

1 Supporting Information of

2 Ambient Measurements of Highly Oxidized Gas Phase Molecules during the
3 Southern Oxidant and Aerosol Study (SOAS) 2013

4

5

6 Paola Massoli¹, Harald Stark^{1,2,3}, Manjula R. Canagaratna¹, Jordan E. Krechmer^{1,2,3}, Lu Xu^{5*}, Nga L.
7 Ng^{4,5}, Roy L. Mauldin III⁶, Chao Yan⁷, Joel Kimmel^{1,8}, Pawel K. Misztal^{9,10,**}, Jose L. Jimenez^{2,3},
8 John T. Jayne¹ and Douglas R. Worsnop¹

9

10

11 ¹ Center for Aerosol and Cloud Chemistry, Aerodyne Research Inc., Billerica, MA 01821

12 ² Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309

13 ³ Department of Chemistry, University of Colorado, Boulder, CO 80309

14 ⁴ School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332

15 ⁵ School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332

16 ⁶ Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO 80309

17 ⁷ Department of Physics, University of Helsinki, FI-00560, Helsinki, Finland

18 ⁸ Tofwerk AG, CH-3600, Thun, Switzerland

19 ⁹ Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720

20 ¹⁰ Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720

21 * Now at the Division of Geological and Planetary Sciences, California Institute of Technology,
22 Pasadena, CA 91125

23 ** Now at Centre for Ecology & Hydrology Edinburgh, Midlothian, EH26 0QB, UK

24

25

26 *Corresponding Author:* Paola Massoli (pmassoli@aerodyne.com)

27 1. Relative transmission measurements

28 Data obtained with the NO_3^- CIMS instrument must be corrected for any m/z dependence that is
29 associated with processes such as transmission of ions from the ion chamber to the mass spectrometer,
30 orthogonal extraction of the ions into the time-of-flight chamber, and detection of ion signal with the
31 multichannel plate. Relative transmission values for the mass spectrometer as a function of m/z were
32 determined in a series of laboratory experiments following methods described in previous studies
33 using perfluorinated alkanic acids with carbon chain lengths of C_5 , C_7 , C_8 , and C_9 .^{1,2} These
34 compounds form clusters with the NO_3^- reagent ion and with deprotonated parent molecules. As
35 shown in Figure S1, the reagent ion clusters and the clusters consisting of 1-3 parent molecules
36 produced a series of signals in the mass spectrum that span a range of m/z (panel a). The experiments
37 involved measurements of the reagent ion signals in the absence of the perfluorinated acid followed by
38 time-series measurements of ion signal after injection of the fluoroacid of interest. The amounts of
39 fluoroacid injected were high enough to lead to a significant decrease in the reagent ion signal. The
40 analysis is based on the idea that in the absence of m/z dependent transmission effects, the total ion
41 current from all the detected ions should remain constant throughout the experiment. Thus, any change
42 in total signal upon addition of the perfluorinated acids provides a measure of the change in ion
43 transmission efficiencies at the m/z 's of the perfluorinated acid ions compared to those of the reagent
44 ions. Two approaches for deriving the relative transmission from the measured time series (panel b)
45 were pursued. First, the 2-D matrix consisting of time series of all reagent and analyte ions was
46 subjected to a matrix inversion algorithm to solve the system of linear equations:

47
$$\text{Constant} = \text{C1} * \text{I1}(t) + \text{C2} * \text{I2}(t) + \dots + \text{CN} * \text{IN}(t) \quad \text{Equation 1}$$

48 where Constant is the constant signal across time series

49 $\text{C1}, \dots, \text{CN}$: fit coefficient

50 $\text{I1}(t), \dots, \text{IN}(t)$: ion signal time series

51 The optimized coefficients C1-CN were converted to relative transmission values by inverting and
52 normalizing to a specific coefficient, typically the coefficient related to the NO_3^- signal at m/z 62.
53 Solutions to the system of linear equations from this method showed a large degree of variation (panel
54 c). This is likely due to the unconstrained nature of the matrix method. A more constrained solution

55 was achieved by assuming that relative transmission follows a functional dependence to m/z , which
56 was derived from an average of the solutions from the matrix inversion:

$$57 \quad RT = 10^{k*(m/z - m/z_0)} \quad \text{Equation 2}$$

58 The optimum value for k of 1.9 ± 0.2 with m/z_0 of 62 was derived from an iterative algorithm that
59 optimized the sum of all ion signals to be constant, following Eq. 1. Much better convergence was
60 achieved using this method (panel d). The relative transmission values were normalized to m/z 62 and,
61 together with the calibration factor C of $1.89 * 10^{10}$ molecules cm^{-3} , converted to absolute sensitivities
62 (in ppb) for all ion signals.

63

64 2. PMF error calculations, diagnostics, and correlation with external tracer

65

66 The Positive Matrix Factorization (PMF) model expresses the measured $m \times n$ data matrix of CIMS
67 mass spectra vs. time (X) as a matrix product of two matrices corresponding to PMF factor time series
68 (G) and PMF factor mass spectra (F):

69

$$70 \quad X = GF + E$$

71 where E is the $m \times n$ data matrix corresponding to the residuals between the measured matrix elements
72 of X and the modelled matrix product of the factor time series and factor mass spectra. The
73 factorization is achieved using non-negatively constrained weighted least squares in which the
74 quantity Q is minimized. Q is defined as

75

$$76 \quad Q = \sum_{i=1}^m \sum_{j=1}^n (E_{ij} / \sigma_{ij})^2$$

77

78 where E_{ij} refers to elements of the residual matrix and σ_{ij} refers to the standard deviation of each data
79 point in the X matrix (X_{ij}). In the limit where the residuals of data points are equal to their standard
80 deviations, each data point contributes a value of 1 to the total Q and the expected value of Q , also
81 referred to as Q_{exp} , is approximately equal to the number of elements of in the X matrix. Q/Q_{exp} is
82 monitored as a diagnostic of the quality of the fits. Previous studies have provided a detailed

