Atmospheric
Chemistry
and Physics

Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization

Chao Yan[1], Wei Nie[2,1], Mikko Äijälä[1], Matti P. Rissanen[1], Manjula R. Canagaratna[3], Paola Massoli[3], Heikki Junninen[1], Tuija Jokinen[1,a], Nina Sarnela[1], Silja A. K. Häme[1], Siegfried Schobesberger[1,b], Francesco Canonaco[4], Lei Yao[5], André S. H. Prévôt[4], Tuukka Petäjä[1,2], Markku Kulmala[1], Mikko Sipilä[1], Douglas R. Worsnop[1,3], and Mikael Ehn[1]

[1]Department of Physics, University of Helsinki, Helsinki, 00140, Finland
[2]Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing University, Nanjing, 210046, China
[3]Aerodyne Research, Inc., Billerica, MA 01821, USA
[4]Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, Villigen, 5232, Switzerland
[5]Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention (LAP3), Department of Environmental Science & Engineering, Fudan University, Shanghai, 200433, China
[a]now at: Department of Chemistry, University of California, Irvine, CA, 92617, USA
[b]now at: Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA

Correspondence to: Chao Yan (chao.yan@helsinki.fi) and Wei Nie (niewei@nju.edu.cn)

Abstract. Highly oxidized multifunctional compounds (HOMs) have been demonstrated to be important for atmospheric secondary organic aerosols (SOA) and new-particle formation (NPF), yet it remains unclear which the main atmospheric HOM formation pathways are. In this study, a nitrate-ion-based chemical ionization atmospheric-pressure-interface time-of-flight mass spectrometer (CI-APi-TOF) was deployed to measure HOMs in the boreal forest in Hyytiälä, southern Finland. Positive matrix factorization (PMF) was applied to separate the detected HOM species into several factors, relating these "factors" to plausible formation pathways. PMF was performed with a revised error estimation derived from laboratory data, which agrees well with an estimate based on ambient data. Three factors explained the majority (> 95 %) of the data variation, but the optimal solution found six factors, including two nighttime factors, three daytime factors, and a transport factor. One nighttime factor is almost identical to laboratory spectra generated from monoterpene ozonolysis, while the second likely represents monoterpene oxidation initiated by NO_3. The exact chemical processes forming the different daytime

factors remain unclear, but they all have clearly distinct diurnal profiles, very likely related to monoterpene oxidation with a strong influence from NO, presumably through its effect on peroxy radical (RO_2) chemistry. Apart from these five "local" factors, the sixth factor is interpreted as a transport related factor. These findings improve our understanding of HOM production by confirming current knowledge and inspiring future research directions and provide new perspectives on using factorization methods to understand short-lived atmospheric species.

1 Introduction

Large amounts of volatile organic compounds (VOCs) are emitted into the atmosphere from both biogenic and anthropogenic sources (Atkinson and Arey, 2003). These VOCs are oxidized in the atmosphere, which leads to thousands of structurally distinct products, containing many functionalities (Hallquist et al., 2009). A subset of these products

become highly oxidized multifunctional (HOM) compounds (HOMs; Ehn et al., 2012) and, although generally considered a minor pathway in VOC oxidation, they play a crucial role in atmospheric aerosol formation (e.g., Kulmala et al., 2013; Ehn et al., 2014; Jokinen et al., 2015) and thereby both air quality (Nel, 2005) and climate (IPCC, 2013).

The existence of HOMs had been suggested by model studies, which assumed that a fraction of the VOC oxidation products was effectively nonvolatile (Spracklen et al., 2011; Riipinen et al., 2011). Only recently, with the development of the APi-TOF (Junninen et al., 2010) and later the chemical ionization atmospheric-pressure-interface time-of-flight mass spectrometer (CI-APi-TOF; Jokinen et al., 2012), has it been possible to directly detect these HOMs (Ehn et al., 2012, 2014), with subsequent studies dedicated to understand the atmospheric implications of HOMs. The systematical investigation of new-particle formation (NPF) events observed at the SMEAR II station in southern Finland suggested a key role of HOMs in NPF (Kulmala et al., 2013). Further laboratory studies have confirmed this finding. Schobesberger et al. (2013) showed that HOMs can participate in the initial steps of NPF by stabilizing sulfuric acid, and the inclusion of this mechanism significantly improves the model prediction of particle number concentration (Riccobono et al., 2014). Ehn et al. (2014) have simulated HOM formation with O_3 and α-pinene (the most abundant biogenic VOC in high latitudes) and shown that these HOMs can explain the majority of the observed particle growth from 5 up to 50 nm at SMEAR II. Though the molar yield of HOMs is only a few percent depending on the VOC structure and oxidant, a global model suggested HOMs play a crucial role in secondary organic aerosol (SOA) burden and cloud condensation nuclei concentrations (Jokinen et al., 2015).

As HOMs are important compounds linking VOCs to SOA, quantitative simulation of SOA formation requires detailed understanding of HOM formation. According to current knowledge, the formation of HOMs consists of two consecutive processes: (1) VOC oxidation forming peroxy radicals (RO_2) able to auto-oxidize through intramolecular H-abstraction, leading to multiple O_2 additions; and (2) termination reactions, which terminate the auto-oxidation by converting RO_2 radicals into closed-shell molecules. Ehn et al. (2014) successfully simulated ambient nighttime HOM spectra by adding O_3 and α-pinene into a chamber, indicating the importance of that O_3-initiated oxidation and the following multi-step H-shift reactions (auto-oxidation). Jokinen et al. (2014, 2015) later expanded the HOM observations to a broader group of VOC precursors and oxidants (O_3 and OH). Similar processes have been confirmed for the NO_3-initiated monoterpene oxidation investigated by Boyd et al. (2015) with chemical ionization mass spectrometry using I^- as the reagent ion. Termination reactions occur in competition with further auto-oxidation and may even prevent it altogether. In the atmosphere, RO_2 termination may happen by reacting with partners ("terminators", i.e., hydroperoxyl radical

(HO_2), RO_2, NO_x) or undergoing self-termination (Orlando and Tyndall, 2012). The large variety of terminators leads to critical branching steps in the atmospheric oxidative pathways, eventually resulting in a large number of different HOM molecules. Despite the new insights acquired from recent chamber studies, HOM formation in the complex atmosphere remains poorly understood. One of the fundamental reasons is the lack of robust methods to analyze the complicated ambient data (e.g., mass spectra containing $\gg 100$ molecular ions) and to link ambient observations and chamber studies.

Positive matrix factorization (PMF) (Paatero and Tapper, 1994) allows for time-resolved mass spectra to be expressed as a linear combination of a finite number of factors, assuming that the factor profiles are constant and unique (Ulbrich et al., 2009). Since this method does not require a priori information about the factors, it is an ideal technique for extracting information from ambient measurements where the detailed chemistry, sources, and atmospheric processes are complex. PMF analysis of aerosol mass spectrometer (AMS) data, for example, has been widely utilized to identify multiple primary organic aerosol sources (i.e., vehicle emissions, biomass burning, cooking) and to characterize SOA aging via factors with varying volatilities and oxidation levels (Lanz et al., 2007; Ulbrich et al., 2009; Ng et al., 2010; Jimenez et al., 2009; Zhang et al., 2011). PMF has also been applied to analyze time-resolved ambient proton transfer reaction mass spectrometer (PTR-MS) measurements of organic species in the gas phase (Vlasenko et al., 2009; Yuan et al., 2012) and to analyze combined AMS-PTR-MS datasets (Slowik et al., 2010; Crippa et al., 2013).

In this work, we report the first success of utilizing PMF on CI-APi-TOF data. We examine the degree to which the PMF factors represent the dominant HOM formation pathways at the observation site and attempt to validate the retrieved factors by comparison to existing chamber data and correlation with other co-located measurements. Our results link the ambient measurement to previous chamber studies and identify needs for future research efforts in this area. This work also provides new perspectives on using PMF to understand the variation of short-lived species, e.g., HOMs.

2 Measurement

2.1 Site description

In this study, the measurement data were obtained at boreal forest research station SMEAR II located in Hyytiälä, southern Finland (Hari and Kulmala, 2005). The station is surrounded by boreal conifer forest and is described as a rural continental background measurement site (e.g., in Manninen et al., 2010). The nearest large cities are Tampere (around 60 km to southwest; 213 000 inhabitants) and Jyväskylä (around 100 km to northeast; 131 000 inhabitants).

SMEAR II is a rural site, but sometimes polluted air masses reach the site and cause relatively high aerosol loadings and high concentrations of gas-phase pollutants. Typical pollutants are from forest fires in Russia, biomass burning from eastern Europe, Tampere urban plume, or a nearby sawmill (southeast of SMEAR II) (e.g., Liao et al., 2011; Ulevicius et al., 2016). Ambient meteorological conditions such as temperature, relative humidity (RH), solar radiation, wind speed and direction, particle concentration and size distribution, and concentrations of aerosol particles and several trace gases, e.g., carbon dioxide (CO_2), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen oxides (NO_x), and ozone (O_3), are continuously monitored at the station.

2.2 Measurement of oxidized organic compounds

A NO_3^--based CI-APi-TOF was deployed to measure the highly oxidized organic compounds as well as sulfuric acid in an intensive observation period in April–May 2012. This state-of-the-art instrument can sensitively and selectively measure many HOMs with high oxygen to carbon ratio. Instrument and measurement details have been described elsewhere (Junninen et al., 2010; Jokinen et al., 2012). The mass spectra were analyzed with the tofTools package developed by Junninen et al. (2010). The quantification of HOM was calculated as

$$[HOM] = \frac{HOM(NO_3^-)}{\sum_{i=0}^{2}(HNO_3)_i \left(NO_3^-\right)} \times C. \tag{1}$$

Here [HOM] is the concentration of the HOM molecule to be quantified, the numerator on the right-hand side is the observed signal of its cluster with NO_3^-, the denominator is the sum of all reagent ion signals, and C is the calibration coefficient representing the detection sensitivity for the HOM molecule. We only include the $HOM \cdot NO_3^-$ adduct in the calculations, since typically only a negligible fraction of the signal is found in the form of a pure deprotonated HOM or a cluster with $(HNO_3) \cdot NO_3^-$. As suggested by Ehn et al. (2014), the calibration coefficient for HOMs is assumed equal to the value used for sulfuric acid within 50 % uncertainty. The calibration coefficient reported by Jokinen et al. (2012) is used in this work, as the tuning of the instrument and the geometry of the sampling tube were similar.

