



Supplement of

Unraveling Arctic submicron organic aerosol sources: a year-long study by H-NMR and AMS in Ny-Ålesund, Svalbard

Marco Paglione et al.

Correspondence to: Kaspar R. Daellenbach (kaspar.daellenbach@psi.ch) and Matteo Rinaldi (m.rinaldi@isac.cnr.it)

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

S.1 Field campaign overview: sampling and chemical analyses

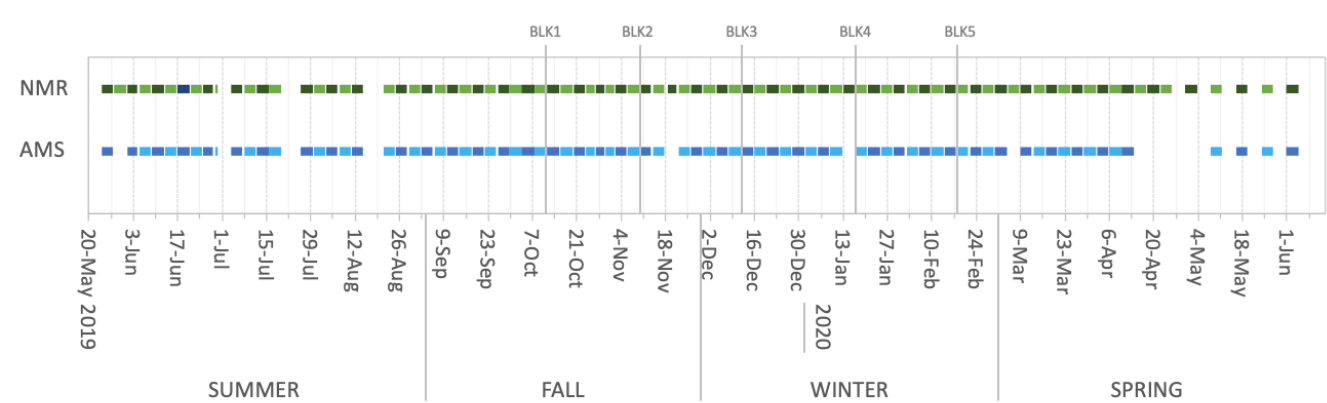


Figure S1. Sampling periods of the PM₁ filters collected at Gruvebadet, Ny-Alesund, and subsequently analyzed by H-NMR (dark-light green bars) and HR-TOF-AMS (dark-light blue bars). Field blank filters collection time is indicated by grey lines and names.

Table S1. Limit of Detection (LOD) for each species measured by Ion Chromatography (IC) and for water-soluble organic carbon (WSOC) by TOC analyzer. The LODs are calculated based on the average and standard deviation of the five field blanks analyzed. In particular $LOD = BLK_mean + 2 * BLK_std.dev$.

LOD	IC ($\mu\text{g/mL}$)																TOC-analyzer ($\mu\text{g/mL}$)
	<i>Na</i>	<i>NH4</i>	<i>ma</i>	<i>K</i>	<i>dma</i>	<i>tma</i>	<i>Mg</i>	<i>Ca</i>	<i>ace</i>	<i>for</i>	<i>MSA</i>	<i>Cl</i>	<i>NO2</i>	<i>NO3</i>	<i>SO4</i>	<i>oxa</i>	<i>WSOC</i>
	0.36	0.03	0.05	0.05	0.05	0.05	0.004	0.06	0.02	0.02	0.02	0.21	0.01	0.03	0.12	0.05	1.16

Table S2. H-NMR identified/measured functional groups/chemical species/categories. *Functional groups are in *italic*. **Categories including some of the other species specifically identified are in underlined italic

name of the species/ functional group*/ category of compounds**	ID of the species/ functional group	chemical shifts used for identification & quantification	examples for molecules	possible origin/source	references
<i>aromatic protons</i>	Ar-H	band 6.5-8.5 ppm	phenols, nitro-phenols [...]	biomass burning, [...]	Decesari et al., 2001; Tagliavini 2006; Decesari et al., 2007; Chalbot and Kavouras, 2014
<i>anomeric and/or vinyl protons</i>	O-CH-O	band 6-6.5 ppm	vinylc protons of not completely oxidized isoprene and terpenes derivatives, of products of aromatic-rings opening (e.g., maleic acid), or anomeric protons of sugars derivatives (glucose, sucrose, levoglucosan, glucuronic acid, etc.)	biogenic marine mostly primary	Decesari et al., 2001; Claeys et al. 2004; Schkolnik & Rudich, 2005; Tagliavini 2006; Decesari et al., 2007; Chalbot and Kavouras, 2014
<i>hydroxyl/alkoxy groups</i>	H-C-O	band 3.2-4.5 ppm	aliphatic alcohols, polyols, saccharides, ethers, and esters	biogenic marine primary	Chalbot and Kavouras, 2014
<i>benzyls and acyls/ amines, sulfonates</i>	H-C-C= / H-C-X (X≠O)	band 1.8-3.2 ppm	protons bound to aliphatic carbon atoms adjacent to unsaturated groups like alkenes (allylic protons), carbonyl or imino groups (heteroallylic protons) or aromatic rings (benzylic protons)	biogenic/anthropogenic mostly secondary	Decesari et al., 2001; Graham et al., 2002; Decesari et al., 2007; Chalbot and Kavouras, 2014
<i>unfunctionalized alkylc protons</i>	H-C	band 0.5-1.8 ppm	methyls (CH ₃), methylenes (CH ₂), and methynes (CH) groups of several possible molecules: fatty acids chains, alkylc portion of biogenic terpenes, etc.	biogenic/anthropogenic primary/secondary	Decesari et al., 2001; Graham et al., 2002; Decesari et al., 2007; Chalbot and Kavouras, 2014
hydroxymethansulfopnic acid	HMSA	singlet at 4.39 ppm		anthropogenic secondary	Suzuki et al., 2001; Gilardoni et al., 2016; Brege et al 2018
methane-sulfonate	MSA	singlet at 2.80 ppm		biogenic marine secondary	Suzuki et al., 2001; Facchini et al., 2008a; Decesari et al., 2020
di-methylamine	DMA	singlet at 2.72 ppm		biogenic marine secondary	Suzuki et al., 2001; Facchini et al., 2008a
tri-methylamine	TMA	singlet at 2.89 ppm		biogenic marine secondary	Suzuki et al., 2001; Facchini et al., 2008a
<u><i>anhydrosugars</i></u>		anomeric singlet between 5.40-5.45 ppm & specific structures between 3.5 and 4.6 ppm	levoglucosan, mannosan, galactosan and anomeric-C anhydroderivatives from cellulose/lignin combustion	biomass burning	Tagliavini et al., 2006; Pietrogrande et al., 2017
levoglucosan	levo	anomeric singlet at 5.45 ppm & specific structures between 3.5 and 4.6 ppm		biomass burning	Tagliavini et al., 2006; Paglione et al., 2014a&b; Pietrogrande et al., 2017
<u><i>saccharides</i></u>	Sac	used synonymously for compounds carrying H-C-O groups in unresolved mixtures but when also anomeric protons (O-CH-O) are present	glucose, sucrose and other sugars structurally similar not unequivocally identified	biogenic marine primary	Graham et al., 2002; Facchini et al., 2008b; Decesari et al., 2011; Decesari et al 2020; Liu et al., 2018; Dall'Osto et al., 2022*; Paglione et al., 2024
glucose	Gls	anomeric doublet at 5.22 ppm & specific structures between 3.5 and 4.2 ppm (not quantified but possibly quantifiable @5.22 ppm)		biogenic marine primary	Decesari et al., 2020; Dall'Osto et al., 2022b
sucrose	Suc	anomeric doublet at 5.40 ppm & specific structures between 3.5 and 4.2 ppm (not quantified but possibly quantifiable @5.40 ppm)		biogenic marine primary	Decesari et al., 2020; Dall'Osto et al., 2022b
ribose	Rib	anomeric doublet at 5.37 and 5.24 ppm & specific structures between 3.6 and 4.2 ppm (not quantified)		biogenic marine primary	Suggested by this study (to be confirmed)
<u><i>polyols</i></u>		unresolved mixture not quantified (including glycerol and D-threitol)	glycerol, threitol, erytritol and structurally similar molecules not unequivocally identified		
glycerol	Gly	specific structures at 3.55, 3.66 & 3.77 ppm (not quantified but possibly quantifiable @ 3.55 ppm)		biogenic marine primary	Decesari et al., 2020; Dall'Osto et al., 2022b
D-threitol	D-th	specific structures between 3.6 - 3.7 ppm (not quantified)		biogenic marine primary	suggested by Paglione et al., 2024 (to be confirmed)
arabitol	Arab	specific structures between 3.6 - 4 ppm (not quantified)		biogenic marine primary	Suggested by this study (to be confirmed)
galacticol	Gal	specific structures between 3.7 - 4 ppm (not quantified)		biogenic marine primary	Suggested by this study (to be confirmed)
<u><i>phenolic compounds</i></u>	PCs	unresolved resonances between 6.5 – 7.2 ppm	Phenol and other compounds consisting of one or more hydroxyl groups (–OH) bonded directly to an aromatic ring (e.g., vanillic acid, etc.)	biomass burning [...]	Decesari et al., 2007; Chalbot and Kavouras, 2014
<u><i>low-molecular weight fatty acids or "lipids"</i></u>	LMW-FA	unresolved complex resonances at 0.9, 1.3, and 1.6 ppm in the H-C spectral region	fatty acids (free or bound) from degraded/oxidized lipids (e.g. caproate, caprylate, suberate, sebacate, etc.) and similar compounds owning a chemical structures of alkanolic acids.	biogenic marine primary	Graham et al., 2002; Facchini et al., 2008b; Decesari et al., 2011; Decesari et al 2020; Liu et al., 2018
<u><i>biogenic SOA</i></u>	BSOA	Series of singlets/doublets between 0.9 – 1.6 ppm	compounds formed from the oxidation of terpenes and isoprene, including terebic acid, MBTCA (Methyl-butanetricarboxylic Acid) and methyl-tetrols	biogenic terrestrial secondary	Finessi et al., 2012; Zanca et al., 2017

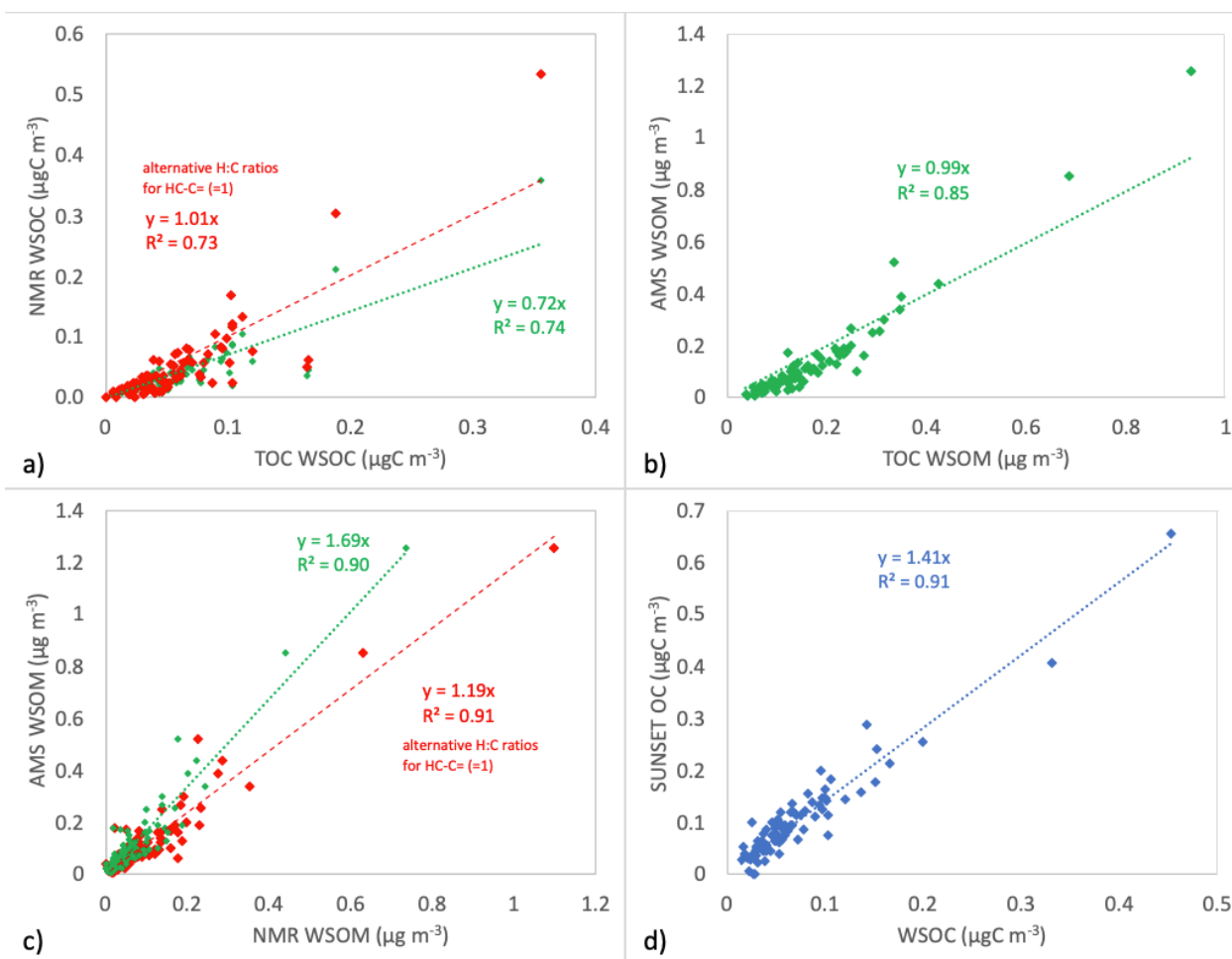


Figure S2. Comparison between the parallel quantifications of WSOC, WSOM and OC measured by the different instruments employed: panel (a) shows the WSOC reconstructed by H-NMR spectra conversion (based on the specific functional-groups-related stoichiometric H:C ratios) against the total WSOC measured by TOC-analyzer; panel (b) reports the water-soluble organic mass (WSOM) measured by AMS against the one calculated by WSOC measured by the TOC-analyzer converted using the AMS OM:OC elemental ratios; panel (c) shows the comparison between WSOM measured by AMS and reconstructed by H-NMR spectra; finally, panel (d) shows the total OC measured by Sunset against the water soluble fraction (WSOC) measured by the TOC-analyzer.

