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

Yiqun Lu[1,2], Yingge Ma[1], Dan Dan Huang[1], Shengrong Lou[1], Sheng'ao Jing[1], Yaqin Gao[1], Hongli Wang[1], Yanjun Zhang[3], Hui Chen[4], Yunhua Chang[5], Naiqiang Yan[2], Jianmin Chen[4], Christian George[3], Matthieu Riva[3], 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 (LAP$^3$), Department of Environmental Science & Engineering, Jiangwan Campus, Fudan University, Shanghai 200438, China

[5] Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), NUIST Center on Atmospheric Environment, Nanjing University of Information Science and Technology, Nanjing 210044, China

## Contents of this file

**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, $C^* < 3 \times 10^{-5}$ μg $m^{-3}$) and low-volatility organic compounds (LVOC, $3 \times 10^{-5} \leq C^* < 3 \times 10^{-1}$ μg $m^{-3}$).

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

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

## S1. Other ancillary measurements

The mass concentration of ambient particles was measured by particle monitor (TEOM 1405DF, Thermo, USA). $SO_2$, $O_3$ and $NO_x$ concentrations were measured using a $SO_2$ analyzer (Model 43i, Thermo, USA), a $O_3$ analyzer (Model 49i, Thermo, USA) and a $NO_x$ analyzer (Model 42i, Thermo, USA) with the detection limits of 0.1 ppbv, 0.5 ppbv and 0.4 ppbv, respectively. The above instruments were pre-calibrated before the campaign. The solar radiation was measured on the rooftop of the building. Atmospheric $N_2O_5$ concentrations were measured by an iodide CI-API-TOF. The concentrations of $NO_3$ radicals were estimated under the assumption that $NO_3$, $NO_2$ and $N_2O_5$ could reach an equilibrium quickly in tropospheric conditions (Brown and Stutz, 2012). The total concentrations of VOC precursors (TVOC) were determined by the measurement of an online GC-MS (7890A-5975C, Agilent, USA).

## S2. Overview of the campaign

An overview of the measurement data, illustrating the air quality as well as the meteorological conditions (global radiation, temperature, wind direction, wind speed, and RH), concentrations of trace gases and pollutants ($PM_{2.5}$, $O_3$, $NO_x$, $N_2O_5$, and TVOCs) during the campaign, is provided in this section as shown in Figure S2 and Table S1. Firstly, the maximum intensities of global radiation on individual days were in a range of 637-867 W $m^{-2}$, indicating strong photochemical activities during the daytime of the campaign. The relative humidity (RH) exhibited a clear diurnal variation pattern with a range of 21-91% . The wind (0-7 m/s) from the north to northeast prevailed during the campaign and frequently resulted in increased $PM_{2.5}$ concentrations due to the transport. The $PM_{2.5}$ concentration were in a range of 6-59 μg $m^{-3}$ (5-95 % percentile). The 5-95 % percentile ranges of [$O_3$], [$NO_x$], and [TVOC] were 4.6-58.6 ppbv, 8.9-69.6 ppbv, and 15.2-77.5 ppbv, respectively. $O_3$ showed an obvious diurnal variation, peaking at 13:00 – 15:00. Diurnal variations of $NO_x$ and TVOC showed high concentrations over the rush hours. A PM episode with mean $PM_{2.5}$ concentration of 56.4 μg $m^{-3}$ occurred from November 6$^{th}$ to November 8$^{th}$, accompanied by a high concentration of both TVOC and $NO_x$, indicating the same origins of air pollutants. While the $N_2O_5$ remained in low concentration levels in general, three peak

concentrations up to about 600 pptv appeared at nighttime during November $6^{th}$-$8^{th}$.

