Journal Pre-proof

Organic molecular markers and source contributions in a polluted municipality of north-east Italy: Extended PCA-PMF statistical approach

Sara Padoan, Alessandro Zappi, Thomas Adam, Dora Melucci, Andrea Gambaro, Gianni Formenton, Olga Popovicheva, Dac-Loc Nguyen, Jürgen Schnelle-Kreis, Ralf Zimmermann

PII: S0013-9351(20)30480-1

DOI: https://doi.org/10.1016/j.envres.2020.109587

Reference: YENRS 109587

To appear in: *Environmental Research*

Received Date: 8 January 2020

Revised Date: 23 March 2020

Accepted Date: 23 April 2020

Please cite this article as: Padoan, S., Zappi, A., Adam, T., Melucci, D., Gambaro, A., Formenton, G., Popovicheva, O., Nguyen, D.-L., Schnelle-Kreis, Jü., Zimmermann, R., Organic molecular markers and source contributions in a polluted municipality of north-east Italy: Extended PCA-PMF statistical approach, *Environmental Research* (2020), doi: https://doi.org/10.1016/j.envres.2020.109587.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc.

Sara Padoan: Validation, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization,

Alessandro Zappi: Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization,

Thomas Adam: Writing - Review & Editing, Supervision

Dora Melucci: Writing - Review & Editing, Supervision

Andrea Gambaro: Resources, Writing - Review & Editing,

Gianni Formenton: Resources, Writing - Review & Editing,

Olga Popovicheva: Writing - Review & Editing,

Dac-Loc Nguyen: Methodology, Validation, Writing - Review & Editing

Jürgen Schnelle-Kreis: Conceptualization, Resources, Writing - Review & Editing, Supervision

Ralf Zimmermann: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

Journal Pre-proof

Journal Pre-proof

Organic molecular markers and source contributions in a polluted municipality of North-East Italy: extended PCA-PMF statistical approach

Sara Padoan[a,b*], Alessandro Zappi[c], Thomas Adam[a,b], Dora Melucci[c], Andrea Gambaro[d], Gianni Formenton[e], Olga Popovicheva[f], Dac-Loc Nguyen[b,g,h], Jürgen Schnelle-Kreis[b], Ralf Zimmermann[b,g]

[a] Universität der Bundeswehr München, Neubiberg, Germany

[b] CMA Comprehensive Molecular Analytics, Helmholtz Zentrum München, München, Germany

[c] Department of Chemistry Ciamician, University of Bologna, Bologna, Italy

[d] Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice-Mestre, Italy

[e] Department of Regional Laboratories, Regional Agency for Environmental Prevention and Protection of Veneto, Mestre, Italy

[f] SINP Moscow State University, Moscow, Russia

[g] Chair of Analytical Chemistry and Joint Mass Spectrometry Centre (JMSC), University of Rostock, D-18051 Rostock, Germany

[h] Institute of Geophysics, Vietnam Academy of Science and Technology, Hanoi, Vietnam

***Corresponding Author:** Sara Padoan

Mailing address: Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany

Telephone number: 00498960043269

E-mail address: sara.padoan@unibw.de

Journal Pre-proof

Journal Pre-proof

Abstract

Exceeding the maximum levels for environmental pollutants creates public and scientific interest for the environmental and human health impact it may have. In Northern Italy, the Po Valley, and in particular the Veneto region, is still a hotspot for air quality improvement. Several monitoring campaigns were carried out in this area to acquire information about sources of pollutants which are considered critical. For the first time, a deep study of the aerosol organic fraction was performed in the town Sernaglia della Battaglia, nearby Treviso. During three seasons of 2017, PM_1 and $PM_{2.5}$ samples were collected simultaneously. Organic molecular markers have been analyzed by in-situ derivatization thermal desorption gas chromatography time-of-flight mass spectrometry (IDTD-GC-TOFMS). Alkanes, polycyclic aromatic hydrocarbons, oxi-polycyclic aromatic hydrocarbons, anhydrous sugars, resins acids, triterpenoids, and acids were considered. The organic chemical composition has been analyzed based on seasonal variation and source contributions. Principal Component Analysis (PCA) and Positive Matrix Factorization (PMF) have been combined to deeply investigate the main sources of particulate organic matter. On the one hand, PCA evaluates the correlations between the organic markers and their seasonal distribution. On the other hand, the source contributions to aerosol composition are estimated by PMF. Four main emission sources were found by PMF: solid fuel combustion (coal, wood), combustion of petroleum distillates (gas and fuel oil) and exhaust gases of vehicles, industrial combustion processes, home heating, and forest fires are evaluated as the most important sources for the air quality and pollution in this municipality of Northern Italy.

Keywords: PM_1, $PM_{2.5}$, particulate organic matter, PCA, PMF, source identification

Journal Pre-proof

Funding sources: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Journal Pre-proof

1. Introduction

Natural and anthropogenic emission sources contribute to different chemical composition of the coarse, fine and ultrafine atmospheric particulate matter (PM) (Chung and Seinfeld, 2002). The chemical composition of PM can drastically change due to local emissions and the impact of human activities (Seinfeld and Pandis, 1998). The organic fraction of PM, both of natural and anthropogenic origins, represent a large fraction, up to 40%, of total PM mass (Chow et al. 2006), and even 75% of the organic carbon mass was associated with submicron-sized PM (Herckes et al., 2006).

Several organic markers are characteristic for specific emission sources and a long lifetime in the atmosphere permits their detection (Cass, 1998). As an example, Li et al. (2016) reported a strong presence of sugars during a haze event in Shangai, in particular levoglucosan. Bi et al. (2008), instead, reported a full characterization of polycyclic aromatic hydrocarbon, oxi-polycyclic aromatic hydrocarbons, and n-alkanes. It is possible to quantify a wide range of other organic compounds, specific for emission sources or chemical processes in the atmosphere (Bi et al., 2008). Qualitative and quantitative analysis of different organic molecular markers can be useful to determine both natural and anthropogenic sources (Gelencsér et al., 2007). Such studies are useful to evaluate the pollution in both rural and urban sites, as reported in the already cited work of Li et al. (2016) for Shangai, or in that of Popovicheva et al. (2020) for Moscow and its surrounding area. Several studies focusing on the chemical characterization of aerosols in large urban environments have been carried out to try to answer both an environmental question (Pietrogrande et al., 2011) and a health question (Boström et al., 2002).

In the Northeastern part of Italy, the Veneto region extends to 18,345 km^2. It is one of the most industrialized areas in Italy and it is recognized by now as an European hotspot for air pollution (Masiol et al., 2013). During two monitoring campaigns carried out in 2012 and 2014, high levels of pollution have been determined in the city of Treviso and in Sernaglia della Battaglia, a small municipality located at 45° 35’ N, 12° 08’ E, 24 km north of Treviso. A further monitoring campaign has been performed in Sernaglia della Battaglia from December 23th, 2016 to November 7th, 2017, to evaluate the possible sources of pollution. The pollution in the Veneto region was found to be largely due to anthropogenic processes including transports, domestic heating, biomass burning, and industrial activities (Valotto et al., 2017).

Chemometrics is a powerful tool for the investigation of polluted atmosphere aerosol composition, and it has been widely used in the environmental field (Mas et al., 2010). A comprehensive monitoring campaign generally produces a great quantity of data that needs to be elaborated. The most used chemometric analyses are principal component analysis (PCA) (Pietrogrande et al., 2011) and cluster analysis (Manta et al., 2002). Moreover, another chemometric technique, positive matrix factorization (PMF), which is suitable for source apportionment modeling, has been applied (Hauke and Kossowski, 2011). While PCA does not consider the temporal variability of variables (pollutants), it can be used to evaluate the correlations between them and their overall changes in the aerosol composition. PMF is more suitable to determine the different sources of pollution and the temporal variability of each pollutant without considering their correlations. PCA and PMF are both factor analyses, but the second one is especially suitable for environmental data due to its non-negativity constraint (Qadir et al., 2013) and to the fact that the computation also considers the errors of the data. PCA and PMF may be used in a complementary way to enhance the data description and the understanding of the environmental problem. Such a procedure was carried out, for example, in the work by Tositti et al. (2014) for the urban environment of Bologna, Italy. The complementary use of these techniques is a valuable approach to better understand the pollutants behavior over time.

Several studies were carried out concerning the pollution of the Veneto region. At the authors' knowledge, however, only two studies analyzed the organic fraction of particulate matter in Veneto. Moreover, the work of Khan et al. (2018) was focused only on polycyclic aromatic hydrocarbons (PAH) in $PM_{2.5}$, while the one of Valotto et al. (2017) analyzed only PAHs and n-alkanes in PM_1.

This study aims to extend the knowledge about air quality in the Po Valley, by extensively study a wide range of organic markers in both size fractions, PM_1 and $PM_{2.5}$, in a highly polluted area as the Veneto region. Moreover, we investigated the seasonal trend and the correlations between the organic pollutants. Finally, we are interested in showing the possibility of obtaining reliable results by PMF to identify natural and anthropogenic emission sources using organic molecules as variables and a reduced number of samples. Indeed, Scerri et al. (2019) and Manousakas et al. (2017) obtained good results with few samples, analyzing the elemental fraction of PM.

Journal Pre-proof

2. Materials and methods

2.1 Sampling and processing

A simultaneous sampling campaign of PM_1 and $PM_{2.5}$ was carried out in Sernaglia della Battaglia, Viale della Rimembranza (GBO coordinates: 45° 52’ 18’’ N, 12° 07’ 58’’ E), in the province of Treviso, Veneto region, Northern Italy (Figure 1). Sernaglia della Battaglia is defined as an urban background according to the indications of Decision 97/101/EC "Exchange of Information" (EOI) and it was identified to evaluate the exposure of receptors sensitive. The municipal territory extends into the central-southern area of the Quartier del Piave, facing Montello hills. The area is almost flat, with altitudes ranging from 89 to 129 m above sea level, going from the southeastern end to the northern end. The main river is the Piave which marks the southern border. The river bed flows a few meters lower than the surrounding land, forming a sort of wooded escarpment. Another important river is the Soligo, which limits the municipality to the east, while some tributaries of the Piave such as the Rosper and the Raboso are much more modest. In the north-eastern area, there are humid environments typical of the high Treviso province. The municipality has about six thousand inhabitants and a population density of 300 inhabitants per square kilometer.

