Can Positive Matrix Factorization help to understand patterns of organic trace gases at the continental GAW site Hohenpeissenberg?

M. Leuchner [1,†,*], S. Gubo[1], C. Schunk[1], C. Wastl[1,§], M. Kirchner[2], A. Menzel[1,†], C. Plass-Dülmer[3],

[1] {Fachgebiet für Ökoklimatologie, Technische Universität München, Hans-Carl-von-Carlowitz-Platz 2, D-85354 Freising, Germany}

[2] {Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Cooperation Group of Comprehensive Molecular Analytics, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany}

[3] {Meteorologisches Observatorium Hohenpeissenberg, Deutscher Wetterdienst, Albin-Schwaiger-Weg 10, D-82383 Hohenpeissenberg, Germany}

[§] {now at: Zentralanstalt für Meteorologie und Geodynamik, Hohe Warte 38, A-1190 Wien, Austria}

[†] {Institute for Advanced Study, Technische Universität München, Lichtenbergstraße 2 a, D-85748 Garching, Germany}

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

Abstract

From the rural Global Atmosphere Watch (GAW) site Hohenpeissenberg in the pre-alpine area of Southern Germany, a dataset of 24 C_2–C_8 non-methane hydrocarbons over a period of seven years was analyzed. Receptor modeling was performed by Positive Matrix Factorization (PMF) and the resulting factors were interpreted with respect to source profiles and photochemical aging. Different to other studies, no direct source attribution was intended as due to chemistry along transport, mass conservation from source to receptor is not given. However, also at remote sites such as Hohenpeissenberg, the observed patterns of NMHC can be

derived from combinations of factors which were determined by PMF. A six factor solution showed a high stability and the most plausible results. In addition to a biogenic and a background factor of very stable compounds, four additional anthropogenic factors were resolved that could be divided into two short- and two long-lived patterns from evaporative sources/natural gas leakage, and incomplete combustion processes. The volume or mass contribution of the different factor categories at the site over the entire period was, in decreasing order: background, gas leakage and long-lived evaporative, residential heating and long-lived combustion, short-lived evaporative, short-lived combustion, and biogenic. The importance with respect to reactivity contribution was generally vice versa with the biogenic and the short-lived combustion factor contributing most. The seasonality of the factors was analyzed and related to results of a simple box model using constant emissions and the photochemical decay calculated from the measured annual cycles of OH radicals and ozone. Two of the factors, short-lived combustion and gas leakage/long-lived evaporative showed winter/summer ratios of about 9 and 7, respectively, as expected from constant source estimations. Contrary, the short-lived evaporative factor showed about factor 3 higher summer time emissions, and the factor residential heating/long-lived combustion about factor 2 higher values in winter.

Keywords

non-methane hydrocarbons (NMHC); receptor modeling; Positive Matrix Factorization (PMF); seasonal cycles, source apportionment; air quality; GAW; photochemical aging

1 Introduction

Tropospheric ozone is an environmental pollutant that causes adverse effects to vegetation, e.g. by reducing and altering physiological processes and plant growth (Matyssek et al., 2010; Nunn et al., 2005), and to humans, where respiratory diseases can be linked to ozone. In addition to these effects, ozone has been one of the most important greenhouse gases since the beginning of industrialization with a large impact on radiative forcing (Gauss et al., 2003). Atmospheric

63 background concentrations of tropospheric ozone are expected to increase in the
64 21st century (Vingarzan, 2004). In contrast to other greenhouse gases, such as
65 carbon dioxide, methane, and nitrous oxide, it is not emitted directly, but
66 produced in the atmosphere by photochemical processes from precursor
67 substances. The main drivers for the production of ozone, besides nitrogen oxides
68 ($NO + NO_2$) and carbon monoxide (CO), are volatile organic compounds (VOC)
69 and amongst them non-methane hydrocarbons (NMHC) (Atkinson, 2000). NMHC
70 are not only important for the photochemical formation of tropospheric ozone, but
71 also for other secondary air pollutants such as peroxycarboxylic nitric anhydrides
72 (PAN), formaldehyde (HCHO) (Rappenglück et al. 2010), and secondary organic
73 aerosols. In addition, many NMHC species act directly as air toxics or hazardous
74 air pollutants. As a consequence, the Gothenburg protocol gave national emission
75 ceilings which for Germany asked for a 73% reduction by 2010 compared to
76 1990. After new evaluation in 2013, the EU commission proposed further
77 reductions which for Germany account for 43% in 2030 compared to 2005
78 (European Commission, 2013). Thus, monitoring and modeling of the
79 spatiotemporal distribution of these species and relating them to source sectors are
80 important for mitigation strategies concerning air quality, radiative forcing, and
81 human health.
82 The most important sources of NMHC are combustion of fossil fuels from road
83 traffic and industrial processes, handling and evaporation of fuels, solvents, and
84 gases, as well as plant emissions, amongst others. Stevenson et al. (2005)
85 expected a strong increase of biogenic NMHC emissions, mainly isoprene,
86 monoterpenes and ethene, caused by temperature stress with future increase of
87 global temperatures. On a global scale, biogenic emissions dominate total VOC
88 emissions (e.g. Sindelarova et al., 2014), while in many urban areas they play a
89 minor role due to high amounts of anthropogenic emissions; however, this is
90 highly dependent on the type and location of the urban area. For major
91 metropolitan areas with large anthropogenic emissions (e.g. Houston, Atlanta) a
92 large impact of the highly reactive biogenic VOC on ozone formation and OH
93 chemistry has been shown (e.g. Chameides et al., 1988; Mao et al., 2010;
94 Leuchner and Rappenglück, 2010).
95 Reliable long-term scientific data of NMHC are gathered by several groups and
96 various networks. Within the framework of the World Meteorological

Organization (WMO), the Global Atmosphere Watch (GAW) program has been developed to achieve global measurements of the chemical composition of the atmosphere with high data quality (WMO, 2007).

In order to quantify impacts of biogenic and anthropogenic origin on photochemical production of ozone, aerosols, and other compounds, an apportionment into specific source categories is necessary (Badol et al., 2008). Several receptor models such as principal component analysis/absolute principal component scores (PCA/APCS) (e.g. Chan and Mozurkewich, 2007; Guo et al., 2004; 2006), chemical mass balance (CMB) (e.g. Badol et al., 2008; Na and Kim, 2007), or UNMIX (e.g. Jorquera and Rappenglück, 2004; Olson et al., 2007) were used for source apportionment. In particular Positive Matrix Factorization (PMF) (e.g. Lingwall and Christensen, 2007; Paatero, 1997; 1999), a multivariate mathematical receptor model, has been shown to be quite reliable at identifying and quantifying source categories. But most studies investigating NMHC composition were concentrated in urban metropolitan areas with mainly anthropogenic emissions (e.g. Brown et al. 2007; Leuchner and Rappenglück 2010; Na and Kim, 2007).

Yuan et al. (2012) stressed the importance of different reactivity of the NMHC compounds and the impact of photochemical aging on the interpretability of the resolved factors as source profiles that has not been considered in most of the studies applying PMF. The impact of photochemical processing increases with longer transport times from source to receptor. Only a few studies have applied PMF receptor modeling at remote sites with a focus on the global or continental background. Lanz et al. (2009) and Sauvage et al. (2009) used PMF analysis for reactive species such as NMHC at remote sites in Switzerland and France despite the PMF assumption of mass conservation from source to measurement site.

In the current work, NMHC data from the GAW global site Hohenpeissenberg, Southern Germany, were used to quantify the impact of different source categories as well as their seasonality at this rural site with PMF analysis. One main objective of this study was to interpret and discuss PMF as a statistical tool to reliably identify sources of reactive trace gases at a rural site though source profiles have been distorted due to photochemical aging.

2 Methods

2.1 Experimental setup

The GAW Observatory Hohenpeissenberg is located about 70 km southeast of Munich (47°48‘ North, 11°02‘ East) at 980 m a.s.l. on top of a hill which is about 300 m above the surrounding countryside (approx. 70% pasture and 30% forest). Sample air was routinely measured daily at 01:00 h (41% of data) and 13:00 h CET (48%), however, 11% of the data were measured at other times of the day. At 13:00 h, the site was generally in a vertically fully developed mixed-layer and local emissions may have affected measurements only low to moderately. For this analysis mainly the 13:00 h data were used to minimize the influence of local sources and shallow boundary layer conditions during nighttime as well as to ensure the homogeneity of the dataset. However, nightly data was analyzed as well to identify differences and help interpreting the results.

C_2-C_8-NMHC were measured with an on-line GC-FID system. It consisted of a 3600 CX Varian gas chromatograph combined with a flame ionization detection (FID) system until January 2008 and was then replaced by a Varian CP-3800 GC-FID. The air intake was 17 m above the ground, and 2 m above a flat roof which was about the same height as the nearby forest canopy (>10 m distance). The intake was a downwards-facing glass-funnel connected to a permanently flushed glass-manifold (375 l/min, 8 m length, 4 cm I.D.). The GC sampling unit was connected to a port on the manifold via a 1/16" Sulfinert line (Restek, length: 2 m, 0.96 mm I.D., 50 ml/min) such that an overall residence time in the lines of 3.5 s was achieved. After the port, the sample gas passed a filter for aerosol and ozone removal (PTFE filter holder: 25 mm I.D., Metron Technology, PTFE coated glass fiber filter, Fiberfilm, Pall Life Sciences, impregnated with sodium thiosulfate ($Na_2S_2O_3$), backed by a PTFE-membrane filter, 20-30 µm pores, Metron Technology), and further downstream a stainless steel-screen (10 µm pores, VICI). A custom-built sampling and gas flow system was used comprising a moisture trap at 228 K (0.5 m 1/8" Sulfinert, Restek), a VOC trap using cryo-adsorption on glass beads (87 K adsorption, 403 K desorption) (SPT-type by Varian, installed in custom built LN_2 dewar), a sample volume determination by measuring the pressure increase in an evacuated reference volume, and the corresponding Valco (VICI) switching valves mounted in a temperature controlled compartment at 293 K. After sampling for 20 min, the cryo-trap was

dry-purged by Helium at 15 ml/min for 15 min. The adsorbed NMHC were then thermally desorbed in a helium carrier gas flow at 5 ml/min injected and separated on a PLOT column (Al_2O_3/ KCl, 50 m x 0.53 mm I.D., Chrompack, The Netherlands). After an initial isothermal phase (313 K for 2 min), the GC column was heated in two phases, first to 345 K (4 K/min), then with a rate of 6 K/min to 473 K. This temperature was kept for 33.7 min. The end of the column extended to the FID system, where the separated compounds were identified. This system was regularly checked and calibrated with helium (zero gas), calibration gas by NPL (certified mixtures of a few ppb in nitrogen of some 30 NMHC), and different reference gases holding synthetic and whole air mixtures in pressurized cylinders (see metadata at WDCGG, 2013). The system has participated in several intercomparisons and proved its ability to measure NMHC on high quality levels (Hörger et al., 2014; Plass-Dülmer et al., 2006; Rappenglück et al, 2006).

