

Unambiguous identification of N-containing oxygenated organic molecules using CI-Orbitrap in an eastern Chinese megacity

Yiqun Lu^{1,2}, Yingge Ma¹, Dan Dan Huang¹, Shengrong Lou¹, Sheng'ao Jing¹, Yaqin Gao¹, Hongli Wang¹, Yanjun Zhang³, Hui Chen⁴, Naiqiang Yan², Jianmin Chen⁴, Christian George³, Matthieu Riva³, Cheng Huang^{1*}

¹ State Environmental Protection Key Laboratory of Formation and Prevention of Urban Air Pollution Complex, Shanghai Academy of Environmental Sciences, Shanghai 200233, China;

² School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

³ Univ. Lyon, Université Claude Bernard Lyon1, CNRS, IRCELYON, 69626 Villeurbanne, France;

⁴ Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention (LAP³), Department of Environmental Science & Engineering, Jiangwan Campus, Fudan University, Shanghai 200438, China

Contents of this file

Section S1 to S3

Figures S1 to S8

Table S1 to S2

Figure S1. The map of the field site (Shanghai Academy of Environmental Sciences) that is a representative urban station.

Figure S2. Timeseries of key measurements during the field campaign.

Figure S3. Mathematical diagnostics of PMF solutions, including the overall changes of Q/Q_{exp} and the explained variation from two-factor to nine-factor solutions. For each number of factors, five seed runs were performed to test the consistency of the solution.

Figure S4. Timeseries of factors in 2-6 factor solutions of PMF. The panels from top to bottom are 2-factor solution, 3-factor solution, 4-factor solution, 5-factor solution and 6-factor solution, respectively.

Figure S5. Diel variation patterns of factors in 2-6 factor solutions of PMF. The panels from top to bottom are 2-factor solution, 3-factor solution, 4-factor solution, 5-factor solution and 6-factor solution, respectively.

Figure S6. The factor profiles in the six-factor solution.

Figure S7. The relative contributions of OOMs with different carbons to the extremely low-volatility organic compounds (ELVOC, $C^* < 3 \times 10^{-5} \mu\text{g m}^{-3}$) and low-volatility organic compounds (LVOC, $3 \times 10^{-5} \leq C^* < 3 \times 10^{-1} \mu\text{g m}^{-3}$).

Figure S8. Timeseries of N_2O_5 concentration (top panel), estimated NO_3 concentration (middle panel), and the nighttime factor-1 (bottom panel).

28 **Figure S9.** Timeseries of PM_{2.5} concentration (top panel), and the episode factor-1 (bottom panel).

29 **S1. Other ancillary measurements**

30 The mass concentration of ambient particles was measured by particle monitor (TEOM 1405DF, Thermo,
31 USA). SO₂, O₃ and NO_x concentrations were measured using a SO₂ analyzer (Model 43i, Thermo, USA), a O₃
32 analyzer (Model 49i, Thermo, USA) and a NO_x analyzer (Model 42i, Thermo, USA) with the detection limits of
33 0.1 ppbv, 0.5 ppbv and 0.4 ppbv, respectively. The above instruments were pre-calibrated before the campaign. The
34 solar radiation was measured on the rooftop of the building. Atmospheric N₂O₅ concentrations were measured by
35 an iodide CI-API-TOF. The concentrations of NO₃ radicals were estimated under the assumption that NO₃, NO₂ and
36 N₂O₅ could reach an equilibrium quickly in tropospheric conditions (Brown and Stutz, 2012). The total
37 concentrations of VOC precursors (TVOC) were determined by the measurement of an online GC-MS (7890A-
38 5975C, Agilent, USA).