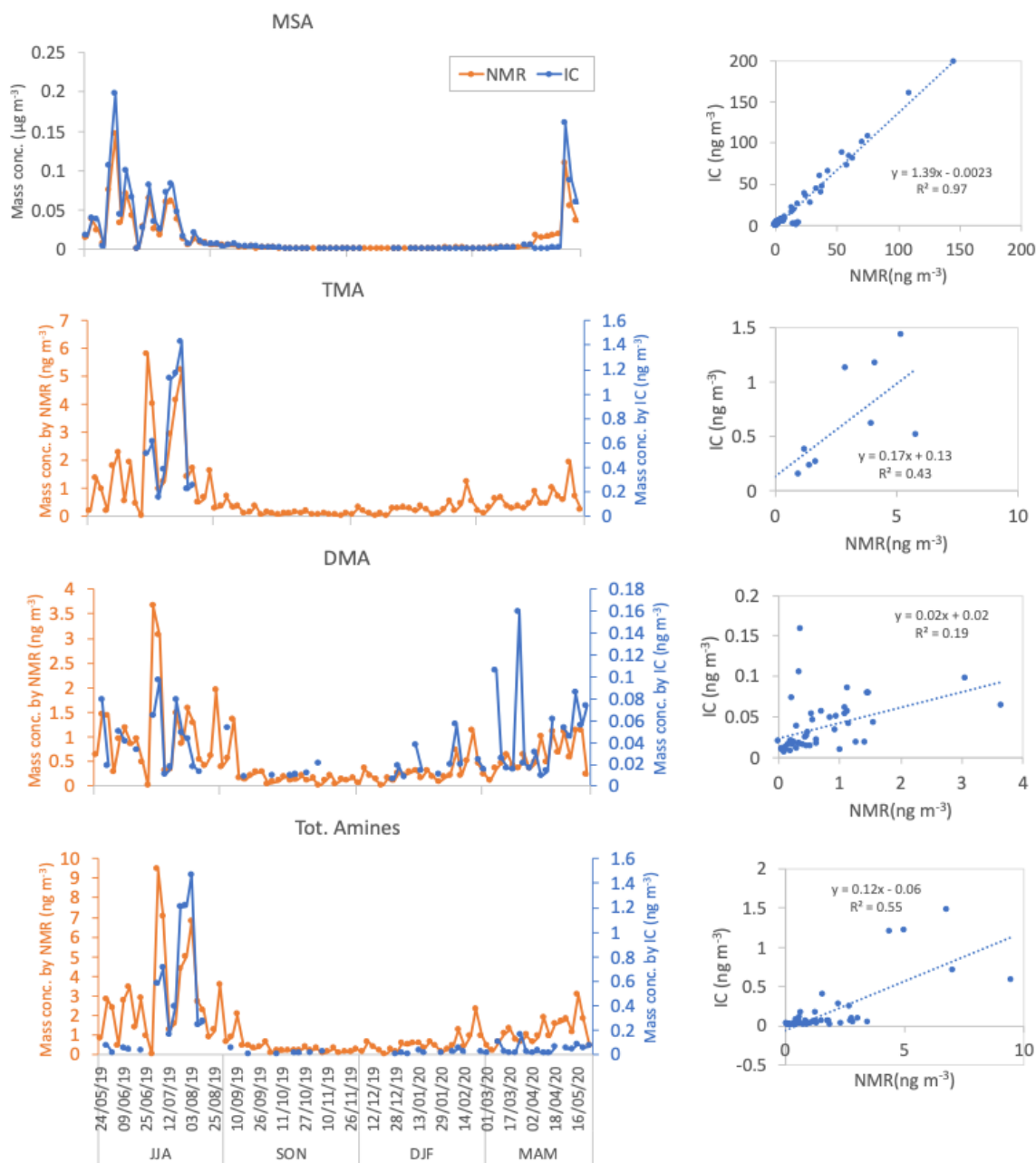


Figure S3. Comparison between mass concentrations of alkylamines and MSA identified and quantified by NMR and IC analyses. Missing data points corresponds to not-detectable values.

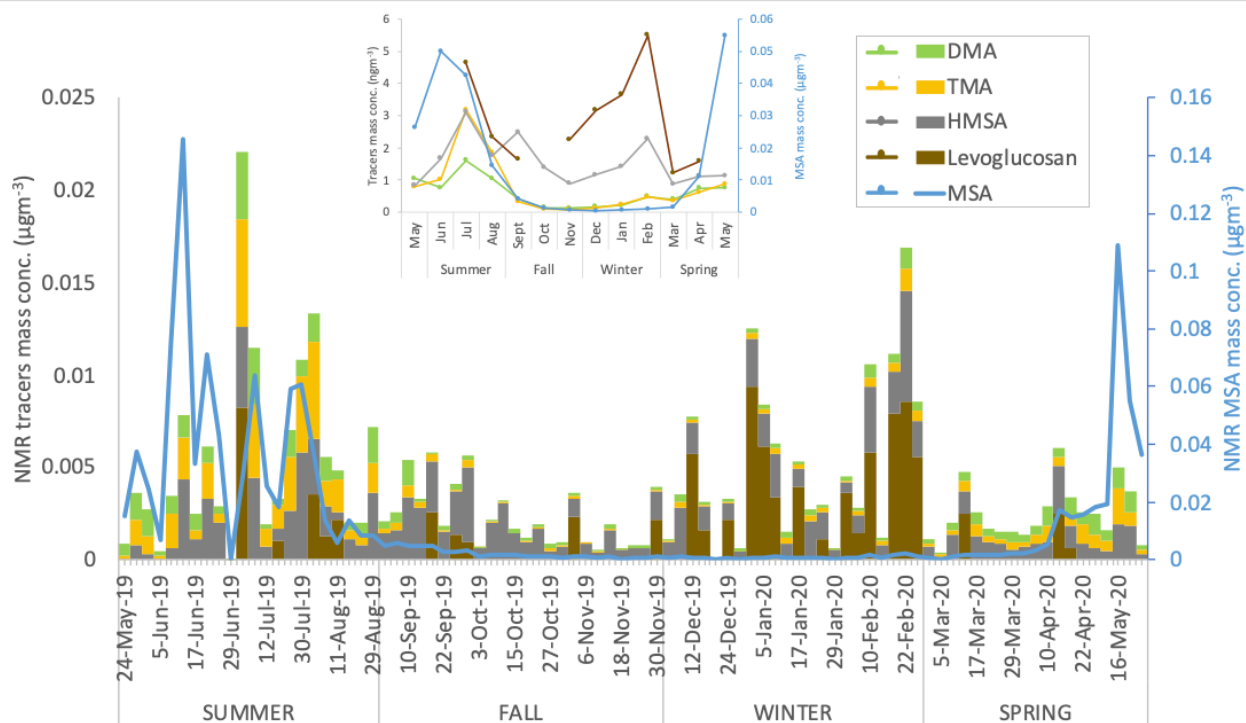


Figure S4. molecular tracers identified and quantified by H-NMR, namely methane sulfonic acid (MSA), di- and tri- methyl amines (DMA and TMA, respectively), hydroxy-methane sulfonic acid (HMS) and levoglucosan. Panel a) shows the concentrations measured in every single samples, while panel b) reports the monthly averages. Missing data points corresponds to not-detectable values (i.e., specific signals of those compounds in the NMR spectra are not emerging from the baseline).

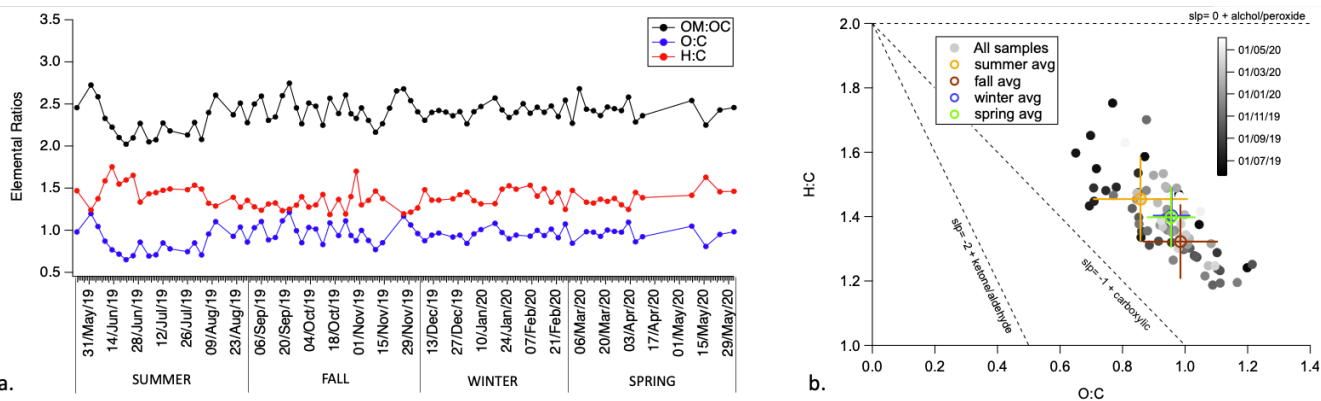


Figure S5. elemental ratios (OM:OC, O:C, H:C) extrapolated by HR-AMS spectra: panel a) time series along the year; panel b) Van Kreveln diagram with highlighted seasonal average values (colored empty circles) and variability (error bars, representing standard deviations)

S.2 Additional information on PMF Analysis of HR-AMS and H-NMR spectra

30 The aim of PMF is to derive a linear combination of components (factors) that can reproduce the observed chemical composition and variations in time of the sampled organic aerosol (OA) and can be possibly linked to specific OA sources and/or OA formation/transformation processes in atmosphere (Zhang et al., 2011).

In particular, PMF attempts to solve the bilinear matrix equation, $x_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij}$, by following the weighted least-squares approach. x_{ij} refers to a particular experimental measurement of concentration species j (here, one point of the mass or NMR spectrum) in one particular sample i . Individual experimental measurements are decomposed into the sum of p components or sources, each one of which is described by the product of two elements; one of these elements, (f_{kj}), defines the relative amount of the considered variable j in the source composition (loading of this variable on the source chemical profile) and the other, (g_{ik}), defines the relative contribution of this source in that sample i (score of the source on this sample). The sum is extended to $k = 1, \dots, p$ factors/sources, leaving the measurement unexplained residual stored in e_{ij} .

40 The mathematical goal of the model is to find values of $g_{i,k}$ (factor contributions), $f_{k,j}$ (factor profiles), and p (number of factors) that best reproduce original data matrix ($x_{i,j}$). For this purpose the values of $g_{i,k}$ and $f_{k,j}$ are iteratively fitted to the data using a least-squares algorithm, minimizing the fit parameter called Q , defined as the sum of squared residuals: $Q = \sum_{i=1}^m \sum_{j=1}^n (\frac{e_{ij}}{s_{ij}})^2$ where s_{ij} is the uncertainty of the j^{th} species concentration in sample i , n is the number of samples, and m is the number of species. The use of a data uncertainties input matrix to scale the residuals is one of the main advantages of PMF with respect to other non-negative factor analysis techniques, making it especially applicable to working with environmental datasets.

In the case of AMS the organic mass spectra were normalized and scaled by the TOC-based WSOC, in order to avoid problems of collection and transmission efficiencies and of nebulization efficiency (Bozzetti et al., 2017; O'Brien et al., 2019). Similarly, also the NMR spectra before the blanks subtraction were normalized to the total WSOC mass reconstructed by stoichiometric H/C ratios conversion applied to the measured functional groups (see Section 2.2.1 of the main text and refer to Tagliavini et al., 2006 and Decesari et al., 2007).

