Supporting Information of

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

Paola Massoli[1], Harald Stark[1,2,3], Manjula R. Canagaratna[1], Jordan E. Krechmer[1,2,3], Lu Xu[5*], Nga L. Ng[4,5], Roy L. Mauldin III[6], Chao Yan[7], Joel Kimmel[1,8], Pawel K. Misztal[9,10,**], Jose L. Jimenez[2,3], John T. Jayne[1] and Douglas R. Worsnop[1]

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

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

[3] Department of Chemistry, University of Colorado, Boulder, CO 80309

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

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

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

[7] Department of Physics, University of Helsinki, FI-00560, Helsinki, Finland

[8] Tofwerk AG, CH-3600, Thun, Switzerland

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

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

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

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

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

1. Relative transmission measurements

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

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

where Constant is the constant signal across time series

C1, … CN: fit coefficient

I1(t), … IN(t): ion signal time series

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

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

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

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

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

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

$$X=GF+E$$

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

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

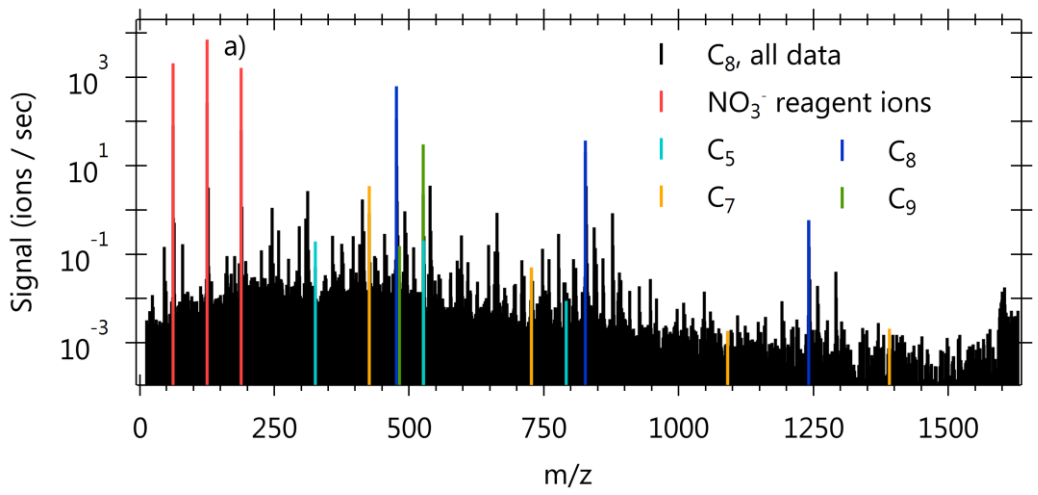
where E_{ij} refers to elements of the residual matrix and σ_{ij} refers to the standard deviation of each data point in the X matrix (X_{ij}). In the limit where the residuals of data points are equal to their standard deviations, each data point contributes a value of 1 to the total Q and the expected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

description of how to calculate the standard deviation values (σ_{ij}) for PMF analysis of unit-mass resolution CIMS data.[3] In this work, PMF analysis was performed on the high resolution dataset in order to utilize the additional chemical information in the mass spectra. Estimation of the uncertainties associated with high-resolution fitting, in which ion intensities are obtained from overlapping peaks, are complicated[4], and methods for approximating them have not yet been fully developed. Thus σ_{ij} values are estimated according to Poisson counting statistics, the dominant source of noise in the unit-mass resolution data[3], as follows:

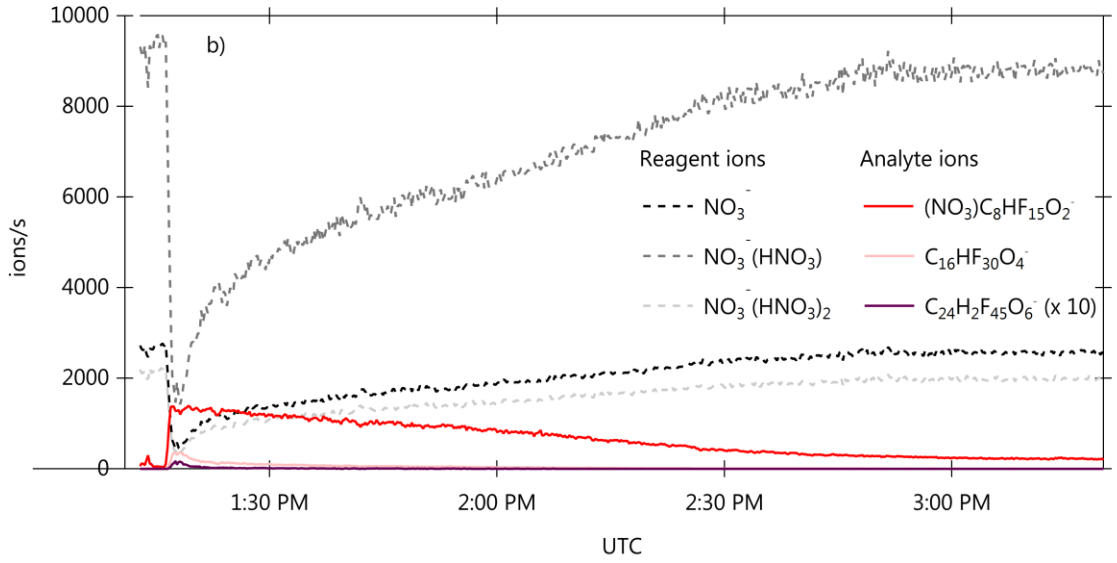
$$\sigma_{ij} = (I/t_s)^{0.5}$$

Where I is the raw ion signal in ions/s, obtained before corrections for ion transmission and extraction are applied, and t_s is the integration time in seconds. We note that the σ_{ij} values estimated using Poisson statistics provide a lower limit for the real noise in the high-resolution data. The PMF analysis in this work utilized the PMF Evaluation Tool (PET)[5] together with the PMF2 algorithm.[6] The signal-to-noise ratio (SNR) for each data point was calculated according to previous studies[5,7] and "bad" variables with SNR < 0.2 were down-weighted by 10 while "weak" variables with SNR between 0.2 and 2 were down-weighted by 3. Figure S4 shows the key diagnostic plots for the Positive Matrix Factorization (PMF) performed on the SOAS 2013 NO_3^- CIMS dataset. The Q/Qexp is shown as a function of the number of factors P (top panel, left). A 6-factor solution (P = 6), yielding a Q/Qexp of 1.1 was chosen to explain this dataset. The top right panel shows the total residuals for the 6-factor solution, where the bottom panels show the