1   # Catching organic vapors between creation and condensation

2   Wei Nie[1,†], Chao Yan[2,3,†], Dan Dan Huang[4,†], Zhe Wang[5,6,†], Yuliang Liu[1], Xiaohui Qiao[7], Yishuo
3   Guo[2], Linhui Tian[8], Penggang Zheng[6], Zhengning Xu[1], Yuanyuan Li[1], Zheng Xu[1], Ximeng Qi[1],
4   Peng Sun[1], Jiaping Wang[1], Feixue Zheng[2], Xiaoxiao Li[7], Rujing Yin[7], Kaspar Dallenbach[3],
5   Federico Bianchi[3], Tuukka Petäjä[1,3], Yanjun Zhang[3], Mingyi Wang[9], Meredith Schervish[9], Sainan
6   Wang[10], Liping Qiao[4], Qian Wang[4], Min Zhou[4], Hongli Wang[4], Chuan Yu[5], Dawen Yao[5], Hai
7   Guo[5], Penglin Ye[11], Shuncheng Lee[5], Yong Jie Li[8], Yongchun Liu[2], Xuguang Chi[1], Veli-Matti
8   Kerminen[3], Mikael Ehn[3], Neil M. Donahue[9], Tao Wang[5], Cheng Huang[4], Markku Kulmala[1,2,3],
9   Douglas Worsnop[3,11], Jingkun Jiang[7,*], Aijun Ding[1,*]

10   **Affiliations:**

11   [1] Joint International Research Laboratory of Atmospheric and Earth System Research, School of
12   Atmospheric Sciences, Nanjing University, Nanjing, China.

13   [2] Aerosol and Haze Laboratory, Beijing Advanced Innovation Center for Soft Matter Science and
14   Engineering, Beijing University of Chemical Technology, Beijing, China

15   [3] Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University
16   of Helsinki, Finland

17   [4] State Environmental Protection Key Laboratory of Formation and Prevention of Urban Air
18   Pollution Complex, Shanghai Academy of Environmental Sciences, Shanghai, China

19   [5] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University,
20   Hong Kong SAR

21   [6] Division of Environment and Sustainability, The Hong Kong University of Science and
22   Technology, Hong Kong SAR, China

23   [7] State Key Joint Laboratory of Environment Simulation and Pollution Control, State
24   Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex,
25   School of Environment, Tsinghua University, Beijing, China

26   [8] Department of Civil and Environmental Engineering, Faculty of Science and Technology,
27   University of Macau, Taipa, Macau, China

28   [9] Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA, USA

29   [10] State Key Laboratory of Organic Geochemistry, Guangzhou Institute of Geochemistry, Chinese
30   Academy of Sciences, Guangzhou, China

31   [11] Aerodyne Research Inc., Billerica, Massachusetts 01821, USA

32

33   *Correspondence to: Aijun Ding (dingaj@nju.edu.cn) and Jingkun Jiang
34   (jiangjk@tsinghua.edu.cn)

35   † These authors contributed equally

Abstract

Secondary organic aerosol (SOA) contributes a significant fraction to aerosol mass and toxicity. Low-volatility organic vapors are critical intermediates connecting the oxidation of volatile organic compounds (VOCs) to SOA formation. However, the direct measurement of intermediate vapors poses a great challenge, further compounded by the difficulty of linking them to specific precursors from a cocktail of complex emission sources in the vast urbanized areas. Here, we present coordinated measurements of low-volatility oxidation products, termed oxygenated organic molecules (OOMs) in three most urbanized regions in China. With a newly-developed analysis methodology, we are able to assign these OOMs to their likely precursors and ultimately connect SOA formation to various VOCs. At all measurement locations, we find similar OOM chemical composition, source distribution, and contribution to SOA. We show that the oxidation of anthropogenic VOCs dominates OOM formation, with ca. 40% contribution from aromatics, and another ca. 40% contribution from alkanes, a previously under-accounted VOC class. Functionalized VOCs either from direct emission or first generation of aromatics or alkanes, are recognized as an important source of observed OOMs. SOA production via the condensation of these anthropogenic OOMs ($pSOA_{oom}$) increases significantly with the accumulation of aerosol pollution and can be up to 4 $\mu g\ m^{-3}\ hr^{-1}$ in highly polluted conditions. $pSOA_{oom}$ constitutes 44 % - 71 % of the SOA mass growth. Overall, by performing the most detailed molecular analysis of in-situ OOMs, we provide the first quantitative assessment of SOA formation from different VOC classes over urbanized eastern China.

One Sentence Summary

Condensation of anthropogenic organic vapors dominates the formation of secondary organic aerosol in Chinese megacities

Main Text

Aerosol particles are important and ubiquitous species in the Earth's atmosphere, strongly affecting climate[1] and damaging human health[2]. Secondary organic aerosols (SOA) are a major component of the overall aerosol loading in various environments around the globe[3,4], and significantly influences the aerosol toxicity by explaining the highest fraction of the reactive oxygen species in megacities[5]. Therefore, major advances in understanding SOA formation, especially the sources and relevant chemical processes, are of central importance for the implementation of effective measures to mitigate SOA pollution.

The lack of the molecular information of the original SOA material is a major obstacle in understanding SOA formation. First, characterizing the chemical composition of SOA inevitably involves evaporation, desorption, or extraction of aerosol samples[6,7], which modifies the organic molecules to a lesser or greater extent. Second, the molecular information obtained in SOA samples may not be source indicative, as various particle-phase reactions, such as functionalization, oligomerization, and fragmentation[8], may have occurred during SOA ageing, in which the molecular features of the precursor molecule are likely not retained.

