Supplement of Atmos. Chem. Phys., 23, 5567–5585, 2023
https://doi.org/10.5194/acp-23-5567-2023-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Chemical identification of new particle formation and growth precursors through positive matrix factorization of ambient ion measurements

Daniel John Katz et al.

Correspondence to: Eleanor C. Browne (eleanor.browne@colorado.edu)

Chemical identification of new particle formation and growth precursors through positive matrix factorization of ambient ion measurements

S1 Pinhole clog

We carefully considered how the clog of the pinhole on September 13th might influence our conclusions and failed to find evidence of artifacts. We expect that the main effect of a clog would be a reduction in total ions measured by the APi-ToF. Following the clog, total positive ion counts decreased slightly (~ 10%) while total negative ion counts were unaffected. One reason for the minimal change in total negative ion counts may be that the clog roughly coincided with a shift in chemical composition of ions towards higher mass species. We believe this change in ion composition is real because it corresponds with changes in the composition of neutral trace gases (see Sect. S9). The transmission function of our instrument results in better transmission of higher *m/z* species, so we would expect to see more total negative ions if charges shift to higher *m/z* clusters even if the ambient ion concentration remains unchanged. Our work is grounded in analysis of changes in ambient ion composition rather than absolute quantification of ions, so by itself a small change in total ion counts has no direct influence on our conclusions. As noted in the main text, the pressure in the small segmented quadrupole (SSQ) was adjusted such that it did not change significantly following the clog. The SSQ is the first chamber in our differentially pumped instrument, so this would also regulate downstream pressures in a similar manner.

While the main effect of the clog would be a reduction total ion counts, it is also possible that a pinhole clog could affect the ion expansion. To investigate this potential effect, the timeseries of several strong peaks in the positive ion spectra consisting of protonated water or water clusters were evaluated. These clusters were selected because they are consistently strong signals. A substantial change in the ratio of these clusters coinciding with the clog would be evidence of changing declustering due to changes in the expansion. The ratios of these clusters change by less than 2% before and after the clog. This finding suggests that the clog has minimal impact on our conclusions.

S2 Fitting of binPMF peaks

In order to evaluate the error introduced by the binPMF and peak-fitting process, synthetic peaks were generated and analyzed. First, Gaussian peaks at selected positions between 200 and 550 *m/z* were generated in time-of-flight (ToF) space. Peak widths were equal to real peaks observed at the selected *m/z*. The peaks were sampled in ToF space at the same interval as the APi-ToF data acquisition. The ToF space to *m/z* space transformation was calculated as:

$$m/z = \left(\frac{(ToF - p1)}{p2}\right)^2 \quad \text{(S.1)}$$

where p1 and p2 are the fit parameters selected for the simulation and ToF is the time of flight (ns). This function was selected for the simulated data because it was found to the best fit function for real data and was used to fit both positive ion and negative ion data throughout the campaign. The values of p1 and p2 depended on the ion polarity being measured. "True" values for p1 and p2 were selected for the simulation. To simulate an upper limit estimate on error in the mass calibration and its impact on the binPMF results, pairs of p1 and p2 values were randomly selected from the set of p1 and p2 values fit from the ambient data using Tofware. Because p1 and p2 do not vary independently, each pair of values consisted of parameters calculated for the same time point in the calibration. This is an upper estimate of our error because it assumes that all shifts in mass calibration contribute error, but there are real shifts in p1 and p2 that result from temperature changes, drift within the instrument, and other factors. Following the ToF to *m/z* space transformation, the synthetic peaks in *m/z* space were binned using the same bins and bin widths used for binPMF and fit with a Gaussian to determine the peak center. Error introduced by the *m/z* calibration were determined using a Monte Carlo method to randomly select many sets of p1 and p2. Root mean squared errors introduced by this method were approximately 50 ppm for both positive ion and negative ion data. The simulation was also repeated using only the "true" fit parameters to determine whether error in the peak positions originated from simulated error in the mass calibration or from the binning and fitting procedure. Error was negligible (<<1 ppm) when using the "true" fit parameters, suggesting that most error is from the mass calibration and not the fitting procedure. Peak broadening was also evaluated. Peaks may be broadened both by the procedure of binning and fitting peaks to bins and by shifts in the mass calibration throughout the campaign. Figure S1 shows the comparison between the peak in a 15-minute average mass spectrum at *m/z* 487 and the Gaussian peak fit to the bins at that mass. Minimal broadening is observed. It should also be noted that peak widths have no direct implications for the conclusions of this work. Peak shape was also investigated. Figure S2 compares the high-resolution peak shape calculated in Tofware and a Gaussian peak shape fit to binPMF peaks. Although minor differences in peak shape are apparent, the shapes are broadly similar and, as with peak widths, peak shapes do not directly impact our results.

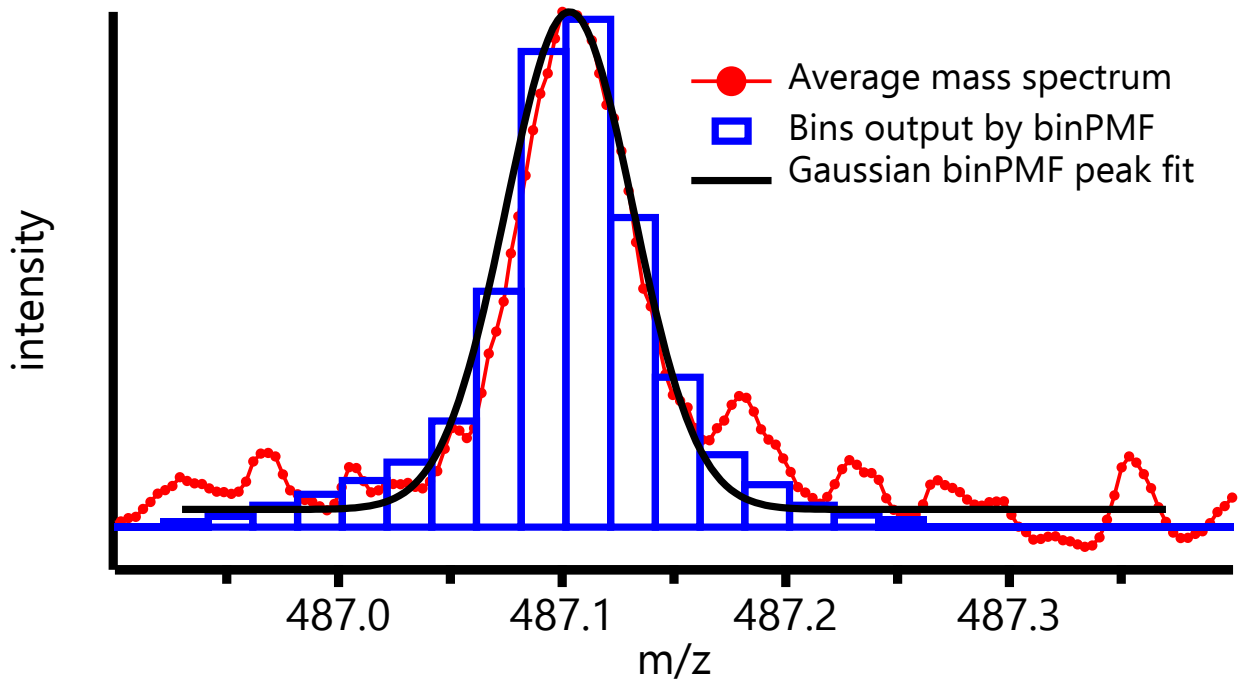


Figure S1: comparison of 15-minute average measured mass spectrum, binPMF binned signal, and Gaussian peak fit. Each bin is centered on the *m/z* value at the midpoint of the bin. The averaging time was 15 minutes.

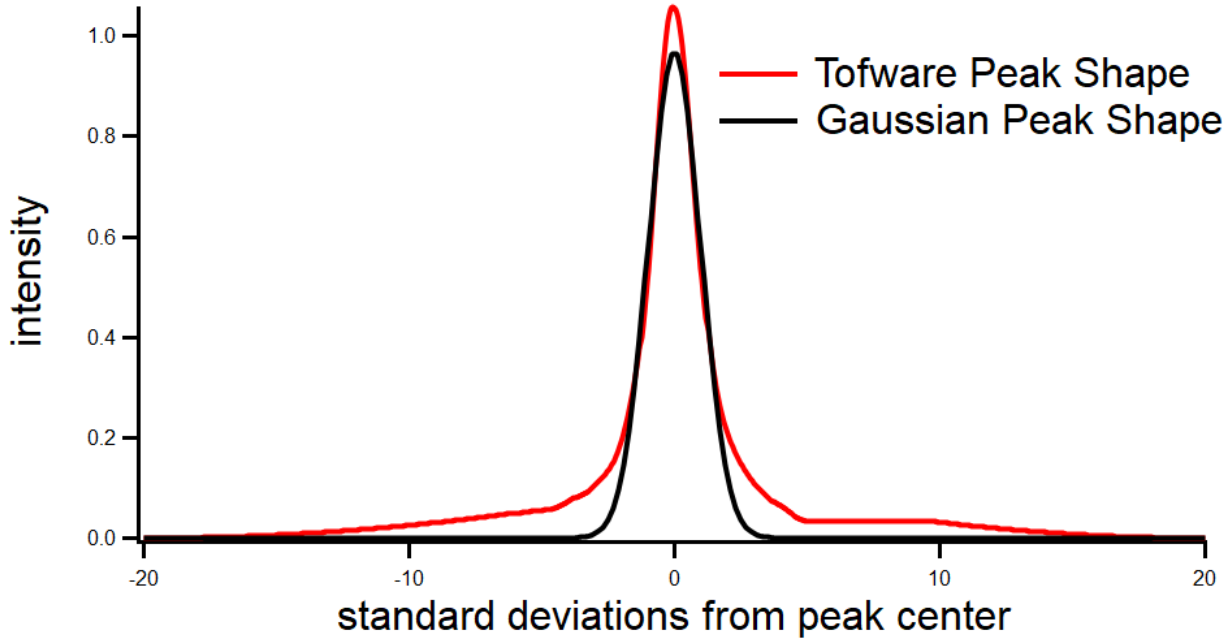


Figure S2: Comparison of Tofware peak shape and Gaussian peak shape.

