Reviewer #3:

This study compares 28 tunnel $PM_{2.5}$ samples with exhaust collected from one gasoline passenger car at idle, 20, and 40 km h^{-1} using UHPLC–ESI–HRMS. The formula-level contrast is potentially useful: CHO and CHON signals overlap substantially, whereas sulfur-containing signals overlap much less. The manuscript also makes welcome efforts to state that the car is a single-vehicle benchmark, to label tire-wear assignments as candidates, and to provide clustering sensitivity analyses.

My principal concern is the step from a detected molecular formula to a source or process. Identical formulas can represent different compounds, ESI peak areas are not particle mass fractions, and the tunnel and dynamometer samples differ in vehicle mix, collection, and probably matrix and detection opportunity. These limitations do not erase the measured contrast, but they constrain what its coverage ratios and clusters can demonstrate. The paper needs a clearer analytical audit, a reproducible statistical workflow, and narrower source-specific conclusions before it is suitable for publication.

Response: We appreciate the reviewer's careful and detailed comments, which have improved the quality of the manuscript. We have revised the manuscript to distinguish the measured formula-level contrasts from compound identification, source attribution and mechanisms. The replies below address the requested sampling records, analytical controls and targeted statistical checks, while retaining the original exploratory HCA partition.

Major comments

1. Define the comparison and its scope

Methods 2.1–2.2, lines 70–90; Introduction, lines 55–66. The three dynamometer filters are three operating conditions of one car, not independent vehicles or replicated runs. The tunnel filters include a mixed fleet (approximately 80% gasoline by the authors' count) and 28 collection periods. Please report the test vehicle's model/year, engine and injection system, emission standard, particulate filter and aftertreatment, odometer, lubricant, sampling date, duration at each speed, exhaust dilution ratio, collection volume and filter load. Clarify the tunnel filter size and the stated 4.3 cm^2 extraction aliquot. Provide the fleet composition and relevant ventilation or background context for the tunnel periods. These details are needed to separate a difference from this particular car and protocol from a general property of gasoline exhaust. Recast claims about 'tailpipe coverage' and traffic-

wide representativeness accordingly. If no further vehicles are available, this should remain a case study with a well-defined comparator.

Response: We have defined the comparison as a case study using a single-vehicle exhaust reference. The three CD samples represent operating conditions of the same car, not independent vehicles or replicate runs. The Introduction now reads as follows:
"This study aimed to quantify the molecular coverage of tunnel aerosol by a controlled **single-vehicle** tailpipe profile and to identify chemically distinct molecular **groups** retained in the tunnel." (Introduction, Lines 64–66)
The lubricant specification, numerical exhaust dilution ratio and exact outer dimensions of the large tunnel filters were not recorded. Filter-matched ventilation measurements and paired outside-tunnel molecular profiles were not available.
The confirmed vehicle configuration and sampling details are summarized in Table S1 (see our response to Reviewer #1, Minor Comment 2). For the tunnel comparison, 28 filters were collected on 2–8 August 2019. The fleet-composition text in Section 2.2 now reads as follows:
"**Gasoline-powered cars accounted for approximately 80% of recorded vehicles; other recorded categories included gasoline motorcycles, gasoline/CNG bi-fuel taxis, electric vehicles and electric motorcycles (Figure S1).**" (Section 2.2, Lines 105–106)
Both extracted portions had a geometric area of approximately 4.3 cm^2. The cutting procedure and particle-deposited areas are specified in our response to Reviewer #1, Major Comment 2. The recorded whole-filter CD net particulate masses are reported in Table S1; the injection-loading estimates are given in Table R1 in our response to Reviewer #1, Major Comment 2.
Tailpipe coverage is qualified as formula-level signal coverage by the tested reference, as detailed in our response to Reviewer #3, Major Comment 2. The vehicle and operating-condition scope is specified in the common Section 4 passage quoted in our response to Reviewer #1, Major Comment 1.

2. Reframe the tailpipe coverage ratio
Section 3.3 and Fig. 4, lines 260–285. TCC means that a formula occurred in at least one tunnel filter and at least one of the three dynamometer filters. TCR is the pooled tunnel ESI peak area of these formulas divided by the pooled area of TCC plus tunnel-only formulas, within a mode and elemental class. It is therefore a signal-overlap statistic, not a tailpipe-derived fraction of $PM_{2.5}$, organic carbon, or even an identified-compound

abundance. A CHO or CHON formula shared across samples need not be the same isomer or come from the same source. Likewise, a low sulfur TCR (0.10–0.15 combined) does not establish that the missing sulfur came from tire wear or in-tunnel chemistry; other fleet exhaust, fuel and lubricants, background aerosol, and resuspension remain possible. Please rename or explicitly qualify the metric in the abstract, figures, and conclusions. Show the distribution of sample-wise TCRs with uncertainty that respects day-level grouping, together with sensitivity to detection frequency, signal threshold, and each of the three CD conditions.

Related sentences from the original submission "TCR therefore quantifies the fraction of the tunnel associated HRMS peak area assigned to formulas also detected in the controlled single vehicle tailpipe profile." (Section 3.3)

"(b) Tailpipe coverage ratio (TCR), calculated from tunnel peak areas as TCC/(TCC + T–O)." (Figure 4 caption)

Response: We agree that TCR quantifies formula-level signal coverage by the tested reference. It uses summed tunnel peak areas, not formula counts, and does not estimate tailpipe-derived $PM_{2.5}$ or OC mass fractions. ESI response is compound-dependent, and shared formulas do not establish identical isomers or common sources. Low coverage identifies a gap in the measured reference without establishing the origin of the unmatched formulas. We retain the abbreviation TCR because its calculation is defined in Section 3.3 and the Figure 4 caption. Section 3.3 now reads as follows:

"The tailpipe coverage ratio (TCR) was calculated separately within each ionization mode and elemental class as TCC peak area divided by the sum of TCC and T–O peak areas **within the comparison set used in Figure 4b**. For each ionization mode and elemental class, TCR was calculated from pooled blank-corrected peak areas across all 28 tunnel samples, with non-detects treated as zero. **Reference membership was assessed from neutral-formula occurrence in either ESI mode; CD peak areas did not enter the ratio. TCR describes formula-level signal coverage, not a tailpipe-derived mass or source fraction.**" (Section 3.3, Lines 298–302)

The Abstract and Section 4 state the result as coverage by the controlled tailpipe profile (see our response to Reviewer #1, Major Comment 1). The Figure 4 caption now reads as follows:

"(b) Tailpipe coverage ratio (TCR), calculated from **pooled blank-corrected** tunnel peak areas as TCC/(TCC + T–O) **within the reported comparison set, separately by ionization mode and elemental class. CD detection determines reference membership; CD peak areas do not enter the calculation.**" (Figure 4 caption, Lines 315–317)

The filter-wise distributions and pooled estimates are shown in Figure R1. Uncertainty was evaluated using 10,000 resamples of the seven sampling days, retaining all four filters from each selected day and holding the original formula sets and CD reference fixed. The combined CHOS+CHONS TCRs remain 0.10 in ESI− and 0.15 in ESI+, with conditional 95% intervals of 0.049–0.241 and 0.142–0.161, respectively. These intervals describe variation within the sampled campaign, not between vehicles.

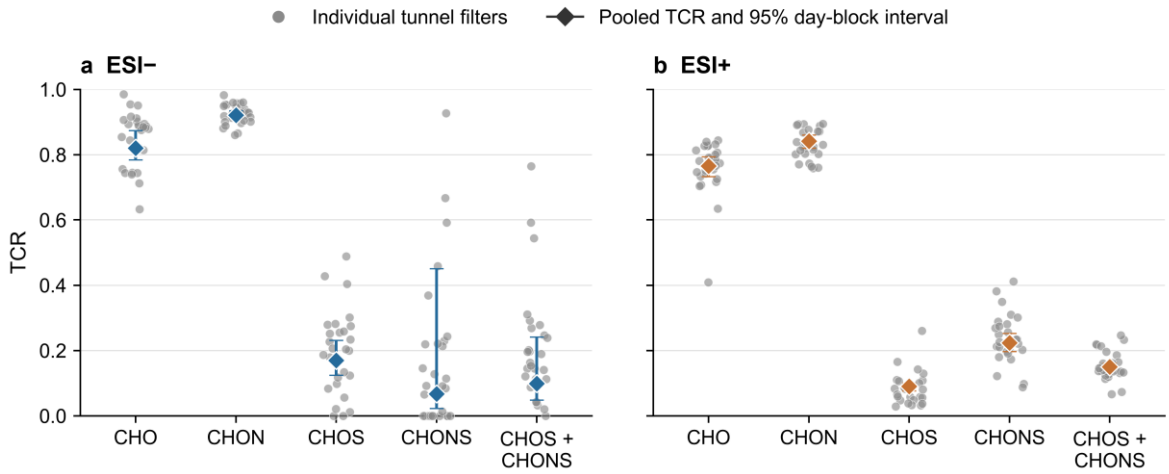


Figure R1. Individual-filter ratios (gray points) and original pooled TCRs (diamonds) within the original TCC + T–O comparison set. Bars show conditional 95% intervals from 10,000 resamples of seven whole-day blocks (four filters per day), with formula sets and the CD reference fixed. Reference membership uses detections in either ESI mode; areas remain mode-specific. All groups have 28 defined ratios except ESI− CHOS (n = 27; one zero denominator).

Table R2 summarizes the combined CHOS+CHONS results for tunnel detection frequency, operational signal thresholds and individual CD conditions. Results for the individual elemental classes and filter-wise TCRs are provided in the accompanying numerical tables.

Table R2. Combined CHOS+CHONS TCR sensitivity within the original comparison set.

Check	ESI− TCR	ESI+ TCR	Retained tunnel signal ESI− / ESI+ (%)
Original three-condition reference	0.10	0.15	100 / 100
Idle reference only	0.05	0.14	100 / 100
20 km h^{-1} reference only	0.02	0.06	100 / 100
40 km h^{-1} reference only	0.10	0.13	100 / 100
Tunnel detection in ≥10/28 filters	0.23	0.15	35.9 / 83.3
Tunnel detection in ≥25/28 filters	0.00	0.13	24.1 / 65.1
Joint processed-area cutoff ≥10^5	0.10	0.15	100 / 100
Joint processed-area cutoff ≥10^6	0.05	0.15	91.4 / 92.6

Individual-CD comparisons use the original tunnel denominator; frequency and joint-threshold comparisons use the retained subset signal. Percentages refer to the original combined sulfur signal in the same comparison set. Cutoffs are operational checks, not measured detection limits. Full-precision values accompany this response.

3. Audit the apparent tunnel-only richness

Methods 2.1–2.4 and Fig. 4a. There are 28 tunnel filters but only three CD filters, with different samplers and likely different particulate loads. The tunnel-only counts (1,108 ESI− and 2,189 ESI+) are sensitive to sampling effort, detection limits, extraction recovery, and matrix effects. Three tunnel field blanks are described, but CD field blanks, procedural blanks, analytical replicates, pooled quality-control injections, recovery checks, and run-order or batch-drift controls are not reported. Please document the blank-subtraction rule and non-detect definition for both sample sets, the time and conditions of storage of the 2019 tunnel filters, and the MZmine thresholds and adduct/isotope handling. Compare richness after matched sampling effort and a common signal threshold (for example, accumulation or rarefaction curves and a load-matched subset). Show whether the main sulfur and coverage contrasts persist after this audit. A neutral formula count must not be presented as a count of identified molecules.

Response: We agree that the cumulative tunnel-only counts depend on sampling effort, so we have documented the analytical controls for both sample sets and compared richness at equal filter number (see our response to Reviewer #1, Major Comment 2). The extracted filter blanks served as procedural blanks, namely three tunnel field blanks and two CD system-background blanks, each correcting its own sample set, whereas the injected solvent blanks were not subtracted. Previous tests of this extraction procedure in our laboratory gave 75%–

105% recoveries for four nitrogen-containing standards ($C_6H_5NO_3$, $C_6H_5NO_5$, $C_9H_9NO_4$ and $C_8H_6N_2O_6$), and re-extraction signals were below three times the corresponding blanks. The present study focuses on molecular diversity across thousands of detected formulas rather than absolute compound concentrations; formula-specific recovery and matrix effects were not quantified.

