability to detect these oxidized products, as well as our understanding of the complicated atmospheric oxidation pathways in which they take part, have been greatly enhanced.

Monoterpenes ($C_{10}H_{16}$), one major group of VOC emitted in forested areas, have been shown to be a large source of atmospheric secondary organic aerosol (SOA). The oxidation of monoterpenes produces an abundance of different oxidation products (Oxygenated VOC, OVOC), including highly oxygenated organic molecules (HOM) with molar yields in the range of a few percent, depending on the specific monoterpene and oxidant (Ehn et al., 2014;Bianchi et al., 2019). Recent chamber studies have greatly advanced our knowledge of formation pathways for monoterpene HOM products, e.g. “monomers” (typically $C_{9\text{-}10}H_{12\text{-}16}O_{6\text{-}12}$) and “dimers” (typically $C_{19\text{-}20}H_{28\text{-}32}O_{8\text{-}18}$). Dimers, as shown by previous studies, can contribute to new particle formation (NPF) (Kirkby et al., 2016;Troestl et al., 2016;Lehtipalo et al., 2018), and are thus of particular interest.

In nearly all atmospheric oxidation chemistry, peroxy radicals (RO_2) are the key intermediates (Orlando and Tyndall, 2012). They form when VOC react with oxidants like ozone, or the hydroxyl (OH) or nitrate (NO_3) radicals, while their termination occurs mainly by bimolecular reactions with nitric oxide (NO), hydroperoxyl (HO_2) and/or other RO_2. RO_2+R’O_2 reactions can form ROOR’ dimers (Berndt et al., 2018a;Berndt et al., 2018b), and this pathway competes with RO_2+NO reactions, meaning that NO, formed by photolysis of NO_2, can efficiently suppress dimer formation, as also seen from atmospheric HOM observations (Ehn et al., 2014;Yan et al., 2016). Mohr et al. (2017) also reported daytime dimers in the boreal forest in Finland, coinciding with NPF events. A better understanding of the formation of these daytime dimers would assist elucidating NPF and particle growth mechanisms.

At night, nitrogen oxides can also impact the oxidation pathways, when NO_2 and O_3 react to form NO_3 radicals that can oxidize monoterpenes. NO_3 radicals are greatly reduced during daytime due to photolysis and reactions with NO reducing their lifetime to a few seconds (Ng et al., 2017). Yan et

al. (2016) reported nighttime HOM initiated by NO_3 in the boreal forest in Finland, but to our knowledge there have been no laboratory studies on HOM formation from NO_3 oxidation of monoterpenes. However, there have been several studies looking into the SOA formation in these systems, finding that certain monoterpenes, like β-pinene, have very high SOA yields, while the most abundant monoterpene, α-pinene, has negligible SOA forming potential. It remains an open question what the role of NO_3 radical oxidation of monoterpenes, and the observed NO_3-derived HOM, in the night-time boreal forest is. Identification of these processes in the ambient environment is fundamental towards better understanding of NPF and SOA.

The recent development of CIMS techniques has allowed researchers to observe unprecedented numbers of OVOC, in real-time (Riva et al., 2019). This ability to measure thousands of compounds is a great benefit, but also a large challenge for the data analyst. For this reason, factor analytical techniques have often been applied to reduce the complexity of the data (Huang et al., 1999), e.g. positive matrix factorization, PMF (Paatero and Tapper, 1994;Zhang et al., 2011). The factors have then been attributed to sources (e.g. biomass burning organic aerosol) or processes (e.g. monoterpene ozonolysis) depending on the application and ability to identify spectral signatures (Yan et al., 2016;Zhang et al., 2017).

In the vast majority of these PMF applications to mass spectra, the mass range of ions has been maximized in order to provide as much input as possible for the algorithm. This approach was certainly motivated in early application of PMF on e.g. offline filters, with chemical information of metals, water-soluble ions, and organic and elemental carbon (OC and EC), where the number of variables is counted in tens, and the number of samples in tens or hundreds (Zhang et al., 2017). However, with gas-phase CIMS, we often have up to a thousand variables, with hundreds or even thousands of samples, meaning that the amount of data itself is unlikely to be a limitation for PMF calculation. In this work, we aimed to explore potential benefits of dividing the spectra into sub-ranges before applying factorization analysis. This approach was motivated by several issues, which we expected to be resolvable by analyzing several mass range separately. Firstly, the loss rate of OVOC by condensation is strongly coupled to the molecular mass (Peräkylä et al., 2020), likely giving very different behaviors for the high and low mass ranges, even when produced by the same source. Second, dimers are a product of two RO_2, which can have different sources, meaning that they may have temporal profiles unlike anything observable for monomers. Finally, if one mass range contains much less signal than another, it will have very little impact on the final PMF results.

In this study, we applied PMF analysis on three different mass ranges of mass spectra of OVOC measured by a chemical ionization atmospheric pressure interface time-of-flight (CI-APi-TOF, Jokinen et al. (2012)) mass spectrometer in the Finnish boreal forest. We utilized our recently

proposed new PMF approach, binPMF, to include as much of the high-resolution information in the mass spectra as possible, in a robust way (Zhang et al., 2019). We show the benefits of the sub-range PMF approach to better separate chemical sources, by reducing disturbance from variable loss terms of the OVOC. Much of the analysis focuses on dimer formation pathways, and the role of different nitrogen oxides in these pathways. We find that both daytime dimers and dimers resulting from the combination of different oxidants can be separated with the sub-range approach, but not with the PMF applied to the full mass range. We believe that this study will provide new perspectives for future studies analyzing gas-phase CIMS data.

2 Methodology

The focus of this work is on retrieving new information from mass spectra by applying new analytical approaches. Therefore, we chose a dataset that has been presented earlier, though without PMF analysis, by Zha et al. (2018), and was also used in the first study describing the binPMF method (Zhang et al., 2019). The measurements are described in more details below in section 2.1, while the data analysis techniques used in this work are presented in section 2.2.

2.1 Measurements

2.1.1 Ambient site

The ambient measurements were conducted at the Station for Measuring Ecosystem–Atmosphere Relations (SMEAR) II in Finland (Hari and Kulmala, 2005) as part of the Influence of Biosphere-Atmosphere Interactions on the Reactive Nitrogen budget (IBAIRN) campaign (Zha et al, 2018). Located in the boreal environment in Hyytiälä, SMEAR II is surrounded with coniferous forest and has limited anthropogenic emission sources nearby. Diverse measurements of meteorology, aerosol and gas phase properties are continuously conducted at the station. Details about the meteorological conditions and temporal variations of trace gases during IBAIRN campaign are presented by Zha et al. (2018) and Liebmann et al. (2018).

2.1.2 Instrument and data

Data were collected with a nitrate (NO_3^-)-based chemical ionization atmospheric pressure interface time-of-flight mass spectrometer (CI-APi-TOF, Jokinen et al., 2012) with about 4000 Th Th^{-1} mass resolving power, at ground level in September, 2016. In our study, the mass spectra were averaged to 1 h time resolution from September 6th to 22nd for further analysis. We use the thomson (Th) as the unit for mass/charge, with 1 Th = 1 Da/e, where *e* is the elementary charge. As all the data discussed in this work are based on negative ion mass spectrometry, we will use the absolute value of the

mass/charge, although the charge of each ion will be negative. The masses discussed in this work includes the contribution from the nitrate ion, 62, unless specifically mentioned. Furthermore, as the technique is based on soft ionization with NO_3^- ions, any multiple charging effects are unlikely, and therefore the reported mass/charge values in thomson can be considered equivalent to the mass of the ion in Da.

The forest site of Hyytiälä is dominated by monoterpene emissions (Hakola et al., 2006). The main feature of previous CI-APi-TOF measurements in Hyytiälä (Ehn et al., 2014;Yan et al., 2016) has been a bimodal distributions of HOM, termed monomers and dimers, as they are formed of either one or two RO_2 radicals, respectively. For the analysis in this study, we chose three mass/charge (*m/z*) ranges of 50 Th each (Figure 1), corresponding to regions between which we expect differences in formation or loss mechanisms. In addition to regions with HOM monomers and HOM dimers, one range was chosen at lower masses, in a region presumably mainly consisting of molecules that are less likely to condense onto aerosol particles (Peräkylä et al., 2020).

2.2 Positive matrix factorization (PMF)

After the model of PMF was developed (Paatero and Tapper, 1994), numerous applications have been conducted with different types of environmental data (Song et al., 2007;Ulbrich et al., 2009;Yan et al., 2016;Zhang et al., 2017). By reducing dimensionality of the measured dataset, PMF model greatly simplifies the data analysis process with no requirement for prior knowledge of sources or pathways as essential input. The main factors can be further interpreted with their unique/dominant markers (elements or masses).

The basic assumption for PMF modelling is mass balance, which assumes that ambient concentration of a chemical component is the sum of contributions from several sources or processes, as shown in equation (1).

$$\mathbf{X} = \mathbf{TS} \times \mathbf{MS} + \mathbf{R} \tag{1}$$

In equation (1), **X** stands for the time series of measured concentration of different variables (*m/z* in our case), **TS** represents the temporal variation of factor contributions, **MS** stands for factor profiles (mass spectral profiles), and **R** is the residual as the difference of the modelled and the observed data. The matrices **TS** and **MS** are iteratively calculated by a least-squares algorithm utilizing uncertainty estimates, to pursue minimized Q value as shown in equation (2), where S_{ij} is the estimated uncertainty, an essential input in PMF model.

$$Q = \sum\sum\left(\frac{R_{ij}}{S_{ij}}\right)^2 \tag{2}$$

PMF model was conducted by multi-linear engine (ME-2) (Paatero, 1999) interfaced with Source Finder (SoFi, v6.3) (Canonaco et al., 2013). Signal-to-noise ratio (SNR) was calculated as SNR*ij* = abs (*Xij*) / abs (*Sij*). When the Signal-to-noise ratio (SNR) is below 1, the signal of X_{ij} will be down-weighted by replacing the corresponding uncertainty S_{ij} by S_{ij}/SNR$_{ij}$ (Visser et al., 2015). Future studies should pay attention to the potential risk when utilizing this method since down-weighting low signals element-wise will create a positive bias to the data. Robust mode was operated in the PMF modelling, where outliers ($\left|\frac{R_{ij}}{S_{ij}}\right| > 4$) were significantly down-weighted (Paatero, 1997).

2.3 binPMF

As a newly developed application of PMF for mass spectral data, binPMF has no requirement for chemical composition information, while still taking advantage of the HR mass spectra, saving effort and time (Zhang et al., 2019). To explore the benefits of analyzing separated mass ranges, we applied binPMF to the three separated ranges. The three ranges were also later combined for binPMF analysis as comparison with the previous results. The PMF model requires both data matrix and error matrix as input, and details of the preparation of data and error matrices are described below.

2.3.1 Data matrix

Different from normal UMR or HR peak fitting, in binPMF, the mass spectra are divided into small bins after baseline subtraction and mass axis calibration. Linear interpolation was first conducted to the mass spectra with a mass interval of 0.001 Th. Then the interpolated data was averaged into bins of 0.02 Th width. We selected three ranges for further analysis based on earlier studies (Ehn et al., 2014;Yan et al., 2016;Bianchi et al., 2019;Peräkylä et al., 2020).