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

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

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

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

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

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


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

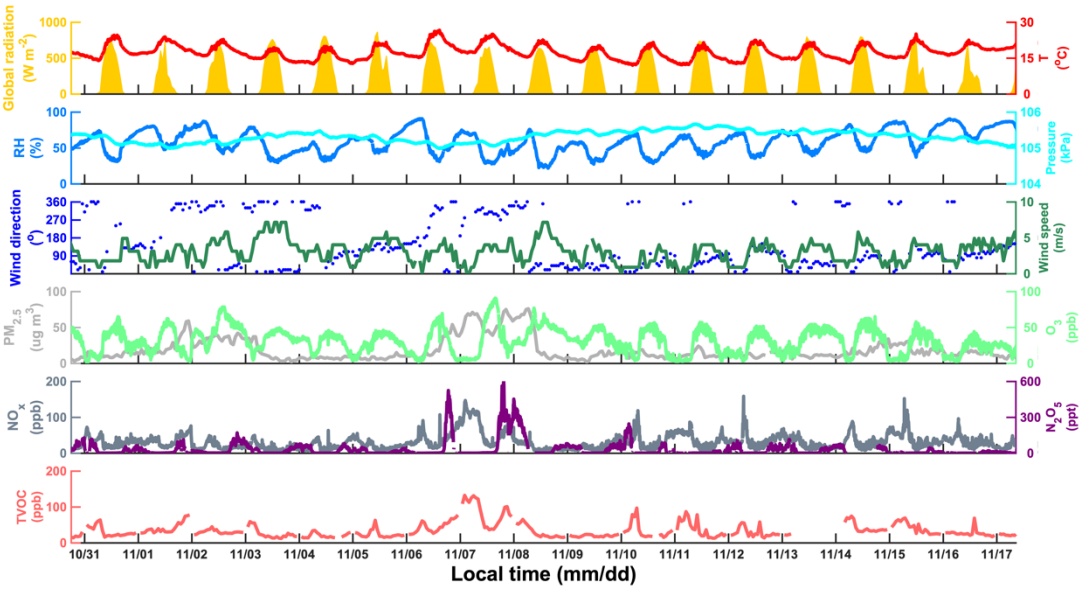


**Figure S2.** Timeseries of key measurements during the field campaign.

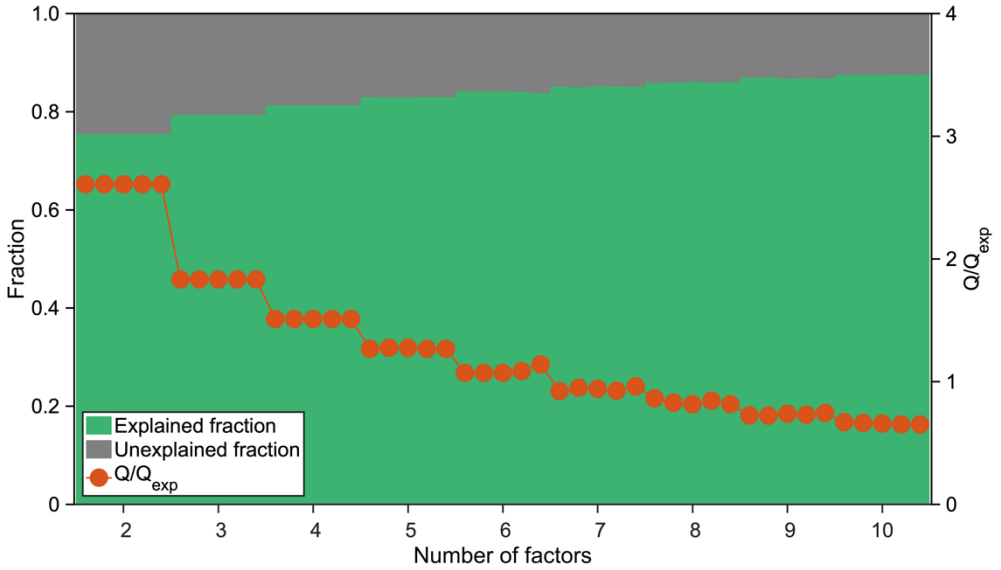


**Figure S3.** Mathematical diagnostics of PMF solutions, including the overall changes of Q/Qexp and the explained variation from two-factor to nine-factor solutions. For each number of factors, five seed runs were performed to test the consistency of the solution.