A key to understanding SOA formation is to obtain the molecular information of the low-volatility gaseous intermediates from VOC oxidation. These vapors can significantly contribute to SOA mass accumulation, meanwhile possess chemical information of the parent molecule, thus serving as molecular tracers for SOA formation processes. Extensive studies using the high resolution of chemical ionization mass spectrometry have identified plenty of highly oxygenated organic vapors from the oxidation of biogenic VOCs, the condensation of which can explain the majority of SOA formation[9-11]. However, comprehensive investigations of low-volatility intermediates remain marginal in urban environments, which hampers the understanding of the dominant processes of SOA formation.

Here, we conducted coordinated comprehensive measurements in three most urbanized regions in China, including Beijing in the North China Plain (NCP), Nanjing and Shanghai in the Yangtze River Delta (YRD), and Hong Kong in the Pearl River Delta (PRD). A map showing the locations of these megacities is provided in Fig.1 and more detailed descriptions of these observation stations are provided in the Method. The measurements of low-volatility gaseous intermediates are conducted with the state-of-the-art mass spectrometry, namely the nitrate ion-based chemical ionization Atmosphere-Pressure-interface Time-of-Flight mass spectrometer (CI-APi-TOF)[9]. In comparison to other chemical ionization techniques (e.g., iodide ionization), nitrate ionization is more sensitive for measuring vapors of low volatility and thus, of high potential contributing to SOA formation (see Extended Data Fig. 2). For all CI-APi-TOFs, we perform calibrations on both charging efficiency[12] and mass-dependent transmission[13] to ensure the data quality.

The obtained mass spectra in these urban atmosphere shows high complexity, implying a large diversity of the precursors and formation pathways of these low-volatility gaseous intermediates. To extract the molecular information from these extraordinarily complicated mass spectra, we conduct the newly developed binPMF[14] in prior to peak identification. Optimal PMF solutions find eight to ten factors at different locations (Extended Data Fig. 3a), out of which seven are common ones in at least two sites. Each retrieved factor by the binPMF can be regarded as a simplified mass spectrum that facilitates more robust peak identification. With the help of binPMF, we identify about 1500 molecular formulae, accounting for more than 80 percent of the spectral signals. These compounds are further characterized according to their carbon number (n_C), effective oxygen number ($n_{Oeff} = n_O - 2n_N$), nitrogen number (n_N) and double bond equivalent (DBE). Many identified organic species are with $n_{Oeff} < 5$, which cannot be defined as highly oxygenated organic molecules (HOM) according to the convention[15]. We thus refer to all our identified species as oxygenated organic molecules (OOMs), to avoid the confliction with the HOM definition. It should be noted that nitrate phenols appear to be the most abundant species in our measurements at all locations. However, they are excluded in this study, since they are known to have little contribution to SOA formation due to their high volatility.

The temporal variation in OOMs, as captured by binPMF, is predominantly caused by the variation of oxidants (e.g., OH and NO_3 radicals) or loss processes. As such, the binPMF itself does not provide sufficient insights in the VOC precursor of OOMs, except for a clear separation of monoterpene products from others (Extended Data Fig. 3b). We therefore develop a workflow

based on integrated knowledge of VOC oxidation and binPMF results (Extended Data Fig. 4). The workflow, exploiting the measured molecular composition, allows us to attribute these OOMs to their likely VOC precursor classes, including alkanes, aromatics, isoprene, and monoterpenes. We refer to the resolved products as alkane-OOMs, aromatic-OOMs, isoprene-OOMs, and monoterpene-OOMs, respectively. This novel framework provides a quick and sensible classification for low-volatility organic vapors, which is tested and validated using existing laboratory studies. The underlying assumptions and associated uncertainties of the framework are discussed in the Method.

The main OOM categories and their relative contributions to SOA are remarkably similar at the four locations - OOMs are dominantly of anthropogenic sources. As shown in Fig. 1, besides the expected high contribution (33% - 41 %) of aromatic-OOMs, the contribution from alkane-OOMs is surprisingly high (38% - 48%). This highlights that alkane is an important SOA source[16]. In addition, a residual of 7 - 11% of the signals can be assigned to anthropogenic sources, but cannot be unambiguously distinguished as aromatics- or alkanes-OOMs. Overall, the anthropogenic sources contributed 84% - 89% of the total observed OOMs. On the other hand, OOMs with biogenic origins, i.e., isoprene-OOMs and monoterpene-OOMs, have much lower contributions (ca. 10%) since the measurements were conducted during the cold seasons. Despite the similar source distribution, the concentration of OOMs varies substantially among locations, with the lowest value in Beijing (2.38×10^7 cm^{-3}), the highest in Hong Kong (2.30×10^8 cm^{-3}), and intermediate values in Shanghai (7.78×10^7 cm^{-3}) and Nanjing (7.74×10^7 cm^{-3}). This difference is very likely an overall result of the intensity of incident solar radiation, air temperature, the concentration of atmospheric oxidants that all together, defines the OOM formation from VOC oxidation[11,17], as well as the OOM removal rate by condensing on particles.