- Range 1, *m/z* 250 – 300 Th, 51 unit masses × 25 bins per unit mass = 1275 bins/variables, consisting mainly of molecules with five to nine carbon atoms and four to nine oxygen atoms in our dataset.
- Range 2, *m/z* 300 – 350 Th, 51×25 = 1275 bins, mainly corresponding to HOM monomer products, featured with nine to ten C- and seven to ten O-atoms.
- Range 3, *m/z* 510 – 560 Th, 51×30 = 1530 bins, mainly corresponding to HOM dimer products, with carbon numbers of sixteen to twenty and eleven to fifteen O-atoms.

To avoid unnecessary computation, only signal regions with meaningful signals in the mass spectra were binned (Zhang et al., 2019). For a nominal mass *N*, the signal region included in further analyses was between *N*-0.2 Th and *N*+0.3 Th for Range 1 and 2, and between *N*-0.2 Th and *N*+0.4 Th for

Range 3. The wider signal regions in Range 3 is due to wider peaks at higher masses. The data were averaged into 1-h time resolution and in total we had 384 time points in the data matrix.

2.3.2 Error matrix

The error matrix represents the estimated uncertainty for each element of the data matrix and is crucial for iterative calculation of the Q minimum. Equation (3) is used for error estimation (Polissar et al., 1998),

$$S_{ij} = \sigma_{ij} + \sigma_{\text{noise}} \tag{3}$$

where S_{ij} represents the uncertainty of *m/z* j at time i, σ_{ij} stands for counting statistics uncertainty and is estimated as follows,

$$\sigma_{ij} = a \times \frac{\sqrt{I_{ij}}}{\sqrt{t_s}} \tag{4}$$

where I is the signal intensity term, in unit of counts per second (cps), t_s stands for length of averaging in seconds, while a is an empirical coefficient to compensate for unaccounted uncertainties (Allan et al., 2003;Yan et al., 2016) and is 1.28 in our study as previously estimated from laboratory experiments (Yan et al., 2016). The σ_{noise} term was estimated as the median of the standard deviations from signals in the bins in the region between nominal masses, where no physically meaningful signals are expected.

3 Results

3.1 General overview of the dataset/spectrum

During the campaign, in autumn, 2016, the weather was overall sunny and humid with average temperature of 10.8 °C and relative humidity (RH) of 87% (Zha et al., 2019). The average concentration of NO_x and O_3 was 0.4 ppbv and 21 ppbv, respectively. The average total HOM concentration was ~ 10^8 molecules cm^{-3}.

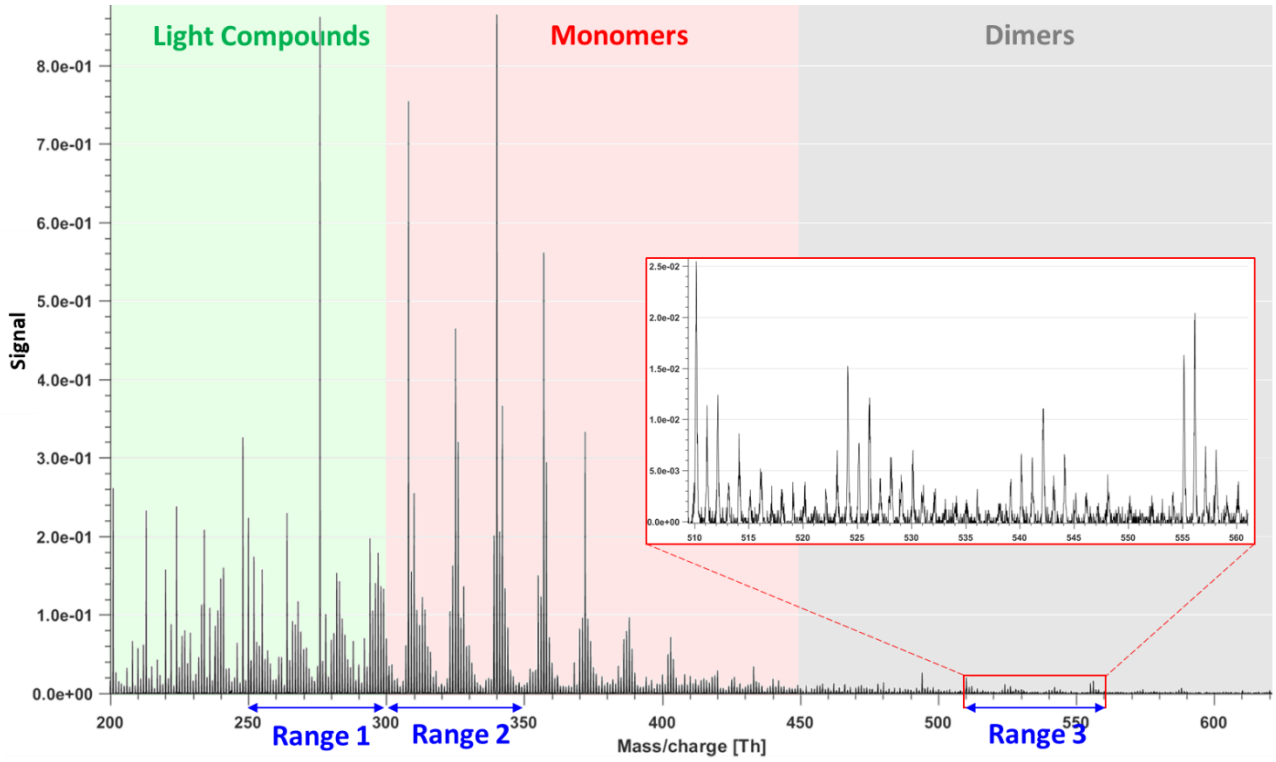

Figure 1. Example of mass spectrum with 1-h time resolution measured from a boreal forest environment during the IBAIRN campaign (at 18:00, Finnish local time, UTC+2). The mass spectrum was divided into three parts and three sub-ranges were chosen from different parts for further analysis in our study. The nitrate ion (62 Th) is included in the mass.

Figure 1 shows the 1 h averaged mass spectrum taken at 18:00 on September 12, as an example of the analyzed dataset. In addition to exploring the benefits of this type of sub-range analysis in relation to different formation or loss pathways, separating into sub-ranges may also aid factor identification for low-signal regions. As shown in Figure 1, there is a difference of 1-2 orders of magnitude in the signal intensity between Range 3 and Ranges 1-2. If all Ranges are run together, we would expect that the higher signals from Ranges 1 and 2 will drive the factorization. While if run separately, separating formation pathways of dimers in Range 3 will likely be easier. As dimers have been shown to be crucial for the formation of new aerosol particles from monoterpene oxidation (Kirkby et al., 2016;Troestl et al., 2016;Lehtipalo et al., 2018), this information may even be the most critical in some cases, despite the low contribution of these peaks to the total measured signal.

binPMF was separately applied to Range 1, 2, 3, and a 'Range combined' which comprised all the three sub-ranges. All the PMF runs for the four ranges were conducted from two to ten factors and repeated three times for each factor number, to assure the consistency of the results. Factorization results and evolution with increasing factor number are briefly described in the following sections, separately for each Range (sections 3.2 – 3.5). It is worth noting that the factor order in factor

evolution does not necessarily correspond to that of the final results. The factor orders displayed in Figures 2-5 have been modified for further comparison between different ranges. More detailed discussion and comparison between the results are presented in Section 4.

3.2 binPMF on Range 1 (250 – 300 Th)

As has become routine (Zhang et al., 2011;Craven et al., 2012), we first examined the mathematical parameters of our solutions. From two to ten factors, Q/Q_{exp} decreased from 2.8 to 0.7 (Fig S1 in supplementary information), and after three factors, the decreasing trend was gradually slowing down and approaching one, which is the ideal value for Q/Q_{exp} as a diagnostic parameter. The unexplained variation showed a decline from 18% to 8% from two to ten factors.

In the two-factor results, two daytime factors were separated, with peak time both at 14:00 - 15:00. One factor was characterized by large signals at 250 Th, 255 Th, 264 Th, 281 Th, 283 Th, 295 Th, 297 Th. The other factor was characterized by large signals at 294 Th, 250 Th, 252 Th, 264 Th, 266 Th, 268 Th, and 297 Th. In Hyytiälä, as reported in previous studies, odd masses observed by the nitrate CI-APi-TOF are generally linked to monoterpene-derived organonitrates during the day (Ehn et al., 2014;Yan et al., 2016). When the factor number increased to three, the two earlier daytime factors remained similar with the previous result, while a new factor appeared with a distinct sawtooth shape in the diurnal cycle. The main marker in the spectral profile was 276 Th, with a clear negative mass defect. When one more factor was added, the previous three factors remained similar as in the three-factor solution, and a new morning factor was resolved, with 264 Th and 297 Th dominant in the mass spectral profile, and a diurnal peak at 11:00.

As the factor number was increased, more daytime factors were separated, with similar spectral profiles to existing daytime factors and various peak times. No nighttime factors were found in the analysis even when the factor number reached ten. We chose the four-factor result for further discussion, and Figure 2 shows the result of Range 1, with spectral profile, time series, diurnal cycle and averaged factor contribution during the campaign. As shown in Figure 2d, Factors 1-3 are all daytime factors, while Factor 4 has no clear diurnal cycle, but a distinct sawtooth shape. Factor 4 comes from a contamination of perfluorinated acids, from the inlet's automated zeroing every three hours during the measurements (Zhang et al., 2019). The zeroing periods have been removed from the dataset before binPMF analysis, but the contamination factor was still resolved. This factor is discussed in more detail in sections 4.1 and 4.4.

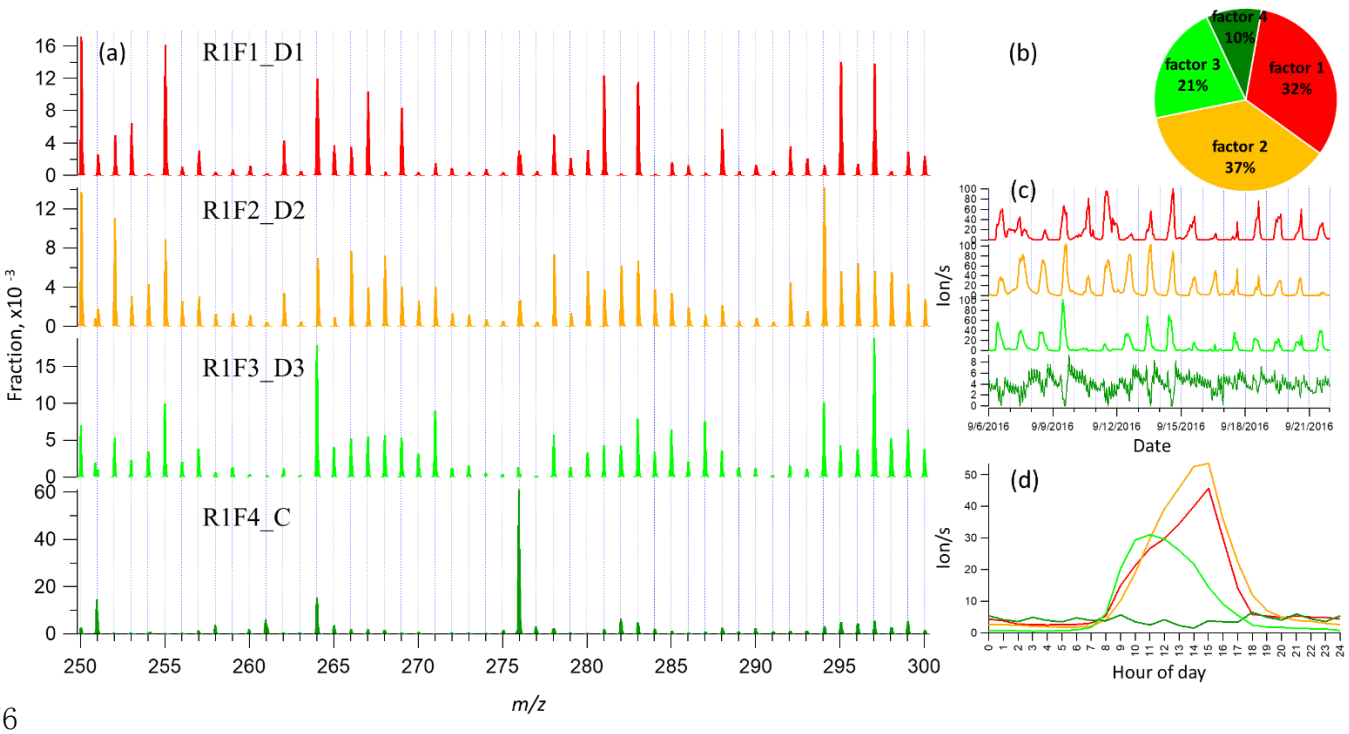

Figure 2 Four-factor result for Range 1, for (a) factor spectral profiles, (b) averaged factor contribution during the campaign, (c) time series and (d) diurnal trend. Details on the factors' naming schemes are shown in Table 1.