A more detailed description of measurement, integration, and error assessment was published by Plass-Dülmer et al. (2002). For this study mixing ratios of 24 substances over the course of seven years from 2003 to 2009, measured daily, were used. An individual uncertainty for each compound and each measurement comprising systematic uncertainty contributions and random factors was estimated and assigned to each value. It considers blank values, peak integration errors (including insufficient chromatographic separation) and detection limit, calibration uncertainties and random fluctuations in the system response (Plass-Dülmer et al., 2002).

2.2 Positive Matrix Factorization (PMF) model description

Next to Chemical Mass Balance (CMB), UNMIX, and PCA (principal components analysis), Positive Matrix Factorization (PMF) has become an accepted and regularly used tool for receptor modeling. In this study, factor analysis was conducted with PMF 3.0 (US-EPA, 2011). Positive Matrix Factorization determines the number of source factors *p*, a species profile *f* for each factor, and the amount *g* that each factor contributes to each sample.

$$X_{ij} = \sum_{k=1}^{p} g_{ik} \times f_{kj} \quad (1)$$

In PMF Eq. (1) is solved by decomposing the matrix X_{ij} of measurement data, with *i* the number of samples and *j* the number of the different chemical species,

into two matrices, factor contributions, and factor profiles. Both factor contributions and factor profiles can then be analyzed.

There are natural and logical physical conditions for such a model: the original data have to be reproduced by the model, the predicted source compositions and contributions must be non-negative, and the sum of the predicted mass contributions for each source must be less than or equal to the total measured mass for each substance (Hopke, 2003).

The multilinear engine ME-2 (Paatero, 1999) gives PMF the ability to solve multilinear problems and implement constraints like the replacement of missing values and individual weighting of data points by associating an uncertainty value u_{ij} to each point. The object function (Eq. (2)) is then minimized using these uncertainties (Norris et al., 2008).

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{X_{ij} - \sum_{k=1}^{p} g_{ik} \times f_{kj}}{u_{ij}} \right]^2 \quad (2)$$

PMF 3.0 also gives the opportunity to test the stability and uncertainty of the computed solutions by using a bootstrap technique. It also provides a tool called Fpeak to control the rotations of the different factors (Norris et al., 2008).

Several studies have compared PMF, CMB, PCA, and UNMIX (Anderson et al., 2002; Miller et al., 2002; Paatero and Tapper, 1994; Willis, 2000) and found limitations and advantages of the different models. Some of the advantages of PMF are the good performance not only with simulated data, the non-negativity constraint, and the possibility of individual treatment for single data points.

Various studies have shown that PMF provides physically reasonable results for source identification of NMHC in environments located in proximity to the sources (e.g. Buzcu and Fraser, 2006; Lanz et al., 2008; Leuchner and Rappenglück, 2010) or particulate matter (e.g. Santoso et al., 2008; Tauler et al., 2009; Yue et al., 2008). But as receptor models are only mathematical models, they do not use pollutant emissions, chemical transformation mechanisms, or meteorological data to identify and quantify the sources at a receptor location. It is difficult to use PMF for data from remote sites, because with chemical reactions during the transport of air masses (Atkinson and Arey, 2003) and the effects of mixing (Parrish et al., 2007) the presumed mass conservation from source to measurement site, necessary for using receptor models (Hopke, 2003), is not

given. Despite these limitations, PMF has been used for remote VOC data (Lanz et al., 2009; Sauvage et al., 2009) and obtained reasonable results in terms of plausible factor compositions and contributions of the source categories. However, it needs to be considered that not only emission profiles but also their different aging determine the factor solutions of PMF.

2.3 Data treatment

A total of 2335 valid day (13:00 h) measurements of 24 substances per measurement, in the following named daytime data, were available for the investigated time span of which 345 values of individual substances were missing (0.6% of the dataset). For all other times a total of 2277 valid measurements of 24 substances, in the following named nighttime data, were used for analysis of which 325 values were missing (0.6% of the dataset). Both datasets were analyzed separately as well as combined. If not indicated explicitly, results refer to the daytime data.

There are different ways to treat missing data. PMF 3.0 provides the option to exclude the entire sample. In that case a loss of 15% of the sample data would have occurred. To avoid such a high loss of data, the missing values were replaced by the respective species geometric mean and the corresponding uncertainties were set to four times the geometric mean according to Sauvage et al. (2009). Five different treatments of missing value replacement and uncertainty assessment were performed for the daytime data as described below and shown in Tab. 1.

Values greater than zero, but below the specific detection limit, were replaced by half the detection limit. The uncertainty for zero values and values below the detection limit was set to the specific detection limit (Sauvage et al. 2009). Fpeak values, indicating the degree of rotation of the solutions, were varied between 5.0 and -5.0 in steps of 0.1.

The impact of different treatments of missing values, of values below the detection limit, and of zero values on the overall results of the PMF analysis was assessed by utilizing five methods of value replacement. In this dataset the number of these values was rather low at 2.3% of all data. Table 1 shows the five different treatments with missing values (treatments 1-5), values below the detection limits (3-5), and zero values (4) replaced by the species median (1) and

geometric mean (2-5). The respective uncertainties were fitted accordingly in treatments 1-4 and with additional 20% to *n*-hexane in treatment 5, due to peak overlaps with an unknown substance in the 2003 and 2004 data. Statistical differences in the treatments were tested with a Levene's test for variances. For the test, data were linearly transformed by normalization with the respective arithmetic mean value of the different treatments. Results are shown in section S4 in the supplement.

The remote character of the research site at Hohenpeissenberg implies that only a few substances were emitted locally, but rather transported from multiple sources in different distances. During the transport time photochemical reactions occur and the original emission pattern is altered, due to different photochemical reactivity of the compounds (Atkinson, 2000; 2008). PMF applied to attribute sources, however, needs inertness of the substances and cannot integrate reactivity into the model. Sauvage et al. (2009) proposed a method to consider photochemical changes by enhancing the uncertainty with increasing compound-reactivity. For each compound a potential error $E_{j(reactivity)}$ was computed with Eq. (3) assuming pseudo first-order reaction kinetics and photochemical reactivity mainly driven by OH radical reactions.

$$E_{j(reactivity)} = X_{ij}(1 - e^{-k_j[OH]\Delta t}) \quad (3)$$

k_j is the second-order rate constant of the reaction between the substance *j* and OH (Atkinson and Arey, 2003), Δt is the source-receptor time of transport. [OH] is the seasonally and spatially averaged OH concentration published by Spivakovsky et al. (2000). The overall uncertainty s_{ij} for the PMF modeling was then computed following the ISO guide rule for uncertainty (ISO 13005, according to Sauvage et al. (2009)) (Eq. (4)) and used for PMF computations.

$$s_{ij} = 2\sqrt{(\frac{U_{ij(measure)}}{2})^2 + (\frac{E_{j(reactivity)}}{\sqrt{3}})^2} \quad (4)$$

Thus applied, the PMF results in factors which correspond to the prevailing factors determined at the receptor site and which are different from the pure emission profiles. They may be seen as aged source profiles with an aging corresponding to some mean transport time from many diverse source locations to the receptor site. It should be pointed out that this approach reduces the impact of the shorter-lived NMHC on the achieved results and that obtained factors have high uncertainty in these short-lived compounds. Taking these complications into

297 account, the aging, the integration over many sources, the discrimination of short-
298 lived compounds, and comparisons with emission profiles are considered as
299 misleading, especially for the short-lived compounds. Nevertheless, the method
300 was tested for our data set and its applicability is further discussed in section S5 in
301 the supplement.
302 Another method to account for photochemical processing is the photochemical
303 age-based parameterization method suggested by Yuan et al. (2012) following de
304 Gouw et al. (2005). It is underlying the assumption that the main sources at the
305 receptor site are either anthropogenic and originate from one major urban
306 settlement with a defined transport time to the receptor site, or of biogenic origin.
307 It is further assumed that the magnitude of urban emissions is proportional to
308 acetylene emissions, the reaction with OH radicals is the dominant form of
309 removal, and photochemical age is defined by the transport from the one urban
310 emission area to the receptor and can be calculated from the ratio of mixing ratios
311 from two NMHC with different lifetimes. These assumptions, in particular the
312 first and last, do not apply at a remote receptor site with many small
313 anthropogenic emission sources around. Again, this approach and its applicability
314 to Hohenpeissenberg data will be further discussed in section S6 in the
315 supplement.
316 There are different guidelines to help determine the number of factors that best
317 model the measured reality. Mathematical variables like *Q* values or the
318 distribution of residuals and stability of the solution can be taken into account, but
319 interpretation of the computed factors by the analyst is a crucial part of selecting
320 the most appropriate solution (Hopke, 2003; Norris et al., 2008). Comparing
321 computed *Q* values as a function of the number of factors to theoretical *Q* values
322 (approximately the number of data points) seems to work only for certain kinds of
323 weighted uncertainties (Hopke, 2003).
324 PMF solutions with 2 to 20 factors were calculated, but only the four most
325 plausible solutions (five to eight factors) were compared in this work. Selection
326 criteria were mathematical indicators such as the *Q* value, residual distribution,
327 explained variance, as well as the plausible explanation of the source categories
328 by expert knowledge of the authors.
329 At the GAW site Hohenpeissenberg other trace gases and particulates have also
330 been measured (Gilge et al., 2010). Resolved factors can be compared to these

independent measurements to verify the apportionment of the factor when the hydrocarbons and the trace gas are emitted from the same source or at the same time but by a different source. Thus, the contributions of factors resolved by PMF were correlated to the additional trace gases NO, NO_2, SO_2, CO, and NO_y. The secondary products ozone and PAN, as well as particulate matter (PM10, PM3, black carbon) data were also analyzed and included in additional PMF runs. Since sometimes correlations can also be coincidental, Lanz et al. (2008) suggested that these correlations can add evidence to the source apportionment of the factors, but should not be used as the only basis for the attribution of sources. In addition to simple correlations between the resolved factors and other substances, the single trace gases or aerosols as well as several combinations of those substances were included in the PMF model to test and help interpreting the apportionment of single factors. However, including further non-NMHC substances did not show clear results. With more than one or two compounds included, the factor solutions did sometimes show alterations in the source profiles and the occurrence of separate factors. Thus the numbers of factors needed to be increased to seven or eight to still retrieve the six identified NMHC factors. In the following source apportionment (section 3), the most important results from inclusion of these substances are noted for each resolved factor. Only if a resolved factor profile significantly correlated with one or more of the other trace gases, it is mentioned in the text.

