Supplement materials to:

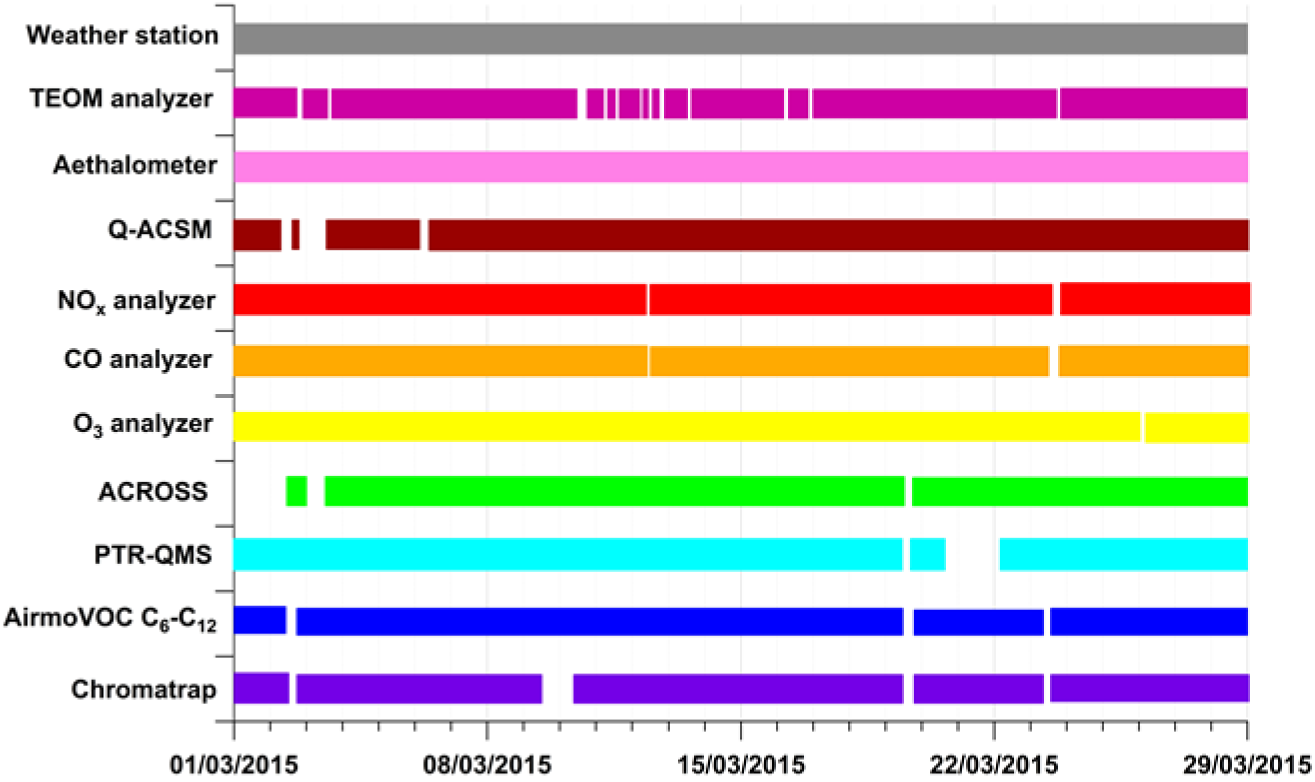
Origin and variability of volatile organic compounds observed at an Eastern Mediterranean background site (Cyprus)

Cécile Debevec.[1,2], Stéphane Sauvage[1], Valérie Gros.[2], Jean Sciare[3,2], Michael Pikridas[3], Iasonas Stavroulas[3], Thérèse Salameh[1], Thierry Leonardis[1], Vincent Gaudion[1], Laurence Depelchin[1], Isabelle Fronval[1], Roland Sarda-Esteve[2], Dominique Baisnée[2], Bernard Bonsang[2], Chrysanthos Savvides[4], Mihalis Vrekoussis[3,5,6], Nadine Locoge[1].

[1]Département SAGE, IMT Lille Douai, Univ. Lille, Lille, 59000, France
[2] Equipe CAE, Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Unité Mixte CEA-CNRS-UVSQ, Gif sur Yvette, 91190, France
[3] Energy, Environment and Water Research Centre, the Cyprus Institute (CyI), Nicosia, 2121, Cyprus
[4] Department of Labour Inspection (DLI), Ministry of Labour, Welfare and Social Insurance, Nicosia, 2121, Cyprus
[5] Institute of Environmental Physics (IUP), University of Bremen, Bremen, 28359, Germany
[6] Center of Marine Environmental Sciences (MARUM), University of Bremen, Bremen, 28359, Germany

Correspondence to: Stéphane Sauvage (stephane.sauvage@imt-lille-douai.fr) – Cecile Debevec (cecile.debevec@imt-lille-douai.fr)

SI-1 Instrument status indicating the period when each gas, aerosol and meteorological instrument was operating



SI-2 Data processing and analysis of the VOCs dataset used as EPA PMF inputs

SI-2.1 Data preparation

The PMF analysis does not support zero and negative values or missing data. Instead, data gaps and concentrations below detection limits have been substituted following an approach developed by Hopke, 2000:

$$x_{ij} = \begin{cases} x_{ij},\ if\ x_{ij} \geq DL_i \\ \frac{DL_i}{2},\ if\ x_{ij} < DL_i \\ \overline{x}_i,\ if\ missing\ x_{ij} \\ \sum_{i=1}^{n} x_{ij}\ ,\ for\ x_{TVOC,j} \end{cases}, \qquad (1)$$

with DL_i and $\overline{x}_i$ respectively the analytical detection limit and geometric mean concentration of the i$^{\mathrm{th}}$ corresponding species. $TVOC$ was defined as total variable and referred to the sum of the n species including in the initial dataset. The proportion of replaced values was estimated to be less than 1 % (especially for compounds with high background mixing ratios) up to 20 %, except for pinenes and 2-methylpentane (26 % and 48 %, respectively) due to their of concentrations often below the detection limit. Additionally, to avoid a too high influence of replaced points, they were downweighted, increasing associated uncertainty values (Polissar et al., 1998; Reff et al., 2007) as:

$$s_{ij} = \begin{cases} 2 \times u(x_{ij}) + \frac{DL_i}{3},\ if\ x_{ij} \geq DL_i \\ 5\ \times \frac{DL_i}{6},\ if\ x_{ij} < DL_i \\ 4 \times\ \overline{x}_i,\ if\ missing\ x_{ij} \\ \sqrt{\sum_{i=1}^{n} {s_{ij}}^2}\ ,\ for\ s_{TVOC,j} \end{cases}, \qquad (2)$$

with $u(x_{ij})$ the absolute analytical uncertainty on the i$^{\mathrm{th}}$ species concentrations in the j$^{\mathrm{th}}$ sample. The method to estimate the uncertainties $u(x_{ij})$ of each sample for each species has been described in Sect. 2.2.1.

SI-2.2 Quality of the VOCs dataset

Besides the amount of replaced values, a compound is characterized by its signal-to-noise (S/N) ratio, which is used as an indicator of the quality of the observations (Paatero and Hopke, 2003). In EPA PMF 5.0, S/N ratios are determined from Eq. 3 (Norris et al., 2014):

$$\left(\frac{S}{N}\right)_i = \frac{1}{m}\sum_{j=1}^{m} d_{ij} \quad with\ \ d_{ij} = \begin{cases} \left(\frac{x_{ij}-s_{ij}}{s_{ij}}\right), & if\ x_{ij} > s_{ij} \\ 0, & if\ x_{ij} < s_{ij} \end{cases}, \qquad (3)$$

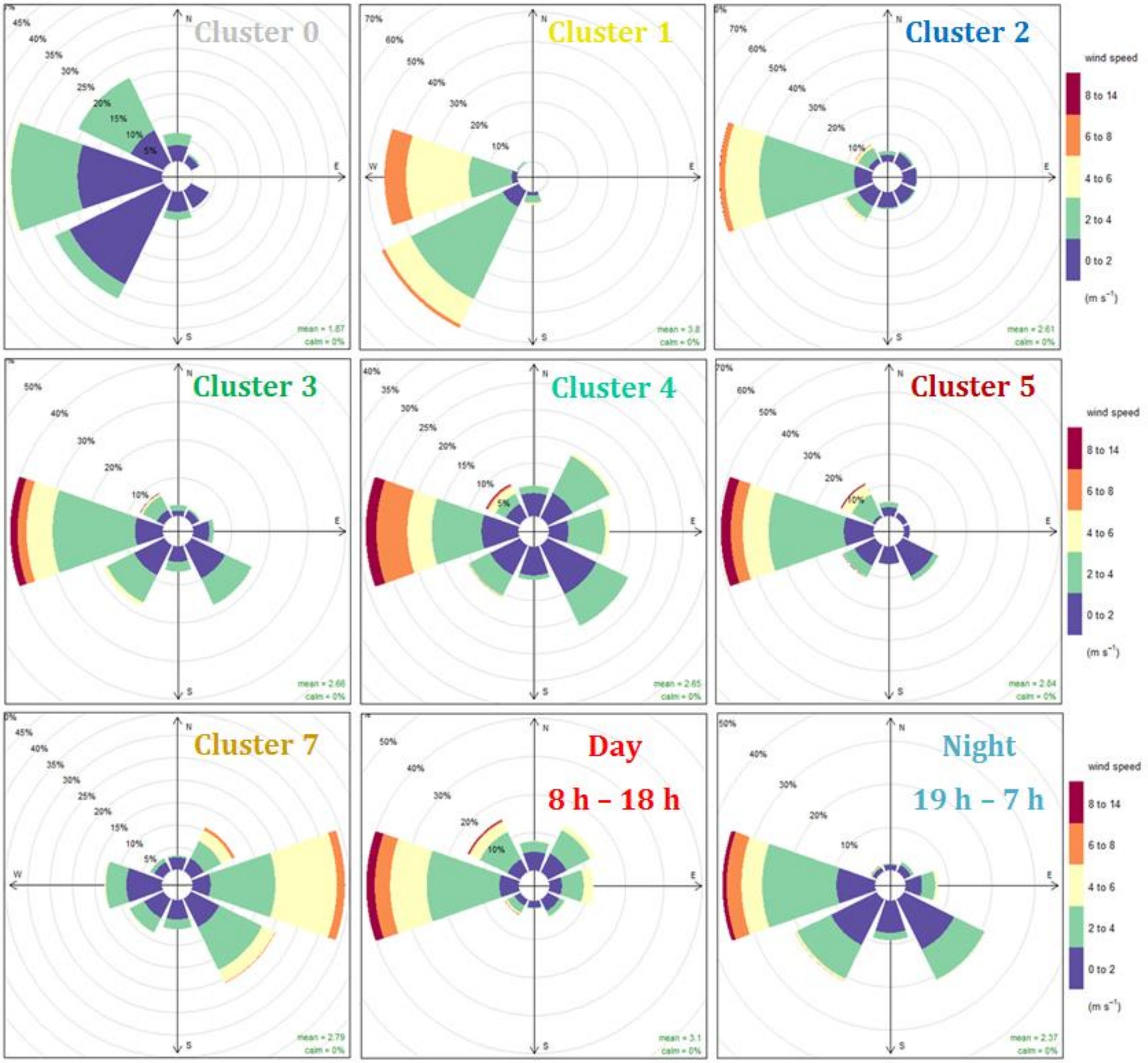
where d_{ij} is the relative difference between species concentration and the corresponding uncertainty in the sample considered. Indeed, only concentration values that exceed their associated uncertainty contribute to the signal portion of the S/N calculation. According to Norris et al., 2014, S/N ratio of 1 corresponds to species with observations twice higher than uncertainties on average and consequently indicates a species with good signal quality. In this study, the data quality is qualified as:

$$\begin{cases} \left(\frac{S}{N}\right)_i \geq 1.2, & strong\ variable\ (stayed\ unchanged) \\ 1 < \left(\frac{S}{N}\right)_i < 1.2, & weak\ variable\ (downweighted) \\ \left(\frac{S}{N}\right)_i \leq 1, & bad\ variable\ (excluded) \end{cases} , \tag{4}$$

Indeed, to minimize the weight of contributions of low quality species, PMF allows declaring these species as "weak" (Paatero and Hopke, 2003) and hence tripling their original uncertainties. Two species have been downweighted here (2-methylpentane and m107) due to analytical issues. No optional extra modeling uncertainty was applied here.