distribution of Q/Qexp as a function of *m/z* and time. The mass spectra (MS) of the 6-factors as output by PMF are shown in Figure S4 as well. It is useful to note that although the total Q/Qexp for the 6-factor solution is close to the "ideal" value of 1, the distribution of Q/Qexp shows large variability. This likely reflects the fact that the PMF assumption of constant factor profiles is not strictly followed for the species studied here due to temporal variations in the distribution of gas phase species with various ambient parameters including oxidation conditions, temperature, and relative humidity. The trend in Q/Qexp with factor number can still provide useful qualitative information for selection of the optimum number of factors. PMF factor exploration showed that the addition of factors beyond 6 did not decrease the residuals or Q/Qexp significantly, meaning that most of the data variability can be explained by these six factors (see Figure S5 for source allocation from 2 to 10 factors).

113 Figures



114



115

116

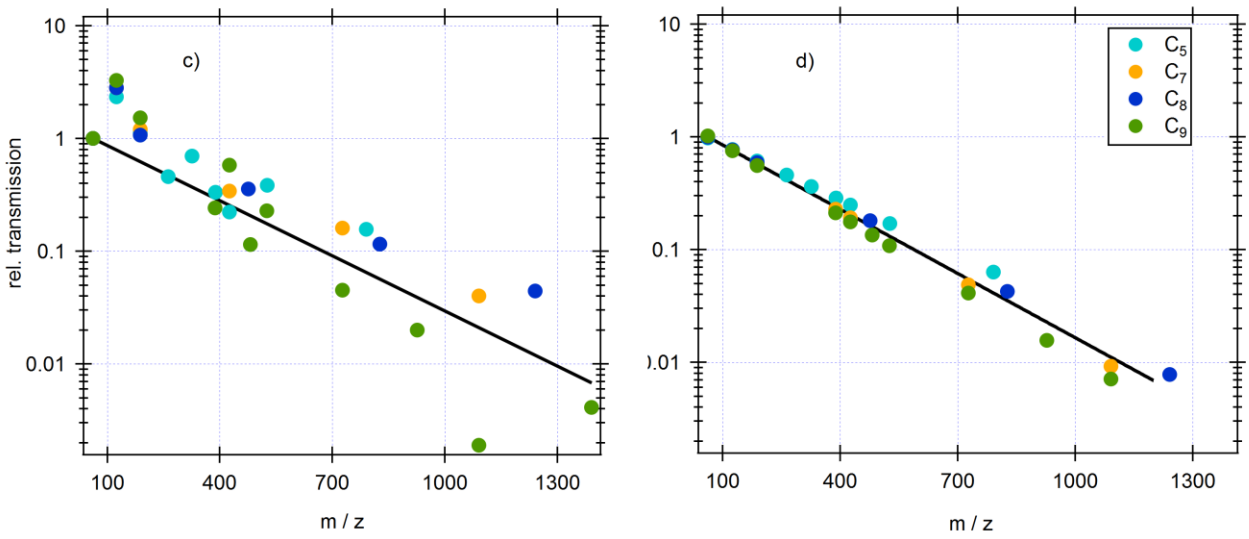
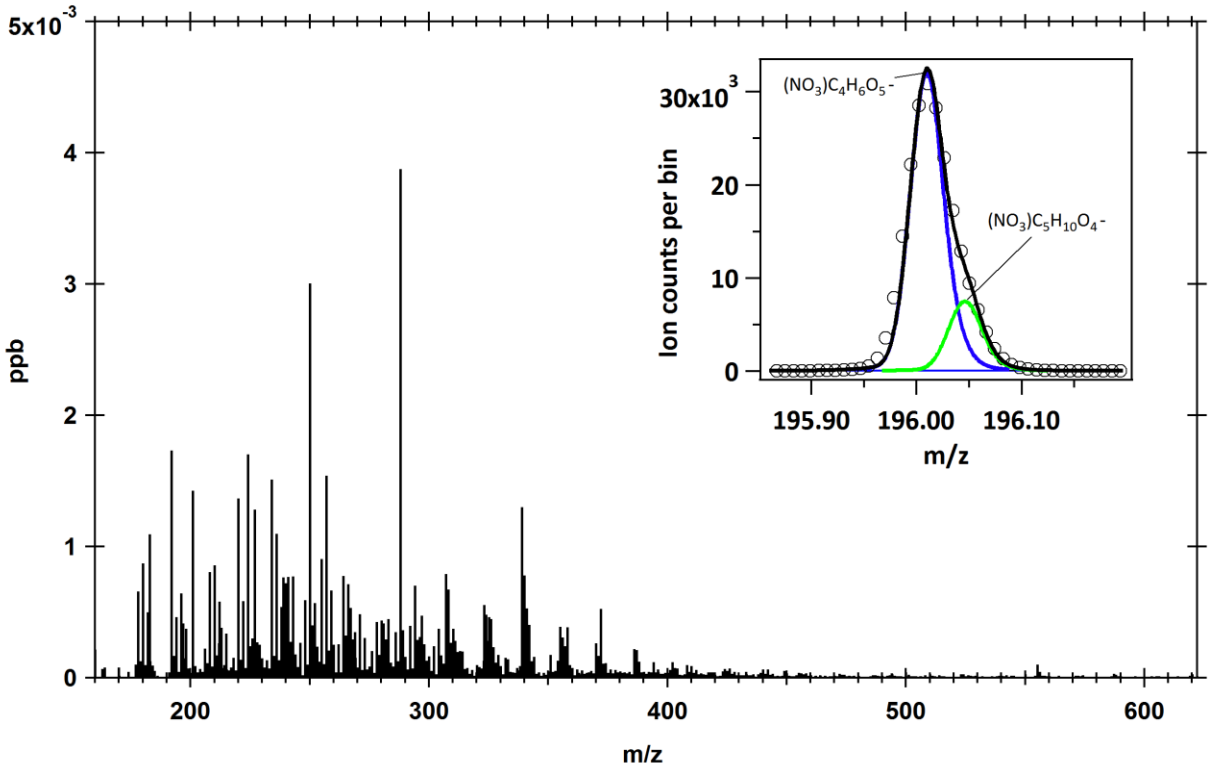