2.3 Positive matrix factorization

2.3.1 Working principle and advantages of PMF

PMF is a well-established model based on the work by Paatero and Tapper (1994). This receptor model is useful for solving functional mixing models when the source number and source profiles are unknown. It fundamentally works on an assumption of mass conservation so that a mass balance analysis can be used to identify and apportion sources of the detected species in the atmosphere. The most important feature that distinguishes PMF from other receptor modeling (e.g., principal component analysis) is that it applies a least-squares algorithm that accounts for data uncertainties. It also constrains the solutions to the non-negative subspace so that they are environmentally reasonable. Due to these advantages, this model is widely used for source apportionment analysis. The PMF analysis in this work uses the IGOR based analyzing interface SoFi (solution finder, version 5.2) and ME-2 as described in Canonaco et al. (2013).

Using PMF on mass spectral data, the mass balance can be described as

$$\mathbf{X} = \mathbf{TS} \cdot \mathbf{MS} + \mathbf{E}. \tag{2}$$

Matrix $\mathbf{X}$ is an $m \times n$ matrix, representing m measurements (in time) of n masses. The sizes of the factor matrices $\mathbf{TS}$ and $\mathbf{MS}$ are $m \times p$ and $p \times n$, respectively, where p is the number of factors. In practice, the matrix $\mathbf{TS}$ is the time series of the p factors representing the source strength, and matrix $\mathbf{MS}$ contains the mass spectra of the p factors. Matrix $\mathbf{E}$ is the residual unexplained by the p factors. It should be noted that the value of p is not pre-fixed, and determination of the value will be based on the interpretability of the solutions.

The PMF runs to seek the minimum Q, the sum of squared residual weighted by the inverse of their respective measurement uncertainty, which can be described as

$$Q = \sum_{i=1}^{m} \sum_{j=1}^{n} (\frac{E_{ij}}{S_{ij}})^2. \tag{3}$$

Here S_{ij} is the estimated measurement uncertainty of mass j at time i, and E_{ij} is the corresponding model residual. In this work, the uncertainty was estimated from laboratory data, which will be discussed in Sect. 2.3.3. Data points where $E_{ij} \gg S_{ij}$ have a large influence on the model iteration, and this needs to be reduced or removed by the model. A robust mode is applied to eliminate the strong outliers determined by α, meaning that any data points yielding $E_{ij}/S_{ij} > \alpha$ will be reduced to this threshold:

$$\text{if } \left| \frac{E_{ij}}{S_{ij}} \right| > \alpha, \left| \frac{E_{ij}}{S_{ij}} \right| = \alpha, \tag{4}$$

where the value of α is a free parameter can be determined by the user, and a value of 4 was suggested by Paatero (1997).

Ideally, the modeled Q value should eventually approach to the expected Q values (Q_{exp}), which is equal to the degree of freedom of the model solution. For mass spectra data, it roughly equals to the size of the matrix:

$$Q_{exp} \approx (n \times m). \tag{5}$$

2.3.2 Data matrix

The nitrate-ion-based (NO_3^-) CI-APi-TOF selectively measures HOMs with a ~ 4000 Th/Th resolving power. In principle, this resolution allows us to fit peaks and in some cases resolve peaks with different composition at the same unit mass.

However, the quality of the peak fitting strongly depends on mass calibration of the spectrum and the smoothness of the peaks. We found that the mass calibration may shift by 5 ppm by using data with 5 min integration time, and some HOM peaks are not smooth enough due to the weak signals. Fitting the peaks beforehand in such circumstances may introduce extra and nonuniform uncertainties that are difficult to estimate. Therefore, the data matrix used in this work is in unit-mass resolution, and peak fitting was performed afterwards to identify the elemental formula of peaks. Some examples of peak fitting are provided in Fig. S10 in the Supplement. The mass range of 201–650 Th was selected for PMF analysis, which covers most of the detectable HOMs. We continuously collected data from 4 April to 7 May 2012, with very few missing time points due to instrumental issues. The input data matrix consists of counts-per-second (cps) values, averaged from raw data using 5 min time resolution, resulting in a total of 9084 mass spectra. Thus, the final data matrix is in the size of 9084 (samples) × 450 (variables).

2.3.3 Error matrix estimation

Due to the abovementioned model principle, the estimation of error matrix (S_{ij}) is crucial. Suggested by Polissar et al. (1998), the error matrix in this work was estimated as

$$S_{ij} = \sigma_{ij} + \sigma_{\text{noise}}. \tag{6}$$

There are two terms contributing to the total measurement uncertainty: σ_{ij} is the analytical uncertainty from counting statistics, and σ_{noise} is the standard deviation of instrument noise, also representing the instrument detection limit.

We estimate both σ_{ij} and σ_{noise} from laboratory data. The schematic of the corresponding experimental setup is provided in the Supplement (Fig. S1). Briefly, we generated stable signals with a temperature-controlled permeation source. A 100 mL min^{-1} (milliliter per minute) N$_2$ gas served as carrier gas flowing through the source, which was then diluted by a 10 L min^{-1} (liter per minute) N$_2$ flow before entering the CI inlet. The experiments were run under the following conditions:

1. Two different chemicals were used, i.e., perfluorobutanoic acid (CF$_3$(CF$_2$)$_2$COOH) and perfluorononanoic acid (CF$_3$(CF$_2$)$_7$COOH), respectively.

2. With each chemical, temperature was changed every hour to create multiple steps of stable signals (Fig. S2).

3. With each chemical, the experiment was repeated twice using different instrumental tunings.

σ_{noise} was calculated as the standard deviation of "blank masses" (800–1000 Th). As shown in Fig. S3, σ_{noise} from laboratory data (two different tunings) and the ambient data agree well. We apply a constant value of 0.035 for σ_{noise} in our analysis, taken as the median of standard deviations of
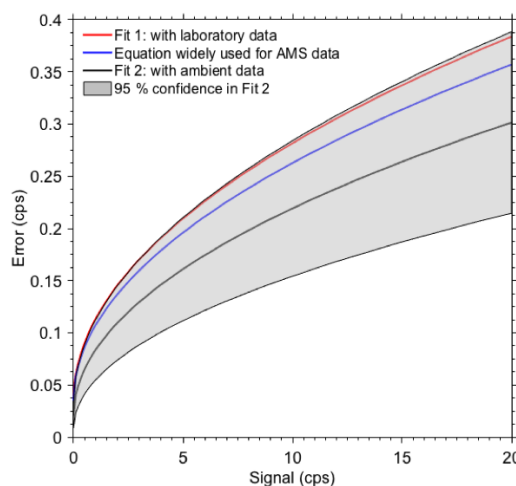

Figure 1. Error matrix estimation by fitting the error to the signal intensity. The red solid line is the best fitted curve from the laboratory experiment data, the blue curve denotes the fitting equation commonly used for AMS data, and the black represents the fitting from the ambient data with a different method (see Supplement Sect. S2) with its fitting uncertainty (95 % confidence) shown as the gray area.

each mass over the mass range, though a weak variation was observed.

The σ_{ij} was estimated based on the assumption that the counting statistics follow the Poisson distribution (Allan et al., 2003):

$$\sigma_{ij} = a \frac{\sqrt{I}}{\sqrt{t_s}}. \tag{7}$$

I is the signal strength (ions s^{-1}) of the ion, t_s is the integration time in seconds, and a is an empirically determined factor incorporating any unaccounted contributions to the uncertainty, for example arising from shifting of mass calibration, or baseline correction when averaging data. It should be noted that the factor a is different from the factor α defined by Allan et al. (2003), which accounts for variability in the size of pulses generated when single ions impact the detector. In this work, our data acquisition card, a time-to-digital converter (TDC), only counted single ions crossing a threshold, and thus the pulse variability did not influence the error estimate, as all signal pulses were large enough to cross the threshold. It should be also noted that most lately manufactured CI-APi-TOFs use an analog-to-digital converter (ADC) as the data acquisition card, the same as AMS does, and in this case the empirical factor a would incorporate the factor α.

With the stable signals during these experiments, the analytical uncertainty was fitted to the signal strength based on Eq. (7). Detailed information and discussion are provided in the Supplement Sect. S1. Briefly, the results suggest that the analytical uncertainty is independent of mass-to-charge and instrument tuning; the a value was fitted as 1.28. In other
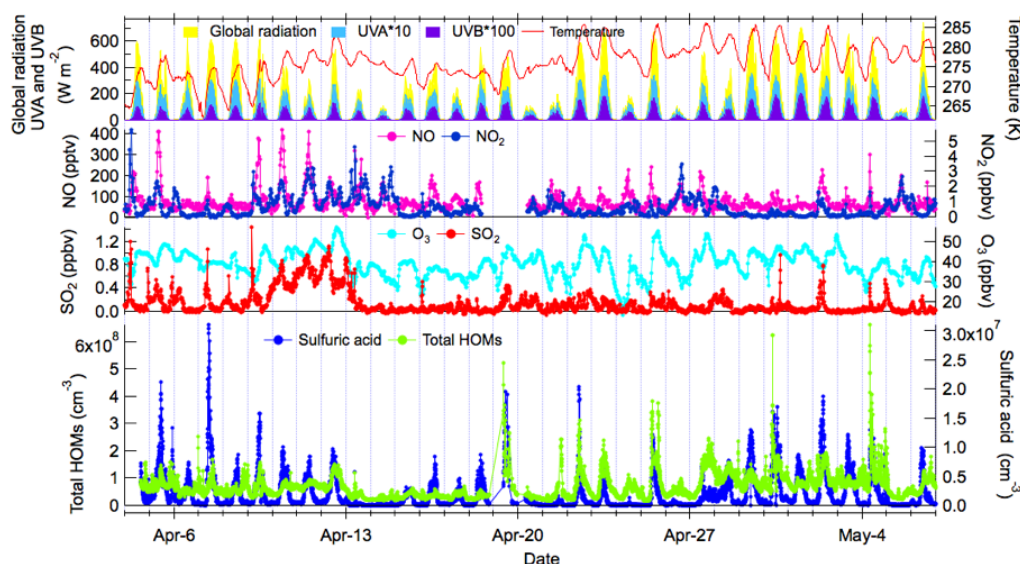
Figure 2. Overview of the measurement from 4 April to 8 May 2012. The top panel shows meteorological parameters, including UVA, UVB, global radiation, and temperature. Co-located measurements of inorganic trace gases, including NO, NO$_2$, SO$_2$, and O$_3$, are shown in middle panels. Highly oxidized species measured by the CI-APi-TOF, i.e., sulfuric acid (SA) and total HOMs, that are shown in the bottom panel.

words, the inclusion of the a parameter increases our uncertainty estimate by 28 %. For 300 s integration time, the overall error was estimated as

$$S_{ij} = 0.074\sqrt{|X_{ij}|} + 0.035. \qquad (8)$$

We also proposed a different statistical method based on ambient data (see Supplement Sect. S2). A comparison of different uncertainty estimation schemes is shown in Fig. 1, where the red curve denotes the revised error estimate in this work, the blue one is the customary estimate for AMS data, and the black one is the error estimated from ambient data with a different estimation scheme with its uncertainty shown as the gray area. Within the fitting uncertainty, all three estimates agree well.

