



*Supplement of*

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

**Xiao He et al.**

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

The copyright of individual parts of the supplement might differ from the article licence.

## 1 **S1. Indicative reaction schemes that are included in the model development**

2 Compounds containing hydrocarbon chains generate a series of ions separated by 14 Da (-CH<sub>2</sub>-) in  
3 their mass spectra, with the specific masses depending on their chemical structures. For example,  
4 aliphatic alkanes, including *n*-/*i*-alkanes, regularly dissociate to form ions with mass-to-charge ratios  
5 (m/z) of 43, 57, 71, and 85, corresponding to C<sub>3</sub>H<sub>7</sub>, C<sub>4</sub>H<sub>9</sub>, C<sub>5</sub>H<sub>11</sub>, and C<sub>6</sub>H<sub>13</sub>, respectively. The M<sup>+</sup>  
6 ion is weakly observed (Lavanchy, Houriet and Gäumann 1979). Aliphatic alkenes exhibit the most  
7 abundant ion fragments at m/z 41, 55, 69, and 83. A short alkyl chain attached to the benzene ring  
8 yields a prominent peak at m/z 91, with the base peak (ion with the highest abundance relative to  
9 other ions in the mass spectrum) gradually shifting to larger ions as the alkyl chain lengthens.  
10 Saturated rings lose one methyl group to produce an abundant [M - 15]<sup>+</sup> signal. The appearance of  
11 the m/z 60 peak for carboxylic acids is a characteristic feature of the McLafferty rearrangement,  
12 providing valuable information about the identification of the compounds (Figure S3b). Aliphatic  
13 alcohols frequently cleave at the β bond, yielding a resonance-stabilized ion at m/z 31 (Figure S3c).  
14 In contrast, aromatic alcohols generate a prominent M<sup>+</sup> ion, as the case for phenols. Esters,  
15 composed of alcohol and acid components, exhibit distinctive fragmentation patterns depending on  
16 the major component. When the acid component predominates, the fragmentation is partially  
17 characterized by typical acid peaks, where the loss of an alkyl radical through α-cleavage occurs  
18 predominantly, producing a discernable peak at m/z 74 (Figure S3f). Conversely, when the alcohol  
19 component dominates, fragmentation patterns resemble those of alcohols (Figure S3g).  
20 Trimethylsilyl derivatization is commonly used in GC analysis, with observed abundant fragments  
21 at m/z 73 and 147 (He et al. 2018). Water elimination, indicated by an 18 Da loss, is common for  
22 hydroxyl groups (Harrison 2012). The McLafferty rearrangement is prominent in the mass spectra  
23 due to γ-hydrogen atom transfer to a keto-group and the β-cleavage through a six-atom ring where  
24 appropriate (McLafferty 1959). Besides, the heteroatoms (e.g., oxygen) suppress M<sup>+</sup> appearance,  
25 while the conjugated aromatic rings intensify the M<sup>+</sup>. Aliphatic amines often undergo cleavage at  
26 the α-C-C bond to produce relatively stable ions: CH<sub>2</sub>NH<sub>2</sub><sup>+</sup> (m/z 30), C<sub>2</sub>H<sub>4</sub>NH<sub>2</sub><sup>+</sup> (m/z 44), and  
27 C<sub>3</sub>H<sub>6</sub>NH<sub>2</sub><sup>+</sup> (m/z 58) for amine groups attached to the primary, secondary, and tertiary carbons,  
28 respectively (Figure S4a) (Mikaia 2022). M<sup>+</sup> is frequently detected for amides, following the  
29 nitrogen rule. α-cleavage or McLafferty rearrangement produces dominant ions in the mass  
30 spectrum, as depicted in Figure S4b. Neutral loss of HC≡N and McLafferty rearrangement produce  
31 abundant [M - 27]<sup>+</sup> and m/z 41 ions for nitriles. C-N and N-O bonds are weak for aliphatic nitro  
32 compounds, and M<sup>+</sup> is rarely observed in the EI spectrum. Aromatic nitro compounds exhibit  
33 additional fragmentation patterns involving ring arrangements and ring-opening reactions. The  
34 dominant ions originate from the loss of NO<sup>+</sup> (m/z 30) or NO<sub>2</sub><sup>+</sup> (m/z 46).

35 When incorporating these rules into the data treatment software (Canvas, version 2.5, J&X  
36 Technologies), several steps are necessary, as depicted in Figure S5. The software offers four built-  
37 in features: ABUND (X) returns the normalized abundance of the input ion mass; HASMASS (X)  
38 indicates whether the input ion exists; ORDER (X) specifies the order of the input ion mass; MASS  
39 (X) returns the mass of the input ion's order. Additionally, the function supports two logical  
40 operators, "And" and "Or". Subsequently, cluster of alkanes can be extracted by the following rules:  
41 ((MASS(1)=43 && (MASS(2)=57 || MASS(2)=71 || MASS(2)=41)) || (MASS(1)=57 &&  
42 (MASS(2)=43 || MASS(2)=71 || MASS(2)=41)))  
43 where "&&" and "||" refers to the logical operators "And" and "Or", respectively. Paste the rules in  
44 Ion Extractor Editor and the cluster of alkanes can be isolated.

