Atmospheric
Chemistry
and Physics

Supplement of

Simultaneous organic aerosol source apportionment at two Antarctic sites reveals large-scale and ecoregion-specific components

Marco Paglione et al.

Correspondence to: Marco Paglione (m.paglione@isac.cnr.it) and Manuel Dall'Osto (dallosto@icm.csic.es)

S.1 Field campaign overview: sampling, aerosol measurements, meteo and air masses

Table S1. Table with temporal period (start, end) of the 23 off-line PM_1 filters collected at Signy (S) and Halley (H). Underscore highlighted at the 42 days overlap period.

| Filter number at Signy station	Date & time Start (Signy)	Date & time Stop (Signy)	Filter number at Halley Station	Date & time start (Halley)	Date & time Stop (Halley)
S1	10/12/18 18:27	14/12/18 15:18			
S2	14/12/18 15:27	18/12/18 20:39			
S3	18/12/18 20:45	21/12/18 20:46			
S4	21/12/18 20:54	24/12/18 20:45			
S5	24/12/18 20:50	28/12/18 19:13	H1	24/12/18 12:54	02/01/19 12:30
S6	28/12/18 19:20	31/12/18 18:10	H2	2/1/19 12:50	06/01/19 10:55
S7	31/12/18 18:18	3/1/19 17:32	H3	6/1/19 10:55	11/01/19 18:44
S8	3/1/19 17:39	11/1/19 15:07	H4	11/1/19 19:02	16/01/19 21:04
S9	11/1/19 15:41	17/1/19 20:59	H5	16/1/19 21:40	21/01/19 09:55
S10	17/1/19 21:17	23/1/19 20:34	H6	21/1/19 14:50	24/01/19 13:09
S11	23/1/19 20:40	31/1/19 21:10	H7	24/1/19 17:21	30/01/19 11:10
S12	31/1/19 21:01	8/2/19 18:07	H8	1/2/19 9:17	04/02/19 09:21
S13	8/2/19 18:14	15/2/19 21:39			
S14	15/2/19 21:45	20/2/19 20:54			

5

10

Table S2. Average meteorological data for the sampling periods at Halley and Signy BAS stations. Wind directions (WD) are divided in four sectors (i.e., North: 315°-45°; East:45°-135°, South:135°-225°, West:225°-315°) and the percentage of sampling time from each sector is reported for each sample.

	Samples number	Meteo variable							
		T (ºC)	RH (%)	P (mbar)	WD (sector % time)				WS (m/s)
					North	East	South	West	
Signy	S1	-0.8±2	90±6	971±8	0	3	1	96	5.5±2
	S2	0.5±2	88±6	992±9	0	13	3	84	4.7±3
	S3	0.6±2	86±5	992±5	0	0	4	96	4.4±3
	S4	-0.2±1	83±6	982±8	1	10	7	82	2.7±2
	S5	1.6±2	94±6	983±8	24	10	18	49	2.1±2
	S6	-0.2±1	92±6	982±9	12	21	4	63	3.5±3
	S7	2.2±1	87±6	985±8	24	14	4	58	2.9±2
	S8	0.9±1	90±4	977±8	10	0	1	89	4.4±4
	S9	-0.1±2	85±6	985±9	23	28	5	44	1.3±1
	S10	1.1±1	91±6	983±9	28	9	5	58	3.2±2
	S11	-0.4±2	90±5	983±8	3	21	15	62	2.6±2
	S12	1.3±1	93±6	981±9	22	9	3	66	3.0±2
	S13	0.5±1	94±7	986±8	13	11	2	75	3.7±2
	S14	0.0±1	92±5	978±8	9	43	8	40	0.8±1
Halley	H1	-4.2±2	81±4	986±8	11	39	6	45	0.5±1
	H2	-5.2±2	86±6	977±8	1	91	8	0	2.8±3
	H3	-5.1±2	80±4	986±8	17	36	23	24	0.2±2
	H4	-5.0±1	83±6	987±8	17	54	3	27	0.9±2
	H5	-2.7±1	82±7	978±8	2	83	15	0	7.1±2
	H6	-4.7±1	76±7	982±8	12	83	5	0	2.7±2
	H7	-5.2±1	83±6	983±8	9	62	26	4	1.8±2
	H8	-6.0±2	88±6	984±8	3	89	8	0	3.3±2
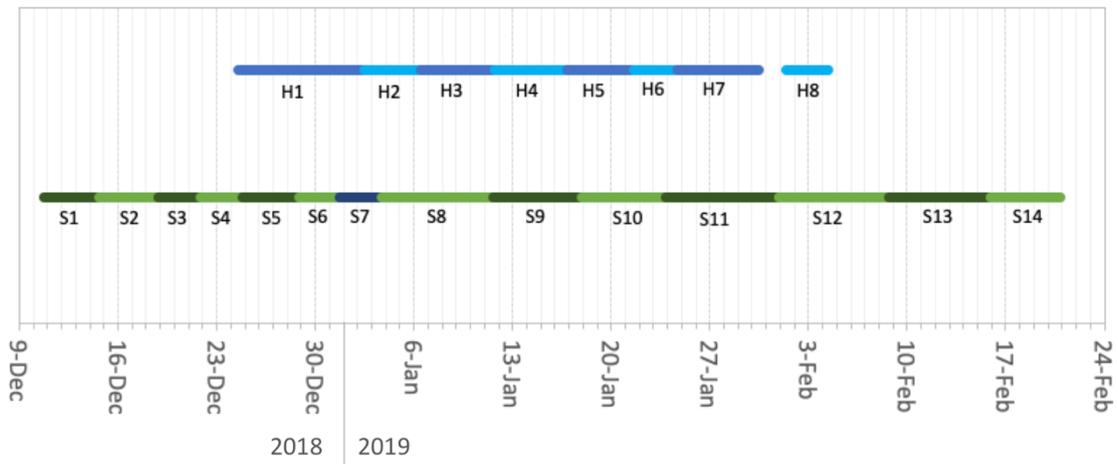
Figure S1. Graphical summary of the measurement periods of the 22 off-line PM$_1$ filters collected at Signy (S, dark-light green bars) and Halley (H, dark-light blue bars) stations.

Table S3. Ion Chromatography measured species list

ions name	ions ID	category	sea-salt components*	non sea-salt components**
acetate	ace	organic anions		
formate	for	organic anions		
methan-sulfonate	MSA	organic anions		
chloride	Cl	inorganic anions	SS_Cl	nSS_Cl
nitrate	NO3	inorganic anions		
sulfate	SO4	inorganic anions	SS_SO4	nSS_SO4
oxalate	oxa	organic anions		
sodium	Na	inorganic cations	SS_Na	nSS_Na
ammonium	NH4	inorganic cations		
methyl-amine	MA	organic cations		
ethyl-amine	EA	organic cations		
potassium	K	inorganic cations	SS_K	nSS_K
di-methyl-amine	DMA	organic cations		
di-ethyl-amine	DEA	organic cations		
tri-methyl-amine	TMA	organic cations		
magnesium	Mg	inorganic cations	SS_Mg	nSS_Mg
calcium	Ca	inorganic cations	SS_Ca	nSS_Ca

*the main ions constituting sea-salt are calculated and grouped based on the global average sea-salt composition found in Seinfel&Pandis, 2016. Briefly, Na concentrations are considered to come entirely from sea-salt. Then, starting from Na concentrations the other sea-salt components are calculated by the relative contribution to the total based on the average global composition of sea-salt (Seinfeld and Pandis, 2016). Finally, the total sea-salt is the sum of the different sea-salt components.
**non sea-salt components are calculated for each species subtracting the sea-salt part from the total concentrations

Table S4. Stoichiometric H/C ratios assigned to functional groups and molecular tracers detected by H-NMR for quantification in μgC m^{-3}

name of the *functional group /species*	ID of the functional group /species	H/C molar ratios for quantification in μgC m^{-3}
aromatic protons	Ar-H	0.4
anomeric and/or vinyl protons	O-CH-O	1
hydroxyl/alkoxy groups	H-C-O	1.1
unsaturated groups/heteroatoms	H-C-C= / H-C-X (X≠O)	2
unfunctionalized alkylic protons	H-C	2
hydroxymethansulfopnic acid	HMSA	2
methane-sufonate	MSA	3
di-methylamine	DMA	3
tri-methylamine	TMA	3

