https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

Non-Target Analysis of Atmospheric Organic Aerosols as a Tool to Discriminate Anthropogenic Contribution in Mixed Air Masses during the ACROSS campaign

Niklas Karbach[1], Pauline Pouyes[2, 3], Emilie Perraudin[3], Eric Villenave[3], Thorsten Hoffmann[1]

[1]Department of Chemistry, Johannes Gutenberg-University, Mainz, 55128, Germany
[2]African Research Center on Air Quality and Climate, University Mohammed VI Polytechnic, Benguerir, 43150, Morocco
[3]University of Bordeaux, EPOC UMR 5803 CNRS, CS50023, 33615 Pessac cedex, France

Correspondence to: Thorsten Hoffmann (hoffmant@uni-mainz.de)

Keywords: organic aerosol, source apportionment, LC-HRMS, non-target analysis, biogenic and anthropogenic markers

Abstract. Organic aerosol is a major component in the particle phase of Earth´s atmosphere and has influences on quality of life, health and climate. In this study, a non-target analysis of the chemical composition of atmospheric organic aerosols using liquid chromatography-Orbitrap mass spectrometry (LC-Orbitrap MS) was conducted to differentiate anthropogenic and biogenic sources through unsupervised KMeans clustering. The ACROSS campaign dataset (consisting of 36 wind-characterized samples) identified 4,916 compounds (in the range 50–400 *m/z*). Due to the location of the sampling site, the samples contain influences from the greater Paris area, as well as biogenic influences from the surrounding forest. K-means clustering, constrained to 2,917 compounds with strong wind-direction correlation, resolved distinct biogenic and anthropogenic clusters. Biogenic aerosols were dominated by CHO compounds (H/C: 1.2–1.7; O/C: 0.15–0.7), consistent with oxidized terpenes, while anthropogenic aerosols featured significant CHOS enrichment (H/C: 1.5–2.2; O/C: 0.2–1.0), including nitrogen-sulfur aromatics (e.g., $C_{10}H_{18}NO_8S^-$ with nitro/sulfonic groups and aromatic fragments). The approach allows to quantify anthropogenic contribution in mixed air masses, demonstrating higher amounts of anthropogenic compounds ratios during Paris-influenced periods. Results validate wind-driven source apportionment for small sample size non-target studies, providing a transferable method for aerosol characterization.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

1 Introduction

1.1 Organic aerosol in the atmosphere

Organic aerosol (OA) is a major component of atmospheric fine particulate matter and frequently dominates the submicron particle mass budget across urban, rural, and remote regions (Murphy et al., 2006; Zhang et al., 2007; Mauderly and Chow, 2008). OA affects climate directly through scattering and absorption of solar radiation and indirectly by altering cloud condensation nuclei activity and cloud microphysics (Gouw and Jimenez, 2009). OA is also linked to adverse health outcomes because fine particles penetrate deep into the respiratory tract and can carry redox active organic functionality and toxic combustion related constituents (Nel, 2005; Pope et al., 2009). Despite decades of research, OA remains a major source of uncertainty in air quality prediction and aerosol radiative forcing because it is chemically complex, multiphase, and dynamically evolving from different physico-chemical aging processes. Ambient OA comprises primary organic aerosol (POA) emitted directly from traffic, biomass burning, cooking, and industrial sources, and secondary organic aerosol (SOA) formed in the atmosphere from oxidation of volatile, intermediate volatility, and semi volatile organic compounds (VOCs, IVOCs, SVOCs) followed by gas particle partitioning and particle phase chemistry (Gouw and Jimenez, 2009).

OA is mostly measured using complementary online and offline approaches that trade temporal resolution against chemical specificity. Aerosol mass spectrometers provide high time resolution bulk composition and oxidation metrics but limited isomer and structure specificity. Filter based sampling enables integration over hours to days and supports chromatographic separation with compound resolved quantification and molecular level characterization by GC MS and LC MS, including polar and thermally labile constituents that are inaccessible to GC based workflows (Laskin et al., 2012; Hildmann and Hoffmann, 2024; Pereira et al., 2026).

High resolution mass spectrometry (HRMS) coupled to liquid chromatography captures thousands of features spanning CHO, CHON, CHOS, and CHONS elemental classes and provides accurate mass and isotopic constraints for formula assignment. Orbitrap based LC HRMS is particularly attractive for non-target analysis (NTA) because it combines high resolving power, stable mass accuracy, and optional fragmentation, enabling both broad molecular fingerprinting and targeted follow up of marker species. NTA workflows typically include peak detection and alignment, adduct grouping, isotope filtering, elemental formula assignment, and annotation supported by MS^2 spectra (Vogel et al., 2019; Giorio et al., 2019; Hildmann and Hoffmann, 2024; Karbach and Hoffmann, 2025; Pereira et al., 2025; Leppla et al., 2026). Because OA datasets are high dimensional and sample series reflect changing sources and oxidation conditions, multivariate statistics and clustering are used to identify chemically coherent subsets and connect them with environmental drivers. The present work applies HPLC Orbitrap MS NTA to day and night filter samples collected during the 2022 ACROSS campaign in France, with clustering used to distinguish anthropogenic and biogenic marker compounds.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

1.2 ACROSS campaign

The ACROSS project (Atmospheric Chemistry of the Suburban Forest), coordinated by Christopher Cantrell and Vincent Michoud (from LISA Créteil), was designed to quantify how urban emissions from a major European megacity (Paris, France) interact with biogenic emissions from surrounding vegetated areas and how this mixing modifies oxidation chemistry, secondary pollutant formation, and aerosol properties. The campaign leveraged the Greater Paris region as a natural laboratory where anthropogenic pollutants (NO_x, SO_2) and ozone are frequently transported toward suburban and forested environments that emit isoprene, monoterpenes, and other BVOCs, especially during heat events (Di Antonio et al., 2025).

1.3 Anthropogenic and biogenic markers in organic aerosol

Molecular markers are individual compounds or compound families that are sufficiently source specific, stable over relevant timescales, and analytically accessible such that they provide interpretable constraints on sources and processes. Mixed anthropogenic biogenic environments add a layer: anthropogenic NOx, sulfate, and ozone concentration levels can amplify or suppress biogenic SOA formation channels and generate hybrid products such as organonitrates and organosulfates that carry both biogenic carbon skeletons and anthropogenic heteroatoms. Hence in this study we tried to identify marker compounds that allow identification and quantification of anthropogenic contributions to given air masses.

Anthropogenic OA markers include primary combustion and fossil related tracers such as polycyclic aromatic hydrocarbons and their oxygenated or nitrated derivatives, levoglucosan and related anhydrosugars as tracers of biomass burning, and organosulfates that formed through interaction of anthropogenic SO_2 and biogenically influenced air masses. In a Van-Krevelen plot, a type of plot that has become standard in non-target analysis of organic aerosol (Kim et al., 2003; Heald et al., 2010; Zhang et al., 2021; Leppla et al., 2026), anthropogenic SOA markers are typically found in the lower left region (aromatics / nitroaromatics) and the upper left region (alkanes, organosulfates, fatty acids, …) (Thoma et al., 2025). Biogenic markers commonly derive from isoprene and terpenes (H/C = 1.6) and can therefore be found to agglomerate in a range of $1.2 \leq H/C \leq 1.6$ (Thoma et al., 2025). For monoterpenes, pinic acid, pinonic acid, and later generation multifunctional acids such as MBTCA are frequently observed, with MBTCA often interpreted as an aging indicator.

1.4 Orbitrap fragmentation

Accurate mass MS^1 alone is rarely sufficient for structural assignment in OA because isomers are abundant and many formulas map to multiple plausible functional group patterns and oxidation histories. Tandem MS provides orthogonal constraints through diagnostic neutral losses and fragment ions that report on functional groups and can connect unknowns to marker families when authentic standards are unavailable. Orbitrap instruments typically implement higher energy

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

collisional dissociation, and fragmentation is commonly acquired as data dependent MS^2 (DDA), which prioritizes abundant precursors however. Data independent acquisition (DIA) strategies address this limitation by collecting fragmentation information more comprehensively, improving retrospective interpretability across large sample sets. DIA experiments are not only applied in atmospheric analysis (Karbach and Hoffmann, 2025), but also in proteomics (Tsou et al., 2016; Ishikawa et al., 2022), oncology (Krasny and Huang, 2021), etc. Basically, wherever a non-target analysis is conducted and complete data for retrospective analysis is needed. The All Ion Fragmentation (AIF) mode fragments all ions, generating systematic MS^2 like information without precursor selection, which is attractive for NTA, as no information is lost and experiments can easily be reevaluated later on. Recent workflows automate fragment annotation and improve precursor fragment linking for full MS plus AIF files, enabling construction of quasi-isolated MS^2 spectra that allow for easy manual evaluation of individual compounds (Karbach and Hoffmann, 2025). Fragment evidence can confirm membership in marker families, support separation of anthropogenic versus biogenic clusters by functional group signatures, and enable robust matching across day and night samples even when precursor intensities vary strongly. Combined with clustering on MS^1 feature tables and targeted confirmation for key markers, Orbitrap DIA can help translate high dimensional NTA outputs into interpretable source and process constraints for mixed urban biogenic field campaigns, like during ACROSS.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

2 Experimental

2.1 Sample preparation

Atmospheric organic aerosol were sampled during the ACROSS campaign from 28.06.2022 to the 19.07.2022 at the ACROSS sampling site which was located in the Rambouillet forest (approx. 45 km SW of Paris). Particles were collected days and nights on 150 mm quartz fiber filters with a Digitel DA-80 high volume sampler in PM2.5 mode. The sampling volume was 30 m^3 h^{-1} and the sampling head was located 4 m above ground level. After sampling, one half of a 47 mm filter punch was taken which was then extracted three times with a mixture of MeOH:H_2O (9:1). The extract was then evaporated to ~100 µL for analysis. Field blanks, once every 2.5 days over the duration of the campaign were stored, extracted and measured in the same way. The blank areas were subtracted from the adjacent filters and signals were only analyzed when the sample signal was at least 3 times higher than the blank signal. Prior to analysis, the filters were stored at -18 °C.