39 **S2. Overview of the campaign**

40 An overview of the measurement data, illustrating the air quality as well as the meteorological
41 conditions (global radiation, temperature, wind direction, wind speed, and RH), concentrations of trace
42 gases and pollutants (PM_{2.5}, O₃, NO_x, N₂O₅, and TVOCs) during the campaign, is provided in this section
43 as shown in Figure S2 and Table S1. Firstly, the maximum intensities of global radiation on individual
44 days were in a range of 637-867 W m⁻², indicating strong photochemical activities during the daytime of
45 the campaign. The relative humidity (RH) exhibited a clear diurnal variation pattern with a range of 21-
46 91% . The wind (0-7 m/s) from the north to northeast prevailed during the campaign and frequently
47 resulted in increased PM_{2.5} concentrations due to the transport. The PM_{2.5} concentration were in a range
48 of 6-59 µg m⁻³ (5-95 % percentile). The 5-95 % percentile ranges of [O₃], [NO_x], and [TVOC] were 4.6-
49 58.6 ppbv, 8.9-69.6 ppbv, and 15.2-77.5 ppbv, respectively. O₃ showed an obvious diurnal variation,
50 peaking at 13:00 – 15:00. Diurnal variations of NO_x and TVOC showed high concentrations over the rush
51 hours. A PM episode with mean PM_{2.5} concentration of 56.4 µg m⁻³ occurred from November 6th to
52 November 8th, accompanied by a high concentration of both TVOC and NO_x, indicating the same origins
53 of air pollutants. While the N₂O₅ remained in low concentration levels in general, three peak

54 concentrations up to about 600 pptv appeared at nighttime during November 6th-8th.

55 **S3. Positive matrix factorization (PMF)**

56 Positive matrix factorization (PMF) allows for time-resolved mass spectra to be expressed as a linear
57 combination of a finite number of factors, assuming that the factor profiles are constant and unique. Since
58 this method does not require a priori information about the factors, it is an ideal technique for extracting
59 information from ambient measurements where the detailed chemistry, sources, and atmospheric
60 processes are complex. PMF has already been used in source apportionment analysis of OOMs in previous
61 studies (Yan et al., 2016; Zhang et al., 2019, 2022). In this study, PMF was performed using the Igor-based
62 interface Source Finder (SoFi, v6.3), run by the multilinear engine (ME-2) (Canonaco et al., 2013). The
63 data for the PMF model inputs were prepared according to the method described in previous studies
64 (Zhang et al., 2022). Note that the orbitrap analyzer does not measure signal below a certain threshold
65 resulting in incomplete time series for species present at low concentration level. Therefore, the species
66 characterized by incomplete time series with more than 90% missing data and the spectra with more than
67 80% missing were removed (Zhang et al., 2022).

68 PMF analysis in this work was performed in 2-10 factors as shown in Figure S3. Five runs for each
69 solution show good consistencies in both Q/Q_{exp} and explained variation, indicating the small model
70 uncertainty. The change of Q/Q_{exp} , which decreases stepwise from 2.61 (assuming two factors) to 0.65
71 (assuming nine factors). Since the absolute value of Q/Q_{exp} might be misleading, the trend of Q/Q_{exp} is
72 useful to determine the minimum factor number (Ulbrich et al., 2009), a large decrease in Q/Q_{exp} indicates
73 that the additional factor may explain a large fraction of unaccounted variability in the data. The third
74 factor significantly decreases the Q/Q_{exp} value from 2.61 to 1.83, suggesting the importance of the third
75 factor. By adding the third factor, the model can explain 79.4 % of the data variation, in comparison to
76 75.4 % when only two factors are assumed. This improvement in model performance also implies the
77 addition of third factor is crucial. The second largest increase in the explained fraction (from 79.4 % to
78 81.3 %) happens when adding the fourth factor and the Q/Q_{exp} value decreases from 1.83 to 1.51. When
79 model contains 5, 6, 7, 8, 9 and 10 factors, the Q/Q_{exp} values are about 1.27, 1.07, 0.94 and 0.83, 0.73

80 and 0.66 respectively while the explained fraction by mode are in a range of 82.9-87.5 %.

