Article

Emissions and potential tracer screening of semivolatile/intermediate-volatility organic compounds from urban vehicle fleets

Check for updates

Yajun Wu[1], Jianfei Peng[1] ✉, Xiaoguo Wang[2], Peiji Liu[1], Yan Liu[1], Fuyang Zhang[1], Jinsheng Zhang[1], Bin Sun[1], Jingqiao Zhang[3], Kai Song[4], Pengfei Song[1], Lin Wu[1], Ting Wang[1], Song Guo[4] & Hongjun Mao[1] ✉

Semi-/intermediate volatile organic compounds (S/IVOCs) are important precursors for secondary organic aerosols (SOA) and ozone formation. Vehicle emissions from real-world vehicle fleets are significant anthropogenic source, but their emission profiles and chemical fingerprints remain inadequately characterized. Here, we combined tunnel observation with comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry to investigate vehicular S/IVOCs emissions. We identified 256 vehicle-related compounds with fleet-average emission factors (EFs) of 16.4 ± 12.1 mg·km^{-1}·veh^{-1}, comprising 67.2% VOCs, 24.3% IVOCs, and 8.5% SVOCs. VOCs accounted for the majority of ozone formation potential (OFP, 84.8%), whereas VOCs, IVOCs, and SVOCs contributed to SOA formation potential at different times. Importantly, our speciated-based SOA estimation enhanced SOA production estimates by 44.1–76.9% compared to traditional approach. We identified eight potential vehicle-related tracers in S/IVOC range through volcano plots and hierarchical clustering analysis, which could benefit future source apportionments. Our work also offers a novel perspective for screening tracers from various sources beyond vehicle-related emissions.

Atmospheric organic compounds serve as critical precursors for secondary organic aerosols (SOA) and ozone formation, profoundly impacting air quality[1–3], climate change[4], and human health[5,6]. The oxidation efficiency and contribution to SOA formation of organics are largely regulated by their distribution between gas and particle phase[7,8], which are further determined by their volatility. Based on effective saturation concentrations (C*), atmospheric organic compounds could be categorized into volatile organic compounds (VOCs), intermediate volatile organic compounds (IVOCs), semi-volatile organic compounds (SVOCs), and low-volatility organic compounds (LVOCs)[9–11]. Laboratory studies have demonstrated that gaseous S/IVOCs, which may encompass hundreds to thousands of chemical species but are often overlooked in traditional inventories, generate SOA with more efficiently than VOCs due to their higher potential to form low-

volatility products upon atmospheric oxidation[7,12–14]. Therefore, a full-volatility framework for organics would significantly improve the accuracy of air quality models and enhance our understanding of their contribution to SOA[15].

Currently, our knowledge on S/IVOCs remains limited due to the analytical challenges associated with accurately identifying and quantifying individual S/IVOCs species. This is particularly evident when compared to the extensive research and emission reduction efforts on VOCs[16]. Conventional analytical techniques, e.g., gas chromatography-mass spectrometry (GC-MS), often struggle to resolve most S/IVOCs at the molecular level, resulting in a substantial fraction of co-eluting compounds as unresolved complex mixture (UCMs)[9,17–24]. Despite efforts to categorize the UCMs—such as the work by Zhao et al.[9] who classified UCMs into

[1]Tianjin Key Laboratory of Urban Transport Emission Research, College of Environmental Science and Engineering, Nankai University, Tianjin, China. [2]Tianjin Eco-Environmental Monitoring Center, Tianjin, China. [3]State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing, China. [4]State Key Joint Laboratory of Environmental Simulation and Pollution Control, International Joint Laboratory for Regional Pollution Control, College of Environmental Sciences and Engineering, Ministry of Education, Peking University, Beijing, China. ✉e-mail: pengjianfei@nankai.edu.cn; Hongjunmao@nankai.edu.cn

unidentified branched alkanes (b-alkanes), n-alkanes, and unidentified cyclic compounds using the volatility bin methods based on retention times of n-alkanes—a significant portion (80–90%) remains unspecified[25]. Considering un-speciated UCMs, the speciation of I/SVOC remains rudimentary, lacking essential structural details, such as carbon skeletons and reactivities. Additionally, the absence of a comprehensive emission inventory that includes chemical composition and volatility profiles poses a major barrier to accurately identifying the sources of SOA in air quality models. Recently, advanced multi-dimensional chromatographic technologies, e.g., comprehensive two-dimensional gas chromatography (GC × GC) coupled with high-resolution mass spectrometry, have offered improved separation capabilities for atmospheric complex mixtures[26]. GC × GC is recognized for its enhanced sensitivity, broad selectivity, and high peak capacity, achieved through the sequential connection of two capillary columns featuring complementary stationary phases[26,27]. GC × GC has been utilized to investigate organic emissions from diverse sources, including cooking emissions[28,29], biomass coal combustion[30–32], and vehicle exhausts[33,34]. Compared to one-dimensional GC, GC × GC significantly improves the resolution and sensitivity while effectively reducing the UCMs and coelution[35].

On-road vehicles are recognized as a major contributor to S/IVOCs in urban environments[36]. To date, most previous research on vehicular S/IVOC emission were conducted through dynamometer tests under well-controlled conditions[9,33,37–39]. These laboratory tests typically involved a limited number of samples for individual vehicle testing, which failed to accurately represent real-world scenarios due to tremendous variations in vehicle fleet composition, driving behaviors, and emission standards[17,19,33,40]. In contrast, tunnel studies can effectively capture emission from a large number of fleet vehicles in motion, providing a more comprehensive view of real-world emissions[41]. However, the research of S/IVOCs from fleet vehicles remains scarce. Only two studies by Tang et al.[42] and Fang et al.[43] have reported total emission of IVOCs in road tunnels, categorizing them into broad groups based on n-alkane using traditional thermal desorption (TD)-GC-MS techniques, without specifying all individual species. The molecular-level characteristics of the volatility of S/IVOCs from on-road vehicles under real-world condition have not yet been reported.

Atmospheric organic components exhibit strong source specificity, serving as crucial tracers for identifying emission sources and assessing their impacts on the atmospheric environment[44]. Compounds such as BTEX (benzene, toluene, ethylbenzene, and xylenes), ethylene, acetylene, isopentane, and n-pentane have been widely utilized as tracers of vehicle exhausts[44]. However, most of the current vehicle exhaust tracers fall within the VOC range, where distinction can often be blurred due to their collinearity and co-emission from other pollution sources. For instance, BTEX could originate from industrial emissions[45], solvent usage[46], and vehicle exhausts[19,45]. Similarly, ethylene, acetylene, isopentane, and n-pentane could serve as typical tracers for vehicle exhausts[19] and coal combustion[47]. These overlaps introduce significant uncertainties in the source apportionments of vehicle emissions, complicating urban air pollution control efforts and regulatory compliance. To address this limitation, it is essential to employ advanced detection technologies to expand the range of source tracers and establish a comprehensive database for various emission sources. Recently, certain S/IVOC species have been identified as tracers to distinguish different sources; for instance, acenaphthylene is recognized as a characteristic indicator of wood combustion, while pyrene and benzo[a]anthracene are used as tracers of coal combustion[24]. Given the significance of vehicle emissions on the urban atmospheric environment, it is imperative to identify new vehicle-related tracers in the S/IVOC range for effectively distinguishing vehicle emissions from other sources and improving the accuracy of source apportionments.

In this study, a tunnel observation in a megacity in China was conducted to characterize the gaseous S/IVOCs emitted by urban vehicle fleets. Vehicular S/IVOCs were identified and quantified at the molecular level through non-targeted analyses with TD-GC × GC-TOFMS. Furthermore, we explored distinct emission characteristics through a comprehensive analysis of the diurnal variations, volatility distributions, and SOA and ozone formation potential of vehicular S/IVOCs. Finally, we screened and identified potential vehicle-related organic tracers by combining various data analytical approaches. These potential tracers, along with conventional VOC tracers, would offer valuable insights into the vehicle emission characteristics and would potentially benefit future source apportionment research.

Results
Emissions of gaseous organic compounds from vehicle fleets in the tunnel

The fleet composition, traffic flow, and average speed in the tunnel exhibit typical urban traffic features. Figure 1b–d illustrates the diurnal traffic flow and vehicle emission standards observed during the measurement. The mean vehicle flow was 10217 ± 2517 vehicles every day, with the peak flow ranging from 492 to 857 vehicles per hour. Traffic intensity was higher during the evening rush hour compared to the morning. As shown in Figs. 1c and S1, gasoline vehicles (GVs) constituted the majority of the vehicle fleet (83.7%) inside the tunnel, especially during the daytime, followed by hybrid electric vehicles (HEV, 13.5%) and diesel vehicles (DVs, 2.6%). However, the proportion of DVs surged to 6.0% at night, significantly higher than in the daytime (2–3%), due to the daytime restrictions on DVs in the urban areas of Tianjin. The average speed of vehicles was 44.7 ± 2.7 km/h during the measurement, higher at midnight and lower during rush hours. These diurnal variations in both traffic density and speed correspond to the travel times of local residents commuting from work.

Besides, temperature and relative humidity during each sampling period were recorded and presented in Figs. S2–S5. Correlation analyses between these meteorological parameters and the concentration differences of organics (outlet minus inlet) were conducted for each time period. The results suggest that elevated temperature and reduced humidity may slightly increase organic compound concentrations in the tunnel; however, the correlations remained weak due to the limited variability of these parameters. Traffic volume and fleet composition were still the dominant factors influencing organic compound emissions in the tunnel.

