

# Final response to the referees' comments for Olin et al.: "Measurement report: Atmospheric new particle formation in a coastal agricultural site explained with binPMF analysis of nitrate CI-API-TOF spectra"

We thank the referees for their very useful comments and have corrected the manuscript according to all of them.

Referee reports are in *black italic* and authors' responses in **blue roman** font. **Bold blue** or ~~struck-through blue~~ fonts highlight changed text parts in some comments. The marked-up manuscript and Supplement highlighting the changes are included at the end of this file.

## Referee comment #1:

*General comment:*

*In this measurement report, "Measurement report: Atmospheric new particle formation in a coastal agricultural site explained with binPMF analysis of nitrate CI-API-TOF spectra" Olin et al. performed nitrate-CIMS measurement at a coastal site. BinPMF was applied to the CIMS data, reporting two factors- F7 and F8, with high mass-to-charge ratios factor, which explained the NPF events well. It is an interesting study, and the results appear to support the conclusion. I have some minor comments the authors may want to consider before publication. My biggest concern is the uncertainties of the measurement as it can lead to some problems when comparing the intensities of different factors.*

*Specifics:*

**1. Abstract.** *You may want to associate F7 or F8 with a more physically meaningful name, as it is hard to comprehend what is the chemistry behind F7/F8.*

We agree that the text in the abstract was not clear enough to understand the chemistry behind the factors 7 and 8, which was actually not very deeply examined in this measurement report. Therefore, we decided to keep their original names, *F7(day-N-HOMs)* and *F8(morning-N-HOMs)*, because they only denote how the factors appeared in the data, which can be provided with much higher certainty than any chemistry-related names with the level of analysis in this measurement report. We also didn't want to mention their names already in the abstract; instead, the text in the abstract is now clarified. The sentence near the middle part of the abstract is now updated to "... both having profiles with patterns of highly oxygenated organic molecules **with one nitrogen atom.**" and the last sentence updated to "It was also observed that the factor 8 acts as ~~a precursor of~~ **like a precursor for** the factor 7 with ~~high solar irradiance~~ **solar radiation and that the formation of the factor 8 is associated with ozone levels.**" The change of wording from "as a precursor" to "like a precursor" is related to the referee #1 comment 19 and the referee #2 comment 3.

**2. Line 73.** *Provide more details about how the instrument works in principle. E.g., how the agent ion was produced, and at what rate? What is the flow rate being sampled into the CIMS?*

Descriptions of the generation of the reagent ion and of the flows of the CI-inlet are now added to the text: " **$\text{NO}_3^-$  was produced by adding nitric acid ( $\text{HNO}_3$ ) vapor to the sheath air which was then exposed to X-ray radiation inside the CI-inlet. The  $\text{HNO}_3$  vapor was produced by directing 10 sccm flow of the sheath air through a glass vial containing liquid  $\text{HNO}_3$ . The sheath air was a 20 slpm flow of outdoor air filtered using an activated carbon filter and a HEPA filter. The outdoor air sample was drawn to the CI-inlet with the flow rate of 8.3 slpm through a side hatch of the van with a 16 mm pipe of 60 cm length, which also acts as a laminarization tube for the CI-inlet.**"

**3. Line 79. Any specific reason for averaging to 600 s? Is the code open-sourced, available at what site?**

The averaging time needs to be long enough to have an adequate signal-to-noise ratio but shorter than the timescale of expected variations in the signal. At this agricultural environment, no rapid variations in the signal are expected and 600 s is a best estimate for the suitable averaging time in this case. This is now clarified in the text with the updated sentence “The data were recorded in 4 s time resolution but were first averaged to 600 s resolution ( $t_a$ ) **in order to increase signal-to-noise ratios (SNRs) but still maintain an adequate time resolution for the campaign.**”

The tofTools code is not fully open-sourced and it has been developed by many authors from many institutions. It has formerly been openly available, but its homepage has not been operated anymore. Therefore, no download link cannot be given for it, but the latest codes may be obtained by contacting the main developer, Prof. Heikki Junninen from University of Tartu.

**4. Line 80. What are the errors for the  $m/z$  calibration?**

The errors of the mass calibration during the campaign were 0.4 ppm, in median. However, there were eight nights near the end of the campaign with the errors up to 80 ppm. This was caused by the high signals from  $\text{H}_2\text{CO}_3$  causing overlapping peaks to the  $\text{NO}_3^-$  and  $\text{NO}_3^- \cdot \text{HNO}_3$  peaks used in the calibration. Nevertheless, that didn't actually cause any excessive errors to the result of the mass calibration, since the mass calibration parameters did not include any significant deviation during those nights. Only the choice of the peaks for the calibration was not very optimal for those nights. These values and this discussion is now included in the text (in the results part, in Sect. 4.2.2) with the added text of “**The median error of the  $m/z$  calibration during the campaign was 0.4 ppm. However, the errors were as high as 80 ppm during eight nights near the end of the campaign due to high  $\text{H}_2\text{CO}_3$  signals causing overlapping signals to the  $\text{NO}_3^-$  and  $\text{NO}_3^- \cdot \text{HNO}_3$  peaks used in the calibration. Nevertheless, the result of the calibration, which is the parameters used to convert time-of-flight to  $m/z$ , did not include any significant deviation during those nights, implying that the calibration itself was successful although the choice of the peaks for the calibration were not optimal for those nights.**”

**5. Line 85-90. I assume the  $C$  value is obtained by your calibration, while  $P$  is from the literature? Just wondering if the  $C$  and  $P$  values would vary a lot over different times and with different instruments.**

The  $C$  value is obtained from three calibration runs, performed once a month at the measurement site. The highest obtained  $C$  was 31% higher than the lowest one. Therefore, the variation in  $C$  over different times is not very high. This level of variation is now mentioned in the text by expressing the value of  $C$  in the form of “ $(2.45 \pm 0.33) \times 10^9 \text{ cm}^{-3}$ ” instead of only “ $2.45 \times 10^9 \text{ cm}^{-3}$ ”. Additionally, “(once a month)” is now mentioned in the text.

The  $P$  value is obtained by calculating the diffusional losses of  $\text{H}_2\text{SO}_4$  inside a 60 cm long tube having the flow rate of 8.3 slpm using the function by Gormley and Kennedy (1948). The obtained penetration efficiency of 0.58 includes a minor uncertainty of 0.03, which results from the uncertainties in the measurements of the sampling pipe length and of the flow rate and from the variation in relative humidity (RH) of the outdoor air, since the diffusion coefficient of  $\text{H}_2\text{SO}_4$  is effectively dependent on RH due to hydration of the  $\text{H}_2\text{SO}_4$  molecules. This information is now included in the text with the updated sentence “ $P$  is the penetration efficiency of  $\text{H}_2\text{SO}_4$  in the sampling pipe (=0.58) **60 cm long sampling pipe having the flow rate of 8.3 slpm (=  $0.58 \pm 0.03$ ) determined with the function by Gormley and Kennedy (1948).**”

**6. Line 90. I am not familiar with nitrate CIMS calibration, can you explicitly explain what “penetration efficiency” is?**

The 60 cm long sampling pipe is used to make the flow to the CI-inlet laminar, required for its optimal operation. The length of the sampling pipe and the flow rate in it vary between campaigns, but the calibration source (producing a known  $\text{H}_2\text{SO}_4$  concentration) is typically placed right at the front of the CI-inlet, i.e., downstream of the sampling pipe. Therefore,  $C$  is used to convert the signal to the  $\text{H}_2\text{SO}_4$  concentration at the front of the CI-inlet. The concentration at the front of the sampling pipe is, however, different because a part of the  $\text{H}_2\text{SO}_4$  molecules are lost onto the inner walls of the sampling tube via diffusion. With the used setup,  $P$  was 0.58, denoting that 58% of  $\text{H}_2\text{SO}_4$  molecules penetrates the sampling tube while 42% are lost onto the walls. This is now clarified in the text with the updated sentence mentioned with the previous comment. Some studies report  $\frac{C}{P}$  as a corrected calibration coefficient, which would be  $4.22 \times 10^9 \text{ cm}^{-3}$  in this study.

*7. Line 105. Elaborate on the uncertainties for the determination of H<sub>2</sub>SO<sub>4</sub>, HIO<sub>3</sub>, and CH<sub>4</sub>SO<sub>3</sub>.*

The uncertainty in the H<sub>2</sub>SO<sub>4</sub> concentration is caused by the uncertainties in *C* and *P*, resulting in the relative uncertainty in the H<sub>2</sub>SO<sub>4</sub> concentration of 15 %, when the uncertainty in *C* is determined with its variation between three calibration runs during the campaign. There is, however, a systematic uncertainty in *C* involved in a single calibration run (Kürten et al., 2012) but it is not easily quantified. A frequently used rough approximation for the uncertainty in the determined H<sub>2</sub>SO<sub>4</sub> concentrations is  $-50\%/+100\%$ . This discussion is now included with the added text “[H<sub>2</sub>SO<sub>4</sub>] has a relative uncertainty of 15 %, originating mostly from the uncertainty in *C*, determined with its variation between three calibration runs during the campaign. However, because a single calibration run also involves its own systematic uncertainty (Kürten et al., 2012)—which is not easily quantified—the uncertainty in [H<sub>2</sub>SO<sub>4</sub>] can be higher (a frequently used rough approximation is  $-50\%/+100\%$ ).”

The uncertainties in the concentrations of HIO<sub>3</sub> and CH<sub>4</sub>SO<sub>3</sub> during the campaign are basically also 15 % (or even  $-50\%/+100\%$ ), similarly to H<sub>2</sub>SO<sub>4</sub>, but their absolute levels can, however, be different in reality because *C* and *P* have not been determined for them. Nevertheless, the absolute levels do not affect the results of this study, since the differences between *C* and *P* for different compounds are expected to be constant over different times. Additionally, in the case of HIO<sub>3</sub> and CH<sub>4</sub>SO<sub>3</sub>, *C* values for them are also expected to be near the *C* value for H<sub>2</sub>SO<sub>4</sub> because they all have collision-limited charging efficiency when reacting with NO<sub>3</sub><sup>-</sup> ions (Simon et al., 2020; Beck et al., 2021; Wang et al., 2021). The *P* value is also not highly dependent on the compound within the range of compounds in question in this study. This discussion is now included in the text with the sentences “Uncertainties in [HIO<sub>3</sub>] and in [CH<sub>4</sub>SO<sub>3</sub>] are not easily quantified but the relative uncertainty of 15 % in [H<sub>2</sub>SO<sub>4</sub>] (or  $-50\%/+100\%$ ) can be used in estimating relative uncertainties during the campaign. However, the absolute levels of [HIO<sub>3</sub>] and [CH<sub>4</sub>SO<sub>3</sub>] can still differ more due to the approximation of using *C* and *P* determined for H<sub>2</sub>SO<sub>4</sub>, but the difference between the calibration coefficients for different compounds is expected to be nearly constant during the campaign.”

*8. Line 108, “Ions smaller than 169 Th were omitted because there are many organic compounds that are unlikely the key compounds in NPF and have relatively high signals possibly causing issues in the binPMF analysis” It is not clear why many organic compounds are unlikely the key compounds in NPF.*

This is now clarified in the text with the added sentences “Compounds with lower masses typically have higher volatilities and are thus more unlikely participating in nucleation and condensation. However, it is not generally true since, e.g., H<sub>2</sub>SO<sub>4</sub> and CH<sub>4</sub>SO<sub>3</sub> are below 169 Th but still known to contribute to NPF. Therefore, they were analyzed separately of the binPMF analysis.”

*9. Line 121. It is not clear why X needs to be normalized and how it is normalized? Is the signal used to calculate H<sub>2</sub>SO<sub>4</sub> concentrations normalized?*

The signal used to calculate the H<sub>2</sub>SO<sub>4</sub> concentration is normalized with the total signal of nitrate ions. The normalization is done with the denominator in Eq. (1). This is needed because signals have always to be expressed relative to the reagent ion signal, which is not very constant due to changes in temperature and the level of liquid HNO<sub>3</sub> in the glass vial. The normalized signal is a dimensionless variable, which is then converted to the concentration-based unit by multiplying it with a calibration coefficient. Similarly to the H<sub>2</sub>SO<sub>4</sub> concentration, the data matrix X needs to be normalized with the reagent ion signal in order to prevent variations in the reagent ion signal to result in erroneous variations in the data of interest. This way, the data matrix X will also be expressed dimensionlessly, and the result of the binPMF run, the time series matrix TS, can also be expressed in a concentration-based unit (after multiplying it with  $\frac{C}{P}$ ). This information is now included in the text with the updated sentences “The data matrix X was normalized before running the binPMF code with the reagent ion signals as is done with the denominator in Eq. (1) for calculating [H<sub>2</sub>SO<sub>4</sub>]. This way the data matrix becomes dimensionless and expressed relative to the NO<sub>3</sub><sup>-</sup> concentration.”

The uncertainty matrix S needs to be expressed in the same unit as X and is thus normalized also with the nitrate ion signal. This is also now mentioned in the text with the updated sentence “Similar to the matrix X, the matrix S was also normalized with the reagent ion signals before running the binPMF code to have them expressed with the same units (dimensionless).”

**10. Line 125.** I am slightly confused as it seems  $S_{ij}$  is associated with moving median, but in equation (4) no median values were used. Is the a value of 1.35 a reference value from literature? Or determined by what method specifically.

We agree that the construction of the uncertainty matrix  $S$  was not described clearly enough. Moving median values were used to determine the values for  $a$  and  $\sigma_{\text{noise}}$ . This was done by, first, determining uncertainties of the signals for all time bins and  $m/z$  bins separately using the deviations between the signals and their moving medians. Then, the obtained uncertainties were fitted to the Poisson distribution for all  $m/z$  bins separately but using all time bins. This results in Eq. (4) where  $a$  and  $\sigma_{\text{noise}}$  are functions of the  $m/z$  bin. Finally, weighted means of  $a$  and  $\sigma_{\text{noise}}$  over all  $m/z$  bins are calculated, which leads to their mentioned values with uncertainties ( $a = 1.35 \pm 0.22$  and  $\sigma_{\text{noise}} = (0.001 \pm 0.003)$  cps). This method is described in more detail in the study by Yan et al. (2016). The paragraph related to this is now rewritten and it now reads “The binPMF code tries to find the matrices  $TS$  and  $MS$  producing the lowest possible value for the sum of the scaled residuals,

$$Q = \sum_{i=1}^{9970} \sum_{j=1}^{6975} (R_{ij}/S_{ij})^2, \quad (3)$$

where  $S$  is the uncertainty matrix.  $S$  was estimated from the ambient data via the method suggested by Yan et al. (2016), i.e., through approximating instrument noise as the difference between the measured signal and its moving median over 5 data points. **The differences are determined separately for all 9970 time bins and for all 6975  $m/z$  bins.** Assuming that the counting statistics follow the Poisson distribution, the function

$$S_{ij} = \sigma_{ij} + \sigma_{\text{noise}} = a \cdot \sqrt{\frac{\max(I, \frac{1}{t_a})}{t_a}} + \sigma_{\text{noise}} \quad (4)$$

where  $a$  is the correcting factor incorporating any unaccounted contributions to the uncertainty (Allan et al., 2003) and  $I$  is the signal intensity in counts per second (cps), **can be described to express the instrument noise. The values for  $a$  and  $\sigma_{\text{noise}}$  were first determined for all  $m/z$  bins separately by fitting the differences between the measured signal and the moving median for different levels of  $I$  using all time bins and Eq. (4). The obtained  $a$  and  $\sigma_{\text{noise}}$  values were then averaged over all  $m/z$  bins using the successes of the fittings as weighting factors. Finally, the values of  $a = 1.35 \pm 0.22$  and  $\sigma_{\text{noise}} = (0.001 \pm 0.003)$  cps were obtained, and the function for calculating the uncertainty matrix became**

$$S_{ij} = 1.35 \cdot \sqrt{\frac{\max(I, \frac{1}{600\text{s}})}{600\text{s}}} + 0.001 \text{ cps}. \quad (5)$$

The value of  $a$  **obtained here**,  $1.35 \pm 0.22$ , is on a similar level to the values in the study by Yan et al. (2016),  $1.28 \pm 0.09$  or  $1.1 \pm 0.3$ . Similar to the matrix  $X$ , the matrix  $S$  was also normalized with the reagent ion signals before running the binPMF code **to have them expressed with the same units (dimensionless).**”

**11. Line 137.** Define “substantially”. A low  $Q/Q_{\text{exp}}$  is not necessarily the best PMF solution.

This sentence is now updated, to better understand how  $p$  was changed, to “The code was run with increasing  $p$  until the ratio  $Q/Q_{\text{exp}}$  ~~did not substantially decrease anymore~~ **set to a level with no further decrease.**”

In this study, the optimal  $p$  value of 8 was selected by considering the behavior of the  $Q/Q_{\text{exp}}$  ratios with different values of  $p$ , by minimizing  $p$  to keep the analysis as simple as possible, and by considering the number of factors having organic patterns. This didn’t lead to the lowest  $Q/Q_{\text{exp}}$  ratio but to near that. Additionally, because the results of PMF analyses, in general, include rotational ambiguity, we added the following sentences to the text: **“The lowest  $Q/Q_{\text{exp}}$  is not necessarily the best solution of PMF because the solution of PMF is not unique due to the nature of PMF. For example, due to its rotational ambiguity, there are basically infinite number of solutions providing equal  $Q$  values (Paatero et al., 2002). Rotational ambiguity is, however, not considered in this study because it is typically not a problem with ambient measurement data, which usually contain enough data points with values near zero (Zhang et al., 2019).”**

12. Line 145. Again, what is the uncertainty for the assumptions of similar calibration coefficient and penetration efficiency?

This is now discussed with the added text “**Penetration efficiencies are not expected to differ notably between  $\text{H}_2\text{SO}_4$  and the compounds with  $m/z$  of 169–450 Th, detected by nitrate CI-API-TOF, but calibration coefficients are.  $C$  for  $\text{H}_2\text{SO}_4$  denotes a value near the highest possible sensitivity, since the charging efficiency of  $\text{H}_2\text{SO}_4$  is collision-limited (Simon et al., 2020). However,  $C$  can be even orders of magnitude higher for compounds with weaker binding energies with  $\text{NO}_3^-$  (Hytinen et al., 2015). Therefore, the sensitivity of the used system can be much lower for some compounds. This is true especially for organic compounds with low oxygen-to-carbon (O:C) ratios, since organic compounds with adequately large number of carbon atoms having  $\text{O:C} \geq 0.6$  are estimated to have similar collision-limited charging efficiencies to  $\text{H}_2\text{SO}_4$  (Simon et al., 2020), but the binPMF factors of this study includes compounds with lower O:C ratios too. As it is practically impossible to determine calibration coefficients for all detectable compounds, this commonly used approximation is used here only to convert the dimensionless variables in TS to a practical concentration-based unit.**”

13. Line 155. Define “strong” and “weaker”. Do you mean particle number concentration?

The definition of the strength of a NPF event, “(the strength of a NPF event is defined as the duration of continuing NPF during a NPF event)”, was mentioned in Sect. 4.4 but is now moved already here to Sect. 4.1.

14. Line 157. How was CS determined? Maybe explain it in the Method section.

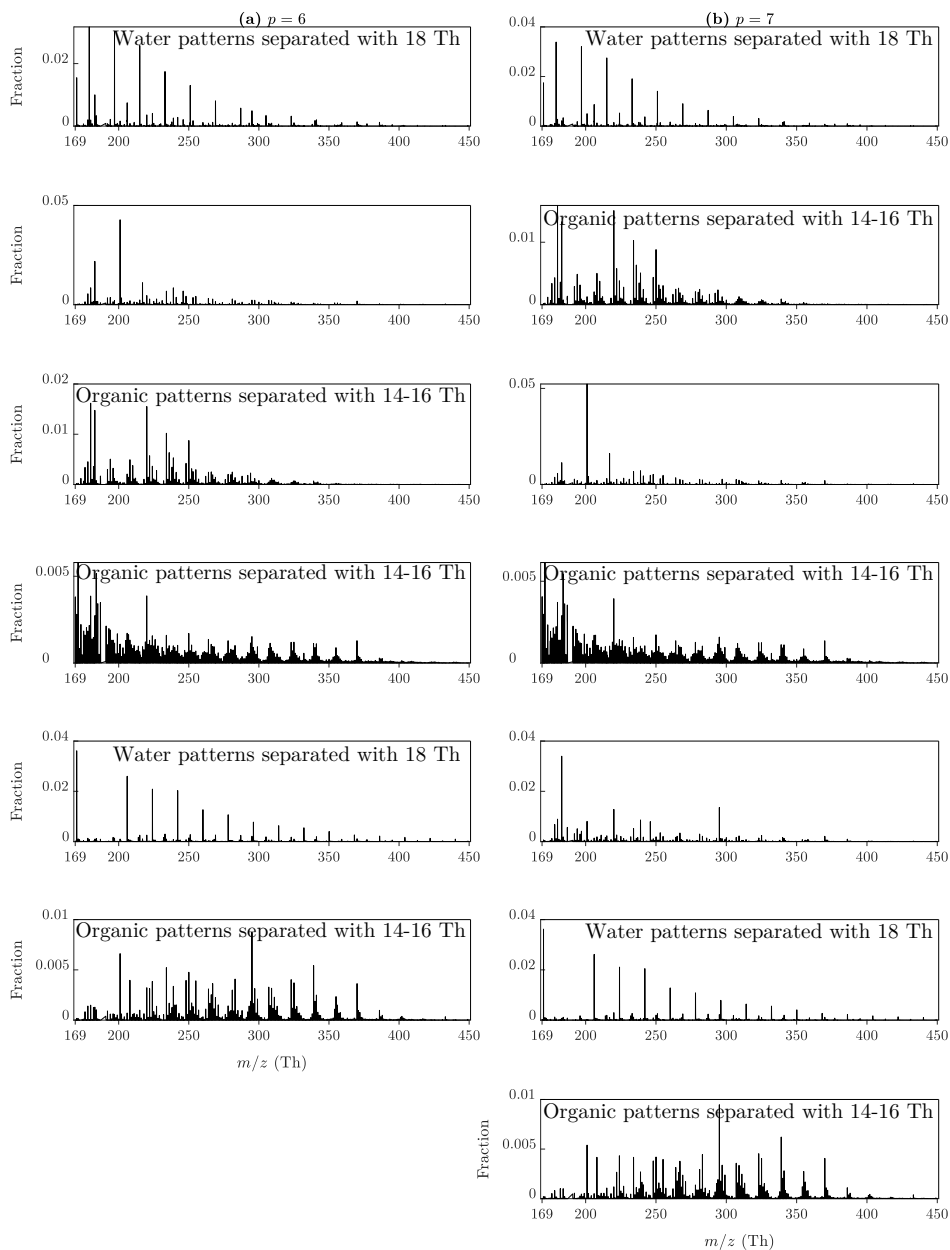
CS was calculated from the DMPS data with the method by Kulmala et al. (2012) using a common approximation of assuming  $\text{H}_2\text{SO}_4$  as the condensing vapor. This is now mentioned already in the “Measurement instruments” section with the updated sentence “DMPS was used to determine the particle size distribution in the size range of 6–823 nm **and CS, which was calculated with the method by Kulmala et al. (2012) using a common approximation of assuming  $\text{H}_2\text{SO}_4$  as the condensing vapor.**”

15. Line 175. It is hard to comprehend this sentence. Define “smallest particles” and rephrase “simultaneously not elevated”

This sentence is now clarified to “They originate probably from direct emission sources, e.g., from tractors harvesting the fields, rather than via NPF because ~~the abundance of the smallest particles is simultaneously not elevated~~ **there is no simultaneous concentration increase in particles smaller than those.**”

16. Line 193. Any data to support the statement of “(3 factor profiles having organic patterns with  $p = 6$  or  $p = 7$ )”.

The asked supporting data is now added to the Supplement as Fig. S1 and the text in question is now updated to “(3 **three** factor profiles having organic patterns with  $p = 6$  or  $p = 7$ , **see Fig. S1**)”.



**Figure S1. Normalized spectra of all binPMF factors if (a)  $p = 6$  or (b)  $p = 7$  were used. Both of these factor sets have three factor profiles having organic patterns, whereas  $p = 8$  results in four factor profiles having organic patterns and is further analyzed in this study.**

17. Line 280. *Has NetRad been mentioned before?*

Yes, it has been mentioned already in Sect. 2.2 with the text “net radiation (NetRad)”.

18. Line 353. *F8 and F7 may have different sensitivity in CIMS. You may need to discuss the uncertainties when comparing their intensities.*

Different sensitivities of the compounds detected by the CI-API-TOF have been discussed already in Sect. 3.2 (related to the referee #1 comment 12). Here, in Sect. 4.4, these possible differences are now considered by updating the text to “**Assuming similar sensitivities of F7(day-N-HOMs) and F8(morning-N-HOMs) in the CI-API-TOF**, the mean intensity of F8(morning-N-HOMs) is 2-fold the mean intensity of F7(day-N-HOMs) with low radiation levels ( $\text{NetRad} < 130 \text{ W m}^{-2}$ ); conversely, the mean intensity of F7(day-N-HOMs) is 3-fold the mean intensity of F8(morning-N-HOMs) with high radiation levels.”