S3 Peak Tables

Formula	Exact Mass	binPMF Peak Center	Factor(s)
HSO_4^-	96.96010	96.963	**Sulfur species,** *Sulfuric acid dimer*
$C_3H_3O_4^-$	103.00368	103.010	**Low *m/z* NO_3^-,** *High* m/z NO_3^-, *Sulfur species*
SO_5^-	111.94719	111.951	**Sulfur species,** *Sulfuric acid dimer*
$(H_2O)HSO_4^-$	114.97067	114.973	**Sulfur species,** *Sulfuric acid dimer*
$(H_2CO_3)NO_3^-$	123.98876	123.987	**Sulfur species,** *Sulfuric acid dimer,* Low *m/z* NO_3^-
$(HNO_3)NO_3^-$	124.98401	124.987	**Low *m/z* NO_3^-,** *Sulfuric acid dimer*
$C_5H_7O_5^-$	147.02990	147.029	**Low *m/z* NO_3^-,** *Sulfur species, High* m/z NO_3^-
$C_2H_3SO_6^-$	154.96558	154.967	**Sulfuric acid dimer,** Sulfur species, Low *m/z* NO_3^-
$(HNO_3)HSO_4^-$	159.95575	159.957	**Sulfuric acid dimer,** *Low* m/z NO_3^-
$(C_3H_4O_4)NO_3^-$	165.99933	166.000	**Low *m/z* NO_3^-,** *Sulfuric acid dimer,* High *m/z* NO_3^-
$(C_5H_6O_4)NO_3^-$	192.01498	192.013	**Low *m/z* NO_3^-,** High *m/z* NO_3^-
$(H_2SO_4)HSO_4^-$	194.92748	194.929	**Sulfuric acid dimer,** Sulfur species
$(H_2SO_4)C_2H_3SO_6^-$	252.93296	252.935	**Sulfuric acid dimer**
$(C_5H_7NO_7)NO_3^-$	255.01062	255.013	**Low *m/z* NO_3^-,** *Sulfur species,* High *m/z* NO_3^-
$(C_5H_9NO_7)NO_3^-$	257.02627	257.027	**Low *m/z* NO_3^-,** *High* m/z NO_3^-, Sulfur species
$(C_5H_{10}N_2O_8)NO_3^-$	288.03208	288.032	**Low *m/z* NO_3^-,** *High* m/z NO_3^-, *Sulfur species*

$(C_{10}H_{10}O_6)NO_3^-$ ***	288.03610	288.032	**Low *m/z* NO_3^-,** *High* m/z *NO_3^-, Sulfur species*
$(H_2SO_4)_2HSO_4^-$	292.89486	292.900	**Sulfuric acid dimer**
$(C_7H_{12}SO_7)NO_3^-$	302.01874	302.017	**Low *m/z* NO_3^-,** *Sulfur species*
$(C_7H_{12}SO_7)HSO_4^-$	336.99048	336.991	**Sulfur species,** *Low* m/z *NO_3^-*
$(C_{10}H_{15}NO_8)NO_3^-$	339.06813	339.064	*High* m/z *NO_3^-, Low* m/z *NO_3^-*
$(C_{10}H_{14}O_9)NO_3^-$	340.05215	340.052	*Low* m/z *NO_3^-, High* m/z *NO_3^-*
$(C_{10}H_{15}NO_9)NO_3^-$	355.06305	355.059	**High *m/z* NO_3^-,** *Low* m/z *NO_3^-*
$(C_{10}H_{14}O_{10})NO_3^-$	356.04706	356.051	**Low *m/z* NO_3^-,** *High* m/z *NO_3^-*
$(C_{10}H_{15}NO_{10})NO_3^-$	371.05796	371.057	**High *m/z* NO_3^-,** Low *m/z* NO_3^-
$(C_{10}H_{14}O_{11})NO_3^-$	372.04198	372.044	*High* m/z *NO_3^-, Low* m/z *NO_3^-*
$(C_{10}H_{15}NO_{11})NO_3^-$	387.05288	387.052	**High *m/z* NO_3^-,** Low *m/z* NO_3^-
$(C_{10}H_{14}O_{12})NO_3^-$	388.03689	388.044	**High *m/z* NO_3^-,** Low *m/z* NO_3^-
$(C_{10}H_{15}NO_{12})NO_3^-$	403.04779	403.047	**High *m/z* NO_3^-,** Low *m/z* NO_3^-
$(C_{10}H_{14}O_{13})NO_3^-$	404.03181	404.043	*High* m/z *NO_3^-, Low* m/z *NO_3^-*
$(C_{10}H_{15}NO_{13})NO_3^-$	419.04271	419.043	*High* m/z *NO_3^-, Low* m/z *NO_3^-*
$(C_{15}H_{23}NO_9)NO_3^-$	423.12565	423.121	**High *m/z* NO_3^-**
$(C_{15}H_{23}NO_{10})NO_3^-$	439.12056	439.118	**High *m/z* NO_3^-**
$(C_{15}H_{23}NO_{11})NO_3^-$	455.11548	455.113	**High *m/z* NO_3^-**
$(C_{15}H_{23}NO_{12})NO_3^-$	471.11039	471.108	**High *m/z* NO_3^-**
$(C_7H_{12}SO_7)C_7H_{11}SO_7^-$	479.05347	479.046	**Low *m/z* NO_3^-,** *Sulfuric acid dimer,* Sulfur species
$(C_{15}H_{23}NO_{13})NO_3^-$	487.10531	487.103	**High *m/z* NO_3^-**
$(C_{15}H_{23}NO_{14})NO_3^-$	503.10022	503.099	**High *m/z* NO_3^-**
$(C_{15}H_{23}NO_{15})NO_3^-$	519.09514	519.095	**High *m/z* NO_3^-**
$(C_{15}H_{23}NO_{16})NO_3^-$	535.09005	535.088	**High *m/z* NO_3^-**

Table S1: Table of negative ion binPMF formulas discussed in the main text with formula exact masses, binPMF peak centers, and the factors which peaks appear in. Formulas show proposed clusters. The listed factors contribute a large (bold), moderate (italicized), or minor (roman) fraction of the signal for each ion. Note that two possible formula assignments are listed for *m/z* 288; for further discussion see Sect 3.1.

Formula	Exact Mass	binPMF Peak Center	Factor(s)
$(C_3H_9N)H^+$	60.08078	60.081	*Nighttime, C18, Alkylpyridinium,* Daytime high *m/z*
$(C_4H_{11}N)H^+$	74.09643	74.095	**Alkylpyridinium**
$(C_2H_6N_2O)H^+$	75.05529	75.057	**Alkylpyridinium**
$(C_3H_{11}NO)H^+$	78.09134	78.093	**Nighttime,** *C18, Alkylpyridinium,* Daytime high *m/z*
$(C_6H_7N)H^+$	94.06513	94.067	**Alkylpyridinium,** C18, Daytime high *m/z*, Nighttime
$(C_6H_{15}N)H^+$	102.12773	102.130	**Nighttime**
$(C_7H_9N)H^+$	108.08078	108.081	**Alkylpyridinium,** *Nighttime,* Daytime high *m/z*, C18
$(C_6H_9NO)H^+$	112.07569	112.077	**Alkylpyridinium,** *C18,* Daytime high *m/z*, Nighttime
$(C_8H_{11}N)H^+$	122.09643	122.097	**Alkylpyridinium,** *Nighttime,* C18, Daytime high *m/z*
$(C_7H_{11}NO)H^+$	126.09134	126.093	**Alkylpyridinium,** *Nighttime,* C18, Daytime high *m/z*
$(C_9H_{13}N)H^+$	136.11208	136.113	**Alkylpyridinium,** *Nighttime,* C18, Daytime high *m/z*
$(C_8H_{13}NO)H^+$	140.10699	140.108	**Alkylpyridinium,** *Nighttime,* C18, Daytime high *m/z*
$(C_{10}H_{15}N)H^+$	150.12773	150.127	**Alkylpyridinium,** *Nighttime,* C18, Daytime high *m/z*
$(C_9H_{15}NO)H^+$	154.12264	154.123	**Alkylpyridinium,** *Nighttime,* C18, Daytime high *m/z*

$(C_{13}H_{21}NO_3)H^+$	240.15942	***	***
$(C_{15}H_{27}NO_3)H^+$	270.20637	270.205	**C18,** Nighttime, Daytime high *m/z*
$(C_{17}H_{31}NO_3)H^+$	298.23767	298.232	**Nighttime,** C18, Daytime high *m/z*
$(C_{18}H_{27}NO_3)H^+$	306.20637	306.197	**C18,** Daytime high *m/z*, Alkylpyridinium
$(C_{18}H_{29}NO_3)H^+$	308.22202	308.219	**C18,** Nighttime
$(C_{18}H_{31}NO_3)H^+$	310.23767	310.235	**C18,** Nighttime, Daytime high *m/z*
$(C_{18}H_{33}NO_3)H^+$	312.25332	312.248	**Nighttime,** *C18,* Daytime high *m/z*
$(C_{21}H_{31}NO_3)H^+$	346.23767	345.235	**C18,** *Daytime high* m/z*,* Nighttime

Table S2: Table of positive ion binPMF formulas discussed in the main text with formula exact masses, binPMF peak centers, and the factors which peaks appear in. The listed factors contribute a large (bold), moderate (italicized), or minor (roman) fraction of the signal for each ion. *m/z* 240 was removed prior to binPMF, and thus there is no binPMF peak center or binPMF factors associated with this ion (see Sect 2.3).