The EFs of measured gaseous organic compounds were obtained using Eq. 1. The EFs for total measured organics within tunnel were 16.4 ± 12.1 mg·km^{-1}·veh^{-1}, with values ranging from 1.3 mg·km^{-1}·veh^{-1} to 83.6 mg·km^{-1}·veh^{-1} (Fig. 2a), comparable to those measured in the Zhujiang tunnel[42] of Guangzhou (16.77 ± 0.94 mg·km^{-1}·veh^{-1}) and the Changjiang tunnel[43] of Shanghai (24.9 ± 7.8 mg·km^{-1}·veh^{-1}) (Fig. 2d). The average measured EFs in tunnel environments were higher than that of GV exhaust but lower than that of DVs measured in chassis dynamometer testing. Besides, the EFs for organics peaked during the midnight period (2:00–4:00), reaching 34.8 ± 21.7 mg·km^{-1}·veh^{-1}, followed by the morning and evening (Fig. 2a). Despite significantly reduced traffic flow (Fig. 1b) and concentration (Fig. S3) at midnight, the EFs for organics were higher compared to daytime levels, primarily due to the increased proportion of DVs (Fig. S1). Furthermore, we reconstructed the fleet EFs based on the vehicle fleet composition within the tunnel and the single-vehicle EFs derived from dynamometer tests (Supporting Information). A comparison between the reconstructed emission factors (EFs_cal) and the measured EFs showed good agreement, as illustrated in Fig. S6. This consistency demonstrates the reliability and effectiveness of capturing the temporal trends in emissions across different time periods in the tunnel. Notably, the EFs_cal (31.9 mg·km^{-1}·veh^{-1}) was slightly lower than the measured EFs (34.8 ± 21.7 mg·km^{-1}·veh^{-1}), indicating that emissions during this period may have been influenced by a higher proportion of high-emitting vehicles or non-vehicular emission sources within the tunnel.

The chemical composition of measured organics in the tunnel is depicted in Fig. 2b, c, with the major species detailed in Fig. S7. Among speciated compounds, single-ring aromatics (SRAs) dominated the total measured organics emission (48.3%), followed by b-alkanes (15.0%), oxygenated compounds (14.4%), nitrogen compounds (8.6%), and n-alkanes (7.8%). Polycyclic aromatic hydrocarbons (PAHs) accounted for only 2.4%
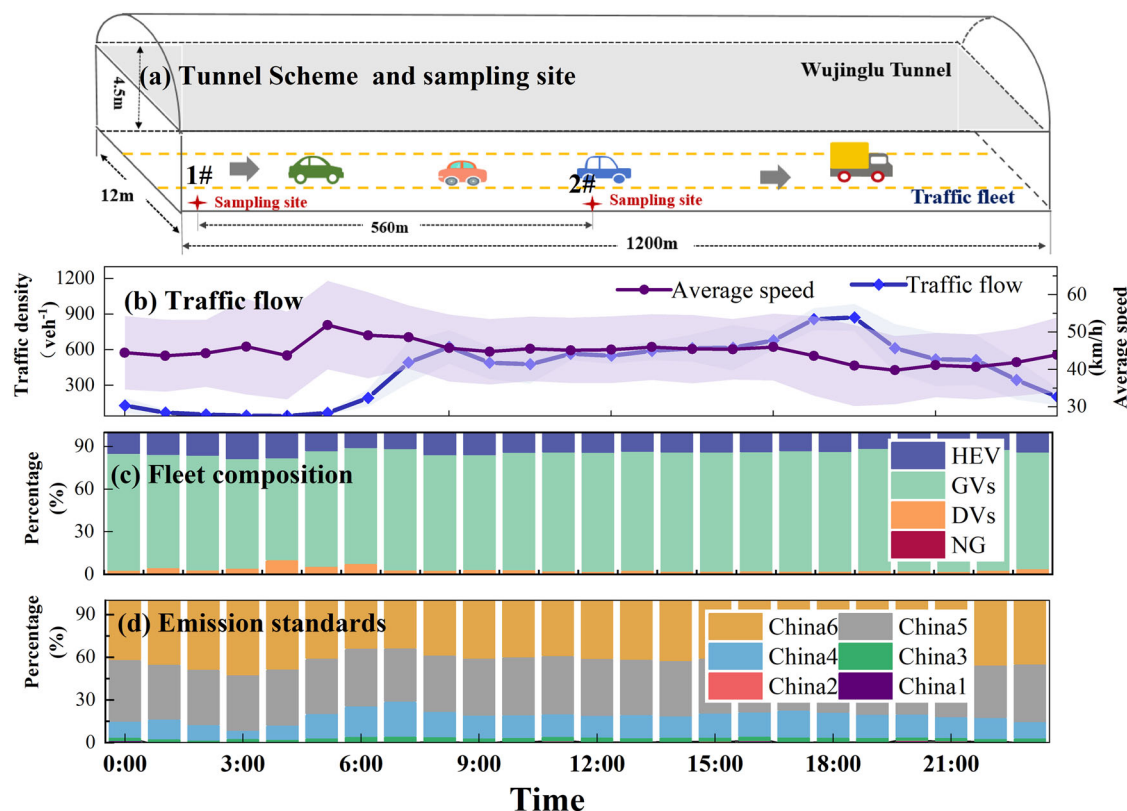
Fig. 1 | Schematic of tunnel and traffic characteristics. a Layout of the Wujinglu tunnel. Site1# and 2# refer to the two sampling locations at the tunnel entrance and exit, respectively. **b** Diurnal variation of traffic flow and vehicle speed; the shaded areas (blue and purple) represent ±1 standard deviation based on data collected over multiple days. **c** Fleet composition and **d** vehicle emission standards observed in the tunnel during the campaign; GVs, HEV, NG, and DVs represent gasoline vehicles, hybrid electric vehicles, natural gas vehicles, and diesel vehicles, respectively.

of total EFs. Besides, ketones were the most abundant oxygenated compounds, contributing 4.1% to the total EFs, followed by esters (2.8%), alcohols (1.9%), and acids (1.9%). The variety of compound categories highlights the inherent chemical complexity. In addition, SRAs and b-alkanes were predominant within VOCs, constituting 69.7% and 11.5% of the VOC emissions, respectively, while b-alkanes (29.8%), oxygenated compounds (28.0%), and n-alkanes (18.3%) were the main IVOC contributors, consistent with previous findings[42,43]. Regarding molecular structure, the majority of organic compounds in the tunnel with an H/C ratio > 0.5 and an O/C ratio < 0.3 (Fig. S8). The double bond equivalent (DBE) offers insights into the unsaturation levels of compound molecules, with higher values indicating greater unsaturation[30]. Species with DBE values ranging from 3 to 8 dominate the organic emissions, suggesting a high level of unsaturation in the tunnel. This was likely ascribed to the higher proportion of aromatics from the vehicle emissions, characterized by a more compact molecular structure, which has a lower H/C ratio and a higher DBE value. Compared with the findings of Zhao et al.[9], where the dominant fraction of organic compounds remained unspeciated, our results greatly complement the molecular fingerprint information of organic emissions from vehicle fleets.

The composition of gaseous organic compounds showed significant diurnal variation (Fig. 2b). For example, nitrogen compounds accounted for 6.0–14.7% of the measured organics at midnight, compared to 0.7% to 3.5% during the daytime. The proportion of oxygen-containing compounds increased from 12.1% during the daytime to 16.8% at night, with notable increases in ketones, esters, and acids. On the other hand, the proportion of SRAs and b-alkanes during the daytime was significantly higher than at night, with compounds such as toluene, m/p-xylene, and b-alkanes previously identified as tracers of GV exhaust. The proportions of n/b-alkanes were markedly higher in the early morning (7:00–9:00) than in other

periods. There was a small difference in components between the EFs at noon (13:00–15:00) and late afternoon (17:00–19:00) due to similar fleet composition and traffic flow, with SRAs constituting the majority, ranging from 58% to 60%, followed by b-alkanes. Besides, due to instrumental limitations, this study focuses solely on the emission characteristics of organic compounds within the C6 to C30 range, excluding those with molecular weights below C5. This limitation may introduce biases in the conclusions regarding the distribution of VOCs and their subcategories. Future studies would incorporate a variety of instruments to extend the analysis to a broader range of VOCs, thereby providing a more comprehensive understanding of emission characteristics.