In this work, two more steps were employed to further modify the error estimation:

1. For variables below $3\sigma_{\text{noise}}$, we fixed the signal as σ_{noise} and the corresponding uncertainty as $6\sigma_{\text{noise}}$. A similar approach was suggested by Polissar et al. (1998), but this practice is criticized by the developer of the PMF model, P. Paatero (Paatero, 2016). The effect of this "data censoring" proved negligible in our work and is discussed in more detail in the Supplement (Sect. 3).

2. A down-weighting scheme was also applied for variables whose mean signal-to-noise ratio (SNR) is low, i.e., X_{ij}/S_{ij}, as defined by Paatero and Hopke (2003). This further increased the error by 2 and 10 folds for "weak" (SNR < 2) and "bad" (SNR < 0.2) signals, respectively. Note that in step 1, all censored data points

are considered as "bad signals" in this step. The distribution of weak and bad signals is shown in Fig. S9. In total 173 variables (masses) were defined as weak signals and 152 variables were defined as bad signals.

3 Data overview

The data were collected at the SMEAR II station from 4 April to 7 May 2012. Figure 2 shows the time series of meteorological conditions (i.e., global radiation, UVA, UVB, and temperature), concentration of trace gases (NO, NO$_x$, O$_3$, SO$_2$), sulfuric acid (SA) concentration, and total HOM concentration. Looking at global radiation or UVA and UVB intensity (global radiation > 400 W m^{-2} or UVA > 15 W m^{-2}, UVB > 0.2 W m^{-2}), 78 % (26 out of 33) of the days in the measurement period had strong photochemical activity, the rest being cloudy days when photochemistry was significantly suppressed. From 9 to 12 April, air mass analysis using backward Lagrangian particle dispersion model (LPDM) (Ding et al., 2013) indicates that the measurement site was influenced by a polluted plume originating from eastern Europe (Fig. S11); clear elevations of anthropogenic pollutants, such as SO$_2$ and NO$_x$, were observed. During the entire period, the measured sum of HOM concentration exhibited clear diurnal variations, with notably higher levels in the daytime. Note this contrasts with lower daytime monoterpene concentrations trend that are typically observed VOCs at the site (Rantala et al., 2014), consistent with photochemical HOM production during daytime.

Apart from the variable concentrations, spectral differences between daytime and nighttime are also evident. The
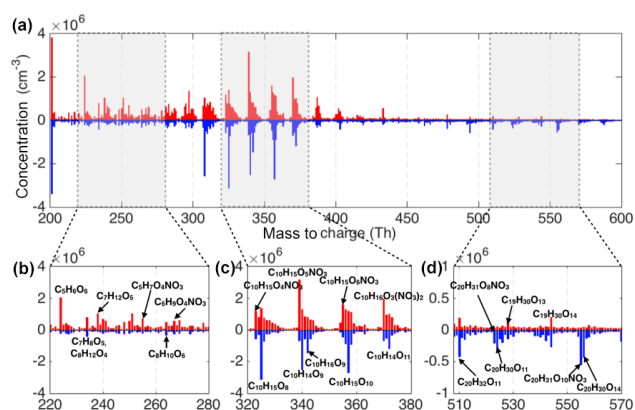
Figure 3. Comparison of spectra measured by CI-APi-TOF between daytime and nighttime. The daytime spectrum (marked in red) is above the zero line and the nighttime spectrum (marked in blue) is below the zero line. Figure 3b, c, and d present expanded mass spectra where major peaks are labeled with their possible elemental formula.

Figure 4. Source allocation from two- to seven-factor PMF solutions.

averaged spectra are presented in Fig. 3a, where night- and daytime spectra are shown below and above the zero line, respectively. As monoterpenes ($C_{10}H_{16}$) are known as the dominant precursors for HOMs at this location (Ehn et al., 2012), we divided the mass range (201–650 Th) into three subranges: (1) 201–290 Th for lighter HOM compounds, mostly containing three to seven carbons; (2) 290–450 Th for HOM "monomer" products, mostly fitting the general formula $C_{9-10}H_{14-16}O_{7-13}N_{0-2}$; and (3) 450–650 Th for HOM "dimer" products with the general formula $C_{16-20}H_{28-32}O_{9-19}N_{0-2}$. Expanded mass spectra are shown in Fig. 3b, c, and d, where some major peaks are labeled with their elemental formula. The lighter HOMs show notably elevated concentrations in the daytime. HOM monomers in the nighttime spectrum are similar to those reported in previous chamber studies (e.g., Ehn et al., 2014), whereas major peaks in the daytime are very likely organonitrates. These plausible organonitrates were identified with high yields when mixing monoterpenes, O_3, and NO_x in the chamber (Ehn et al., 2014; Jokinen et al., 2014), and they are also suggested to be important to NPF (Kulmala et al., 2013). Higher signals of HOM dimers are observed in the nighttime, with many major peaks similar to those have been reported by Ehn et al. (2014). However, there are also peaks likely containing nitrogen, which are produced through different reaction pathways.

Below, all elemental formulas for molecules containing N atoms will be expressed as NO_3 groups, since such organonitrate functionality is the only expected form of NO_3 (-ONO_2) in HOM species.
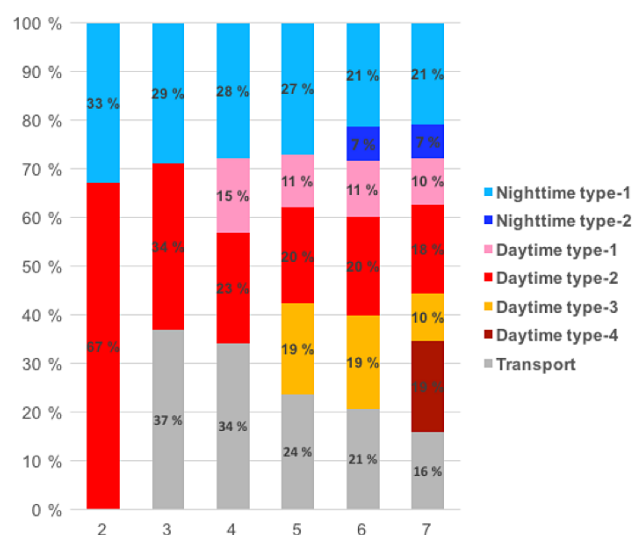
4 Results and discussion

4.1 Evolution of PMF solutions

Since the PMF analysis is performed without any a priori knowledge, the choice of the proper number of factors is the most critical decision towards interpreting the PMF results. Choosing the best factor number is a compromise. More factors give the model more freedom to explain subtle variations of the data but too many factors can force the model to split a physically meaningful factor into unrealistic ones. In this work, PMF analysis was initially done for two factors, and followed with a step-wise addition of one factor until the additional factor could no longer be interpreted based on the unique mass spectral feature or comparison of its time trend with auxiliary data. Figure 4 shows the average contribution of PMF solutions to HOM concentration assuming two to seven factors. Our main analysis focuses on the six-factor solution, but a short discussion of factor evolution is included below (factor profile and time series is shown in Fig. S12).

The two factor solution leads to distinct day- and nighttime factors. The spectral difference is also obvious: daytime factor contains more light HOM molecules but few HOM dimer products, while the nighttime factor contains very few light HOM molecules but most of the HOM dimer products. In addition, peaks with odd masses, which are likely nitrate-containing HOMs, dominate the daytime factor, while the major peaks in the nighttime factor have even masses and are unlikely to contain organic nitrogen.

In the three-factor case, the profile of two factors (daytime factor and nighttime factor) are more or less the same as those in the two-factor case, while the new factor is featured by a prominent peak at 201 Th, which is identified as nitrophenol ($C_6H_5NO_3$), although this species is detected as an adduct
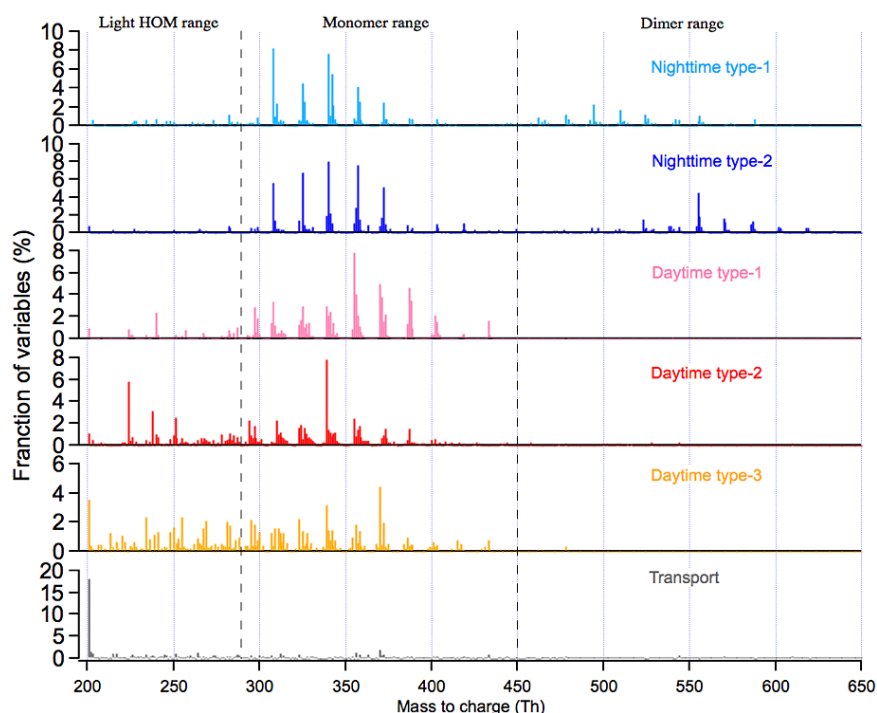
Figure 5. Factor profiles in six-factor PMF. The total signal of each factor is normalized to unity, and y axis is the fraction of variables in the factor in percentage.

with NO_3^-. Since the new factor exhibits a weak diurnal cycle, we temporally name it with its prominent peak, "201 Th factor".