Journal Pre-proof

Figure 1. The sampling site, Sernaglia della Battaglia, with the main possible sources of pollution.

The winter set of aerosol samples was collected from 30.01.2017 to 11.02.2017 (9 samples), the spring set from 16.04.2017 to 06.05.2017 (21 samples) and the early summer set from 12.06.2017 to 16.06.2017 (5 samples). The PM_1 and $PM_{2.5}$ samples were simultaneously collected using low-volume samplers (Skypost Tecora) with a nominal flow of 38.3 L min^{-1}. All samples were collected daily (24 h) on quartz fiber filters (Ø = 47 mm; Whatman QMA) with low volume samplers following the European standard EN 12341:2014 (2.3 m^3 h^{-1}). Prior and after the collection, filters were conditioned for 48 h in a chamber provided with a control system for constant temperature and humidity (20 ± 1°C and of 50 ± 5% relative humidity) (Emerson S05KA Emerson Network Power, Italy), and then weighted twice with an analytical balance (Sartorius series Genius, mod. SE2, Germany, sensitivity of 0.0001 mg). The final weight was the average of two subsequent measures. Samples were stored in clean Petri slides at a temperature of −20°C to avoid sample degradation and losses of the more volatile compounds.

2.2 Instrumental analysis

PM_1 and $PM_{2.5}$ samples were analyzed by in-situ derivatization thermal desorption-gas chromatography time-of-flight mass spectrometry (IDTD-GC-TOFMS), as described in details in Orasche et al. (2011). Filters were cut into punches with diameters from 3 mm to 8 mm. Each filter punch was spiked with two internal standard mixtures. The first (non-polar) internal standard consisted of fifteen deuterated polycyclic aromatic hydrocarbons (PAH), two deuterated oxidized-polycyclic aromatic hydrocarbons o-PAH and four deuterated n-alkanes. The second (polar) internal standard mixture contained $^{13}C_6$-levoglucosan, $^{13}C_6$-vanillin, and D_{31}-palmitic acid. The quantification of the chemical compounds was carried out by external standard calibration. The samples were placed into goose-neck glass-liners for thermal desorption which were sealed with Polytetrafluoroethylene (PTFE) caps. Before analysis, liners were opened by a sampling robot (Focus, Atas GL, Netherlands) and 10 µL N-Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) was added in each liner to moisten the sample surface. After the liners were placed into the cold injector (60 °C) by a sampling robot a steady flow of carrier gas (helium) removed the air. During the venting step, the carrier gas flow was 0.7 mL min^{-1} and the split flow was 50 mL min^{-1}. After 180 s the carrier gas flow was increased splitless to 4 mL min^{-1}. The carrier gas was lead to a bypass with a cartridge filled with MSTFA before entering the injector. The injector was heated to 300°C with a heating rate of 2°C s^{-1} after the bypass was opened. During 16 min of reaction and desorption time, the carrier gas was continuously enriched with derivatization reagent. Subsequently, the bypass was closed and the column flow was set to 0.7 mL min^{-1} with a split flow of 50 mL min^{-1} after the thermal desorption was finished. Desorbed molecules were focused at 60°C on the head of the capillary column (BPX5, 25 m, 0.22 mm ID, 0.25 µm film, SGE, Australia) which was installed in an Agilent 6890 gas chromatograph (Agilent, USA). The thermal desorption step was followed by heating up the GC oven to 130°C with a rate of 80°C min^{-1}. Then the rate was lowered to 8°C min^{-1} until a temperature of 330°C was reached followed by an isothermal time of 30 min. Identification and quantification of target compounds

Journal Pre-proof

were carried out on a Pegasus III TOFMS using Chroma TOF software package (LECO, USA) being capable of peak deconvolution. The data acquisition range was 35 to 500 m/z with an acquisition frequency of 25 spectra per second which is necessary for reliable peak deconvolution.

The measurement of a sample series (10 samples) was followed by a measure of reference material to check the intensity of the two internal standard solutions. Then, an empty liner was measured as blank. No considerable memory effects were observed in the reference samples or the blank runs.

2.3 Statistical analysis

To evaluate the relations between the variables (chemical species) and their behavior over time, the GC-MS data were analyzed with several chemometric methods. The correlations between the variables were evaluated computing Spearman's correlation indexes (Press and Press, 2013). Spearman correlation is a nonparametric way of computing correlation that is more suitable than Pearson's correlation when there is no linear relationship between variables or variables are not normally distributed (Hauke and Kossowski, 2011). Principal Components Analysis (PCA) (Bro and Smilde, 2014) was then used to evaluate the distribution of the variables over the three considered seasons. PCA is a well-known chemometric procedure that allows rotating the space spanned by the original variables to a new space, spanned by the Principal Components (PCs), in which most of the information contained in the original data is reported in the first (generally two or three) PCs. Therefore, visualizing the 2- or 3D plot of PC1 vs PC2 (and/or PC3) allows studying the behavior of samples (in the scores plot) and variables (in the loadings plot).

Variables have been further investigated by PMF (Paatero and Tapper, 1994). PMF, as well as PCA, rotates the space spanned by the original variables. However, unlike PCA, the calculated factors are not orthogonal to each other. This means that the solution obtained by PMF is not unique, but the model may be further rotated to obtain a different solution, that can be better or worse than the previous one. The computation, therefore, needs also the standard deviations associated with each measurement to choose the most suitable number of factors and to evaluate the rotational stability of the obtained solution (Paatero and Tapper, 1994). Thus, the proposed PMF model is used to evaluate the inner characteristics of the data. It was demonstrated (Cruz and Pandis, 1997; Scerri et al., 2019) that reliable PMF models can be obtained with proper precautions even with datasets containing a small number of samples and by using PM_1 and $PM_{2.5}$ datasets together (Scerri et al., 2019). Rotational stability of the PMF model was evaluated by displacement, bootstrapping and a combination of both (Scerri et al., 2019).

While PCA is useful to evaluate the correlation between variables and observations, it can be also used to describe each observation by itself whereas PMF is more focused on the description of the

Journal Pre-proof

whole dataset. Therefore, PCA may be used to describe how every single variable influences the observations and if there are observations or groups of observations that have peculiar characteristics. PMF, instead, loses the focus on every single observation or variable and considers the temporal trend of all variables to look for a possible common source for them. Therefore, the complementary use of PCA and PMF is a good way to obtain an overall statistical description of data, both considering each observation as "independent" and as part of a temporal trend.

The data matrices were used as they are, which means that values below the detection limit were included as such without substituting them with some indicators. Standard deviations were estimated as the regression standard deviations.

Spearman correlations and PCA were computed with the software R v3.5.2 (R Core Team, Vienna, Austria), while PMF was computed with EPA PMF v5.0 (Sonoma Technology Inc., Petaluma, CA).

Journal Pre-proof

3. Results and discussion

3.1 Analyzed pollutants overview

Table 1 shows an overview of the 62 analyzed compounds in PM_1 and $PM_{2.5}$. In this table, the mean and median value, minima and maxima concentration and limit of quantification (LoQ) are reported for each variable in each PM size class. These markers were chosen based on the ones already identified as characteristic for sources of pollution in Pietrogrande et al. (2011) and the reported literature. LoQs were calculated as $6s_{y/x}/b$ where $s_{y/x}$ is the standard error and b is the slope of the regression line specific for that pollutant.

			PM1				PM2.5				PM1/ PM2.5 mean ratio
Molecules	**Code**	**LOQ**	**Mean**	**Median**	**Min***	**Max**	**Mean**	**Median**	**Min***	**Max**	
Acids											
Eicosanoic acid	EI-AC	0.03	3.29	2.08	0.88	12.4	4.29	2.64	0.00	32.0	0.77
Adipic acid	AD-AC	0.01	0.11	0.00	0.00	1.50	0.25	0.00	0.00	2.48	0.43
Alkanes											
Docosane	C22	0.01	1.58	0.84	0.41	7.28	2.30	1.20	0.41	14.4	0.69
Tricosane	C23	0.01	1.82	0.95	0.24	10.5	2.81	1.35	0.24	19.6	0.65
Tetracosane	C24	0.01	2.93	1.82	0.00	11.6	3.28	2.01	0.87	15.4	0.89
Pentacosane	C25	0.01	2.13	1.18	0.65	9.89	2.49	1.47	0.71	12.8	0.85
Hexacosane	C26	0.01	2.17	1.48	0.00	8.50	2.68	1.64	0.64	14.6	0.81
Heptacosane	C27	0.01	1.71	1.24	0.62	7.31	2.19	1.49	0.18	11.6	0.78
Octacosane	C28	0.01	1.24	0.71	0.14	4.93	1.50	0.92	0.03	7.50	0.82
Nonacosane	C29	0.01	1.79	1.13	0.59	8.31	2.56	1.75	0.63	14.6	0.70
Triacontane	C30	0.02	0.89	0.67	0.00	2.58	1.19	0.78	0.49	4.57	0.75
Hentriacontane	C31	0.02	1.19	0.82	0.00	4.73	1.84	1.32	0.57	8.68	0.64
Dotriacontane	C32	0.02	0.98	0.46	0.00	7.60	1.11	0.55	0.31	7.32	0.89
Tritriacontane	C33	0.02	0.62	0.48	0.00	2.15	0.96	0.65	0.30	4.31	0.65
PAH											
Fluoranthene	FLU	0.01	0.52	0.11	0.00	4.84	1.03	0.15	0.03	10.7	0.51
Acephenanthrylene	ACE	0.01	0.16	0.03	0.00	1.58	0.31	0.05	0.00	3.68	0.50
Pyrene	PYR	0.01	0.73	0.18	0.02	6.48	1.36	0.20	0.00	15.0	0.54
Benzo[c]phenanthrene	BcFA	0.01	0.12	0.03	0.00	1.05	0.21	0.03	0.00	2.42	0.59
Benzo[ghi]fluoranthene	BghiF	0.02	1.01	0.16	0.00	7.21	1.56	0.16	0.00	13.2	0.64
Benz[a]anthracene	BaA	0.01	0.84	0.08	0.00	7.69	1.28	0.09	0.00	15.1	0.66
Cyclopenta[cd]pyrene	CPP	0.03	1.90	0.17	0.00	17.6	2.66	0.21	0.00	30.7	0.71
Chrysene	CRY	0.01	1.73	0.17	0.00	15.6	2.69	0.22	0.00	28.0	0.64
Sum of Benzofluoranthenes	BbkF	0.001	3.56	0.77	0.00	32.8	4.44	0.91	0.00	32.8	0.80
Benz[e]pyrene	BeP	0.01	2.53	0.73	0.00	16.9	3.42	0.87	0.05	25.8	0.74