45 Similar, the cluster of aliphatic amines can be extracted by the following rules:

46 (MASS(1) = 30 && ABUND(MASS(2)) < 20) || (MASS(1) = 58 && ABUND(MASS(2)) < 40)||  
47 (MASS(1) = 58 && MASS(2) = 59)|| (MASS(1) = 30 && (MASS(2) = 31 || MASS(2) =  
48 28))||((ABUND(30) + ABUND(44)) > 100)

49 In total, the extraction rules for 26 compound clusters are constructed with high accuracy and  
50 repeatability.

## 51 **S2. The description of unresolved peaks**

52 The unidentified peaks accounted for 15%, 18%, and less than 1% for HDDV vapor, HDDV aerosol,  
53 and ambient aerosol samples, respectively. These identified peaks can be divided into three sub-  
54 categories:

55 The first category comprises column bleedings that regularly appear in the GC×GC plot underneath  
56 alkane groups. Most column bleedings occur towards the end of GC thermal program when the  
57 temperature exceed 240 °C. These column bleedings were promptly identified and removed at the

58 outset of the data treatment process. Some column bleedings with aliphatic siloxane skeletons  
 59 appeared regularly throughout the GC×GC tests. The number of silicon atoms within the skeletons  
 60 varied much, resulting in different characteristic ions. These bleeding peaks often overlapped with  
 61 chromatographic peaks. Their mass spectra and example chemical structures are shown in Figure  
 62 S6.

63 The second category includes ester compounds scattering in the GC×GC plot. We have observed  
 64 some aliphatic ester peaks in the middle of the GC×GC plot, one example given in Figure S7. Their  
 65 mass spectra were compared to the National Institute of Standards and Technology (NIST) library,  
 66 and these peaks were identified with high match scores. However, these ester peaks were infrequent,  
 67 contributing to less than 1% of the total signal, and were not consistently detected in every sample.  
 68 In this regard, the ester compounds were not further discussed and classified as unidentified.

69 The third category pertains to the unresolved peaks literally. The mass spectra of these peaks were  
 70 disorganized and lacked distinct patterns. Some fragment ions originated from their own peaks,  
 71 while others were interferences from the matrix. These peaks could not be categorized into any  
 72 specific groups within the model and were labeled as unidentified.

73 Table S1. Information of the two test HDDVs.

Vehicle ID	Emission standard	Aftertreatment	Model year	Gross vehicle weight (kg)	Vehicle type	Mileage ( $\times 10^3$ km)	Engine model
#1	China V	Selective catalytic reduction	2021	25000	Semi-trailer tractor	22.2	dCi450-51
#2	China V	Selective catalytic reduction	2021	25000	Semi-trailer tractor	34.8	MC13.54-50

74

75 Table S2. List of ambient samples and the corresponding PM concentration.

Sample ID	Collection date	PM <sub>2.5</sub> concentration ( $\mu\text{g}/\text{m}^3$ )	PM <sub>10</sub> concentration ( $\mu\text{g}/\text{m}^3$ )
#1	1-Nov-2023	17	38
#2	5-Nov-2023	14	28
#3	10-Nov-2023	16	35
#4	11-Nov-2023	13	28
#5	17-Nov-2023	19	66
Blank	17-Nov-2023	19	66

76

77 Table S3. List of the deuterated internal standards.

Internal standard	Formula
<i>n</i> -Dodecane-d <sub>26</sub>	C <sub>12</sub> D <sub>26</sub>
<i>n</i> -Hexadecane-d <sub>34</sub>	C <sub>16</sub> D <sub>34</sub>
<i>n</i> -Eicosane-d <sub>42</sub>	C <sub>20</sub> D <sub>42</sub>
<i>n</i> -Tetracosane-d <sub>50</sub>	C <sub>24</sub> D <sub>50</sub>
<i>n</i> -Octacosane-d <sub>58</sub>	C <sub>28</sub> D <sub>58</sub>
Naphthalene-d <sub>8</sub>	C <sub>10</sub> D <sub>8</sub>
Acenaphthene-d <sub>10</sub>	C <sub>12</sub> D <sub>10</sub>
Phenanthrene-d <sub>10</sub>	C <sub>14</sub> D <sub>10</sub>
Chrysene-d <sub>12</sub>	C <sub>18</sub> D <sub>12</sub>

78

79 Table S4. The list and linear correlation of the authentic standards.

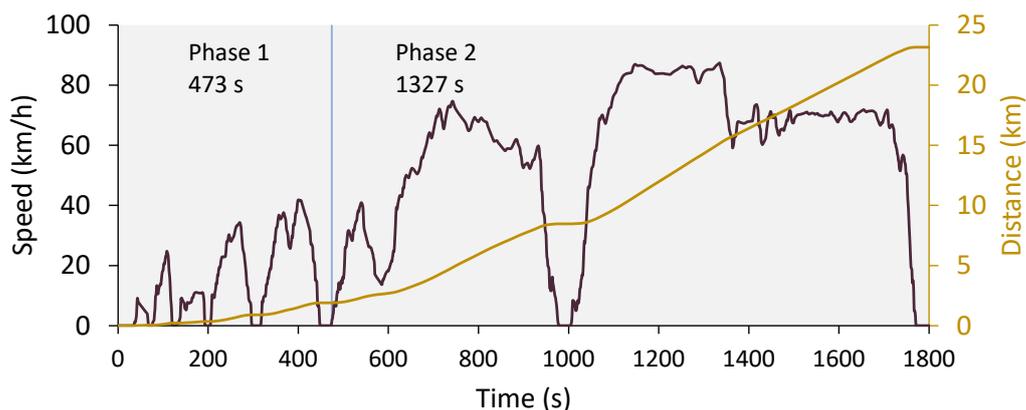
Group	Name	1 <sup>st</sup> RT (min)	2 <sup>nd</sup> RT (s)	R <sup>2</sup>
Acid	Isobutyric Acid	4.67	1.25	0.99
Acid	2-Methyl butyric acid	7.07	1.36	0.99
Alcohol	1-Decanol	19.67	1.42	0.98
Alcohol	1-Hexadecanol	34.14	1.38	0.96