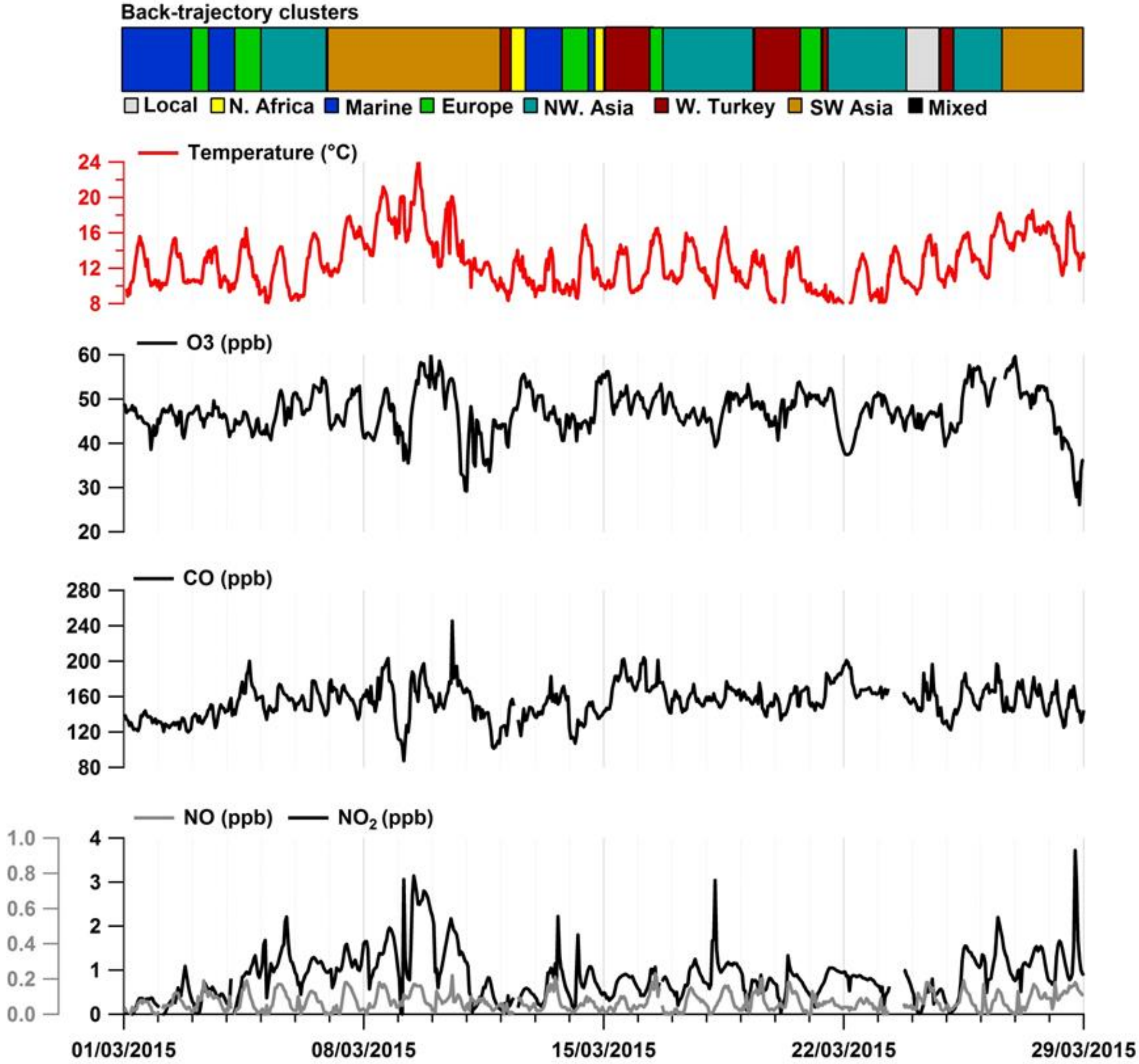
SI-3 Wind roses in function of air mass origins and time periods.

Contribution expressed in % corresponds to the frequency of occurrence in a wind direction. Day and night wind roses don’t include data associated to C7. Time is given as local time. C0 – Local; C1 – N. Africa; C2 – marine air masses; C3 – Europe; C4 – NW. Asia; C5 – W. of Turkey; C7 – SW. Asia