19. Line 354. *I don't see why it confirms the transformation from F8 to F7 as F8 decreases while F7 increases. There are other possibilities e.g., transport of pollutants with different intensities, and changes in air masses.*

It is true that F8 decreasing and F7 increasing with radiation does not necessarily confirm the role of radiation in this process. However, by looking at the time series of F7, F8 and  $\text{O}_3$  in Fig. 7, it can be observed that when NetRad increases, F8 decreases and almost the same amount is added to F7, while their sum behaves very similar to the trend of the  $\text{O}_3$  concentration. Therefore, and due to the behavior of F7 and F8 with NetRad levels (see the previous comment), there is clearly something related to radiation and the abundance of F7 and F8. As mentioned by the referee, there are other possibilities, such changes in air masses, causing the behavior of F7 and F8 with radiation levels. The word “confirms” in the sentence in question is now replaced with the word “supports” to clarify that it was not fully confirmed and also the text “**It cannot be proven with these measurements that this transformation is actually a real chemical process since the behavior of the factors with NetRad levels can also be resulted, e.g., from possibly different origins of the air masses between the daytime and the nighttime. Nevertheless, the current knowledge cannot at least exclude the transformation via radiation, because the compounds in F7(day-N-HOMs) are basically more oxidized than the ones in F8(morning-N-HOMs) (discussed later in Sect. 4.4) being physically reasonable since photochemistry typically leads to oxidation.**” is now added to Sect. 4.3. See also the referee #2 comment 3, which is related to this as well.

20. Line 379. “... temperature to disfavoring them...” do you specifically mean high or low temperatures?

High temperatures were meant. The word “high” is now added.

21. Conclusion. *You may want to make the conclusion section short, highlighting the new findings in this study.*

The conclusions section is now shortened from ~44 lines to ~31 lines. Discussions about Pearson’s correlation coefficients with different particle sizes and about estimating the behavior of F7 and F8 in other studies using UMR data are now removed from there (see the marked-up manuscript for the changes in it).

22. Figure 1. *What is the red ban between Dp 3-8 nm?*

The red banners in that particle size range are caused by that the data for this particle size range is obtained from the NAIS, which has a general tendency to overestimate concentrations at those particle sizes (Gagné et al., 2011; Mirme and Mirme, 2013; Kangasluoma et al., 2020). This is now mentioned in the caption of Fig. 1 with the text “**The data from the NAIS show red banners due to its general tendency to overestimate concentrations at these particle sizes (Gagné et al., 2011; Mirme and Mirme, 2013; Kangasluoma et al., 2020).**”



23. Figure 7. The label/legends in this figure are small, you may want to make it visible at a font size of at least 8.

All the texts in Figs. 1 and 7 are now increased. Due to that, the legend for the bottom panels of Fig. 7 was needed to be moved from the (a) plot to the (c) plot; thus, the text “**the legend is shown in the bottom panel of (c)**” was added to the caption of Fig. 7.

## Referee comment #2:

### General comments:

*In this paper, the authors investigated the new particle formation (NPF) events in a coastal agricultural site in Southwestern Finland by using a combination of a nitrate ion-based chemical-ionization mass spectrometer, and gas analyzers as well as aerosol samplers. The binned positive matrix factorization method (binPMF) was applied to the measured mass spectra, showing that eight factors could describe the time series of ambient gas and cluster composition during the NPF events. Before publication, I think there are several comments that the authors may need to consider.*

1. There are several uncertainties in this study that may lead to some problems or make this study not really convincing. First, the mass errors ranged from -10 ppm to 50 ppm (Line 211), so the identification of compounds with a high molecular weight maybe not be correct. How did the authors determine the confidence levels of the identifications in Table 1? Second, the authors said that “it cannot be certainly proved that a variable is actually forming new particles or growing them by examining the correlations. There is always a possibility that a variable is only observed simultaneously with NPF events due to the similarity of its source and the source of the precursor really causing the NPF events.” I agree with the authors about this point, but does it also mean the results of this study are also based on this uncertainty?

The mass errors before the binPMF run were 0.4 ppm, in median (see the referee #1 comment 4). However, because the  $m/z$  bin width in the binPMF needs to be a finite value, summing all signal within a  $m/z$  bin results in losing of the information on how the signal behaves within the bin. In this study, the bin width of 0.02 Th was used, as by Zhang et al. (2019). Choosing narrower bin width would result in better mass resolution but in lower signal-to-noise ratio, because the sum within a bin would be smaller. Therefore, we didn't want to alter the bin width for this study. Based on the mass errors of the water clusters, the mass errors after the binPMF run were between -10 ppm and 0 ppm when  $m/z$  is below 280 Th and increase from 0 ppm to +50 ppm when  $m/z$  increases from 280 Th to 400 Th. Fortunately, the main interest in the spectra in the binPMF factors is actually quite near 280 Th, for which the mass error is near 0 ppm.

Additionally, the mass errors approaching +50 ppm for the larger masses is eventually not as problematic as it may sound like. That is because the mass errors are known for the water clusters. These known errors can be used to estimate what the mass errors for any compound with a specific  $m/z$  value should be. For example, the peak at 201.0171 Th in F3 is  $\text{NO}_3^- \cdot \text{C}_6\text{H}_4\text{OHNO}_2$  (nitrophenol) with the mass error of -9.3 ppm, while the nearest water cluster peak, at 197.0278 Th in F2, is  $\text{NO}_3^- \cdot \text{HNO}_3(\text{H}_2\text{O})_4$  with the mass error of -7.8 ppm. Because these errors (-9.3 ppm vs. -7.8 ppm) are so close to each other, it can be identified with a high certainty that the peak at 201.0171 Th in F3 is really  $\text{NO}_3^- \cdot \text{C}_6\text{H}_4\text{OHNO}_2$ . Similarly, in a case of a compound with a higher molecular weight, for example the peak at 339.0585 Th in F7 having the error of +28 ppm for  $\text{C}_{10}\text{H}_{15}\text{O}_8\text{N}$ , can also be identified with a quite high certainty although the errors sound large (the mass errors of water cluster peaks near this  $m/z$  range are +25 ... +35 ppm).

This is now clarified in the text, first, with the added text “**aided by observed errors of  $m/z$  ratios of water clusters (see Sect. 4.2.3)**” to the sentence “Nevertheless, identifying possible chemical formulae from the observed peaks was still done with relatively high confidence for several peaks.” in Sect. 4.2.2 and, secondly, with the added text “**The confidence levels are based on subjective estimations on the success of peak fitting (aided by the errors of water clusters at specific  $m/z$  ratios and by the known isotopic patterns) and on the expectation of the compound to be detected with the used instrumentation.**” to the caption of Table 1. This addition answers also to the question about how the confidence levels of the identifications were defined.

As for the second concern in this referee comment, the uncertainty in the knowledge of the precursor really causing NPF events is always present in similar studies. Therefore, the observed correlations between any variable and NPF should also



be assessed by considering whether they are physically reasonable or not. For the compounds found in F7 leading to NPF events, there is currently no proof against it since it is not implausible that organic compounds with presumably low volatilities could form new particles and grow them. We have used wordings that do not strictly say that, e.g., “F7 causes NPF”. Instead, wordings like “It was observed that the factor 7 had elevated intensities during the NPF events.” in the abstract and “a suggested explanation for particle formation and growth observed in the studied area” in Sect. 4.5 have been used to emphasize the involved uncertainty.

The sentence “**This kind of uncertainty is, however, present in all similar studies as well; and therefore, the observed correlations should also be assessed by considering whether they are physically reasonable or not.**” is now added after the sentence “There is always a possibility that a variable is only observed simultaneously with NPF events due to the similarity of its source and the source of the precursor really causing the NPF events.” in Sect. 4.3. Additionally, there is a sentence in the conclusions section which is now updated to “In conclusion, NPF events observed at the studied coastal agricultural environment **seem to** follow this routing: ozone levels elevate which causes F8(morning-N-HOMs) intensity to elevate, which is then transformed to F7(day-N-HOMs) via radiation; if F7(day-N-HOMs) is the major form in the spectra, a NPF event is observed.”

In addition to this discussion on F7 and NPF, some too strict wordings used to describe the transformation of F8 to F7 are also now loosened (see the answers to the referee #1 comment 19 and the referee #2 comment 3).

*2. Where do the F8 compounds come from? I also think the authors need to give a map showing the sampling site and the meteorological information such as the wind speed and direction is also required to illustrate the sources of measured aerosols and gases.*

In this measurement report, we do not examine the actual origins of the detected compounds and binPMF factors. The idea here is to report what has been observed but not to deeply examine the root causes of the observations. Therefore, we decided not to examine where the compounds in the binPMF factors come from. The sentence “**The actual origins of the detected compounds and binPMF factors are not examined in this measurement report.**” is now added to the end of the introduction section to highlight this. Because of that, we also didn’t include the map and any more detailed meteorological information of the sampling site. However, we have now included a better description of the sampling site to Sect. 2.1 with the updated text “The measurement site is located in the middle of fields and has the shortest distance to the sea of 500 m **and to the nearest forest of 100 m. The nearest town, Parainen, with ~15 000 inhabitants, is located 5 km to the west and a larger city, Turku, with ~195 000 inhabitants, is located ~18 km toward the inland.**” to give an idea on what the sources of the compounds, in addition to the field itself, could be.

*3. The time profiles of F7 compounds did not correlate with F8 compounds (Figure 7), I do not understand why the F7 formed from the F8?*

By looking at the time series of F7, F8 and O<sub>3</sub> in Fig. 7, it can be observed that when NetRad increases, F8 decreases and almost the same amount is added to F7, while their sum behaves very similar to the trend of the O<sub>3</sub> concentration. Therefore, they don’t need to correlate for F7 to be formed from F8, but rather to anti-correlate (but only with NetRad).

However, because the actual (chemical) transformation of F8 to F7 (where F8 is the precursor for F7) cannot be proven, we have loosened some wordings related to it. We have now used wordings like “**seemingly** transforms” instead of “transforms” and “acts **like** a precursor” instead of “acts as a precursor” in all corresponding locations to highlight that this transformation is only what has been observed from the time series but the chemical mechanism hasn’t been examined. Additionally, the sentence “It is evident that F7(day-N-HOMs) is connected to particle formation and growth process and F8(morning-N-HOMs) acts as a precursor for F7(day-N-HOMs).” in Sect. 4.4 is now updated to “It is evident that F7(day-N-HOMs) is connected to particle formation and growth process and F8(morning-N-HOMs) acts **like** a precursor for F7(day-N-HOMs) **because the F8(morning-N-HOMs) level decreases with increasing F7(day-N-HOMs).**”

See also the referee #1 comment 19, which is related to this as well.

4. Did the authors detect halogenated organics due to the proximity of the measurement site to the sea?

Halogenated organics were not examined from this nitrate CI-API-TOF data because the instrument is typically not used to detect them (at least with nitrate ionization) and, according to our knowledge, there is currently no expectation that they would be connected to NPF events.

*Specific comments:*

5. Line 9: “Values of  $f_{F7}$  higher than 0.5 were typically observed during the NPF events”. However, Figure 7c showed this value is lower than 0.5 during the NPF events on May 8-11 and 17.

The critical value of  $f_{F7}$  for NPF events is not exactly 0.5. Therefore, we have used the “ $\sim 0.5$ ” notation. However, the uncertainty of the critical value is now better defined and all “ $\sim 0.5$ ” notations are now replaced with “ $0.50 \pm 0.05$ ”. Thus, the minimum of the critical value is about 0.45, which is exceeded during the NPF events on the days mentioned in the comment.

6. Line 75: What is the mass resolution of CIMS during the field observation?

The mass resolution is now mentioned with the added sentence “**The API-TOF mass spectrometer provided the mass resolving power of 3500–4000 Th/Th for the studied mass range with the used voltage settings.**”

7. Line 108: “Ions smaller than 169 Th were omitted because there are many organic compounds that are unlikely the key compounds in NPF”. However, methanesulfonic acid can also efficiently initiate NPF in the presence of small alkylamines and water (Chen et al., 2016; Dawson et al., 2012).

See the referee #1 comment 8. The citation to the article by Dawson et al. (2012) is now added to the text about methanesulfonic acid and NPF in the introduction section.

*References:*

Chen, H., Varner, M. E., Gerber, R. B. and Finlayson-Pitts, B. J.: Reactions of Methanesulfonic Acid with Amines and Ammonia as a Source of New Particles in Air, *J. Phys. Chem. B*, 120(8), 1526–1536, doi:10.1021/acs.jpcc.5b07433, 2016.

Dawson, M. L., Varner, M. E., Perraud, V., Ezell, M. J., Gerber, R. B. and Finlayson-Pitts, B. J.: Simplified mechanism for new particle formation from methanesulfonic acid, amines, and water via experiments and *ab initio* calculations, *Proc. Natl. Acad. Sci. U. S. A.*, 109(46), 18719–18724, doi:10.1073/pnas.1211878109, 2012.

## 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 I. Techniques of data interpretation and error analysis, *J. Geophys. Res.-Atmos.*, 108, 4090, <https://doi.org/https://doi.org/10.1029/2002JD002358>, 2003.
- Beck, L. J., Sarnela, N., Junninen, H., Hoppe, C. J., Garmash, O., Bianchi, F., Riva, M., Rose, C., Peräkylä, O., Wimmer, D., Kausiala, O., Jokinen, T., Ahonen, L., Mikkilä, J., Hakala, J., He, X. C., Kontkanen, J., Wolf, K. K., Cappelletti, D., Mazzola, M., Traversi, R., Petroselli, C., Viola, A. P., Vitale, V., Lange, R., Massling, A., Nøjgaard, J. K., Krejci, R., Karlsson, L., Zieger, P., Jang, S., Lee, K., Vakkari, V., Lampilahti, J., Thakur, R. C., Leino, K., Kangasluoma, J., Duplissy, E. M., Siivola, E., Marbouti, M., Tham, Y. J., Saiz-Lopez, A., Petäjä, T., Ehn, M., Worsnop, D. R., Skov, H., Kulmala, M., Kerminen, V. M., and Sipilä, M.: Differing Mechanisms of New Particle Formation at Two Arctic Sites, *Geophys. Res. Lett.*, 48, e2020GL091334, <https://doi.org/10.1029/2020GL091334>, 2021.
- Dawson, M. L., Varner, M. E., Perraud, V., Ezell, M. J., Gerber, R. B., and Finlayson-Pitts, B. J.: Simplified mechanism for new particle formation from methanesulfonic acid, amines, and water via experiments and ab initio calculations, *P. Natl. Acad. Sci. USA*, 109, 18719–18724, <https://doi.org/10.1073/pnas.1211878109>, 2012.
- Gagné, S., Lehtipalo, K., Manninen, H. E., Nieminen, T., Schobesberger, S., Franchin, A., Yli-Juuti, T., Boulon, J., Sonntag, A., Mirme, S., Mirme, A., Hörrak, U., Petäjä, T., Asmi, E., and Kulmala, M.: Intercomparison of air ion spectrometers: an evaluation of results in varying conditions, *Atmos. Meas. Tech.*, 4, 805–822, <https://doi.org/10.5194/amt-4-805-2011>, 2011.
- Gormley, P. G. and Kennedy, M.: Diffusion from a Stream Flowing through a Cylindrical Tube, *P. Roy. Irish Acad. A*, 52, 163–169, <https://doi.org/10.2307/20488498>, 1948.
- Hytinen, N., Kupiainen-Määttä, O., Rissanen, M. P., Muuronen, M., Ehn, M., and Kurtén, T.: Modeling the Charging of Highly Oxidized Cyclohexene Ozonolysis Products Using Nitrate-Based Chemical Ionization, *J. Phys. Chem. A*, 119, 6339–6345, <https://doi.org/10.1021/acs.jpca.5b01818>, 2015.
- Kangasluoma, J., Cai, R., Jiang, J., Deng, C., Stolzenburg, D., Ahonen, L. R., Chan, T., Fu, Y., Kim, C., Laurila, T. M., Zhou, Y., Dada, L., Sulo, J., Flagan, R. C., Kulmala, M., Petäjä, T., and Lehtipalo, K.: Overview of measurements and current instrumentation for 1–10 nm aerosol particle number size distributions, *J. Aerosol Sci.*, 148, 105584, <https://doi.org/10.1016/j.jaerosci.2020.105584>, 2020.
- Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P., Junninen, H., Paasonen, P., Riipinen, I., Lehtinen, K. E. J., Laaksonen, A., and Kerminen, V.-M.: Measurement of the nucleation of atmospheric aerosol particles, *Nature Protocols*, 7, 1651 – 1667, 2012.
- Kürten, A., Rondo, L., Ehrhart, S., and Curtius, J.: Calibration of a Chemical Ionization Mass Spectrometer for the Measurement of Gaseous Sulfuric Acid, *J. Phys. Chem. A*, 116, 6375–6386, <https://doi.org/10.1021/jp212123n>, 2012.
- Mirme, S. and Mirme, A.: The mathematical principles and design of the NAIS – a spectrometer for the measurement of cluster ion and nanometer aerosol size distributions, *Atmos. Meas. Tech.*, 6, 1061–1071, <https://doi.org/10.5194/amt-6-1061-2013>, 2013.
- Paatero, P., Hopke, P. K., Song, X.-H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, *Chemometr. Intell. Lab.*, 60, 253–264, [https://doi.org/10.1016/S0169-7439\(01\)00200-3](https://doi.org/10.1016/S0169-7439(01)00200-3), 2002.
- Simon, M., Dada, L., Heinritzi, M., Scholz, W., Stolzenburg, D., Fischer, L., Wagner, A. C., Kürten, A., Rörup, B., He, X.-C., Almeida, J., Baalbaki, R., Baccarini, A., Bauer, P. S., Beck, L., Bergen, A., Bianchi, F., Bräkling, S., Brilke, S., Caudillo, L., Chen, D., Chu, B., Dias, A., Draper, D. C., Duplissy, J., El-Haddad, I., Finkenzeller, H., Frege, C., Gonzalez-Carracedo, L., Gordon, H., Granzin, M., Hakala, J., Hofbauer, V., Hoyle, C. R., Kim, C., Kong, W., Lamkaddam, H., Lee, C. P., Lehtipalo, K., Leiminger, M., Mai, H., Manninen, H. E., Marie, G., Marten, R., Mentler, B., Molteni, U., Nichman, L., Nie, W., Ojdanic, A., Onnela, A., Partoll, E., Petäjä, T., Pfeifer, J., Philippov, M., Quéléver, L. L. J., Ranjithkumar, A., Rissanen, M. P., Schallhart, S., Schobesberger, S., Schuchmann, S., Shen, J., Sipilä, M., Steiner, G., Stozhkov, Y., Tauber, C., Tham, Y. J., Tomé, A. R., Vazquez-Pufleau, M., Vogel, A. L., Wagner, R., Wang, M., Wang, D. S., Wang, Y., Weber, S. K., Wu, Y., Xiao, M., Yan, C., Ye, P., Ye, Q., Zauner-Wieczorek, M., Zhou, X., Baltensperger, U., Dommen, J., Flagan, R. C., Hansel, A., Kulmala, M., Volkamer, R., Winkler, P. M., Worsnop, D. R., Donahue, N. M., Kirkby, J., and Curtius, J.: Molecular understanding of new-particle formation from  $\alpha$ -pinene between  $-50$  and  $+25$  °C, *Atmos. Chem. Phys.*, 20, 9183–9207, <https://doi.org/10.5194/acp-20-9183-2020>, 2020.
- Wang, M., He, X. C., Finkenzeller, H., Iyer, S., Chen, D., Shen, J., Simon, M., Hofbauer, V., Kirkby, J., Curtius, J., Maier, N., Kurtén, T., Worsnop, D. R., Kulmala, M., Rissanen, M., Volkamer, R., Tham, Y. J., Donahue, N. M., and Sipilä, M.: Measurement of iodine species and sulfuric acid using bromide chemical ionization mass spectrometers, *Atmos. Meas. Tech.*, 14, 4187–4202, <https://doi.org/10.5194/amt-14-4187-2021>, 2021.
- Yan, C., Nie, W., Äijälä, M., Rissanen, M. P., Canagaratna, M. R., Massoli, P., Junninen, H., Jokinen, T., Sarnela, N., Häme, S. A. K., Schobesberger, S., Canonaco, F., Yao, L., Prévôt, A. S. H., Petäjä, T., Kulmala, M., Sipilä, M., Worsnop, D. R., and Ehn, M.: Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization, *Atmos. Chem. Phys.*, 16, 12715–12731, <https://doi.org/10.5194/acp-16-12715-2016>, 2016.

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, *Atmos. Meas. Tech.*, 12, 3761–3776, <https://doi.org/10.5194/amt-12-3761-2019>, 2019.

# Measurement report: Atmospheric new particle formation in a coastal agricultural site explained with binPMF analysis of nitrate CI-APi-TOF spectra

Miska Olin<sup>1</sup>, Magdalena Okuljar<sup>2</sup>, Matti P. Rissanen<sup>1</sup>, Joni Kalliokoski<sup>1</sup>, Jiali Shen<sup>2</sup>, Lubna Dada<sup>2,3</sup>, Markus Lampimäki<sup>2</sup>, Yusheng Wu<sup>2</sup>, Annalea Lohila<sup>2,4</sup>, Jonathan Duplissy<sup>2,5</sup>, Mikko Sipilä<sup>2</sup>, Tuukka Petäjä<sup>2</sup>, Markku Kulmala<sup>2,6,7</sup>, and Miikka Dal Maso<sup>1</sup>

<sup>1</sup>Aerosol Physics Laboratory, Faculty of Engineering and Natural Sciences, Tampere University, Tampere, Finland

<sup>2</sup>Institute for Atmospheric and Earth System Research (INAR)/Physics, Faculty of Science, University of Helsinki, Helsinki, Finland

<sup>3</sup>Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, Villigen, Switzerland

<sup>4</sup>Climate System Research, Finnish Meteorological Institute, Helsinki, Finland

<sup>5</sup>Helsinki Institute of Physics (HIP)/Physics, Faculty of Science, University of Helsinki, Helsinki, Finland

<sup>6</sup>Aerosol and Haze Laboratory, Beijing Advanced Innovation Center for Soft Matter Sciences and Engineering, Beijing University of Chemical Technology (BUCT), Beijing, China

<sup>7</sup>Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing University, Nanjing, China

**Correspondence:** Miska Olin (miska.olin@tuni.fi)

**Abstract.** The occurrence of new particle formation (NPF) events detected in a coastal agricultural site, at Qvidja, in South-western Finland, was investigated using the data measured with a nitrate ion-based chemical-ionization mass spectrometer (CI-APi-TOF). The binned positive matrix factorization method (binPMF) was applied to the measured spectra. It resulted in eight factors describing the time series of ambient gas and cluster composition at Qvidja during spring 2019. The most interesting factors related to the observed NPF events were the two factors with the highest mass-to-charge ratios, numbered 7 and 8, both having profiles with patterns of highly oxygenated organic molecules with one nitrogen atom. It was observed that the factor 7 had elevated intensities during the NPF events. A variable with an even better connection to the observed NPF events is  $f_{F7}$ , which denotes the fraction of the total spectra within the studied mass-to-charge ratio range between 169 and 450 Th being in a form of the factor 7. Values of  $f_{F7}$  higher than ~~0.5 were typically~~  $0.50 \pm 0.05$  were observed during the NPF events, of which durations also correlated with the duration of  $f_{F7}$  exceeding ~~the critical value of 0.5~~ this critical value. It was also observed that the factor 8 acts ~~as a precursor of~~ like a precursor for the factor 7 with ~~high solar irradiance~~ solar radiation and that the formation of the factor 8 is associated with ozone levels.

## 1 Introduction

Atmospheric new particle formation (NPF) and primary emissions are both important sources of the total particle number concentration and cloud condensation nuclei in the global troposphere and in the continental boundary layer (Merikanto et al., 2009; Fountoukis et al., 2012; Posner and Pandis, 2015; Dunne et al., 2016; Kulmala et al., 2016; Gordon et al., 2017; Kerminen

et al., 2018; Olin et al., 2022). NPF occurring in different environments has a very diverse behavior and is also not well quantified (Kerminen et al., 2018).

Atmospheric particle measurements and related analyses of NPF events, i.e., events of the formation of new small particles and their subsequent growth, have been performed in a wide range of environments. NPF events seem to be occurring almost everywhere in the world at all seasons, but their occurrence varies a lot between the sites and seasons (Kerminen et al., 2018). Probably the most intensively studied environments in terms of NPF are boreal forests and urban areas. In this study, we present the gas, particle, and cluster data measured in a coastal agricultural site and search for parameters favoring or disfavoring NPF events. NPF is often occurring as regional NPF events over large spatial areas of even hundreds of kilometers. On the other hand, it can be more local (Junninen et al., 2022).