S4 binPMF error estimation

Electronic noise contributes more to error than counting statistics because of the low signal-to-noise of APi-ToF data. Thus, our initial investigation of error used values independent of signal intensity. We investigated two alternative error estimation methods in addition to the one described in the main text. In one of these we used only the very high *m/z* bins where no peaks were observed. For both positive ion and negative ion datasets no peaks were observed beyond *m/z* 700 and all *m/z* 700-1400 were used to estimate error. As with the technique described in the main text, the noise in this region was binned in the same manner as the signal at lower *m/z*. The standard deviation was then calculated for each noise bin throughout the campaign. All standard deviations for bins in the range *m/z* 700-1400 were then averaged. This results in one error value for all *m/z* and all time points.

We also investigated time-dependent error calculated using the high *m/z* noise bins. With this method one error value was calculated for all *m/z*, but it was allowed to vary over time. The standard deviation of all bins at *m/z* 700-1400 was calculated for each time point. This value was then used as the error for every *m/z* at that time point. These differing error estimation techniques produced similar distributions of error values (Fig. S3) with individual values within an order of magnitude of each other. The different error estimates did not result in significantly different binPMF solutions. We found PMF solutions to be insensitive to exact error values for error values of the correct order of magnitude.

Signal-to-noise ratios (SNRs) were not used to weight final binPMF results. While we did perform some PMF calculations with downweighting of low SNR bins, the solutions were very similar to those calculated without weighting. We found that baseline correction was far more important than signal weighting in achieving binPMF solutions that are not obfuscated by excess noise. With a poor baseline correction, a factor containing mostly noise was produced. Bins that did not contain chemical information were sorted into the noise factor, and the intensity of the factor varied to capture the drift in electronic error. With better baseline correction electronic noise was subtracted from the spectra prior to binning and all the binPMF factors contained chemically meaningful information. Since bins lacking in chemical information were addressed by the baseline correction, we determined that weighting was not necessary for our analysis.

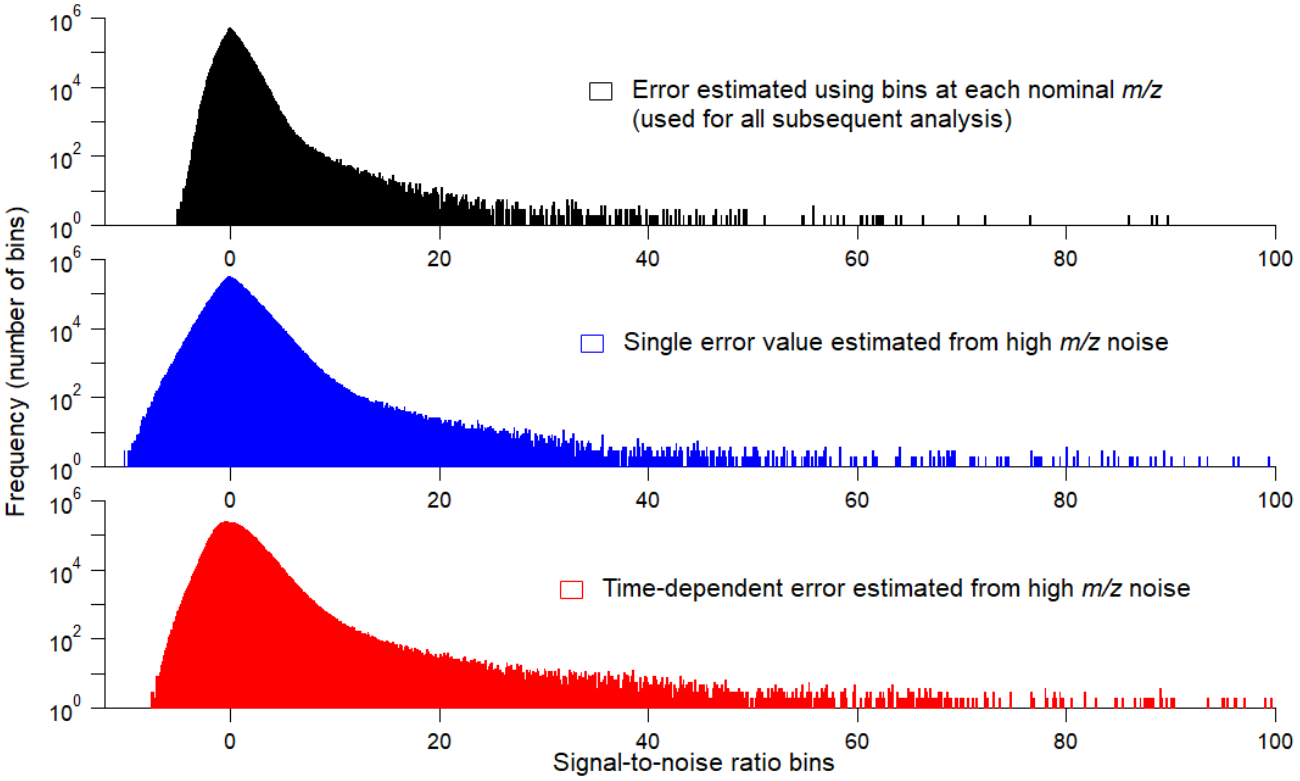


Figure S3: Histograms of the SNR calculated for each bin at each time point using the three different error estimation techniques.

S5 Sulfuric acid proxy calculation

The sulfuric acid concentration proxy was calculated using the method described by Mikkonen et al. (2011) using the formula:

$$[H_2SO_4] = 8.21 \times 10^{-3} \times k \times Radiation \times [SO_2]^{0.62} \times (CS \times RH)^{-0.13} \quad \text{(S2)}$$

where k is the effective rate coefficient of sulfur dioxide oxidation by hydroxyl radicals in cm^3/molec·s, radiation is a measurement of shortwave radiation (0.4-4 μm) in W/m^2, $[SO_2]$ is the measured concentration of sulfur dioxide in molec/cm^3, CS is the condensation sink in s^{-1}, and RH is the relative humidity. The constant is an empirical value derived from fitting measured sulfuric acid data. While Mikkonen et al. (2011) fit data from several different campaigns to calculate a constant that should be applicable under a wide variety of conditions, we cannot be certain that this value effectively reproduces the sulfuric acid concentration at SGP. Regardless, our sulfuric acid proxy was used only to understand the trends in sulfuric acid, and the absolute magnitude of the sulfuric acid concentration is not relevant to our conclusions. The effective rate coefficient of sulfur dioxide oxidation, k, was calculated by:

$$k = \frac{A \cdot k_3}{(A+k_3)} exp\left\{k_4 \left[1 + \log_{10}\left(\frac{A}{k_3}\right)^2\right]^{-1}\right\} \quad \text{(S3)}$$

$$A = k_1[M] \cdot \left(\frac{300}{T}\right)^{k_2} \quad \text{(S4)}$$

where $k_1 = 4\times10^{-31}$, $k_2 = 3.3$, $k_3 = 2\times10^{-12}$, $k_4 = 0.8$, [M] is the density of air in molec/cm^3, and T is the temperature in Kelvin (Mikkonen et al., 2011). The condensation sink was calculated according to the equation:

$$CS = 2\pi D \sum_i D_{pi}\beta_i N_i \quad \text{(S5)}$$

where D is the diffusion coefficient of sulfuric acid in m^2/s, D_{pi} is the diameter in m of particles in class size i, β_i is the Fuchs-Sutugin correction factor, and N is the number concentration of particles in number/m^3. Measurements of SO_2, particle size distribution, RH, radiation, and the parameters required to calculate the condensation sink are made routinely at the site (Trojanowski, 2016; Zhang, 1997).

S6 Negative ion binPMF solution selection

Compared to the four-factor solution presented in the main text, the three-factor solution (Fig S4), essentially combines the "sulfur species factor" and the "low *m/z* NO_3^- factor." While the sulfuric acid dimer is still represented fairly well by this solution, nearly every other daytime sulfur species at low *m/z* (including HSO_4^-, $(H_2O)HSO_4^-$, SO_5^-) has a very high residual and is not well-captured. Adding another factor allows the low *m/z* sulfur species to be sorted into a separate factor that reflects their behavior. In the five-factor solution (Fig S5), only the high *m/z* NO_3^- factor changes significantly. It is split into two factors with very similar mass spectra and diel behavior. One of the two new factors has a time series nearly identical to the time series of the high *m/z* NO_3^- factor except for a few short periods where it dips nearly to zero, and the other new factor has a time series that is nearly always close to zero except for brief spikes which correspond to the dips in the other factor (i.e. the sum of the two new time series is very similar to the time series of the high *m/z* NO_3^- factor in the four-factor solution). There does not appear to be any chemical explanation for these spikes, and they are likely due to small shifts in the mass calibration throughout the campaign. Combined with the similarity of the spectra, the unusual behavior of the time series implies that the splitting of the high *m/z* NO_3^- factor into two new factors has no chemical significance. Therefore, a fifth factor is not included and the four-factor solution is selected for further analysis.