2.2 LC/HRMS method & analysis

The measurements were performed on a Dionex UltiMate 3000 UHPLC system coupled to a heated ESI source (spray voltage: -4 kV, heated to 150 °C, capillary temperature: 350 °C) and a Q Exactive Orbitrap (both Thermo Fisher Scientific). All MS spectra were acquired with a resolution of 140000 between m/z ratios of 50-400 in the negative ionization mode. The MS method was a DIA experiment (full MS/AIF) with fragmentation energies of 11, 30, 90 NCE (normalized collision energy). A Waters Acquity UPLC CSH Fluoro Phenyl (PFP) column with 100 mm x 2.1 mm (1.7 µm particle size) located in a 30 °C column oven with a flowrate of 0.5 mL/min was used. The chromatographic method lasted 13 min and a gradient of two mobile phases A (ultra pure water with 2 % ACN and 0.04 % formic acid) and B (ultra pure ACN with 2 % water) was used. The gradient used for all separations was as follows: 0-11 min: 10 % B linear increase to 99 % B; 11-12 min: holding 99 % B; 12-12.5 min: 99 % B decrease to 10 % B; 12.5-13 min: holding 10 % B. The measurements were analyzed with the tool presented in (Karbach and Hoffmann, 2025), yielding formulas of molecular ions and related fragmentation information. The measured intensities were normalized on the individual volume sampled, yielding intensities proportional to the concentration of the substance in the sampled air mass.

2.3 Data processing

The measured raw data were imported into mzmine4 where a workflow created with the built-in *mzwizzard* processed the data (mass detection, chromatogram builder, smoothing, local minimum resolver, 13C isotope filter, isotopic peak finder, join aligner, duplicate peak filter) to yield a single aligned list of all compounds that could be detected in the raw data files.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

This list of compounds (samples and blanks) was then exported as .csv file. After exporting the data, the blank was subtracted from the data three times, and then rejecting all the compounds whose intensity was below 0. This list of compounds, which is unique for every raw data file (mass and retention time) was then passed to the analysis tool detailed in (Karbach and Hoffmann, 2025) where the chemical formula and corresponding fragments were determined. This yields a summary file for every sample that was measured containing only compounds that are above the detection limit (i.e. > 3 times the blank value) and do show at least one additional isotopic signal. All further analysis was then done with the predicted formula, identified fragments, and the calculated peak area.

2.4 HYSPLIT

Wind data for the individual samples were acquired using the HYSPLIT model (Stein et al., 2015). For each sample, one archive trajectory was calculated in the ensemble mode for a single source. The GFS (0.25 degree, global) model was used for the calculations, and each trajectory was calculated 24 h backwards from the end of the sampling period. The height of the source was set to 10 m AGL. With the HYSPLIT data, a predominant wind direction was manually assigned to every sample. The determined wind direction based on the back trajectory should better represent the true origin of the sampled air mass rather than just the wind direction during the time of sampling (as was also measured during the ACROSS campaign), as also the history of the air mass is considered. The determined wind direction of the air mass does however generally match the measured wind direction at the time of sampling (Di Antonio et al., 2025). A summary of the results can be seen in Figure 1.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

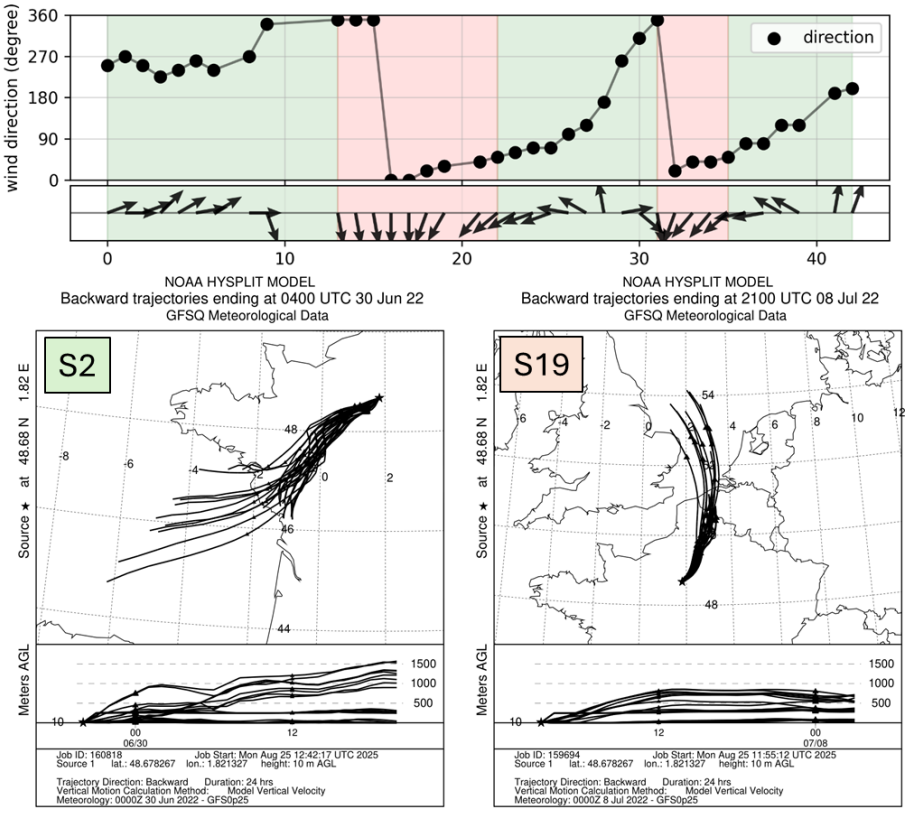

Figure 1: (top): wind direction for every sample during the ACROSS 2022 campaign. When the air mass was coming from over the greater Paris area, the air mass was considered to be anthropogenic (red shaded areas) and for other directions the air mass was considered not to be anthropogenically influenced and therefore referred to as biogenic (green shaded areas). The two maps on the bottom show the output of HYSPLIT calculations for sample S2 (biogenic) and sample S19 (anthropogenic) which represent typical trajectories for the two cases.

2.5 Apportionment of the sum formulas origin

The wind data, determined as described in chapter 2.4, is used to assign sources to individual compounds, where compounds whose intensity peaks during N-NE (red shaded areas) influenced wind conditions likely originate from the greater Paris area and are therefore likely to be of anthropogenic origin or anthropogenically influenced, whereas samples taken when the airmass was coming from the other directions are likely to be mostly of mixed or of biogenic origin (green shaded areas). A Python script was written to automatically perform the clustering based on wind direction (more specifically naturally influenced air mass or anthropogenic influenced air mass) and compound intensity.

Peak areas per assigned molecular formula were normalized to sampled air volume (to account for different sampling durations) and to total signal per sample (to account for different total particle concentrations, e.g. due to rain). The normalization helps to remove meteorological influences in the dataset and keep directional effects preserved, generally

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

improving directional clustering results. Each detected compound´s normalized intensity profile was partitioned into user labeled natural and anthropogenic periods, and, the mean normalized contribution during natural and anthropogenic periods and their difference (natural minus anthropogenic) was calculated. Compounds with insufficient summed signal, or when more than 15 samples were below detection limit, were excluded. The compounds were then clustered with KMeans (k=3, random_state=0). Cluster membership was interpreted as source related groups with cluster 0 being the biogenic, and cluster 2 being the anthropogenic cluster.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

3 Results & Discussion

3.1 Whole dataset

The ACROSS dataset that was analyzed in this work consists of 37 individual samples (18 nighttime samples and 19 daytime samples) in which a total of 4916 individual compounds (between *m/z* = 59.0133, and *m/z* = 398.3272) have been identified. Most of the compounds are in the range *m/z* = 250 – 355 (see Figure 2 (left)). The distribution of compounds across the retention times is rather constant with decreasing numbers of identified compounds towards the end of the chromatographic run (Figure 2 (right)).

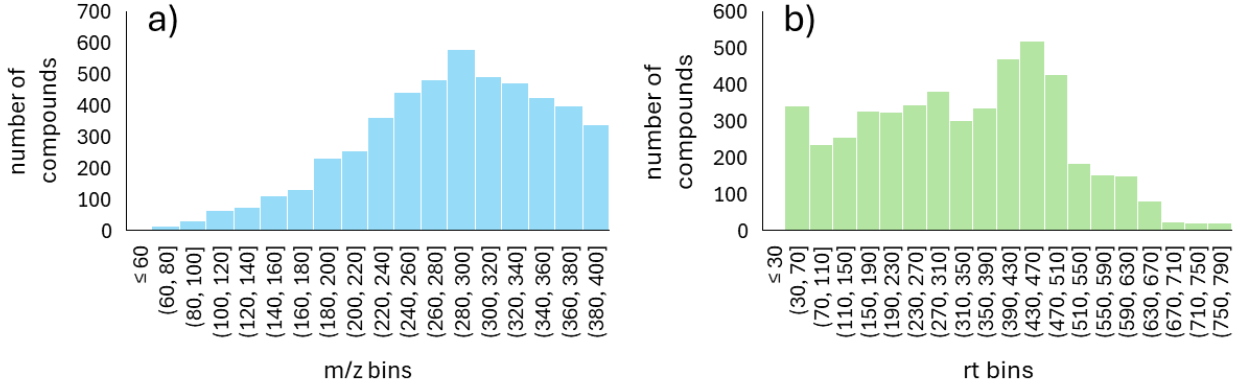

Figure 2: a) Histogram of the distribution of compounds across m/z bins (bin size: 20 amu, total number of compounds. 4916, m/z range: 50-400 amu). b) Histogram of the retention times (bin size: 40 s, total number of compounds: 4916, LC runtime: 390 s).

In the CHO subset 1406 compounds could be identified. In the CHOS subset 1235, in the CHON subset 1015, and 1260 compounds in the "others" subset. Comparison with subset distributions from samples taken for instance in the Amazon rainforest (Leppla et al., 2026) show higher numbers for the CHON and CHOS subsets (or respectively lower abundances of the CHO subset), which can be explained by the higher anthropogenic influence in the ACROSS dataset. Later in this study it is shown, that sulfur and nitrogen containing compounds (most dominant in the CHOS subset) are dominant in anthropogenically influenced air masses, which explains the observed high abundances in the CHOS and CHON subset (e.g. (Rincón et al., 2012; Wang et al., 2017)). As can be seen in Figure 3 (left), most of the compounds are below an O/C ratio of 1, and between an H/C ratio of 1.2 – 2. The Kroll plot shown in Figure 3 (right (Kroll et al., 2011)) highlights the evolution of organics in the atmosphere, with fragmentation, functionalization and oligomerization, leading to the typically shaped curve. Oxidation of organics leads to functional groups being introduced into the molecule, ultimately increasing COS (vertical movement). When oxidation progresses, molecules ultimately lose parts of its structure (fragmentation), leading to

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

horizontal movement in the right direction. It is therefore evident that with increasing time of the organics in the atmosphere the compounds contained in the air mass move upwards and to the right in the Kroll plot, which for mixed air masses lead to the typical curved shape.

