Automated compound speciation, cluster analysis, and quantification of organic vapors and aerosols using comprehensive two-dimensional gas chromatography and mass spectrometry

Xiao He[1], Xuan Zheng[1*], Shuwen Guo[1], Lewei Zeng[1], Ting Chen[1], Bohan Yang[1], Shupei Xiao[1], Qiongqiong Wang[2], Zhiyuan Li[3], Yan You[4], Shaojun Zhang[5,6,7,8], and Ye Wu[5,6,7,8]

[1]College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen 518060, China
[2]Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan 430074, China
[3]School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, China
[4]National Observation and Research Station of Coastal Ecological Environments in Macao, Macao Environmental Research Institute, Macau University of Science and Technology, Macao SAR 999078, China
[5]School of Environment, State Key Joint Laboratory of Environment Simulation and Pollution Control, Tsinghua University, Beijing 100084, China
[6]State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, Beijing 100084, China
[7]Beijing Laboratory of Environmental Frontier Technologies, School of Environment, Tsinghua University, Beijing 100084, China
[8]Laboratory of Transport Pollution Control and Monitoring Technology, Transport Planning and Research Institute, Ministry of Transport, Beijing 100028, China

Correspondence to: Xuan Zheng (x-zheng11@szu.edu.cn)

Abstract: The advancement of analytical techniques, such as comprehensive two-dimensional gas chromatography coupled with mass spectrometry (GC×GC-MS), enables the efficient separation of complex organics. Developing innovative methods for data processing and analysis is crucial to unlock the full potential of GC×GC-MS in understanding intricate chemical mixtures. In this study, we proposed an innovative method for the semi-automated identification and quantification of complex organic mixtures using GC×GC-MS. The method was formulated based on self-constructed mass spectrum patterns and the traversal algorithms and was applied to organic vapor and aerosol samples collected from the tailpipe emissions of heavy-duty diesel vehicles and the ambient atmosphere. Thousands of compounds were filtered, speciated, and clustered into 26 categories, including aliphatic and cyclic hydrocarbons, aromatic hydrocarbons, aliphatic oxygenated species, phenols and alkyl-phenols, and heteroatom containing species. The identified species accounted for over 80% of all the eluted chromatographic peaks at the molecular level. A comprehensive analysis of quantification uncertainty was undertaken. Using representative compounds, quantification uncertainties were found to be less than 37.67%, 22.54%, and 12.74% for alkanes, polycyclic aromatic hydrocarbons (PAHs), and alkyl-substituted benzenes, respectively, across the GC×GC space, excluding the first and the last time intervals. From source apportionment perspective, adamantane was clearly isolated as a potential tracer for heavy-duty diesel vehicles (HDDVs) emission. The systematic distribution of nitrogen-containing compounds in oxidized and reduced valences was discussed and many of them served as critical tracers for secondary nitrate formation processes. The results highlighted the benefits of developing self-constructed models for the enhanced peak identification, automated cluster analysis, robust uncertainty estimation, and source apportionment and achieving the full potential of GC×GC-MS in atmospheric chemistry.

1 Introduction

Improved sampling strategies, coupled with innovative measurement techniques, are imperative to capture the dynamic nature of atmospheric chemistry, particularly in the context of climate change and health risks (Franklin et al. 2023, Franklin et al. 2022, Huo et al. 2021, Phillips et al. 2018). Comprehensive two-dimensional gas chromatography coupled with mass spectrometry (GC×GC-MS) has emerged as a powerful tool for compound detection and identification, benefiting from the combination of two columns with orthogonal selectivity (Alam et al. 2013, Franklin et al. 2022).

Despite its capabilities, GC×GC-MS encounters formidable challenges in data analysis, which can be extremely complicated and demanding. Efforts have been made to handle the deluge of data generated by GC×GC-MS. Traditionally, mass spectra were deconvoluted and compared to spectra from the National Institute of Standards and Technology (NIST) library for peak identification with pre-defined criteria (Guo et al. 2016, Piotrowski et al. 2018). Retention indices (RI) were further introduced to distinguish homologous compounds with resembling mass spectra. A pioneering and instructive work for searching criteria to classify GC×GC peaks was published in 2003 (Welthagen, Schnelle-Kreis and Zimmermann 2003). Welthagen (2003) incorporated the mass fragmentation patterns to classify compounds in atmospheric aerosol samples. Compounds belonging to the same chemical group related to one another in the GC×GC space and distributed in a structured pattern. They successfully identified seven groups of compounds, including alkanes, alkenes and cycloalkanes, alkyl substituted benzenes, alkyl substituted polar benzenes, hydrated naphthalenes and alkenyl benzenes, alkylated naphthalenes, and alkane acids, occupying more than 60% of the total peak area. This work set a good example of how user-defined rules could facilitate the identification of specific compound groups.

Recent advances in chemometric tools for GC×GC-MS analysis involving machine learning and deep learning renovate multi-dimensional chromatography fields (Stefanuto, Smolinska and Focant 2021). Bendik (2021) developed a programming suite for high-confidence and fast compound identification using GC×GC coupled with time-of-flight mass spectrometry (TOF-MS) (Bendik et al. 2021). He (2022) extracted featured mass spectrometric information of the intermediate-volatility and semi-volatile organic compounds (I/SVOCs) by integrating algorithmic approaches into GC×GC-MS data (He et al. 2022a, He et al. 2022b). A novel pixel-based multiway principal component analysis method was employed in Song (2023) to identify key tracers during incense burning (Song et al. 2023). Nevertheless,

interpreting GC×GC-MS data requires advanced computational tools and expertise, and the investigation of unknown compounds remains challenging due to the inadequate validation procedures, overreliance on manual data processing, limited access to computational resources, and the insufficient expertise in handling complex chromatographic data effectively.

Bridging this gap requires further development of sophisticated algorithms and analytical approaches to unlock the full potential of GC×GC. This study proposes a bottom-up method for cluster analysis and quantification of organic vapors and aerosols within complex atmospheric mixtures. The scripts were initiated with the recognition of the common mass spectral features of specific species and were tailored to a wide range of compound clusters. The scripts were then trained, iterated, and optimized using real sample data until robust outputs were achieved. The new strategy reduces the ambiguity often associated with identifying compounds in complex mixtures.

The proliferation of heavy-duty diesel vehicles (HDDVs) has raised significant concerns due to their increasing role in freight transport and in various industrial operations (Yan et al. 2022, Cheng et al. 2022). Despite their low retention rate, HDDVs release substantial amounts of particulate matter, nitrogen oxides, ammonia, and carbon monoxide into the atmosphere compared with other vehicle types (Wang et al. 2023, Silva et al. 2023, Chang et al. 2022, Stanimirova et al. 2023, Hamilton and Harley 2021, 2021, Kruve et al. 2014). To address this, gas and aerosol samples were collected from representative HDDV tailpipes and the ambient environment, then analyzed using GC×GC-MS. The proposed bottom-up method was employed for a comprehensive analysis of the complex organic mixtures, resulting in the identification of 26 compound categories, including hydrocarbons in multiple forms, oxygenated components, and species containing heteroatoms. Over 80% of all the chromatographic peaks were identified and assigned to a compound cluster using the proposed method, leaving a minor portion of organic matrix unresolved. Different compound clusters occupied separate positions in the GC×GC space, and distinctive distribution patterns within diverse samples and their contribution fractions were revealed. Quantification uncertainties were addressed thoroughly and the significant potential deviation when using n-alkanes as semi-quantification surrogates was highlighted. Overall, integrating automated algorithms with GC×GC data analysis holds significant implications for advancing our understanding of atmospheric chemistry, improving secondary organic aerosol (SOA) estimation, and guiding the formulation of environmental policies.