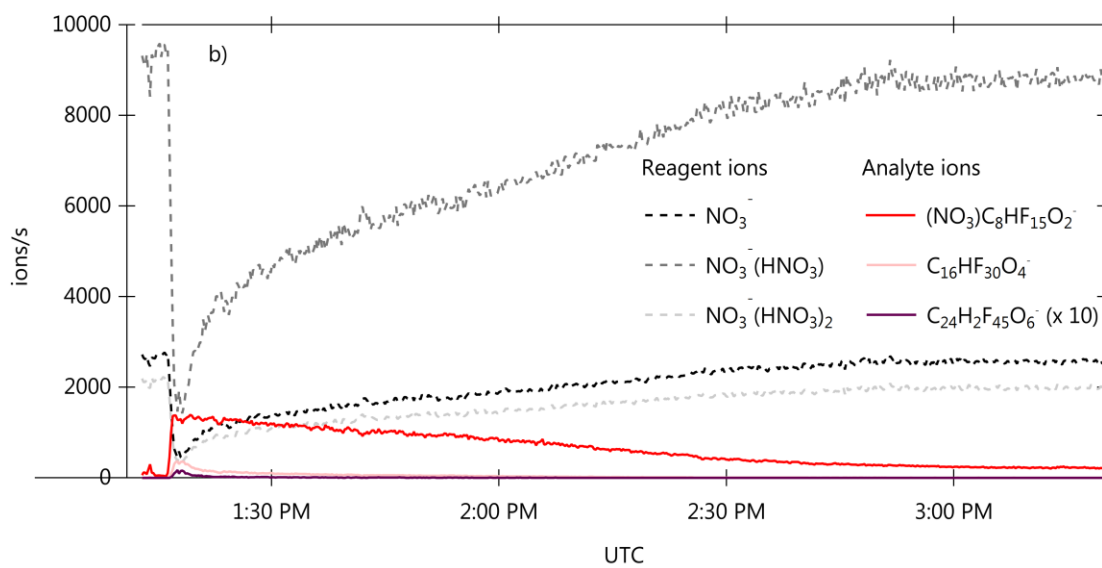
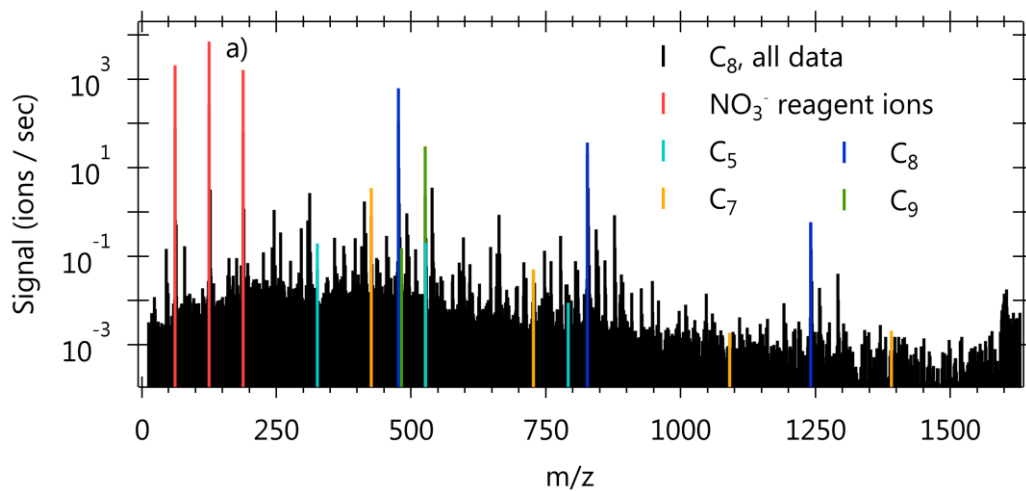
83 description of how to calculate the standard deviation values (σ_{ij}) for PMF analysis of unit-mass
84 resolution CIMS data.³ In this work, PMF analysis was performed on the high resolution dataset in
85 order to utilize the additional chemical information in the mass spectra. Estimation of the uncertainties
86 associated with high-resolution fitting, in which ion intensities are obtained from overlapping peaks,
87 are complicated⁴, and methods for approximating them have not yet been fully developed. Thus σ_{ij}
88 values are estimated according to Poisson counting statistics, the dominant source of noise in the unit-
89 mass resolution data³, as follows:

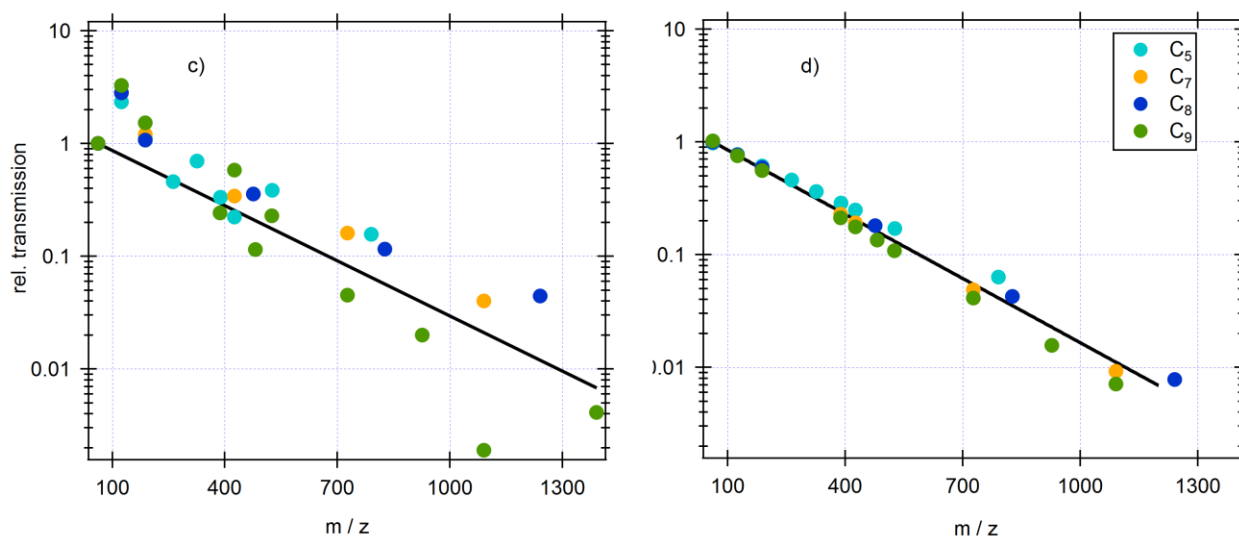
90

91

$$\sigma_{ij} = (I/t_s)^{0.5}$$

92 Where I is the raw ion signal in ions/s, obtained before corrections for ion transmission and extraction
93 are applied, and t_s is the integration time in seconds. We note that the σ_{ij} values estimated using
94 Poisson statistics provide a lower limit for the real noise in the high-resolution data. The PMF analysis
95 in this work utilized the PMF Evaluation Tool (PET)⁵ together with the PMF2 algorithm.⁶ The signal-
96 to-noise ratio (SNR) for each data point was calculated according to previous studies^{5,7} and “bad”
97 variables with SNR < 0.2 were down-weighted by 10 while “weak” variables with SNR between 0.2
98 and 2 were down-weighted by 3. Figure S4 shows the key diagnostic plots for the Positive Matrix
99 Factorization (PMF) performed on the SOAS 2013 NO₃⁻ CIMS dataset. The Q/Qexp is shown as a
100 function of the number of factors P (top panel, left). A 6-factor solution (P = 6), yielding a Q/Qexp of
101 1.1 was chosen to explain this dataset. The top right panel shows the total residuals for the 6-factor
102 solution, where the bottom panels show the distribution of Q/Qexp as a function of m/z and time. The
103 mass spectra (MS) of the 6-factors as output by PMF are shown in Figure S4 as well. It is useful to
104 note that although the total Q/Qexp for the 6-factor solution is close to the “ideal” value of 1, the
105 distribution of Q/Qexp shows large variability. This likely reflects the fact that the PMF assumption of
106 constant factor profiles is not strictly followed for the species studied here due to temporal variations
107 in the distribution of gas phase species with various ambient parameters including oxidation
108 conditions, temperature, and relative humidity. The trend in Q/Qexp with factor number can still
109 provide useful qualitative information for selection of the optimum number of factors. PMF factor
110 exploration showed that the addition of factors beyond 6 did not decrease the residuals or Q/Qexp
111 significantly, meaning that most of the data variability can be explained by these six factors (see
112 Figure S5 for source allocation from 2 to 10 factors).





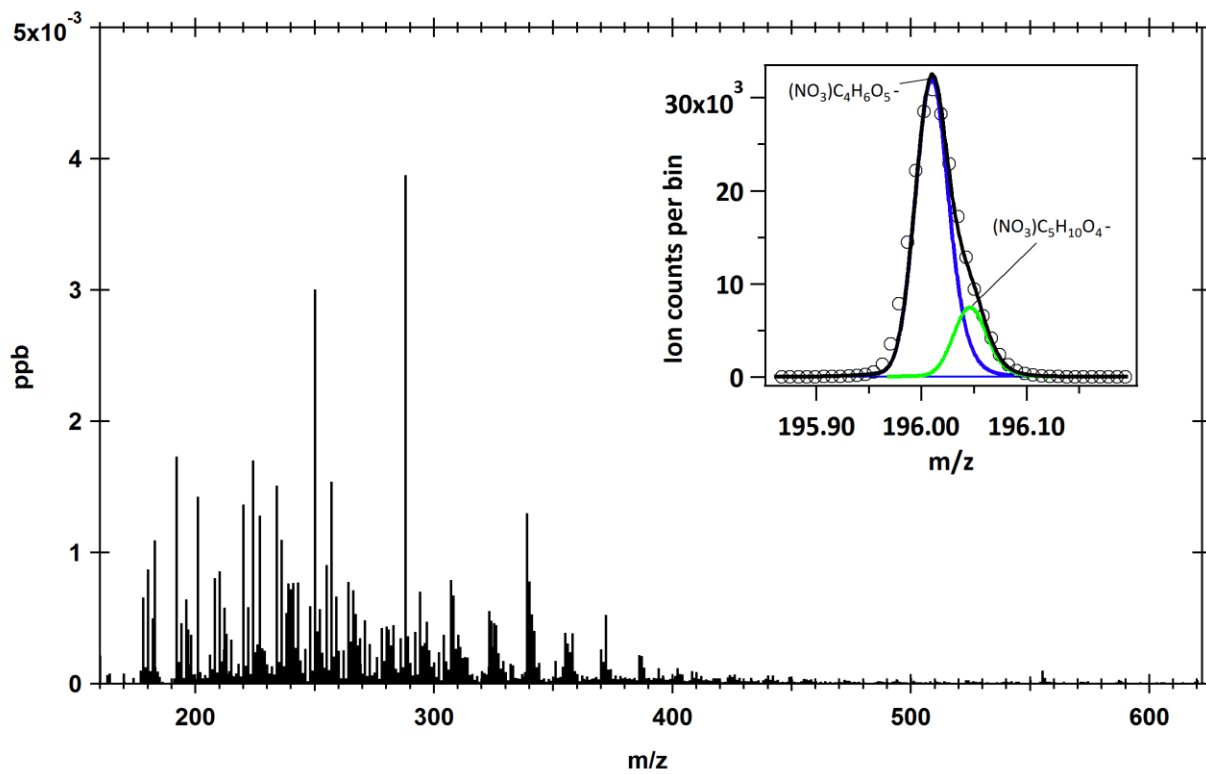
117

118 Figure S1: Results of laboratory experiments conducted to obtain transmission corrections for the NO_3^-
 119 CIMS data. Panel a): mass spectrum showing ion signals from perfluorinated acids; clusters and
 120 signals from deprotonated acids are shown. Panel b): time series for reagent ions and ions related to
 121 perfluoroheptanoic acid ($\text{C}_7\text{HF}_{13}\text{O}_2$) during transmission experiments. Panels c) and d): relative
 122 transmission values from matrix inversion and functional dependence methods, respectively.