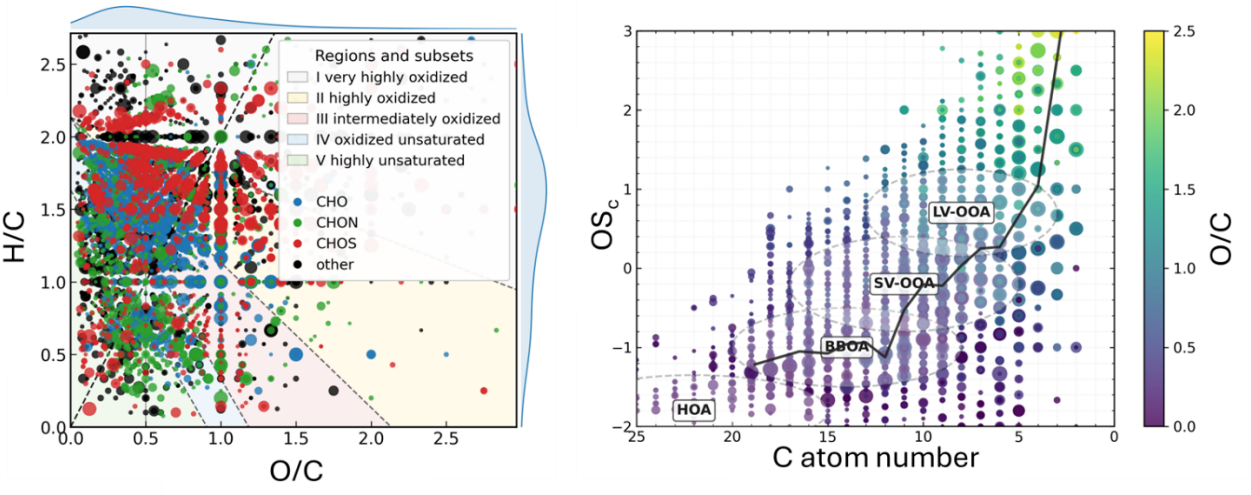

Figure 3: Left: Van-Krevelen plot of all compounds detected in the ACROSS dataset. The 5 shaded areas correspond to very highly oxidized compounds (I, light blue), highly oxidized compounds (II, yellow), intermediately oxidized compounds (III, red), oxidized unsaturated compounds (IV, blue), and highly unsaturated compounds (V, green). The theory behind those shaded areas was taken from (Zhang et al., 2021). The Van-Krevelen plot visualizes the large abundance of compounds in the whole dataset, spanning three weeks of day and night samples with variable metrological conditions. Right: Kroll plot of all detected compounds in the ACROSS dataset. The shape of identified compounds follows the expected curvature for atmospheric organic aerosol samples which results in a combination of functionalization and fragmentation, sequentially leading to higher OS_c and lower C atom numbers as explained by (Kroll et al., 2011). The oval overlays visualize different regions of different groups like hydrocarbon like (HOA), biomass burning (BBOA), semi-volatile (SV-OOA) or low-volatile (LV-OOA) organic aerosol.

3.2 Individual results

As can be seen when considering the back trajectories, the samples taken during the campaign can be clustered into two types. One where the wind was coming from over the greater Paris area, and one where the wind was *not*. Figure 4 (top part) shows the sample S6 where the wind was coming from the W-SW direction with the air mass not being anthropogenically influenced, and sample S22 during which the wind was coming from NE direction and therefore anthropogenically influenced. Both samples are nighttime samples, so differences can solely be attributed to different meteorological conditions during sampling. It can be seen that for biogenically influenced air masses (S6), a large accumulation of compounds from the CHO subset can be found between $1.2 \leq H/C \leq 1.7$, and $0.15 \leq O/C \leq 0.7$. In S22 (anthropogenically influenced) there is however another distinct accumulation of CHOS compounds ($1.5 \leq H/C \leq 2.2$; $0.2 \leq O/C \leq 1.0$) that is missing in the biogenically influenced sample. Comparison with sample S21 (daytime sample, lower right of Figure 4) shows no significant decrease in the size of the CHOS cluster, indicating no influence of the diurnal cycle on the occurrence

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

of compounds in this cluster. However, the CHO cluster does show a significant increase during daytime as can also be seen in the comparison of sample S21 and S22, suggesting increased production of those compounds during the day. Looking at sample S5 (daytime sample, lower left of Figure 4), it can be seen, that even during biogenically influenced time periods, the amount and intensity of CHOS compounds increases, which as later discussed, might be explained by increased anthropogenic activity and therefore anthropogenic emissions during the day.

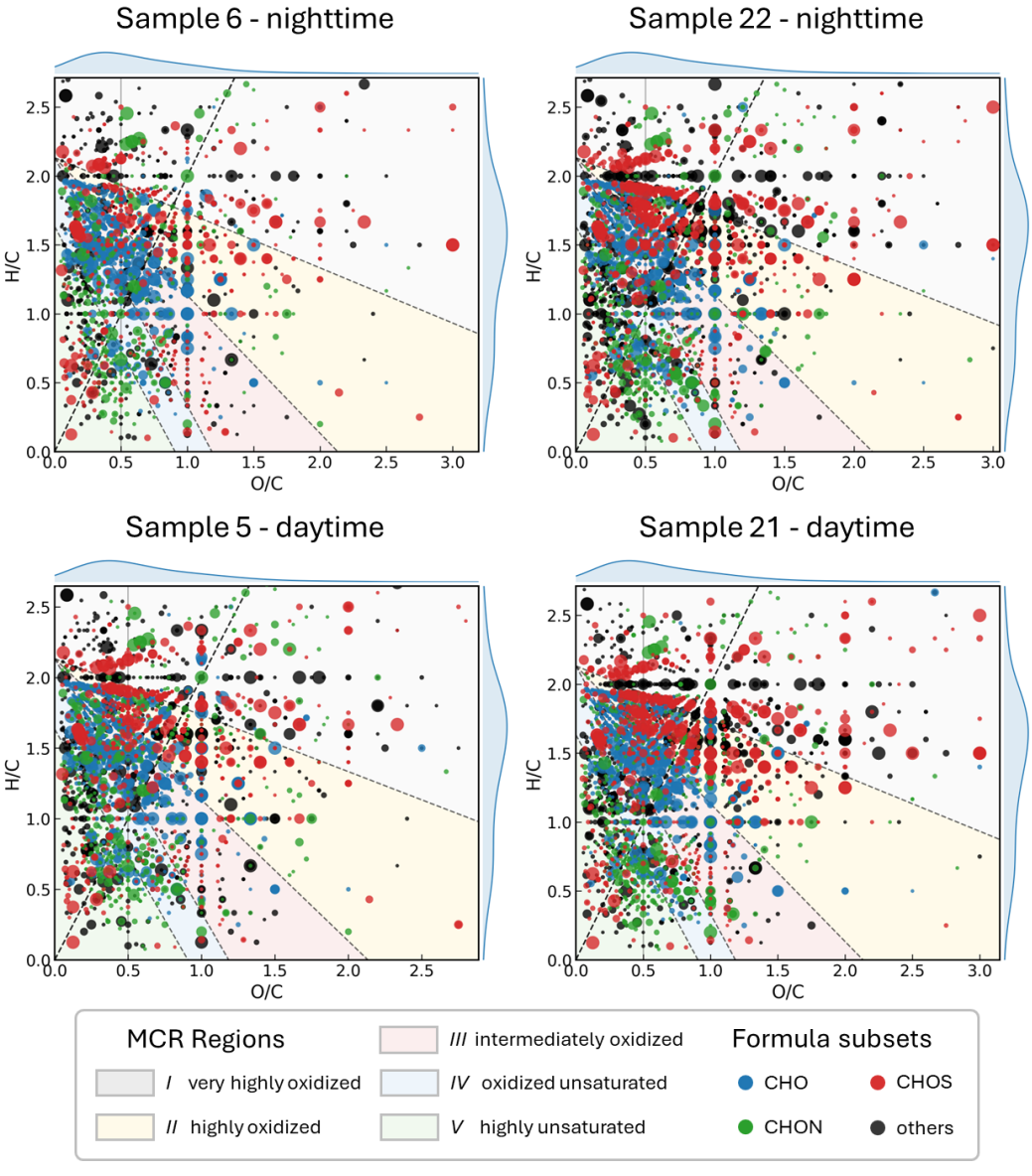

Figure 4: Van-Krevelen plots of 4 individual samples. The plots on the left side are during biogenically influenced time periods (S5 and S6) and the ones on the right side are during anthropogenically influenced time periods (S21 and S22). The top ones show nighttime samples, whereas the ones on the bottom show daytime samples. The difference between the biogenically and the anthropogenically influenced samples is most obvious during the nighttime, as can be seen by the accumulation of CHO

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

compounds at 1.2 ≤ H/C ≤ 1.7, and 0.15 ≤ O/C ≤ 0.7 in the biogenically influenced sample S6. This accumulation is missing in the nighttime sample of the anthropogenically influenced sample S22. On the other hand, sample S22 contains an accumulation of CHOS compounds at 1.5 ≤ H/C ≤ 2.2; 0.2 ≤ O/C ≤ 1.0 which is missing in sample S6. Differences between the anthropogenic and biogenic samples at daytime are less obvious as can be seen in the lower part of the figure. Here both accumulations specific to the anthropogenic and the biogenic cluster are still visible in the individual plots, however it seems more mixed in the sense that sample S21 (anthropogenic) contains also elements of the biogenic cluster, and sample S5 (biogenic) does also contain compounds from the anthropogenic cluster.