Several parameters influencing the occurrence of NPF events have been proposed in the literature, but a consistent and universal parameterization is still lacking since the optimal parameters predicting NPF events in different locations vary significantly (Kerminen et al., 2018). Variables for predicting NPF most frequently reported are solar radiation (Birmili and Wiedensohler, 2000; Birmili et al., 2003; Guo et al., 2012; Jun et al., 2014; Pierce et al., 2014; Qi et al., 2015), pre-existing particle loadings (Birmili et al., 2003; Dal Maso et al., 2007; Pikridas et al., 2012; Salma et al., 2016; Dada et al., 2017), relative humidity (RH) (Birmili et al., 2003; Wu et al., 2007; Guo et al., 2012; Jun et al., 2014), temperature ( $T$ ) (Paasonen et al., 2013; Dunne et al., 2016; Dada et al., 2017), and the concentration of sulfuric acid ( $\text{H}_2\text{SO}_4$ ) (Birmili et al., 2003; Kulmala et al., 2006; Wang et al., 2011; Qi et al., 2015; Yao et al., 2018) or its main precursor, sulfur dioxide ( $\text{SO}_2$ ) (Birmili and Wiedensohler, 2000; Woo et al., 2001; Guo et al., 2012). Of those, high solar irradiances and high  $\text{H}_2\text{SO}_4$  concentrations ( $[\text{H}_2\text{SO}_4]$ ) typically favor the occurrence of NPF events. Instead, high pre-existing particle loadings, such as condensation sink (CS) consuming possibly nucleating precursors, and high RH are typically disfavoring. Instead,  $T$  and  $\text{SO}_2$  concentration ( $[\text{SO}_2]$ ) are inconclusive, as their roles in NPF events are ambiguous between different data sets. Also organic compounds with low enough volatilities or highly oxygenated organic molecules (HOMs), have recently been connected to atmospheric NPF (Dada et al., 2017; Bianchi et al., 2019). Specific for coastal environments are the roles of iodine compounds (O'Dowd et al., 2002), such as iodic acid ( $\text{HIO}_3$ ) (Sipilä et al., 2016; Baccarini et al., 2020; He et al., 2021), and methanesulfonic acid ( $\text{CH}_3\text{SO}_3\text{H}$ ) (~~Beck et al., 2021~~) ([Dawson et al., 2012](#); [Beck et al., 2021](#)) in NPF. Ammonia ( $\text{NH}_3$ ), as a base, is known to enhance  $\text{H}_2\text{SO}_4$ -driven nucleation through stabilizing critical clusters (Kulmala et al., 2014), which can be important in agricultural environments since it is a substantial component in fertilizers. Studies on NPF from agricultural land emissions have usually been related to  $\text{NH}_3$ , but also, e.g., skatole is suggested (Ciuraru et al., 2021), which is an organic compound in sewage sludge used in fertilizers.

Additionally, general criteria for the occurrence of NPF events have been searched for, but still, no suitable estimator covering a variety of different environments has been found. Another difficulty with these criteria is that they usually require specific information on multiple variables, such as particle size distributions or growth rates, or precursor concentrations. In this study, we propose another parameter (named here  $f_{\text{F7}}$ ) for predicting NPF events, which can be used in examining the occurrence of NPF events in coastal or agricultural areas or even in searching for the general criteria. Surprisingly,  $f_{\text{F7}}$  does not include any particle variables but is solely based on mass spectrometric data of airborne molecules analyzed using the binned positive matrix factorization method (binPMF) instead. More specifically,  $f_{\text{F7}}$  denotes how much of the observed mass spectra is

in a specific form of the mixture of organic compounds. Additionally, the suggested explanation or routing of the observed NPF events gives insights into particle formation mechanisms occurring in the studied environment. The actual origins of the detected compounds and binPMF factors are not examined in this measurement report.

## 2 Measurement methods

### 2.1 Measurement site and time range

Measurements were performed at a pilot agricultural farm for regenerative farming, at Qvidja, located in a coastal environment in Turku Archipelago, in Southwestern Finland. The measurement site is located in the middle of fields and has the shortest distance to the sea of 500 m and to the nearest forest of 100 m. The nearest town, Parainen, with ~15 000 inhabitants, is located 5 km to the west and a larger city, Turku, with ~195 000 inhabitants, is located ~18 km toward the inland. The fields have clayey soil and they consist of several grass and clover species. Since 2017, sustainable and environmentally friendly field management practices have been conducted at the farm. More detailed information on the location, species, and management practices of Qvidja can be found in Heimsch et al. (2021).

The data used in this study are a part of long-term measurements provided by the instruments installed in a container in the middle of the fields. Additionally, a laboratory van, ATMo-Lab, was parked next to the container for the time range between 2 Apr 2019 and 26 Jun 2019. As the key instrument of this study, a mass spectrometer, was located in the van, only the data from this time range from the longer time series are utilized in this study.

### 2.2 Measurement instruments

The measurement instruments in the container include several gas analyzers, devices measuring environmental parameters, and aerosol samplers. Gas analyzers utilized in this study measure the concentrations of SO<sub>2</sub>, ozone (O<sub>3</sub>), carbon monoxide (CO), nitric oxide (NO), nitrogen oxides (NO<sub>x</sub>), and NH<sub>3</sub>. Utilized environmental parameters are temperature (*T*), relative humidity (RH), and net radiation (NetRad). The used aerosol samplers were two Particle Size Magnifiers (PSM,A and PSM,B), a Neutral Cluster and Air Ion Spectrometer (NAIS), and a Differential Mobility Particle Sizer (DMPS). PSM,A and PSM,B were used to determine the particle size distributions in the size ranges of 1.15–2.8 nm and 1.3–2.8 nm, respectively. NAIS was used here to detect particles in the size range of 2.7–6.5 nm using its negative-polarity charger. DMPS was used to determine the particle size distribution in the size range of 6–823 nm and CS, which was calculated with the method by Kulmala et al. (2012) using a common approximation of assuming H<sub>2</sub>SO<sub>4</sub> as the condensing vapor.

A nitrate-ion-based (NO<sub>3</sub><sup>-</sup>-based) chemical-ionization atmospheric-pressure-interface time-of-flight (CI-APi-TOF) mass spectrometer (Aerodyne Research Inc.; USA and ToFwerk AG, Switzerland; Jokinen et al. 2012) was installed in the laboratory van. It consists of a chemical-ionization (CI) inlet (Eisele and Tanner, 1993) and an APi-TOF mass spectrometer (Junninen et al., 2010). The-NO<sub>3</sub><sup>-</sup> was produced by adding nitric acid (HNO<sub>3</sub>) vapor to the sheath air which was then exposed to X-ray radiation inside the CI-inlet. The HNO<sub>3</sub> vapor was produced by directing 10 sccm flow of the sheath air through a glass vial



containing liquid  $\text{HNO}_3$ . The sheath air was a 20 slpm flow of outdoor air filtered using an activated carbon filter and a HEPA filter. The outdoor air sample was drawn to the CI-inlet with the flow rate of 8.3 slpm through a side hatch of the van with a 16 mm pipe of 60 cm length, which also acts as a laminarization tube for the CI-inlet. The API-TOF mass spectrometer provided the mass resolving power of 3500–4000 Th/Th for the studied mass range with the used voltage settings.

### 3 CI-API-TOF data processing

The data were recorded in 4 s time resolution but were first averaged to 600 s resolution ( $t_a$ ) in order to increase signal-to-noise ratios (SNRs) but still maintain an adequate time resolution for the campaign. The averaging was performed with a TOF-mass spectrometer data processing code for Matlab, tofTools, resulting in 9970 valid time bins. The averaged TOF-spectra were then calibrated with the tofTools code against known mass-to-charge ratios ( $m/z$ ) of always existing ions with  $\text{NO}_3^-$ -based CI-inlet,  $\text{NO}_3^-$ ,  $\text{NO}_3^- \cdot \text{HNO}_3$ , and  $\text{NO}_3^- \cdot (\text{HNO}_3)_2$  or  $\text{NO}_3^- \cdot \text{nitrophenol}$  ( $\text{C}_6\text{H}_4\text{OHNO}_2$ ).

#### 3.1 Potential new particle-forming acids

Compounds previously most associated with NPF observed in a coastal environment measured using the CI-API-TOF are the acids  $\text{H}_2\text{SO}_4$ ,  $\text{HIO}_3$ , and  $\text{CH}_4\text{SO}_3$ .

$[\text{H}_2\text{SO}_4]$  is calculated as by Olin et al. (2020) with the equation

$$[\text{H}_2\text{SO}_4] = \frac{C}{P} \cdot \frac{\{\text{HSO}_4^-\} + \{\text{HSO}_4^- \cdot \text{HNO}_3\} + \{\text{HSO}_4^- \cdot \text{H}_2\text{SO}_4\}}{\{\text{NO}_3^-\} + \{\text{NO}_3^- \cdot \text{HNO}_3\} + \{\text{NO}_3^- \cdot (\text{HNO}_3)_2\}} \quad (1)$$

where  $C$  is the calibration coefficient for  $\text{H}_2\text{SO}_4$  ( $= 2.45 \times 10^9 \text{ cm}^{-3} = (2.45 \pm 0.33) \times 10^9 \text{ cm}^{-3}$ ) determined with known  $\text{H}_2\text{SO}_4$  concentrations (details of the method are well described in Kürten et al. (2012)) at the measurement site (once a month),  $P$  is the penetration efficiency of  $\text{H}_2\text{SO}_4$  in the sampling pipe ( $= 0.58$ ) 60 cm long sampling pipe having the flow rate of 8.3 slpm ( $= 0.58 \pm 0.03$ ) determined with the function by Gormley and Kennedy (1948), and the curly brackets denote the areas of the peaks at the corresponding mass-to-charge ratios in the high-resolution mass spectra obtained from the tofTools code.  $[\text{H}_2\text{SO}_4]$  has a relative uncertainty of 15 %, originating mostly from the uncertainty in  $C$ , determined with its variation between three calibration runs during the campaign. However, because a single calibration run also involves its own systematic uncertainty (Kürten et al., 2012)—which is not easily quantified—the uncertainty in  $[\text{H}_2\text{SO}_4]$  can be higher (a frequently used rough approximation is  $-50\%/+100\%$ ).

Although generally untypical, there was another peak overlapping with the bisulfate ion,  $\text{HSO}_4^-$ , peak (96.96 Th) at 96.97 Th in these data. The overlapping peak corresponds to  $\text{H}_2\text{CO}_3 \cdot \text{Cl}^-$  (hydrochloric acid, HCl, and carbonic acid,  $\text{H}_2\text{CO}_3$ , were also detected separately), and it covered 10–90 % of the area of the total peak at 97 Th. Therefore, high-resolution fitting was a necessity in determining  $[\text{H}_2\text{SO}_4]$  in this case, whereas fitting with the unit mass resolution (UMR) is sufficient in many cases. The detection of  $\text{H}_2\text{CO}_3$  with  $\text{NO}_3^-$  ionization is untypical and it usually implies insufficient  $\text{HNO}_3$  in the ionizer. This potentially leads to detecting less bound product- $\text{NO}_3^-$  adducts, i.e., decreasing the selectivity (Hyttinen et al., 2015), yet it is

not an easy task to estimate how much. Nothing such was observed but it should be kept in mind that the data from the last 3 weeks (during which the  $\text{H}_2\text{CO}_3$  signal was at its highest) has to be interpreted with caution.

The concentrations of  $\text{HIO}_3$  ( $[\text{HIO}_3]$ ) and  $\text{CH}_4\text{SO}_3$  ( $[\text{CH}_4\text{SO}_3]$ ) were calculated as for  $\text{H}_2\text{SO}_4$ , with the exception of the peaks  $\{\text{IO}_3^-\} + \{\text{IO}_3^- \cdot \text{HNO}_3\}$  and  $\{\text{CH}_3\text{SO}_3^-\} + \{\text{CH}_3\text{SO}_3^- \cdot \text{HNO}_3\}$ , respectively, as a numerator in Eq. (1). Due to the lack of calibration methods for compounds other than  $\text{H}_2\text{SO}_4$ , we used a common approximation in which the values of  $C$  and  $P$  determined for  $\text{H}_2\text{SO}_4$  were used for these compounds too. That is because these species ( $\text{HIO}_3$  and  $\text{CH}_4\text{SO}_3$ ) have collision-limited charging efficiency when reacting with  $\text{NO}_3^-$  ions (Simon et al., 2020; Beck et al., 2021; Wang et al., 2021). Uncertainties in  $[\text{HIO}_3]$  and in  $[\text{CH}_4\text{SO}_3]$  are not easily quantified but the relative uncertainty of 15 % in  $[\text{H}_2\text{SO}_4]$  (or  $-50\%/+100\%$ ) can be used in estimating relative uncertainties during the campaign. However, the absolute levels of  $[\text{HIO}_3]$  and  $[\text{CH}_4\text{SO}_3]$  can still differ more due to the approximation of using  $C$  and  $P$  determined for  $\text{H}_2\text{SO}_4$ , but the difference between the calibration coefficients for different compounds is expected to be nearly constant during the campaign.

### 3.2 binPMF analysis

The mass spectra between 169 and 450 Th obtained from the tofTools code were analyzed using the binPMF analysis method described by Zhang et al. (2019). Ions smaller than 169 Th were omitted because there are many organic compounds that are unlikely the key compounds in NPF and have relatively high signals possibly causing issues in the binPMF analysis. Compounds with lower masses typically have higher volatilities and are thus more unlikely participating in nucleation and condensation. However, it is not generally true since, e.g.,  $\text{H}_2\text{SO}_4$  and  $\text{CH}_4\text{SO}_3$  are below 169 Th but still known to contribute to NPF. Therefore, they were analyzed separately of the binPMF analysis. One of the highest peaks in the spectra, malonic acid- $\text{NO}_3^-$ -cluster (166 Th), with its isotopes, were the largest of the omitted ions. Ions larger than 450 Th were also omitted due to their reduced transmission efficiency inside the APi-TOF device with the used voltage settings. Additionally, ions between 188 and 190 Th were omitted because the nitrate trimer—one of the reagent ions—with its isotopes falls in that range.

As by Zhang et al. (2019), the spectra were binned to bins with a width of 0.02 Th between  $N-0.2$  and  $N+0.3$  Th where  $N$  is an integer mass, resulting in 25 bins per integer, i.e., in 6975 bins in total in this case. The basis of PMF, in general, is to express the measured data matrix ( $X$ ) with a time series matrix of factor intensities ( $TS$ ), a factor mass spectra matrix ( $MS$ ), and a residual matrix ( $R$ ). The residuals are tried to be minimized so that the factor construction would describe the measured spectra most realistically. The number of factors,  $p$ , is a free parameter and its optimal value can be estimated by minimizing the residuals. In this case of binPMF, this is mathematically expressed with the equation

$$X_{9970 \times 6975} = TS_{9970 \times p} \times MS_{p \times 6975} + R_{9970 \times 6975}. \quad (2)$$

The data matrix  $X$  was normalized before running the binPMF code with the reagent ion signals as is done with the denominator in Eq. (1) for calculating  $[\text{H}_2\text{SO}_4]$ . This way the data matrix becomes dimensionless and expressed relative to the  $\text{NO}_3^-$  concentration.

The binPMF code tries to find the matrices TS and MS producing the lowest possible value for the sum of the scaled residuals,

$$Q = \sum_{i=1}^{9970} \sum_{j=1}^{6975} (R_{ij}/S_{ij})^2, \quad (3)$$

where S is the uncertainty matrix. S was estimated from the ambient data via the method suggested by Yan et al. (2016), i.e., through approximating instrument noise as the difference between the measured signal and its moving median over 5 data points. The differences are determined separately for all 9970 time bins and for all 6975  $m/z$  bins. Assuming that the counting statistics follow the Poisson distribution, the function

$$\begin{aligned} S_{ij} &= \sigma_{ij} + \sigma_{\text{noise}} = a \sqrt{\frac{\max(I, \frac{1}{t_a})}{t_a}} + \sigma_{\text{noise}} \\ &= (1.35 \pm 0.22) \sqrt{\frac{\max(I, \frac{1}{600\text{s}})}{600\text{s}}} + (0.001 \pm 0.003) \text{ cps} \end{aligned}$$

$$S_{ij} = \sigma_{ij} + \sigma_{\text{noise}} = a \cdot \sqrt{\frac{\max(I, \frac{1}{t_a})}{t_a}} + \sigma_{\text{noise}} \quad (4)$$

where  $a$  is the correcting factor incorporating any unaccounted contributions to the uncertainty (Allan et al., 2003) and  $I$  is the signal intensity in counts per second (cps), ~~was obtained. The correcting factor can be described to express the instrument noise. The values for  $a$  incorporates any unaccounted contributions to the uncertainty (Allan et al., 2003), and  $\sigma_{\text{noise}}$  were first determined for all  $m/z$  bins separately by fitting the differences between the measured signal and the moving median for different levels of  $I$  using all time bins and Eq. (4). The obtained  $a$  and  $\sigma_{\text{noise}}$  values were then averaged over all  $m/z$  bins using the successes of the fittings as weighting factors. Finally, the values of  $a = 1.35 \pm 0.22$  and  $\sigma_{\text{noise}} = (0.001 \pm 0.003) \text{ cps}$  were obtained, and the function for calculating the uncertainty matrix became~~

$$S_{ij} = 1.35 \cdot \sqrt{\frac{\max(I, \frac{1}{600\text{s}})}{600\text{s}}} + 0.001 \text{ cps}. \quad (5)$$

The value of  $a$  obtained here,  $1.35 \pm 0.22$ , is on a similar level to the values in the study by Yan et al. (2016),  $1.28 \pm 0.09$  or  $1.1 \pm 0.3$ . Similar to the matrix X, the matrix S was also normalized with the reagent ion signals before running the binPMF code to have them expressed with the same units (dimensionless).

The binPMF code used in this study (customized Matlab's nnmf function) includes similar down-weighting schemes for signals with low ~~signal-to-noise-ratios SNRs~~ and for outliers with  $|R_{ij}/S_{ij}| > 4$  as the code used by Zhang et al. (2019). A theoretically expected  $Q$  value ( $Q_{\text{exp}}$ ) can be calculated as the number of non-down-weighted elements in X subtracted by the number of elements in TS and MS in total. The code was run with increasing  $p$  until the ratio  $Q/Q_{\text{exp}}$  ~~did not substantially decrease anymore~~ set to a level with no further decrease. Because a PMF analysis depends on the initial guesses of TS and MS (seeds), the code was run with 50 random seeds per every  $p$  to find the optimal  $p$  and finally with 300 random seeds

with the optimal  $p$ , of which the seed providing the lowest  $Q$  value was selected for further analysis. The lowest  $Q/Q_{\text{exp}}$  is not necessarily the best solution of PMF because the solution of PMF is not unique due to the nature of PMF. For example, due to its rotational ambiguity, there are basically infinite number of solutions providing equal  $Q$  values (Paatero et al., 2002). Rotational ambiguity is, however, not considered in this study because it is typically not a problem with ambient measurement data, which usually contain enough data points with values near zero (Zhang et al., 2019).

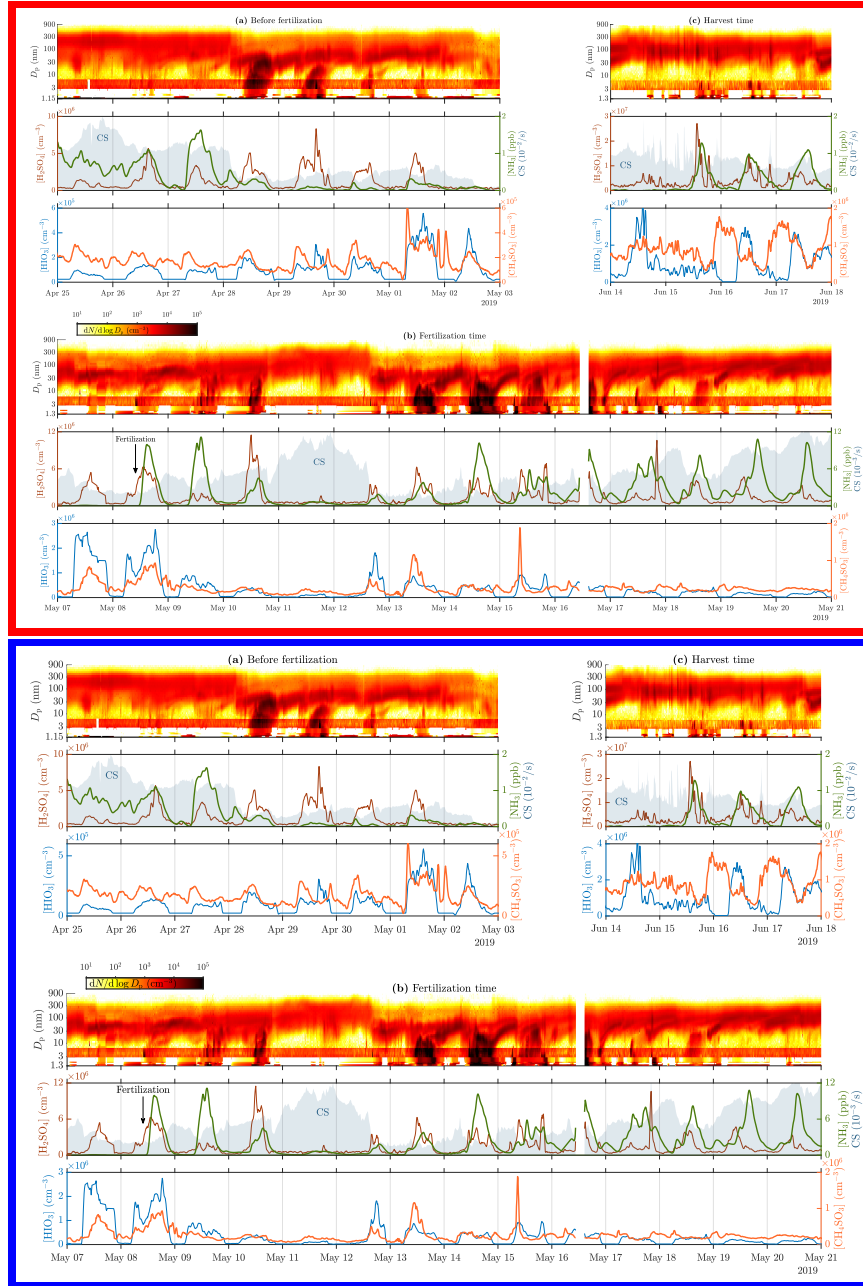
Because the factor intensity matrix TS is a dimensionless variable, it is further converted to a practical variable using the calibration coefficient and penetration efficiency of  $\text{H}_2\text{SO}_4$  also for the factor intensities. Hence, the factor intensities denote hereafter the total concentrations of all compounds within a factor. However, it should be noted that a total concentration of a factor would be true only if all the compounds would have responses between concentrations and the normalized signals and the penetration efficiencies similar to  $\text{H}_2\text{SO}_4$ . Penetration efficiencies are not expected to differ notably between  $\text{H}_2\text{SO}_4$  and the compounds with  $m/z$  of 169–450 Th, detected by nitrate CI-API-TOF, but calibration coefficients are.  $C$  for  $\text{H}_2\text{SO}_4$  denotes a value near the highest possible sensitivity, since the charging efficiency of  $\text{H}_2\text{SO}_4$  is collision-limited (Simon et al., 2020). However,  $C$  can be even orders of magnitude higher for compounds with weaker binding energies with  $\text{NO}_3^-$  (Hytinen et al., 2015). Therefore, the sensitivity of the used system can be much lower for some compounds. This is true especially for organic compounds with low oxygen-to-carbon (O:C) ratios, since organic compounds with adequately large number of carbon atoms having  $\text{O:C} > 0.6$  are estimated to have similar collision-limited charging efficiencies to  $\text{H}_2\text{SO}_4$  (Simon et al., 2020), but the binPMF factors of this study includes compounds with lower O:C ratios too. As it is practically impossible to determine calibration coefficients for all detectable compounds, this commonly used approximation is used here only to convert the dimensionless variables in TS to a practical concentration-based unit.

## 4 Results and discussion

From all 84 measurements days, 25 days show clear features of NPF events, during which a new particle mode appears by growing from the very small particle sizes near 1 nm toward a larger-sized background mode and finally merges with it. Instead, 37 days have hardly any NPF event features and 19 days have unclear features and cannot thus be classified as event or non-event days. The remaining 3 days remain unsure due to gaps in the measurement data. These classifications are listed in Table S1.