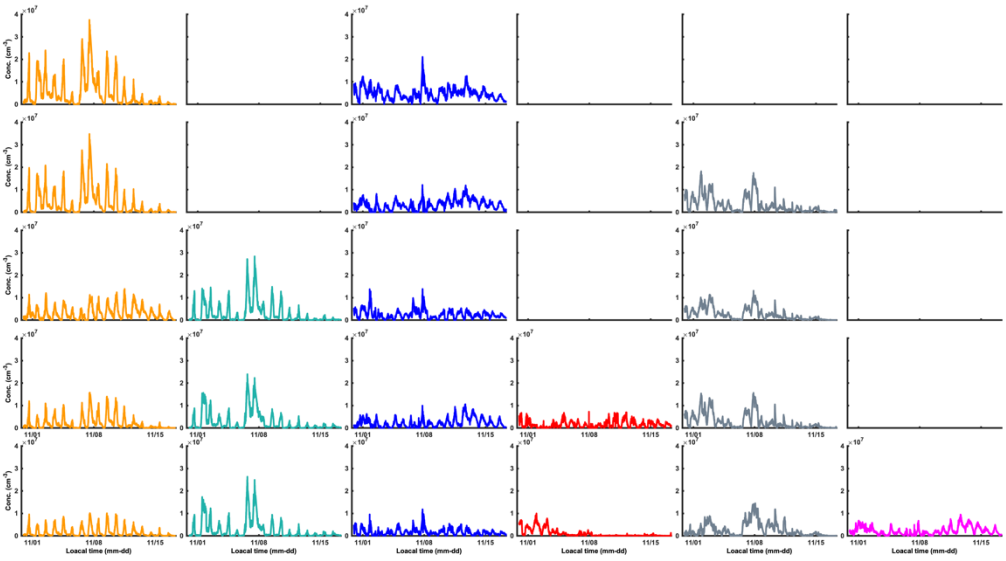

**Figure S4.** Timeseries of factors in 2-6 factor solutions of PMF. The panels from top to bottom are 2-factor solution, 3-factor solution, 4-factor solution, 5-factor solution and 6-factor solution, respectively.

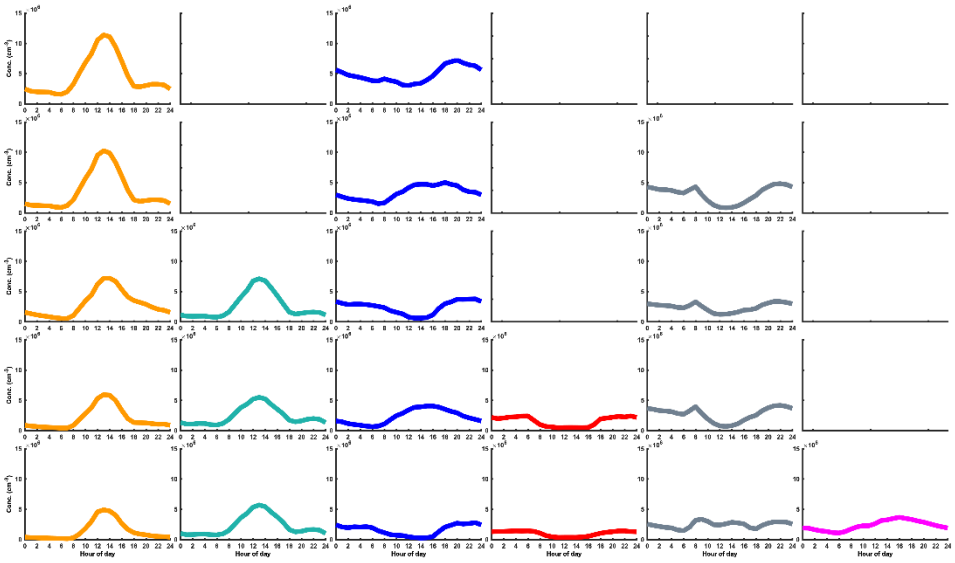

**Figure S5.** Diel variation patterns of factors in 2-6 factor solutions of PMF. The panels from top to bottom are 2-factor solution, 3-factor solution, 4-factor solution, 5-factor solution and 6-factor solution, respectively.

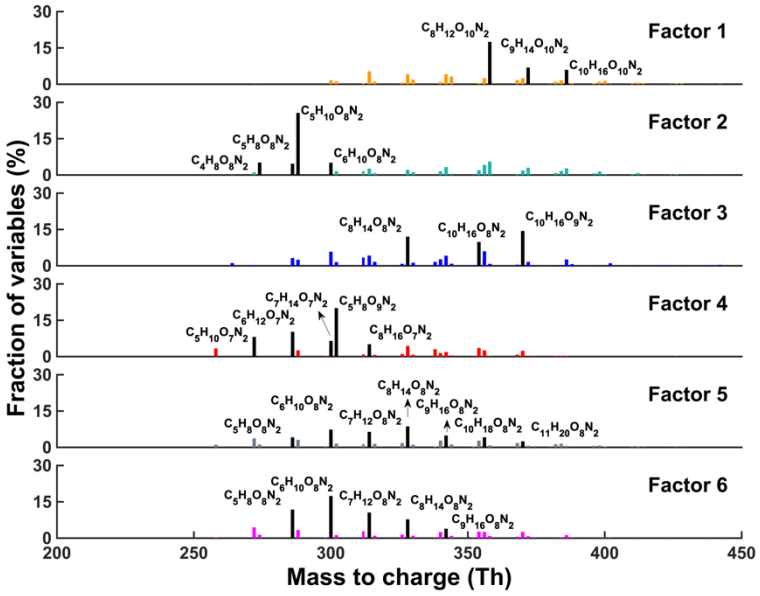


**Figure S6.** The factor profiles in the six-factor solution, the black ones represent fingerprint molecules.

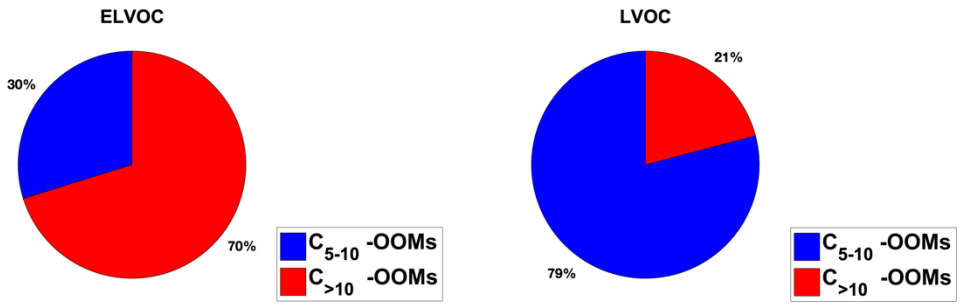


**Figure S7.** The relative contributions of OOMs with different carbons to the extremely low-volatility organic compounds (ELVOC, C* < 3 × $10^{-5}$ μg $m^{-3}$) and low-volatility organic compounds (LVOC, 3 × $10^{-5}$ ≤ C* < 3 × $10^{-1}$ μg $m^{-3}$).

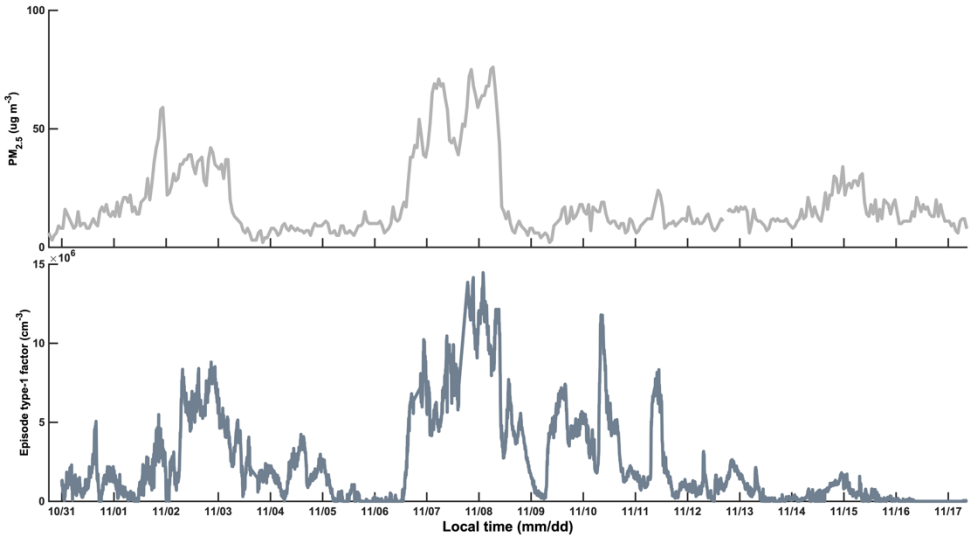


**Figure S8.** Timeseries of $PM_{2.5}$ concentration (top panel), and the episode factor-1 (bottom panel).

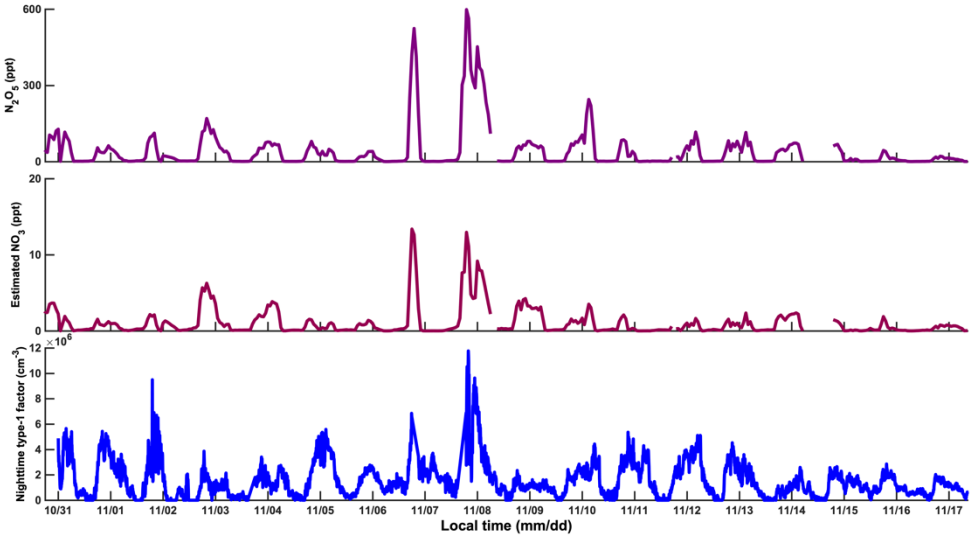


**Figure S9.** Timeseries of $N_2O_5$ concentration (top panel), estimated $NO_3$ concentration (middle panel), and the nighttime factor-1 (bottom panel).

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 | $\overline{[2N-OOM_{Aro}]}$ | $\overline{[2N-OOM_{Ali}]}$ | $\overline{[2N-OOM_{MT}]}$ | $\overline{[2N-OOM_{Total}]}$ |
|---|---|---|---|---|
| $CL_{day}$ | 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.