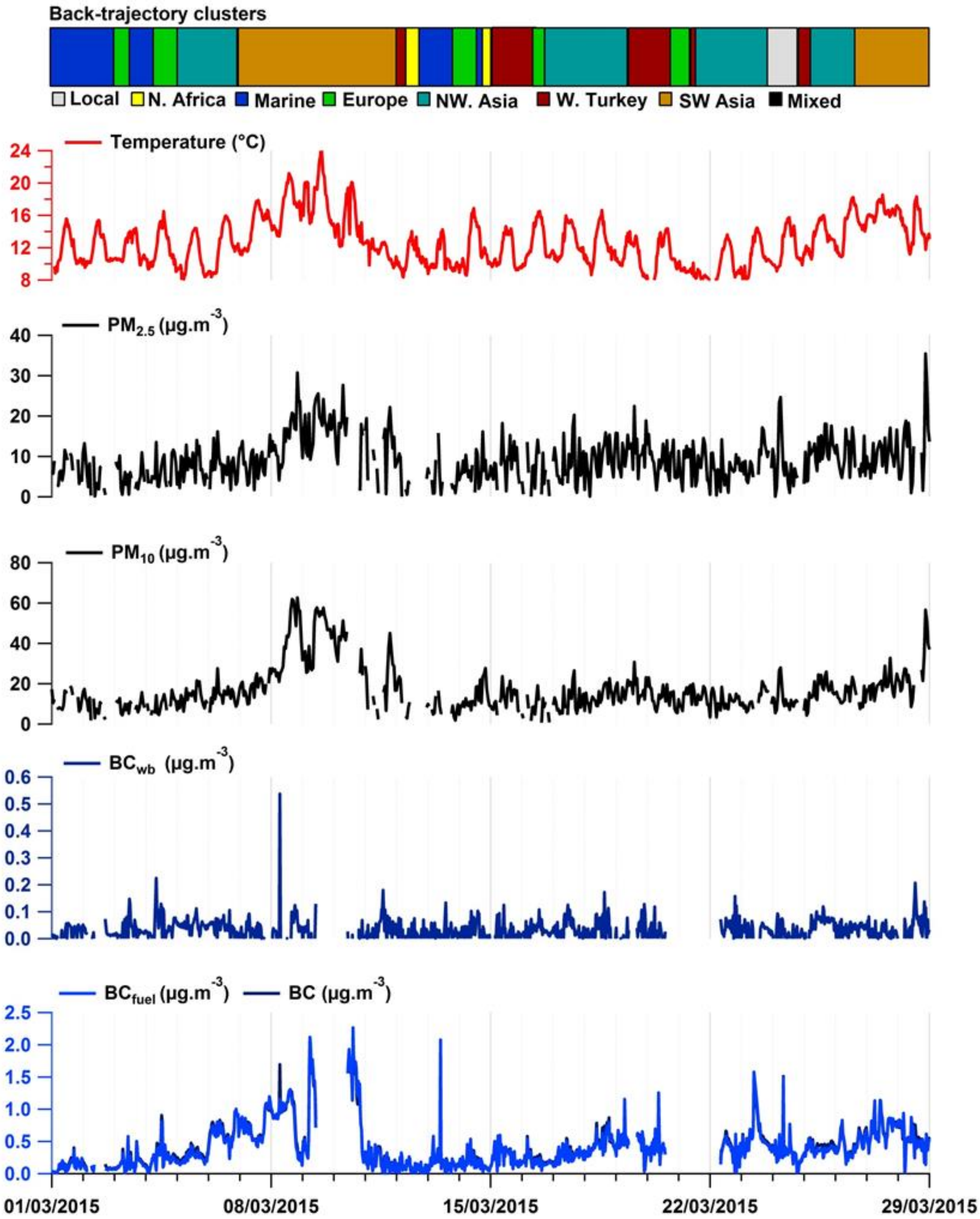


SI-4 Gas and aerosol tracers

SI-4.1Time serie of gas tracers (Ozone, CO, NO and NO_2 – black and grey colors) in relation to air mass origins and temperature.



SI-4.2Time of aerosol tracers ($PM_{2.5}$, PM_{10}, BC, BC fuel and BC wood burning – blue and black colors) in relation to air mass origins and temperature.