In the four-factor solution, the daytime factor in the two-factor case splits into two new factors, termed daytime type-1 and daytime type-2, respectively. Their diurnal patterns are different – the daytime type-1 factor starts to increase at 04:00 and reaches the peak at $10{:}00\,\mathrm{UTC} + 3$, while the daytime type-2 factor starts to increase at around 06:00, and reaches the peak around 11:00–15:00. The major peaks in both new factors are organonitrates but in different masses: 355 Th ($C_{10}H_{15}O_6NO_3$) and 387 Th ($C_{10}H_{15}O_8NO_3$) are the most prominent peaks in the daytime type-1 factor, and 339 Th ($C_{10}H_{15}O_5NO_3$) is the highest peak in the daytime type-2 factor.

Introducing a fifth factor retrieves a third daytime factor. The other two daytime factors remain similar to those in the four-factor solution in respect to their diurnal patterns and major peaks, with their contributions to total HOM concentration reduced from 15 and 23 to 11 and 20 %, respectively (Fig. 4). The contribution of the "201 Th factor" also has a pronounced decrease from 34 to 24 % (Fig. 4), and its diurnal pattern has a clear change – peaking time changed from 12:00 to 09:00. The new daytime type-3 factor starts to increase at 06:00 in the morning and reach its peak value at 14:00 Fingerprint peaks in this factor are 213 Th ($C_3H_5O_3NO_3$), 241 Th ($C_4H_5O_4NO_3$),

255 Th ($C_5H_7O_4NO_3$), 269 Th ($C_6H_9O_4NO_3$), and 281 Th ($C_7H_9O_4NO_3$).

The six-factor solution separates nighttime factor into two different factors, namely nighttime type-1 and nighttime type-2, with the remaining factors are almost unchanged with respect to the five-factor solution (Figs. 5 and 6). Both new factors show elevated concentrations in the nighttime. The dominant peaks in the nighttime type-1 factor contain even masses in both HOM monomer and dimer mass ranges. In the nighttime type-2 factor, however, more intense odd-mass peaks are present, such as 403 Th ($C_{10}H_{15}O_9NO_3$) and 419 Th ($C_{10}H_{15}O_{10}NO_3$) in the monomer range, as well as 523 Th ($C_{20}H_{31}O_8NO_3$), 554 Th ($C_{20}H_{32}O_6(NO_3)_2$), and 555 Th ($C_{20}H_{31}O_{10}NO_3$) in the dimer range.

When seven factors are assumed, an additional daytime type factor appears. The new factor contains peaks that are mostly identified as nitrogen-containing organic compounds with 4–10 carbon atoms. Since there is no strong correlation with any independent tracer, we choose to limit our further analysis to the six-factor solution. Note that, without such correlations, it is not possible to distinguish the identification of "real" factors.

4.2 Mathematical diagnostics of PMF solutions

Mathematical diagnostics is important in evaluating PMF model performance. It usually includes the Q/Q_{exp} value, the distribution of Q over time and variables, the fraction of
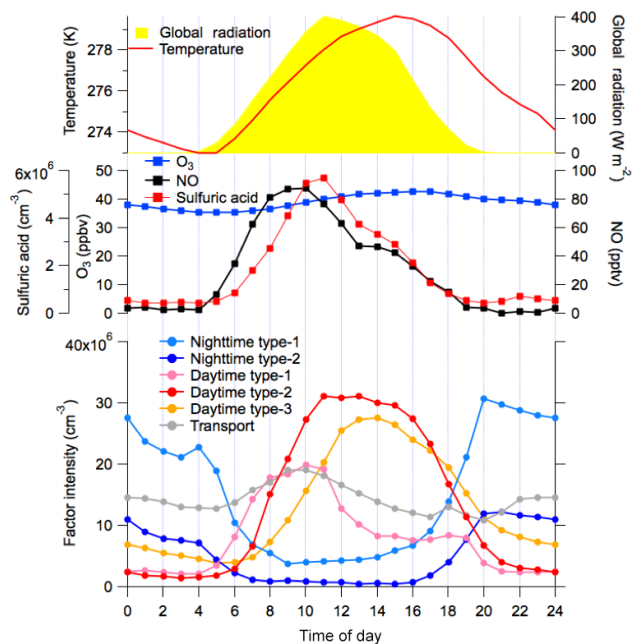
Figure 6. The diurnal cycle of PMF factors, selected meteorological parameters, and trace gas concentration.

explained variation in the data, and the consistency of seed runs.

Figure 7a shows the change of Q/Q_{exp}, which decreases stepwise from 2.44 (assuming two factors)(to 0.76 (assuming seven factors). For the six-factor PMF solution (the chosen optimal solution; see Sects. 4.1 and 4.3), the distribution of Q/Q_{exp} values over masses and time is shown in Fig. 7b and c, respectively. The distribution of Q/Q_{exp} exhibits a large variation (from 0.01 to 6.46) over masses (Fig. 7b). This is much larger than the theoretical variations around Q/Q_{exp} of 1 observed for synthetic datasets where random error is the only source of noise in the input data. For this dataset, the very low Q/Q_{exp} values may be explained by error amplification when censoring and down-weighting data, but the exact reason for the large Q/Q_{exp} values is more difficult to determine. Additionally, the Q/Q_{exp} variation over time is also quite large. Such large variations of Q/Q_{exp} suggest that the assumption of PMF did not perfectly hold; i.e., the factor profiles were not constant (Paatero, 2016). There could be a few reasons for the inconstancy of factor profiles, for example the change in the distribution of different monoterpene species and temperature and RH effects on monoterpene oxidation. Overall, the variation in the Q/Q_{exp} distribution clearly reveals that small and large values canceled each other out, causing the overall Q/Q_{exp} value close to 1. Therefore, it must be noted that the overall Q/Q_{exp} value must not be used on its own to judge the quality of PMF results. Instead, the temporal and mass spectral variation of the Q/Q_{exp} must also be examined in detail in order to ap-

propriately interpret the overall Q/Q_{exp} value and the PMF results.

Though the absolute value of Q/Q_{exp} might be misleading, the trend of Q/Q_{exp} is useful to determine the minimum factor number. As suggested by Ulbrich et al. (2009), a large decrease in Q/Q_{exp} indicates that the additional factor may explain a large fraction of unaccounted variability in the data. As shown in Fig. 7a, the third factor significantly decreases the Q/Q_{exp} value from 2.44 to 1.53, suggesting the importance of the third factor. By adding the third factor, the model can explain 95 % of the data variation, in comparison to 92 % when only two factors are assumed. This improvement in model performance also implies the third factor is crucial. The second largest increase in the explained fraction (from 95.5 to 97 %) happens when adding the sixth factor, suggesting the separation of the two nighttime type factors is significant, as mentioned in Sect. 4.1.

In order to evaluate the consistency of the PMF results, we run PMF from five different random starting points for each number of factors (seed runs; Paatero, 2007). As shown in Fig. 7a, the five seed runs for each factor number show good consistencies in both Q/Q_{exp} and explained variation, indicating the small model uncertainty. The only exception is the five-factor PMF, where the results in five seed runs show two groups with small discrepancies. This can indicate that there are likely two factorizations that generate equally valid solutions, suggesting that one more factor is required to resolve both factorizations.

4.3 Interpretation of PMF results

The mathematical diagnostics characterize the technical aspects of PMF. However, they are not guaranteed to give the most realistic solution. PMF is a descriptive model; thus the "interpretability" or "meaningfulness" is the most critical criterion in determining the best solution. Interpretation of PMF results needs careful examination of each retrieved factor, which usually requires many considerations:

– Comparison between the profile of retrieved factors and reference spectra from laboratory studies. The uncentered correlations (UC, Eq. 9, Ulbrich et al., 2009) is used to quantitatively assess the similarity:

$$\text{UC} = \frac{x \cdot y}{\|x\| \|y\|}, \tag{9}$$

where x and y denote a pair of time series or factor profile as vectors. In fact, as a new measurement technique, only a few of reference spectra have been reported for monoterpene oxidation (Jokinen et al., 2014; Ehn et al., 2014; Mutzel et al., 2015).