123

124

125



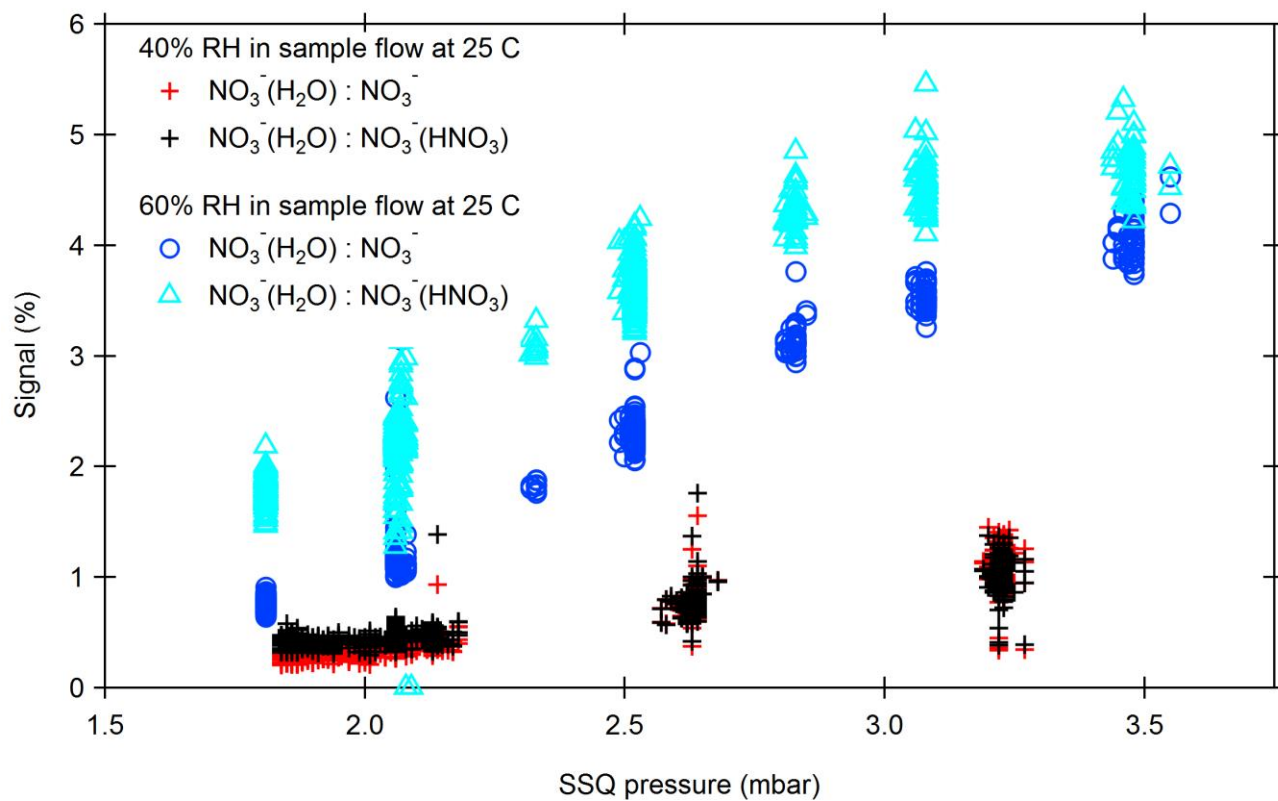
126

127 Figure S2: Campaign average high resolution mass spectrum with an example of molecular separation
128 using the Tofware capabilities. The mass spectral data are corrected from instrument transmission.

129

130

131



132

133 Figure S3: Percentage change of the ratio between the water cluster $\text{NO}_3^-(\text{H}_2\text{O})$, m/z 80, and the nitrate
 134 reagent ions NO_3^- , m/z 62, and $\text{NO}_3^-(\text{HNO}_3)$, m/z 125, with respect to the pressure in the first-stage
 135 small quadrupole (SSQ). The ratio depends on SSQ pressure and changes faster at higher RH.