45

Table S5. 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
aromatic protons	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
anomeric and/or vinyl protons	O-CH-O	band 6-6.5 ppm	vinylic 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
hydroxyl/alkoxy groups	H-C-O	band 3.2-4.5 ppm	aliphatic alcohols, polyhols, saccharides, ethers, and esters	biogenic marine primary	Chalbot and Kavouras, 2014
benzyls and acyls/ amines, sulfonates	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
unfunctionalized alkylic protons	H-C	band 0.5-1.8 ppm	methyls (CH3), methylenes (CH2), and methynes (CH) groups of several possible molecules: fatty acids chains, alkylic 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-sufonate	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
N-osmolytes		singlets between 3.1 and 3.3	betaine, choline and other structurally similar N-containing compounds not unequivocally identified (e.g., phosphocholine)	biogenic marine primary	Cleveland et al., 2012; Chalbot et al., 2013; Decesari et al., 2020; Dall'Osto et al., 2022b
betaine	Bet	singlet at 3.25 ppm (not quantified here but possibly quantifiable)		biogenic marine primary	Cleveland et al., 2012; Chalbot et al., 2013; Decesari et al., 2020; Dall'Osto et al., 2022b
choline	Cho	singlet at 3.18 ppm (not quantified here but possibly quantifiable)		biogenic marine primary	Cleveland et al., 2012; Chalbot et al., 2013; Decesari et al., 2020; Dall'Osto et al., 2022b
saccharides	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., 2022a
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
polyols		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		specific structures between 3.6 - 3.7 ppm (not quantified)		biogenic marine primary	suggested in this study (to be confirmed)
acidic-sugars / sulfonate esters		band 4-4.3 ppm (not quantified)	uronic acids, sulfonate-derivatives of polyols	biogenic marine primary/secondary	suggested in this study (to be confirmed)
neutral sugars (saccharides) and polyols		band 3.5-3.9 ppm (not quantified)	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., 2022a
low-molecular weight fatty acids or "lipids"	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 alkanoic acids.	biogenic marine primary	Graham et al., 2002; Facchini et al., 2008b; Decesari et al., 2011; Decesari et al 2020; Liu et al., 2018
lactic acid	Lac	doublet 1.37-1.36 ppm & quadruplet at 4.23 ppm (not quantified but possibly quantifiable @1.37-1.36 ppm)		biogenic marine primary	Suzuki et al., 2001; Decesari et al., 2020; Dall'Osto et al., 2022a
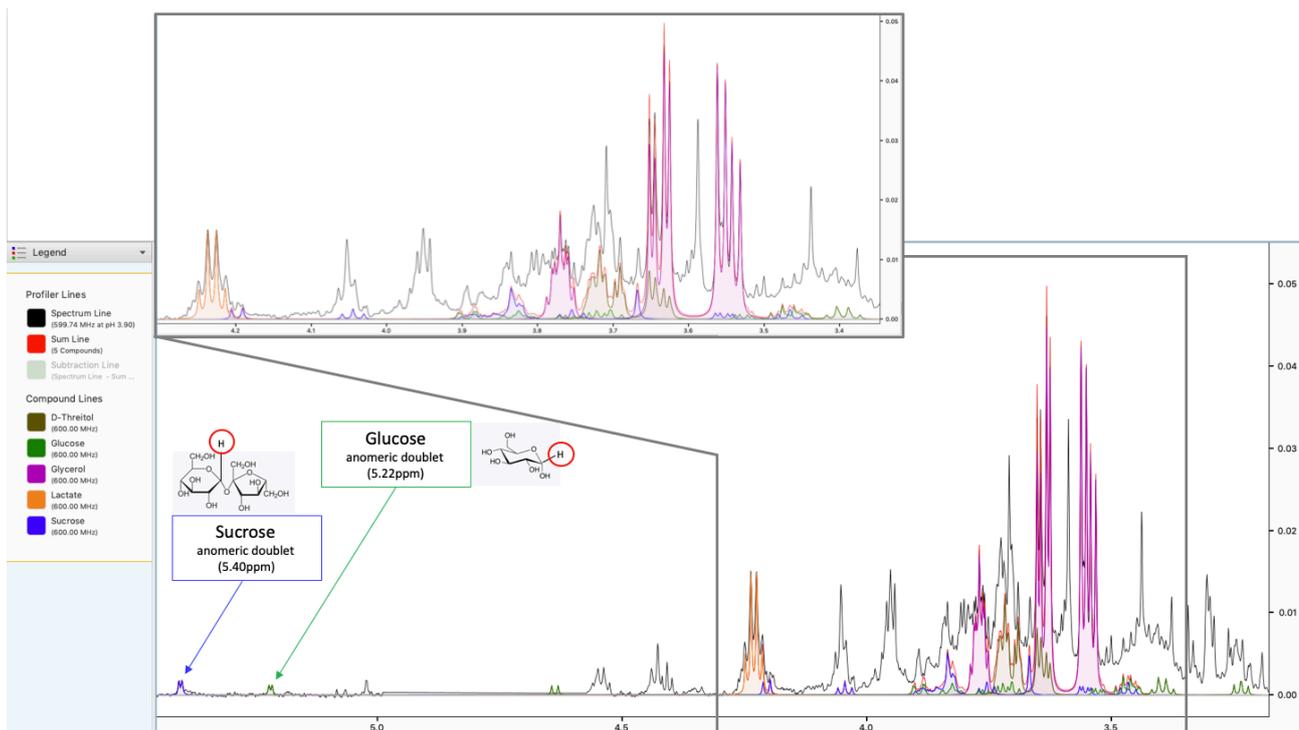
Figure S2. Example of identification of possible tracers using the extensive libraries of compounds offered by Chenomx NMR suite (Chenomx inc., evaluation version 9.0). In this figure are shown the expected NMR spectral patterns of some sugars and polyols, specifically sucrose (blue line), glucose (green line), glycerol (magenta line), D-threitol (brownish line) and lactate (orange line), against the NMR spectrum of PM1 sample S4 (black line). Sucrose and glucose molecular structures are also drawn in the figure, highlighting (with the red circles) the anomeric hydrogen used for their identification.
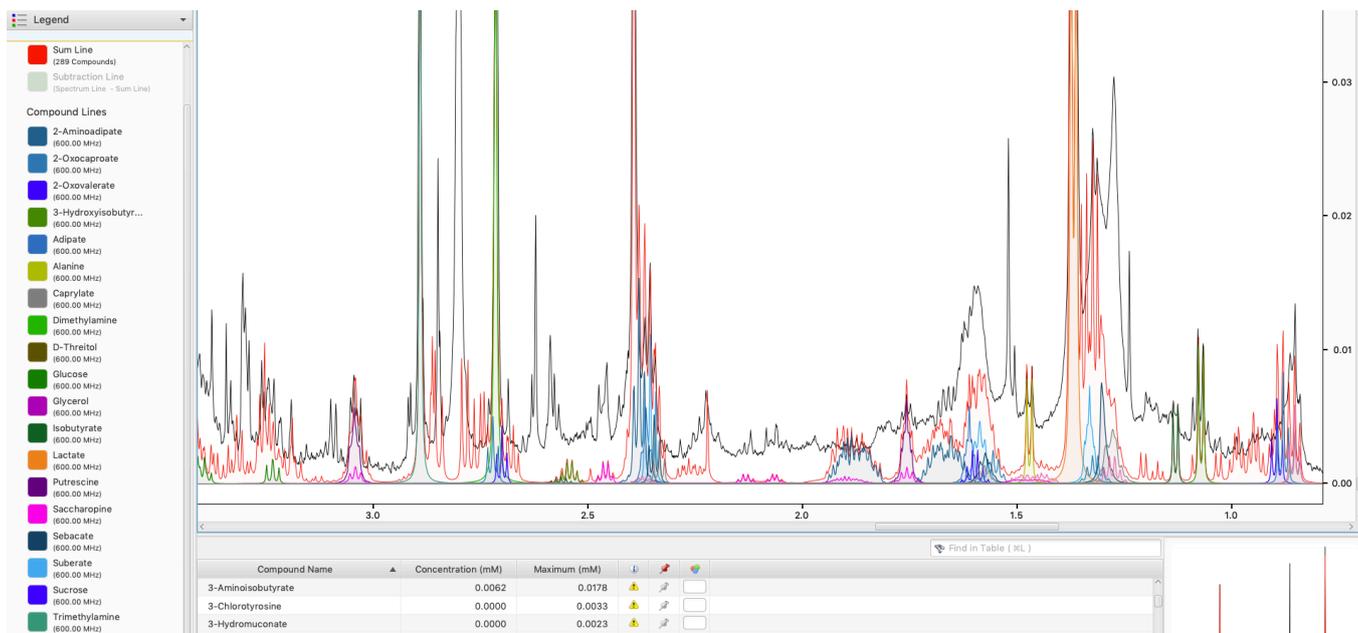
Figure S3. Another example, similar to previous figure, of identification of possible tracers using the extensive libraries of compounds offered by Chenomx NMR suite (Chenomx inc., evaluation version 9.0). Here it is reported an attempt of fitting the ambient PM1 spectrum of sample S4 with the signals 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 such as caproate, caprylate, suberate, sebacate, etc.
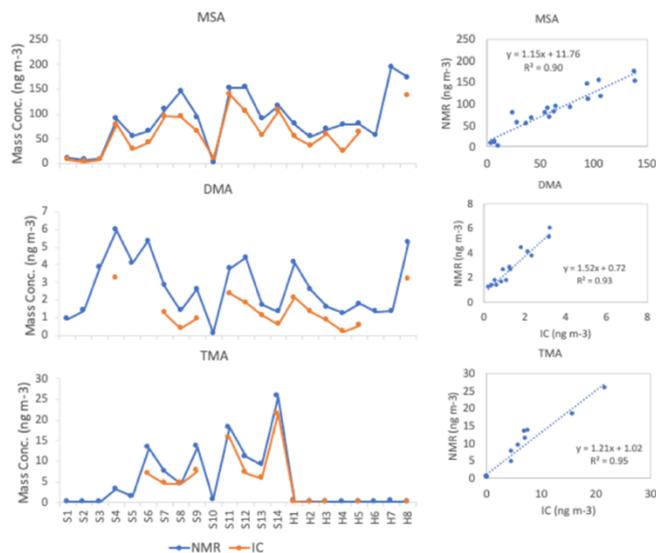
Figure S4. Comparison between mass concentrations of alkylamines and MSA identified and quantified by NMR and IC analyses

S.2 Factor Analysis of H-NMR Spectra

Factor analysis, in the broad sense, includes several multivariate statistical techniques that have been extensively used in the last years in atmospheric sciences for aerosol source apportionment on the basis of the internal correlations of observations at a receptor site, or receptor modelling (Viana et al., 2008; Belis et al., 2019). Starting with the principal component analysis (PCA), recent developments are designed to be especially applicable to working with environmental data by forcing all the values in the solutions to be non-negative, which is more realistic and meaningful from a physical point of view.

The application of non-negative factor analysis techniques to NMR spectral datasets is relatively new for atmospheric sciences, even though being widely employed in other fields, especially in biochemistry. In the present study, we employed factor analysis to analyze the collection of 22 NMR spectra of PM1 samples collected at both Signy and Halley, following the method already described in previous publications (Decesari et al., 2011; Finessi et al., 2012; Paglione et al., 2014a) and briefly reported also here below.