### 4.1 Time series of measured variables

Figure 1 presents time series of particle size distribution contours with variables most likely promoting NPF for three example time ranges. The time range before the fertilization (Fig. 1a) represents a great example of a quite constant background particle mode around 100 nm during the first three days followed by a clear and strong (the strength of a NPF event is defined as the duration of continuing NPF during a NPF event) NPF event (Apr 28th). The subsequent three days also have NPF event features but they become weaker every day and the last one (May 1st) has just hints of a very short NPF event. The reason for the inexistent NPF during the first three days is probably the suppression due to a high CS level (0.01 – 0.021/s). The reason



**Figure 1.** Time series of particle size distribution contours (top panels),  $[H_2SO_4]$ ,  $[NH_3]$ , CS (shaded areas in the middle panels),  $[HIO_3]$ , and  $[CH_4SO_3]$  (bottom panels) for different example time ranges, **(a)** before the fertilization, **(b)** during and right after the fertilization, and **(c)** during the harvest. Particle data below 3 nm are from a PSM, over 6 nm are from the DMPS, and the remaining part between them is from the NAIS. [The data from the NAIS show red banners due to its general tendency to overestimate concentrations at these particle sizes \(Gagné et al., 2011; Mirme and Mirme, 2013; Kangasluoma et al., 2020\).](#) Note that the PSM data shown in **(a)** are from the PSM,A whereas the data shown in **(b)** and **(c)** are from the PSM,B due to different coverage of the devices. Also, note different scales of the y-axes between the subfigures.

for the strong NPF event on Apr 28th and the subsequent weakening events cannot be clearly explained with any of the measured variables. Nevertheless, the most promising particle-forming compound measured is  $\text{H}_2\text{SO}_4$  because its concentration is elevated approximately at the same time as the events. However, it does not show a decreasing trend during the weakening trend of the NPF events and CS also remains on a lower level ( $\sim 0.005 \text{ 1/s}$ ).  $[\text{HIO}_3]$  is also elevated during the daytime but its levels do not explain these NPF events because it is its highest on the last two days, having no clear NPF events.  $[\text{NH}_3]$  and  $[\text{CH}_4\text{SO}_3]$  seem to not clearly explain any changes in particle size distribution data.

The time range during the fertilization (May 8th) and right after it (Fig. 1b) includes one of the strongest NPF events (May 14th) during this measurement. Noteworthy is that the fertilization increases daytime  $[\text{NH}_3]$  remarkably (100-fold compared to the time right before the fertilization and 10-fold compared to the average of the preceding days). However, the occurrence of NPF events does not increase after the fertilization even though  $[\text{NH}_3]$  remains high, implying that, on one hand,  $\text{NH}_3$  is not a key compound or is not a limiting factor in NPF events during these measurements. On the other hand,  $[\text{NH}_3]$  seems to be higher during the days showing NPF features, with the exception of the last days, during which CS begins to approach and exceed the level of  $0.01 \text{ 1/s}$  showing high  $[\text{NH}_3]$  but no NPF.  $[\text{H}_2\text{SO}_4]$  seems promising here too but still no clear criterion for the occurrence of an NPF event at this site cannot be constructed from any of the measured variables.

The time range during the harvest (Fig. 1c) shows another type of changes in particle size distribution data. There are short spikes of increased particle concentrations within the size range of 20–200 nm and thus in CS too, occurring several times a day. They originate probably from direct emission sources, e.g., from tractors harvesting the fields, rather than via NPF because ~~the abundance of the smallest particles is simultaneously not elevated~~there is no simultaneous concentration increase in particles smaller than those. It can be observed that also the  $[\text{H}_2\text{SO}_4]$  time series contains spikes at the coincident time moments, highlighting the possibility of tractors as the source of 20–200 nm particles.

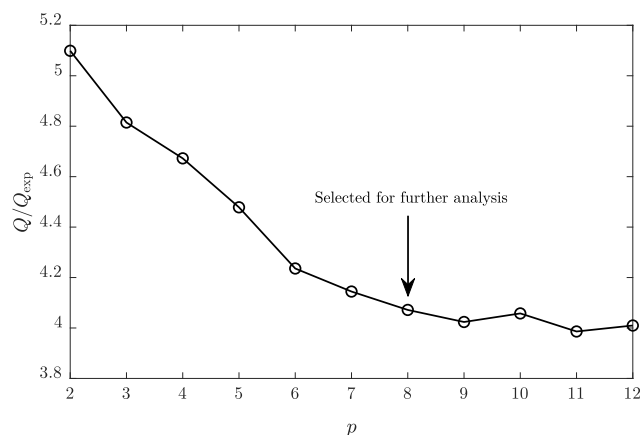
## 4.2 binPMF results

Previously presented analysis on the measured variables did not provide any clear formula for the occurrence of the NPF events in this measurement site. Next, we concentrate on the results from the binPMF analysis and how they can be related to the occurrence of the events.

### 4.2.1 Determining the optimal number of factors

Figure 2 presents the lowest obtained ratios  $Q/Q_{\text{exp}}$  with different random seeds and with different numbers of factors ( $p$ ). It can be seen that  $Q/Q_{\text{exp}}$  decreases with increasing  $p$ , as expected. The decrease rate is at its greatest when  $p$  increases from 2 to 6. This denotes that there is a significant improvement in every step in which the measured spectra are described with an additional factor beginning from the set of two factors only. It also denotes that there could be six main sources resulting in the measured spectra. However,  $Q/Q_{\text{exp}}$  still decreases slightly from  $p = 6$  to  $p = 9$ , denoting that taking more than six sources into account provides a slight improvement in describing the spectra with more factors. The decreasing rate of  $Q/Q_{\text{exp}}$  diminishes almost totally after  $p = 9$  and  $Q/Q_{\text{exp}}$  sets to 4. Theoretically,  $Q/Q_{\text{exp}}$  should set to 1 if the uncertainty matrix  $S$  has been constructed ideally. Thus,  $Q/Q_{\text{exp}}$  of 4 being over 1 denotes that underestimations exist in estimating the uncertainties

of the measured spectra. According to the decrease rates, the optimal  $p$  would be in the range of 6–9. The lowest sensible  $p$  is preferred because it simplifies further analysis. Because organic compounds are in a key role in this study, the value  $p = 8$  was selected for further analysis because it is the lowest value producing the MS matrix which includes at least 4 four different factor profiles having organic patterns separated with 14–16 Th (3 three factor profiles having organic patterns with  $p = 6$  or  $p = 7$ , see Fig. S1). These separations correspond to the increments of mass due to additional oxygen molecules or due to different carbon chain lengths.



**Figure 2.** The lowest obtained  $Q/Q_{\text{exp}}$  as a function of the number of factors ( $p$ ).

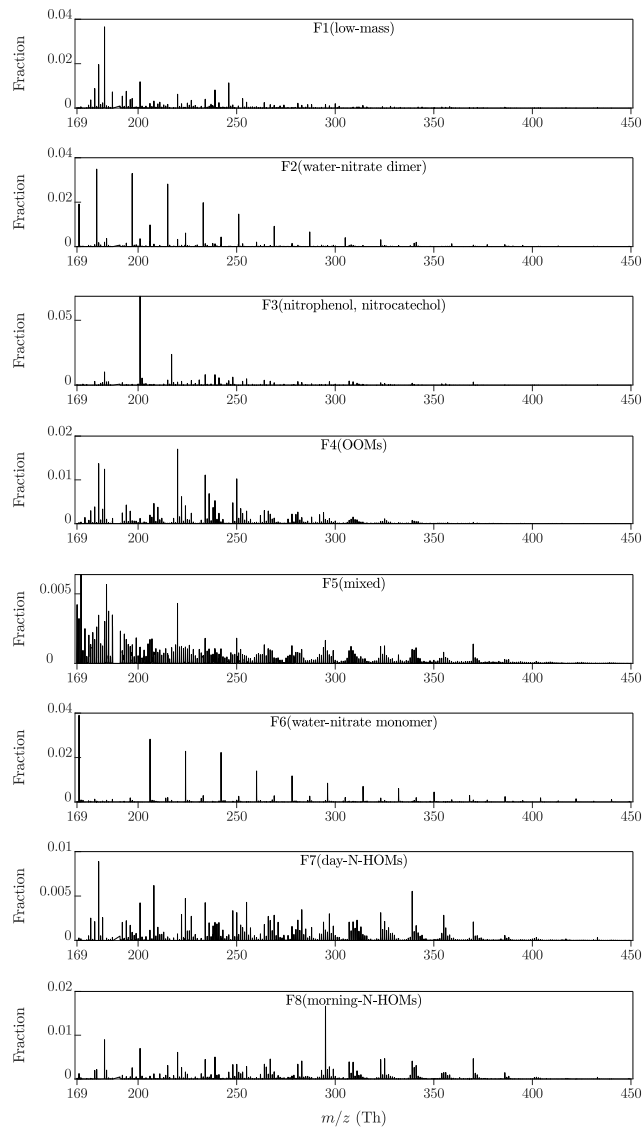
#### 4.2.2 Selected set of factors

Figure 3 presents an overview of the factor profiles (arranged according to their average masses) and Fig. 4 their diurnal variations obtained with  $p = 8$ . It can be observed with a rough examination that the factors 4, 5, 7, and 8 have organic patterns. In addition to those, factors 2 and 6 have sharp peaks repeating with 18 Th, which correspond to the nitrate monomer and dimer clustered with different numbers of water molecules. Conversely, the remaining factors, 1 and 3, have peaks located with no clear patterns. The diurnal variations show that the factors 1, 3, 4, and 7 are typically encountered in the daytime, while the rest are at their highest during mornings.

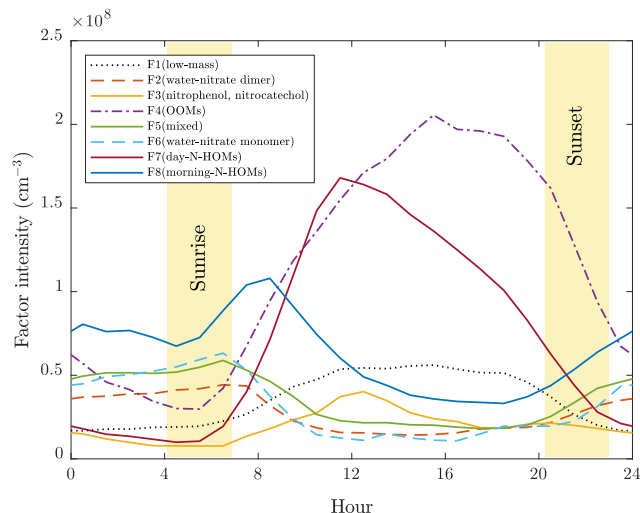
The key strength of a binPMF analysis is its applicability in examining the factor profiles with high mass resolution. A more thorough examination of the factor profiles and their naming is performed next. Several peaks were identified from the profiles using the tofTools code and the highest ones in every profile are presented in Table 1. A more extensive peak list can be found in Table S2.

It is worthwhile to note that the  $m/z$  axis was calibrated using the known compounds within the range of 62–201 Th, causing that the peaks at the higher end of the spectra are systematically slightly shifted toward larger values. The median error of the  $m/z$  calibration during the campaign was 0.4 ppm. However, the errors were as high as 80 ppm during eight nights near the end of the campaign due to high  $\text{H}_2\text{CO}_3$  signals causing overlapping signals to the  $\text{NO}_3^-$  and  $\text{NO}_3^- \cdot \text{HNO}_3$  peaks used





**Figure 3.** Factor profiles normalized with the total factor signal, i.e., the sum of the signal from every  $m/z$  bin of a factor is 1. Zoomed views of these spectra can be found in Figs. [S1–S8](#) [S2–S9](#) and their raw spectra are also available at Olin (2022).



**Figure 4.** Mean diurnal variations of the factor intensities. The yellow areas denote the variation in sunrise and sunset times during the experiments.

in the calibration. Nevertheless, the result of the calibration, which is the parameters used to convert time-of-flight to  $m/z$ , did not include any significant deviation during those nights, implying that the calibration itself was successful although the choice of the peaks for the calibration were not optimal for those nights. Additionally, due to the limited mass resolution of a binPMF analysis, the spectra in the factor profiles can be further dispositioned slightly. Nevertheless, identifying possible chemical formulae from the observed peaks was still done with relatively high confidence for several peaks, aided by observed errors of  $m/z$  ratios of water clusters (see Sect. 4.2.3). Note that determining correct isomers of identified chemical formulae is generally impossible with the TOF-mass spectrometer.

#### 4.2.3 Characteristics of the factors 2 and 6

The factors 2 and 6 are the factor profiles which include the clearest and most confidently identified peaks and their repeating pattern. The factor profiles consist almost solely of the clusters of nitrate dimer (the factor 2) or nitrate monomer (the factor 6) and 3–16 (the factor 2) or 6–21 (the factor 6) water molecules. They also have similar morning-type diurnal patterns but are still separated into different factors because the factor 2 has a higher intensity in the beginning part of the measurement period and the factor 6 vice versa. These water clusters are instrumental impurities caused by humidity inside the CI-inlet, where water molecules cluster with newly formed nitrate ions. In this article, these factors are named *F2(water-nitrate dimer)* and *F6(water-nitrate monomer)*.

Because the water cluster-peaks in *F2(water-nitrate dimer)* and *F6(water-nitrate monomer)* are detected with very high confidence up to 400 Th, their errors can be used as another type of mass calibration, performed after the binPMF analysis. The water cluster-peaks have errors between -10 ppm and 0 ppm when  $m/z$  is below 280 Th and increase from 0 ppm to +50

**Table 1.** The highest peaks in the binPMF factors. Observed  $m/z$  ratios<sup>a</sup>, chemical formulae corresponding best to them (in neutral forms, i.e., without the nitrate ion or with an added proton)<sup>b</sup>, errors between the observed and exact  $m/z$  ratios<sup>c</sup>, possible compound names<sup>d</sup>, and confidence levels of the identifications<sup>e</sup> are presented. The confidence levels are based on subjective estimations on the success of peak fitting (aided by the errors of water clusters at specific  $m/z$  ratios and by the known isotopic patterns) and on the expectation of the compound to be detected with the used instrumentation. A more extensive peak list can be found in Table S2.

F1 (low-mass)	F2 (water-nitrate dimer)	F3 (nitrophenol, nitrocatechol)	F4 (OOMs)	F5 (mixed)	F6 (water-nitrate monomer)	F7 (day-N-HOMs)	F8 (morning-N-HOMs)
<sup>a</sup> 182.9960 Th	179.0170 Th	201.0171 Th	220.0465 Th	171.0606 Th	170.0531 Th	180.0156 Th	295.0749 Th
<sup>b</sup> C <sub>7</sub> H <sub>4</sub> O <sub>6</sub>	(H <sub>2</sub> O) <sub>3</sub> HNO <sub>3</sub>	C <sub>6</sub> H <sub>4</sub> OHNO <sub>2</sub>	C <sub>7</sub> H <sub>10</sub> O <sub>4</sub>	unidentified	(H <sub>2</sub> O) <sub>6</sub>	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	C <sub>9</sub> H <sub>15</sub> O <sub>6</sub> N
<sup>c</sup> -14 ppm	-7.7 ppm	-9.3 ppm	-1.5 ppm		-8.4 ppm	-4.0 ppm	+11 ppm
<sup>d</sup> chelidonic acid	water-nitrate dimer	nitrophenol	very high		water-nitrate	succinic acid	moderate
<sup>e</sup> low	very high	very high			very high	very high	
180.0171 Th	197.0278 Th	217.0109 Th	180.0173 Th	183.9966 Th	206.0747 Th	208.0159 Th	182.9943 Th
C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	(H <sub>2</sub> O) <sub>4</sub> HNO <sub>3</sub>	C <sub>6</sub> H <sub>3</sub> (OH) <sub>2</sub> NO <sub>2</sub>	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	unidentified	(H <sub>2</sub> O) <sub>8</sub>	C <sub>5</sub> H <sub>6</sub> O <sub>5</sub>	C <sub>7</sub> H <sub>4</sub> O <sub>6</sub>
-12 ppm	-7.8 ppm	-3.3 ppm	-13 ppm		-9.0 ppm	-29 ppm	-4.9 ppm
succinic acid	water-nitrate dimer	nitrocatechol	succinic acid		water-nitrate	ketoglutaric acid	chelidonic acid
high	very high	high	high		very high	low	moderate
201.0153 Th	215.0387 Th	182.9971 Th	182.9899 Th	220.0135 Th	224.0849 Th	339.0585 Th	201.0159 Th
C <sub>6</sub> H <sub>4</sub> OHNO <sub>2</sub>	(H <sub>2</sub> O) <sub>5</sub> HNO <sub>3</sub>	C <sub>7</sub> H <sub>4</sub> O <sub>6</sub>	C <sub>7</sub> H <sub>4</sub> O <sub>6</sub>	unidentified	(H <sub>2</sub> O) <sub>9</sub>	C <sub>10</sub> H <sub>15</sub> O <sub>8</sub> N	C <sub>6</sub> H <sub>4</sub> OHNO <sub>2</sub>
-0.4 ppm	-8.9 ppm	-20 ppm	+19 ppm		-6.5 ppm	+28 ppm	-3.4 ppm
nitrophenol	water-nitrate dimer	chelidonic acid	chelidonic acid		water-nitrate	moderate	nitrophenol
very high	very high	moderate	low		very high		very high
246.0000 Th	(H <sub>2</sub> O) <sub>6...16</sub> HNO <sub>3</sub>	234.0615 Th	C <sub>8...9</sub> H <sub>12...14</sub>	OOMs, HOMs,	(H <sub>2</sub> O) <sub>10...21</sub>	C <sub>7...10</sub> H <sub>11...15</sub>	C <sub>7...10</sub> H <sub>11...15</sub>
unidentified		C <sub>8</sub> H <sub>12</sub> O <sub>4</sub>	O <sub>4...6</sub>	and N-HOMs		O <sub>6...9</sub> N	O <sub>6...9</sub> N
		+1.7 ppm		240–390 Th			
		terpenylic acid					
		moderate					

ppm when  $m/z$  increases from 280 Th to 400 Th. Hence, also the peaks in the other factors are expected to have similar error levels.

#### 4.2.4 Characteristics of the factor 3

The factor 3 comprises mainly of a peak typically encountered with nitrate ionization,  $\text{NO}_3^- \cdot \text{C}_6\text{H}_4\text{OHNO}_2$  (nitrophenol,  $\text{C}_6\text{H}_4\text{OHNO}_2$ , clustered with a nitrate ion), and a peak at 217.0109 Th. The latter one is possibly from  $\text{C}_6\text{H}_3(\text{OH})_2\text{NO}_2$ , which corresponds to nitrocatechol (or its isomer). Both nitrophenol and nitrocatechol are connected at least to biomass burning (Iinuma et al., 2010). Additionally, other peaks located with no clear pattern, however, with much lower intensities exist in the factor 3. The ones of these with the highest intensities include peaks at 182.9971 Th and 234.0615 Th, which correspond best with  $\text{C}_7\text{H}_3\text{O}_6^-$  and with  $\text{NO}_3^- \cdot \text{C}_8\text{H}_{12}\text{O}_4$ , respectively. One possibility is that they are from chelidonic and terpenylic acid, both connected to vegetation. This factor is named here *F3(nitrophenol, nitrocatechol)*.

#### 4.2.5 Characteristics of the factor 4

The next clearest of the factor profiles is the factor 4. It consists mainly of oxidized organic molecules (OOMs) within the  $m/z$  range of 200–350 Th (including clustered nitrate ions, which are hereafter omitted from the shown chemical formulae for convenience) having the chemical formulae of a form of  $\text{C}_x\text{H}_y\text{O}_z$ , of which the peaks identified with high confidence have  $x$  of 7, 8, or 9,  $y$  of 10, 12, or 14, and  $z$  of 4, 5, or 6. The largest of these molecules can also be considered as HOMs, by the definition of a HOM having at least six oxygen atoms (Bianchi et al., 2019). Additionally, the factor 4 includes peaks corresponding likely with succinic acid and possibly with chelidonic acid (at 182.9899 Th). Succinic acid together with peaks with lower intensities and lower confidence levels also likely found, fumaric and malic acid, belong in the citric acid cycle, which is a metabolic pathway of aerobic organisms, including plants. However, the peak at 182.9899 Th can potentially be something other than chelidonic acid, at least a different compound than in other factors having a peak at  $\sim 183$  Th, because the peak in this factor falls on the other side of the exact mass of deprotonated chelidonic acid (182.9935 Th) compared to the other factors. A lower error than for chelidonic acid (+19 ppm) is achieved with  $\text{C}_2\text{H}_3\text{O}_5\text{N}$  (-3 ppm), which could be peroxyacetyl nitrate but its detection with  $\text{NO}_3^-$ -based ionization has not been reported before. This factor is named here *F4(OOMs)*.

#### 4.2.6 Characteristics of the factors 7 and 8

The factors 7 and 8 have repeating organic patterns (mainly N-HOMs which are HOMs with an extra nitrogen atom) in the  $m/z$  range of 230–410 Th. These N-HOMs have the chemical formulae of a form of  $\text{C}_x\text{H}_y\text{O}_z\text{N}$ , of which the peaks identified with high confidence have  $x$  of 7, 8, 9, or 10,  $y$  of 11, 13, or 15, and  $z$  of 6, 7, 8, or 9. Though the transmission efficiency decreases with increasing  $m/z$ , a high peak at 339.0585 Th being likely  $\text{C}_{10}\text{H}_{15}\text{O}_8\text{N}$  in the factor 7 and a very high peak at 295.0749 Th being likely  $\text{C}_9\text{H}_{15}\text{O}_6\text{N}$  in the factor 8 are distinguished clearly from the profiles. The first indication of the connection between  $\text{C}_{10}\text{H}_{15}\text{O}_8\text{N}$  and atmospheric nucleation was reported by Kulmala et al. (2013).

The profiles of the factors 7 and 8 differ also in a way that the peaks at the lower end of the  $m/z$  range are of higher intensity in the factor 7 than in the factor 8. These peaks in the factor 7 include also compounds that can be related to the citric acid cycle, e.g., succinic, ketoglutaric, and malic acid. Instead, no citric acid cycle-related compounds exist in the factor 8 but nitrophenol and possible chelidonic acid do. According to their diurnal patterns, these factors are named here *F7(day-N-HOMs)* and *F8(morning-N-HOMs)*.

#### 4.2.7 Characteristics of the factor 5

The factor 5 is elevated at nighttime and consists of several unidentified peaks in the lower end of the  $m/z$  range and of OOMs, HOMs, and N-HOMs mixed with each other in the  $m/z$  range of 240–390 Th. No clear pattern can be observed from this mixture.

One key feature in the factor 5 is also its clearly increasing intensity toward the summer, while the other factors do not show that kind of behavior. Although with low intensity and confidence, there are also peaks suggesting citric acid cycle-related compounds, e.g., fumaric, malic, and aconitic acid. Additionally, although not directly included in the factor 5, pyruvic acid (at 87.0088 Th in the spectra before binPMF), a key compound in the beginning of the citric acid cycle, seems to be encountered simultaneously with the factor 5. This factor is named here *F5(mixed)*.

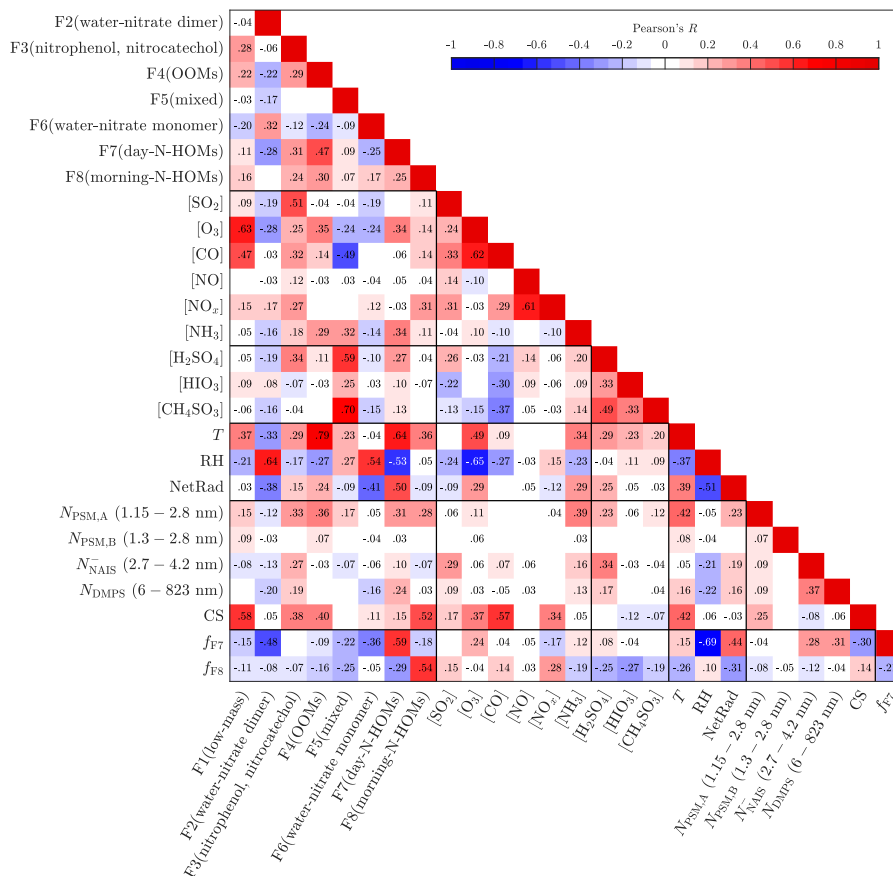
#### 4.2.8 Characteristics of the factor 1

The last one of the factor profiles, the factor 1, is elevated at daytime and consists mostly of low-mass compounds, including nitrophenol, succinic acid, probably chelidonic and malic acid, and several other peaks without any clear pattern between them. This factor is named here *F1(low-mass)*.

