Supplement of Atmos. Chem. Phys., 15, 1221–1236, 2015
http://www.atmos-chem-phys.net/15/1221/2015/
doi:10.5194/acp-15-1221-2015-supplement

Atmospheric Chemistry and Physics
Open Access

Supplement of

Can positive matrix factorization help to understand patterns of organic trace gases at the continental Global Atmosphere Watch site Hohenpeissenberg?

M. Leuchner et al.

Correspondence to: M. Leuchner (leuchner@wzw.tum.de)

Supplement

S1. Comparison to alternative factor solutions

In Fig. S1 the additional factors that were resolved in the seven and eight factor solutions and one different factor for the five factor solution are shown. In general, all factors of the six factor solution (Fig. 3) could also be found in these alternative computations. For the five factor computation considerable differences to the solution above could be found in Factors II, IV, and V. Factor II also contained large amounts of ethane, smaller amounts of acetylene, benzene, and toluene. The longer-lived Factors IV and V were combined to one factor. The seven and eight factor solutions provided an additional common factor that contained more than one third of the modeled isobutene and considerable amounts of 2-methylpentane, 3-methylpentane, benzene, toluene and some other substances. Thus, the amounts of these substances in other factors, especially Factor III, decreased accordingly, compared to the six factor computation. Ethane in Factor V only occurred for the six factor solution; it was distributed in other factors, e.g. Factor VI, for the seven and eight factor solutions. Isopentane and *n*-pentane accounted for the second additional factor in the eight factor solution, thus, their contribution to Factor III was much lower than in the six factor solution.

The five factor solution combined the two long-lived Factors IV and V from the six factor solution that should appear separately when compared e.g. to Sauvage et al. (2009) as well as to the source apportionment in section 3.2. The seven and eight factor solutions showed an additional factor with comparatively small contributions of many substances that could not be apportioned to proper source categories. Since the statistical values of the PMF analysis did not show a clear indication for the ideal number of factors, the decision was made to choose six factors as the most reasonable and explainable solution that PMF calculated.

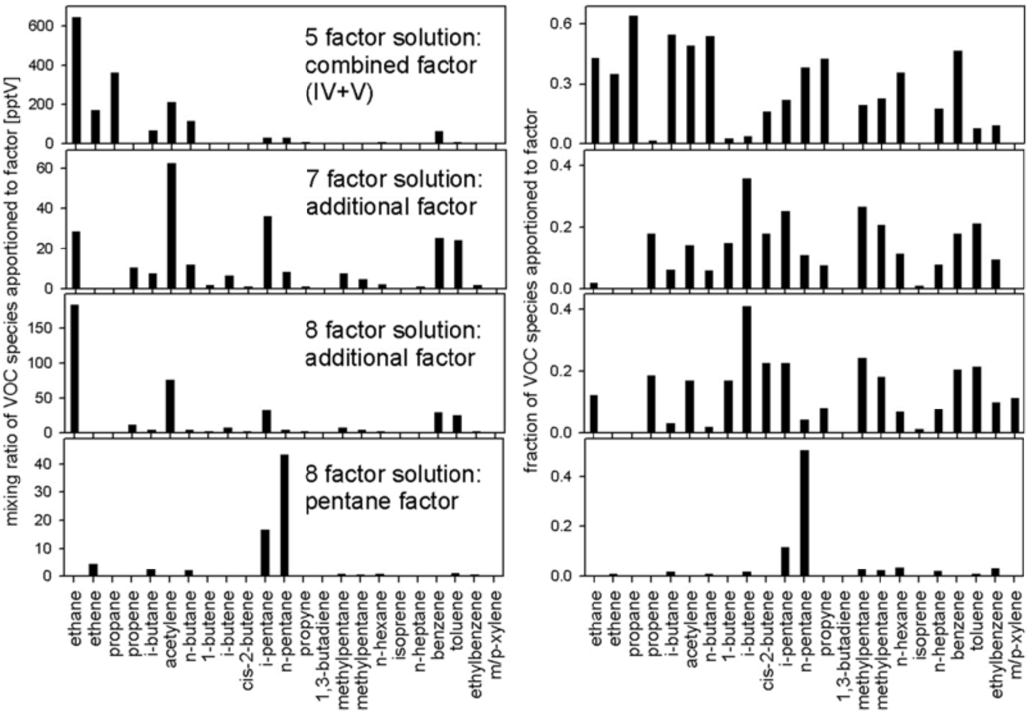

Figure S1. Factor profiles of alternative solutions calculated by PMF, left: mixing ratio of each species apportioned to each factor [pptv], right: contribution of each factor to the species.

S2. Resolved factors including nighttime data

Figure S2 shows factor profiles for the six factor solution calculated by PMF for all measured data (daytime and nighttime). The same six factors (compared to Fig. 3) were resolved and only slight differences to the daytime data solution were found which supports the stability of the PMF solution in extracting profiles reliably also at this remote site. One main difference could be seen for the biogenic factor. The relative contributions remained basically identical, but the absolute amount was lower by a factor of around 2. This can be explained by very low to zero emissions of isoprene, the dominant compound, at night. The five anthropogenic factors were also basically identical. Only small additional abundances of ethane and propane as well as some higher contribution of the dominating ethene in the short-lived combustion factor could be observed.