Preparation of input matrices: NMR spectra pre-processing

The original NMR spectra were subjected to several preprocessing steps in order to remove spurious sources of variability prior to the application of factor analysis. 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. Signals associated to blanks (Ar-H at 8.14-8.10, 7.69-7.62, and 7.38-7.36 ppm; vynil-anomeric at 6.43-6.39, 6.20-6.16, and 5.98-5.96ppm; HC-C=O at 2.38-2.36 ppm) were removed because considered not environmentally relevant. 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 factor analysis. The factor analysis techniques used in this study include two different algorithms: the "multivariate curve resolution" (MCR), according to the classical alternating least-square approach (Jaumot et al., 2005; Tauler 1995) and the "Positive Matrix Factorization" approach (PMF, Paatero and Tapper, 1994) by applying the Multilinear Engine 2 solver (ME-2, Paatero, 2000) controlled within the Source Finder software (SoFi v4.8, Canonaco et al., 2013; Crippa et al., 2014). Regardless of the specific algorithms, the methods of factor analysis are based on the same bilinear model that can be described by the following equation (S1):

$$x_{ij} = \sum_{k=1}^{P} g_{ik} f_{kj} + e_{ij}$$

(Eq. S1)

where x_{ij} refers to a particular experimental measurement of concentration of species j (one of the analytes or, here, one signal of the NMR spectrum) in one particular sample i. Individual experimental measurements are decomposed into the sum of p contributions or sources, each one of which is described by the product of two elements, one (f_{kj}) defining the relative amount

of the considered variable j in the source composition (loading of this variable on the source) and another (g_{ik}) defining the relative contribution of this source in that sample i (score of the source on this sample). The sum is extended to k=1,…,p factors (or "sources"), leaving the measurement unexplained residual stored in e_{ij} (so, with $e_{ij}= x_{ij}-g_{i,k}*f_{k,j}$).

The mathematical goal of every 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. Q may be defined in different ways depending on model's approach but it is substantially always the sum of squared residuals:

$$Q_{MCR} = \sum_{i=1}^{m} \sum_{j=1}^{n} (e_{i,j})^2 \qquad \text{(Eq. S2)}$$

where e_{ij} is the measurement unexplained residual, n is the number of samples and m is the number of species.

PMF incorporates in the calculation an evaluation of the "uncertainties" ($s_{i,j}$) associated with every measurement and so defines Q as:

$$Q_{PMF} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{e_{i,j}}{s_{i,j}}\right)^2 \qquad \text{(Eq. S3)}$$

where s_{ij} is the uncertainty of the jth species concentration in sample i,

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 and 2014b). Briefly, the uncertainty is calculated for each sample as 7 times the standard deviation of the signal in a portion of the spectrum containing only noise/spurious signal (usually between 6.5 and 7ppm).

Choice of the best solution, residuals analysis and interpretation of the factors as OA components

Solutions with different number of factors (p= from two up to eight) were explored for the spectral datasets. Eventually, a five-factors solution was chosen because of the best separation of interpretable spectral features and of the best agreement between the two algorithms applied with respect to both spectral profiles and contributions. The 4-factors solution (p=4) was also considered, but rejected in the end because not able to separate the POA enriched of lipids, polyols and saccharides from the POA-SOA 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 S5) and in disagreement between the two methodologies of factor analysis applied.
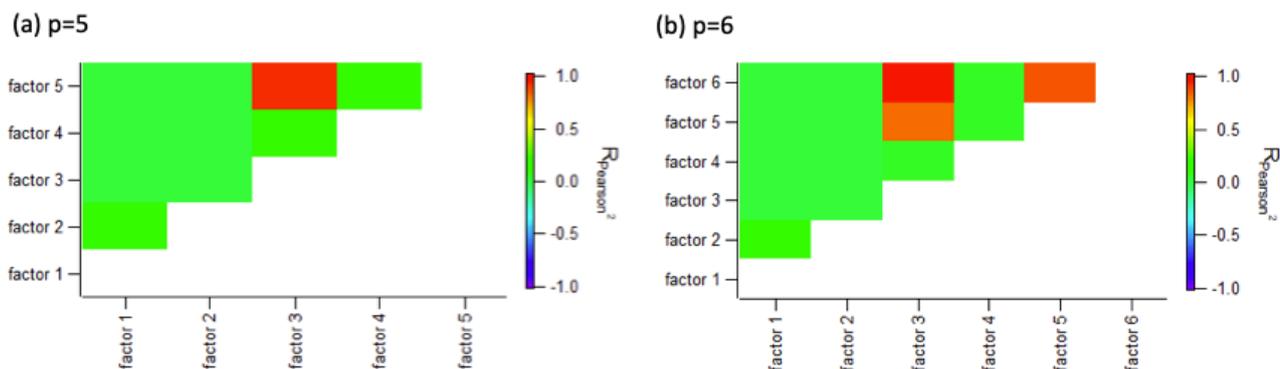
Figure S5. Correlation coefficients (Pearson R^2) between NMR factor profiles of different solutions by PMF ME-2: (a) five-factors solution (p=5), eventually chosen as the best solution; (b) six-factors solution.

Mathematical metrics were also used to support the determination of the best number of factors. A first standardized criterion was the inspection of Q-values, i.e., the total sum of squares residuals (Paatero et al., 2002). Q is expected to decrease when increasing the number of factors. However, spurious solutions provide only minor decreases in Q, whereas genuine factors explain a significant fraction of the total variance and their inclusion is generally reflected by a marked decrease in Q. Therefore, the visual inspection of the curve Q-values versus number of factors often provides a straightforward manner to highlight to number of "genuine factors" (Paatero and Tapper, 1993). In PMF the Q-value is usually evaluated as the ratio between total sum of scaled residuals (Q_{PMF} in equation S3) and the theoretical Q-value, also called "Q-expected" (Q_{exp}). The theoretical Q-value (Q_{exp}) is considered to be approximately equal to the number of degrees of freedom and can be calculated by:

$$Q_{exp} = nm - p(n+m) \qquad \text{(Eq. S4)},$$

where n is again the number of samples, m is the number of species/variables in the dataset, and p is the number of factors fitted by the model (Paatero and Hopke, 2009). In this study, the Q/Q_{exp} values for the NMR factor analysis (averaged between the two methods, Figure S6(a)) suggest that a number of factors higher than five does not significantly improve the goodness of fit. It is worth noting that to have comparable numbers for both the factor analysis methods, in Figure S6(a) the Q/Q_{exp} values for MCR were calculated using the Q_{PMF} definition (Eq. S3), starting from the residuals by the MCR model outputs. Figure S6 shows also the residuals of PMF ME-2 and MCR-ALS modelled 5-factor solutions, in term of their frequency distribution (panel (b)) and their values among samples (panel (c)) and variables (NMR spectral signals, panel (d)). The scaled residuals resulted to be mostly symmetrically distributed within a range of -3 to +3, as expected for a good solution. Moreover, residuals look to be quite randomly distributed between samples and variables, without any clear structures/patterns.
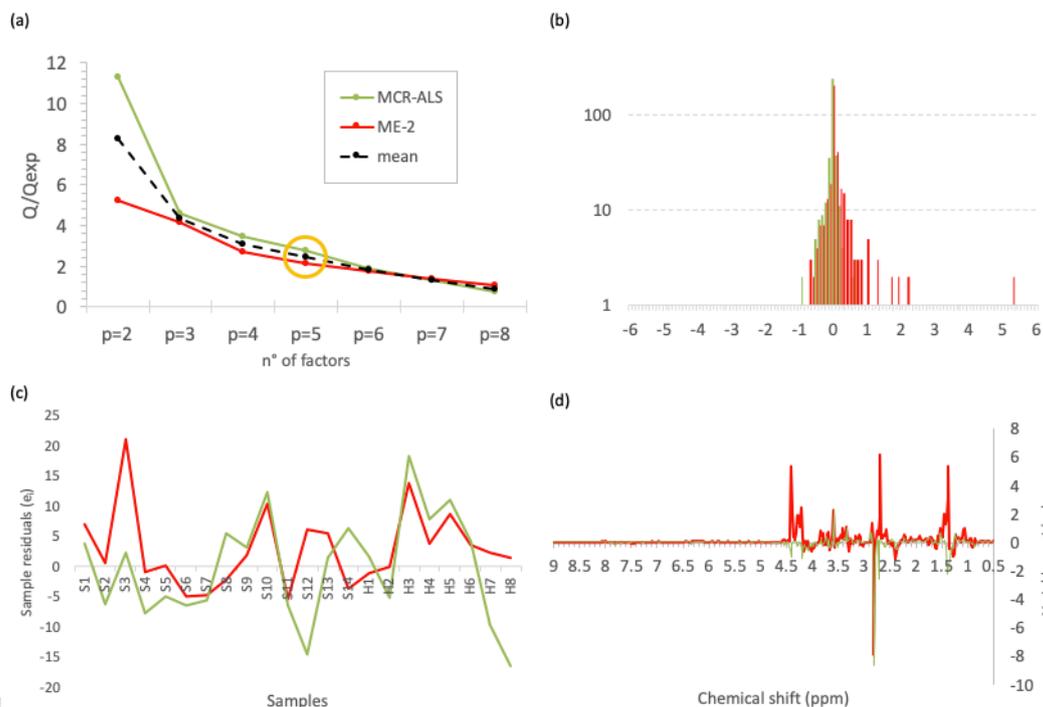
Figure S6. NMR factor analysis Q-values and residuals plots: (a) Q/Q$_{exp}$ ratio versus the number of factors p. Black dashed line represents average values between the two methods applied (MCR-ALS in green and PMF ME-2 in red). Yellow circle denotes the chosen solution (p=5); (b) frequency distribution of the scaled residuals; (c) residual values among samples; (d) residual values among variables (NMR spectral signals).