136

137

138

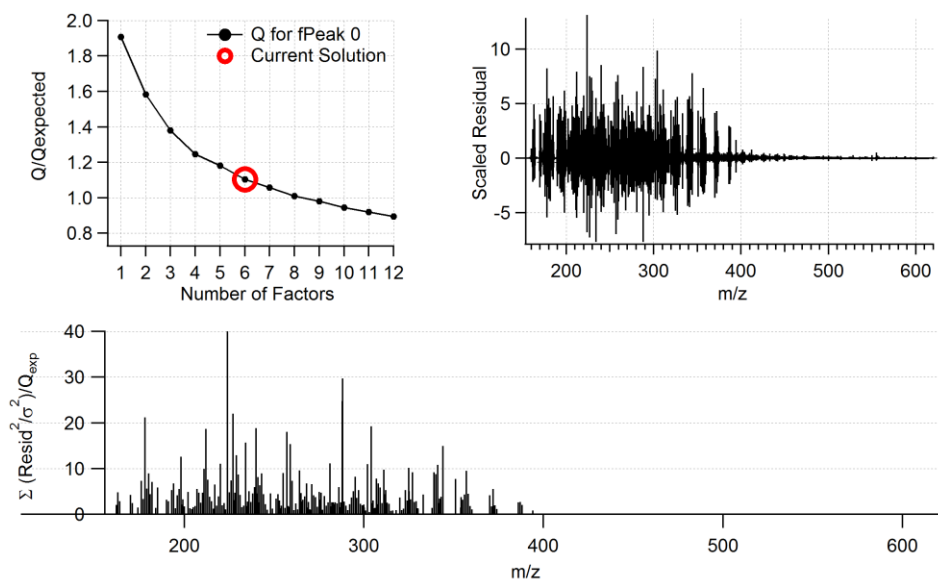
139

140

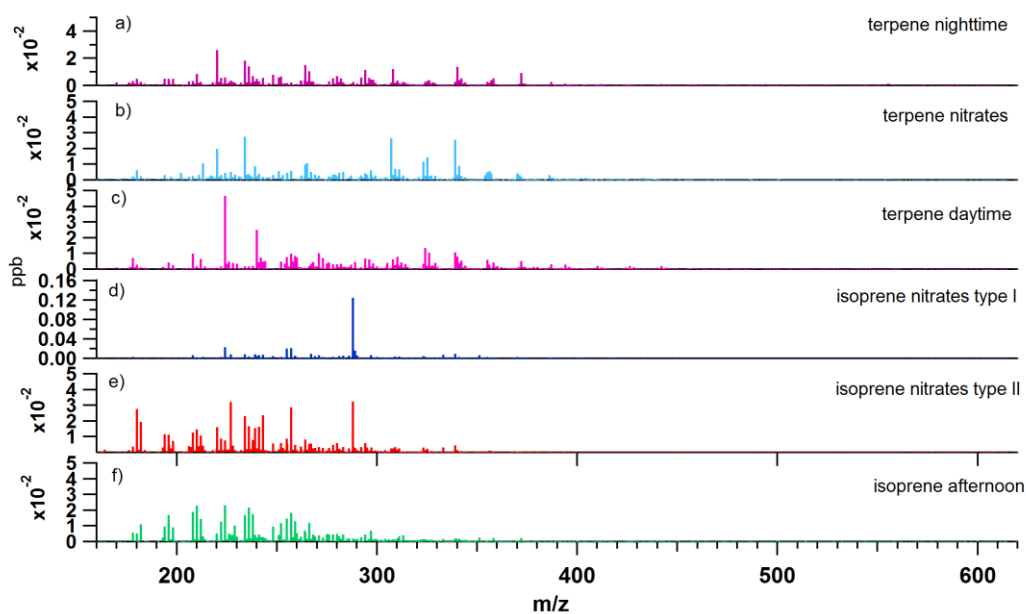
141

142

143



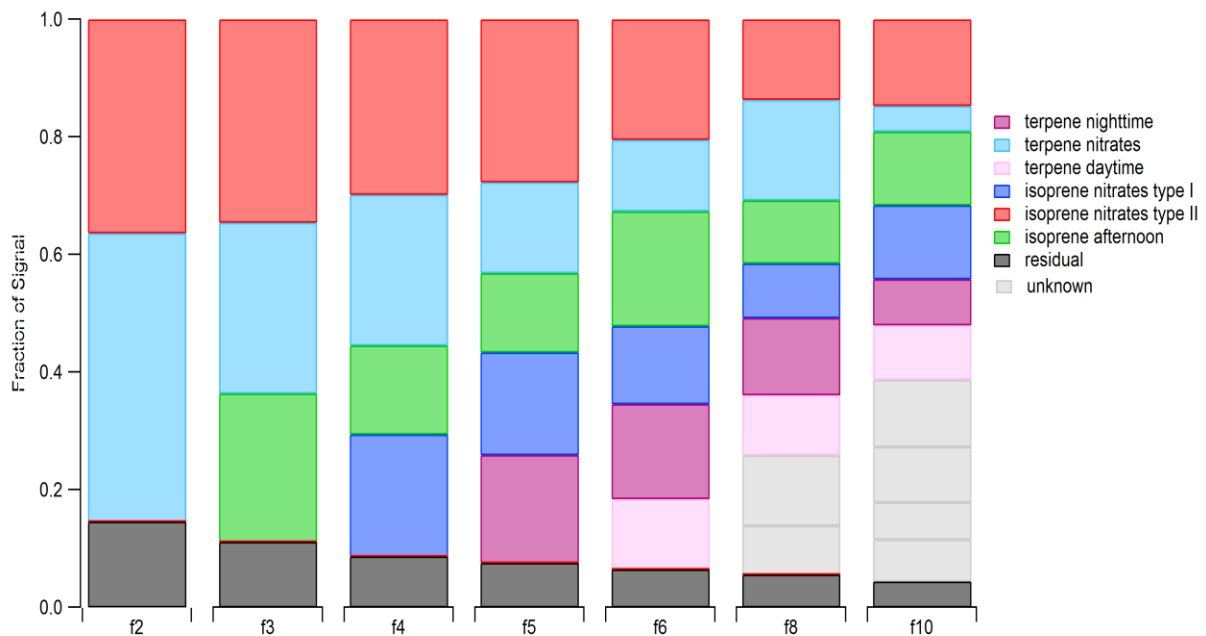
144



145

146 Figure S4: Upper panel: Summary of key diagnostic plots (Q/Q_{exp} , scaled residuals and scaled
 147 residuals over Q_{exp} for each m/z) for the PMF analysis performed on the NO_3^- CIMS data collected
 148 during SOAS 2013. A 6-factor solution yielding $Q/Q_{\text{exp}} = 1.1$ was chosen. Lower panel: Mass spectra
 149 (MS) of the 6 PMF factors before splitting the MS into C_n families, labeled a) to f) from top to bottom.
 150 Panels a) to c) are the MS for the terpene-related factors, and panels d) to f) are the MS for the
 151 isoprene-related factors.

152



153

154 Figure S5: Source allocation from 2-10 factors PMF solution. We chose the 6-factor solution for this
 155 dataset because it gave the best combination of number of factors that could be explained by chemical
 156 processes and low residual. The grey sections indicated the additional factors that split at higher
 157 solutions. Residuals are shown in black.