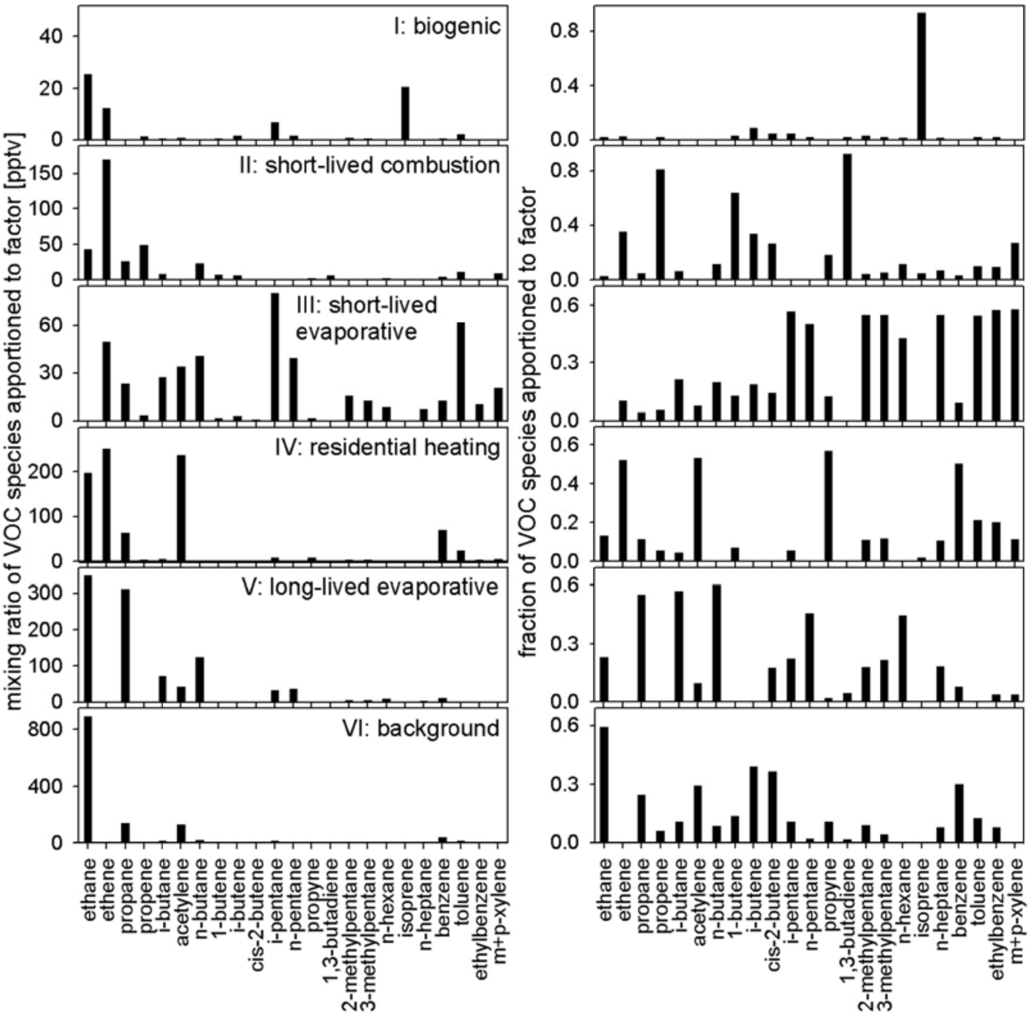

Figure S2. Factor profiles of the six factor solution calculated by PMF for all data (daytime and nighttime), left: mixing ratio of each species apportioned to each factor [pptv], right: contribution of each factor to the species. Note that the scales for each subplot are different due to large variations in absolute mixing ratios.

S3. Resolved factors from nighttime data only

In Fig. S3 the PMF results from exclusively nighttime data are plotted. Only five reasonable factors were extracted, because the biogenic factor disappeared due to low or no isoprene contributions at nighttime. However, all five anthropogenic factors were resolved and exhibited quite similar profiles as for daytime (Fig. 3) or daytime and nighttime (Fig. S2) data, respectively. Small differences could be found in the short-lived combustion factor with higher contribution of ethene and some small additional contribution of propane. The residential heating/long-lived combustion factor showed higher values of ethane and acetylene and slightly lower values of ethene. Ethane was higher in the long-lived evaporative factor.

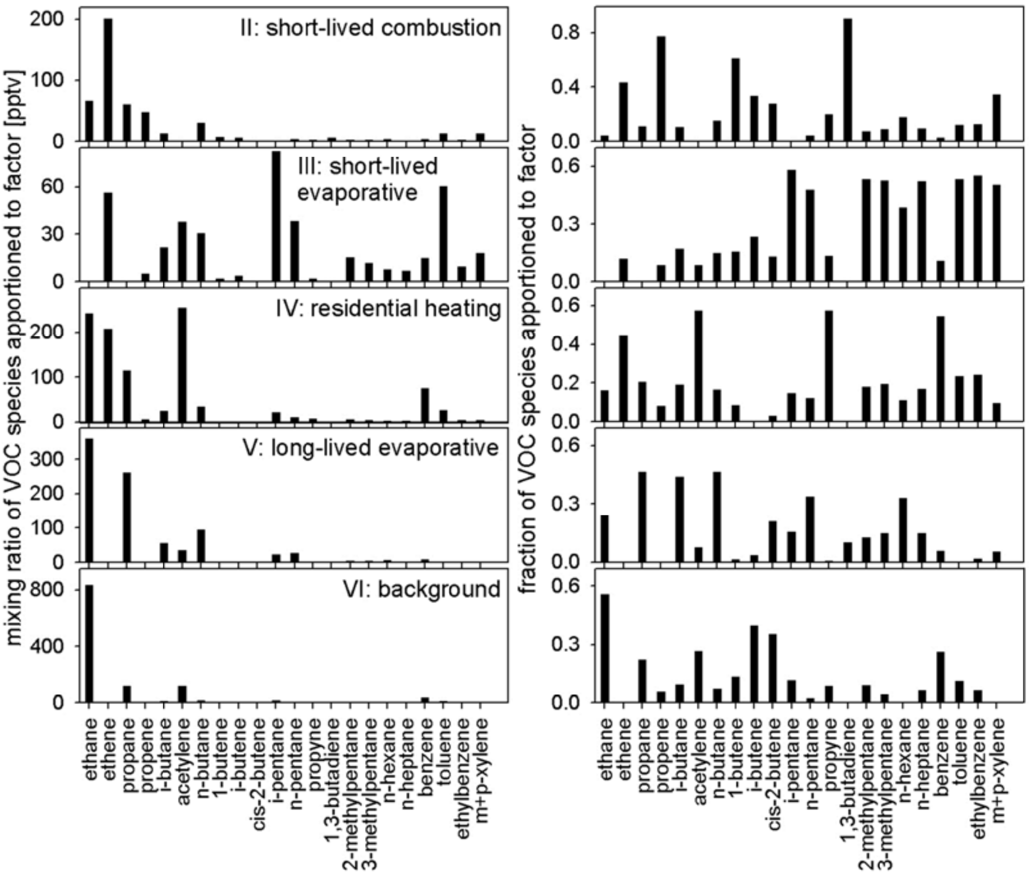

Figure S3. Factor profiles of the five factor solution calculated by PMF for nighttime data only, left: mixing ratio of each species apportioned to each factor [pptv], right: contribution of each factor to the species. Note that the scales for each subplot are different due to large variations in absolute mixing ratios. At nighttime only five factors were resolved, since isoprene contributions were too low for a separate factor.

S4. Effects of different data treatment methods

Five methods to deal with missing values, values below the detection limit and uncertainty values due to interference with other substances or other measurement problems were applied to the six factor solution to analyze the effects of different treatments to the overall solution (cf. Tab. 1). Figure S4 shows the results of the five different scenarios displayed on a relative scale. Only the results for Factor V are shown, since within this factor the largest differences between the computed results of different datasets can be found.

No significant statistical difference for any profile was found for any of the five treatments with a Levene’s test for variances. Thus, in this study different treatment of these values had only a very small quantitative impact on the final result. An additional reason for the differences between the computations could be different starting points in PMF that led to slightly different resulting numbers, but had no influence on the overall stable result.

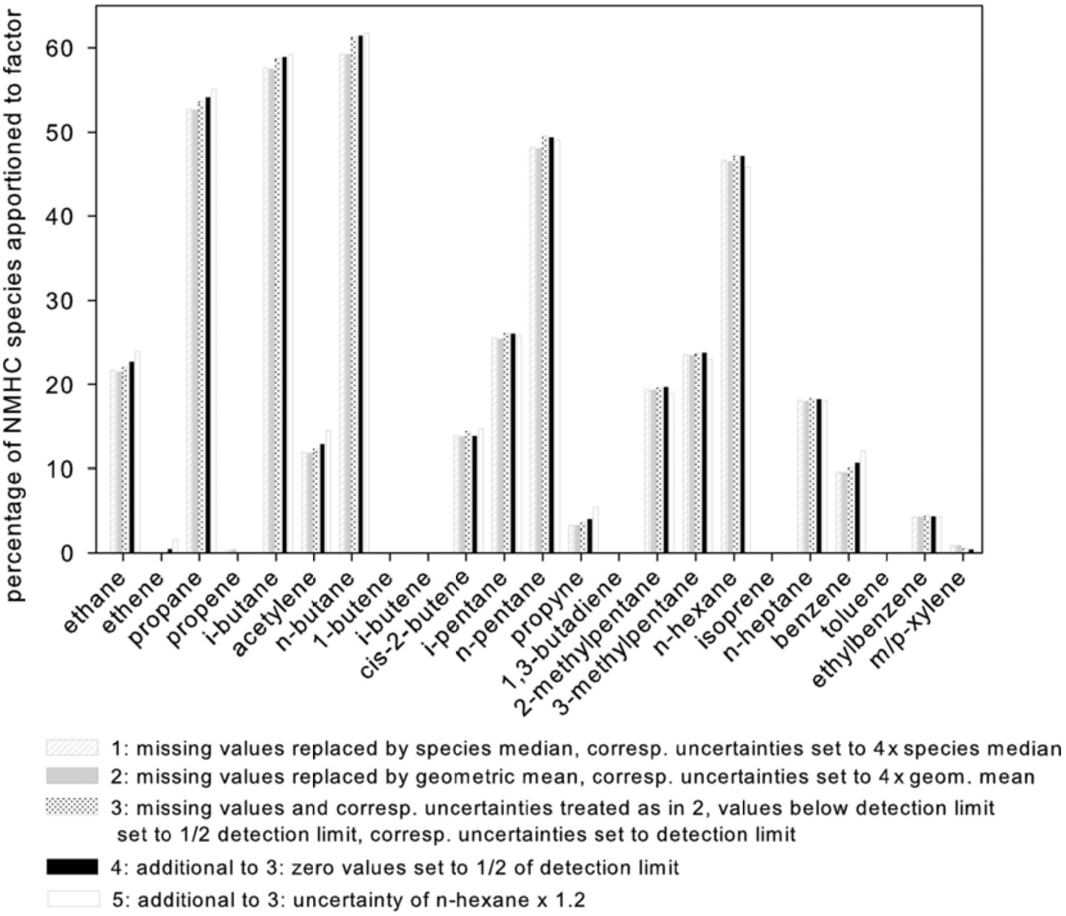

Figure S4. Composition profiles of Factor V, showing the relative contribution of each substance to the factor, for five different treatments of missing and small values (see Tab. 1).

S5. Inclusion of a rough proxy for profile distortion due to photochemical aging

Based on the rate constants for the respective substances (Atkinson, 2000), a reactivity error value depending on different transport or reaction time was calculated for every sample and every species (cf. section 2.3) following the approach by Sauvage et al. (2009). Reactivity error contributions were calculated for up to 72 hours. However, it should be pointed out that the variation of the transport times around characteristic values should be considered in relation to the respective factor and the lifetimes of compounds present in the factor. Accordingly, for Factors II and III with compound lifetimes of hours to few days, it does not make sense to vary the transport times by more than 12-24 hours, otherwise compounds vanish. Furthermore, it should be considered that a factor is distorted due to aging depending on the different lifetimes of the VOC compounds present in the respective factor. But the factors determined by PMF in this paper have been demonstrated to represent compounds with similar lifetimes. Thus, the distorting effect is rather small, and in fact the uncertainty due to aging should be scaled to lifetime differences rather than absolute lifetimes. However, we want to use and discuss the Sauvage et al. (2009) approach, and Tab. S1 shows the average values of the calculated uncertainties for modeled reaction times of 0 and 36 hours exemplarily. Uncertainties for compounds with low reactivity such as some alkanes and aromatic species remained at relatively low levels; ethane as the most stable substance changed its uncertainty from 7.1 to 8.0%, benzene from 8.6 to 16.3%. A stronger influence of reactivity existed for alkenes, e.g. for 1-butene from 29.3 to 92.4% or 1,3-butadiene from 43.6 to 108.9%. The uncertainty for the highly reactive isoprene rose from 33.3% to 101.7% after 12 h (not shown) and to 115.8% after 36 h of assumed reaction time. These high uncertainties made proper modeling in PMF impossible and for the reasons discussed before this demonstrates the limits of this approach. To partly overcome this, the isoprene uncertainty due to reactivity was not considered as the biogenic factor mainly depends on isoprene alone.

The attempt to compensate the mass conservation assumption of the PMF model by changing the uncertainty values of the reactive species had an impact on the resulting factors. The model reacted to the change of uncertainty data by changing the quantitative composition of each factor; the Q values for the model decreased with increasing uncertainties. The largest change of substance contributions to each factor was found between zero and twelve hours where the uncertainties changed the most. Despite the sometimes large changes of some substance contributions, especially ethane and ethene, the relative profiles of the factors remained more or less recognizable for the different tested times. This demonstrates the robustness of the results even though this approach is rather crude for the reasons discussed above. The absolute and relative factor profiles of the PMF solution with six factors and 36 h reaction time for all substances except isoprene are shown in Fig. S5 to illustrate the changes in the solutions with integrated reactivity. However, due to the discussion above the apparently small deviations to the original solution without introducing this enhanced uncertainty due to reactivity were not further discussed.

In terms of model stability, bootstrap runs for the 0 hours solution were very stable with only small variations in the contributions of the substances to the factors. Small variability could only be seen for Factors II and III in regards to some alkenes. Bootstrap runs including reaction times of 12-72 hours were still stable for most factors, in particular for longer-lived factors (Factor IV, VI). Factor V also exhibited a quite stable behavior with only some variability of ethane. The bootstrap process attributed factor profiles with various different contributions of aromatic compounds, *n*-heptane, and, to some extent, methylpentanes and hexane, to both Factors II and III. For Factor II the median of the contributing percentages of *n*-heptane, toluene, ethylbenzene, and xylenes of the 200 bootstrap profiles was much higher than the base run value. The stabilities from the bootstrap runs with the inclusion of the photochemical aging showed a tendency that factors with high relative contributions of long-lived substances were more stable than those with high contributions of substances with higher rate constants such as ethene, propene, and toluene, as expected.

In addition to the less stable bootstrap runs, transport times from major source areas such as the surrounding cities of Munich, Augsburg, and Weilheim were on average shorter than twelve hours. Additionally, the patterns of the factors were shifted towards longer-lived compounds, reducing the influence of species with higher reactivity in a way that the new factors do not well resemble the original emission profiles at the sources. Aging can thus better be taken into account when the profiles are interpreted rather as aged profiles than source emission profiles unaltered by higher uncertainties due to photochemical reactivity. In conclusion, an application of the method proposed by Sauvage et al. (2009) seems not suitable to better interpret the results at this site.

Table S1. Mean values of calculated uncertainties for assumed time of reaction with OH of 0 hours and 36 hours.

Compound	Average uncertainty [%]	
	0 h	36 h
Ethane	7.1	8.0
Ethene	11.0	56.3
Propane	6.2	13.5
Propene	12.6	82.7
Isobutane	6.9	21.6
Acetylene	21.0	25.2
n-butane	26.0	39.2
1-butene	29.3	92.4
Isobutene	16.9	96.2
cis-2-butene	39.1	105.3
Isopentane	11.2	33.8
n-pentane	6.8	32.4
Propyne	33.5	55.1
1,3-butadiene	43.6	108.9
2-methylpentane	15.0	42.8
3-methylpentane	25.7	48.9
n-hexane	18.3	45.3
Isoprene	33.3	115.8
n-heptane	12.9	48.4
Benzene	8.6	16.3
Toluene	10.9	44.8
Ethylbenzene	30.7	59.5
m+*p*-xylene	25.2	72.8

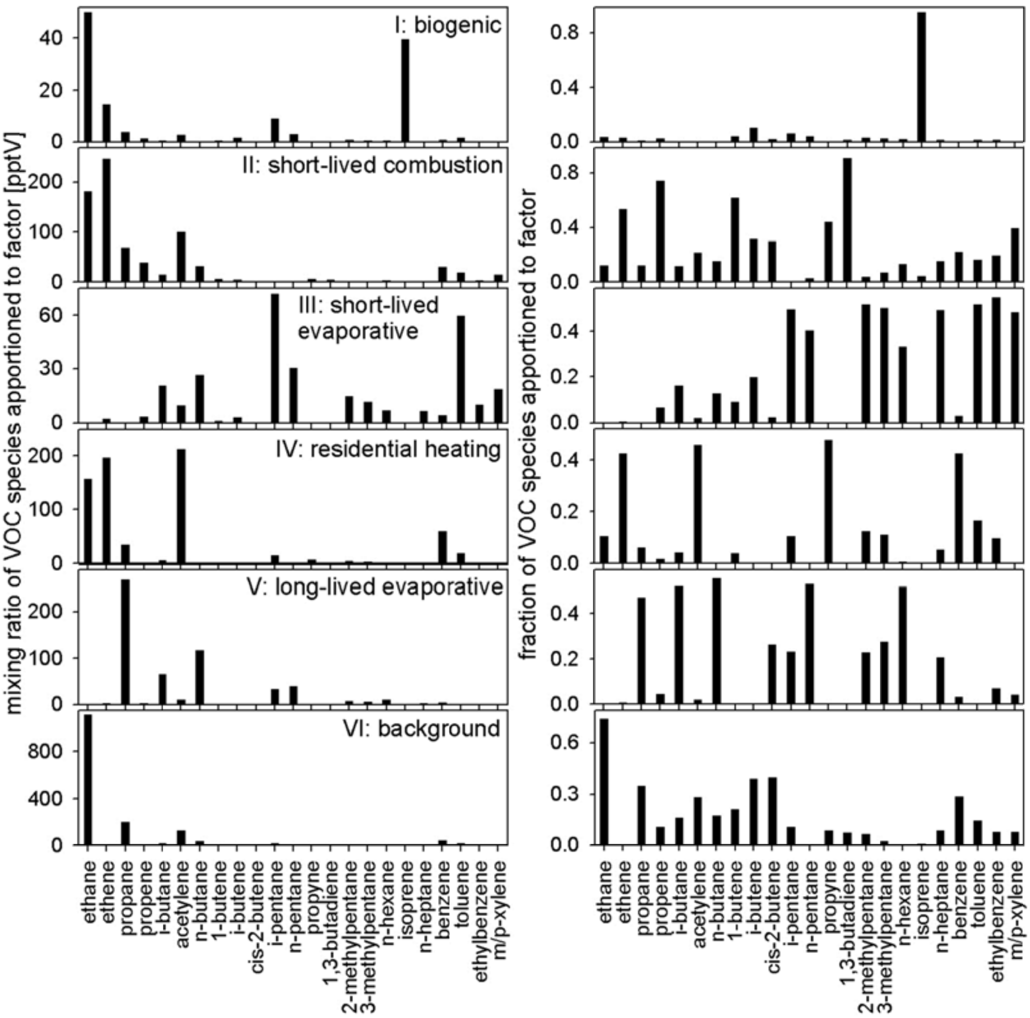

Figure S5. Factor profiles for the six factor solution calculated by PMF, assumed reaction time: 36 h, for isoprene: 0 h. left: mixing ratio of each species apportioned to each factor [pptv], right: contribution of each factor to the species.

S6. Photochemical age-based parameterization method

S6.1 Methods

The parameterization method uses the ratio of two NMHC that react at different rates with OH to represent photochemical age of an air mass and thus photochemical reaction time (de Gouw et al., 2005). OH exposure as product of the OH radical concentration [OH] and reaction time Δt was determined from Eq. (S1).

$$[OH]\Delta t = \frac{1}{(k_A - k_B)} \times \left[ln\frac{[A]}{[B]}|_{t=0} - ln\frac{[A]}{[B]}\right] \qquad (S1)$$

The parameters k_A and k_B are rate constants of the respective NMHC (Atkinson and Arey, 2003), $\frac{[A]}{[B]}|_{t=0}$ is the initial ratio of the NMHC mixing ratios in fresh emissions, and $\frac{[A]}{[B]}$ the measured NMHC ratio. For this approach the ratios of benzene and toluene, *m*+*p*-xylene and benzene, as well as *m*+*p*-xylene and ethylbenzene were used for the determination of photochemical age and OH exposure (de Gouw et al., 2005; Yuan et al., 2012; 2013). The objective was to evaluate if this method is capable to help interpreting the resolved factors in terms of photochemical degradation and source apportionment rather than to determine emission ratios. However, emission ratios were calculated to check for plausibility of the method following de Gouw et al. (2005) and Yuan et al. (2012) using Eq. (S2):

$$\frac{[NMHC]}{[C_2H_2]} = ER \times e^{[-(k_{NMHC}-k_{C_2H_2})[OH]\Delta t]} \qquad \text{(S2)}$$

Here, [NMHC] and [C_2H_2] are mixing ratios of the particular NMHC and acetylene, respectively, k_{NMHC} and k_{C2H2} are OH rate constants, ER is the emission ratio of the respective NMHC to acetylene. OH exposure ([OH]Δt) was calculated with Eq. (S1). ER and k_{NMHC} are unknown and were determined from regression fits as described in de Gouw et al. (2005). The results and applicability of the photochemical age-based parameterization method are discussed in sections S6.2 and S.6.3

Yuan et al. (2012) further suggested two methods to test if the PMF factors are derived from independent sources or rather from different stages of photochemical processing from one common source. The first test is for dependency of the factor fractions of the respective factors to the OH reaction rate constant (k_{OH}). In case that a factor extracted from PMF derived from a specific source category, no dependency should be observed. The second test takes into account the relationship of two extracted factor profiles in terms of photochemical processing. The ratio of the profiles *i* and *j* ($R_{profile(i)/profile(j)}$) should follow Eq. (S3), if two factors are photochemically derived from one another with *A* being a scaling factor accounting for normalization procedures in the PMF model (Yuan et al., 2012).

$$R_{profile(i)/profile(j)} = A \times e^{-k_{OH}[OH]\Delta t} \qquad \text{(S3)}$$

The results of the factor dependency tests are shown and discussed in section S6.3.

S6.2 Results and Discussion

Yuan et al. (2012) used the ratio *m*+*p*-xylene/benzene (M/B), Yuan et al. (2013) *m*+*p*-xylene/ethylbenzene (M/E), and de Gouw et al. (2005) benzene/toluene (B/T) for the determination of the photochemical age. In this study all three ratios were applied and OH exposure and photochemical age compared to test for plausibility. The differences in the resulting photochemical age were very large up to a factor of 6.5 (median of photochemical age calculated for all samples) and 14.1 (arithmetic mean) between the three methods with large outliers at certain times of factors up to 1000. A reliable calculation of photochemical age out of the ratio of two compounds with different reactivity was not possible with our dataset and thus the derived emission ratios also varied largely depending on the NMHC ratio applied for the photochemical age calculation. The choice of a single ratio for the determination of the photochemical age seems quite arbitrary. This is due to several properties at the site Hohenpeissenberg: we cannot hold the assumption (assumption 4 from Yuan et al. (2012)) that the initial emission ratios of the different contributing sources are constant for all sources or air masses that reached the site that in addition have very different reaction times with OH, since the site is exposed to the impact of many overlying sources in different distances. Thus, it is not possible to separate aging, mixing, and source type from one specific ratio. Therefore, the photochemical age-based parameterization method is not applicable at our site.

S6.3 Evaluation of photochemical dependency of the factors

Yuan et al. (2012) suggested tests to check for photochemical dependency of the resolved factors. In case that the factors extracted from PMF are derived mainly from photochemical processing a dependency on the rate constants should be seen. Figure S6 illustrates relative factor contributions to each NMHC species as a function of the OH rate constants. Within the two evaporative categories (short-lived and long-lived) and also the two combustion categories (short-lived and residential heating/long-lived) a split of short and long-lived compounds could be observed between the two respective factors. Not a clear photochemical relationship, as shown by Yuan et al. (2012) for Beijing data, could be seen in the Hohenpeissenberg analyses. The factors discussed here showed a certain separation by photochemical processing into short and long-lived factors, but were still separated into the source categories evaporative and combustion, supporting the source apportionment from section 3.2.

The second method tests the relationship of two factor profiles in terms of photochemical aging. If one factor was photochemically derived from another one, a dependency of the factor ratio to the rate constants should be found. Figure S7 illustrates results of the following factor ratios: background and short-lived combustion, background and residential heating/long-lived combustion, background and long-lived evaporative, background and short-lived evaporative, long-lived evaporative and short-lived evaporative, residential heating/long-lived combustion and short-lived combustion. None of the tested factor ratios was explained by a relation according to Eq. (S3), not even between short-lived and long-lived profiles within the evaporative or combustion categories. This indicates independency of the factor profiles in term of photochemical processing.

Both tests did not show clear signs of photochemical relationship between the different extracted profiles. As has been discussed before, this is attributed to the complex situation of multiple sources with different transport times overlaying in the receptor profiles.