The PMF AMS input matrix here included the 81 organics mass spectra with their 741 HR fragments until m/z up to 208. Before PMF, the organic data matrix was arranged according to the Ulbrich et al. (2009) recommendations. Isotope ions were removed and a minimum counting error was applied to create the data uncertainty input matrix. All fragments with a signal-to-noise ratio (SNR) below 0.2 were removed from the matrices, and those with a signal-to-noise ratio below 2.0 were down-weighted, according to the recommendations of Paatero and Hopke (2003), increasing their uncertainty by a factor 2. Finally, the fragments related to ion CO_2^+ were also down-weighted since they are calculated as a constant fraction of the ion CO_2^+ (Allan et al., 2004). Unconstrained PMF was performed for $p = 2-8$ factors to choose a “best” solution before the model statistical and rotational uncertainty analysis (described later). Five random seed runs were performed for each p (35 runs in total). Some diagnostics were produced (Figure S6) to investigate the optimum p based on the explained variability of the input

matrix (Q/Q_{exp} , scaled residuals) and the stability and/or interpretability of the solutions among different runs for each p . Eventually, a 4-factors solution was chosen because of the best separation of interpretable spectral features and contributions. All random seed runs provided essentially identical results (that is, the lowest Q/Q_{exp} relative standard deviation) especially for the 4-factor solution (the factor profiles are shown in Figure S7). The Q/Q_{exp} of this average 4-factor solution exhibited a random pattern in both dimensions (time series and variables) of the reconstructed PMF output matrix (81 samples and 741 HR fragments up to m/z 208).

The 3-factors solution ($p=3$) was also considered, but eventually rejected because not able to separate a specific factor related to the Arctic Haze period from the background mixed factor (see description in the main text, Section 3.3). Going to 5-factors instead, the solutions start to be less robust producing multiple factors for the same constituents (see correlation coefficients reported in Figure S8).

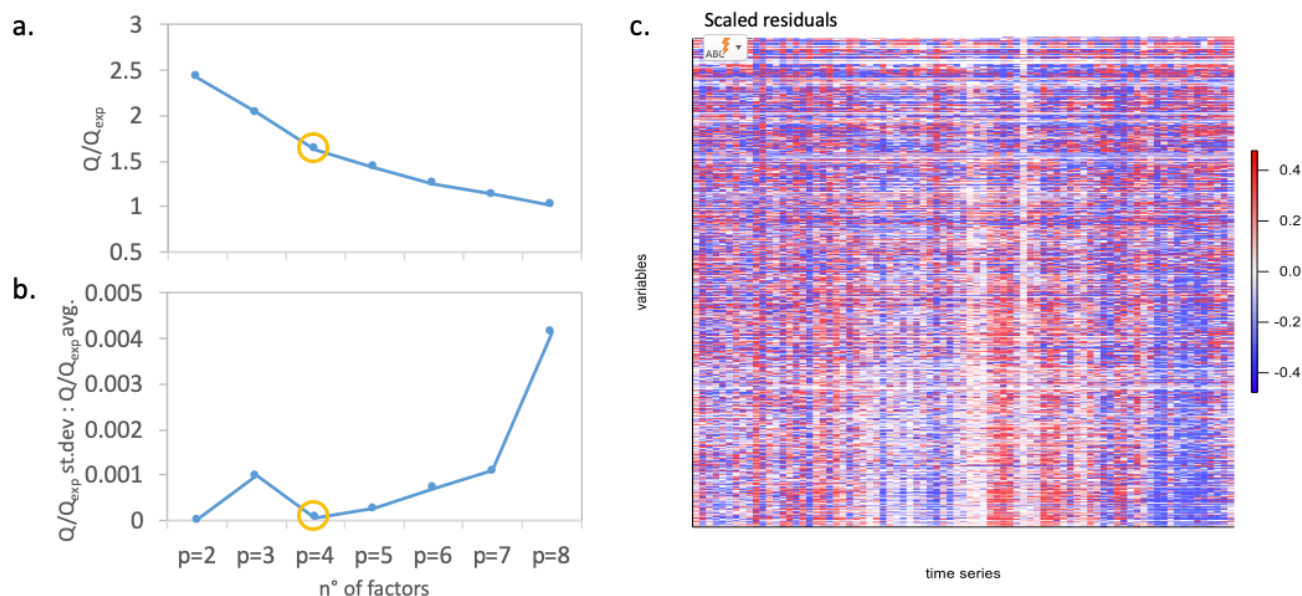
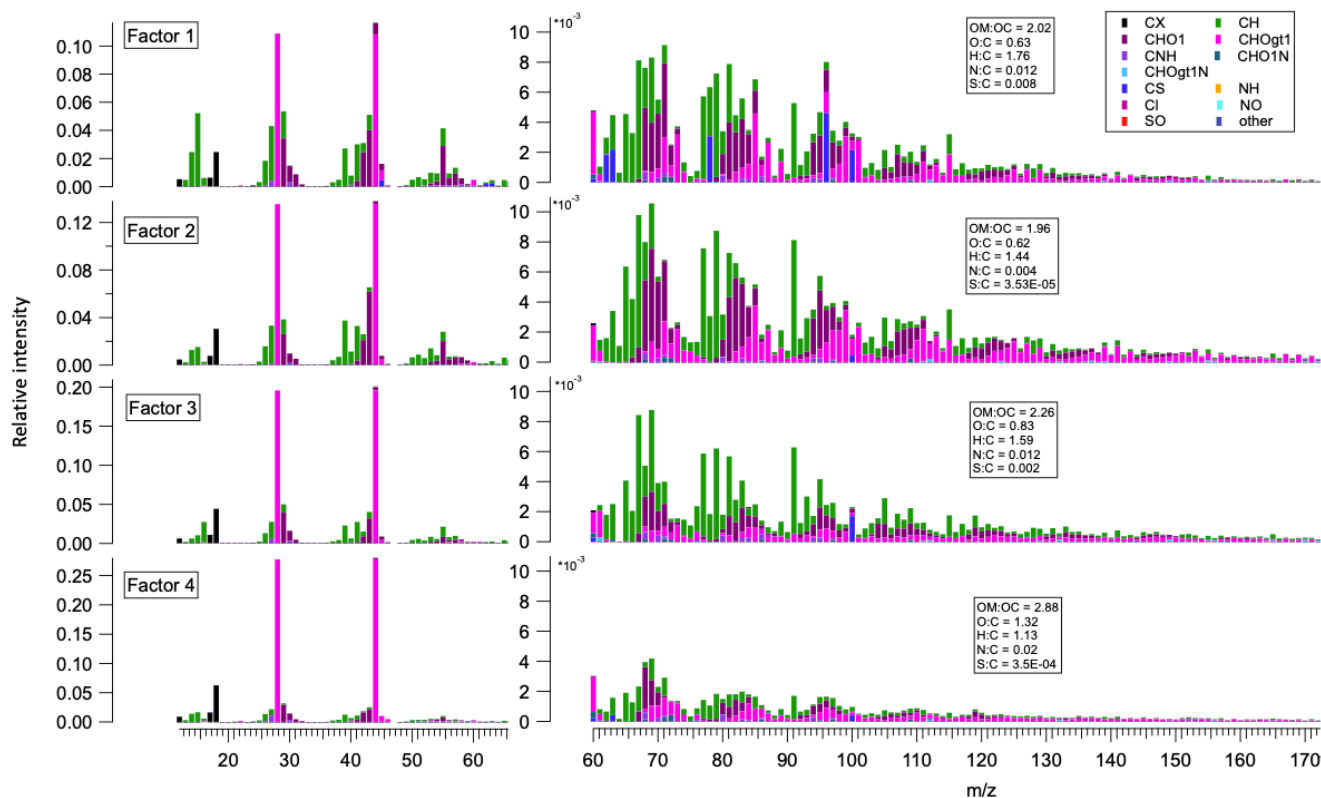


Figure S6. AMS PMF Q-values and residuals plots: (a) Q/Q_{exp} ratio versus the number of factors p as the average between the different 5 random runs executed for each p (i.e., Q/Q_{exp} avg.). (b) Q/Q_{exp} avg. to evaluate the stability of the different random runs for each p . Both in (a) and (b) the yellow circle denotes the chosen solution ($p=4$); (c) distribution of the scaled residuals among samples and variables (AMS spectra m/z fragments).

For factor identification (Zhang et al., 2011), we used a combination of criteria. These include in particular the factor seasonal cycle, fragmentation pattern, characteristic fragments, time series correlation with external markers, time series correlation with environmental parameters and the BT analysis.

The atomic ratios (O:C, H:C, N:C and S:C) and OM:OC ratios shown in Figure S7 were calculated for each AMS-PMF factor (for m/z up to 208), using the Analytical Procedure for Elemental Separation (EAlight version 1.06) within Igor.



85 **Figure S7.** Mass spectra profiles of the 4-factors by AMS-PMF. The profiles are shown as normalized fragment intensities in HR with average atomic ratios. The fragments are color-coded with the family (as detailed by the legend).

(a)

p=4

R2 (Pearson)

		F1	F2	F3	F4
Profiles		Factor1	Factor2	Factor3	Factor4
F1	Factor1	1			
F2	Factor2	0.86	1		
F3	Factor3	0.92	0.96	1	
F4	Factor4	0.92	0.76	0.83	1

		F1	F2	F3	F4
Contributions		Factor1	Factor2	Factor3	Factor4
F4	Factor4	0.06	0.00	0.00	1
F3	Factor3	0.00	0.13	1	
F2	Factor2	0.09	1		
F1	Factor1	1			

(b)

p=5

R2 (Pearson)

		F1	F2	F3	F4	F5
Profiles		Factor1	Factor2	Factor3	Factor4	Factor5
F1	Factor1	1				
F2	Factor2	0.90	1			
F3	Factor3	0.52	0.42	1		
F4	Factor4	0.86	0.85	0.71	1	
F5	Factor5	0.94	0.98	0.50	0.92	1

		F1	F2	F3	F4	F5
Contributions		Factor1	Factor2	Factor3	Factor4	Factor5
F5	Factor5	0.44	0.02	0.02	0.01	1
F4	Factor4	0.02	0.10	0.07	1	
F3	Factor3	0.00	0.04	1		
F2	Factor2	0.02	1			
F1	Factor1	1				

90 **Figure S8.** Correlation coefficients (Pearson R2) between AMS factor profiles and contributions of different solutions by PMF ME-2: (a) 4-factors solution ($p=4$), eventually chosen as the best solution; (b) 5-factors solution.

The PMF NMR input matrix included the full collection of 87 NMR spectra of PM1 samples object of the study. The application of PMF analysis to NMR spectral datasets is relatively new for atmospheric sciences, even though non-negative factor analysis being widely employed in other fields, especially in biochemistry. In the present study, we followed the procedure already described in previous publications (Paglione et al., 2014a; Paglione et al., 2024).

Before the statistical analysis, the original NMR spectra were subjected to several preprocessing steps in order to remove spurious sources of variability. A polynomial fit was applied to baselines and subtracted from the spectra. Careful horizontal alignment of the spectra was performed using the Tsp-d4 and buffer singlets as reference positions (at 0.00ppm and 8.45ppm, respectively). The spectral regions containing only noise or sparse signals of solvent/buffer ($H < 0.5$ ppm; $4.7 < H < 5.2$ ppm; and $8.15 < H < 8.60$ ppm) were omitted. The five blanks spectra were averaged together and the corresponding mean blank-spectrum was subtracted to all the sample-spectra. Binning over 0.02 ppm of chemical shift intervals was applied to remove the effects of peak position variability caused by matrix effects. Low-resolution spectra (~400-points) were finally obtained and processed by applying the Multilinear Engine 2 solver (ME-2, Paatero, 2000) controlled within the Source Finder software (SoFi v8.6, Canonaco et al., 2013; Crippa et al., 2014).

The uncertainty input matrix required by PMF was derived in this study from the signal-to-noise ratios of the NMR spectra (as already described in previous publications, Paglione et al., 2014a, 2014b and 2024). In particular, the uncertainty was calculated for each sample as 7 times the standard deviation of the signal intensity in a portion of the spectrum containing only noise/baseline values (between 6.5 and 7ppm).

Solutions with different number of factors (p = from two up to eight) were explored for the spectral dataset. Also for NMR five random seed runs were performed for each p (35 runs in total). Eventually, a five-factors solution was chosen because of the best separation of interpretable spectral features and contributions. The 4-factors solution ($p=4$) was also considered, but rejected in the end because not able to separate a specific factor related to the Arctic Haze period from the background mixed factor (see later description). Going to 6-factors instead, the solutions start to be less robust producing multiple factors for the same constituents (see correlation coefficients reported in Figure S9) and in disagreement between different runs.