– Identification of key molecules as specific fingerprints of factors, as listed in Table 1. These molecules are chosen either if they are the most visible ones in the profile or if they are mostly (usually > 70 %) allocated to one
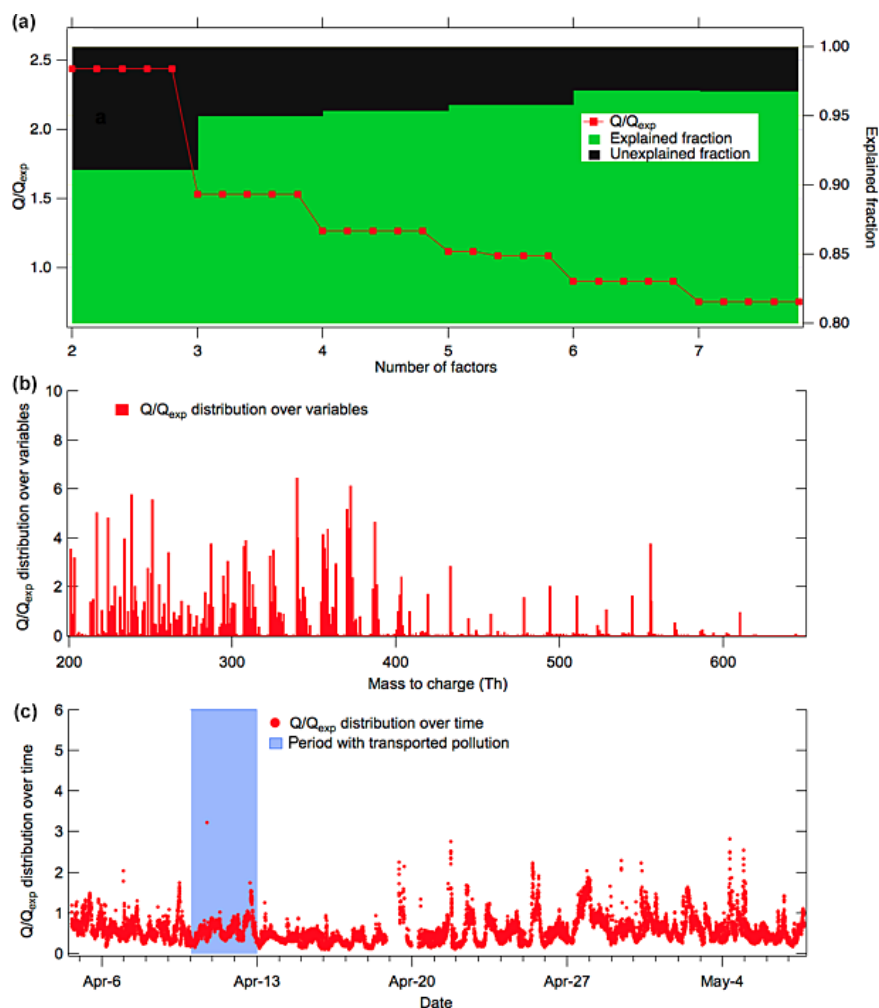
Figure 7. Mathematical diagnostics of PMF solutions, including the overall changes of Q/Q_{exp} and the explained variation from two-factor to seven-factor solutions. For each number of factors, five seed runs were performed to test the consistency of the solution.

specific factor. This method is rationalized by the fact that much molecular information is retained in the spectra, which helps to deduce the plausible reaction pathways.

– Temporal correlation of factors with other tracers which represent specific sources or atmospheric processes.

– Other information such as meteorology (e.g., air mass trajectories).

Based on these considerations, we concluded that the PMF solution with six factors is the optimal solutions. Figure 5 shows the spectra of the six factors, and their diurnal patterns are shown in Fig. 6, together with some relevant trace gases and meteorological parameters. It should be noted that all the mass spectra and diurnal profiles are very distinct, indicative of a realistic PMF solution. In the following subsections, each factor is discussed in detail.

4.3.1 Nighttime factors

Nighttime type-1 factor

The nighttime type-1 factor is the largest contributor to nighttime HOM concentration. It exhibits elevated intensity during 20:00–04:00 and is less intense (about five times lower) in the daytime. The major peaks in this factor are identified as $C_{10}H_{14-16}O_{6-13}$ and $C_{19-20}H_{28-32}O_{10-18}$. As shown in Fig. 8, the profile of this factor is very similar to the reference spectrum from previous laboratory studies reported by Ehn et al. (2014), where only ozone and α-pinene were mixed. It should be noted that, in the atmosphere, there is always a mixture of monoterpenes likely contributing to these signals, in contrast to a single monoterpene precursor was used in the chamber experiments. Also humidity and temperature were typically different and changing constantly, and all these together can explain the minor difference in individual peak intensities; for example, 372 Th ($C_{10}H_{14}O_{11}$)

Table 1. Suggested elemental composition of fingerprint molecules of the six factors. * Peak fitting are shown in Fig. S10.

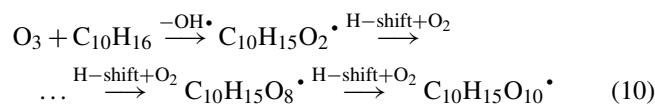| Factor	Fingerprint molecules
Nighttime type-1	$C_{10}H_{14}O_7$, $C_{10}H_{15}O_8$, $C_{10}H_{14}O_9$, *$C_{10}H_{15}O_{10}$, $C_{20}H_{32}O_{11}$
Nighttime type-2	$C_{20}H_{31}O_8NO_3$, *$C_{20}H_{31}O_{10}NO_3$, $C_{20}H_{32}O_6(NO_3)_2$
Daytime type-1	*$C_{10}H_{15}O_6NO_3$, $C_{10}H_{15}O_8NO_3$
Daytime type-2	*$C_{10}H_{15}O_5NO_3$, $C_5H_6O_7$
Daytime type-3	*$C_3H_5O_3NO_3$, $C_4H_5O_4NO_3$, $C_5H_7O_4NO_3$, $C_6H_9O_4NO_3$, $C_7H_9O_4NO_3$
Transport	*$C_6H_5NO_3$
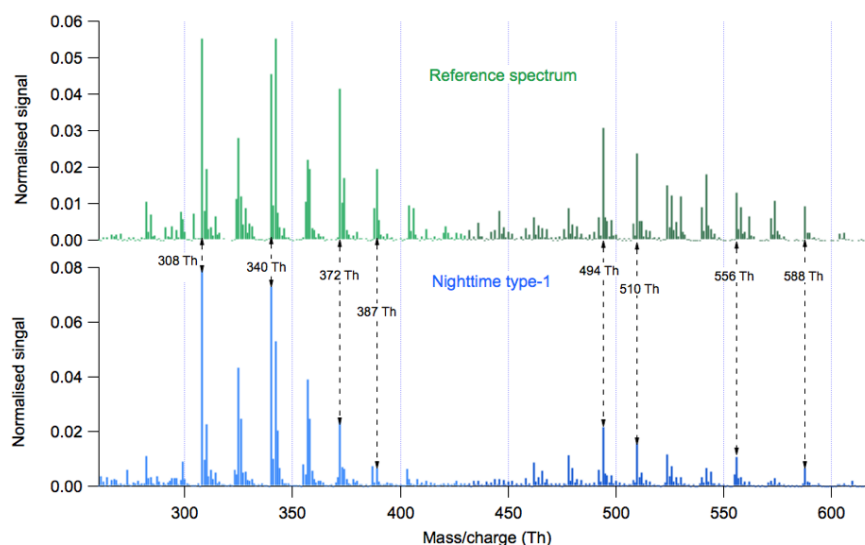
Figure 8. Comparison between the reference spectrum (Ehn et al., 2014) and the O_3 + monoterpene factor.

and 389 Th ($C_{10}H_{15}O_{12}$) are higher in the reference spectrum than in the factor profile. The coefficient of uncentered correlations between the factor profile and the reference spectrum was calculated to be 0.91, confirming the high similarity between them. Thus, the source of this factor is very likely the ozonolysis of monoterpenes.

Nighttime type-2 factor

The diurnal variation of the nighttime type-2 factor has a similar pattern to that of the nighttime type-1 factor. Its intensity is about 30 % of nighttime type-1 factor during the nighttime and almost decreases to 0 during the day (Fig. 6). To our knowledge, no reference spectrum that matches the profile of this factor (shown in Fig. 5) has been reported. However, a set of masses can represent a new fingerprint. Figure 9a shows these fingerprint peaks in the dimer range, which are categorized and marked in different colors. In general, the vast majority of compounds contain nitrogen and we divide dimer peaks in this factor into six groups according to their elemental composition, i.e., $C_{20}H_{31}O_{7-15}NO_3$, $C_{20}H_{32}O_{4-11}(NO_3)_2$, $C_{19}H_{29}O_{6-13}NO_3$, $C_{19}H_{31}O_{8-11}NO_3$, $C_{18}H_{29}O_{8-11}NO_3$, and other non-nitrogen-containing dimers. As dimers are

closed-shell molecules, assumed to be formed through the reaction between two peroxy radicals (RO_2) (Rissanen et al., 2014), the nitrogen atom(s) in the dimer molecule must come from its parent RO_2 radical, suggesting NO_3-initiated oxidation. Note that the possibility of NO_x involvement can be ruled out, because when NO_x reacts with RO_2 it either ends up with an organonitrate HOM monomer or forms an alkoxy radical (RO) so that the nitrogen atom will not retain in the molecule. The fractions of different groups are shown in Fig. 9b. About 61 % of HOM dimers in this factor contain one nitrogen atom, suggesting that the major dimer formation process involves reaction between two RO_2 radicals initiated by NO_3 and O_3, respectively. Also, about 22 % of these dimers contain two nitrogen atoms, meaning that both reacting RO_2 radicals are NO_3-initiated. The schematic illustrations given below show two examples of dimer formation containing one nitrogen atom ($C_{20}H_{31}NO_{13}$, 555 Th including NO_3^-) and two nitrogen atoms ($C_{10}H_{32}N_2O_{12}$, 554 Th including NO_3^-), respectively.