Alcohol	n-Nonadecanol	39.94	1.39	0.95
Alcohol	1-Docosanol	45.00	1.43	0.99
Aldehyde	Octanal	11.60	1.48	1.00
Aldehyde	Decanal	17.80	1.46	0.99
Aldehyde	Pentanal	3.73	1.17	0.97
Aldehyde	Hexanal	5.67	1.41	1.00
Aldehyde	Decanal	17.80	1.55	0.99
Alkane	Heptane	3.74	0.74	0.97
Alkane	Octane	5.60	0.87	0.98
Alkane	Nonane	8.34	0.94	1.00
Alkane	Decane	11.47	0.97	1.00
Alkane	Undecane	14.67	0.98	0.99
Alkane	Dodecane	17.67	0.97	0.97
Alkane	Tridecane	20.47	0.99	0.97
Alkane	Tetradecane	23.20	0.98	0.99
Alkane	Pentadecane	25.67	1.00	0.99
Alkane	Hexadecane	28.07	0.98	0.99
Alkane	Heptadecane	30.34	1.00	0.97
Alkane	Octadecane	32.54	1.04	0.97
Alkane	Nonadecane	34.60	1.03	0.99
Alkane	Eicosane	36.54	1.07	0.96
Alkane	Heneicosane	38.40	1.08	0.96
Alkane	Docosane	40.20	1.11	0.98
Alkane	Tricosane	41.94	1.12	0.97
Alkane	Tetracosane	43.60	1.12	0.97
Alkane	Pentacosane	45.20	1.17	0.99
Alkane	Hexacosane	46.74	1.22	0.99
Alkane	Heptacosane	48.20	1.20	0.98
Alkane	Octacosane	49.60	1.25	0.98
Alkane	Nonacosane	51.00	1.27	0.97
Alkane	Triacontane	52.34	1.33	0.96
Alkane	Hentriacontane	53.60	1.41	0.97
Alkane	Dotriacontane	54.87	1.53	0.96
Alkane	Tritriacontane	56.14	1.77	0.96
Alkane	Tetracontane	57.60	2.09	0.96
Alkane	Pentatriacontane	59.27	2.46	0.99
Alkene	1-Octene	5.39	0.90	0.99
Alkene	1-Octadecene	32.39	1.07	0.99
Alkene	1-Docosene	40.13	1.13	0.99
Alkene	1-Decene	11.19	0.99	0.98
Alkene	1-Dodecene	17.39	1.03	0.96
Alkene	1-Tetradecene	22.99	1.02	0.98
Alkyl-benzene	Ethylbenzene	7.27	1.46	1.00
Alkyl-benzene	p-Xylene	7.54	1.42	0.99

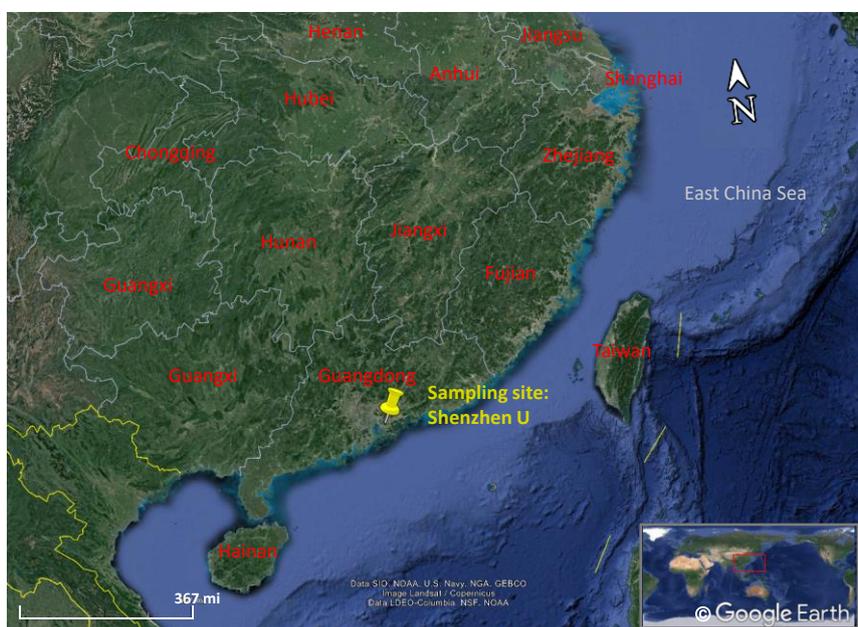
Alkyl-benzene	o-Xylene	8.14	1.57	0.99
Alkyl-benzene	Isopropylbenzene	9.07	1.51	1.00
Alkyl-benzene	1,3,5-Trimethylbenzene	10.54	1.55	0.99
Alkyl-benzene	Butyl-benzene	13.27	1.55	0.99
Alkyl-benzene	Pentyl-benzene	16.40	1.54	0.97
Alkyl-benzene	Hexyl-benzene	19.40	1.50	0.97
Alkyl-benzene	4-Ethyltoluene	10.40	1.61	0.98
Alkyl-benzene	p-Cymene	12.33	1.61	0.97
Alkyl-PAHs	1-Methylnaphthalene	20.87	2.44	1.00
Alkyl-PAHs	1-Ethyl-naphthalene	23.14	2.41	0.97
Amide	Acetamide	5.07	2.51	1.00
Amide	Propanamide	7.00	2.79	0.96
Amide	N,N-Dibutylformamide	20.54	1.95	0.95
Amine	Triethylamine	3.60	0.86	0.99
Amine	Dibutylamine	10.27	1.18	0.99
Amine	Aniline	10.74	2.58	0.98
Amine	1,2-Benzenediamine	17.80	3.50	1.00
Amine	1-Naphthalenamine	26.60	3.57	0.96
Amine	4-Biphenylamine	31.54	3.56	0.99
Cycloalkane	Ethyl-cyclohexane	6.59	0.99	0.95
Cycloalkane	Butyl-cyclohexane	12.59	1.12	0.99
Cycloalkane	Hexyl-cyclohexane	18.86	1.16	0.97
Cycloalkane	Octyl-cyclohexane	24.46	1.13	0.95
Cycloalkane	Decyl-cyclohexane	29.46	1.15	0.95
Cycloalkane	Dodecyl-cyclohexane	33.93	1.21	0.97
Cycloalkane	Tetradecyl-cyclohexane	37.99	1.24	0.97
Cycloalkane	Hexadecyl-cyclohexane	41.66	1.29	0.98
Ester	Butyl acetate	5.93	1.39	0.98
Ester	Amyl Acetate	8.73	1.51	0.97
Ester	Isoamyl Acetate	7.66	1.39	0.99
Furan	Furan	12.46	1.40	0.96
Ketone	2-Nonanone	14.34	1.50	0.98
Ketone	2-Dodecanone	23.00	1.43	0.97
Ketone	2-Pentadecanone	30.34	1.38	0.99
Ketone	2-Pentanone	3.53	1.09	0.99
Ketone	3-Heptanone	7.93	1.49	0.99
Oxy-PAH	1,4-Naphthalenedione	23.60	3.36	0.99
Oxy-PAH	9,10-Anthracenedione	36.07	3.85	0.99
Oxy-PAH	Benz(a)anthracene-7,12-dione	46.74	4.62	0.98
PAHs	Indene	12.94	2.12	0.99
PAHs	Naphthalene	17.26	2.95	1.00
PAHs	Acenaphthylene	24.53	3.38	1.00
PAHs	Acenaphthene	25.39	3.27	0.99
PAHs	Fluorene	27.79	3.30	1.00