158

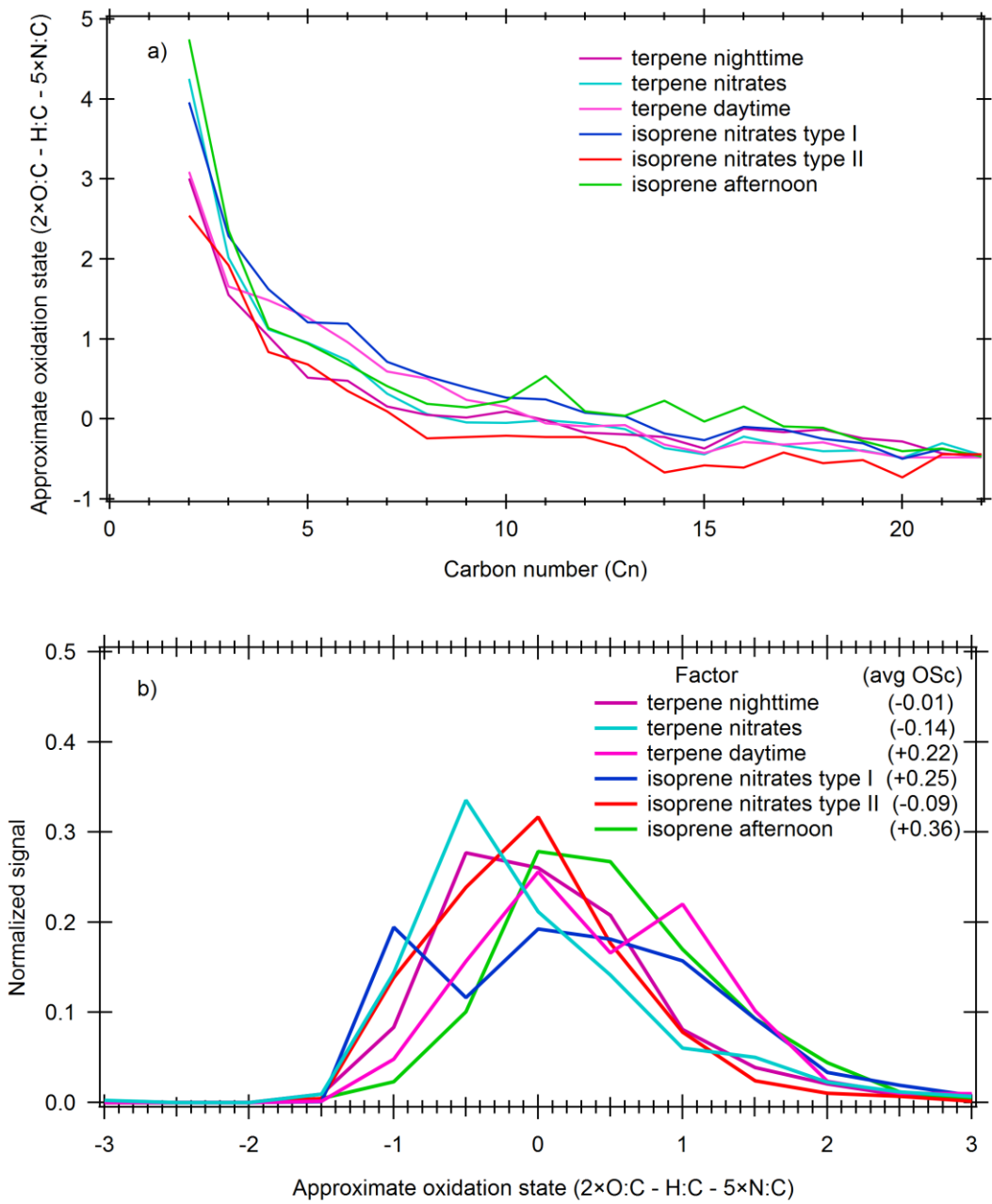
159

160

161

162

163



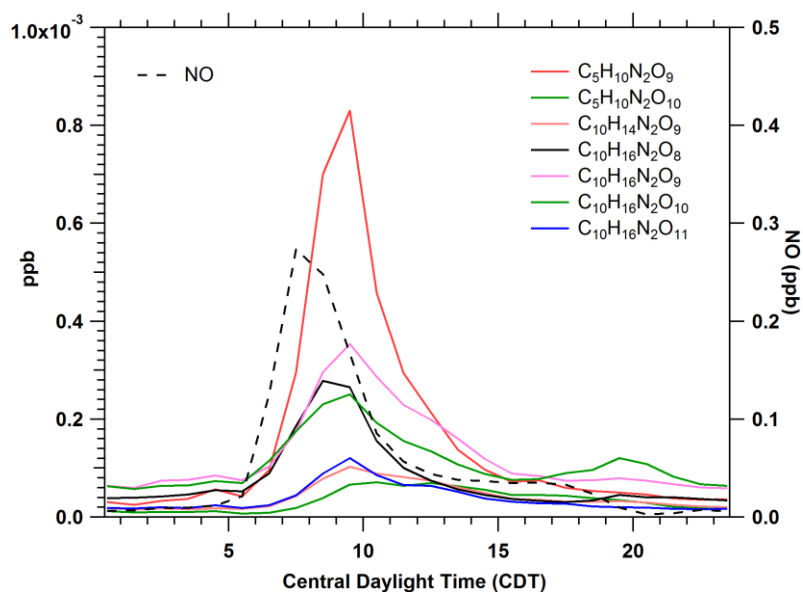
164

165 Figure S6: Panel a) shows the approximate oxidation state (OSc) for each factor as a function of the
 166 carbon number Cn, while panel b) shows the normalized factor signal as a function of OSc. The
 167 average OSc for each factor is also reported in the legend.