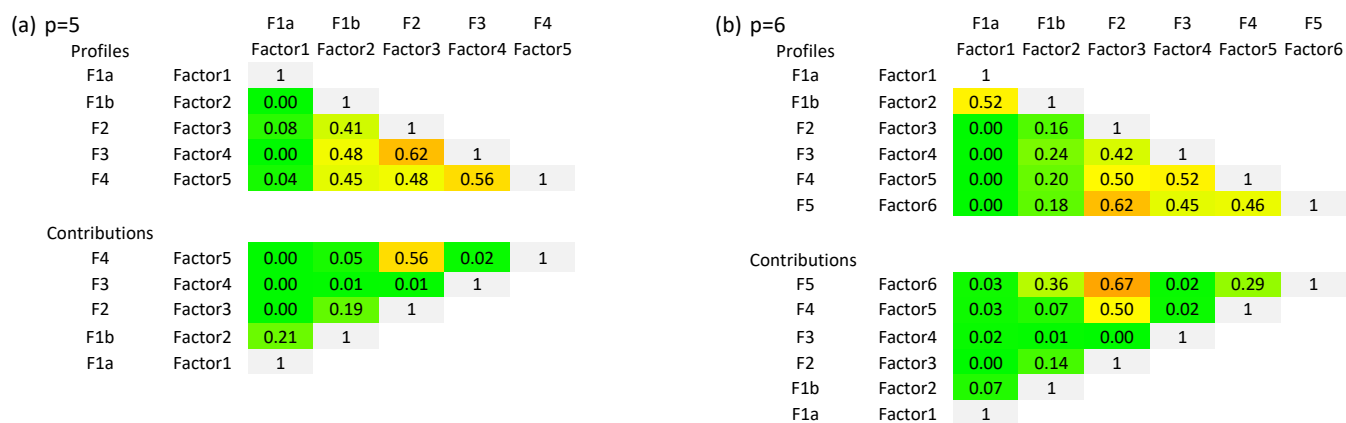


Figure S9. Correlation coefficients (Pearson R2) between NMR factor profiles and contributions of different solutions by PMF ME-2: (a) five-factors solution (p=5), eventually chosen as the best solution; (b) six-factors solution.

The same mathematical diagnostics used for AMS-PMF solutions evaluation were produced also for NMR (Figure S10). The Q/Qexp values for the NMR-PMF suggest that a number of factors higher than five does not significantly improve the goodness of fit (panel a). All random seed runs provided essentially identical results (that is, the lowest Q/Qexp relative standard deviation) only starting from the 5-factor solution (panel b). The scaled residuals resulted to be randomly distributed between samples and variables, without any clear structures/patterns (panel c).

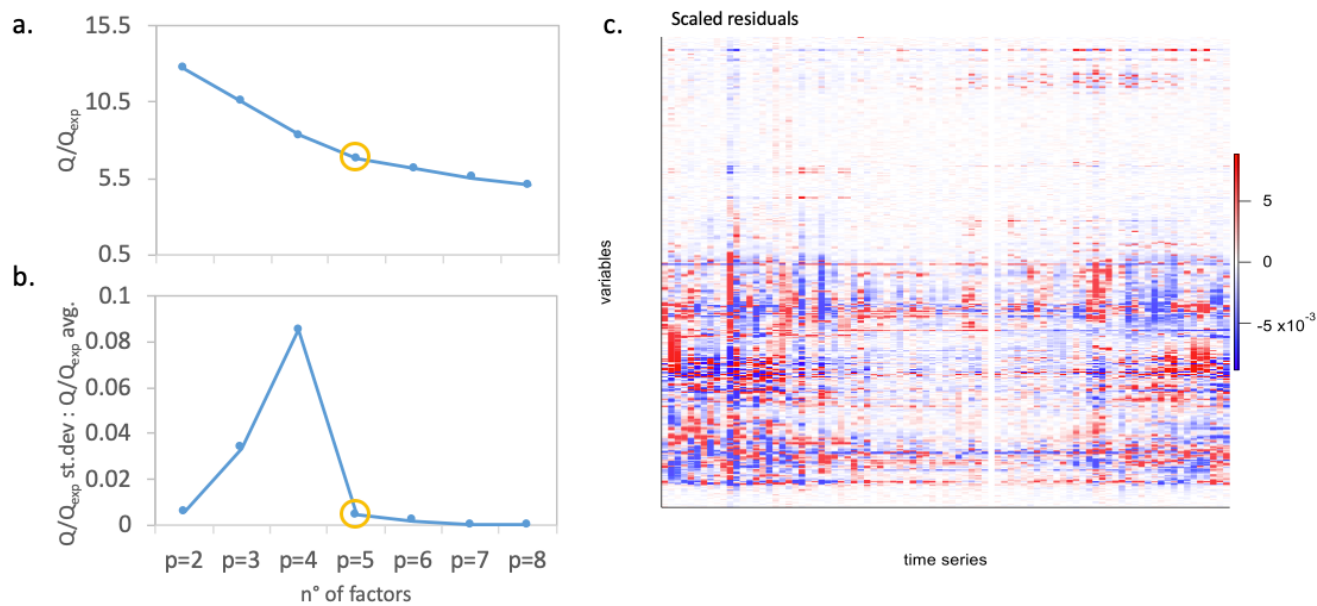


Figure S10. NMR factor analysis Q-values and residuals plots: (a) Q/Q_{exp} ratio versus the number of factors p as the average between the different 5 random runs executed for each p (i.e., Q/Q_{exp} avg.). (b) Q/Q_{exp} avg. to evaluate the stability of the different random runs for each p . Both in (a) and (b) the yellow circle denotes the chosen solution ($p=4$); (c) distribution of the scaled residuals among samples and variables (AMS spectra m/z fragments).

The interpretation of NMR factor spectral profiles was based on the presence of molecular resonances of tracer compounds, and on the comparison with a library of reference spectra recorded in laboratory/chamber experiments or in the field during near-source studies (Suzuki et al., 2001; Schmitt-Kopplin et al., 2012; Paglione et al., 2014a, 2014b; Decesari et al., 2014, 2020; Paglione et al., 2024). The identification was also supported by elaboration tools/software providing extensive libraries of biogenic compounds, such as Chenomx NMR suite (Chenomx inc., evaluation version 9.0), or allowing theoretical simulations of H-NMR spectra of atmospheric relevant molecules, such as ACD/Labs (Advanced Chemistry Developments inc., version 12.01). Two examples of tracers identified in the NMR spectra and used for interpretation of NMR factor profiles are reported in Figure S11 and S12). Further comparisons for the interpretation of Factor 1b and Factor 2 are instead reported in Figure S13 and S15.

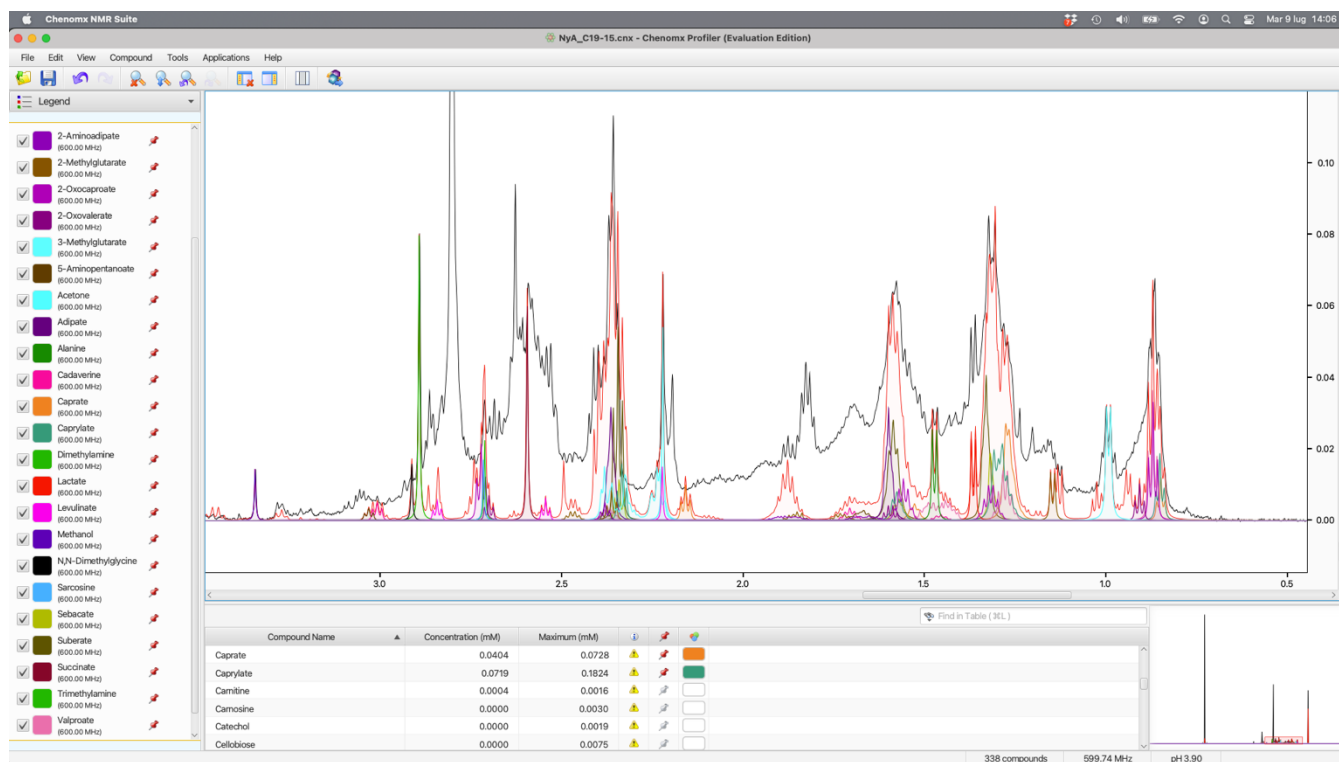


Figure S11. Example of identification of possible tracers using the extensive libraries of compounds offered by Chenomx NMR suite (Chenomx inc., evaluation version 9.0). Here it is reported an attempt of fitting the ambient PM1 spectrum of sample 26-Jul-2019, with the signals expected for the molecules available in the database. Red line is the fitting line using the sum of the possible molecules available in the database. Legend reports a list of compounds identified in this spectrum. Especially noteworthy are the signals of some fatty acids esters from degraded/oxidized lipids such as caproate, caprylate, suberate, sebacate, etc. and similar compounds owning a chemical structure of alkanolic acids.

155

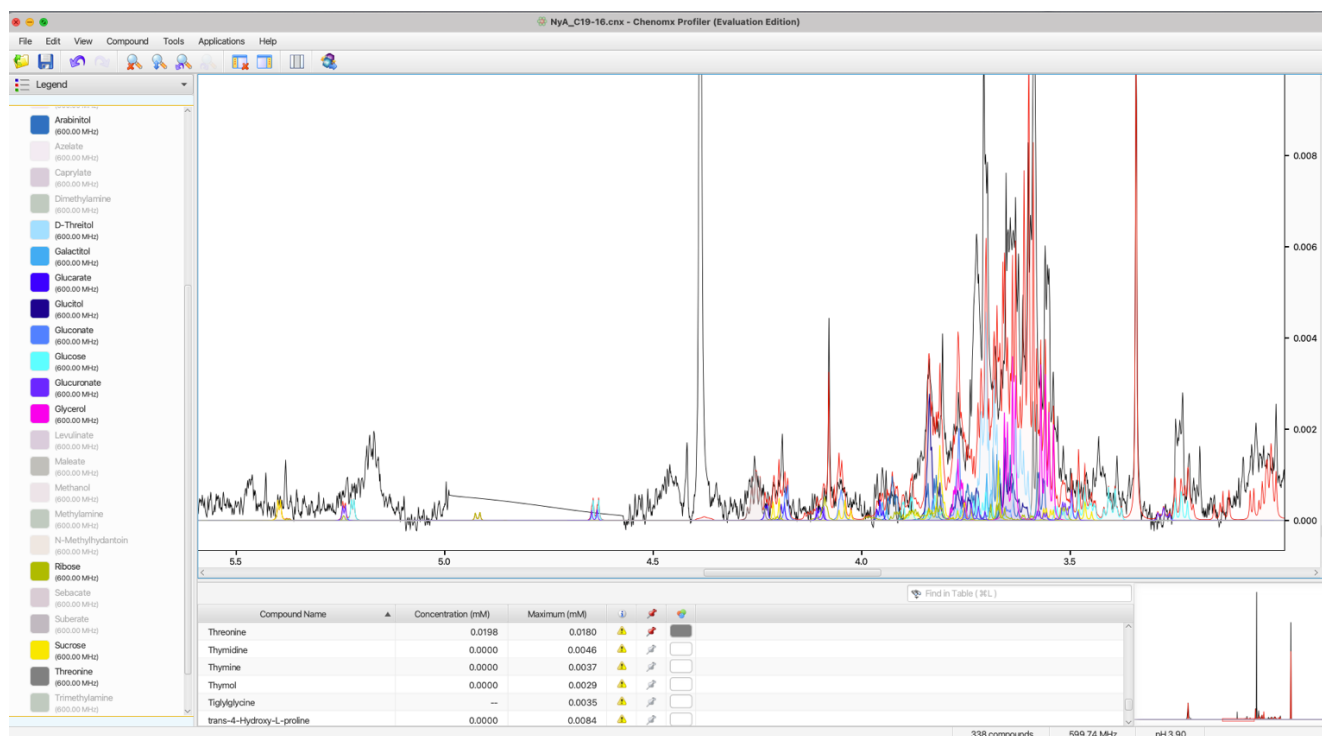


Figure S12. Example of identification of possible tracers using the extensive libraries of compounds offered by Chenomx NMR suite (Chenomx inc., evaluation version 9.0). In this figure are shown the expected NMR spectral patterns of some sugars and polyols, specifically sucrose (yellow line), glucose (cyan line), ribose (greenish line), glycerol (magenta line), D-threitol (light-blue line) and others (as specified in the color legend), against the NMR spectrum of PM1 sample 30-Jul-2019 (black line). The spectral region distorted by H₂O signal is removed in the ambient spectrum