Three tunnel extracts (Fri-AF, Sun-AF and Wed-AF), about one in ten of the 31 samples, were reanalyzed in both ionization modes, and their raw peak areas correlated with r = 0.925–0.998 (mean 0.980 across six comparisons, Figure R2).

No pooled-QC sequence or quantitative batch-drift assessment is documented in the available records. The repeat analyses characterize three tunnel extracts, and the three solvent-blank injections separated the tunnel and CD batches; neither quantifies inter-batch drift.

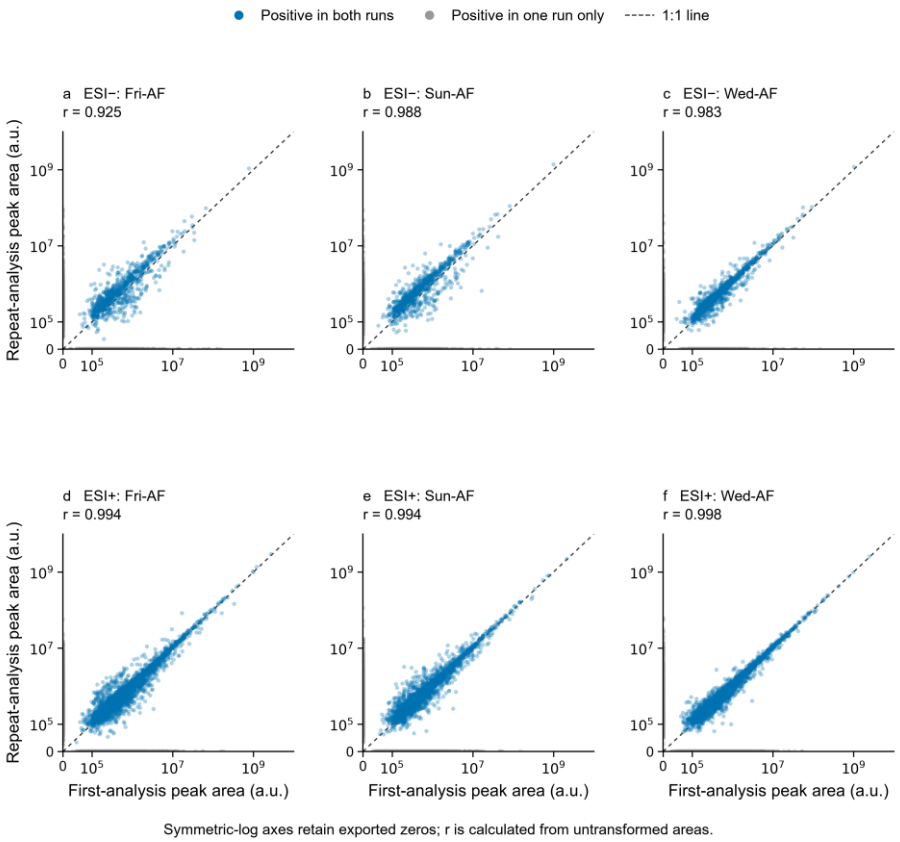


Figure R2. Repeat-analysis peak-area profiles. Three tunnel samples (Fri-AF, Sun-AF and Wed-AF) were reanalyzed in ESI− (a–c) and ESI+ (d–f). Points are aligned feature peak areas before blank correction, formula filtering or normalization. Blue points have positive areas in both runs, and gray points have a positive area in one run only. Features with two zero areas are omitted. Symmetric-log axes preserve exported zeros, and the dashed line is 1:1. Pearson r was calculated on untransformed areas using all plotted points. The six r values have an arithmetic mean of 0.980 and range 0.925–0.998, and they describe six mode-specific comparisons of three samples. The paired numerical data are provided in the accompanying repeat-injection peak-area table (Repeat_injection_peak_areas.csv).

Section 2.2 now states the storage conditions and duration of the 2019 filters (about 23 months), and Section 2.3 now states the run order and the repeat analyses together with the blank-subtraction rule and the non-detect definition, which apply identically to both sample sets. The revised text now reads as follows:

"**The high-volume filters were wrapped in foil to protect them from light and stored at −20 °C until UHPLC–HRMS analysis in July 2021, about 23 months after sampling.**" (Section 2.2, Lines 101–103)

"… Detailed chromatographic and ion source parameters followed Li et al. (2024a). **Tunnel samples were analyzed first, followed by three solvent-blank injections and then the CD samples. Three tunnel sample extracts (Fri-AF, Sun-AF and Wed-AF) were reanalyzed in both ionization modes, and the raw peak areas of the repeat analyses were highly correlated (Pearson r = 0.925–0.998, mean 0.980 across six comparisons).** Three field blanks and two CD system-background blanks were extracted and analyzed using the same procedure as the samples. Peak areas were corrected **using the corresponding tunnel** field **blanks or CD system-background blanks** before all subsequent analyses. **For each formula and ionization mode, b was the mean area in the blanks of the corresponding sample set. After formula-level aggregation, an uncorrected area A was retained when A ≥ 50,000 and A ≥ 5b, and its corrected area was A − b. Formulas failing these criteria were treated as non-detects with zero area in that sample.**" (Section 2.3, Lines 116–125)

Table S2 lists the recorded deconvolution, isotope-grouping and alignment settings, whereas the initial mass-detection and chromatogram-builder thresholds could not be recovered from the retained records. Isotopic peaks were grouped into single features, and areas assigned to the same neutral formula were summed across features and adduct forms and counted once per mode, although a complete adduct-search list was not retained. Section 2.4 now reads as follows:

"Raw UHPLC–HRMS data were processed using MZmine (v2.53**; recorded parameters in Table S2**) following the peak picking, chromatogram deconvolution, and retention time alignment workflow described by Zhang et al. (2025) and Lin et al. (2022)."

(Section 2.4, Lines 127–128)

Table S2. Recorded MZmine processing parameters.

Step	Recorded parameter	ESI−	ESI+
ADAP deconvolution	S/N threshold	10	10
ADAP deconvolution	Minimum feature height	30,000	60,000
ADAP deconvolution	Coefficient/area threshold	10	10
Isotope grouping	m/z tolerance	0.001 m/z or 2 ppm	0.001 m/z or 3 ppm
Isotope grouping	RT tolerance / maximum charge	0.3 min / 2	0.3 min / 2
Isotope grouping	Representative isotope	Most intense	Most intense
Join alignment	m/z tolerance	0.001 m/z or 2 ppm	0.001 m/z or 3 ppm
Join alignment	RT tolerance / m/z–RT weights	0.3 min / 75–25	0.3 min / 75–25

"Most intense" is the representative-isotope option used during isotope grouping.

The cumulative T–O and CD–O counts are unchanged. The Figure 4a caption now reads as follows:

"(a) Formula number and number fraction by elemental class**, based on cumulative detections across 28 tunnel filters and three CD operating-condition samples. Counts are neutral molecular formulas.**" (Figure 4 caption, Lines 313–314)

The additional common-threshold checks likewise gave lower mean tunnel union counts for all formulas (Table R3). Sulfur-containing counts were also lower in the tunnel, except in ESI− at the 10^6 cutoff, so the sulfur contrast concerns composition and signal coverage and not richness.

Table R3**. Formula counts at equal filter number under common signal cutoffs.**

Common processed-area cutoff	ESI− all formulas T / CD	ESI−CHOS+ CHONS T / CD	ESI+ all formulas T / CD	ESI+CHOS+ CHONS T / CD
Original retention criteria	612.2 / 978	236.7 / 313	1,596.6 / 2,948	306.5 / 742
$\geq 10^5$	609.2 / 967	235.1 / 308	1,596.3 / 2,948	306.5 / 742
$\geq 10^6$	209.4 / 323	88.9 / 86	1,134.2 / 2,263	183.2 / 524

T is the mean union count across all 3,276 combinations of three tunnel filters; CD is the three-condition union. The combinations overlap and are not independent replicates. Original retention required A ≥ 50,000 and A ≥ 5b, where A is the aggregated formula area before blank correction and b is the corresponding blank mean; retained areas were A − b. No additional common-area cutoff was applied in the first row.

In Table R1, all 28 tunnel injections exceeded the CD loading range, so a load-matched subset could not be formed and coverage was assessed under common cutoffs instead. Low sulfur coverage persisted, with combined CHOS+CHONS TCRs of 0.10 in ESI− and 0.15 in ESI+ in the original comparison set and 0.05 and 0.15 at the joint 10^6 cutoff, which retained 91.4% and 92.6% of the combined sulfur signal (Table R2).

4. Limit tire-wear attribution to what the measurements identify

Section 3.3–3.4, Fig. 5, and Supplement Text S1 and Table S7. The 62 'Primary TWP' and 24 'Secondary TWP' candidates are selected from elemental constraints, correlations with reference-formula time series, and CH_2 Kendrick coherence. Neither a reference formula nor a homologous series establishes a compound identity or a tire source; common traffic timing can generate the correlations. C3 contains 10 candidate Primary TWP assignments among 50, while $C_4H_{11}N$ and $C_4H_{10}N_2$ alone account for 41.29% and 21.69% of its ESI+ peak area in Table S7. These low-mass N formulas are not source-specific. The reported one-feature removal only tests $C_4H_{11}N$. Several high-ranked 'Primary TWP' candidates in C3 are CHO formulas, which underscores the non-specificity of the label. Please provide chromatographic retention and MS/MS or authentic-standard evidence for a small set of influential markers, ideally alongside tire, lubricant, road dust, and exhaust reference

materials. If these are unavailable, label the categories as formula associations throughout, remove any implied primary/secondary transformation assignment, and avoid treating C3 as evidence of a tire-wear source. Repeat the clustering and peak-area summaries after excluding both dominant low-mass N assignments.

Response: The original 62 and 24 formula sets were renamed Association groups I and II, and their reported memberships were retained. The terminology is revised in the Abstract, Sections 3.3–3.4 and 4, Text S1, Figures 5–6 and Table S14. The classification in Section 3.3 now reads as follows:

"This procedure yielded 62 **formulas in Association group I** and 24 **in Association group II**." (Section 3.3, Lines 334–335)

The category definition in Text S1 and the Figure 6 caption now reads as follows:

"**Association groups I and II denote formula-level associations with their respective reference sets.**" (Text S1 / Figure 6 caption, Lines 392–393)

The Figure 5 caption now reads as follows:

"Figure 5. CH_2-based Kendrick mass defect map for formulas associated with tire wear **reference formulas**. Gray open symbols denote formulas outside KMD series, and colored symbols denote members of CH_2 spaced formula series. Inset bars show elemental class number fractions for **Association groups I and II**." (Figure 5 caption, Lines 344–346)

The influential C3 formulas $C_4H_{11}N$, $C_4H_{10}N_2$, $C_{17}H_{22}O$ and $C_{16}H_{20}O$ have not been structurally confirmed by authentic-standard retention-time matching or diagnostic MS/MS evidence. The formula-association interpretation requested by the reviewer is therefore adopted. The simultaneous-exclusion check separates the influence of the two high-response CHN formulas from the grouping of the remaining C3 members (Table S13), and most remaining members still clustered together.

Available RT-specific MS/MS evidence for four tentative annotations is reported in Table S16. These features occur in C1 or C2, or outside the HCA input, and do not identify C3 as a tire-wear source.

The revised C3 description in Section 3.4 now reads as follows:

"**C3 contained 50 formulas**, including **34 T–O formulas (68.0%), six TCC formulas (12.0%)**, and **10 formulas in Association group I** (20.0%). **The associated formulas contributed 12.95% of the pooled normalized C3 peak area in ESI+.** … Within this mode, $C_4H_{11}N$ **and $C_4H_{10}N_2$ together** accounted for **62.98%** of the summed C3 peak **area (Table S14). $C_4H_{11}N$** is consistent with C_4 alkyl amines, including diethylamine reported in air influenced by traffic (Brean et al., 2025). After **both formulas were** excluded and HCA

was repeated, **47 of the 48** remaining C3 formulas remained **together (Table S13).** … C3 **is** a **small group** characterized by a pronounced **low-molecular-mass** CHN response together with **a subset of** formulas associated with **tire wear reference formulas.**"
(Section 3.4, Lines 403–404, 407–410, 412–414)

Table S13**. Simultaneous exclusion of** $C_4H_{11}N$ **and** $C_4H_{10}N_2$ from the HCA input.