### 4.3 Correlations between variables

Figure 5 presents Pearson's correlation coefficients ( $R$ ) between all possible pairs of measured variables and binPMF factors in 10 min time resolution. Connecting any variable to NPF events can be estimated with correlations to particle variables, such as to the number concentration ( $N$ ) measured by PSMs, NAIS, or DMPS, and to CS. It should be noted that this is only an estimation on the connections between variables and NPF events. Thus, it cannot be certainly proved that a variable is actually forming new particles or growing them by examining the correlations. There is always a possibility that a variable is only observed simultaneously with NPF events due to the similarity of its source and the source of the precursor really causing the NPF events. This kind of uncertainty is, however, present in all similar studies as well; and therefore, the observed correlations should also be assessed by considering whether they are physically reasonable or not. In Fig. 5, the number concentrations from the different devices represent different particle size ranges; additionally,  $N_{\text{NAIS}}^-$  refers only to particles detected using the negative-polarity charger. It can be seen that clearly positive correlations for all particle sizes are observed with F3(nitrophenol, nitrocatechol), F7(day-N-HOMs),  $[\text{SO}_2]$ ,  $[\text{NH}_3]$ ,  $[\text{H}_2\text{SO}_4]$ ,  $T$ , and NetRad, of which F3(nitrophenol, nitrocatechol) and  $[\text{H}_2\text{SO}_4]$  have the strongest ones. It is clear why  $T$  and NetRad are positively correlated with particles because

NPF events almost exclusively occur during the daytime. The connections of  $[\text{NH}_3]$  and  $[\text{H}_2\text{SO}_4]$  to NPF were observed already with Fig. 1b, which also explains the correlation of  $[\text{SO}_2]$  because  $\text{H}_2\text{SO}_4$  is formed from  $\text{SO}_2$ . The strongest negative correlations are observed with RH, which is obvious due to the fact that NPF events generally occur mostly with low RH.



**Figure 5.** Pearson's correlation coefficient ( $R$ ) matrix between measured variables and binPMF factors. Values of  $R$  are shown as colors and numerically. They are shown only for statistically significant pairs ( $p$ -value below 0.05).

The correlations between the binPMF factors and the other variables give hints for the contents and sources of the factors. F1(low-mass) correlates relatively well with  $[\text{O}_3]$ , suggesting that F1(low-mass) includes compounds related to ozonolysis. F2(water-nitrate dimer) and F6(water-nitrate monomer) correlate well with RH because they consist mainly of the water clusters of nitrate, which are more abundant in high RH. F3(nitrophenol, nitrocatechol) correlates well with  $[\text{SO}_2]$ ,  $[\text{CO}]$ ,  $[\text{NO}_x]$ , and  $[\text{H}_2\text{SO}_4]$ , which suggests that F3(nitrophenol, nitrocatechol) could be originated from combustion sources, such as from tractors, as already discussed before. F4(OOMs) has the highest correlations with  $[\text{O}_3]$  and  $[\text{NH}_3]$  and also a very high correlation with  $T$ . A deeper examination between F4(OOMs) and  $T$  shows that an even higher correlation is achieved with a vapor pressure-type function of  $T$ ,  $\exp(AT + B)$ , which implies that the origin of F4(OOMs) could be related to the evaporation

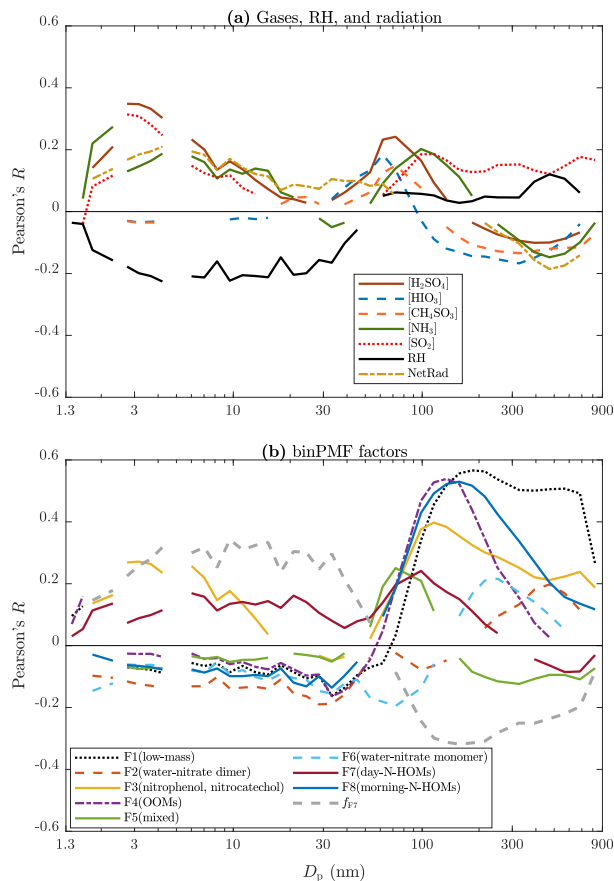
of some compounds, possibly requiring  $O_3$  to be detected as HOMs with the CI-API-TOF. F5(mixed) correlates well with the concentrations of inorganic compounds, such as with  $[NH_3]$ ,  $[H_2SO_4]$ ,  $[HIO_3]$ , and  $[CH_4SO_3]$ .

F7(day-N-HOMs) correlates well with  $[O_3]$ ,  $[NH_3]$ , and  $[H_2SO_4]$  whereas F8(morning-N-HOMs) correlates well with  $[NO_x]$ . Their strongest correlations, however, are against  $T$ , RH, and NetRad, of which  $T$  correlates positively with both F7(day-N-HOMs) and F8(morning-N-HOMs), but RH and NetRad behave differently. F7(day-N-HOMs) has a strong negative correlation with RH and a strong positive correlation with NetRad, whereas F8(morning-N-HOMs) has weak correlations with RH and NetRad in the opposite directions. F7(day-N-HOMs) and F8(morning-N-HOMs) are further examined by defining variables  $f_{F7}$  and  $f_{F8}$ , which refer to their fractions in the total spectrum. In other words, e.g.,  $f_{F7}$  denotes how much of the spectrum is in an F7-like form. A positive correlation of  $f_{F7}$  with NetRad and a negative correlation of  $f_{F8}$  with NetRad denote that the spectrum prefers the F7-like form over the F8-like form with higher NetRad levels. This can be interpreted so that N-HOMs are in the F8-like form in the mornings but solar radiation seemingly transforms them into the F7-like form for daytime. It cannot be proven with these measurements that this transformation is actually a real chemical process since the behavior of the factors with NetRad levels can also be resulted, e.g., from possibly different origins of the air masses between the daytime and the nighttime. Nevertheless, the current knowledge cannot at least exclude the transformation via radiation, because the compounds in F7(day-N-HOMs) are basically more oxidized than the ones in F8(morning-N-HOMs) (discussed later in Sect. 4.4) being physically reasonable since photochemistry typically leads to oxidation.

Particle concentrations are expressed with only 4 size ranges (total concentrations from PSM,A, PSM,B, NAIS, and DMPS) and with CS in Fig. 5. The connection between NPF events and the measured variables can be further examined by expressing the correlation coefficients for all particle sizes (Fig. 6a). Notable is that the correlation coefficients for  $[H_2SO_4]$ ,  $[NH_3]$ , and  $[SO_2]$  are positive from the smallest particle sizes up to around 30 nm. It suggests that they are involved in forming new particles and growing them to those sizes but not to larger sizes because the correlation coefficients approach zero near 30 nm. It is known that high NetRad and low RH favor NPF events. This can be observed in Fig. 6a as positive and negative correlation coefficients, respectively, from the smallest sizes up to  $\sim 60$  nm, which can be interpreted as a rough validation of this method in examining connections between variables and NPF events. The most interesting particle size range in terms of a NPF event in this case is around 10–40 nm because the background aerosol is usually in larger sizes and the newly formed particles in smaller sizes. Particle concentrations within this size range increase mainly during NPF events only; the size distributions of the newly formed particles sweep past this size range (see Fig. 1). In Fig. 6a, only the correlation coefficient for NetRad stays clearly positive (and clearly negative for RH) within the size range of 10–40 nm.

Figure 6b presents the correlation coefficients for the binPMF factors as a function of particle size. F3(nitrophenol, nitrocatechol) seems to be somehow involved in forming new particles but not in growing them past the 10–40 nm size range. The correlation coefficient for F7(day-N-HOMs), instead, stays on the positive side even up to 300 nm, while for all other factors they are on the negative side within the lower half of the whole particle size range. Within the size range of 10–40 nm, F7(day-N-HOMs) is the only one of the factors having positive correlation coefficients. Even stronger correlations are achieved when examining the fraction of F7(day-N-HOMs) in the total spectrum,  $f_{F7}$ . From these plots (Fig. 6), it can be hypothesized that  $[H_2SO_4]$ ,  $[NH_3]$ , F3(nitrophenol, nitrocatechol) and F7(day-N-HOMs) participate in forming new particles but only F7(day-



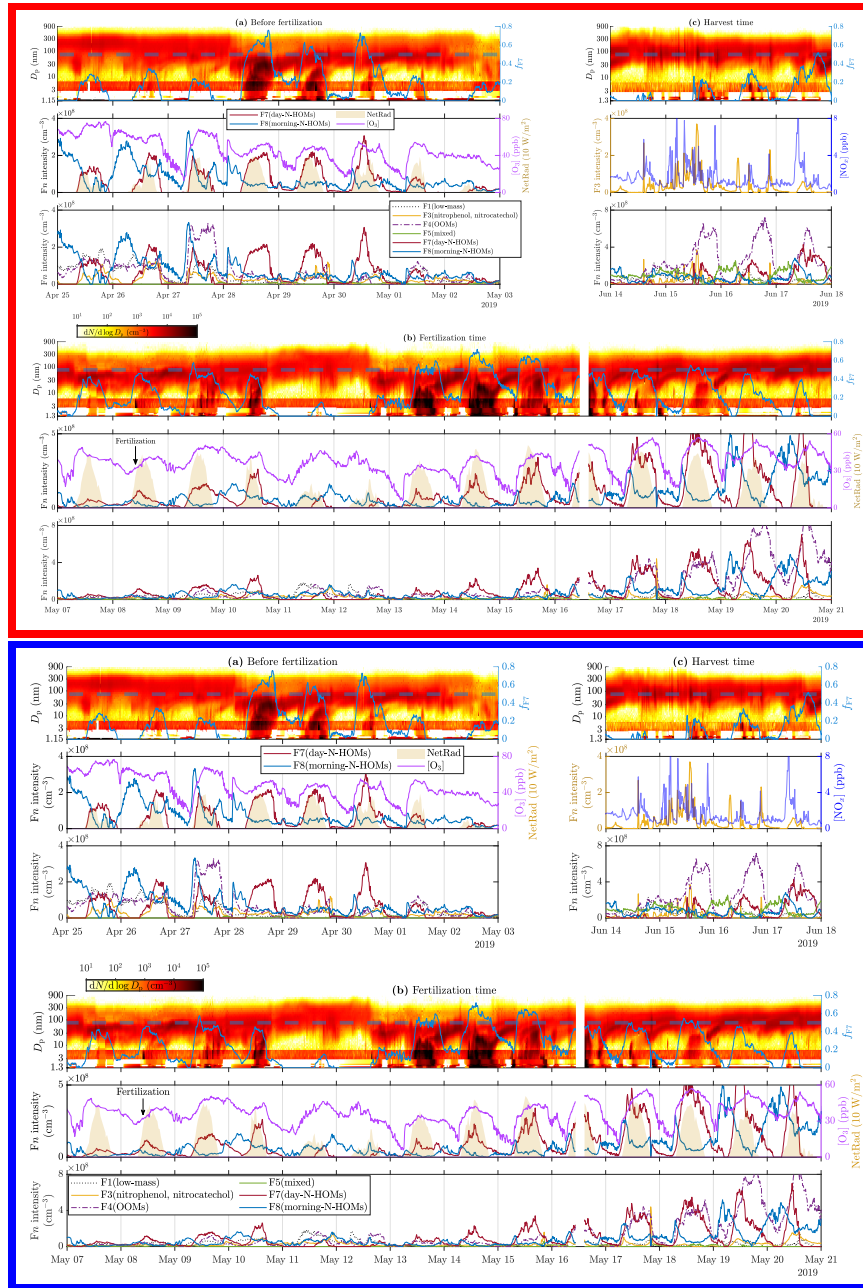


**Figure 6.** Pearson's correlation coefficient ( $R$ ) between all particle size bins and (a) measured variables and (b) binPMF factors. Data below 3 nm are from a PSM, over 5 nm are from the DMPS, and the remaining part between them are from the NAIS. Lines are shown only for statistically significant pairs (p-value below 0.05).

N-HOMs) is related to growing the particles large enough to provide a full NPF event. Figure S9-S10 presents similar plots for  $[\text{H}_2\text{SO}_4]$ ,  $[\text{NH}_3]$ , F3(nitrophenol, nitrocatechol), F7(day-N-HOMs), and  $f_{F7}$  separately for days with and without NPF events. It can be seen that correlations are stronger during the days with NPF events for particle sizes up to  $\sim 40$  nm.

#### 4.4 Time series with binPMF factors

Figure 7 presents time series of particle size distribution contours with factor intensities and  $[\text{O}_3]$  or  $[\text{NO}_x]$  for the three example time ranges. It can be observed from Fig. 7a,b that F8(morning-N-HOMs), firstly, increases simultaneously with  $[\text{O}_3]$  and, secondly, seemingly transforms to F7(day-N-HOMs) with increased NetRad levels. NPF events seem to occur when the intensity of F7(day-N-HOMs) is high enough but it is not simply controlling the strength of a NPF event (**defined as the duration of continuing new particle formation during a NPF event**). For example in Fig. 7b, the daytime intensity of F7(day-N-HOMs)



**Figure 7.** Time series of particle size distribution contours and  $f_{F7}$  (top panels),  $F7$ (day-N-HOMs) and  $F8$ (morning-N-HOMs) intensities,  $[O_3]$ , NetRad (shaded areas in the middle panels), and other factor intensities except for the water cluster-based ones (bottom panels; the legend is shown in the bottom panel of (c)) for different example time ranges, (a) before the fertilization, (b) during and right after the fertilization, and (c) during the harvest. Note that, in (c), the middle panel presents  $F3$ (nitrophenol, nitrocatechol) intensity and  $[NO_x]$  instead. The dashed lines shown in the top panels denote the critical  $f_{F7}$  value of  $\sim 0.50 \pm 0.05$ . Particle data below 3 nm are from a PSM, over 6 nm are from the DMPS, and the remaining part between them are from the NAIS. Note that the PSM data shown in (a) are from the PSM,A whereas the data shown in (b) and (c) are from the PSM,B due to different coverage of the devices. Also note different scales of the y-axes between the subfigures.

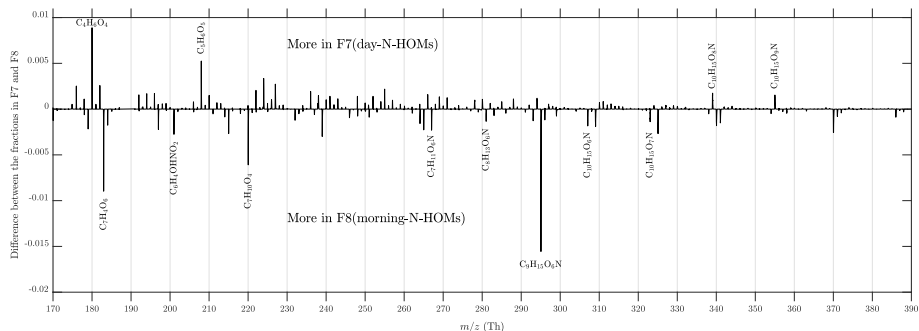
has an increasing trend beginning on May 12th but the NPF event strength has a decreasing trend beginning already on May 14th. Notable is that also other factors, especially F4(OOMs), have increasing trends during that time range. Therefore, when investigating the fraction  $f_{F7}$ , it can be seen that it has a decreasing trend beginning on May 14th as well. For the whole studied three-month time range,  $f_{F7}$  seems to correlate well with the existence or the strength of a NPF event. There seems to be a critical value for  $f_{F7}$  (near  $0.50 \pm 0.05$ ), exceeding of which induces a full NPF event. Additionally, when  $f_{F7}$  decreases below the critical value during a NPF event, the event is usually terminated. The diurnal average of  $f_{F7}$  is 0.31 on the days showing clear features of NPF events and 0.10 on the days showing hardly any features of NPF events. If considering only a typical time range of observed NPF events, from 10:00 to 20:00, the average of  $f_{F7}$  on the days with NPF events is 0.46. In conclusion, other factors than F7(day-N-HOMs) seem to act as an inhibitor for a NPF event, similarly to CS; however, the mechanism behind disfavoring NPF events by the sum of other factors is unknown. Notable is that  $f_{F7}$  does not include any particle variables even though particles should act as a sink for particle forming vapors and thus disfavor NPF events. This could be explained partly by the fact that CS and  $f_{F7}$  are negatively correlated ( $R = -0.30$ ).

Particle spikes during the harvest time in Fig. 7c occur simultaneously with the F3(nitrophenol, nitrocatechol) intensity and  $[\text{NO}_x]$ . As discussed before in Sect. 4.1, the spikes may be related to tractor emissions. The spikes in  $[\text{NO}_x]$  support the hypothesis of these particles being emitted by tractors but F3(nitrophenol, nitrocatechol) consists mostly of compounds related to vegetation and biomass burning; thus, it is not obvious why it spikes too. The reasons could be that fossil fuel-combusting vehicles emit some compounds found in F3(nitrophenol, nitrocatechol) or that they are released from cut grass during the harvesting process. As already seen from Fig. 6b, increased F3(nitrophenol, nitrocatechol) intensity is connected to increased particle concentrations below 20 nm but not to growing them toward larger sizes and thus not to inducing a NPF event.

It is evident that F7(day-N-HOMs) is connected to particle formation and growth process and F8(morning-N-HOMs) acts ~~as~~ like a precursor for F7(day-N-HOMs) ~~because the F8(morning-N-HOMs) level decreases with increasing F7(day-N-HOMs).~~ Assuming similar sensitivities of F7(day-N-HOMs) and F8(morning-N-HOMs) in the CI-APi-TOF, the mean intensity of F8(morning-N-HOMs) is 2-fold the mean intensity of F7(day-N-HOMs) with low radiation levels ( $\text{NetRad} < 130 \text{ W m}^{-2}$ ); conversely, the mean intensity of F7(day-N-HOMs) is 3-fold the mean intensity of F8(morning-N-HOMs) with high radiation levels. This ~~confirms-supports~~ the role of radiation in transforming N-HOMs from the F8-like form into the F7-like form. The difference in their profiles is further examined in Fig. 8. It can be seen that the clearest differences are, e.g., more  $\text{C}_4\text{H}_6\text{O}_4$  (possibly succinic acid) and  $\text{C}_5\text{H}_6\text{O}_5$  (possibly ketoglutaric acid) existent in F7(day-N-HOMs) and more  $\text{C}_7\text{H}_4\text{O}_6$  (possibly chelidonic acid),  $\text{C}_7\text{H}_{10}\text{O}_4$ , and  $\text{C}_9\text{H}_{15}\text{O}_6\text{N}$  existent in F8(morning-N-HOMs). In general, the compounds existing more in F8(morning-N-HOMs) have ~~oxygen-to-carbon (O:C )~~ ratios of around 0.7, but the ones in F7(day-N-HOMs) have O:C ratios of around 0.9, indicating that solar radiation transforming F8(morning-N-HOMs) to F7(day-N-HOMs) is related to oxidation of organic compounds.

#### 4.5 Suggested explanation for particle formation

Figure 9 presents a suggested explanation for particle formation and growth observed in the studied area. The green arrows denote the route leading to a NPF event, which requires ozone to form F8(morning-N-HOMs) which then transforms to F7(day-



N-HOMs) with available solar radiation. Finally, the fraction  $f_{F7}$  needs to be high enough ( $>0.5$   $>0.50 \pm 0.05$ ) for a NPF event to be induced.

**Figure 9.** Suggested explanation for particle formation and growth observed in the studied area. The green arrows, including ozone, radiation, and high  $f_{F7}$ , denote the route leading to a NPF event. The red arrows present which variables decrease the probability of a NPF event. The + and − signs after  $f_{F7}$  denote the direction and the strength of the change of  $f_{F7}$ , which needs to be high ( $>0.50 \pm 0.05$ ) for a NPF event to be induced. Additionally, increased  $[\text{NO}_x]$ ,  $[\text{H}_2\text{SO}_4]$ , and F3(nitrophenol, nitrocatechol) levels can be seen as particle spikes at the diameters of 20–200 nm. Moreover, increased  $[\text{NH}_3]$ ,  $[\text{H}_2\text{SO}_4]$ , and F3(nitrophenol, nitrocatechol) levels can lead to sub-10 nm particles, which can be grown larger with the assistance of F7(day-N-HOMs).

Because the occurrence of NPF events seems to be controlled by  $f_{F7}$ , other factors than F7(day-N-HOMs) can inhibit NPF events if their intensities increase. Additionally, increasing temperature (or its exponential form) leads to more F4(OOMs), which does not favor a NPF event but more likely disfavors it. Hence, radiation is connected to inducing NPF events but [high](#) temperature to disfavoring them. These are typically connected also to each other but their different roles in NPF may explain why NPF events are generally observed in all seasons, with high radiation levels but with high  $T$  in summers and vice versa in winters. Other variables (F1(low-mass), F5(mixed), F8(morning-N-HOMs), [HIO<sub>3</sub>], and [CH<sub>4</sub>SO<sub>3</sub>]) were observed to not induce a NPF event; instead, F1(low-mass), F5(mixed), and F8(morning-N-HOMs) are slightly connected to lowering  $f_{F7}$ .

As demonstrated in Figs. 1c and 7c, spikes in [NO<sub>x</sub>], [H<sub>2</sub>SO<sub>4</sub>], and F3(nitrophenol, nitrocatechol) lead to particle spikes at the diameters of 20–200 nm, probably due to nearby tractor emissions. The particle spikes are, however, different from NPF events in terms of their time scales and particle sizes. Figure 6 suggests that [NH<sub>3</sub>], [H<sub>2</sub>SO<sub>4</sub>], and F3(nitrophenol, nitrocatechol) can form sub-10 nm particles but not larger. In theory, there is a possibility that the smallest particles are, in some cases, formed via those but their growth occurs via F7(day-N-HOMs), and a NPF will eventually be observed. Also possible is that particles are both formed and grown via F7(day-N-HOMs). It should be noted that the studied farm may not be the actual source of particle forming precursors as the site is close to the sea and forest as well and that F7(day-N-HOMs) may not include the actual compound(s) behind the particle formation and growth but is observed simultaneously with the actual compound(s) instead.

#### 4.6 Applications in other studies with nitrate CI-API-TOF spectra

Estimating the binPMF factors extracted from this study in analyses of other studies can be done without performing a binPMF analysis by using unit mass resolution (UMR) tracers. Table S3 presents Pearson's correlation coefficients between the most important binPMF factors and UMR data. UMR data are simply the sums from the spectra between, e.g., N-0.5 Th and N+0.5 Th. However, UMR data can lack important information on multiple compounds overlapping at the same UMR mass-to-charge ratio.

For example, one can estimate how F7(day-N-HOMs) would behave in another study by the time series of the spectra between 284.5 Th and 285.5 Th (UMR285). Pearson's  $R$  between UMR285 and F7(day-N-HOMs) is 0.91, whereas between UMR285 and F8(morning-N-HOMs)  $R$  is only 0.36 (for comparison,  $R = 0.25$  between F7(day-N-HOMs) and F8(morning-N-HOMs)) and between UMR285 and F4(OOMs)  $R$  is only 0.65 ( $R = 0.47$  between F7(day-N-HOMs) and F4(OOMs)). The most selective UMR tracer would be one that has  $R = 1$  for the factor in question but near the ones in Fig. 5 for the all other factors. However, this is not the case for any factor, but the ones in Table S3 denote the most promising ones with their correlation coefficients. If one needs to estimate the variable  $f_{F7}$ , the most promising tracers are UMR271 divided by UMR260 ( $R = 0.86$  between the division and  $f_{F7}$ ), UMR285 divided by UMR197 ( $R = 0.84$ ), and UMR313 divided by UMR260 ( $R = 0.84$ ). However, it should be noted that the critical value depends on the selected UMR tracers.

## 5 Conclusions

New particle formation events occurring in a coastal agricultural site were examined by performing measurements of gases, molecular clusters, particles, and environmental parameters. The area is a pilot agricultural farm for regenerative farming, Qvidja, located in Southwestern Finland. This study covers roughly three months of the measurement data recorded between April and June in 2019, ~~when a laboratory van, ATMo-Lab, was parked next to a stationary measurement container located in the middle of the farming fields.~~ 2019. The CI-API-TOF mass spectrometer was used to measure potential new particle-forming acids, sulfuric acid, iodic acid, and methanesulfonic acid, together with a multitude of other compounds. In addition to these acids, the high resolution-mass spectra between 169 and 450 Th were elaborated via the binned positive matrix factorization (binPMF) method. Eight binPMF factors were selected as the optimal set of factors.