3.3 binPMF on Range 2 (300-350 Th)

This range covers the monoterpene HOM monomer range, and binPMF results have already been discussed by Zhang et al. (2019) as a first example of the application of binPMF on ambient data. Our input data here is slightly different. In the previous study, the 10 min automatic zeroing every three hours was not removed before averaging to 1 hour time resolution while here, we have removed this data. Overall, the results are similar as in our earlier study, and therefore the result are just briefly summarized below for further comparison and discussion in Section 4. Similar to Range 1, both the Q/Q_{exp} (2.2 to 0.6) and unexplained variation (16% to 8%) declined with the increased factor number from two to ten.

When the factor number was two, one daytime and one nighttime factor were separated, with diurnal peak times at 14:00 and 17:00, respectively. The nighttime factor was characterized by masses at 340 Th, 308 Th and 325 Th (monoterpene ozonolysis HOM monomers (Ehn et al., 2014)) and remained stable throughout the factor evolution from two to ten factors. With the addition of more factors, no more nighttime factors got separated while the daytime factor was further separated and more daytime factors appeared, peaking at various times in the morning (10:00 am), at noon or in the early afternoon (around 14:00 pm and 15:00 pm). High contribution of 339 Th can be found in all the daytime factor

profiles. As the factor number reached six, a contamination factor appeared, characterized by large signals at 339 Th and 324 Th, showing negative mass defects (Figure S2 in the Supplement). The factor profile is nearly identical to the contamination factor determined in Zhang et al. (2019), where the zeroing periods were not removed, causing larger signals for the contaminants. In our dataset, where the zeroing periods were removed, no sawtooth pattern was discernible in the diurnal trend, yet it could still be separated even though it only contributed 3% to Range 2. More about the contamination factors from different ranges will be discussed in Section 4.4. We chose to show the four-factor result below, to simplify the later discussion and comparison. Figure 3 shows four-factor result of Range 2, with spectral profile, time series, diurnal cycle and averaged factor contribution during the campaign.

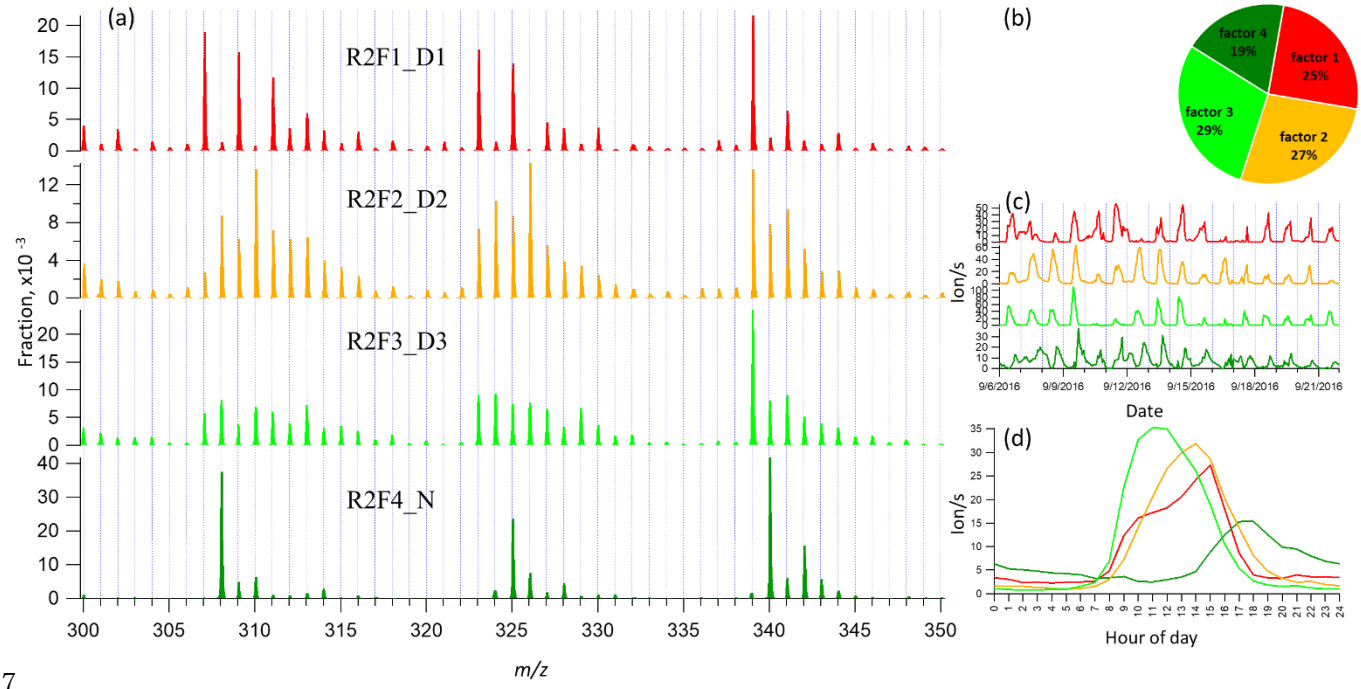

Figure 3 Four-factor result for Range 2, for (a) factor spectral profiles, (b) averaged factor contribution during the campaign, (c) time series and (d) diurnal trend. Details on the factors' naming schemes are shown in Table 1.

3.4 binPMF on Range 3 (510-560 Th)

Range 3 represents mainly the monoterpene HOM dimers (Ehn et al., 2014). Similar to Range 1 and 2, both the Q/Q_{exp} (1.5 to 0.6) and unexplained variation (18% to 15%) showed decreasing trend with the increased factor number (2-10). As can be seen from Figure 1, data in Range 3 had much lower signals, compared to that of the Range 1 and 2, explaining the higher unexplained variation for Range 3.

In the two-factor result for Range 3, one daytime and one nighttime factor appeared, with diurnal peak times at noon and 18:00, respectively. The nighttime factor was characterized by ions at 510 Th, 524 Th, 526 Th, 542 Th, and 555 Th, 556 Th, while the daytime factor showed no dominant marker masses, yet with relatively high signals at 516 Th, 518 Th and 520 Th. As the number of factors increased to three, one factor with almost flat diurnal trend was separated, with dominant masses of 510 Th, 529 Th, 558 Th. Most peaks in this factor had negative mass defects, and this factor was again linked to a contamination factor. The four-factor result resolved another nighttime factor with a dominant peak at 555 Th, and effectively zero contribution during daytime. As the factor number was further increased, the new factors seemed like splits from previous factors with similar spectral profiles. We therefore chose four-factor result also for Range 3 (results shown in Fig. 4) for further discussion.

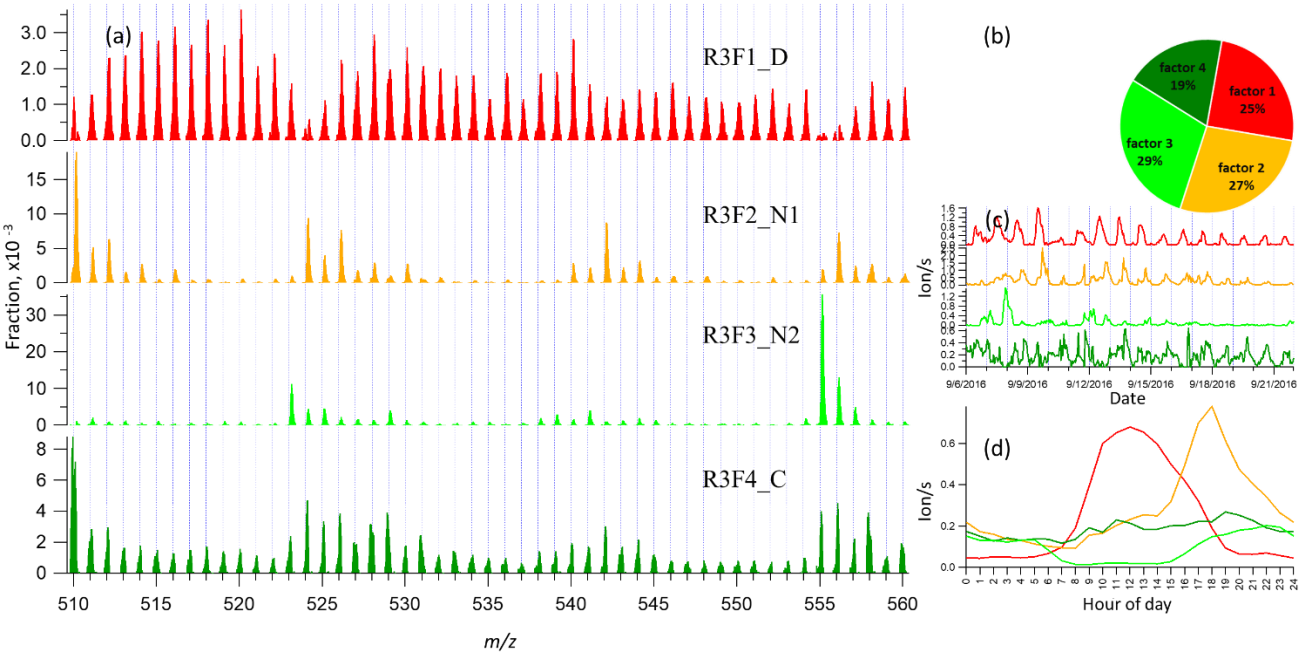

Figure 4 Four-factor result for Range 3, for (a) factor spectral profiles, (b) averaged factor contribution during the campaign, (c) time series and (d) diurnal trend. Details on the factors' naming schemes are shown in Table 1.

3.5 binPMF on Range Combined (250-350 Th & 510-560 Th)

As comparison to the previous three ranges, we conducted the binPMF analysis on Range Combined, which is the combination of the three ranges. The results of this range are fairly similar to those of Ranges 1 and 2, as could be expected since the signal intensities in these ranges were much higher than in Range 3. As the number of factors increased (2-10), both the Q/Q_{exp} (1.3 to 0.6) and unexplained variation (16% to 8%) showed a decreasing trend.

