

Toolbox for accurate estimation and validation of PMF solutions in PM source apportionment

Vy Ngoc Thuy Dinh¹, Gaëlle Uzu¹, Pamela Dominutti¹, Stéphane Sauvage^{1,2}, Rhabira Elazzouzi¹, Sophie Darfeuil¹, Céline Voiron¹, Abdoulaye Samaké³, Shouwen Zhang⁴, Stéphane Socquet⁵, Olivier Favez^{6,7}, Jean-Luc Jaffrezo¹

¹ Université Grenoble Alpes, CNRS, IRD, INP-G, IGE (UMR 5001), 38000 Grenoble, France

²IMT Nord Europe, Institut Mines-Télécom, Université de Lille, Centre for Energy and Environment, 59000, Lille, France

³ Atmo Normandie, 76000 Rouen, France

⁴ Atmo Hauts de France, 59044 Lille, France

⁵Atmo AuRA, 69500 Bron, France

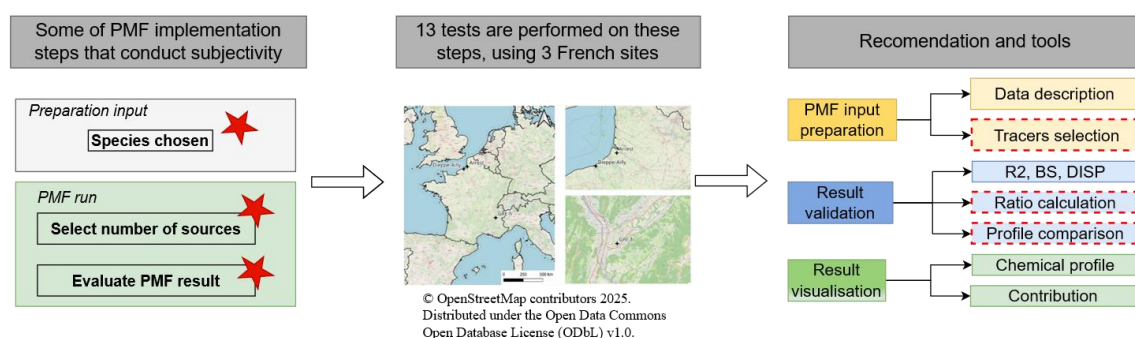
⁶INERIS, Parc Technologique Alata, BP 2, 60550 Verneuil-en-Halatte, France

⁷Laboratoire central de surveillance de la qualité de l'air (LCSQA), 60550 Verneuil-en-Halatte, France

Correspondence to: Gaëlle Uzu, gaelle.uzu@ird.fr

Abstract

Positive matrix factorization (PMF) is the most commonly used approach for particulate matter source apportionment; however, the implementation steps of the model require considerable user experience. Most studies apply PMF according to the recommendations of the Environmental Protection Agency and the European Commission, while relatively few studies focus on further developing the PMF methodology. This study aims to develop a systematic method that reduces some subjective aspects when performing a PMF study, providing recommendations and tools for its application and validation. A total of 13 targeted tests were conducted to address key sources of subjectivity in PMF, categorized into three critical aspects: preparation of the input matrix, selecting the number of sources, and validation of the PMF solution. The results of the first step highlighted that using a single source tracer reduces the tracer's dispersion into other sources, leading to more accurate results. The second stage tests suggested that the selection of a source tracer should be based on low uncertainty and specific temporal evolution, in order to facilitate the determination of a new source without compromising the PMF solution. Finally, the validation step was set up as an advanced comparison of the PMF-derived source profiles with those in the literature, including SPECIEUROPE database, using the ratio of chemicals and distance metrics. All outcomes of this study are compiled into a Python package providing essential tools to support the work from PMF implementation to solution validation, leading to less subjective solutions and more rigorous and reliable source apportionment.



Key words: PM source apportionment, PMF methodology development, PMF toolkits.

1. Introduction

Receptor models are widely used for atmospheric particulate matter source apportionment (PM SA), assigning the PM sources using measurements at a receptor site. Positive matrix factorization (PMF), developed by Paatero and Tappert (1994), has been the most popular receptor model for PM SA (Hopke, 2016). PMF is frequently performed using EPA PMF software developed by the US - Environmental Protection Agency. A Fundamentals and User Guide was launched along with the software, providing a general guide for implementing PMF for all environment domains (Norris et al., 2014). Focusing especially on the air pollution SA, the Forum for Air Quality Modeling (FAIRMODE), which is chaired by the European Commission Joint Research Centre, reported a harmonized receptor model protocol that recommended different aspects of PMF implementation, including input matrix, uncertainty calculation, and the number of sources (Belis et al., 2014). Apart from these reports, most studies essentially use PMF, with few studies considering the improvement of the methodology (Hopke et al., 2020). There remains a lack of recommendations regarding aspects that require user decisions, as some of these have not yet been systematically tested to provide 'educated guesses'.

However, some efforts to improve the PMF methodology have been established over the years, including (1) developing tools to estimate the uncertainties of PMF results and apply PMF for big data (Brown et al., 2015; Hopke et al., 2023), (2) developing tools to compare chemical profiles (Belis et al., 2015), (3) adding new tracers representing the organic sources (Borlaza et al., 2021; Glojek et al., 2024; Hu et al., 2010; Lu et al., 2018; Mardoñez et al., 2023; Samaké et al., 2019b; Wang et al., 2012), (4) standardizing the PMF implementation methodology (Chen et al., 2022; Weber et al., 2019), (5) Combining multiple sites to reduce the rotational ambiguity of the PMF (Hernández-Pellón and Fernández-Olmo, 2019; Pandolfi, 2020; Pietrodangelo et al., 2024; Dai et al., 2020), (6) incorporating meteorological data to the PMF (Dai et al., 2021) and, (7) improving result visualizations (PyPMF-<https://pypmf.readthedocs.io>, (Weber et al., 2019)).

Indeed, PMF analysis involves several steps that require subjective choices from the user. Since prior information is not required to perform a PMF study, the results obtained with the model strongly depend on these users' decisions. One example of this is the input data set chosen, which orientates the determination of the sources, as reported by Amato et al. (2024), demonstrating that the lack of tracers of a source obviously prevents its identification. It raises questions on the basic data set for the mandatory input for meaningful work. Further, the calculation method of input data uncertainties could influence the PMF solution stability, including the scaled residual and bootstrap results (Waked et al., 2014). As another example, Belis et al. (2015) indicated that the number of sources obtained can broadly vary according to the expertise of the group performing a PMF study, as

demonstrated from the comparison of PM SA results generated by 38 research groups on the same dataset. Finally, constraints that can be applied to initial results to obtain the final solution rely on user experience, and their application criteria are rarely documented in the literature; in the same way, comparison of the chemical profiles obtained in the studies is rarely benchmarked towards previous results.

The present study describes some further developments and tests in an attempt to improve the robustness of the PMF methodology by proposing pathways for narrowing the subjective choices of the users by testing some of the limits of the methodology, particularly when it comes to the choice of variables and the number of sources. We also propose some toolkits for quicker preparation, evaluation, and benchmarking of the results during the trial and error phases of the implementation of a PMF study. These different steps are presented for several studies conducted with the EPA PMF 5.0 software using databases of PM₁₀ sample series analyzed in recent years within various programs conducted in France.

2. Methodology

2.1. General organization of the approach

Performing a PMF study is a complex process, including a large number of steps with trial and error testing and feedback loops, where many implementation steps require subjectivity in the user's choices, as presented in Figure S1. These choices include, among others, input matrix preparation, selection of the number of factors, and result validation criteria. In this study, we performed tests on these different steps, with (1) Tests on the input matrix, evaluating the variations induced in the PMF solution by including or excluding specific tracers (2) Tests performed on the selection of the number of sources, providing recommendations for selecting the tracers of sources (3) PMF output evaluation by comparing PMF-derived chemical profile with that reported in the literature. The details of each test and the data used for each of them are presented in Table 1 and sections 2.1.1, 2.1.2, and 2.1.3, respectively. This work is based on the series of PM₁₀ samples collected in several projects, also detailed in Table 1.

Table 1. Tests performed, as described in sections 2.1.1, 2.1.2, and 2.1.3.

Steps	Test	Details test	Study site	Tools
Input matrix	Test 1	Using OC instead of OC*	Arrest, Ailly	PMF
	Test 2	Using only levoglucosan instead of levoglucosan + mannosan	Arrest, Ailly	PMF
	Test 3	Using Arabitol and Mannitol instead of the sum of both species	Arrest, Ailly	PMF
	Test 4	Adding Oxalate and HULIS in the input	Ailly	PMF
Number of factors selection	Tests 5, 6,7,8,9	Adding 2-MT by varying the concentration and uncertainty	Grenoble	PMF
	Tests 10, 11	Generate fake tracers (Fake 1, Fake 2) to test the capacity of	Grenoble	PMF, distance metric

		PMF in the separation of source.		
PMF solution accuracy evaluation	Test 12	Validation of the ratio between variables in the profile	Grenoble	Ratio calculation
	Test 13	Chemical profile comparison	Grenoble	Distance metric

2.1.1. Tests on input matrix

Four tests were performed to evaluate the PMF solution sensitivity by modifying the input matrix (Table 1). These tests were performed on a large dataset with about 220 samples analyzed for an extensive array of chemical components (see section 2.2.1). Tests 1 to 3 test the sensitivity of some changes in the input data for some commonly used compounds (OC, levoglucosan, polyols). Conversely, test 4 incorporates chemical species that are rarely added to PMF (oxalate, HUmic LIke Substances - HULIS) to both explore their impact on the results and investigate their sources in PM₁₀.

Many PMF studies make use of OC* instead of OC - as an input variable to avoid double-counting part of the total OC mass (Borlaza et al., 2021; Dominutti et al., 2024; Srivastava et al., 2018). To do so, authors generally retrieve from OC the carbon contents included in the used input organic variables. Such a prior data handling is notably included as a recommendation in the European guidance for PM SA (Belis et al., 2014). Test 1 compares the results of using this approach or not. In test 2, a single tracer of biomass Burning (BB) (levoglucosan) is considered instead of two (levoglucosan and mannosan), to evaluate the stability of the PMF solution to such changes. In test 3, a similar approach is adopted, comparing PMF results when using a "Polyols" input (sum of concentrations of arabitol and mannitol), instead of each chemical species separately. It is well known that these sugar alcohols are associated with biogenic activities from fungi (Samaké et al., 2019a; Yttri et al., 2007), and can then be considered as tracers for primary biogenic aerosols. In test 4, the PMF input incorporates two organic compounds that have a crucial mass contribution to OC, (i.e., HULIS and Oxalate) and are also mainly known as secondary products of primary emissions from different sources. This test aims to provide information on the stability of the factors obtained with more straightforward data sets, and also investigate the sources of these secondary organic components in the atmosphere. The inputs of Test 1 to 4 and the PMF base run are presented in Table S1.