2 Materials and methods

2.1 Sample collection, treatment, and instrumental analysis

For the collection of HDDVs tailpipe emissions, chassis dynamometer experiments were conducted at the China Automotive Technology & Research Center (CATARC) in Guangzhou, China. Exhaust emissions from HDDVs were diluted in a constant volume sampler (CVS, CVS-ONE-MV-HE, Horiba), following the China heavy-duty commercial vehicle test cycle for tractor trailers (CHTC-TT) driving cycles. Two HDDVs equipped with the selective catalytic reduction (SCR) system were recruited. The two HDDVs met the China IV national emission standard and were manufactured in 2021. More information is summarized in Table S1. The average temperature in the sampling train was precisely controlled at 47 °C, while airflow, relative humidity, and pressure were monitored simultaneously. The speed trace and characteristics of CHTC-TT are shown in Figure S1.

Gaseous exhausts were collected using two adsorbent thermal desorption (TD) tubes in series (Tenax TA, C1-AXXX-5003, Markes International) after passing through a Teflon filter. Particulate exhausts were deposited on a 47 mm quartz filter (Grade QM-A, Whatman). Ambient $PM_{2.5}$ filter samples were collected on the rooftop of a 5-story building on the campus of Shenzhen University (22.60°N, 114.00°E) during November 2023 in western Shenzhen, approximately 25 m above the ground. The sampling site was surrounded by campus, residential areas, greenbelts, and a golf park, as shown in Figure S2. Previous studies demonstrated that the $PM_{2.5}$ concentration in this aera represented the average pollution scheme in Shenzhen (Huang et al. 2018, Yu et al. 2020). The sampling strategy followed a regular schedule of one 24-h sample every day using a high-volume sampler (Th-1000c II, Wuhan Tianhong Environmental Protection Industry Co., Ltd). In total, 55 TA tube samples (including 11 field blank samples), 20 HDDV aerosol samples (including 3 field blank samples), and 6 ambient aerosol samples (including one blank sample) were collected. The list of ambient samples and the relevant PM concentrations are listed in Table S2. The sorbent tubes were well sealed and stored dry at room temperature, and quartz filters were frozen at −18 °C before analysis. All sampling materials were pre-baked thoroughly to remove potential carbonaceous contamination.

TD samples were injected with 2 μL of deuterated internal standard (IS) mixing solution through a mild N_2 blow (CSLR, Markes International). The list of deuterated IS is shown in Table S3. A precise portion of 1 cm^2 (1 cm × 1 cm) filter sample was isolated and cut into strips. They were spiked with 2 μL of IS

mixing solution and inserted into a passivated quartz tube. All TD samples and quartz tubes were loaded onto a thermal desorption autosampler (ULTRA-xr, Markes International), thermally desorbed (UNITY-xr, Markes International), and subjected to GC×GC separation (Agilent 8890, Agilent Technologies; Solid State Modulator1810, *J&X* Technologies) and mass spectrometry detection (Agilent 5977B, Agilent Technologies).

The thermal desorption system heated the TD tubes to 320 °C (quartz tubes to 330 °C) for 20 min, while the trap remained at 20 °C. Following tube desorption, the trap temperature was raised to 330 °C (340 °C for quartz tubes) for 5 min at the maximum heating rate, and the vaporized analytes were purged into the 1st GC column with a desorb split flow of 6 mL/min. Separation of the analytes was carried out using a DB-5ms capillary column (30 m × 0.25 mm × 0.25 μm, Agilent Technologies) as the primary column and a DB-17ms capillary column (1.2 m × 0.18 mm × 0.18 μm, Agilent Technologies) as the secondary column. The modulation column consisted of a VF-1ms capillary column (0.7 m × 0.25 mm × 0.10 μm, Agilent Technologies) connecting to the 1st column and an Ultimate Plus deactivated fused silica tubing (0.6 m × 0.25 mm, Agilent Technologies) connecting to the 2nd column.

Initially, the GC oven was set at 50 °C for 3 min, followed by a gradual increase at a rate of 5 °C/min until it reached 310 °C, where it was maintained for an additional 5 min. The entry and exit hot zones were set +10 °C higher than the GC oven temperature, while the trap zone was maintained at -50 °C. The modulation cycle had a period of 4 s. Carrier gas flow was set at 1.2 mL/min. The MS had an integer resolution and was conducted in electron impact positive (EI+) mode (70 eV). It was operated over a range of 20–350 amu, and the temperature of the transfer line, ion source, and MS quadrupole was 300 °C, 250 °C, and 170 °C, respectively.

2.2 Data collection, alignment, and parsing

GC×GC-MS data acquisition was performed using Enhanced MassHunter (version 10.0, Agilent Technologies) and SSCenter (version 2.4.0.0, *J&X* Technologies). All data utilized to develop and test the scripts were processed by Canvas Browser (version 2.5, *J&X* Technologies), which included baseline correction, mass spectra deconvolution, and peak smoothing. Baseline correction and peak smoothing enhanced the signal-to-noise ratio (S/N) and improved overall data quality.

Chromatographic peaks were filtered using the following criteria: baseline noise = 150, S/N $>$ 50. For each individual sample, after isolating all compounds of interest, a peak table was exported with 1st

retention time (RT) and 2nd RT, peak area, peak height, peak width, and deconvoluted mass spectra, arranged in 1st RT sequential order. These quantitative variables were further processed for targeted and non-targeted "omics"-oriented analysis.

As expected, the chromatographic variables experienced RT shifts due to column degradation, routine maintenance (e.g., cutting column), and system fluctuations (e.g., variations in carrier gas pressure). The initial tolerance for RT shifts in adaptive cluster matching was set at 1 period of modulation in the 1st dimension and 0.1 s in the 2nd dimension. Additionally, a 2D shift cluster consisting of $C_{16}D_{34}$, $C_{24}D_{50}$, and $C_{32}D_{66}$, was configured, with the merit of correcting 2nd RT shift. Data correction or data alignment is critical for accurate and consistent peak integration.

2.3 Algorithmic development

EI spectra are typically characterized by a molecular ion (M^+) peak plus a collection of fragment ion peaks. The M^+ may dominate the mass spectrum in some cases (e.g., unsubstituted polycyclic aromatic hydrocarbons (PAHs)), but more frequently presents at a relatively low intensity. The EI spectra are highly comparable among different instrument systems and experimental conditions, making them an excellent measure for identifying compounds. The characteristic ions and their relative intensities depend on the intrinsic nature of the targeted compounds, necessitating knowledge of basic rules and common fragmentation routes to interpret EI mass spectra. Figure 1 illustrates the workflow for establishing computational strategies for robust and reproducible GC×GC-MS data processing.

Functional groups significantly affect the fragmentation patterns observed in mass spectrometry, and some ions are typical of given structures. Isotopic peaks (e.g., hydrogen and chlorine) provide additional information about the molecules (Du and Angeletti 2006, Fernandez-de-Cossio et al. 2004). These pieces of information form the foundation for building up the model for cluster analysis, which is addressed in greater detail in the supporting information (S1). These indicative reaction schemes have been incorporated into the model development. Each critical step of model construction and validation is described thoroughly. The quantitative variables in the data alignment table, combining the chromatographic and MS information, are properly exploited and determine the overall speciation capacities.