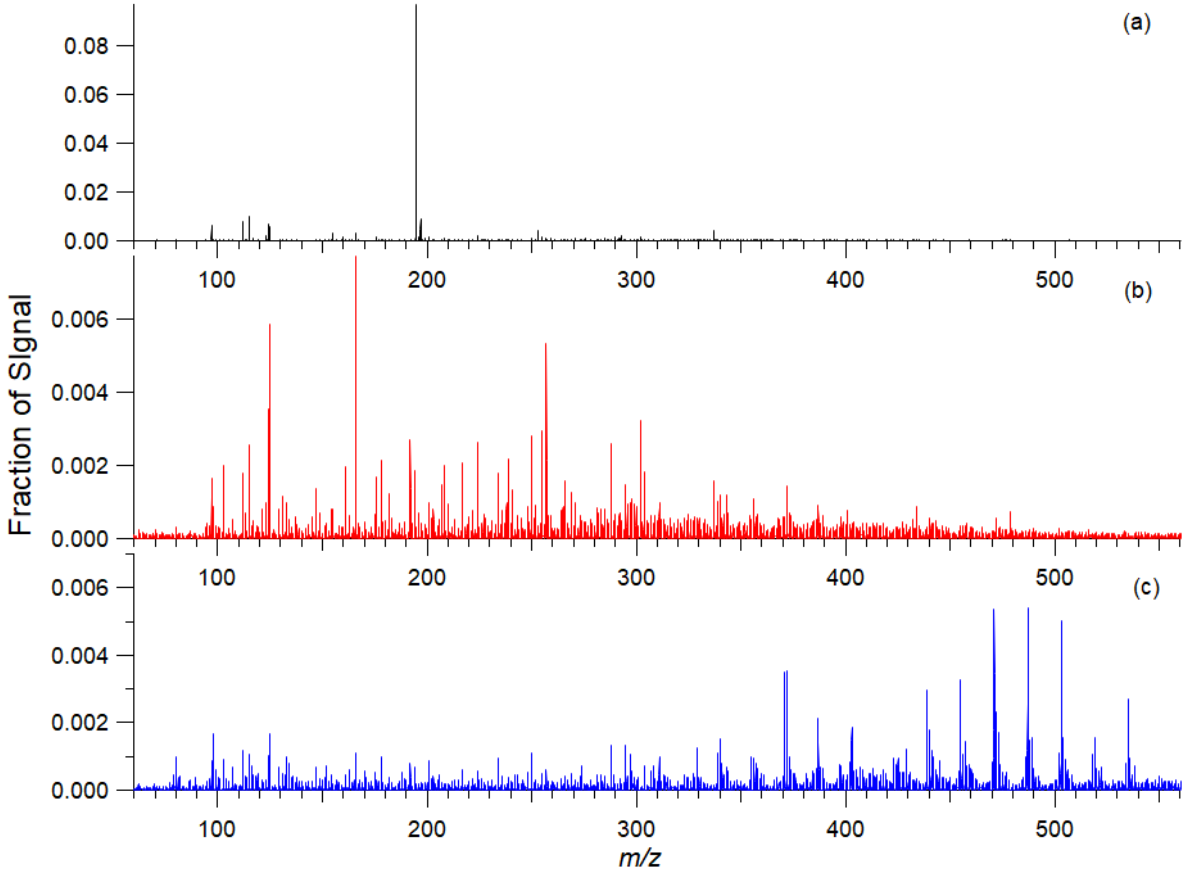


Figure S4: Mass spectra of the factors in the three-factor negative ion binPMF solution.

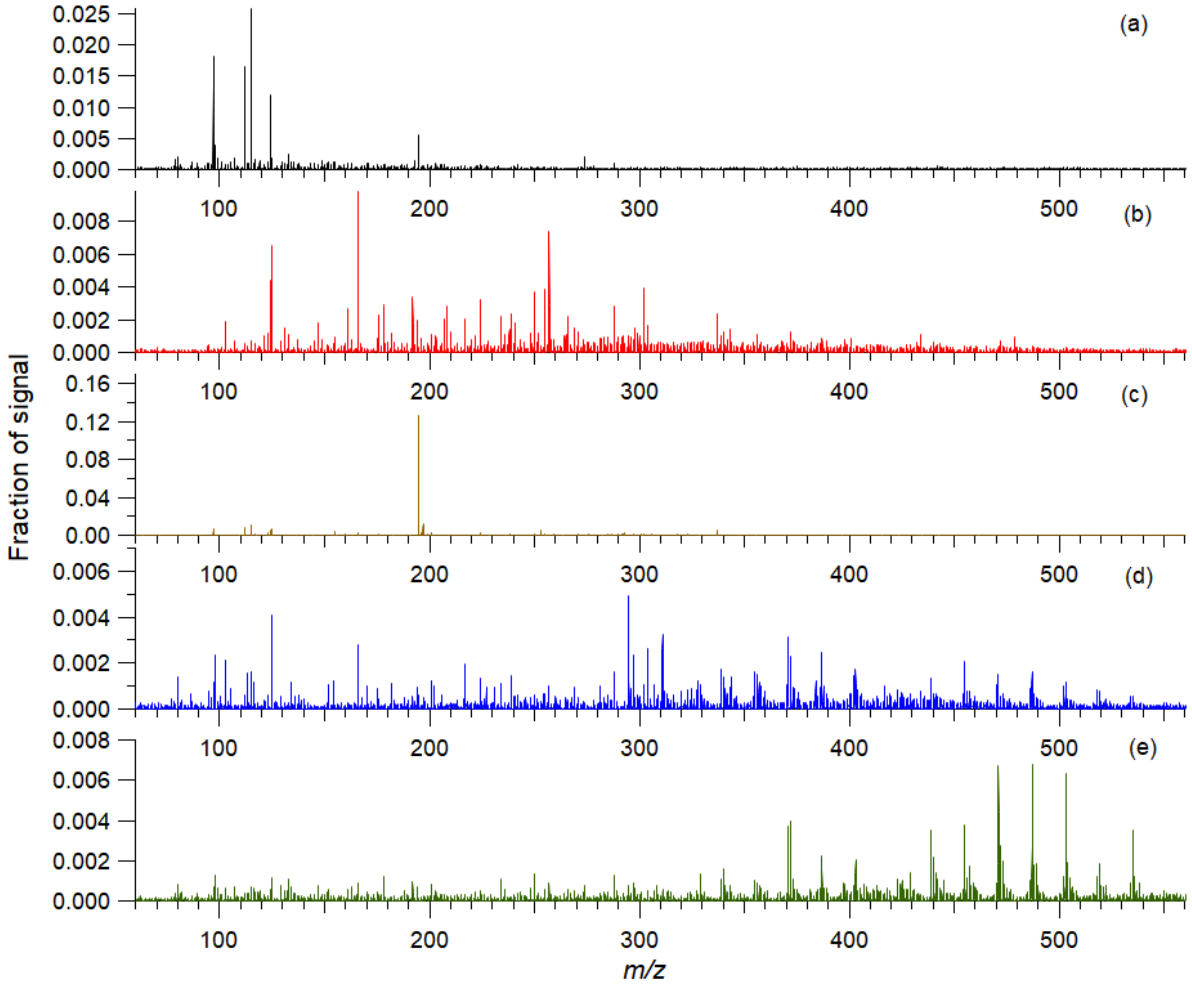


Figure S5: Mass spectra of the factors in the five-factor negative ion binPMF solution.