In the two-factor result, one daytime factor and one nighttime factor were separated. In the nighttime factor, most masses were found at even masses, and the fraction of masses in Range 3 was much higher than that in daytime factor. In contrast, in the daytime factor, most masses were observed at odd masses and the fraction of signal in Range 3 was much lower. During the day, photochemical reactions as well as potential emissions increase the concentration of NO, which serves as peroxy radical (RO_2) terminator and often outcompetes RO_2 cross reactions in which dimers can be formed (Ehn et al., 2014). Thus, the production of dimers is suppressed during the day, yielding instead a larger fraction of organic nitrates, as has been shown also previously (Yan et al., 2016).

With the increase of the number of factors, more daytime factors were resolved with different peak times. When the factor number reached seven, a clear sawtooth-shape diurnal cycle occurred, i.e. the contamination factor, caused by the zeroing. As more factors were added, no further nighttime factors were separated, and only more daytime factors appeared. To simplify the discussion and inter-range comparison, we also here chose the four-factor result for further analysis. Figure 5 shows the four-factor result of Range Combined, with spectral profile, time series, diurnal cycle and averaged factor contribution during the campaign. The signals in range of 510-560 Th were enlarged 100-fold to be visible.

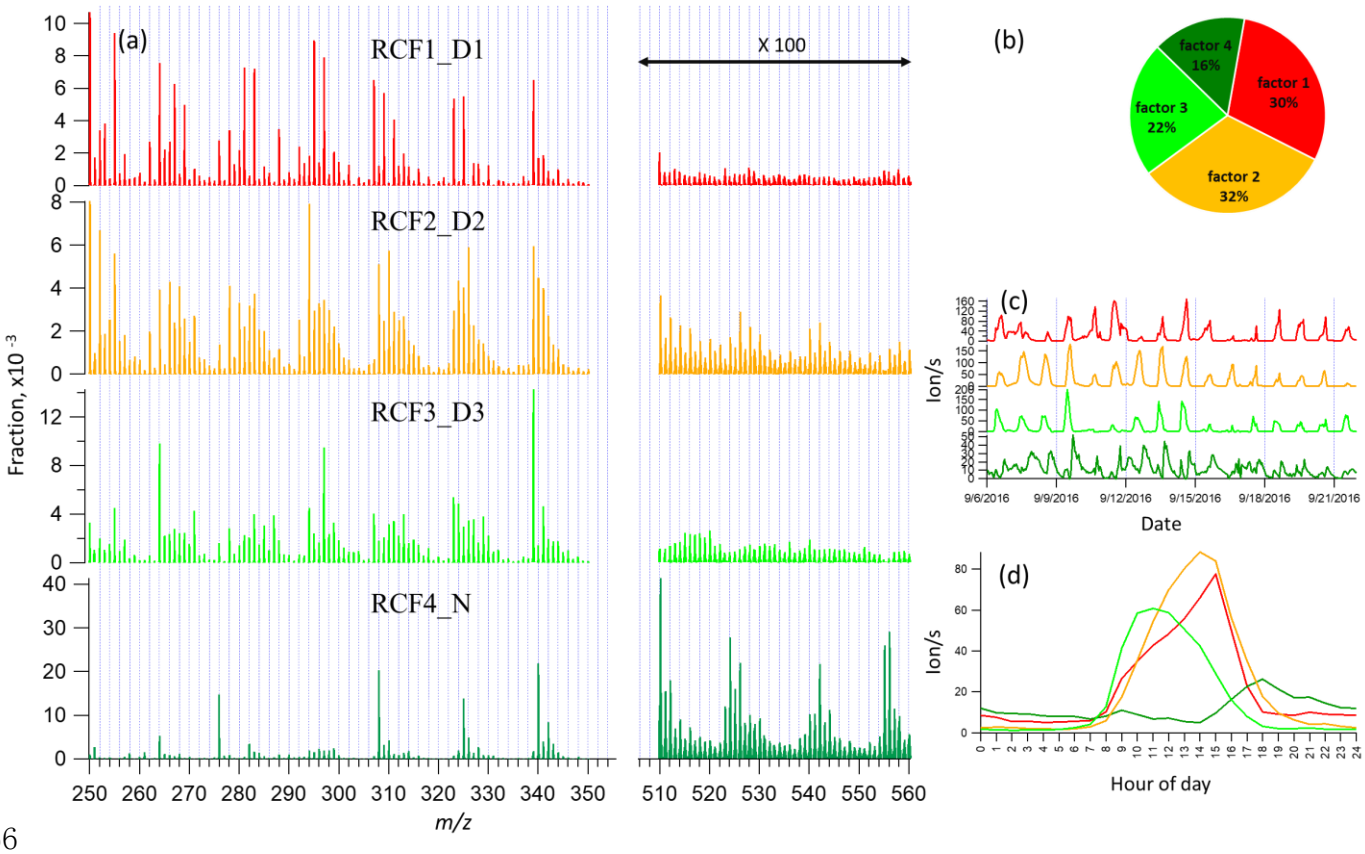

Figure 5 Four-factor result for Range Combined, for (a) factor spectral profiles, (b) averaged factor contribution during the campaign, (c) time series and (d) diurnal trend. Details on the factors' naming schemes are shown in Table 1.

4 Discussion

In Section 3, results by binPMF analysis were shown for Ranges 1, 2, 3 and Combined. In this section, we discuss and compare the results from the different ranges. To simplify the inter-range comparison, we chose four-factor results for all the four ranges, with the abbreviations shown in Table 1. From Range 1, three daytime factors and a contaminations factor were separated. In Range 2, three daytime factors and one nighttime factor (abbreviated as R2F4_N) were resolved. The R2F4_N factor was characterized by signals at 308 Th ($C_{10}H_{14}O_7 \cdot NO_3^-$), 325 Th ($C_{10}H_{15}O_8 \cdot NO_3^-$), and 340 Th ($C_{10}H_{14}O_9 \cdot NO_3^-$), and can be confirmed as monoterpene ozonolysis products (Ehn et al., 2014;Yan et al., 2016). With the increase of factor number to six, the contamination factor got separated also in this mass range. In Range 3, one daytime factor, two nighttime factors and a contamination factor were separated. The first nighttime factor (R3F2_N1) had large peaks at 510 Th ($C_{20}H_{32}O_{11} \cdot NO_3^-$) and 556 Th ($C_{20}H_{30}O_{14} \cdot NO_3^-$), dimer products that have been identified during chamber studies of monoterpene ozonolysis (Ehn et al., 2014). The molecule observed at 510 Th has 32 H-atoms, suggesting that one of the RO_2 involved would have been initiated by OH, which is formed during the ozonolysis of alkenes such as monoterpenes at nighttime (Atkinson et al., 1992;Paulson and Orlando, 1996). The other nighttime factor (R3F3_N2) was dominated by ions at 523 Th ($C_{20}H_{31}O_8NO_3 \cdot NO_3^-$) and 555 Th ($C_{20}H_{31}O_{10}NO_3 \cdot NO_3^-$), representing nighttime monoterpene oxidation involving NO_3. As these dimers contain only one N-atom, and 31 H-atoms, we can assume that they are formed from reactions between an RO_2 formed from NO_3 oxidation and another RO_2 formed by ozone oxidation. These results match well with the profiles in a previous study by Yan et al. (2016). The results of Range Combined are very similar to Range 2, with one nighttime factor and three daytime factors. The contamination factor was separated with increase of factor number to seven.

Table 1. Summary of PMF results for the different mass ranges

Range	Factor number	Factor name[a]	Dominant peaks	Peak time
1 (250 - 300 Th)	1	R1F1_D1	250, 255, 295, 297	15:00
	2	R1F2_D2	250, 252, 294	15:00
	3	R1F3_D3	264, 297	11:00
	4	R1F4_C	276	-[b]
2 (300 - 350 Th)	1	R2F1_D1	307, 309, 323, 325, 339,	15:00
	2	R2F2_D2	310, 326, 339,	14:00
	3	R2F3_D3	339	11:00
	4	R2F4_N	308, 325, 340	18:00

3 (510 – 560 Th)	1	R3F1_D	516, 518, 520, 528, 540	12:00
	2	R3F2_N1	510, 524, 542, 556	18:00
	3	R3F3_N2	523, 555	22:00
	4	R3F4_C	510, 558	-[b]
Combined (1, 2, 3)	1	RCF1_D1	250, 255, 295, 339	15:00
	2	RCF2_D2	250, 252, 294, 339	14:00
	3	RCF3_D3	264, 297, 339	11:00
	4	RCF4_N	308, 340, 510, 524, 555, 556	18:00

[a] Factor name is defined with range name, factor number and name. For example, RxFy represents Factor y in Range x. RC stands for Range Combined. For the factor name, D is short for daytime, N for Nighttime, C for contamination.
[b] The contamination factor in Range 1 shows sawtooth pattern; while in Range 3 shows no diurnal pattern.

4.1 Time series correlation

In Figure 6, the upper panels show the time series correlations among the first three ranges. As expected based on the results above, generally the daytime factors, and the two nighttime monoterpene ozonolysis factors (R2F4_N and R3F2_N1) correlated well, respectively. However, the contamination factors did now show strong correlation between different ranges, even though undoubtedly from the same source. More about the contamination factors will be discussed in Section 4.4. The lower panels in Figure 6 displays the correlations between the first three ranges and the Range Combined, and clearly demonstrates that the results of Range Combined is mainly controlled by high signals from Range 1 and 2. More detailed aspects of the comparison between factors in different ranges is given in the following sections. The good agreements between factors from different subranges also help to verify the robustness of the solutions.

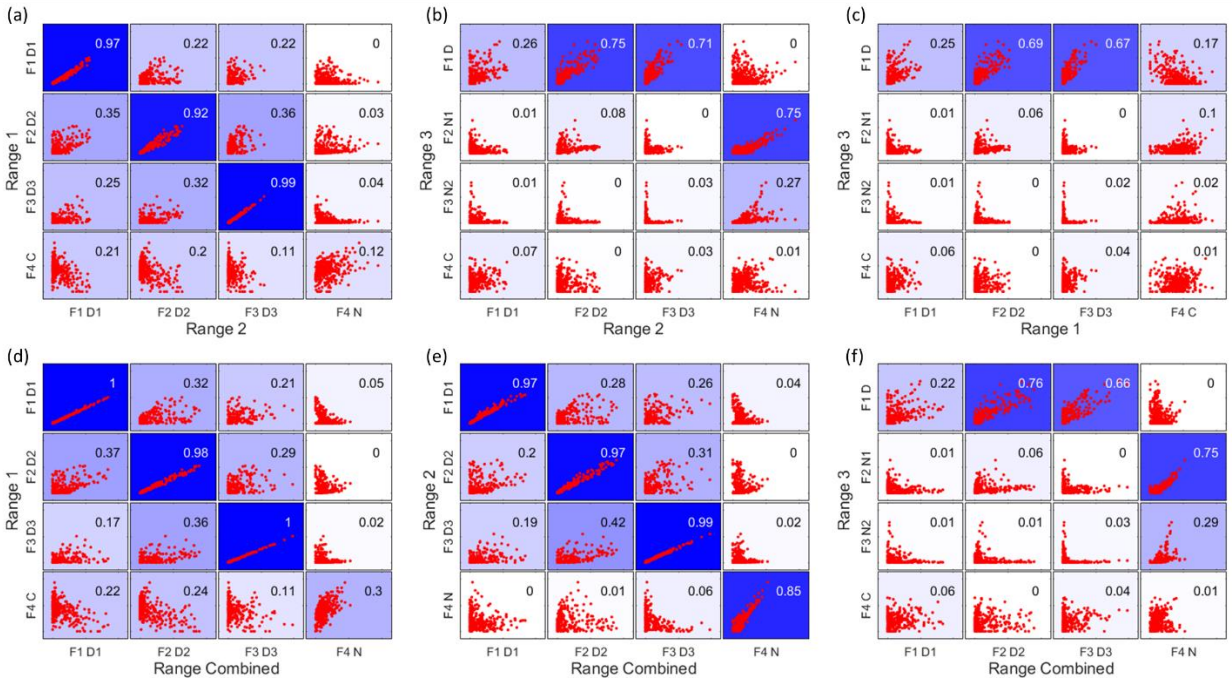

Figure 6 Time series correlations among Range 1, 2, 3 (upper panels a-c), and between the first three ranges and the Range Combined (lower panels d-f). The abbreviations for different factors are the same in Table 1, with F for factor, D for daytime, N for nighttime and C for contamination, e.g. F1D1 for Factor 1 daytime 1.