Analysis	Total n	C1 / C2 / C3	C3 associated formulas (%)	C3 ESI+ associated area (%)	C3 ESI+ CHN area (%)
Principal HCA	709	585 / 74 / 50	10 (20.0%)	12.95	63.50
Both CHN removed, fixed membership	707	585 / 74 / 48	10 (20.8%)	34.99	1.40
Both CHN removed, HCA repeated	707	555 / 105 / 47	10 (21.3%)	35.65	1.42

Area fractions use summed normalized areas within the indicated group in ESI+. Fixed-membership and reclustered summaries use their respective remaining-member denominators. Higher associated fractions after exclusion reflect the reduced denominator. The principal 709-formula partition is the reference; the fixed-membership and repeated-HCA results are sensitivity checks. In the primary partition, the CHN column includes all three CHN formulas (63.50%); the two excluded formulas together contribute 62.98%.

The corresponding Abstract description now reads as follows:
"… two smaller **groups** concentrated in tunnel samples, characterized respectively by **sulfur-containing** formulas and by formulas associated with tire wear **reference formulas** together with a low-molecular-mass CHN response in **positive-ion mode**." (Abstract, Lines 24–26)
The corresponding Section 4 target statement is quoted in our response to Reviewer #1, Major Comment 1.

5. Resolve the sulfur chemistry before inferring mechanism
Sections 3.2–3.4, especially lines 199–205 and 344–355. CHOS and CHONS formulas do not by themselves establish organosulfate functional groups, sulfur-bearing lubricants, a tire source, or secondary formation. The low overlap and the C2 enrichment support a descriptive statement about sulfur-containing formulas in these extracts. Please report which specific high-signal C2 features drive this result and examine their chromatograms, mass errors, isotope patterns, and diagnostic MS/MS fragments where available. Relate them to measured inorganic sulfate, EC/OC, and tunnel operating conditions with uncertainty and appropriate day-level treatment. An independent source test, such as matched lubricant/tire/road-dust spectra or sulfur isotope composition of suitable separated fractions, would strengthen attribution, but the present conclusions should not

depend on an unmeasured pathway. Distinguish oxidation, partitioning, and true secondary synthesis rather than grouping all differences as 'near-road processing'.

Response: We have replaced the earlier general statement with a mode-specific description. Section 3.4 now reads as follows:

"**C2 contained 39 sulfur-containing formulas (52.7%), compared with 25.98% in C1 and 28.35% across all 709 selected formulas. In ESI−, $C_9H_{19}NOS$ and C_7H_5NOS were the only sulfur-containing formulas and contributed 33.28% and 15.33% of the C2 normalized peak area in ESI−, respectively (48.61% combined). … In ESI+, 37 sulfur-containing formulas contributed 62.35% of the C2 normalized peak area in ESI+, and the largest single contribution was 6.30%.**" (Section 3.4, Lines 373–375, 377–378)

The two ESI− formulas and the four largest ESI+ sulfur-containing contributors were examined; their recorded retention intervals, mass errors, isotope patterns and available MS/MS evidence are summarized in Tables R4 and R5. Mass errors (−1.52 to 2.67 ppm) were within the mode-specific tolerances. Isotope patterns of $C_9H_{19}NOS$, C_7H_5NOS and $C_{11}H_{12}N_2OS$ partially agreed with the modelled values; the other three envelopes remained unresolved.

Table R4. Selected C2 sulfur features.

Mode / formula label	Area (%)	Ion convention / observed m/z	RT interval (min)	Mass error (ppm)
ESI− $C_9H_{19}NOS$	33.28	$[M-H]^-$ 188.111376	8.474–8.819	−1.00 to 0.14
ESI− C_7H_5NOS	15.33	$[M-H]^-$ 150.001744	8.545–8.855	−1.52 to −0.15
ESI+ $C_{11}H_{12}N_2OS$	6.30	$[M+H]^+$ 221.074244	9.537–9.730	−0.96 to −0.18
ESI+ $C_{10}H_{19}NO_7S$	3.23	† 344.075842	17.519–17.666	1.04 to 2.67
ESI+ $C_7H_4N_2O_6S$	3.02	$[M+H]^+$ 244.986232	13.741–13.948	−1.07 to 0.70
ESI+ $C_{12}H_{17}NO_7S$	2.97	$[M+H]^+$ 320.080204	16.579–16.795	0.87 to 1.77

Area % is each formula's share of pooled normalized C2 area within its mode; other columns refer to selected chromatographic features.

†$C_{10}H_{19}NO_7S$ derives from source annotation $C_{10}H_{19}NNa_2O_7S$; its ppm error uses the recorded plus-H convention, without independent adduct/charge confirmation.

The saved MS1 traces for the six selected features are shown in Figure R3.

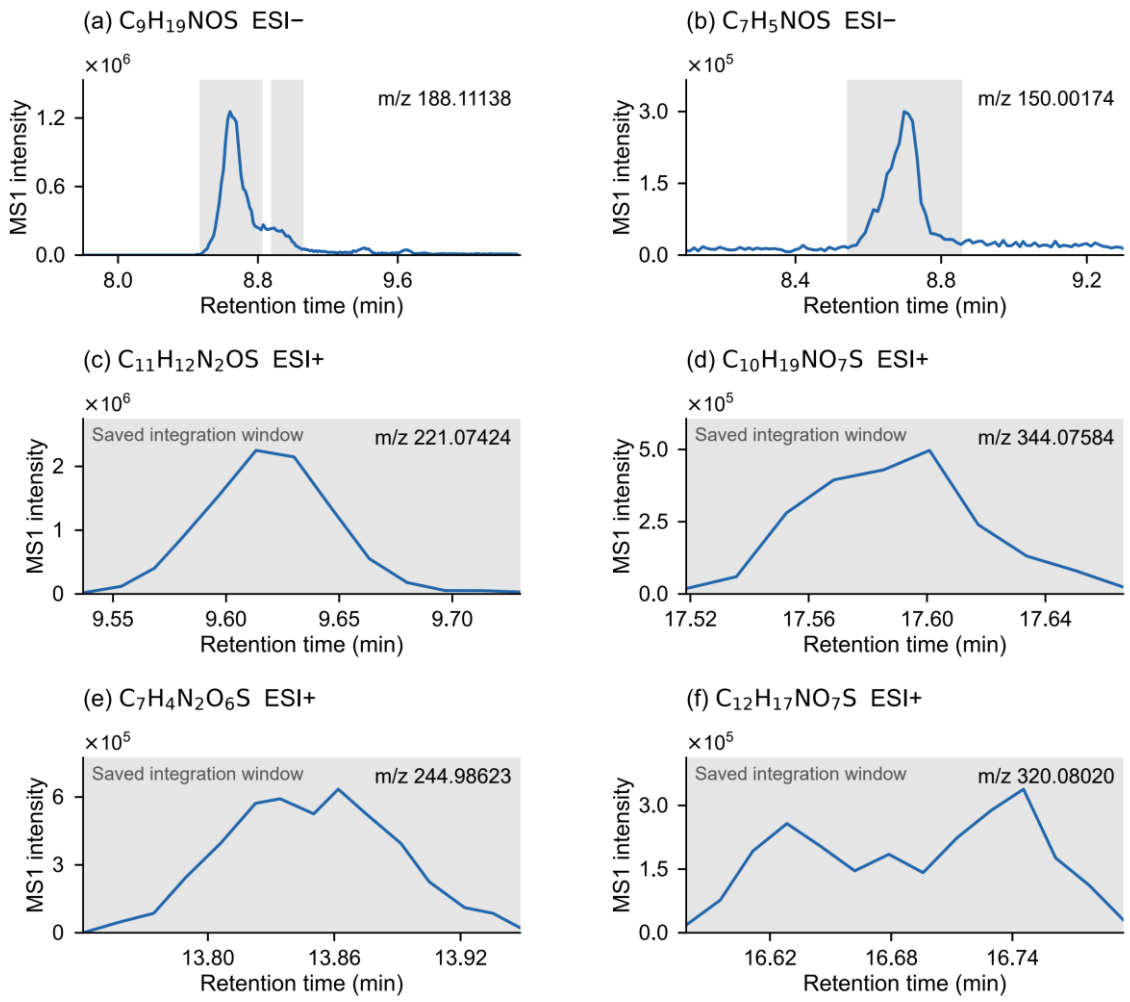


Figure R3. Saved MS1 traces for the six selected C2 sulfur-containing features. Panels (a–f) correspond to $C_9H_{19}NOS$ (ESI−, Sat-M), C_7H_5NOS (ESI−, Sat-N), $C_{11}H_{12}N_2OS$ (ESI+, Wed-T), $C_{10}H_{19}NO_7S$ (ESI+, Wed-T), $C_7H_4N_2O_6S$ (ESI+, Wed-T), and $C_{12}H_{17}NO_7S$ (ESI+, Thu-M), respectively. Shading denotes the recorded integration windows. The two ESI− traces include expanded retention-time windows; the four ESI+ traces cover only saved integration windows and do not establish baseline separation. Intensities are unscaled saved MS1 signals. The separate 24MoBT MS/MS spectrum in Figure S8 was obtained from Wed-M, rather than the Wed-T injection in panel (c). These representative records do not establish unique structures or source identity.

Table R5. Isotope and MS/MS evidence for the same six features.

Mode / formula label	Isotope pattern (observed / modelled)	MS/MS interpretation
ESI− $C_9H_{19}NOS$	M+1 10.33 / 11.13%; M+2 4.56 / 5.24%; ^{34}S 4.17 / 4.47%. Partial agreement.	Limited compatible products; no supported named structure.
ESI− C_7H_5NOS	^{13}C 7.51 / 7.57%; ^{34}S 4.67 / 4.47%. Nominal M+1 is background-limited.	No supported named structure from the available spectra.
ESI+ $C_{11}H_{12}N_2OS$	M+1 13.61 / 13.61%; M+2 4.43 / 5.54%. M+2 is below the model.	m/z 177.0482, 150.0373, 136.0217 and 109.0108; library-compatible pattern, tentative 24MoBT.
ESI+ $C_{10}H_{19}NO_7S$	M+1 7.09 / 12.47%; M+2 45.90 / 6.63%. Excess M+2; envelope unresolved.	No source-linked MS/MS available for interpretation.
ESI+ $C_7H_4N_2O_6S$	M+1 8.82 / 9.38%; M+2 28.12 / 6.10%. Excess M+2; envelope unresolved.	Available MS/MS does not establish a named structure or sulfate-ester functionality.
ESI+ $C_{12}H_{17}NO_7S$	M+1 6.36 / 14.61%. A coherent isotope envelope could not be reliably evaluated.	No source-linked MS/MS available for interpretation.

Isotope entries are observed/modelled percentages and do not establish a unique structure. Product-ion m/z values for tentative 24MoBT are median observed masses across the ten eligible spectra.