Figure S1: Results of laboratory experiments conducted to obtain transmission corrections for the NO_3^- CIMS data. Panel a): mass spectrum showing ion signals from perfluorinated acids; clusters and signals from deprotonated acids are shown. Panel b): time series for reagent ions and ions related to perfluoroheptanoic acid ($C_7HF_{13}O_2$) during transmission experiments. Panels c) and d): relative transmission values from matrix inversion and functional dependence methods, respectively.



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

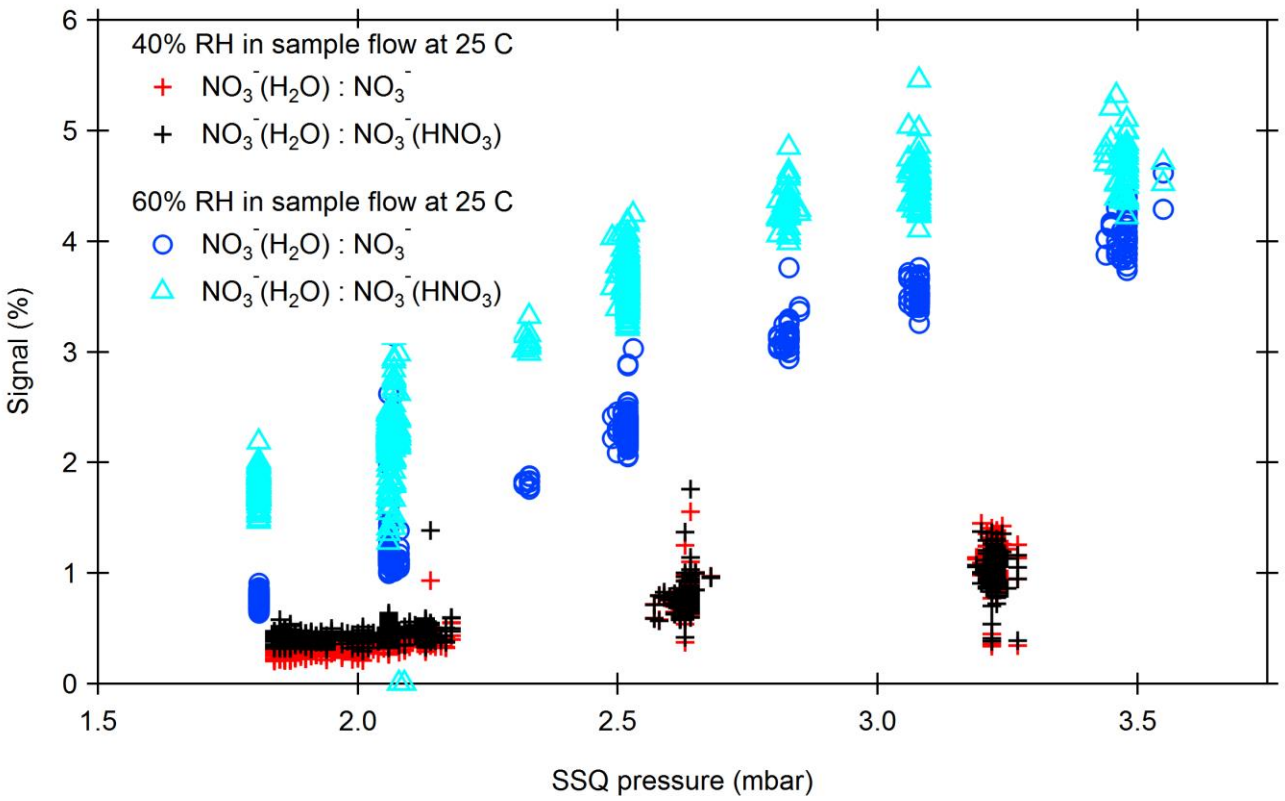
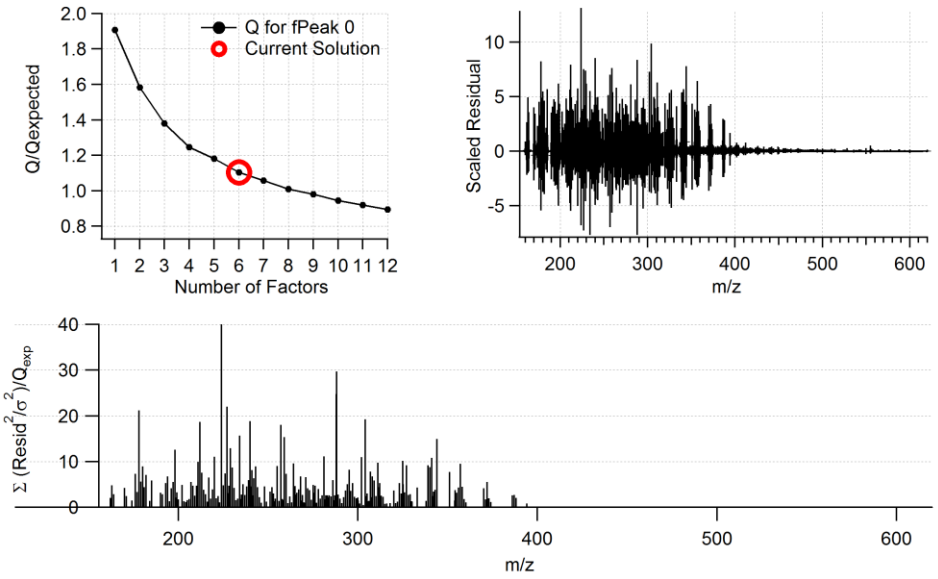
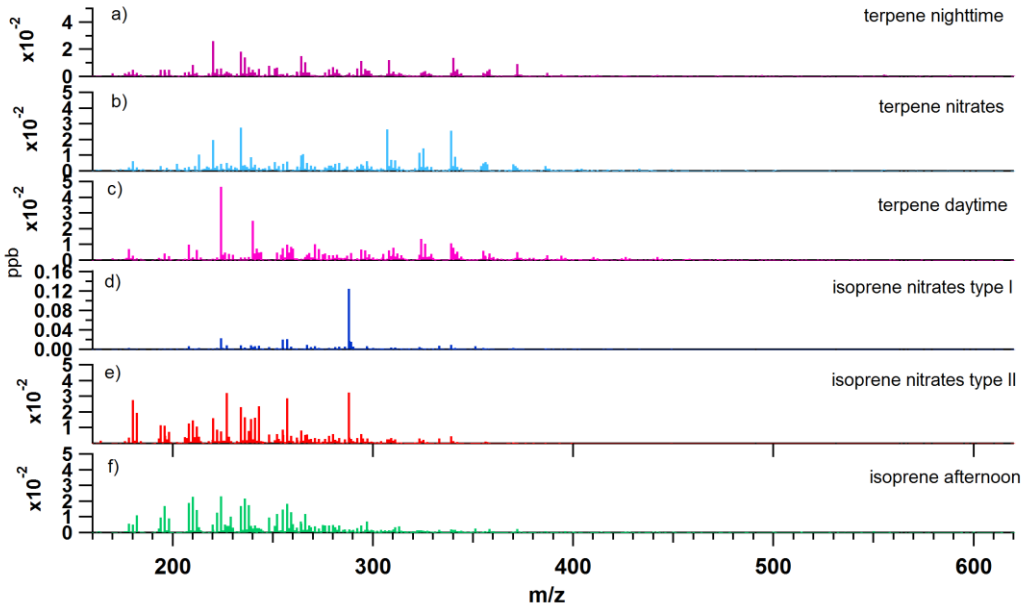


Figure S3: Percentage change of the ratio between the water cluster $NO_3^-(H_2O)$, *m/z* 80, and the nitrate reagent ions NO_3^-, *m/z* 62, and $NO_3^-(HNO_3)$, *m/z* 125, with respect to the pressure in the first-stage small quadrupole (SSQ). The ratio depends on SSQ pressure and changes faster at higher RH.