PMF 3.0 provides a bootstrap function that selects blocks of input samples and creates new input files from them. These files have the same dimensions as the original input files. Then PMF is run and the resulting factors are mapped to the base factor they correlate with best. In this study, 200 bootstrap runs were performed on the base of the final solution with six factor profiles.

2.4 Seasonal cycles calculated from a simple box model

Anthropogenic NMHC generally show pronounced seasonal cycles with maxima in winter and minima in summer. Apart from changing emissions over the course of the year and often associated with temperature, e.g. evaporation generally has a summer maximum and residential heating a winter maximum, the photochemical cycle determines the chemical removal with an order of magnitude higher concentrations of OH in summer than in winter, at mid latitudes. As this

365 photochemical signal is often stronger than the seasonal cycle of emissions, we
366 first analyzed the seasonal variation solely due to photochemistry with assumed
367 constant emissions. Then, deviations from this seasonality in the PMF derived
368 factors gave information on the seasonality of the corresponding emissions.
369 It is assumed in this simple box calculation that constant NMHC emissions of 5 *
370 10^3 molec cm^{-3} s^{-1} are fed into the atmosphere. Emissions are balanced by the
371 atmospheric removal due to OH radicals and ozone. The OH and ozone
372 concentrations are taken from the monthly averaged measurements of these
373 compounds at Hohenpeissenberg, OH by chemical ionization mass spectrometry
374 (CIMS) (Berresheim et al., 2000) and ozone by UV-absorption (Gilge et al.,
375 2010). Sine fits through the monthly averages are used to drive the removal with
376 24 hour time-resolution. The NMHC start concentrations are adjusted such that
377 stable annual cycles are established.
378 The different seasonal signals are demonstrated by the 95-/5-percentile ratios as a
379 function of the OH rate constants in Fig. 1 for compounds which only react with
380 OH radicals. No annual cycles could be seen at reaction rate constants below 10^{-14}
381 cm^3 $molec^{-1}$ s^{-1}, and an increase from winter/summer ratios of 2 to 12 was
382 depicted for OH reaction rate constants from 10^{-13} to 10^{-11} cm^3 $molec^{-1}$ s^{-1},
383 respectively. Accordingly, for ethane with k_{OH} = 2.1 * 10^{-13} cm^3 $molec^{-1}$ s^{-1}, a
384 "damped" seasonal variability of factor 3.7 could be expected if sources were
385 constant over the year, whereas for reactive compounds like heptane with k_{OH} =
386 6.9 * 10^{-12} cm^3 $molec^{-1}$ s^{-1} the seasonal variation of factor 11 had about the range
387 as for OH (Tab. 2; rate constants at 283K; Atkinson, 2000; Atkinson and Arey,
388 2003). Figure 2 shows the resulting seasonal variation for different anthropogenic
389 NMHC including their reactions with OH and ozone. Rate constants were
390 computed in dependence of the monthly mean temperatures for ozone, for OH the
391 monthly averaged half width of the OH diurnal cycle was determined and the
392 corresponding average temperature was used. The reactive alkenes like *cis*-2-
393 butene showed in this box calculation seasonal cycles which were less
394 pronounced than for compounds not reacting with ozone because the seasonal
395 variation of ozone had only a summer/winter ratio of 1.8.
396 The winter/summer ratios of all considered NMHC in this study were calculated
397 in this simple box and compared to the observed ratios (Tab. 2). The observed
398 ratios were derived from sine fits to the monthly mean concentrations over the last

ten years after compensating for trends in the data. Apparently, only few NMHC show similar winter/summer ratios as expected from constant emissions.

3 Results and Discussion

3.1 Determination of the number of factors

The calculated Q values of four different factor solutions with five to eight factors and other diagnostic parameters are shown in Tab. 3. The values decreased with increasing number of factors, due to a better explanation of the variability of the measured NMHC mixing ratios by a higher number of factors, and lower global minima of the object functions. All scaled residuals should be within ± 3σ (Willis, 2000) and symmetrically distributed. With this dataset all solutions from five to eight factors had normally distributed residuals. The number of residuals beyond three standard deviations decreased for an increasing number of factors (Tab. 3) down to four for 20 factors. An Fpeak value of 0.0 showed the lowest Q values for all solutions displayed in Tab. 3, additional rotation of the factors did not improve the results. Coefficients of determination of modeled to measured mixing ratios and the mean ratios were high for all solutions and increased with the number of resolved factors from 0.86 (R^2) and 0.97 (ratio) to 0.89 and 0.98, for five to eight factors, respectively.

An indication for an appropriate number of factors is the stability of the factors after performing the analysis at least five times with the same parameters but randomized starting points. No multiple solutions should be found (Hopke, 2003). In this analysis the tested five, six, seven, and eight factor solutions were very stable and showed the same distribution of factors for all five computations (not shown here).

On the basis of these statistical indicators no final decision about the optimal number of factors could be made. Thus, it remained a process based on plausibility arguments mainly by checking the resulting factors versus reasonable source or aged source profiles (cf. section 3.2), also in comparison to previous studies (Lanz et al., 2009; Lanz et al., 2008; Sauvage et al., 2009). Careful consideration led to the decision of choosing six factors as the most reasonable solution. A comparison to alternative solutions with five, seven, and eight factors can be found in the supplement (section S1).

3.2 Source apportionment

In the following, the six factor solution (Fig. 3) of the daytime data is presented and discussed. Including nighttime data or use of exclusively nighttime data resulted only in very slight differences in the derived factors (see sections S2 and S3 in the supplement).

Figure 3 shows the absolute and relative contributions of substances to the six factor solution. Absolute values are the mixing ratios of each substance that PMF apportioned to each factor. Relative contributions are the fractions of each substance attributed to each factor, therefore, the sum of all factors for each substance is 1.

The apportionment of factors to source categories and the interpretation of chemical aging for the respective factors were performed by comparison to source profiles from the literature. Because of the complexity of the atmospheric system with transport, mixing, and chemistry, each individual factor cannot be attributed exclusively to one source category. The factors should then be seen as aged profiles originating from different sources belonging to similar source categories (Sauvage et al., 2009). In addition to a change of composition patterns during transport, emission profiles in literature vary due to measurement uncertainty associated with the use of different techniques, e.g. on-line or canister samples, the number of substances measured, the experimental set-up and associated conditions, and the published units. These variations made direct quantitative comparisons of the profiles difficult, but allowed a qualitative assessment for identification of possible source profiles.

For a plausibility check of the attribution of the factors to source categories, the annual courses including the winter/summer ratios of the retrieved factors are compared to corresponding box model calculations in section 3.3.

3.2.1 Biogenic sources

Factor I explained 34 pptv or 94% of the measured isoprene which was the only biogenic NMHC included in this analysis. On the absolute scale ethane (49 pptv), ethene (21 pptv) and isopentane (14 pptv) also contributed to this factor, but on the relative scale this factor only contains about 4% of the total amounts of ethane

and ethene, thus, this factor is apportioned to biogenic sources. The short lifetime of isoprene in the atmosphere excludes distant sources. Small amounts of ethene found in this factor might also be of biogenic origin, since it is an important plant hormone (Fall, 1999). Factor I was the only factor with a distinct maximum in summer (Fig. 4). Since isoprene emissions depend on the photosynthesis of plants, temperature, and solar radiation (Fuentes and Wang, 1999), the maximal source strength and thus the maximal mixing ratios were found in July. The ethane contribution to this factor might have derived from biomass burning (Stein and Rudolph, 2007) with a maximum during the summer.

Similar profiles for this factor, including amounts of ethane, *n*-pentane and isopentane were also found by Sauvage et al. (2009) for three remote sites in France. The alkane contributions in this factor could be attributed to artifacts from the PMF model, temperature related emissions like evaporation from fuel, or mixing with other sources. Thus, the biogenic factor is possibly slightly overestimated, but on the other hand the whole set of biogenically emitted monoterpenes was not considered in this study. The inclusion of additional trace gases and aerosols apportioned up to 25% of total ozone to this factor, which could be due to correlation of both variables with temperature and radiation in the course of the year.

3.2.2 Short-lived incomplete combustion sources

Factors II and III both showed a large contribution of short-lived substances. Factor II contained rather short-lived alkenes such as ethene (mixing ratio: 133 pptv, fractions of the substance attributed to this factor: 27%), propene (46 pptv, 79%), 1,3-butadiene (6 pptv, 90%), and some butenes, typical for incomplete combustion processes. The average lifetime of Factor II was 1.6 days, calculated from annual mean OH and ozone and factor loading-weighted mean lifetime. This factor is attributed to short-lived combustion sources mainly from vehicular exhaust.

Literature profiles of vehicle emissions show similarly high contributions of ethene and propene (e.g. Badol et al., 2008; Friedrich and Obermeier, 1999; Hellen et al., 2003; Liu et al., 2008; Pang et al., 2014; Sagebiel et al., 1996; Thijsse et al., 1999). Sauvage at al. (2009) provided a vehicle exhaust factor very similar for one of the French remote sites. Factor II correlated well with NO_2

(r=0.84), NO (0.75), NO_y (0.73), and CO (0.66), all associated with traffic emissions. Included in the PMF runs, only significant fractions of NO_2 (~10%) were apportioned to this factor.

This factor does not resemble vehicular emissions alone, since longer-lived combustion compounds such as benzene, acetylene, and aromatics contributed to Factor IV as described below.

3.2.3 Short- to medium-lived evaporative sources

Factor III explained large parts of the measured toluene (63 pptv, 55%), ethylbenzene (11 pptv, 60%), and *m*+*p*-xylene (23 pptv, 65%). Alkanes such as pentanes, hexanes, and heptane contributed to this factor with 40-60%, butanes with 20%. The lifetimes of the compounds were in the short- to medium-lived range of mostly 1-6 days. The aromatics were the only substances measured in this study that are found in evaporative solvent emission profiles, e.g. in paint or wood coating. The alkanes are found in the gasoline composition and evaporation source. This factor is thus further referred to as short- to medium-lived evaporative sources, additionally containing some proportions of short to medium-lived combustion compounds (ethane, acetylene, isopentane, alkenes, benzene and toluene), which also explains its correlation with combustion tracers NO_2 (r=0.74), NO (0.74), and NO_y (0.75). The largest fraction of NO_2 (> 70%) was abundant in a factor similar to this, when included in the PMF runs.