PAHs	Phenanthrene	32.26	3.73	0.99
PAHs	Anthracene	32.53	3.70	0.99
PAHs	Fluoranthene	37.00	3.04	0.99
PAHs	Pyrene	38.86	4.38	0.99
PAHs	Benz[a]anthracene	44.59	4.58	0.97
PAHs	Chrysene	44.73	4.76	0.97
PAHs	Benzo[b]fluoranthene	49.26	5.15	0.96
PAHs	Benzo[k]fluoranthene	49.39	5.15	0.98
PAHs	Benzo[a]pyrene	50.53	5.61	0.97
PAHs	Indeno[1,2,3-cd]fluoranthene	55.46	3.02	0.99
Pinene	$\alpha$ -Pinene	9.46	1.19	0.95
Pinene	$\beta$ -Pinene	10.86	1.35	0.99
Polyphenyl	Biphenyl	22.74	2.44	0.95

80

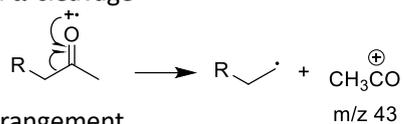


81  
 82 Figure S1. Speed trace for China heavy-duty commercial vehicle test cycle for tractor trailers  
 83 (CHTC-HT) cycle. The total duration is 1800 s and divided into two phases of 344 s and 1330 s.  
 84 The average speed is 46.5 km/h, and the idle ratio is 8.6%.  
 85

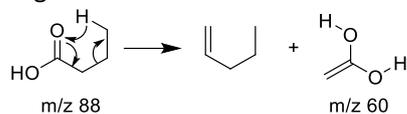


86  
 87 Figure S2. Sampling site location on Shenzhen University campus. PM<sub>2.5</sub> high-volume sampler was  
 88 installed on the rooftop of a 5-story building surrounded by university campus, residential areas,  
 89 greenbelts, and a golf park.

(a) Radical initiated  $\alpha$ -cleavage



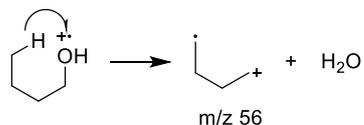
(b) McLafferty rearrangement



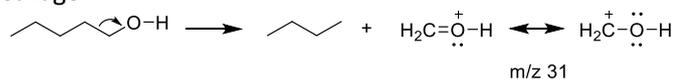
(c) Loss of  $\text{CH}_2\text{OH}$



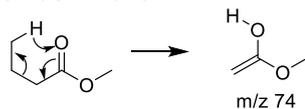
(d) Loss of  $\text{H}_2\text{O}$



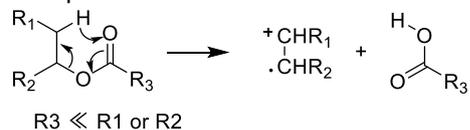
(e)  $\beta$ -cleavage



(f) Ester with acid portion dominant



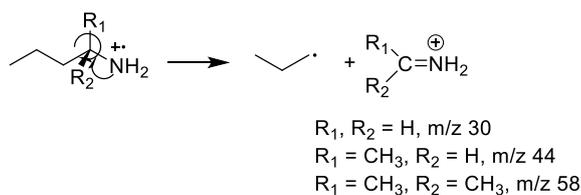
(g) Ester with alcohol portion dominant



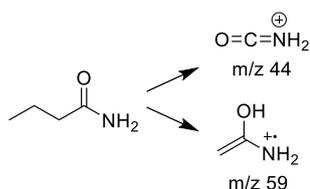
90

91 Figure S3. Interpretation of characteristic ions in the EI spectrum for O-containing species,  
92 including (a) aliphatic ketones, (b-d) aliphatic alcohols, (e) carboxylic acids, (f) esters with acid  
93 portion dominant, and (g) esters with alcohol portion dominant.