2.1.2. Tests on the number of factors and the number of chemical tracers

The number of factors chosen for an optimal solution is commonly selected based on the statistical parameters of initial PMF runs. The basic parameter considered in this approach is $Q_{\text{true}}/Q_{\text{expect}}$, which is the ratio between the goodness of fit of the solution and the expected one. The Q_{expect} is calculated using the number of samples, factors, and number of chemicals included in the input matrix (Belis et al., 2014). The optimal number of sources is selected when this ratio approaches 1 (Belis et al., 2014). Other criteria are then further evaluated, which are: the geochemical likelihood of the solutions, the uncertainty of the solution (bootstrap, displacement), and the statistical parameters of the solution (residual, R^2). However, the number of factors observed in the literature is rarely larger than 10 to 12, for well-documented PMF studies. Tests 5 to 11 investigate whether such a number of factors present an inherent limit of the PMF process or if it is related to the structure and content of the database. We tried to investigate the intrinsic capability of the PMF methodology to delineate further sources when proper tracers are included in the input data set.

On the basis of a 11 factors' PMF solution obtained with a yearly time series of observations in Grenoble (France) (Borlaza et al., 2021), we conducted tests 5 to 9 by adding a tracer of biogenic oxidation products (2-methyltetrols, (2MT), coming from the oxidation of isoprene (Carlton et al., 2009; Edney et al., 2005)) on top of the 3-MBTCA (tracer of the Secondary Organic Aerosol (SOA) formed from the alpha-pinene oxidation (Claeys et al., 2007; Kourtchev et al., 2008)), aiming to separate these two SOA sources. The time series of their respective concentrations differ, but the initial 11 factors of PMF work could not properly separate the two factors. The concentrations and uncertainty of 2-MT were changed conditionally (resulting in different S/N) to assess how these parameters influence the source separation. This was further investigated with tests 10 and 11, where a synthetic series of tracers was generated using the criteria defined in tests 5-9 to validate their efficiency and delineate additional sources. The inputs of each test and the PMF solutions are presented in Table S2.

2.1.3. Tools to evaluate the PMF result

To evaluate PMF outputs, each obtained factor is questioned as possibly being representative of a given emission source and/or secondary formation process based on its chemical profile and timeseries, generally according to the operator's knowledge and experience. There are indeed few guidelines and tools that allow the benchmarking of these choices, which are rarely fully backed up in the presentation of PMF works. In order to provide a more objective way of testing the reliability of the results from a PMF study, we implemented two tools to evaluate the chemical profiles obtained. The first one is based on a compilation of diagnostic ratios available in the literature concerning some common chemical species used in PMF studies in order to test automatically if the results of PMF are in agreement with these ratios. The detailed expected range of the ratio, the reference, as well as the type of source according to ratio value are presented in Table S3, Supplement. The Python code for this tool is proposed in a Python package, called PMF_toolkits (https://github.com/DinhNgocThuyVy/PMF_toolkits), and an example of the application is proposed in Test 12.

A second tool is proposed to automatically compare the full chemical profiles obtained in the PMF output with banks of known chemical profiles. The database is extracted from the European Commission sources profile repository SPECIEUROPE (Table S4) (Pernigotti et al., 2016). It has been augmented with some recent sources reported in the literature that are not yet available in the repository, including all the ones generated from past studies in France from our group (IGE database) that are currently being transferred (Table S5).

2.2. Study site and chemical analysis

2.2.1. Study sites

PM₁₀ series of samples from three different French sites were used for these 15 tests: Arrest, Ailly, and Grenoble (Table 1). The sampling date, frequency, and number of samples are shown in Table 2. The chemical analyses of these samples are briefly described in section 2.2.2. They are detailed, along with their corresponding initial PMF outputs (used as the basis for the present paper), in Zhang et al. (2024) for the Arrest and Ailly sites and in Borlaza et al. (2021) for the Grenoble sites.

Arrest and Ailly are located in coastal rural sites in the north of France (Figure 1b), which are 50 km apart. Arrest covers an area of 11.15 km² and has a population of 855, representing a density of 77 inhabitants/km². The agglomeration of Ailly, with a population of 46,223, occupies an area of 5.7 km². Eleven sources were identified

in the PM₁₀ for each site: Aged sea salt, biomass burning (BB), MSA-rich, heavy oil combustion (HFO), industrial, mineral dust, nitrate-rich, primary biogenic, primary traffic, sea salt, secondary biogenic.

Grenoble is an urban background site located in an Alpine valley southeast of France (Figure 1c), sprawling over 18.13 km² with 154,018 inhabitants in 2023 with a density of 800 inhabitants/km². The initial PMF work in Grenoble identified 10 sources of PM₁₀, including aged sea salt, BB, industrial, mineral dust, nitrate-rich, primary biogenic, primary traffic, sea/road salt, secondary biogenic, and sulfate-rich.

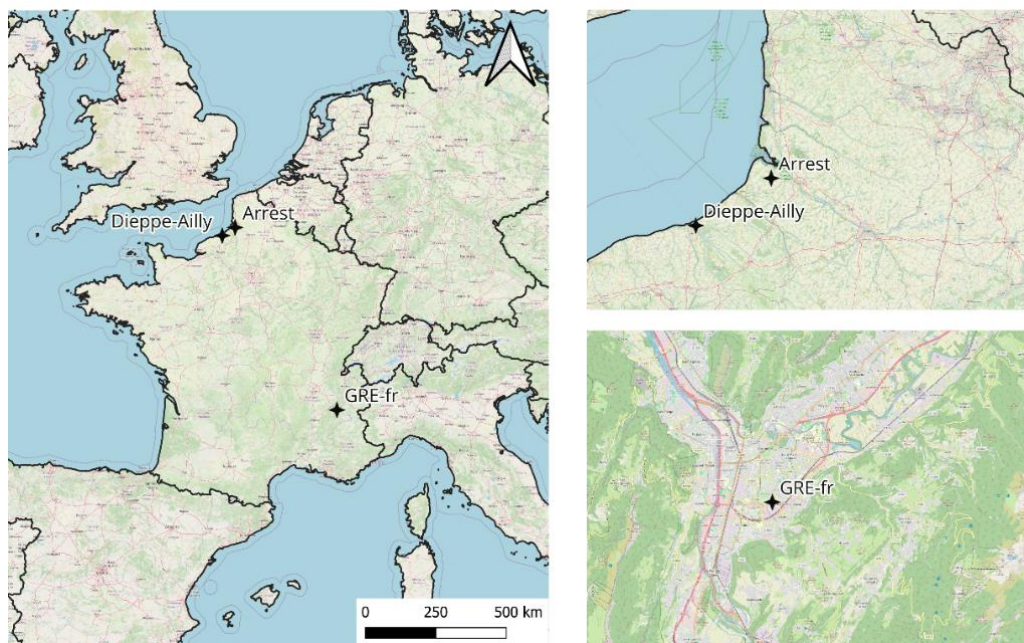


Figure 1. Location of Ailly, Arrest and Grenoble © OpenStreetMap contributors 2025. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

Table 2. Sites description

Sites	Date sampling	Sampling frequency	Typology	Number samples	Coordinates
Arrest	01/02/2021 – 20/6/2021	Daily	Rural background	140	50°07'48.0"N 1°37'12.0"E
Ailly	11/02/2021 – 29/6/2021	Daily	Rural background	139	49°55'12.0"N 1°06'00.0"E
Grenoble	28/2/2017 – 10/03/2018	Every third day	Urban background	128	45°09'41"N, 5°44'07"E

2.2.2. Chemical analysis

The daily PM₁₀ sampling was performed using high-volume samplers (Digitel DA80, 30 m³ h⁻¹) using 150 mm-diameter pure quartz fibre filter (Tissu-quartz PALL QAT-UP 2500 diameter 150 mm). A standard protocol was applied for cleaning, unloading, packing, and storing filters to avoid contamination. Field blank filters were collected (about 8 – 12 % of the number of samples) to estimate the detection limit and control the filter contaminations. This protocol is presented in Weber et al. (2019) and Borlaza et al. (2021).

The PM₁₀ filter samples were extracted to perform chemical analysis for quantifying different PM components, including the main components by mass and many tracers of sources, including carbonaceous fractions (OC, EC),

major ionic species (Cl^- , NO_3^- , SO_4^{2-} , Na^+ , NH_4^+ , K^+ , Mg^{2+} , Ca^{2+}), methanesulfonic acid (MSA), Oxalate, HULIS, 3-methyl-1,2,3-butanetricarboxylic acid (3-MBTCA), anhydrous sugar and saccharides (levoglucosan, mannosan, arabinol, mannitol) and metals (Al, As, Ba, Cd, Co, Cu, Fe, Mn, Ni, Pb, Rb, Se, Sr, Ti, V, Zn). Details of chemical analysis are presented in Borlaza et al. (2021).

In brief, the elemental carbon (EC) and organic carbon (OC) were analyzed using a Sunset Lab analyzer (Birch and Cary, 1996) based on the EUSAAR2 thermo-optical protocol.

The filters were punched using 11.34 cm² punches, soaked in 10mL of ultra-pure water, and filtered after 20 mins of agitation using a 0.25 μm porosity filter. These water extracts were used for the following analysis. The major ionic components, MSA, and Oxalate, were analyzed using an ICS3000 dual-channel chromatograph (Thermo-Fisher) with a CS12 column for cations analysis and an AS11HC column for anions analysis. The anhydrous-sugar and saccharides analysis was carried out on the aqueous phase by high-performance liquid chromatography (HPLC) with Pulsed Amperometric Detection (PAD) (model Dionex DX500 + ED40). HPLC-PAD was performed using a Thermo-Fisher ICS 5000+ HPLC equipped Metrosep Carb column and precolumn in isocratic mode. 3-MBTCA was analyzed using the HPLC coupled mass spectroscopy (HPLC-MS) with negative mode electrospray ionization.

HULIS measurement was conducted according to the protocol reported by Baduel et al. (2009, 2010). Briefly, the analysis was performed on the water-soluble fraction of filter samples. The neutral components, hydrophobic bases, inorganic anions, mono- and di-acids are removed by passing through a weak anion exchange resin. After extraction, HULIS quantification is performed using a TOC analyzer (Shimadzu).

Finally, the filters were punched into a 38mm diameter, mineralized and used to conduct the analysis of major and trace elements. The analysis was performed using inductively coupled plasma mass spectroscopy (ICP-MS) (ELAN 6100 DRC II PerkinElmer or NEXION PerkinElmer) (Alleman et al., 2010).

2.3. PMF

The PMF methodology is used to perform Tests 1 to 11, using EPA PMF 5.0 software. The PMF input variables of all tests are shown in Table S1 for Test 1 to Test 4 and Table S2 for Test 5 to Test 11. The PMF theory, briefly described in sections 2.3.1 and 2.3.2, points out the crucial parameters in PMF execution. In Tests 5 to 11, the inputs are conditionally generated, as presented in section 2.3.3.