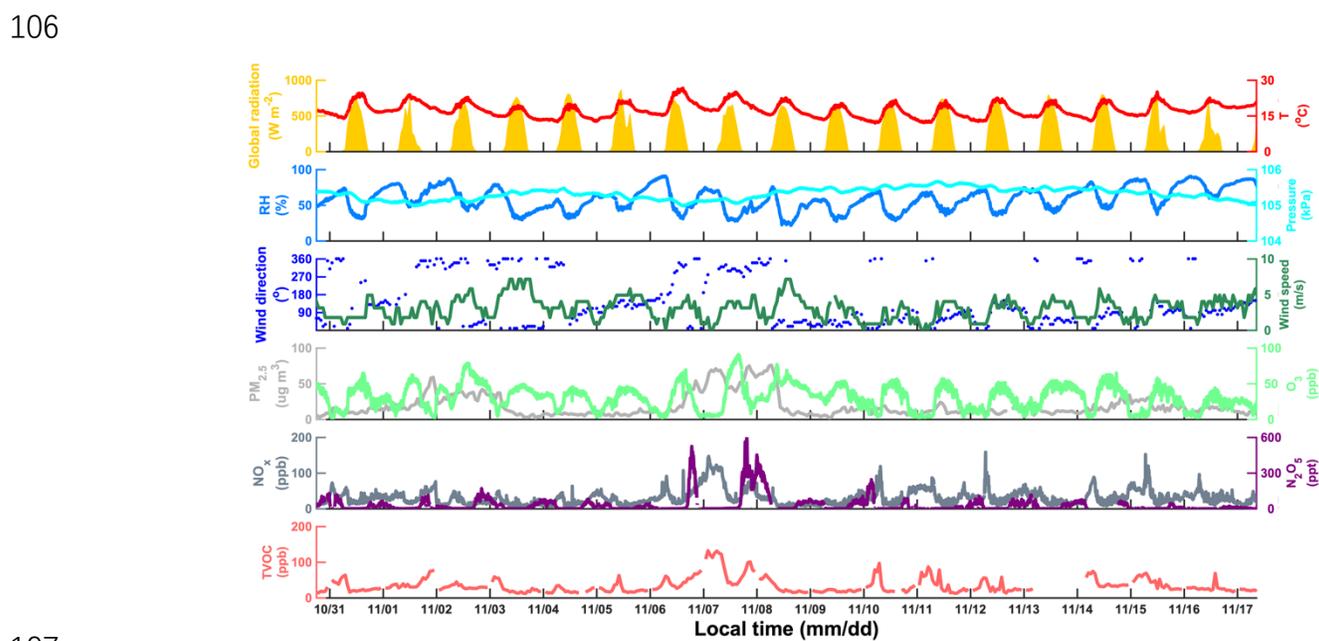
81 Since the PMF analysis is a pure mathematical method without any prior physical or chemical
82 assumptions, choosing the best factor number is critical before describing the PMF results. In terms of
83 trends, more factors would get more freedom to follow subtle variations of the matrix, however, artificially
84 choosing too many factors will over analyze the matrix, resulting in the split of physically meaningful
85 source apportionment into meaningless ones. The timeseries and diurnal variations of factors are shown
86 in Figure S4 and Figure S5. The two-factor solution leads to a distinct daytime factor and a night factor.
87 In the three-factor solution, the timeseries of first two factors are more or less the same as those in the
88 two-factor case, but the variation pattern of second factor has changed in the daytime, the new factor tracks
89 the PM_{2.5} concentration well in two PM episodes, and exhibits a ush-hour peak in the morning. The four-
90 factor solution results in two daytime factors originated from the old daytime factor. When five factors are
91 assumed, an additional nighttime factor appears. When six factors are assumed, an afternoon rush-hour
92 factor appears. For seven factors, the derived new factor has no strong correlation with any independent
93 tracer. Herein, we concluded that the PMF solution with six factors is the optimal solutions and chose
94 to limit our further analysis to the six-factor solution because it is not possible to distinguish the
95 identification of “real” factors without significant correlations. The factor profiles in the six-factor solution
96 could be seen in Figure S6.

97 In the aspect of variation patterns, we classify the six factors into three types. The first two factors
98 are related to the daytime photochemical activities and defined as daytime factor-1 and daytime factor-2.
99 The third factor and fourth factor show clear nocturnal patterns and defined as nighttime factor-1 and
100 nighttime factor-2. The fifth factor and the sixth factor are more related to the emission episode and thus
101 defined as episode factor-1 and episode factor-2. Table S1 shows the peak times and fingerprint molecules
102 of the factors.