The detailed molecular information of OOMs reflects prominent features of photochemistry in the polluted urban atmosphere. First, the most distinct feature of the spectrum is a regular spacing corresponding to a difference of CH_2, which can be seen from, for instance, the series of $C_xH_{2x}O_7N_2$ (x= 4-13) in alkane-OOMs and $C_xH_{2x-3}O_8N$ (x= 5-14) in aromatic-OOMs (Fig. 2a and 2b). As no known gas-phase reaction can lead to either addition or subtraction of CH_2, this clearly reflects the nature of anthropogenic VOC emissions; homologous compounds are usually co-emitted, such as benzene, toluene and xylene, or sequences of alkanes. Second, NO_x dominates the termination reaction with RO_2 radicals, especially for alkane-RO_2 radicals, resulting in a major fraction of alkane-OOMs (84%) and aromatic-OOMs (52%) consisting of nitrate-containing compounds (Fig. 2a). Third, multi-step oxidation plays a significant role in OOM formation. This is supported by: 1) the overall fraction of nitrate-containing compounds is much higher than any known branching ratio of the reaction between RO_2 and NO forming organonitrates; and 2) di-nitrate OOMs comprise a considerable fraction for both alkane-OOMs (34%) and aromatic-OOMs (7%). Fourth, molecular characteristics reflect that auto-oxidation plays an important role in OOMs formation. For aromatic-OOMs, in addition to confirming that autoxidation occurs more efficiently for aromatics substituted by larger alkyl groups, i.e., with larger n_C[18-21], our observations also show an increasing trend of n_{Oeff} from north to south (Extended Data Fig. 5). Such temperature dependence is in line with that autoxidation is promoted by high temperature. For alkane-OOMs,

156   autoxidation can also contribute to 1 to 3 n_{Oeff} (Extended Data Fig. 7). These results underscore
157   the necessity of implementing the autoxidation scheme in SOA models even in high NO_x
158   environments[18].

159   We use two Volatility Basis Set (VBS)[22] parameterizations to estimate OOM volatility, a key
160   property influencing the condensation of OOMs to form SOA. By grouping different molecules
161   into several volatility bins, VBS has been successfully and widely used to represent the volatility
162   distribution of organic vapors, especially for complex multi-component systems[10]. As suggested
163   by a recent study[21], we apply different volatility parameterizations to OOMs from different
164   precursor classes. This is because that the effectiveness of oxygen atoms in reducing volatility
165   differ significantly in different functional groups, such as peroxy vs hydroxyl and carbonyl groups,
166   which are mainly formed from different oxidation pathways[23,24].Therefore, separating OOMs from
167   different sources is crucial, demonstrating another virtue of our workflow. In this study, we adopt
168   parameterizations used by Mohr et al.[25] and Wang et al.[21], to estimate the volatility of
169   monoterpene-OOMs and other types of OOMs, respectively. The results were also validated and
170   compared with another different parameterization proposed by Li et al., [26]. Although the volatility
171   distribution is not identical, the results show excellent consistence in total concentration of
172   condensable vapors (Extended Data Fig. 8).

173   In general, OOM "intrinsic" volatility distributions (i.e., at 300 K) are similar among all locations:
174   SVOC and LVOC account for the majority of OOMs, with a considerable fraction of ELVOC and
175   ULVOC (Fig. S14). The bulk volatility of alkane-OOMs appears to be higher than that of aromatic-
176   OOMs (Fig. 2d). However, both alkane-OOMs and aromatic-OOMs comprise considerable
177   fractions of LVOC and ELVOC that are able to contribute to SOA formation. Aside from the
178   intrinsic volatility, ambient temperature influences the volatility distribution; the substantial
179   temperature differences in our measurement locations lead to large shifts in the ambient volatility
180   distribution of approximately 1 decade (1 volatility bin) per 15 ^{o}C. As shown in Extended Data
181   Fig. 9, the bulk OOM volatility is lowest in Beijing, intermediate in Nanjing and Shanghai, and
182   highest in Hong Kong.

183   The source-segregated OOM contribution to SOA formation is a vital basis for formulating
184   effective strategies to mitigate aerosol pollution. For this reason, we investigate the contribution
185   of OOMs from different source classes to SOA mass accumulation over a wide range of $PM_{2.5}$
186   concentrations. We calculate the net condensation mass flux of OOMs towards the particulate
187   phase using the same approach as reported in Trostl et al.[10] (see Method). This flux defines the
188   SOA formation rate via the irreversible condensation of OOMs, termed as $pSOA_{oom}$ in this work.
189   Note that this term does not include the possible contribution of SVOC to SOA, which is
190   dominated by equilibrium partitioning. We show the calculation based on the observation in
191   Nanjing in Fig.3, and results for other locations are provided in Extended Data Fig. 10. As shown
192   in Fig 3c, the total $pSOA_{oom}$ can reach up to ~ 4 $\mu g\ m^{-3}\ h^{-1}$ at the $PM_{2.5}$ concentration of 220 $\mu g\ m^{-}$
193   ^{3} in Nanjing (Fig. 3c). Aromatic-OOMs are the largest contributor (ca. 50 – 70 %) to the total
194    $pSOA_{oom}$ in all the locations, and alkane-OOMs are the second-largest, contributing to ca. 20 – 30
195   % of the total $pSOA_{oom}$. In contrast, the contribution from biogenic OOMs to $pSOA_{oom}$ is very

small, especially at high PM$_{2.5}$ levels. This is because isoprene-OOMs are mostly SVOC, and monoterpene-OOMs are of too low concentrations.

Unexpectedly, the pSOA$_{oom}$ contributed by alkane-OOMs, aromatic-OOMs and isoprene-OOMs shows an increasing trend from very clean to highly polluted atmospheric conditions (Fig. 3c, Extended Data Fig. 11). As the photochemical oxidation is usually thought to be reduced during haze condition, this increasing trend indicates that the weakened photochemistry due to the aerosol dimming effect is outcompeted by the simultaneously elevated VOC concentration (Fig. 3a). Consequently, the VOC oxidation rate is enhanced during pollution, leading to high concentrations of all types of OOMs even in the coincidence of a high particle surface-area concentration (Fig. 3b), with the only exception being the relatively less important monoterpene-OOMs (Extended Data Fig. 11). In short, our results demonstrate a positive feedback loop between OOM formation and pollution, in which OOM condensation significantly favors pollution accumulation by forming SOA, and the weakened air dispersion in haze condition, in turn, favors the OOM formation and condensation.