Volatility distributions of gaseous organic compounds and their contribution to secondary pollutants

The volatility of organics largely determines their existing form in the atmosphere, affecting their oxidation efficiency and contribution to the pollutants formation[11,38]. The observed vehicular organics displayed a broad range of volatilities, with VOCs (67.2%) and IVOCs (24.3%) being the predominant groups (Figs. 2a and 3g–j). The average volatility distribution closely resembles that observed in gasoline exhausts in previous research[38]. The distribution of volatiles across different sampling periods showed obvious differences in the tunnel. On one hand, semi-volatility components were significantly more abundant at night (Fig. 3g) compared to the daytime, with the proportion of SVOCs at night (9.1 ± 6.3%) being nearly double that during the day (1.9–5.8%). On the other hand, the proportion of IVOCs during the morning peak (59.9 ± 11.5%) was slightly higher than during other periods, whereas VOCs components significantly increased during the afternoon (17:00–19:00) and noon hours (13:00–15:00), with VOCs constituting 77.6 ± 9.2% and 75.2 ± 9.8%, respectively. Recent studies have reported that gasoline-related emissions[33,38] are mainly composed of
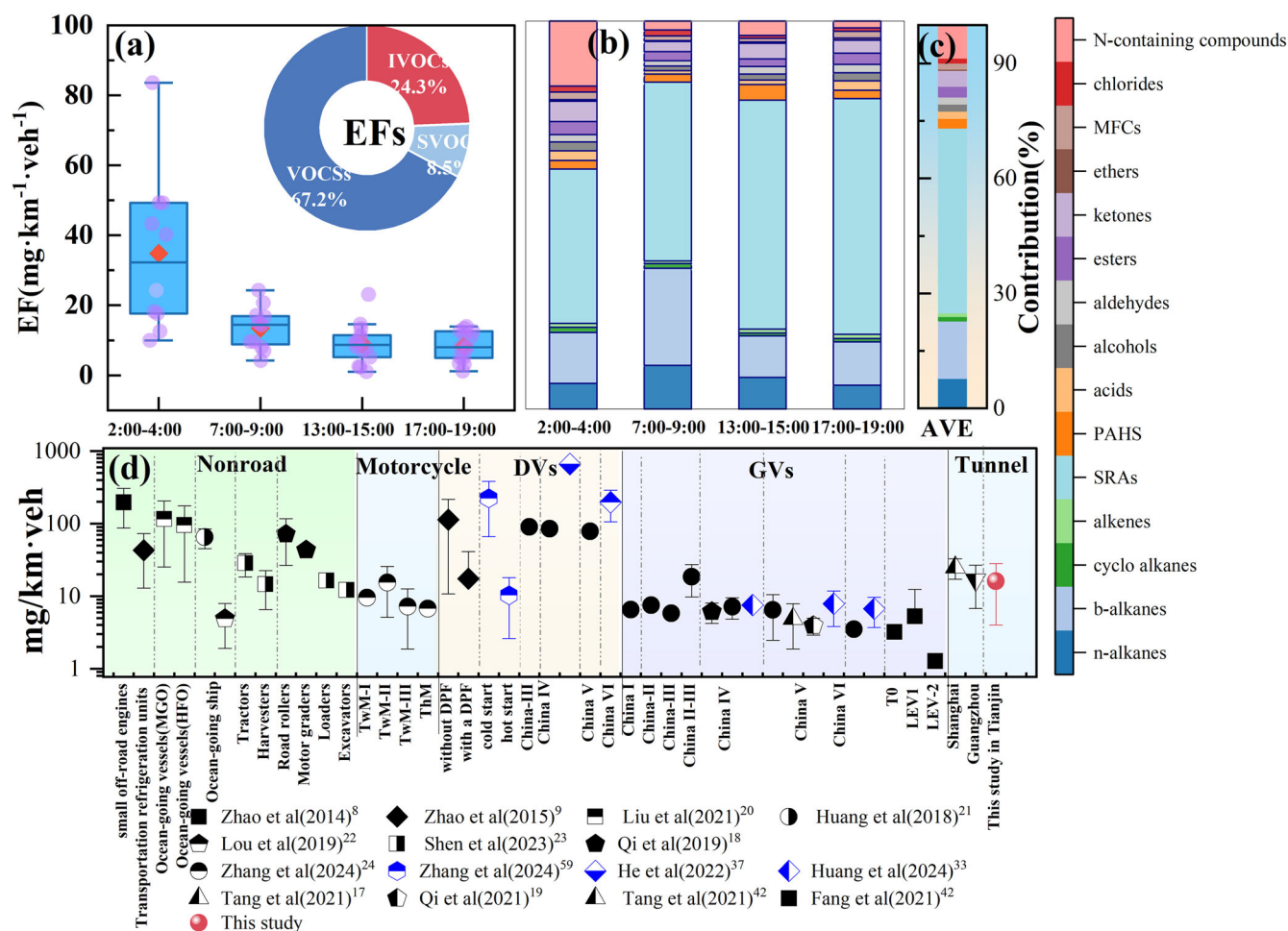
Fig. 2 | Emission profile of gaseous organics within the tunnel. a EFs for different sampling periods; purple dots represent individual sample EFs, and red dots indicate the mean values for each period. Boxes represent the interquartile range (25–75th percentiles), the centerlines indicate the medians, and the whiskers denote the full data range. The pie chart shows the volatility distribution of organics during the entire campaign. **b** Composition of organics during different sampling periods.

SRAs and MFCs refer to single-ring aromatics and multifunctional organics. **c** Average composition of organics over the entire campaign. **d** Comparison of tunnel-derived EFs with those from other mobile sources; TwM, ThM, and DPF indicate two-wheeled motorcycles, three-wheeled motorcycles, and diesel particulate filters, respectively. Blue dots denote measurements obtained by GC × GC-MS, while black dots were obtained by GC-MS.

VOCs, while DVs[9,34,37,38], and lubricants[48] typically contain a higher proportion of relatively LVOCs. Our results were consistent with the fleet composition during the sampling period, indicating a significant relationship between vehicle fleets and volatility distribution.

Although there was little variation in total EFs among different sampling periods (Fig. 2a), significant variations were observed in the distribution and composition of volatile compounds. Notably, the variations in chemical species across different sampling periods were predominantly found within the S/IVOCs range (Fig. 3a–d). For example, the proportions of b/n-alkanes in the volatility of 4–6 (IVOCs) were markedly higher in the morning (7:00–9:00) compared to other periods (Fig. 3b). It is noteworthy that these differences in n/b-alkanes were consistently observed across nearly all sampling days. Although traffic volume during the morning rush hour was slightly lower than in the evening, cold-start conditions were more prevalent in the morning, as most vehicles had been parked overnight. The resulting lower engine and exhaust system temperatures likely contributed to the more pronounced increase in alkane concentrations observed during this period[17,19]. Research has demonstrated that cold-start operation results in higher IVOC EFs compared to hot-start operation, primarily due to the reduced efficiency of the catalytic converter during cold starts, particularly with respect to alkanes that are not oxidized[17,19]. As shown in Fig. 3h, j, compared to the evening rush hour (17:00–19:00), which aligns with traffic volume and fleet composition, the proportion of IVOCs during the morning

(38.2%) was significantly higher than that during the evening peak (20.5%). Meanwhile, the carbon number distribution during the morning rush hour exhibits a marked increase in both the light alkane fraction (C5–C8) and mid-to-high carbon species (C16–C19) (Fig. S9). Lubricating oil, which contains substantial amounts of C18–C25 alkanes[49], further supports the influence of cold starts, unburnt fuel, and lubricating oil evaporation on emissions. These combined factors result in a significantly higher proportion of alkane emissions during the morning rush hour compared to other time periods. Additionally, nitrogen compounds (10.3%) and acids (2.5%) were found to increase significantly at 2:00–4:00 am, particularly in the volatility of 2 (SVOCs), with phthalimide as the primary contributor. To clarify the origin of elevated nitrogen-containing compounds during 2:00–4:00 a.m., we incorporated multiple supporting observations. Video surveillance (Fig. S10) confirmed the presence of barbecue carts with open-flame grills in the tunnel around 3:30 a.m., during periods of near-zero traffic (Fig. 1), suggesting a non-vehicular source. Concurrently, particle number concentration (PNC) peaked sharply at 3:00 a.m., with a size distribution in the Aitken mode (Fig. S11)[50]—indicative of incomplete combustion[51]. Moreover, GC × GC-TOFMS results showed nitrogenous heterocycles and amines consistent with biomass burning profiles reported in the literature[30]. This evidence strongly indicates that biomass combustion is the primary driver of the elevated nitrogen-containing compounds observed during midnights. The burning charcoal likely produced a large
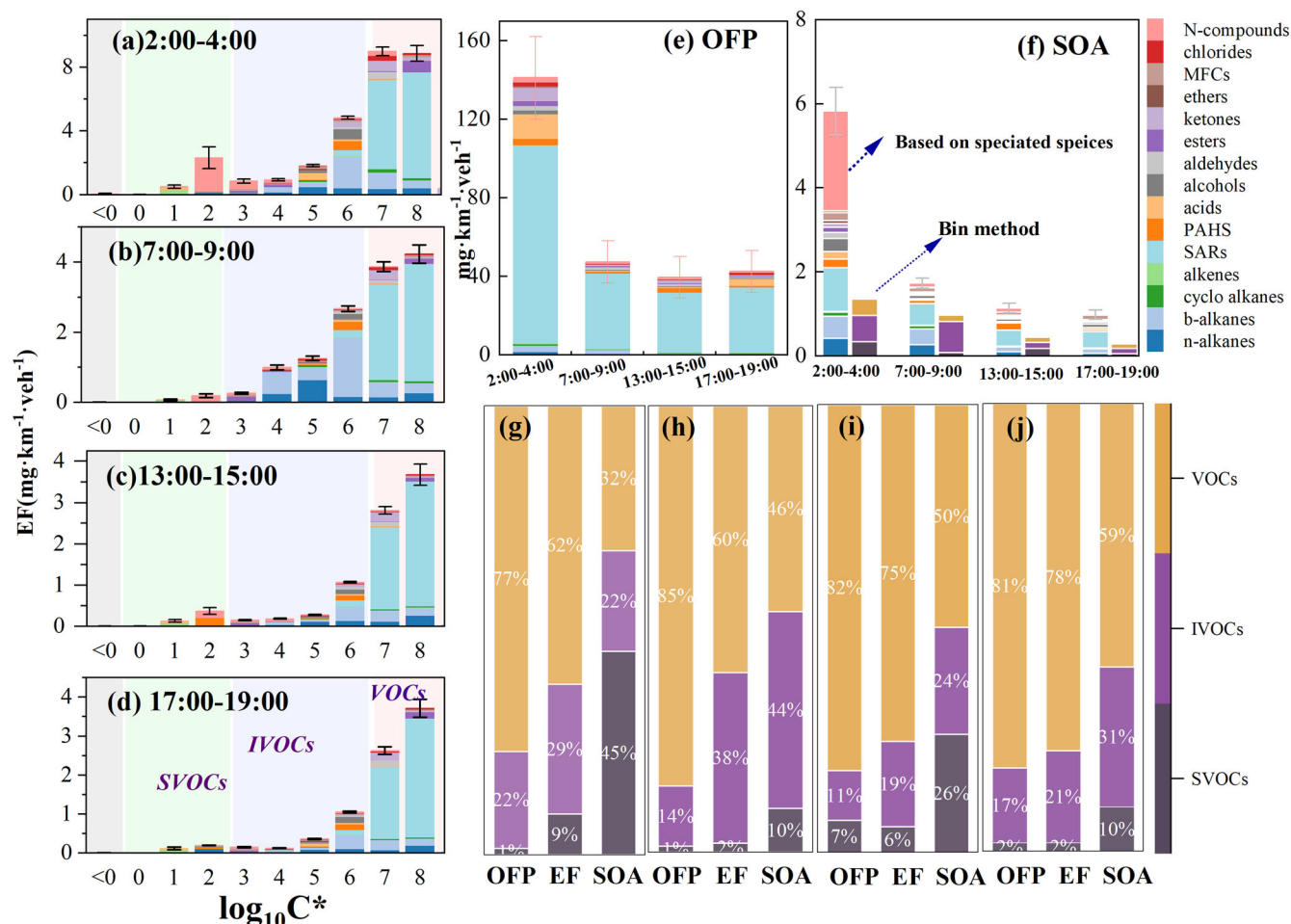
Fig. 3 | The volatility distribution and contribution to secondary pollutants of organics in the tunnel. a–d Represent the volatility distribution of S/IVOC during different sampling periods; Background shading denotes distinct volatility ranges. **e** Contribution of organics to ozone formation potential during sampling periods. **f** Contribution of organics to SOA during sampling periods; The method based on speciated species represents the SOA estimation quipped with detailed chemical parameters based on GC × GC-TOFMS. Bin method represents the calculation using the parameters of n-alkane corresponding to each bin grouped with retention time of n-alkane. **g–j** Represent the contributions of VOC, IVOC, and SVOC to EFs, SOA, and ozone formation during the sampling period of 2:00–4:00, 7:00–9:00, 13:00–15:00, and 17:00–19:00, respectively.