103

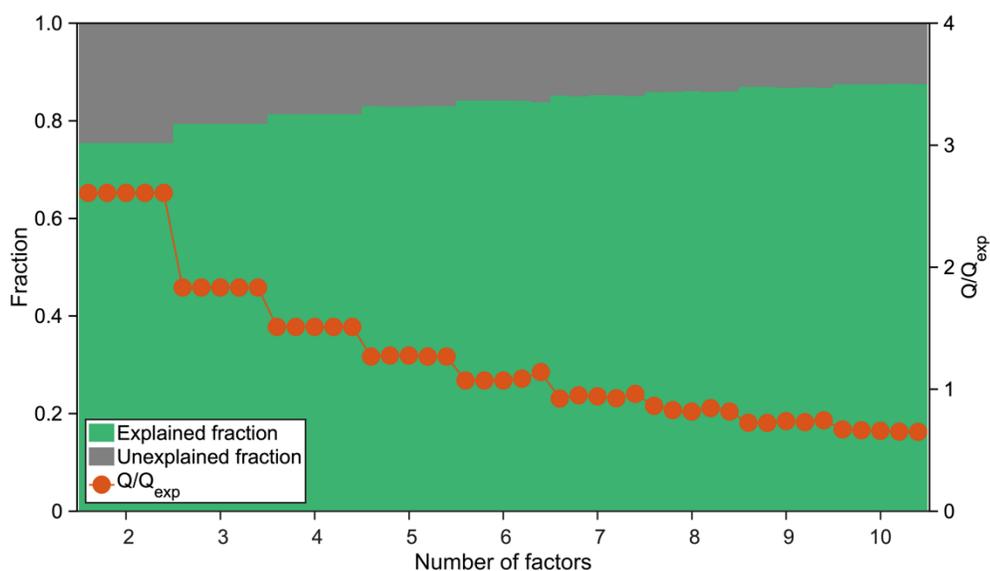


104
105 **Figure S1.** The map of the field site (Shanghai Academy of Environmental Sciences) that is a representative urban station.



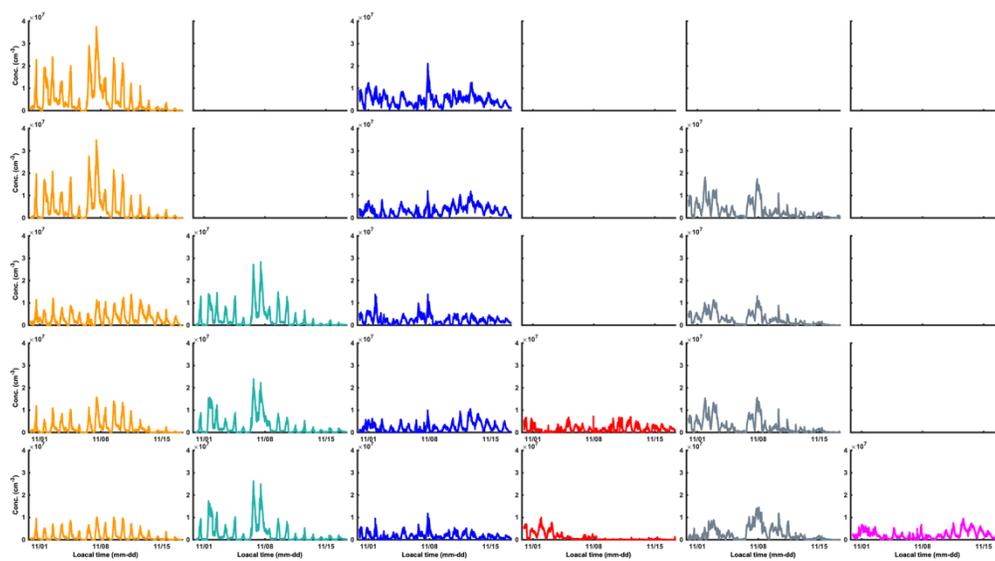
107
108 **Figure S2.** Timeseries of key measurements during the field campaign.