The interpretation of 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 (Facchini et al., 2008b; Schmitt-Kopplin et al., 2012; Decesari et al., 2020).

Figure S7 reports profiles and contributions of the H-NMR PMF factors identified. In particular, the Factor 1 is mainly characterized by the presence in the spectral profile of bands at 0.9, 1.3 and 1.6 ppm, corresponding to aliphatic chains with terminal methyl moieties typical of fatty acid esters (such as caproate, caprylate, azelate, suberate, sebacate etc.) which are interpreted as degradation products of lipids, and at 3.2-3.8 ppm characteristics of sugars and polyols. Fatty acids/lipids and polyols enrichment has already been documented in sea-spray aerosol from bubble-bursting experiments by previous studies reporting NMR compositional data (Facchini et al., 2008b; Schmitt-Kopplin et al., 2012; Decesari et al., 2020) as also shown in Figure S8 comparing Factor 1 profile with NMR spectra by bubble bursting experiments from previous studies (Decesari et al., 2020; Dall'Osto et al., 2022a).
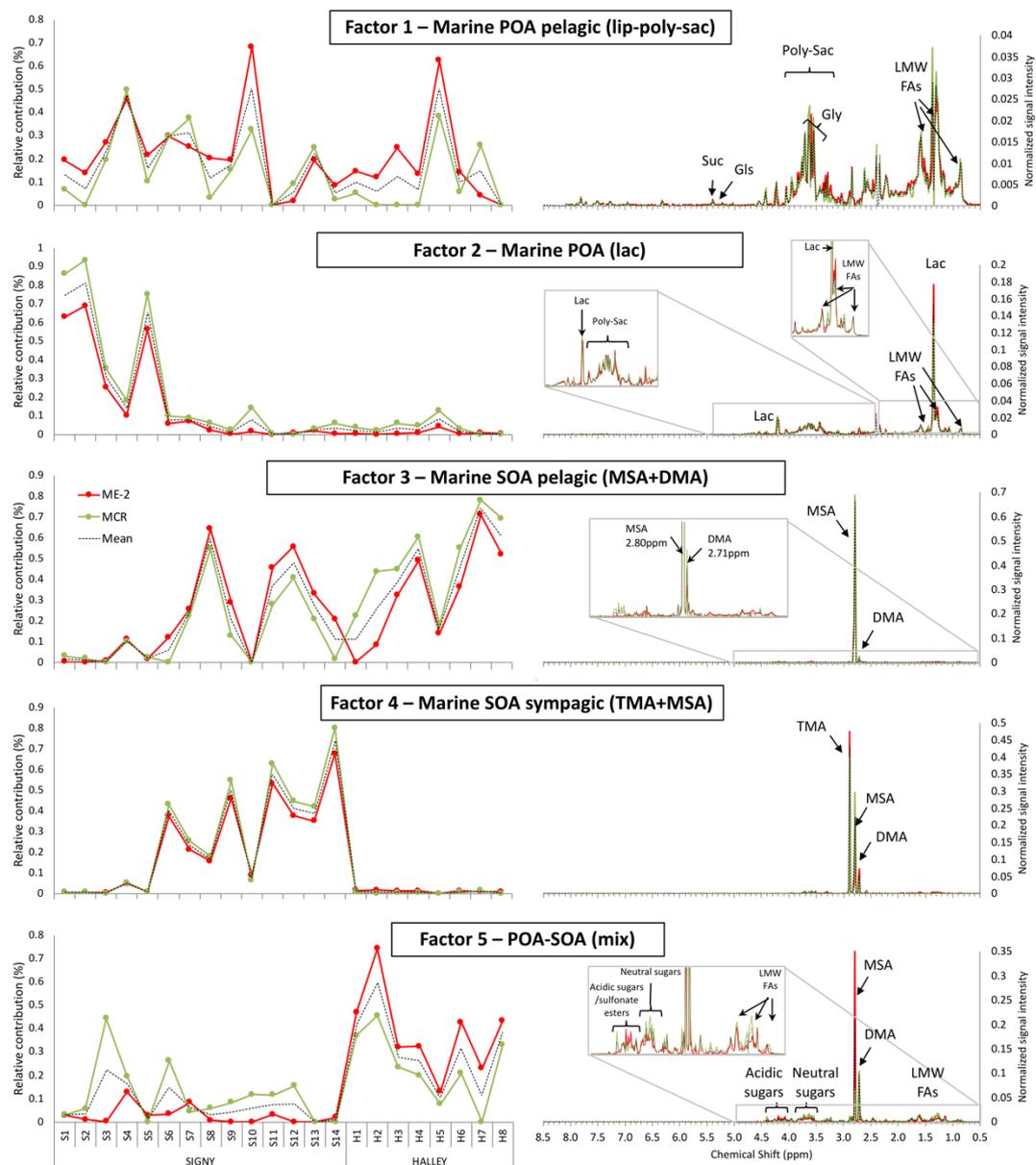
Figure S7: Profiles and contributions of the 5-factors solution from NMR spectra factor analysis. Results from the two different algorithms and the average between them are reported: PMF ME-2 (red line), MCR-ALS (green line), and average value (black dashed line) in each graph. H-NMR peaks of individual compounds (MSA: methane-sulfonate; DMA & TMA: di- and tri- methylamines; Lac: lactic acid; Gly: glycerol; Suc: sucrose; Gls: glucose) are specified in the profiles, along with the band of unresolved mixtures: LMW-FAs (low-molecular weight fatty acids), acidic and neutral sugars and generic Poly-Sac (polyols-saccharides).
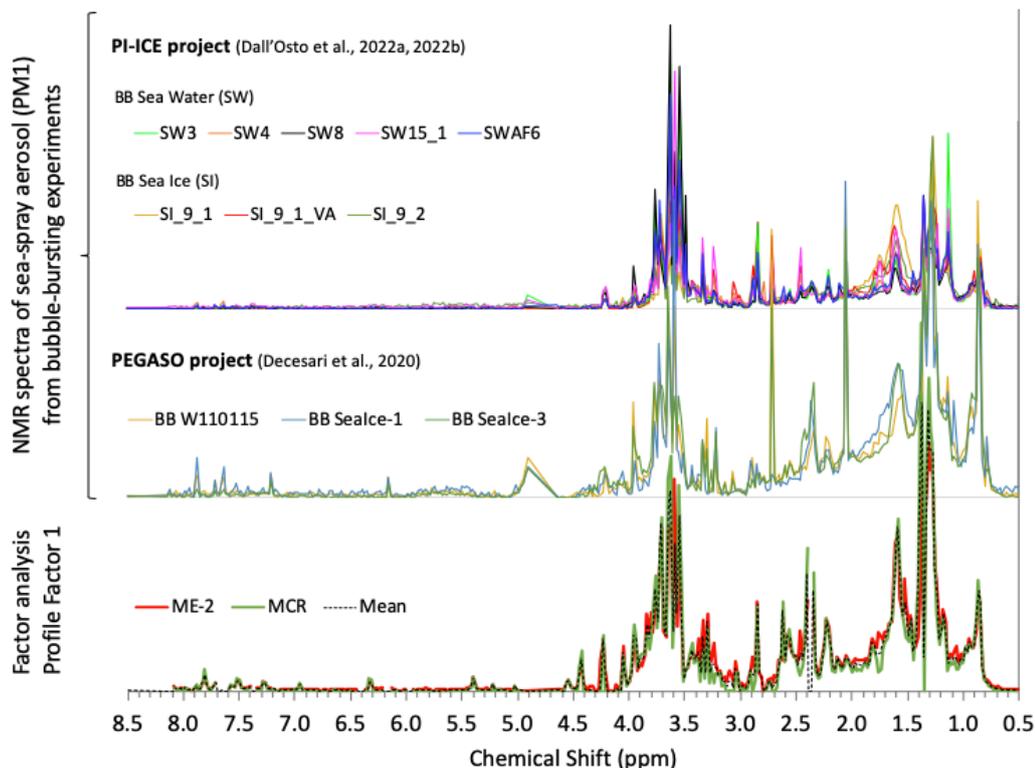
13

Figure S8: comparison between the profile of Factor 1 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).

It is plausible that glycerol, and other polyols or sugars (i.e., sucrose, glucose) together with some osmolytes (such as betaine) identifiable in the NMR spectra, have a chemical bond to lipids, making glycolipids and phospholipids.

For this reason, Factor 1 is considered as a Marine Primary Organic Aerosol (POA) factor impacting both Signy and Halley and so representing a background component in the region. Moreover, looking at the CWT maps (Figure 8) this POA component is more associated with air-masses coming from the pelagic open ocean regions (North-Western from Signy and Eastern from Halley. For all these reasons Factor 1 is called "Marine POA pelagic (lipids-polyols-saccharides)".

Factor 2 then, representing a significant portion (up to ~70%) of some samples especially in Signy (i.e., S3-S5), shows a mixture of lipids and polyols, similar to Factor 1 even if in lower proportion. But it shows also important differences, with a substantial contribution of lactic acid signals (at 1.35 and 4.21ppm). Lactic acid - a major product of sugars fermentation common to many microorganisms (Miyazaki et al., 2014) - was already identified in sea-water and sea-spray aerosol samples of the region and considered of primary biogenic origin (Decesari et al., 2020). For these reasons, and given that this factor was characterizing especially the first sampling period at Signy dominated by primary components (both organic and inorganic – sea-salt), Factor 2 is considered as another marine POA component more characteristic of specific areas around Antarctic

Peninsula (as highlighted by CWT maps in Figure 8) and in fact influencing only few Signy samples (i.e., S1-S5). This factor was so called "Marine POA (lac)".