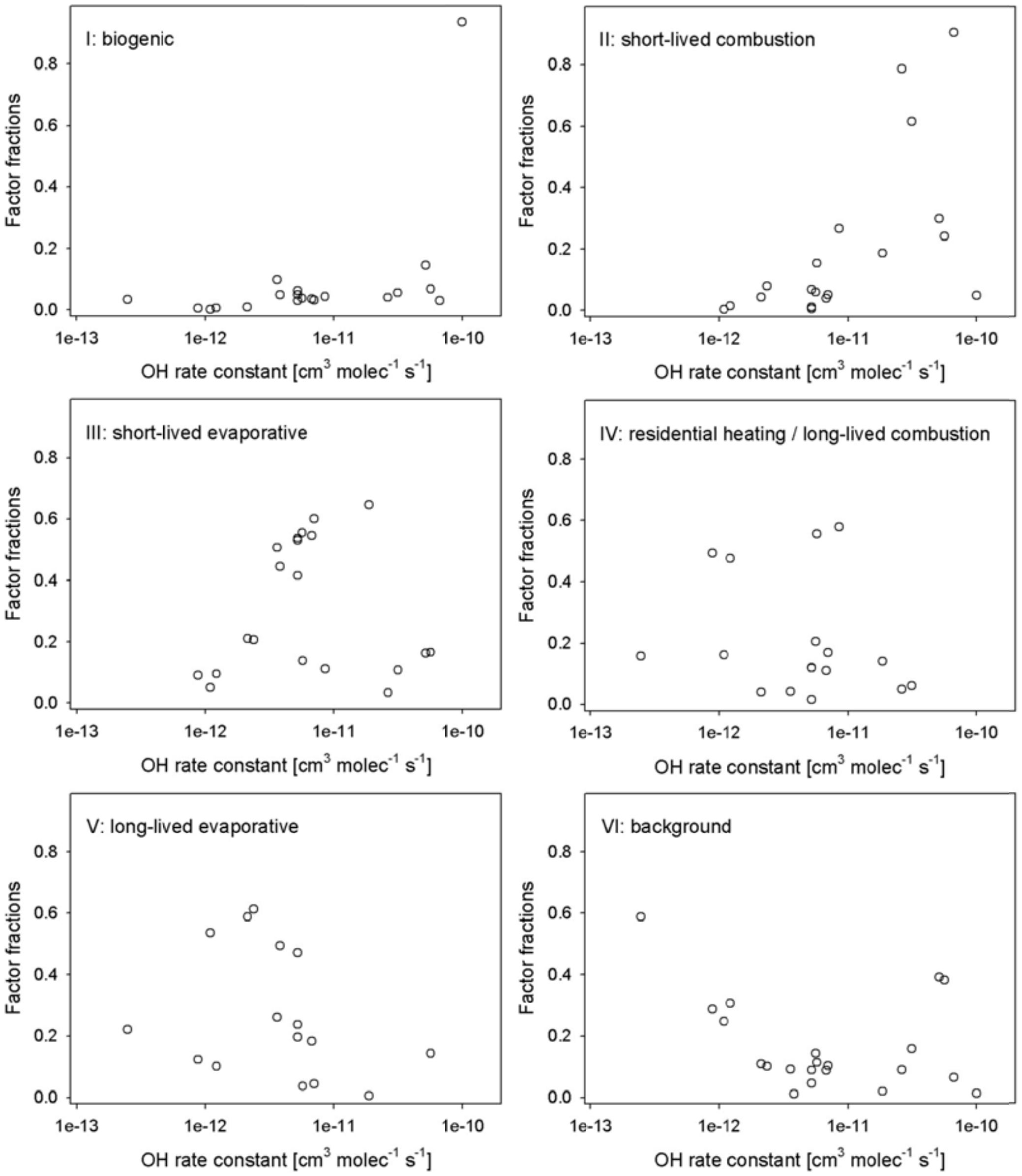

Figure S6. Factor contributions to each NMHC species as a function of the rate constants (k_{OH}). Each data point represents one compound. No clear dependence on k_{OH} can be seen.

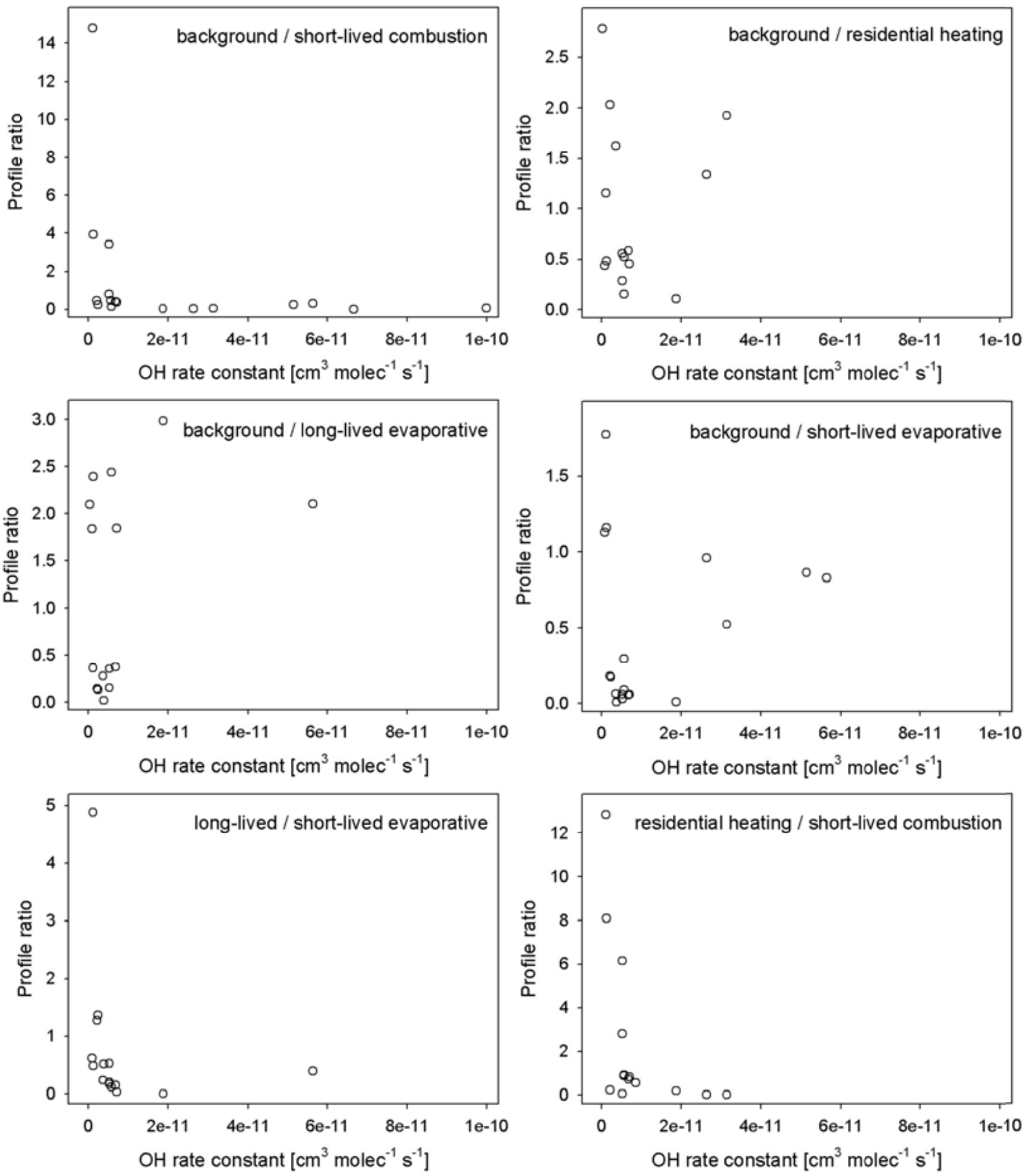

Figure S7. Abundance ratios of the compounds from two profiles as a function of k_{OH} values. No dependence on k_{OH} can be seen which means that the factors are not directly photochemically derived from each other.