S7 Time series of negative ion binPMF factors

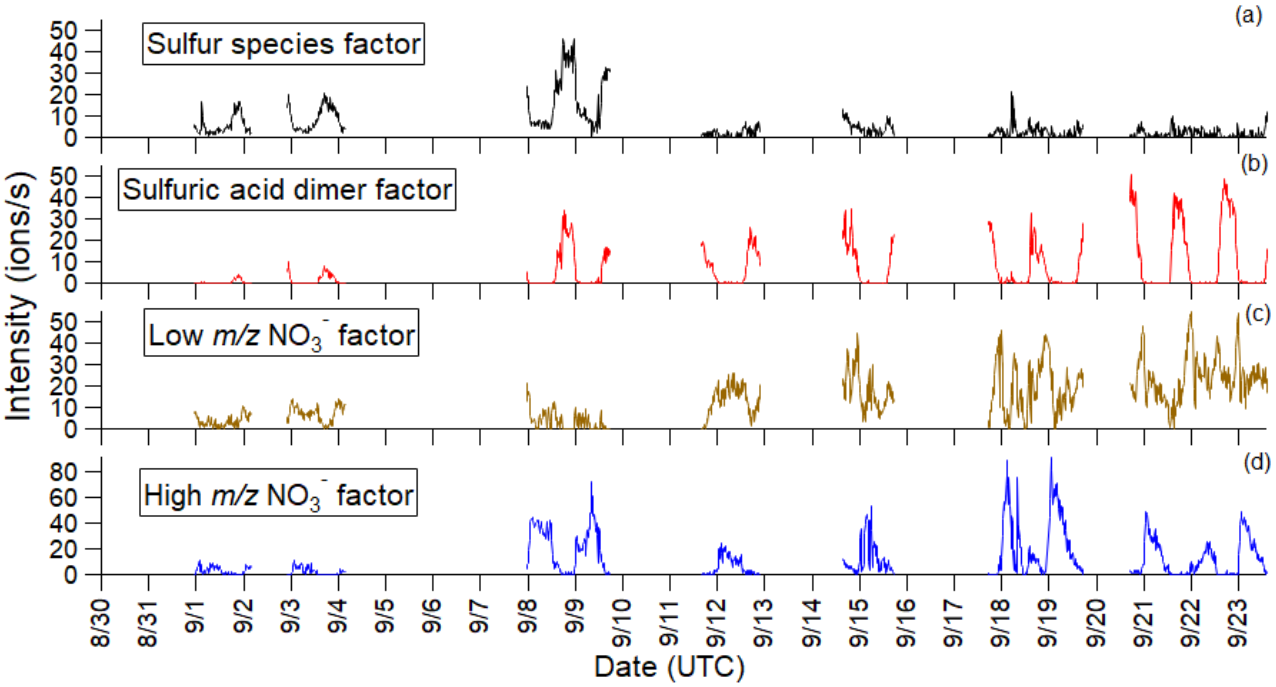
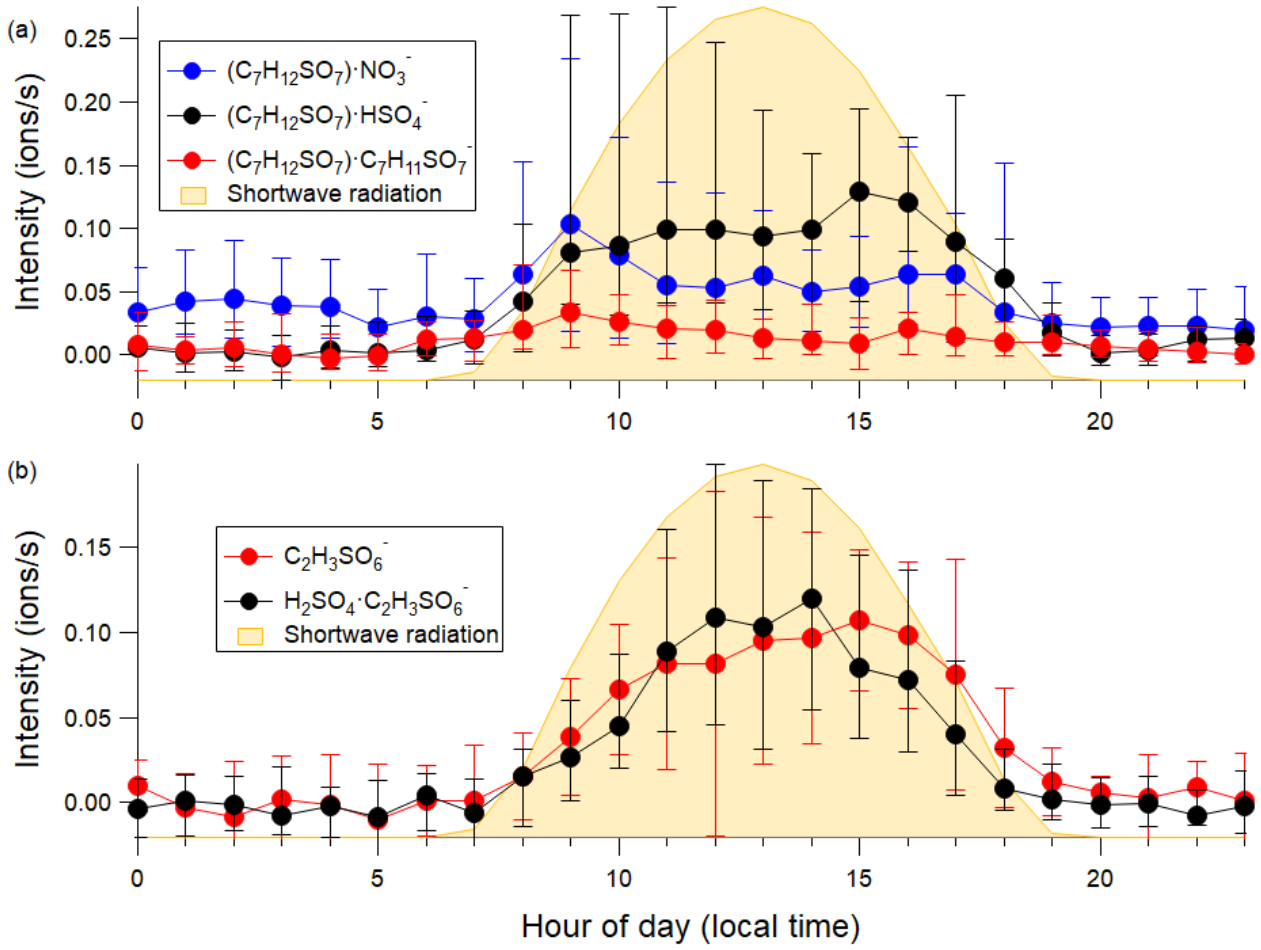


Figure S6: Time series of the four-factor negative ion binPMF solution throughout the campaign – (a) sulfur species factor, (b) sulfuric acid dimer factor, (c) low *m/z* NO_3^- factor, and (d) high *m/z* NO_3^- factor.

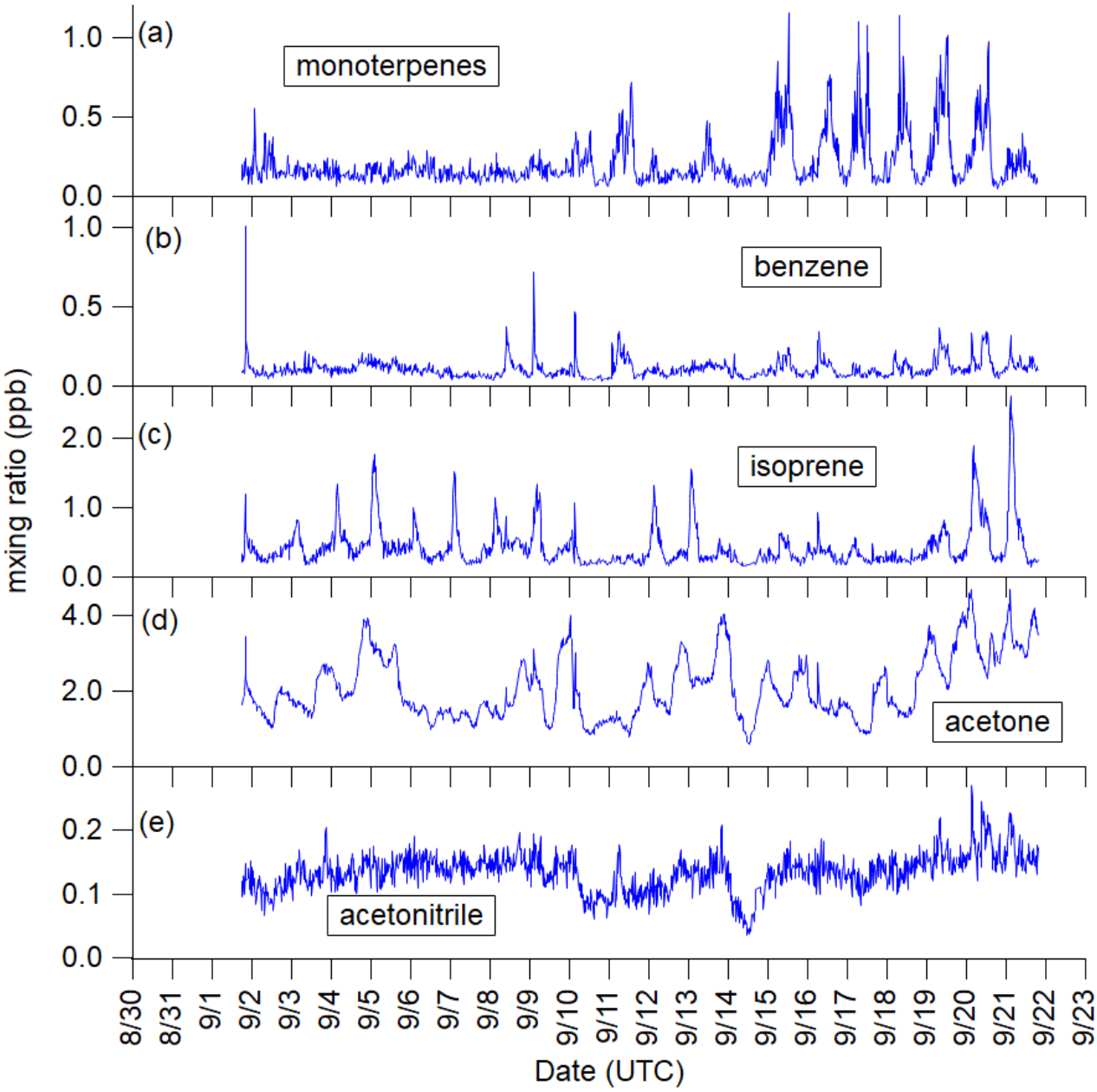
S8 Diel profiles of organosulfates



150

Figure S7: Diel profiles of selected organosulfates (a) Clusters of $C_7H_{12}SO_7$ (C7) with various anions and, (b) glycolic acid sulfate (GAS, $C_2H_3SO_6^-$) and its cluster with HSO_4^-.

S9 Time series of selected PTRMS ions



155

Figure S8: Time series of selected trace gases measured by the proton transfer reaction mass spectrometer through the campaign – (a) monoterpenes, (b) benzene, (c) isoprene, (d) acetone, and (e) acetonitrile. Note that the range of the y-axis varies between panels.

The PTRMS measurements are external tracers that substantiate the changes observed in binPMF factors. The time series of monoterpenes (MTs) measured by the PTRMS (Fig. S8a) shows a strong increase in the second half of the campaign. Several

binPMF factors in both the negative ion and positive ion modes show similar increases midway through the campaign. Although it does not coincide exactly with the increases in intensity of binPMF factors, the increase in MTs measured by the PTRMS suggests that the changes in the intensity of the binPMF factors is not due to instrument artifacts. The instrument did become clogged at a similar point in the campaign (13 September). The primary effect of a clog, however, would be to reduce measured signal rather than alter the signal composition. Furthermore, any change in composition induced by a clog would likely have a fairly constant diel profile. Combined with the evidence from the PTRMS tracers, we conclude that the most likely explanation for changing binPMF results is due to real atmospheric variation rather than an instrument artifact.