$$O_3 + C_{10}H_{16} \xrightarrow{-OH^\bullet} C_{10}H_{15}O_2^\bullet \xrightarrow{H-shift+O_2}$$

$$\dots \xrightarrow{H-shift+O_2} C_{10}H_{15}O_8^\bullet \xrightarrow{H-shift+O_2} C_{10}H_{15}O_{10}^\bullet \quad (10)$$
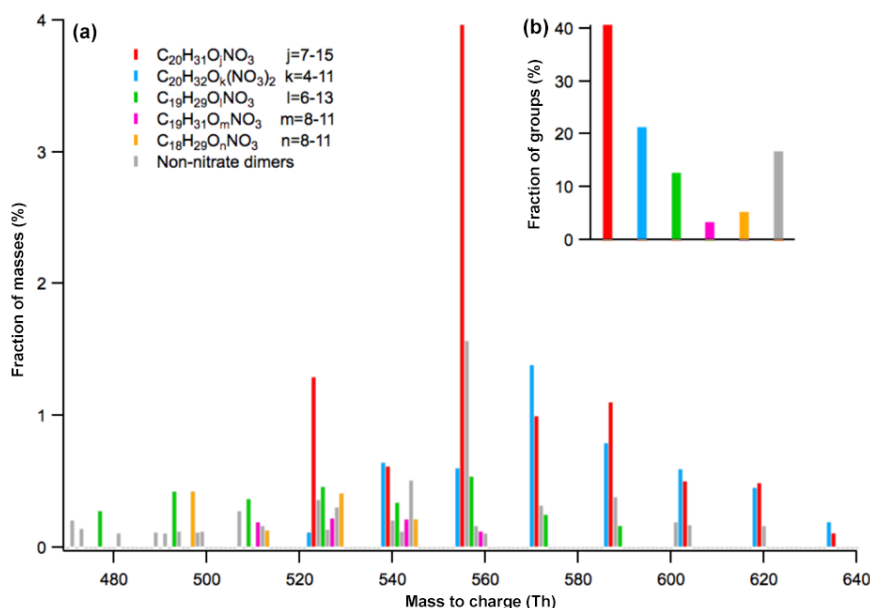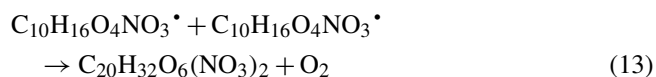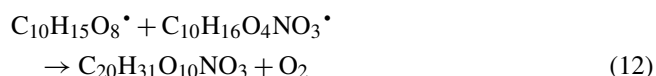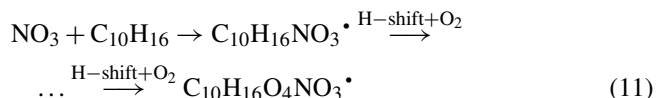
Figure 9. Dimer profile of the nighttime type-2 factor. All dimer peaks are assigned to six groups based on their elemental formula and marked with different colors. Figure 9a shows the location and mass fraction of individual peaks, and Fig. 9b gives the fraction of these groups.

$$NO_3 + C_{10}H_{16} \rightarrow C_{10}H_{16}NO_3{}^\bullet \xrightarrow{\text{H-shift+O}_2}$$

$$\ldots \xrightarrow{\text{H-shift+O}_2} C_{10}H_{16}O_4NO_3{}^\bullet \qquad (11)$$

$$C_{10}H_{15}O_8{}^\bullet + C_{10}H_{16}O_4NO_3{}^\bullet$$

$$\rightarrow C_{20}H_{31}O_{10}NO_3 + O_2 \qquad (12)$$

$$C_{10}H_{16}O_4NO_3{}^\bullet + C_{10}H_{16}O_4NO_3{}^\bullet$$

$$\rightarrow C_{20}H_{32}O_6(NO_3)_2 + O_2 \qquad (13)$$

As NO_3 is involved in the formation of more than 80 % dimer molecules, the nighttime type-2 factor is likely representing monoterpene oxidation by NO_3.

Comparison of the two nighttime factors

As mentioned above, the nighttime factors are interpreted as representing nighttime oxidation of monoterpene initiated by the two major nocturnal atmospheric oxidants – O_3 and NO_3, respectively. Their nighttime patterns are similar, exhibiting an increase at 20:00 and a decrease at 04:00 in the next morning (Fig. 6). However, as the O_3 concentration is relatively stable throughout day while NO_3 is much lower in the daytime, the O_3-initiated factor has finite level during the daytime while the NO_3-initiated factor goes almost to 0. In general, the O_3-initiated factor is a larger contributor than the NO_3-initiated factor, suggesting that O_3 is a more important nighttime oxidant for HOM formation at this measurement location. However, as shown in Fig. 10a, during the period (from 9 to 12 April) when polluted air masses containing high NO_x were transported to this area, the NO_3-initiated
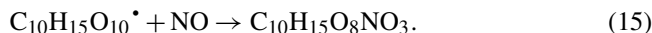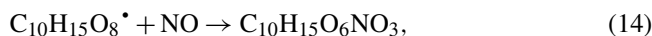
oxidation was significantly enhanced and became dominant. Since the NO_3 chemistry could be one important pathway of forming HOMs, future laboratory study of this reaction channel is required.

4.3.2 Daytime factors

The interpretation of daytime factors is more difficult, likely reflecting more complex daytime photochemistry. Nevertheless, certain conclusions can be drawn from spectral characteristics and temporal behavior of the three daytime factors.

Daytime type-1 factor

As shown in Fig. 6, this factor concentration starts to increase in the early morning (around 04:00), concurrent with the increase of NO and the decrease of the two nighttime factors. The very similar temporal behavior of this factor and NO (Fig. 10b) indicates that NO reaction is likely plays an important role in this factor. The two highest peaks in this new factor are 355 Th ($C_{10}H_{15}O_6NO_3$) and 387 Th ($C_{10}H_{15}O_8NO_3$), which are likely formed through the reaction between the two most abundant (O_3-initiated) RO_2 radicals and NO, as shown below:

$$C_{10}H_{15}O_8{}^\bullet + NO \rightarrow C_{10}H_{15}O_6NO_3, \qquad (14)$$

$$C_{10}H_{15}O_{10}{}^\bullet + NO \rightarrow C_{10}H_{15}O_8NO_3. \qquad (15)$$

We hereby interpret this factor as products from $RO_2 + NO$ reaction, which is also consistent with the observation that no dimer HOMs are present because NO is the dominating RO_2 terminator in this pathway.
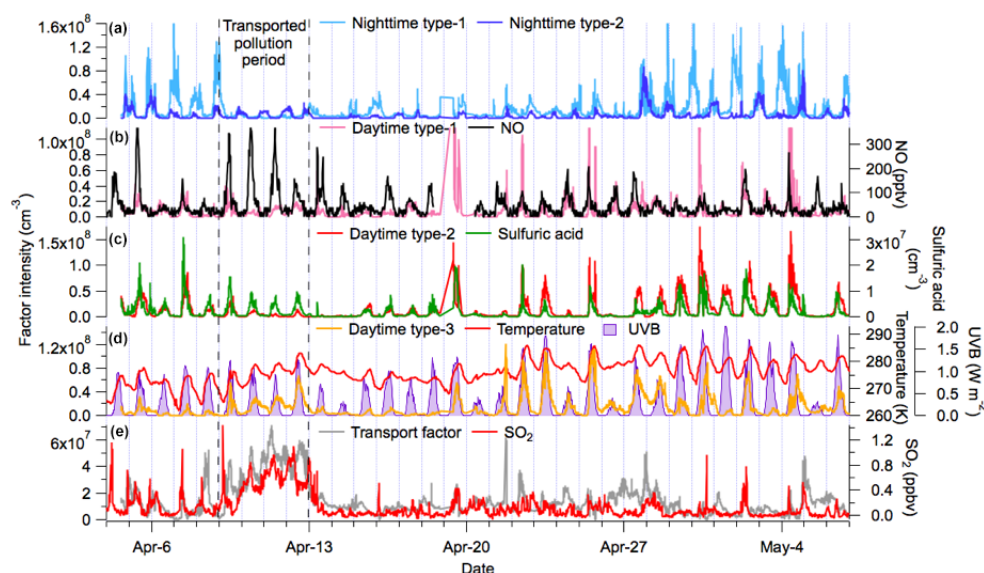
Figure 10. Temporal behaviors of PMF factors and relevant tracer gases as well as meteorological conditions. The period with transported pollution is marked by the dashed lines. Panel **(a)** depicts the temporal variation of the two nighttime factors. Panel **(b)** shows the time series of daytime type-1 factor together with NO. Panel **(c)** demonstrates the similar temporal behavior of the daytime type-2 factor and sulfuric acid. Panel **(d)** shows the time series of the daytime type-3 factor together with the relevant meteorological conditions (i.e., UVB and temperature). Panel **(e)** depicts the temporal variation of the transport factor, together with SO_2, a tracer for transported pollution.

Daytime type-2 factor

The daytime type-2 factor is one of the main daytime HOM contributors. The major peak in this factor is found at 339 Th ($C_{10}H_{15}O_5NO_3$), the single highest organonitrate molecule observed at this site and the representative of daytime HOMs previous reported by Kulmala et al. (2013). Another major peak in this factor is 224 Th ($C_5H_6O_6$), possibly a fragment of monoterpene oxidation as observed in laboratory experiments (e.g., Tröstl et al., 2016). Besides these major peaks, this factor contains many other HOM monomer peaks.

This factor rises at around 05:00 and reaches a maximum between 11:00 and 15:00 (Fig. 6). Figure 10c shows that the time series of this factor and sulfuric acid are very similar. In cloudy days (UVB $< 0.2\,\mathrm{W\,m^{-2}}$), the intensity of this factor is near 0. Note that this factor tracks sulfuric acid better than solar radiation. For example, the solar radiation was similar on 7 and 8 April, whereas the factor's intensity was much lower on 8 April, similar to the variation of sulfuric acid. Due to this reason, we interpret this factor as daytime oxidation of monoterpene controlled by OH, though NO must also be involved because the single highest peak is an organonitrate. Also, note that the participation of O_3 cannot be entirely excluded.

Daytime type-3 factor

The daytime type-3 factor shows maximum intensity in the afternoon around 14:00 (Fig. 6). Fingerprint peaks in this factor are organonitrate HOMs with smaller molecule weight, such as 213 Th ($C_3H_5O_3NO_3$), 241 Th ($C_4H_5O_4NO_3$), 255 Th ($C_5H_7O_4NO_3$), 269 Th ($C_6H_9O_4NO_3$), and 281 Th ($C_7H_9O_4NO_3$). Indicated by the smaller carbon number in the molecules, these light HOMs could come from anthropogenic VOCs (i.e., benzene and toluene). However, this possibility seems unlikely since the intensity of this factor does not show any significant increase during the period with transported pollution (9–12 April) when presumably benzene and toluene concentration were elevated. Another possibility is that these compounds are fragments from the oxidation of larger VOCs (e.g., monoterpene), and the presence of some HOM monomer peaks in this factor seems to support this assumption. This factor shows a good correlation with UVB (see Fig. 10d, and Table 2), indicating the HOM formation pathway represented by this factor is probably OH-initiated. Though the fingerprint peaks in this factor are organonitrates, the temporal variation of this factor shows no dependence on NO concentration. Instead, it exhibits a similar pattern with temperature, as shown in Fig. 10d. One possible explanation is that these small HOM molecules are relatively more volatile, so that their aerosol–gas partitioning is strongly affected by temperature – higher temperature leads to less condensation and high gas-phase concentration.

Table 2. Suggested HOM formation pathways represented by each factor and the uncentered correlation coefficients between factors and other relevant conditions. For each factor or relevant parameters, 1491 data points (30 min time resolution) are used. * These species cannot be ruled out.

Factors	Suggested main oxidant	Suggested main RO_2 terminator	correlation coefficient (UC, $n = 1491$)			
			NO	H_2SO_4	UVB	T
Nighttime type-1	O_3	RO_2	−0.05	0.19	0.14	0.26
Nighttime type-2	NO_3	RO_2	−0.07	0.18	0.06	0.33
Daytime type-1	O_3	NO (*HO_2)	0.38	0.56	0.50	0.42
Daytime type-2	OH (*O_3)	NO (*HO_2)	0.22	0.76	0.84	0.69
Daytime type-3	OH (*O_3)	NO (*HO_2)	0.29	0.56	0.66	0.80
Transport factor	–	–	0.41	0.48	0.44	0.51

4.3.3 Transport factor

According to the mathematical diagnostics discussed in Sect. 4.2, the third factor is important for the model to account for a significant fraction of the variability in the ambient data. The only prominent peak in this factor is nitrophenol ($C_6H_5NO_3$, 201 Th), a tracer for biomass burning suggested by previous studies (e.g., Mohr et al., 2013). The temporal behavior of this factor is similar to SO_2, both showing a significant enhancement during the period of 9–12 April, when the measurement site was influenced by polluted air masses coming from eastern Europe (see Fig. S11). We therefore suggest that this factor is a signature of transported pollution from biomass burning from continental areas.

4.4 Implication for atmospheric chemistry

In principle, in the atmosphere the formation pathway of HOM molecules involves addition of multiple O_2 molecules via autoxidation, including one oxidation initiation (by O_3, NO_3, or OH) and one termination reaction (mainly by NO, HO_2, or RO_2). Each pathway serves as a HOM source, leading to distinct profiles of HOM products for a specific VOC, with the overall HOM profile being a superposition of multiple pathways, depending on each source intensity. In practice, the relative importance of these pathways is highly dependent on atmospheric conditions. Table 2 lists suggested formation pathways for each factor, together with their correlation coefficients with other relevant measurements. Nighttime type-1 and nighttime type-2 likely represent monoterpene oxidation initiated by two major nighttime atmospheric oxidants, O_3 and NO_3, respectively. Indicated by high dimer concentrations from $RO_2 + RO_2$ reaction, RO_2 is the main terminator for both of them, probably because HO_2 and NO concentration is comparatively low in the nighttime. Both nighttime factors anti-correlate with NO. The daytime type-1 factor probably represents O_3-initiated oxidation followed by NO termination, and among all "local" factors it has the best correlation with NO. Daytime type-2 factor has the best correlation with H_2SO_4 and UVB and preassembly also with

OH. Though the exact chemistry producing the daytime type-2 factor is unclear, its clear dependence on OH indicates the oxidative pathways have been shifted from dark chemistry (O_3- or NO_3-initiated oxidation) to photochemistry (OH-initiated oxidation). Some initiator–terminator combinations are not found in PMF solutions, which may indicate their minor contributions to HOM production. For example, the combination of "OH-initiation" and "RO_2-termination" may not exist because, in the daytime, NO and HO_2 are much more efficient in terminating RO_2. Similarly, a pathway of "NO_3-initiation" followed by "NO termination" might be less likely, probably because NO is titrated by O_3 in the night and NO_3 hardly exists in the daytime.

5 Conclusion

HOMs have been confirmed by recent studies as significant sources of secondary organic aerosol, and thus understanding their formation pathways is relevant to atmospheric aerosol chemistry. This paper reports the success of applying PMF to differentiate HOMs originated from different sources in a boreal forest environment.

HOMs were measured with a CI-APi-TOF using nitrate ions for charging. Since the high-resolution peak fitting may introduce uncertainties that are not well quantified, we input unit-mass-resolution data as the data matrix and identify certain peaks with high-resolution afterwards. The error matrix is equally important to the data signal levels as an input parameter in PMF. In this work, errors were estimated from laboratory data by fitting the statistical uncertainty to the signal strength. The estimate shows good agreement with both that derived from an independent statistical analysis of the ambient data and with an approach widely used for aerosol mass spectrometrical data.

Mathematical diagnostics suggest that the error estimation is proper and that the model results are robust, although we did observe large variation of the Q/Q_{exp} value over masses and time, which suggests that some variation in the data was still not fully captured by the model. We note that the ab-

solute value of Q/Q_{exp} may not be a good parameter to alone evaluate the PMF performance, but relative changes are still very useful. For example, we observed a large decrease in Q/Q_{exp} when using a three-factor solution compared to a two-factor solution, suggesting the importance of the third factor, identified as the "transport" factor. This was supported by the three-factor solution being able to explain most ($> 95\%$) of the observed spectral and temporal variations.

In respect to the interpretability, the data are optimally explained by six factors. In the six-factor solution, two nighttime factors likely represent the oxidation of monoterpene initiated by O_3 and NO_3, respectively. The profile of the O_3 + monoterpene factor is similar to the reference spectrum in previous chamber studies where only O_3 and monoterpenes were injected, and the uncentred correlation coefficient between the factor and the reference spectrum is 0.91. The NO_3 + monoterpene reaction channel is supported by the detection of nitrogen-containing dimer compounds. In the early morning, both nighttime chemistry channels are suppressed by NO reaction, shown by the appearance of factors representing RO_2 + NO reactions. The major peaks in the first daytime factor are $C_{10}H_{15}O_{6,8}NO_3$, whose parent RO_2 radicals are likely from O_3 + monoterpene. Two other daytime factors are retrieved, though the underlying chemical processes forming those components are not clearly understood. One daytime factor correlated well with sulfuric acid, suggesting the chemistry represented by this factor could be controlled by the OH radical. The third daytime factor contained many smaller HOM molecules and showed notable correlation with UVB and temperature. The interpretation is that the formation of these smaller HOM molecules are OH-initiated, and their gas-phase concentration is affected by temperature probably through particle-gas partitioning. Apart from these five "local" factors, the sixth factor is interpreted as a transport factor due to its similar temporal variation to SO_2 and its prominent peak $C_6H_5NO_3$, a reported tracer of biomass burning.

Among the six factors retrieved by PMF, only the nighttime type-1 factor (O_3 + monoterpene) has been confirmed in the laboratory. However, the retrieval of this factor also strongly supports the validity of the model results. The deduced chemical processes for the nighttime type-2 factor (NO_3 + monoterpene) and the daytime type-1 factor (RO_2 + NO) are supported by their correlations with other co-located measurements. To confirm and better understand these two factors, laboratory experiments are needed to investigate the yields and dependence on other parameters. The daytime factors are harder to interpret. However, testing the hypotheses suggested by PMF solutions will be a good starting point for future studies. In summary, running PMF on CI-APi-TOF data was successful, and the results presented in this paper improve our understanding of HOM production by confirming current knowledge and inspiring future research directions.

6 Data availability

The PMF input data, output data for the optimal solution, sulfuric acid concentration, and total HOM concentration are available at https://etsin.avointiede.fi/dataset/urn-nbn-fi-csc-kata20161006183507547233. The trace gas data, e.g., SO_2 and O_3 concentrations, and meteorological data can be downloaded from http://avaa.tdata.fi/web/smart/smear/download. For the raw mass spectrometer data, please contact the first author via email: chao.yan@helsinki.fi.

The Supplement related to this article is available online at doi:10.5194/acp-16-12715-2016-supplement.

References

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, J. Geophys. Res.-Oc. Atmos., 108, 4091, doi:10.1029/2002JD002359, 2003.

Atkinson, R. and Arey, J.: Atmospheric degradation of volatile organic compounds, Chem. Rev., 103, 4605–4638, 2003.

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 + NO_3 system: effect of humidity and peroxy radical fate, Atmos. Chem. Phys., 15, 7497–7522, doi:10.5194/acp-15-7497-2015, 2015.

Canonaco, F., Crippa, M., Slowik, J. G., Baltensperger, U., and Prévôt, A. S. H.: 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, Atmos. Meas. Tech., 6, 3649–3661, doi:10.5194/amt-6-3649-2013, 2013.

Crippa, M., Canonaco, F., Slowik, J. G., El Haddad, I., DeCarlo, P. F., Mohr, C., Heringa, M. F., Chirico, R., Marchand, N., Temime-Roussel, B., Abidi, E., Poulain, L., Wiedensohler,

A., Baltensperger, U., and Prévôt, A. S. H.: Primary and secondary organic aerosol origin by combined gas-particle phase source apportionment, Atmos. Chem. Phys., 13, 8411–8426, doi:10.5194/acp-13-8411-2013, 2013.

Ding, A., Wang, T., and Fu, C.: Transport characteristics and origins of carbon monoxide and ozone in Hong Kong, South China, J. Geophys. Res.-Oc. Atmos., 118, 9475–9488, 2013.

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, Th. F.: Gas phase formation of extremely oxidized pinene reaction products in chamber and ambient air, Atmos. Chem. Phys., 12, 5113–5127, doi: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., Kurtén, T., Nielsen, L. B., Jørgensen, S., Kjaergaard, H. G., Canagaratna, M., Maso, M. D., Berndt, T., Petäjä, 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, 2014.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, Th. F., Monod, A., Prévôt, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmos. Chem. Phys., 9, 5155–5236, doi:10.5194/acp-9-5155-2009, 2009.

Hari, P. and Kulmala, M.: Station for measuring ecosystem-atmosphere relations, Boreal Environ. Res., 10, 315–322, 2005.

IPCC: Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK, and New York, NY, USA, 2013.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of Organic Aerosols in the Atmosphere, Science, 326, 1525–1529, doi:10.1126/science.1180353, 2009.

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, Atmos. Chem. Phys., 12, 4117–4125, doi:10.5194/acp-12-4117-2012, 2012.