From all 84 measurement days, 25 days showed clear features of NPF events. The NPF event days can be partly explained with ammonia, sulfuric acid, and condensation sink levels. The concentrations of ammonia and sulfuric acid were generally higher and condensation sink lower during the NPF events, but still without a clear formula on the occurrence of the events. Iodic acid and methanesulfonic acid were observed to not correlate with the NPF events, although the site is located near the sea. An even better explanation is, instead, achieved when examining the levels of the binPMF factors. It was observed that the factor F7(day-N-HOMs) is high during the NPF events. Further examination shows that the events can be explained very accurately using a single variable,  $f_{F7}$ , which denotes the fraction of the spectra between 169 and 450 Th that is in the form of F7(day-N-HOMs). In all NPF events,  $f_{F7}$  exceeded a critical value of  $\sim 0.5$   $0.50 \pm 0.05$  and the total time of the exceeding corresponds to the length of the event. Surprisingly, no particle values are needed to predict NPF events in this case even though the condensation sink generally disfavors the events.

~~Pearson's correlation coefficient between F7(day-N-HOMs) and particle concentrations at every measured particle size bin is positive from near 1 nm up to 300 nm, but all other factors behave differently. F3(nitrophenol, nitrocatechol) is positive for the smallest particle sizes but approaches zero near 20 nm, which can be interpreted so that F3(nitrophenol, nitrocatechol) may be involved in NPF but not in growing them past the size of 20 nm and thus not inducing NPF events. All other factors, instead, have negative correlation coefficients for particles smaller than 40 nm. F3(nitrophenol, nitrocatechol) together with sulfuric acid and nitrogen oxides was also connected to the appearance of short particle spikes in the size range of 20–200 nm, which are presumably due to the emissions of tractors harvesting the fields.~~

Examination of time series revealed that the intensity of F7(day-N-HOMs) is elevated with solar radiation when F8(morning-N-HOMs) exists, like F8(morning-N-HOMs) is transformed to F7(day-N-HOMs). Instead, F8(morning-N-HOMs) seems to be formed through ozonolysis because its intensity is elevated simultaneously with the ozone concentration. In conclusion, NPF events observed at the studied coastal agricultural environment seem to follow this routing: ozone levels elevate which causes F8(morning-N-HOMs) intensity to elevate, which is then transformed to F7(day-N-HOMs) via radiation; if F7(day-N-HOMs) is the major form in the spectra, a NPF event is observed.

Investigation of the high resolution-spectra of F7(day-N-HOMs) and F8(morning-N-HOMs) shows that they both consist mainly of HOMs with an extra nitrogen atom, but the compounds existing more in F7(day-N-HOMs) have higher oxygen-to-

carbon ratios than the ones in F8(morning-N-HOMs). Additionally, there seem to be compounds related to the citric acid cycle in F7(day-N-HOMs) but not in F8(morning-N-HOMs). Notable compounds in their spectra are  $C_{10}H_{15}O_8N$  (339.0585 Th) in F7(day-N-HOMs) and  $C_9H_{15}O_6N$  (295.0749 Th) in F8(morning-N-HOMs).  $C_{10}H_{15}O_8N$  is the compound with which the first indication of the connection between HOMs and atmospheric nucleation was observed (Kulmala et al., 2013). Another factor containing HOMs was also obtained from the binPMF analysis, F4(OOMs). It was observed to not favor NPF events, but more likely to disfavor instead, due to its elevated intensity with higher temperatures causing lower  $f_{F7}$ .

~~In estimating the behavior of F7(day-N-HOMs) and F8(morning-N-HOMs) in any other study with a nitrate ion-based mass spectrometer without performing the binPMF analysis, time series of those compounds could be useful in estimating. If only unit mass resolution data is available, the time series of 220 or 236 Th could be examined in estimating F4(OOMs), of 285, 271, or 339 Th in estimating F7(day-N-HOMs), and of 295 or 265 Th in estimating F8(morning-N-HOMs). The time series of 271 Th divided by the time series of 260 Th is the best estimate for examining the (relative) behavior of the fraction  $f_{F7}$ .~~

*Data availability.* Time series data measured at the Quidja farm from Apr to Jun 2019 together with the raw spectra of the binPMF factors are freely available at <https://doi.org/10.5281/zenodo.6394454> (Olin, 2022).

*Author contributions.* MDM, MS, TP, and MS designed the research. MOI, MOk, JK, ML, YW, and JD performed the measurements. MOI and MOk handled the measurement data. MOI performed the PMF analysis and prepared the paper with contributions from all co-authors.

*Competing interests.* Some authors are members of the editorial board of Atmospheric Chemistry and Physics. The peer-review process was guided by an independent editor, and the authors have also no other competing interests to declare.

*Acknowledgements.* University of Helsinki and Quidja technical staff and Quidja farm owners are acknowledged for their support on the site. We thank the tofTools team for providing tools for mass spectrometry analysis. This research has been supported by the Academy of Finland through the Condenz project (grant nos. 310627 & 326437), through the ACCC Flagship grant (no. 337551), and through the Academy Research Fellow grant (no. 331207).



## 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.-Atmos.*, 108, 4090, <https://doi.org/https://doi.org/10.1029/2002JD002358>, 2003.
- Baccarini, A., Karlsson, L., Dommen, J., Duplessis, P., Vüllers, J., Brooks, I. M., Saiz-Lopez, A., Salter, M., Tjernström, M., Baltensperger, U., Zieger, P., and Schmale, J.: Frequent new particle formation over the high Arctic pack ice by enhanced iodine emissions, *Nature Comm.*, 11, 1–11, <https://doi.org/10.1038/s41467-020-18551-0>, 2020.
- Beck, L. J., Sarnela, N., Junninen, H., Hoppe, C. J., Garmash, O., Bianchi, F., Riva, M., Rose, C., Peräkylä, O., Wimmer, D., Kausiala, O., Jokinen, T., Ahonen, L., Mikkilä, J., Hakala, J., He, X. C., Kontkanen, J., Wolf, K. K., Cappelletti, D., Mazzola, M., Traversi, R., Petroselli, C., Viola, A. P., Vitale, V., Lange, R., Massling, A., Nøjgaard, J. K., Krejci, R., Karlsson, L., Zieger, P., Jang, S., Lee, K., Vakkari, V., Lampilahti, J., Thakur, R. C., Leino, K., Kangasluoma, J., Duplissy, E. M., Siivola, E., Marbouti, M., Tham, Y. J., Saiz-Lopez, A., Petäjä, T., Ehn, M., Worsnop, D. R., Skov, H., Kulmala, M., Kerminen, V. M., and Sipilä, M.: Differing Mechanisms of New Particle Formation at Two Arctic Sites, *Geophys. Res. Lett.*, 48, e2020GL091334, <https://doi.org/10.1029/2020GL091334>, 2021.
- 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, *Chem. Rev.*, 119, 3472–3509, <https://doi.org/10.1021/acs.chemrev.8b00395>, 2019.
- Birmili, W. and Wiedensohler, A.: New particle formation in the continental boundary layer: Meteorological and gas phase parameter influence, *Geophys. Res. Lett.*, 27, 3325–3328, <https://doi.org/10.1029/1999GL011221>, 2000.
- Birmili, W., Berresheim, H., Plass-Dülmer, C., Elste, T., Gilge, S., Wiedensohler, A., and Uhrner, U.: The Hohenpeissenberg aerosol formation experiment (HAFEX): A long-term study including size-resolved aerosol, H<sub>2</sub>SO<sub>4</sub>, OH, and monoterpenes measurements, *Atmos. Chem. Phys.*, 3, 361–376, <https://doi.org/10.5194/ACP-3-361-2003>, 2003.
- Ciuraru, R., Kammer, J., Decuq, C., Vojkovic, M., Haider, K., Carpentier, Y., Lafouge, F., Berger, C., Bourdat-Deschamps, M., Ortega, I. K., Levavasseur, F., Houot, S., Loubet, B., Petitprez, D., and Focsa, C.: New particle formation from agricultural recycling of organic waste products, *npj Climate and Atmospheric Science*, 4, 1–10, <https://doi.org/10.1038/s41612-021-00160-3>, 2021.
- Dada, L., Paasonen, P., Nieminen, T., Mazon, S. B., Kontkanen, J., Peräkylä, O., Lehtipalo, K., Hussein, T., Petäjä, T., Kerminen, V. M., Bäck, J., and Kulmala, M.: Long-term analysis of clear-sky new particle formation events and nonevents in Hyytiälä, *Atmos. Chem. Phys.*, 17, 6227–6241, <https://doi.org/10.5194/ACP-17-6227-2017>, 2017.
- Dal Maso, M., Sogacheva, L., Aalto, P. P., Riipinen, I., Komppula, M., Tunved, P., Korhonen, L., Suur-Uski, V., Hirsikko, A., Kurtén, T., Kerminen, V.-M., Lihavainen, H., Viisanen, Y., Hansson, H.-C., and Kulmala, M.: Aerosol size distribution measurements at four Nordic field stations: identification, analysis and trajectory analysis of new particle formation bursts, *Tellus B*, 59, 350–361, <https://doi.org/10.1111/j.1600-0889.2007.00267.x>, 2007.
- Dawson, M. L., Varner, M. E., Perraud, V., Ezell, M. J., Gerber, R. B., and Finlayson-Pitts, B. J.: Simplified mechanism for new particle formation from methanesulfonic acid, amines, and water via experiments and ab initio calculations, *P. Natl. Acad. Sci. USA*, 109, 18719–18724, <https://doi.org/10.1073/pnas.1211878109>, 2012.
- Dunne, E. M., Gordon, H., Kürten, A., ao Almeida, J., Duplissy, J., Williamson, C., Ortega, I. K., Pringle, K. J., Adamov, A., Baltensperger, U., Barmet, P., Benduhn, F., Bianchi, F., Breitenlechner, M., Clarke, A., Curtius, J., Dommen, J., Donahue, N. M., Ehrhart,

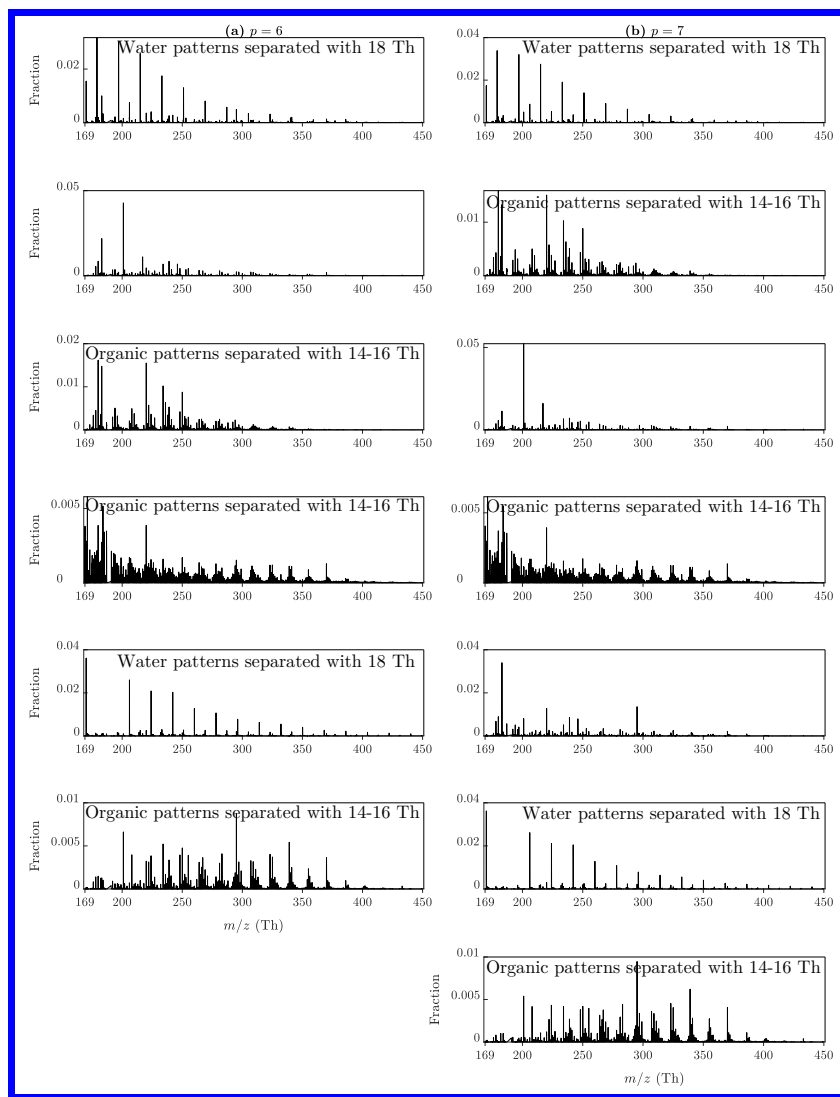
- S., Flagan, R. C., Franchin, A., Guida, R., Hakala, J., Hansel, A., Heinritzi, M., Jokinen, T., Kangasluoma, J., Kirkby, J., Kulmala, M., Kupc, A., Lawler, M. J., Lehtipalo, K., Makhmutov, V., Mann, G., Mathot, S., Merikanto, J., Miettinen, P., Nenes, A., Onnela, A., Rap, A., Reddington, C. L. S., Riccobono, F., Richards, N. A. D., Rissanen, M. P., Rondo, L., Sarnela, N., Schobesberger, S., Sengupta, K., Simon, M., Sipilä, M., Smith, J. N., Stozkhov, Y., Tomé, A., Tröstl, J., Wagner, P. E., Wimmer, D., Winkler, P. M., Worsnop, D. R., and Carslaw, K. S.: Global atmospheric particle formation from CERN CLOUD measurements, *Science*, 354, 1119–1124, <https://doi.org/10.1126/science.aaf2649>, 2016.
- Eisele, F. L. and Tanner, D. J.: Measurement of the gas phase concentration of  $\text{H}_2\text{SO}_4$  and methane sulfonic acid and estimates of  $\text{H}_2\text{SO}_4$  production and loss in the atmosphere, *J. Geophys. Res. Atmos.*, 98, 9001–9010, <https://doi.org/10.1029/93JD00031>, 1993.
- Fountoukis, C., Riipinen, I., Denier van der Gon, H. A. C., Charalampidis, P. E., Pilinis, C., Wiedensohler, A., O'Dowd, C., Putaud, J. P., Morerman, M., and Pandis, S. N.: Simulating ultrafine particle formation in Europe using a regional CTM: contribution of primary emissions versus secondary formation to aerosol number concentrations, *Atmos. Chem. Phys.*, 12, 8663–8677, <https://doi.org/10.5194/acp-12-8663-2012>, 2012.
- Gagné, S., Lehtipalo, K., Manninen, H. E., Nieminen, T., Schobesberger, S., Franchin, A., Yli-Juuti, T., Boulon, J., Sonntag, A., Mirme, S., Mirme, A., Hörrak, U., Petäjä, T., Asmi, E., and Kulmala, M.: Intercomparison of air ion spectrometers: an evaluation of results in varying conditions, *Atmos. Meas. Tech.*, 4, 805–822, <https://doi.org/10.5194/amt-4-805-2011>, 2011.
- Gordon, H., Kirkby, J., Baltensperger, U., Bianchi, F., Breitenlechner, M., Curtius, J., Dias, A., Dommen, J., Donahue, N. M., Dunne, E. M., Duplissy, J., Ehrhart, S., Flagan, R. C., Frege, C., Fuchs, C., Hansel, A., Hoyle, C. R., Kulmala, M., Kürten, A., Lehtipalo, K., Makhmutov, V., Molteni, U., Rissanen, M. P., Stozkhov, Y., Tröstl, J., Tsagkogeorgas, G., Wagner, R., Williamson, C., Wimmer, D., Winkler, P. M., Yan, C., and Carslaw, K. S.: Causes and importance of new particle formation in the present-day and preindustrial atmospheres, *J. Geophys. Res.-Atmos.*, 122, 8739–8760, <https://doi.org/10.1002/2017JD026844>, 2017.
- Gormley, P. G. and Kennedy, M.: Diffusion from a Stream Flowing through a Cylindrical Tube, *P. Roy. Irish Acad. A*, 52, 163–169, <https://doi.org/10.2307/20488498>, 1948.
- Guo, H., Wang, D. W., Cheung, K., Ling, Z. H., Chan, C. K., and Yao, X. H.: Observation of aerosol size distribution and new particle formation at a mountain site in subtropical Hong Kong, *Atmos. Chem. Phys.*, 12, 9923–9939, <https://doi.org/10.5194/ACP-12-9923-2012>, 2012.
- He, X. C., Tham, Y. J., Dada, L., Wang, M., Finkenzeller, H., Stolzenburg, D., Iyer, S., Simon, M., Kürten, A., Shen, J., Rörup, B., Rissanen, M., Schobesberger, S., Baalbaki, R., Wang, D. S., Koenig, T. K., Jokinen, T., Sarnela, N., Beck, L. J., Almeida, J., Amanatidis, S., Amorim, A., Ataei, F., Baccarini, A., Bertozzi, B., Bianchi, F., Brilke, S., Caudillo, L., Chen, D., Chiu, R., Chu, B., Dias, A., Ding, A., Dommen, J., Duplissy, J., Haddad, I. E., Carracedo, L. G., Granzin, M., Hansel, A., Heinritzi, M., Hofbauer, V., Junninen, H., Kangasluoma, J., Kemppainen, D., Kim, C., Kong, W., Krechmer, J. E., Kvashin, A., Laitinen, T., Lamkaddam, H., Lee, C. P., Lehtipalo, K., Leiminger, M., Li, Z., Makhmutov, V., Manninen, H. E., Marie, G., Marten, R., Mathot, S., Mauldin, R. L., Mentler, B., Möhler, O., Müller, T., Nie, W., Onnela, A., Petäjä, T., Pfeifer, J., Philippov, M., Ranjithkumar, A., Saiz-Lopez, A., Salma, I., Scholz, W., Schuchmann, S., Schulze, B., Steiner, G., Stozkhov, Y., Tauber, C., Tomé, A., Thakur, R. C., Väisänen, O., Vazquez-Pufleau, M., Wagner, A. C., Wang, Y., Weber, S. K., Winkler, P. M., Wu, Y., Xiao, M., Yan, C., Ye, Q., Ylisirniö, A., Zauner-Wieczorek, M., Zha, Q., Zhou, P., Flagan, R. C., Curtius, J., Baltensperger, U., Kulmala, M., Kerminen, V. M., Kurtén, T., Donahue, N. M., Volkamer, R., Kirkby, J., Worsnop, D. R., and Sipilä, M.: Role of iodine oxoacids in atmospheric aerosol nucleation, *Science*, 371, <https://doi.org/10.1126/science.abe0298>, 2021.

- Heimsch, L., Lohila, A., Tuovinen, J.-P., Vekuri, H., Heinonsalo, J., Nevalainen, O., Korkiakoski, M., Liski, J., Laurila, T., and Kulmala, L.: Carbon dioxide fluxes and carbon balance of an agricultural grassland in southern Finland, *Biogeosciences*, 18, 3467–3483, <https://doi.org/10.5194/bg-18-3467-2021>, 2021.
- Hytinen, N., Kupiainen-Määttä, O., Rissanen, M. P., Muuronen, M., Ehn, M., and Kurtén, T.: Modeling the Charging of Highly Oxidized Cyclohexene Ozonolysis Products Using Nitrate-Based Chemical Ionization, *J. Phys. Chem. A*, 119, 6339–6345, <https://doi.org/10.1021/acs.jpca.5b01818>, 2015.
- Iinuma, Y., Böge, O., Gräfe, R., and Herrmann, H.: Methyl-Nitrocatechols: Atmospheric Tracer Compounds for Biomass Burning Secondary Organic Aerosols, *Environ. Sci. Technol.*, 44, 8453–8459, <https://doi.org/10.1021/es102938a>, 2010.
- 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, <https://doi.org/10.5194/acp-12-4117-2012>, 2012.
- Jun, Y.-S., Jeong, C.-H., Sabaliauskas, K., Richard Leaitch, W., and Evans, G. J.: A year-long comparison of particle formation events at paired urban and rural locations, *Atmos. Pollut. Res.*, 5, 447–454, <https://doi.org/10.5094/APR.2014.052>, 2014.
- Junninen, H., Ehn, M., Petäjä, T., Luosujärvi, L., Kotiaho, T., Kostianinen, 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, <https://doi.org/10.5194/amt-3-1039-2010>, 2010.
- Junninen, H., Ahonen, L., Bianchi, F., Quéléver, L., Schallhart, S., Dada, L., Manninen, H., Leino, K., Lampilahti, J., Mazon, S., Rantala, P., Rätty, M., Kontkanen, J., Negri, S., Aliaga, D., Garmash, O., Alekseychik, P., Lipp, H., Tamme, K., Levula, J., Sipilä, M., Ehn, M., Worsnop, D., Zilitinkevich, S., Mammarella, I., Rinne, J., Vesala, T., Petäjä, T., Kerminen, V.-M., and Kulmala, M.: Previously unrecognized climatic effect of wetlands through new particle formation, manuscript accepted for publication, 2022.
- Kangasluoma, J., Cai, R., Jiang, J., Deng, C., Stolzenburg, D., Ahonen, L. R., Chan, T., Fu, Y., Kim, C., Laurila, T. M., Zhou, Y., Dada, L., Sulo, J., Flagan, R. C., Kulmala, M., Petäjä, T., and Lehtipalo, K.: Overview of measurements and current instrumentation for 1–10 nm aerosol particle number size distributions, *J. Aerosol Sci.*, 148, 105 584, <https://doi.org/10.1016/j.jaerosci.2020.105584>, 2020.
- Kerminen, V.-M., Chen, X., Vakkari, V., Petäjä, T., Kulmala, M., and Bianchi, F.: Atmospheric new particle formation and growth: review of field observations, *Environ. Res. Lett.*, 13, 103 003, <https://doi.org/10.1088/1748-9326/aadf3c>, 2018.
- Kulmala, M., Lehtinen, K., and Laaksonen, A.: Cluster activation theory as an explanation of the linear dependence between formation rate of 3 nm particles and sulphuric acid concentration, *Atmos. Chem. Phys.*, 6, 787–793, <https://doi.org/10.5194/acp-6-787-2006>, 2006.
- Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P., Junninen, H., Paasonen, P., Riipinen, I., Lehtinen, K. E. J., Laaksonen, A., and Kerminen, V.-M.: Measurement of the nucleation of atmospheric aerosol particles, *Nature Protocols*, 7, 1651 – 1667, 2012.
- Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H., Nieminen, T., Petäjä, T., Sipilä, M., Schobesberger, S., Rantala, P., Franchin, A., Jokinen, T., Järvinen, E., Äijälä, M., Kangasluoma, J., Hakala, J., Aalto, P., Paasonen, P., Mikkilä, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., Ruuskanen, T., Mauldin III, R., Duplissy, J., Vehkamäki, H., Bäck, J., Kortelainen, A., Riipinen, I., Kurtén, T., Johnston, M., Smith, J., Ehn, M., Mentel, T., Lehtinen, K., Laaksonen, A., Kerminen, V.-M., and Worsnop, D.: Direct observations of atmospheric aerosol nucleation, *Science*, 339, 943–946, <https://doi.org/10.1126/science.1227385>, 2013.
- Kulmala, M., Petäjä, T., Ehn, M., Thornton, J., Sipilä, M., Worsnop, D., and Kerminen, V.-M.: Chemistry of Atmospheric Nucleation: On the Recent Advances on Precursor Characterization and Atmospheric Cluster Composition in Connection with Atmospheric New Particle Formation, *Annu. Rev. Phys. Chem.*, 65, 21–37, <https://doi.org/10.1146/annurev-physchem-040412-110014>, 2014.