The $C_{11}H_{12}N_2OS$ feature had fragments compatible with a library spectrum of 2-(4-morpholinyl)benzothiazole (24MoBT), supporting a tentative annotation (Figure S8). No named structure was established for the other five features (Table R5). The Association group II link between C_7H_5NOS and 2-hydroxybenzothiazole is based on formula identity. The chemical interpretation in Section 3.4 now reads as follows:

"**The two ESI− sulfur formulas contain only one oxygen atom each and cannot represent intact sulfate esters or sulfonates. C_7H_5NOS is linked to 2-hydroxybenzothiazole at the formula level.**" (Section 3.4, Lines 375–377)

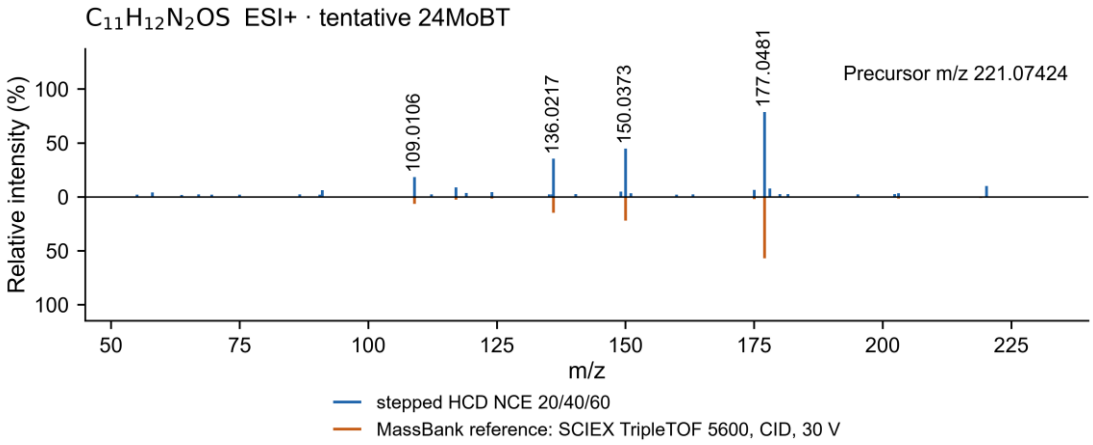


Figure S8. Representative MS/MS comparison for tentative 24MoBT. Representative ESI+ MS/MS spectrum of the $C_{11}H_{12}N_2OS$ feature (blue; stepped HCD NCE 20/40/60) and the MassBank 24MoBT reference (orange; SCIEX TripleTOF 5600, CID, 30 V). The comparison supports a tentative annotation without authentic-standard confirmation.

The requested Spearman comparisons with inorganic sulfate and EC/OC are provided in Table R6. All 16 day-block 95% intervals included zero; these comparisons did not resolve a specific relationship with the bulk measurements or establish a source or formation mechanism.

Table R6**. Eight C2 sulfur signals versus sulfate and EC/OC.**

Mode / feature or sum	Sulfate, ρ [95% interval]	EC/OC, ρ [95% interval]
ESI− $C_9H_{19}NOS$	−0.11 [−0.60, 0.41]	0.12 [−0.32, 0.48]
ESI− C_7H_5NOS	−0.14 [−0.60, 0.33]	−0.05 [−0.55, 0.34]
ESI− C2 sulfur sum	−0.18 [−0.65, 0.41]	0.04 [−0.38, 0.45]
ESI+ $C_{11}H_{12}N_2OS$	0.10 [−0.46, 0.58]	0.18 [−0.38, 0.52]
ESI+ $C_{10}H_{19}NO_7S$	−0.22 [−0.58, 0.11]	0.06 [−0.35, 0.44]
ESI+ $C_7H_4N_2O_6S$	−0.13 [−0.41, 0.15]	0.14 [−0.26, 0.50]
ESI+ $C_{12}H_{17}NO_7S$	0.00 [−0.66, 0.49]	0.33 [−0.10, 0.66]
ESI+ C2 sulfur sum	−0.15 [−0.77, 0.40]	0.24 [−0.19, 0.62]

Signals are blank-corrected, unnormalized peak areas; each sum includes all C2 sulfur formulas in its mode. Spearman ρ uses 28 filters; 95% percentile intervals use 10,000 paired bootstrap resamples of the seven complete sampling days, retaining all four within-day filters together. Sulfate was measured in water extracts by Dionex-600 ion chromatography; OC and EC were measured with a DRI 2001 analyser using the IMPROVE_A thermal–optical reflectance protocol.

Filter-matched operational records were unavailable. The source hypotheses are addressed in our response to Reviewer #2, Comment 3. Gas–particle partitioning redistributes compounds

between phases; oxidation alters chemical composition; secondary formation produces compounds from precursors after emission. The reference-development sentence in the Abstract now reads as follows:

"**Molecular reference profiles for traffic PM2.5 should therefore combine controlled tailpipe measurements, tunnel or on-road profiles, and reference materials from lubricants, tire wear and road dust, while oxidation, gas–particle partitioning and secondary formation need to be assessed separately.**" (Abstract, Lines 27–30)

6. Correct and stabilize the statistical workflow

Methods 2.5; Supplement Texts S2–S3, Tables S4, S5, and S9. Welch tests compare 28 tunnel samples with only three non-independent operating conditions from one vehicle. Within-mode relative areas are compositional; a change in one abundant feature changes other relative areas without a change in their absolute loading. Zero replacement and the fixed 10^{-9} pseudocount can drive large fold changes at low signal. More immediately, Text S2 says fold changes use arithmetic group means, while Eq. S2-1 displays medians and even repeats 'log2 FC' on the right-hand side; Fig. S5 says all 5,666 assignments were tested, while the text says testing was restricted to assignments detected in at least 10 tunnel samples. Please reconcile these statements, specify the family used for each FDR adjustment, and publish the feature-by-sample matrix and executable analysis. Recompute the contrast under a consistent fold-change definition, plausible pseudocounts, detection thresholds, and alternative normalizations. Treat the CD condition comparison descriptively unless its uncertainty can be estimated from independent runs or vehicles.

Related sentences from the original submission "Statistical testing was restricted to formula and ionization mode assignments detected in at least 10 of the 28 tunnel samples." (Text S2)

Response: The three CD samples represent operating conditions of one vehicle, not independent vehicles or replicate runs. Their comparison is therefore interpreted descriptively. Nominal p values were used for HCA input selection; BH q values were supplementary diagnostics, not a selection criterion. Neither is used to establish population-level tunnel–exhaust differences. The single-vehicle limitation is stated in Section 4, which now reads as follows:

"**The present single-vehicle reference, obtained under three steady-state conditions, does not characterize between-vehicle emission variability or transient emissions during acceleration, deceleration and stop-and-go operation.**" (Section 4, Lines 439–441)

The median notation and repeated $\log_2$FC term in Eq. S2-1 have been corrected. Text S2 now reads as follows:

"**For** formula **f, Y denotes signal on the stated normalization scale. The descriptive contrast is the log$_2$ ratio of arithmetic means when both means are positive. The** tunnel **mean includes all 28 filters,** and the **chassis dynamometer (CD) mean weights the three measured conditions equally. Non-detections contribute zero. No positive pseudocount is added** to **these shared-signal** means **(Eq. S2-1).**" (Text S2)

$$\log_2 FC_f = \log_2\left(\frac{\overline{Y_{T,f}}}{\overline{Y_{CD,f}}}\right) \qquad \text{(S2-1)}$$

The tests preceded the ≥10-filter screen. For HCA input selection, $\varepsilon = 10^{-9}$ was added to both group means, whereas the descriptive shared-signal ratios in Eq. S2-1 use no positive pseudocount. The revised Text S2 uses "formulas" consistently within each ionization mode. The FDR families are specified in Text S2 and summarized in Table S9. The Text S2 passage now reads as follows:

"**Two-sided Welch tests for HCA input selection used log$_{10}$(RA + 10^{-9}) (Welch, 1947) and preceded** the tunnel **detection-frequency filter. …** The **comparison datasets used for HCA input selection contained 5,666 formulas across the two modes. Benjamini–Hochberg adjustment** used **all p** values separately within each **mode-specific HCA selection dataset, without elemental-class stratification (5,137 ESI+ formulas; 529 ESI− formulas)** (Benjamini and Hochberg, **1995).**" (Text S2)

The 529-formula ESI− export comprises formulas shared by the two mode-specific catalogs across all 31 samples. Its q values apply to that saved export, whereas the descriptive sensitivity assessments use the complete 2,086-formula ESI− matrix. The revised Figure S5 separates shared-signal contrasts from one-sided detections. Its caption now reads as follows:

"**(a, b) Formulas detected in both datasets are shown by log$_2$ ratios of arithmetic mean** tunnel **and CD relative areas,** with the **three CD conditions weighted equally; dashed** lines mark **±2. (c, d) Formulas without a retained signal** in **any corresponding-mode CD sample** are shown **by tunnel detection frequency** and **mean tunnel relative area, without finite pseudocount-derived ratios.**" (Figure S5 caption)

Table R7 applies the same arithmetic-mean contrast without a p-value gate. The displayed normalization, area and detection-frequency settings retained 576–709 original members. Among shared formulas in the original subset, probabilistic quotient normalization (PQN) and median-ratio scaling each reversed the direction of 5 of 14 ESI− contrasts and none of 136 ESI+ contrasts. Thus, some individual ESI− contrast directions depend on normalization. The individual CD references retained 693–703 members; the mode-specific and individual-condition results are supplied in the accompanying files.

Table R7. Descriptive sensitivity of the original subset and complete candidate pool.

Setting	Original 709 retained	Complete-matrix candidates
Total-area baseline	709	1,118
PQN	654	1,004
Positive-value median ratio	646	1,000
Processed area ≥100,000	709	1,118
Processed area ≥250,000	692	1,067
Tunnel detections ≥7/28	709	1,357
Tunnel detections ≥14/28	576	917

Complete-matrix candidates use 5,137 ESI+ and 2,086 ESI− formulas. Baseline settings are total-area normalization, processed area ≥50,000, tunnel detections ≥10/28 and the equally weighted three-CD mean. One setting changes per row. Shared signals require $|\log_2$ mean ratio$| \geq 2$; one-sided detections use occurrence and frequency. No p-value gate is used. Area floors are not instrumental detection limits.

The six saved pseudocount diagnostics span 10^{-12}–10^{-6}. For the one-sided $C_4H_{11}N$ signal in ESI+, the finite $\log_2$ ratio changes from 31.24 to 11.30 across those endpoints. Such signals are therefore described by occurrence rather than finite fold changes. Under the baseline area and detection-frequency rules, removing $\varepsilon = 10^{-9}$ caused no shared-formula crossings of $|\log_2 FC| = 2$; with the p-value gate removed, occurrence-based inclusion introduced no further membership changes among the original 709.

The complete matrices yield 1,118 baseline candidates and 917–1,357 across the displayed settings, so the original 709 are retained as an exploratory subset. These checks concern the HCA input screen and Figure S5; TCC/T–O membership and the peak-area-based TCR do not use fold change. Fold change and nominal p contributed to input selection but not to clustering distances (Text S3); clustering sensitivity is addressed in our response to Reviewer #3, Major Comment 7.

The normalized peak-area matrix of the 709 HCA input formulas, the sample metadata and the code that reproduces the primary HCA partition are deposited in a Zenodo repository. References to volcano screening have been replaced by HCA input selection, and q values are reported as supplementary selection diagnostics in Text S2. Data Availability now reads as follows:

"Data Availability. Essential data used in this study, including the normalized peak-area matrix of the 709 formulas selected for HCA, the sample metadata and the code that reproduces the primary HCA partition, are available at https://zenodo.org/records/24141410. Detailed mass spectrometry datasets are available upon request to the corresponding author (yuemei.han@ieecas.cn)." (Data Availability, Lines 455–458)

7. Show what clustering adds beyond the selection step

Section 3.4 and Supplement Text S3, Tables S6–S11. HCA uses 709 features already selected for a strong tunnel/CD contrast, so the resulting tunnel-enriched domains are not independent confirmation of that contrast. The seven daily values are smoothed over three days, which can increase similarity among features responding to shared day effects. The reported 82–84% mapped agreement under descriptor rescaling is dominated by the very large C1 group; adjusted Rand indices fall to 0.463–0.499 (Table S9). Report C2 and C3 membership overlap separately under rescaling, removal of smoothing, and day-block resampling, along with uncertainty in the day-level profiles. State the domain's magnitude as well as its formula count: from Table S11, C1 accounts for about 98.2% of the mean normalized ESI+ signal and 95.4% of ESI− signal within the selected HCA assignments (these are neither mass nor all detected ions). This context is essential when discussing the smaller domains.