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_{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

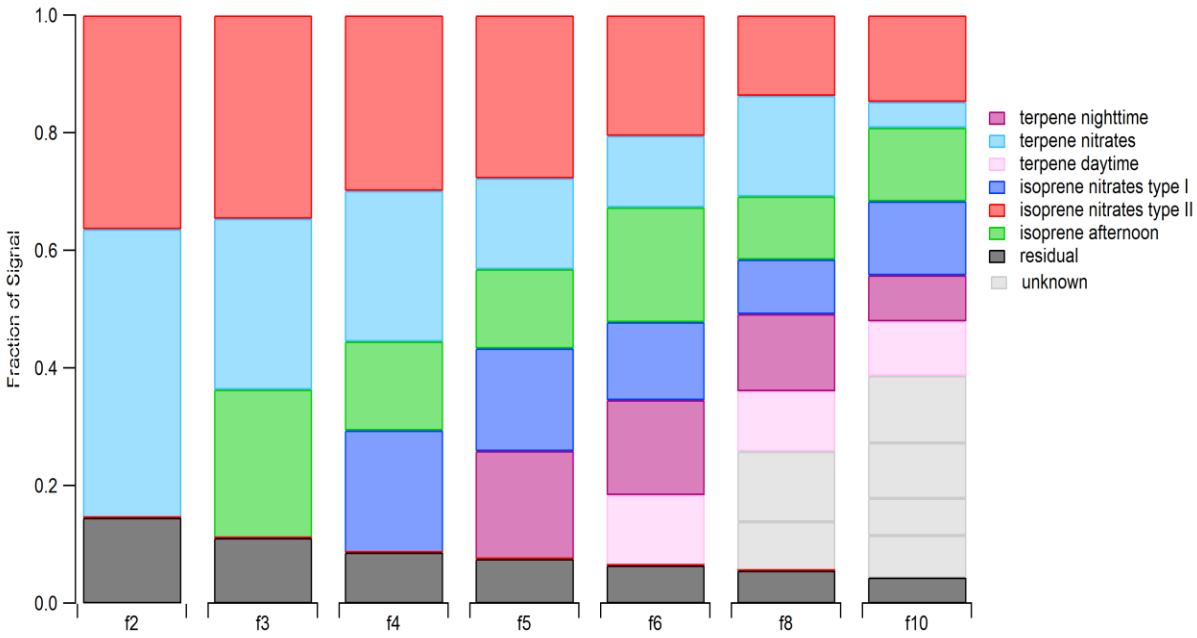
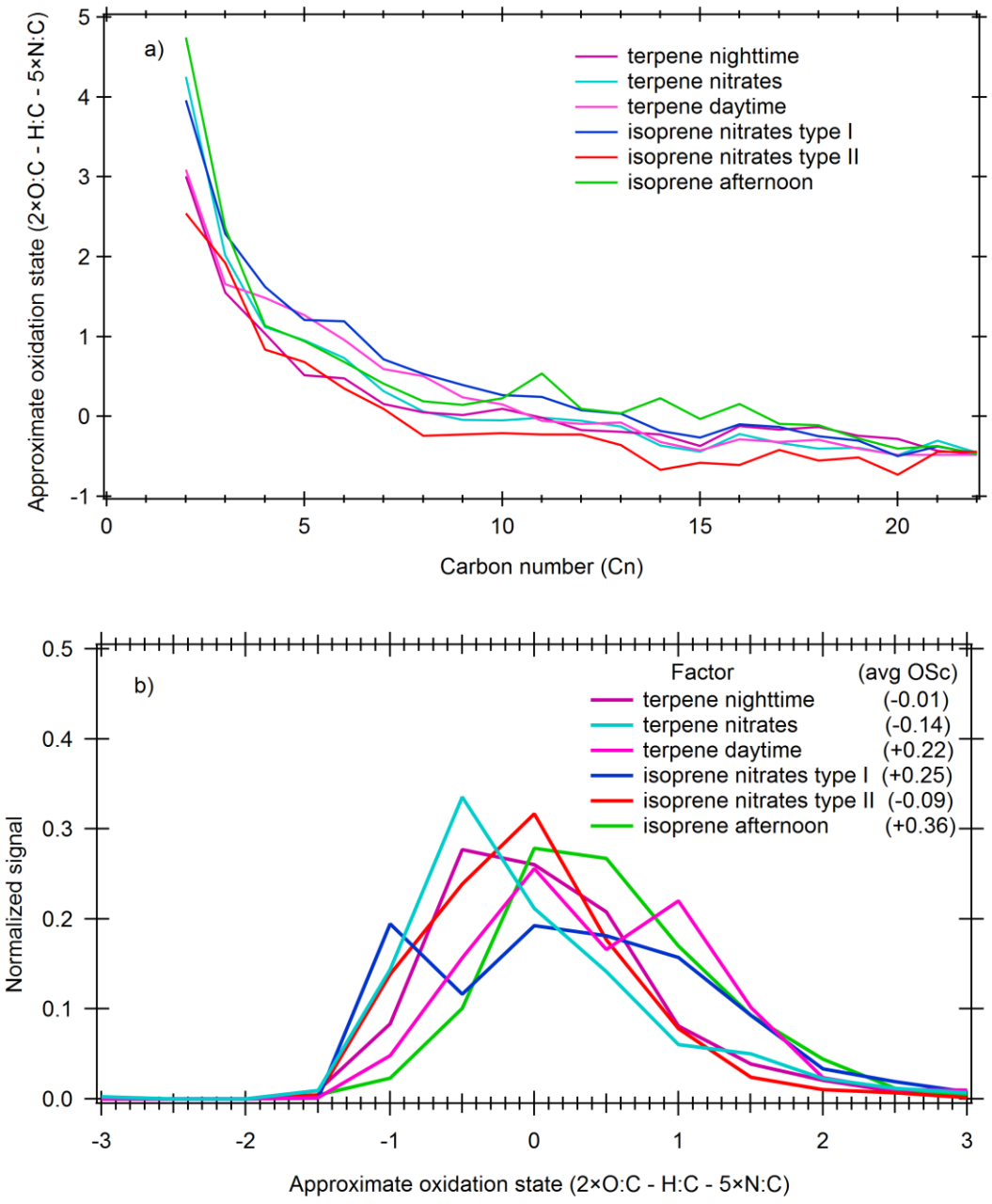