(a) Radical initiated  $\alpha$ -cleavage



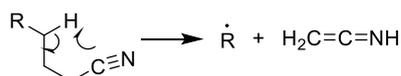
(b) McLafferty rearrangement



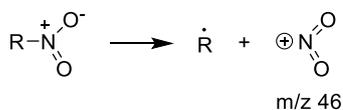
(c) Loss of HCN



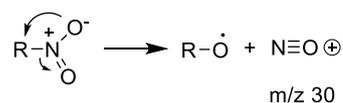
(d) McLafferty rearrangement



(e) Loss of NO<sub>2</sub>

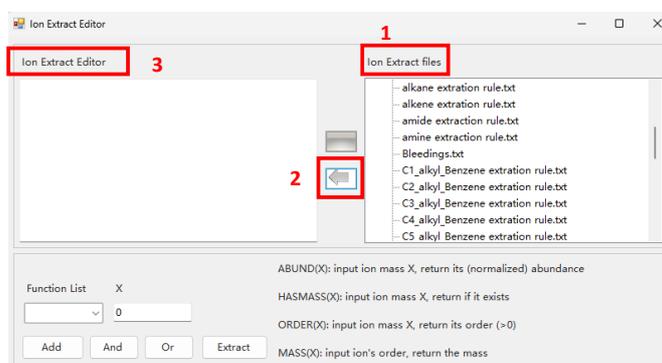


(f) Loss of NO

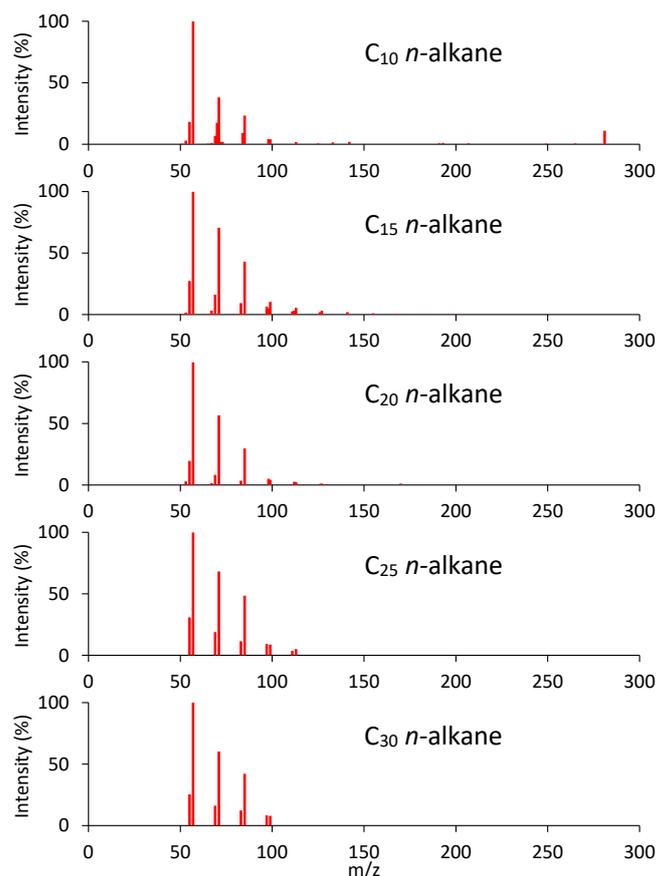


94  
95 Figure S4. Interpretation of the characteristic ions in the EI spectrum for N-containing species,  
96 including (a) aliphatic amines, (b) amides, (c) and (d) nitriles, and (e) organic nitrates, and (e)  
97 organic nitrites.  
98

Open Canvas → Open Browser → File → Load Data (load a sample) →  
Speciation → Find All Peaks → Mass Spectrum → Extraction Rules

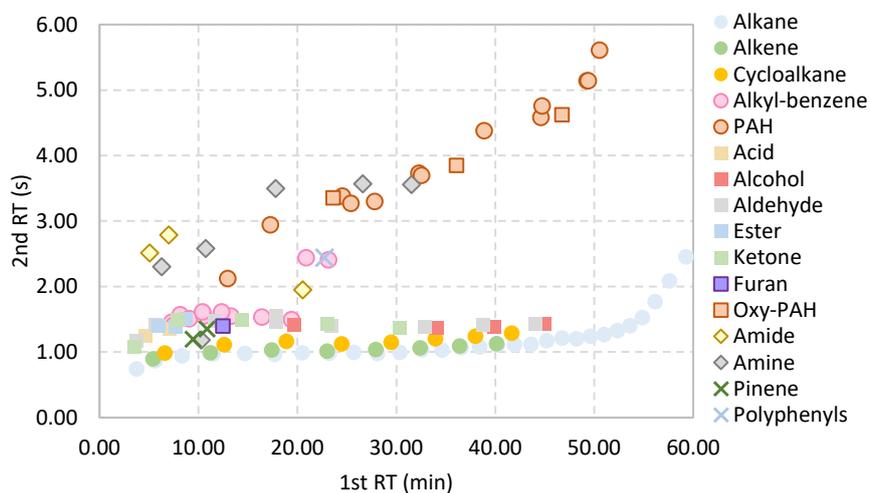


99  
100 Figure S5. The steps to enable the ion extract function built in Canvas.



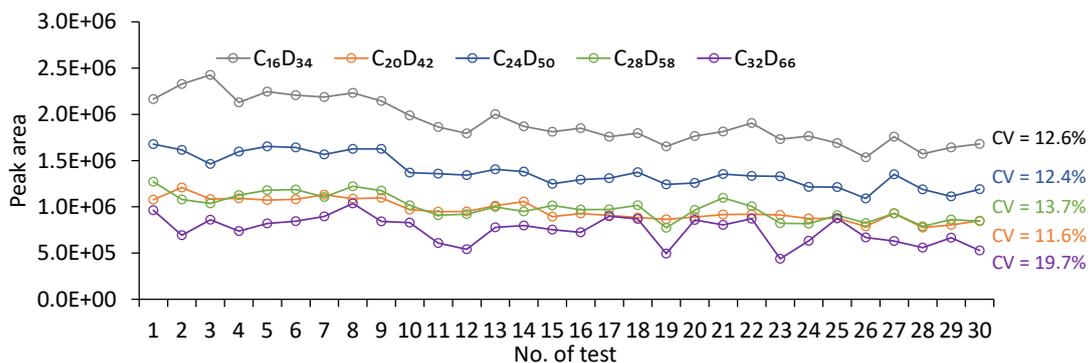
101  
102  
103  
104  
105

Figure S6. The mass spectra of homologous *n*-alkane standards from C<sub>10</sub> to C<sub>30</sub>, which exhibit high similarity. In such cases, the similarity score could be elevated incorrectly and mislead compound identification results.