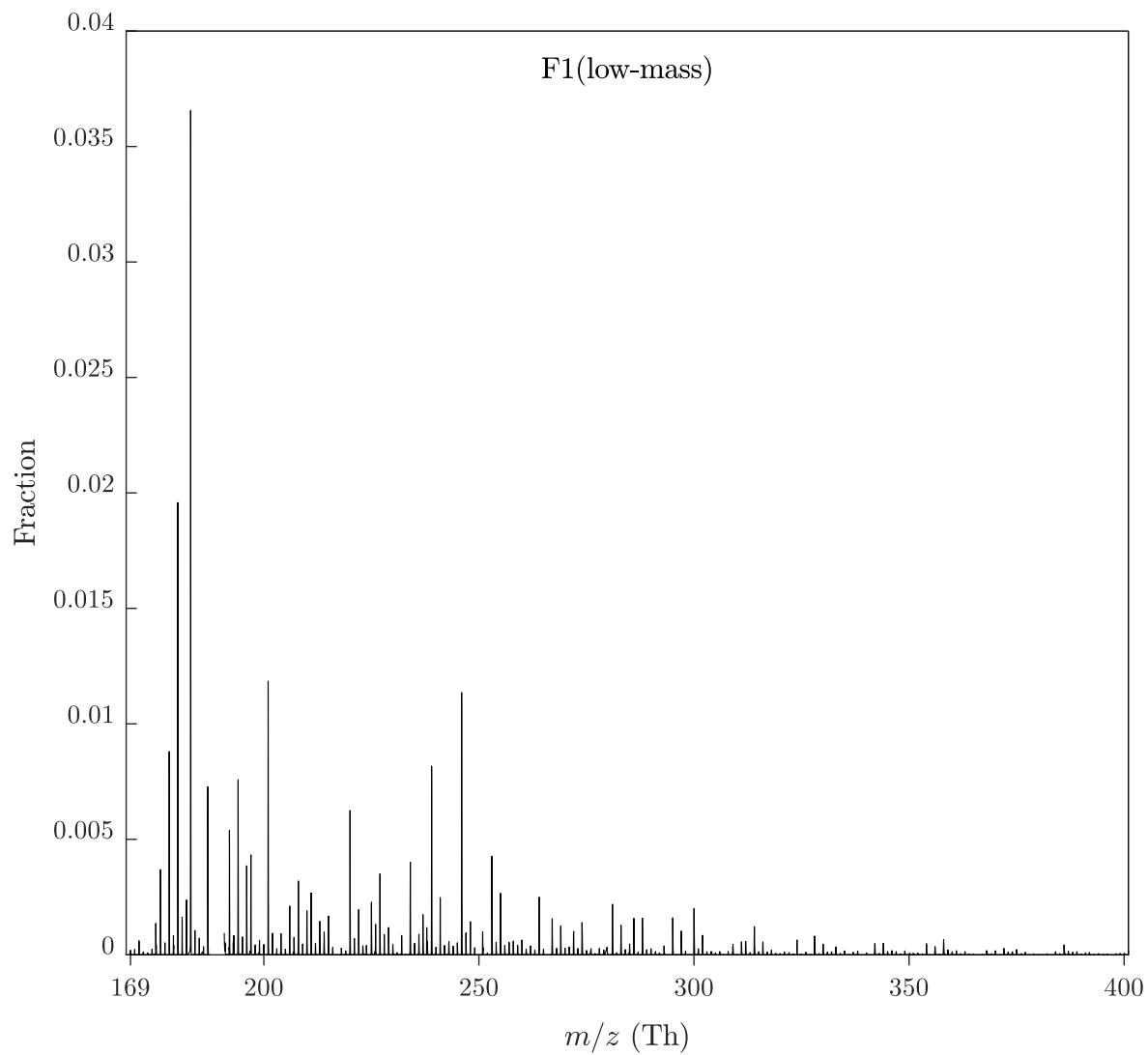
- Kulmala, M., Luoma, K., Virkkula, A., Petäjä, T., Paasonen, P., Kerminen, V.-M., Nie, W., Qi, X., Shen, Y., Chi, X., and Ding, A.: On the mode-segregated aerosol particle number concentration load: contributions of primary and secondary particles in Hyytiälä and Nanjing, *Boreal Environ. Res.*, 21, 319–331, <http://www.borenv.net/BER/archive/ber213-4.htm#319>, 2016.
- Kürten, A., Rondo, L., Ehrhart, S., and Curtius, J.: Calibration of a Chemical Ionization Mass Spectrometer for the Measurement of Gaseous Sulfuric Acid, *J. Phys. Chem. A*, 116, 6375–6386, <https://doi.org/10.1021/jp212123n>, 2012.
- Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J., and Carslaw, K. S.: Impact of nucleation on global CCN, *Atmos. Chem. Phys.*, 9, 8601–8616, <https://doi.org/10.5194/acp-9-8601-2009>, 2009.
- Mirme, S. and Mirme, A.: The mathematical principles and design of the NAIS – a spectrometer for the measurement of cluster ion and nanometer aerosol size distributions, *Atmos. Meas. Tech.*, 6, 1061–1071, <https://doi.org/10.5194/amt-6-1061-2013>, 2013.
- O'Dowd, C. D., Jimenez, J. L., Bahreini, R., Flagan, R. C., Seinfeld, J. H., Hämeri, K., Pirjola, L., Kulmala, M., and Hoffmann, T.: Marine aerosol formation from biogenic iodine emissions, *Nature*, 417, 632–636, <https://doi.org/10.1038/nature00775>, 2002.
- Olin, M.: Atmospheric data measured in Quidja, Finland, from Apr to Jun 2019, Zenodo [data set], <https://doi.org/10.5281/zenodo.6394454>, 2022.
- Olin, M., Kuuluvainen, H., Aurela, M., Kalliokoski, J., Kuittinen, N., Isotalo, M., Timonen, H. J., Niemi, J. V., Rönkkö, T., and Dal Maso, M.: Traffic-originated nanocluster emission exceeds H<sub>2</sub>SO<sub>4</sub>-driven photochemical new particle formation in an urban area, *Atmos. Chem. Phys.*, 20, 1–13, <https://doi.org/10.5194/acp-20-1-2020>, 2020.
- Olin, M., Patoulias, D., Kuuluvainen, H., Niemi, J. V., Rönkkö, T., Pandis, S. N., Riipinen, I., and Dal Maso, M.: Contribution of traffic-originated nanoparticle emissions to regional and local aerosol levels, *Atmos. Chem. Phys.*, 22, 1131–1148, <https://doi.org/10.5194/acp-22-1131-2022>, 2022.
- Paasonen, P., Asmi, A., Petäjä, T., Kajos, M. K., Äijälä, M., Junninen, H., Holst, T., Abbatt, J. P., Arneth, A., Birmili, W., Denier Van Der Gon, H., Hamed, A., Hoffer, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Plass-Dülmer, C., Pryor, S. C., Räisänen, P., Swietlicki, E., Wiedensohler, A., Worsnop, D. R., Kerminen, V. M., and Kulmala, M.: Warming-induced increase in aerosol number concentration likely to moderate climate change, *Nature Geoscience*, 6, 438–442, <https://doi.org/10.1038/ngeo1800>, 2013.
- Paatero, P., Hopke, P. K., Song, X.-H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, *Chemometr. Intell. Lab.*, 60, 253–264, [https://doi.org/10.1016/S0169-7439\(01\)00200-3](https://doi.org/10.1016/S0169-7439(01)00200-3), 2002.
- Pierce, J. R., Westervelt, D. M., Atwood, S. A., Barnes, E. A., and Leaitch, W. R.: New-particle formation, growth and climate-relevant particle production in Egbert, Canada: analysis from 1 year of size-distribution observations, *Atmos. Chem. Phys.*, 14, 8647–8663, <https://doi.org/10.5194/acp-14-8647-2014>, 2014.
- Pikridas, M., Riipinen, I., Hildebrandt, L., Kostenidou, E., Manninen, H., Mihalopoulos, N., Kalivitis, N., Burkhardt, J. F., Stohl, A., Kulmala, M., and Pandis, S. N.: New particle formation at a remote site in the eastern Mediterranean, *J. Geophys. Res.-Atmos.*, 117, 12 205, <https://doi.org/10.1029/2012JD017570>, 2012.
- Posner, L. N. and Pandis, S. N.: Sources of ultrafine particles in the Eastern United States, *Atmos. Environ.*, 111, 103–112, <https://doi.org/https://doi.org/10.1016/j.atmosenv.2015.03.033>, 2015.
- Qi, X. M., Ding, A. J., Nie, W., Petäjä, T., Kerminen, V.-M., Herrmann, E., Xie, Y. N., Zheng, L. F., Manninen, H., Aalto, P., Sun, J. N., Xu, Z. N., Chi, X. G., Huang, X., Boy, M., Virkkula, A., Yang, X.-Q., Fu, C. B., and Kulmala, M.: Aerosol size distribution and new particle formation in the western Yangtze River Delta of China: 2 years of measurements at the SORPES station, *Atmos. Chem. Phys.*, 15, 12 445–12 464, <https://doi.org/10.5194/acp-15-12445-2015>, 2015.

- Salma, I., Németh, Z., Kerminen, V.-M., Aalto, P., Nieminen, T., Weidinger, T., Molnár, A., Imre, K., and Kulmala, M.: Regional effect on urban atmospheric nucleation, *Atmos. Chem. Phys.*, 16, 8715–8728, <https://doi.org/10.5194/acp-16-8715-2016>, 2016.
- Simon, M., Dada, L., Heinritzi, M., Scholz, W., Stolzenburg, D., Fischer, L., Wagner, A. C., Kürten, A., Rörup, B., He, X.-C., Almeida, J., Baalbaki, R., Baccarini, A., Bauer, P. S., Beck, L., Bergen, A., Bianchi, F., Bräkling, S., Brilke, S., Caudillo, L., Chen, D., Chu, B., Dias, A., Draper, D. C., Duplissy, J., El-Haddad, I., Finkenzeller, H., Frege, C., Gonzalez-Carracedo, L., Gordon, H., Granzin, M., Hakala, J., Hofbauer, V., Hoyle, C. R., Kim, C., Kong, W., Lamkaddam, H., Lee, C. P., Lehtipalo, K., Leiminger, M., Mai, H., Manninen, H. E., Marie, G., Marten, R., Mentler, B., Molteni, U., Nichman, L., Nie, W., Ojdanic, A., Onnela, A., Partoll, E., Petäjä, T., Pfeifer, J., Philippov, M., Quéléver, L. L. J., Ranjithkumar, A., Rissanen, M. P., Schallhart, S., Schobesberger, S., Schuchmann, S., Shen, J., Sipilä, M., Steiner, G., Stozhkov, Y., Tauber, C., Tham, Y. J., Tomé, A. R., Vazquez-Pufleau, M., Vogel, A. L., Wagner, R., Wang, M., Wang, D. S., Wang, Y., Weber, S. K., Wu, Y., Xiao, M., Yan, C., Ye, P., Ye, Q., Zauner-Wieczorek, M., Zhou, X., Baltensperger, U., Dommen, J., Flagan, R. C., Hansel, A., Kulmala, M., Volkamer, R., Winkler, P. M., Worsnop, D. R., Donahue, N. M., Kirkby, J., and Curtius, J.: Molecular understanding of new-particle formation from  $\alpha$ -pinene between  $-50$  and  $+25$  °C, *Atmos. Chem. Phys.*, 20, 9183–9207, <https://doi.org/10.5194/acp-20-9183-2020>, 2020.
- Sipilä, M., Sarnela, N., Jokinen, T., Henschel, H., Junninen, H., Kontkanen, J., Richters, S., Kangasluoma, J., Franchin, A., Peräkylä, O., Rissanen, M. P., Ehn, M., Vehkamäki, H., Kurten, T., Berndt, T., Petäjä, T., Worsnop, D., Ceburnis, D., Kerminen, V. M., Kulmala, M., and O'Dowd, C.: Molecular-scale evidence of aerosol particle formation via sequential addition of HIO<sub>3</sub>, *Nature*, 537, 532–534, <https://doi.org/10.1038/nature19314>, 2016.
- Wang, M., He, X. C., Finkenzeller, H., Iyer, S., Chen, D., Shen, J., Simon, M., Hofbauer, V., Kirkby, J., Curtius, J., Maier, N., Kurtén, T., Worsnop, D. R., Kulmala, M., Rissanen, M., Volkamer, R., Tham, Y. J., Donahue, N. M., and Sipilä, M.: Measurement of iodine species and sulfuric acid using bromide chemical ionization mass spectrometers, *Atmos. Meas. Tech.*, 14, 4187–4202, <https://doi.org/10.5194/amt-14-4187-2021>, 2021.
- Wang, Z. B., Hu, M., Yue, D. L., Zheng, J., Zhang, R. Y., Wiedensohler, A., Wu, Z. J., Nieminen, T., and Boy, M.: Evaluation on the role of sulfuric acid in the mechanisms of new particle formation for Beijing case, *Atmos. Chem. Phys.*, 11, 12 663–12 671, <https://doi.org/10.5194/ACP-11-12663-2011>, 2011.
- Woo, K. S., Chen, D. R., Pui, D. Y. H., and McMurry, P. H.: Measurement of Atlanta Aerosol Size Distributions: Observations of Ultrafine Particle Events, *Aerosol Sci. Tech.*, 34, 75–87, <https://doi.org/10.1080/02786820120056>, 2001.
- Wu, Z., Hu, M., Liu, S., Wehner, B., Bauer, S., Maßling, A., Wiedensohler, A., Petäjä, T., Dal Maso, M., and Kulmala, M.: New particle formation in Beijing, China: Statistical analysis of a 1-year data set, *J. Geophys. Res.-Atmos.*, 112, 9209, <https://doi.org/10.1029/2006JD007406>, 2007.
- Yan, C., Nie, W., Äijälä, M., Rissanen, M. P., Canagaratna, M. R., Massoli, P., Junninen, H., Jokinen, T., Sarnela, N., Häme, S. A. K., Schobesberger, S., Canonaco, F., Yao, L., Prévôt, A. S. H., Petäjä, T., Kulmala, M., Sipilä, M., Worsnop, D. R., and Ehn, M.: Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization, *Atmos. Chem. Phys.*, 16, 12 715–12 731, <https://doi.org/10.5194/acp-16-12715-2016>, 2016.
- Yao, L., Garmash, O., Bianchi, F., Zheng, J., Yan, C., Kontkanen, J., Junninen, H., Mazon, S. B., Ehn, M., Paasonen, P., Sipilä, M., Wang, M., Wang, X., Xiao, S., Chen, H., Lu, Y., Zhang, B., Wang, D., Fu, Q., Geng, F., Li, L., Wang, H., Qiao, L., Yang, X., Chen, J., Kerminen, V.-M., Petäjä, T., Worsnop, D. R., Kulmala, M., and Wang, L.: Atmospheric new particle formation from sulfuric acid and amines in a Chinese megacity, *Science*, 361, 278–281, <https://doi.org/10.1126/science.aao4839>, 2018.

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, *Atmos. Meas. Tech.*, 12, 3761–3776, <https://doi.org/10.5194/amt-12-3761-2019>, 2019.

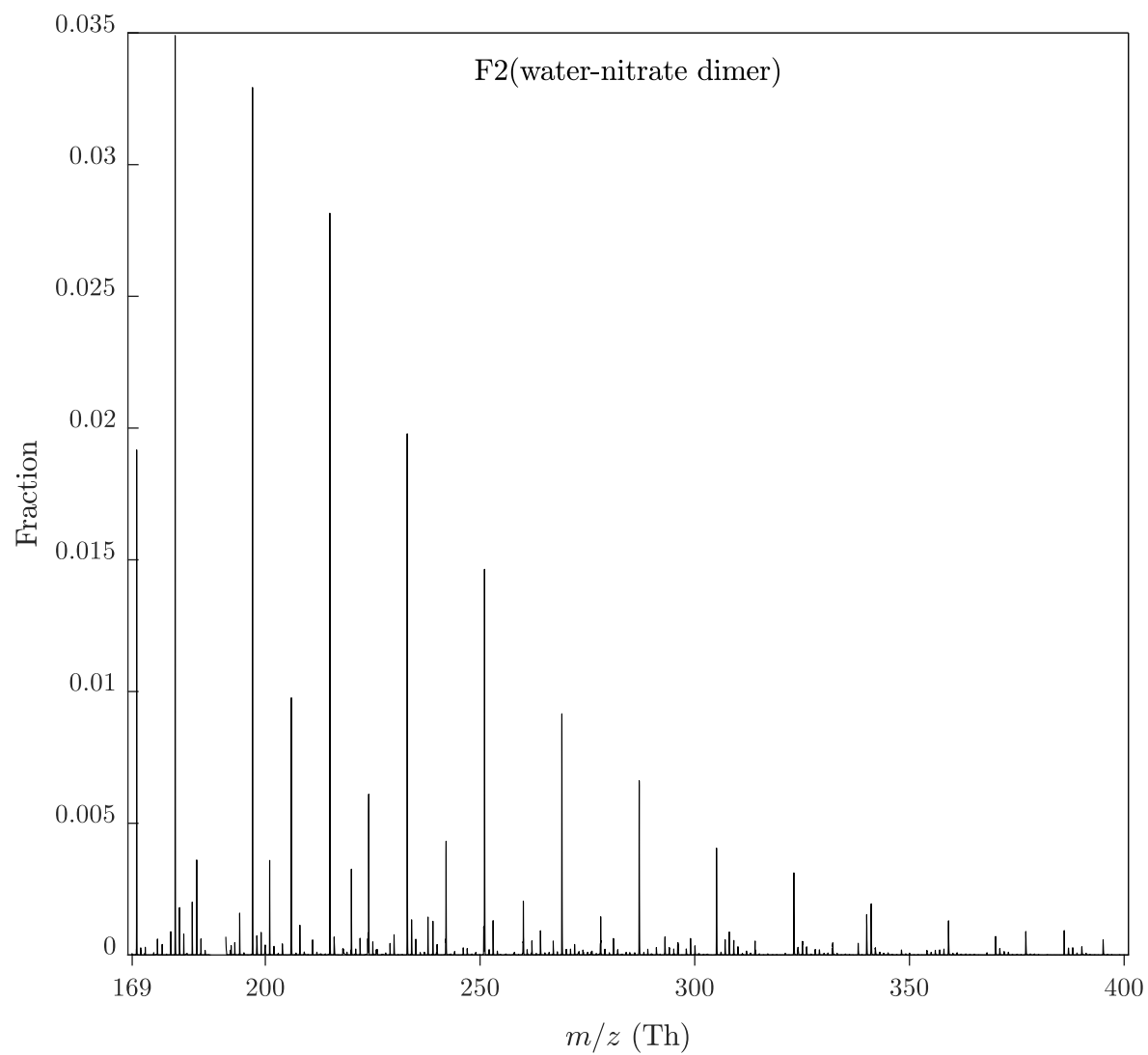


**Figure S1.** Normalized spectra of all binPMF factors if (a)  $p = 6$  or (b)  $p = 7$  were used. Both of these factor sets have three factor profiles having organic patterns, whereas  $p = 8$  results in four factor profiles having organic patterns and is further analyzed in this study.