160

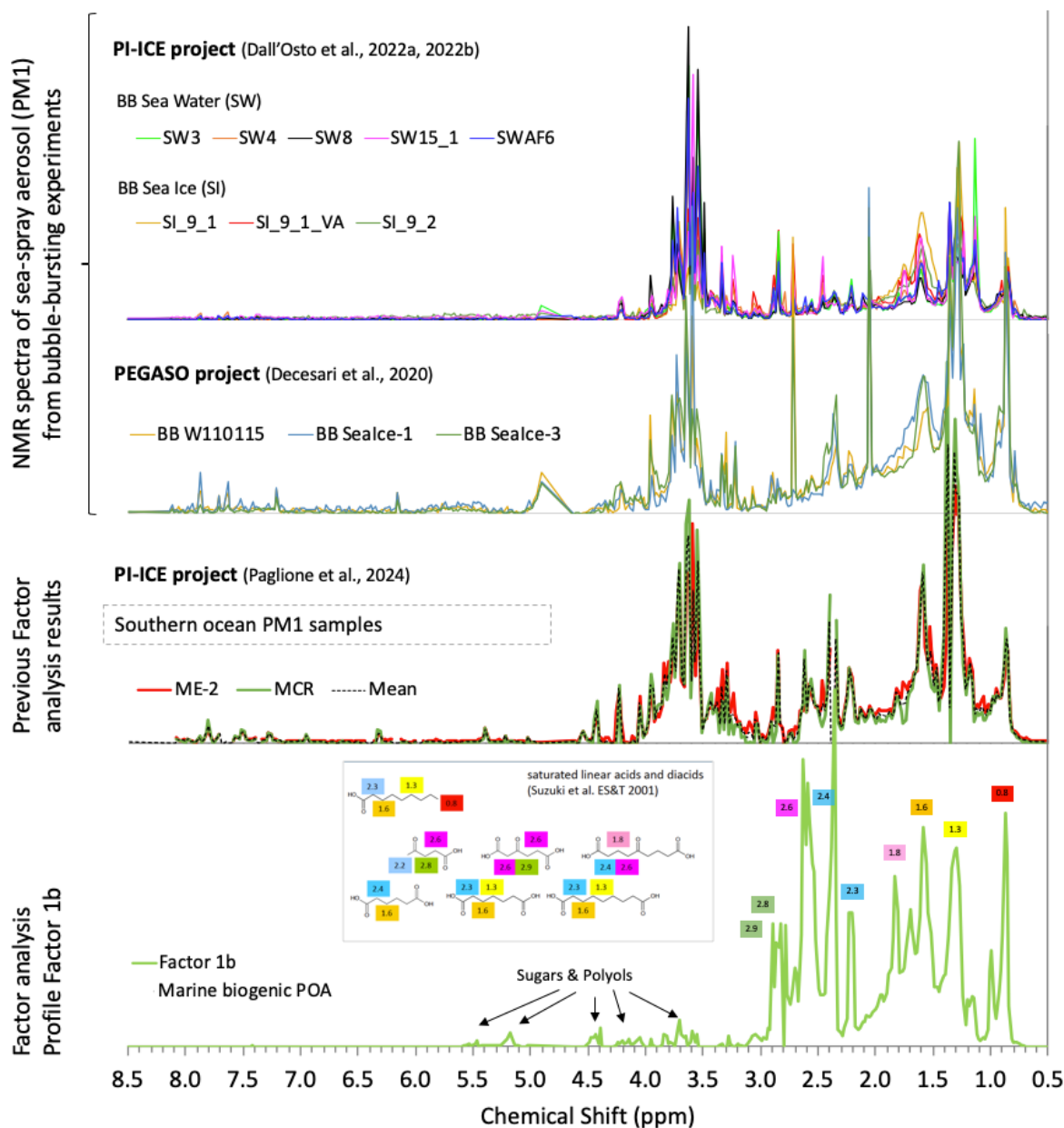
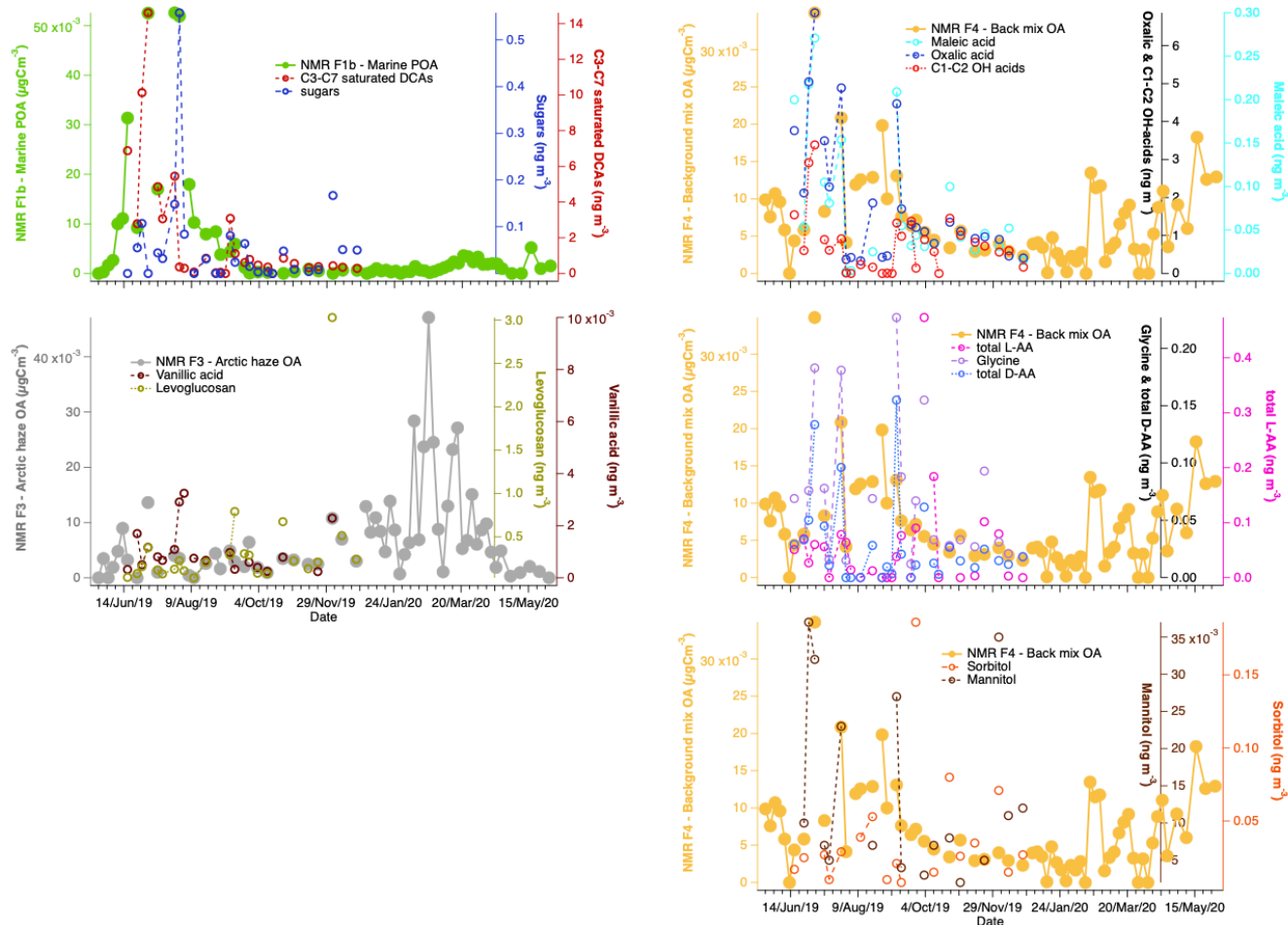
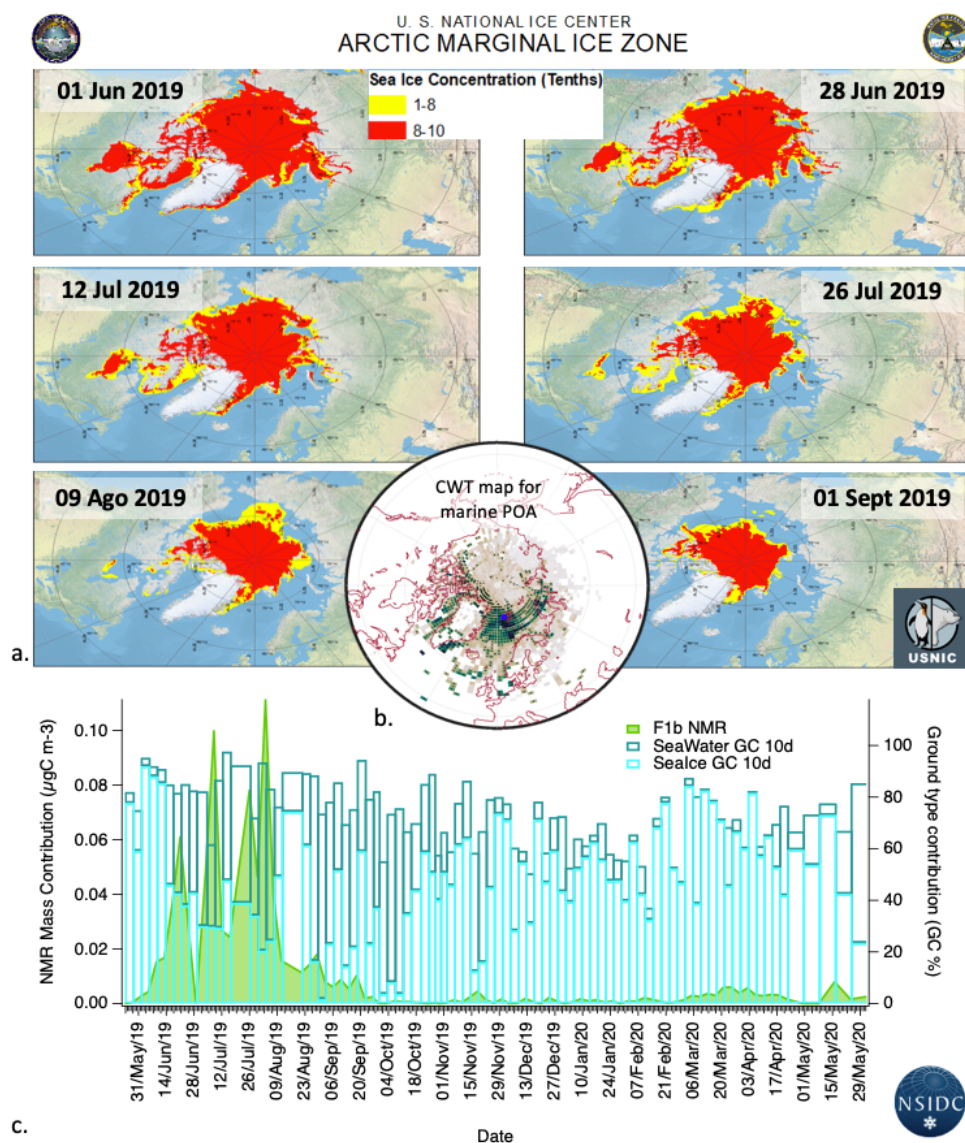


Figure S13. Comparison between the profile of Factor 1b (attributed to marine POA) and some NMR spectra of sea-spray generated during bubble bursting experiments from previous studies (PEGASO and PI-ICE projects, Decesari et al., 2020; Dall'Osto et al., 2022a) and with Factor analysis results on Southern Ocean ambient OA in PM1 samples (Paglione et al., 2024).



170 **Figure S14.** Comparison between the time trends of selected H-NMR factors and of molecular tracers from IC/HPLC-MS analysis:(top-left)
H-NMR factor for marine aged POA vs. C3-C7 saturated dicarboxylic acids (DCAs) (= malonic + succinic + malic + glutaric + adipic +
pimelic acids) and sugars (= glucose + sucrose + xylose + ribose); (left-center) H-NMR factor for Arctic haze and wood burning tracers
(levoglucosan and vanillic acid); (top-right) H-NMR Factor 4 (background) vs oxalic acid, C1-C2 monocarboxylic and hydrocarboxylic
acids (OH-MCAs = acetic + glycolic acid) and maleic acid; (right-center) H-NMR Factor 4 (background) vs glycine, total L-aminoacids (L-
175 AA) and total D-aminoacids (D-AA). The trends of the main C6 polyols (mannitol, sorbitol) are shown on the bottom-right corner.