2.3.1. PMF formula

PMF algorithm based on observed concentration data (matrix X) to identify the contribution of the factors (matrix G) and factors profiles (matrix F) as described in the equations below:

$$X_{ij} = \sum_{k=1}^p G_{ik} * F_{kj} + E_{ij} \quad (\text{E.q 1})$$

$$G_{ik} \geq 0, F_{kj} \geq 0 \quad (\text{E.q 2})$$

Where X is a (i×j) matrix of j chemical species in measured period i (generally daily sampling) into p factors with a matrix (i × k) representing the source contribution (G) and a matrix (k × j) representing the factor composition (F). E is the residual for each species. All the factor matrices G and F elements are constrained to be non-negative. The solutions are selected by finding F and G to obtain the minimum value of the quality of fit parameter (Q), which is calculated by:

$$Q = \sum_{i=1}^n \sum_{j=1}^m E_{ij}^2 / \sigma_{ij}^2 \quad (\text{E.q 3})$$

$$\{G, F\} = \arg \min Q \quad (\text{E.q 4})$$

Where σ_{ij}^2 is the known uncertainties for each data value x_{ij} (arranged in uncertainty matrix of input data). The higher the uncertainty, the lower the ratio E_{ij}^2 / σ_{ij}^2 , consequently lesser influence on Q calculation. Thus, chemical species with high uncertainty have a diminished effect on the PMF model.

2.3.2. Signal-to-noise ratio

The signal-to-noise ratio (S/N) represents the relationship between the concentration and uncertainty of a species (EQ 5,6,7). A species with $S/N < 0.2$ means concentrations are almost equal to uncertainties, and it has to be removed from the model to avoid unreliable results. Conversely, a S/N value > 2 (where concentration > 3 times uncertainty) is considered satisfactory enough to consider the corresponding variable as a "Strong" one. Between 0.2 and 2, the specie will be categorized as "Weak", and the uncertainties is set to 3 times the original one (Belis et al., 2014).

$$\left(\frac{S}{N}\right)_j = \frac{1}{n} \sum_{i=1}^n d_{ij} \quad (\text{E.q 5})$$

$$d_{ij} = \left(\frac{x_{ij} - s_{ij}}{s_{ij}}\right), \text{ if } x_{ij} > s_{ij} \quad (\text{E.q 6})$$

$$d_{ij} = 0 \text{ if } x_{ij} \leq s_{ij} \quad (\text{E.q 7})$$

with x_{ij} is concentration and s_{ij} is the uncertainty of measured period i in chemicals j

Therefore, S/N is an indicator of the influence of a chemical specie on the Q calculation. This indicator is used in Tests 5 to 9 to elucidate the importance of uncertainty in the separation of a factor in PMF.

2.3.3. The generation of variable characteristics of 2-MT for sensitivity tests

The uncertainty is calculated using the formula of Gianini et al. (2012), presented in Table S6. The uncertainties of the test component (2-MT) for the sensitivity tests in section 2.1.2 are generated by varying the limit of quantification (QL), coefficient variation (CV), and additional coefficient (a) (Table 3), resulting in the S/N value ranging from 1 to 8. The concentration and time series of the component are also changed to evaluate the effects on the PMF result, as presented in Figure S2.

Table 3. S/N value corresponding to parameters used for uncertainty calculation of 2-MT and Fake1, Fake2

	S/N value	QL ($\mu\text{g}/\text{m}^3$)	CV	Alpha
2-MT1	S/N = 2	0.000352	15%	10%
2-MT2	S/N = 2	0.00352	15%	10%
2-MT3	S/N = 3.8	0.000176	10%	10%
2-MT4	S/N = 8	0.0000704	5%	5%
2-MT5	S/N = 1.4	0.000352	20%	10%
Fake 1	S/N = 4	20%		
Fake 2	S/N = 3	30%		

2.4. Distance metrics

The comparisons between the PMF-derived chemical profiles and the SPECIATE + IGE database are performed with the two distance metrics Pearson Distance (PD) and Similarity Identity Distance (SID). These distances are reported by Belis et al. (2015), and are calculated by EQ8 and EQ9. The PD is sensitive to the main composition of PM, while SID represents the similarity of all common components of the paired comparison of chemical profile and, therefore, also includes the influence of minor chemical species by mass. The chemical profiles are declared similar if $0 < \text{SID} < 1$ and $\text{PD} < 0.4$.

$$PD = 1 - R^2 \text{ (E.q 8)}$$

where R^2 is the Pearson coefficient of the relative mass to PM of all components between 2 chemical profiles

$$SID = \frac{\sqrt{2}}{m} \sum_{j=1}^m \frac{|x_j - y_j|}{x_j + y_j} \text{ (E.q 9)}$$

Where:

m is the number of common chemical components of 2 chemical profiles

x and y are the relative mass to PM of 2 chemical profiles.

3. Result and discussion

3.1. Tests on the PMF inputs

3.1.1. Tests on the sensitivity of the PMF of changes in commonly used chemical species

Tests 1, 2, and 3 evaluate the efficiency and usefulness of some choices for the variables in the PMF inputs for the PMF work in the two sites, Ailly and Arrest. The comparison of source contribution to PM between the base run solutions and Tests 1, 2, 3 (Figure 2) shows a good stability PMF over all tests and both sites, with a consistency for the number of factors (remaining at 11 sources) and a maximum difference in source contribution to PM among tests being limited to 1% for Arrest and 3% for Ailly.

Test 1 shows negligible variation in all factors compared to the base run PMF solution in both sites when using OC or OC*. The difference between the two concentrations is relatively low, only 6% and 3% of the OC mass for Arrest and Ailly, respectively. The stable total reconstructed PM concentrations, the unchanging number of sources, and the stable contributions of each factor suggest that it may not be necessary to use OC*, at least in cases when the mass difference between OC and OC* is below 6%. Similar results were obtained for another study in an Alpine Valley site (Glojek et al., 2024), where the difference between the two concentrations is 10 %.

Test 2, comparing the results using a single BB tracer (levoglucosan) vs using both tracers, levoglucosan and mannosan (in the base run), also indicates a relatively low change in PM contribution. The highest changes are observed in the contribution of BB (increases from 7% to 10%). A closer examination of the chemical profiles shows that with a single tracer, there is less dispersion of chemical species of interest in the other sources (in this case, levoglucosan, OC, EC). Especially for the traffic source, where the presence of levoglucosan is not really justified based on geochemical knowledge (Borlaza, 2021; Weber et al., 2019), a reduction of 5% of levoglucosan is observed (compared to the base PMF run). This reduction results in a lower contribution of OC and EC in the traffic profile, eventually decreasing the contribution of traffic to PM_{10} . The bootstrap values from this test indicated a reduced traffic swap with the other sources compared to the base run PMF (Table S7). In addition, test 2 conducted a more homogenous PM contribution of BB between Arrest and Ailly (15% vs 11%), located 50 km

apart. Consequently, test 2 strongly suggests the benefits of using levoglucosan instead of levoglucosan and mannosan.

Test 3 considers using the sum of concentrations of arabitol and mannitol (base run) instead of the two chemical species separately as proper tracers of primary biogenic emission associated with fungal spores (Bauer et al., 2002; Rogge et al., 2007; Samaké et al., 2019b; Yttri et al., 2011). Similar to test 2, test 3 reveals only minimal change in the factor associated with these tracers, with a decrease of 0.5% in the contribution of the primary biogenic source to PM. This change also results from more dispersion of arabitol and mannitol to the other sources, compared to the use of their sum (as "Polyols"). Although the contribution demonstrates minor variation, this result again suggests the advantage of using a single tracer. As mentioned above, it seems again that using two separate tracers can lead to their larger distribution in some other sources, resulting in a decrease in the contribution of their source and some degree of mixing.

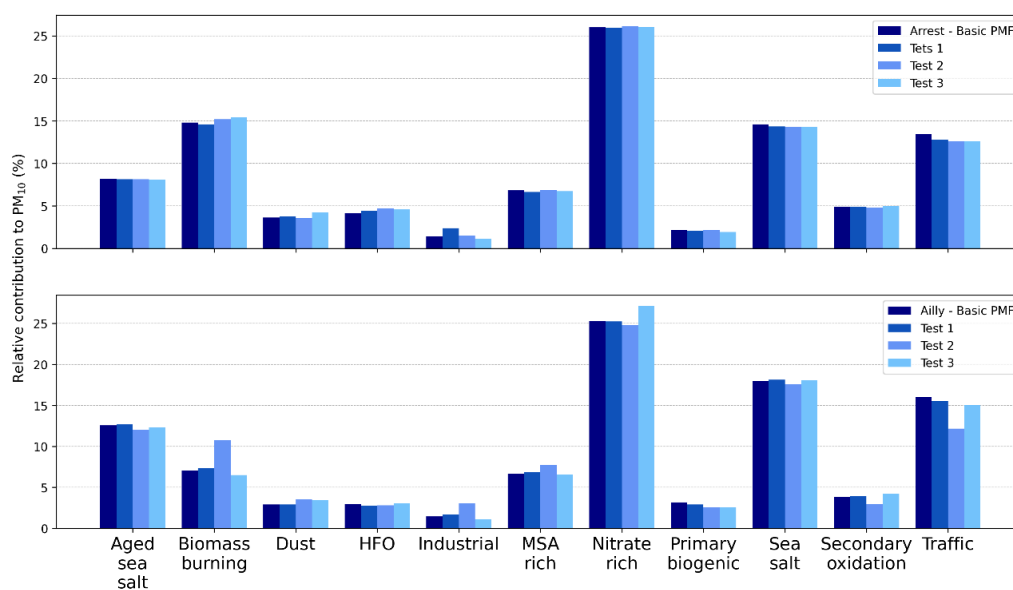


Figure 2. PM₁₀ source contributions of base run PMF and Test 1, 2, 3 in Arrest (Upper) and Ailly (Lower).

3.1.2. Tests on adding new chemical species

Test 4 incorporated HULIS and Oxalate simultaneously into PMF to evaluate the impacts of these changes in the input data and investigate the possibility of determining their fate in the atmosphere. Indeed, these two compounds are known as secondary organic products, and ubiquitous. They are generally the main components identified in the OM. Developing a better knowledge of the sources of their main precursor, along with the evaluation of the ability of the PMF method to deliver such information, is therefore of interest for AQ plans. Test 4 indicates that the optimal number of sources when adding these chemicals remains stable at 11 sources. The main statistical parameters (BS values and reconstruction of PM) are not impacted (Table S7), and the chemical profiles remain unchanged except for these two chemical species, which seems in itself a good appraisal of the base run PMF solution stability. Hence, these results potentially provide some novel insights into the relation to the primary sources for these secondary organic fractions, which are rarely discussed.