109

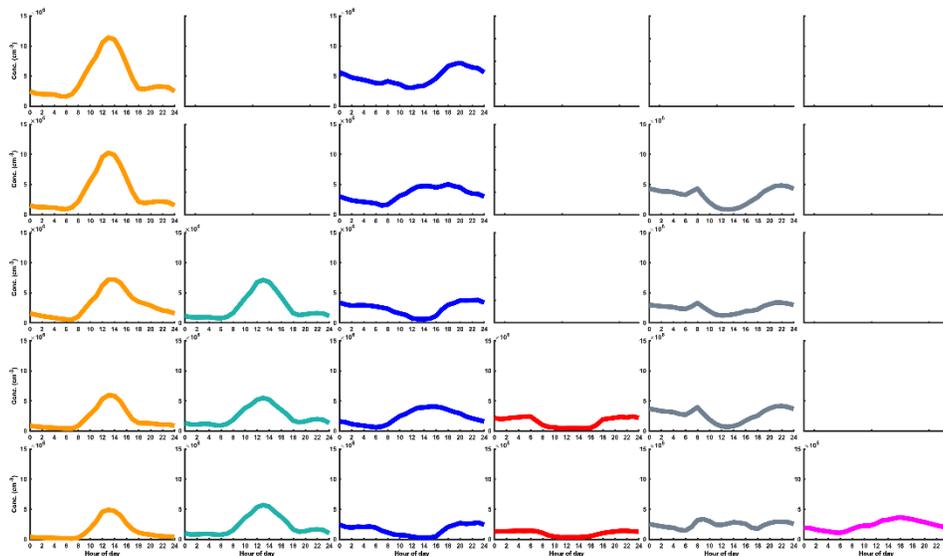


110
 111 **Figure S3.** Mathematical diagnostics of PMF solutions, including the overall changes of Q/Q_{exp} and the explained variation from two-factor
 112 to nine-factor solutions. For each number of factors, five seed runs were performed to test the consistency of the solution.