Response: HCA adds a temporal and compositional description of the selected tunnel formulas. The primary partition contains a CHO/CHON-rich C1 group and two smaller groups characterized by sulfur-containing formulas (C2) and a pronounced low-molecular-mass CHN response (C3). C2 has a higher sulfur-containing formula fraction than C1 and the full selected set, although sulfur atom number was not used in the clustering distance. Section 3.4 now reads as follows:

"**C2 contained 39 sulfur-containing formulas (52.7%), compared with 25.98% in C1 and 28.35% across all 709 selected formulas.**" (Section 3.4, Lines 373–374)

The primary $\alpha = 0.80$ specification is not the maximum-silhouette solution over the tested grid, and the higher coefficient at $\alpha = 0.90$ is reported in Table S11. Text S3 now reads as follows:

"**The primary $\alpha = 0.80$ specification assigns 80% of the distance weight to the sampling-day profiles and 20% to chemical composition. It is not the maximum-silhouette solution in the grid; at $K = 3$, $\alpha = 0.90$ gave a higher mean silhouette coefficient (0.647; Table S11).**" (Text S3)

Removing the three-point moving average changed membership substantially, giving an ARI of 0.298 and mapped agreement of 75.5%, the lowest among the six settings in Table S10. Figure R4 and Table R8 examine the profiles of the original members. With primary C2 membership held fixed, the sampled Monday–Thursday versus Friday–Sunday mean contrast was already visible before smoothing. This observation and recovery of the same membership by unsmoothed HCA are separate questions. The observed within-day ranges describe variation across four distinct sampling periods; daily-mean confidence intervals were not estimated because those periods are not exchangeable replicates.

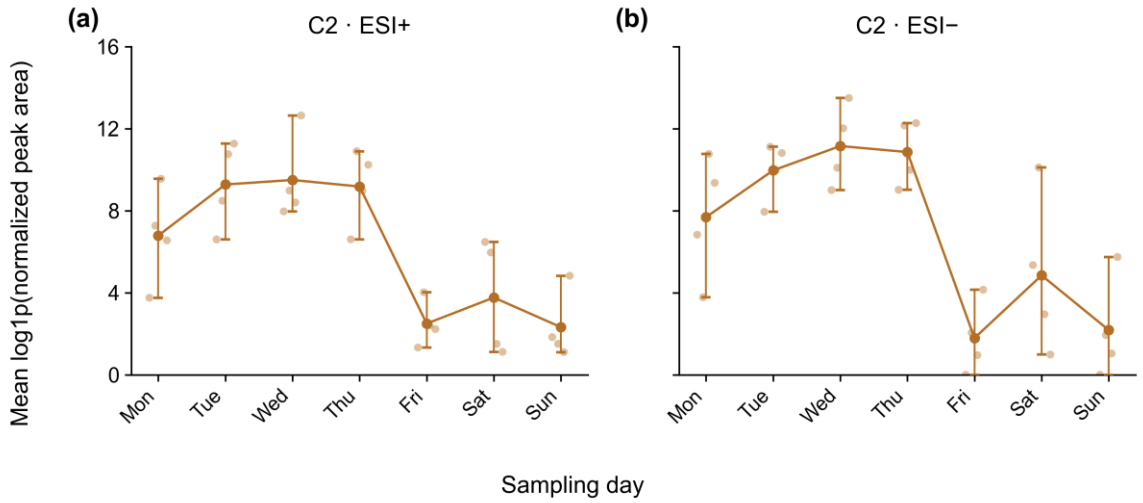


Figure R4. Unsmoothed C2 profiles in ESI+ (a) and ESI− (b). Reproduced from Figure S6b and S6e. Small points are sampling-period values, connected points are daily means, and bars show the observed within-day minimum–maximum. Primary C2 membership is fixed.

Table R8**. Unsmoothed C2 daily means and observed within-day ranges.**

Sampled day	ESI+ mean	ESI+ range	ESI− mean	ESI− range
Mon	6.78	3.76–9.56	7.68	3.78–10.77
Tue	9.28	6.60–11.27	9.96	7.94–11.12
Wed	9.50	7.97–12.64	11.15	9.00–13.49
Thu	9.18	6.60–10.90	10.86	9.02–12.27
Fri	2.50	1.33–4.02	1.79	0.00–4.15
Sat	3.77	1.12–6.48	4.85	0.99–10.11
Sun	2.32	1.10–4.83	2.18	0.00–5.74

Each daily mean averages four sampling-period values, each the mean log1p(normalized peak area) over the primary C2 members (61 in ESI+ and 13 in ESI−).

The sampled-day description in Section 3.4 now reads as follows:

"**For the primary C2 group, unsmoothed daily means of log1p(normalized peak area) ranged from 6.78 to 9.50 on the sampled Monday–Thursday and from 2.32 to 3.77 on Friday–Sunday in ESI+. In ESI−, daily means ranged from 7.68 to 11.15 and from 1.79 to 4.85, respectively (Figure S6).**" (Section 3.4, Lines 380–382)

Overall agreement is strongly influenced by C1's 82.5% share of the members. Under z-score global matching, the 585 matched formulas comprise 552 C1 members, 33 C2 members and none from C3. Group-specific silhouettes are therefore reported alongside the overall statistics. Signal shares are reported separately as a measure of magnitude. Section 3.4 now reads as follows:

"The partition had a mean silhouette coefficient of 0.622, a median silhouette coefficient of 0.729, and a negative silhouette fraction of 4.2%. **Because C1 contained 82.5% of the members, these overall silhouette summaries and the 82.5%–83.9% mapped agreement under descriptor rescaling were dominated by this group. The mean silhouette widths of C2 and C3 were 0.377 and 0.326, respectively (Table S15).**" (Section 3.4, Lines 354–357)

"**Within the selected HCA set, the C1/C2/C3 shares of pooled normalized peak area were 98.22%/0.56%/1.22% in ESI+ and 95.35%/3.91%/0.74% in ESI− (Table S6).**" (Section 3.4, Lines 357–359)

Membership overlap was higher when only signal normalization changed than under descriptor rescaling or removal of smoothing (Table S15). All 50 original C3 members remained together under z-score rescaling but within a 611-member group, so C3 was not recovered as a separate small group.

Table S10. Overall partition diagnostics under changes in preprocessing.

Setting	Mapped agreement (%)	ARI	Mean silhouette	Negative silhouette (%)
MR baseline	100.00	1.0000	0.622	4.2
Total-area (TSS)	96.47	0.8592	0.615	5.9
PQN	97.88	0.9116	0.628	4.7
Descriptor z-score	82.51	0.4634	0.246	10.7
Descriptor min–max	83.92	0.4989	0.243	19.0
No smoothing	75.46	0.2976	0.375	15.7

The 709 members, K = 3 and temporal weight = 0.80 are fixed. Mapped agreement uses globally optimal one-to-one label matching over all 709 members. Silhouette coefficients are computed in each variant's own distance space. ARI denotes adjusted Rand index; MR, median-ratio normalization; TSS, total-area normalization; PQN, probabilistic quotient normalization.

Table S15**.** Group-specific **membership and composition under changes in preprocessing.**

Setting	**C2 overlap / group n**	**C2 J**	**C2 S %**	**C3 overlap / group n**	**C3 J**
MR baseline	74/74 in 74	1.0000	52.70	50/50 in 50	1.0000
Total-area (TSS)	72/74 in 77	0.9114	51.95	50/50 in 68	0.7353
PQN	71/74 in 76	0.8987	51.32	50/50 in 57	0.8772
Descriptor z-score	32/74 in 32	0.4324	65.63	50/50 in 611	0.0818
Descriptor min–max	70/74 in 172	0.3977	37.21	38/50 in 43	0.6909
No smoothing	64/74 in 139	0.4295	41.73	29/50 in 102	0.2358

The 709 members, K = 3 and temporal weight = 0.80 are fixed. Group overlaps give captured/reference members and matched-group size, using independent maximum-Jaccard matching; J is intersection/union. S % is the sulfur-containing formula count fraction of the matched group. For z-score rescaling, the independent matching selects a 32-member group, whereas the globally optimal matching used for mapped agreement in Table S10 pairs C2 with a 66-member group. MR, median-ratio normalization; TSS, total-area normalization; PQN, probabilistic quotient normalization. The primary mean silhouette widths were 0.377 for C2 and 0.326 for C3.

The matched groups also change in composition as original C2 members are retained and other formulas enter. Across all five altered settings, the retained fraction was higher for the original sulfur-containing members than for non-sulfur members, although the differences were modest except under z-score rescaling (Table R9). Under min–max rescaling and without smoothing, incoming members contained 26/102 (25.5%) and 22/75 (29.3%) sulfur-containing formulas, respectively, close to 28.35% in the full selected set. The higher sulfur fractions of these matched groups largely reflect retained original C2 members. These are composition summaries of groups matched to C2, not an independent recovery test.

Table R9. Retention of the original C2 members by sulfur content.

Setting	Matched group n	Original S retained (n/39)	Original non-S retained (n/35)	S from other groups (n)
MR baseline	74	39/39	35/35	0
Total-area (TSS)	77	39/39	33/35	1
PQN	76	38/39	33/35	1
Descriptor z-score	32	21/39	11/35	0
Descriptor min–max	172	38/39	32/35	26
No smoothing	139	36/39	28/35	22

The matched groups are the independent maximum-Jaccard matches in Table S15. Retention counts refer only to the original C2 sulfur-containing (n = 39) or non-sulfur members (n = 35). Incoming sulfur members from other primary groups do not enter those denominators.

Day-block resampling showed variable recovery of the smaller groups, with median C2/C3 Jaccard overlaps of 0.47/0.38 when normalization and the primary smoothed profiles were fixed and 0.22/0.19 when normalization was re-estimated and unsmoothed profiles were used, which also changes the processing. C2 temporal silhouettes were negative in all six held-out-period splits (Table R10). These results limit claims of reproducible temporal separation. Text S3 now reads as follows:

"**Each of two day-block approaches used 500 resamples of seven days with replacement, jointly across modes. The first fixed full-data normalization and the primary smoothed profiles, weighting daily coordinates by their resampling frequencies. The second retained four filters per sampled day, re-estimated within-mode normalization and used unsmoothed profiles. Comparison of the latter with the primary smoothed groups combines day and processing sensitivity. Jaccard overlaps and member-retention fractions are summarized by medians with empirical 2.5–97.5% ranges; sulfur fractions are medians. Co-assignment is the eligible-resample fraction placing a pair in the same group, averaged over member pairs. Member retention is the captured fraction of computable reference members; undefined zero-profile distances were excluded. Held-out checks used six overlapping choices of two training periods and the complementary two validation periods, with seven days and 14 filters per side. Normalization references were fitted to training filters, training profiles were smoothed, and training labels were evaluated using unsmoothed held-out temporal distances and centroid correlations. Ranges across these six non-independent splits are minima and maxima.**" (Text S3)

Table R10**. Day-block resampling and held-out-period diagnostics.**

Diagnostic	C2	C3
Primary mean silhouette	0.377	0.326
Day-block, fixed profiles, Jaccard	0.4685 [0.1005–0.8210]	0.3788 [0.0784–0.8475]
Day-block, fixed profiles, member retention	0.8108 [0.1622–1.0000]	0.7800 [0.3000–1.0000]
Day-block, fixed profiles, within-group co-assignment	0.724	0.719
Day-block, fixed profiles, sulfur fraction median (%)	45.46	24.58
Day-block, refitted normalization, Jaccard	0.2197 [0.1018–0.5754]	0.1886 [0.0721–0.4174]
Day-block, refitted normalization, member retention	0.6892 [0.1622–0.9595]	0.5800 [0.1600–0.9905]
Held-out periods, temporal mean silhouette	−0.316 to −0.101	−0.175 to 0.367†
Held-out periods, centroid Pearson r	0.675–0.952	0.177–0.964

†The upper C3 held-out silhouette belongs to a 597-member group jointly matched to C1 and C3, not a separately recovered C3 group.

The primary partition is retained as a descriptive summary of the selected tunnel dataset. Because the formulas were preselected for tunnel–CD contrasts, HCA describes the grouping within this selected set and does not independently confirm that contrast. Membership was more consistent under changes in signal normalization than under descriptor rescaling or removal of smoothing, with Jaccard overlaps of 0.40–0.43 for C2 and 0.08–0.69 for C3 under the latter settings (Table S15), and day-block resampling and held-out-period checks also showed variable membership (Table R10). We therefore interpret the groups only within the selected dataset.