Factor 3 and Factor 4 profiles are instead dominated by methane-sulfonate (MSA), with its specific singlet at 2.80ppm, and by low molecular methylamines (especially DMA and TMA), characterized by singlets at 2.71 and 2.89ppm, respectively. The predominance of these compounds indicates marine biogenic secondary formation processes for these factors (both representing Marine SOA). But interestingly, Factor 3 is strongly dominated by MSA and retains especially DMA, while Factor 4 profile shows a higher impact of methylamines, especially TMA. Noticeably, looking at the contributions time series, whilst Factor 3 is present at both sites showing more or less the same trends of MSA concentrations, Factor 4 is instead characteristic of Signy only and in particular of the second sampling period, the one characterized by air masses recirculating over sympagic waters of the Weddell Sea. This confirm our previous findings in the same area pointing out to sympagic Weddell sea region as a source of biogenic organic nitrogen and in particular amines in ambient aerosols (Dall'Osto et al., 2017; Dall'Osto et al., 2019; Decesari et al., 2020; Brean et al., 2021).

Factor 5 is very characteristic of Halley samples and it is specifically identified by the signals at chemical shift between 4 - 4.5 ppm. These signals have never been observed before in ambient aerosol samples (at least for our best knowledge) and are largely missing in the Signy samples. They can be possibly attributed to acidic sugars (e.g., uronic acids, such as gluconic, glucuronic or galacturonic), having sharp signals in that region, but not perfectly matching enough to explain the band observed in Halley spectra. Considering the high abundance of nSS-SO4 and the likely corresponding acidic nature of the aerosol in Halley, a hypothesis for the occurrence of these spectral features can be the esterification of common polyols (such as glycerol) to organic sulfates. To test this hypothesis, we simulated with ACD/Labs (Advanced Chemistry Developments inc., version 12.01) the theoretical NMR shifts of glycerol and possible products of its esterification with sulfonic groups (as shown in Figure S9): this hypothetical esterification seems to confirm the appearance of NMR signals in the region 4-4.5 ppm. Adding other possible common polyols (such as erythritol and arabitol) again with their hypothetical esterification products, the simulated spectra are even more enriched of signals in the 4-4.5ppm region (Figure S10). So, we tentatively attributed those signals to a mixture of acidic sugars (e.g., uronic acids, such as gluconic, glucuronic or galacturonic) and organic sulfate (sulfate-esters). However, this attribution remains just speculative at this stage and possibly needs confirmation from additional analysis/data.

In any case, as already mentioned, alkoxyl groups even if not unequivocally identified at molecular level are usually considered as primarily emitted (confirmed also by the presence of low-molecular-weight fatty acids chains, possibly from degraded/oxidized lipids, signals in the alkyls region at 0.9, 1.3, and 1.6 ppm). But the factor 5 profile shows contemporary also some secondary features, such as MSA and DMA signals which makes the source associated with this factor of difficult interpretation. For this reason, we consider this factor 5 as a mixture of primary and secondary OA specifically characterizing Halley site and worth of a deeper investigation. Considering that this component seems to be present just in Halley samples,

we could speculate that it is a mixture of primary and secondary components coming partially from a specific local source not influencing Signy and partially from marine very processed air masses. The fact that the air masses coming to Halley had previously travelled almost entirely above the PBL (Figure S16b), supports this second possible hypothesis of Factor 5 as influenced by marine emissions transported and re-processed following a free-tropospheric circulation above Antarctica.
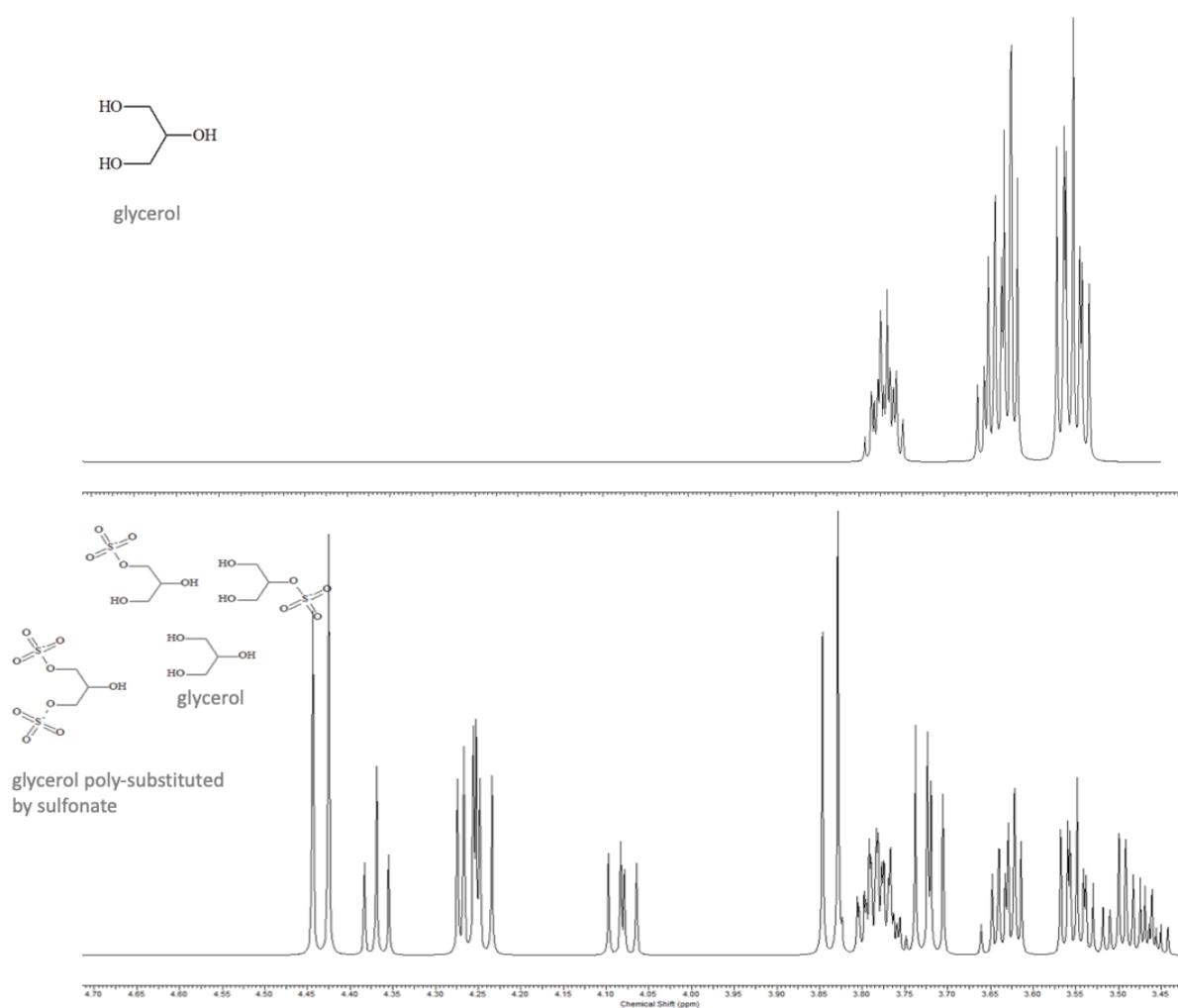

Figure S9. Simulation of theoretical NMR shifts of glycerol (upper panel) and some of its possible sulfonate esters (lower panel) using ACD/Labs tools (Advanced Chemistry Developments inc., version 12.01).
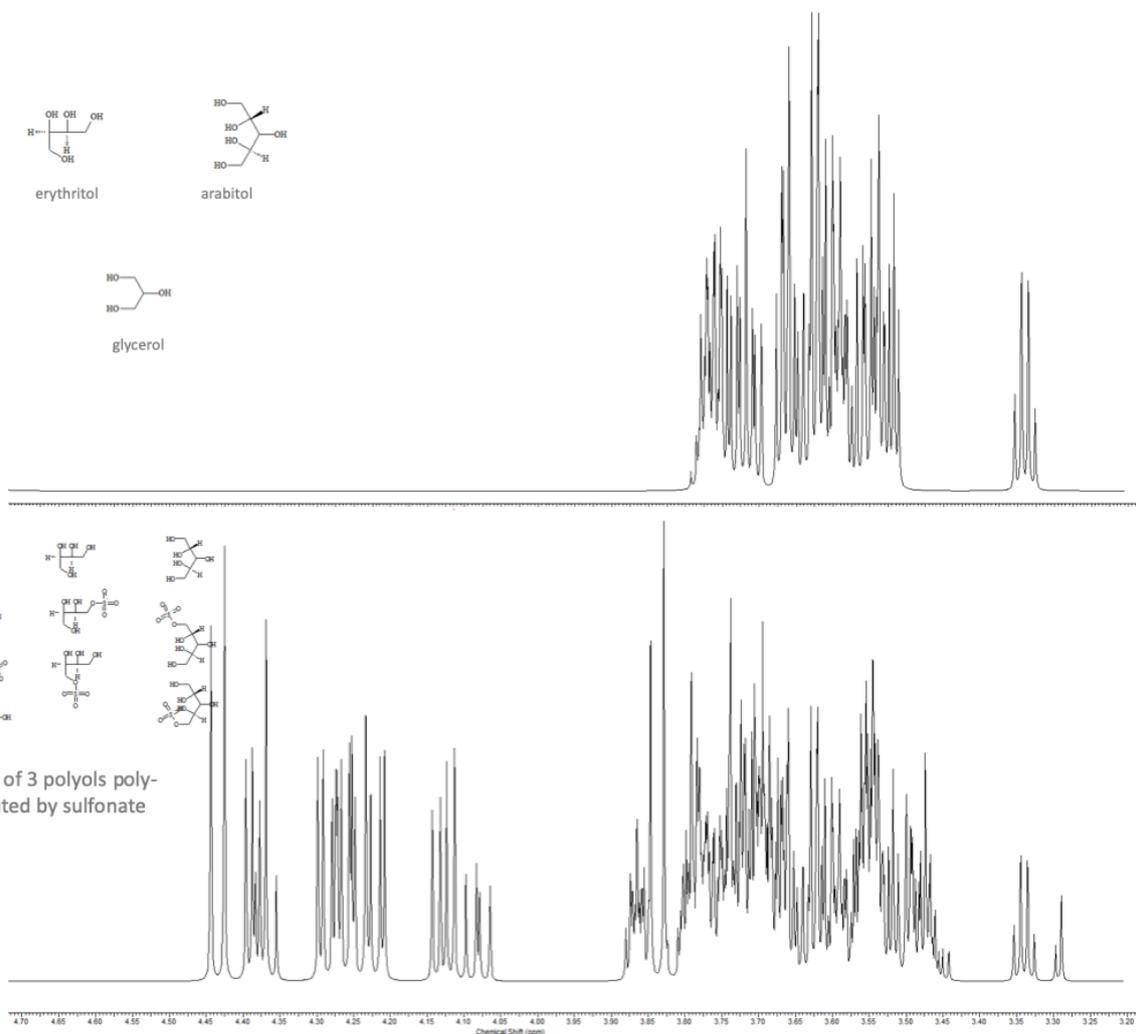
Figure S10. ACD/Labs simulation of theoretical NMR shifts of a mixture of common polyols (i.e., glycerol, erythritol and arabitol, upper panel) and some of their possible sulfonate esters (lower panel).