3.3 Clustering

As the whole dataset is quite large, looking and identifying individual compounds or compound classes directly is hardly possible. Clustering reduces the complexity of the dataset, creating smaller subsets, making it significantly easier to identify interesting compounds and compound classes. Upon looking at individual samples (namely sample S6 and sample S22), two different accumulations of compounds have been identified, one of which originates from anthropogenically influenced air masses and one of which originates from biogenically influenced air masses. Therefore, KMeans clustering was performed on the whole dataset with an input being the wind direction of the individual samples. The clustering results can be seen in Figure 5. It is evident that the two manually identified accumulations (the anthropogenic and the biogenic) in chapter 3.2 are also found in the results of the automatic KMeans clustering. To reduce false positives and increase confidence of the clustering results, compounds that do not show a strong wind correlation (namely having a directional dominance of their intensity of < 0.4) are removed prior to clustering, leaving a total of 2917 compounds. 1894 of which are in the biogenic cluster and 1023 of which being in the anthropogenic cluster. As can be seen in Table 1, the clustering results show a higher proportion of sulfur containing compounds in the anthropogenic dataset, resulting in formation of organic sulfates and sulfonates by interaction of biogenic air masses with anthropogenic emissions (Brüggemann et al., 2020; Glasius et al., 2022; Ma et al., 2025). Upon comparison with a recent study, our results show similar clustering results, despite having significantly fewer samples (352 filter samples in (Thoma et al., 2025), only 37 filter samples in this study). This can be achieved through rigorous filtering of compounds that do not show a strong wind correlation, and excluding them from the clustering results. A larger dataset would increase the number of compounds that are assigned to each cluster as small anomalies in the compounds concentration profiles would be averaged out.

As a further validation experiment, an additional new clustering run has been performed, where samples S6, S21, S22 and S38 (for compound selection see chapter 3.3.4) have been removed from the training dataset prior to clustering. The clustering results show very similar behavior compared to the first clustering run where all available samples were used. In this run, the anthropogenic cluster consisted of 996 compounds and the biogenic cluster 1, and 2 of 695, and 1138 compounds respectively. The individual Van Krevelen plots show the same accumulation of compounds as identified in Figure 5. These Van Krevelen plots, and the bootstrapping plot of the validation clustering run can be found in the Supplemental Information.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

Table 1: Number of identified chemical formulas and relative amount for each cluster.

	BIOGENIC CLUSTER 1	BIOGENIC CLUSTER 2	ANTHROPOGENIC CLUSTER
CHO	409 (33.4%)	253 (37.9 %)	158 (15.4 %)
CHOS	248 (20.2 %)	148 (22.2 %)	313 (30.6 %)
CHON	295 (24.1 %)	134 (20.1 %)	181 (17.7 %)
OTHERS	276 (22.3 %)	133 (19.9 %)	371 (36.3 %)
TOTAL	1226 (100 %)	668 (100 %)	1023 (100 %)

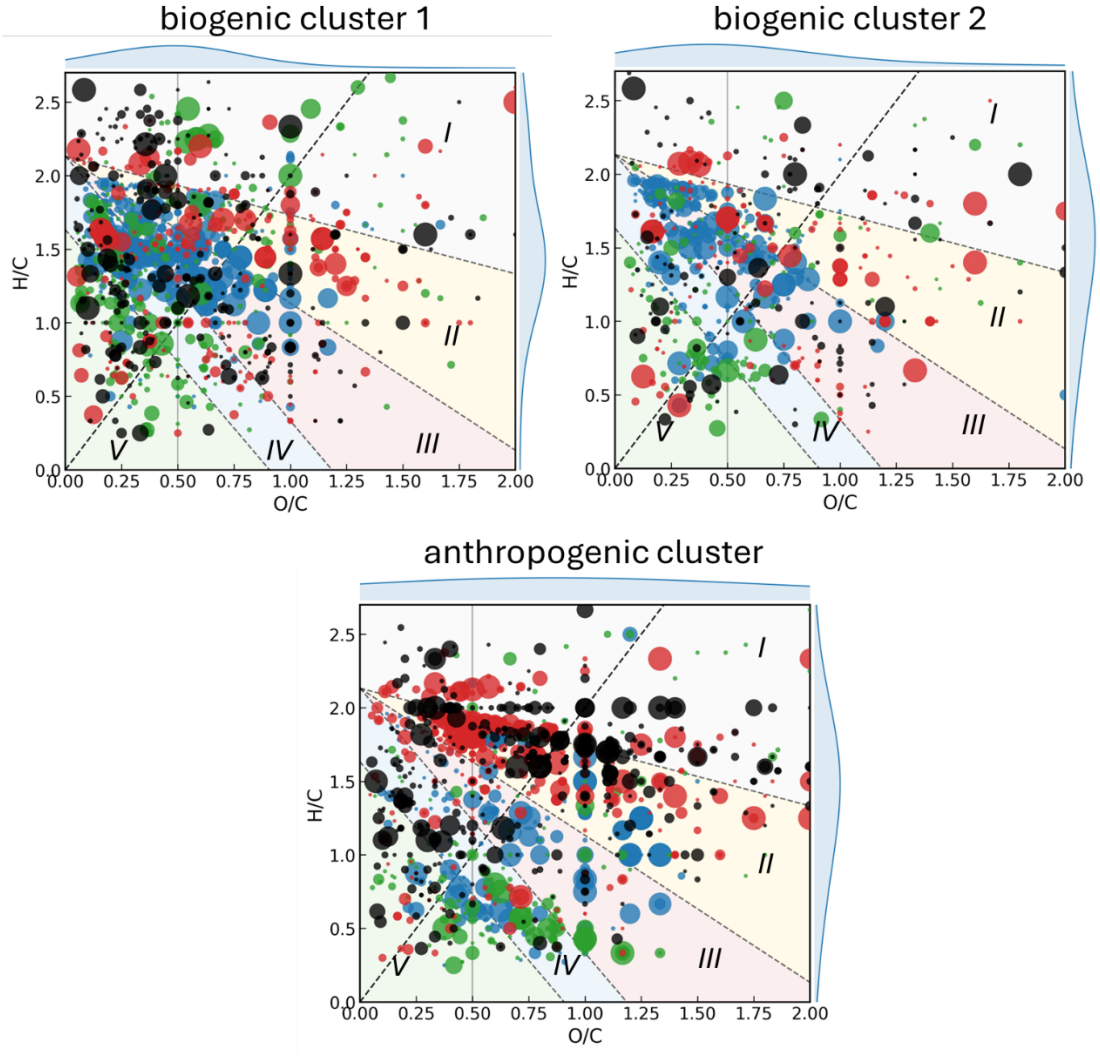

Figure 5: Van-Krevelen plots of the compounds assigned to the biogenic (left) and the anthropogenic (right) cluster. The color of the individual datapoints represents the subset of the corresponding chemical formula (blue: CHO, red: CHOS, green: CHON, black: others). The 5 shaded areas correspond to very highly oxidized compounds (I, light blue), highly oxidized compounds (II,

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

yellow), intermediately oxidized compounds (III, red), oxidized unsaturated compounds (IV, blue), and highly unsaturated compounds (V, green). The theory behind those shaded areas was taken from (Zhang et al., 2021). It can be seen that the biogenic cluster shows an accumulation of compounds of the CHO subset between 1.2 ≤ H/C ≤ 1.7 and 0.15 ≤ O/C ≤ 0.5 corresponding to oxidization products of terpenes (H/C = 1.6). The anthropogenic cluster shows an accumulation of CHOS compounds at significantly higher H/C ratios and is missing the accumulation of CHO compounds that is present in the biogenic cluster. Despite the significantly lower amount of filter samples analyzed in our dataset, the cluster compositions look similar to the cluster compositions presented recently in (Thoma et al., 2025).

3.3.1 Bootstrapping

Bootstrapping was performed to assess the repeatability of the clustering results. The analysis showed that the clustering was highly robust, with only a small number of compounds exhibiting an alternative cluster assignment in some bootstrap iterations (Figure 6). Most compounds were assigned to the same cluster in every iteration, indicating a high degree of cluster stability. Inspection of the clustering output further suggested that clusters 0 and 1 predominantly represent biogenic compounds, whereas cluster 2 represents anthropogenic compounds. This interpretation is supported by the Van Krevelen plots, in which clusters 0 and 1 show similar compositional patterns, while cluster 2 exhibits a clearly distinct distribution (Figure 6). To investigate why the algorithm separated the biogenic compounds into two clusters, the directional dominance was examined for clusters 0 and 1. Here, directional dominance is defined as the contribution of the dominant wind direction bin to the sum of contributions across all wind direction bins over the entire campaign. The histograms shown in Figure 7 indicate that compounds in cluster 0 generally exhibit higher directional dominance than those in cluster 1. This suggests that the separation into two biogenic clusters is driven, at least in part, by differences in the strength of directional structure within the signal. To further assess the stability of the anthropogenic cluster, which is central to the identification of anthropogenic marker compounds in this study, KMeans clustering was repeated with the number of clusters reduced from three to two, while all other preprocessing steps and settings were kept unchanged. Under these conditions, the anthropogenic cluster remained identical to that obtained in the three cluster solution. Together, the bootstrap analysis and the two cluster control experiment demonstrate that the identification of the anthropogenic cluster is highly robust. The high stability of the bootstrap assignments, the pronounced directional dominance observed for compounds in the anthropogenic cluster, and the identical recovery of this cluster in the two cluster analysis support a high level of confidence in the anthropogenic marker compounds identified in this study.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