The two earlier concluding sentences on incomplete representation and molecular targets have been removed; the specific targets remain in Section 4. The closing paragraph of Section 3.4 now reads as follows:

"Taken together, molecular correspondence between the controlled gasoline tailpipe and tunnel profiles was dominated by the large CHO/CHON-rich C1 baseline, whereas C2 and C3 represented **groups with different compositional characteristics concentrated** in tunnel aerosol. Because C1 accounted for 82.5% of the selected formulas **and 98.22% and 95.35% of the selected normalized ESI+ and ESI− signal, respectively,** comparisons based on overall profile similarity would largely reflect this shared composition and provide limited resolution of the smaller C2 and C3 **groups.**" (Section 3.4, closing paragraph, Lines 416–420)

The corresponding wording in the Abstract and Section 4 now reads as follows:
"… Hierarchical clustering of selected formulas showed that **tunnel-only formulas** contributed 33%–41% of the signal of the dominant shared CHO/CHON baseline and formed most of two smaller **groups** concentrated in tunnel samples, characterized respectively by **sulfur-containing** formulas and by formulas associated with tire wear **reference formulas** together with a low-molecular-mass CHN response in **positive-ion mode**."
(Abstract, Lines 22–26)
"… HCA **assigned the selected formulas to three groups. C1, the largest group, was dominated by CHO and CHON formulas and contained most formulas shared with the CD profile, C2 was enriched in sulfur-containing formulas, and the ESI+ signal of C3 was dominated by two low-molecular-mass CHN formulas.**" (Section 4, Lines 425–427)

Minor comments

1. Section 3.1, lines 185–195: 'emergent' and 'attenuated' use ratios above 10^4 or below 10^{-4} with a nominal area of 1 for nondetections, but each speed has one filter. Describe these as detection/response contrasts and provide raw area and blank distributions for the influential features; avoid implying repeatable speed effects.

Response: We now describe differences between the measured CD profiles as detection/response contrasts. "Emergent" and "attenuated" are replaced by "higher-speed-only" and "idle-only", determined using the retention criteria in revised Section 2.3 (see our response to Reviewer #3, Major Comment 3), without nominal areas for non-detections. Reapplying the original ratio rule to the archived processed matrices confirmed identical category membership; Table S3 counts and descriptor summaries are unchanged. Section 3.1 now reads as follows:
"Individual formulas were evaluated **by their retention state in each operating condition (Section 2.3)** to identify formulas **that differed between each higher-speed profile and idle (20/0 and 40/0)** in the controlled tailpipe molecular profile. **Formulas retained in only one profile of a pair were labelled idle-only or higher-speed-only. These categories describe detection/response contrasts between the three measured profiles.** Relative to idle, the 20 km h^{-1} profile showed **idle-only to higher-speed-only** formula **count** ratios of 1.6 in ESI− and 1.8 in ESI+, while the 40 km h^{-1} profile showed ratios of 0.7 and 0.8, respectively (Table **S3**). CHON formed a major fraction of the **higher-speed-only** formulas. In ESI−, **higher-speed-only** CHON formulas **had** lower **mean** m/z and higher **mean** DBE/C **than**

idle-only CHON formulas in both speed contrasts. In ESI+, the 20/0 comparison was dominated by **idle-only** CHON formulas, while the 40/0 comparison contained **comparable numbers of idle-only and higher-speed-only CHON formulas**, with **higher-speed-only** formulas again **having** lower **mean** m/z and higher **mean** DBE/C. **Figure S7 gives the raw integrated areas and the two CD blank responses of 20 influential CHON formulas.** These contrasts show that the detected distribution of nitrogen-containing formulas differed among the three tested conditions and define the compositional range of the controlled tailpipe profile." (Section 3.1, Lines 212–223)

The revised Table S3 wording now reads as follows:

"Table **S3**. **Formula counts** and molecular **descriptors (m/z, OSc and DBE/C)** for **idle-only** and **higher-speed-only** formulas in each **20/0 or 40/0 comparison, separately for** ESI− and **ESI+. Descriptor values** are mean ± **SD across formulas within each category.**" (Table S3)

The Figure 2 description reports the observed contrasts between the measured profiles. The revised text now reads as follows:

"In ESI−, CHON **formulas ranged** from C5 to C15 with a maximum **formula count** near C8, and **their** number in this range was approximately 60% higher **in the measured** 40 km h^{-1} **profile** than at idle. In ESI+, CHON **formulas ranged** from C15 to C25, and the fraction below C20 **was** 15% **higher in the measured** 40 km **h^{-1} profile than at idle.**" (Section 3.1, Lines 202–205)

Figure R5 provides an overview of the uncorrected feature-area profiles across m/z in the three CD samples and the two CD system-background blanks, on common scales within each mode. Blank responses are generally lower than sample responses. The two blank records used for correction are shown individually. Formula retention follows the criteria in Section 2.3 after area aggregation.

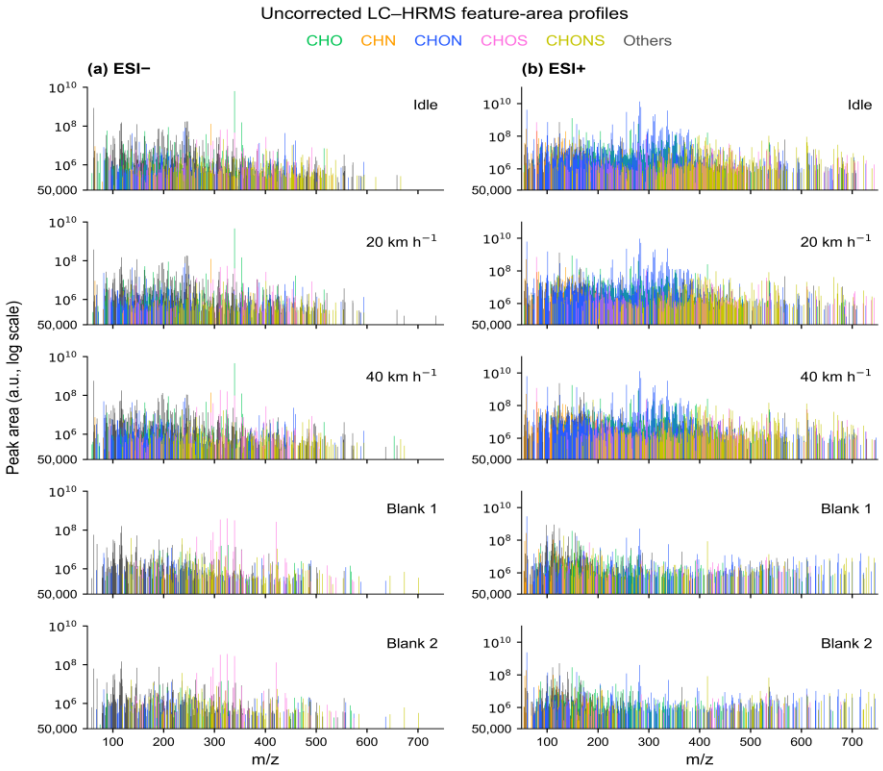


Figure R5. m/z-resolved sample and blank responses. Uncorrected, unnormalized integrated areas of aligned LC–HRMS features in the three CD samples and the two CD system-background blanks, in (a) ESI− and (b) ESI+. Each stick represents one chromatographic feature at its exported m/z, colored by its elemental-family label; Others includes unannotated features. Records share a logarithmic area scale within each mode, and only areas ≥50,000 are shown. These are integrated-area profiles, not single-scan intensity spectra.

Figure S7 provides formula-by-formula comparisons for the 20 influential CHON formulas underlying the m/z and DBE/C contrasts, and Table R11 lists their raw areas, blank responses and retention states. Sixteen formulas have no returned peak in either blank; four (C_4H_5NO, $C_8H_{10}N_2O_2$, $C_{24}H_{33}NO_6$ and $C_4H_8N_2O$) have positive blank responses. All formulas were evaluated using $A \geq 50{,}000$ and $A \geq 5b$, where b is the mean of the two CD blanks. Thus, Figure R5 supplies the overall background context, while Figure S7 provides matched sample–blank evidence for the reported detection/response contrasts.

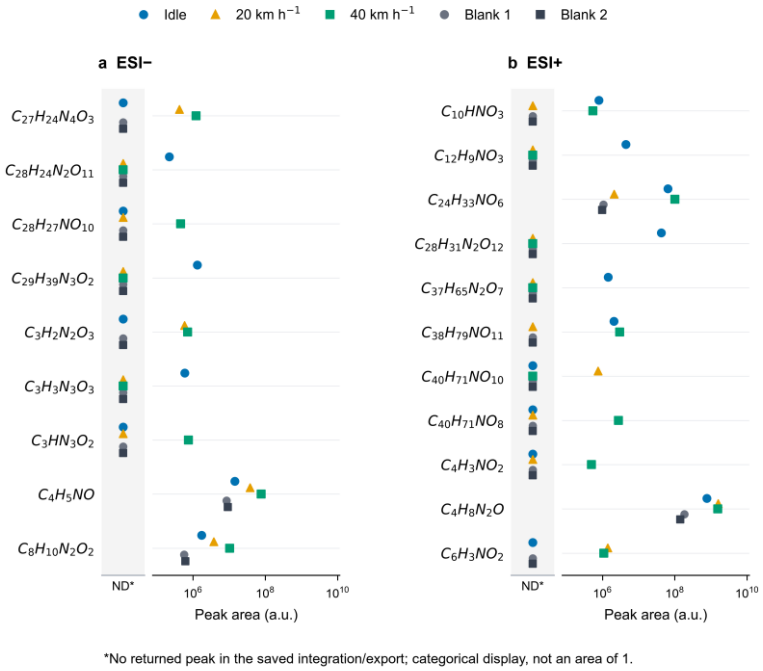


Figure S7. Raw areas and blank responses of influential CD CHON formulas. Uncorrected, unnormalized formula-aggregated areas of 20 influential CHON formulas in the three CD profiles and the two CD blank records, in (a) ESI− and (b) ESI+. Selection was based on corrected signal contribution and influence on the m/z and DBE/C summaries. Individual blank responses show the observed spread for each formula. The ND strip marks no returned peak in the saved export; it is a categorical position, not an area of 1. Positive points may still fail the retention criteria in Section 2.3.

Table R11. Raw areas, blank responses and retention states of the 20 influential CD CHON formulas.

Mode	Formula	Idle	20 km h^{-1}	40 km h^{-1}	Blank 1	Blank 2	R/NR (idle/20/40)
ESI−	$C_{27}H_{24}N_4O_3$	0	4.22E+05	1.22E+06	0	0	NR/R/R
ESI−	$C_{28}H_{24}N_2O_{11}$	2.23E+05	0	0	0	0	R/NR/NR
ESI−	$C_{28}H_{27}NO_{10}$	0	0	4.55E+05	0	0	NR/NR/R
ESI−	$C_{29}H_{39}N_3O_2$	1.32E+06	0	0	0	0	R/NR/NR
ESI−	$C_3H_2N_2O_3$	0	5.81E+05	7.13E+05	0	0	NR/R/R
ESI−	$C_3H_3N_3O_3$	5.94E+05	0	0	0	0	R/NR/NR
ESI−	$C_3HN_3O_2$	0	0	7.50E+05	0	0	NR/NR/R
ESI−	C_4H_5NO	1.45E+07	3.78E+07	7.73E+07	8.56E+06	9.25E+06	NR/NR/R
ESI−	$C_8H_{10}N_2O_2$	1.73E+06	3.76E+06	1.04E+07	5.73E+05	6.16E+05	NR/R/R
ESI+	$C_{10}HNO_3$	7.99E+05	0	5.42E+05	0	0	R/NR/R
ESI+	$C_{12}H_9NO_3$	4.43E+06	0	0	0	0	R/NR/NR
ESI+	$C_{24}H_{33}NO_6$	6.44E+07	2.10E+06	1.01E+08	1.06E+06	9.87E+05	R/NR/R
ESI+	$C_{28}H_{31}N_2O_{12}$	4.25E+07	0	0	0	0	R/NR/NR
ESI+	$C_{37}H_{65}N_2O_7$	1.44E+06	0	0	0	0	R/NR/NR
ESI+	$C_{38}H_{79}NO_{11}$	2.08E+06	0	2.99E+06	0	0	R/NR/R
ESI+	$C_{40}H_{71}NO_{10}$	0	7.53E+05	0	0	0	NR/R/NR
ESI+	$C_{40}H_{71}NO_8$	0	0	2.75E+06	0	0	NR/NR/R
ESI+	$C_4H_3NO_2$	0	0	4.93E+05	0	0	NR/NR/R
ESI+	$C_4H_8N_2O$	7.67E+08	1.58E+09	1.54E+09	1.85E+08	1.41E+08	NR/R/R
ESI+	$C_6H_3NO_2$	0	1.40E+06	1.09E+06	0	0	NR/R/R

Areas are uncorrected, unnormalized formula-aggregated responses. b is the mean of the two CD background records. Retention requires A ≥ 50,000 and A ≥ 5b, with corrected area A − b. A positive raw response may be non-retained. R and NR denote retained and not retained. The 20 formulas were selected by corrected response and influence on the m/z and DBE/C summaries, without an additional blank-magnitude rule.