3.2.4 Residential heating and long-lived incomplete combustion sources

In comparison to Factors II and III, the compounds in Factors IV and V exhibited longer atmospheric lifetimes (21 days (IV) and 26 days (V)). Large parts of the alkynes acetylene (216 pptv, 49%) and propyne (8 pptv, 56%), as well as ethene (287 pptv, 58%) and benzene (65 pptv, 48%) were explained in Factor IV that also contained some C_7-C_8 aromatics (15-20%), and ethane and propane (16% each). This factor could be attributed to residential heating and wood burning and maybe other (incomplete) combustion processes, in particular from road traffic. The major compounds resembled in this factor had atmospheric lifetimes of 1-17 days. Only a few source profiles for domestic combustion emissions were found

in literature. A wood burning profile by Friedrich and Obermeier (1999), adjusted to the substances measured for this study, showed a similar composition to Factor IV, as does a "residential heating" profile by Klemp et al. (2002) and Mannschreck et al. (2002). Ethene and acetylene contributions of the literature profiles differed from those found in Factor IV. These differences indicated that not only wood burning but also other sources contributed to this factor. Similar profiles to Factor IV were found by Lanz et al. (2008) and Sauvage et al. (2009), who attributed the French profiles to hot water generation and building heating by burning fossil fuels and wood. The benzene/toluene ratio at those French sites was 2, in Mannschreck et al. (2002) 3.2, and here for factor IV 2.7. Evtyugina et al. (2014) determined similar benzene/toluene ratios between 1.7 and 5.0 for emission factors from different woods from fireplaces and woodstoves. Aromatics, ethene, and acetylene could, however, also be contributed from vehicle exhaust (e.g. Pang et al., 2014; Badol et al., 2008). Factor IV correlated well with the combustion tracer CO (r=0.87). A correlation to SO_2 (0.62) was found for this factor, which is reasonable, since household emissions contribute to SO_2 in the atmosphere (UBA, 2011). SO_2 out of the additional PMF runs was almost exclusively apportioned to this factor. Also the largest fractions of NO and all aerosols (> 80%) as well as amounts of CO (~20%) contributed to this factor, supporting residential heating and other long-lived combustion sources as the main contributors.

3.2.5 Long-lived gas leakage and evaporative sources

In Factor V alkanes from natural or liquid gas leakage and gasoline evaporation were the predominant substances, with ethane (331 pptv, 22%), propane (306 pptv, 54%), isobutane (75 pptv, 59%), *n*-butane (128 pptv, 61%), isopentane (37 pptv, 26%), *n*-pentane (38 pptv, 50%), *n*-hexane (9 pptv, 47%), and small amounts of acetylene and benzene (approx. 10% each) contributing to it. These were typical gas leakage and evaporative sources, but generally longer-lived (> 3 days, average 26 days) than in Factor III (< 6 days, average 4.9 days). Literature emission profiles for crude oil production, natural gas, gasoline, and LPG evaporation (e.g. Friedrich and Obermeier, 1999; Lanz et al., 2008; Leuchner and Rappenglück, 2010; Mannschreck et al., 2002; Thijsse et al., 1999; Watson et al., 2001) closely resembled this factor. Although emission profiles and PMF results

of urban datasets presented high isopentane/*n*-pentane ratios, Sauvage et al. (2009) found ratios close to 1 for remote sites in France for evaporative sources, similar to our study. Factor V can be viewed as an aged combined profile of evaporative losses of natural gas, gasoline, and LPG with atmospheric lifetimes of 4-60 days (C_2-C_4 alkanes). The correlation with CO (r=0.80) may on first view indicate a relation to fossil fuel burning, but it appears more likely that it is due to the similar (large) foot print areas associated with the similar lifetimes and the similar source areas which are mainly related to high population density.

3.2.6 Background sources

Factor VI can be apportioned to remote sources showing the continental background. It explained most of the measured ethane (881 pptv, 59%) and quite large amounts of benzene (42 pptv, 31%), acetylene (29%) and propane (25%). Ethane is the most abundant and longest-living compound measured in this study. The high abundance in combination with the lack of shorter-lived compounds were indicators for aged air masses (average lifetime of factor: 45 days). Propane, acetylene, and benzene also are very stable substances (lifetimes 11-17 days at annual average OH concentration of $9.4 * 10^5$ molec cm^{-3}) that underline the background character of this factor. Hellen et al. (2003) provided a measured profile of similar composition attributed to distant sources in Finland. Sauvage et al. (2009) identified a remote source factor with high ethane loadings and amounts of propane in France and Lanz et al. (2008) found an ethane factor in Switzerland. There is also very high resemblance with compound composition and mixing ratios, when compared to other remote data. Similar levels of ethane and propane as apportioned to this factor study were found at Pico Mountain, Azores, and also Mauna Loa, Hawaii, (Helmig et al., 2008) as well as in Mace Head, Ireland, (Yates et al., 2010) from ground-based measurements. Aircraft data from the central North Atlantic (Lewis et al., 2007) showed a similar average profile from all campaigns to our PMF resolved data. The mixing ratios of the most important substances in this factor from our study and the one from Lewis et al. (2007) are for ethane 881 (this study) and 870 pptv (Lewis et al, 2007), propane 141 and 100 pptv, acetylene 126 and 110 pptv, benzene 42 and 38 pptv, and toluene 16 and 10 pptv, respectively. Factor VI did not correlate well with any of the additional trace gases. More than 50% of CO and the largest part of

601 ozone (> 60%) were apportioned to this factor from the PMF model, when
602 included.
603
604 ### 3.3 Seasonality and total contributions
605 The winter/summer amplitude was already used for better attribution of source
606 categories to the individual factors in the previous sections (Fig. 4). The total
607 monthly variation of factorial NMHC mixing ratios is summed up in Fig. 5. The
608 maximum was encountered in February (7.1 ppbv), the minimum in July (1.8
609 ppbv). During fall and winter, chemical reaction rates decreased due to lower OH
610 concentrations as a consequence of lower UV light available and lower
611 temperatures. Thus, the NMHC were not depleted as rapidly as during the
612 summer months.
613 Within this dataset, biogenic sources only contributed to the measured and
614 modeled NMHC in summer, but are responsible for 20% of the total modeled
615 amount of hydrocarbons in July. Due to the high reactivity of isoprene, this
616 overall small factor plays an important role in atmospheric processes during
617 summertime.
618 In section 2.4, we calculated the expected seasonal variation of the individual
619 NMHC within a simple box model with constant emissions over the year and
620 losses due to measured annual cycles of OH radicals and ozone. In order to
621 analyze the impact of different sources on the seasonality, we now calculated the
622 expected seasonal variation of the factors derived in this PMF in the constant
623 source scenario by combining the calculated seasonality of the individual NMHC
624 weighted by their factor loadings. Thus determined winter/summer ratios are
625 compared to the PMF derived seasonal variations in Tab. 4. Larger
626 winter/summer ratios than in the constant source scenario indicate stronger
627 sources in winter than in summer, smaller indicate stronger sources in summer.
628 Similar winter/summer ratios existed for the short lived combustion (II), and the
629 long-lived gas leakage and evaporative (V) sources pointing towards constant
630 emissions by these sources over the year. This is reasonable since Factor II is
631 related mainly to traffic exhaust and Factor V to natural gas and LPG losses, and
632 gasoline evaporation emissions, both expected to be fairly constant throughout the
633 year. Factor III, short lived evaporative, is associated to temperature driven
634 emissions and accordingly expected to be stronger in summer, as seen by the

observed ratio of 3.6 compared to the constant emission box calculation with a ratio of 9.5. Contrary, the factor associated with residential heating has twice as high emissions in winter than in summer (15.8) compared to a constant source scenario with winter/summer ratio of 7.1. The background Factor VI indicated slightly higher emissions in summer than in the constant source scenario (ratios of 3.1 compared to 5.0), which possibly might be due to biomass burning contributions to ethane which are higher in summer, however, also inter-hemispheric transport reduces the winter maxima of these long-lived gases (Rudolph, 1995).

Though the seasonal comparison to our simple box calculations showed reasonable results, it should be pointed out that this approach is rather simplistic and might not appropriately describe the situation at Hohenpeissenberg. These include, among others, different footprint areas for compounds of different lifetime and in different seasons, the representativeness of OH and ozone for such different footprint areas, and the seasonally changing mixed-layer height and intensity of vertical mixing. However, such factors are considered in this first approximation of minor importance on the seasonal variability compared to the changing photochemical loss and changing regional emissions.

Over the entire seven year period background sources contributed most to the NMHC volume (Fig. 6a) and mass at the remote site (31.3 vol%, 25.8 mass%) followed by gas leakage/long-lived evaporative sources (24.6 vol%, 25.4 mass%), residential heating/long-lived combustion sources (23.4 vol%, 21.0 mass%), short-lived evaporative sources (11.3 vol%, 17.8 mass%), short-lived combustion sources (6.0 vol%, 5.9 mass%), and the biogenic sources (3.5 vol%, 4.0 mass%). Weighted with the corresponding OH reaction rate constants, the mean OH-reactivity of the factors was essentially vice versa (Fig. 6b): biogenic – 24.2%, short-lived combustion – 22.5%, residential heating – 20.6%, short-lived evaporative – 15.1%, background sources – 9.0%, long-lived evaporative – 8.5%. Despite low volume and mass fractions the biogenic isoprene factor has the highest contribution for the total reactivity and ozone formation. The impact is even higher during maximum emissions in the summer.

4 Conclusions and Outlook

Positive Matrix Factorization (PMF) was used for characterizing the impact of source categories and photochemical aging on this dataset of NMHC measurements at the remote GAW site Hohenpeissenberg. This new approach does not aim for strict source apportionment and consequently does not require mass conservation or explicit treatment of photochemical aging in the factor profiles, however, on the cost of losing a quantitative understanding of the source contributions. For the determination of the number of factors, statistics on the provided results were not decisive. Interpretability of the computed factors was the most important parameter in the analysis and interpretation of the results. Treatment of missing values and uncertainties had no substantial influence on the solutions. It could be interesting for other datasets to evaluate the number of missing values that the model can compensate for before an effect on the result is seen.

The stability of the modeled factors depended on the individual uncertainties of the contributing substances. Especially short-lived Factors II (incomplete combustion) and III (evaporative) showed lower stability in the bootstrap runs performed by PMF. These two factors were the ones that contained compounds with high reactivity and thus higher variability at the receptor site. The other factors could be attributed to biogenic sources, residential heating including long-lived incomplete combustion, long-lived gas leakages/evaporative sources, and sources that reflect the continental background. The measured anthropogenic NMHC have maxima in winter when photochemical removal by OH or ozone is lowest. Compared to simple box calculations assuming constant emissions, it could be demonstrated that the short-lived combustion factor (II) and the gas leakage/long-lived evaporative factor (V) indicated constant emissions over the year as expected. The short-lived evaporative (III) emissions are higher in summer by factor 2-3 in line with higher temperatures, the residential heating and wood burning (IV) factor indicated about twofold higher emissions in winter, and the background (VI) was smoother than the constant source scenario, indicating additional sources in summer, like biomass burning, or enhanced losses due to inter-hemispheric transport in northern-hemispheric winter. The overall influence of biogenic isoprene sources (I) and other short-lived factors (II and III) on reactivity is substantial. In terms of reactivity and chemical processes like ozone production, these factors dominate with on average more than 60% compared to

their low volume contribution of less than 20%. When considering other biogenic emissions such as monoterpenes, sesquiterpenes and OVOC, the impact of the biogenic factor on chemical processes in summer during daytime will even increase.