S10 Positive ion binPMF solution selection

In the positive ion binPMF solution with only three factors (Fig S7), species which peak either in the morning or evening but are not consistently high through the day or night are not captured well. The C18 species at *m/z* 306, 308, and 310 have high residuals, as do some species in the nighttime factor including *m/z* 312. Adding a fourth factor does a better job of accounting for this signal because distinct factors that peak either in the morning or evening are resolved. When a fifth factor is added (Fig S8), the alkylpyridinium *m/z* factor is split into two new factors. These new factors are very similar in their spectra with peaks at the same *m/z* and only minor differences in relative intensities. As with the splitting of the high *m/z* NO_3^- factor in the negative ion mode binPMF solution (described in Sect S4), one time series contains dips that correspond to spikes in the other time series, and the sum of the two is similar to the time series of the alkylpyridinium factor. Again, this appears to be splitting of a single factor into two new ones without any chemical significance. Therefore, the four-factor solution was selected.

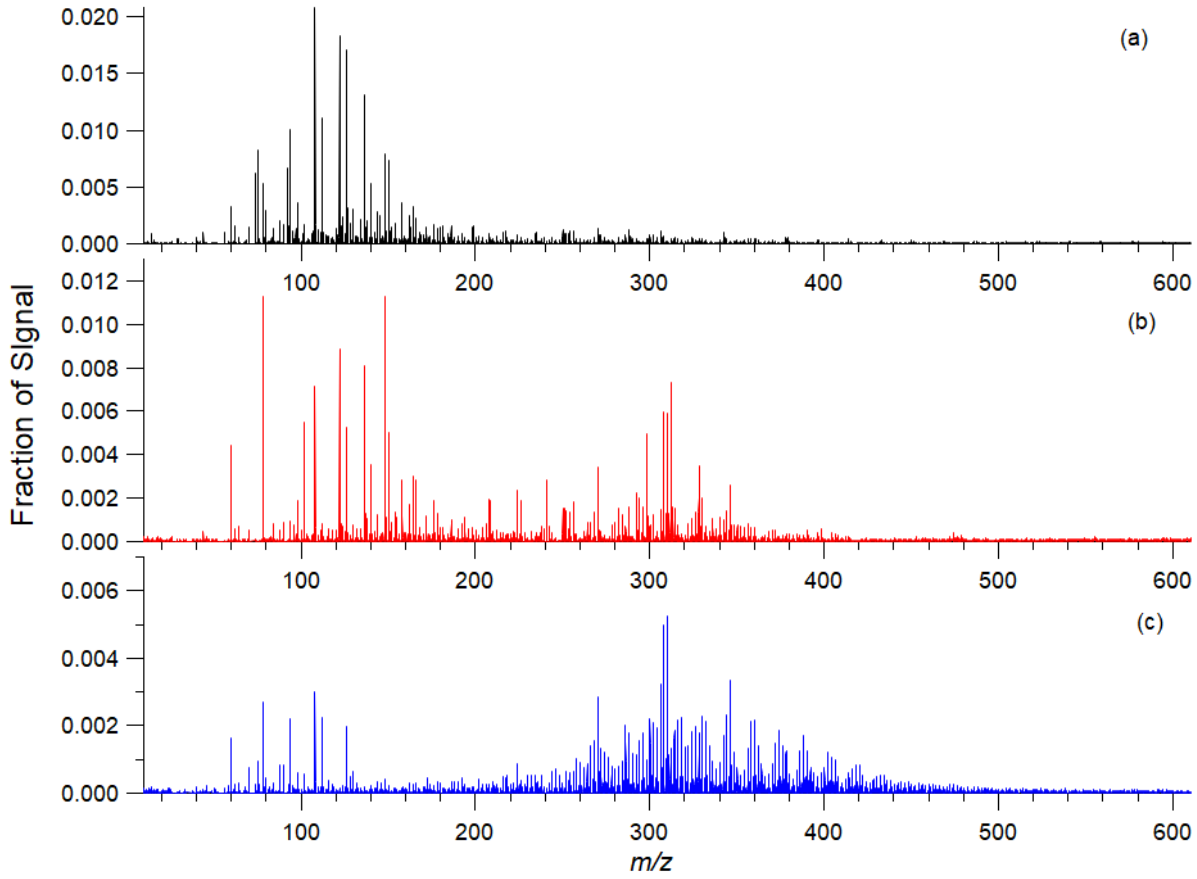


Figure S9: Mass spectra of the factors in the three-factor positive ion binPMF solution.

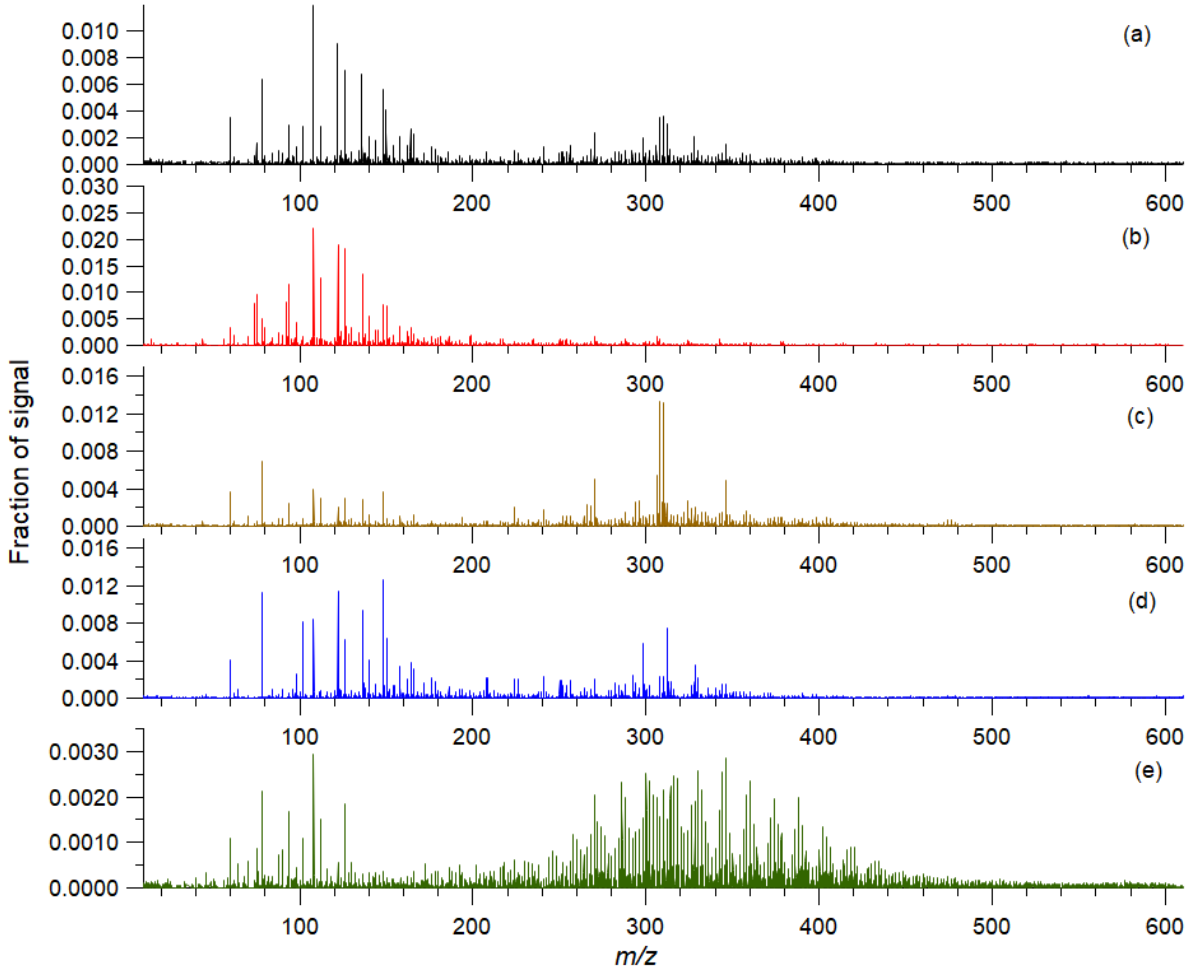


Figure S10: Mass spectra of the factors in the five-factor positive ion binPMF solution.