number of nitrogen compounds, further emphasizing the considerable impact of open-air barbecuing on urban organic emissions. The result also aligns with the previous conclusion that the presence of high-emitting vehicles or non-vehicular emission sources within the tunnel during the midnights. Our findings clearly indicate that S/IVOCs provide valuable insights into the sources of pollutant emissions, which can aid in accurate source identification and the development of effective emission control strategies.

The Eqs. S5 and S6 were used to calculate the formation potential to secondary pollutants in this study. The contributions of various groups to ozone and SOA formation potential are illustrated in Fig. 3e–j. Overall, VOCs accounted for 84.8% of the estimated ozone production and 46.7% of SOA formation potential, compared to 67.2% in EFs during the whole campaign (Fig. S12). Aromatics took up 81.4% of the total OFP, followed by oxygenated compounds (7.7%) and PAHs (2.6%). Benzene, 1,3-dimethyl-, o-xylene, benzene, 1,2,4-trimethyl- were the main VOC species that contribute most to OFP. Additionally, benzene, 1,2,3,5-tetramethyl-, naphthalene, 1-Hexanol, 2-ethyl-, benzoic acid, and benzo[b]fluoranthene were the primary IVOCs contributors for OFP (Fig. S13). Although S/IVOCs accounted for only 32.8% of EFs, they predominant the SOA formation potential (53.3%) in the tunnel (Figs. 2a and S12). The substantial contribution of S/IVOCs to SOA could be attributed to their relatively higher yields and reaction rates with OH radicals, as evidenced in various emissions

studies, including cooking emissions[28], gasoline exhaust[8,19,33,52], diesel exhausts[9], and coal combustion[53,54]. Aromatics accounted for 25.4% to the SOA estimation, closely followed by nitrogen compounds (23.7%), oxygenated compounds (21.4%), and b-alkanes (11.7%) (Fig. 3f). Alcohols (5.5%), multi-functionality compounds (5.3%), and acids (3.2%) were the main oxygenated compounds for SOA generation in the tunnel. From the perspective of species, 1-hexanol, 2-ethyl, bis(2-ethylhexyl) phthalate, benzo[b] fluoranthene, naphthalene, and n-Hexadecanoic acid were found as significant S/IVOCs contributors to SOA formation potential.

VOCs dominated ozone formation potential, while VOCs, IVOCs, and SVOCs played varying roles in SOA estimation throughout different times of the day. SOA formation potential at 2:00–4:00 was much higher than in other periods, due to the larger EFs and elevated S/IVOCs proportion (38.1%). Nitrogen compounds (40.8%) were the main contributors to SOA formation potential during 2:00–4:00, followed by aromatics (21.3%). Despite minimal differences in EFs in the morning and evening sampling time (Fig. 2a and Fig. 3h, j), SOA formation during the morning peak was 52.6% higher than during the evening peak due to the high proportion of IVOCs, particularly the b-alkanes and n-alkanes with volatility between 4 and 5 (Fig. 3b). As discussed above, there might be high-emitting DVs during morning rush hours, which likely emit more IVOCs. These IVOCs have a higher yield to form LVOCs during atmospheric oxidation, potentially contributing to the abundant SOA formation in the succeeding

daytime. This finding indicates that we might underscore the pivotal role of high-emitting DVs in atmospheric SOA formation, which could be an important aspect of future fine particulate control.

To investigate the influence of chemical speciation on the accuracy of SOA estimation, we applied the parameterization method (Eq. S6) using two different types of input data: a bin-based approach using surrogate n-alkanes and a species-resolved approach incorporating compound-specific reactivity and yield data. The bin-based approach, commonly used in previous studies, estimates SOA formation by assigning the OH reactivity and SOA yield of a representative n-alkane to all compounds within the same volatility bin. In contrast, the species-resolved approach calculates SOA formation potential for each identified VOC and S/IVOC species individually. This method enables the use of a more accurate volatility distribution, a broader range of chemical species, and compound-specific OH reaction rate constants and SOA yields, thereby improving the fidelity of the estimation. As shown in Fig. 3f, the speciated compounds method significantly enhanced SOA production estimates by 44.1–76.9%. Previous laboratory studies have found that the measured substantial SOA formation potential from vehicle exhaust cannot be fully explained by the SOA yield of total VOCs[12,52,55], even with some S/IVOC precursors[41]. Our results elaborate that UCM proportion through appropriate techniques could partially address the underestimation of SOA production. Compared to the significance of VOCs in ozone formation, enhancing the control of S/IVOCs would be more effective in mitigating fine particle pollution.

Screening of organic markers under real-world driving conditions

Atmospheric organic concentration and fleet organic emissions in the tunnel are utilized for identifying vehicle emission tracers. We first use partial least squares discriminant Analysis (PLS-DA), a supervised method for classifying grouped data, and key components identification[56,57]. The results showed that samples from different periods in the tunnel flock together closely but are far away from ambient atmosphere samples (Fig. S14), revealing considerable differences in the chemical groups of these samples. Therefore, we considered the tunnel emissions as a whole for further analysis.

Secondly, a volcano plot is employed to visualize the *P*-value from *t*-test alongside the fold change (FC) values, facilitating the identification of differential species between vehicle fleet emissions and the urban atmosphere. The details are available in the Supporting Information. In this study, organic compound meeting the criteria of FC > 1.5 and *P*-value < 0.05 depicted as red dots in Fig. 4a, with their size proportional to the EFs. A total of 20 species are identified, comprising 12 IVOC species and 8 VOC species, including 7 SRAs, 2 aldehydes, 3 b-alkanes, 1 nitrogen compounds, 2 alkenes, 1 cycloalkane, 1 alcohol, 2 esters, and 1 multi-functionality compounds (Fig. 4 and Table S1). These compounds exhibit larger FC and lower *P*-values, highlighting their potential as significant molecular tracers for distinguishing vehicle fleet emission from other sources. We further analyze the diurnal trends of these species to testify their homology (Fig. 4b). It is found that these potential tracer compounds exhibit distinct diurnal features. Some compounds were significantly higher at 2:00–4:00, while others showed relatively stable EFs among different sampling periods. We want to emphasize that vehicle emission might not be the only emission source in the tunnel; unknown emission sources might exist in the tunnel. For example, as discussed above and illustrated in Fig. S10, emissions during 2:00–4:00 in the tunnel were not only from vehicles but also from burning charcoal.