Fractions of the factor composition profiles were often in agreement with source profiles from literature and calculated results from PMF studies in urban areas, some of the factors of this study even showed high agreement with those from other studies. However, some of the literature source profiles exhibit large uncertainties and only low substance resolution, thus the comparisons to these profiles leave a lot of room for interpretation (Theloke and Friedrich, 2007). Improvement and updates of emission profiles, particularly in regard to new legal limitations of emissions, are highly needed.

Since Yuan et al. (2012) emphasized the non-negligible influence of photochemical aging on PMF results at urban sites, the pronounced inclusion of aging into the interpretation of the factors (in this case Factors II-V) rather than a typical source apportionment seems necessary for remote sites that cannot fulfill the assumption of mass conservation during transport. Factors II and IV were thus attributed to incomplete combustion originating mainly from vehicular exhaust and residential heating (lead components: light alkenes, alkynes, benzene), and Factors III and V to evaporative losses from fuel (including fuel not burned e.g. during cold-start), natural gas and solvents as they predominantly contain alkanes and aromatics. Differences between Factors II and IV, and between III and V, respectively, were the different life-times of the compounds of on average 1.6 days (II) and 21 days (IV), and 4.9 days (III) and 26 days (V), at annual average OH and ozone. In view of interpreting patterns in the atmosphere, the regional background is resembled by the longer-lived patterns (IV and V) and fresh, local impacts by the shorter lived patterns (II and III), the superposition reflects the different degrees of aging and mixing. This interpretation with respect to only two lumped source categories is plausible in the light of mainly two factors: measurements were made at times of well mixed conditions (13:00 h CET) and local, more pronounced impacts of individual sources were damped. Secondly, emissions are usually correlated with population density for major source types especially in the rural surroundings of Hohenpeissenberg with no major local sources but multiple, wide-spread small sources, e.g. the closest motorway is

some 30 km east. Thus, signatures of different individual sources were not expected to show up in different air masses in a pronounced way.

The factors resolved for the Hohenpeissenberg dataset were very similar to those found by Sauvage et al. (2009) for remote sites in France; PMF seems to be able to calculate reasonable results for reactive species without including reactivity into the uncertainty for the PMF model. It should be emphasized that the factors are not expected to resemble emission profiles but nevertheless are related to emission categories, and they offer a good tool to characterize source impacts and remoteness of different stations and regions.

In terms of implementing photochemical processing by adding to uncertainties or applying photochemical age-based parameterization, neither did the method suggested by Sauvage et al. (2009) improve the interpretability, nor was the photochemical age-based parameterization method suggested by Yuan et al. (2012) well applicable for a site like Hohenpeissenberg. The most important underlying restriction for the application of the latter method to our dataset is that we cannot calculate photochemical age reliably from the scaling of the ratio of two compounds with different reactivity.

With the inclusion of nighttime data PMF still resolved the same six factors 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. Using another receptor model like CMB or UNMIX on this dataset could further confirm the resolved factors.

Acknowledgements

We gratefully acknowledge the technical work by Katja Michl and Erasmus Tensing in set-up of instruments, measurements, and data evaluation. We further thank Stefan Gilge and the team at the GAW station Hohenpeissenberg for non-organic trace gas data and Tim Sparks for editing the language of our manuscript. This work was supported by the German Research Foundation (DFG) and the Technische Universität München within the funding program Open Access Publishing.

References

Anderson, M.J., Daly, E.P., Miller, S.L., and Milford, J.B.: Source apportionment of exposures to volatile organic compounds: II. Application of receptor models to TEAM study data, Atmos. Environ., 36, 3643-3658, 2002.

Atkinson, R.: Atmospheric chemistry of VOCs and NO_x, Atmos. Environ., 34, 2063-2101, 2000.

Atkinson, R.: Our present understanding of the gas-phase atmospheric degradation of VOCs, in: Simulation and Assessment of Chemical Processes in a Multiphase Environment, edited by Barnes, I. and Kharytonov, M.M., 1-19, Springer, Netherlands, Dordrecht, 2008.

Atkinson, R. and Arey, J.: Atmospheric degradation of volatile organic compounds, Chem. Rev., 103, 4605-4638, 2003.

Badol, C., Locoge, N., and Galloo, J.C.: Using a source-receptor approach to characterise VOC behaviour in a French urban area influenced by industrial emissions. Part II: Source contribution assessment using the chemical mass balance (CMB) model, Sci. Total. Environ., 389, 429-440, 2008.

Berresheim, H., Elste, T., Plass-Dülmer, C., Eiselebb, F.L., Tanner, D.J.: Chemical ionization mass spectrometer for long-term measurements of atmospheric OH and H_2SO_4, Int. J. Mass Spectrom., 220, 91-109, doi:10.1016/S1387-3806(00)00233-5, 2000.

Borbon, A., Coddeville, P., Locoge, N., and Galloo, J.C.: Characterising sources and sinks of rural VOC in eastern France, Chemosphere, 57, 931-942, 2004.

Brown, S.G., Frankel, A., and Hafner, H.R.: Source apportionment of VOCs in the Los Angeles area using positive matrix factorization, Atmos. Environ., 41, 227-237, 2007.

Buzcu, B. and Fraser, M.P.: Source identification and apportionment of volatile organic compounds in Houston, TX, Atmos. Environ., 40, 2385-2400, 2006.

Chameides, W.L., Lindsay, R.W., Richardson, J., and Klang, C.S.: The Role of Biogenic Hydrocarbons in Urban Photochemical Smog: Atlanta as a Case Study, Science, 241, 1473–1475, doi:10.1126/science.3420404, 1988.

Chan, T.W. and Mozurkewich, M.: Application of absolute principal component analysis to size distribution data: identification of particle origins, Atmos. Chem. Phys., 7, 887-897, 2007.

de Gouw, J.A., Middlebrook, A.M., Warneke, C., Goldan, P.D., Kuster, W.C., Roberts, J.M., Fehsenfeld, F.C., Worsnop, D.R., Canagaranta, M.R., Pszenny, A.A.P., Keene, W.C., Marchewka, M., Bertman, S.B., and Bates, T.S.: Budget of organic carbon in a polluted atmosphere: Results from the New England Air Quality Study in 2002, J. Geophys. Res., 110, D16305, 1-22, doi:10.1029/2004JD005623, 2005.

European Commission: Annexes to the Proposal for a Directive of the European Parliament and of the Council on the reduction of national emissions of certain atmospheric pollutants and amending Directive 003/35/EC, Brussels, 18.12.2013, http://ec.europa.eu/environment/air/pdf/com2013_920/COM_2013_920_F1_ANNEX_EN.pdf, 2013.

Evtyugina, M., Alves, C., Calvo, A., Nunes, T., Tarelho, L., Duarte, M., Prozil, S.O., Evtuguin, D.V., and Pio, C.: VOC emissions from residential combustion of Southern and mid-European woods, Atmos. Environ. 83, 90–98. doi:10.1016/j.atmosenv.2013.10.050, 2014.

Fall, R.: Biogenic Emissions of Volatile Organic Compounds from Higher Plants, in: Reactive Hydrocarbons in the Atmosphere, edited by Hewitt, N., Academic Press, London, 41-95, 1999.

Friedrich, R. and Obermeier, A.: Anthropogenic emissions of volatile organic compounds, in: Reactive Hydrocarbons in the Atmosphere, edited by Hewitt, N., Academic Press, London, 1-39, 1999.

Fuentes, J.D. and Wang, D.: On the seasonality of isoprene emissions from a mixed temperate forest, Ecol. Appl., 9, 1118-1131, 1999.

Guo, H., Wang, T., and Louie, P.K.K.: Source apportionment of ambient non-methane hydrocarbons in Hong Kong: Application of a principle component analysis/absolute principle component scores (PCA/APCS) receptor model, Environ. Pollut., 129, 489-498, 2004.

Guo, H., Wang, T., Blake, D.R., Simpson, I.J., Kwok, Y.H., and Li, Y.S.: Regional and local contributions to ambient non-methane volatile organic compounds at a polluted rural/coastal site in Pearl River Delta, China, Atmos. Environ., 40, 2345-2359, 2006.

Gauss, M., Myhre, G., Pitari, G., Prather, M.J., Isaksen, I.S.A., Berntsen, T.K., Brasseur, G.P., Dentener, F.J., Derwent, R.G., Hauglustaine, D.A., Horowitz, L.W., Jacob, D.J., Johnson, M., Law, K.S., Mickley, L.J., Müller, J.F., Plantevin, P.H., Pyle, J.A., Rogers, H.L., Stevenson, D.S., Sundet, J.K., van Weele, M., and Wild, O.: Radiative forcing in the 21st century due to ozone changes in the troposphere and the lower stratosphere, J. Geophys. Res., 108, 4292, doi:10.1029/2002JD002624, 2003.

Gilge, S., Plass-Duelmer, C., Fricke, W., Kaiser, A., Ries, L., Buchmann, B., and Steinbacher, M.: Ozone, carbon monoxide and nitrogen oxides time series at four alpine GAW mountain stations in central Europe, Atmos. Chem. Phys., 10, 12295-12316, doi:10.5194/acp-10-12295-2010, 2010.

Hellen, H., Hakola, H., and Laurila, T.: Determination of source contributions of NMHCs in Helsinki (60°N, 25°E) using chemical mass balance and the Unmix multivariate receptor models, Atmos. Environ., 37, 1413-1424, 2003.

Helmig, D., Tanner, D.M., Honrath, R.E., Owen, R.C., and Parrish, D.D.: Nonmethane hydrocarbons at Pico Mountain, Azores: 1. Oxidation chemistry in the North Atlantic region, J. Geophys. Res., 113, 1-16, D20S91, doi:10.1029/2007JD008930, 2008.

Hoerger, C. C., Werner, A., Plass-Duelmer, C., Reimann, S., Eckart, E., Steinbrecher, R., Aalto, J., Arduini, J., Bonnaire, N., Cape, J. N., Colomb, A., Connolly, R., Diskova, J., Dumitrean, P., Ehlers, C., Gros, V., Hakola, H., Hill, M., Hopkins, J. R., Jäger, J., Junek, R., Kajos, M. K., Klemp, D., Leuchner, M., Lewis, A. C., Locoge, N., Maione, M., Martin, D., Michl, K., Nemitz, E., O'Doherty, S., Pérez Ballesta, P., Ruuskanen, T. M., Sauvage, S., Schmidbauer, N., Spain, T. G., Straube, E., Vana, M., Vollmer, M. K., Wegener, R., and Wenger, A.: ACTRIS non-methane hydrocarbon intercomparison experiment in Europe to support WMO-GAW and EMEP observation networks, Atmos. Meas. Tech. Discuss., 7, 10423-10485, doi:10.5194/amtd-7-10423-2014, 2014.