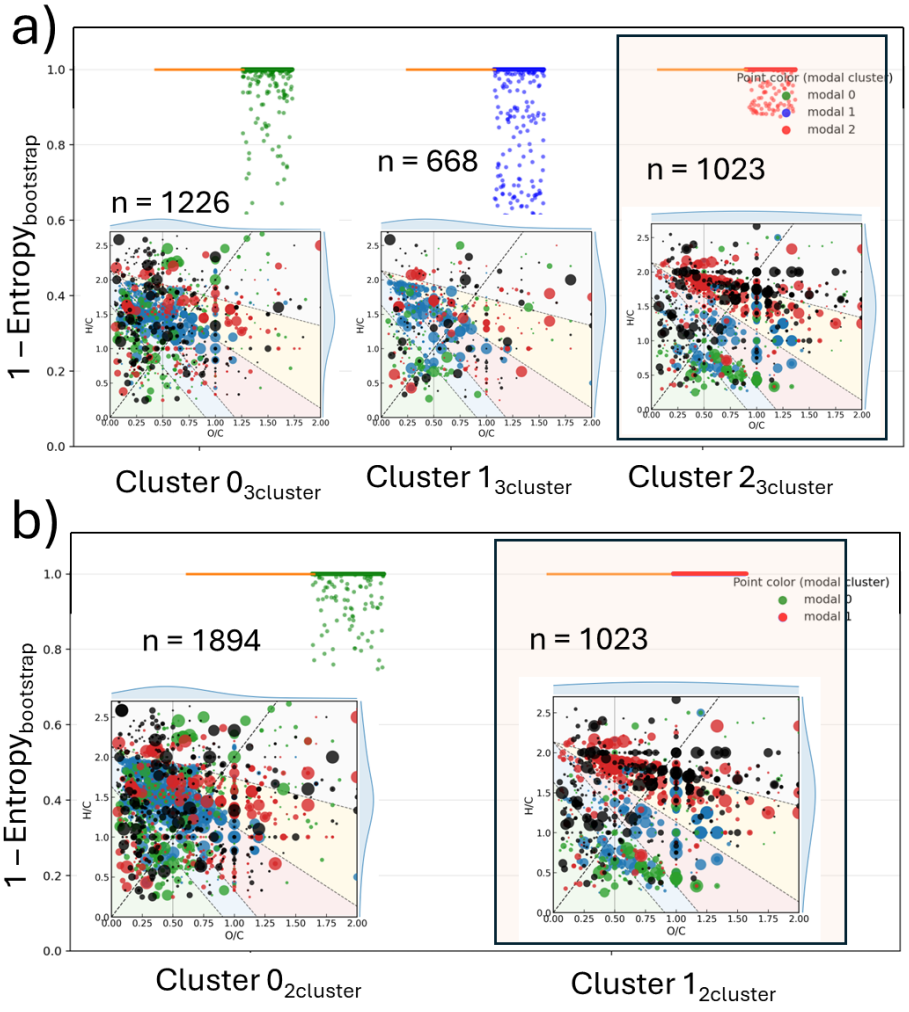

Figure 6: Bootstrapping results of the KMeans clustering for a clustering experiment with three clusters (a, top), and two clusters (b, bottom). The anthropogenic cluster, highlighted in red, has the same composition for the experiment with two and three clusters, highlighting the reliability and repeatability of the prediction of anthropogenic compounds.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

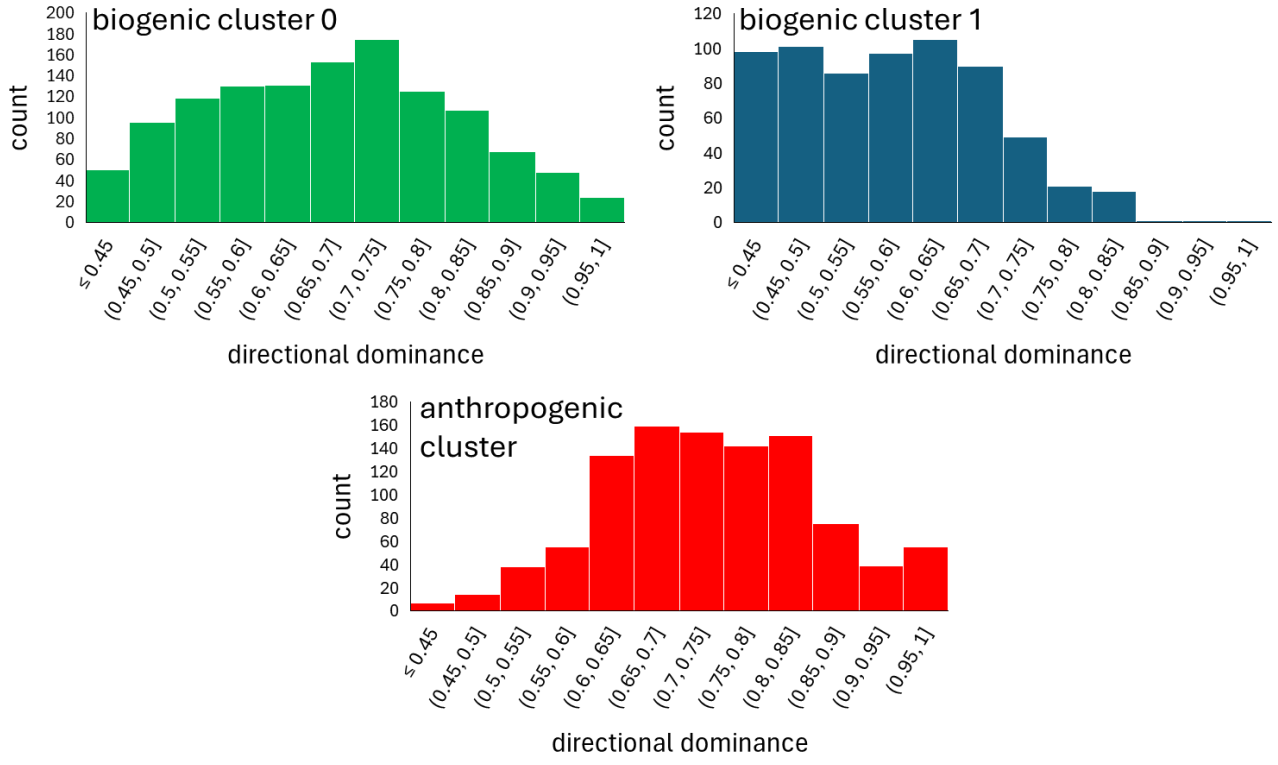

Figure 7: Directional dominance of the compounds in the individual clusters. The directional dominance is defined as the contribution of the dominant wind direction bin to the sum of contributions across all wind direction bins over the whole duration of the campaign. It can be seen that biogenic cluster 0 and the anthropogenic cluster show a high directional dominance of one wind direction, indicating a highly localized source of the compounds in those two clusters. The biogenic cluster 1 on the other hand does not show such a strong directional dominance, suggesting that the compounds in this cluster have multiple sources all around the sampling site.

3.3.2 Anthropogenic compounds

A closer look at the clustered data has revealed interesting compounds that were all assigned to the anthropogenic cluster. The compounds are $C_9H_{16}NO_8S^-$, $C_{10}H_{18}NO_8S^-$, and $C_{11}H_{20}NO_8S^-$, and likely related to each other, as can be seen by the molecular formula (all spaced by one CH_2 group), and as can be seen by their similar concentration profile over the duration of the campaign (see Figure 8). The concentration of those compounds is strongly increased during anthropogenically influenced time periods, which validates the automatic clustering results. Fragmentation data of the compounds (specifically fragmentation data for $C_{10}H_{18}NO_8S^-$) reveals that the compound (and therefore likely the whole compound class) carries a nitro group which is shown by the neutral loss of HNO_3 (with measured fragment: $C_{10}H_{17}O_5S_1^-$) (Surratt et al., 2007). An HSO_3^- fragment was also found, which is indicative of a sulfonic acid (Levsen et al., 2007). Lastly, and probably most indicative of the anthropogenic nature of the compounds, a fragment with formula $C_7H_7O^-$ was found, which, by looking at

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

possible structure candidates for this sum formula, consists most likely of an aromatic ring. Cresol does possess this formula, and consists of an aromatic ring with an hydroxy, and an additional methyl group. The close relationship to phenol explains the acidity, which in turn explains why we could see the fragment in ESI negative. Aromatic compounds are largely considered to be of anthropogenic origin with motor vehicles / industrial processes / fuel evaporation being the primary emitter of volatile compounds like BTEX (benzene, toluene, ethylbenzene and xylene) (Liu et al., 2025; Le Bras et al., 2026), incomplete combustion processes of fossil fuels and biomass burning being an emitter of PAHs (Ravindra et al., 2008; Drotikova et al., 2020). Both of the compound classes undergo oxidation and transformation in the atmosphere, leading to various reaction products (Zimmermann et al., 2013; Cheng et al., 2021; Drotikova et al., 2021; Priestley et al., 2021).

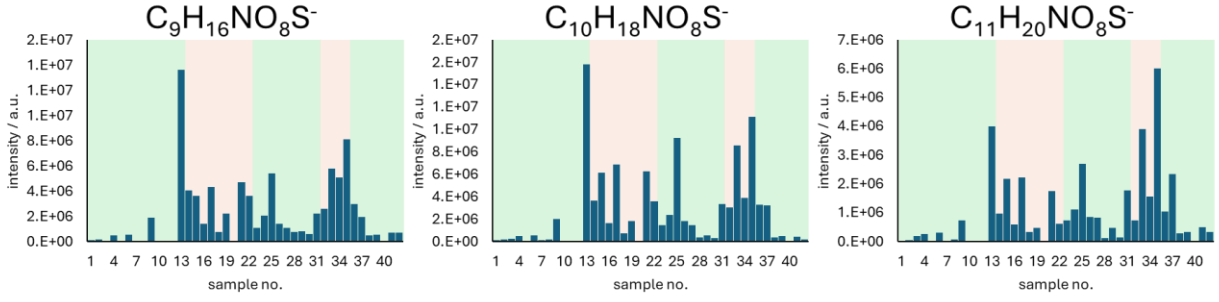

Figure 8: Measured intensities for the three compounds presented. The compounds belong to the "others" subset (black) in the provided Van Krevelen plots. Due to very similar intensity patterns all along the ACROSS campaign, a relationship can be assumed between these compounds. For all of them, the intensity is elevated during the anthropogenic influenced time periods, and the elevated intensity remains for approximately 1 day after the biogenic period has already started.

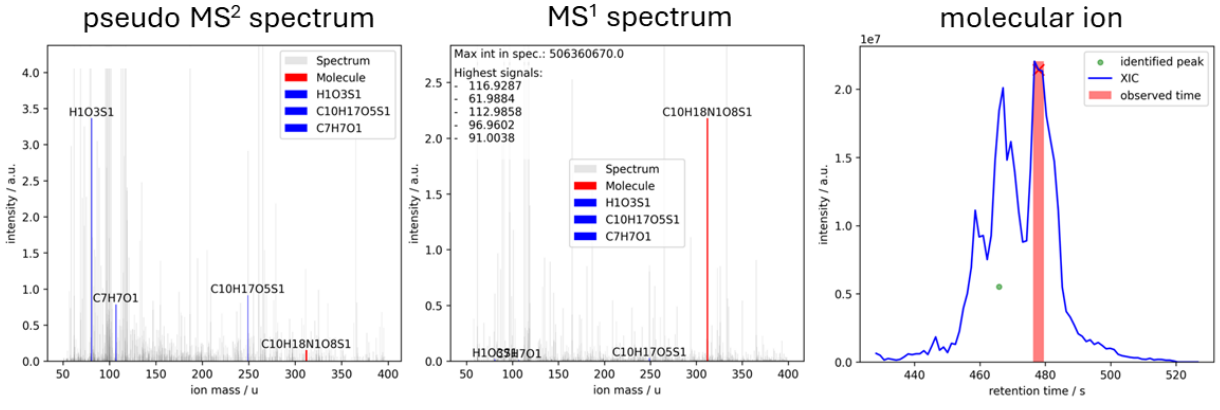

Figure 9: MS results of $C_{10}H_{18}NO_8S^-$ as analyzed in the highest intensity sample S13. The left plot shows the pseudo MS^2 spectrum of the compound (@ rt = 479 s) which was extracted from the available AIF spectrum. The MS spectrum in the middle shows the MS^1 spectrum. The plot on the right shows the XIC for m/z = 312.0758.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

3.3.3 Biogenic compounds

As an example, a compound with the formula $C_{11}H_{17}O_6^-$, that was assigned to the biogenic cluster was investigated in deeper detail. In contrast to the compounds presented in chapter 3.3.2, this compound does show the highest concentration during the last two biogenically influenced time periods and is significantly lower in concentration during anthropogenic periods. Strangely, the concentration of this compound is also low during the first biogenically influenced time period. This can be explained by the different meteorological conditions during this period, as is stated in (Di Antonio et al., 2025). There, this period is defined as a "clean period" with lower temperatures compared to the two following biogenically influenced time periods. The chemical formula has been of interest in various other SOA studies (Kundu et al., 2012; Hamilton et al., 2013; Inomata et al., 2025; Cai et al., 2026). As can be seen in Figure 10, various C10 fragments were found, suggesting monoterpenes as a potential precursor of this compound, and therefore also suggesting its biogenic origin, being in accordance with (Kundu et al., 2012; Inomata et al., 2025; Cai et al., 2026).

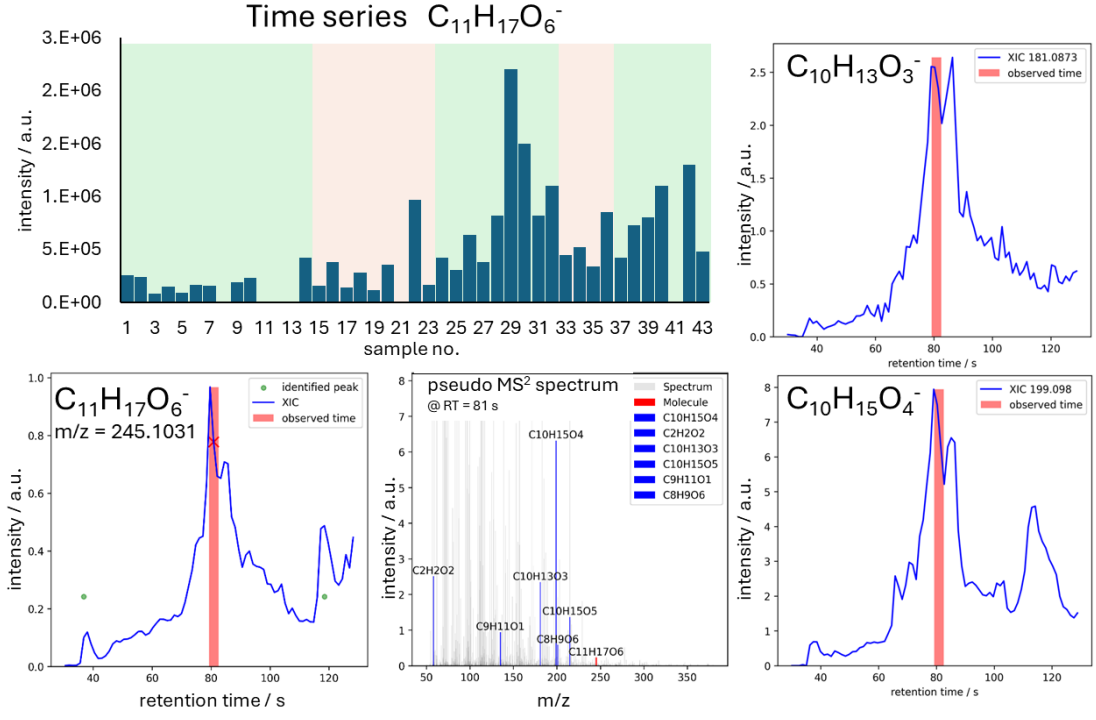

Figure 10: Top-left: measured intensities for m/z = 245.1031 (± 10 ppm) during the ACROSS campaign. The green shaded areas show the biogenically influenced time periods and the red shaded areas show the anthropogenic ones. As can be seen, the intensity of the compound is significantly higher during the second and third biogenic period, which might be explained by the lower temperature in the first biogenic period (see the meteorological information provided in (Di Antonio et al., 2025)). Lower left: XIC of the target compound. Lower middle: pseudo MS^2 spectrum of the target compound. As can be seen, three C10 compounds are created in the fragmentation experiments, suggesting a monoterpene as a potential atmospheric precursor molecule. Left side: XICs of two example fragments with a C10 formula. As can be seen, the XICs of the fragment masses, that were taken from the AIF spectra, match with the peak-shape of the precursor ion.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

3.3.4 Calculation of anthropogenic influence on the chemical content of a given aerosol sample.

Once obtained a set of compounds of anthropogenic origin, the anthropogenic influence of unknown organic aerosol samples can now be assessed just by calculating the relative contribution of "anthropogenic" compounds on the whole sample. Figure 11 shows the contribution of the different clusters to the total composition of 4 different samples. As expected, the samples taken during NE wind, generally do show a larger contribution of anthropogenic compounds as samples taken during other wind directions. As explained in chapter 3.3 a validation clustering run has been conducted where samples S6, S21, S22, and S38 were removed from the training dataset. Next to the complete clustering shown in Figure 11, this "reduced" clustering attempt was used to also calculate the anthropogenic influence of the 4 samples. For this reduced clustering attempt, it is important to stress that the algorithm has never seen the 4 individual samples, therefore representing actual new data. The calculated cluster contributions for this "reduced" clustering attempt are strikingly similar to the original cluster contributions calculated with all available samples, clearly showing the applicability to determine cluster contributions of truly unknown organic aerosol samples. The plots of the cluster contributions can be found in the Supplemental Information.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

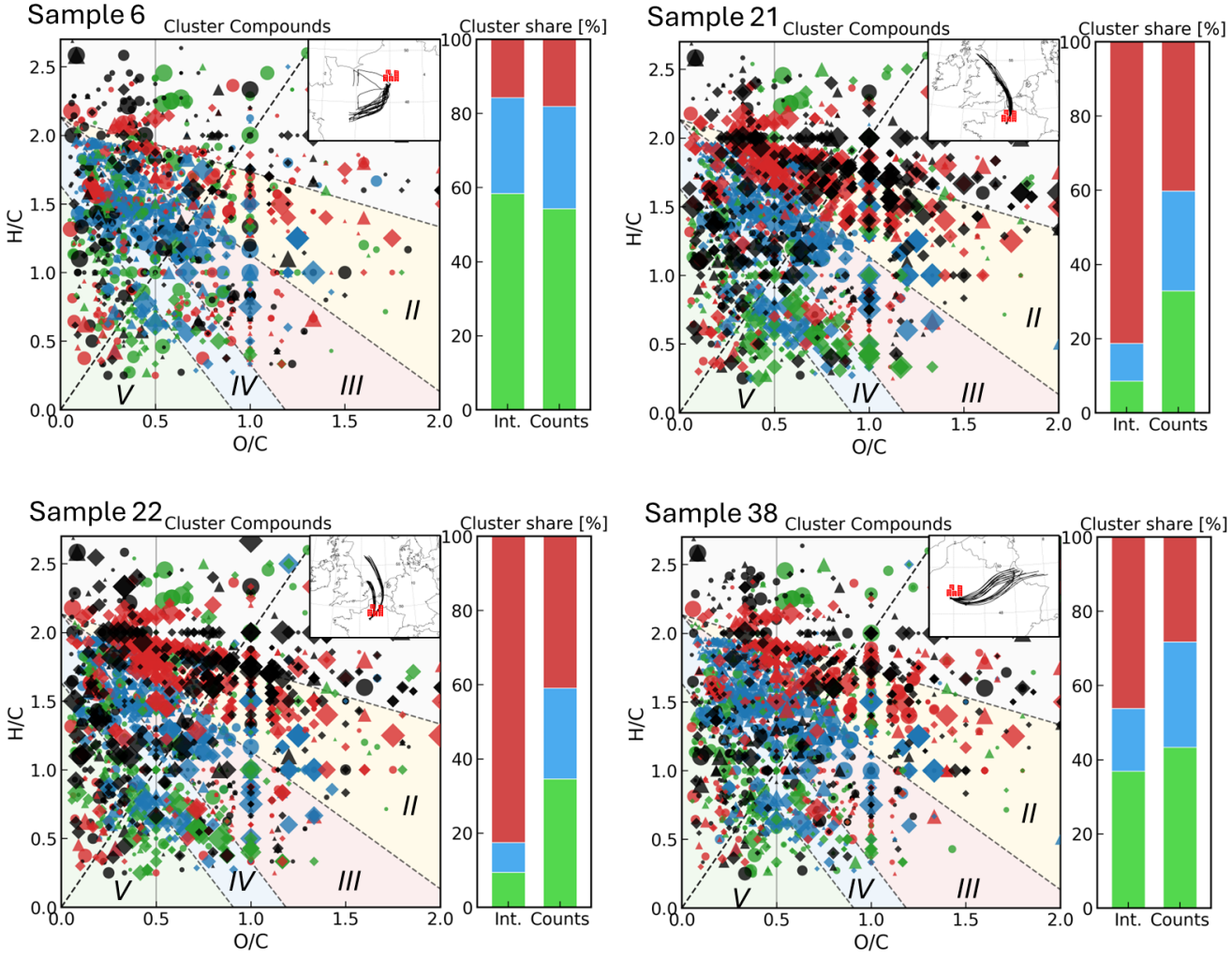

Figure 11: Van-Krevelen plots for 4 different samples (top left: S6, top right: S21, bottom left: S22, bottom right: S38) collected during the ACROSS campaign. On the right a typical Van-Krevelen plot is shown with 5 shaded areas corresponding to very highly oxidized compounds (I, light blue), highly oxidized compounds (II, yellow), intermediately oxidized compounds (III, red), oxidized unsaturated compounds (IV, blue), and highly unsaturated compounds (V, green). The theory behind those shaded areas was taken from (Zhang et al., 2021). The size of the individual datapoints corresponds to the measured intensity, the form of the datapoints corresponds to the identified cluster (round: biogenic, square: anthropogenic, triangular: unidentified). The color of the datapoints corresponds to the subset of the chemical formula (blue: CHO, red: CHOS, green: CHON, black: others). To enhance visibility, the plot is zoomed in on the majority of compounds with borders $0 \leq H/C \leq 2.7$; $0 \leq O/C \leq 1.5$. In the top right corner of the Van-Krevelen plots the HYSPLIT trajectory of the 24 h before the sample was taken is shown. As can clearly be seen, for sample S21 and S22, the back trajectories pass over the greater Paris area (red marker in the plot), therefore explaining the elevated anthropogenic influence of the air masses collected in those samples. Next to the Van-Krevelen plot the cluster share of the compounds that make up the corresponding sample is shown (left side of cluster share: intensity distribution between different clusters; right side of cluster share: number distribution between different cluster shares). Here the elevated anthropogenic influence for sample S21 and S22 can be seen as well (anthropogenic cluster: red).

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

4 Conclusions

This study successfully demonstrates an approach for determining anthropogenic influence on the chemical composition of mixed organic aerosol samples through wind-direction-constrained non-target analysis. In contrast to previous studies, the number of samples required could be significantly reduced to only 37 samples while still providing stable results. By applying K-means clustering (k=3) to 2917 compounds with strong wind correlation from the ACROSS 2022 campaign dataset, we resolved distinct biogenic and anthropogenic molecular signatures with high chemical specificity. The biogenic cluster was dominated by CHO compounds (H/C: 1.2–1.7; O/C: 0.15–0.5), consistent with oxidized terpene products, while the anthropogenic cluster featured significant CHOS enrichment (H/C: 1.5–2.2; O/C: 0.2–1.0), including nitrogen-containing aromatic compounds such as $C_{10}H_{18}NO_8S^-$ with diagnostic fragments indicating nitro groups (HNO_3 loss), sulfonic acid moieties (HSO_3^-), and aromatic structures ($C_7H_7O^-$).

Our methodological contribution in this work lies in the development of a robust wind-correlation filtering approach that enables reliable source apportionment even with relatively small sample sizes (n=37), addressing a critical limitation in field studies where extensive sampling is often impractical, time consuming and expensive. A key strength of this approach is its transferability to other regions without requiring extensive prior knowledge of local emission profiles. The clustering results allow evaluating anthropogenic contribution, for a given sample, allowing e.g. to quickly identify new polluters within the trajectory of the air mass, or detect anomalies and changes in anthropogenic emissions. This might help to evaluate the effectiveness of measures that should protect the environment.

5 Data availability

The clustering results can be found in the supplement. The raw data and code is available upon request from the corresponding author.

6 Author contributions

NK conceptualized the analysis idea, extracted the filter samples used in this study, developed the analysis method, and performed the analysis. PP, EP and EV planned and conducted the filter sampling during the ACROSS campaign. All authors contributed to the discussion of the results and participated in finalizing the manuscript.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

7 Competing interests

The authors do not have to declare any competing interests.

8 Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

9 Acknowledgements

The authors would like to gratefully acknowledge Vincent Michoud and Christopher Cantrell (both from LISA at Créteil (France)) as co-coordinators of the ACROSS 2022 campaign.

10 Financial support

Open Access funding enabled and organized by Project DEAL. This research has been supported by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG: HO 1748/19-1; HO 1748/23-1; TRR 301 – Project-ID 428312742; BMBF Project ATTO+, 01LK1602D). The ACROSS project received funding from the French National Research Agency (ANR) under the program ANR-17-MPGA-0002 and was also supported by the French National program LEFE of CNRSINSU.

11 References

Brüggemann, M., Xu, R., Tilgner, A., Kwong, K. C., Mutzel, A., Poon, H. Y., Otto, T., Schaefer, T., Poulain, L., Chan, M. N., and Herrmann, H.: Organosulfates in Ambient Aerosol: State of Knowledge and Future Research Directions on Formation, Abundance, Fate, and Importance, Environmental science & technology, 54, 3767–3782, https://doi.org/10.1021/acs.est.9b06751, 2020.

Cai, M., Yuan, B., Hu, W., Chenshuo, Y., Huang, S., Yang, S., Chen, W., Peng, Y., Deng, Z., Zhao, J., Chen, D., Sun, J., and Shao, M.: New insight into the formation and aging processes of organic aerosol from positive matrix factorization (PMF) analysis of ambient FIGAERO-CIMS thermograms, Atmos. Chem. Phys., 26, 769–788, https://doi.org/10.5194/acp-26-769-2026, 2026.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

Cheng, X., Chen, Q., Jie Li, Y., Zheng, Y., Liao, K., and Huang, G.: Highly oxygenated organic molecules produced by the oxidation of benzene and toluene in a wide range of OH exposure and NO x conditions, Atmos. Chem. Phys., 21, 12005–12019, https://doi.org/10.5194/acp-21-12005-2021, 2021.

Di Antonio, L., Beekmann, M., Siour, G., Michoud, V., Cantrell, C., Bauville, A., Bergé, A., Cazaunau, M., Chevaillier, S., Cirtog, M., Brito, J. F. de, Formenti, P., Gaimoz, C., Garret, O., Gratien, A., Gros, V., Haeffelin, M., Hawkins, L. N., Kotthaus, S., Noyalet, G., Pereira, D. L., Petit, J.-E., Pronovost, E. D., Riffault, V., Yu, C., Foret, G., Doussin, J.-F., and Di Biagio, C.: Modelling of atmospheric variability in gas and aerosols during the ACROSS campaign 2022 of the greater Paris area: evaluation of the meteorology, dynamics and chemistry, Atmos. Chem. Phys., 25, 4803–4831, https://doi.org/10.5194/acp-25-4803-2025, 2025.

Drotikova, T., Dekhtyareva, A., Kallenborn, R., and Albinet, A.: Polycyclic aromatic hydrocarbons (PAHs) and their nitrated and oxygenated derivatives in the Arctic boundary layer: seasonal trends and local anthropogenic influence, Atmos. Chem. Phys., 21, 14351–14370, https://doi.org/10.5194/acp-21-14351-2021, 2021.

Drotikova, T., Ali, A. M., Halse, A. K., Reinardy, H. C., and Kallenborn, R.: Polycyclic aromatic hydrocarbons (PAHs) and oxy- and nitro-PAHs in ambient air of the Arctic town Longyearbyen, Svalbard, Atmos. Chem. Phys., 20, 9997–10014, https://doi.org/10.5194/acp-20-9997-2020, 2020.

Giorio, C., Bortolini, C., Kourtchev, I., Tapparo, A., Bogialli, S., and Kalberer, M.: Direct target and non-target analysis of urban aerosol sample extracts using atmospheric pressure photoionisation high-resolution mass spectrometry, Chemosphere, 224, 786–795, https://doi.org/10.1016/j.chemosphere.2019.02.151, 2019.

Glasius, M., Thomsen, D., Wang, K., Iversen, L. S., Duan, J., and Huang, R.-J.: Chemical characteristics and sources of organosulfates, organosulfonates, and carboxylic acids in aerosols in urban Xi'an, Northwest China, The Science of the total environment, 810, 151187, https://doi.org/10.1016/j.scitotenv.2021.151187, 2022.

Gouw, J. de and Jimenez, J. L.: Organic aerosols in the Earth's atmosphere, Environmental science & technology, 43, 7614–7618, https://doi.org/10.1021/es9006004, 2009.

Hamilton, J. F., Alfarra, M. R., Robinson, N., Ward, M. W., Lewis, A. C., McFiggans, G. B., Coe, H., and Allan, J. D.: Linking biogenic hydrocarbons to biogenic aerosol in the Borneo rainforest, Atmos. Chem. Phys., 13, 11295–11305, https://doi.org/10.5194/acp-13-11295-2013, 2013.

Heald, C. L., Kroll, J. H., Jimenez, J. L., Docherty, K. S., DeCarlo, P. F., Aiken, A. C., Chen, Q., Martin, S. T., Farmer, D. K., and Artaxo, P.: A simplified description of the evolution of organic aerosol composition in the atmosphere, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL042737, 2010.

Hildmann, S. and Hoffmann, T.: Characterisation of atmospheric organic aerosols with one- and multidimensional liquid chromatography and mass spectrometry: State of the art and future perspectives, TrAC Trends in Analytical Chemistry, 175, 117698, https://doi.org/10.1016/j.trac.2024.117698, 2024.

Inomata, S., Fukuyama, D., and Sekimoto, K.: Chemical analysis of new particles with diameters less than 10 nm, generated from β-pinene ozonolysis, Chemistry Letters, 54, https://doi.org/10.1093/chemle/upaf061, 2025.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

Ishikawa, M., Konno, R., Nakajima, D., Gotoh, M., Fukasawa, K., Sato, H., Nakamura, R., Ohara, O., and Kawashima, Y.: Optimization of Ultrafast Proteomics Using an LC-Quadrupole-Orbitrap Mass Spectrometer with Data-Independent Acquisition, Journal of proteome research, 21, 2085–2093, https://doi.org/10.1021/acs.jproteome.2c00121, 2022.

Karbach, N. and Hoffmann, T.: CID fragment annotation from data-independent experiments in non-target organic aerosol analysis: presenting an easy-to-use tool, Analytical and bioanalytical chemistry, 417, 6961–6971, https://doi.org/10.1007/s00216-025-06184-5, 2025.

Kim, S., Kramer, R. W., and Hatcher, P. G.: Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram, Analytical chemistry, 75, 5336–5344, https://doi.org/10.1021/ac034415p, 2003.

Krasny, L. and Huang, P. H.: Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology, Molecular omics, 17, 29–42, https://doi.org/10.1039/d0mo00072h, 2021.

Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E., and Worsnop, D. R.: Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, Nature chemistry, 3, 133–139, https://doi.org/10.1038/nchem.948, 2011.