The coefficient of determination, R^2, is marked in each subplot by a number shown in the right upper corners and by the blue colors, with stronger blue indicating higher R^2.

4.2 Daytime processes

4.2.1 Factor comparison

As mentioned above, with increasing number of factors, usually more daytime factors will be resolved, reflecting the complicated daytime photochemistry. The three daytime factors between Range 1 and 2 agreed with each other quite well (Figure 6a). However, R1F1_D1 and R2F1_D1 did not show strong correlation with the only daytime factor in Range 3 (R3F1_D), while the other two daytime factors in both Range 1 and 2, i.e. R1F2_D2, R1F3_D3, R2F2_D2, R2F3_D3, correlated well with R3F1_D from Range 3.

The 1st daytime factors from Range 1 and 2, R1F1_D1 and R2F1_D1, were mainly characterized by odd masses 255 Th, 281 Th, 283 Th, 295 Th, 297 Th, 307 Th, 309 Th, 311 Th, 323 Th, 325 Th, 339 Th. The factors are dominated by organonitrates. Organic nitrate formation during daytime is generally associated with the termination of RO_2 radicals by NO. This termination step is mutually exclusive with the termination of RO_2 with other RO_2, which can lead to dimer formation. If the NO concentration is the limiting factor for the formation of these factors, the low correlations between the NO-terminated monomer factors, and the dimer factors, is to be expected. In contrast, if the other daytime factors mainly depend on oxidant and monoterpene concentrations, some correlation between those, and the daytime dimer factor, is to be expected, as shown in Figure 6b, c.

All the spectral profiles resolved from Range Combined binPMF analysis inevitably contained mass contribution from 510 – 560 Th, even the daytime factor from Range Combined (RCF1_D1) which did not show clear correlation with R3F1_D from Range 3 (Figure 6e).

The 2nd and 3rd daytime factors in Range 1 and 2, R1F2_D2, R1F3_D3, R2F2_D2, R2F3_D3, had high correlations with R3F1_D in Range 3. Daytime factors in Range Combined (RCF2_D2 and RCF3_D3) also showed good correlation with R3F1_D in Range 3. However, if we compare R3F1_D and the mass range of 510 – 560 Th of the daytime factors in Range Combined, just with a quick look, we can readily see the difference. The daytime factor separated in Range 3 (R3F1_D) has no obvious markers in the profile. With the increase of factor number (up to ten factors), no clearly new factors were separated in Range 3, but instead the previously separated factors were seen to split into several factors. However, the spectral pattern in R3F1_D is different from that in the mass range of 510 – 560 Th in RCF2_D2. The factorization of Range Combined was mainly controlled by low masses due to their high signals. The signals at high masses were forced to be distributed according to the

time series determined by small masses. Ultimately, this will lead to failure in factor separation for this low-signal range.

4.2.2 Daytime dimer formation

Dimers are primarily produced during nighttime, due to NO suppressing RO_2 + RO_2 reactions in daytime (Ehn et al., 2014;Yan et al., 2016). However, in this study, we found one clear daytime factor in Range 3 (R3F1_D, peak at local time 12:00, UTC+2) by sub-range analysis. With high loadings from even masses including 516, 518, 520, 528, 540 Th, this only daytime factor in dimer range correlated very well with two daytime factors in Ranges 1 and 2 (R1F2_D2, R1F3_D3, R2F2_D2, R2F3_D3) (Figure 6b and c). Table 2 include the correlation matrix of all PMF and factors and selected meteorological parameters. Strong correlation between R3F1_D with solar radiation was found, with R = 0.79 (Table 2). This may indicate involvement of OH oxidation in producing this factor.

Table 2 Correlation between factors and meteorological parameters and gases

	R1F1_D1	R1F1_D2	R1F1_D3	R1F1_C	R2F1_D1	R2F2_D2	R2F3_D3	R2F4_N	R3F1_D	R3F2_N1	R3F3_N2	R3F4_C	RCF1_D1	RCF2_D2	RCF3_D3	RCF4_N
O_3	0.51	0.59	0.35	-0.18	0.47	0.57	0.36	0.43	0.55	0.33	0.27	0.22	0.49	0.57	0.33	0.34
NO	0.13	-0.01	0.24	-0.03	0.18	-0.02	0.24	-0.22	0.13	-0.19	-0.17	0.03	0.13	0.00	0.26	-0.18
NOx	-0.05	-0.22	-0.10	0.09	-0.01	-0.23	-0.11	-0.13	-0.16	-0.21	-0.04	0.04	-0.04	-0.22	-0.09	-0.11
RH	-0.46	-0.80	-0.63	0.30	-0.43	-0.82	-0.64	-0.27	-0.78	-0.39	-0.07	-0.07	-0.43	-0.82	-0.60	-0.21
T	0.66	0.72	0.40	-0.24	0.65	0.66	0.41	0.39	0.65	0.30	0.14	0.19	0.66	0.68	0.38	0.24
UVB	0.52	0.63	0.82	-0.40	0.52	0.68	0.84	-0.30	0.79	-0.08	-0.27	0.08	0.49	0.68	0.83	-0.29

As previous studies have shown, dimers greatly facilitate new particle formation (NPF) (Kirkby et al., 2016;Troestl et al., 2016;Lehtipalo et al., 2018), and this daytime dimer factor may represent a source of dimers that would impact the initial stages of NPF in Hyytiälä. Mohr et al. (2017) reported a clear diel pattern of dimers (sum of about 60 dimeric compounds of $C_{16\text{-}20}H_{13\text{-}33}O_{6\text{-}9}$) during NPF events in 2013 in Hyytiälä, with minimum at night and maximum after noon, and estimated these dimers can contribute ~5% of the mass of sub-60 nm particles. The link between the dimers presented in that paper and those reported here will require further studies, as will the proper quantification of the dimer factor identified here.

4.3 Nighttime processes

4.3.1 Factor comparison

Since high-mass dimers are more likely to form at night due to photochemical production of NO in daytime, which inhibits RO_2 + RO_2 reactions, Range 3 had the highest fraction of nighttime signals

of all the sub-ranges. While Range 3 produced two nighttime factors, Ranges 2 and Combined showed one, and Range 1 had no nighttime factor. The difference between the two results also indicates the advantage of analyzing monomers and dimers separately.

The two nighttime factors in Range 3 can be clearly identified as arising from ozonolysis (R3F2_N1) and a mix of ozonolysis and NO_3 oxidation (R3F2_N2) based on the mass spectral profiles, as described above. The organonitrate at 555 Th, $C_{20}H_{31}O_{10}NO_3 \cdot NO_3^-$, is a typical marker for NO_3 radical initiated monoterpene chemistry (Yan et al., 2016). However, several interesting features become evident when comparing to the results of Range 2 and Combined. Firstly, only one nighttime factor (R2F4_N, RCF4_N) was separated in each of these ranges, and that shows clear resemblance with ozonolysis of monoterpenes as measured in numerous studies, e.g. Ehn et al. (2012);(2014). Secondly, the high correlation found in Figure 6b between the ozonolysis factors (i.e., R2F4_N, R3F2_N1, RCF4_N), further supports the assignment. However, factor R2F4_N is the only nighttime factor in the monomer range, suggesting that NO_3 radical chemistry of monoterpenes in Hyytiälä does not form substantial amounts of HOM monomers. The only way for the CI-APi-TOF to detect products of monoterpene-NO_3 radical chemistry may thus be through the dimers, where one highly oxygenated RO_2 radical from ozonolysis reacts with a less oxygenated RO_2 radical from NO_3 oxidation.

In the results by Yan et al. (2016) the combined UMR-PMF of monomers and dimers did yield a considerable amount of compounds in the monomer range also for the NO_3 radical chemistry factor. There may be several reasons for this discrepancy. One major cause for differences between the spring dataset of Yan et al. (2016) and the autumn dataset presented here, is that nighttime concentrations of HOM was greatly reduced during our autumn campaign. The cause may have been fairly frequent fog formation during nights, and also the concentration of e.g. ozone decreased nearly to zero during several nights (Zha et al., 2018). It is also possible that the NO_3 radical-related factor by Yan et al. (2016) is probably a mixture of NO_3 and O_3 radical chemistry, while the monomer may thus be attributed to the O_3 part. Alternatively, the different conditions during the two measurement periods, as well as seasonal difference in monoterpene mixtures (Hakola et al., 2012), caused variations in the oxidation pathways.

4.3.2 Dimers initiated by NO_3 radicals

Previous studies show that NO_3 oxidation of α-pinene, the most abundant monoterpene in Hyytiälä (Hakola et al., 2012), produces fairly little SOA mass (yields <4 %), while β-pinene shows yields of up to 53 % (Bonn and Moorgat, 2002;Nah et al., 2016). The NO_3+β-pinene reaction results in low volatile organic nitrate compounds with carboxylic acid, alcohol, and peroxide functional groups (Fry

et al., 2014;Boyd et al., 2015), while NO_3+α-pinene reaction will typically lose the nitrate functional group and form oxidation products with high vapor pressures (Spittler et al., 2006;Perraud et al., 2010). Most monoterpene-derived HOM, including monomers, are low-volatile (Peräkylä et al., 2020) and thus a low SOA yield indicates a low HOM yield. Thus, while there are to our knowledge no laboratory studies on HOM formation from NO_3 oxidation of α-pinene, a low yield can be expected based on SOA studies.