Journal Pre-proof

Benz[a]pyrene	BaP	0.02	0.86	0.08	0.00	7.54	0.91	0.10	0.00	9.33	0.94
Perylene	PER	0.02	0.22	0.06	0.00	1.65	0.32	0.07	0.00	2.72	0.70
Anthanthrene	AN	0.01	0.09	0.00	0.00	0.92	0.13	0.00	0.00	1.23	0.69
Dibenz[ah]anthracene	DiBaA	0.01	0.06	0.00	0.00	0.58	0.09	0.00	0.00	0.71	0.69
Indeno[1,2,3-cd]pyrene	IND	0.01	0.52	0.12	0.00	3.90	0.76	0.14	0.00	6.47	0.69
Picene	PIC	0.04	0.08	0.00	0.00	0.72	0.12	0.00	0.00	0.96	0.66
Benzo[ghi]perylene	BaPy	0.02	1.30	0.32	0.00	9.40	1.80	0.39	0.00	13.5	0.72
Coronene	COR	0.04	0.36	0.00	0.00	3.05	0.45	0.00	0.00	3.74	0.81
Retene	RET	0.03	0.06	0.00	0.00	0.51	0.11	0.00	0.00	1.02	0.57
PAH 302:1	PAH-302:1	0.04	0.01	0.00	0.00	0.20	0.03	0.00	0.00	0.31	0.56
PAH 302:2	PAH-302:2	0.04	0.01	0.00	0.00	0.20	0.02	0.00	0.00	0.21	0.68
1-Methyl-benz[a]anthracene	1-MBaA	0.02	0.08	0.00	0.00	1.08	0.14	0.00	0.00	2.02	0.56
Naphtho[1,2-kb]fluoranthene	N-12-kbF	0.01	0.30	0.00	0.00	3.85	0.61	0.00	0.00	5.96	0.48
Dibenz[al]pyrene	DBalP	0.01	0.25	0.00	0.00	2.92	0.31	0.00	0.00	3.83	0.80
Naphtho[2,3-b]fluoranthene	N-23-bF	0.01	0.17	0.00	0.00	2.32	0.33	0.00	0.00	3.96	0.51
Dibenz[ae]pyrene	DBaeP	0.01	0.08	0.00	0.00	1.18	0.18	0.00	0.00	1.71	0.42
Naphtho[2,1-a]pyrene	N-12-aP	0.01	0.17	0.00	0.00	2.22	0.28	0.00	0.00	3.11	0.61
Naphtho[2,3-a]pyrene	N-23-aP	0.01	0.07	0.00	0.00	1.08	0.12	0.00	0.00	1.42	0.59
Dibenz[ai]pyrene	DBaiP	0.02	0.08	0.00	0.00	1.10	0.14	0.00	0.00	1.70	0.62
Dibenz[ah]pyrene	DBahP	0.03	0.05	0.00	0.00	0.95	0.09	0.00	0.00	1.50	0.57
o-PAH											
1H-Phenalen-1-one	1HPhe1one	0.2	2.77	0.00	0.00	26.3	4.90	0.00	0.00	55.9	0.57
Xanthone	Xan	0.1	0.08	0.00	0.00	1.35	0.16	0.00	0.00	2.13	0.49
9,10-Anthracenedione	ANQ-DO	0.1	0.58	0.30	0.00	3.47	0.89	0.34	0.00	7.72	0.66
Cyclopenta(def)phenanthrenone	CPPH-O	0.03	0.20	0.00	0.00	1.61	0.30	0.00	0.00	4.00	0.65
1,8-Naphthalic anhydride	NAP-AN	0.1	3.09	0.92	0.00	23.2	4.08	1.21	0.00	35.4	0.76
1,8-Naphthalaldehydic acid	NAP-AC	0.1	6.20	2.49	0.00	35.4	7.90	3.58	0.00	47.5	0.78
11H-Benzo[a]fluoren-11-one	11HBaFone	0.03	0.44	0.09	0.00	3.80	0.60	0.09	0.00	6.90	0.73
7H-Benzo[c]fluorene-7-one	11HBCFone	0.03	0.22	0.00	0.00	1.86	0.30	0.06	0.00	3.35	0.73
11H-Benzo[b]fluoren-11-one	11HBbFone	0.03	0.89	0.21	0.00	7.36	1.16	0.22	0.00	12.8	0.77

Journal Pre-proof

7H-Benzo[de]anthracen-7-one	11HBdeFone	0.03	1.79	0.31	0.00	15.4	2.44	0.32	0.00	28.7	0.74
Anhydrous Sugars											
Galactosan	GAL	0.1	5.31	1.75	0.35	45.6	7.30	1.79	0.33	74.6	0.73
Mannosan	MAN	0.1	22.3	12.1	0.68	169	27.8	11.66	0.74	238	0.80
Levoglucosan	LEV	0.1	456	182	19.3	2350	546	203	37.9	2700	0.83
Resins Acids											
Isopimaric acid	IS-AC	0.01	0.01	0.00	0.00	0.05	0.01	0.00	0.00	0.07	0.76
Dehydroabietic acid, methyl ester	DEH-AC-M	0.01	0.15	0.08	0.01	0.73	0.16	0.08	0.00	0.99	0.93
Dehydroabietic acid	DEH-AC	0.01	2.00	0.84	0.04	13.8	2.68	1.08	0.00	15.6	0.74
Triterpenoids											
α-sitosterol	a-sit	0.7	13.1	0.00	0.00	156	19.8	0.00	0.00	233	0.66
Stigmasta-3,5-dien-7-one	Stig	0.7	1.02	0.00	0.00	11.09	1.28	0.00	0.00	12.6	0.80

Table 1. Statistical summary of variable concentrations. All values are reported in ng m^3

* A minimum value of 0 indicates that in the chromatogram that specific analyte has not been found.

Journal Pre-proof

In **PM_1**, anhydrous sugars are found to be the most concentrated compounds during winter and spring, while, during summer, acids is the class of compounds most present in PM_1. Levoglucosan showed the highest concentration during all three seasons. It had a median concentration of 1080 ng m^{-3}, 162 ng m^{-3} and 25.0 ng m^{-3} during winter, spring and summer, respectively. *Acids* represented 3%, 32%, and 59% of the chemical markers during winter, spring, and summer respectively. Among the acids investigated, eicosanoic acid was the most concentrated during all three seasons, followed by the adipic acid. *Alkanes* represented 2%, 3% and 13% of the chemical markers during winter, spring and summer respectively. The detailed composition of alkanes showed a homogenous distribution where tetracosane was the most concentrated during the investigated seasons. During winter, *PAHs* represented 4% of the chemical markers. During spring and summer, they were found only in a few traces. Benzo fluoranthenes were the most concentrated compounds during winter and spring, replaced as the most concentrated by benzo[e]pyrene during summer. o-*PAHs* represented 3%, 1% and 1% of the chemical markers during winter, spring and summer respectively. Among all o-PAHs, 1,8-naphthalaldehydic acid was the most concentrated compound, during all the three periods. *Resin acids* were present only in traces. Among three studied resin acids, dehydroabietic acid was the most concentrated during all periods. It showed its maximum during winter with a median concentration equal to 4.23 ng m^{-3} and its minimum during the early summer with a median concentration equal to 0.1 ng m^{-3}. *Triterpenoids* were present only in traces and showed significant concentrations during winter only, when they represented 3% of the chemical markers with α-sitosterol the most concentrated compound.