Figure S5: Source allocation from 2-10 factors PMF solution. We chose the 6-factor solution for this dataset because it gave the best combination of number of factors that could be explained by chemical processes and low residual. The grey sections indicated the additional factors that split at higher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

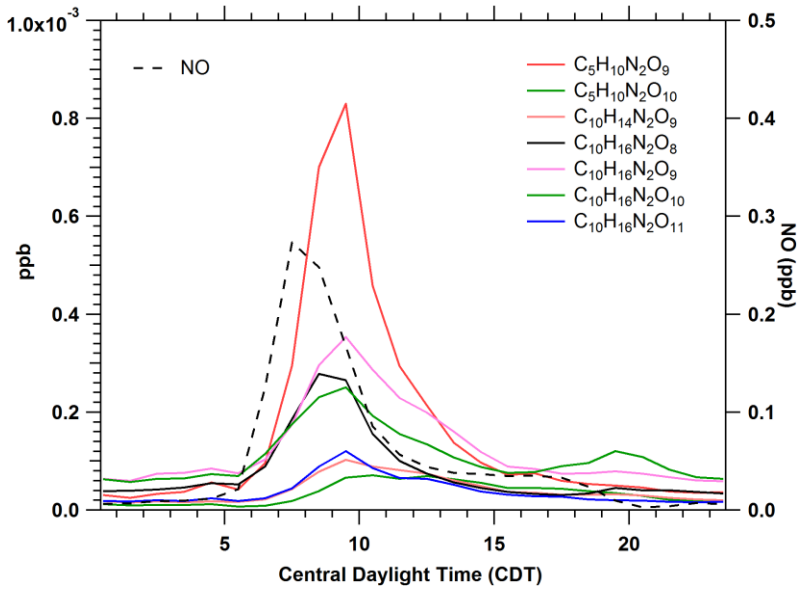


164

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



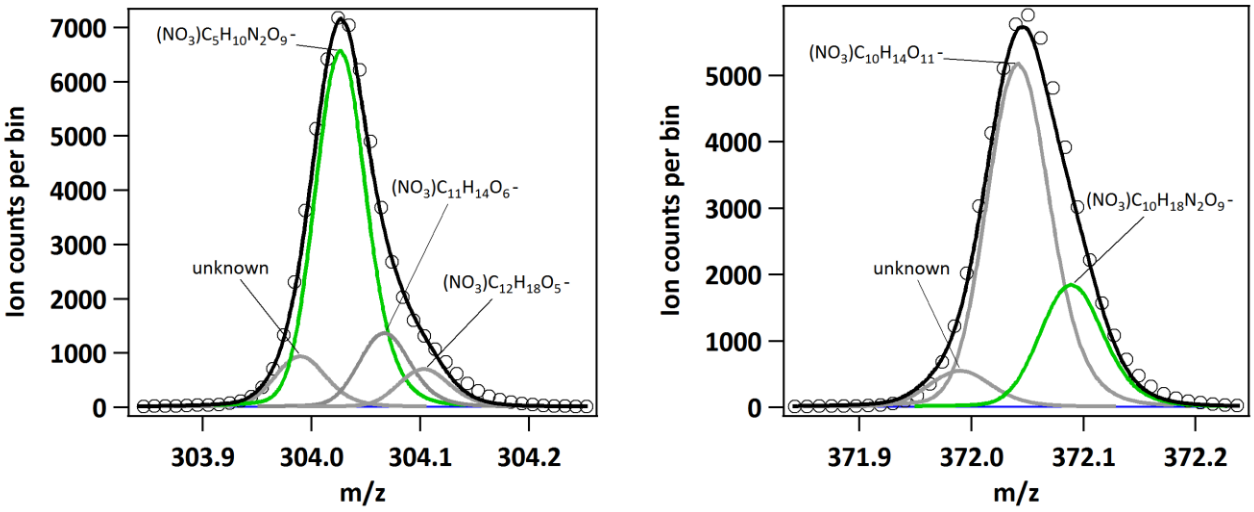


Figure S7: Top panel: diurnal cycles of C_5 isoprene dinitrate and C_{10} monoterpene dinitrate ions. All dinitrate ions peak at ~1000 CDT, approximately 2 hours after the morning NO peak. The clustering reagent ion NO_3^- was omitted from the formulas to make the labels more readable. Bottom panel: high resolution fits of *m/z* 304 and *m/z* 372, where a C_5 dinitrate ($C_5H_{10}N_2O_{9,}$) and a C_{10} dinitrate ($C_{10}H_{18}N_2O_9$) are respectively detected.