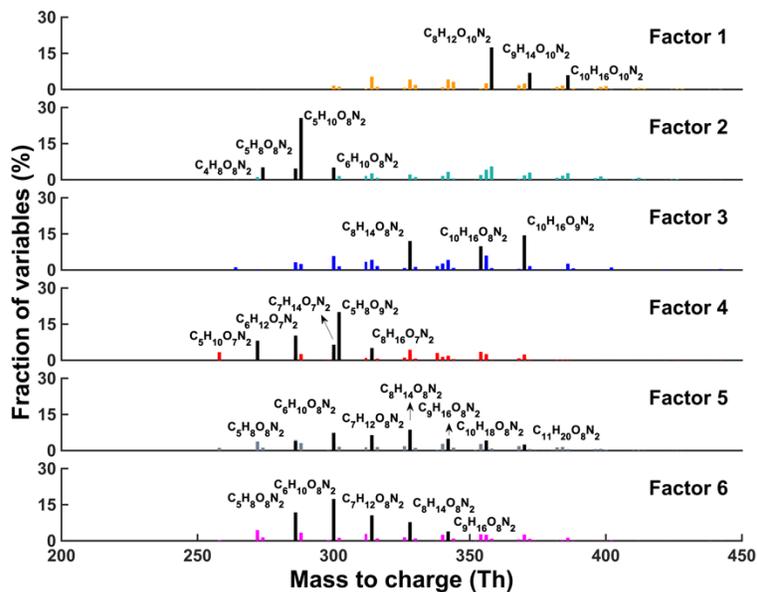
113
 114



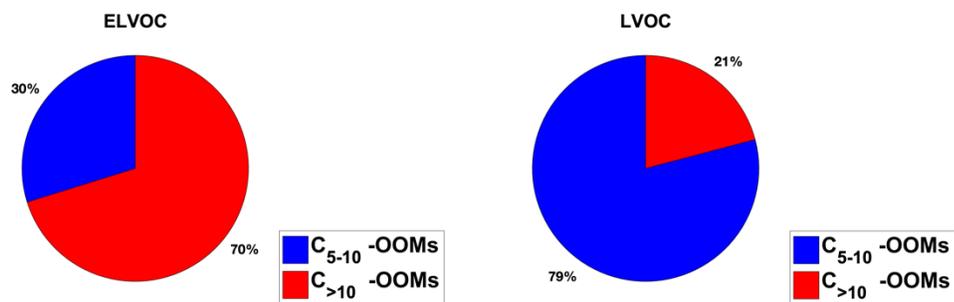
115
 116 **Figure S4.** Timeseries of factors in 2-6 factor solutions of PMF. The panels from top to bottom are 2-factor solution, 3-factor solution, 4-
 117 factor solution, 5-factor solution and 6-factor solution, respectively.



118
 119 **Figure S5.** Diel variation patterns of factors in 2-6 factor solutions of PMF. The panels from top to bottom are 2-factor solution, 3-factor
 120 solution, 4-factor solution, 5-factor solution and 6-factor solution, respectively.

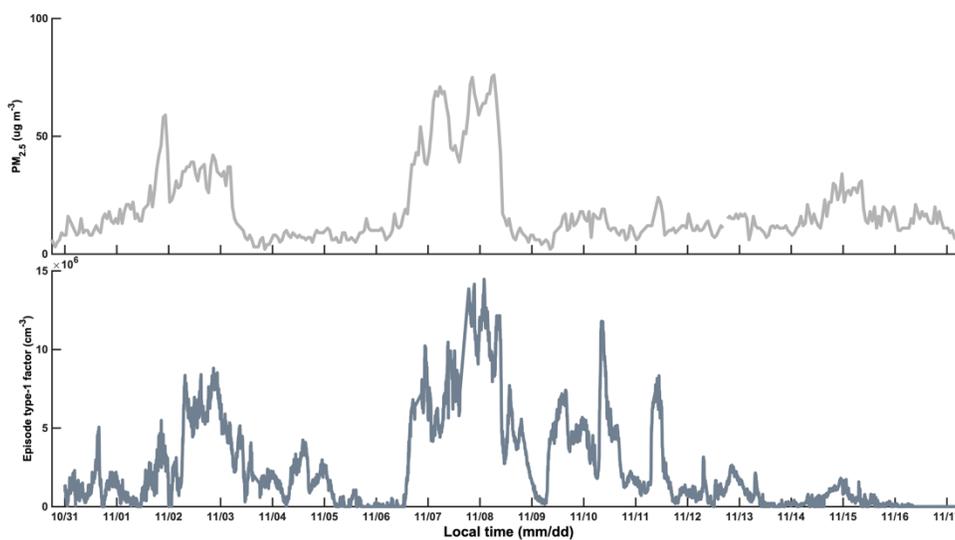


121
 122 **Figure S6.** The factor profiles in the six-factor solution, the black ones represent fingerprint molecules.



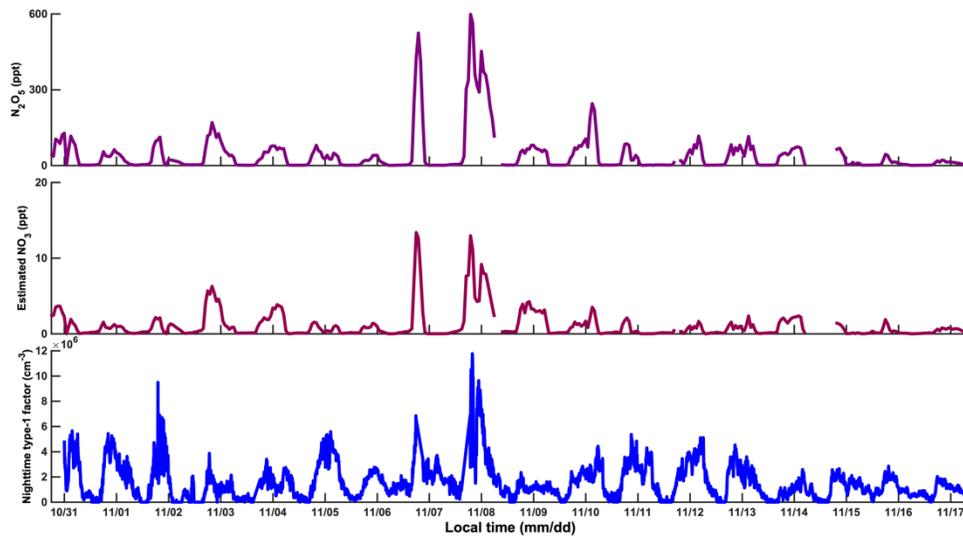
123
 124 **Figure S7.** The relative contributions of OOMs with different carbons to the extremely low-volatility organic compounds (ELVOC, $C^* < 3$
 125 $\times 10^{-5} \mu\text{g m}^{-3}$) and low-volatility organic compounds (LVOC, $3 \times 10^{-5} \leq C^* < 3 \times 10^{-1} \mu\text{g m}^{-3}$).