2. Section 3.2, lines 215–240: requiring detection in all 28 tunnel samples makes the 'persistent core' especially dependent on abundance and detection threshold. Show sensitivity at 25/28 or a similar prevalence threshold. A non-significant ANOVA of CoreFrac is not evidence that its contribution is invariant.

Response: We have clarified the core definition and added the requested prevalence sensitivity comparison. Detection in all three CD samples was required in the original calculation but was omitted from the previous definition. The analyzed subset is now termed the CD-shared persistent core. The corresponding wording in the Introduction and Section 3.2 now reads as follows:

"The controlled gasoline tailpipe profile was first characterized across the three operating conditions, while formulas detected in all 28 tunnel samples **and all three CD samples** defined the **CD-shared** persistent core." (Introduction, Lines 68–70)

"**The CD-shared persistent core was** defined separately in each ionization mode as formulas detected in all 28 **tunnel samples and all three CD** samples. **It** comprised 25 formulas in ESI− **and 156** in **ESI+. CoreFrac, the total peak area of core formulas divided by the total peak area of all assigned formulas in each sample and ionization mode, was 27.9 ± 5.5% and 39.6 ± 4.5%, respectively (mean ± SD).**"
(Section 3.2, Lines 246–249)

CoreFrac was calculated for each sample and ionization mode as the summed blank-corrected, unnormalized core peak area divided by the corresponding total assigned peak area. Only tunnel prevalence changed; CD 3/3, processed areas, signal-retention criteria and sample-specific denominators were held fixed. The supporting comparison is provided here in Table R12. Relaxing tunnel prevalence to at least 25/28 increased membership from 25 to 46 formulas in ESI− and from 156 to 281 in ESI+. Mean CoreFrac increased from 27.93% to 45.13% and from 39.60% to 46.18%, respectively. CHO and CHON remained predominant by formula number in ESI− (76.0% and 71.7%), as did nitrogen-containing formulas in ESI+ (69.2% and 65.5%). Thus, set size and signal share depend on prevalence, particularly in ESI−, while the predominant elemental groups remain.

Table R12**. Sensitivity of the CD-shared core to tunnel prevalence.**

Mode	Tunnel criterion	Formula count	CoreFrac (%) Mean ± SD	Observed range (%)	Main classes by number (%)
ESI−	28/28	25	27.93 ± 5.50	9.47–35.07	76.0
ESI−	≥25/28	46	45.13 ± 8.13	15.37–58.14	71.7
ESI+	28/28	156	39.60 ± 4.49	31.20–50.03	69.2
ESI+	≥25/28	281	46.18 ± 3.99	36.88–54.99	65.5

All rows require CD 3/3. CoreFrac summaries use 28 sample-wise ratios. The last column gives CHO + CHON in ESI− and all nitrogen-containing formulas (N ≥ 1) in ESI+, by formula number rather than peak area.

We agree that a non-significant ANOVA does not establish invariant contribution. We have removed the interpretation in Section 3.2 that the non-significant result demonstrates invariant or limited CoreFrac variation. Persistent detection describes recurrence under the applied retention criteria; it does not imply an unchanged relative signal contribution. Figure 3 retains the strict-core members and uses Non-core for the gray category. Its caption now reads as: "Colored triangles denote **CD-shared** persistent core formulas detected in **all 28 tunnel samples and all three CD samples within the corresponding ionization mode** and

are colored by elemental class (CHO, CHN, CHON, CHOS, **CHONS; Other comprises CH).** Gray circles denote **non-core** formulas from all tunnel samples. Density contours show the distribution of all formulas, and stacked bars show the elemental class composition **by formula number** of each core group." (Figure 3 caption, Lines 240–244)

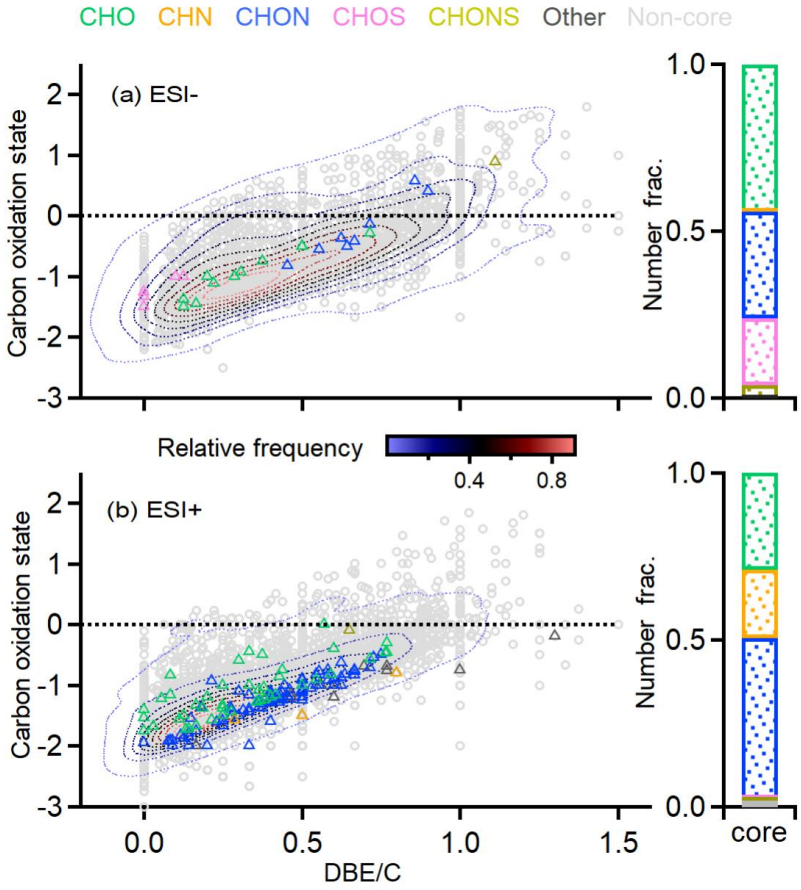


Figure 3. Colored triangles denote CD-shared persistent core formulas detected in all 28 tunnel samples and all three CD samples within the corresponding ionization mode and are colored by elemental class (CHO, CHN, CHON, CHOS, CHONS; Other comprises CH). Gray circles denote non-core formulas from all tunnel samples. Density contours show the distribution of all formulas, and stacked bars show the elemental class composition by formula number of each core group.

3. Supplement Table S2 and manuscript lines 230–242: rainy, cloudy, and sunny each correspond to one day with four within-day filters. The text acknowledges this; keep day-specific language everywhere and avoid weather-effect inference from these ANOVAs.

Response: We agree. The weather labels identify the recorded conditions on the three sampled days only. The four time-resolved filters of each day support the reported mean

response and within-day variation, whereas weather cannot be separated from other day-specific factors.

Section 3.2 therefore describes the ordering among the sampled days without F and p values, which remain in Table S5, and no longer refers to sampling-period summaries of formula number and total peak area. The Figure 1 caption identifies the sampled rainy and sunny days. The CoreFrac table cited in the comment, now Table S4, labels its groups as sampling period and sampling day, and Text S4 states the scope of each comparison. The revised text now reads as follows:

"Formula number and total HRMS peak area were **summarized for the three sampled campaign days, each with four time-resolved filters** (**Table S5**). In ESI−, mean total peak area was highest on the **sampled** cloudy day, lowest on the **sampled** rainy day, and intermediate on the **sampled** sunny day. In ESI+, mean total peak area was highest on the **sampled** sunny day and was approximately 50 % higher than on either the **sampled** cloudy or rainy day." (Section 3.2, Lines 261–265)

"Because each condition was represented by only one campaign day with four sampling periods, these comparisons describe day-specific variation during the campaign rather than weather effects (Tong et al., 2016)." (Section 3.2, Lines 265–266)

"Shaded bands indicate **the sampled** rainy **day** (blue) and **the sampled** sunny **day** (orange), and all other intervals are cloudy." (Figure 1 caption, Lines 179–180)

"Table S4. Descriptive statistics (mean ± SD) and ordinary one-way ANOVA results (F and p) for the core fraction (CoreFrac = core peak area / total peak area) across four diurnal periods and three **selected sampling days**, reported separately for ESI+ and ESI− modes. **Sampling periods contain seven filters each and sampling days four filters each. Rainy, cloudy and sunny describe the weather on the three selected sampling days.**" (Table S4 title and note)

"**CoreFrac was compared among four sampling periods (**seven **filters** per **period) and three selected sampling days (four filters per day; Table S4)**. **Formula number and summed HRMS peak area were summarized for** the three **selected sampling** days **(Table S5). …** Because each **condition was represented by only one** campaign **day, the sampling-day** tests quantify differences among the three **sampled** days rather than weather effects. **All comparisons used ordinary one-way ANOVA.**" (Text S4)

Table S5. Selected-day summaries across four periods.

Mode	Metric	Sampled rainy day	Sampled cloudy day	Sampled sunny day	F	p
ESI+	Formula number	1026 ± 119	1064 ± 196	1170 ± 46	1.21	0.343
ESI−	Formula number	309 ± 65	379 ± 78	284 ± 69	1.93	0.201
ESI+	Peak area ($\times 10^{10}$)	1.1 ± 0.2	1.1 ± 0.2	1.7 ± 0.2	10.27	0.005
ESI−	Peak area ($\times 10^{8}$)	6.0 ± 1.0	8.5 ± 0.8	6.7 ± 1.4	5.67	0.026

Each sampled day is represented by four 3 h filter samples (n = 4). Values are mean ± SD; F and p are from ordinary one-way ANOVA among the three sampled days. Weather labels describe conditions on the individual sampled days.

4. Figure 6 and Table S7: explain the different peak-area denominators for class, domain, mode, and all-assignment comparisons directly in the legend. Include the unscaled and normalized values for the influential features in a machine-readable table.

Response: We agree and have revised both legends. The Figure 6 caption and the Table S14 caption now state the denominator for the class, group and mode comparisons, and both refer to Table S6 for the share of each group in the total signal of all formulas selected for HCA. Panel (b) shows unweighted medians, and the earlier reference to peak-area weighting has been removed.

The requested machine-readable table is provided in the Supplementary Data file. For the 83 influential formulas (counted per ionization mode), it lists the blank-corrected unnormalized area and the normalized area in each of the 28 samples, with the normalization factor of each sample. The influential formulas are all formulas listed in Table S14 and all sulfur-containing formulas of group C2. For all 709 formulas selected for HCA, the Supplementary Data file lists the pooled areas, denominators, shares and heights, so that the Table S14 shares and the Figure 6(d) heights can be recalculated. The revised caption text now reads as follows:

"**Figure 6. Group composition and molecular descriptors.** (a) **Group** composition **by formula number. Association groups I and II denote formula-level associations with their respective reference sets. Each number fraction uses all formulas in that group as its denominator; a formula observed in both modes is counted once in each mode**. (b) **Unweighted group medians of** carbon oxidation state and modified aromaticity index**, with bars spanning the 25th–75th percentiles. (c) Van Krevelen distributions; circles** and

triangles denote ESI+ and **ESI−**. (d) **Group-**resolved **spectra. Mean** normalized areas **across all 28 samples, including non-detection zeros, are divided by the largest mean within the same group and mode and multiplied by 100. These heights are base-peak-scaled profiles. Group shares of the pooled signal of all formulas selected for HCA in each mode are reported separately in Table S6.** Inset **pies** show elemental**-class number fractions; Other combines CH, CHNS and CHS. n is the group formula count. Underlying occurrence** and **association fields, sample-level values and denominators are supplied in the Supplementary Data file.**" (Figure 6 caption, Lines 392–401)

"Table **S14**. Highest-response formulas by group and mode. **… Formulas** are **ranked separately** within **each group** and mode. **Peak-area percentages use normalized, non-log-transformed signals pooled across the 28 tunnel samples. The numerator is the listed formula's pooled area; the denominator is the pooled area of all members in the same group and mode, including those not displayed. Class contributions within a group use the same total. ESI+ and ESI− areas are not pooled. Comparison m/z values are formula-based values used for plotting. The Supplementary Data file supplies unnormalized and normalized values, normalization factors and complete totals. Group shares of the pooled signal of all formulas selected for HCA in each mode are reported in Table S6.**" (Table S14 caption)

Table S8. Formula numbers and structural summaries.