Hopke, P.K.: Recent developments in receptor modeling, J. Chemometr., 17, 255-265, 2003.

Jorquera, H. and Rappenglück, B.: Receptor modeling of ambient VOC at Santiago, Chile, Atmos. Environ., 38, 4243-4263, 2004.

Klemp, A., Mannschreck, K., Pätz, H.W., Habram, M., Matuska, and Slemr, F.: Determination of anthropogenic emission ratios in the Augsburg area from concentration ratios: results from long-term measurements, Atmos. Environ., 36, 61-80, 2002.

871 Lanz, V.A., Hueglin, C., Buchmann, B., Hill, M., Locher, R., Staehelin, J., and
872 Reimann, S.: Receptor modeling of C-2-C-7 hydrocarbon sources at an urban
873 background site in Zurich, Switzerland: changes between 1993-1994 and
874 2005-2006, Atmos. Chem. Phys., 8, 2313-2332, doi:10.5194/acp-8-2313-
875 2008, 2008.
876 Lanz, V.A., Henne, S., Staehelin, J., Hueglin, C., Vollmer, M.K., Steinbacher, M.,
877 Buchmann, B., and Reimann, S.: Statistical analysis of anthropogenic non-
878 methane VOC variability at a European background location (Jungfraujoch,
879 Switzerland), Atmos. Chem. Phys., 9, 3445-3459, doi:10.5194/acp-9-3445-
880 2009, 2009.
881 Lingwall, J.W. and Christensen, W.F.: Pollution source apportionment using a
882 priori information and positive matrix factorization, Chemometr. Intell. Lab.,
883 87, 281-294, 2007.
884 Liu, Y., Shao, M., Fu, L., Lu, L., Zeng, L., Tang, D.: Source profiles of volatile
885 organic compounds (VOCs) measured in China: Part I, Atmos. Environ., 42,
886 6247–6260, doi:10.1016/j.atmosenv.2008.01.070, 2008.
887 Leuchner, M. and Rappenglück, B.: VOC source-receptor relationships in
888 Houston during TexAQS-II, Atmos. Environ., 44, 4056-4067, 2010.
889 Lewis, A.C., Evans, M.J., Methven, J., Watson, N., Lee, J.D., Hopkins, J.R.,
890 Purvis, R.M., Arnold, S.R., McQuaid, J.B., Whalley, L.K., Pilling, M.J.,
891 Heard, D.E., Monks, P.S., Parker, A.E., Reeves, C.E., Oram, D.E., Mills, G.,
892 Bandy, B.J., Stewart, D., Coe, H., Williams, P., and Crosier, J.: Chemical
893 composition observed over the mid-Atlantic and the detection of pollution
894 signatures far from source regions, J. Geophys. Res., 112, D10S39,
895 doi:10.1029/2006JD007584, 2007.
896 Mannschreck, K., Klemp, A., Kley, D., Friedrich, R., Kühlwein, J., Wickert, B.,
897 Matuska, P., Habram, M., and Slemr, F.: Evaluation of an emission inventory
898 by comparisons of modelled and measured emission ratios of individual HCs,
899 CO and NOx, Atmos. Environ., 36, 81-94, 2002.
900 Mao, J., Ren, X., Chen, S., Brune, W.H., Chen, Z., Martinez, M., Harder, H.,
901 Lefer, B., Rappenglück, B., Flynn, J., and Leuchner, M.: Atmospheric
902 oxidation capacity in the summer of Houston 2006: Comparison with summer
903 measurements in other metropolitan studies. Atmos. Environ., 44, 4107-4115,
904 doi:10.1016/j.atmosenv.2009.01.013, 2010.
905 Matyssek, R., Wieser, G., Ceulemans, R., Rennenberg, H., Pretzsch, H., Haberer,
906 K., Löw, M., Nunn, A.J., Werner, H., Wipfler, P., Oßwald, W., Nikolova, P.,
907 Hanke, D.E., Kraigher, H., Tausz, M., Bahnweg, G., Kitao, M., Dieler, J.,
908 Sandermann, H., Herbinger, K., Grebenc, T., Blumenröther, M., Deckmyn,
909 G., Grams, T.E.E., Heerdt, C., Leuchner, M., Fabian, P., and Häberle, K.H.:
910 Enhanced ozone strongly reduces carbon sink strength of adult beech (*Fagus*
911 *sylvatica*) – Resume from the free-air fumigation study at Kranzberg Forest,
912 Environ. Pollut., 158, 2527-2532, 2010.
913 Miller, S.L., Anderson, M.J., Daly, E.P., and Milford, J.B.: Source apportionment
914 of exposures to volatile organic compounds: I. Evaluation of receptor models
915 using simulated exposure data, Atmos. Environ., 36, 3629-3641, 2002.
916 Na, K. and Kim, Y.P.: Chemical mass balance receptor model applied to ambient
917 C2–C9 VOC concentration in Seoul, Korea: Effect of chemical reaction
918 losses, Atmos. Environ., 41, 6715-6728, 2007.
919 Norris, G., Vedantham, R., Wade, K., Brown, S., Prouty, J., and Foley, C.: EPA
920 Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide, U.S.
921 Environmental Protection Agency,

http://www.epa.gov/heasd/documents/EPA_PMF_3.0_User_Guide.pdf (last access: 13 March 2014), 2008.
Nunn, A.J., Kozovits, A.R., Reiter, I.M., Heerdt, C., Leuchner, M., Lütz, C., Liu, X., Löw, M., Winkler, J.B., Grams, T.E.E., Häberle, K.H., Werner, H., Fabian, P., Rennenberg, H., and Matyssek, R.: Comparison of ozone uptake and sensitivity between a phytotron study with young beech and a field experiment with adult beech (*Fagus sylvatica*), Environ. Pollut., 137, 494-506, 2005.
Olson, D.A., Norris, G.A., Seila, R.L., Landis, M.S., and Vette, A.F.: Chemical characterization of volatile organic compounds near the World Trade Center: Ambient concentrations and source apportionment, Atmos. Environ., 41, 5673-5683, 2007.
Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemometr. Intell. Lab., 37, 23-35, 1997.
Paatero, P.: The Multilinear-Engine – A Table-Driven, Least Squares Program for Solving Multilinear Problems, Including the n-Way Parallel Factor Analysis Model, J. Comput. Graph. Stat., 1, 854-888, 1999.
Paatero, P. and Tapper, U.: Analysis of positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values, Environmetrics, 5, 111-126, 1994.
Pang, Y., Fuentes, M., Rieger, P.: Trends in the emissions of Volatile Organic Compounds (VOCs) from light-duty gasoline vehicles tested on chassis dynamometers in Southern California. Atmos. Environ., 83, 127–135. doi:10.1016/j.atmosenv.2013.11.002, 2014.
Parrish, D.D., Stohl, A., Forster, C., Atlas, E.L., Blake, D.R., Goldan, P.D., Kuster, W.C., and de Gouw, J.A.: Effects of mixing on evolution of hydrocarbon ratios in the troposphere, J. Geophys. Res.-Atmos., 112, doi: 10.1029/2006JD007583, 2007.
Plass-Dülmer, C., Michl, K., Ruf, R., and Berresheim, H.: C-2-C-8 hydrocarbon measurement and quality control procedures at the Global Atmosphere Watch Observatory Hohenpeissenberg, J. Chromatogr. A, 953, 175-197, 2002.
Plass-Dülmer, C., N. Schmidbauer, J. Slemr, F. Slemr, and H.D'Souza: European hydrocarbon intercomparison experiment AMOHA part 4: Canister sampling of ambient air, J. Geophys. Res., 111, D04306, doi:10.1029/2005JD006351, 2006.
Rappenglück, B., E. Apel, M. Bauerfeind, J. Bottenheim, P. Brickell, P. Cavolka, J. Cech, L. Gatti, H. Hakola, J. Honzak, R. Junek, D. Martin, C. Noone, C. Plass-Dülmer, D. Travers, and D. Wang: The first VOC intercomparison exercise within the Global Atmosphere Watch (GAW), Atmos. Environ. 40, 7508-7527, 2006.Rappenglück, B., Dasgupta, P., Leuchner, M., Li, Q., and Luke, W.: Formaldehyde and its relation to CO, PAN, and SO2 in the Houston-Galveston airshed, Atmos. Chem. Phys., 10, 2413-2424, doi:10.5194/acp-10-2413-2010, 2010.
Rudolph, J., The tropospheric distribution and budget of ethane, J. Geophys. Res. 100, 11269-11381, 1995.
Sagebiel, J.C., Zielinska, B., Pierson, W.R., and Gertler, A.W.: Real-world emissions and calculated reactivities of organic species from motor vehicles, Atmos. Environ., 30, 2287-2296, 1996.
Santoso, M., Hopke, P.K., Hidayat, A., and Diah, D.L.: Sources identification of the atmospheric aerosol at urban and suburban sites in Indonesia by positive matrix factorization, Sci. Total. Environ., 397, 229-237, 2008.