Traditionally, compound identification relies on the electron ionization-based fragmentogram and the deconvoluted mass spectra. Empirically, one-by-one compound identification can be greatly intervened

by neighbouring peaks, especially those with similar structures, and introduce considerable uncertainties. A good example is the assignment of homologous *n*-alkanes, of which the fragmentograms bear a close resemblance (Figure S6). In such cases, the similarity score (the measure of similarity between the observed mass spectrum and the NIST library hit) could be erroneously inflated to 850 (out of 999) or higher. In contrast, cluster analysis involves the comprehensive analysis of a specific type of compounds on a large scale, aiming to provide a holistic understanding of the distribution and transformation of the specific compound cluster being investigated.

Due to the complexity and remarkable peak capacities, sophisticated and detailed scripts for cluster identification were constructed. Heteroatom-containing species, e.g., amides and amines, were carefully examined. The scripts began by recognition of the common mass spectra features of compound cluster of interest and are addressed in more details in the following descriptive framework:

1. The Boolean value of characteristic ions.
2. The intensity sequence of abundant ions in the whole spectra.
3. The retention time window restriction for certain compound groups.
4. The pattern of mass spectrometry variation with the increased number of substituents or the extension of the carbon chain.
5. An iteration framework that involved repetitive cycles among all the tested samples.

The scripts were then trained, iterated, and optimized incorporating real sample data, and the parameters were adjusted accordingly until a robust output was achieved. The extractor function built into the Canvas software was activated, and all the scripts were imported to facilitate automated cluster analysis. The scripts parsed all the files in the given directory into the required structure and generated three reports in the form of .pdf, .csv, and .bmp. The .csv file contained key information including the compound name, compound cluster, 1st and 2nd RTs, and peak area (based on total ion current (TIC)).

```
For (i = 1 to m) # m equals the number of all tested samples.
   Load the sample
   Peak identification
   Baseline correction
   Mass spectra deconvolution
   Peak smoothing
      For (j = 1 to 26) # In total, 26 compound clusters were isolated with high accuracy and repeatability.
         Execute the extraction rule of cluster (j)
```

```
        Export peak number, 1st RT and 2nd RT, peak area, peak height, peak width, and deconvoluted mass spectra
    Next j
Next i
```

Once exported, the peaks were further processed for quantification/semi-quantification following the steps below. First, calibration curves were prepared by spiking different volumes of the standard solution mixture onto the blank TD tubes and blank filters, respectively. Peak area ratios, i.e., peak area of authentic standards over that of the internal standards, were used to build the linear relationship, with the merit of correcting system fluctuations. The selection of authentic standards prioritized their wide distribution across the entire chromatogram space, ranging from high to low volatility and weak to strong polarity, and meanwhile encompassing a broader range of functional groups and heteroatoms. The distribution and performance of all authentic standards are summarized in Table S4 and Figure S7. Second, for the un-quantified peaks, their complied information (X, Y, Z) corresponding to (1st RT, 2nd RT, compound cluster) is looped through the list of all authentic standards in the following descriptive pseudo-codes until the optimal authentic standard to semi-quantify the target peak is exported. It should be emphasized that the un-quantified peak and the corresponding authentic standard to semi-quantify it must belong to the same group due to their physicochemical similarities.

```
For (i = 1 to n) # n equals the number of authentic standards and is a known variable.
  If (ZM = Zi) # M is the un-quantified peak and i refers to the authentic standard that is selected in a certain loop.
    Ai = Min (an array of ((XM – Xi)^2 + (YM – Yi)^2)) # This sentence dose not conform to the grammar rule of Visual Basic for Applications in Excel, and it is for illustrative purposes only.
    Export Zi, (Xi, Yi, Zi), its peak area, and its linear calibration relationship.
  End if
Next i
```

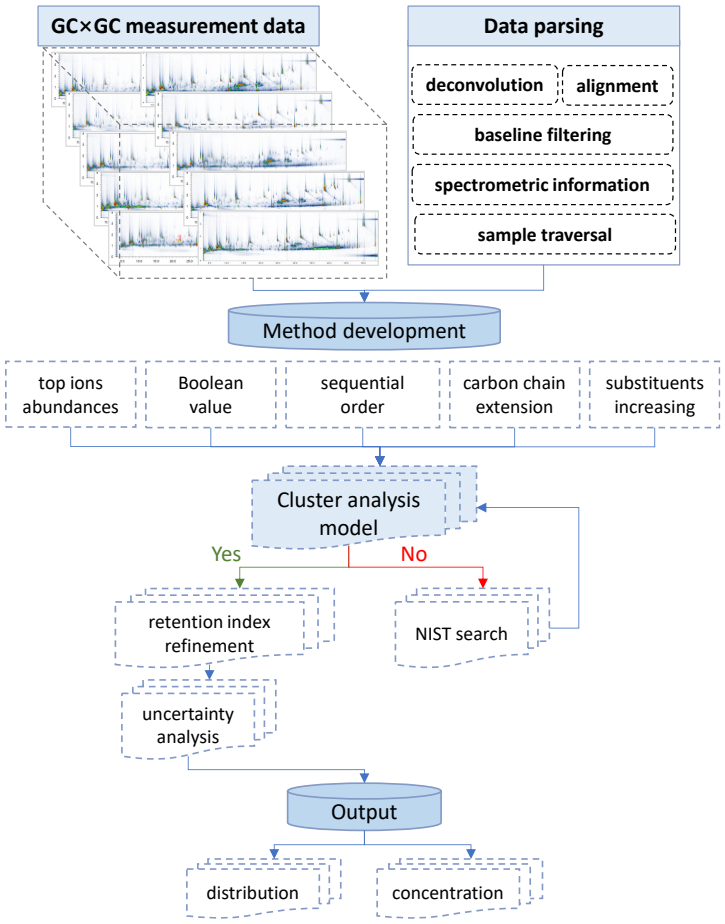

Figure 1. Flow diagram illustrating the multistep data processing for establishing computational strategies for cluster analysis and quantification of organic vapors and aerosols using GC×GC-MS data.

2.4 Quality assurance/control and uncertainty evaluation

It is common for thermal decomposition to occur in analytical methods involving heating processes, potentially leading to the erroneous detection of compounds that are either not present in real samples or present in low concentrations. Such artifacts need careful scrutiny, and the availability of authentic standards covering the GC×GC space range is essential for validation. Nevertheless, the possibility that some observed analytes are decomposition products cannot be entirely ruled out. Peaks of ISs were traced across all samples to monitor the variations across several modules, and the results are presented in Figure S8. Excellent stability was clearly observed, demonstrating the robustness of the testing system. Strong linear correlations were achieved for this set of authentic standards between the peak area ratio and the spiked mass, with Pearson's R ranging from 0.97 to 0.99.

3 Results and discussion

3.1 Overall performance of the algorithm and compound identification

The optimization of component identification remains challenging, and this work involves converting known chemical compounds into molecular descriptors and utilizing cluster analysis to predict the relationship between these descriptors and structural information. After continuous trials to improve reliability and data processing speed, a final solution of 26 compound clusters stands out with high accuracy and repeatability:

- Aliphatic hydrocarbons, including *n*-/*i*-alkanes and alkenes
- Cycloalkanes
- Alkyl-substituted benzenes, including C_1 – C_6 alkyl-substituted benzenes
- Adamantanes
- Hopanes
- 2 – 5 ring PAHs
- Acids
- Aliphatic alcohols
- Aliphatic aldehydes and ketones
- Oxy-PAHs
- Phthalates
- Phenols and alkyl-substituted phenols
- Phenol ethers
- Amides
- Amines
- Pyridines
- Nitro compounds, including organic nitrates and organic nitrites

Validation of the model output using field diesel samples has been conducted and has shown high estimation accuracy and integrity. Generally, over 82% of the peaks have been successfully classified and assigned to the corresponding compound groups, and their distribution in an example GC×GC plot is shown in Figure 2. To confirm the tentatively identified heteroatom groups, their raw chromatogram, mass spectra, and chemical structures of representative species are displayed in Figures S10-S16. Less

than 18% of the chromatographic peaks were identified as unresolved components. Aliphatic hydrocarbons were generally located in the lowest positions in the GC×GC space, except for column bleedings (Figure 2a-c and Figure S9), and their 2^{nd} RT drifted less than 1s from the far-left to the far-right side. Nitrogen-containing compounds in oxidized and reduced valences, including amides, amines, pyridines, and nitro compounds, were resolved simultaneously under respective filtering rules and occupied slightly higher positions in the GC×GC space (Figure 2f). Amines and pyridines, being more volatile, eluted at early stages, whereas nitro compounds and amides eluted at middle and late stages sequentially. Due to their high volatility, C_2-C_6 alkyl-substituted benzenes also appeared at the beginning of the GC×GC space and predominantly partitioned into the gas phase. Their 2^{nd} RTs were comparable to those of pyridines and amides, with negligible drift in 2^{nd} RT. Aliphatic oxygen-containing compounds, including acids, alcohols, and ketones, were found to be in the middle region and covered a wide volatility range. There aliphatic oxygen-containing compounds affect the acidity of the atmosphere, participate in aqueous phase reactions, and contribute significantly to the formation of SOA (Cope et al. 2021, Xu et al. 2022). Phenols with one or more hydroxyl groups attached to an aromatic benzene ring were observed in the middle of the GC×GC space. Oxy-PAHs and PAHs were present in the upper-middle of the GC×GC space, with their volatility range extending towards the low volatility end. A clear trend tilting towards the upper right corner was observed, suggesting that aromaticity plays a significant role in the retention in the secondary dimension.

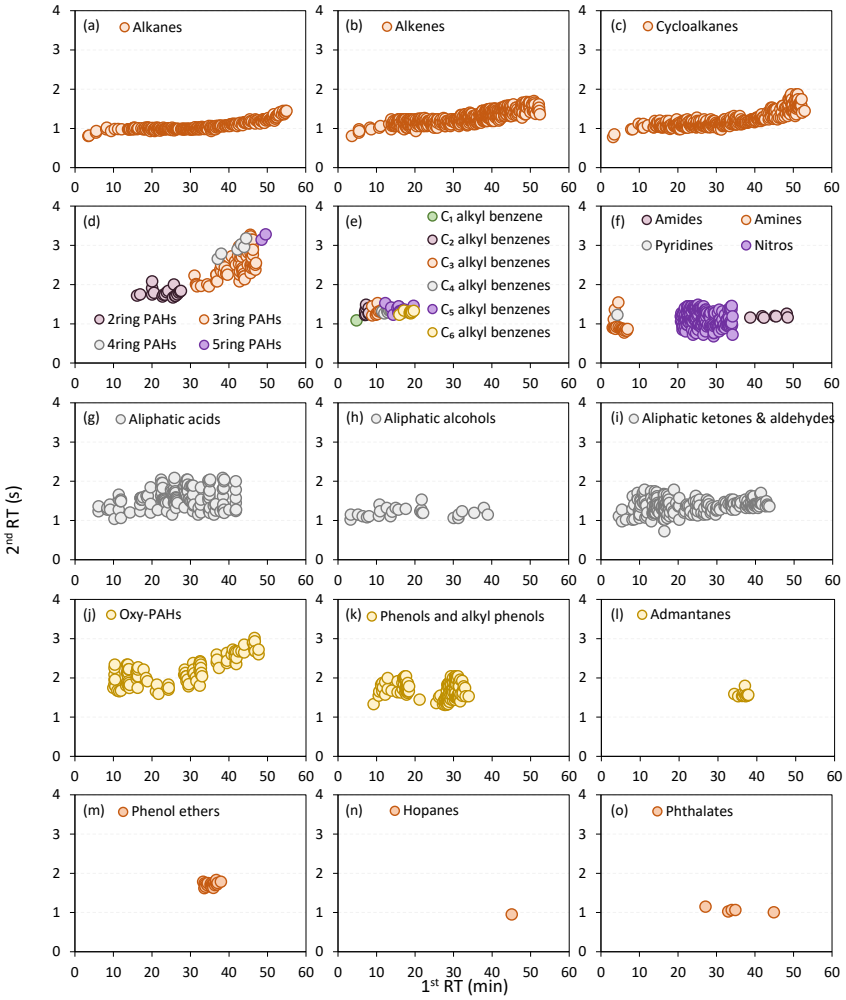

Figure 2. The distribution of the 26 compound groups in an example GC×GC plot. For clear visualization, different compound groups are displayed separately, except for 2–5 ring PAHs, C_2–C_6 alkyl-substituted benzenes, and nitrogen-containing species. Nitro compounds include organic nitrates and organic nitrites, due to the co-existence of the characteristic ions at m/z 30 (NO^+) and m/z 46 (NO_2^+).

3.2 Estimation of the uncertainty associated with the (semi-) quantification

We conducted a systematic evaluation of the model output, and the results are shown in Figure 3 and Figure 4. To address this issue comprehensively and accurately, we selected three types of standards including C_7–C_{37} *n*-alkanes, C_2–C_6 alkyl-substituted benzenes, and 2–4 ring PAHs, representing a full range of polarities and functionalities. The quantification deviation was computed according to the principles of the model. Chromatographic peaks were quantified either by their authentic standards or the surrogates within the same compound category after being classified into one of the 26 compound classes. For example, if the mass spectrum of a chromatographic peak resembled the pattern of the compound class of alkanes, it would be assigned to the alkane group and quantified by its authentic standard if available, or by the *n*-alkane (*n*-alkane serving as the semi-quantification surrogate in this

303 case) that was closest to it spatially. Similarly, if the mass spectrum of a chromatographic peak followed
304 the pattern of C_x alkyl-substituted benzenes, it would be assigned to the C_x alkyl-substituted benzene
305 group and quantified by its authentic standard if available, or by the alkyl-substituted benzene (with
306 alkyl-substituted benzenes serving as the semi-quantification surrogate) that was closest to it spatially.
307 In light of this explanation, the deviation of the slopes of the calibration curves of any pair of adjacent
308 authentic standards within the same compound category was computed to represent the ceiling of the
309 semi-quantification uncertainty. Uncertainties are calculated using the following Eq. (1):

310 $$Uncertainty\ (\%) = \frac{Abs(S_p - S_s)}{Smaller\ (S_p, S_s)} * 100 \quad (1)$$

311 where S_p and S_s are the slopes of the previous and subsequent compounds, respectively.
312 The slopes increased rapidly from 3.13 (C_7 *n*-alkane) to 8.21 (C_9 *n*-alkane), fluctuated slightly from 8.85
313 to 11.8 in the range of C_9 to C_{27} *n*-alkanes, and decreased gradually after C_{28} *n*-alkane to the end of C_{37}
314 *n*-alkane. Throughout the volatility range of C_9–C_{37} *n*-alkanes, uncertainties were less than 37.67%,
315 except for one interval between C_8 and C_9 *n*-alkanes, where the quantification deviation reached 142%.
316 A similar trend was observed for PAHs, with uncertainties less than 22.54%, except for the first and last
317 intervals, where the quantification deviations were 55.44% and 81.59%, respectively, as shown in Figure
318 3. Stable responses of C_2–C_6 alkyl-substituted benzenes were monitored, and the uncertainties were less
319 than 12.74%. In other words, for any given peak, it would be quantified or semi-quantified by one
320 authentic standard, and the upper limit of quantification uncertainty, originating from any pair of adjacent
321 authentic standards, was as discussed earlier.
322 It is reasonable that the uncertainty ranges of alkyl-substituted benzenes were less than those of *n*-alkanes
323 and PAHs, given that alkyl-substituted benzenes eluted early in the front half, whereas alkanes and PAHs
324 covered the entire volatility range. These trends illustrated that the responses of GC×GC to the analysts
325 were sensitive to the volatility distribution, with accurate quantification being more reliable in the middle
326 region. This also highlighted the utility of introducing more authentic standards and the benefits of
327 enriching compound categories. It can be speculated that the quantification uncertainty would be further
328 reduced with the addition of more standard compounds.
329 Furthermore, we explored the uncertainty estimation of dividing the whole chromatogram into bins based
330 on retention time, and all the species in the same bin were quantified, referring to the mass-to-signal
331 responses of the C_n *n*-alkanes (Zhao et al. 2015, Zhao et al. 2014). This approach corrected the signal

variation of hydrocarbons in the GC-MS and was widely adopted for quantifying unresolved complex mixtures (UCMs) (Shen et al. 2023, Zhao et al. 2022). We chose four types of standards belonging to different compound categories with similar 1st RTs and different 2nd RTs, including $C_{19}H_{40}$ (1st RT = 34.6 min, 2nd RT = 1.03 s), 9,10-anthracenedione (1st RT = 36.07 min, 2nd RT = 3.85 s), $C_{19}H_{40}$ (1st RT = 36.54 min, 2nd RT = 1.07 s), and fluoranthene (1st RT = 37.00 min, 2nd RT = 3.04 s), and assessed the deviation of slopes between each pair of the standards. Results shown in Figure 4 indicate that the deviation between the three pairs of standards was 1809% (C_{19} *n*-alkane vs. 9,10-anthracenedione), 1903% (9,10-anthracenedione vs. C_{20} *n*-alkane), and 105% (C_{20} *n*-alkane vs. fluoranthene), respectively. Quantitative errors in measuring unidentified chromatographic peaks using *n*-alkanes responses could reach three orders of magnitude, especially for oxygen-containing species. Errors in quantifying aromatic components, e.g., PAHs, also exceeded 100% in some cases.

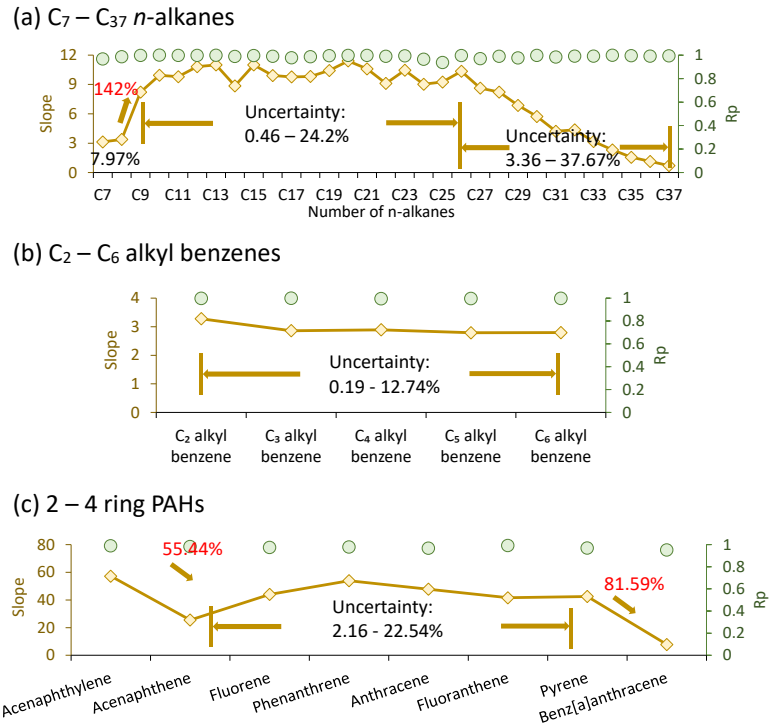

Figure 3. Slope and Pearson correlation variation of (a) C_7–C_{37} *n*-alkanes, (b) C_2–C_6 alkyl-substituted benzenes, and (c) 2–4 ring PAHs. Brown diamond dots represent slopes of different species and are referenced to the left axis. Green circles denote the Pearson correlation of individual species and are referenced to the right axis. Pearson correlation values for *n*-alkanes, C_2–C_6 alkyl-substituted benzenes, and 2–4 ring PAHs range from 0.936 to 0.999, 0.994 to 0.998, and 0.952 to 0.992, respectively. Uncertainties are computed using the equation provided in the main text.

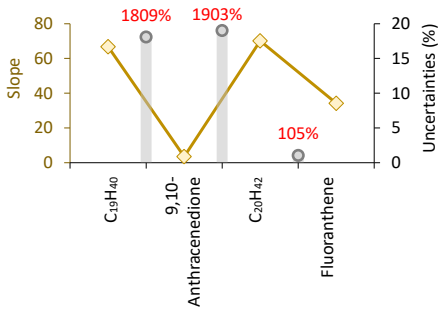

Figure 4. Slopes and uncertainty estimation for example compounds with close 1st RTs and different 2nd RTs: $C_{19}H_{40}$ (1st RT = 34.6 min, 2nd RT = 1.03 s), 9,10-anthracenedione (1st RT = 36.07 min, 2nd RT = 3.85 s), $C_{20}H_{42}$ (1st RT = 36.54 min, 2nd RT = 1.07 s), and fluoranthene (1st RT = 37.00 min, 2nd RT = 3.04 s). Brown diamond dots represent the slopes of different species and are referenced to the left axis. Gray bars denote the uncertainty estimation for these compounds and are referenced to the right axis.

3.3 Cluster analysis in organic vapor and aerosol samples

The model was applied to organic vapor samples from HDDV tailpipe emissions (referred to as HDDV vapors), aerosol samples from HDDV tailpipe emissions (referred to as HDDV aerosols), and atmospheric aerosol samples (referred to as ambient aerosols) for cluster analysis. The results are shown in Figure S17, which displays the distribution of the top few species with a contribution fraction exceeding 5%, and in Figure 5, which shows the mass stacking. Overall, the identified chromatographic peaks accounted for 85%, 82%, and 99% for HDDV vapors, HDDV aerosols, and ambient aerosol samples, respectively. The unidentified peaks were less than 20% and are addressed in greater detail in the supporting information (S2).

Distinct cluster distribution features can be extracted. For ambient aerosol samples, aliphatic ketones were the most abundant cluster, contributing to 27% of all the peak signals, followed by alkanes and alkenes. A notable fraction of 15.2% of nitro compounds was observed exclusively in ambient samples, indicating significant secondary nitrate formation processes under atmospheric conditions. Aliphatic acids and oxy-PAHs were also detected at high levels, with the top six groups accounting for over 95% of the total classified peak signals. Minor but non-negligible fractions included cycloalkane, aliphatic alcohols, and phenols and alkyl-substituted phenols.

Similarly, aliphatic ketones ranked first for HDDV aerosol samples, with mass intensity reaching 46% of the total signals, followed by alkanes. Aliphatic alcohols and oxy-PAHs were also detected at high levels, and the top four groups accounted for over 88% of the total classified peak signals. Cycloalkanes,

amides, phenols and alkyl-substituted phenols, and alkenes were compound clusters with lower abundance, ranging from 1% to 4%.

For HDDV vapors, the most abundant group was phenols and alkyl-substituted phenols, constituting 34% of the total peak signals. Compared with previous results, where the most abundant group was reported to be alkanes (Wang et al. 2022, Alam et al. 2019), the adoption of the innovative model contributed to resolving the oxygenated fractions and reduced inaccuracies in SOA simulation due to the lack of species information. The compound cluster is confirmed by 1) the retention time window, including 1^{st} RT and 2^{nd} RT, and 2) the mass spectra. Detailed information is displayed in Figure S16. The 2^{nd} RTs of the identified phenols and alkyl-substituted phenols range from 1.45 to 1.78 s, well above the hydrocarbon regions, where the 2^{nd} RTs fall within the range of approximately 1.0 to 1.15 s. Their mass spectra also feature the typical phenol ions at m/z = 94, 107, 121, 135, 149, and 191. Alkanes ranked as the second top species, followed by C_1 alkyl-substituted benzene. C_1–C_6 alkyl-substituted benzenes were negligible in both ambient and HDDV aerosol samples but were present in notable abundance in HDDV vapor samples. This distribution aligned with their placement in the GC×GC plot, indicating they were relatively volatile species and partitioned predominantly into the gas phase. Oxy-PAHs and aliphatic ketones contributed 6% of the total identified peak intensities, followed by minor fractions including C_2 alkyl-substituted benzene, cycloalkanes, and alkenes.

The model output illustrates the overall distribution of compound clusters in various gas and aerosol samples, providing comparative insights. Carboxylic acids indicated a higher oxidation state than other compound clusters and were exclusively observed at a notable level in ambient samples compared with "freshly emitted" source samples. The oxidation state of dominant compounds in HDDV samples was comparatively low. For example, a significant ketone fraction was observed in HDDV samples, with the majority partitioning into the aerosol phase due to the long carbon chain skeleton and thus low volatility. Phenols and alkyl-substituted phenols were the leading species in HDDV gas samples. He (2022) reported that the oxygenated I/SVOCs accounted for over 20% of the total I/SVOCs mass in HDDV tailpipe emissions (He et al. 2022a). With the refinement and improvement of model performance, e.g., further splitting mixed mixtures, the oxygenated fraction was elevated to over 50%.

This study highlighted the systematic presence and distribution of nitrogen-containing compounds in both oxidized valences (including nitro compounds) and reduced valences (including amides, amines, and pyridines). Among them, amines and amides were key precursors for new particle formation

processes in a polluted atmosphere (Saeki et al. 2022, Cai et al. 2021), and pyridines, with the nitrogen atom in the aromatic ring, were readily dissolved in water, participating in the global nitrogen cycle in ecosystems (Kosyakov et al. 2020). Nitro compounds, which include a wide range of organic compounds with NO or NO_2 substituents, served as critical tracers for secondary nitrate formation processes. Amines and pyridines were volatile species occupying the early section of the GC×GC space, while nitro compounds and amides were distributed in the middle and rear space. Individual nitrogen-containing species were present at trace levels under atmospheric conditions and were difficult to detect. Moreover, authentic standards or high-resolution mass spectrometry were required to identify and quantify each compound (Zhang et al. 2018). With the establishment of an algorithmic solution, we were able to conduct a comprehensive scan of nitrogen-containing compound clusters.

In addition to common features, specific compounds were identified in separate samples and could potentially serve as markers or tracers for primary emissions. Adamantane and its derivatives, with the fusion of three cyclohexane rings (chemical structure and mass spectrum shown in Figure S18a), are natural products in petroleum (Stout and Douglas 2004). They were volatile and had previously been isolated using GC×GC-ToF-MS in crude oil (Wang et al. 2013). Adamantanes were observed in HDDV vapor samples, contributing 1.4% to the identified peaks. Hopane (chemical structure and mass spectrum shown in Figure S18b) is also a natural product in petroleum and bitumen and serves as an important marker for vehicle emissions due to its persistency and stability (He et al. 2022b, Wong et al. 2021). Hopane was reported to survive heat treatment up to 460 °C and was exclusively detected in HDDV aerosol samples, with an intensity fraction of 0.3% (Wu and Geng 2016).

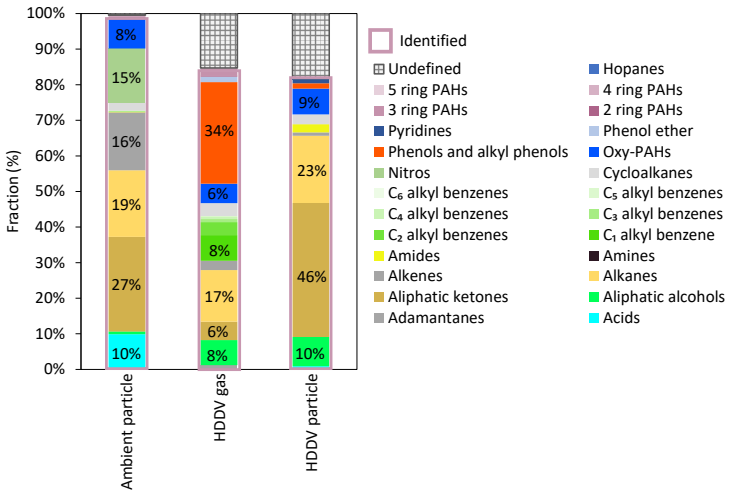

Figure 5. Fractional distributions of different compound clusters in ambient aerosol samples, HDDV tailpipe vapors, and HDDV tailpipe aerosols. Numbers labelled on each column represent the fractions of the top few groups in different samples. Identified clusters are outlined in light purple.

4 Conclusions and outlook

We presented an innovative method for optimizing the separation and identification of organic vapors and aerosols, focusing on establishing molecular descriptors and cluster analysis algorithms. The model outputs were validated using field samples with high accuracy and integrity. Less than 20% of the peaks were unresolved components. The retention patterns of various compound groups and their distribution in the GC×GC plot were resolved, and the influence of functional groups on fragmentation was thoroughly addressed. We also provided a comprehensive analysis of the quantification uncertainties of this new approach and highlighted the significant quantitative errors when using *n*-alkanes as semi-quantification surrogates. This model was applied to various types of field samples, and the results revealed distinctive distribution patterns of compound clusters and contribution fractions, providing valuable insights into the compositions of organic vapors and aerosols, and offering potential markers for specific emission sources.

Compound speciation in atmospheric chemistry continues to be a dynamic and challenging field. Speciated compounds enable models to consider the diversity of organic species and dynamic chemical transformations in the atmosphere, contributing to more accurate SOA simulation results. This approach also allows for a more refined description of the dispersion of pollutants, thereby assisting in the development of localized air quality management strategies as we strive for a more accurate and comprehensive understanding of atmospheric chemistry.

Supplement link:

Data availability

The measurement data and codes used in this study are available on request.

Author contribution:

X.H.: Conceptualization, formal analysis, model development, data validation, writing−original draft, funding acquisition; X.Z.: Writing−reviewing and editing, project administration, supervision, funding acquisition; S.G: Experiment; L.Z., T.C., B.Y., and S.X.: Experiment; Q.W., Z.L., Y.Y., S.Z., and Y.W.: Data validation, writing−reviewing and editing.

Competing interests:

The authors declare that they have no conflict of interest.

Acknowledgements

The authors acknowledge the financial support of the National Natural Science Foundation of China (Grant No. 42105100 and 42261160645), Scientific Research Fund at Shenzhen University (868-000001032089 and 827-000907), Shenzhen Outstanding Science and Technology Innovation Program (RCBS20231211090534051), and Macao Science and Technology Development Fund (0023/2022/AFJ).

References

2021. Ministry of Ecology and Environment of the People's Republic of China. China Mobile Source Environmental Management Annual Report 2021.

Alam, M. S., C. E. West, A. G. Scarlett, S. J. Rowland & R. M. Harrison Application of 2D-GCMS reveals many industrial chemicals in airborne particulate matter. *Atmospheric Environment*, 65**,** 101-111, 2013.

Alam, M. S., S. Zeraati-Rezaei, H. M. Xu & R. M. Harrison Characterization of Gas and Particulate Phase Organic Emissions (C_9-C_{37}) from a Diesel Engine and the Effect of Abatement Devices. *Environmental Science & Technology*, 53**,** 11345-11352, 2019.

Bendik, J., R. Kalia, J. Sukumaran, W. H. Richardot, E. Hoh & S. T. Kelley Automated high confidence compound identification of electron ionization mass spectra for nontargeted analysis. *J Chromatogr A*, 1660**,** 462656, https://www.ncbi.nlm.nih.gov/pubmed/34798444, 2021.

Cai, R. L., C. Yan, D. R. Worsnop, F. Bianchi, V. M. Kerminen, Y. C. Liu, L. Wang, J. Zheng, M. Kulmala & J. K. Jiang An indicator for sulfuric acid-amine nucleation in atmospheric environments. *Aerosol Science and Technology,* 55**,** 1059-1069, 2021.

Chang, Y. H., K. Cheng, Y. Q. Kuang, Q. Y. Hu, Y. Q. Gao, R. J. Huang, C. Huang, W. W. Walters & M. F. Lehmann Isotopic Variability of Ammonia (δ^{15}N-NH_3) Slipped from Heavy-Duty Vehicles under Real-World Conditions. *Environ Sci Technol Letters,* 9**,** 726-732, 2022.

Cheng, S. F., Y. B. Zhao, B. B. Zhang, P. Peng & F. Lu Structural decomposition of heavy-duty diesel truck emission contribution based on trajectory mining. *Journal of Cleaner Production,* 380**,** 135172, 2022.

Cope, J. D., K. A. Abellar, K. H. Bates, X. Fu & T. B. Nguyen Aqueous Photochemistry of 2-Methyltetrol and Erythritol as Sources of Formic Acid and Acetic Acid in the Atmosphere. *ACS Earth and Space Chemistry,* 5**,** 1265-1277, https://doi.org/10.1021/acsearthspacechem.1c00107, 2021.

Du, P. & R. H. Angeletti Automatic deconvolution of isotope-resolved mass spectra using variable selection and quantized peptide mass distribution. *Anal Chem,* 78**,** 3385-3392, https://www.ncbi.nlm.nih.gov/pubmed/16689541, 2006.

Fernandez-de-Cossio, J., L. J. Gonzalez, Y. Satomi, L. Betancourt, Y. Ramos, V. Huerta, V. Besada, G. Padron, N. Minamino & T. Takao Automated interpretation of mass spectra of complex mixtures by matching of isotope peak distributions. *Rapid Commun Mass Spectrom,* 18**,** 2465-2472, https://www.ncbi.nlm.nih.gov/pubmed/15384131, 2004.

Franklin, E. B., S. Amiri, D. Crocker, C. Morris, K. Mayer, J. S. Sauer, R. J. Weber, C. Lee, F. Malfatti, C. D. Cappa, T. H. Bertram, K. A. Prather & A. H. Goldstein Anthropogenic and Biogenic Contributions to the Organic Composition of Coastal Submicron Sea Spray Aerosol. *Environ Sci Technol,* 56**,** 16633-16642, https://www.ncbi.nlm.nih.gov/pubmed/36332100, 2022.

Franklin, E. B., L. D. Yee, R. Wernis, G. Isaacman-VanWertz, N. Kreisberg, R. Weber, H. Zhang, B. B. Palm, W. Hu, P. Campuzano-Jost, D. A. Day, A. Manzi, P. Artaxo, R. A. F. Souza, J. L. Jimenez, S. T. Martin & A. H. Goldstein Chemical Signatures of Seasonally Unique Anthropogenic Influences on Organic Aerosol Composition in the Central Amazon. *Environ Sci Technol,* 57**,** 6263-6272, https://www.ncbi.nlm.nih.gov/pubmed/37011031, 2023.

Guo, Q., J. Yu, K. Yang, X. Wen, H. Zhang, Z. Yu, H. Li, D. Zhang & M. Yang Identification of complex septic odorants in Huangpu River source water by combining the data from gas chromatography-olfactometry and comprehensive two-dimensional gas chromatography using retention indices. *Sci Total Environ,* 556**,** 36-44, https://www.ncbi.nlm.nih.gov/pubmed/26974564, 2016.

Hamilton, S. D. & R. A. Harley High-Resolution Modeling and Apportionment of Diesel-Related Contributions to Black Carbon Concentrations. *Environ Sci Technol,* 55**,** 12250-12260, 2021.

He, X., X. Zheng, Y. You, S. Zhang, B. Zhao, X. Wang, G. Huang, T. Chen, Y. Cao, L. He, X. Chang, S. Wang & Y. Wu 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, https://www.ncbi.nlm.nih.gov/pubmed/35436508, 2022a.

He, X., X. Zheng, S. Zhang, X. Wang, T. Chen, X. Zhang, G. Huang, Y. Cao, L. He, X. Cao, Y. Cheng, S. Wang & Y. Wu 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, https://dx.doi.org/10.5194/acp-22-13935-2022, 2022b.

Huang, X. F., B. B. Zou, L. Y. He, M. Hu, A. S. H. Prévôt & Y. H. Zhang Exploration of PM2.5 sources on the regional scale in the Pearl River Delta based on ME-2 modeling. *Atmos. Chem. Phys.*, 18**,** 11563-11580, https://acp.copernicus.org/articles/18/11563/2018/, 2018.

Huo, Y., Z. Guo, Q. Li, D. Wu, X. Ding, A. Liu, D. Huang, G. Qiu, M. Wu, Z. Zhao, H. Sun, W. Song, X. Li, Y. Chen, T. Wu & J. Chen Chemical Fingerprinting of HULIS in Particulate Matters Emitted from Residential Coal and Biomass Combustion. *Environ Sci Technol*, 55**,** 3593-3603, https://www.ncbi.nlm.nih.gov/pubmed/33656861, 2021.

Kosyakov, D. S., N. V. Ul'yanovskii, T. B. Latkin, S. A. Pokryshkin, V. R. Berzhonskis, O. V. Polyakova & A. T. Lebedev Peat burning - An important source of pyridines in the earth atmosphere. *Environ Pollut*, 266**,** 115109, https://www.ncbi.nlm.nih.gov/pubmed/32622216, 2020.

Kruve, A., K. Kaupmees, J. Liigand & I. Leito Negative electrospray ionization via deprotonation: predicting the ionization efficiency. *Anal Chem*, 86**,** 4822-30, https://www.ncbi.nlm.nih.gov/pubmed/24731109, 2014.

Phillips, K. A., A. Yau, K. A. Favela, K. K. Isaacs, A. McEachran, C. Grulke, A. M. Richard, A. J. Williams, J. R. Sobus, R. S. Thomas & J. F. Wambaugh Suspect Screening Analysis of Chemicals in Consumer Products. *Environ Sci Technol*, 52**,** 3125-3135, https://www.ncbi.nlm.nih.gov/pubmed/29405058, 2018.

Piotrowski, P. K., B. A. Weggler, D. A. Yoxtheimer, C. N. Kelly, E. Barth-Naftilan, J. E. Saiers & F. L. Dorman Elucidating Environmental Fingerprinting Mechanisms of Unconventional Gas Development through Hydrocarbon Analysis. *Anal Chem*, 90**,** 5466-5473, https://www.ncbi.nlm.nih.gov/pubmed/29580048, 2018.

Saeki, K., K. Ikari, H. Yokoi, S. I. Ohira, H. Okochi & K. Toda Biogenic Diamines and Their Amide Derivatives Are Present in the Forest Atmosphere and May Play a Role in Particle Formation. *ACS Earth and Space Chemistry*, 6**,** 421-430, 2022.

Shen, X., H. Che, Z. Yao, B. Wu, T. Lv, W. Yu, X. Cao, X. Hao, X. Li, H. Zhang & X. Yao Real-World Emission Characteristics of Full-Volatility Organics Originating from Nonroad Agricultural Machinery during Agricultural Activities. *Environ Sci Technol*, 57**,** 10308-10318, https://www.ncbi.nlm.nih.gov/pubmed/37419883, 2023.

Silva, L. F. M., A. R. H. De La Cruz, A. H. M. Nunes & A. Gioda Real-Time Monitoring of Nitrogen Oxides Emission Factors Using Sensors in the Exhaust Pipes of Heavy Vehicles in the Metropolitan Region of Rio de Janeiro. *Journal of the Brazilian Chemical Society*, 2023.

Song, K., R. Tang, J. Zhang, Z. Wan, Y. Zhang, K. Hu, Y. Gong, D. Lv, S. Lu, Y. Tan, R. Zhang, A. Li, S. Yan, S. Yan, B. Fan, W. Zhu, C. K. Chan & S. Guo. 2023. Molecular fingerprints and health risks of home-use incense burning smoke. Copernicus GmbH.

Stanimirova, I., D. Q. Rich, A. G. Russell & P. K. Hopke A long-term, dispersion normalized PMF source apportionment of PM2.5 in Atlanta from 2005 to 2019. *Atmospheric Environment*, 312**,** 120027, 2023.

Stefanuto, P.-H., A. Smolinska & J.-F. Focant Advanced chemometric and data handling tools for GC×GC-TOF-MS Application of chemometrics and related advanced data handling in chemical separations. *TrAC Trends in Analytical Chemistry*, 139**,** 116251, 2021.

Stout, S. A. & G. S. Douglas Diamondoid hydrocarbons - Application in the chemical fingerprinting of natural gas condensate and gasoline. *Environmental Forensics*, 5**,** 225-235, 2004.

Wang, A. Q., Z. B. Yuan, X. H. Liu, M. L. Wang, J. Yang, Q. E. Sha & J. Y. Zheng Measurement-based intermediate volatility organic compound emission inventory from on-road vehicle exhaust in China. *Environmental Pollution*, 310, 2022.

Wang, G. L., S. B. Shi, P. R. Wang & T. G. Wang Analysis of diamondoids in crude oils using comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Fuel*, 107**,** 706-714, 2013.

Wang, H., S. J. Zhang, X. M. Wu, Y. F. Wen, Z. H. Li & Y. Wu Emission Measurements on a Large Sample of Heavy-Duty Diesel Trucks in China by Using Mobile Plume Chasing. *Environ Sci Technol*, 57**,** 15153-15161, 2023.

Welthagen, W., J. Schnelle-Kreis & R. Zimmermann Search criteria and rules for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry analysis of airborne particulate matter. *J Chromatogr A*, 1019**,** 233-49, https://www.ncbi.nlm.nih.gov/pubmed/14650618, 2003.

Wong, Y. K., X. H. H. Huang, Y. Y. Cheng & J. Z. Yu Estimating primary vehicular emission contributions to PM2.5 using the Chemical Mass Balance model: Accounting for gas-particle partitioning of organic aerosols and oxidation degradation of hopanes. *Environmental Pollution*, 291**,** 118131, 2021.

Wu, L. L. & A. S. Geng Differences in the thermal evolution of hopanes and steranes in free and bound fractions. *Organic Geochemistry*, 101**,** 38-48, 2016.

Xu, B., G. Zhang, O. Gustafsson, K. Kawamura, J. Li, A. Andersson, S. Bikkina, B. Kunwar, A. Pokhrel, G. Zhong, S. Zhao, J. Li, C. Huang, Z. Cheng, S. Zhu, P. Peng & G. Sheng Large contribution of fossil-derived components to aqueous secondary organic aerosols in China. *Nat Commun*, 13**,** 5115, https://www.ncbi.nlm.nih.gov/pubmed/36045131, 2022.

Yan, J. Z., G. Wang, S. Y. Chen, H. Zhang, J. Q. Qian & Y. X. Mao Harnessing freight platforms to promote the penetration of long-haul heavy-duty hydrogen fuel-cell trucks. *Energy*, 254**,** 124225, 2022.

Yu, D., Z. Tan, K. Lu, X. Ma, X. Li, S. Chen, B. Zhu, L. Lin, Y. Li, P. Qiu, X. Yang, Y. Liu, H. Wang, L. He, X. Huang & Y. Zhang An explicit study of local ozone budget and NOx-VOCs sensitivity in Shenzhen China. *Atmospheric Environment*, 224**,** 117304, https://www.sciencedirect.com/science/article/pii/S1352231020300467, 2020.

Zhang, Y., R. Li, J. Fang, C. Wang & Z. Cai Simultaneous determination of eighteen nitro-polyaromatic hydrocarbons in PM(2.5) by atmospheric pressure gas chromatography-tandem mass spectrometry. *Chemosphere*, 198**,** 303-310, https://www.ncbi.nlm.nih.gov/pubmed/29421744, 2018.

Zhao, Y., C. J. Hennigan, A. A. May, D. S. Tkacik, J. A. de Gouw, J. B. Gilman, W. C. Kuster, A. Borbon & A. L. Robinson Intermediate-volatility organic compounds: a large source of secondary organic aerosol. *Environ Sci Technol*, 48**,** 13743-137550, https://www.ncbi.nlm.nih.gov/pubmed/25375804, 2014.

Zhao, Y., N. T. Nguyen, A. A. Presto, C. J. Hennigan, A. A. May & A. L. Robinson 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, https://www.ncbi.nlm.nih.gov/pubmed/26322746, 2015.

Zhao, Y., D. S. Tkacik, A. A. May, N. M. Donahue & A. L. Robinson Mobile Sources Are Still an Important Source of Secondary Organic Aerosol and Fine Particulate Matter in the Los Angeles Region. *Environ Sci Technol*, 56**,** 15328-15336, https://www.ncbi.nlm.nih.gov/pubmed/36215417, 2022.

600 For Table of Contents Only

601