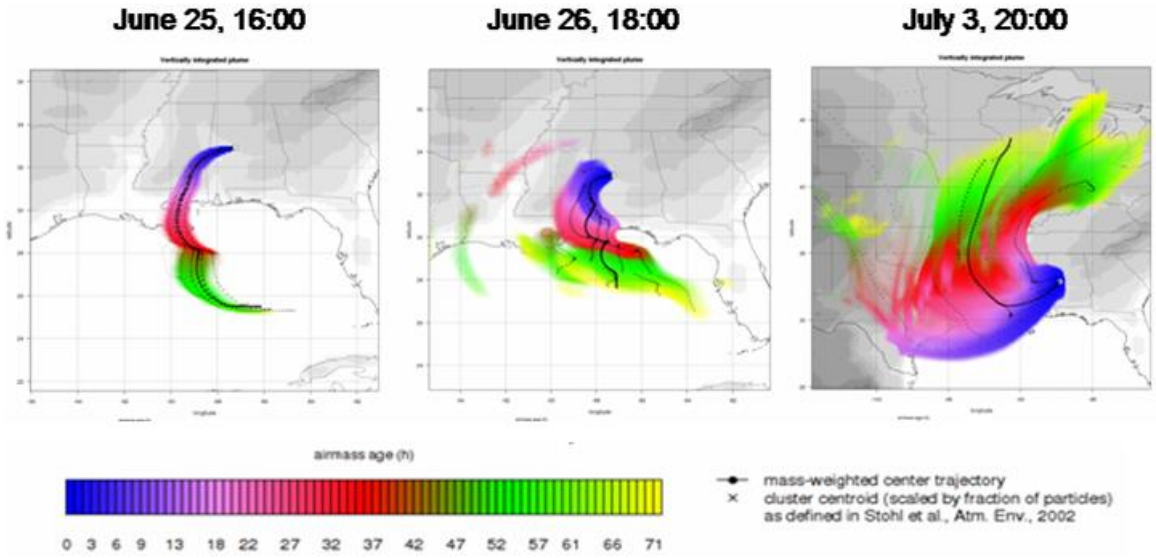


Figure S8: FLEXPART back-trajectories of the air masses reaching the CTR site during June 25-26 and July 3, when high levels of SO_2 were observed. These air masses most likely carried emissions from coal fired power plants to the site.

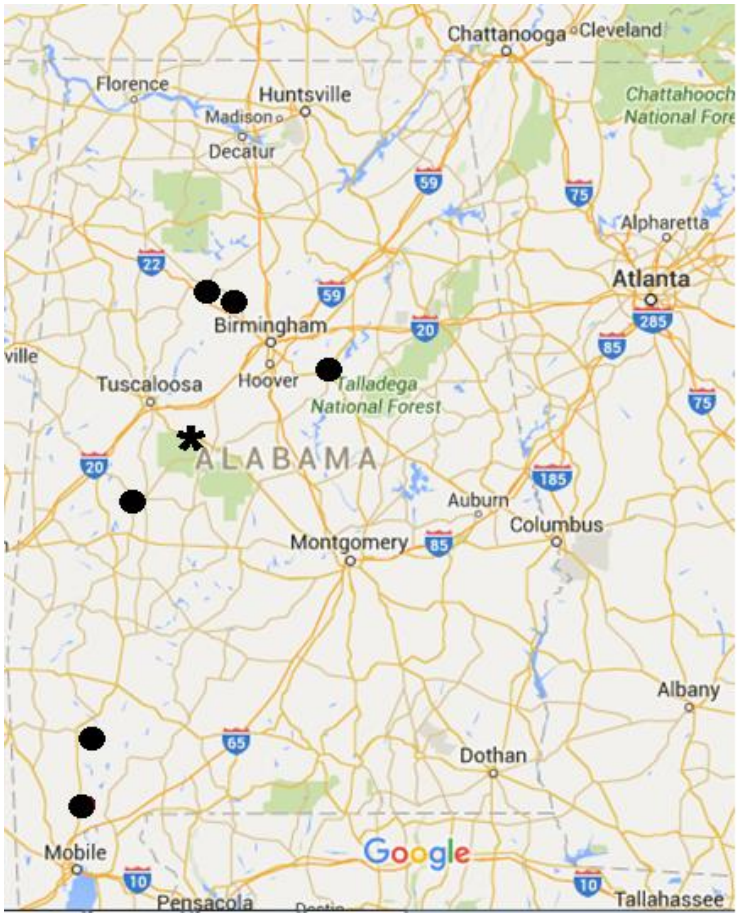


Figure S9: Map of Alabama with major coal-fired power plants (black circles) and location of Centreville, where the CTR and SEARCH sites were located (black star)