The interpretation of factors and their attribution to specific sources is further supported by the correlation of factors contributions with the available chemical tracers (i.e., sea salt and other inorganic ions, MSA and amines) showed in Table S6. As expected (and partially already discussed), POA components correlate with sea-salt and its main constituents of clear primary origin, while SOA factors correlate with tracers of secondary processes such as MSA and alkyl-amines.

Table S6. Pearson correlation coefficients between NMR factor contributions and ions/tracers measured by IC.

	R^2 (Pearson Coeff^{x2})	WSOM	SO4	NO3	NH4	Na	Cl	K	Mg	Ca	MSA	DMA	TMA	SeaSalt	nSS-otherIons	nSS-SO4	NO3	NH4	WS PM1
Whole dataset																			
POA_pelagic (LipoSac)	F1	0.45	0.01	0.70	0.11	0.72	0.68	0.10	0.66	0.62	0.06	0.11	0.04	0.71	0.40	0.10	0.70	0.11	0.68
POA (Lac)	F2	0.53	0.00	0.63	0.04	0.58	0.60	0.01	0.77	0.55	0.08	0.02	0.04	0.59	0.46	0.05	0.63	0.04	0.62
SOA_pelagic (MSA+DMA)	F3	0.14	0.23	0.12	0.11	0.02	0.05	0.04	0.03	0.04	0.88	0.01	0.09	0.02	0.03	0.23	0.12	0.11	0.00
SOA_sympagic (TMA+MSA)	F4	0.02	0.03	0.05	0.36	0.00	0.05	0.02	0.00	0.12	0.13	0.02	1.00	0.00	0.02	0.04	0.05	0.36	0.01
POA-SOA (Mix)	F5	0.03	0.02	0.01	0.08	0.00	0.27	0.03	0.00	0.02	0.01	0.08	0.23	0.00	0.00	0.02	0.01	0.08	0.00
Signy																			
POA_pelagic (LipoSac)	F1	0.49	0.14	0.69	0.23	0.92	0.93	0.05	0.79	0.87	0.05	0.24	0.17	0.92	0.42	0.63	0.69	0.23	0.93
POA (Lac)	F2	0.52	0.02	0.68	0.21	0.56	0.60	0.00	0.76	0.84	0.10	0.02	0.15	0.57	0.54	0.39	0.68	0.21	0.64
SOA_pelagic (MSA+DMA)	F3	0.27	0.04	0.37	0.01	0.03	0.03	0.15	0.06	0.22	0.93	0.04	0.25	0.03	0.14	0.16	0.37	0.01	0.04
SOA_sympagic (TMA+MSA)	F4	0.27	0.14	0.31	0.54	0.11	0.10	0.00	0.11	0.20	0.43	0.01	1.00	0.10	0.04	0.61	0.31	0.54	0.10
POA-SOA (Mix)	F5	0.35	0.14	0.57	0.12	0.82	0.83	0.00	0.74	0.79	0.03	0.19	0.10	0.83	0.27	0.43	0.57	0.12	0.80
Halley																			
POA_pelagic (LipoSac)	F1	0.64	0.03	0.93	0.11	0.83	N/D	0.87	0.67	0.26	0.05	0.03	0.00	0.88	0.44	0.04	0.93	0.11	0.02
POA (Lac)	F2	0.68	0.00	0.97	0.08	0.88	N/D	0.94	0.76	0.33	0.00	0.02	0.00	0.93	0.54	0.01	0.97	0.08	0.06
SOA_pelagic (MSA+DMA)	F3	0.00	0.57	0.02	0.27	0.00	N/D	0.03	0.04	0.08	0.86	0.00	0.06	0.00	0.10	0.58	0.02	0.27	0.53
SOA_sympagic (TMA+MSA)	F4	0.04	0.00	0.42	0.22	0.45	N/D	0.37	0.42	0.38	0.02	0.23	0.51	0.44	0.45	0.00	0.42	0.22	0.01
POA-SOA (Mix)	F5	0.09	0.08	0.13	0.02	0.17	N/D	0.07	0.15	0.26	0.00	0.73	0.14	0.15	0.25	0.07	0.13	0.02	0.07

Tests on robustness of the results

In order to check the possible influence of single species or single samples on the factor analysis, a series of sensitivity tests were run (using only the PMF ME-2 algorithm) and the corresponding results were compared between each other in order to find the most robust factorization. Figure S11 shows the comparison of the results on the complete dataset (already discussed) and 2 other runs in which we excluded from the PMF-input matrix: 1- the MSA signal (i.e., singlet at 2.80ppm) and, 2- the sample S3, characterized by very specific spectral features possibly influencing the factorization. Removing MSA signal the PMF best solution became a 4-factor solution, because was not possible to isolate the Factor 3 representing the marine SOA pelagic (dominated by MSA signal), but all the other factors looked in good agreement. These sensitivity analyses showed that removing single samples or variables did not change the main results, confirming the apportionment of the different factors/sources already presented. Likewise, in order to specifically check the separation between primary and secondary sources, we applied the factor analysis adding to the ambient aerosol spectra also 16 H-NMR spectra of Sea-Spray Aerosol (SSA) generated in bubble bursting tank experiments by local Antarctic sea-waters and melted sea-ice during PI-ICE project, as described by Dall'Osto et al. (2022a; 2022b and *in prep.*). Figure S11 reports the full comparison in term of both factor profiles and contributions. The results strongly confirmed the attribution of POA factors identified to primarily emitted particles resembling very well the SSA from bubble bursting experiments. Particularly significant in this regard is the fact that looking at the relative contributions of the different factors (showed in Figure S12) all the SSA samples are entirely (almost) explained by Factor 1 and 2, which are the components interpreted as POA in the solution presented in the main text.
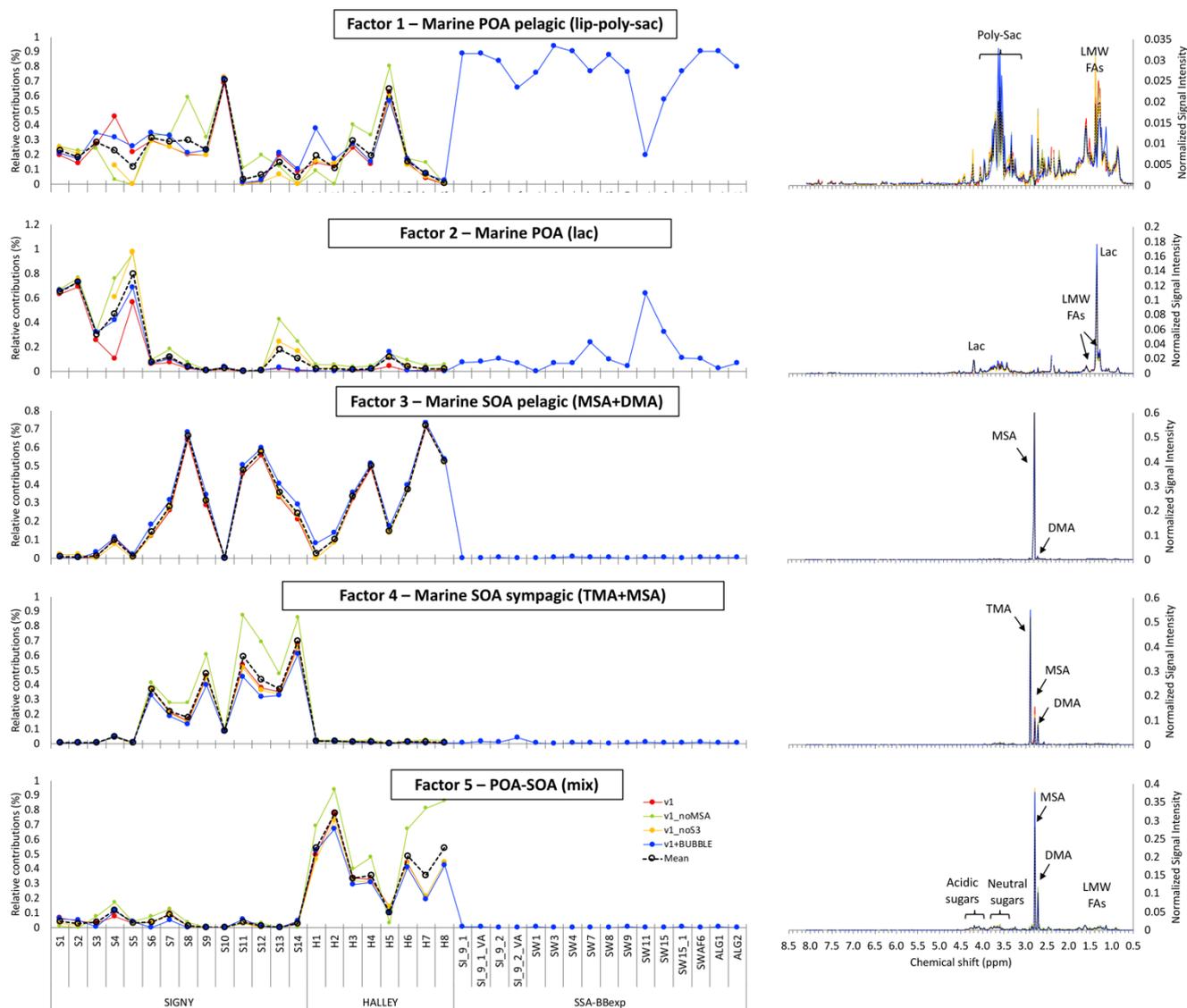
18