180 **Figure S15.** Panel (a): maps of the Arctic Marginal Ice Zone (MIZ) from the archive of the US National Ice Center (USNIC) for some dates
 of the summer 2019 (<https://usicecenter.gov/Products>, last access: 16 Jan. 2025). Panel (b): concentration weighted trajectories (CWT) map
 of Factor 1b, showing the most probable source area for that factor, to be compared with panel (a). Panel (c): bars show the sea-water/sea-
 ice fractional influence on the backtrajectories at GVB, overlaid on the time series of PMF contributions of F1b (marine POA) as apportioned
 185 by NMR analysis (green line and area). Ground condition maps were obtained from the National Ice Center's Interactive Multisensor Snow
 and Ice Mapping System (IMS) (Helfrich et al., 2007; National Ice Center, 2008), National Snow & Ice Data Center (NSIDC; <https://nsidc.org/>, last access: 16 Jan. 2025). We used the daily Northern Hemisphere maps with a resolution of 4 km. The ground types considered
 were "sea-water", "sea-ice", "land", and "snow". Seawater indicates passage of the air mass over open seawaters, while sea ice indicates
 passage over icecovered seawaters. For each back-trajectory endpoint, we applied nearest-neighbor interpolation in space and time to find
 the closest pixels on the satellite map and associated the endpoint with the corresponding ground type. Combining the information obtained
 190 along the whole back-trajectory (or group of back-trajectories for PM1 samples) allowed estimation of the contribution of each ground type
 to each PM1 sample.

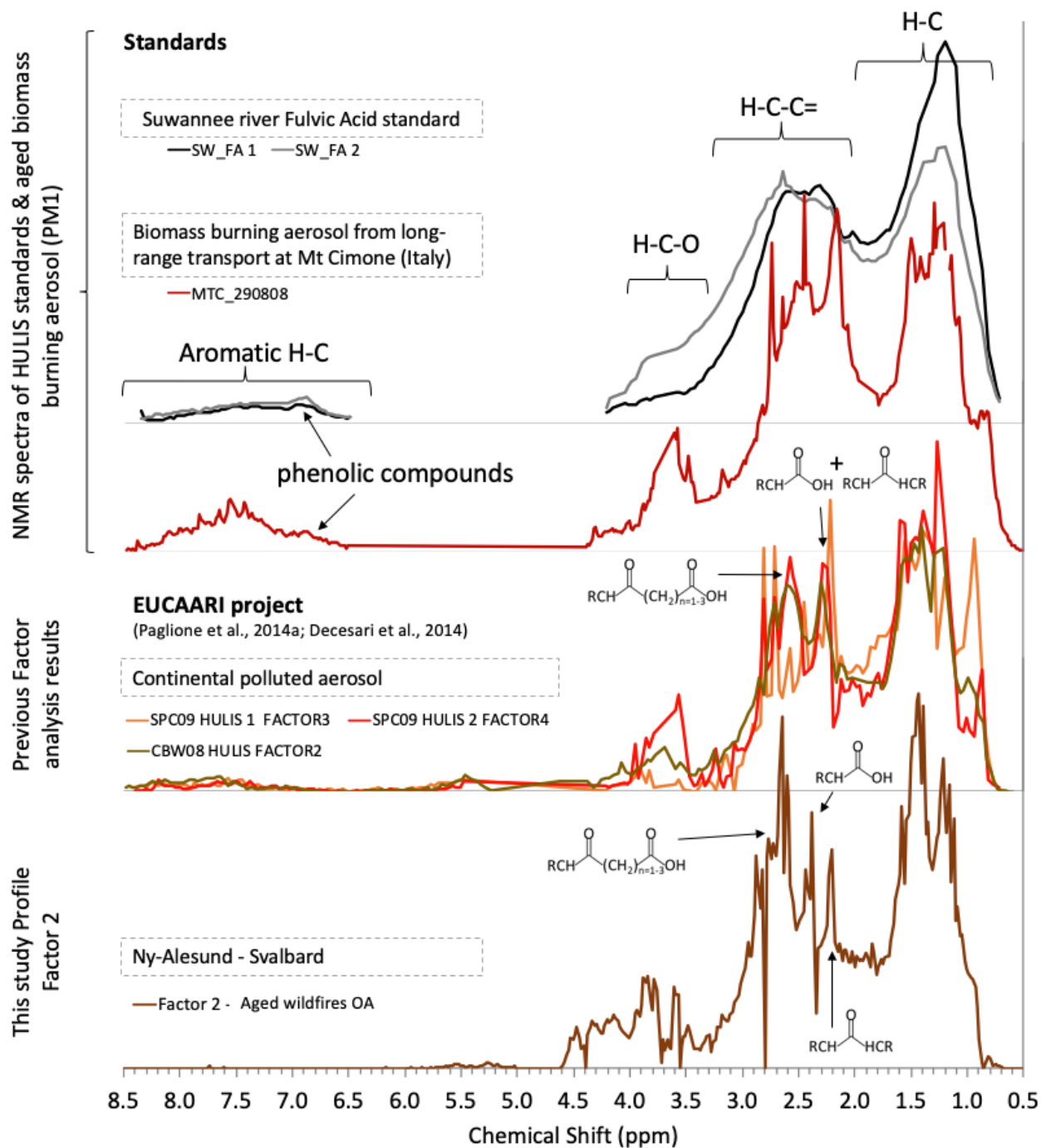


Figure S16. Comparison between the profile of Factor 2 and some NMR spectra of HULIS by laboratory analysis of standards (Suwannee river Fulvic acid), ambient PM1 samples and factor analysis results from previous studies (Paglione et al., 2014a; 2014b; Decesari et al., 2014).

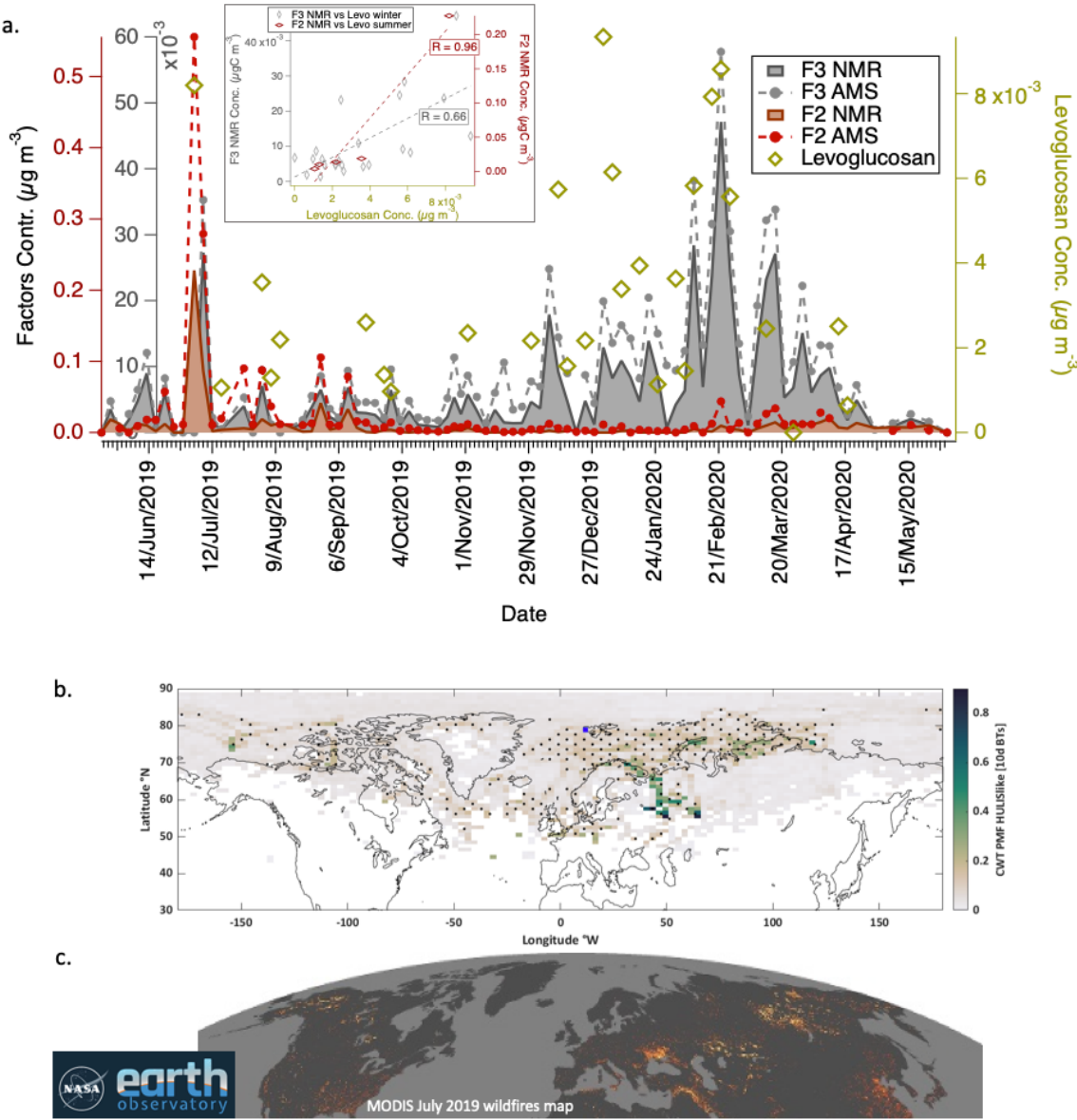
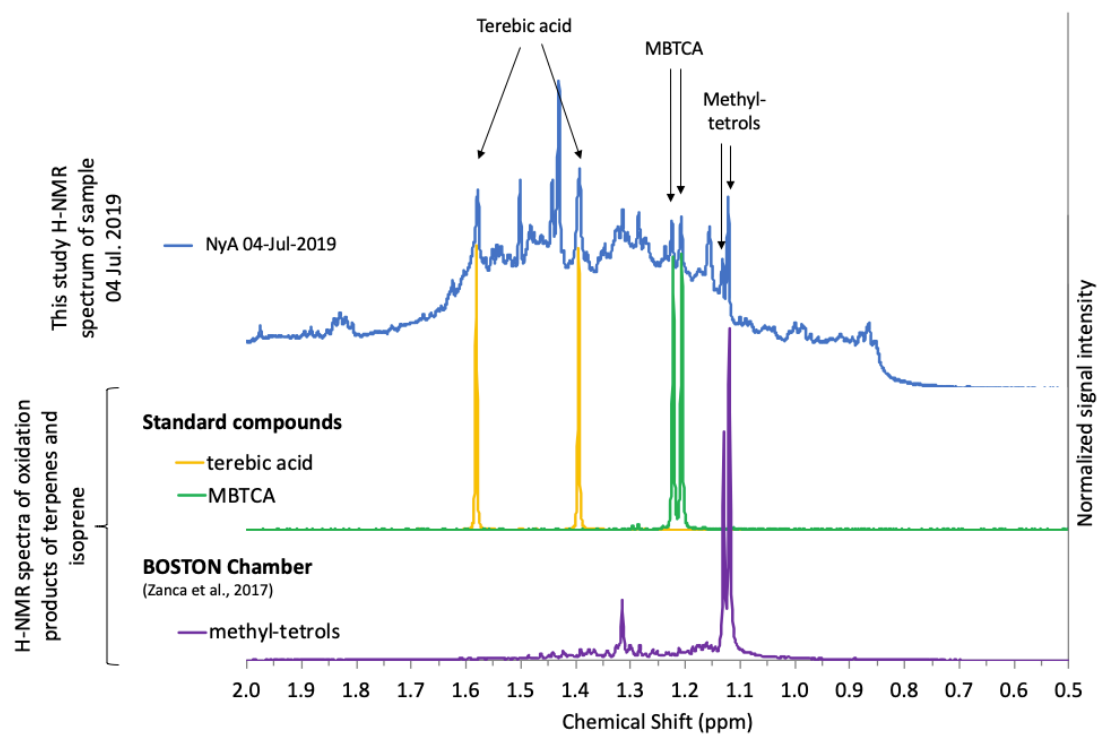


Figure S17. Panel (a): levoglucosan concentration measured by NMR (greenish markers) overlaid on the time series of PMF contributions of Factor 2 and Factor 3 (redish and greysh lines and backgrounds); in the subpanel, scatterplots of levoglucosan concentrations (x-axis) against F2 (right y-axis) and F3 (left y-axis) contributions during summer (jun.-sept.) and winter (dec.-mar.), respectively. Panel (b): concentration weighted trajectories (CWT) map of Factor 2, showing the most probable source area for that factor; to be compared with panel (c), reporting the wildfires map by NASA MODIS satellite for July 2019.

205



210 **Figure S18.** Comparison between the H-NMR spectrum of the sample of July 04th 2019 at full resolution and H-NMR spectra of oxidation products of terpenes (i.e., terebic acid and MBTCA by standard solutions) and isoprene (i.e., methyl-tetrols, by PAM chamber experiments).

215

Table S3. Pearson Correlation coefficients (R) between the time series of AMS and NMR factor-contributions for the chosen solutions.