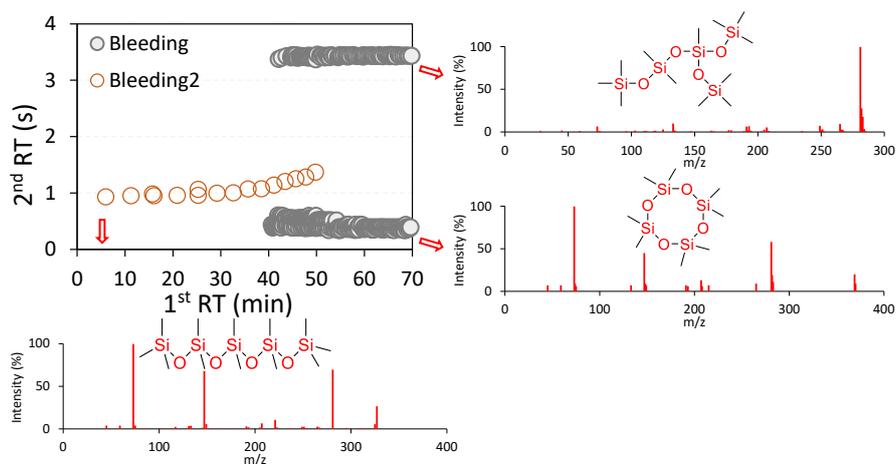
106  
107

Figure S7. The distribution of all authentic standards, coded by colors and shapes.



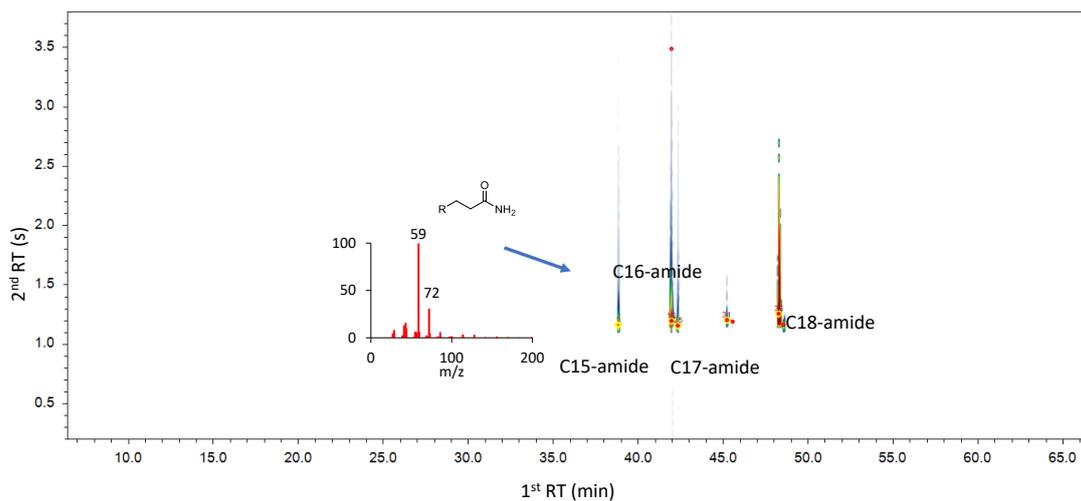
108  
109  
110  
111

Figure S8. The variation of peak areas of internal standards tested throughout the whole measurement.



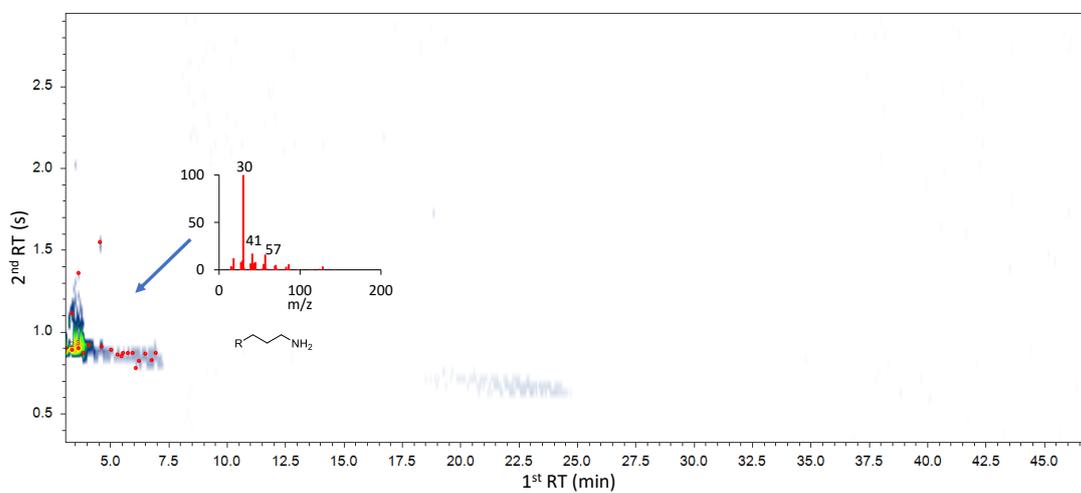
112  
113  
114

Figure S9. The distribution of the column bleedings in an example GCxGC plot and their respective mass spectra and chemical structures.



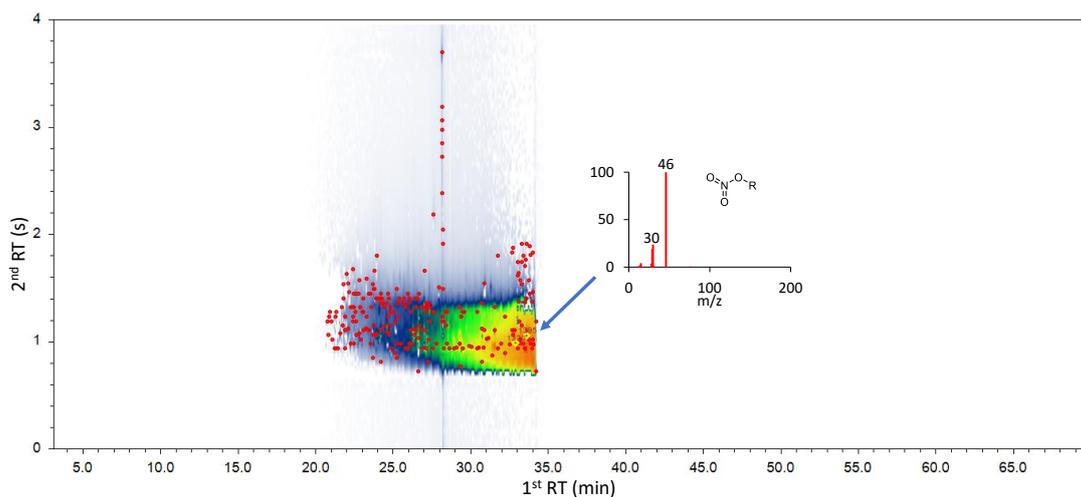
115  
116  
117

Figure S10. The chromatogram and mass spectra of tentatively identified amide representatives.



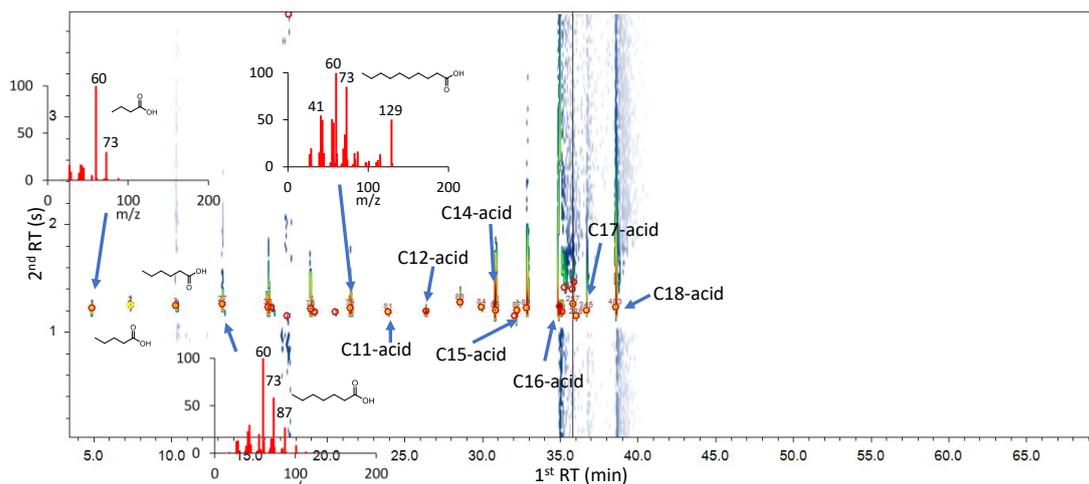
118  
119  
120

Figure S11. The chromatogram and mass spectra of tentatively identified amine representatives.



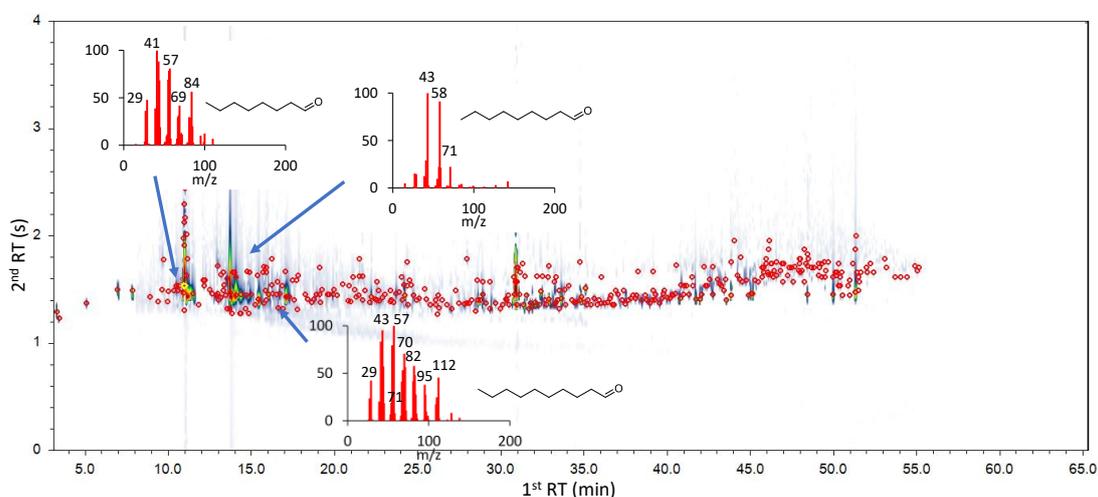
121  
122

Figure S12. The chromatogram and mass spectra of tentatively identified nitro representatives.



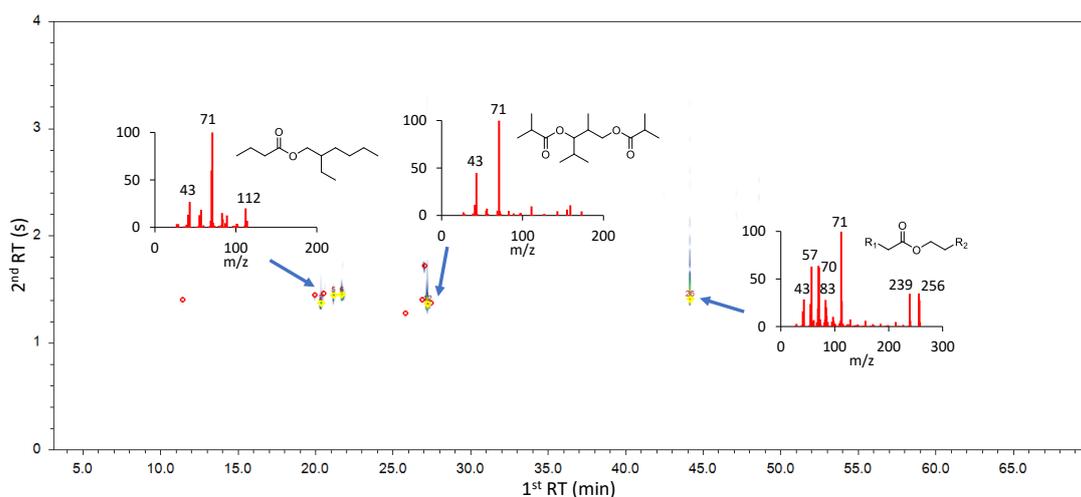
123  
124  
125  
126

Figure S13. The chromatogram and mass spectra of tentatively identified aliphatic acid representatives.



127  
128  
129  
130

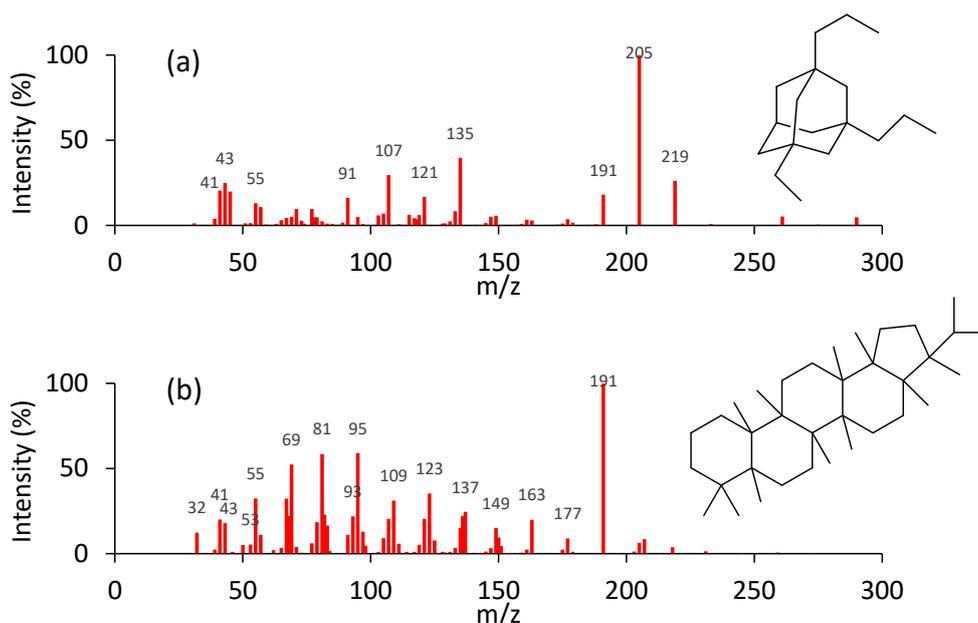
Figure S14. The chromatogram and mass spectra of tentatively identified aliphatic aldehyde and ketone representatives.



131  
132  
133

Figure S15. The chromatogram and mass spectra of tentatively identified aliphatic ester representatives.





142  
 143 Figure S18. The chemical structures and mass spectra of (a) 1.1-Ethyl-3,5-dipropyladamantane  
 144 adamantane and (b) 17 $\alpha$ (H),21 $\beta$ (H)-hopane.  
 145

146 **References:**

- 147 Harrison A. G.: Pathways for water loss from doubly protonated peptides containing serine or threonine.  
 148 *J Am Soc Mass Spectrom*, 23, 116-123, <https://www.ncbi.nlm.nih.gov/pubmed/22065406>, 2012.  
 149 He X., Huang X. H. H., Chow K. S., Wang Q., Zhang T., Wu D. and Yu J. Z.: Abundance and Sources of  
 150 Phthalic Acids, Benzene-Tricarboxylic Acids, and Phenolic Acids in PM<sub>2.5</sub> at Urban and Suburban Sites  
 151 in Southern China. *ACS Earth and Space Chemistry*, 2, 147-158, 2018.  
 152 Lavanchy A., Houriet R. and Gäumann T.: The mass spectrometric fragmentation of n-alkanes. *Organic*  
 153 *Mass Spectrometry*, 14, 79-85, <https://doi.org/10.1002/oms.1210140205>, 1979.  
 154 McLafferty F. W.: Mass Spectrometric Analysis. Molecular Rearrangements. *Anal Chem*, 31, 82-87,  
 155 <https://dx.doi.org/10.1021/ac60145a015>, 1959.  
 156 Mikaia A.: Protocol for Structure Determination of Unknowns by EI Mass Spectrometry. I. Diagnostic  
 157 Ions for Acyclic Compounds with up to One Functional Group. *Journal of Physical and Chemical*  
 158 *Reference Data*, 51, 031501, 2022.  
 159