973 Sauvage, S., Plaisance, H., Locoge, N., Wroblewski, A., Coddeville, P., and
974 Galloo, J.C.: Long term measurement and source apportionment of non-
975 methane hydrocarbons in three French rural areas, Atmos. Environ., 43, 2430-
976 2441, 2009.
977 Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T.,
978 Müller, J.-F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic
979 VOC emissions calculated by the MEGAN model over the last 30 years.
980 Atmos. Chem. Phys., 14, 9317–9341, doi:10.5194/acp-14-9317-2014, 2014.
981 Spivakovsky, C.M., Logan, J.A., Montzka, S.A., Balkanski, Y.J., Foreman-
982 Fowler, M., Jones, D.B.A., Horowitz, L.W., Fusco, A.C., Brenninkmeijer,
983 C.A.M., Prather, M.J., Wofsy, S.C., and McElroy, M.B.: Three-dimensional
984 climatological distribution of tropospheric OH: Update and evaluation, J.
985 Geophys. Res.,-Atmos., 105, 8931-8980, 2000.
986 Stein, O. and Rudolph, J.: Modeling and interpretation of stable carbon isotope
987 ratios of ethane in global chemical transport models, J. Geophys. Res., 112,
988 D14308, 1-18, doi:10.1029/2006JD008062, 2007.
989 Stevenson, D.S., Doherty, R., Sanderson, M., Johnson, C.E., Collins, W.J., and
990 Derwent, R.G.: Impacts of climate change and variability on tropospheric
991 ozone and its precursors, Faraday Discuss., 130, 1-17, 2005.
992 Tauler, R., Viana, M., Querol, X., Alastuey, A., Flight, R.M., Wentzell, P.D., and
993 Hopke, P.K.: Comparison of the results obtained by four receptor modelling
994 methods in aerosol source apportionment studies, Atmos. Environ., 43, 3989-
995 3997, 2009.
996 Theloke, J. and Friedrich, R.: Compilation of a database on the composition of
997 anthropogenic VOC emissions for atmospheric modeling in Europe, Atmos.
998 Environ., 41, 4148-4160, 2007.
999 Thijsse, T.R., van Oss, R.F., and Lenschow, P.: Determination of source
1000 contributions to ambient volatile organic compound concentrations in Berlin,
1001 JAPCA J. Air Waste Ma., 1394-1404, 1999.
1002 UBA: Nationale Trendtabellen für die deutsche Berichterstattung atmosphärischer
1003 Emissionen seit 1990 (Stand: 15. April 2011), available at:
1004 http://www.umweltbundesamt-daten-zur-
1005 umwelt.de/umweltdaten/public/document/downloadPrint.do?ident=20681
1006 (last access: 13 March 2014), 2011.
1007 US-EPA: Research in Action – EPA Positive Matrix Factorization (PMF) Model,
1008 available at: http://www.epa.gov/heasd/research/pmf.html. (last access: 6
1009 August 2013), 2011.
1010 Vingarzan, R.: A review of surface ozone background levels and trends, Atmos.
1011 Environ., 38, 3431-3442, 2004.
1012 Watson, J.G., Chow, J.C., and Fujita, E.M.: Review of volatile organic compound
1013 source apportionment by chemical mass balance, Atmos. Environ., 35, 1567-
1014 1584, 2001.
1015 WDCGG: World Data Centre for Greenhouse Gases: Hohenpeissenberg – DWD,
1016 available at: http://ds.data.jma.go.jp/gmd/wdcgg/cgi-
1017 bin/wdcgg/accessdata.cgi?cntry=Germany&index=HPB647N00-
1018 DWD&para=VOCs&param=200612120859&select=parameter&parac=proce
1019 ssing (last access: 13 March 2014), 2013.
1020 Willis, R.D.: Final Report. Workshop on UNMIX and PMF as applied to $PM_{2.5}$,
1021 Research Triangle Park, NC, 14-16 February 2000, 2000.
1022 WMO: GAW Report No. 172, WMO Global Atmosphere Watch (GAW)
1023 Strategic Plan: 2008–2015, 2007.

1024 Yates, E.L., Derwent, R.G., Simmonds, P.G., Greally, B.R., O'Doherty, S., and
1025 Shallcross, D.E.: The seasonal cycles and photochemistry of C_2 - C_5 alkanes at
1026 Mace Head, Atmos. Environ., 44, 2705-2713, 2010.
1027 Yuan, B., Shao, M., de Gouw, J., Parrish, D.D., Lu, S., Wang, M., Zeng, L.,
1028 Zhang, Q., Song, Y., Zhang, J., and Hu, M.: Volatile organic compounds
1029 (VOCs) in urban air: How chemistry affects the interpretation of positive
1030 matrix factorization (PMF) analysis, J. Geophys. Res., 117, D24302, 1-17,
1031 doi:10.1029/2012JD018236, 2012.
1032 Yuan, B., Hu, W.W., Shao, M., Wang, M., Chen, W.T., Lu, S.H., Zeng, L.M., and
1033 Hu, M.: VOC emissions, evolutions and contributions to SOA formation at a
1034 receptor site in eastern China, Atmos. Chem. Phys., 13, 8815-8832,
1035 doi:10.5194/acp-13-8815-2013, 2013.
1036 Yue, W., Stolzel, M., Cyrys, J., Pitz, M., Heinrich, J., Kreyling, W.G., Wichmann,
1037 H.E., Peters, A., Wang, S., and Hopke, P.K.: Source apportionment of
1038 ambient fine particle size distribution using positive matrix factorization in
1039 Erfurt, Germany, Sci. Total. Environ., 398, 133-144, 2008.
1040
1041

Tables

Table 1. Different treatments of missing values, values below the detection limit, and zero values.

#	Treatment
1	missing values replaced by species median; corresponding uncertainties set to four times the species median
2	missing values replaced by geometric mean; corresponding uncertainties set to four times the geometric mean
3	additional to 2: values below detection limit (dl) set to half the detection limit; corresponding uncertainties set to detection limit
4	additional to 3: zero values set to half the detection limit
5	additional to 3: uncertainty of n-hexane increased by 20%

1047 Table 2. Observed winter/summer ratios for the considered NMHC compared to
1048 ratios obtained from box calculations using constant emissions.

in pptv	winter/summer (observed)	box results 95-/5-percentile
Ethane	2.7	3.7
Ethene	5.5	9.0
Propane	4.7	7.9
Propene	4.9	7.3
Isobutane	4.2	9.5
Acetylene	5.7	6.9
n-butane	4.6	9.8
1-butene	3.9	7.1
Isobutene	1.5	7.6
cis-2-butene	2.7	3.3
Isopentane	2.1	10.1
n-pentane	2.6	10.8
Propyne	6.1	10.6
1,3-butadiene	5.0	9.1
2-methylpentane	2.6	10.5
3-methylpentane	2.9	10.6
n-hexane	3.3	10.8
Isoprene	0.0	8.7
n-heptane	3.0	11.0
Benzene	4.8	7.8
Toluene	3.0	9.5
Ethylbenzene	2.9	10.8
m+*p*-xylene	4.1	11.2

1049

1050 Table 3. Mathematical diagnostics for the results of PMF computations for
1051 different numbers of factors.

	number of factors (k)			
	5	6	7	8
number of samples	2,335	2,335	2,335	2,335
number of species	24	24	24	24
Q-value from PMF	150,678	117,832	94,444	75,804
Fpeak	0.0	0.0	0.0	0.0
mean ratio NMHC modeled / NMHC measured	0.97	0.97	0.98	0.98
mean coefficient of determination (R^2)	0.86	0.88	0.89	0.89
number of species unexplained > 25 %	5	3	3	3
scaled residuals beyond 3 σ	4,354	2,937	1,971	1,391

1052
1053
1054

Table 4. Comparison of the calculated winter/summer ratios from the box model with constant emissions, and the observed winter/summer ratios for the PMF-factors.

	II: short-lived combustion	III: short-lived evaporative	IV: residential heating / long-lived combustion	V: gas leakage / long-lived evaporative	VI: background
winter/summer (calculated for constant emissions)	8.7	9.5	7.1	7.1	5.0
winter/summer (observed)	9.0	3.6	15.8	7.0	3.1

Figures

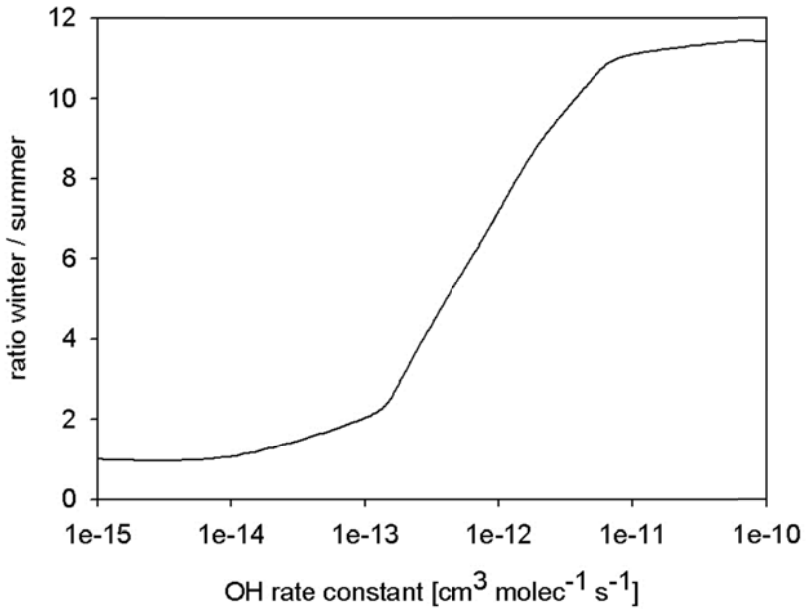


Figure 1. The expected winter/summer ratios (95-/5-percentile) as function of the OH rate constants for compounds with assumed constant sources in a simplified box-model and the mean OH profile determined at Hohenpeissenberg.

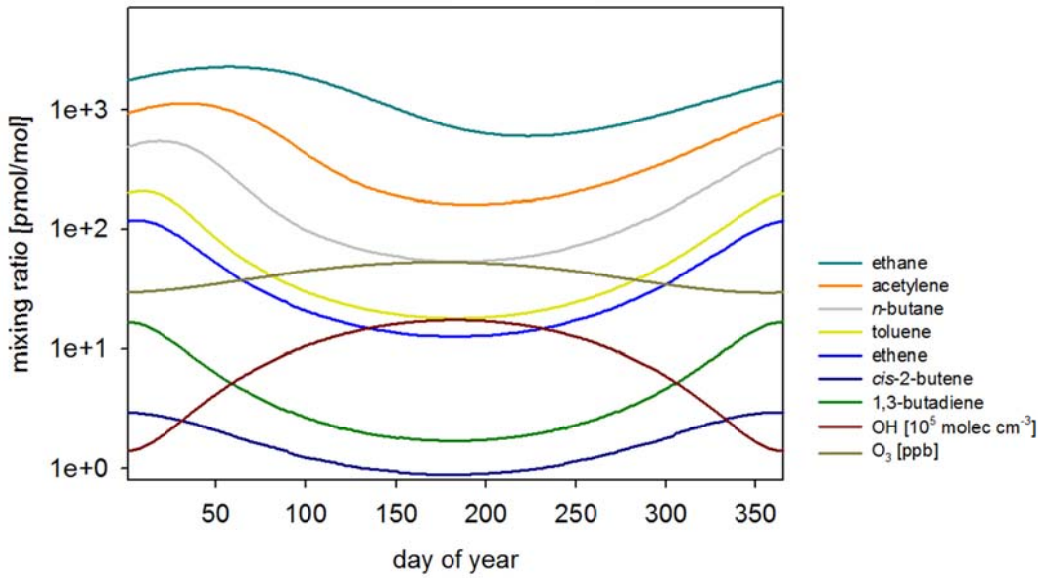
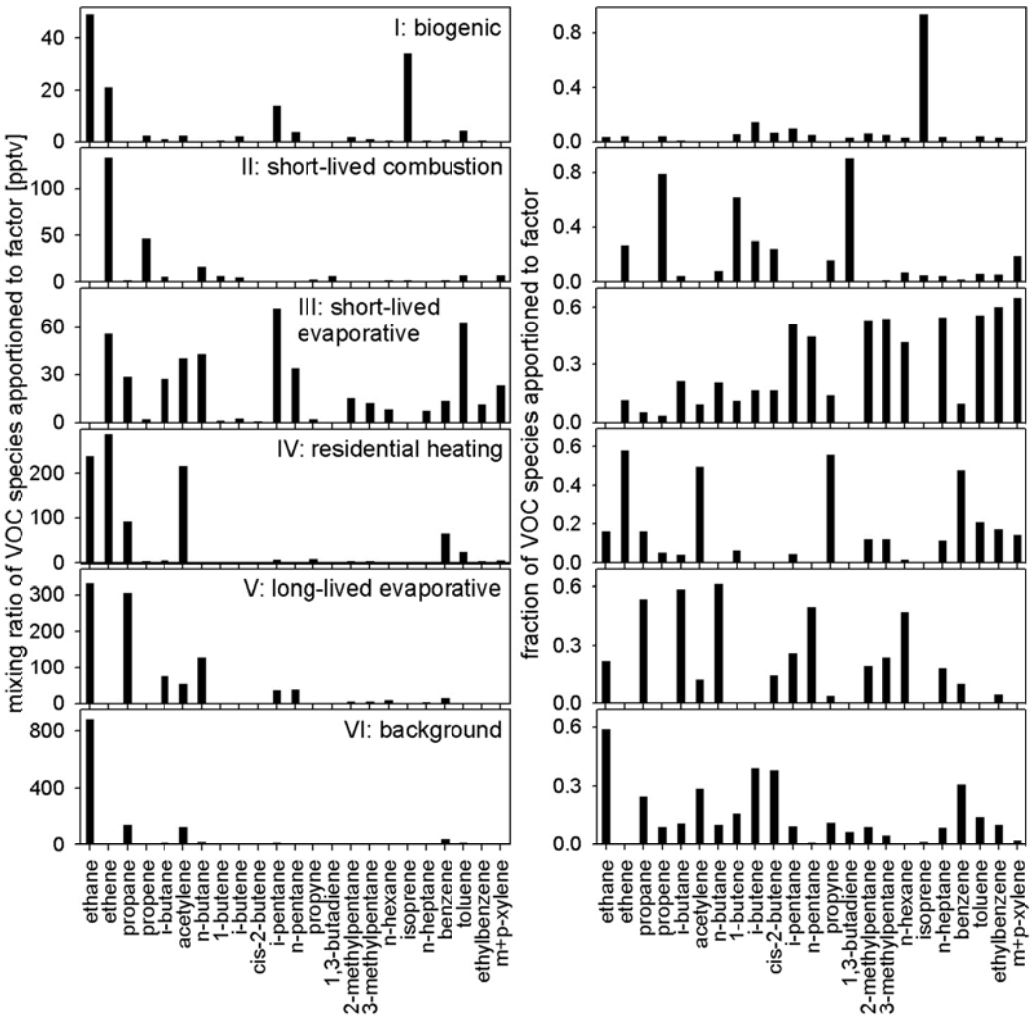