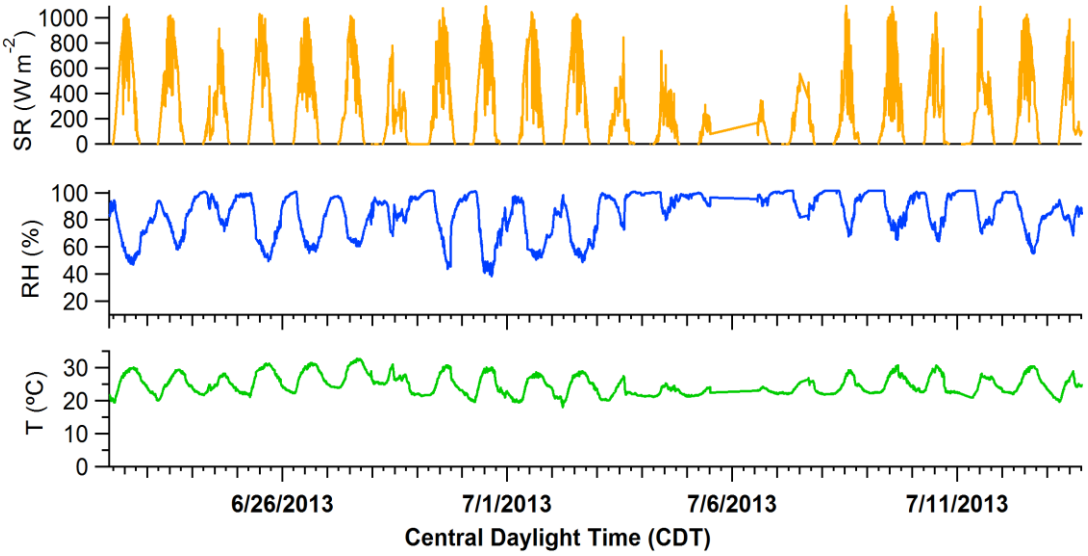


Figure S10: Temporal series of temperature (T), relative humidity (RH) and solar radiation (SR) at the CTR site during SOAS 2013.

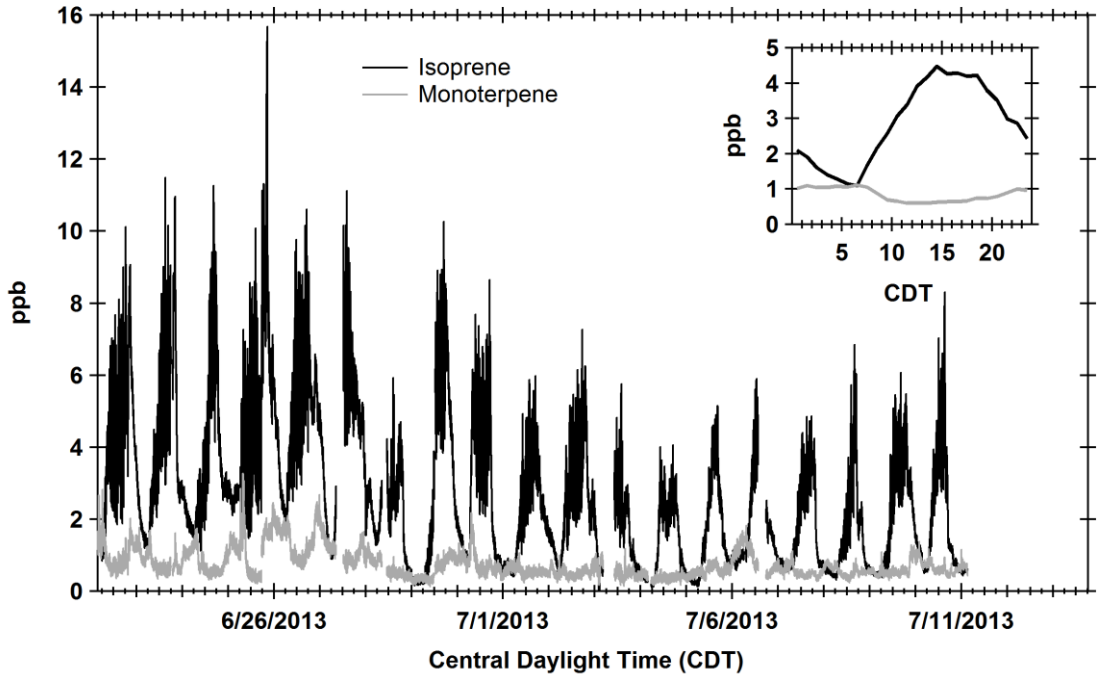


Figure S11: Time series of the monoterpene and isoprene signal (ppb) as measured by PTR-TOF-MS along with their diurnal cycles (insert). The monoterpene had small diurnal oscillations and were higher at nighttime while the isoprene signal reached higher daytime values, with a peak at 15:00. The campaign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

References

(1) Huey, L. G.; Hanson, D. R.; Howard, C. J. Reactions of SF_6- and I- with atmospheric trace gases. *J. Phys. Chem*. **1995**, 99(14), 5001–5008, DOI: 10.1021/j100014a021.

(2) Heinritzi, M.; Simon, M.; Steiner, G.; Wagner, A. C.; Kürten, A.; Hansel, A.; Curtius, J. Characterization of the mass-dependent transmission efficiency of a CIMS. *Atmos. Meas. Tech.* **2016**, 9, 1449-1460, DOI: 10.5194/amt-9-1449-2016.

(3) 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.; Ehn, M. Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization. *Atmos. Chem. Phys*. **2016**, 16, 12715-12731, DOI: 10.5194/acp-16-12715-2016.

(4) Cubison, M.J., Jimenez, J.-L. Statistical precision of the intensities retrieved from constrained fitting of overlapping peaks in high-resolution mass spectra. *Atmos. Meas. Tech*. **2015**, 8, 2333-2345, DOI: 10.5194/amt-8-2333-2015.

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

(6) Paatero, P. Least squares formulation of robust non-negative factor analysis. *Chemometr. Intell. Lab*. **1997**, 37, 23-35.

(7) Paatero, P.; Hopke, P.K. Discarding or downweighting high-noise variables in factor analytic models. *Anal. Chim. Acta.* **2003**, 490,277-289, DOI: 10.1016/s0003-2670(02)01643-4.