HULIS are consistently associated with traffic, secondary biogenic oxidation, BB, and industrial activities during the whole observation period (Figure 3a), all sources that are mainly involved in the emission of OM. HULIS are also presented in the nitrate-rich and sea salt sources, with a lower contribution. This partially aligns with a PM source apportionment that incorporated HULIS in a study in Beijing and reported that the most important source

of HULIS is coal/ BB and traffic (Dominutti et al., 2024; Li et al., 2019; Srivastava et al., 2018). In addition, current research has suggested that HULIS are predominantly linked to combustion processes during the winter period, either by direct emissions or by the secondary formation in the processing of these emissions, and are supposed to be mainly anthropogenic (Baduel et al., 2009, 2010; Graber and Rudich, 2006; Zheng et al., 2013). Our results align with the concepts reported in these studies, which demonstrated that the primary sources in the cold period are traffic and BB (Figure 3c). On the other hand, in summer, the main sources are the production of secondary biogenic organic aerosol and traffic (Figure 3b). During summer, polymerization or oligomerization processes activated by higher temperature, light intensity, high volatile organic compounds (VOCs) and O₃ or OH radical concentration can produce photochemical reactions leading to HULIS formations (Hoffer et al., 2006; Wu et al., 2018). In addition, Zheng et al. (2013) suggested that the VOCs generated by vehicles are critical precursors of HULIS through heterogeneous reactions.

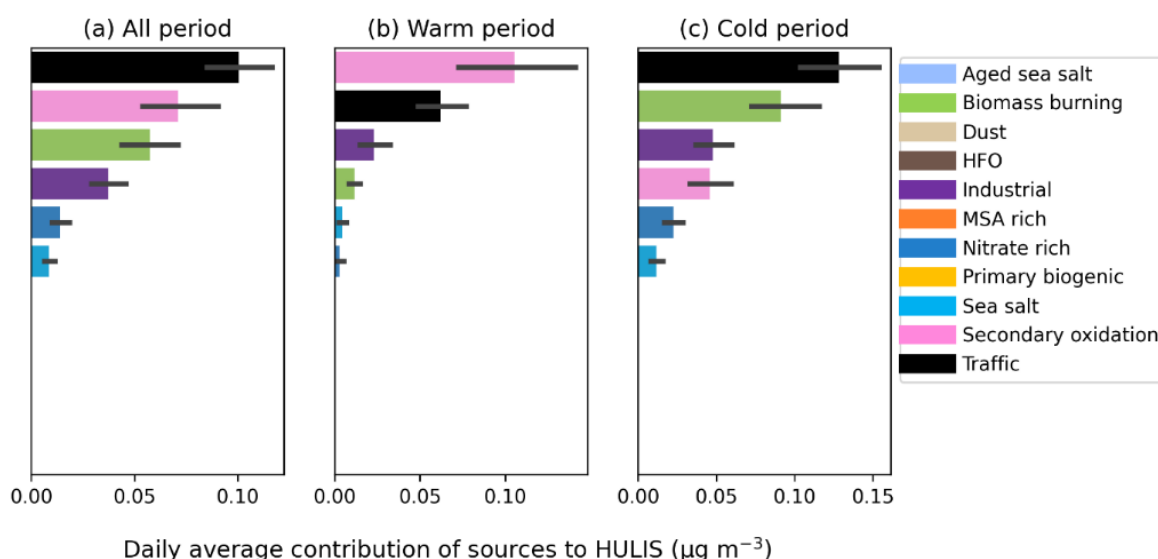


Figure 3. Daily average sources contribution to HULIS ($\mu\text{g m}^{-3}$) in (a) The study period (b) the warm period (April, May, June), and cold period (February, Mars) in Ailly. The bars represent the mean, and the error bars represent the 90% confidence interval of the mean.

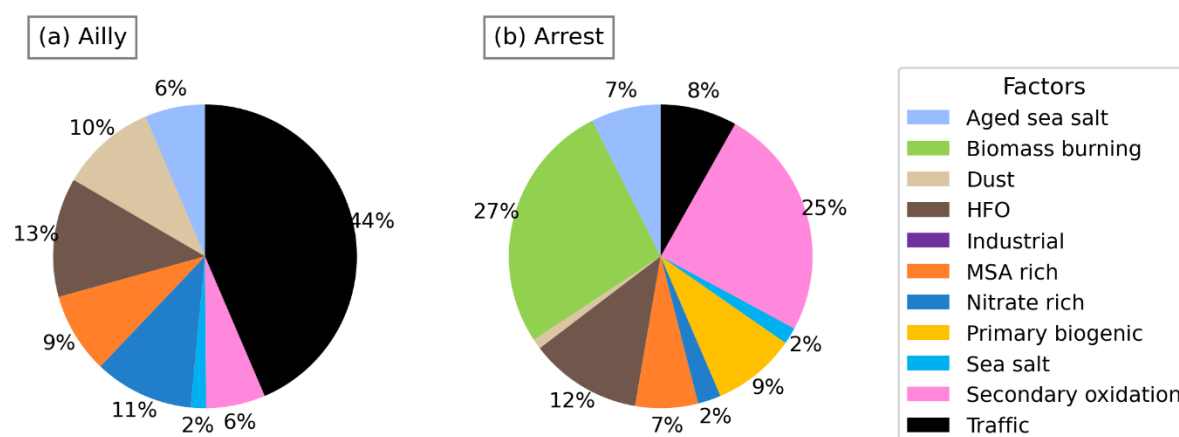


Figure 4. The relative contribution of the source to Oxalate (a) in Ailly and (b) in Arrest.

In Ailly, oxalate is mainly associated with traffic (44%), HFO (13%), nitrate-rich (11%), and dust (10%) (Figure 4a). In contrast, in Arrest, oxalate is predominantly associated with BB (27%) and secondary biogenic organic aerosol traced by 3-MBTCA (25%). This discrepancy can be attributed to the sampling locations, as the site in

Ailly is closer to roads and the parking of a cultural landmark and also immediately next to the seashore. Therefore, the contributions of oxalate are primarily related to traffic and marine sources (including intense shipping emissions in the English Channel). In contrast, Arrest's sampling site is entirely separated from the main roads and surrounded by large natural areas (marsh area in the Baie de Somme). Therefore, the origins of oxalate depicted in our PMF results could be plausible. Indeed, these sources are in line with several studies that discussed the main origins of oxalate, like Zhou et al. (2015), who indicated that oxalate is related to secondary processes of BB emission aging, or Kawamura & Bikkina (2016) who suggested that oxalate may be derived from traffic emissions, associated with oxidation of compounds derived from gasoline combustion engines and residual oil combustion. The same authors also found a linear relationship between oxalate and methane sulfonic acid (MSA), explaining it by the fact that aged sea salt includes unsaturated fatty acids that can form oxalate by further photochemical oxidation processes.

Finally, this test is interesting because it confirms that introducing secondary species that we know come from several sources does not lead to new independent sources despite their significant contribution to the total carbon mass. Further, this does not disturb the previous sources from the base run but leads to the repartition of these species into reasonable sources, opening the door to understanding their formation process in more detail.

However, on the contrary, the introductions of some other organic species (not reported in this specific work) were not so successful, and we failed to obtain reasonable solutions or new factors with the introduction in the input PMF data of specific species like MSA (a known tracer of oxidation products from marine VOC emissions (Li et al., 1993)), cellulose (a known tracer of plant debris (Brighty et al., 2022)), or 2-methyltetrols (2-MT) (a known tracer of oxidation products from isoprene (Edney et al., 2005)), among others. Some of these trials are reported in Glojek et al. (2024). Obviously, the capacity of PMF to separate a source does not depend solely on the introduction of a proper tracer, even with a specific time series (Glojek et al., 2024), but also on other aspects, some of which are tentatively investigated in the next section 3.2.

3.2. The ultimate determination of source number

The following series of tests (5 to 9) are performed with the base case of the Grenoble times series (Borlaza et al., 2021) (Table 1). They investigate the impact of some changes in the characteristics of the 2-MT time series of concentrations in order to evaluate the conditions necessary to obtain a specific factor associated with secondary oxidation from the isoprene emission. This essentially includes changes in concentrations and uncertainties (S/N values), as shown in Table 4. In addition, Tests 10 and 11 try to evaluate the capability of PMF to generate more than 11 or 12 factors when appropriate tracers are introduced in the input data. This is tested with the generation of "fake components" (Fake 1 and Fake 2), with characteristics (time series, concentrations, S/N) guided by the previous tests 5 to 9 for defining "a proper tracer".

It should first be noted that the time series of 2-MT (based on real measurements), Fake 1 and Fake 2 are quite different from that of all other chemical species from the input matrix (Figure S3), and particularly all of the biogenic tracers. The best correlation is obtained between 2-MT and MSA, with an R^2 of 0.30, for 128 samples, preventing collinearity. Table 4 includes some of the results from these tests, including the optimal number of sources identified, the number of sources including a substantial fraction of the 2-MT (for test 5 to test 9), Fake 1 and Fake 2 (test 10, 11, respectively), and the performance of reproduction of their concentrations by the PMF. It

should be noted that all 20 runs performed for each test converged, and constraints "Pull up maximally" for the source tracer were applied to improve the base run results (Table S8).

Table 4. The description of input and output statistical parameters of Test 5 to Test 11

	Test5	Test6	Test7	Test8	Test9	Test10	Test11
Mean concentration (ng m ⁻³)	6.7	67	6.7	6.7	6.7	31.3	24.5
S/N	2	2	3	8	1	4	2
Maximum R ² with the other chemicals	0.30	0.30	0.30	0.30	0.30	0.22	0.11
Number of sources identified	11	12	12	12	11	13	14
Number of factors where the source of 2-MT can be identified	13	12	12	5	x	5	13
The number of sources that contribute to 2-MT	3	2	1	1	7	1	1
R ² between tracer _{measured} and tracer _{predicted}	0.38	0.99	1.00	1.00	0.25	1.00	1.00
Q _{true} / Q _{expected}	2.3	2.1	2.1	2.1	2.1	2.2	2.2

For Tests 5 and 9, with the actual measured concentrations of 2-MT and S/N values < 2, the PMF does not allow the identification of a specific source for 2-MT, with a number of sources identical to the base run (11 sources). 2-MT contributes to 3 sources in Test 5 (higher S/N) and 7 in Test 9 (lower S/N) (Figure S4). Although the principal attribution is to primary biogenic in both tests, the reconstruction of the species is better in test 5, resulting in an R² between observed and predicted concentration of 0.38 vs 0.25 for test 5 and test 9, respectively. Even though the performances are better for the higher S/N and the tracer has no strong collinearity with other factors, the results in Tests 5 and 9 do not provide a separate factor.

The PMF solution is improved in test 6, with the combination of S/N=2 and a higher concentration (10 times higher than the measured concentration of 2-MT), allowing the identification of a new source comprising most of the 2-MT (98%). Tests 7 and 8, which have the measured concentrations but higher S/N values, also facilitated the separation of 2-MT into an additional 12th source, comprising 100% of the tracer in both tests (Figure S4). Due to very high seasonality, the mass fraction associated with this source is only meaningful during a short period, reaching about 2-3 % only in July – August. Interestingly, we observed that the S/N also affects the ability to identify the source: in the PMF process, the source of 2-MT appears when five sources are selected with S/N = 8, while with S/N ≤ 3 in (Tests 6 and 7), it is identified later when 12 sources are selected. Finally, it is worth noting that the Q_{true} / Q_{expected} ratio is relatively stable and not a good indicator for the choice of the proper number of sources at this stage of the resolution of the PMF.

All these tests tend to indicate that the S/N value is the most important criterion that allows the separation of a specific factor in a complex PMF with about 8-10 factors, with a value of S/N above 2-3 being a threshold above which separation can be effective. It seems that the higher the value, the easier the separation of the specific factor. High concentration may play a role for source identification, but it is not essential in source contribution (contribution of 2-MT source to PM is similar for test 6 and test 7). However, this is also conditional on a time series of the concentration of a specific tracer that presents low collinearity with other tracers, a value of R² < 0.3 for about 100 or more samples being advisable.

With Tests 11 and 12, we tried to verify if these criteria are valid for increasing the number of factors that are possible to delineate for a PMF study in this range of 10 – 15 sources, using reconstructed (fake) time series. The

Fake 1 and Fake 2 components are generated with criteria as in Table 4, and then successively incorporated into the input database of Test 7. Table 4 indicates that the PMF enables the identification of 13 and 14 sources for Test 10 and Test 11, respectively, with excellent reconstructions of the fake tracers ($R^2 = 1$).

The extensive descriptions of the PMF solution, including the reconstructed level of PMF (R^2 between predicted and observed concentration), bootstrap values, and displacement (DISP) in Test 10 and Test 11 (Table S9 and S10), demonstrate the robustness of solutions. The narrow band of DISP values of 2-MT, Fake 1 and Fake 2 in their source (Table S10), highlights the low rotation variation of these tracers. The variation of the ratio $Q_{\text{true}}/Q_{\text{expected}}$ and R^2 of $\text{PM}_{10}^{\text{obs}}$ and $\text{PM}_{10}^{\text{predicted}}$ (figure 5) shows that the solutions reached the lowest $Q_{\text{true}}/Q_{\text{expected}}$ and highest R^2 at 13 factors for test 10 and 14 for test 11. Notably, the $Q_{\text{true}}/Q_{\text{expected}}$ close to 2 is observed as the most appropriate, indicating the identical ratio for all tests (see Table 4).

In addition, the chemical profile of the sources from Test 5 to 11 and the base run PMF 11 factors solution is compared using PD and SID, indicating that the chemical profiles are all within an excellent homogeneity range (Figure S5). This last point emphasized that incorporating 2-MT, Fake 1, and Fake 2 does not affect the initial solution. Finally, all these tests clearly indicate that valid solutions with at least up to 14 different sources are possible with the PMF, providing that the input includes tracers with low collinearity and S/N above three.

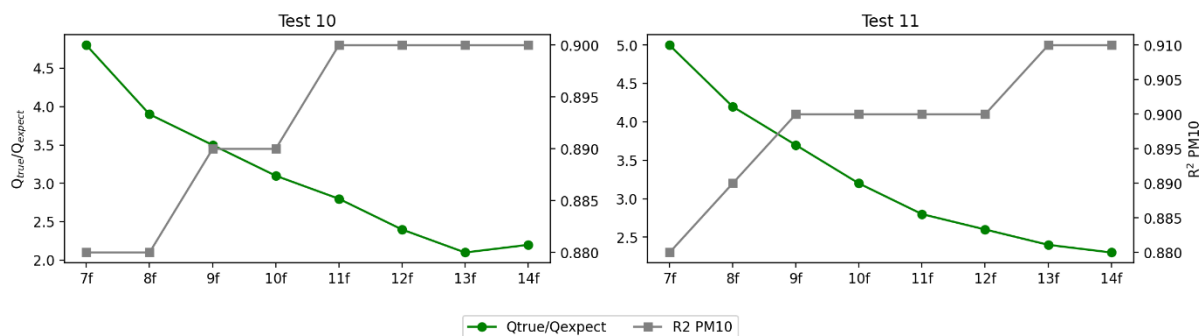


Figure 5. The variation of R^2 between observed PM_{10} and predicted PM_{10} and ratio $Q_{\text{true}}/Q_{\text{robust}}$ over the number of factors of Test 10 (left) and Test 11(right).

3.3. Chemical profile comparison

3.3.1. Validation of the ratio between chemicals/tracers in the profile

While the ratio of specific chemical species has long been used in many studies (particularly concerning trace elements) for the tentative evaluation of the sources of PM components, these diagnostic ratios are very rarely checked in PMF studies for the validation of the PMF solutions presented. We tried to implement an automatic tool for easily performing such validations, using an extended table of plausible ranges of ratios obtained from the literature for many common sources obtained in PMF studies.

This step is preliminary to the evaluation of the appropriateness of a PMF solution. A solution with a chemical profile that significantly exceeds the accepted range should be further investigated to confirm the geochemical meaningfulness of sources for the specific sampling site. The tools to compare the specific ratio of chemicals in the profile of PMF-derived sources and literature have been developed (detailed in S1, Supplement) and can be accessed at https://github.com/DinhNgocThuyVy/PMF_toolkits.

Table 5 presents the results of some specific ratios for the sources obtained in the 11-factor base run solution of the study in Grenoble used in the previous sections. It indicates that the PMF solution is globally well aligned with the diagnostic ratios reported in the literature. Some deviation can be seen for the mineral dust (Al/Ca) and primary

traffic factor (OC/EC), which is supposed to be negligible since the discrepancy is relatively low (the ratio in Grenoble is lower than 8% compared to that reported in the references).

Table 5. Comparison of the ratio between chemicals in the source profile of Grenoble and the literature

Source	Ratio	Grenoble	Accepted range	Reference
BB	Levoglucosan/K ⁺	4.37	0.38 – 58.8	(Li et al., 2021)
	OC/EC	3.58	3 – 10.6	(Fine et al., 2001; Li et al., 2021)
	Levoglucosan/Mannosan	8.92	3-10 (softwood)	(Fine et al., 2002)
Mineral dust	OC/EC	3.56	2.3-17.7	(Amato et al., 2009)
	Ti/Ca	0.014	0.012-0.018	(Nicolás et al., 2008)
	Al/Ca	0.23	0.251-0.253	(Nicolás et al., 2008)
Primary traffic	OC/EC	1.26	1.34-3.7	(Giugliano et al., 2005)
	Cu/Sn	4.99	4.7-5.4	(Amato et al., 2011)
	Fe/Cu	21.74	17-29	(Amato et al., 2011)
Primary biogenic	OC/(Arabitol + Mannitol)	0.18	0.01 - 0.32	(Weber et al., 2019)
Secondary oxidation	3-MBTCA/OC	0.009	0.010 – 0.014	(Borlaza et al., 2021)
MSA rich	MSA/OC	0.26	0.1-0.7	(Weber et al., 2019)
Nitrate rich	NO ₃ ⁻ /NH ₄ ⁺	0.8 (molar ratio)	0.6 - 1 (molar ratio)	(Heo et al., 2013)
Sulfate rich	SO ₄ ²⁻ /NH ₄ ⁺	0.5 (molar ratio)	0.4 - 0.5 (molar ratio)	(Heo et al., 2013)

3.3.2. Chemical profile comparison – biomass burning source

Following the previous test about diagnostic ratio, the results obtained for the sources from a PMF study could be benchmarked for their full chemical profiles against those obtained in the literature. We developed another tool for the automatic comparison with chemical profiles from 2 databases (SPECIEUROPE and SPECIEUROPE + IGE). The comparison aims to point out if the profiles obtained in a PMF work are similar to those reported in the literature, but also to provide insight into naming as properly as possible a factor identified by the PMF (which is also a critical step, particularly for an air quality regulation perspective). The comparison is performed using the Pearson distance (PD) and Similarity Identity Distance (SID) (Belis et al., 2015). The criteria for selecting similar profiles in the database are as follows: (1) PD<1, (2) 0<SID<1, (3) number of chemicals used for comparison (n) > 50% of the chemical species in the PMF-derived source. Ultimately, the most similar chemical profile (compared to the profile to be evaluated) is selected by sorting the lowest PD ∩ the lowest SID ∩ the highest n.

In addition, the tool observed not only the PD and SID, but also the number of chemicals used for comparison, supposing that the comparison with less than half of the chemical species in the profile is insignificant. The outputs of this tool generate a list of sources whose chemical profiles are the most similar to the tested one (Figure S7). Finally, the information about country and year tests, references, as well as the ratio between different chemicals in the profile is exported, providing complete information. The code also generates figures of the chemical composition of the profiles (like in Figure 6) and PD-SID figures (like in Figure S7), which allow the user to select the number of profiles that are the most similar to the tested one (eight profiles in the case of figure 6). The Python code for this tool is proposed in the same previous Python package as above and is presented in S1.

An example of using this tool is provided for the biomass burning (BB) factor of the same 11-factor base run of the Grenoble PMF used above. BB, is one of the most well-known sources in the database (76 biomass combustion

profiles over a total of 287 profiles in SPECIEUROPE). Figures S7 and S8 present the PD-SID for the comparisons with SPECIEUROPE and SPECIEUROPE + IGE databases, respectively. Using the latter yielded a lower average of PD and SID compared to using only SPECIEUROPE (PD: 0.2 vs 0.1, SID: 0.8 vs 0.6). Indeed, the chemical profiles from the SPECIEUROPE database are barely used in the comparison of Figure S8, which mainly uses profiles from France (included in the IGE database) due to their proximity, as reported by Weber et al. (2019). Only one chemical profile in SPECIEUROPE appeared in the comparison, which is the Larch closed fireplace. However, while using SPECIEUROPE led to a higher distance for the returned profiles, PD and SID remained within the homogeneous limits in both comparisons.

A closer examination of the composition used to compare the BB chemical profile in Grenoble and the eight most similar ones exported by the tool in the SPECIEUROPE is provided in Figure 6. Most of the sources are related to wood combustion activities, with the main compositions of the source being OC, Levoglucosan, and EC. The "non-comparison" composition (total mass of the chemicals that were not used for the comparison) is relatively low (under 10%), giving some confidence in the proximity of the profiles. Surprisingly, the "Tyre wear" source appears in this list of the most similar profiles to BB in Grenoble; however, the contribution of non-comparison species is 95%, indicating the comparison did not consider the main composition of this source. Indeed, the species used for this comparison are mainly trace metals (Antimony, Arsenic, Barium, Cadmium, Calcium, Chromium, Copper, Lead, Magnesium, Manganese, Nickel, Vanadium, Zinc), which are not the proper tracers nor the high contributors to the BB source. Consequently, the number of species used to compare and the contribution of these species should be investigated to ensure the robustness of the comparison.

These results emphasize that it is crucial to consider the SID and PD values together with the number of chemical species compared, but also to observe in more detail the chemical profiles in order to evaluate the PMF solutions. Hence, with this added possibility of rapidly accessing these comparisons, this tool efficiently generates meaningful results in the process of performing a PMF study. However, it would be interesting to keep implementing more chemical profiles in the SPECIEUROPE database, and alternative methods like creating a joint database with SPECIATE (<https://www.epa.gov/air-emissions-modeling/speciate>, Date Accessed: 02/04/2025) could be considered.

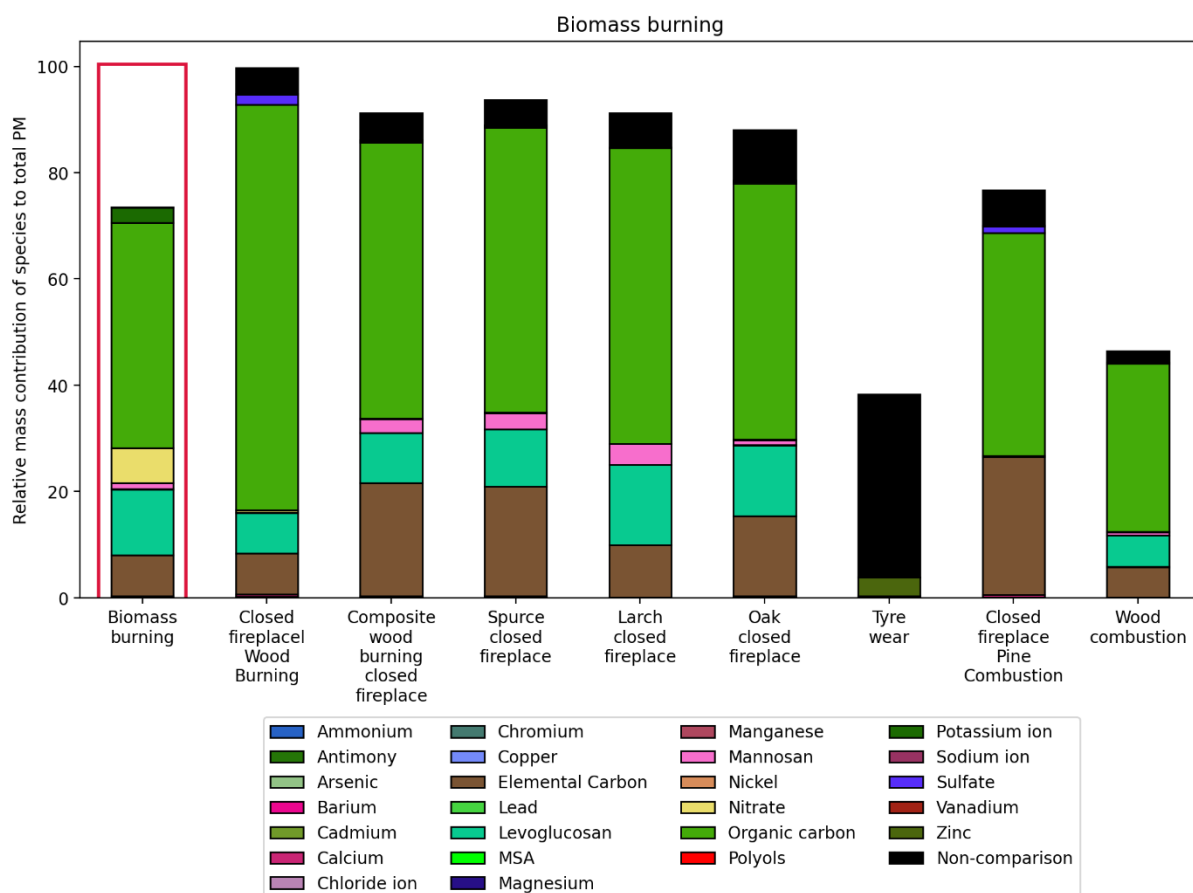


Figure 6. Composition of BB in Grenoble (framed in red) and its eight most similar sources extracted from SPECIEUROPE. The color denotes the chemical species. The y-axis is the mass of the given species relative to the PM total mass. Non-comparison is the total mass of the chemicals not used for the comparison.

4. Conclusion

This study tries to establish a more robust methodology to minimize the subjectivity in performing source-receptor modelling using the Positive Matrix Factorization methodology. This provides to both the scientific community and first-time users with a reliable framework for obtaining accurate solutions. Through extensive testing, we identified key steps that mitigate subjectivity in the PMF model and proposed clear recommendations for tracer selection and PMF solution validation.

Tests performed on the input chemical species demonstrated that changes in input data does not significantly influence the PMF solution. With a disparity between OC and OC* of less than 6%, the PMF solution remained unchanged, suggesting that subtracting the carbon mass is unnecessary in this case. Furthermore, these tests point out that using a proper tracer, (i.e. emitted by a single source only) yields more accurate results than using several, improving source separation and reducing tracer dispersion across other sources.

In determining the optimal number of sources, we emphasized key criteria for tracer selection in source identification. Important criteria for a tracer to delineate a distinct source requires $S/N \geq 3$ and a quite independent temporal evolution 'with R^2 between the tracer and the other chemicals in the input matrix lower than 0.3. Incorporating such a tracer in the PMF model enables the identification of a new source without distorting the chemical profiles of existing sources.

Finally, specific tools for PMF result validation were set up, including validation on key chemical ratios and comparisons with reference chemical profiles from archival databases. Comparing with well-known chemical ratios representative of a source helps in confirming that the chemical profile is appropriate, ensuring the main components contribute as expected. In addition, the test on comparing PMF-derived source and chemical profiles in SPECIEUROPE and IGE database enlightened the importance of considering not only SID and PD distances but also the number of species and the accuracy of mass reconstruction. These factors are crucial for reliable and meaningful comparisons.

Many tools used in this study have been integrated into a dedicated Python package – PMF_toolkits, providing the atmospheric science community with some more advanced and easy tools for PMF input preparation, source tracer selection, solution validation, and visualization. This contribution paves the way for more standardized and objective PMF analyses, ultimately enhancing source apportionment studies and air quality research.

Code availability

PMF_toolkits package and its documentation are available on GitHub, which are free to access and can be downloaded at: https://github.com/DinhNgocThuyVy/PMF_toolkits.

Data availability

The datasets could be made available upon request by contacting the corresponding author.

Author contributions

VNTD: Data curation, visualization, writing-original draft, writing-review & editing. GU, JJ: Conceptualization, mentoring, supervision, validation of the methodology and results, acquired fundings for the original PM sampling and analysis, writing-review & editing. PD, SS: Conceptualization, writing-review & editing. OF: Writing-review & editing. RA, SD, CV: Analysis, writing-review & editing. AS, SZ, SS: In charge of the air quality monitoring project, writing-review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgment

The authors would like to express their sincere gratitude to many people of the Air-O-Sol analytical platform at IGE and to all the personnel within the AASQA in charge of the sites for their contribution in conducting the dedicated sample collection.

Financial support

The sampling and chemical analyses performed at the Ailly, Arrest and Grenoble have been mainly funded by the French Ministry of Environment in the frame of the CARA program and Mobilair for GRE (grant no. ANR-15-

IDEX-02) and performed on the AirOsol Platform facility at IGE. Vy Dinh PhD grant has been co-funded by ANR ABS (ANR-21-CE01-0021-01) and by the Predictair project PR-PRE-2021-UGA and UGA 2022-16 FUGA.

References

- Alleman, L. Y., Lamaison, L., Perdrix, E., Robache, A., and Galloo, J.-C.: PM10 metal concentrations and source identification using positive matrix factorization and wind sectoring in a French industrial zone, *Atmospheric Research*, 96, 612–625, <https://doi.org/10.1016/j.atmosres.2010.02.008>, 2010.
- Amato, F., Pandolfi, M., Viana, M., Querol, X., Alastuey, A., and Moreno, T.: Spatial and chemical patterns of PM10 in road dust deposited in urban environment, *Atmospheric Environment*, 43, 1650–1659, <https://doi.org/10.1016/j.atmosenv.2008.12.009>, 2009.
- Amato, F., Viana, M., Richard, A., Furger, M., Prévôt, A. S. H., Nava, S., Lucarelli, F., Bukowiecki, N., Alastuey, A., Reche, C., Moreno, T., Pandolfi, M., Pey, J., and Querol, X.: Size and time-resolved roadside enrichment of atmospheric particulate pollutants, *Atmos. Chem. Phys.*, 11, 2917–2931, <https://doi.org/10.5194/acp-11-2917-2011>, 2011.
- Amato, F., Drooge, B. L. van, Jaffrezo, J. L., Favez, O., Colombi, C., Cuccia, E., Reche, C., Ippolito, F., Ridolfo, S., Lara, R., Uzu, G., Ngoc, T. V. D., Dominutti, P., Darfeuil, S., Albinet, A., Srivastava, D., Karanasiou, A., Lanzani, G., Alastuey, A., and Querol, X.: Aerosol source apportionment uncertainty linked to the choice of input chemical components, *Environment International*, 184, 108441, <https://doi.org/10.1016/j.envint.2024.108441>, 2024.
- Baduel, C., Voisin, D., and Jaffrezo, J.-L.: Water-soluble atmospheric HULIS in urban environments Water-soluble atmospheric HULIS in urban environments Water-soluble atmospheric HULIS in urban environments, *Chem. Phys. Discuss*, 2009.
- Baduel, C., Voisin, D., and Jaffrezo, J. L.: Seasonal variations of concentrations and optical properties of water soluble HULIS collected in urban environments, *Atmospheric Chemistry and Physics*, 10, 4085–4095, <https://doi.org/10.5194/acp-10-4085-2010>, 2010.
- Bauer, H., Kasper-Giebl, A., Löflund, M., Giebl, H., Hitzenberger, R., Zibuschka, F., and Puxbaum, H.: The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols, *Atmospheric Research*, 64, 109–119, [https://doi.org/10.1016/S0169-8095\(02\)00084-4](https://doi.org/10.1016/S0169-8095(02)00084-4), 2002.
- Belis, C. A., Favez, O., Harrison, R. M., Larsen, B. R., Amato, F., El Haddad, I., Hopke, P. K., Nava, S., Paatero, P., Prévôt, A., Quass, U., Vecchi, R., Viana, M., and European Commission (Eds.): European guide on air pollution source apportionment with receptor models, Publications Office, Luxembourg, 1 pp., <https://doi.org/10.2788/9332>, 2014.
- Belis, C. A., Karagulian, F., Amato, F., Almeida, M., Artaxo, P., Beddows, D. C. S., Bernardoni, V., Bove, M. C., Carbone, S., Cesari, D., Contini, D., Cuccia, E., Diapouli, E., Eleftheriadis, K., Favez, O., Haddad, I. E., Harrison, R. M., Hellebust, S., Hovorka, J., Jang, E., Jorquera, H., Kammermeier, T., Karl, M., Lucarelli, F., Mooibroek, D., Nava, S., Nøjgaard, J. K., Paatero, P., Pandolfi, M., Perrone, M. G., Petit, J. E., Pietrodangelo, A., Pokorná, P., Prati, P., Prevot, A. S. H., Quass, U., Querol, X., Saraga, D., Sciare, J., Sfetsos, A., Valli, G., Vecchi, R., Vestenius, M., Yubero, E., and Hopke, P. K.: A new methodology to assess the performance and uncertainty of source apportionment models II: The results of two European intercomparison exercises, *Atmospheric Environment*, 123, 240–250, <https://doi.org/10.1016/j.atmosenv.2015.10.068>, 2015.

567 Birch, M. E. and Cary, R. A.: Elemental Carbon-Based Method for Monitoring Occupational Exposures to
 568 Particulate Diesel Exhaust, *Aerosol Science and Technology*, 25, 221–241,
 569 <https://doi.org/10.1080/02786829608965393>, 1996.

570 Borlaza: Disparities in particulate matter (PM10) origins and oxidative potential at a city scale (Grenoble, France)
 571 - Part 2: Sources of PM10 oxidative potential using multiple linear regression analysis and the predictive
 572 applicability of multilayer perceptron n, *Atmospheric Chemistry and Physics*, 21, 9719–9739,
 573 <https://doi.org/10.5194/acp-21-9719-2021>, 2021.

574 Borlaza, L., Weber, S., Uzu, G., Jacob, V., Cañete, T., Micallef, S., Trébuchon, C., Slama, R., Favez, O., and
 575 Jaffrezo, J.-L.: Disparities in particulate matter (PM10) origins and oxidative potential at a city scale (Grenoble,
 576 France) - Part 1: Source apportionment at three neighbouring sites, *Atmospheric Chemistry and Physics*, 21, 5415–
 577 5437, <https://doi.org/10.5194/acp-21-5415-2021>, 2021.

578 Brighty, A., Jacob, V., Uzu, G., Borlaza, L., Conil, S., Hueglin, C., Grange, S. K., Favez, O., Trébuchon, C., and
 579 Jaffrezo, J. L.: Cellulose in atmospheric particulate matter at rural and urban sites across France and Switzerland,
 580 *Atmospheric Chemistry and Physics*, 22, 6021–6043, <https://doi.org/10.5194/acp-22-6021-2022>, 2022.

581 Brown, S. G., Eberly, S., Paatero, P., and Norris, G. A.: Methods for estimating uncertainty in PMF solutions:
 582 Examples with ambient air and water quality data and guidance on reporting PMF results, *Science of the Total*
 583 *Environment*, 518–519, 626–635, <https://doi.org/10.1016/j.scitotenv.2015.01.022>, 2015.

584 Carlton, A. G., Wiedinmyer, C., and Kroll, J. H.: A review of Secondary Organic Aerosol (SOA) formation from
 585 isoprene, *Atmos. Chem. Phys.*, 2009.

586 Chen, G., Canonaco, F., Tobler, A., Aas, W., Alastuey, A., Allan, J., Atabakhsh, S., Aurela, M., Baltensperger,
 587 U., Bougiatioti, A., De Brito, J. F., Ceburnis, D., Chazéau, B., Chebaicheb, H., Daellenbach, K. R., Ehn, M., El
 588 Haddad, I., Eleftheriadis, K., Favez, O., Flentje, H., Font, A., Fossum, K., Freney, E., Gini, M., Green, D. C.,
 589 Heikkinen, L., Herrmann, H., Kalogridis, A.-C., Keernik, H., Lhotka, R., Lin, C., Lunder, C., Maasikmets, M.,
 590 Manousakas, M. I., Marchand, N., Marin, C., Marmureanu, L., Mihalopoulos, N., Močnik, G., Nęcki, J., O’Dowd,
 591 C., Ovadnevaite, J., Peter, T., Petit, J.-E., Pikridas, M., Matthew Platt, S., Pokorná, P., Poulain, L., Priestman, M.,
 592 Riffault, V., Rinaldi, M., Róžański, K., Schwarz, J., Sciare, J., Simon, L., Skiba, A., Slowik, J. G., Sosedova, Y.,
 593 Stavroulas, I., Styszko, K., Teinmaa, E., Timonen, H., Tremper, A., Vasilescu, J., Via, M., Vodička, P.,
 594 Wiedensohler, A., Zografou, O., Cruz Minguillón, M., and Prévôt, A. S. H.: European aerosol phenomenology –
 595 8: Harmonised source apportionment of organic aerosol using 22 Year-long ACSM/AMS datasets, *Environment*
 596 *International*, 166, 107325, <https://doi.org/10.1016/j.envint.2022.107325>, 2022.

597 Claey, M., Szmigielski, R., Kourtev, I., Van Der Veken, P., Vermeylen, R., Maenhaut, W., Jaoui, M.,
 598 Kleindienst, T. E., Lewandowski, M., Offenberg, J. H., and Edney, E. O.: Hydroxycarboxylic Acids: Markers
 599 for Secondary Organic Aerosol from the Photooxidation of α -Pinene, *Environ. Sci. Technol.*, 41, 1628–1634,
 600 <https://doi.org/10.1021/es0620181>, 2007.

601 Dai, Q., Hopke, P. K., Bi, X., and Feng, Y.: Improving apportionment of PM2.5 using multisite PMF by
 602 constraining G-values with a priori information, *Science of the Total Environment*, 736,
 603 <https://doi.org/10.1016/j.scitotenv.2020.139657>, 2020.

604 Dai, Q., Ding, J., Song, C., Liu, B., Bi, X., Wu, J., Zhang, Y., Feng, Y., and Hopke, P. K.: Changes in source
 605 contributions to particle number concentrations after the COVID-19 outbreak: Insights from a dispersion
 606 normalized PMF, *Science of the Total Environment*, 759, <https://doi.org/10.1016/j.scitotenv.2020.143548>, 2021.

607 Dominutti, P. A., Mari, X., Jaffrezo, J. L., Dinh, V. T. N., Chifflet, S., Guigue, C., Guyomarc'h, L., Vu, C. T.,
608 Darfeuil, S., Ginot, P., Elazzouzi, R., Mhadhbi, T., Voiron, C., Martinot, P., and Uzu, G.: Disentangling fine
609 particles (PM_{2.5}) composition in Hanoi, Vietnam: Emission sources and oxidative potential, *Science of the Total*
610 *Environment*, 923, <https://doi.org/10.1016/j.scitotenv.2024.171466>, 2024.

611 Edney, E. O., Kleindienst, T. E., Jaoui, M., Lewandowski, M., Offenber, J. H., Wang, W., and Claeys, M.:
612 Formation of 2-methyl tetrols and 2-methylglyceric acid in secondary organic aerosol from laboratory irradiated
613 isoprene/NOX/SO₂/air mixtures and their detection in ambient PM_{2.5} samples collected in the eastern United
614 States, *Atmospheric Environment*, 39, 5281–5289, <https://doi.org/10.1016/j.atmosenv.2005.05.031>, 2005.

615 Fine, P. M., Cass, G. R., and Simoneit, B. R. T.: Chemical Characterization of Fine Particle Emissions from
616 Fireplace Combustion of Woods Grown in the Northeastern United States, *Environ. Sci. Technol.*, 35, 2665–2675,
617 <https://doi.org/10.1021/es001466k>, 2001.

618 Fine, P. M., Cass, G. R., and Simoneit, B. R. T.: Chemical Characterization of Fine Particle Emissions from the
619 Fireplace Combustion of Woods Grown in the Southern United States, *Environ. Sci. Technol.*, 36, 1442–1451,
620 <https://doi.org/10.1021/es0108988>, 2002.

621 Gianini, M. F. D., Fischer, A., Gehrig, R., Ulrich, A., Wichser, A., Piot, C., Besombes, J. L., and Hueglin, C.:
622 Comparative source apportionment of PM₁₀ in Switzerland for 2008/2009 and 1998/1999 by Positive Matrix
623 Factorisation, *Atmospheric Environment*, 54, 149–158, <https://doi.org/10.1016/j.atmosenv.2012.02.036>, 2012.

624 Giugliano, M., Lonati, G., Butelli, P., Romele, L., Tardivo, R., and Grosso, M.: Fine particulate (PM_{2.5}–PM₁) at
625 urban sites with different traffic exposure, *Atmospheric Environment*, 39, 2421–2431,
626 <https://doi.org/10.1016/j.atmosenv.2004.06.050>, 2005.

627 Glojek, K., Thuy, V. D. N., Weber, S., Uzu, G., Manousakas, M., Elazzouzi, R., Džepina, K., Darfeuil, S., Ginot,
628 P., Jaffrezo, J. L., Žabkar, R., Turšič, J., Podkoritnik, A., and Močnik, G.: Annual variation of source contributions
629 to PM₁₀ and oxidative potential in a mountainous area with traffic, biomass burning, cement-plant and biogenic
630 influences, *Environment International*, 189, 108787, <https://doi.org/10.1016/j.envint.2024.108787>, 2024.

631 Graber, E. R. and Rudich, Y.: Atmospheric Chemistry and Physics Atmospheric HULIS: How humic-like are
632 they? A comprehensive and critical review, *Atmos. Chem. Phys.*, 6, 729–753, 2006.

633 Heo, J., Dulger, M., Olson, M. R., McGinnis, J. E., Shelton, B. R., Matsunaga, A., Sioutas, C., and Schauer, J. J.:
634 Source apportionments of PM_{2.5} organic carbon using molecular marker Positive Matrix Factorization and
635 comparison of results from different receptor models, *Atmospheric Environment*, 73, 51–61,
636 <https://doi.org/10.1016/j.atmosenv.2013.03.004>, 2013.

637 Hernández-Pellón, A. and Fernández-Olmo, I.: Using multi-site data to apportion PM-bound metal(loid)s: Impact
638 of a manganese alloy plant in an urban area, *Science of the Total Environment*, 651, 1476–1488,
639 <https://doi.org/10.1016/j.scitotenv.2018.09.261>, 2019.

640 Hoffer, A., Gelencser, A., Guyon, P., Kiss, G., Schmid, O., Frank, G. P., Artaxo, P., and Andreae, M. O.: Optical
641 properties of humic-like substances (HULIS) in biomass-burning aerosols, *Atmos. Chem. Phys.*, 2006.

642 Hopke, P. K.: Review of receptor modeling methods for source apportionment, *Journal of the Air and Waste*
643 *Management Association*, 66, 237–259, <https://doi.org/10.1080/10962247.2016.1140693>, 2016.

644 Hopke, P. K., Dai, Q., Li, L., and Feng, Y.: Global review of recent source apportionments for airborne particulate
645 matter, *Science of The Total Environment*, 740, 140091, <https://doi.org/10.1016/j.scitotenv.2020.140091>, 2020.

Hopke, P. K., Chen, Y., Rich, D. Q., Mooibroek, D., and Sofowote, U. M.: The application of positive matrix factorization with diagnostics to BIG DATA, *Chemometrics and Intelligent Laboratory Systems*, 240, 104885, <https://doi.org/10.1016/j.chemolab.2023.104885>, 2023.

Hu, D., Bian, Q., Lau, A. K. H., and Yu, J. Z.: Source apportioning of primary and secondary organic carbon in summer PM_{2.5} in Hong Kong using positive matrix factorization of secondary and primary organic tracer data, *J. Geophys. Res.*, 115, 2009JD012498, <https://doi.org/10.1029/2009JD012498>, 2010.

Kawamura, K. and Bikkina, S.: A review of dicarboxylic acids and related compounds in atmospheric aerosols: Molecular distributions, sources and transformation, *Atmospheric Research*, 170, 140–160, <https://doi.org/10.1016/j.atmosres.2015.11.018>, 2016.

Kourtchev, I., Warnke, J., Maenhaut, W., Hoffmann, T., and Claeys, M.: Polar organic marker compounds in PM_{2.5} aerosol from a mixed forest site in western Germany, *Chemosphere*, 73, 1308–1314, <https://doi.org/10.1016/j.chemosphere.2008.07.011>, 2008.

Li, S.-M., Barrie, L. A., Talbot, R. W., Harriss, R. C., Davidson, C. I., and Jaffrezo, J.-L.: Seasonal and geographic variations of methanesulfonic acid in the arctic troposphere, *Atmospheric Environment. Part A. General Topics*, 27, 3011–3024, [https://doi.org/10.1016/0960-1686\(93\)90333-T](https://doi.org/10.1016/0960-1686(93)90333-T), 1993.

Li, W., Ge, P., Chen, M., Tang, J., Cao, M., Cui, Y., Hu, K., and Nie, D.: Tracers from Biomass Burning Emissions and Identification of Biomass Burning, *Atmosphere*, 12, 1401, <https://doi.org/10.3390/atmos12111401>, 2021.

Li, X., Yang, K., Han, J., Ying, Q., and Hopke, P. K.: Sources of humic-like substances (HULIS) in PM 2.5 in Beijing: Receptor modeling approach, *Science of the Total Environment*, 671, 765–775, <https://doi.org/10.1016/j.scitotenv.2019.03.333>, 2019.

Lu, Z., Liu, Q., Xiong, Y., Huang, F., Zhou, J., and Schauer, J. J.: A hybrid source apportionment strategy using positive matrix factorization (PMF) and molecular marker chemical mass balance (MM-CMB) models, *Environmental Pollution*, 238, 39–51, <https://doi.org/10.1016/j.envpol.2018.02.091>, 2018.

Mardoñez, V., Pandolfi, M., Borlaza, L. S. J., Jaffrezo, J.-L., Alastuey, A., Besombes, J.-L., Moreno, I. R., Perez, N., Močnik, G., Ginot, P., Krejci, R., Chrastny, V., Wiedensohler, A., Laj, P., Andrade, M., and Uzu, G.: Source apportionment study on particulate air pollution in two high-altitude Bolivian cities: La Paz and El Alto, <https://doi.org/10.5194/acp-2022-780>, 2023.

Nicolás, J., Chiari, M., Crespo, J., Orellana, I. G., Lucarelli, F., Nava, S., Pastor, C., and Yubero, E.: Quantification of Saharan and local dust impact in an arid Mediterranean area by the positive matrix factorization (PMF) technique, *Atmospheric Environment*, 42, 8872–8882, <https://doi.org/10.1016/j.atmosenv.2008.09.018>, 2008.

Norris, G., Duvall, R., Brown, S., and Bai, S.: EPA Positive Matrix Factorization (PM F) 5.0 Fundamentals and User Guide, 2014.

Paatero, P. and Tappert, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, <https://doi.org/10.1002/env.3170050203>, 1994.

Pandolfi, M.: Supplement of Long-range and local air pollution: what can we learn from chemical speciation of particulate matter at paired sites? The copyright of individual parts of the supplement might differ from the CC BY 4.0 License, *Supplement of Atmos. Chem. Phys.*, 20, 409–429, <https://doi.org/10.5194/acp-20-409-2020-supplement>, 2020.

Pernigotti, D., Belis, C. A., and Spanò, L.: SPECIEUROPE: The European data base for PM source profiles, *Atmospheric Pollution Research*, 7, 307–314, <https://doi.org/10.1016/j.apr.2015.10.007>, 2016.

686 Pietrodangelo, A., Bove, M. C., Forello, A. C., Crova, F., Bigi, A., Brattich, E., Riccio, A., Becagli, S., Bertinetti,
 687 S., Calzolari, G., Canepari, S., Cappelletti, D., Catrambone, M., Cesari, D., Colombi, C., Contini, D., Cuccia, E.,
 688 De Gennaro, G., Genga, A., Ielpo, P., Lucarelli, F., Malandrino, M., Masiol, M., Massabò, D., Perrino, C., Prati,
 689 P., Siciliano, T., Tositti, L., Venturini, E., and Vecchi, R.: A PM10 chemically characterized nation-wide dataset
 690 for Italy. Geographical influence on urban air pollution and source apportionment, *Science of The Total*
 691 *Environment*, 908, 167891, <https://doi.org/10.1016/j.scitotenv.2023.167891>, 2024.
 692 Rogge, W. F., Medeiros, P. M., and Simoneit, B. R. T.: Organic marker compounds in surface soils of crop fields
 693 from the San Joaquin Valley fugitive dust characterization study, *Atmospheric Environment*, 41, 8183–8204,
 694 <https://doi.org/10.1016/j.atmosenv.2007.06.030>, 2007.
 695 Samaké, A., Jaffrezo, J. L., Favez, O., Weber, S., Jacob, V., Canete, T., Albinet, A., Charron, A., Riffault, V.,
 696 Perdrix, E., Waked, A., Golly, B., Salameh, D., Chevrier, F., Oliveira, D. M., Besombes, J. L., Martins, J. M. F.,
 697 Bonnaire, N., Conil, S., Guillaud, G., Mesbah, B., Rocq, B., Robic, P. Y., Hulin, A., Meur, S. L., Descheemaeker,
 698 M., Chretien, E., Marchand, N., and Uzu, G.: Arabitol, mannitol, and glucose as tracers of primary biogenic organic
 699 aerosol: The influence of environmental factors on ambient air concentrations and spatial distribution over France,
 700 *Atmospheric Chemistry and Physics*, 19, 11013–11030, <https://doi.org/10.5194/acp-19-11013-2019>, 2019a.
 701 Samaké, A., Jaffrezo, J. L., Favez, O., Weber, S., Jacob, V., Albinet, A., Riffault, V., Perdrix, E., Waked, A.,
 702 Golly, B., Salameh, D., Chevrier, F., Miguel Oliveira, D., Bonnaire, N., Besombes, J. L., Martins, J. M. F., Conil,
 703 S., Guillaud, G., Mesbah, B., Rocq, B., Robic, P. Y., Hulin, A., Le Meur, S., Descheemaeker, M., Chretien, E.,
 704 Marchand, N., and Uzu, G.: Polyols and glucose particulate species as tracers of primary biogenic organic aerosols
 705 at 28 French sites, *Atmospheric Chemistry and Physics*, 19, 3357–3374, [https://doi.org/10.5194/acp-19-3357-](https://doi.org/10.5194/acp-19-3357-2019)
 706 2019, 2019b.
 707 Srivastava, D., Tomaz, S., Favez, O., Lanzafame, G. M., Golly, B., Besombes, J. L., Alleman, L. Y., Jaffrezo, J.
 708 L., Jacob, V., Perraudin, E., Villenave, E., and Albinet, A.: Speciation of organic fraction does matter for source
 709 apportionment. Part 1: A one-year campaign in Grenoble (France), *Science of the Total Environment*, 624, 1598–
 710 1611, <https://doi.org/10.1016/j.scitotenv.2017.12.135>, 2018.
 711 Waked, A., Favez, O., Alleman, L. Y., Piot, C., Petit, J. E., Delaunay, T., Verlinden, E., Golly, B., Besombes, J.
 712 L., Jaffrezo, J. L., and Leoz-Garziandia, E.: Source apportionment of PM10 in a north-western Europe regional
 713 urban background site (Lens, France) using positive matrix factorization and including primary biogenic
 714 emissions, *Atmospheric Chemistry and Physics*, 14, 3325–3346, <https://doi.org/10.5194/acp-14-3325-2014>, 2014.
 715 Wang, Y., Hopke, P. K., Xia, X., Rattigan, O. V., Chalupa, D. C., and Utell, M. J.: Source apportionment of
 716 airborne particulate matter using inorganic and organic species as tracers, *Atmospheric Environment*, 55, 525–
 717 532, <https://doi.org/10.1016/j.atmosenv.2012.03.073>, 2012.
 718 Weber, S., Salameh, D., Albinet, A., Alleman, L. Y., Waked, A., Besombes, J. L., Jacob, V., Guillaud, G.,
 719 Meshbah, B., Rocq, B., Hulin, A., Dominik-Sègue, M., Chrétien, E., Jaffrezo, J. L., and Favez, O.: Comparison of
 720 PM10 sources profiles at 15 french sites using a harmonized constrained positive matrix factorization approach,
 721 *Atmosphere*, 10, <https://doi.org/10.3390/atmos10060310>, 2019.
 722 Wu, G., Wan, X., Gao, S., Fu, P., Yin, Y., Li, G., Zhang, G., Kang, S., Ram, K., and Cong, Z.: Humic-Like
 723 Substances (HULIS) in Aerosols of Central Tibetan Plateau (Nam Co, 4730 m asl): Abundance, Light Absorption
 724 Properties, and Sources, *Environmental Science and Technology*, 52, 7203–7211,
 725 https://doi.org/10.1021/ACS.EST.8B01251/SUPPL_FILE/ES8B01251_SI_001.PDF, 2018.

Yttri, K. E., Dye, C., and Kiss, G.: Atmospheric Chemistry and Physics Ambient aerosol concentrations of sugars and sugar-alcohols at four different sites in Norway, *Atmos. Chem. Phys.*, 7, 4267–4279, 2007.

Yttri, K. E., Simpson, D., Nøjgaard, J. K., Kristensen, K., Genberg, J., Stenström, K., Swietlicki, E., Hillamo, R., Aurela, M., Bauer, H., Offenberg, J. H., Jaoui, M., Dye, C., Eckhardt, S., Burkhardt, J. F., Stohl, A., and Glasius, M.: Source apportionment of the summer time carbonaceous aerosol at Nordic rural background sites, *Atmos. Chem. Phys.*, 11, 13339–13357, <https://doi.org/10.5194/acp-11-13339-2011>, 2011.

Zhang, S., Samaké, A., Alleman, L., Favez, O., Dinh Ngoc, T. V., and Jaffrezo, J.-L.: Identification des sources des particules sur la zone littorale des Hauts de France et de la Normandie, *Atmo Hauts-de-France ; Atmo Normandie ; LCSQA ; IGE – Institut des Géosciences de l’Environnement*, 2024.

Zheng, G., He, K., Duan, F., Cheng, Y., and Ma, Y.: Measurement of humic-like substances in aerosols: A review, *Environmental Pollution*, 181, 301–314, <https://doi.org/10.1016/j.envpol.2013.05.055>, 2013.

Zhou, Y., Huang, X. H., Bian, Q., Griffith, S. M., Louie, P. K. K., and Yu, J. Z.: Sources and atmospheric processes impacting oxalate at a suburban coastal site in Hong Kong: Insights inferred from 1year hourly measurements, *Journal of Geophysical Research*, 120, 9772–9788, <https://doi.org/10.1002/2015JD023531>, 2015.