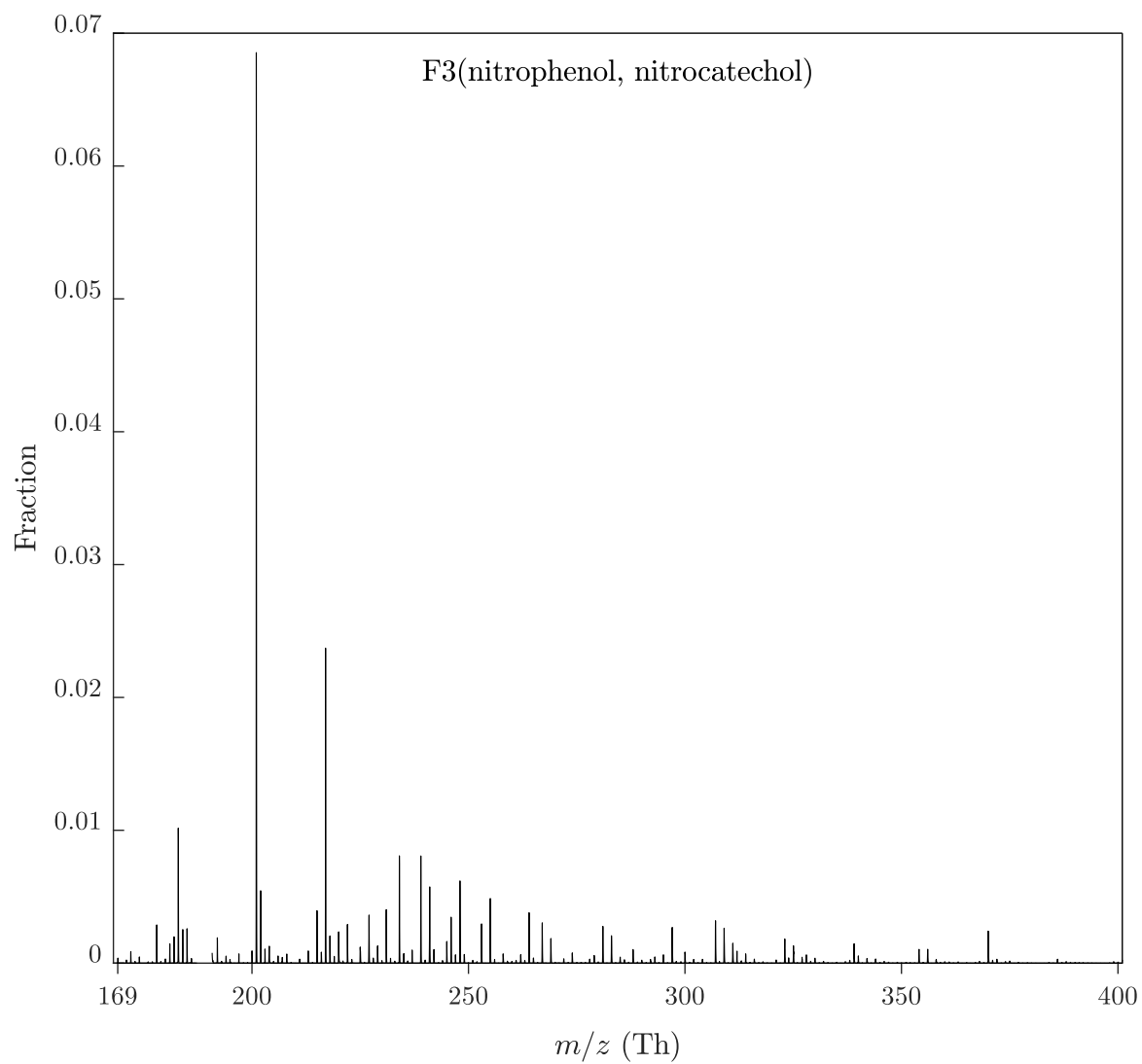


**Figure S2.** Normalized spectrum of the binPMF factor F1(low-mass) up to 400 Th.

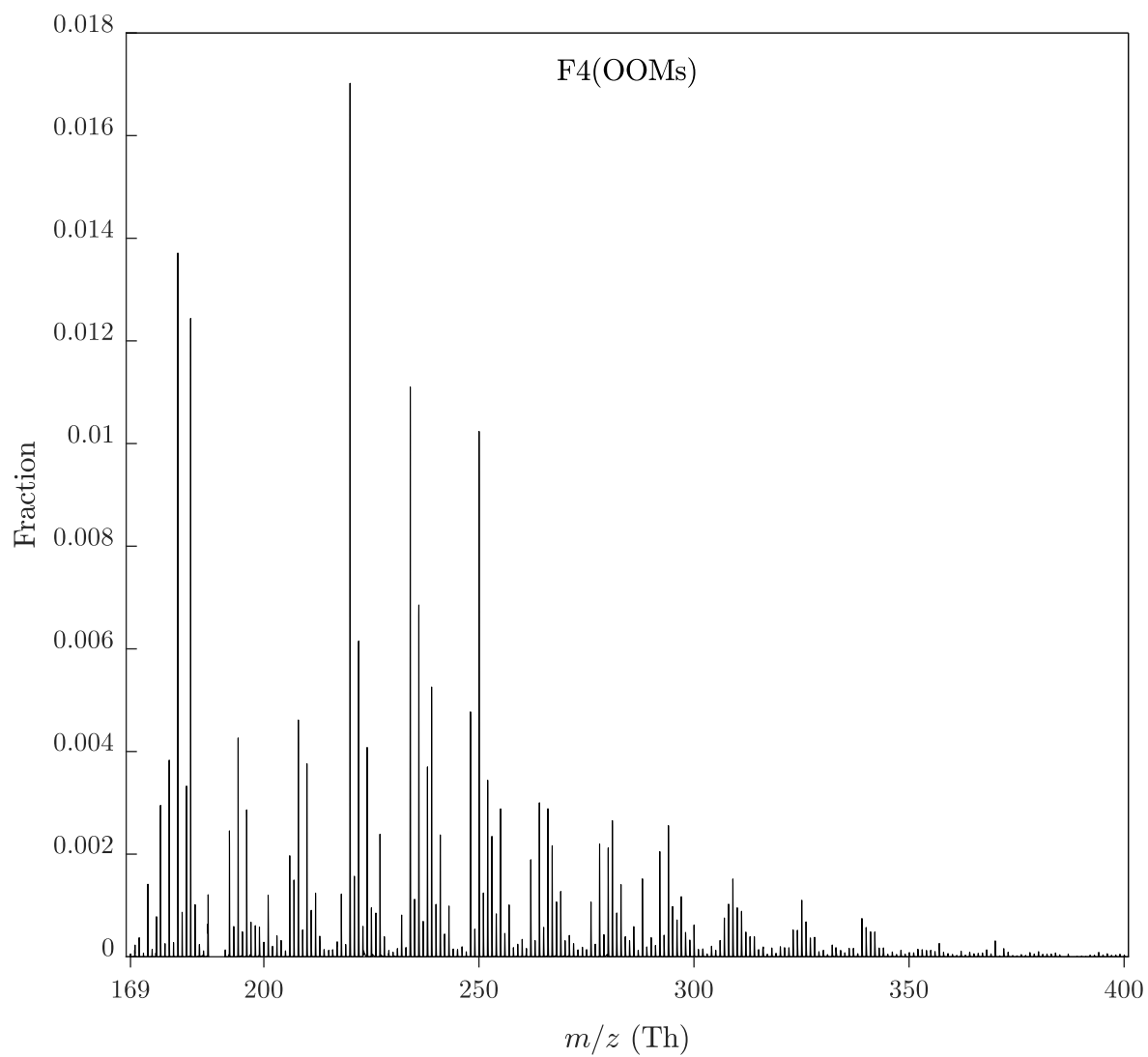




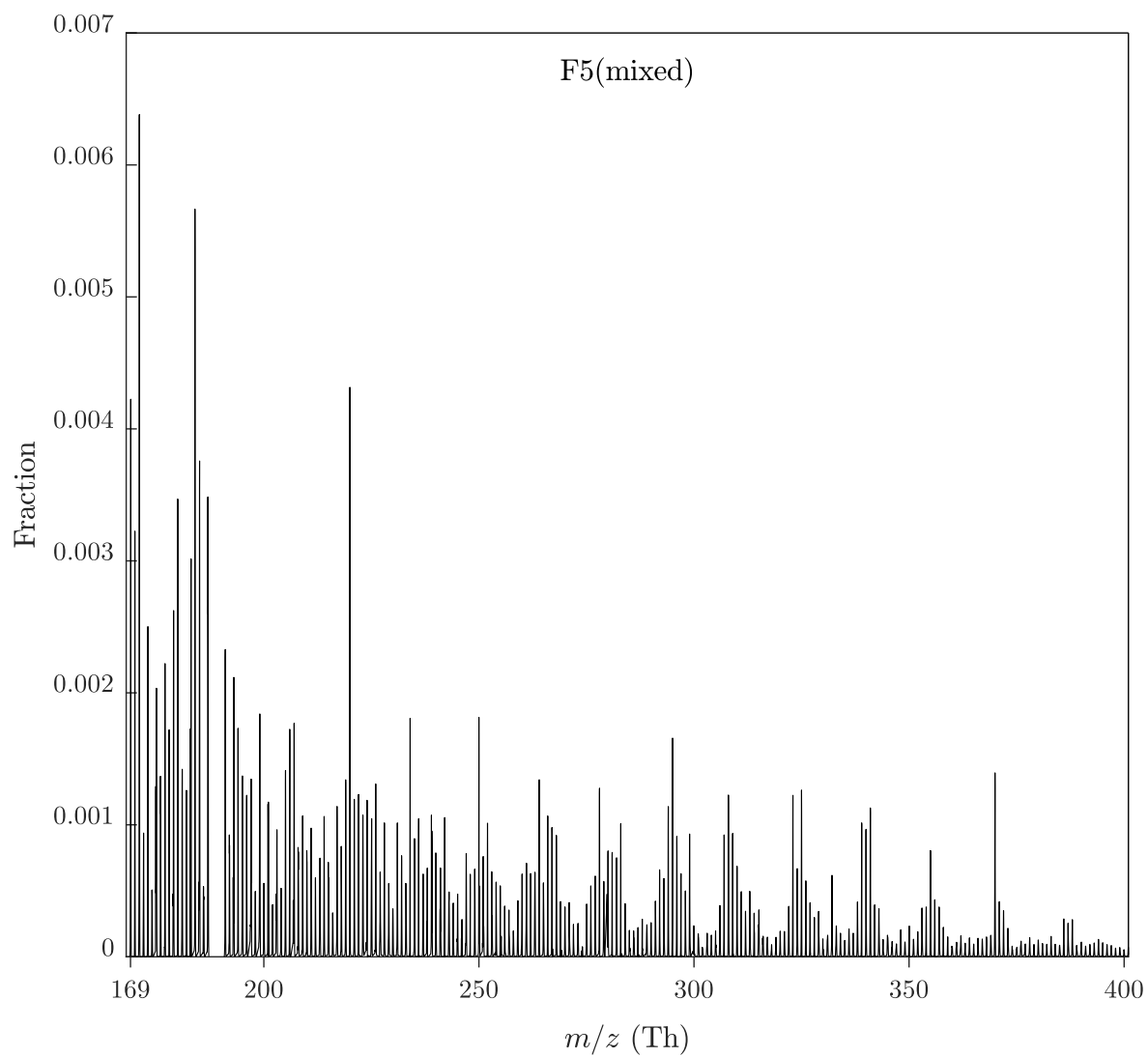
**Figure S3.** Normalized spectrum of the binPMF factor F2(water-nitrate dimer) up to 400 Th.



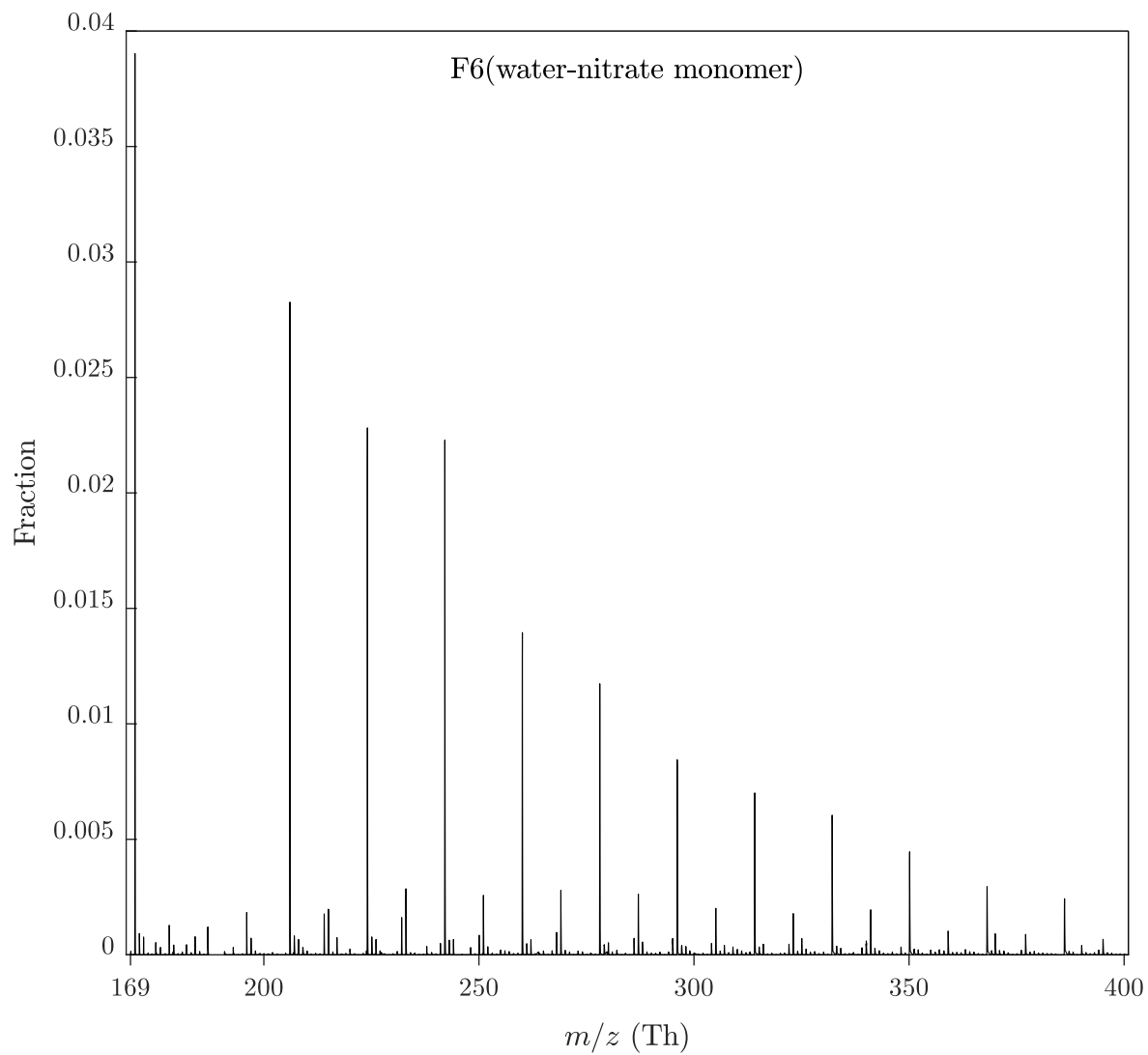
**Figure S4.** Normalized spectrum of the binPMF factor F3(nitrophenol, nitrocatechol) up to 400 Th.



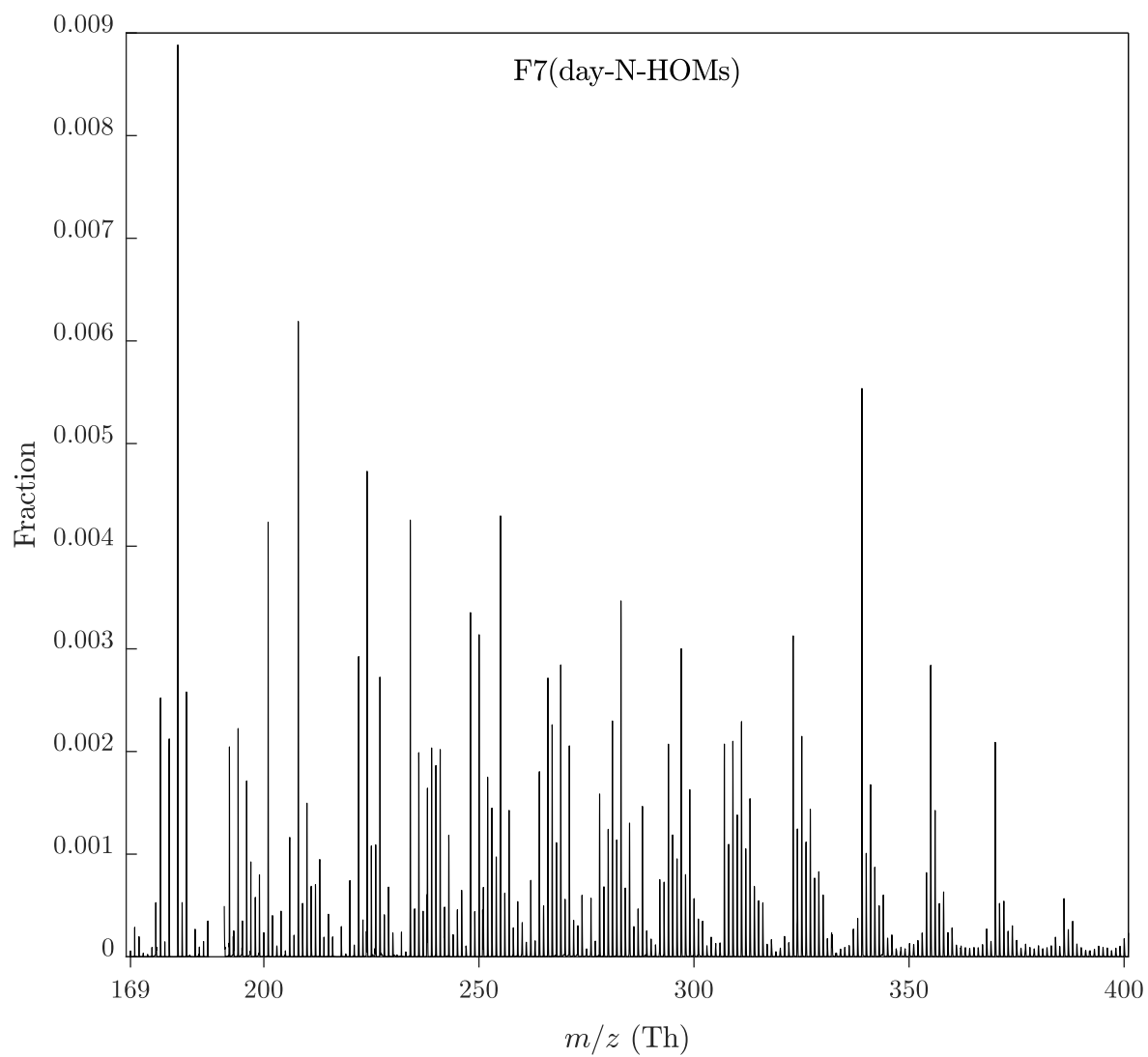
**Figure S5.** Normalized spectrum of the binPMF factor F4(OOMs) up to 400 Th.



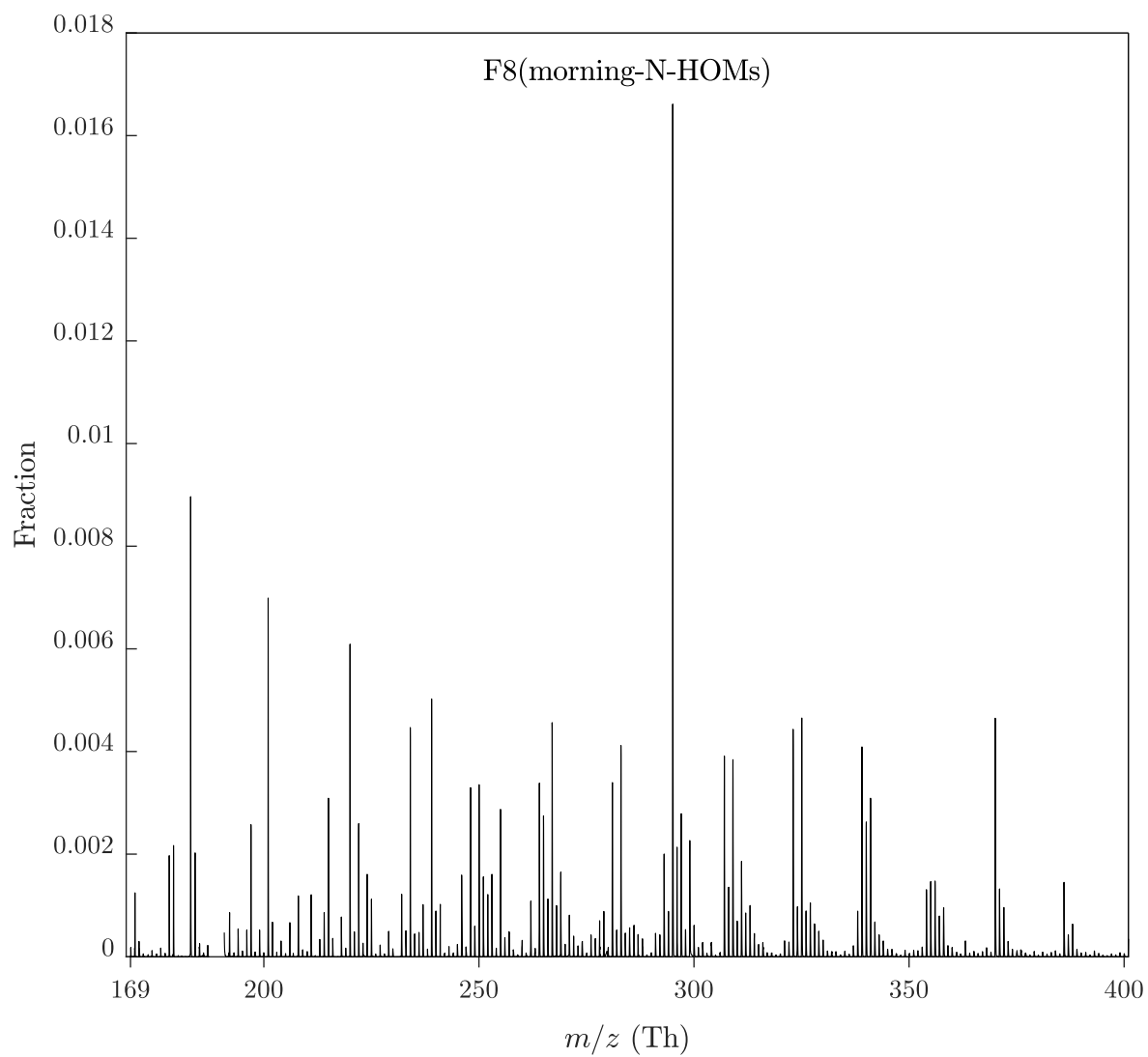
**Figure S6.** Normalized spectrum of the binPMF factor F5(mized) up to 400 Th.



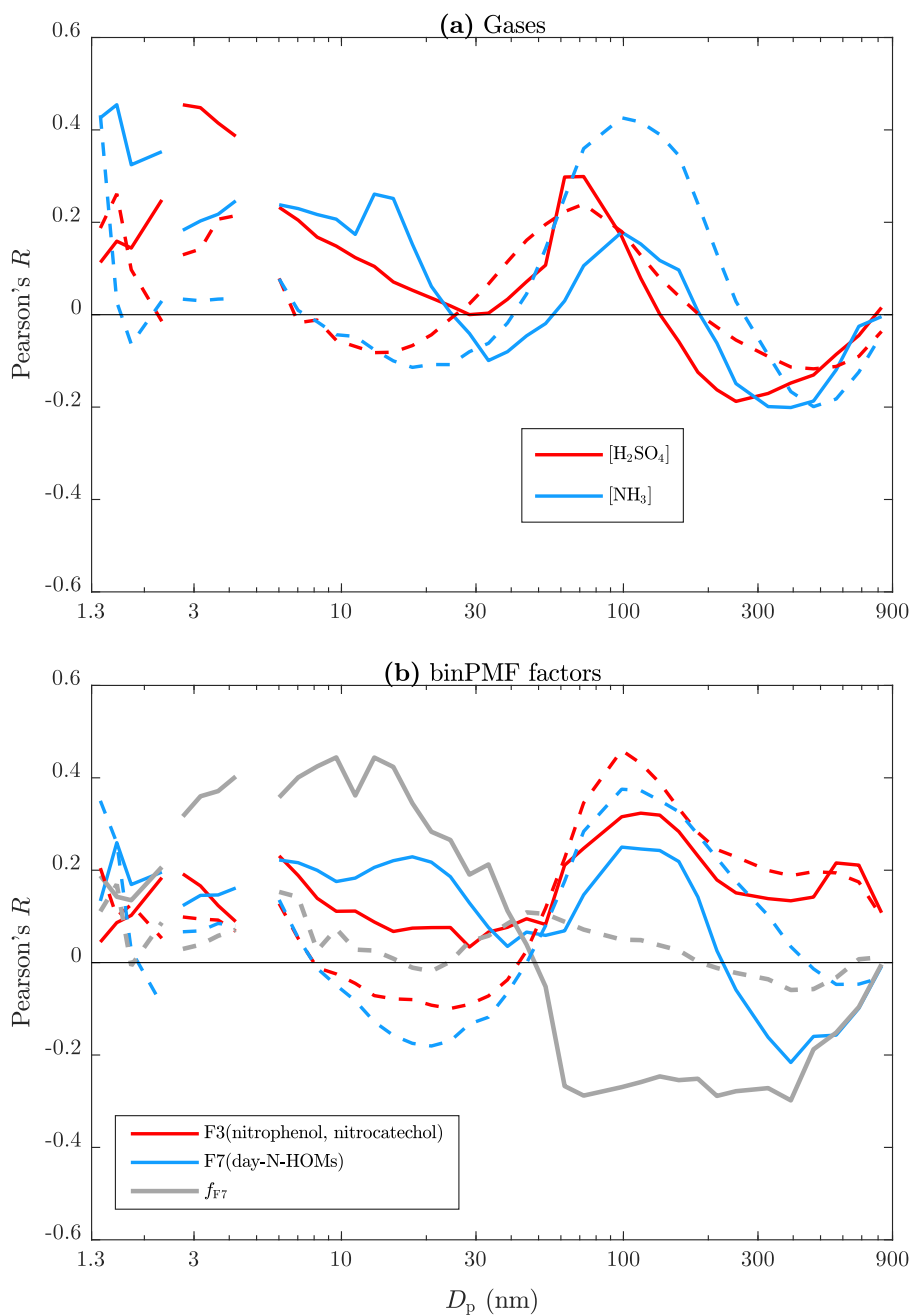
**Figure S7.** Normalized spectrum of the binPMF factor F6(water-nitrate monomer) up to 400 Th.



**Figure S8.** Normalized spectrum of the binPMF factor F7(day-N-HOMs) up to 400 Th.



**Figure S9.** Normalized spectrum of the binPMF factor F8(morning-N-HOMs) up to 400 Th.



**Figure S10.** Pearson's correlation coefficient ( $R$ ) between all particle size bins and **(a)** the concentrations of sulfuric acid ( $[H_2SO_4]$ ) and ammonia ( $[NH_3]$ ) and **(b)** binPMF factors 3 and 7, together with the variable  $f_{F7}$ . Data below 3 nm are from a PSM, over 5 nm are from the DMPS, and the remaining part between them are from the NAIS. Solid lines denote data from the days with NPF events and dashed lines from the days with no NPF event. Due to a relative low amount of the data in these plots, the lines are shown regardless of their statistic significance.



**Table S1.** Classifications of the measurement days in terms of new particle formation events, according to the behavior of the measured particle size distributions.

Day	Classification	Day	Classification	Day	Classification
		May 1	unclear	Jun 1	no event
		May 2	no event	Jun 2	no event
Apr 3	no event	May 3	event	Jun 3	no event
Apr 4	no event	May 4	unclear	Jun 4	no event
Apr 5	no event	May 5	unsure	Jun 5	no event
Apr 6	no event	May 6	no event	Jun 6	no event
Apr 7	no event	May 7	no event	Jun 7	no event
Apr 8	no event	May 8	event	Jun 8	no event
Apr 9	no event	May 9	event	Jun 9	unclear
Apr 10	no event	May 10	event	Jun 10	event
Apr 11	event	May 11	no event	Jun 11	event
Apr 12	event	May 12	no event	Jun 12	unclear
Apr 13	unsure	May 13	event	Jun 13	unclear
Apr 14	unsure	May 14	event	Jun 14	no event
Apr 15	event	May 15	event	Jun 15	no event
Apr 16	event	May 16	event	Jun 16	no event
Apr 17	event	May 17	event	Jun 17	unclear
Apr 18	event	May 18	event	Jun 18	unclear
Apr 19	unclear	May 19	no event	Jun 19	unclear
Apr 20	unclear	May 20	no event	Jun 20	unclear
Apr 21	no event	May 21	no event	Jun 21	unclear
Apr 22	no event	May 22	unclear	Jun 22	no event
Apr 23	no event	May 23	no event	Jun 23	unclear
Apr 24	no event	May 24	unclear	Jun 24	event
Apr 25	unclear	May 25	no event	Jun 25	event
Apr 26	no event	May 26	unclear		
Apr 27	no event	May 27	event		
Apr 28	event	May 28	no event		
Apr 29	event	May 29	unclear		
Apr 30	event	May 30	unclear		
		May 31	event		

**Table S2.** List of the 10 tallest peaks in the binPMF factors. The values denote  $m/z$  (in Th) of the peaks in ascending order. The tallest peaks are marked in bold.

F1 (low-mass)	F2 (water-nitrate dimer)	F3 (nitrophenol, nitrocatechol)	F4 (OOMs)	F5 (mixed)	F6 (water-nitrate monomer)	F7 (day-N-HOMs)	F8 (morning-N-HOMs)
178.0124	170.0531	182.9971	180.0173	169.0690	<b>170.0531</b>	<b>180.0156</b>	182.9943
180.0171	<b>179.0170</b>	<b>201.0171</b>	182.9899	170.0254	206.0747	201.0157	201.0159
<b>182.9960</b>	197.0278	202.0200	208.0300	<b>171.0606</b>	224.0849	208.0159	220.0477
186.9861	206.0718	217.0109	<b>220.0465</b>	179.0216	242.0948	224.0156	220.0477
192.0182	215.0387	234.0615	222.0456	179.9922	260.1054	234.0614	239.0192
194.0262	224.0849	239.0162	234.0535	183.0787	278.1149	248.0509	267.0465
201.0153	233.0494	241.0585	236.0449	183.9966	296.1236	255.0276	<b>295.0749</b>
220.0494	251.0581	248.0504	239.0190	185.0550	314.1326	283.0417	323.0654
239.0191	269.0695	255.0434	248.0564	187.0077	332.1426	323.0650	325.0703
246.0000	287.0777	264.0093	250.0571	220.0135	350.1475	339.0585	370.0555

**Table S3.** Pearson's correlation coefficients ( $R$ ) between the most important binPMF factors and UMR data. The bold values denote the most suitable UMR tracers for the binPMF factors. A perfectly selective UMR tracer for a factor would have  $R$  equal to the ones in footnotes (and  $R = 1$  between the factor and the tracer).  $R$  for the other factors not shown here for these selected UMR tracers are less than 0.40.

UMR	220	236	285	271	339	295	265
F4(OOMs)	<b>0.98</b>	<b>0.97</b>	0.65 <sup>a</sup>	0.64 <sup>a</sup>	0.59 <sup>a</sup>	0.43 <sup>b</sup>	0.62 <sup>b</sup>
F7(day-N-HOMs)	0.54 <sup>a</sup>	0.63 <sup>a</sup>	<b>0.91</b>	<b>0.89</b>	<b>0.87</b>	0.37 <sup>c</sup>	0.54 <sup>c</sup>
F8(morning-N-HOMs)	0.47 <sup>b</sup>	0.36 <sup>b</sup>	0.36 <sup>c</sup>	0.42 <sup>c</sup>	0.52 <sup>c</sup>	<b>0.95</b>	<b>0.82</b>

<sup>a</sup>  $R = 0.47$  between F4(OOMs) and F7(day-N-HOMs)

<sup>b</sup>  $R = 0.30$  between F4(OOMs) and F8(morning-N-HOMs)

<sup>c</sup>  $R = 0.25$  between F7(day-N-HOMs) and F8(morning-N-HOMs)