In order to quantify the importance of the pSOA$_{oom}$ to the total SOA formation, we compare the pSOA$_{oom}$ to the actual augmentation rate of SOA mass ($d(SOA)/dt$). We perform this analysis in selected cases with stable meteorological conditions, e.g., no dramatic change in boundary layer height, wind speed, and wind direction, to minimize the meteorology interferences (see Extended Data Fig. 12 for an example). As shown in Fig.4a, the pSOA$_{oom}$ exhibits a good linear correlation with the $d(SOA)/dt$. The estimated contributions of the pSOA$_{oom}$ to the overall $d(SOA)/dt$ (as the fitted slopes) are 44%, 70%, 71%, and 68% in Beijing, Nanjing, Shanghai, and Hong Kong, respectively. The relatively lower contribution found in Beijing provides a possible explanation that the haze formation in northern China shows a weaker dependence on photochemistry than in southern China[27-29].

Benefiting from our comprehensive observation and novel workflow, we are able to estimate the precursor-segregated pSOA$_{oom}$ and, in turn, the apparent SOA yield (γ SOA$_{oom}$= pSOA$_{oom}$ / k[VOC][OH]) of some selected species, e.g. C6-C9 aromatics (Fig. 4b). Interestingly, not only the SOA mass yields of these species, but also the HOMs molar yields of C6 aromatics are higher than the values derived from laboratory studies (Fig. 4c). Our results concur with previous reports that the lack of accounting the low-volatility OOMs in previous laboratory investigations may significantly bias the SOA yields low[30]. More importantly, our results may suggest that the aromatic precursors, other than the traditional species we usually measured, exist yet not counted in the apparent SOA yield calculation. The apparent yields of C6-C9 aromatic OOMs yields were calculated without pre-assumption on the exact molecular composition of its precursor molecule, e.g., whether the C6 aromatic OOM is from a benzene or a phenol molecule. The results indicate that other aromatic precursors, especially functionalized aromatic precursors, exist yet under measured. These unknown precursors, either from direct emission or first generation products from oxidation of precursor VOCs, likely have much higher SOA forming potential than their peer hydrocarbon species. This also partly explains the observed high concentrations of OOMs, and in turn, significant contribution to SOA formation.

In summary, by conducting coordinated comprehensive measurements in four megacities in east China and performing novel and detailed molecular analyses, we perform a species-level source appointment of condensable organic vapors (i.e., OOMs) and gain the insights to eventually connect the emission inventory to SOA formation. This is a critical advantage in the investigation of the SOA sources in the urban atmosphere of high chemical complexity. We demonstrate that the condensation of anthropogenic-OOM is a dominant source of SOA even under severe haze conditions. These new insights shed light into the full course of SOA formation, including precursor (i.e., VOCs) recognition, products identification, key oxidation processes, and SOA formation via vapor condensation. More importantly, the OOM distribution and formation pathways is largely the same across three most urbanized regions of Beijing-Tianjin-Hebei, YRD and PRD, where more than 800 million people live and suffer from air pollution. This implies a possibility of solving air pollution issues with a uniform and effective mitigation strategy.

References

1   Edenhofer, O. & Seyboth, K. in *Encyclopedia of Energy, Natural Resource, and Environmental Economics* (ed Jason F. Shogren) 48-56 (Elsevier, 2013).

2   Heal, M. R., Kumar, P. & Harrison, R. M. Particles, air quality, policy and health. *Chemical Society Reviews* **41**, 6606-6630, doi:10.1039/c2cs35076a (2012).

3   Jimenez, J. L. *et al.* Evolution of Organic Aerosols in the Atmosphere. *Science* **326**, 1525-1529, doi:10.1126/science.1180353 (2009).

4   Huang, R. J. *et al.* High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **514**, 218-222, doi:10.1038/nature13774 (2014).

5   Zhou, J. *et al.* Predominance of secondary organic aerosol to particle-bound reactive oxygen species activity in fine ambient aerosol. *Atmos. Chem. Phys.* **19**, 14703-14720, doi:10.5194/acp-19-14703-2019 (2019).

6   Laskin, A., Laskin, J. & Nizkorodov, S. Mass spectrometric approaches for chemical characterisation of atmospheric aerosols: Critical review of the most recent advances. *Environmental Chemistry - ENVIRON CHEM* **9**, 163-189, doi:10.1071/EN12052 (2012).

7   Nozière, B. *et al.* The Molecular Identification of Organic Compounds in the Atmosphere: State of the Art and Challenges. *Chemical Reviews* **115**, 3919-3983, doi:10.1021/cr5003485 (2015).

8   Isaacman-VanWertz, G. *et al.* Chemical evolution of atmospheric organic carbon over multiple generations of oxidation. *Nature chemistry* **10**, 462-468, doi:10.1038/s41557-018-0002-2 (2018).

9   Ehn, M. *et al.* A large source of low-volatility secondary organic aerosol. *Nature* **506**, 476-497, doi:10.1038/nature13032 (2014).

10   Troestl, J. *et al.* The role of low-volatility organic compounds in initial particle growth in the atmosphere. *Nature* **533**, 527-531, doi:10.1038/nature18271 (2016).

11   Stolzenburg, D. *et al.* Rapid growth of organic aerosol nanoparticles over a wide tropospheric temperature range. *Proceedings of the National Academy of Sciences* **115**, 9122-9127, doi:10.1073/pnas.1807604115 (2018).

12   Kürten, A., Rondo, L., Ehrhart, S. & Curtius, J. Calibration of a Chemical Ionization Mass Spectrometer for the Measurement of Gaseous Sulfuric Acid. *The Journal of Physical Chemistry A* **116**, 6375-6386, doi:10.1021/jp212123n (2012).

13   Heinritzi, M. *et al.* Characterization of the mass-dependent transmission efficiency of a CIMS. *Atmos. Meas. Tech.* **9**, 1449-1460, doi:10.5194/amt-9-1449-2016 (2016).

14   Zhang, Y. *et al.* A novel approach for simple statistical analysis of high-resolution mass spectra. *Atmos. Meas. Tech.* **12**, 3761-3776, doi:10.5194/amt-12-3761-2019 (2019).

15   Bianchi, F. *et al.* Highly Oxygenated Organic Molecules (HOM) from Gas-Phase Autoxidation Involving Peroxy Radicals: A Key Contributor to Atmospheric Aerosol. *Chemical Reviews* **119**, 3472-3509, doi:10.1021/acs.chemrev.8b00395 (2019).

16   Wang, Z. *et al.* Efficient alkane oxidation under combustion engine and atmospheric conditions. *Communications Chemistry* **4**, 18, doi:10.1038/s42004-020-00445-3 (2021).

17   Simon, M. *et al.* Molecular understanding of new-particle formation from alpha-pinene between -50 °C and 25 °C. *Atmos. Chem. Phys. Discuss.* **2020**, 1-42, doi:10.5194/acp-2019-1058 (2020).

18   Wang, S., Wu, R., Berndt, T., Ehn, M. & Wang, L. Formation of Highly Oxidized Radicals and Multifunctional Products from the Atmospheric Oxidation of Alkylbenzenes. *Environmental Science & Technology* **51**, 8442-8449, doi:10.1021/acs.est.7b02374 (2017).

19   Molteni, U. *et al.* Formation of highly oxygenated organic molecules from aromatic compounds. *Atmospheric Chemistry and Physics* **18**, 1909-1921, doi:10.5194/acp-18-1909-2018 (2018).

20   Garmash, O. *et al.* Multi-generation OH oxidation as a source for highly oxygenated organic molecules from aromatics. *Atmospheric Chemistry and Physics* **20**, 515-537, doi:10.5194/acp-20-515-2020 (2020).

21   Wang, M. *et al.* Photo-oxidation of Aromatic Hydrocarbons Produces Low-Volatility Organic Compounds. *Environmental science & technology* **54**, 7911-7921, doi:10.1021/acs.est.0c02100 (2020).

22   Donahue, N. M., Epstein, S. A., Pandis, S. N. & Robinson, A. L. A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics. *Atmospheric Chemistry and Physics* **11**, 3303-3318, doi:10.5194/acp-11-3303-2011 (2011).

23   Wang, L., Wu, R. & Xu, C. Atmospheric Oxidation Mechanism of Benzene. Fates of Alkoxy Radical Intermediates and Revised Mechanism. *The Journal of Physical Chemistry A* **117**, 14163-14168, doi:10.1021/jp4101762 (2013).

24   Pankow, J. F. & Asher, W. E. SIMPOL.1: a simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds. *Atmospheric Chemistry and Physics* **8**, 2773-2796, doi:10.5194/acp-8-2773-2008 (2008).

25   Mohr, C. *et al.* Molecular identification of organic vapors driving atmospheric nanoparticle growth. *Nature Communications* **10**, 4442, doi:10.1038/s41467-019-12473-2 (2019).

26   Li, Y., Pöschl, U. & Shiraiwa, M. Molecular corridors and parameterizations of volatility in the chemical evolution of organic aerosols. *Atmos. Chem. Phys.* **16**, 3327-3344, doi:10.5194/acp-16-3327-2016 (2016).

27   Tilmes, S. *et al.* Climate Forcing and Trends of Organic Aerosols in the Community Earth System Model (CESM2). *Journal of Advances in Modeling Earth Systems* **11**, 4323-4351, doi: https://doi.org/10.1029/2019MS001827 (2019).

28   Zhang, F. *et al.* An unexpected catalyst dominates formation and radiative forcing of regional haze. *Proc Natl Acad Sci U S A* **117**, 3960-3966, doi:10.1073/pnas.1919343117 (2020).

29   Huang, X. *et al.* Amplified transboundary transport of haze by aerosol–boundary layer interaction in China. *Nature Geoscience* **13**, 428-434, doi:10.1038/s41561-020-0583-4 (2020).

30   Zhang, X. *et al.* Influence of vapor wall loss in laboratory chambers on yields of secondary organic aerosol. *Proceedings of the National Academy of Sciences* **111**, 5802-5807, doi:10.1073/pnas.1404727111 (2014).

31   Odum, J. R. *et al.* Gas/Particle Partitioning and Secondary Organic Aerosol Yields. *Environmental Science & Technology* **30**, 2580-2585, doi:10.1021/es950943 (1996).

32   Cocker Iii, D. R., Mader, B. T., Kalberer, M., Flagan, R. C. & Seinfeld, J. H. The effect of water on gas–particle partitioning of secondary organic aerosol: II. m-xylene and 1,3,5-trimethylbenzene photooxidation systems. *Atmospheric Environment* **35**, 6073-6085, doi: https://doi.org/10.1016/S1352-2310(01)00405-8 (2001).

33   Song, C., Na, K. & Cocker, D. R. Impact of the Hydrocarbon to NOx Ratio on Secondary Organic Aerosol Formation. *Environmental Science & Technology* **39**, 3143-3149, doi:10.1021/es0493244 (2005).

34   Ng, N. L. *et al.* Secondary organic aerosol formation from m-xylene, toluene, and benzene. *Atmos. Chem. Phys.* **7**, 3909-3922, doi:10.5194/acp-7-3909-2007 (2007).

35   Sato, K. *et al.* AMS and LC/MS analyses of SOA from the photooxidation of benzene and 1,3,5-trimethylbenzene in the presence of NO_x: effects of chemical structure on SOA aging. *Atmos. Chem. Phys.* **12**, 4667-4682, doi:10.5194/acp-12-4667-2012 (2012).

36   Li, L., Tang, P., Nakao, S., Chen, C. L. & Cocker Iii, D. R. Role of methyl group number on SOA formation from monocyclic aromatic hydrocarbons photooxidation under low-NOx conditions. *Atmos. Chem. Phys.* **16**, 2255-2272, doi:10.5194/acp-16-2255-2016 (2016).

37   Li, L., Tang, P., Nakao, S. & Cocker Iii, D. R. Impact of molecular structure on secondary organic aerosol formation from aromatic hydrocarbon photooxidation under low-NOx conditions. *Atmos. Chem. Phys.* **16**, 10793-10808, doi:10.5194/acp-16-10793-2016 (2016).

Author contributions WN, CY, DH, ZW, AD, JJ, MK designed the study; WN, CY, DH, ZW, YL, XQ, YG, LT, PZ analyzed the data; WN, CY, DH, ZW wrote the manuscript; YL, ZX, YL, XQ, collected other research materials; all authors participated in relevant scientific discussion and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Aijun Ding (dingaj@nju.edu.cn) and Jingkun Jiang (jiangjk@tsinghua.edu.cn).
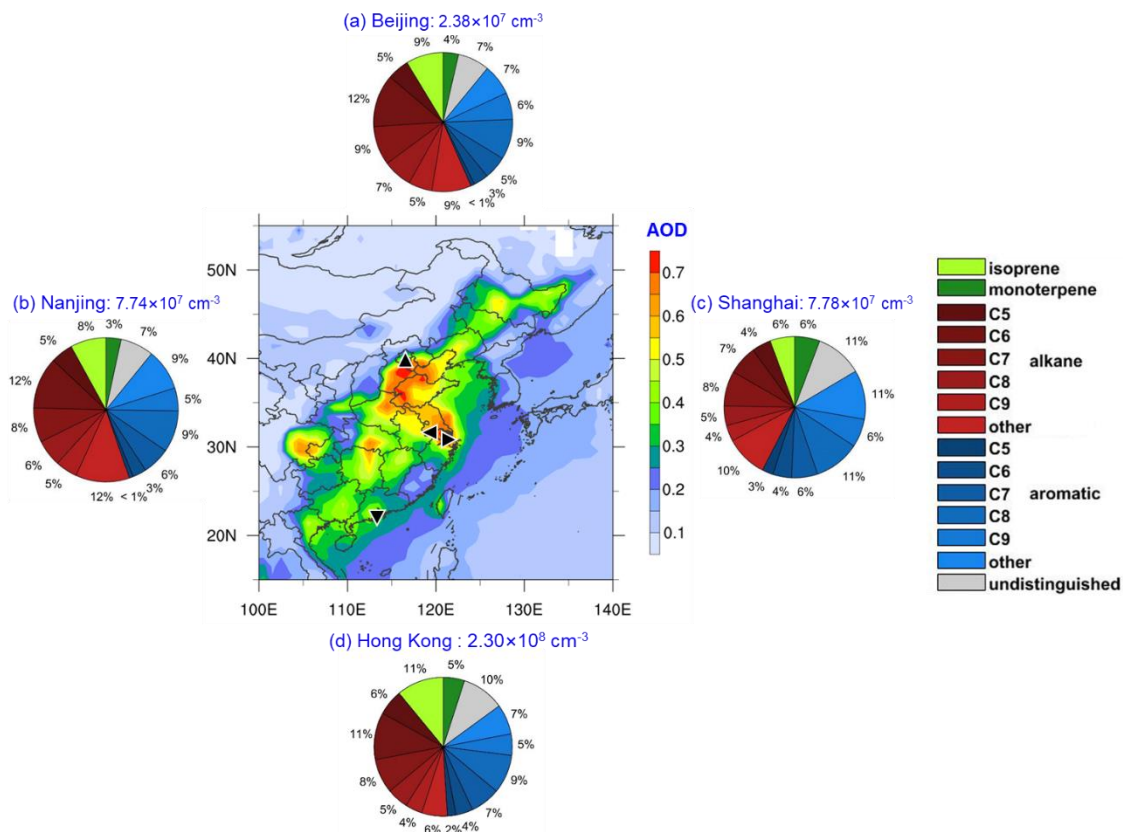
**Fig. 1. Distribution of oxidized oxygenated molecules (OOMs) in (a) Beijing, (b) Nanjing, (c) Shanghai, and (d) Hong Kong.** OOMs are assigned to precursor classes by a framework described in the text and indicated on the pie chart by the color hue, while the carbon number is given by saturation. Colors in the pie chart of bright green, red, blue and dark green represent isoprene-OOMs, alkane-OOMs, aromatic-OOMs and monoterpene-OOMs, respectively. The map shows the general aerosol pollution level in eastern China. The color code denotes the average aerosol optical depth (AOD) during October to December from 2014 to 2018.
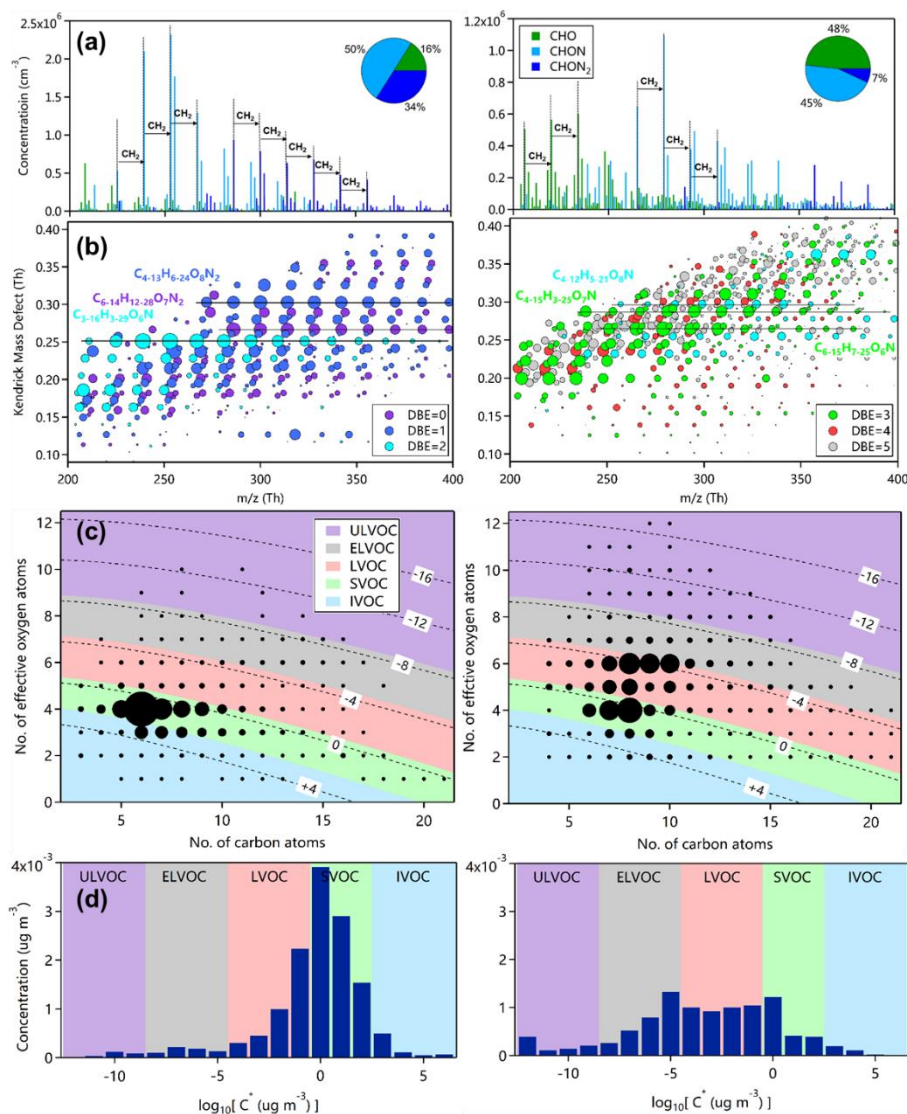
Fig. 2. Characteristics of alkane-OOMs (left) and aromatic-OOMs (right) in Nanjing. (a) The spectra of OOMs colored by their types with nitrate-OOMs (CHON) in light blue, dinitrate-OOMs (CHON$_2$) in dark blue, and non-nitrate-OOMs (CHO) in green. OOMs with a regular shift corresponding to CH$_2$ are marked in the plot. (b) Kendrick mass defect plot showing OOM composition. The x-axis is the exact mass of HOMs, and the y-axis is the Kendrick mass defect. The symbol color denotes the OOM double-bond equivalent (DBE), and the symbol size is proportional to the logarithm of the signal strength. (c) Observed OOMs plotted as a function of the number of carbon and oxygen atoms per molecule. The symbol size is linearly proportional to the total measured concentration of all molecules at a given point. Contour lines indicate the log of the saturation concentration (C*) and background colors correspond to VBS volatility classes. (d) The same OOM compounds as shown in Fig. 2b and 2c are binned into a one-dimensional volatility basis set (VBS) at 300 K.
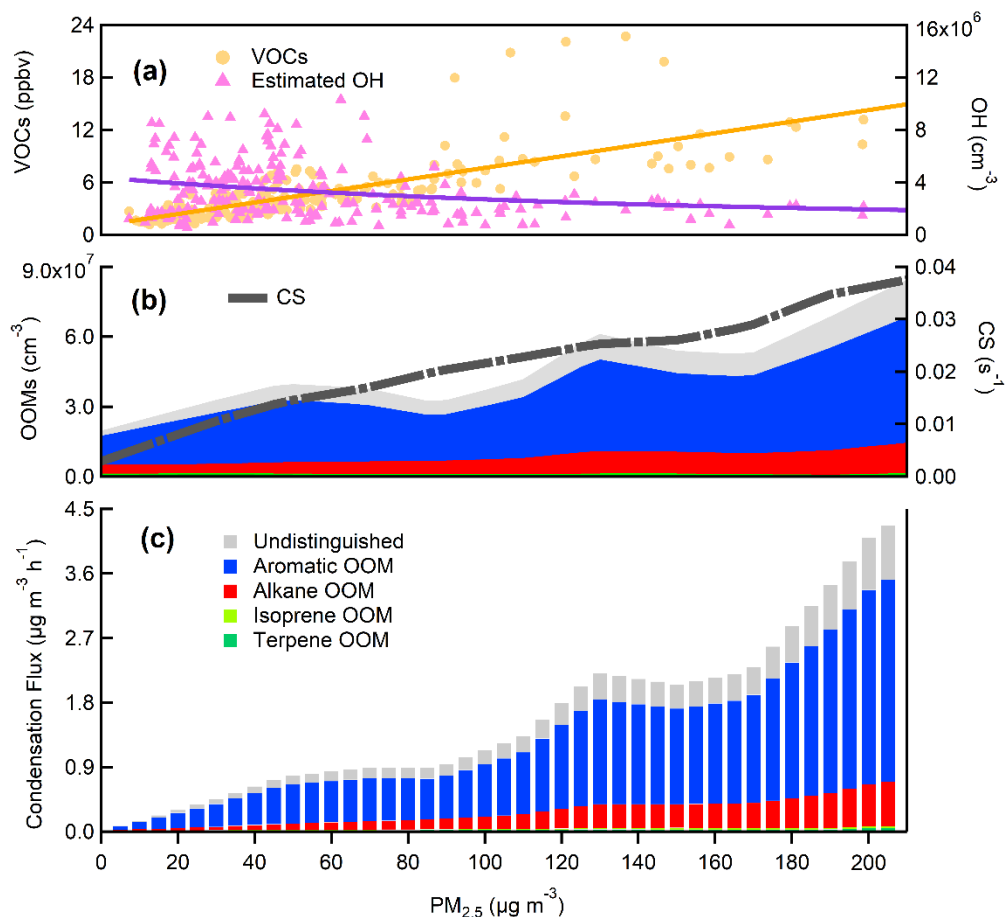
Fig. 3 Influencing factors on OOMs formations and its subsequent impacts on PM$_{2.5}$ pollutions. (a) Concentrations of VOCs and estimated OH as a dependent of PM$_{2.5}$ from near zero to 220 µg/m^3 in Nanjing. Light yellow circles represent the sum of more than 40 VOC species with the number of carbon atoms equal or higher than 5 measured using a PTR-TOF-MS. Light purple triangles represent the estimated OH via the measured sulfuric acid. (b) OOM concentration and condensation sink as a dependent of PM$_{2.5}$ from near zero to 220 µg/m^3 in Nanjing. Sampling hours with UVB higher than 0.1 were selected for both panels (a) and (b) to guarantee that sulfuric acids were mainly formed from the reaction of SO$_2$ and OH. (c) OOM condensation flux as a dependent of PM$_{2.5}$ from near zero to 220 µg/m^3 in Nanjing. Blue bars represent aromatic-OOMs, red bars represent alkane-OOMs, and dark green bars represent monoterpene-OOMs.
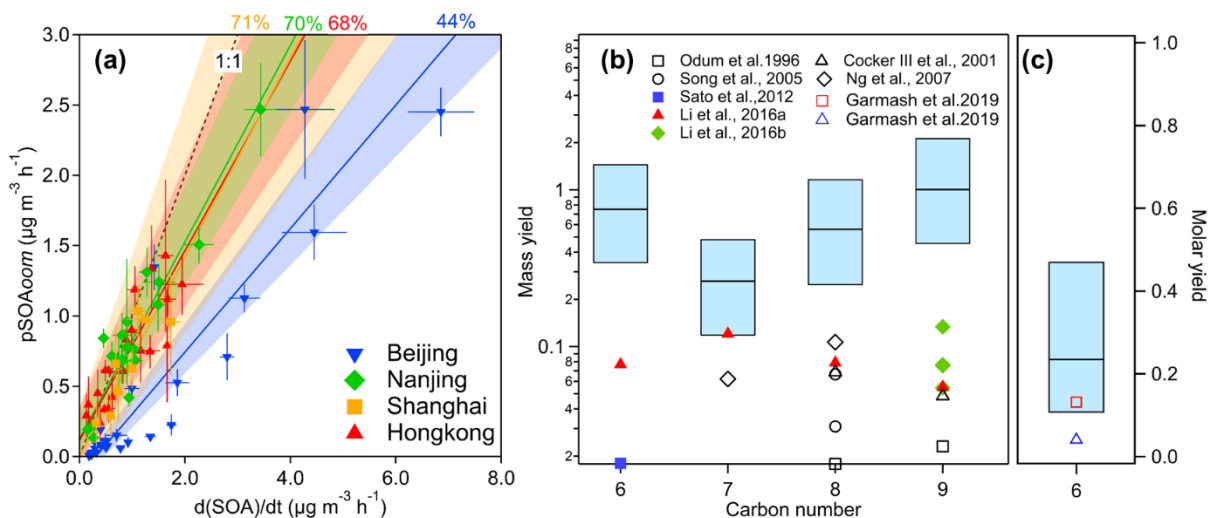
Fig. 4. (a) Contribution of OOMs condensation to the formation of secondary organic aerosol (SOA). Blue triangles, green diamonds, brown squares, and red triangles represent Beijing, Nanjing, Shanghai and Hong Kong, respectively. Error bars indicate the statistical (1σ) uncertainties in fitting the SOA formation rate during each episode, as well as averaging the OOM condensation flux in the corresponded episode. Shaded areas represent the 90 present confidence intervals of a linear fit to the data for each megacity. (b) **Mass yield of C6-C9 aromatic-derived OOMs. (c) Molar yield of C6 aromatic-derived OOMs.** Boxes represent percentiles of the calculated mass/molar yield obtained in this study. The lower, middle, and upper horizontal lines of the boxes represent percentiles of 25, 50, and 75. The makers are mass/molar yield of SOA got from labs[31-37].