To eliminate the impacts of other potential sources and identify the vehicle-related emission tracers, we further perform hierarchical clustering analysis (HCA) analysis, a method that organizes data into a hierarchy of clusters, typically visualized by a dendrogram[58]. We clustered the normalized EFs of selected species using HCA, with further details provided in the Supporting Information. Additionally, we included known tracers from previous research[19,33,34,59] as auxiliary species, including specific PAHs (benzo[b]fluoranthene, naphthalene, benzo[a]pyrene), 1-hexadecanol, and

n-alkanes to enhance the discrimination of the screening results. It is important to recognize that vehicle emissions are not the sole source of these compounds; for instance, biomass burning and biological processes significantly contribute to naphthalene emissions[30], while fatty alcohols are common indicators of cooking emissions[29,60]. However, in the semi-enclosed tunnel environment, vehicle emissions are likely the predominant source of these species. The dendrogram effectively illustrates the clustering relationships and similarities among the species[61]. The calculated silhouette score (SC)[61] shows that three clusters represent the optimal grouping (Supporting information and Fig. S15). During the clustering process, similar compounds are progressively grouped together (Fig. 5b). For example, heptane, 3-methyl-, benzaldehyde, 2-methyl-, benzene, pentyl-, and p-Cymene cluster closely, indicating a high degree of similarity. These species are associated with b-alkanes and naphthalene, which are recognized tracers of GV exhaust. Thus, this group of compounds is likely derived from GV emissions. Additionally, benzene, (1-methylbutyl)- clusters with 1-hexadecanol, benzo[a]pyrene, and benzo[b]fluoranthene, which are primarily associated with DV exhaust. This suggests that benzene, (1-methylbutyl)- may serve as a potential organic marker for DVs. Notably, other species, i.e., 2-propenoic acid, butyl ester, and 5-undecene, 4-methyl-, et al. are also grouped together. As shown in Fig. 4b, the diurnal patten of these species are similar, with significantly higher emission observed at night. This variability and nocturnal peak make them unreliable as markers for vehicle emissions.

Considering that the organic tracers should remain relatively stable in the environment, the species with low reactivity and long lifetime were typically prioritize in our study. OH· radical is one of the significant oxidants in the atmosphere, which dominates the degradation of organics. Therefore, we excluded highly reactive species in the atmosphere based on the reaction rates with OH radicals, e.g., α-Methylstyrene, piperidine, and 3-phenyl. In addition to the uniqueness of source emissions, measurability is also a key aspect for effective tracers. Therefore, an EF threshold of 0.05% is also utilized to exclude low-emission species. Eventually, heptane, 3-methyl-, 3-phenylpropyl benzoate, benzaldehyde, 2-methyl-, benzene,1-methyl-4-(1-methylpropyl)-, Benzene, 2,4-dimethyl-1-(1-methylethyl)-, and 3-Ethyl-3-methylheptane are confirmed as potential organic markers for real-world fleet vehicles (Fig. 5c). Notably, 2-Propanoic acid, butyl ester exhibited a relatively high contribution to the total EFs in this study and showed a diurnal variation pattern that closely mirrored traffic flow, suggesting a potential association with vehicular emissions. However, the species more likely to originate from industrial activities such as the use of coatings, solvents, and plastic additives, introducing uncertainty in its source attribution. Given that our analysis is based on the concentration difference between tunnel exit and entrance—under relatively stable background conditions and dominated by vehicular sources—the influence of non-mobile emissions (e.g., industrial emission) has been effectively minimized. Therefore, we speculate that 2-Propanoic acid, butyl ester is more likely linked to non-exhaust vehicle-related emissions, including tire and brake wear or off-gassing from interior materials. As current understanding of the emission pathways of such compounds remains limited, further targeted investigations are warranted to clarify their source characteristics and diagnostic value in the context of vehicle emissions in the future. These species are not conventional VOCs such as toluene, ethylbenzene, and p-xylene, but instead fall within the IVOCs and SVOCs range. The key potential organic markers identified in this study, combined with traditional VOC tracers, could significantly enhance source apportionment efforts. To evaluate these compounds as vehicle emission tracers, we compared the concentrations these tracers at an urban site (Nankai University campus, 38.98 N, 117.34 E, Fig. S16), a tunnel entrance, and inside the tunnel. Detailed sampling procedures are provided in the Supplementary Information. Results (Fig. S17) showed the highest concentrations inside the tunnel, followed by the entrance, with the lowest at the urban site, consistent with the spatial pattern of vehicle emissions.

In our field observation of the tunnel, conventional vehicular VOC species, including toluene, i-pentane, n-pentane, and ethylbenzene, were
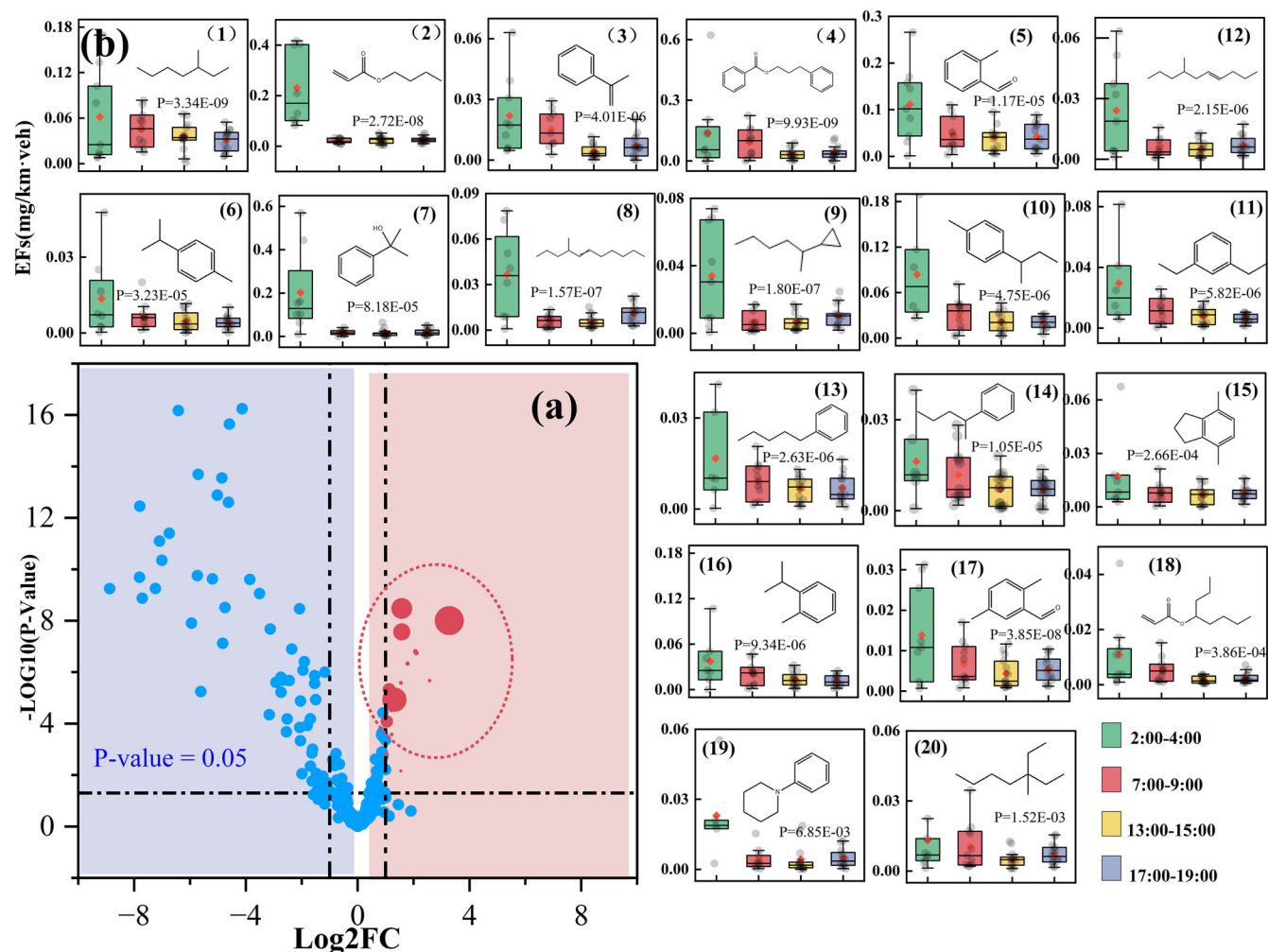
Fig. 4 | Difference analysis and tracers screening of fleet vehicles. a Volcano plot showing the fold-change (FC) and *P*-value from *t*-test comparing tunnel samples to ambient atmosphere. Red points represent species that meet FC > 1.5 and *P*-value < 0.05, with their size proportional to the EFs. **b** Diurnal variation of EFs, molecular structure, and *P*-value of selected species. Serial numbers and compound details are provided in Table S2. Bar colors correspond to the four sampling periods. Boxes represent the interquartile range (25–75th percentiles), the centerlines indicate the medians, and the whiskers denote the full data range.

simultaneously collected alongside S/IVOCs and analyzed using GC-MS[41,62]. To further validate the representativeness and reliability of the selected S/IVOC species as vehicular emission indicators, correlation analyses were conducted between these compounds and conventional vehicular VOC tracers, including i-pentane, n-pentane, toluene, and ethylbenzene[44]. As shown in Fig. S18, the T/B ratio observed in the tunnel samples was approximately 1.93, which is consistent with the typical value (~2.0) reported for urban vehicle emissions[63], indicating that the sampled fleet emissions were representative. Additionally, correlation analyses were conducted between the selected S/IVOC species and conventional vehicular VOC tracers, including i-pentane, n-pentane, toluene, and ethylbenzene. The results (Table S6) showed moderate correlation coefficients (*r* values ranging from 0.4 to 0.7). These findings suggest that, despite differences in volatility and emission dynamics, the selected S/IVOC species exhibit emission patterns consistent with typical vehicle-related VOCs. Moreover, synchronized diurnal variation trends observed between i-pentane and the selected S/IVOCs (Fig. S19) further reinforce their association with vehicular sources. Additionally, correlation analysis (Table S6) revealed a positive relationship between the tracers and NOx in tunnel (correlation coefficient > 0.5, *P* < 0.05), further supporting their potential as vehicle emission tracers preliminary. Furthermore, the potential influence of non-vehicular emission sources within the tunnel, as discussed in the results, may introduce additional uncertainties in the identification of these markers.

Therefore, future research should focus on cross-validating these tracers through controlled laboratory experiments and independent field studies, ensuring their robustness and applicability in source apportionment efforts.

Discussion

This study highlights the significance of measuring S/IVOCs from real-world emissions, a factor often overlooked in previous research. The vehicular S/IVOC profiles obtained in this study serve as a crucial addition to the existing S/IVOC emission database and can be utilized in future composition-based emission inventory studies and volatility-based SOA simulation research. For instance, the ongoing development of the volatility basis set (VBS) approach necessitates more comprehensive S/IVOC emission data[10,15].

Source identification and apportionment are essential for developing effective emission control strategies. However, the overlap of traditional markers among different sources introduces considerable uncertainty into source apportionment results[44,45]. We emphasize that S/IVOCs, which encompass hundreds to thousands of chemical species with a broader range of volatility, provide more comprehensive information regarding origins and emission sources. For instance, this study revealed cold start emissions and the evaporation of unburned fuel and lubricating oil during the morning rush hour, indicated by the presence of large n/b-alkanes within the IVOC range. More importantly, we identified eight potential tracers for
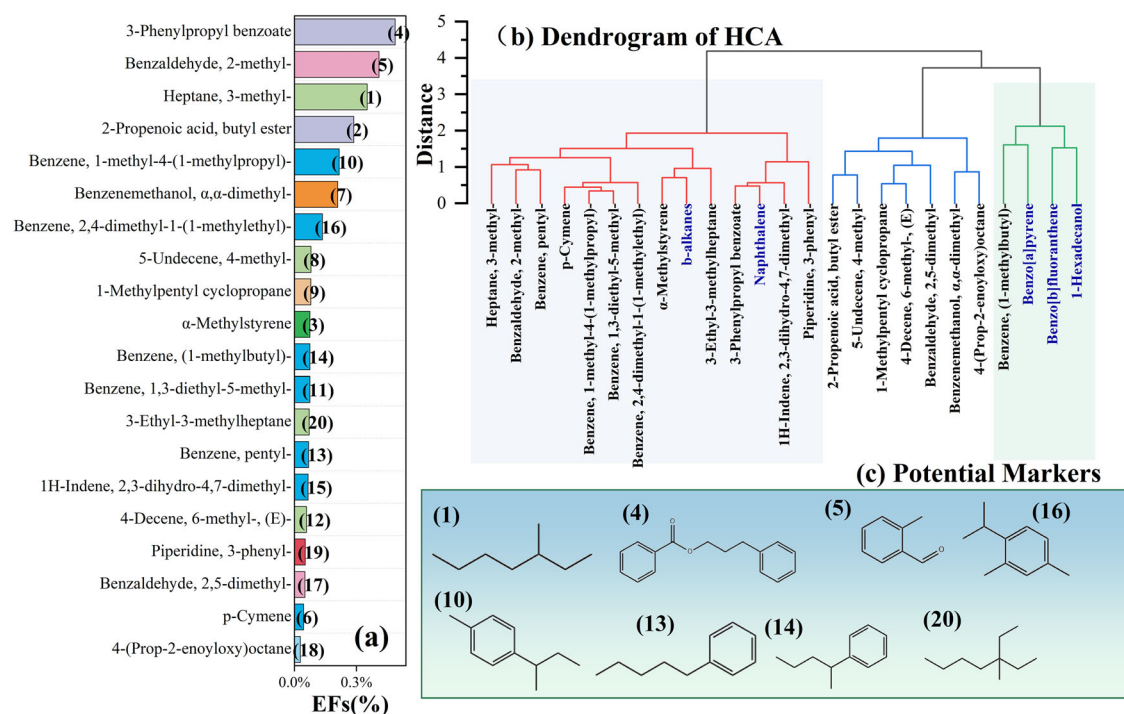
Fig. 5 | Screening and classification of potential marker compounds for tunnel fleet emissions. a EFs of selected species measured in the tunnel; Serial numbers and detailed compound information were provided in Table S2. **b** The dendrogram of HCA. Colors represent different clusters, with blue indicating auxiliary species referenced from previous studies. The horizontal axis corresponds to individual compounds, and the vertical axis indicates the relative linkage distance (dimensionless). **c** Potential markers recommended for fleet vehicles in real world.

vehicular emissions, all of which most of them fall within the S/IVOC category. The distribution of these selected tracers aligns with the spatial pattern of vehicle emissions, with the highest concentrations observed inside the tunnel, followed by the roadside site, and the lowest at the urban background site. This pattern, coupled with their positive correlation with NOx, preliminarily confirms their reliability and potential for future organic source apportionment efforts. Future research should focus on cross-validating these tracers through controlled laboratory experiments and independent field studies, ensuring their robustness and applicability in source apportionment efforts. We believe that these key tracers, when combined with widely recognized VOC tracers, could be instrumental in source apportionment and in supporting targeted strategies for air pollution control and public health protection. Additionally, our tracer screening approach that combines full-volatility measurements with comprehensive statistical methods could be employed to identify and quantify markers from various sources, thereby extending its applicability beyond vehicle emissions. These potential tracers, in conjunction with conventional VOC tracers, could facilitate a more nuanced understanding of source attribution and strategies for reducing vehicular pollution. In addition, it also offers a novel framework for screening tracers from various sources, extending its relevance beyond just vehicle emissions.

Furthermore, we emphasize that detailed S/IVOC information obtained through advanced techniques not only provides more precise emission fingerprints but also significantly addresses the underestimation of SOA production noted in previous studies. Laboratory research has indicated a 2-4 times underestimation of SOA estimation when relying solely on VOCs as precursors[12].Our findings reveal that VOCs represent only a small fraction of the potential for SOA formation, and even the inclusion of unspeciated S/IVOCs still leads to an underestimation of SOA production. This underscores the necessity of incorporating speciated S/IVOCs in future laboratory studies of SOA formation. However, it is important to acknowledge that the oxidation mechanisms and SOA yields of S/IVOCs remain significantly less understood compared to those of VOCs and relevant studies could be important research topics in the future.

Besides, we still need to stress the limitations of TD-GC × GC-TOF-MS. It does not provide a comprehensive measurement of all organic compounds. The TD process may lead to underestimation of some S/IVOC compounds due to incomplete desorption or thermal degradation. Highly polar compounds may not elute effectively from the GC column, leading to biases in volatility and polarity distribution estimates. To address these limitations and achieve a more comprehensive characterization of vehicle-emitted organics, complementary techniques such as PTR-MS and CIMS should be employed alongside GC × GC in future studies.

Materials and methods
Tunnel measurements
Urban tunnels offer an enclosed and well-defined environments for quantifying emissions from a mixed vehicle fleets under real-world driving condition[64]. A two-week observation was carried out within the Wujinglu Tunnel located in the downtown of Tianjin city, China (longitude 117.21, latitude 39.14), from August 30 to September 14, 2023. The sampling site, situated in the city center, reflects typical urban traffic characteristics, providing a reliable estimate of fleet vehicle emissions.

Two sampling sites inside the tunnel 560 meters apart from each other are designated as the entrance and exit sites[65], according to the traffic direction (Fig. 1a). The airflow within the tunnel aligns with the traffic direction. Numerous instruments were installed on the sidewalk at an elevation of 1.5 meters above the pavement at both the entrance and exit sampling sites[41]. For instance, meteorological parameters, such as wind speed, temperature, atmospheric pressure, and relative humidity were measured at both the entrance and exit sites using a portable meteorological station (VAISALA WXT 520, Helsinki, Finland). Additionally, a high-definition vehicle license plate recognition system (DS-TCG225-KN, Hik-vision Inc.) was installed for vehicle classification to ascertain the detailed vehicle information from a local traffic database using the license plate information[41,65], including fuel type, emission standards, etc.

Gaseous organics were collected at both entrance and exit site using two pre-conditioned Tenax-TA tubes (Markes, UK) in series after passing

through quartz filters, at a flow rate of 0.6 L/min (Fig. S20). Samples were gathered during the following times: 2:00–4:00, 7:00–9:00, 13:00–15:00, and 17:00–19:00. Eight tube samples would be obtained each day. A daily blank sample (Fig. S21) from the tunnel was collected and subsequently subtracted during the quantification procedure. The time series of concentrations are shown in Fig. S2.

Besides, a Tenax TA breakthrough experiment was conducted prior to formal sampling by connecting two adsorbent tubes in series. The first tube served as the sample tube, and the second as the backup tube. Both tubes were sampled simultaneously over a 6-h period to assess potential breakthrough under actual sampling conditions. No significant breakthrough was observed after 6 h of sampling (Fig. S22). After subtracting the blank volume, the amount detected in the backup tube was less than 6% of that in the sample tube, indicating that breakthrough during sampling was negligible and the retention performance of the Tenax TA tubes was sufficient for the target compounds.

Chemical analysis and quality control

Samples were analyzed using a TD system (Markes, TD-100) at a desorption temperature of 330 °C, combined with GC×GC-TOF-MS (EI-TOFMS 0620, Guangzhou Hexin Instrument Co., Ltd., Guangzhou, China) operating with a 6 s modulation period. The GC×GC-TOF-MS system consisted of a solid-state modulator (SSM1810, J&X Technologies, China) integrated with a gas chromatograph (7890 A, Agilent Technologies) and connected to a time-of-flight mass spectrometer. The first dimension GC column utilized a non-polar DB-5 MS column, while the second dimension GC column employed a mid-polarity DB-17 MS column. The GC temperature program began at 30 °C for 5 min, then ramped up to 300 °C at a rate of 3 °C/min, and was maintained for 5 min. The detailed parameters of TD and GC×GC-TOF-MS are provided in Table S3 and Table S4. Blank samples were analyzed following the same methodology, and their signals were subtracted from those of the actual samples.

The collected high-dimensional GC×GC data were imported into Canvas (V2.5, J&X Technologies, China) for visualization and further data preprocessing. Chemicals were resolved blob-by-blob at the molecular level using authentic standards, retention indices libraries (RI), the National Institute of Standards and Technology library (NIST17), and the characteristic patterns of homologs eluted in the two-dimensional chromatogram[60,66]. Detailed information was given in the Supporting Information. The actual calibrated quantities (ng) were determined using the standard curve and quantifier response. Calibration curves for all the authentic standards were thoroughly established, with R^2 values ranging from 0.95 to 0.99, highlighting the robustness of quantification (Table S3). Besides, the recoveries of the standards ranged from 79% to 105% (Table S1). Compounds lacking standards were semi-quantified using n-alkanes from the same volatility bin[8] or surrogates from the same chemical class. The uncertainties associated with semi-quantification in this study ranged from 21% to 41% (Supporting Information). Detailed information on the authentic standards, quality control, and discussions of uncertainty are provided in Supporting information. A total of 256 chemicals were (semi-)quantified, encompassing aromatics, alkanes, oxygenated compounds (including ketones, alcohols, esters, acids, and aldehydes), and other chemicals. The typical chromatograms of samples were given in the Fig. S21.

Estimation of EFs and volatility distribution

The average emission factor (EF, mg/(km·veh)) of vehicular exhaust pollutants within the tunnel could be obtained through the following formula[41]:

$$EF_s = \frac{(C_{exit} - C_{entrance}) \times A \times V_{air} \times t}{N \times L} \quad (1)$$

Where C_{exit} and $C_{entrance}$ is the concentrations of pollutants from the exit and entrance, respectively; V_{air} (m/s) represent the air velocity parallel to the tunnel; A denotes the tunnel cross-section area(m^2); t is sampling time(s); L

is the distance between sampling points (km); N is the number of vehicles that passed through the sampling points.

The volatility distribution of identified gaseous organics was determined from the effective saturation concentration (C*, μg/m^3)[38,54]. In this method, every volatility distribution of 10^n spans a range from 0.3×10^n to 3×10^n in logarithmic space, with n varying from −1 to 8. This includes LVOCs (C* < 0.3), SVOCs (0.3 < C* < 300), IVOCs (300 < C* < 3×10^6), and VOCs(C* > 3×10^6). C* values were calculated as follows[11,38,54]:

$$C_i^* = \frac{M_i 10^6 \xi_i P_{L,i}^0}{760 RT} \quad (2)$$

Where, M_i is the molecular weight of species i (g/mol), while ξ_i represents the activity coefficient of compound i in the condensed phase, which is assumed to be 1. $P_{L,i}^0$ is the subcooled liquid saturation vapor pressure of pure compound i at 298 K (Pa), deriving from the Estimation Programs Interface (EPI) Suite developed by the U.S. EPA, accessible at www.epa.gov/oppt/exposure/pubs/episuitedl.htm.

For comparison, the traditional bin method[8,9] for determining organics involved binning based on the retention time of n-alkane was also used for dividing the measured organics and further for the calculation of SOA formation. This method, typically applied to high UCM ratios and each bin was defined by the carbon number of a specific n-alkane. SVOCs, IVOCs, and VOCs correspond to the retention time ranges, >B22, B12–B22, and <B12, respectively.

Screening and classification of the gaseous organic tracers

The classical volcano plot is a visual tool that displays both fold-change (FC) and t-statistic, commonly used for identifying differential metabolites or molecules in metabolomics[67]. In a volcano plot, the x-axis represents the log$_2$FC, which indicates the magnitude of the difference between two conditions (e.g., treatment vs control). The y-axis displays the P-value of the t-test, usually transformed for –log$_{10}$(P-value)), reflecting the statistical significance of the observed differences. This dual-axis approach allows for the simultaneous assessment of both the magnitude of the change and the confidence in that change. The details are given in the Supporting Information. The emission from vehicles in the tunnel and the atmospheric concentration (reconstructed based on entrance data, Supporting Information) were used for classification and screening potential organic tracers for fleet vehicles in our study. The selection of FC > 1.5 as the screening threshold was based on a sensitivity analysis of different cutoffs (FC > 1.2, 1.5, and 2.0) in our study, which is provided in the Supplementary Information.

HCA[58] was used to further classify and confirm selected species in this study. HCA was employed to investigate the similarity and potential covariation among organic compounds, with the aim of identifying species that exhibit consistent temporal trends and emission behaviors. HCA is a technique used to group similar objects into clusters based on their characteristics, forming a hierarchical tree called a dendrogram. HCA begins by treating each object as an individual cluster, then iteratively merges the nearest pairs based on a selected distance metric, such as Euclidean distance or Pearson correlation. This merging process continues until all objects are combined into a single cluster, resulting in a tree-like structure that illustrates the nested groupings of the data. This method allows for the grouping of compounds based on their statistical relationships, helping to reveal underlying chemical associations and potential common sources. In this study, HCA was used as a supportive approach to assess the coherence of selected tracer candidates and to enhance the understanding of their emission patterns within the tunnel environment. Additional details are available in the Supporting Information.

Data availability

The datasets associated with the current study are available from the corresponding author (pengjianfei@nankai.edu.cn) on reasonable request.

References

1. Peng, J. et al. Explosive secondary aerosol formation during severe haze in the North China Plain. *Environ. Sci. Technol.* **55**, 2189–2207 (2021).
2. Murphy, B. N. et al. Reactive organic carbon air emissions from mobile sources in the United States. *Atmos. Chem. Phys.* **23**, 13469–13483 (2023).
3. Zhao, B., Wang, S. & Hao, J. Challenges and perspectives of air pollution control in China. *Front. Environ. Sci. Eng.* **18**, 68 (2024).
4. Shrivastava, M. et al. Recent advances in understanding secondary organic aerosol: implications for global climate forcing. *Rev. Geophys.* **55**, 509–559 (2017).
5. Dong, T. et al. Human indoor exposome of chemicals in dust and risk prioritization using EPA's ToxCast database. *Environ. Sci. Technol.* **53**, 7045–7054 (2019).
6. Qiao, L. et al. Screening of ToxCast chemicals responsible for human adverse outcomes with exposure to ambient air. *Environ. Sci. Technol.* **56**, 7288–7297 (2022).
7. Robinson, A. L. et al. Rethinking organic aerosols: semivolatile emissions and photochemical aging. *Science* **315**, 1259–1262 (2007).
8. Zhao, Y. et al. Intermediate-volatility organic compounds: a large source of secondary organic aerosol. *Environ. Sci. Technol.* **48**, 13743–13750 (2014).
9. Zhao, Y. et al. Intermediate volatility organic compound emissions from on-road diesel vehicles: chemical composition, emission factors, and estimated secondary organic aerosol production. *Environ. Sci. Technol.* **49**, 11516–11526 (2015).
10. Chang, X. et al. Full-volatility emission framework corrects missing and underestimated secondary organic aerosol sources. *One Earth* **5**, 403–412 (2022).
11. Donahue, N. M., Kroll, J. H., Pandis, S. N. & Robinson, A. L. A two-dimensional volatility basis set—Part 2: diagnostics of organic-aerosol evolution. *Atmos. Chem. Phys.* **12**, 615–634 (2012).
12. Peng, J. et al. Gasoline aromatics: a critical determinant of urban secondary organic aerosol formation. *Atmos. Chem. Phys.* **17**, 10743–10752 (2017).
13. Presto, A. A. et al. Intermediate-volatility organic compounds: a potential source of ambient oxidized organic aerosol. *Environ. Sci. Technol.* **43**, 4744–4749 (2009).
14. Jathar, S. H. et al. Unspeciated organic emissions from combustion sources and their influence on the secondary organic aerosol budget in the United States. *Proc. Natl. Acad. Sci. USA* **111**, 10473–10478 (2014).
15. Zhao, J. et al. Comprehensive assessment for the Impacts of S/IVOC emissions from mobile sources on SOA formation in China. *Environ. Sci. Technol.* **56**, 16695–16706 (2022).
16. Wu, Y. et al. Source apportionment of VOCs based on photochemical loss in summer at a suburban site in Beijing. *Atmos. Environ.* **293**, 119459 (2023).
17. Tang, R. et al. Measurement report: distinct emissions and volatility distribution of intermediate-volatility organic compounds from on-road Chinese gasoline vehicles: implication of high secondary organic aerosol formation potential. *Atmos. Chem. Phys.* **21**, 2569–2583 (2021).
18. Qi, L. et al. Intermediate-volatility organic compound emissions from nonroad construction machinery under different operation modes. *Environ. Sci. Technol.* **53**, 13832–13840 (2019).
19. Qi, L. et al. Primary organic gas emissions from gasoline vehicles in China: factors, composition and trends. *Environ. Pollut.* **290**, 117984 (2021).
20. Liu, Y. et al. Identification of two main origins of intermediate-volatility organic compound emissions from vehicles in China through two-phase simultaneous characterization. *Environ. Pollut.* **281**, 117020 (2021).
21. Huang, C. et al. Intermediate volatility organic compound emissions from a large cargo vessel operated under real-world conditions. *Environ. Sci. Technol.* **52**, 12934–12942 (2018).
22. Lou, H. et al. Emission of intermediate volatility organic compounds from a ship main engine burning heavy fuel oil. *J. Environ. Sci.* **84**, 197–204 (2019).
23. Shen, X. et al. Real-world emission characteristics of full-volatility organics originating from nonroad agricultural machinery during agricultural activities. *Environ. Sci. Technol.* **57**, 10308–10318 (2023).
24. Zhang, Z. et al. VOC and IVOC emission features and inventory of motorcycles in China. *J. Hazard. Mater.* **469**, 133928 (2024).
25. Alam, M. S., Zeraati-Rezaei, S., Xu, H. & Harrison, R. M. Characterization of gas and particulate phase organic emissions (C9-C37) from a diesel engine and the effect of abatement devices. *Environ. Sci. Technol.* **53**, 11345–11352 (2019).
26. Dallüge, J., Beens, J. & Brinkman, U. A. T. Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. *J. Chromatogr. A* **1000**, 69–108 (2003).
27. An, Z. et al. Frontier review on comprehensive two-dimensional gas chromatography for measuring organic aerosol. *J. Hazard. Mater. Lett.* **2**, 100013 (2021).
28. Song, K. et al. Impact of cooking style and oil on semi-volatile and intermediate volatility organic compound emissions from Chinese domestic cooking. *Atmos. Chem. Phys.* **22**, 9827–9841 (2022).
29. Guo, Z. et al. Higher toxicity of gaseous organics relative to particulate matters emitted from typical cooking processes. *Environ. Sci. Technol.* **57**, 17022–17031 (2023).
30. Huo, Y. et al. Chemical fingerprinting of HULIS in particulate matters emitted from residential coal and biomass combustion. *Environ. Sci. Technol.* **55**, 3593–3603 (2021).
31. Zhang, H. et al. Monoterpenes are the largest source of summertime organic aerosol in the southeastern United States. *Proc. Natl. Acad. Sci. USA* **115**, 2038–2043 (2018).
32. Cai, S. et al. Time-resolved intermediate-volatility and semivolatile organic compound emissions from household coal combustion in Northern China. *Environ. Sci. Technol.* **53**, 9269–9278 (2019).
33. Huang, D. D. et al. Obscured contribution of oxygenated intermediate-volatility organic compounds to secondary organic aerosol formation from gasoline vehicle emissions. *Environ. Sci. Technol.* **58**, 10652–10663 (2024).
34. He, X. et al. Comprehensive chemical characterization of gaseous I/SVOC emissions from heavy-duty diesel vehicles using two-dimensional gas chromatography time-of-flight mass spectrometry. *Environ. Pollut.* **305**, 119284 (2022).
35. An, Z., Ren, H., Xue, M., Guan, X. & Jiang, J. Comprehensive two-dimensional gas chromatography mass spectrometry with a solid-state thermal modulator for in-situ speciated measurement of organic aerosols. *J. Chromatogr. A* **1625**, 461336 (2020).
36. Nault, B. A. et al. Secondary organic aerosols from anthropogenic volatile organic compounds contribute substantially to air pollution mortality. *Atmos. Chem. Phys.* **21**, 11201–11224 (2021).
37. He, X. et al. Comprehensive characterization of particulate intermediate-volatility and semi-volatile organic compounds (I/SVOCs) from heavy-duty diesel vehicles using two-dimensional gas chromatography time-of-flight mass spectrometry. *Atmos. Chem. Phys.* **22**, 13935–13947 (2022).
38. Lu, Q., Zhao, Y. & Robinson, A. L. Comprehensive organic emission profiles for gasoline, diesel, and gas-turbine engines including intermediate and semi-volatile organic compound emissions. *Atmos. Chem. Phys.* **18**, 17637–17654 (2018).
39. Zhao, Y., Lambe, A. T., Saleh, R., Saliba, G. & Robinson, A. L. Secondary organic aerosol production from gasoline vehicle exhaust: effects of engine technology, cold start, and emission certification standard. *Environ. Sci. Technol.* **52**, 1253–1261 (2018).
40. Drozd, G. T. et al. Detailed speciation of intermediate volatility and semivolatile organic compound emissions from gasoline vehicles: effects of cold-starts and implications for secondary organic aerosol formation. *Environ. Sci. Technol.* **53**, 1706–1714 (2019).