In **$PM_{2.5}$**, anhydrous sugars were the most concentrated compounds during all three seasons. Anhydrous sugars represented 82%, 67% and 39% of the chemical markers during winter, spring and summer respectively. Among anhydrous sugars, levoglucosan was the most concentrated during all three seasons. During winter, anhydrous sugars were followed by triterpenoids (median concentration of 43.7 ng m^{-3}), PAH (52.7 ng m^{-3}) and o-PAH (43.4 ng m^{-3}), with a respective total concentration of 4% of the chemical markers. Acids (46.8 ng m^{-3}), represented respectively 3% of the chemical markers. During spring, acids (77.6 ng m^{-3}) represented 26% of the chemical markers, followed by alkanes (13.3 ng m^{-3}), o-PAH (4.91 ng m^{-3}), and PAH (3.30 ng m^{-3}) which represented 4%. 2% and 1%, respectively. During summer, acids (57.5 ng m^{-3}) represented 46% of the chemical markers, alkanes 14% (18.2 ng m^{-3}) and o-PAH 1% (1.10 ng m^{-3}). All the chemical compounds, except acids, showed higher concentrations during winter with a remarkable decrease during spring and even more during early summer sampling periods. Among *acids*, eicosanoic was the most concentrated. Among *alkanes*, during the winter, tricosane (4.21 ng m^{-3}) was the most concentrated alkane. During spring, tetracosane (1.63 ng m^{-3}) was the most concentrated alkane and during summer, nonacosane (2.31 ng m^{-3}) was the most concentrated alkane. Among *o-PAH*s, 1,8-

Journal Pre-proof

naphthalaldehydic acid was the most concentrated during all seasons, showing a median concentration of 18.2 ng m^{-3} during winter, of 3.54 ng m^{-3} during spring and of 0.97 ng m^{-3} during the early summer. Among *resin acids*, dehydroabietic acid was the most concentrated during all seasons. It showed a median concentration of 5.86 ng m^{-3} during winter, of 0.955 ng m^{-3} during the spring and of 0.114 ng m^{-3} during the early summer. *Triterpenoids* were present only during winter and α-sitosterol was the most concentrated compound, showing a median concentration of 38.8 ng m^{-3}, while stigmasta-3,5-dien-7-one showed a median concentration of 4.87 ng m^{-3}.

The last column of Table 1 shows the ratio between the mean values of organic constituents in PM_1 (submicron-sized mode) over those in $PM_{2.5}$ (fine-sized mode). The ratios from organic groups of anhydro-sugars, resin acids, alkanes range respectively from 0.73 - 0.83, 0.74 – 0.93, and 0.65 – 0.89, showing that they are more found in submicron particles, consistent with results from other studies (Herckes et al., 2006; Li et al., 2016). On the other hand, other groups of acids, PAHs, o-PAHs, and triterpenoid ranges from 0.43 – 0.77, 0.51 – 0.94, 0.49 – 0.76, and 0.66 – 0.80, respectively, indicating that they are more shifted toward fine-sized mode.

3.2 Statistical analysis

3.2.1 Analytical results

The most general chemometric analysis carried out on PM data is the computation of the Spearman's correlation indexes (ρ). A description of the dataset, considering a general temporal trend of the variables, considering also the correlation indexes, is useful to extract some information about the data even before any other statistical analysis.

Acids were investigated by a short chain acid, AD-AC (C6:0) and a long chain one, EI-AC (C20:0). All samples showed a higher concentration in EI-AC than AD-AC. Taking into account the correlations, it is clear that the two considered acids have different emission sources. Indeed, their correlation is poor ($\rho < 0.6$). AD-AC is an oxidation product from other organic compounds. As reported by Cruz and Pandis (1997), AD-AC can be considered as a secondary organic aerosol (SOA) compound, reflecting the long-range transport of the aerosols. It is an ozonolysis product of cyclohexene products of the urban environment (Cruz and Pandis, 1997). AD-AC concentrations were higher during warmer periods than colder periods, highlighting a major contribution of SOA during spring and summer than winter. Barbaro et al. (2020) suggested that the higher concentration in summer can be also due to enhanced photochemical production in the warmer period.

N-alkanes showed higher concentrations during colder periods than warmer periods with the maximum concentration during winter decreasing during spring and early summer. The abundance distribution of different n-alkane chains showed seasonal variations highlighting changes in n-alkanes

Journal Pre-proof

predominant emission sources. During winter, C22-C29 alkanes showed higher concentrations highlighting anthropogenic emission sources as dominant emission sources, mostly diesel fuel and lubrication oil (Popovicheva et al., 2016). Precisely, gasoline emissions are mostly characterized by high contents of C4 to C10 n-alkanes, while diesel fuel consists mostly of C8 to C25 hydrocarbons (Pietrogrande et al., 2011). Diesel and gasoline engine exhaust is a more complex mixture. The common emission source(s) of the group C22-C29 n-alkanes is also confirmed by the generally high correlation indexes ($\rho > 0.7$) between them. The high correlation indexes were found also in another work (Valotto et al., 2017), which obtained two groups of n-alkanes with correlations higher than 0.9: C23-C26 and C27-C33. During spring and early summer, n-alkanes showed a little shift of composition with lower concentrations of shorter chains C22 and C23 and higher concentrations of C24-C31, with the presence of heavier n-alkanes ($C > C27$). High molecular weight n-alkanes are likely derived from terrestrial higher plants (Rissanen et al., 2006)and are introduced into the aerosol atmosphere from plant abrasion.

Ten *PAHs* were discussed as the main relevant markers of anthropogenic emissions sources: FLU, PYR, BaA, CRY, BbkF, BeP, BaP, IND, BaPy, and COR. These variables show strong correlations ($\rho > 0.8$) between each other in all PM. COR, however, has a slight correlation ($\rho \approx 0.6$) with all the other PAHs. PAHs show also a strong seasonality. Higher PAH concentrations were found during winter. This abundance is mostly related to the increased emissions from primary sources depending on domestic heating (Okuda et al., 2010). This seasonal emission is coupled with the increasing emissions coming from motor vehicles. Furthermore, BbkF, the most concentrated during the winter and spring seasons (30.01-11.02 and 16.04-6.05) suggested also a coal combustion origin, coming from coal-burning for residential heating (Chen et al., 2005). Furthermore, the ratio BeP/BaP ($\rho > 0.9$ in all PMs) showed higher values during the colder period, than the warmer periods. During winter, the ratio was equal to 2, while during the spring period it was equal to 3 and during the early summer period, it was equal to 4. A previous work (Schnelle-Kreis et al., 2005) suggests that lower values indicate a higher influence of non-traffic sources. The results obtained for PAH in this work are in very good agreement with the ones obtained by Khan et al. (2018), who analyzed in particular BaA, CRY, BaP, DiBaA, BaPy, IND, and the fluoranthenes in $PM_{2.5}$. The annual means reported in that study are not significantly different from our ones (at significance level 0.05).

o-PAH were investigated because of their toxicological importance. Indeed, they might be involved in the occurrence of oxidative stress (Pietrogrande et al., 2011). These compounds show strong correlations ($\rho > 0.7$) between each other in all PM. O-PAHs were more concentrated during the colder periods than warmer periods, suggesting the increasing impact of traffic and domestic heating. The hypothesis, supported by the previous work of Pietrogrande et al. (2011), that o-PAHs, during the colder months are mainly emitted from primary sources seems to be confirmed by the fact that during the colder months the photodegradation of PAHs is slowed down by the less photochemical activity and lower ozone concentrations (Schnelle-Kreis et al., 2005). Nineteen PAH and o-PAH molecules (AN, DiBaA, PIC, RET, PAH-302:1, PAH-302:2, 1-MBaA, N-12-kbF, DBalP, N-23-bF, DBaeP, N-12-aP, N-23-aP, DBaiP, DBahP, 1HPhe1one, Xan, CPPH-O, IS-AC) were found only in winter samples. Xan and CPPH-O are o-PAH involved in oxidative stress (Ladji et al., 2009). These winter variables show high correlations in PM_1 ($\rho > 0.7$) between them and the correlations generally increase in $PM_{2.5}$ ($\rho > 0.8$). The only exception is 1-MBaA, which has poor correlations ($\rho < 0.6$) with all the other variables in both PM_1 and $PM_{2.5}$.

Triterpenoids were found only during winter. Usually, triterpenoids are associated with the use of oils in cooking. α-Sitosterol (a-sit) has been mainly found in emissions from corn oil, olive oil, and wheat germ oil (Sims et al., 1972). Stigmasta-3,5-dien-7-one (Stig) is one of the principal sterols in the plant foodstuffs and it has been found in soybean and rapeseed oils. Besides, Stig has been found in oxidized soybean oils at a temperature between 30°C and 44°C for to 13 days (Finocchiaro and Richardson, 1983). The quantification of both a-sit. and Stig shows an influence of cooking emission on PM1 and $PM_{2.5}$ mostly during the coldest period. Their correlation is slight ($\rho \approx 0.6$) in PM_1, but is very strong in $PM_{2.5}$ ($\rho > 0.9$).

The biomass burning markers, levoglucosan (LEV) and dehydroabietic acid (DEH-AC), showed quite similar temporal trends. Due to a large amount of wildfire and residential wood heating emissions, LEV has been accepted as a specific marker of biomass burning (Hoffmann et al., 2010). LEV and DEH-AC showed higher concentration during winter than during spring and early summer. This is a clear indicator of high emissions of smoke particles, associated with residential wood-burning (Hoffmann et al., 2010; Nielsen, 1996). The higher concentrations of these compounds during the winter season highlight the impact of the massive usage of local residential heating. Furthermore, the ratio between LEV, emitted by the burning of several types of cellulose-containing biomass, and DEH-AC, only emitted by gymnosperm combustion, can be used to distinguish between different types of wood combustion contributions. Ratio values found in these samples are consistent with the composition of smoke emissions from residential wood burning in urban sites (Oliveira et al., 2007). Spearman correlation indexes in all PM datasets highlighted a high correlation ($\rho > 0.7$) between LEV and DEH-AC, giving evidence of their same origin from biomass burning sources. Moreover, statistical analysis showed a high correlation ($\rho > 0.7$) also between LEV and its isomers galactosan (GAL) and mannosan (MAN). Dehydroabietic acid methyl ester (DEH-AC-M) (partially formed by esterification of dehydroabietic acid with wood alcohol during the combustion) and DEH-AC have $\rho > 0.7$ too, showing a similar temporal course of markers coming from the biomass burning processes.