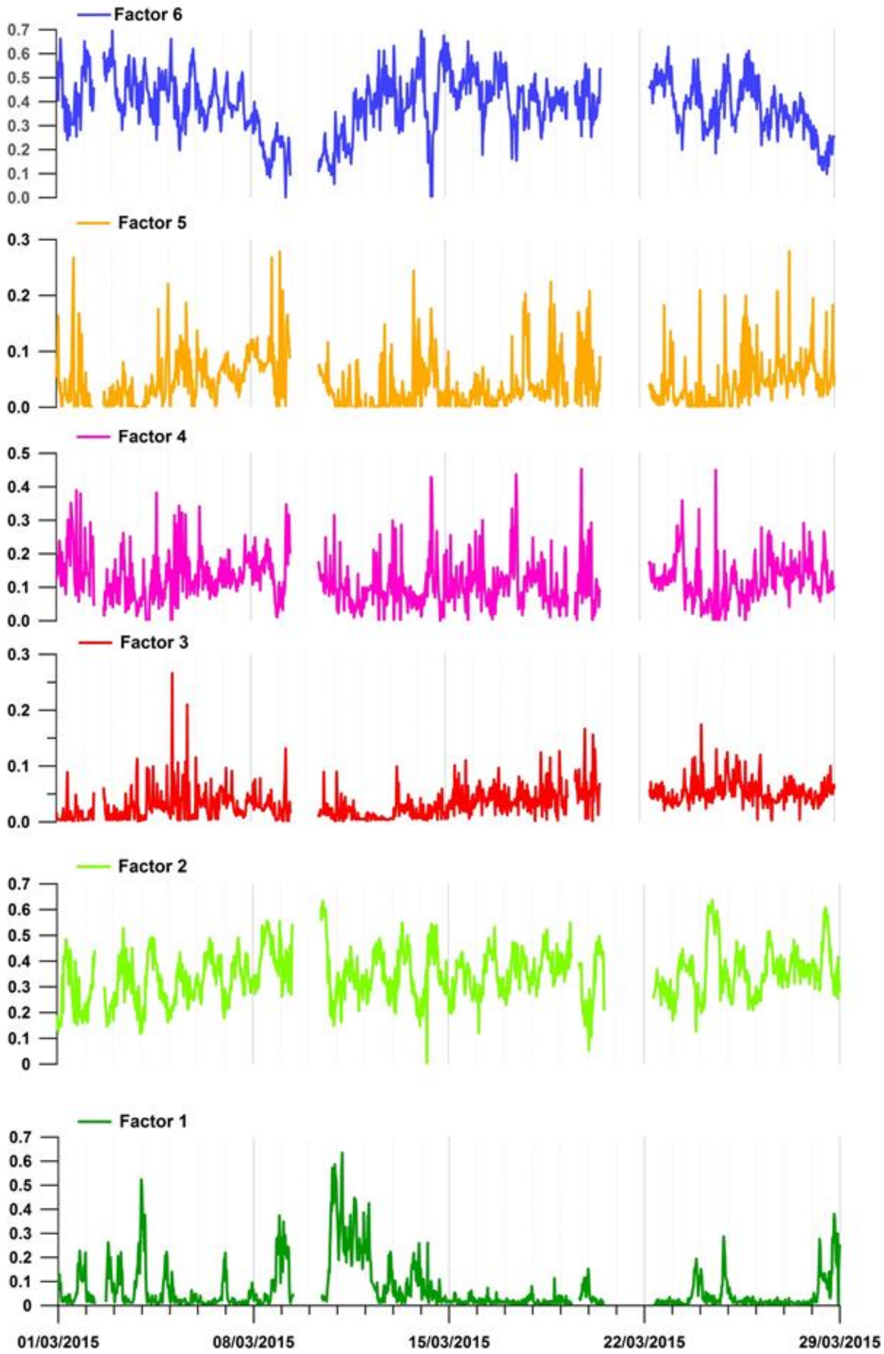


SI-5 Statistics (µg.m^{-3}), detection limits (DL - µg.m^{-}3) and relative uncertainties u(X)/X (Unc. - %) of selected VOC concentrations measured at the site.

	Species	Min	25 %	50 %	Mean	75 %	Max	σ	DL	Unc.
ALKANES	**Ethane**	1.41	2.62	2.92	3.05	3.38	6.80	0.66	0.13	6
	Propane	0.10	1.89	2.04	2.20	2.39	6.24	0.62	0.19	14
	i-Butane	0.03	0.16	0.27	0.32	0.41	1.83	0.23	0.06	8
	n-Butane	0.02	0.29	0.44	0.54	0.67	6.77	0.43	0.04	19
	i-Pentane	0.02	0.06	0.15	0.25	0.32	3.98	0.31	0.03	19
	n-Pentane	0.01	0.09	0.15	0.21	0.28	2.14	0.21	0.02	24
	2-Methylpentane	0.03	0.03	0.05	0.09	0.11	2.27	0.13	0.05	9
ALKENES	**Ethylene**	0.01	0.14	0.32	0.35	0.49	2.85	0.27	0.06	12
	Propene	0.02	0.11	0.17	0.19	0.25	1.28	0.13	0.05	31
ALKYNE	**Acetylene**	0.17	0.57	0.70	0.71	0.83	2.32	0.21	0.07	15
DIENE	**Isoprene**	0.01	0.08	0.11	0.13	0.16	0.62	0.09	0.02	11
TERPENES	**α-Pinene**	0.04	0.04	0.10	0.33	0.33	10.63	0.74	0.09	10
	β-Pinene	0.03	0.03	0.10	0.34	0.32	11.14	0.80	0.07	12
AROMATICS	**Benzene**	0.03	0.29	0.37	0.37	0.44	0.94	0.12	0.06	6
	Toluene	0.03	0.07	0.15	0.19	0.24	3.35	0.22	0.06	9
	Ethylbenzene	0.02	0.02	0.02	0.04	0.02	1.76	0.09	0.04	11
	m,p-Xylenes	0.02	0.02	0.02	0.06	0.06	3.82	0.16	0.04	12
	o-Xylene	0.02	0.02	0.02	0.03	0.02	1.77	0.08	0.03	19
ALCOHOL	**Methanol**	0.87	2.23	3.29	3.84	4.73	12.82	2.13	0.08	21
CARBONYL COMPOUNDS	**Acetaldehyde**	0.19	0.52	0.74	0.83	1.03	2.81	0.44	0.03	10
	Acetone	1.02	2.08	2.57	2.72	3.03	6.46	0.96	0.01	9
	MVK+MACR	0.01	0.06	0.08	0.09	0.11	0.41	0.06	0.01	12
	MEK	0.18	0.47	0.59	0.66	0.74	2.11	0.30	0.01	9
NITRILE	**Acetonitrile**	0.06	0.13	0.17	0.20	0.21	1.97	0.16	0.01	13

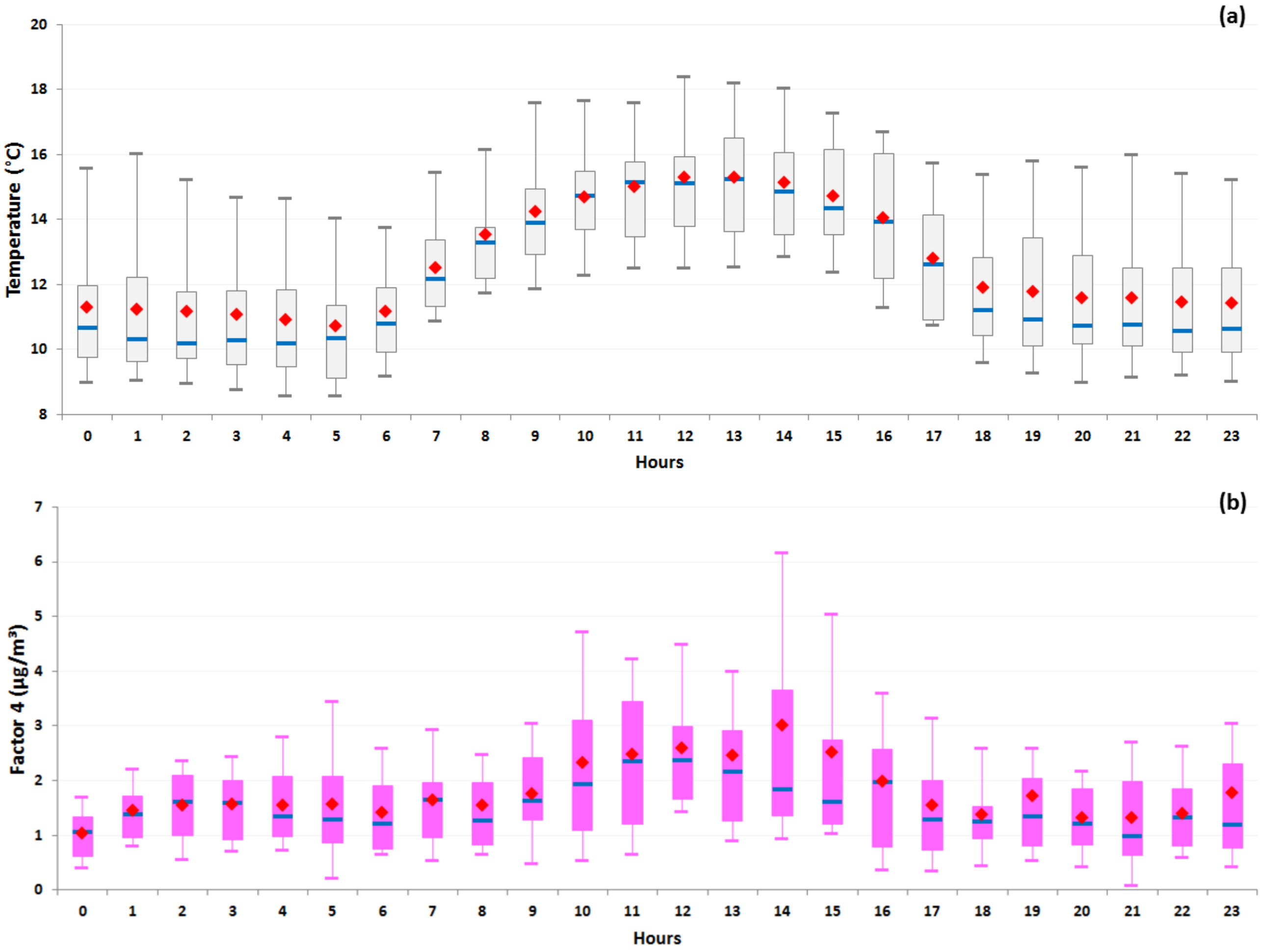
SI-6 Time series of PMF factor relative contribution.

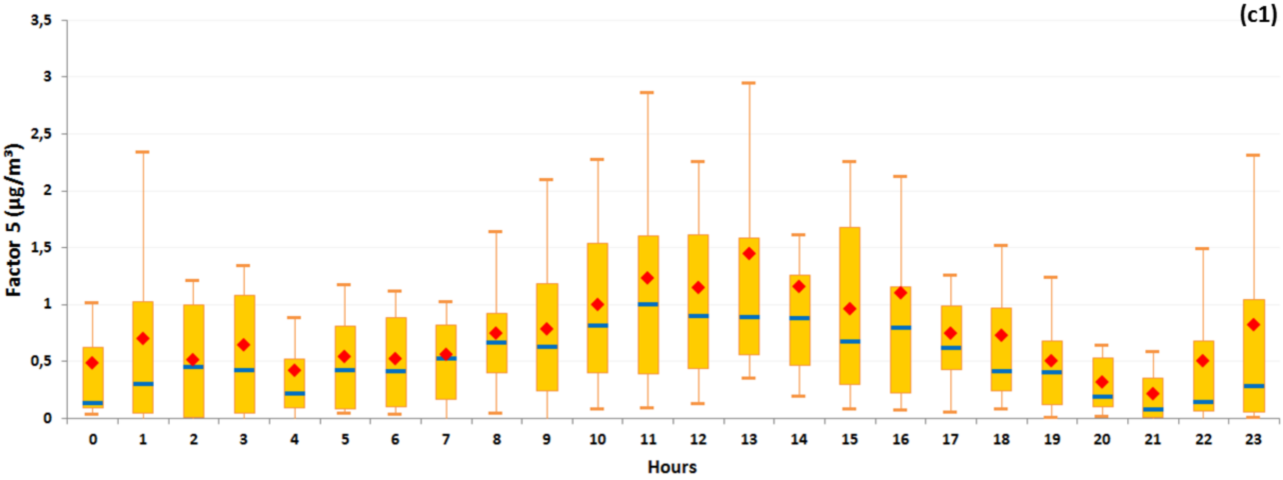
Factor 1 - biogenic source 1; factor 2 - biogenic source 2; factor 3 – short-lived combustion source; factor 4 – evaporative sources; factor 5 – industrial and evaporative sources; factor 6 – regional background.

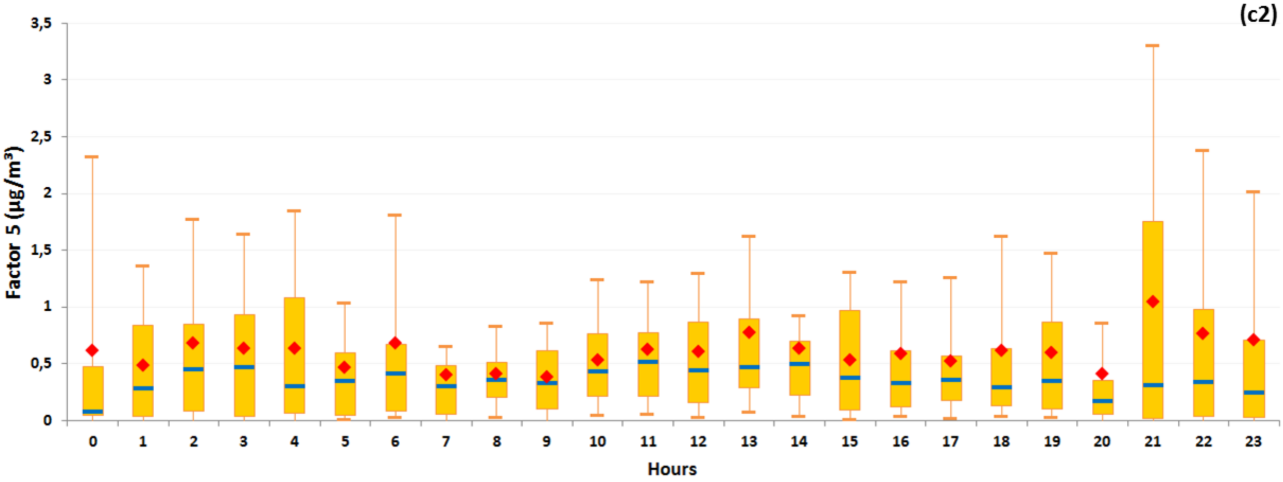


SI-7 Diel variation of the factor contributions (b and c) and temperature (a) represented by hourly box plots.

Factor 5 results are represented in two figures in function of wind direction sectors: c1 only with South and Southeast directions and c2 with the others. Diel profiles don't include contributions obtained when the site was under the influence of air masses categorized in "Southwest Asia" cluster (see section 3.5.2.). Blue solid line represents the median contribution, the red marker represents the mean contribution and the box shows the InterQuartile Range (IQR). The bottom and the top of box depict the first and the third quartiles (i. e. Q1 and Q3). The ends of the whiskers correspond to first and the ninth deciles (i. e. D1 and D9). Time is given as local time.

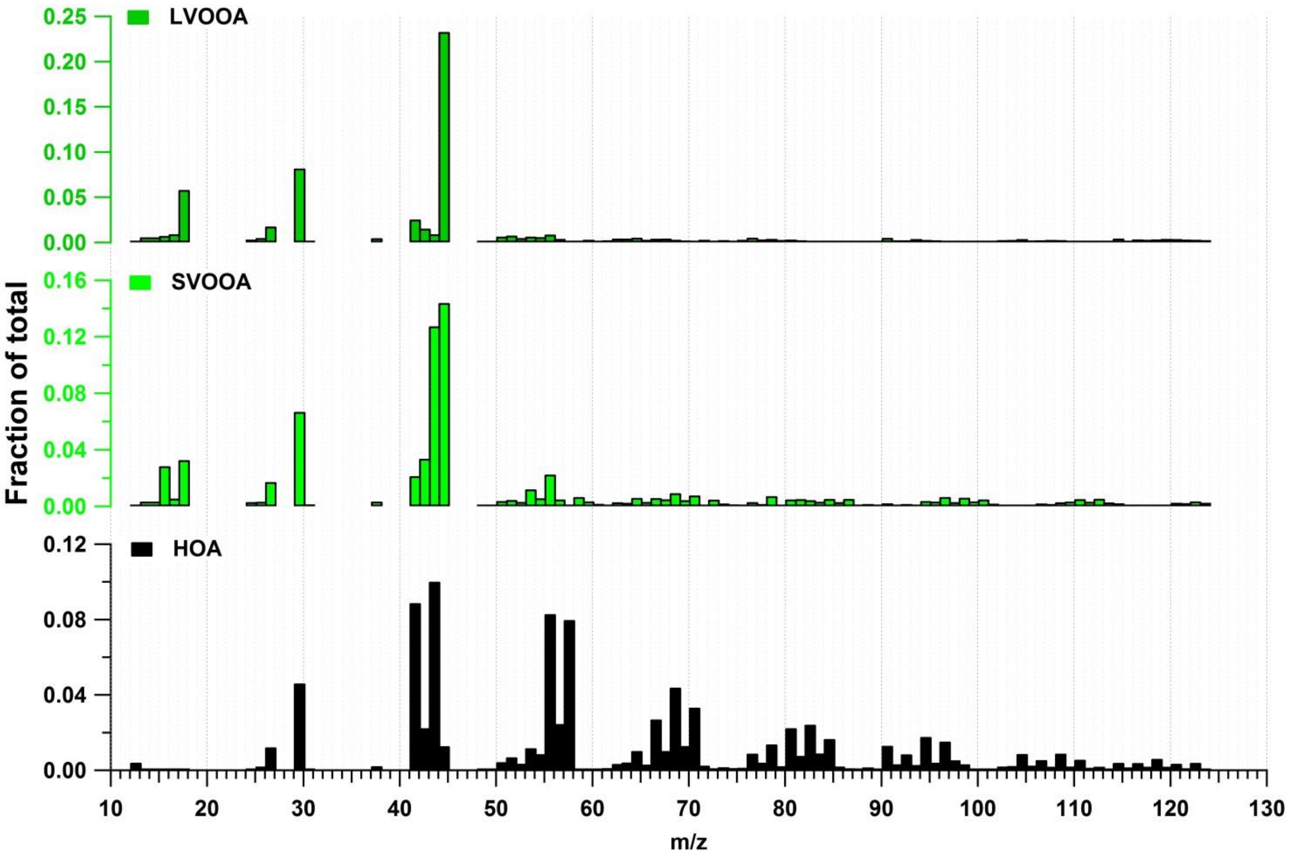



(c1)
Factor 5 (µg/m³)
3,5
3
2,5
2
1,5
1
0,5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hours


(c2)
Factor 5 (µg/m³)
3,5
3
2,5
2
1,5
1
0,5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hours

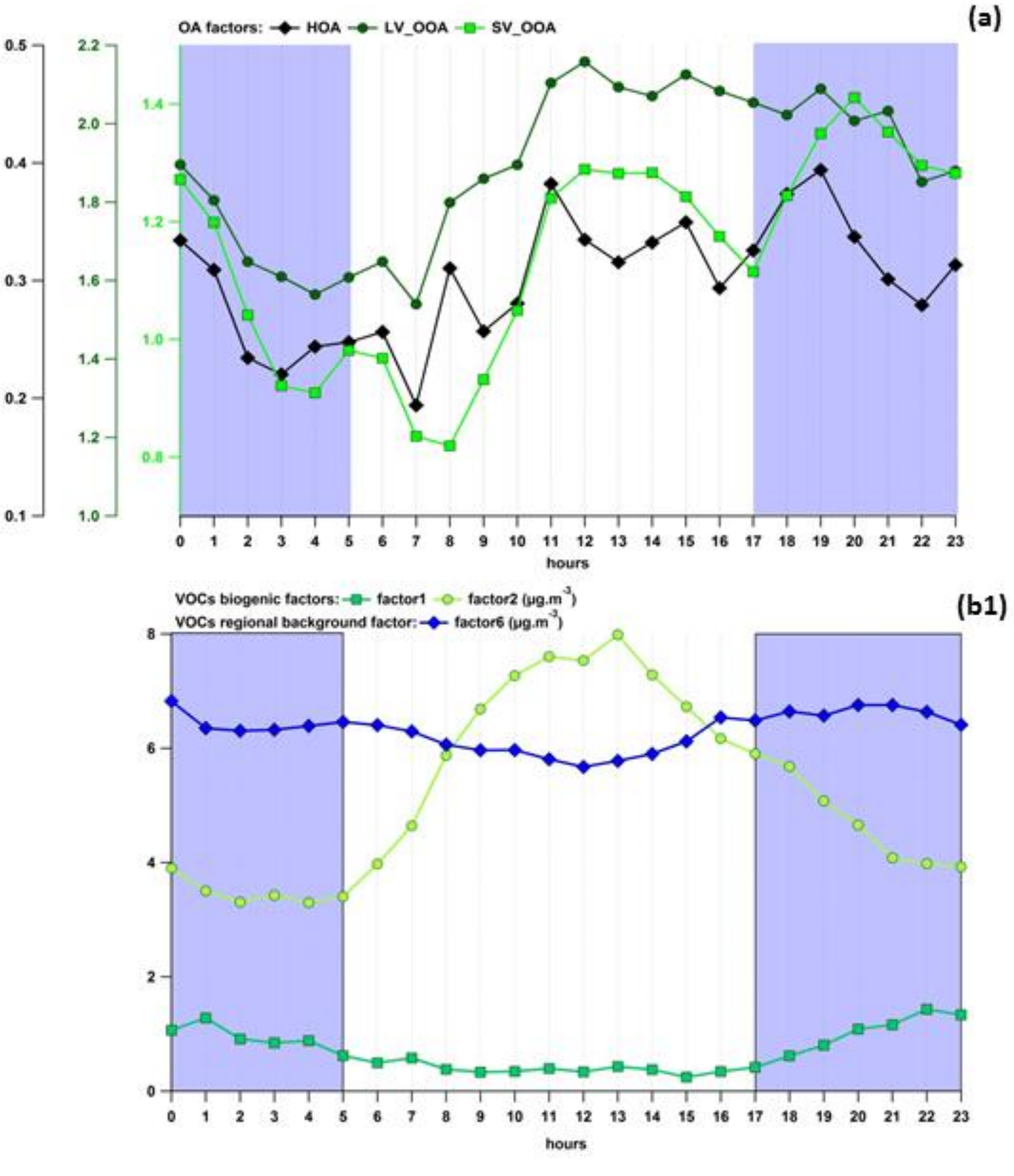
SI-8 Mass spectra profile obtained for the 3 factor constrained PMF solution.

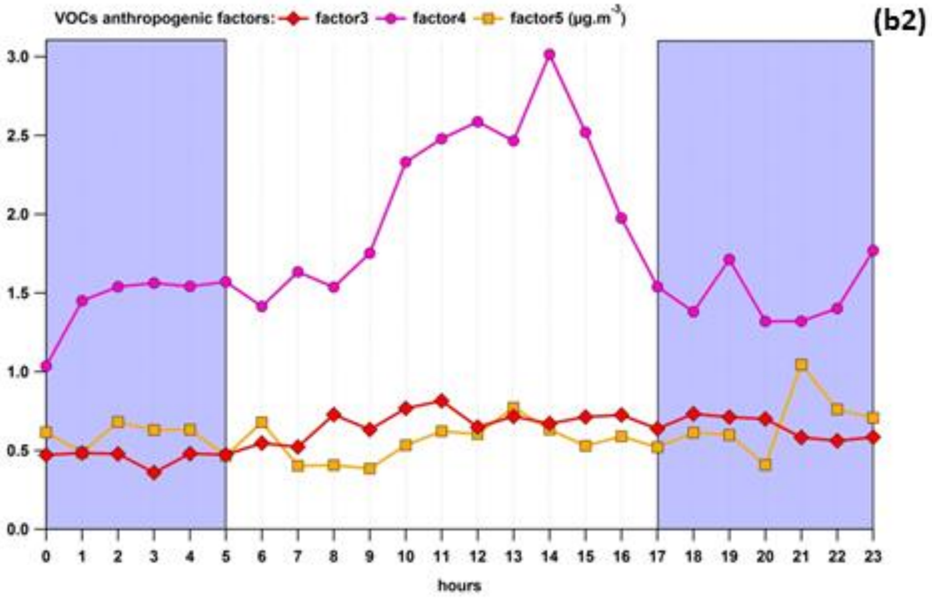
OA factors: HOA - hydrogen-like OA; SV-OOA – semi-volatile oxygen-like OA; LV-OOA – low-volatile oxygen-like OA.



SI-9 Diel variation of the OA and VOCs factor contributions represented by average contributions.

Time is given as local time. OA factors (a): HOA - hydrogen-like OA; SV-OOA – semi-volatile oxygen-like OA; LV-OOA – low-volatile oxygen-like OA. VOCs factors (b1 and b2): Factor 1 - biogenic source 1; factor 2 - biogenic source 2; factor 3 – short-lived combustion source; factor 4 – evaporative sources; factor 5 – industrial and evaporative sources; factor 6 – regional background.




VOCs anthropogenic factors: factor3 factor4 factor5 (µg.m^{-3})
(b2)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
hours