41. Zhang, J. et al. Secondary organic aerosol formation potential from vehicular non-tailpipe emissions under real-world driving conditions. *Environ. Sci. Technol.* **58**, 5419–5429 (2024).

42. Tang, J. et al. Intermediate volatile organic compounds emissions from vehicles under real world conditions. *Sci. Total Environ.* **788**, 147795 (2021).

43. Fang, H. et al. Measurement report: emissions of intermediate-volatility organic compounds from vehicles under real-world driving conditions in an urban tunnel. *Atmos. Chem. Phys.* **21**, 10005–10013 (2021).

44. Liu, Y. et al. Source profiles of volatile organic compounds (VOCs) measured in China: part I. *Atmos. Environ.* **42**, 6247–6260 (2008).

45. Sha, Q. E. et al. A newly integrated dataset of volatile organic compounds (VOCs) source profiles and implications for the future development of VOCs profiles in China. *Sci. Total Environ.* **793**, 148348 (2021).

46. Yuan, B., Shao, M., Lu, S. & Wang, B. Source profiles of volatile organic compounds associated with solvent use in Beijing, China. *Atmos. Environ.* **44**, 1919–1926 (2010).

47. Shi, Y., Xi, Z., Simayi, M., Li, J. & Xie, S. Scattered coal is the largest source of ambient volatile organic compounds during the heating season in Beijing. *Atmos. Chem. Phys.* **20**, 9351–9369 (2020).

48. Worton, D. R. et al. Lubricating oil dominates primary organic aerosol emissions from motor vehicles. *Environ. Sci. Technol.* **48**, 3698–3706 (2014).

49. Liang, Z. et al. Comprehensive chemical characterization of lubricating oils used in modern vehicular engines utilizing GC × GC-TOFMS. *Fuel* **220**, 792–799 (2018).

50. Fuyang Zhang, J. P. et al. Real world particle size distribution from vehicle fleet and implication on emission control. *npj Clean Air*. https://www.nature.com/articles/s44407-025-00007-8 (2025).

51. Vörösmarty, M., Hopke, P. K. & Salma, I. Attribution of aerosol particle number size distributions to main sources using an 11-year urban dataset. *Atmos. Chem. Phys.* **24**, 5695–5712 (2024).

52. Du, Z. et al. Comparison of primary aerosol emission and secondary aerosol formation from gasoline direct injection and port fuel injection vehicles. *Atmos. Chem. Phys.* **18**, 9011–9023 (2018).

53. Hatch, L. E. et al. Identification and quantification of gaseous organic compounds emitted from biomass burning using two-dimensional gas chromatography–time-of-flight mass spectrometry. *Atmos. Chem. Phys.* **15**, 1865–1899 (2015).

54. Huo, Y. et al. Addressing unresolved complex mixture of I/SVOCs emitted from incomplete combustion of solid fuels by nontarget analysis. *JGR Atmos*. **126**, e2021JD035835 (2021).

55. Zhang, J. et al. Marked impacts of transient conditions on potential secondary organic aerosol production during rapid oxidation of gasoline exhausts. *npj Clim. Atmos. Sci.* **6**, 59 (2023).

56. Ballabio, D. & Consonni, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal. Methods* **5**, 3790–3798 (2013).

57. Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660.e19 (2019).

58. Gordon, A. D. A review of hierarchical classification. *J. R. Stat. Soc. Ser. A* **150**, 119–137 (1987).

59. Zhang, X. et al. Comprehensive characterization of speciated volatile organic compounds (VOCs), gas-phase and particle-phase intermediate- and semi-volatile volatility organic compounds (I/S-VOCs) from Chinese diesel trucks. *Sci. Total Environ.* **912**, 168950 (2024).

60. Song, K. et al. Non-target scanning of organics from cooking emissions using comprehensive two-dimensional gas chromatography-mass spectrometer (GC × GC-MS). *Appl. Geochem.* **151**, 105601 (2023).

61. Yu, H. & Hou, X. Hierarchical clustering in astronomy. *Astron. Comput.* **41**, 100662 (2022).

62. Song, C., Liu, Y., Sun, L., Zhang, Q. & Mao, H. Emissions of volatile organic compounds (VOCs) from gasoline- and liquified natural gas (LNG)-fueled vehicles in tunnel studies. *Atmos. Environ.* **234**, 117626 (2020).

63. Wu, Y. et al. Evaluating vehicular exhaust and evaporative emissions via VOC measurement in an underground parking garage. *Environ. Pollut.* **333**, 122022 (2023).

64. Zhang, J. et al. Vehicular non-exhaust particulate emissions in Chinese megacities: Source profiles, real-world emission factors, and inventories. *Environ. Pollut.* **266**, 115268 (2020).

65. Ho, C. S. et al. Tunnel measurements reveal significant reduction in traffic-related light-absorbing aerosol emissions in China. *Sci. Total Environ.* **856**, 159212 (2023).

66. Song, K. et al. Addressing new chemicals of emerging concern (CECs) in an indoor office. *Environ. Int* **181**, 108259 (2023).

67. Chen, Y. et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. *Nat. Commun.* **15**, 1657 (2024).

Acknowledgements

Author contributions

Y.W. and J.P. wrote the main manuscript text. X.W., Y.L., and F.Z. performed methodology validation and formal analysis. P.L., P.S., L.W., and T.W. contributed to investigation and conceptualization. B.S. and J.Z. provided resources and supported investigations. H.M. supervised the project and managed funding acquisition. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-025-01078-w.

Correspondence and requests for materials should be addressed to Jianfei Peng or Hongjun Mao.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.