Metric	C1	C2	C3
Description	Shared CHO/CHON baseline	Sulfur-rich tunnel-retained composition	Low-mass CHN response
Formulas (ESI+ / ESI−)	585 (518/67)	74 (61/13)	50 (47/3)
TCC, n (%)	124 (21.20)	5 (6.76)	6 (12.00)
T–O, n (%)	414 (70.77)	63 (85.14)	34 (68.00)
Association group I, n (%)	33 (5.64)	3 (4.05)	10 (20.00)
Association group II, n (%)	14 (2.39)	3 (4.05)	0 (0.00)
CHO + CHON, n (%)	356 (60.85)	29 (39.19)	36 (72.00)
Sulfur-containing, n (%)	152 (25.98)	39 (52.70)	10 (20.00)
CHOS + CHONS, n (%)	120 (20.51)	35 (47.30)	7 (14.00)
Median carbon oxidation state	−1.07	−1.19	−1.09
Median modified aromaticity index	0.29	0.22	0.35

Formulas in Association groups I and II are not counted again as TCC or T–O. Number fractions use each group total. Sulfur-bearing includes CHOS, CHONS, CHNS and CHS; CHOS+CHONS is the oxygen-containing subset.

Table S7. Annotation-specific normalized area fractions (%).

Group	Mode	TCC	T–O	Group I	Group II
C1	ESI+	61.40	33.14	4.61	0.85
C1	ESI−	58.63	41.37	0.00	0.00
C2	ESI+	4.80	89.92	4.10	1.18
C2	ESI−	4.84	79.84	0.00	15.33
C3	ESI+	4.35	82.70	12.95	0.00
C3	ESI−	36.64	63.36	0.00	0.00

Each denominator includes all members of the indicated group and mode. Association labels take precedence over TCC/T–O in the display, so the displayed T–O shares exclude association-labelled formulas. Values sum to 100% before rounding.

Table S14. Highest-response formulas by group and mode.

Mode	Group	Rank	Formula	Comparison m/z	Annotation	Peak area (%)
ESI+	C1	1	$C_{12}H_{14}N_2O_2$	219.1128	TCC	11.10
ESI+	C1	2	$C_6H_8O_2$	113.0597	T–O	10.13
ESI+	C1	3	$C_{12}H_{23}NO$	198.1852	TCC	9.76
ESI+	C1	4	C_4H_8O	73.0648	TCC	6.50
ESI+	C1	5	$C_5H_{12}N_2O$	117.1022	TCC	4.35
ESI+	C1	6	$C_{12}H_{17}NO$	192.1383	TCC	3.98
ESI+	C1	7	$C_{10}H_{14}N_2$	163.1230	TCC	3.48
ESI+	C1	8	$C_{10}H_{16}O_5$	217.1071	TCC	1.79
ESI+	C1	9	$C_{10}H_{22}O_4$	207.1591	TCC	1.56
ESI+	C1	10	$C_{12}H_{16}N_2O_2$	221.1285	TCC	1.48
ESI−	C1	1	$C_{14}H_{21}NO_3$	250.1449	TCC	29.36
ESI−	C1	2	$C_8H_9NO_3$	166.0510	TCC	9.34
ESI−	C1	3	$C_{18}H_{34}O_3$	297.2435	T–O	6.37
ESI−	C1	4	$C_9H_{18}O_5S$	237.0802	T–O	5.10
ESI−	C1	5	$C_{16}H_{30}O_4$	285.2071	TCC	4.39
ESI−	C1	6	$C_{11}H_{15}NO_3$	208.0979	TCC	3.73
ESI−	C1	7	$C_{18}H_{32}O_3$	295.2279	TCC	3.70
ESI−	C1	8	$C_{19}H_{30}O_4$	321.2071	T–O	2.78
ESI−	C1	9	$C_{12}H_{14}N_2O_3$	233.0932	T–O	2.16
ESI−	C1	10	$C_{12}H_{14}N_2O_2$	217.0983	T–O	2.14
ESI+	C2	1	$C_{11}H_{12}N_2OS$	221.0743	T–O	6.30
ESI+	C2	2	$C_{20}H_{26}O_3$	315.1955	T–O	4.03
ESI+	C2	3	$C_{39}H_{50}N_2O$	563.3996	T–O	3.62
ESI+	C2	4	$C_{28}H_{43}N$	394.3468	T–O	3.52
ESI+	C2	5	$C_{10}H_{19}NO_7S$	298.0955	T–O	3.23
ESI+	C2	6	$C_7H_4N_2O_6S$	244.9863	T–O	3.02
ESI+	C2	7	$C_{12}H_{17}NO_7S$	320.0798	T–O	2.97

Mode	Group	Rank	Formula	Comparison m/z	Annotation	Peak area (%)
ESI+	C2	8	$C_{13}H_{24}N_2O_2S$	273.1631	T–O	2.77
ESI+	C2	9	$C_{21}H_{39}NO_{15}S_2$	610.1834	T–O	2.72
ESI+	C2	10	$C_{15}H_{32}N_2OS$	289.2308	T–O	2.39
ESI−	C2	1	$C_9H_{19}NOS$	188.1115	T–O	33.28
ESI−	C2	2	C_7H_5NOS	150.0019	Group II	15.33
ESI−	C2	3	$C_{10}H_8O_4$	191.0350	T–O	7.15
ESI−	C2	4	$C_{17}H_{30}O_3$	281.2122	T–O	6.36
ESI−	C2	5	$C_{18}H_{34}O_2$	281.2486	T–O	6.24
ESI−	C2	6	$C_{20}H_{26}O_3$	313.1809	T–O	5.54
ESI−	C2	7	$C_{12}H_{10}N_2O_2$	213.0670	T–O	5.10
ESI−	C2	8	$C_{18}H_{32}O_4$	311.2228	TCC	4.84
ESI−	C2	9	$C_{15}H_{10}O_5$	269.0455	T–O	3.90
ESI−	C2	10	$C_{18}H_{30}O_4$	309.2071	T–O	3.44
ESI+	C3	1	$C_4H_{11}N$	74.0964	T–O	41.29
ESI+	C3	2	$C_4H_{10}N_2$	87.0917	T–O	21.69
ESI+	C3	3	$C_{17}H_{22}O$	243.1743	Group I	4.70
ESI+	C3	4	$C_{16}H_{20}O$	229.1587	Group I	4.04
ESI+	C3	5	$C_{18}H_{22}O$	255.1743	T–O	2.89
ESI+	C3	6	$C_{19}H_{30}O$	275.2369	TCC	2.15
ESI+	C3	7	$C_{19}H_{28}O$	273.2213	T–O	2.08
ESI+	C3	8	$C_{19}H_{34}N_2O_2$	323.2693	T–O	1.09
ESI+	C3	9	$C_{32}H_{30}N_2O_2$	475.2380	Group I	1.09
ESI+	C3	10	$C_{16}H_{29}NO_6$	332.2068	T–O	0.95
ESI−	C3	1	$C_{18}H_{28}O_4$	307.1915	T–O	45.66
ESI−	C3	2	$C_{12}H_{20}O_4$	227.1289	TCC	36.64
ESI−	C3	3	$C_{18}H_7NO_6$	332.0201	T–O	17.71

Formulas are ranked separately within each group and mode. Peak-area percentages use normalized, non-log-transformed signals pooled across the 28 tunnel samples. The numerator is the listed formula's pooled area; the denominator is the pooled area of all members in the same group and mode, including those not displayed. Class contributions within a group use the same total. ESI+ and ESI− areas are not pooled. Comparison m/z values are formula-based values used for plotting. The Supplementary Data file supplies unnormalized and normalized values, normalization factors and complete totals. Group shares of the pooled signal of all formulas selected for HCA in each mode are reported in Table S6.

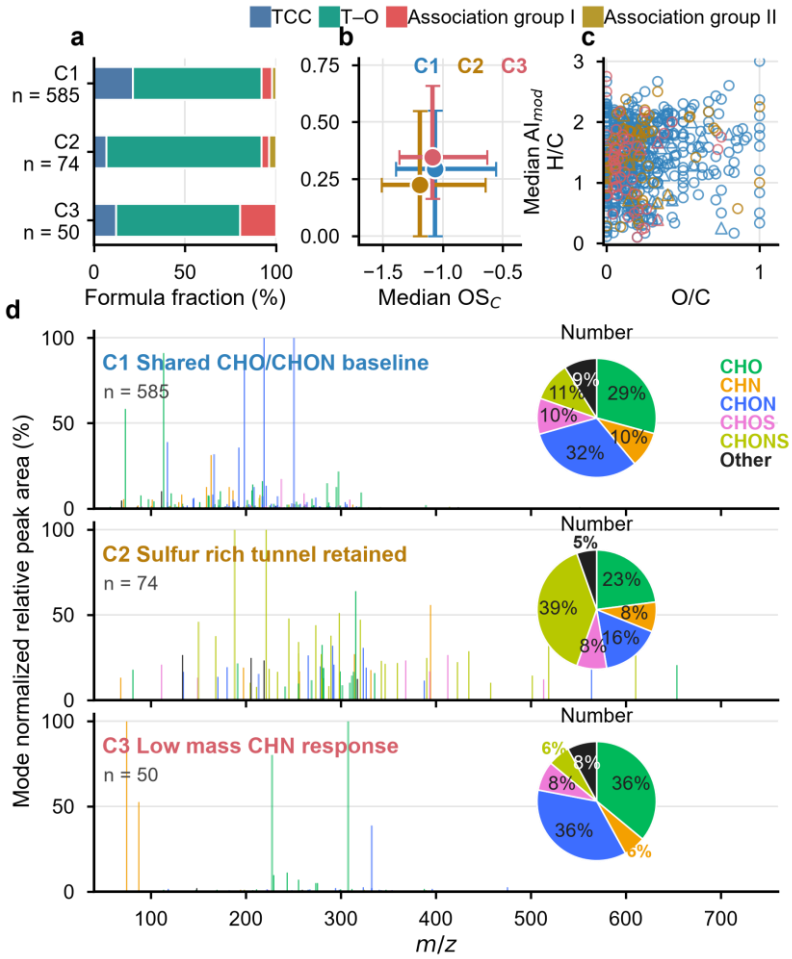


Figure 6. Group composition and molecular descriptors. (a) Group composition by formula number. Association groups I and II denote formula-level associations with their respective reference sets. Each number fraction uses all formulas in that group as its denominator; a formula observed in both modes is counted once in each mode. (b) Unweighted group medians of carbon oxidation state and modified aromaticity index, with bars spanning the 25th–75th percentiles. (c) Van Krevelen distributions; circles and triangles denote ESI+ and ESI−. (d) Group-resolved spectra. Mean normalized areas across all 28 samples, including non-detection zeros, are divided by the largest mean within the same group and mode and multiplied by 100. These heights are base-peak-scaled profiles. Group shares of the pooled signal of all formulas selected for HCA in each mode are reported separately in Table S6. Inset pies show elemental-class number fractions; Other combines CH, CHNS and CHS. n is the group formula count. Underlying occurrence and association fields, sample-level values and denominators are supplied in the Supplementary Data file.