Figure 2. The calculated seasonal cycles of different NMHC assuming constant source of 5 * 10^3 molec cm^{-3} s^{-1} and the average annual cycles of OH and ozone as measured at Hohenpeissenberg.



1072

1076 Figure 3. Factor profiles for the six factor solution calculated by PMF, left:
1077 mixing ratio of each species apportioned to each factor [pptv], right: contribution
1078 of each factor to the species. Note that the scales for each subplot are different due
1079 to large variations in absolute mixing ratios.

1077

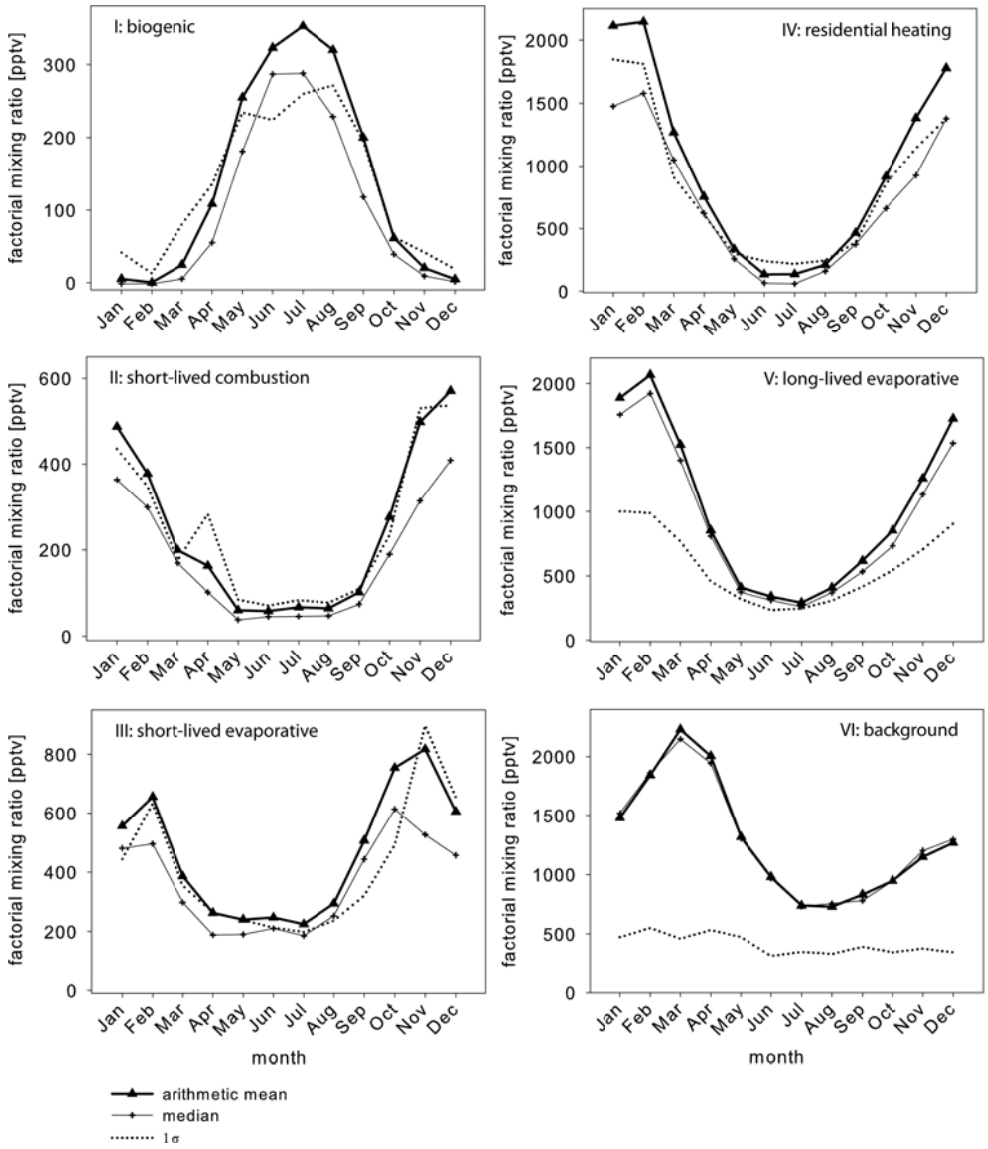


Figure 4. Annual pattern of the mean, median, and standard deviation of the factorial mixing ratio [pptv] of the respective source category.

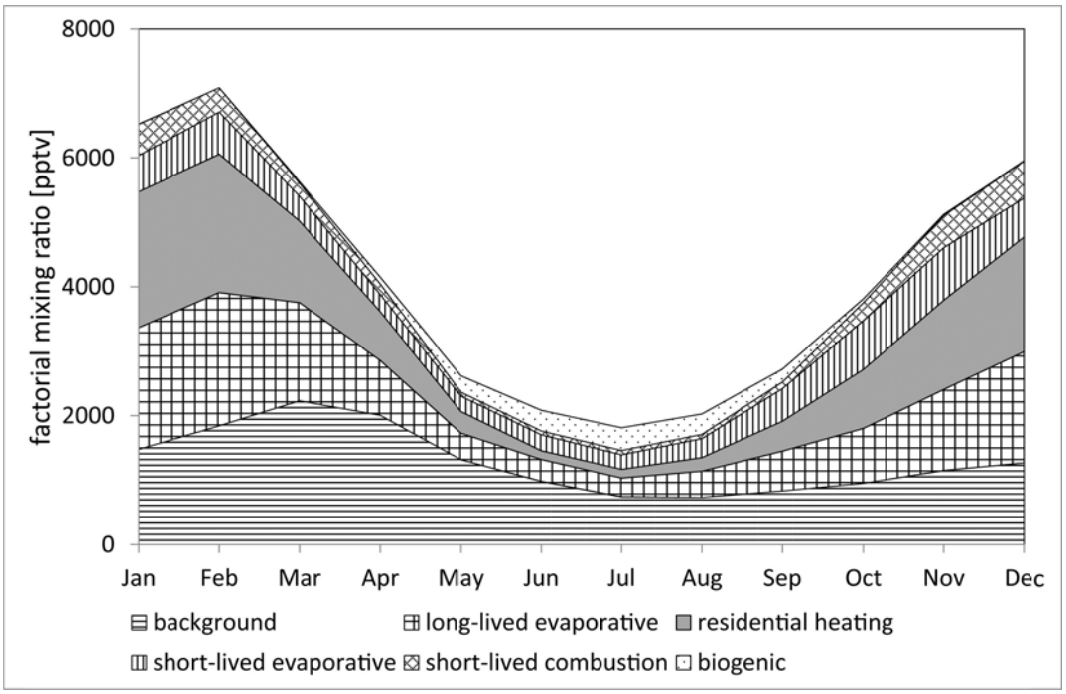
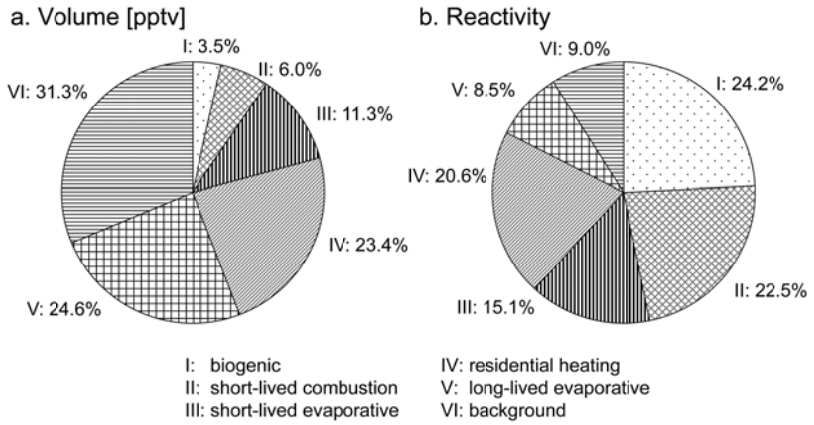


Figure 5. Mean monthly variability of the contributions from the six factors.



1089

1092 Figure 6. Contributions of individual factors, left: to the total amount of modeled
1093 NMHC [pptv], right: to the total OH reactivity of modeled NMHC [ppt cm^3
1094 $molec^{-1}$ s^{-1}].

1093