168

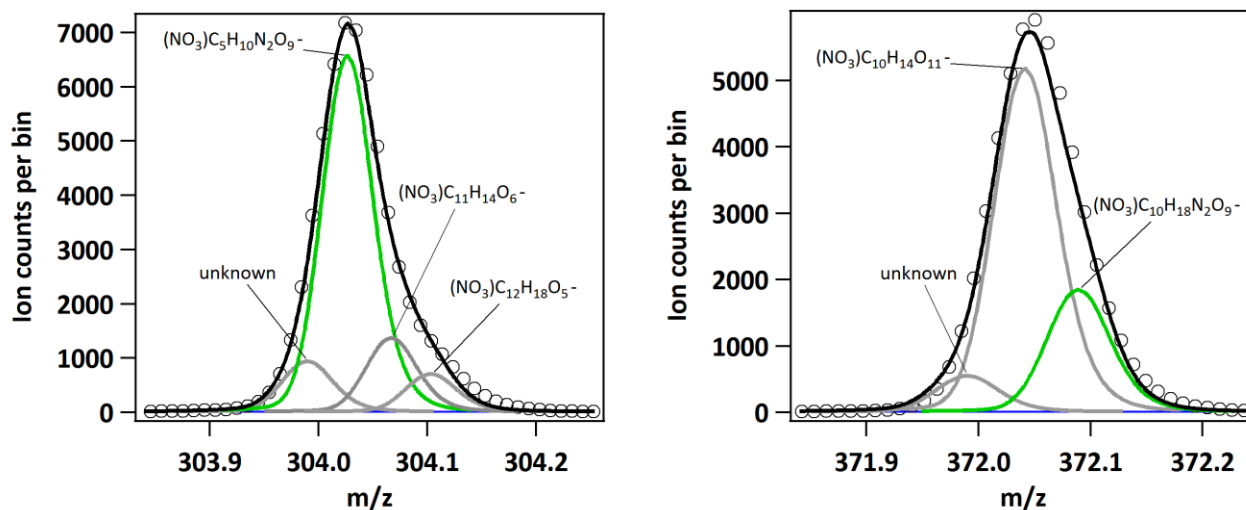
169

170



171

172

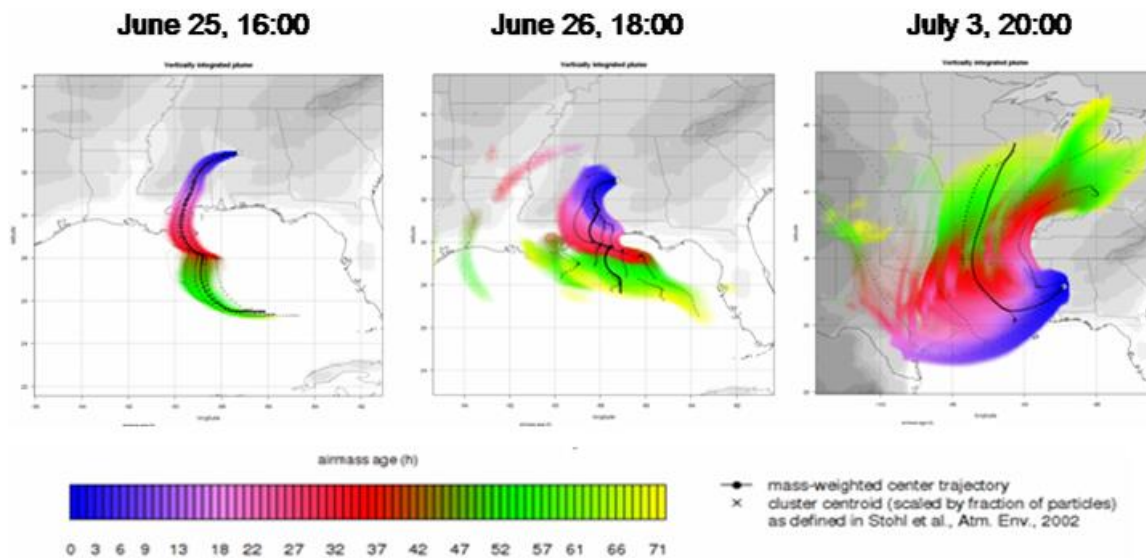


173

174 Figure S7: Top panel: diurnal cycles of C₅ isoprene dinitrate and C₁₀ monoterpene dinitrate ions. All
175 dinitrate ions peak at ~1000 CDT, approximately 2 hours after the morning NO peak. The clustering
176 reagent ion NO₃⁻ was omitted from the formulas to make the labels more readable. Bottom panel: high
177 resolution fits of *m/z* 304 and *m/z* 372, where a C₅ dinitrate (C₅H₁₀N₂O₉) and a C₁₀ dinitrate
178 (C₁₀H₁₈N₂O₉) are respectively detected.

179

180



181

182

183 Figure S8: FLEXPART back-trajectories of the air masses reaching the CTR site during June 25-26
184 and July 3, when high levels of SO₂ were observed. These air masses most likely carried emissions
185 from coal fired power plants to the site.

186

187

188

189

190

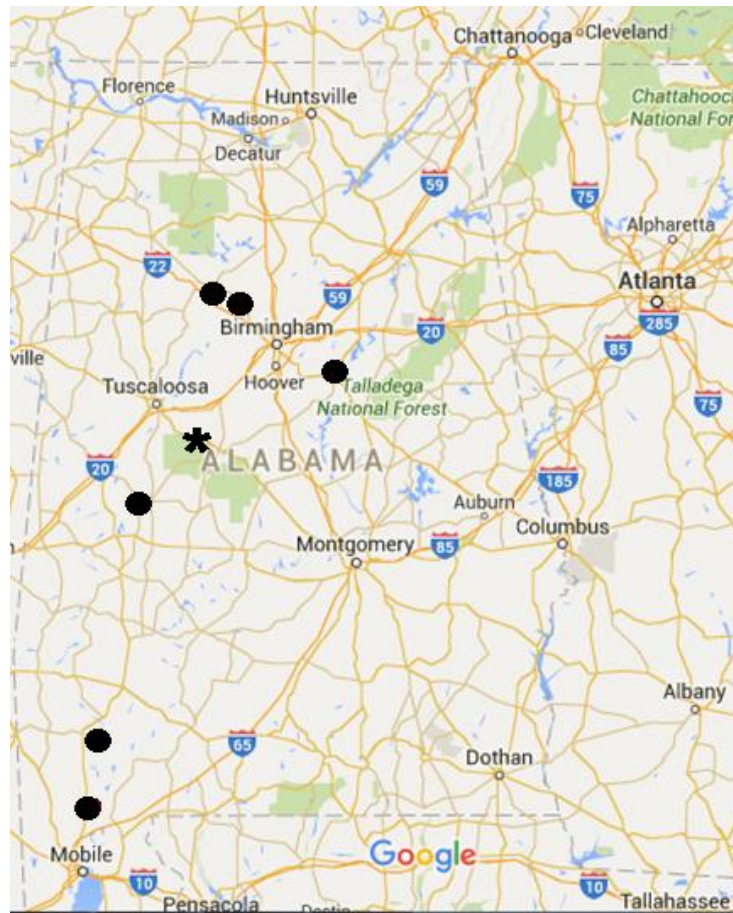
191

192

193

194

195



196

197

198 Figure S9: Map of Alabama with major coal-fired power plants (black circles) and location of
199 Centreville, where the CTR and SEARCH sites were located (black star)

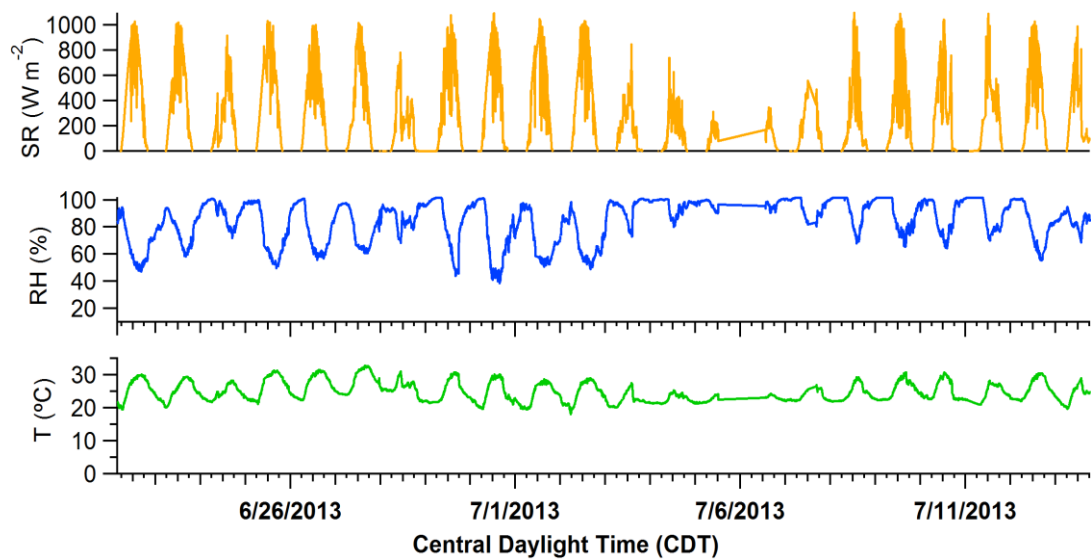
200

201

202

203

204



205

206

207 Figure S10: Temporal series of temperature (T), relative humidity (RH) and solar radiation (SR) at the
208 CTR site during SOAS 2013.