185

S11 Time series of positive ion binPMF factors and *m/z* 240

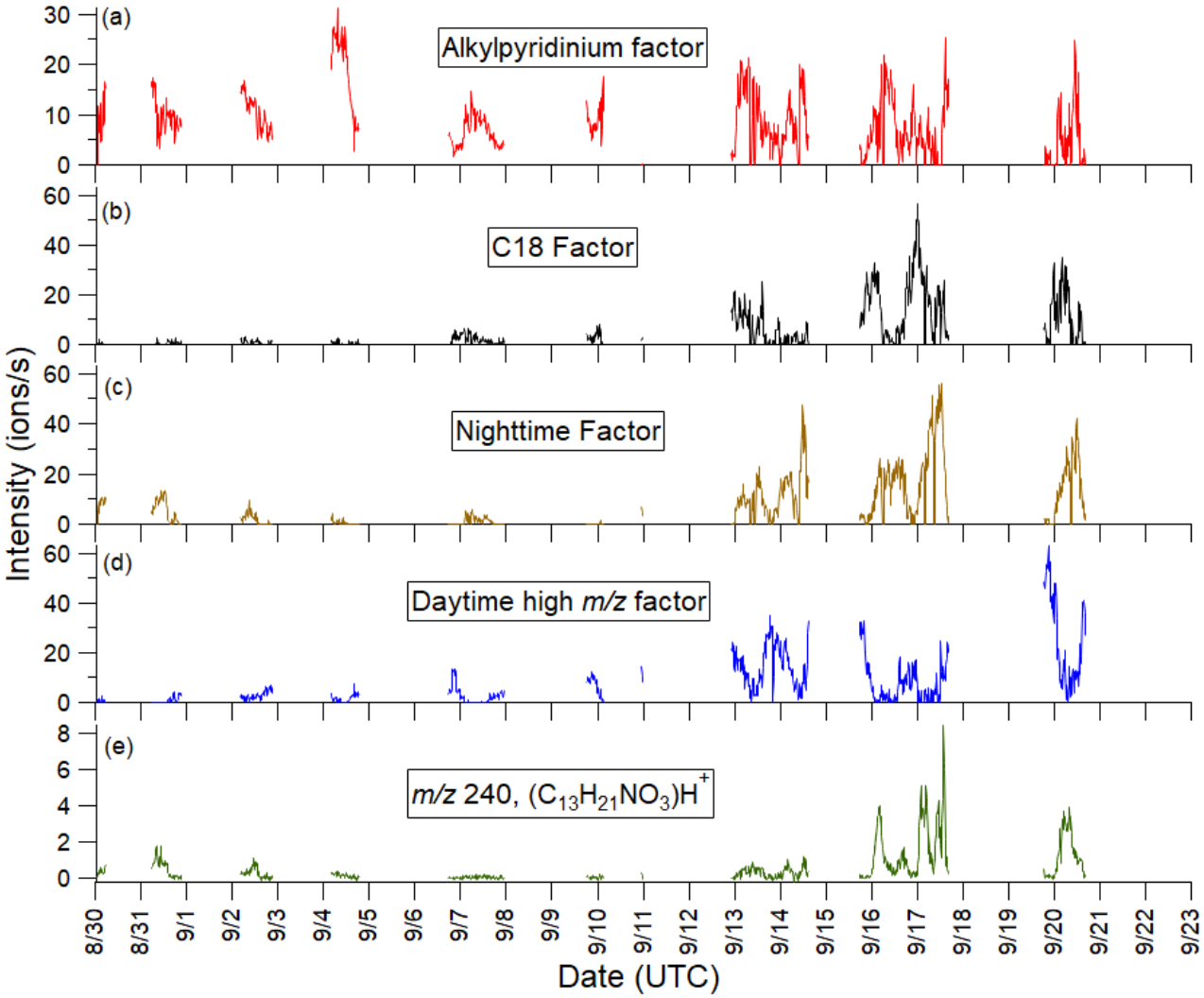


Figure S11: Time series of the four factors and selected ion in the positive ion binPMF solution (a) alkylpyridinium factor, (b) C18 factor, (c) nighttime factor, (d) daytime high *m/z* factor, and (e) the ion at *m/z* 240, $(C_{13}H_{21}NO_3)H^+$.

190

S12 Diel profiles of alkylpyridiniums

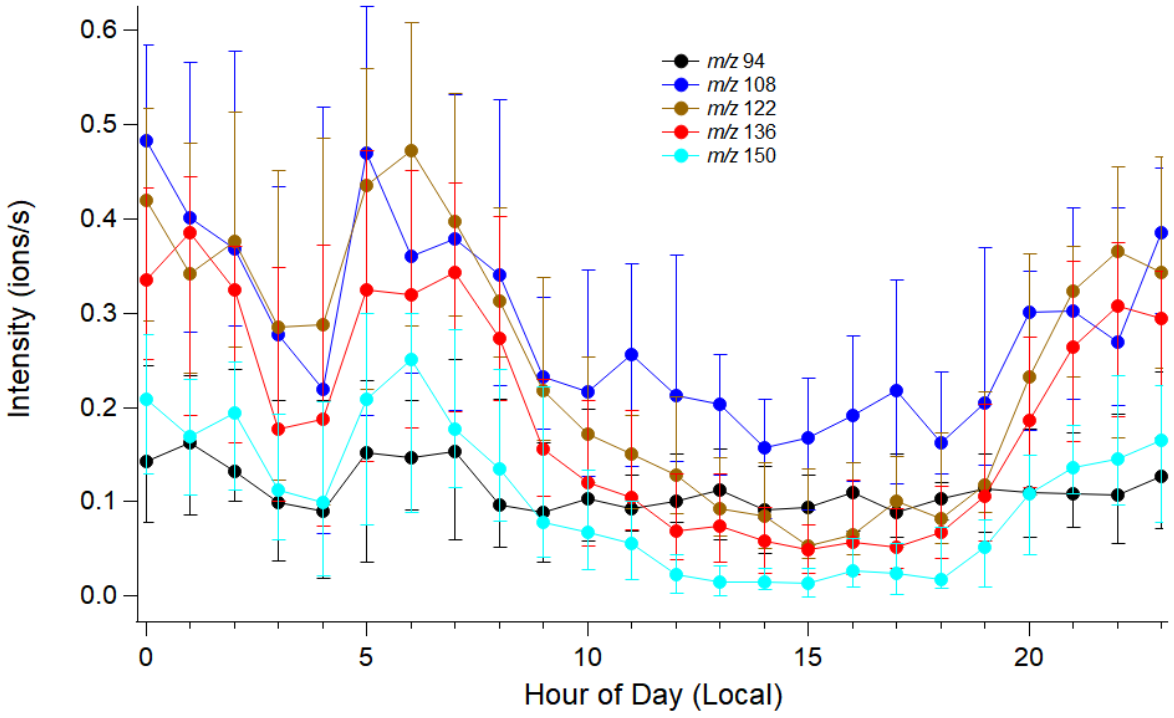


Figure S12: Diel plots of the series of five alkylpyridinium cations (*m/z* 94-150, $(C_5H_5(CH_2)_xN)H^+$, $1 \leq x \leq 5$) measured in positive ion mode.

195

S13 Diel profiles of C3 and C4 amines

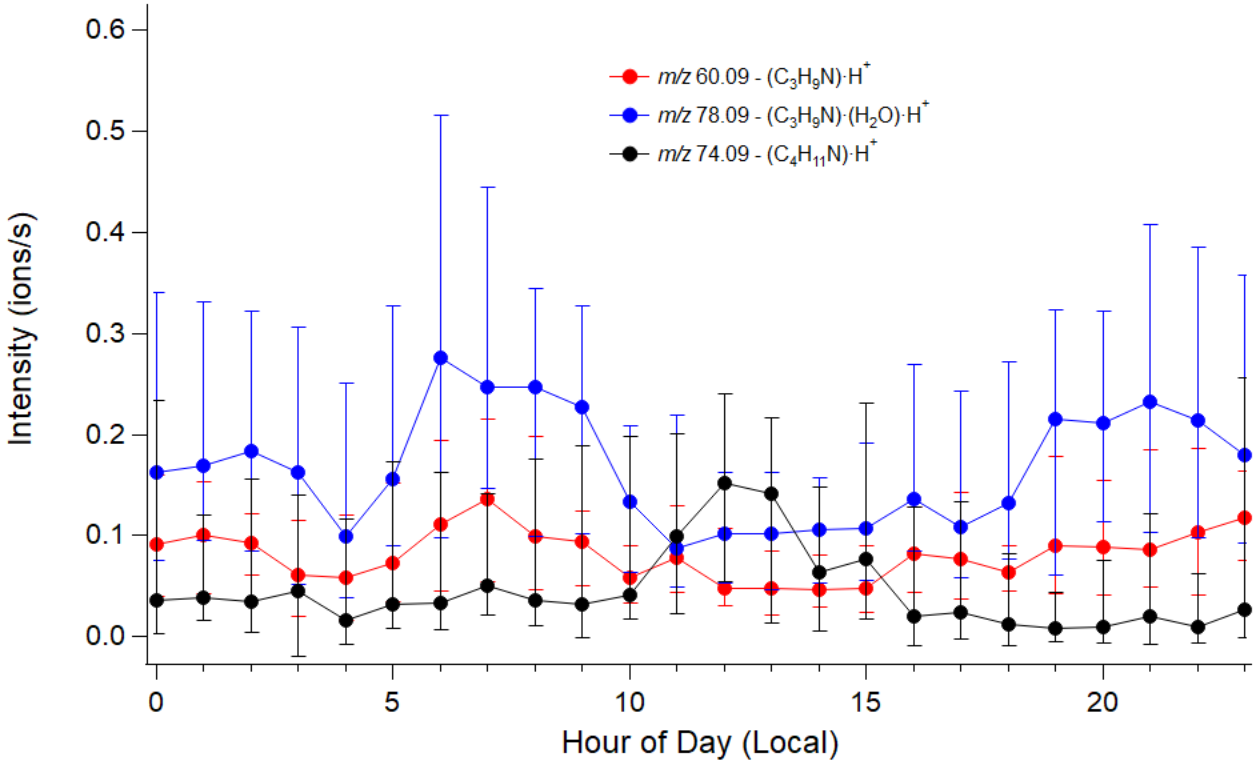


Figure S13: Diel plots of the bins containing C3 amine (C_3H_9N), C3 amine clustered with water, and C4 amine ($C_4H_{11}N$) signals. While the C3 amine and its water cluster peak early in the morning, the C4 amine is most intense in the middle of the day.