Jokinen, T., Sipilä, M., Richters, S., Kerminen, V. M., Paasonen, P., Stratmann, F., Worsnop, D., Kulmala, M., Ehn, M., and Herrmann, H.: Rapid Autoxidation Forms Highly Oxidized RO_2 Radicals in the Atmosphere, Angew. Chem., Int. Ed., 53, 14596–14600, 2014.

Jokinen, T., Berndt, T., Makkonen, R., Kerminen, V.-M., Junninen, H., Paasonen, P., Stratmann, F., Herrmann, H., Guenther, A. B., and Worsnop, D. R.: Production of extremely low volatile organic compounds from biogenic emissions: Measured yields and atmospheric implications, P. Natl. Acad. Sci. USA, 112, 7123–7128, 2015.

Junninen, H., Ehn, M., Petäjä, T., Luosujärvi, L., Kotiaho, T., Kostiainen, R., Rohner, U., Gonin, M., Fuhrer, K., Kulmala, M., and Worsnop, D. R.: A high-resolution mass spectrometer to measure atmospheric ion composition, Atmos. Meas. Tech., 3, 1039–1053, doi:10.5194/amt-3-1039-2010, 2010.

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, Science, 339, 943–946, 2013.

Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prévôt, A. S. H.: Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, Atmos. Chem. Phys., 7, 1503–1522, doi:10.5194/acp-7-1503-2007, 2007.

Liao, L., Maso, M. D., Taipale, R., Rinne, J., Ehn, M., Junninen, H., Äijälä, M., Nieminen, T., Alekseychik, P., and Hulkkonen, M.: Monoterpene pollution episodes in a forest environment: indication of anthropogenic origin and association with aerosol particles, Boreal Environ. Res., 16, 288–303, 2011.

Manninen, H. E., Nieminen, T., Asmi, E., Gagné, S., Häkkinen, S., Lehtipalo, K., Aalto, P., Vana, M., Mirme, A., Mirme, S., Hõrrak, U., Plass-Dülmer, C., Stange, G., Kiss, G., Hoffer, A., Töro, N., Moerman, M., Henzing, B., de Leeuw, G., Brinkenberg, M., Kouvarakis, G. N., Bougiatioti, A., Mihalopoulos, N., O'Dowd, C., Ceburnis, D., Arneth, A., Svenningsson, B., Swietlicki, E., Tarozzi, L., Decesari, S., Facchini, M. C., Birmili, W., Sonntag, A., Wiedensohler, A., Boulon, J., Sellegri, K., Laj, P., Gysel, M., Bukowiecki, N., Weingartner, E., Wehrle, G., Laaksonen, A., Hamed, A., Joutsensaari, J., Petäjä, T., Kerminen, V.-M., and Kulmala, M.: EUCAARI ion spectrometer measurements at 12 European sites – analysis of new particle formation events, Atmos. Chem. Phys., 10, 7907–7927, doi:10.5194/acp-10-7907-2010, 2010.

Mohr, C., Lopez-Hilfiker, F. D., Zotter, P., Prévôt, A. S., Xu, L., Ng, N. L., Herndon, S. C., Williams, L. R., Franklin, J. P., and Zahniser, M. S.: Contribution of nitrated phenols to wood burning brown carbon light absorption in Detling, United Kingdom during winter time, Environ. Sci. Technol., 47, 6316–6324, 2013.

Mutzel, A., Poulain, L., Berndt, T., Iinuma, Y., Rodigast, M., Böge, O., Richters, S., Spindler, G., Sipilä, M., and Jokinen, T.: Highly oxidized multifunctional organic compounds observed in tropospheric particles: A field and laboratory study, Environ. Sci. Technol., 49, 7754–7761, 2015.

Nel, A.: Air pollution-related illness: effects of particles, Science, 308, 804–806, 2005.

Ng, N. L., Canagaratna, M. R., Zhang, Q., Jimenez, J. L., Tian, J., Ulbrich, I. M., Kroll, J. H., Docherty, K. S., Chhabra, P. S., Bahreini, R., Murphy, S. M., Seinfeld, J. H., Hildebrandt, L., Donahue, N. M., DeCarlo, P. F., Lanz, V. A., Prévôt, A. S. H., Dinar, E., Rudich, Y., and Worsnop, D. R.: Organic aerosol components observed in Northern Hemispheric datasets from Aerosol Mass Spectrometry, Atmos. Chem. Phys., 10, 4625–4641, doi:10.5194/acp-10-4625-2010, 2010.

Orlando, J. J. and Tyndall, G. S.: Laboratory studies of organic peroxy radical chemistry: an overview with emphasis on recent issues of atmospheric significance, Chem. Soc. Rev., 41, 6294–6317, 2012.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemom. Intell. Lab. Syst., 37, 23–35, 1997.

Paatero, P.: User's guide for positive matrix factorization programs PMF2 and PMF3, Helsinki, University of Helsinki, 2007.

Paatero, P.: Interactive comment on "Source characterization of Highly Oxidized Multifunctional Compounds in a Boreal Forest Environment using Positive Matrix Factorization" by Chao Yan et al., Atmos. Chem. Phys. Discuss., C1–C12, doi:10.5194/acp-2016-367-RC2, 2016.

Paatero, P. and Hopke, P. K.: Discarding or downweighting high-noise variables in factor analytic models, Anal. Chim. Acta, 490, 277–289, 2003.

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.

Polissar, A. V., Hopke, P. K., Paatero, P., Malm, W. C., and Sisler, J. F.: Atmospheric aerosol over Alaska: 2. Elemental composition and sources, J. Geophys. Res.-Oc. Atmos., 103, 19045–19057, 1998.

Rantala, P., Taipale, R., Aalto, J., Kajos, M. K., Patokoski, J., Ruuskanen, T. M., and Rinne, J.: Continuous flux measurements of VOCs using PTR-MS—reliability and feasibility of disjunct-eddy-covariance, surface-layer-gradient, and surface-layer-profile methods, Boreal Environ. Res., 19, 87–107, 2014.

Riccobono, F., Schobesberger, S., Scott, C. E., Dommen, J., Ortega, I. K., Rondo, L., Almeida, J., Amorim, A., Bianchi, F., and Breitenlechner, M.: Oxidation products of biogenic emissions contribute to nucleation of atmospheric particles, Science, 344, 717–721, 2014.

Riipinen, I., Pierce, J. R., Yli-Juuti, T., Nieminen, T., Häkkinen, S., Ehn, M., Junninen, H., Lehtipalo, K., Petäjä, T., Slowik, J., Chang, R., Shantz, N. C., Abbatt, J., Leaitch, W. R., Kerminen, V.-M., Worsnop, D. R., Pandis, S. N., Donahue, N. M., and Kulmala, M.: Organic condensation: a vital link connecting aerosol formation to cloud condensation nuclei (CCN) concentrations,

Atmos. Chem. Phys., 11, 3865–3878, doi:10.5194/acp-11-3865-2011, 2011.

Rissanen, M. P., Kurtén, T., Sipilä, M., Thornton, J. A., Kangasluoma, J., Sarnela, N., Junninen, H., Jørgensen, S., Schallhart, S., and Kajos, M. K.: The Formation of Highly Oxidized Multifunctional Products in the Ozonolysis of Cyclohexene, J. Am. Chem. Soc., 136, 15596–15606, 2014. doi:10.5194/acp-10-1969-2010, 2010.

Schobesberger, S., Junninen, H., Bianchi, F., Lönn, G., Ehn, M., Lehtipalo, K., Dommen, J., Ehrhart, S., Ortega, I. K., and Franchin, A.: Molecular understanding of atmospheric particle formation from sulfuric acid and large oxidized organic molecules, P. Natl. Acad. Sci. USA, 110, 17223–17228, 2013.

Spracklen, D. V., Jimenez, J. L., Carslaw, K. S., Worsnop, D. R., Evans, M. J., Mann, G. W., Zhang, Q., Canagaratna, M. R., Allan, J., Coe, H., McFiggans, G., Rap, A., and Forster, P.: Aerosol mass spectrometer constraint on the global secondary organic aerosol budget, Atmos. Chem. Phys., 11, 12109–12136, doi:10.5194/acp-11-12109-2011, 2011.

Tröstl, 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., Kürten, A., Laaksonen, A., Lawler, M., Leiminger, M., Mathot, S., Möhler, O., Nieminen, T., Onnela, A., Petäjä, T., Piel, F. M., Miettinen, P., Rissanen, M. P., Rondo, L., Sarnela, N., Schobesberger, S., Sengupta, K., Sipilä, M., Smith, J. N., Steiner, G., Tomè, 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, doi: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, doi:10.5194/acp-9-2891-2009, 2009.

Ulevicius, V., Bycenkiene, S., Bozzetti, C., Vlachou, A., Plauškaite, K., Mordas, G., Dudoitis, V., Abbaszade, G., Remeikis, V., Garbaras, A., Masalaite, A., Blees, J., Fröhlich, R., Dällenbach, K. R., Canonaco, F., Slowik, J. G., Dommen, J., Zimmermann, R., Schnelle-Kreis, J., Salazar, G. A., Agrios, K., Szidat, S., El Haddad, I., and Prévôt, A. S. H.: Fossil and non-fossil source contributions to atmospheric carbonaceous aerosols during extreme spring grassland fires in Eastern Europe, Atmos. Chem. Phys., 16, 5513–5529, doi:10.5194/acp-16-5513-2016, 2016.

Vlasenko, A., Slowik, J., Bottenheim, J., Brickell, P., Chang, R. W., Macdonald, A., Shantz, N., Sjostedt, S., Wiebe, H., and Leaitch, W.: Measurements of VOCs by proton transfer reaction mass spectrometry at a rural Ontario site: Sources and correlation to aerosol composition, J. Geophys. Res.-Oc. Atmos., 114, D21305, doi:10.1029/2009JD012025, 2009.

Yuan, B., Shao, M., Gouw, J., Parrish, D. D., Lu, S., Wang, M., Zeng, L., Zhang, Q., Song, Y., and Zhang, J.: Volatile organic compounds (VOCs) in urban air: How chemistry affects the interpretation of positive matrix factorization (PMF) analysis, J. Geophys. Res.-Oc. Atmos., 117, D24302, doi:10.1029/2012JD018236, 2012.

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, Anal. Bioanal. Chem., 401, 3045–3067, 2011.