126



127
 128 **Figure S8.** Timeseries of PM_{2.5} concentration (top panel), and the episode factor-1 (bottom panel).

129



130

131 **Figure S9.** Timeseries of N_2O_5 concentration (top panel), estimated NO_3 concentration (middle panel), and the nighttime factor-1 (bottom
132 panel).

Table S1 Summary of the Factors in six-factor solution

Factor	Factor	Peak time	Fingerprint molecules
Daytime	Daytime factor-1	12:00-14:00	$C_nH_{2n-4}O_{10}N_2$ (n=8-10)
	Daytime factor-2	12:00-14:00	$C_nH_{2n}O_8N_2$ (n=4-5), $C_nH_{2n-2}O_8N_2$ (n=5-6)
Nighttime	Nighttime factor-1	19:00-23:00	$C_{10}H_{16}O_9N_2$, $C_{10}H_{16}O_8N_2$, $C_8H_{14}O_8N_2$
	Nighttime factor-2	20:00-06:00	$C_5H_8O_9N_2$, $C_nH_{2n}O_7N_2$ (n=5-8)
Episode	Episode factor-1	PM episode	$C_nH_{2n-2}O_8N_2$ (n=5-11)
	Episode factor-2	Afternoon rush-hour (16:00)	$C_nH_{2n-2}O_8N_2$ (n=5-9)

Table S2 Averaged nO_{eff} of 2N-OOMs in the four cases

Case	$\overline{[2N - OOM_{Aro}]}$	$\overline{[2N - OOM_{Alu}]}$	$\overline{[2N - OOM_{MT}]}$	$\overline{[2N - OOM_{Total}]}$
CL _{day}	5.6	4.0	4.8	4.6
CL _{night}	4.8	3.9	4.6	4.2
PL _{day}	5.3	4.0	4.9	4.3
PL _{night}	4.8	3.9	4.5	4.1

Reference

- Brown, S. S. and Stutz, J.: Nighttime radical observations and chemistry, *Chem. Soc. Rev.*, 41, 6405–6447, doi:10.1039/c2cs35181a, 2012.
- Canonaco, F., Crippa, M., Slowik, J. G., Baltensperger, U. and Prévôt, A. S. H. H.: SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data, *Atmos. Meas. Tech.*, 6(12), 3649–3661, doi:10.5194/amt-6-3649-2013, 2013.
- Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R. and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, *Atmos. Chem. Phys.*, 9(9), 2891–2918, doi:10.5194/acp-9-2891-2009, 2009.
- 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, doi: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(7), 3761–3776, doi:10.5194/amt-12-3761-2019, 2019.
- Zhang, Y., Li, D., Ma, Y., Dubois, C., Wang, X., Perrier, S., Chen, H., Wang, H., Jing, S., Lu, Y., Lou, S., Yan, C., Nie, W., Chen, J., Huang, C., George, C. and Riva, M.: Field Detection of Highly Oxygenated Organic Molecules in Shanghai by Chemical Ionization–Orbitrap, *Environ. Sci. Technol.*, doi:10.1021/acs.est.1c08346, 2022.