R (Pearson)			AMS				NMR					
			F1	F2	F3	F4	F1a	F1b	F2	F3	F4	F1a+F1b
			marine OA	Aged wildfires OA	Arctic- haze OA	background mix OA	marine SOA	marine POA	Aged wildfires OA	Arctic haze OA	background mix OA	marine OA_TOT
AMS	F1	marine OA	1	0.23	0.00	0.01	0.84	0.80	0.11	-0.05	0.05	0.93
	F2	Aged wildfires OA	0.23	1	0.14	0.42	0.21	0.58	0.98	0.10	0.73	0.49
	F3	Arctic haze OA	0.00	0.14	1	0.34	-0.07	0.07	0.09	0.84	-0.05	0.02
	F4	background mix OA	0.01	0.42	0.34	1	0.08	0.14	0.45	0.38	0.57	0.13
NMR	F1a	marine SOA	0.84	0.21	-0.07	0.08	1	0.50	0.17	-0.07	0.13	0.80
	F1b	marine POA	0.80	0.58	0.07	0.14	0.50	1	0.43	0.04	0.24	0.92
	F2	Aged wildfires OA	0.11	0.98	0.09	0.45	0.17	0.43	1	0.06	0.77	0.38
	F3	Arctic haze OA	-0.05	0.10	0.84	0.38	-0.07	0.04	0.06	1	-0.09	-0.01
	F4	background mix OA	0.05	0.73	-0.05	0.57	0.13	0.24	0.77	-0.09	1	0.22
	F1a+F1b	marine OA_TOT	0.93	0.49	0.02	0.13	0.80	0.92	0.38	-0.01	0.22	1

220

225

Model uncertainty

230 The PMF model statistical and rotational uncertainty was assessed via a bootstrapping (BS) approach for both AMS and NMR dataset. The BS approach randomly resample the spectral datasets, generating new input matrices from the original input matrix for each new run. Each newly generated PMF input matrix had a total number of samples equal to those of the original matrices, although some of the original filter samples were represented several times, and others were not represented at all. The resulting variability intervals represent possible temporal variations of the factors profiles, random measurement inaccuracies and errors in the modelling process, such as rotational ambiguity and a mis-specified number of factors (Reff et al, 2007). 100 BS runs have been performed to AMS 4-factors solution and NMR 5-factors solution respectively. In both the
235 analyses, BS resamples reproduced 100% of the base factors and only a low change in Q was observed (6% and 9% for AMS and NMR respectively). The average estimated concentration of each factor to the total observed OA varied by less than 30 % of its mean value. Profiles and contributions reported in Figure S19 are the averages (lines) and the standard deviations (error bars) from the 100 BS runs.

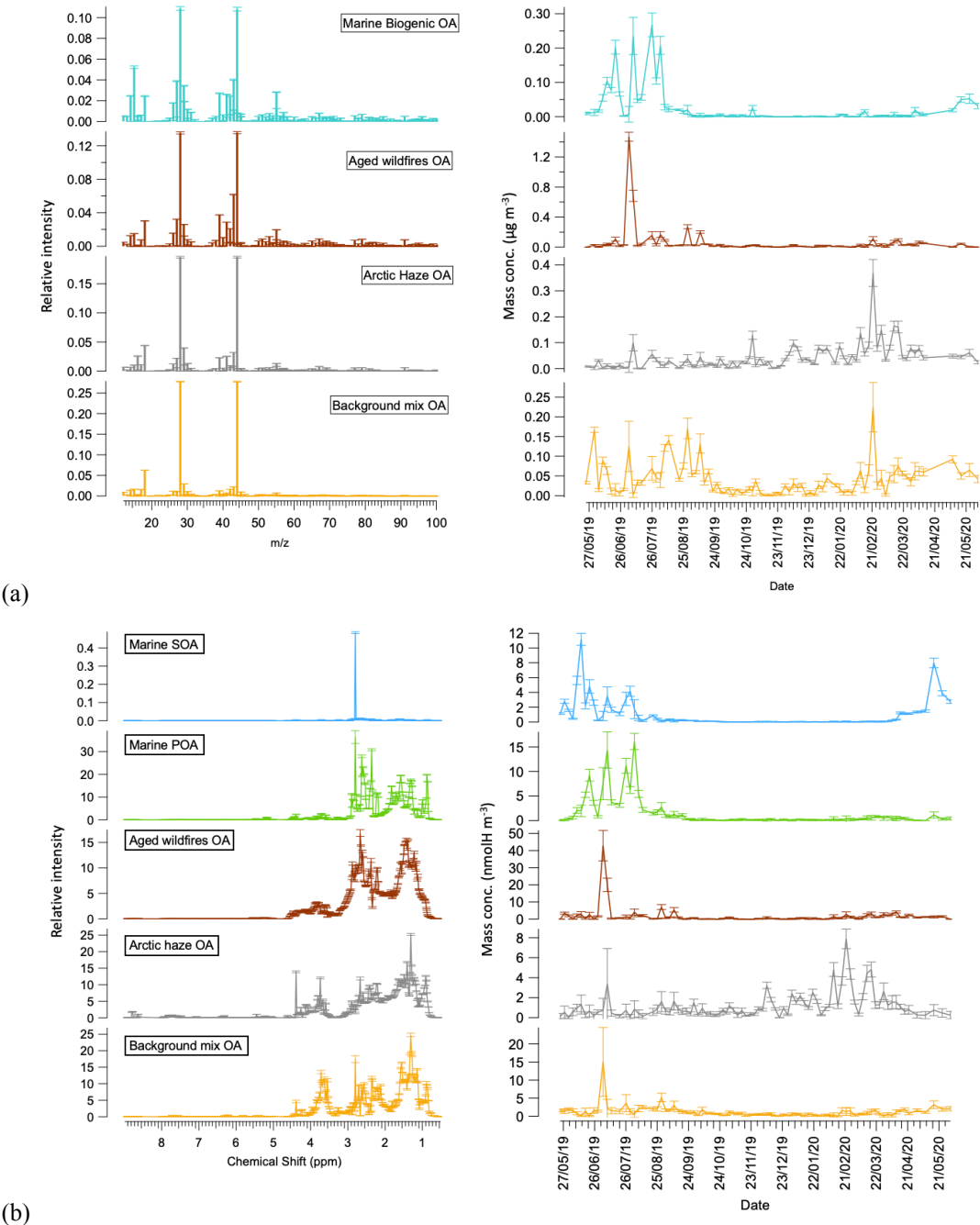


Figure S19. Results of the model uncertainty evaluation through bootstrap analysis (100 runs) for AMS (panel a) and NMR (panel b): lines represent the average factor profiles and contributions; error bars are the standard deviation of the 100 solutions.

250

Table S4. Pearson correlation coefficients between AMS & NMR factor contributions and ions/tracers measured by IC and other ancillary measurements.

R (Pearson)			AMS				NMR					
			F1	F2	F3	F4	F1a	F1b	F2	F3	F4	F1a+F1b
			Marine OA	Aged wildfires OA	Arctic-haze OA	Background mix OA	Marine SOA	Marine POA	Aged wildfires OA	Arctic-haze OA	background mix OA	Marine OA_TOT
Main tracers	IC	SO4	-0.05	0.07	0.79	0.43	0.01	-0.08	0.07	0.78	-0.04	-0.06
		NO3	0.02	0.08	0.11	0.11	-0.11	0.10	0.04	0.20	0.05	0.01
		NH4	0.35	0.37	0.59	0.58	0.37	0.28	0.34	0.55	0.22	0.35
		Na	-0.21	-0.09	0.42	0.27	-0.08	-0.28	-0.03	0.48	-0.04	-0.24
		Br ⁺	0.85	0.60	0.57	0.21	0.89	0.87	0.49	0.34	0.47	0.89
		SeaSalt	-0.21	-0.09	0.42	0.27	-0.08	-0.28	-0.03	0.48	-0.04	-0.24
		nss-SO4	-0.04	0.08	0.79	0.43	0.01	-0.07	0.08	0.79	-0.04	-0.05
		nss-K	-0.15	-0.08	0.25	-0.09	-0.15	-0.13	-0.08	0.20	-0.08	-0.16
		nss-other_ions	-0.16	-0.05	0.36	-0.04	-0.18	-0.13	-0.05	0.32	-0.08	-0.17
	AMS	Organics	0.35	0.94	0.35	0.61	0.31	0.63	0.90	0.30	0.68	0.55
		Sulfate	0.08	0.12	0.80	0.52	0.12	0.04	0.10	0.73	0.00	0.08
		Nitrate	0.33	0.30	0.07	0.44	0.17	0.35	0.24	0.06	0.30	0.31
		Ammonium	0.09	0.13	0.80	0.53	0.13	0.05	0.11	0.73	0.01	0.09
	PSAP	eBC	-0.15	0.08	0.75	0.48	-0.15	-0.08	0.10	0.83	0.08	-0.13
	SUNSET	EC	-0.09	0.14	0.80	0.36	-0.18	0.09	0.10	0.81	0.00	-0.03
Organic tracers	IC	MSA	0.82	0.16	-0.11	0.04	0.98	0.46	0.13	-0.09	0.10	0.77
		Amines_TOT	0.64	0.41	0.09	0.10	0.32	0.85	0.29	-0.05	0.24	0.75
	NMR	MSA	0.84	0.22	-0.08	0.09	1.00	0.50	0.18	-0.08	0.16	0.81
		Levoglucosan	-0.14	0.36	0.54	0.40	-0.06	0.03	0.34	0.57	0.38	0.00
		HMSA	0.38	0.44	0.39	0.36	0.35	0.43	0.37	0.43	0.33	0.45
		TMA	0.67	0.73	0.14	0.38	0.59	0.84	0.67	0.08	0.50	0.84
		DMA	0.50	0.83	0.20	0.60	0.50	0.66	0.78	0.17	0.63	0.66
		Amines_TOT	0.64	0.80	0.17	0.48	0.58	0.81	0.74	0.12	0.57	0.81

* Bromide is measured by UniVE

255

260

Source-marker AMS fragments.

We provide here in Table S5 specific HR-AMS mass fragments identified in our dataset as characteristic of specific sources, which were also identified in previous studies. These fragments were selected based on their highest contribution to these factors and the dominant contribution of these factors to these fragments.

Table S5.

Factor	Characterizing peaks (up to mz 150) (*)
Factor 1 – Marine biogenic OA	CHS, CH2SO, CH3SO, CH2SO2, CH3SO2, CH4SO3
Factor 2 – Aged wildfires OA	CH5O2, C3H7O, C5H4, C4H2O, C3H7O2, C5H6O, C3H3O3, C6H7O, C5H5O2, C5H6O2, C5H7O2, C4H4O3, C5H8O2, C5H9O2, C7H8O, C6H7O2, C6H8O2, C5H5O3, C6H9O2, C5H6O3, C5H7O3, C4H9O4, C8H10O, C6H5NO2, C7H7O2, C7H8O2, C3H9O5, C7H9O2, C7H10O2, C6H7O3, C7H11O2, C6H9O3, C9H8O, C5H9O4, C8H8O2, C7H7NO2, C8H9O2, C8H10O2, C6H3O4, C7H7O3, C8H11O2, C7H8O3, C7H9O3, C6H7O4, C5H8O5
Factor 3 – Arctic haze OA	CH4, C2H6O, CH3NO2, C5H12, C5O, C3H8O2, C4O2, C5H12O, C7H16, C7H3O, C7H4O, C8H16, C8H18, C8H5O2, C9H18O, C10H22, C8H3O3, C8H4O3
Factor 4 – Background OA	C8H3O

(*) The fragment is explained by the factor for more than 60%

S3. Source contributions to OC: methods

The water soluble fraction of OC, accounting on average for 71±18% of the total OC, was quantified by AMS and NMR (compared with TOC-analyzer measurements) and further apportioned to different factors/sources using PMF. Furthermore, to quantify factors/sources contribution to the total OC, we employed a multilinear regression model (MLR). Specifically, we assume each factor has a constant recovery coefficient (RC), i.e., the reciprocal of water solubility. The sum of the product of these recovery coefficients and the corresponding factor contributions was then fitted to the total OC concentrations via Bayesian-based statistical framework Stan (Carpenter et al., 2017). This can be expressed as:

$$OC \sim N(\sum_i RC_i \cdot C_i, \sigma_{OC}^2) \tag{4}$$

where RC_i represents the recovery coefficient for factor i , C_i represents the concentrations of factor i , σ_{OC}^2 represents the variance, calculated as 10% of measured OC in our case, accounting for the residual error in the model. The Stan model was set with 4 Markov chain with 2000 iterations each to explore the solution set effectively. Detailed results with both AMS and NMR factors as input can be seen in Table S4.

The approach was applied both to the AMS/NMR factors expressed in term of $\mu\text{gC m}^{-3}$ (as already done in previous studies (Casotto et al., 2023; Cui et al., 2024) but also (for the first time) directly to NMR factors quantified in term of $\mu\text{molH m}^{-3}$, in order to try an approach to convert μmolH to μmolC without arbitrary assumptions (on the stoichiometric ratio of the different functional groups, as mentioned in Section 2.2 of the main text). It should be noted that the fitting coefficients starting from

concentrations in $\mu\text{molH m}^{-3}$ are all much higher than 1 because they include also the stoichiometric H:C conversion necessary to fit OC mass. Given the fact that the OC-fitting results are very similar using both the approaches (see the comparison in Fig. S20), we eventually discuss in the main text only the version of the calculation starting from $\mu\text{gC m}^{-3}$ (Figure 6 and related discussion). As a side consideration, this consistency of results supports the reliability of the stoichiometric ratio chosen for the C quantification of the functional groups from NMR spectra (Section 2.2).

Considering the reasonable agreement between AMS and NMR reconstruction of total marine WSOC fraction (see Figure 4), we considered marine OC_{AMS} to correspond to the sum of marine primary and secondary components as apportioned by NMR (marine POC_{NMR} and SOC_{NMR}) and we then split the marine OC_{AMS} based on the NMR relative contributions of the two components, as expressed in following equations:

$$\text{marine } \text{SOC}_{\text{AMS}} = \text{marine } \text{OC}_{\text{AMS}} * \frac{\text{marine } \text{SOC}_{\text{NMR}}}{\text{marine } \text{POC}_{\text{k,NMR}} + \text{marine } \text{SOC}_{\text{k,NMR}}} \tag{5.1}$$

$$\text{marine } \text{POC}_{\text{AMS}} = \text{marine } \text{OC}_{\text{AMS}} * \frac{\text{marine } \text{POC}_{\text{NMR}}}{\text{marine } \text{POC}_{\text{k,NMR}} + \text{marine } \text{SOC}_{\text{k,NMR}}} \tag{5.2}$$

Table S6. OC fitting coefficients resulting by the multilinear regression of WSOC factors apportioned both by AMS and NMR

			OC recovery coefficients (RC) by WSOC		
			by μgC	by μmolH	water-solubility
AMS	F1	Marine OA	1.72	12.08	0.58
	F2	Aged wildfires OA	1.21	10.06	0.82
	F3	Arctic-haze OA	1.56	13.05	0.64
	F4	Background OA	2.21	20.78	0.45
NMR	F1a	Marine SOA	1.33	8.55	0.75
	F1b	Marine POA	1.55	14.44	0.65
	F2	Aged wildfires OA	2.22	20.84	0.45
	F3	Arctic-haze OA	2.04	18.80	0.49
	F4	Background OA	1.56	13.66	0.64

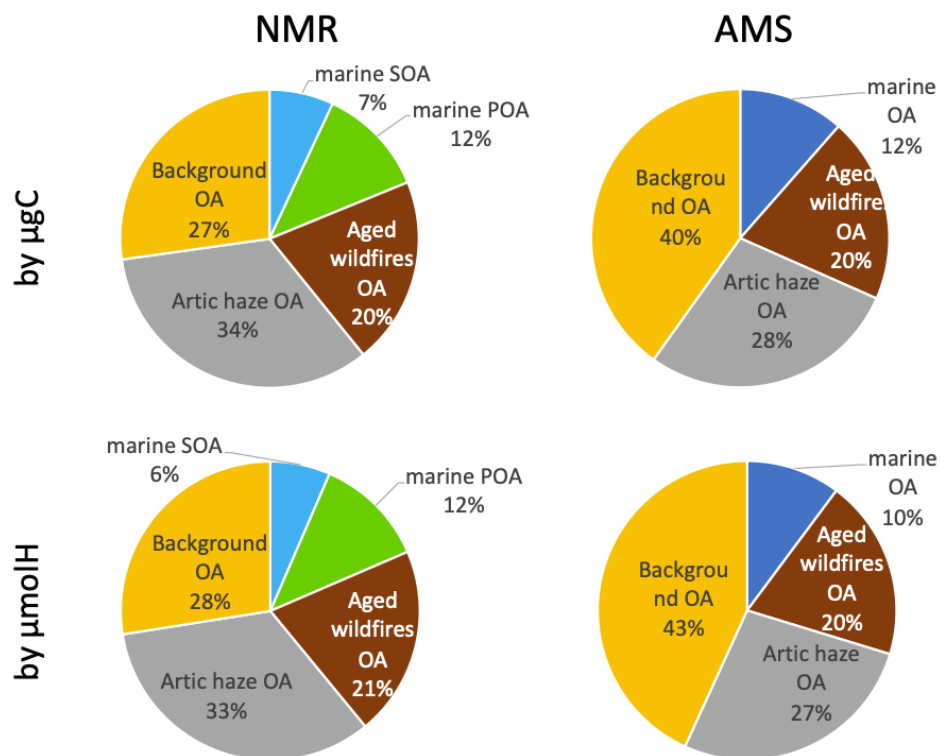


Figure S20. Comparison between results of multilinear regression model applied both to the NMR and AMS factors (left and right side, respectively), expressed in term of $\mu\text{gC m}^{-3}$ (upper charts) or alternatively in term of $\mu\text{molH m}^{-3}$ (lower charts). Pie charts report the relative contributions of the PMF-factors as annual averages.

Supplementary References

- Allan, J. D., Coe, H., Bower, K. N., Alfarra, M. R., Delia, A. E., Jimenez, J. L., Middlebrook, A. M., Drewnick, F., Onasch,
310 T. B., Canagaratna, M. R., Jayne, J. T., and Worsnop, D. R. : Technical note: A generalized method for the extraction of
chemically resolved mass spectra from Aerodyne aerosol mass spectrometer data, *J. Aerosol Sci.*, 35, 909–922, 2004.
- Bozzetti, C.; Sosedova, Y.; Xiao, M.; Daellenbach, K. R.; Ulevicius, V.; Dudoitis, V.; Mordas, G.; Byčenkienė, S.; Plauškaitė,
K.; Vlachou, A.: Argon Offline-AMS Source Apportionment of Organic Aerosol over Yearly Cycles for an Urban, Rural, and
315 Marine Site in Northern Europe. *Atmos. Chem. Phys.* 2017, 17 (1), 117– 141, DOI: 10.5194/acp-17-117-2017, 2017.
- Canonaco, F., Crippa, M., Slowik, J. G., Baltensperger, U., and Prévôt, A. S. H.: SoFi, an IGOR-based interface for the efficient
use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer
data, *Atmos. Meas. Tech.*, 6, 3649–3661, <https://doi.org/10.5194/amt-6-3649-2013>, 2013.
320
- Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, Riddell A. Stan: A
Probabilistic Programming Language. *J Stat Softw.* 2017;76:1. doi: 10.18637/jss.v076.i01, 2017.
- Casotto, R., Skiba, A., Rauber, M., Strähle, J., Tobler, A., Bhattu, D., Lamkaddam, H., Manousakas, M.I., Salazar, G., Cui, T.
325 and Canonaco, F.: Organic aerosol sources in Krakow, Poland, before implementation of a solid fuel residential heating ban.
Science of the Total Environment, 855, p.158655, 2023.
- Crippa, M., Canonaco, F., Lanz, V. A., Äijälä, M., Allan, J. D., Carbone, S., Capes, G., Ceburnis, D., Dall'Osto, M., Day, D.
A., DeCarlo, P. F., Ehn, M., Eriksson, A., Freney, E., Hildebrandt Ruiz, L., Hillamo, R., Jimenez, J. L., Junninen, H., Kiendler-
330 Scharr, A., Kortelainen, A.-M., Kulmala, M., Laaksonen, A., Mensah, A. A., Mohr, C., Nemitz, E., O'Dowd, C., Ovadnevaite,
J., Pandis, S. N., Petäjä, T., Poulain, L., Saarikoski, S., Sellegri, K., Swietlicki, E., Tiitta, P., Worsnop, D. R., Baltensperger,
U., and Prévôt, A. S. H.: Organic aerosol components derived from 25 AMS data sets across Europe using a consistent ME-2
based source apportionment approach, *Atmos. Chem. Phys.*, 14, 6159–6176, <https://doi.org/10.5194/acp-14-6159-2014>, 2014.
- Cui, T., Manousakas, M.I., Wang, Q., Uzu, G., Hao, Y., Khare, P., Qi, L., Chen, Y., Han, Y., Slowik, J.G. and Jaffrezo, J.L.:
335 Composition and Sources of Organic Aerosol in Two Megacities in Western China Using Complementary Mass Spectrometric
and Statistical Techniques. *ACS Es&t Air*, 1(9), pp.1053-1065, 2024.
- Decesari, S., Mircea, M., Cavalli, F., Fuzzi, S., Moretti, F., Tagliavini, E., and Facchini, M. C.: Source attribution of water-
340 soluble organic aerosol by Nuclear Magnetic Resonance spectroscopy, *Environ. Sci. Technol*, 41, 2479–2484, 2007.

- Decesari, S., Allan, J., Plass-Duelmer, C., Williams, B. J., Paglione, M., Facchini, M. C., O'Dowd, C., Harrison, R. M., Gietl, J. K., Coe, H., Giulianelli, L., Gobbi, G. P., Lanconelli, C., Carbone, C., Worsnop, D., Lambe, A. T., Ahern, A. T., Moretti, F., Tagliavini, E., Elste, T., Gilge, S., Zhang, Y., and Dall'Osto, M.: Measurements of the aerosol chemical composition and mixing state in the Po Valley using multiple spectroscopic techniques, *Atmos. Chem. Phys.*, 14, 12109–12132, <https://doi.org/10.5194/acp-14-12109-2014>, 2014.
- Decesari, S., Paglione, M., Rinaldi, M., Dall'Osto, M., Simó, R., Zanca, N., Volpi, F., Facchini, M. C., Hoffmann, T., Götz, S., Kampf, C. J., O'Dowd, C., Ceburnis, D., Ovadnevaite, J., and Tagliavini, E.: Shipborne measurements of Antarctic submicron organic aerosols: an NMR perspective linking multiple sources and bioregions, *Atmos. Chem. Phys.*, 20, 4193–4207, <https://doi.org/10.5194/acp-20-4193-2020>, 2020.
- O'Brien, R. E., Ridley, K. J., Canagaratna, M. R., Jayne, J. T., Croteau, P. L., Worsnop, D. R., Budisulistiorini, S. H., Surratt, J. D., Follett, C. L., Repeta, D. J., and Kroll, J. H.: Ultrasonic nebulization for the elemental analysis of microgram-level samples with offline aerosol mass spectrometry, *Atmos. Meas. Tech.*, 12, 1659–1671, <https://doi.org/10.5194/amt-12-1659-2019>, 2019.
- Paatero, P.: User's guide for the multilinear engine program “ME2” for fitting multilinear and quasimultilinear models, University of Helsinki, Finland, 2000.
- Paatero, P. & Hopke, P. K. Discarding or downweighting high-noise variables in factor analytic models. *Anal. Chim. Acta* 490, 277–289, 2003.
- Paglione, M., Saarikoski, S., Carbone, S., Hillamo, R., Facchini, M. C., Finessi, E., Giulianelli, L., Carbone, C., Fuzzi, S., Moretti, F., Tagliavini, E., Swietlicki, E., Eriksson Stenström, K., Prévôt, A. S. H., Massoli, P., Canaragatna, M., Worsnop, D., and Decesari, S.: Primary and secondary biomass burning aerosols determined by proton nuclear magnetic resonance (1H-NMR) spectroscopy during the 2008 EUCAARI campaign in the Po Valley (Italy), *Atmos. Chem. Phys.*, 14, 5089–5110, <https://doi.org/10.5194/acp-14-5089-2014>, 2014a.
- Paglione, M., Kiendler-Scharr, A., Mensah, A. A., Finessi, E., Giulianelli, L., Sandrini, S., Facchini, M. C., Fuzzi, S., Schlag, P., Piazzalunga, A., Tagliavini, E., Henzing, J. S., and Decesari, S.: Identification of humic-like substances (HULIS) in oxygenated organic aerosols using NMR and AMS factor analyses and liquid chromatographic techniques, *Atmos. Chem. Phys.*, 14, 25–45, <https://doi.org/10.5194/acp-14-25-2014>, 2014b.

- 375 Paglione, M., Beddows, D. C. S., Jones, A., Lachlan-Cope, T., Rinaldi, M., Decesari, S., Manarini, F., Russo, M., Mansour, K., Harrison, R. M., Mazzanti, A., Tagliavini, E., and Dall'Osto, M.: Simultaneous organic aerosol source apportionment at two Antarctic sites reveals large-scale and ecoregion-specific components, *Atmos. Chem. Phys.*, 24, 6305–6322, <https://doi.org/10.5194/acp-24-6305-2024>, 2024.
- 380 Reff, A., Eberly, S. I. & Bhawe, P. V.: Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. *J. Air Waste Manag. Assoc.* 57, 146–154, 2007.
- Schmitt-Kopplin, P., Liger-Belair, G., Koch, B. P., Flerus, R., Kattner, G., Harir, M., Kanawati, B., Lucio, M., Tziotis, D., Hertkorn, N., and Gebefügi, I.: Dissolved organic matter in sea spray: a transfer study from marine surface water to aerosols, *Biogeosciences*, 9, 1571–1582, <https://doi.org/10.5194/bg-9-1571-2012>, 2012.
- 385 Suzuki, Y., Kawakami, M., and Akasaka, K.: ¹H NMR Application for Characterizing Water-Soluble Organic Compounds in Urban Atmospheric Particles, *Environ. Sci. Technol.*, 35, 2656–2664, <https://doi.org/10.1021/es001861a>, 2001.
- 390 Tagliavini, E., Moretti, F., Decesari, S., Facchini, M. C., Fuzzi, S., and Maenhaut, W.: Functional group analysis by H NMR/chemical derivatization for the characterization of organic aerosol from the SMOCC field campaign, *Atmos. Chem. Phys.*, 6, 1003–1019, doi:10.5194/acp-6-1003-2006, 2006.
- Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, *Atmos. Chem. Phys.*, 9, 2891–2918, <https://doi.org/10.5194/acp-9-2891-2009>, 2009.
- 395 Zhang, Q. et al. Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review. *Anal. Bioanal. Chem.* 401, 3045–3067, 2011.