Figure S11: the same as Figure S7 but including results of different runs of ME-2 starting from slightly different input datasets: v1 is the solution already presented in Figure S3 and discussed in the text; v1_noMSA is the p=5 solution using NMR-spectra without the MSA signal (removing 2.79 & 2.81ppm from the input matrix); v1_noS3 is the p=6 solution using a dataset without sample S3; finally, v1+BUBBLE is the p=6 solution starting from the combined dataset of ambient-aerosol samples + sea-spray aerosol samples generated in bubble bursting experiments.
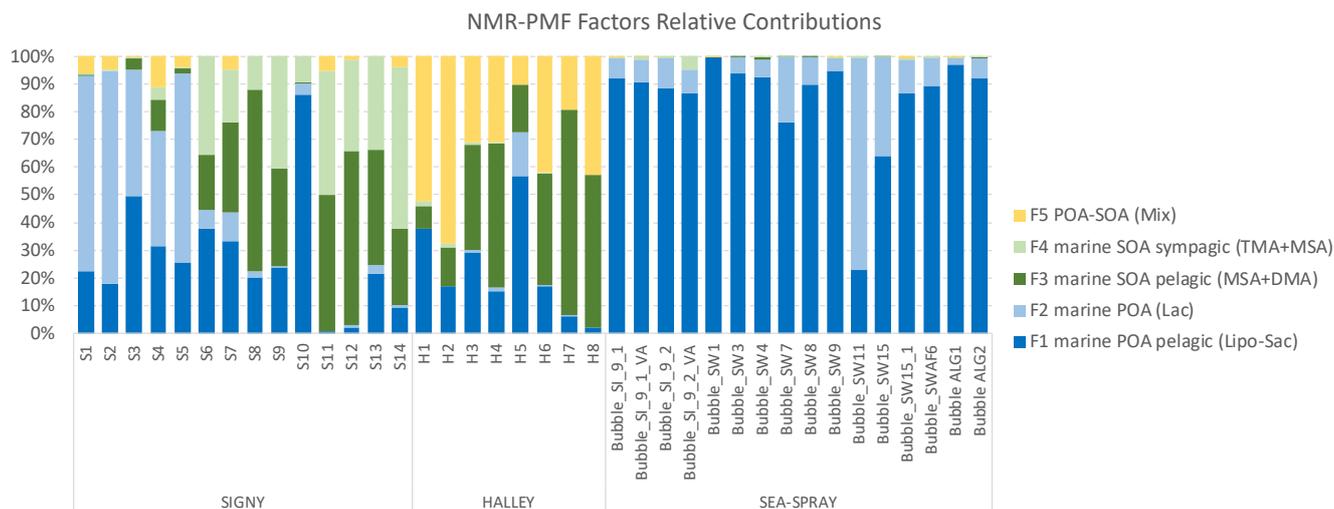
Figure S12: Factors relative contributions for the 5-factors solution using both ambient PM$_1$ samples from Signy and Halley and sea-spray samples from PI-ICE Bubble Bursting experiments (labelled as "Bubble_x").

S.3 Supplementary results and discussion
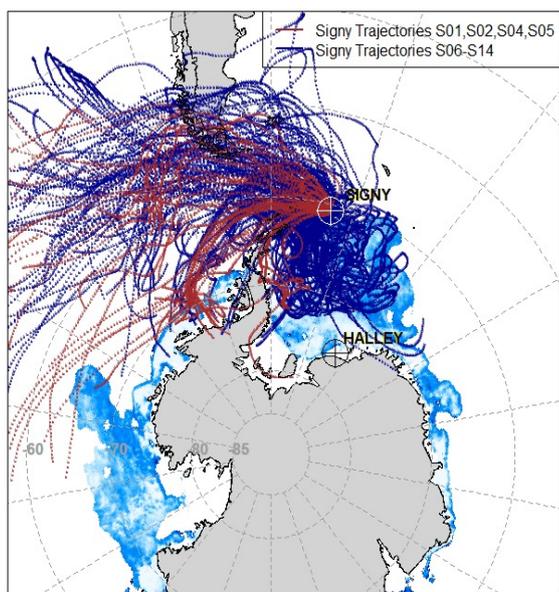
Figure S13. Air mass back trajectories for the two distinct periods sampled at Signy : first period (n=5, S1-S5) and second period (n=9, S6-S14)
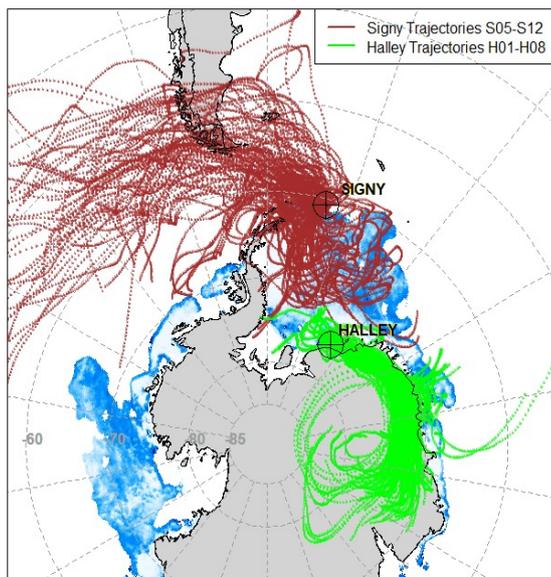
Figure S14. Air mass back trajectories for the overlapped Signy and Halley aerosol samples during approximately the same time period for Signy (n=8, S5-S12) and second period (n=8, H1-H8)
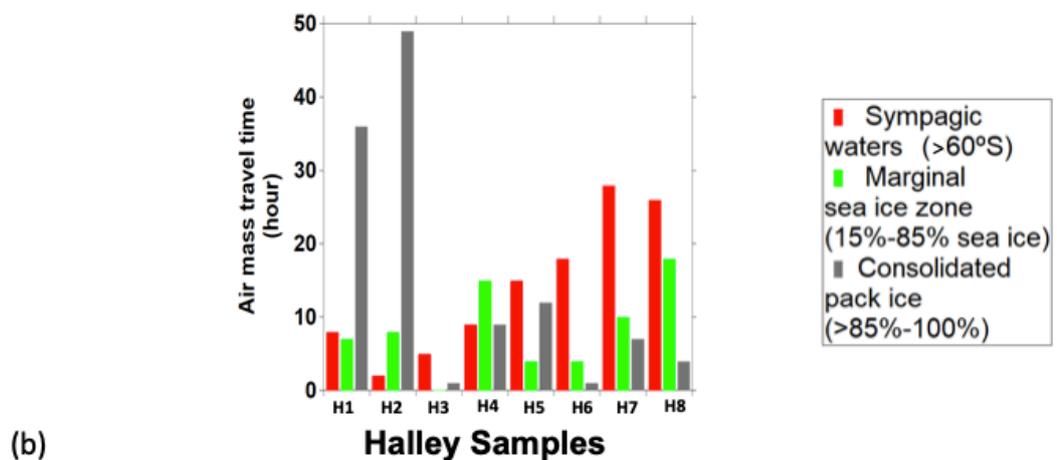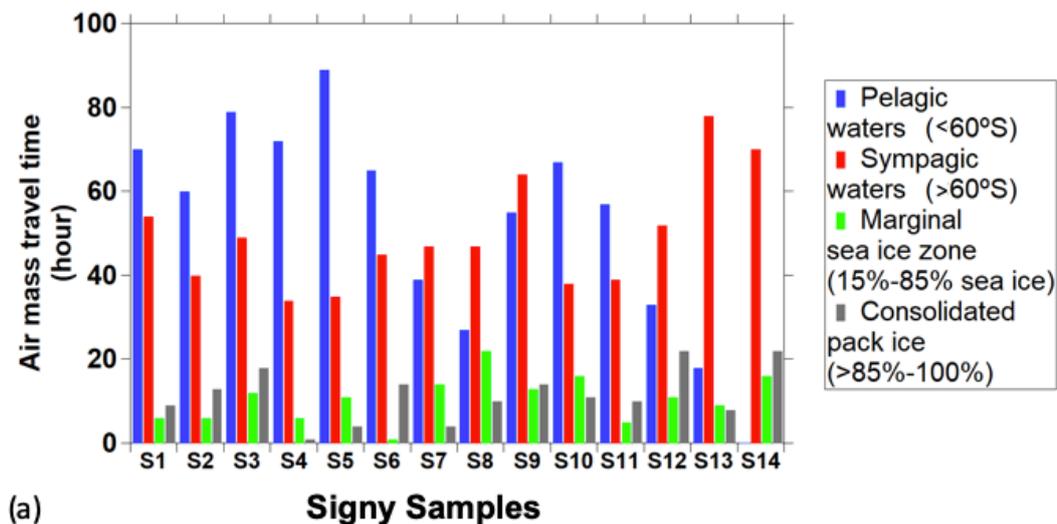
Figure S15 air mass travel time over different surfaces corresponding to (a) Signy and (b) Halley samples. All the waters more South than 60°S of latitude are defined Sympagic waters because considered influenced by sea ice even if at a lower extent than marginal sea ice zone.
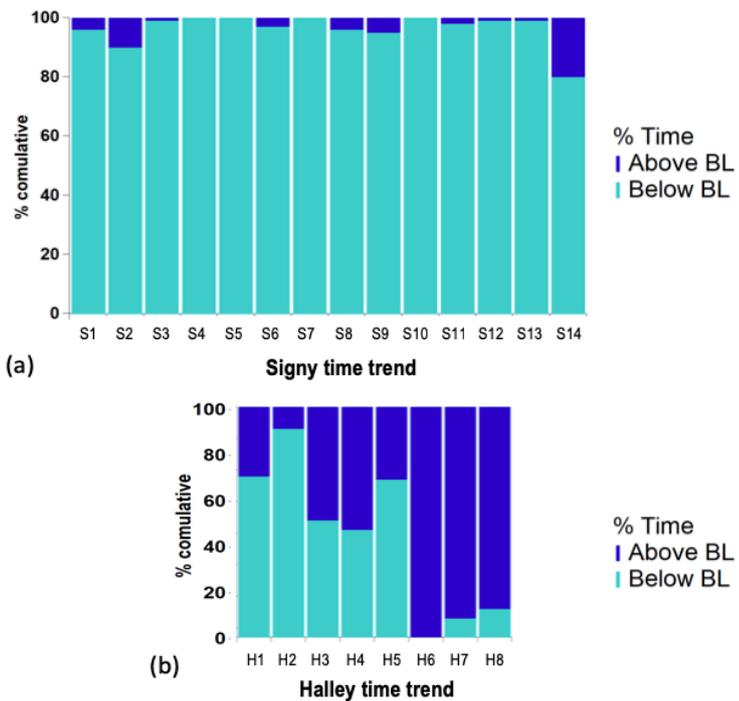
Figure S16. Air mass back trajectories time spent above and below the marine boundary layer for (a) Signy and (b) Halley.
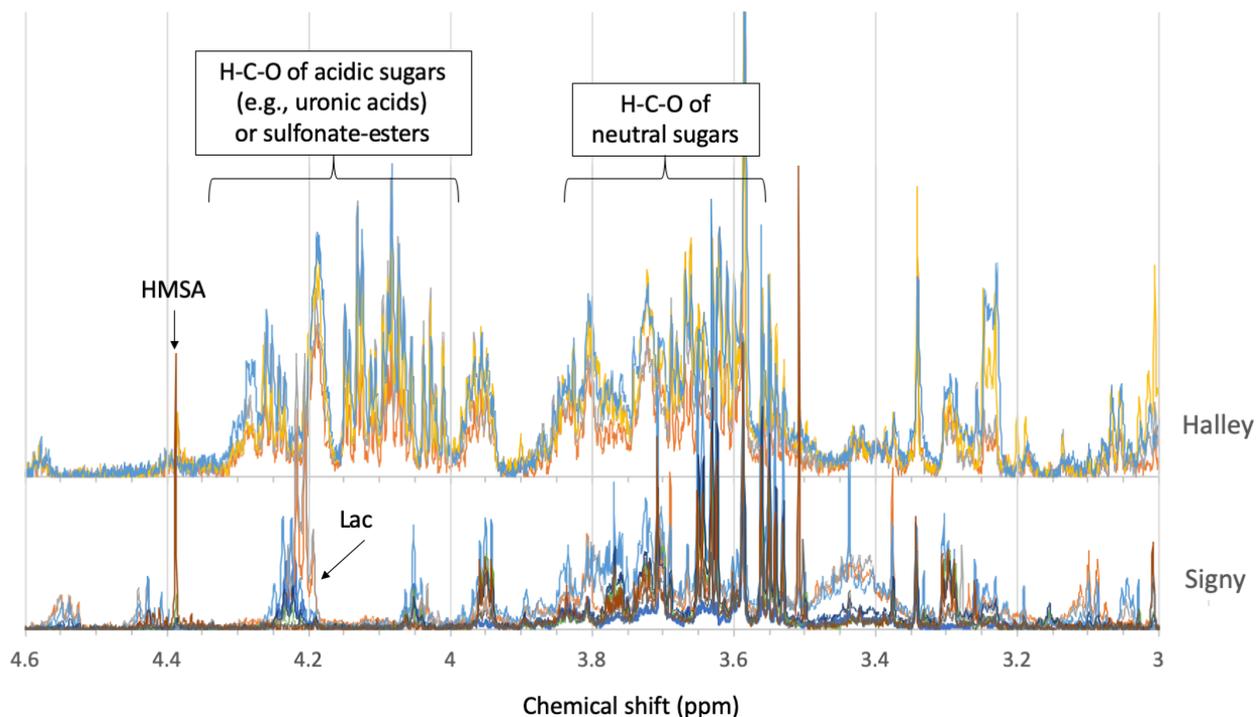
Figure S17. Alcoxy region of the H-NMR spectra of three Halley and eight Signy PM$_1$ samples. Specific NMR resonances were assigned to lactic acid (Lac) and hydroxymethane-sulphonate (HMSA). Highlighted are also the two systems of peaks tentatively linked to neutral sugars (3.5 - 3.9 ppm, such as glycerol) and acidic sugars and/or sulfate-esters (4 - 4.3 ppm).