3.2.2 Principal component analysis

Journal Pre-proof

Some variables were found to be present only during the winter period: AN, DiBaA, PIC, RET, PAH-302:1, PAH-302:2, 1-MBaA, N-12-kbF, DBalP, N-23-bF, DBaeP, N-12-aP, N-23-aP, DBaiP, DBahP, 1HPhe1one, Xan, CPPH-O, IS-AC, a-Sit, Stig. In **Table 1**, these are characterized by a median value of 0. Such variables were excluded from the PCA analysis. This choice has been made because the influence of such variables could hide the information given by the others (most of all on winter samples), making not reliable the interpretation of the chemometric results. Once removed those 21 variables from the dataset, the experimental data were divided into two datasets, corresponding to the two analyzed PM size classes reported in **Table 1**: PM_1 and $PM_{2.5}$. The subsequent chemometric analyses were then applied to each dataset separately. Both PM_1 and $PM_{2.5}$ datasets have the dimensions of 35 x 44. Data were auto-scaled before PCA analysis. This means that each digit of the dataset has been subtracted by the column mean and the product divided by the column standard deviation.

Figure 2 shows the bi-plots obtained by the PCA carried out on the two datasets. It reports both scores and loadings in the same PC1 vs PC2 graph. PC1 and PC2 carry 72.4% of explained variance for PM_1 (**Figure a**) and 81.6% for $PM_{2.5}$ (**Figure b**). The two PMs have the same behavior: the scores indicate evident discrimination of the winter samples (green points) from the spring (black) and summer (red) ones. In PM_1, and partially in $PM_{2.5}$, good discrimination of spring and summer samples can also be observed. Most of the variables are in the positive region of PC1, where also the winter aerosol composition is present. It indicates that the winter composition is well discriminated because most of the studied compounds have higher concentrations in this period rather than in spring or summer. These PCA models were computed after removing three winter days from both datasets: 30.01.2017, 31.01.2017 and 01.02.2017. The scores of these samples were found at even higher values of PC1 and “pulled” also most of the variables at higher PC1 producing a graph difficult to interpret. The only variables that show high concentration in spring and summer as well are AD-AC and C30 for PM_1 and DEH-AC-M for $PM_{2.5}$.

To further improve the study on spring and summer aerosol composition, PCAs were recomputed excluding all winter data. Results (bi-plots) are reported in **Figure 3**. The explained variances carried by these models (PC1+PC2) are 62.9% for PM_1 (**Figure a**) and 60.8% for $PM_{2.5}$ (**Figure b**). In this case too, the behavior of the samples is the same for the two PMs. The summer composition is well discriminated from the spring one, and the variables that mostly describe these compositions (i.e. the variables mostly concentrated in summer) are AD-AC and alkanes. This distribution of AD-AC and alkanes with respect to the other variables also explains the low Spearman’s correlation previously observed between these variables and all the others. The other variables (the remaining PAH and all o-PAH, anhydrous sugars, and resin acids) are generally higher in spring samples.

Journal Pre-proof

From **Figure 3** it is interesting to note a group of spring samples in both PCA models at high positive values of PC2 and PC1 close to zero. No variables are present in such a region, indicating a general low concentration of organic particulate. After checking for the weather conditions in these days, the temperatures in these days resulted to be higher than in the other spring days: while the mean temperature of the other days was around 5°C, the temperature of these was at least 7°C. By considering also that the three winter samples excluded from the previous PCA (being too concentrated) were also the coldest of the entire dataset (their temperature were found around -2°C), a general consideration, even if not demonstrated, can be made from PCA models: a higher temperature, even of few degrees only, may be sufficient to reduce the pollution due to organic molecules.

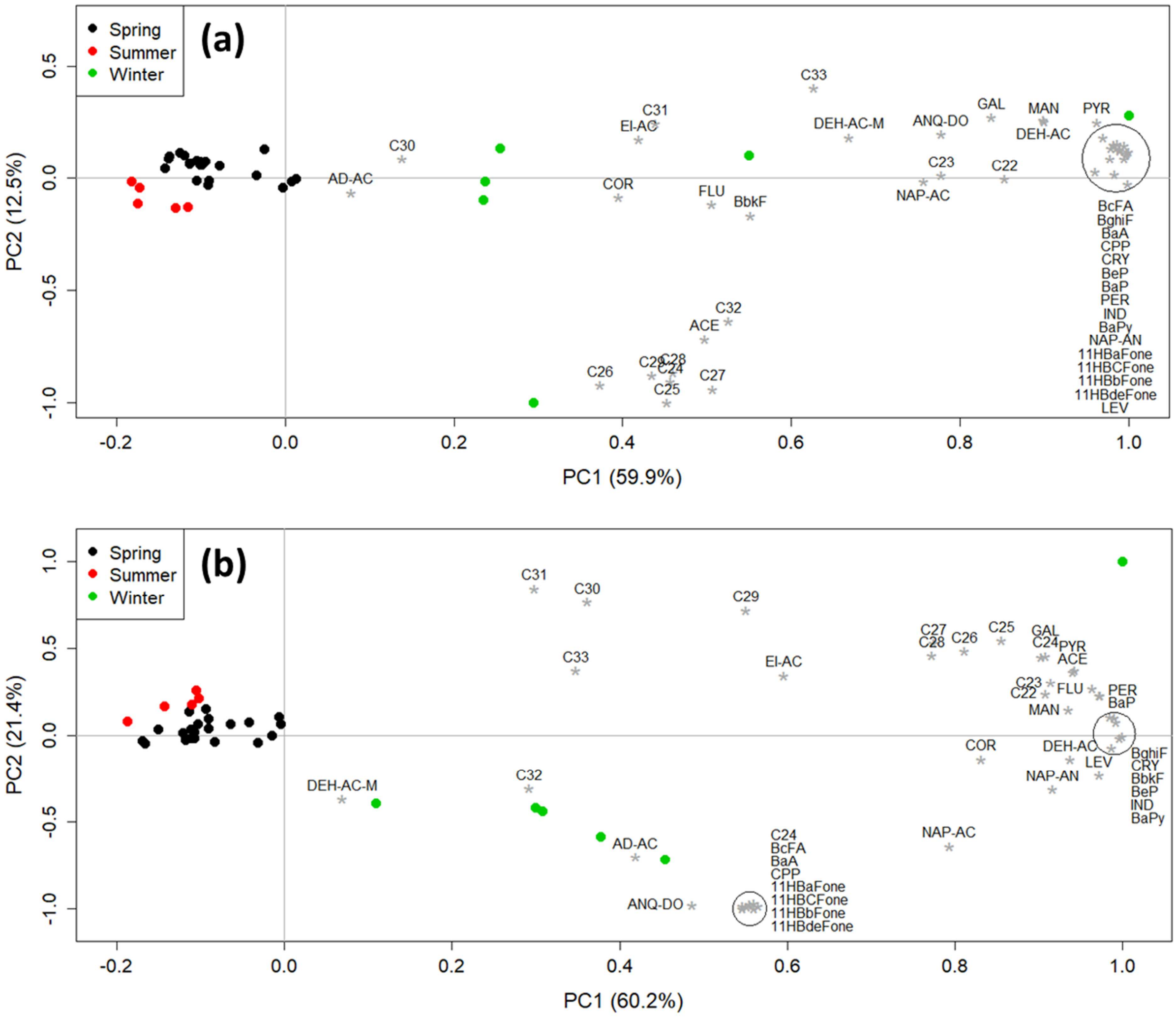

Figure 2. PCA bi-plot of the two datasets: PM_1 (a) and $PM_{2.5}$ (b). Black points represent daily spring aerosol composition, red points summer samples, and green points winter samples. Loadings are represented by asterisks (*). Variables codes are reported on top of loadings.