209

210

211

212

213

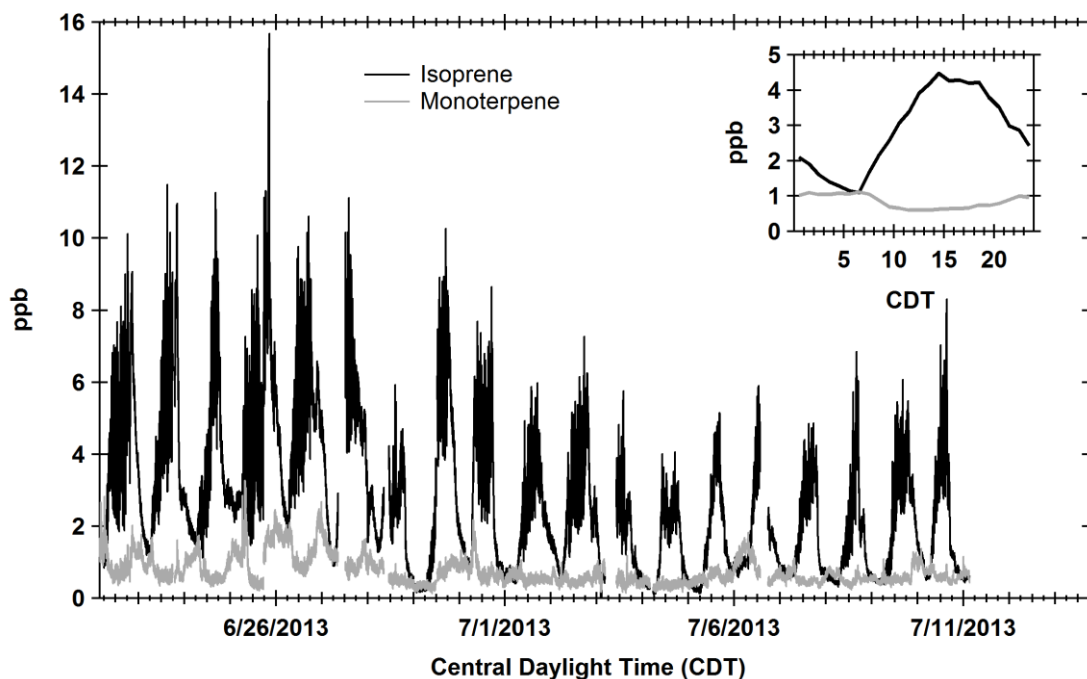
214

215

216

217

218



219

220 Figure S11: Time series of the monoterpene and isoprene signal (ppb) as measured by PTR-TOF-MS
 221 along with their diurnal cycles (insert). The monoterpene had small diurnal oscillations and were
 222 higher at nighttime while the isoprene signal reached higher daytime values, with a peak at 15:00. The
 223 campaign average mass loadings of monoterpene and isoprene precursors were 0.85 and 2.8 ppb,
 224 respectively.

225

226

227

228

229

230

231

232

233 References

- 234 (1) Huey, L. G.; Hanson, D. R.; Howard, C. J. Reactions of SF₆⁻ and I⁻ with atmospheric trace gases.
235 *J. Phys. Chem.* **1995**, 99(14), 5001–5008, DOI: 10.1021/j100014a021.
- 236 (2) Heinritzi, M.; Simon, M.; Steiner, G.; Wagner, A. C.; Kürten, A.; Hansel, A.; Curtius, J.
237 Characterization of the mass-dependent transmission efficiency of a CIMS. *Atmos. Meas. Tech.* **2016**,
238 9, 1449-1460, DOI: 10.5194/amt-9-1449-2016.
- 239 (3) Yan, C.; Nie, W.; Äijälä, M.; Rissanen, M. P.; Canagaratna, M. R.; Massoli, P.; Junninen, H.;
240 Jokinen, T.; Sarnela, N.; Häme, S. A. K.; Schobesberger, S.; Canonaco, F.; Yao, L.; Prévôt, A. S. H.;
241 Petäjä, T.; Kulmala, M.; Sipilä, M.; Worsnop, D. R.; Ehn, M. Source characterization of highly
242 oxidized multifunctional compounds in a boreal forest environment using positive matrix
243 factorization. *Atmos. Chem. Phys.* **2016**, 16, 12715-12731, DOI: 10.5194/acp-16-12715-2016.
- 244 (4) Cubison, M.J., Jimenez, J.-L. Statistical precision of the intensities retrieved from constrained
245 fitting of overlapping peaks in high-resolution mass spectra. *Atmos. Meas. Tech.* **2015**, 8, 2333-2345,
246 DOI: 10.5194/amt-8-2333-2015.
- 247 (5) Ulbrich, I. M.; Canagaratna, M. R.; Zhang, Q.; Worsnop, D. R.; and Jimenez, J. -L. Interpretation
248 of organic components from Positive Matrix Factorization of aerosol mass spectrometric data. *Atmos.*
249 *Chem. Phys.* **2009**, 9, 2891-2918, DOI: 10.5194/acp-9-2891-2009.
- 250 (6) Paatero, P. Least squares formulation of robust non-negative factor analysis. *Chemometr. Intell.*
251 *Lab.* **1997**, 37, 23-35.
- 252 (7) Paatero, P.; Hopke, P.K. Discarding or downweighting high-noise variables in factor analytic
253 models. *Anal. Chim. Acta.* **2003**, 490,277-289, DOI: 10.1016/s0003-2670(02)01643-4.