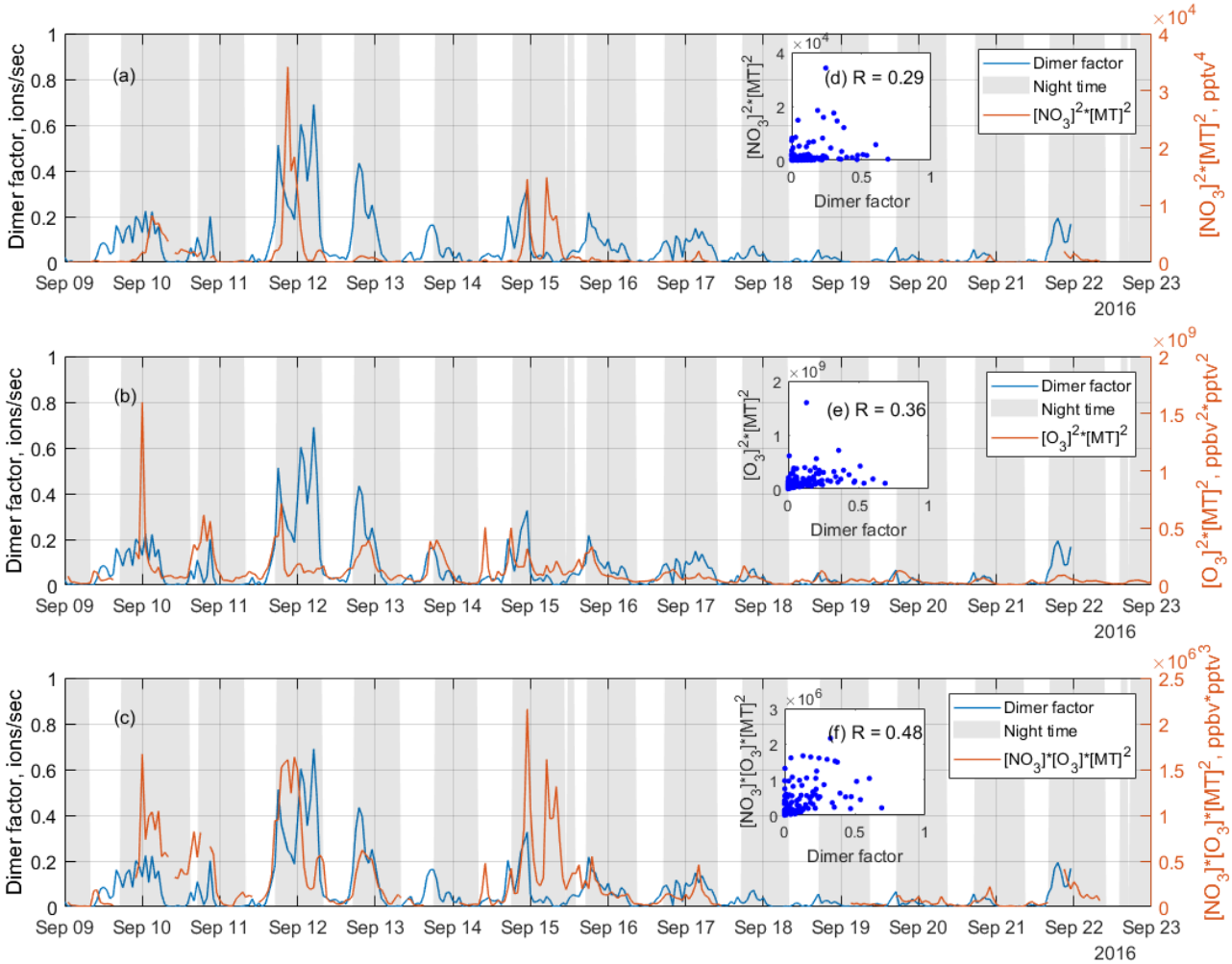

Figure 7 Time series of the NO_3 oxidation dimer factor (blue line), and the product of (a) $[NO_3]^2$ × [monoterpene]^2, (b) $[O_3]^2$ × [monoterpene]^2, and (c) $[NO_3]$ × $[O_3]$ × [monoterpene]^2, where [] represents concentration in unit of pptv for NO_3 radicals and monoterpene, ppbv for O_3, while the scatter plots are shown as inserts, (d), (e), (f), respectively. The scatter plots and correlation coefficients R are only calculated from nighttime data, which is selected based on solar radiation, to eliminate the influence from daytime oxidation processes.

As discussed above, a dimer factor (R3F2_N2) was identified as being a crossover between NO_3 radical initiated and O_3 initiated RO_2 radicals. Figure 7 shows the time series of this factor, as well as the product of $[NO_3]^2$ × [monoterpene]^2, $[O_3]^2$ × [monoterpene]^2, and $[NO_3]$ × $[O_3]$ × [monoterpene]^2. These products are used to mimic the formation rates of the RO_2 radicals reacting to form the dimers,

either from pure NO_3 oxidation (Fig. 7a), pure O_3 oxidation (7b), or the mixed reaction between RO_2 from the two oxidants (7c). The NO_3 concentration was estimated in Liebmann et al. (2018) for the same campaign. Monoterpenes were measured using a proton transfer reaction time of flight mass spectrometer (PTR-TOF-MS). More details on measurement of NO_3 proxy and monoterpene can be found in in Liebmann et al. (2018).

As shown in Figure 7, the time series of the dimer factor tracks those of [NO_3] × [monoterpene] and [O_3] × [monoterpene] reasonably well, but shows the highest correlation with the product of [NO_3] × [O_3] × [monoterpene]^2. This further supports this dimer formation as a mixed processes of ozonolysis and NO_3 oxidation. The heterogeneity of the monoterpene emissions in the forest, and the fact that no dimer loss process is included, partly explain the relatively low correlation coefficients. The sampling inlets for PTR-TOF were about 170 m away from the NO_3 reactivity measurement (Liebmann et al., 2018), which in turn was some tens of meters away from the HOM measurements. Thus, this analysis should be considered qualitative only.

The nitrate dimer factor (R3F2_N2) was dominated by the organonitrate at 555 Th, $C_{20}H_{31}O_{10}NO_3 \cdot NO_3^-$. However, unlike the pure ozonolysis dimer factor which had a corresponding monomer factor (R = 0.86 between factor R2F4_N and R3 F2_N1), this NO_3-related dimer factor did not have an equivalent monomer factor. This suggests that the NO_3 oxidation of the monoterpene mixture in Hyytiälä does not by itself form much HOM, but in the presence of RO_2 from ozonolysis, the RO_2 from NO_3 oxidation can take part in HOM dimer formation. This further implies that, different from previous knowledge based on single-oxidant experiments in chambers, NO_3 oxidation may have a larger impact on SOA formation in the atmosphere where different oxidants exist concurrently. This highlights the need for future laboratory studies to consider systems with multiple oxidants during monoterpene oxidation experiments, to truly understand the role and contribution of different oxidants, and NO_3 in particular.

4.4 Fluorinated compounds

During the campaign, an automated instrument zeroing every three hours was conducted. While the zeroing successfully removed the low-volatile HOM and H_2SO_4, the process also introduced contaminants into the inlet lines, e.g. perfluorinated organic acids from Teflon tubing. Each zeroing process lasted for 10 min. In the data analysis, we removed all the 10-min zeroing periods, and averaged the data to 1-h time resolution, but contaminants were still identified in all ranges by binPMF. However, the correlation between contamination factors from different ranges is low (Figure 6c).

To future investigate the low factor correlations of the same source, three fluorinated compounds with different volatilities, $(CF_2)_3CO_2HF{\cdot}NO_3^-$ (275.9748 Th), $(CF_2)_5C_2O_4H^-$ (338.9721 Th), and $(CF_2)_6CO_2HF{\cdot}NO_3^-$ (425.9653 Th), were examined in fine time resolution, i.e. 1 min. The time series and 3-h cycle of the three fluorinated compounds were shown in Figure S3 and S4 in Supplement. The correlation coefficients dropped greatly before and after the zero period was removed, from 0.9 to 0.3 for R^2 between 276 Th and 339 Th, and 0.8 to 0.1 between 276 Th and 426 Th (Fig. S5a, b). Similar effect is also found with the 1 h averaged data (Fig. S5c, d). It is evident that the three fluorinated compounds were from the same source (zeroing process), but due to their different volatilities, they were lost at different rates. This, in turn, means that the spectral signature of this source will change as a function of time, at odds with one of the basic assumptions of PMF.

The analysis of the fluorinated compounds in our system was here merely used as an example to show that volatility can impact source profiles over time. In Figure S5, it can be clearly seen that the profile of Range Combined is noisier than that of Range 3, probably due to the varied fractional contributions of contamination compounds to the profile. In ambient data, products from different sources can have undergone atmospheric processing, altering the product distribution. This analysis highlighted the importance of differences in the sink terms due to different volatilities of the products. This may be an important issue for gas phase mass spectrometry analysis, potentially underestimated by many PMF users, as it is likely only a minor issue for aerosol data, for which PMF has been applied much more routinely. If failing to achieve physically meaningful factors using PMF on gas phase mass spectra, our recommendation is to try applying PMF to sub-ranges of the spectrum, where IVOC, SVOC and (E)LVOC could be analyzed separately.

4.5 Atmospheric insights

Based on the new data analysis technique binPMF applied in sub-ranges of mass spectra, we were able to separate two particularly intriguing atmospheric processes, the formation of daytime dimers as well as dimer formation involving NO_3 radicals, which otherwise could not have been identified in our study.

With a diurnal peak around noon time, the daytime dimers identified in this study correlate very well with daytime factors in monomer range. Strong correlation between this factor and solar radiation indicate the potential role of OH oxidation in the formation of daytime dimers. By now, very few studies have reported the observations of daytime dimers. As dimers are shown to be able to take part in new particle formation (NPF) (Kirkby et al., 2016), this daytime dimer may contribute to the early stages of NPF in the boreal forest.

The second process identified in our study is the formation of dimers that are a crossover between NO_3 and O_3 oxidation. Such dimers have been identified before (Yan et al., 2016). However, we were not able to identify corresponding HOM monomer compounds. This finding indicates that while NO_3 oxidation of the monoterpenes in Hyytiälä may not undergo autoxidation to form HOM by themselves, they can contribute to HOM dimers when the NO_3-derived RO_2 react with highly oxygenated RO_2 from other oxidants. Multi-oxidant systems should be taken into consideration in future experimental studies on monoterpene oxidation processes.

5 Conclusions

The recent developments in the field of mass spectrometry, combined with factor analysis techniques such as PMF, have greatly improved our understanding of complicated atmospheric processes and sources. In this study, we applied the new binPMF approach (Zhang et al., 2019), to separate sub-ranges of mass spectra measured using a chemical ionization mass spectrometer in the Finnish boreal forest. By using this method, we were able to identify a daytime dimer factor, presumably initiated by OH/O_3 oxidation of monoterpenes, forming from RO_2+RO_2 reactions despite competition from daytime NO. This compound group, showing a diurnal peak around noon, may contribute to new particle formation at the site. In addition, we successfully separated NO_3–related dimers which would not have been identified from this dataset without utilizing the different sub-ranges. The NO_3-related factor was consistent with earlier observations (Yan et al., 2016), with the exception that we did not observe any corresponding monomer factor. This may be explained by the observed nitrate-containing dimers being formed from two RO_2, where one is initiated by oxidation by O_3, and the other by NO_3. If the NO_3-derived RO_2 are not able to form HOM by themselves, there will not be any related monomers observed. To validate this hypothesis, future laboratory experiments that target more complex oxidation systems will be useful in order to understand the role of NO_3 oxidation in SOA formation under different atmospheric conditions.

Apart from these two major findings, we also found several other benefits of applying PMF on separate sub-ranges of the mass spectra. First, different compounds from the same source can have variable loss rates due to differences in volatilities. This leads to increased difficulty for PMF to separate this source, but if the PMF analysis is run separately on lighter masses (with higher volatility) and heavier masses (with lower volatility), the source may become easier to distinguish. Secondly, chemistry or sources contributing only to one particular mass range, e.g. dimers, can be better separated. Thirdly, mass ranges with small, but informative, signals can be more accurately assigned as their contribution becomes larger than if the entire mass range was analyzed at once. Finally, running PMF on separate mass ranges also allows comparing the factors between the different ranges,

helping to verify the results. In summary, while we do not suggest that this type of sub-range analysis should always be utilized, we recommend other analysts of gas-phase mass spectrometer data to test this approach in order to see whether additional useful information can be obtained. In our dataset, this method was crucial for identifying different types of dimers and dimer formation pathways, which are of great importance for the formation of both new particles and SOA.

Data availability. The data used in this study are available from the first author upon request: please contact Yanjun Zhang (yanjun.zhang@helsinki.fi).

Author contribution. ME and YZ designed the study. QZ and MR collected the data; data analysis and manuscript writing were done by YZ. All coauthors discussed the results and commented the manuscript.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements. We thank the tofTools team for providing tools for mass spectrometry data analysis. The personnel of the Hyytiälä forestry field station are acknowledged for help during field measurements.

Financial support. This research was supported by the European Research Council (Grant 638703-COALA), the Academy of Finland (grants 317380 and 320094), and the Vilho, Yrjö and Kalle Väisälä Foundation. K.R.D. acknowledges support by the Swiss National Science postdoc mobility grant P2EZP2_181599.

Reference

Allan, J. D., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer 1. Techniques of data interpretation and error analysis, Journal of Geophysical Research: Atmospheres, 108, 2003.

Atkinson, R., Aschmann, S. M., Arey, J., and Shorees, B.: Formation of OH radicals in the gas phase reactions of O3 with a series of terpenes, 97, 6065-6073, 10.1029/92jd00062, 1992.

Berndt, T., Mentler, B., Scholz, W., Fischer, L., Herrmann, H., Kulmala, M., and Hansel, A.: Accretion Product Formation from Ozonolysis and OH Radical Reaction of α-Pinene: Mechanistic Insight and the Influence of Isoprene and Ethylene, Environmental Science & Technology, 52, 11069-11077, 10.1021/acs.est.8b02210, 2018a.

Berndt, T., Scholz, W., Mentler, B., Fischer, L., Herrmann, H., Kulmala, M., and Hansel, A.: Accretion Product Formation from Self- and Cross-Reactions of RO2 Radicals in the Atmosphere, Angewandte Chemie International Edition in English 57, 3820-3824, 10.1002/anie.201710989, 2018b.

Bertram, T. H., Kimmel, J. R., Crisp, T. A., Ryder, O. S., Yatavelli, R. L. N., Thornton, J. A., Cubison, M. J., Gonin, M., and Worsnop, D. R.: A field-deployable, chemical ionization time-of-flight mass spectrometer, Atmospheric Measurement Techniques, 4, 1471-1479, 10.5194/amt-4-1471-2011, 2011.

Bianchi, F., Kurtén, T., Riva, M., Mohr, C., Rissanen, M. P., Roldin, P., Berndt, T., Crounse, J. D., Wennberg, P. O., Mentel, T. F., Wildt, J., Junninen, H., Jokinen, T., Kulmala, M., Worsnop, D. R., Thornton, J. A., Donahue, N., Kjaergaard, H. G., and Ehn, M.: Highly Oxygenated Organic Molecules (HOM) from Gas-Phase Autoxidation Involving Peroxy Radicals: A Key Contributor to Atmospheric Aerosol, Chemical Reviews, 119, 3472-3509, 10.1021/acs.chemrev.8b00395, 2019.

Bonn, B., and Moorgat, G. K.: New particle formation during a- and b-pinene oxidation by O3, OH and NO3, and the influence of water vapour: particle size distribution studies, Atmos. Chem. Phys., 2, 183-196, 10.5194/acp-2-183-2002, 2002.

Boyd, C. M., Sanchez, J., Xu, L., Eugene, A. J., Nah, T., Tuet, W. Y., Guzman, M. I., and Ng, N. L.: Secondary organic aerosol formation from the β-pinene+NO3 system: effect of humidity and peroxy radical fate, Atmos. Chem. Phys., 15, 7497-7522, 10.5194/acp-15-7497-2015, 2015.

Canagaratna, M., Jayne, J., Jimenez, J., Allan, J., Alfarra, M., Zhang, Q., Onasch, T., Drewnick, F., Coe, H., and Middlebrook, A.: Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer, Mass Spectrometry Reviews, 26, 185-222, 2007.

Canonaco, F., Crippa, M., Slowik, J., Baltensperger, U., and Prévôt, A.: SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data, Atmospheric Measurement Techniques, 6, 3649, 2013.

Craven, J. S., Yee, L. D., Ng, N. L., Canagaratna, M. R., Loza, C. L., Schilling, K. A., Yatavelli, R. L. N., Thornton, J. A., Ziemann, P. J., Flagan, R. C., and Seinfeld, J. H.: Analysis of secondary organic aerosol formation and aging using positive matrix factorization of high-resolution aerosol mass spectra: application to the dodecane low-NO_x system, Atmos. Chem. Phys., 12, 11795-11817, 10.5194/acp-12-11795-2012, 2012.

Ehn, M., Kleist, E., Junninen, H., Petäjä, T., Lönn, G., Schobesberger, S., Dal Maso, M., Trimborn, A., Kulmala, M., Worsnop, D. R., Wahner, A., Wildt, J., and Mentel, T. F.: Gas phase formation of extremely oxidized pinene reaction products in chamber and ambient air, Atmos. Chem. Phys., 12, 5113-5127, 10.5194/acp-12-5113-2012, 2012.

Ehn, M., Thornton, J. A., Kleist, E., Sipila, M., Junninen, H., Pullinen, I., Springer, M., Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I.-H., Rissanen, M., Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurten, T., Nielsen, L. B., Jorgensen, S., Kjaergaard, H. G., Canagaratna, M., Dal Maso, M., Berndt, T., Petaja, T., Wahner, A., Kerminen, V.-M., Kulmala, M., Worsnop, D. R., Wildt, J., and Mentel, T. F.: A large source of low-volatility secondary organic aerosol, Nature, 506, 476-479, 10.1038/nature13032, 2014.

Fry, J. L., Draper, D. C., Barsanti, K. C., Smith, J. N., Ortega, J., Winkler, P. M., Lawler, M. J., Brown, S. S., Edwards, P. M., Cohen, R. C., and Lee, L.: Secondary Organic Aerosol Formation and Organic Nitrate Yield from NO3 Oxidation of Biogenic Hydrocarbons, Environmental Science & Technology, 48, 11944-11953, 10.1021/es502204x, 2014.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P.: A GLOBAL-MODEL OF NATURAL VOLATILE ORGANIC-COMPOUND EMISSIONS, Journal of Geophysical Research-Atmospheres, 100, 8873-8892, 10.1029/94jd02950, 1995.

Hakola, H., Tarvainen, V., Bäck, J., Ranta, H., Bonn, B., Rinne, J., and Kulmala, M.: Seasonal variation of mono- and sesquiterpene emission rates of Scots pine, Biogeosciences, 3, 93-101, 10.5194/bg-3-93-2006, 2006.

Hakola, H., Hellén, H., Hemmilä, M., Rinne, J., and Kulmala, M.: In situ measurements of volatile organic compounds in a boreal forest, Atmos. Chem. Phys., 12, 11665-11678, 10.5194/acp-12-11665-2012, 2012.

Hari, P., and Kulmala, M.: Station for Measuring Ecosystem–Atmosphere Relations (SMEAR II), Boreal Environment Research, 10, 315-322, 2005.

Huang, S., Rahn, K. A., and Arimoto, R.: Testing and optimizing two factor-analysis techniques on aerosol at Narragansett, Rhode Island, Atmospheric Environment, 33, 2169-2185, https://doi.org/10.1016/S1352-2310(98)00324-0, 1999.

Jokinen, T., Sipilä, M., Junninen, H., Ehn, M., Lönn, G., Hakala, J., Petäjä, T., Mauldin Iii, R. L., Kulmala, M., and Worsnop, D. R.: Atmospheric sulphuric acid and neutral cluster measurements using CI-APi-TOF, Atmospheric Chemistry and Physics, 12, 4117-4125, 10.5194/acp-12-4117-2012, 2012.

Kirkby, J., Duplissy, J., Sengupta, K., Frege, C., Gordon, H., Williamson, C., Heinritzi, M., Simon, M., Yan, C., Almeida, J., Troestl, J., Nieminen, T., Ortega, I. K., Wagner, R., Adamov, A., Amorim, A., Bernhammer, A.-K., Bianchi, F., Breitenlechner, M., Brilke, S., Chen, X., Craven, J., Dias, A., Ehrhart, S., Flagan, R. C., Franchin, A., Fuchs, C., Guida, R., Hakala, J., Hoyle, C. R., Jokinen, T., Junninen, H., Kangasluoma, J., Kim, J., Krapf, M., Kuerten, A., Laaksonen, A., Lehtipalo, K., Makhmutov, V., Mathot, S., Molteni, U., Onnela, A., Peraekylae, O., Piel, F., Petaejae, T., Praplan, A. P., Pringle, K.,

Rap, A., Richards, N. A. D., Riipinen, I., Rissanen, M. P., Rondo, L., Sarnela, N., Schobesberger, S., Scott, C. E., Seinfeld, J. H., Sipilae, M., Steiner, G., Stozhkov, Y., Stratmann, F., Tome, A., Virtanen, A., Vogel, A. L., Wagner, A. C., Wagner, P. E., Weingartner, E., Wimmer, D., Winkler, P. M., Ye, P., Zhang, X., Hansel, A., Dommen, J., Donahue, N. M., Worsnop, D. R., Baltensperger, U., Kulmala, M., Carslaw, K. S., and Curtius, J.: Ion-induced nucleation of pure biogenic particles, Nature, 533, 521-526, 10.1038/nature17953, 2016.

Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H. E., Nieminen, T., Petäjä, T., Sipilä, M., Schobesberger, S., Rantala, P., Franchin, A., Jokinen, T., Järvinen, E., Äijälä, M., Kangasluoma, J., Hakala, J., Aalto, P. P., Paasonen, P., Mikkilä, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., Ruuskanen, T., Mauldin, R. L., Duplissy, J., Vehkamäki, H., Bäck, J., Kortelainen, A., Riipinen, I., Kurtén, T., Johnston, M. V., Smith, J. N., Ehn, M., Mentel, T. F., Lehtinen, K. E. J., Laaksonen, A., Kerminen, V.-M., and Worsnop, D. R.: Direct Observations of Atmospheric Aerosol Nucleation, 339, 943-946, 10.1126/science.1227385 %J Science, 2013.

Lamarque, J. F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmos. Chem. Phys., 10, 7017-7039, 10.5194/acp-10-7017-2010, 2010.

Lee, B. H., Lopez-Hilfiker, F. D., Mohr, C., Kurtén, T., Worsnop, D. R., and Thornton, J. A.: An Iodide-Adduct High-Resolution Time-of-Flight Chemical-Ionization Mass Spectrometer: Application to Atmospheric Inorganic and Organic Compounds, Environmental Science & Technology, 48, 6309-6317, 10.1021/es500362a, 2014.

Lehtipalo, K., Yan, C., Dada, L., Bianchi, F., Xiao, M., Wagner, R., Stolzenburg, D., Ahonen, L. R., Amorim, A., Baccarini, A., Bauer, P. S., Baumgartner, B., Bergen, A., Bernhammer, A.-K., Breitenlechner, M., Brilke, S., Buchholz, A., Mazon, S. B., Chen, D., Chen, X., Dias, A., Dommen, J., Draper, D. C., Duplissy, J., Ehn, M., Finkenzeller, H., Fischer, L., Frege, C., Fuchs, C., Garmash, O., Gordon, H., Hakala, J., He, X., Heikkinen, L., Heinritzi, M., Helm, J. C., Hofbauer, V., Hoyle, C. R., Jokinen, T., Kangasluoma, J., Kerminen, V.-M., Kim, C., Kirkby, J., Kontkanen, J., Kürten, A., Lawler, M. J., Mai, H., Mathot, S., Mauldin, R. L., Molteni, U., Nichman, L., Nie, W., Nieminen, T., Ojdanic, A., Onnela, A., Passananti, M., Petäjä, T., Piel, F., Pospisilova, V., Quéléver, L. L. J., Rissanen, M. P., Rose, C., Sarnela, N., Schallhart, S., Schuchmann, S., Sengupta, K., Simon, M., Sipilä, M., Tauber, C., Tomé, A., Tröstl, J., Väisänen, O., Vogel, A. L., Volkamer, R., Wagner, A. C., Wang, M., Weitz, L., Wimmer, D., Ye, P., Ylisirniö, A., Zha, Q., Carslaw, K. S., Curtius, J., Donahue, N. M., Flagan, R. C., Hansel, A., Riipinen, I., Virtanen, A., Winkler, P. M., Baltensperger, U., Kulmala, M., and Worsnop, D. R.: Multicomponent new particle formation from sulfuric acid, ammonia, and biogenic vapors, 4, eaau5363, 10.1126/sciadv.aau5363 %J Science Advances, 2018.

Liebmann, J., Karu, E., Sobanski, N., Schuladen, J., Ehn, M., Schallhart, S., Quéléver, L., Hellen, H., Hakola, H., Hoffmann, T., Williams, J., Fischer, H., Lelieveld, J., and Crowley, J. N.: Direct measurement of NO3 radical reactivity in a boreal forest, Atmos. Chem. Phys., 18, 3799-3815, 10.5194/acp-18-3799-2018, 2018.

Mohr, C., Lopez-Hilfiker, F. D., Yli-Juuti, T., Heitto, A., Lutz, A., Hallquist, M., D'Ambro, E. L., Rissanen, M. P., Hao, L., Schobesberger, S., Kulmala, M., Mauldin III, R. L., Makkonen, U., Sipilä, M., Petäjä, T., and Thornton, J. A.: Ambient observations of dimers from terpene oxidation in the gas phase: Implications for new particle formation and growth, 44, 2958-2966, 10.1002/2017gl072718, 2017.

Nah, T., Sanchez, J., Boyd, C. M., and Ng, N. L.: Photochemical Aging of α-pinene and β-pinene Secondary Organic Aerosol formed from Nitrate Radical Oxidation, Environmental Science & Technology, 50, 222-231, 10.1021/acs.est.5b04594, 2016.

Ng, N. L., Brown, S. S., Archibald, A. T., Atlas, E., Cohen, R. C., Crowley, J. N., Day, D. A., Donahue, N. M., Fry, J. L., Fuchs, H., Griffin, R. J., Guzman, M. I., Herrmann, H., Hodzic, A., Iinuma, Y., Jimenez, J. L., Kiendler-Scharr, A., Lee, B. H., Luecken, D. J., Mao, J., McLaren, R., Mutzel, A., Osthoff, H. D., Ouyang, B., Picquet-Varrault, B., Platt, U., Pye, H. O. T., Rudich, Y., Schwantes, R. H., Shiraiwa, M., Stutz, J., Thornton, J. A., Tilgner, A., Williams, B. J., and Zaveri, R. A.: Nitrate radicals and biogenic volatile organic compounds: oxidation, mechanisms, and organic aerosol, Atmospheric Chemistry and Physics, 17, 2103-2162, 10.5194/acp-17-2103-2017, 2017.

Orlando, J. J., and Tyndall, G. S.: Laboratory studies of organic peroxy radical chemistry: an overview with emphasis on recent issues of atmospheric significance, J Chemical Society Reviews, 41, 6294-6317, 2012.

Paatero, P., and Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, Environmetrics, 5, 111-126, 1994.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemometrics and Intelligent Laboratory Systems, 37, 23-35, https://doi.org/10.1016/S0169-7439(96)00044-5, 1997.

Paatero, P.: The Multilinear Engine—A Table-Driven, Least Squares Program for Solving Multilinear Problems, Including the n-Way Parallel Factor Analysis Model, Journal of Computational and Graphical Statistics, 8, 854-888, 10.1080/10618600.1999.10474853, 1999.

Paulson, S. E., and Orlando, J. J.: The reactions of ozone with alkenes: An important source of HOx in the boundary layer, 23, 3727-3730, 10.1029/96gl03477, 1996.

Peräkylä, O., Riva, M., Heikkinen, L., Quéléver, L., Roldin, P., and Ehn, M.: Experimental investigation into the volatilities of highly oxygenated organic molecules (HOM), Atmospheric Chemistry and Physics, 20, 649–669, 10.5194/acp-2019-620, 2020.

Perraud, V., Bruns, E. A., Ezell, M. J., Johnson, S. N., Greaves, J., and Finlayson-Pitts, B. J.: Identification of Organic Nitrates in the NO3 Radical Initiated Oxidation of α-Pinene by Atmospheric Pressure Chemical Ionization Mass Spectrometry, Environmental Science & Technology, 44, 5887-5893, 10.1021/es1005658, 2010.

Polissar, A. V., Hopke, P. K., Paatero, P., Malm, W. C., and Sisler, J. F.: Atmospheric aerosol over Alaska: 2. Elemental composition and sources, Journal of Geophysical Research: Atmospheres, 103, 19045-19057, 1998.

Pope III, C. A., Ezzati, M., and Dockery, D. W.: Fine-particulate air pollution and life expectancy in the United States, New England Journal of Medicine, 360, 376-386, 2009.

Riva, M., Rantala, P., Krechmer, J. E., Peräkylä, O., Zhang, Y., Heikkinen, L., Garmash, O., Yan, C., Kulmala, M., Worsnop, D., and Ehn, M.: Evaluating the performance of five different chemical ionization techniques for detecting gaseous oxygenated organic species, Atmospheric Measurement Techniques, 12, 2403-2421, 10.5194/amt-12-2403-2019, 2019.

Shiraiwa, M., Ueda, K., Pozzer, A., Lammel, G., Kampf, C. J., Fushimi, A., Enami, S., Arangio, A. M., Fröhlich-Nowoisky, J., Fujitani, Y., Furuyama, A., Lakey, P. S. J., Lelieveld, J., Lucas, K., Morino, Y., Pöschl, U., Takahama, S., Takami, A., Tong, H., Weber, B., Yoshino, A., and Sato, K.: Aerosol Health Effects from Molecular to Global Scales, Environmental Science & Technology, 51, 13545-13567, 10.1021/acs.est.7b04417, 2017.

Song, Y., Shao, M., Liu, Y., Lu, S., Kuster, W., Goldan, P., and Xie, S.: Source apportionment of ambient volatile organic compounds in Beijing, Environmental science & technology, 41, 4348-4353, 2007.

Spittler, M., Barnes, I., Bejan, I., Brockmann, K. J., Benter, T., and Wirtz, K.: Reactions of NO3 radicals with limonene and α-pinene: Product and SOA formation, Atmospheric Environment, 40, 116-127, https://doi.org/10.1016/j.atmosenv.2005.09.093, 2006.

Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.: IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 1535 pp, in, Cambridge Univ. Press, Cambridge, UK, and New York, 2013.

Troestl, J., Chuang, W. K., Gordon, H., Heinritzi, M., Yan, C., Molteni, U., Ahlm, L., Frege, C., Bianchi, F., Wagner, R., Simon, M., Lehtipalo, K., Williamson, C., Craven, J. S., Duplissy, J., Adamov, A., Almeida, J., Bernhammer, A.-K., Breitenlechner, M., Brilke, S., Dias, A., Ehrhart, S., Flagan, R. C., Franchin, A., Fuchs, C., Guida, R., Gysel, M., Hansel, A., Hoyle, C. R., Jokinen, T., Junninen, H., Kangasluoma, J., Keskinen, H., Kim, J., Krapf, M., Kuerten, A., Laaksonen, A., Lawler, M., Leiminger, M., Mathot, S., Moehler, O., Nieminen, T., Onnela, A., Petaejae, T., Piel, F. M., Miettinen, P., Rissanen, M. P., Rondo, L., Sarnela, N., Schobesberger, S., Sengupta, K., Sipila, M., Smith, J. N., Steiner, G., Tome, A., Virtanen, A., Wagner, A. C., Weingartner, E., Wimmer, D., Winkler, P. M., Ye, P., Carslaw, K. S., Curtius, J., Dommen, J., Kirkby, J., Kulmala, M., Riipinen, I., Worsnop, D. R., Donahue, N. M., and Baltensperger, U.: The role of low-volatility organic compounds in initial particle growth in the atmosphere, Nature, 533, 527-531, 10.1038/nature18271, 2016.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, Atmos. Chem. Phys., 9, 2891-2918, 10.5194/acp-9-2891-2009, 2009.

Yan, C., Nie, W., Aijala, M., Rissanen, M. P., Canagaratna, M. R., Massoli, P., Junninen, H., Jokinen, T., Sarnela, N., Hame, S. A. K., Schobesberger, S., Canonaco, F., Yao, L., Prevot, A. S. H., Petaja, T., Kulmala, M., Sipila, M., Worsnop, D. R., and Ehn, M.: Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization, Atmospheric Chemistry and Physics, 16, 12715-12731, 10.5194/acp-16-12715-2016, 2016.

Zha, Q., Yan, C., Junninen, H., Riva, M., Sarnela, N., Aalto, J., Quéléver, L., Schallhart, S., Dada, L., Heikkinen, L., Peräkylä, O., Zou, J., Rose, C., Wang, Y., Mammarella, I., Katul, G., Vesala, T., Worsnop, D. R., Kulmala, M., Petäjä, T., Bianchi, F., and Ehn, M.: Vertical characterization of highly oxygenated molecules (HOMs) below and above a boreal forest canopy, Atmos. Chem. Phys., 18, 17437-17450, 10.5194/acp-18-17437-2018, 2018.

Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Ulbrich, I. M., Ng, N. L., Worsnop, D. R., and Sun, Y.: Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review, Analytical and Bioanalytical Chemistry, 401, 3045-3067, 10.1007/s00216-011-5355-y, 2011.

Zhang, Y., Lin, Y., Cai, J., Liu, Y., Hong, L., Qin, M., Zhao, Y., Ma, J., Wang, X., and Zhu, T.: Atmospheric PAHs in North China: spatial distribution and sources, Science of the Total Environment, 565, 994-1000, 2016.

Zhang, Y., Cai, J., Wang, S., He, K., and Zheng, M.: Review of receptor-based source apportionment research of fine particulate matter and its challenges in China, Science of the Total Environment, 586, 917-929, 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.: A novel approach for simple statistical analysis of high-resolution mass spectra, Atmospheric Measurement Techniques, 12, 3761-3776, 10.5194/amt-12-3761-2019, 2019.