Supplementary references

Belis C.A., Favez O., Mircea M., Diapouli E., Manousakas M-I., Vratolis S., Gilardoni S., Paglione M., Decesari S., Mocnik G., Mooibroek D., Salvador P., Takahama S., Vecchi R., Paatero P., European guide on air pollution source apportionment with receptor models - Revised version 2019, EUR 29816 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-09001-4, doi:10.2760/439106, JRC117306, 2019.

Brege, M., Paglione, M., Gilardoni, S., Decesari, S., Facchini, M. C., and Mazzoleni, L. R.: Molecular insights on aging and aqueous-phase processing from ambient biomass burning emissions-influenced Po Valley fog and aerosol, Atmos. Chem. Phys., 18, 13197–13214, https://doi.org/10.5194/acp-18-13197-2018, 2018.

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.

Chalbot, M.C., and Kavouras, I.G.: Nuclear magnetic resonance spectroscopy for determining the functional content of organic aerosols: a review; Environ Pollut. 2014, Aug:191:232-49. doi: 10.1016/j.envpol.2014.04.034.

Cleveland, M.J., L.D. Ziemba, R.J. Griffin, J.E. Dibb, C.H. Anderson, B. Lefer, B. Rappenglück Characterisation of urban aerosol using aerosol mass spectrometry and proton nuclear magnetic resonance spectroscopy, Atmos. Environ., 54 (2012), pp. 511-518, 10.1016/j.atmosenv.2012.02.074

Crippa, M., Canonaco, F., Lanz, V. A., Äijälä, M., Allan, J. D., Carbone, S., Capes, G., Ceburnis, D., Dall'Osto, M., Day, D. A., De-Carlo, P. F., Ehn, M., Eriksson, A., Freney, E., Hildebrandt Ruiz, L., Hillamo, R., Jimenez, J. L., Junninen, H., Kiendler-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.

Gilardoni, S., Massoli, P., Paglione, M., Giulianelli, L., Carbone, C., Rinaldi, M., Decesari, S., Sandrini, S., Costabile, F., Gobbi, G. P., Pietrogrande, M. C., Visentin, M., Scotto, F., Fuzzi, S., and Facchini, M. C.: Direct observation of aqueous secondary organic aerosol from biomass burning emissions, P. Natl. Acad. Sci. USA, 113, 10013–10018, 2016.

Graham B, Mayol-Bracero OL, Guyon P, Roberts GC, Decesari S, Facchini MC, Artaxo P, Maenhaut W, Koll P, Andreae MO (2002) J Geophys Res 107(D20):8047

Liu, J., Dedrick, J., Russell, L. M., Senum, G. I., Uin, J., Kuang, C., Springston, S. R., Leaitch, W. R., Aiken, A. C., and Lubin, D.: High summertime aerosol organic functional group concentrations from marine and seabird sources at Ross Island, Antarctica, during AWARE, Atmos. Chem. Phys., 18, 8571– 8587, https://doi.org/10.5194/acp-18-8571-2018, 2018.

Paatero, P.: User's guide for the multilinear engine program "ME2" for fitting multilinear and quasimultilinear models, University of Helsinki, Finland, 2000.

Paatero P. and Tapper U.: Analysis of different modes of factor analysis as least squares fit problems, Chemom. Intell. Lab. Syst., 18, 183-194, 1993.

Paatero P., Hopke P.K., Song X.H., and Ramadan Z.: Understanding and controlling rotations in factor analytic models, Chemom. Intell. Lab. Syst., 60, 253-264, 2002.

Paatero P. and Hopke P.K.: Rotational Tools for Factor Analytic Models, J. Chemom., 23, 91-100, 10.1002/cem.1197, 2009.

Schkolnik, G., Rudich, Y. Detection and quantification of levoglucosan in atmospheric aerosols: a review. Anal Bioanal Chem 385, 26–33 (2006). https://doi.org/10.1007/s00216-005-0168-5

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.

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.

Viana M., Kuhlbusch T.A.J., Querol X., Alastuey A., Harrison R.M., Hopke P.K., Winiwarter W., Vallius M., Szidat S., Prévôt A.S.H., Hueglin C., Bloemen H., Wåhlin P., Vecchi R., Miranda A.I., Kasper-Giebl A., Maenhaut W., Hitzenberger R., Source apportionment of particulate matter in Europe: a review of methods and results, J. Aerosol Sci. 39: 827–849, 2008.