Journal Pre-proof

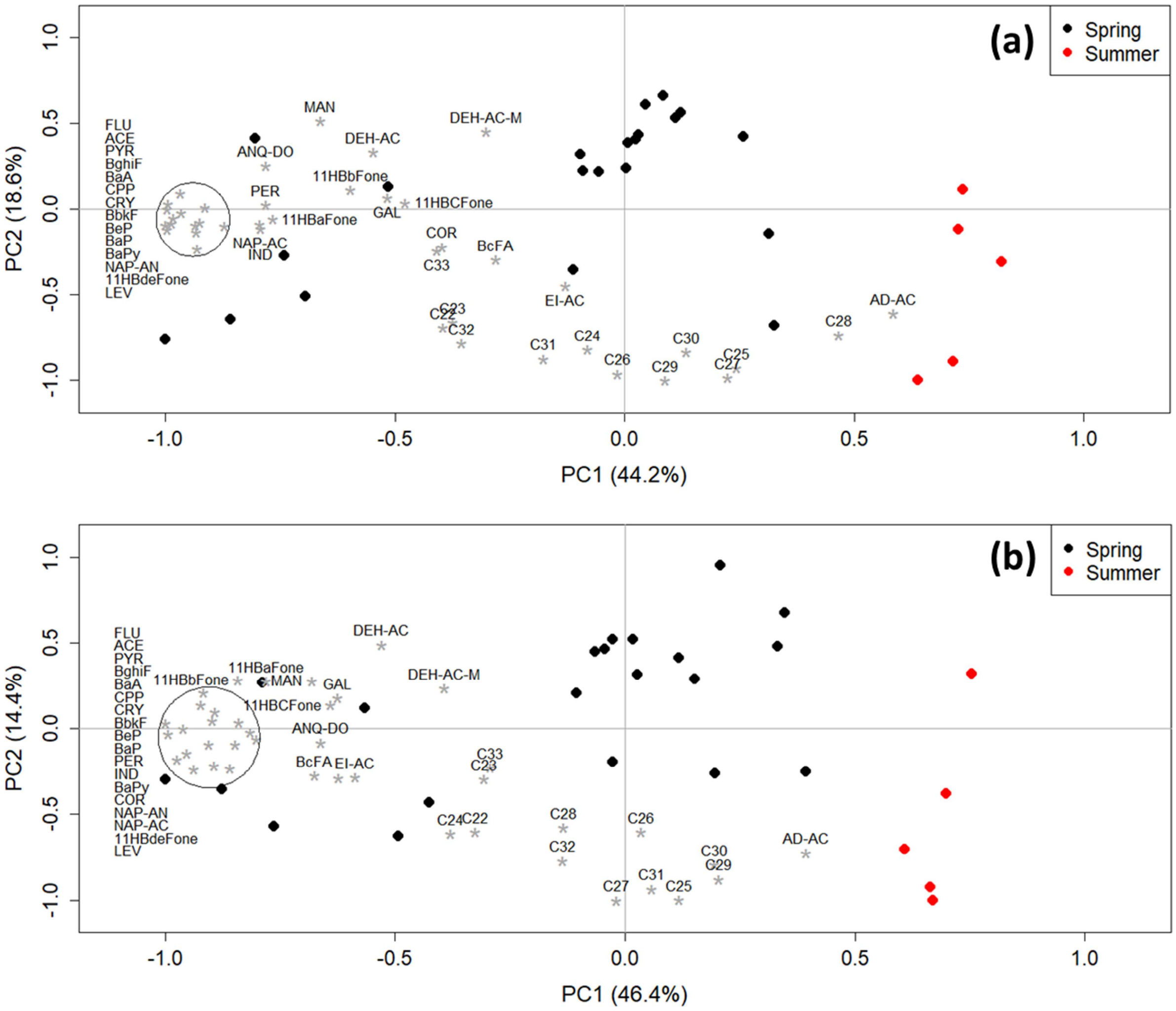

Figure 3. PCA carried out on two datasets with spring (black points) and summer (red points) aerosol composition only. PM_1 (a) and $PM_{2.5}$ (b) plot. Loadings are represented by asterisks (*). Variables codes are reported on top of loadings and sample dates are reported above the scores.

3.2.3 Positive matrix factorization (PMF)

The main possible emission sources of the particulate organic matter have been investigated and identified computing PMF with all samples (70). To have enough samples for a reliable PMF model, PM_1 and $PM_{2.5}$ datasets were gathered in a single dataset (Press and Press, 2013). The possibility of merging the two datasets was confirmed by a MANOVA (Multivariate Analysis of Variance) test (Stahle and Wold, 1990). It showed that the two datasets can be considered a unique statistical population because no significant differences were found (p-value = 0.730). After a first PMF computation with all 62 analyzed variables, it was decided to remove the twelve PAH variables only present in winter from the dataset: PAH-302:1, PAH-302:2, 1-MBaA, N-12-kbF, DBalP, N-23-bF, DBaeP, N-12-aP, N-23-aP, DBaiP, DBahP, 1HPhe1one. These variables added difficulty in interpreting the factors. The other variables present only in winter samples (AN, DiBaA, PIC, RET, IS-AC, a-sit, Stig) were considered “weak” in EPA-PMF computation, thus their uncertainties (standard deviations) were automatically tripled. The final dataset was composed of 48 variables, 7 of which were considered "weak". Standard deviations were estimated as the regression standard deviations.

A three-factor solution was selected for this study (**Figure 4**). A 2-factor solution mixed the emission sources and a 4- or 5-factor solutions split the factors with no reasonable interpretation, thus these were discarded. Each one of the three factors was characterized by a clear pattern of organic species which could be associated with one specific or a group of specific emission sources.

Errors in the solution and rotational ambiguities of the PMF model were tested by displacement (DISP), bootstrap (BS), and the combination of displacement and bootstrap (BS-DISP) (Scerri et al., 2019). The computation of such methods is included in EPA-PMF version 5.0 and was applied to the base-run solution (BS was applied with 100 runs and a minimum correlation value of 0.6). DISP analysis showed no swaps between factors and BS showed that the factors were reproduced by 96%. BS-DISP accepted 96% of cases, too. However, the percentage change in the Q parameter for the best-calculated solution was 0.09%, which is less than 0.5%, which is the maximum change recommended. Thus, it was decided to accept the PMF model without any further rotational constrain.

Factor 1 has been associated partially to the coniferous wood combustion as it was mostly dominated by DEH-AC (55.4% of species) and DEH-AC-M (65.3%). DEH-AC and DEH-AC-M have been proposed as markers for coniferous wood combustion (Leithead et al., 2006). The significant presence of LEV (42.5%), MAN (49.7%), and GAL (40.0%), confirm the general wood combustion source.

Factor 2 has been associated with the biomass burning source and was highly dominated by PAHs, o-PAHs, and anhydrous sugars. LEV (49.0%) is a specific marker for biomass burning, accompanied by low quantities of its isomers: MAN (43.4%) and GAL (46.9%). Furthermore, LEV is specifically formed during the pyrolysis process of cellulose. PAHs (in particular FLU, 73.9%, ACE, 78.5%, and PYR, 74.2%) and o-PAHs are well associated with biomass combustion (Schnelle-Kreis et al., 2007). Alkanes with carbon chains between C22 and C25 contributed also highly to factor 2 (around 30% for each analyte). In particular, C23 contributed to 60.7% of species. n-alkanes between C23 and C25 are markers for diesel exhaust.

Factor 3 has been associated to biogenic emission and cooking emission and was highly dominated by alkanes with carbon chains ≥ 26 (percentage of species range between 52.7% for C33 and 69.5% for C30) and acids. An n-alkane profile with a carbon chain ≥ C27 carbon atoms indicates inputs arising from higher plant waxes (Yadav et al., 2013). Fatty acids between C10 and C26 indicate terrestrial higher plant waxes (Rogge et al., 1993). The high contribution of adipic acid (89.8%) can be caused by secondary organic aerosol formation, or meat cooking emissions (Rogge et al., 1991).

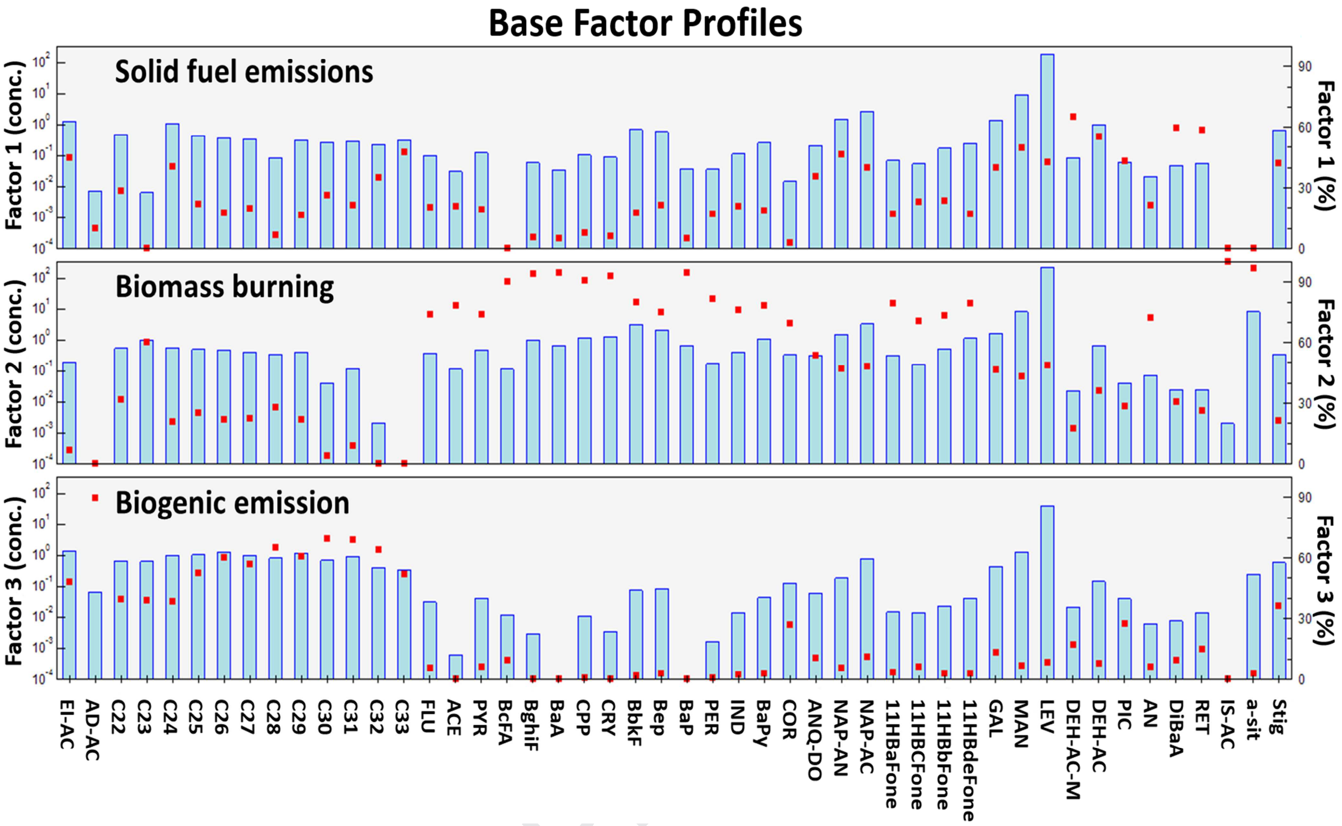

Figure 4. PM1 and PM2.5 PMF factor profile. Solid bars correspond to concentrations and red points correspond to the percentage of species.