S14 Positive ion GKA with possible formulas

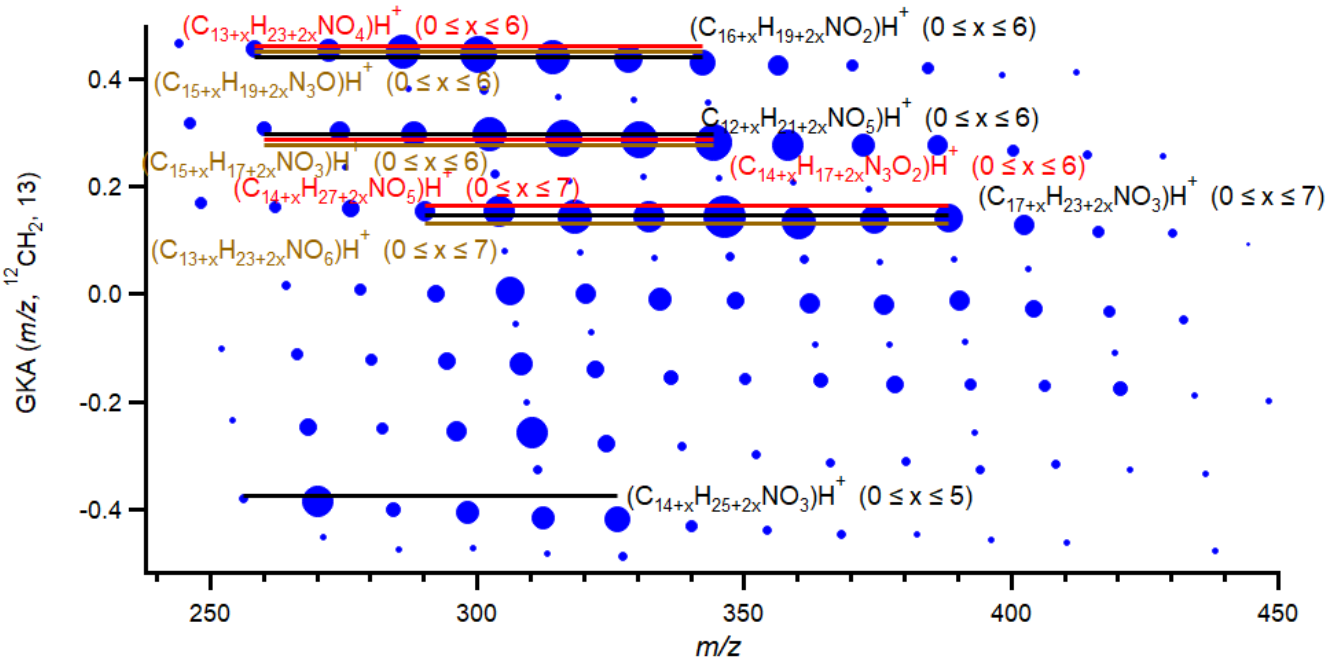


Figure S14: GKA plot of high *m/z* daytime factor with CH_2 base unit and integer of 13. The most probable formulas are shown in black. Formulas in red and brown are considered less likely.

The generalized Kendrick analysis (GKA) plot of the high *m/z* daytime factor (Fig. S14) shows that some peaks fall along horizontal lines and are likely related by CH_2 units, but this does not hold true for all observed species. This result is confirmed by high-resolution peak fitting, which also suggests that the ions resulting in the observed peaks are not separated only by units of CH_2. For example, the two most intense peaks in the series of peaks with the highest GKA values (top line on the plot) are found at *m/z* 286 and 300. Since they are separated by Δ*m/z* of 14, it is possible that they are related by a CH_2 unit. Both the GKA plot of binPMF peaks and high-resolution peak fitting suggest that $(C_{18}H_{23}NO_2)H^+$ and $(C_{19}H_{25}NO_2)H^+$ are reasonable formulas for these peaks. However, in the series of peaks with the most negative GKA values (bottom line of the plot) the most intense peak is at *m/z* 270. Both the GKA plot and high-resolution peak fitting suggest that $(C_{15}H_{27}NO_3)H^+$ is a reasonable formula for this peak. There is a peak present at *m/z* 284, but the deviation of that point from the horizontal line implies that it is not related to the peak at *m/z* 270 by a CH_2 unit. High-resolution peak fitting confirms that $(C_{16}H_{29}NO_3)H^+$, the formula which corresponds to the addition of a CH_2 group, is not a likely formula for this peak (the formula $(C_{15}H_{25}NO_4)H^+$ fits better). Therefore, both GKA analysis and high-resolution peak fitting agree that the observed species are not related only by CH_2 units, and it is unlikely that the shift in the plot is an artifact caused by the binning and fitting procedure.

S15 Positive ion HYSPLIT back trajectory cluster analysis

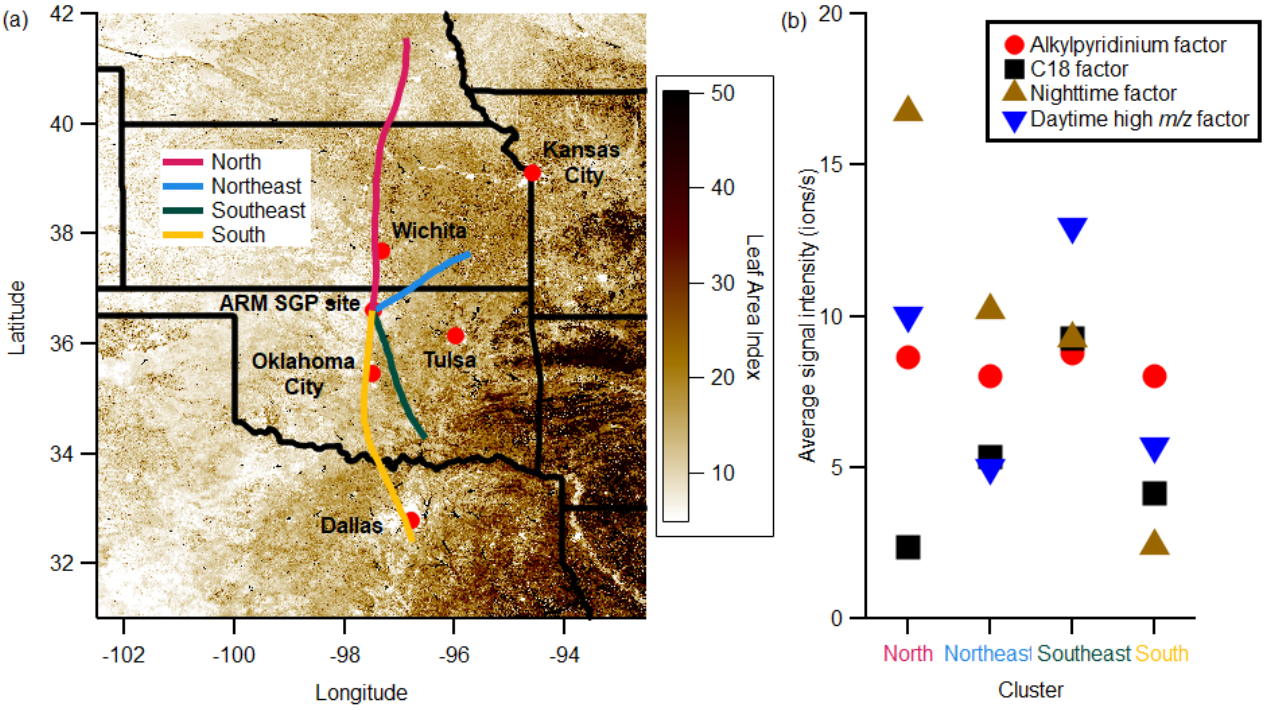


Figure S15: (a) HYSPLIT clusters calculated from back trajectories. The color scale shows leaf area index measured by MODIS. (b) Average signal intensities of factors for each HYSPLIT back trajectory cluster

As with negative ion factors, HYSPLIT back trajectory cluster analysis was performed for positive ion binPMF factors. Figure S14 shows back trajectory clusters for 24-hour back trajectories calculated each hour overlaid on a map of leaf area index. Figure S14(a) shows the back trajectories grouped into each cluster. The north (n = 12 trajectories), northeast (n = 61), and southeast (n = 80) clusters respectively have 58%, 54%, and 48% of their trajectories arriving at the site during the day (8:00-18:00 local time). The south cluster (n = 54) is the only cluster which has a significantly different number of daytime and nighttime trajectories with 33% of trajectories arriving during the day.

Figure S14(b) shows the average signal of each binPMF factor when the HYSLPIT clusters arrive at the site. The alkylpyridinium factor is almost constant among the clusters, which is consistent with the long atmospheric lifetime expected for these species. This consistency could also be due to local sources distributed approximately uniformly around the site. For the other factors, interpreting the back trajectories is challenging since the sources and chemistry of the observed ions is not yet understood. Additionally, it should be noted that, especially for species with shorter lifetimes, local processes such as crop harvesting may contribute more to the changes in observed intensity than long range transport. Possible evidence that local emissions and chemistry may control these factors comes from the nighttime factor which is enhanced in the north cluster and

is lowest in the south cluster despite both of these trajectories passing over urban areas (Wichita, Kansas and Oklahoma City, Oklahoma respectively).

240

References

Mikkonen, S., Romakkaniemi, S., Smith, J. N., Korhonen, H., Petäjä, T., Plass-Duelmer, C., Boy, M., McMurry, P. H., Lehtinen, K. E. J., Joutsensaari, J., Hamed, A., Mauldin, R. L., Birmili, W., Spindler, G., Arnold, F., Kulmala, M., and Laaksonen, A.: A statistical proxy for sulphuric acid concentration, Atmos. Chem. Phys., 11, 11319–11334, https://doi.org/10.5194/acp-11-11319-2011, 2011.

Trojanowski, R.: Sulfur Dioxide Monitor (AOSSO2), 2016-08-23 to 2016-09-21, Southern Great Plains (SGP) Lamont, OK (Extended and Co-located with C1) (E13). ARM Data Center. Data set accessed 2020-07-10 at http://dx.doi.org/10.5439/1250820.

Zhang, D.: Radiative Flux Analysis (RADFLUXBRS1LONG). 2016-08-30 to 2016-10-01, Southern Great Plains (SGP) Central Facility, Lamont, OK (C1). ARM Data Center. Data set accessed 2020-04-23 at http://dx.doi.org/10.5439/1395069.