Kundu, S., Fisseha, R., Putman, A. L., Rahn, T. A., and Mazzoleni, L. R.: High molecular weight SOA formation during limonene ozonolysis: insights from ultrahigh-resolution FT-ICR mass spectrometry characterization, Atmos. Chem. Phys., 12, 5523–5536, https://doi.org/10.5194/acp-12-5523-2012, 2012.

Laskin, A., Laskin, J., and Nizkorodov, S. A.: Mass spectrometric approaches for chemical characterisation of atmospheric aerosols: critical review of the most recent advances, Environmental Chemistry, 9, 163–189, https://doi.org/10.1071/EN12052, 2012.

Le Bras, Z., Rubli, P., Hueglin, C., and Reimann, S.: Measurement report: 30 years of BTEX monitoring at a suburban site in Switzerland supported by additional urban VOC observations, Atmos. Chem. Phys., 26, 869–878, https://doi.org/10.5194/acp-26-869-2026, 2026.

Leppla, D., Hildmann, S., Zannoni, N., Kremper, L. A., Holanda, B. A., Williams, J., Pöhlker, C., Wolff, S., Sà, M., Solci, M. C., Pöschl, U., and Hoffmann, T.: Comprehensive non-targeted molecular characterization of organic aerosols in the Amazon rainforest, Atmos. Chem. Phys., 26, 365–390, https://doi.org/10.5194/acp-26-365-2026, 2026.

Levsen, K., Schiebel, H.-M., Terlouw, J. K., Jobst, K. J., Elend, M., Preiss, A., Thiele, H., and Ingendoh, A.: Even-electron ions: a systematic study of the neutral species lost in the dissociation of quasi-molecular ions, Journal of mass spectrometry JMS, 42, 1024–1044, https://doi.org/10.1002/jms.1234, 2007.

Liu, X., Zhang, X., Dufresne, M., Wang, T., Wu, L., Lara, R., Seco, R., Monge, M., Yáñez-Serrano, A. M., Gohy, M., Petit, P., Chevalier, A., Vagnot, M.-P., Fortier, Y., Baudic, A., Ghersi, V., Gille, G., Lanzi, L., Gros, V., Simon, L., Héllen, H., Reimann, S., Le Bras, Z., Müller, M. J., Beddows, D., Hou, S., Shi, Z., Harrison, R. M., Bloss, W., Dernie, J., Sauvage, S., Hopke, P. K., Duan, X., An, T., Lewis, A. C., Hopkins, J. R., Liakakou, E., Mihalopoulos, N., Zhang, X., Alastuey,

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

A., Querol, X., and Salameh, T.: Measurement report: Exploring the variations in ambient BTEX in urban Europe and their environmental health implications, Atmos. Chem. Phys., 25, 625–638, https://doi.org/10.5194/acp-25-625-2025, 2025.

Ma, J., Reininger, N., Zhao, C., Döbler, D., Rüdiger, J., Qiu, Y., Ungeheuer, F., Simon, M., D'Angelo, L., Breuninger, A., David, J., Bai, Y., Li, Y., Xue, Y., Li, L., Wang, Y., Hildmann, S., Hoffmann, T., Liu, B., Niu, H., Wu, Z., and Vogel, A. L.: Unveiling a large fraction of hidden organosulfates in ambient organic aerosol, Nature communications, 16, 4098, https://doi.org/10.1038/s41467-025-59420-y, 2025.

Mauderly, J. L. and Chow, J. C.: Health effects of organic aerosols, Inhalation toxicology, 20, 257–288, https://doi.org/10.1080/08958370701866008, 2008.

Murphy, D. M., Cziczo, D. J., Froyd, K. D., Hudson, P. K., Matthew, B. M., Middlebrook, A. M., Peltier, R. E., Sullivan, A., Thomson, D. S., and Weber, R. J.: Single-particle mass spectrometry of tropospheric aerosol particles, JGR Atmospheres, 111, https://doi.org/10.1029/2006JD007340, 2006.

Nel, A.: Atmosphere. Air pollution-related illness: effects of particles, Science (New York, N.Y.), 308, 804–806, https://doi.org/10.1126/science.1108752, 2005.

Pereira, D. L., Gratien, A., Giorio, C., Mebold, E., Bertin, T., Gaimoz, C., Doussin, J.-F., and Formenti, P.: Two optimized methods for the quantification of anthropogenic and biogenic markers in aerosol samples using liquid chromatography mass spectrometry and gas chromatography mass spectrometry, Atmos. Meas. Tech., 19, 1–19, https://doi.org/10.5194/amt-19-1-2026, 2026.

Pereira, D. L., Giorio, C., Gratien, A., Zherebker, A., Noyalet, G., Chevaillier, S., Alage, S., Almarj, E., Bergé, A., Bertin, T., Cazaunau, M., Coll, P., Di Antonio, L., Harb, S., Heuser, J., Gaimoz, C., Guillemant, O., Language, B., Lauret, O., Macias, C., Maisonneuve, F., Picquet-Varrault, B., Torres, R., Triquet, S., Zapf, P., Hawkins, L., Pronovost, D., Riley, S., Flaud, P.-M., Perraudin, E., Pouyes, P., Villenave, E., Albinet, A., Favez, O., Aujay-Plouzeau, R., Michoud, V., Cantrell, C., Cirtog, M., Di Biagio, C., Doussin, J.-F., and Formenti, P.: Molecular characterization of organic aerosols in urban and forested areas of Paris using high-resolution mass spectrometry, Atmos. Chem. Phys., 25, 4885–4905, https://doi.org/10.5194/acp-25-4885-2025, 2025.

Pope, C. A., Ezzati, M., and Dockery, D. W.: Fine-particulate air pollution and life expectancy in the United States, The New England journal of medicine, 360, 376–386, https://doi.org/10.1056/NEJMsa0805646, 2009.

Priestley, M., Bannan, T. J., Le Breton, M., Worrall, S. D., Kang, S., Pullinen, I., Schmitt, S., Tillmann, R., Kleist, E., Zhao, D., Wildt, J., Garmash, O., Mehra, A., Bacak, A., Shallcross, D. E., Kiendler-Scharr, A., Hallquist, Å. M., Ehn, M., Coe, H., Percival, C. J., Hallquist, M., Mentel, T. F., and McFiggans, G.: Chemical characterisation of benzene oxidation products under high- and low-NO x conditions using chemical ionisation mass spectrometry, Atmos. Chem. Phys., 21, 3473–3490, https://doi.org/10.5194/acp-21-3473-2021, 2021.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

Ravindra, K., Sokhi, R., and van Grieken, R.: Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation, Atmospheric Environment, 42, 2895–2921, https://doi.org/10.1016/j.atmosenv.2007.12.010, 2008.

Rincón, A. G., Calvo, A. I., Dietzel, M., and Kalberer, M.: Seasonal differences of urban organic aerosol composition – an ultra-high resolution mass spectrometry study, Environmental Chemistry, 9, 298–319, https://doi.org/10.1071/EN12016, 2012.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System, Bulletin of the American Meteorological Society, 96, 2059–2077, https://doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Surratt, J. D., Kroll, J. H., Kleindienst, T. E., Edney, E. O., Claeys, M., Sorooshian, A., Ng, N. L., Offenberg, J. H., Lewandowski, M., Jaoui, M., Flagan, R. C., and Seinfeld, J. H.: Evidence for organosulfates in secondary organic aerosol, Environmental science & technology, 41, 517–527, https://doi.org/10.1021/es062081q, 2007.

Thoma, M., Bachmeier, F., Knauf, K., David, J., Simon, M., and Vogel, A. L.: Seasonal analysis of organic aerosol composition resolves anthropogenic and biogenic sources at a rural background station in central Europe, Environ. Sci.: Atmos., 5, 703–713, https://doi.org/10.1039/d4ea00163j, 2025.

Tsou, C.-C., Tsai, C.-F., Teo, G. C., Chen, Y.-J., and Nesvizhskii, A. I.: Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers, Proteomics, 16, 2257–2271, https://doi.org/10.1002/pmic.201500526, 2016.

Vogel, A. L., Lauer, A., Fang, L., Arturi, K., Bachmeier, F., Daellenbach, K. R., Käser, T., Vlachou, A., Pospisilova, V., Baltensperger, U., Haddad, I. E., Schwikowski, M., and Bjelić, S.: A Comprehensive Nontarget Analysis for the Molecular Reconstruction of Organic Aerosol Composition from Glacier Ice Cores, Environmental science & technology, 53, 12565–12575, https://doi.org/10.1021/acs.est.9b03091, 2019.

Wang, X., Hayeck, N., Brüggemann, M., Yao, L., Chen, H., Zhang, C., Emmelin, C., Chen, J., George, C., and Wang, L.: Chemical Characteristics of Organic Aerosols in Shanghai: A Study by Ultrahigh-Performance Liquid Chromatography Coupled With Orbitrap Mass Spectrometry, JGR Atmospheres, 122, https://doi.org/10.1002/2017JD026930, 2017.

Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Allan, J. D., Coe, H., Ulbrich, I., Alfarra, M. R., Takami, A., Middlebrook, A. M., Sun, Y. L., Dzepina, K., Dunlea, E., Docherty, K., DeCarlo, P. F., Salcedo, D., Onasch, T., Jayne, J. T., Miyoshi, T., Shimono, A., Hatakeyama, S., Takegawa, N., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Williams, P., Bower, K., Bahreini, R., Cottrell, L., Griffin, R. J., Rautiainen, J., Sun, J. Y., Zhang, Y. M., and Worsnop, D. R.: Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced Northern Hemisphere midlatitudes, Geophysical Research Letters, 34, https://doi.org/10.1029/2007GL029979, 2007.

Zhang, Y., Wang, K., Tong, H., Huang, R.-J., and Hoffmann, T.: The maximum carbonyl ratio (MCR) as a new index for the structural classification of secondary organic aerosol components, Rapid communications in mass spectrometry RCM, 35, e9113, https://doi.org/10.1002/rcm.9113, 2021.

https://doi.org/10.5194/egusphere-2026-2473
Preprint. Discussion started: 13 May 2026

Zimmermann, K., Jariyasopit, N., Massey Simonich, S. L., Tao, S., Atkinson, R., and Arey, J.: Formation of nitro-PAHs from the heterogeneous reaction of ambient particle-bound PAHs with N2O5/NO3/NO2, Environmental science & technology, 47, 8434–8442, https://doi.org/10.1021/es401789x, 2013.