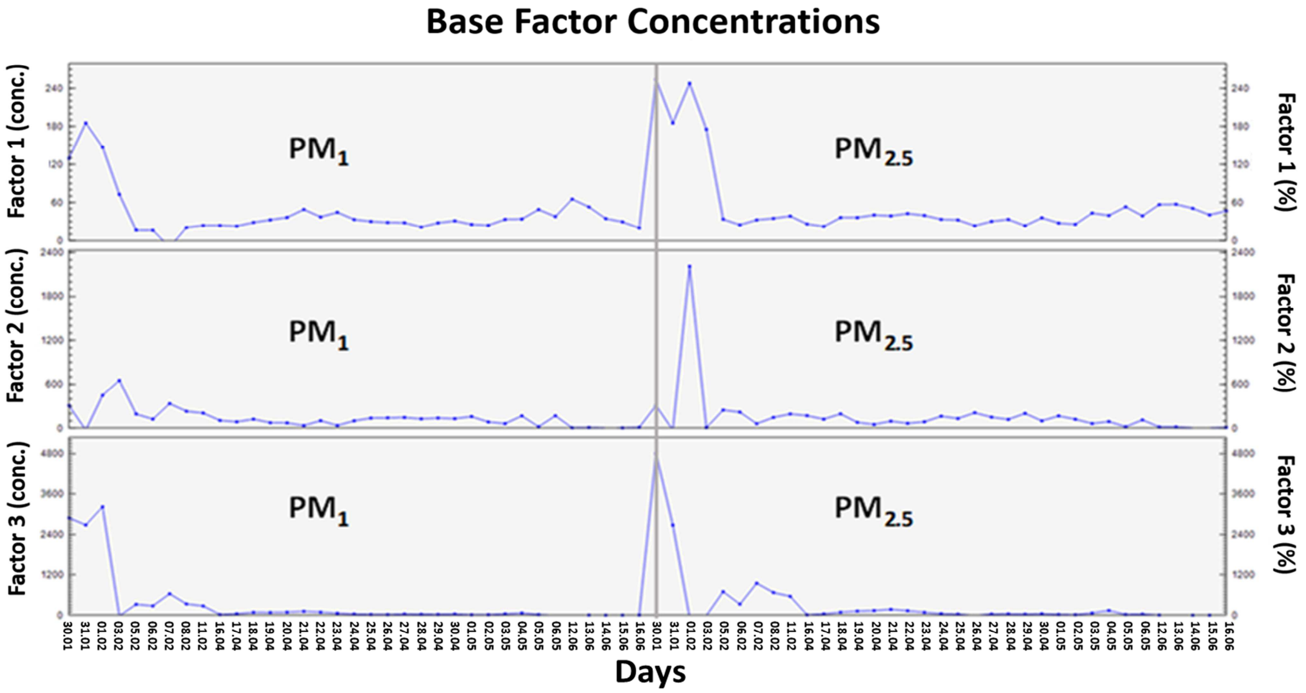

Figure 5. PM_1 and $PM_{2.5}$ PMF factor contribution.

The three-factor solution partially agrees with what found in previous work (Masiol et al., 2020). In that case, a six-factor solution was obtained in a study of $PM_{2.5}$ of several cities in Veneto. Masiol et al. analyzed the inorganic ions of PM2.5, thus the PMF reported factors due to resuspended dust and secondary nitrate and sulfate, which can not be detected by our study. The fifth and sixth of their PMF, however, was assigned to biomass burning and fossil fuel combustion, confirming that these are two important pollution sources in the Veneto region.

Factor contribution (**Figure 5**), illustrates how the three different factors, identified as possible emissions sources of the quantified chemical markers, characterize the different sampling periods. All three different factors show their maximum contribution during the winter period. This confirms what has already been seen with PCA, where most of the variables showed a general higher concentration during the winter period. Factor 2 and factor 3, especially, seem to be higher due to the greater contribution from anthropogenic sources such as heating systems and diesel car emissions. Factor 1, related to the natural emission source, shows an increasing contribution during the spring and summer period (compared to Factor 2 and 3), highlighting the decrease of the impact of anthropogenic emission sources in favor of the growth of natural emission sources. This may be due to the reduced use of home heating, going towards the warm season and the parallel increase in the presence of natural materials due to the spring-summer season.

The results achieved in the present work indicates a strong seasonality of the analyzed pollutants. Winter presents the highest concentrations of most of the species, due, most of all, to fuel emissions and biomass combustion. This confirms what already found in Veneto in previous works, like the one of Khan et al. (2016), which reports a higher concentration of organic carbon in colder months. Masiol et al. (2017) too reported a higher concentration of $PM_{2.5}$ in winter rather than in warmer months. The higher winter concentration of pollutants is generally ascribed to domestic heating (Masiol et al., 2017). The high concentration of organic molecules indicates that the domestic heating is produced not only by methane but also by more polluting sources as, for example, wood. The species that had a higher concentration in warmer seasons (spring and early summer) were AD-AC and the long-chain n-alkanes. PMF indicates a common source that can be due to biogenic emissions. Another possible explanation for AD-AC is that reported by Barbaro et al. (2020): AD-AC, instead, can be mostly produced in warmer seasons due to the higher presence of OH radicals that photo-oxydizes, for example, α-pinene.

4. Conclusion

An extensive study of the organic fraction of particulate matter in a small municipality of the Veneto region, northeastern Italy, was carried out. Three seasons were considered: winter, spring, and early summer. A strong seasonality on the pollution behavior was observed.

Journal Pre-proof

IDTD–GC–TOFMS technique was applied to obtain a large dataset of chemical markers for air pollution. The statistical analyses, Principal Component Analysis (PCA) and Positive Matrix Factorization (PMF) enabled a deeper discussion of the contribution of each class of chemical compound and the identification of the main factors representing different sources. Emission sources discriminating between natural and anthropogenic emissions are found. The temporal trend of these environmental markers shows the main changes during the year. Higher concentrations of most of the chemical compounds are observed during winter than spring and summer, highlighting the greater impact of human and anthropogenic emission sources during the coldest period of the year. During winter, aerosol compositions were mostly characterized by chemical compounds from combustion sources such as short chains of n-alkanes from fossil fuel emissions, and most of the PAHs and o-PAHs, as well as levoglucosan and dehydroabietic acid as biomass burning markers. During spring and early summer, PM_1 and $PM_{2.5}$ composition are explained by the same chemical compounds, but with an important role of acids and longer n-alkanes chains, C29-C33, indicating a more important role of biogenic sources.

Journal Pre-proof

Acknowledgments

This work was carried out thanks to the support of ARPAV, Ca´ Foscari University of Venice and Helmholtz Zentrum Munchen.
O.P. thanks RSF project No. 19773004 financial support for developing the urban pollution methodology.

References

Barbaro, E., Morabito, E., Gregoris, E., Feltracco, M., Gabrieli, J., Vardè, M., Cairns, W.R.L., Dallo, F., De Blasi, F., Zangrando, R., Barbante, C., Gambaro, A., 2020. Col Margherita Observatory: A background site in the Eastern Italian Alps for investigating the chemical composition of atmospheric aerosols. Atmos. Environ. 221. https://doi.org/10.1016/j.atmosenv.2019.117071

Bi, X., Simoneit, B.R.T., Sheng, G., Fu, J., 2008. Characterization of molecular markers in smoke from residential coal combustion in China. Fuel 87, 112–119.

Boström, C.E., Gerde, P., Hanberg, A., Jernström, B., Johansson, C., Kyrklund, T., Rannug, A., Törnqvist, M., Victorin, K., Westerholm, R., 2002. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. Environ. Health Perspect. 110, 451–488.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6, 2812–2831. https://doi.org/10.1039/c3ay41907j

Cass, G.R., 1998. Organic molecular tracers for particulate air pollution sources. TrAC - Trends Anal. Chem. 17, 356–366. https://doi.org/10.1016/S0165-9936(98)00040-5

Chen, Y., Sheng, G., Bi, X., Feng, Y., Mai, B., Fu, J., 2005. Emission factors for carbonaceous particles and polycyclic aromatic hydrocarbons from residential coal combustion in China. Environ. Sci. Technol. 39, 1861–1867.

Chung, S.H., Seinfeld, J.H., 2002. Global distribution and climate forcing of carbonaceous aerosols. J. Geophys. Res. Atmos. 107. https://doi.org/10.1029/2001JD001397

Cruz, C.N., Pandis, S.N., 1997. A study of the ability of pure secondary organic aerosol to act as cloud condensation nuclei. Atmos. Environ. 31, 2205–2214. https://doi.org/10.1016/S1352-2310(97)00054-X

Finocchiaro, E., Richardson, T., 1983. Sterol oxides in foodstuffs: a review. J. Food Prot. 46, 917–925.

Gelencsér, A., May, B., Simpson, D., Sánchez-Ochoa, A., Kasper-Giebl, A., Puxbaum, H., Caseiro, A., Pio, C.A., Legrand, M., 2007. Source apportionment of PM2.5 organic aerosol over Europe: Primary/secondary, natural/anthropogenic, and fossil/biogenic origin. J. Geophys. Res. Atmos. 112, 1–12. https://doi.org/10.1029/2006JD008094

Hauke, J., Kossowski, T., 2011. Comparison of Values of Pearson’s and Spearman’s Correlation Coefficients on the Same Sets of Data. Quaest. Geogr. 30, 87–93. https://doi.org/10.2478/v10117-011-0021-1

Herckes, P., Engling, G., Kreidenweis, S.M., Collett, J.L., 2006. Particle size distributions of organic aerosol constituents during the 2002 yosemite aerosol characterization study. Environ. Sci. Technol. 40, 4554–4562. https://doi.org/10.1021/es0515396

Hoffmann, D., Tilgner, A., Iinuma, Y., Herrmann, H., 2010. Atmospheric stability of levoglucosan: A detailed laboratory and modeling study. Environ. Sci. Technol. 44, 694–699. https://doi.org/10.1021/es902476f

Khan, M.B., Masiol, M., Bruno, C., Pasqualetto, A., Formenton, G.M., Agostinelli, C., Pavoni, B., 2018. Potential sources and meteorological factors affecting PM2.5-bound polycyclic aromatic hydrocarbon levels in six main cities of northeastern Italy: an assessment of the related carcinogenic and mutagenic risks. Environ. Sci. Pollut. Res. 25, 31987–32000. https://doi.org/10.1007/s11356-018-2841-1

Journal Pre-proof

Khan, M.B., Masiol, M., Formenton, G., Di Gilio, A., de Gennaro, G., Agostinelli, C., Pavoni, B., 2016. Carbonaceous PM2.5 and secondary organic aerosol across the Veneto region (NE Italy). Sci. Total Environ. 542, 172–181. https://doi.org/10.1016/j.scitotenv.2015.10.103

Ladji, R., Yassaa, N., Balducci, C., Cecinato, A., Meklati, B.Y., 2009. Annual variation of particulate organic compounds in PM10 in the urban atmosphere of Algiers. Atmos. Res. 92, 258–269.

Leithead, A., Li, S.M., Hoff, R., Cheng, Y., Brook, J., 2006. Levoglucosan and dehydroabietic acid: Evidence of biomass burning impact on aerosols in the Lower Fraser Valley. Atmos. Environ. 40, 2721–2734. https://doi.org/10.1016/j.atmosenv.2005.09.084

Li, X., Jiang, L., Hoa, L.P., Lyu, Y., Xu, T., Yang, X., Iinuma, Y., Chen, J., Herrmann, H., 2016. Size distribution of particle-phase sugar and nitrophenol tracers during severe urban haze episodes in Shanghai. Atmos. Environ. 145, 115–127. https://doi.org/10.1016/j.atmosenv.2016.09.030

Liao, H., Seinfeld, J.H., Adams, P.J., Mickley, L.J., 2004. Global radiative forcing of coupled tropospheric ozone and aerosols in a unified general circulation model. J. Geophys. Res. D Atmos. 109, 1–33. https://doi.org/10.1029/2003JD004456

Manousakas, M., Papaefthymiou, H., Diapouli, E., Migliori, A., Karydas, A.G., Bogdanovic-Radovic, I., Eleftheriadis, K., 2017. Assessment of PM2. 5 sources and their corresponding level of uncertainty in a coastal urban area using EPA PMF 5.0 enhanced diagnostics. Sci. Total Environ. 574, 155–164.

Manta, D.S., Angelone, M., Bellanca, A., Neri, R., Sprovieri, M., 2002. Heavy metals in urban soils: a case study from the city of Palermo (Sicily), Italy. Sci. Total Environ. 300, 229–243.

Mas, S., de Juan, A., Tauler, R., Olivieri, A.C., Escandar, G.M., 2010. Application of chemometric methods to environmental analysis of organic pollutants: A review. Talanta. https://doi.org/10.1016/j.talanta.2009.09.044

Masiol, M., Formenton, G., Pasqualetto, A., Pavoni, B., 2013. Seasonal trends and spatial variations of PM10-bounded polycyclic aromatic hydrocarbons in Veneto Region, Northeast Italy. Atmos. Environ. 79, 811–821. https://doi.org/10.1016/j.atmosenv.2013.07.025

Masiol, M., Squizzato, S., Formenton, G., Harrison, R.M., Agostinelli, C., 2017. Air quality across a European hotspot: Spatial gradients, seasonality, diurnal cycles and trends in the Veneto region, NE Italy. Sci. Total Environ. 576, 210–224. https://doi.org/10.1016/j.scitotenv.2016.10.042

Masiol, M., Squizzato, S., Formenton, G., Khan, M.B., Hopke, P.K., Nenes, A., Pandis, S.N., Tositti, L., Benetello, F., Visin, F., Pavoni, B., 2020. Hybrid multiple-site mass closure and source

Journal Pre-proof

apportionment of PM2.5 and aerosol acidity at major cities in the Po Valley. Sci. Total Environ. 704. https://doi.org/10.1016/j.scitotenv.2019.135287

Nielsen, T., 1996. Traffic contribution of polycyclic aromatic hydrocarbons in the center of a large city. Atmos. Environ. 30, 3481–3490.

Okuda, T., Okamoto, K., Tanaka, S., Shen, Z., Han, Y., Huo, Z., 2010. Measurement and source identification of polycyclic aromatic hydrocarbons (PAHs) in the aerosol in Xi ' an , China , by using automated column chromatography and applying positive matrix factorization (PMF). Sci. Total Environ. 408, 1909–1914. https://doi.org/10.1016/j.scitotenv.2010.01.040

Oliveira, C., Pio, C., Alves, C., Evtyugina, M., Santos, P., Gonçalves, V., Nunes, T., Silvestre, A.J.D., Palmgren, F., Wåhlin, P., 2007. Seasonal distribution of polar organic compounds in the urban atmosphere of two large cities from the North and South of Europe. Atmos. Environ. 41, 5555–5570.

Paatero, P., Tapper, U., 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 111–126. https://doi.org/10.1002/env.3170050203

Pietrogrande, M.C., Abbaszade, G., Schnelle-Kreis, J., Bacco, D., Mercuriali, M., Zimmermann, R., 2011. Seasonal variation and source estimation of organic compounds in urban aerosol of Augsburg, Germany. Environ. Pollut. 159, 1861–1868. https://doi.org/10.1016/j.envpol.2011.03.023

Popovicheva, O., Kistler, M., Kireeva, E., Persiantseva, N., Timofeev, M., Kopeikin, V., Kasper-Giebl, A., 2014. Physicochemical characterization of smoke aerosol during large-scale wildfires: Extreme event of August 2010 in Moscow. Atmos. Environ. 96, 405–414.

Popovicheva, O., Padoan, S., 2019. Spring aerosol in urban atmosphere of megacity: analytical and statistical assessment for source impacts. Aerosol Air Qual. Res.

Popovicheva, O.B., Engling, G., Diapouli, E., Saraga, D., Persiantseva, N.M., Timofeev, M., Kireeva, E.D., Shonija, N.K., Chen, S.-H., Nguyen, D.L., 2016. Impact of smoke intensity on size-resolved aerosol composition and microstructure during the biomass burning season in Northwest Vietnam. Aerosol Air Qual. Res 16, 2635–2654.

Press, I., Press, I., 2013. The Proof and Measurement of Association between Two Things Author (s): C . Spearman Source : The American Journal of Psychology , Vol . 15 , No . 1 (Jan ., 1904), pp . 72-101 15, 72–101.

Journal Pre-proof

Qadir, R.M., Abbaszade, G., Schnelle-Kreis, J., Chow, J.C., Zimmermann, R., 2013. Concentrations and source contributions of particulate organic matter before and after implementation of a low emission zone in Munich, Germany. Environ. Pollut. 175, 158–167. https://doi.org/10.1016/j.envpol.2013.01.002

Rissanen, T., Hyötyläinen, T., Kallio, M., Kronholm, J., Kulmala, M., Riekkola, M.-L., 2006. Characterization of organic compounds in aerosol particles from a coniferous forest by GC–MS. Chemosphere 64, 1185–1195.

Rogge, W.F., Cass, G.R., Hlldemann, L.M., Mazurek, M.A., Slmoneit, B.R.T., 1991. Sources of Fine Organic Aerosol. 1. Charbroilers and Meat Cooking Operations. Environ. Sci. Technol. 25, 1112–1125. https://doi.org/10.1021/es00018a015

Rogge, W.F., Mazurek, M.A., Hildemann, L.M., Cass, G.R., Simoneit, B.R.T., 1993. Quantification of urban organic aerosols at a molecular level: identification, abundance and seasonal variation. Atmos. Environ. Part A. Gen. Top. 27, 1309–1330.

Scerri, M.M., Genga, A., Iacobellis, S., Delmaire, G., Giove, A., Siciliano, M., Siciliano, T., Weinbruch, S., 2019. Investigating the plausibility of a PMF source apportionment solution derived using a small dataset: A case study from a receptor in a rural site in Apulia-South East Italy. Chemosphere 236, 124376.

Schnelle-Kreis, J., Sklorz, M., Orasche, J., Stölzel, M., Peters, A., Zimmermann, R., 2007. Semi Volatile Organic Compounds in Ambient PM2.5. Seasonal Trends and Daily Resolved Source Contributions. Environ. Sci. Technol. 41, 3821–3828. https://doi.org/10.1021/es060666e

Schnelle-Kreis, J., Sklorz, M., Peters, A., Cyrys, J., Zimmermann, R., 2005. Analysis of particle-associated semi-volatile aromatic and aliphatic hydrocarbons in urban particulate matter on a daily basis. Atmos. Environ. 39, 7702–7714.

Seinfeld, J.H., Pandis, S.N., 1998. Chemistry of the atmospheric aqueous phase. Atmos. Chem. Phys. From Air Pollut. to Clim. Chang. 337407.

Sims, R.J., Fioriti, J.A., Kanuk, M.J., 1972. Sterol additives as polymerization inhibitors for frying oils. J. Am. Oil Chem. Soc. 49, 298–301.

Stahle, L., Wold, S., 1990. Multivariate analysis of variance (MANOVA). Chemom. Intell. Lab. Syst. 9, 127–141.

Tositti, L., Brattich, E., Masiol, M., Baldacci, D., Ceccato, D., Parmeggiani, S., Stracquadanio, M., Zappoli, S., 2014. Source apportionment of particulate matter in a large city of southeastern Po

Journal Pre-proof

Valley (Bologna, Italy). Environ. Sci. Pollut. Res. 21, 872–890.

Valotto, G., Rampazzo, G., Gonella, F., Formenton, G., Ficotto, S., Giraldo, G., 2017. Source apportionment of PAHs and n-alkanes bound to PM1 collected near the Venice highway. J. Environ. Sci. (China) 54, 77–89. https://doi.org/10.1016/j.jes.2016.05.025

Yadav, S., Tandon, A., Attri, A.K., 2013. Monthly and seasonal variations in aerosol associated n-alkane profiles in relation to meteorological parameters in New Delhi, India. Aerosol Air Qual. Res. 13, 287–300.

Journal Pre-proof

1. Extensive analysis of the Particulate Matter in a hot spot for air quality in Europe, the Po Valley;
2. Chemical and environmental characterization of Particulate Matter;
3. Complementary use of two statistical techniques: PCA and PMF.

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof