Final Author response

Response to Anonymous Referee #1

Reviewer comment:
1 Overall comment
This study focuses on the source apportionment of VOCs measurements at a suburban site in the North-West Indo-Gangetic Plain. The period studied is the month of May 2012. Authors use a Positive Matrix Factorization Model (PMF) to resolve source contributions to VOCs, perform a conditional probability functional analysis to locate the different sources and calculate the ozone and secondary organic aerosols formation potential. Moreover, results of PMF are compared with the source apportionment of three different emission inventory estimates.
Overall, the analysis performed is interesting and valuable. However, the manuscript needs improvements in the logical framing of the work with respect to its contribution and implications to the field. Also the introduction and the results need to be improved in this sense. I recommend publication after the authors have addressed the following substantive concerns/comments on their manuscript.

Author response: We sincerely thank the reviewer for his encouragement and the in-depth comments and suggestions which have greatly improved the clarity of the manuscript and have helped us to emphasize the implications of this study to the field more clearly.
The detailed response to each comment and changes made in the manuscript are listed below.

2 Major comments
1. ABSTRACT - the abstract is a bit too technical. I recommend to focus more on the big picture and major findings and implication of the paper (as outlined in the conclusions).
Author response: We appreciate this feedback and have revised the abstract in accordance with it. Revised abstract reads as follows:

Changes in the manuscript:
"In this study we undertook quantitative source apportionment for 32 volatile organic compounds (VOCs) measured at a suburban site in the densely populated North-West Indo-Gangetic Plain using the US EPA PMF 5.0 Model. Six sources were resolved by the PMF model. In descending order of their contribution to the total VOC burden these are "biofuel use and waste disposal" (23.2%), "wheat-residue burning" (22.4%), "cars" (16.2%), "mixed daytime sources" (15.7%), "industrial emissions and solvent use" (11.8%) and "two-wheelers" (8.6%).
Wheat residue burning is the largest contributor to the total ozone formation potential (26.2%). For the emerging contaminant isocyanic acid, photochemical formation from precursors (37%) and wheat residue burning (25%) were the largest contributors to human exposure. Wheat residue burning was also the single largest source of the photochemical precursors of isocyanic acid, namely, formamide, acetamide and propanamide, indicating that this source must be most urgently targeted to reduce human concentration exposure to isocyanic acid in the month of May. Our results highlight that for accurate air quality forecasting and modelling it is essential that emissions are attributed only to the months in

which the activity actually occurs. This is important both for emissions from crop residue burning (which occur in May and from Mid-October to the end of November).
The SOA formation potential is dominated by "cars" (36.9%) and "two-wheelers" (21.1%), which also jointly account for 47% of the human class I carcinogen benzene in the PMF model. This stands in stark contrast to various emission inventories which estimate the transport sector contribution to the benzene exposure as (~10%) and consider residential biofuel use, agricultural residue burning and industries to be more important benzene sources. Overall it appears that none of the emission inventories represent the regional emissions in an ideal manner. Our PMF solution suggests that transport sector emissions may be underestimated by GAINSv5.0 and EDGARv4.3.2 and overestimated by REASv2.1, while the combined effect of residential biofuel use and waste disposal emissions as well as the VOC burden associated with solvent use and industrial sources may be overestimated by all emission inventories. The agricultural waste burning emissions of some of the detected compound groups (ketones, aldehydes and acids) are missing in the EDGARv4.3.2 inventory."

Reviewer comment: 2. INTRODUCTION - the introduction should better frame the background of the study, its motivation and what is the new contribution of the work. In particular:
• Only one source receptor modelling study that has been cited is in the region of the study (Srivastava et al., 2005). Are there any source receptor modelling or more general studies that focus on VOCs over the IGP? If yes, they should be acknowledged. If no, this should be underlined.

Author response: We have cited all the source receptor modelling studies for VOC performed in India that are available in the peer reviewed literature. Beyond the studies we cited, there is one PMF study from the Kathmandu valley in Nepal (Sarkar et al. 2017) which represents a different environment and one attempted PMF study from the Eastern Himalayas (Sarkar et al. 2014, acpd-14-32133-2014), which did not make it into ACP. We have now made it more clear that no other VOC Source apportionment study in the IGP exists.
Changes in the manuscript: We inserted the following sentence after line 12 on page 2
"The only other source receptor modelling study in South Asia was conducted using a positive matrix factorisation model (EPA PMF5.0) with data collected in the Kathmandu valley, Nepal, as part of the SUSKAT campaign and attributed a negligible fraction of the anthropogenic VOC burden to residential biofuel usage (~14%). Instead different industrial sources including brick kilns (jointly 52%) and the transport sector (21%) were identified as the dominant VOC sources in the Kathmandu valley."

Reviewer comment: VOCs source apportionment estimates for the region under study are presented for different emissions inventories. However, it is claimed 'Considering the large discrepancies between bottom up inventories and different source receptor modelling studies', when 2/3 source receptor models studies presented so far are out of the understudied region. This claim need to be justified, or more appropriate studies need to be cited.
Author response: We agree with the reviewer that comparing performance of emissions inventories for the entire NW-IGP with a source receptor model study conducted in a specific megacity could introduce a bias. Since we have no other source receptor studies to fall back on, we have now made the comparison between the inventory and the source receptor modelling studies more specific and spatially accurate limiting it to the Delhi National Capital Region, Greater Mumbai and Greater Kolkata, when comparing with the source receptor

modelling studies of these cities, respectively and have revised the text from line 5-11 accordingly.

In response to a later comment by the same reviewer about the mismatch between the emission inventory year and the year of our study we have now substituted EDGAR v4.2 with the latest EDGAR v4.3.2 version (Huang et al. Atmos. Chem. Phys., 17, 7683–7701, 2017)**,** with emissions for the year 2012, which appears to represent a significant improvement over the previous version. We have not included MIX Asia (Li et al. Atmos. Chem. Phys., 17, 935–963, 2017) since the NMVOC data of this inventory for India has been sourced from REASv2.1 without any changes. As a result of this update we have also revised the text of lines 11-17 in the introduction (not just the results and discussion section 3.8 and Figure 8).

Changes in the manuscript: The revised paragraphs now read:

"Previous source receptor modelling studies of VOC emission from India (Srivastava, 2004; Srivastava et al., 2005; Majumdar et al., 2009) produced results that conflicted strongly with the bottom up emission inventories, all of which contain significant emissions from residential fuel usage even when filtered for the New Delhi National Capital Region (41-45%), Greater Mumbai (32-36%) and Greater Kolkatta (33-59%). Transport sector emissions, according the bottom up emission inventories contribute only 15-35%, 17-43% and 6-14% to the total VOC emissions in New Delhi National Capital Region, Greater Mumbai and Greater Kolkatta, respectively. All previous studies from India employed a chemical mass balance (CMB) technique for ambient VOC source attribution and identified the transport sector as the main source of NMVOCs in the form of evaporative emissions (40-87%) in Mumbai (Srivastava, 2004), diesel internal combustion engines (26-58%) in Delhi (Srivastava et al., 2005) and roadway/refuelling exhaust (~40%) in Kolkata city (Majumdar et al., 2009). Except for the study performed in Kolkata which found a contribution of <10% from wood combustion, residential fuel usage was not identified as a potential VOC source in those source receptor modelling studies. The observed discrepancy could be partially caused by the fact a CMB is not necessarily an ideal tool for conducting source receptor modelling study in understudied environments as the model needs to be initialized with locally measured source profiles of all potentially significant sources. However, it is unlikely that this is the only reason for the discrepancies between source receptor modelling outcomes and emission inventories.

Different bottom up emission inventories also have large discrepancies between each other when extracted for the NW-IGP. For our study region (27.4-34.9 °N and 72-79.8 °E), EDGAR v4.3.2 estimates that the road transport sector contributes only 18% of the total anthropogenic VOC emissions (440 Gg y^{-1}), while REAS v2.1 (and MIX Asia) attribute 35.8% of the total anthropogenic VOC emissions (1230 Gg y^{-1}) to this sector. For industrial emissions and solvent use, GAINS has the lowest (540 Gg y^{-1}) and EDGAR v4.3.2 the highest absolute emissions (900 Gg y^{-1}). Crop residue burning as VOC source is missing in REAS but accounts for a 6% (145 Gg y^{-1}) and 7% (163 Gg y^{-1}) share of the annual VOC emissions in EDGAR and GAINS, respectively."

Reviewer comment: The study takes into consideration a specific month, May 2012. It is needed to explain why this month is important for the region under study and which general conclusions can be made from it.

Author response: The month of May is of specific interest for the NW-IGP as it is strongly affected by a seasonal source in the form of wheat residue burning. Crop residue burning activity from the NW-IGP appear prominently in various fire count products such as MODIS or VIIRS fire counts. Our study provides the first in-situ observations which allow to assess whether VOC emissions from this pyrogenic source are properly represented in the available emission inventories.

Changes in the manuscript: We have inserted the following sentence at the end of the paragraph:
"The month of May is of special interest, as it is affected by widespread wheat residue burning in the NW-IGP. In the present study, we quantify the contribution of this important area source to the VOC burden at a downwind site. Our analysis includes several rarely reported nitrogen containing compounds which appear to have strong pyrogenic sources in this particular study region. Compounds such as amines, amides and isocyanic acid are presently not included in global emission inventories and the default atmospheric chemistry mechanisms, despite their potential importance for secondary aerosol formation and human health."

Reviewer comment: The aims of the paper need to be better outlined (e.g. in the last paragraph of the introduction the comparison with emissions inventories is not mentioned in the objectives).
Changes in the manuscript: We added the following at the end of the paragraph:
"We compare our source-receptor modelling output with several emission inventories such as REAS v2.1, EDGAR v4.3.2 and GAINS v5 to assess which emission inventory is most consistent with the results of our source receptor modelling study that employs in-situ observations."

3. METHODS - The description of methods should be revised in its content. In particular:

Reviewer comment: Why have authors chosen to use the US EPA PMF 5.0 model? A brief motivation and description of the model need to be provided along with relevant references.
Changes in the manuscript: We added the following brief description and motivation:
"The EPA PMF 5.0 receptor model (Paatero et al. 2014, Norris et al. 2014) is multivariate factor analysis tool (Paatero & Tapper 1994, Paatero 1997), which decomposes the data matrix x_{ij} with i number of samples and j number of measured VOCs into two matrices, the factor contribution matrix g_{ik} (which provides the mass g contributed by each factor to the individual sample) and the factor profiles matrix f_{kj} (which provides the source profile/fingerprint of each individual source). Both matrices are established for a user defined number of sources p from the existing intrinsic variability in the dataset leaving behind a matrix of residuals e_{ij}.

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij}$$

The two primary advantages of the PMF over other source receptor modelling tools are its inherent non-negative constraints (Hopke 2016) and its capability of optimally weighing individual data points and assigning uncertainties which makes it possible to include less robust species that can be useful for defining real sources. The EPAv5.0 model is superior when compared to other source receptor modelling tools as contains advanced rotational features (Paatero & Hopke) which allow to constrain the rotational ambiguity in a manner that pushes the PMF solution toward the real world space."

Reviewer comment: Almost the entire part of the methodology in Section 2.4 and 2.5 is left to the supplement or to other studies. Since it is a fundamental part of the methodology used in this the study, I would suggest to expand these sections. On the other hand, the detailed description in section 2.2. is not really relevant for this study, and should be cut/shortened.
Author response: We have expanded section 2.4 and 2.5 in the main manuscript and removed the relevant sections from the supplement. We have also shortened section 2.2 but retained the technical details of how the input data was generated.
Changes in the manuscript: Page 3 line 6 was shortened to:

~~"Since the technical details of the measurements and the QA/QC protocol have already been~~ As described in greater detail in Sinha et al.,(2014), ~~we provide only a quick summary here.~~ ambient […]"

Section 2.4 was expanded to:

"**2.4 Conditional Probability Function analysis**

We perform a conditional probability function (CPF) analysis (Leuchner and Rappenglück, 2010) which aids in identifying physical locations of different PMF source factors without using back trajectories (Xie and Berkowitz, 2006). The CPF is computed using the factor contribution of the PMF model in combination with the wind direction at the receptor site. It quantifies the probability of factor contributions surpassing a certain threshold (75^{th} percentile) for a particular wind direction sector thereby highlighting directional dependency of source factors and is defined as follows:

$$CPF = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \quad (2)$$

Here, $m_{\Delta\theta}$ refers to number of samples exceeding the criterion value in a certain wind sector and $n_{\Delta\theta}$ counts the total number of data points in that respective wind sector. $\Delta\theta$ was assigned a value of 30°. "

Section 2.5 was expanded to:

"2.5 Calculation of the ozone formation potential and SOA formation potential

Ozone production potential for each of the PMF derived source factors was calculated based on the method used in Sinha et al., (2012) using the following equation:

$$Ozone\ production\ potential = \left(\sum_i k_{(VOC_i + OH)}\ [VOC_i]\right) \times\ [OH]\ \times n \quad (3)$$

Here, n = 2 and [OH] = 10^6 molecules/cm^3. The values were summed up for all the VOCs for obtaining the ozone production potential corresponding to each of the PMF derived factors for the daytime hours (07:00-18:00LT).

Secondary organic aerosol (SOA) potential was calculated for the PMF source factors using the literature SOA yields (Derwent et al., 2010) under low NO_x conditions for benzene, toluene, ethylbenzene, trimethylbenzene, styrene, methanol, isoprene, formaldehyde, acetaldehyde, acetone, formic acid and acetic acid using the equation given below for 07:00-18:00LT:

$$SOA\ potential = \ (\sum_i [VOC_i][SOAP_i]) \quad (4)$$

Reviewer comment: The description of the methodology used to compare results of this study with the emission inventories estimates should be outlined.

Author response: done

Changes in the manuscript: We inserted a section 2.6 to describe this methodology which was earlier described in the results section (3.8) and have removed the method from section 3.8 to avoid repetition. Now section 3.8 only discusses the results.

"2.6 Methodology for the comparison of PMF source factors with existing emission inventories

Global Emission Database for Global Atmospheric Research (EDGARv4.3) inventory for the year 2012 (Huang et al. 2017), and two regional emission inventories: Regional Emission inventory in Asia (REAS v2.1) for the year 2008 (Kurokawa et al., 2013) and the Greenhouse Gas and Air Pollution Interactions and Synergies model (GAINS) (Amann et al., 2011) for the year 2010 (Stohl et al., 2015) were compared with our PMF output. The gridded inventory was filtered for Latitude: 27.4-34.9 N and Longitude: 72-79.8 E, i.e. the fetch region from which the air mass trajectories reach the receptor site within one day. This

filtering is required because compounds with photochemical lifetimes of less than a day (e.g. styrene, C-8 and C-9 aromatics) feature prominently in several source profiles indicating that most of the transport sector emission were less than a day old when they reached the receptor site. Other compounds with longer lifetimes such as toluene (2 days), benzene (6 days) or acetonitrile (months) can reach the site from more distant sources. The wheat residue burning source shows the highest cross correlation with the regional fire counts for a lag time of 2 days indicating that emissions from distant sources can and do impact the site with a time lag. The chosen fetch region includes the areas where the maximum number of wheat residue burning fire counts are observed while avoiding a size that is too large to be consistent with the relatively unaltered signature of some of the other PMF source profiles.
Annual emissions were available for EDGAR (2012) and GAINS (2010), whereas, REAS provided monthly data (May 2008). However, Figure S5 shows that despite providing monthly data, the REAS emission inventory has very little seasonality for any of the sources. To facilitate the comparison of the PMF output of the month of May which is affected by a strongly seasonal source (crop residue burning) with emission inventories that provide only annual data as of now, we calculate hypothetical pie charts which attribute annual crop residue burning emissions over the region only to the 2.5 months when crop residue burning actually occurs (middle of October to end of November and May)."

Reviewer comment: 4. RESULTS AND DISCUSSIONS -
Overall results are too descriptive, and there are repetitions of information that figures already provide. I suggest to focus more on what can be deduced from the analysis rather than on its description.
Author response: We appreciate this advice by the anonymous reviewer #1 and have restructured our results and discussion section. The former section 3.1 has been combined with some details regarding the model output validation which were spread out over sections 3.2-3.6 and has been shifted to a new section "2.4 Validation of the PMF output" in response to one of the comments of the anonymous reviewer #2. Our Results and discussion section now starts with the content of the former section 3.7 (now shifted to 3.1) "Split up of VOC Emission Sources in Mohali and their contribution to Ozone and SOA Formation Potential". Sections 3.2-3.6 containing the description of the PMF results for the individual factors have been re-written to focus on what the analysis means rather than on describing the results.
Changes in the manuscript:
Section 3.2 now reads:
"The biofuel use and waste disposal factor combines two sources with similar source profiles and high spatio-temporal overlap into one factor. As discussed previously for other South Asian atmospheric environments (Sarkar et al., 2017), the source contributions of domestic biofuel use and domestic waste burning are difficult to segregate. Figure 5 shows a weak bimodal behaviour with an early morning and late evening peak for this factor, as both domestic biofuel use and waste disposal fires peak in the early morning and in the evening hours (Nagpure et al., 2015).
The highest conditional probability for this factor is from the North (>0.4), the direction of the Dadu Majra landfill in Chandigarh, followed by the wind direction NW where a large village (Mauli Baidwan) can be found within 1 km of the receptor and NE, the direction of Panchkula's garbage dump in Sector 23. This and the fact that the average contribution of this factor remains above 30 μgm-3 throughout the night indicates that garbage burning contributes significantly to the biofuel use & waste disposal factor.
Figure 3 and Figure 6 show that this factor explains a significant share of the mass of acetonitrile (a biomass burning tracer), aldehydes, ketones, acids, propyne, and propene in the PMF model. For propene (60%), aldehydes (85%) and ketones (68%) the residential sector is

the dominant source in the most recent speciated emission inventory EDGARv4.3.2. The percentage share for aldehydes and ketones in the inventory is higher than its share in the PMF because the agricultural residue burning source of these compounds is currently missing in the inventory. For acids, however, the residential fuel usage source in the inventory (0.5%) is dwarfed by solvent use associated emissions (96%), while in the PMF the two biomass burning sources (residential biofuel use and waste disposal and wheat residue burning) account for almost 69% of the total acids in the model. High emission of oxygenated VOCs have been reported previously for source profiles of biofuel-stoves (Wang et al., 2009; Paulot et al., 2011; Stockwell et al., 2016) open waste burning (Sharma et al., 2019) and PMF factors' results of residential biofuel use and waste disposal factor in Kathmandu, Nepal (Sarkar et al., 2017).

It should be noted, that this factor is responsible for approximately 25% of the total benzene emissions in our PMF model, while emission inventories attribute a larger share (39-74%) of this compound to this source. Since benzene is an identified Group-1 carcinogen (IARC, 1987) and emissions occur within the household itself (domestic cooking) or within close proximity of the house (waste disposal) this factor deserves special attention in programs targeted at emission reductions. However, the impact of such emission reductions in the residential and waste management sector on human benzene exposure are likely to be overestimated by modelling studies using present day emission inventories, as the inventories attribute 39-74% of the benzene emissions to residential fuel usage and waste disposal, while the PMF suggests the transport sector is the largest benzene source. Direct emission of isocyanic acid, a highly toxic emerging contaminant and its photochemical precursors (Alkyl amines and Amides) was observed from this source and explained 18% of the isocyanic acid mass concentration and 7-15% of all the alkyl amines and amides in the PMF model, respectively."

Section3 3.3 now reads:

Wheat residue burning takes place every year in the NW- IGP in the post-harvest season and generally peaks in the month of May. It has been shown that wheat residue burning has a major impact on both ozone mixing ratios (Kumar et al., 2016) and VOC mixing ratios and hydroxyl radical reactivity (Kumar et al., 2018), and results in a large suite of unknown (40%) and poorly quantified reactive gaseous emissions. Wheat residue burning and emissions are transported to the receptor site from a large fetch region and often with a significant lag time. Hence, there is no strong conditional probability for enhancements from any specific wind direction (Figure 5).

Figure 3 and Figure 6 shows that the wheat residue burning factor explains a significant share of all acids, amines/amides, several ketones, and aldehydes, isoprene/furan, monoterpenes, acetonitrile, propene, styrene and phenol in the PMF model. This makes wheat residue burning the largest contributor to the human exposure to isocyanic acid in the month of May both through direct emissions of isocyanic acid and by virtue of being the largest source for its photochemical precursors.

In the EDGARv4.3.2 the agricultural residue burning source of ketones, aldehydes and acids is missing. On the other hand, agricultural waste burning appears to be the dominant anthropogenic isoprene source (94%) in the EDGARv4.3.2 inventory while in our PMF model residential biofuel usage and the transport sector are equally important contributors to the isoprene/furan mass. The monoterpene emissions from agricultural residue burning (6%) in the EDGARv4.3.2 inventory are dwarfed by emissions from solvent use (90%), while in our PMF solution wheat residue burning and the transport sector appear to be the dominant anthropogenic monoterpene sources.

Section3 3.4 now reads:

“The source fingerprint of the industrial emissions and solvent use factor is dominated by methanol (7.3 μgm-3), acetic acid (3.9 μgm-3) and acetone (2.9 μgm-3). This points towards solvent use (Gaimoz et al., 2011) and/or polymer manufacturing (Sarkar et al., 2017) contributing to the industrial emission and solvent use factor. In addition, Figure 3 and Figure 6 show that this factor explains a significant fraction of the benzene (20%) and acetonitrile (17%) mass in the PMF model. While both are known for their use as solvents (Brown et al., 2007), they can also be emitted from the combustion. The EDGARv4.3.2 emission inventory has a strong industrial and solvent source of toluene, xylenes, acids, formaldehyde and monoterpenes which is not reflected with equal strength in our PMF solution.

The correlation of the industrial emissions and solvent use factor with the SO2 time series (R=0.6), indicates that the emissions of coal or biofuel burning in industrial units and/or coal fired power plants may also be contributing to this factor profile. Figure 5 shows that the highest conditional probability of this factor is to the South East direction (120° -150° wind sector). The receptor site is downwind of a 600 MW coal fired power plant located in Jagadhri (80 km SE) as well as downwind of several industrial areas and brick kiln clusters located around Dera Bassi (15 km), Lalru (20 km) and Jagadhari (80 km) when the wind blows from this direction. In the Kathmandu valley, biofuel co-fired brick kilns explained a significant fraction of the benzene and acetonitrile mass (Sarkar et al., 2017) and the factor profile shows a moderate correlation with the source signature of brick kiln emissions (R=0.5), hence a combustion contribution from brick kilns to the factor profile cannot be ruled out. The diel profile broadly reflects boundary layer dynamics with factor contributions increasing continuously throughout the night indicating a buildup of constant emissions in the nocturnal boundary layer. Factor contributions peak in the early morning (32-49 μg m-3 between 5-9 am local time) and the factor contribution of this factor decreases from 9 am onwards after the breakup of the nocturnal boundary layer. This factor has higher average than the median factor contributions at night due to strong plumes (max 375 μgm-3) reaching the receptor when it is downwind of the industrial sector but not during other nights when the wind direction is from rural Punjab (NW) or the urban sector (NE).”

Section3 3.5 now reads:

The factor profile of the 4-wheeler factor explains a significant share of all aromatic compounds in the PMF model. The factor represents a mixture of multiple components contributed by fuel exhaust and fuel evaporative running losses from vehicles and resembles ambient air samples from a busy traffic intersection. Similar profiles have been observed during field measurements in Beirut, Lebanon (Salameh et al., 2014, 2016) and Hong Kong (Ho et al., 2004). The highest conditional probability is observed for the Chandigarh wind sector (0-90°). As reported previously from Mexico City during the Milagro campaign (Bon et al., 2011), a significant mass of methanol (4.3 μgm-3) and other oxygenated VOCs were present in the traffic emissions factor. The fact that this factor explains 28% of the total m/z 57 is consistent with the gasoline additive MTBE being detected at this m/z ratio as an interference to acrolein/methylketone (Karl et al., 2003; Warneke et al., 2003, 2005; Rogers et al., 2006). Signals at m/z 31, 47, 59, 61, 73, 87 in aged traffic plumes can be attributed to formaldehyde, formic acid , glyoxal, acetic acid, methylglyoxal and 2-butanedione which are products of the gas phase oxidation of toluene, C-8 and C-9 aromatic compounds (Bethel et al., 2000; Ervens et al., 2004). In addition, car exhaust also explained 34% of the propyne mass in the model.

Factor 5, 2-wheeler exhaust, explains 50% of the total toluene mass as well as 17%, 12% and 9%, of the total C-8 aromatics, benzene and C-9 aromatics in the PMF model, respectively. The factor shows a signal at m/z 61 (acetic acid) which may partially be due to fragmentation of octane or ethyl acetate (Warneke et al., 2003; Rogers et al., 2006) which could be present in fuel. The mass has also been attributed to acetic acid in a previous study of diesel tailpipe emissions (Jobson et al., 2005). Nevertheless, it still seems that the 2-wheeler factor profile has

a higher contribution from oxidised compounds compared to the car factor profile indicating that the plumes are typically more aged. Figure 7 shows that this factor displays higher conditional probability than the car factor towards the towns Kharar (8 km N), Dera Bassi (15 km SE) and Lalru (20 km SE), and a lower conditional probability than the car factor towards Chandigarh (NE) indicating 2-wheelers are more abundant in small towns, while cars dominate the traffic emissions in urban Chandigarh.

Figure 7 illustrates that both the traffic factors show bimodal peaks in morning (10.3 μgm-3 at 5-9 am local time) and evening (20 μgm-3 at 7-9 pm local time) during peak traffic hours. When the wind blows from the urban sector (0-90°) during peak traffic hour (7-9 pm) peak factor contributions of >260 μgm-3 for cars and > 150 μgm-3 for 2-wheelers are observed.

As can be seen from Figure 6, the two traffic factors jointly explain 47%, 80%, 70% and 67% of the total benzene, toluene, C-8 and C-9 aromatic compounds in the model consistent with findings from the Kathmandu valley that traffic, not residential biofuel use and waste disposal is the more important source of aromatic compounds in South Asia. It is also clear that despite stringent regulations, the transport sector in the region is still the largest contributor to human benzene exposure. It can be seen from Figure S8a-d that various emission inventories consider the transport sector to be a minor source of benzene (10-16%). The EDGAR v4.3.2 emission inventory also considers the transport sector to be only a minor source of toluene (11-15%) and xylenes (17-22%). Residential fuel usage, industries and solvent use are considered to be the most significant year around source of benzene, toluene and xylenes. Agricultural residue burning becomes the most significant source of all aromatic compounds in the EDGAR v4.3.2 emission inventory, when crop residue burning emissions are treated as occurring during crop residue burning season only, which may imply that the annual emissions of aromatic compounds from the stubble burning may be overestimated. REAS v.2.1 appears to be overestimating the residential fuel burning contribution to benzene and toluene emissions and the solvent usage contribution to toluene emissions. However, it captures the contribution of the transport sector to xylenes and trimethylbenzenes well."

Section3.3.6 now reads:

"Figures 4 and 6 show that mixed daytime sources comprising of biogenic emissions and photochemically formed compounds explained 22% of the monoterpenes and 25% of the measured isoprene, respectively. Isoprene has a short chemical lifetime of 1.5 hours during the day and 16% and 11% of its first generation oxidation products MVK and MEK (Kesselmeier and Staudt, 1999) were also attributed to this factor. In addition, the mixed daytime factor explains 41%, 44%, 24% and 22% of the total formaldehyde, formic acid/ethanol, methanol and acetone mass, respectively. Photochemically formed isocyanic acid, formamide, acetamide and propanamide explain a slightly lower fraction (27-37%) of the total mass concentration of these compounds compared to what has been reported from wintertime Kathmandu valley (36-41%). Figure 7 illustrates that the mixed daytime factor peaks between 9 am and 4 pm and shows a slightly enhanced conditional probability for the 180 -330° rural wind sector (0.2-0.3) due to agroforestry plantations of poplar in the rural landscape."

Reviewer comment: • Section 3.8 presents the comparison between the source apportionment study and emission inventories estimates, i.e. a point vs gridded data. Is it sufficient to filter gridded data for LAT LONG from which air mass trajectories reach the site within one day to make the comparison reliable?

Author response: Air is a rapidly moving medium, in particular in May when the average wind speed is 5.6 ms^{-1}. Hence, the comparison of a receptor point with a much larger gridded area of an emission inventory should not be a concern. In fact, Sofowote et al. 2015 (Atmos. Environ. 108:151–57) used the PMF to source apportion the impact of distant sources on the

$PM_{2.5}$ aerosol burden at 5 remote locations in Ontario, Canada. We think that the more pertinent question is: How large should that gridded area be for a meaningful comparison? Many of the very specific tracers have short photochemical lifetimes of less than a day (e.g. styrene, C-8 and C-9 aromatics). Since these short lived compounds feature prominently in several source profiles, rather than being absent, this indicates that e.g. the 4-wheeler emissions on average have been subjected to photochemical aging for less than 4-10 hours prior to reaching the site. On the other hand, other compounds e.g. toluene (2 days), benzene (6 days) or acetonitrile (months) could have been transported much further away. The wheat residue burning source shows the greatest cross correlation for a lag time of 2 days indicating that emissions from distant sources can and do impact the site with a time lag. Hence we chose a compromise between the two sets of compounds in terms of lifetimes and delineated a fetch region of 1 day for the comparison with the emission inventories. This fetch region includes the areas where the maximum number of wheat residue burning fire counts are observed by satellites while avoiding a size that is too large to be consistent with the relatively unaltered signature of some of the other PMF source profiles.

Changes in the manuscript: We have inserted the following text into the newly created section 2.6

"This filtering is required because compounds with photochemical lifetimes of less than a day (e.g. styrene, C-8 and C-9 aromatics) feature prominently in several source profiles indicating that most of the transport sector emission were less than a day old when they reached the receptor site. Other compounds with longer lifetimes such as toluene (2 days), benzene (6 days) or acetonitrile (months) can reach the site from more distant sources. The wheat residue burning source shows the greatest cross correlation for a lag time of 2 days indicating that emissions from distant sources can and do impact the site with a time lag. The fetch region chosen for comparison with the emission inventories includes the areas where the maximum number of wheat residue burning fire counts are observed by satellites while avoiding a size that is too large to be consistent with the relatively unaltered signature of some of the other PMF source profiles. "

Reviewer comment: Moreover, the study considers May 2012, while emissions inventory data are for 2008/2010. Which are the uncertainties in using these approaches in the comparison? Authors should justify and better describe these choices.

Author response: We have reduced the uncertainties of the comparison by switching from EDGARv4.2 to the more recent version 4.3.2 for the year 2012. As far as REASv2.1 for the year 2008 is concerned, we could not improve the comparison as the NMVOC dataset of the MIX Asia 2010 inventory is identical to the NMVOC dataset of the REAS 2008 inventory. When it comes to the uncertainties introduced by comparing one month's data with an annual average emission inventory is concerned there are two parts to the answer.

1)The first part of the answer is that at present the only inventory that gives monthly data is in no way better than the inventories which provide only annual average data as the monthly data hardly differs from the sum of annual emissions divided by 12. Methane emissions from rice paddies in Punjab persist in the REAS emission inventory throughout the year even in months in which rice is not grown. Other sources do not appear to have been treated differently. Hence de facto there is no seasonality in any of the emission inventories available at present, a short coming that must be overcome in the long run but is beyond the scope of this work.

2) For emission inventories that do not provide monthly data, we have facilitated the comparison of the PMF output of the month of May which is affected by a strongly seasonal source (crop residue burning). To do so, we calculate hypothetical pie charts which attribute annual crop residue burning emissions over the region only to the 2.5 months when crop residue burning actually occurs (middle of October to end of November and May). This should reduce the uncertainty of the comparison. It allows to assess whether the model has the correct annual total emissions of the crop residue burning source and just lacks the proper distribution in the form of monthly data or is off with respect to the total annual emissions itself.

Changes in the manuscript: The following two text segments have been included in section 2.6

"Annual emissions were available for EDGAR (2012) and GAINS (2010), whereas, REAS provided monthly data (May 2008). However, Figure S6 shows that despite providing monthly data, the REAS emission inventory has very little seasonality for any of the sources."

"To facilitate the comparison of the PMF output of the month of May which is affected by a strongly seasonal source (crop residue burning) with emission inventories that provide only annual data, we calculate hypothetical pie charts which attribute annual crop residue burning emissions over the region only to the 2.5 months when crop residue burning actually occurs (middle of October to end of November and May)."

Figure 8 has been changed – so has the accompanying text.

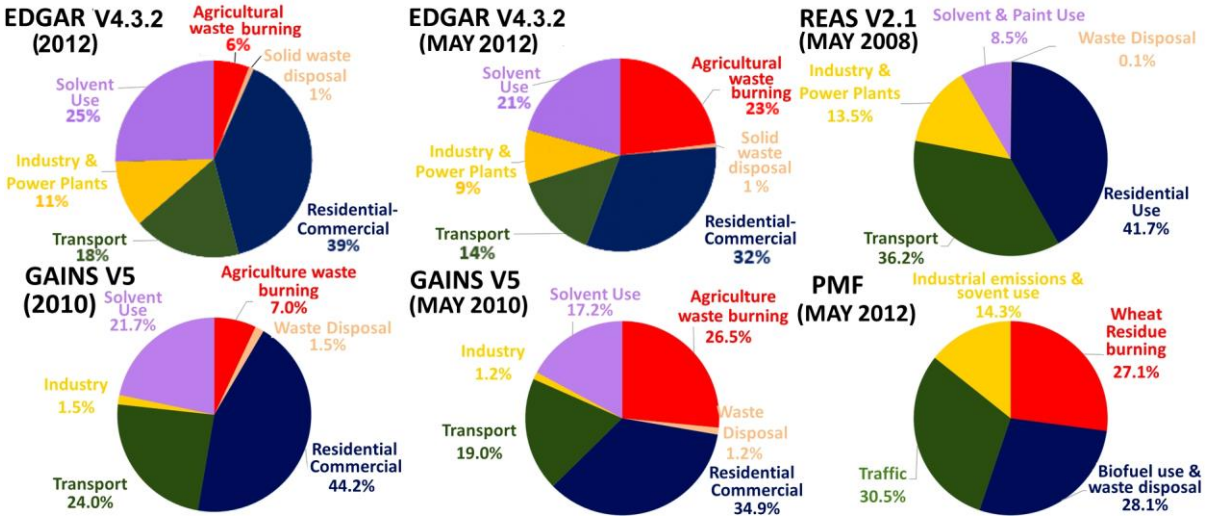

Figure 8 has been revised and now includes EDGAR v4.3.2 (2012) instead of v4.2 (2008) and have updated the discussion accordingly. The latest EDGAR represents a significant improvement over the EDGAR HTAP and v4.2.

We have also added supplementary figures to compare speciated emission inventories with the PMF output for individual aromatic compounds

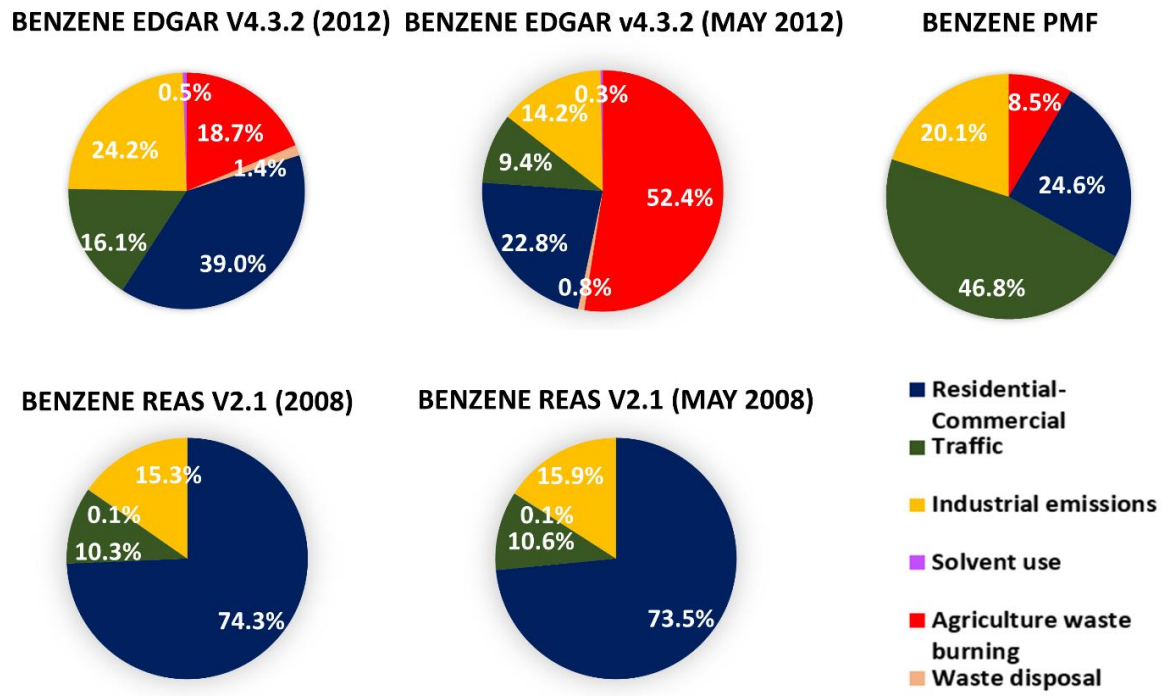

Figure S8a: Comparison of the PMF output with benzene emission inventories for the study region.

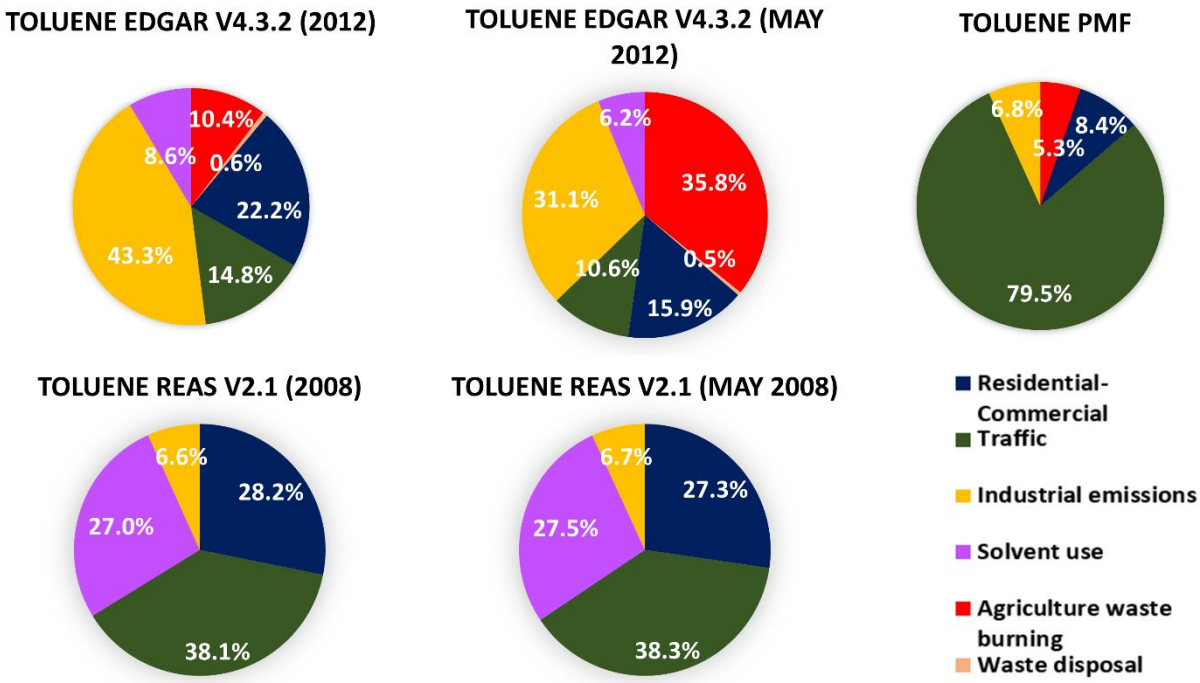

Figure S8b: Comparison of the PMF output with toluene emission inventories for the study region.

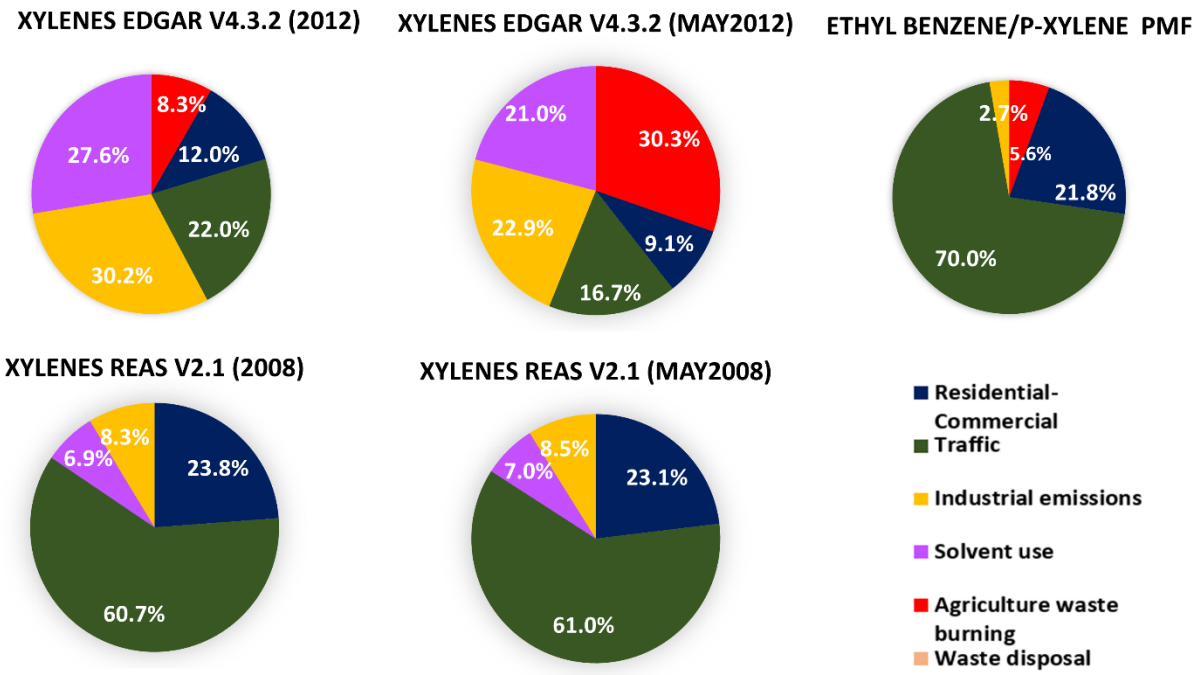

Figure S8c: Comparison of the PMF output with xylenes in the emission inventories for the study region.

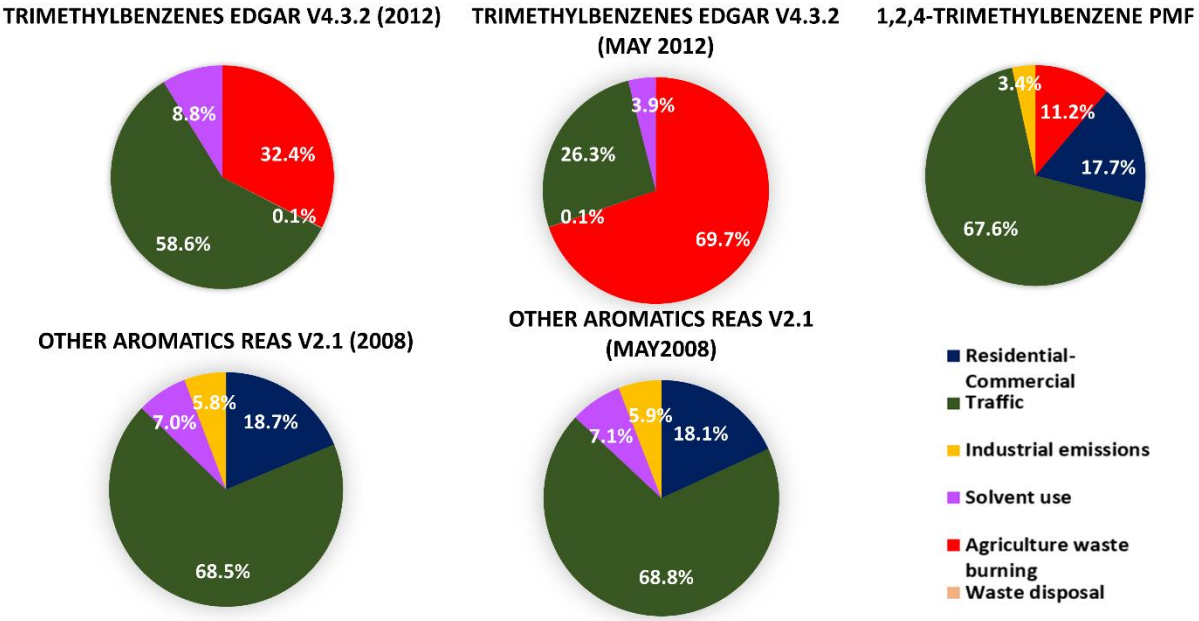

Figure S8d: Comparison of the PMF output of C-9 aromatic compounds with the class "other aromatic compounds" in the emission inventories for the study region.

Reviewer comment: 5. CONCLUSIONS - It would be more valuable for the reader if the authors focused more on the achievements and implications of the results. The last paragraph of 3.8 may be included in the conclusions rather than in results.

Done we have shifted the paragraph and have re-written the conclusions. It now reads as follows:

"Our results highlight that for accurate air quality forecasting and modelling it is essential that emissions are attributed only to the months in which the activity actually occurs. This is important both for emissions from crop residue burning (which occur in May and from Mid-October to the end of November). Annually averaged emissions are unlikely to yield accurate air quality forecast in regions affected by such seasonal events. At present, more specialized fire emission inventories such as FINN (Wiedinmyer et al., 2011) must be used to account for the full seasonality and day to day variations of open burning emissions. We also demonstrate, that the source profiles obtained as PMF output can be validated and matched against samples collected at the potential sources to validate the factor identification.
For the human class I carcinogen benzene, the traffic factor alone contributed to 47% of the total benzene mass at this receptor site followed by residential biofuel use and waste disposal (25%) and industrial emissions and solvent use (20%). This stands in stark contrast to various emission inventories which consider domestic biofuel usage (39%), agricultural residue burning (19%) and industries (24%) to be the most important sources of benzene emissions. Since the annual NAAQS for benzene is exceeded at this receptor site (Chandra and Sinha, 2016), all three sectors must be targeted for emission reductions.
For the emerging contaminant isocyanic acid, photochemical formation from precursors (37%), wheat residue burning (25%) and biofuel usage and waste disposal (18%) were the largest contributors to human exposure. The monthly average isocyanic mixing ratio of 1.4 ppb exceeds concentrations that can, after dissociation at blood pH, result in blood cyanate ion concentrations (Roberts et al., 2011) high enough to produce significant health effects in humans (Wang et al., 2007) such as atherosclerosis, cataracts and rheumatoid arthritis due to protein damage. Peak mixing ratios of this compound exceed 3 ppb in some night time wheat residue burning plumes. Wheat residue burning was also the single largest source of the photochemical precursors of isocyanic acid, namely, formamide, acetamide and propanamide, indicating that this source must be most urgently targeted to reduce human concentration exposure to isocyanic acid.
Overall it appears that none of the emission inventories is ideal at the present. Our PMF solution suggests that transport sector emissions may be underestimated by GAINSv5.0 and EDGARv4.3.2, while the combined effect of residential biofuel use and waste disposal emissions as well as the VOC burden associated with solvent use may be overestimated by all emission inventories. Agricultural waste burning emissions of some of the detected compound groups (ketones, aldehydes and acids) are missing in the EDGARv4.3.2 inventory while aromatic emissions from the same source appear to be overestimated. Thus, large improvements are required in existing emission inventories for correct source attribution and inclusion of missing compounds over this densely populated region of the world."

3 Minor comments
Reviewer comment: 1. First author name (Pallavi) is missing.
Pallavi is a single name author. Her orcid is https://orcid.org/0000-0003-3664-6260

Reviewer comment: 2. Page 2 line 5 '...deserve further study' this sentence need citation.
Author response: This sentence refers to the previous sentence. Citations have been added to the previous sentence (Pawar et al. 2015, Sinha et al. 2014, Kumar et al. 2016)

Reviewer comment: 3. Page 2 line 31 '...and strong photochemistry' this sentence need citation.
Author response & changes in the manuscript: A citation to Sinha et al. 2014 has been added

Reviewer comment: 4. Section 2.3: need to add cross references to Table S3, Figure S4 a, b c.
Author response: done, we have added the cross reference in line 26 page 3
Figures S4 a, b c show how the factor profile, percentage of each VOC originating from a certain source, and the factor contribution change while increasing the number of factors in the model.
and line 1 page 4

A list of the constraints applied is provided in the supplementary table S3

Reviewer comment: 5. Page 9 line 16 'However, Figure S5..'. It is Figure S6 in the Supplement.
Author response: we have changed the numbering of several figures in the supplement as Reviewer #2 asked us to include an additional plot. The numbers are now consistent with the numbering in the manuscript.
Reviewer comment: 6. Figure 1 (b): add lat - long grid. It may be worth to add in the caption the exact coordinates of the site.
Author response: We have added the exact coordinates of the site instead.
We don't agree that adding a grid to the bottom figure is a good idea. It becomes a mess since Google Earth does not seem to allow us to define the grid spacing. It doesn't even seem to allow us to choose a different font size for the location labels and the grid labels. We are dealing with an area of less than 1 x 1 degree, so the figure with grid on looks ugly.
Figure 1b with grid on:

Other minor corrections: While preparing the new supplementary Figure S7 a small mistake in the calculation of the factor time series in μg/m^3 was spotted and corrected in Figure 5,7, S5c and throughout the manuscript.

Response to Anonymous Referee #2

Reviewer comment: The article titled "Source apportionment of volatile organic compounds in the northwest Indo–Gangetic Plain using positive matrix factorisation model" by Pallavi et al., is generally well written and contains some useful information. VOC source apportionment studies are sparse in India, and this study could encourage more such studies in future which is required to understand the VOCs impact on air quality. However, on many occasions, authors seem to over-interpret the results and have drawn some rather farfetched conclusions. I would recommend publication provided my concerns are being addressed satisfactorily.

Author response:

We thank the anonymous reviewer for his/her critical feedback and have addressed the comments individually as detailed below.

Abstract:

Discussion paper

Reviewer comment: Numbers can be presented in a better way for ease of reading. Authors can put the percentage contribution of different factors/parameters in the parenthesis beside them.

Author response: We thank the anonymous reviewer for this valuable suggestion. The anonymous reviewer #1, suggested to focus the abstract more on the big picture. In response to both comments we have reduced the numbers in the abstract and added the percentage contribution of different factor/parameters in parenthesis beside them. It now reads as follows:

Changes in the manuscript:

"In this study we undertook quantitative source apportionment for 32 volatile organic compounds (VOCs) measured at a suburban site in the densely populated North-West Indo-Gangetic Plain using the US EPA PMF 5.0 Model. Six sources were resolved by the PMF model. In descending order of their contribution to the total VOC burden these are "biofuel use and waste disposal" (23.2%), "wheat-residue burning" (22.4%), "cars" (16.2%), "mixed daytime sources" (15.7%), "industrial emissions and solvent use" (11.8%) and "two-wheelers" (8.6%).

Wheat residue burning is the largest contributor to the total ozone formation potential (32.4%). For the emerging contaminant isocyanic acid, photochemical formation from precursors (37%) and wheat residue burning (25%) were the largest contributors to human exposure. Wheat residue burning was also the single largest source of the photochemical precursors of isocyanic acid, namely, formamide, acetamide and propanamide, indicating that this source must be most urgently targeted to reduce human concentration exposure to isocyanic acid in the month of May. Our results highlight that for accurate air quality forecasting and modelling it is essential that emissions are attributed only to the months in which the activity actually occurs. This is important for emissions from crop residue burning (which occur in May and from Mid-October to the end of November).

The SOA formation potential is dominated by "cars" (36.9%) and "two-wheelers" (22.1%), which also jointly account for 47% of the human class I carcinogen benzene in the PMF model. This stands in stark contrast to various emission inventories which estimate the transport sector contribution to the benzene exposure as (~10%) and consider residential biofuel use, agricultural residue burning and industries to be more important benzene sources. Overall it appears that none of the emission inventories represent the regional emissions in an ideal manner. Our PMF solution suggests that transport sector emissions may be underestimated by GAINSv5.0 and EDGARv4.3.2 and overestimated by REASv2.1, while

the combined effect of residential biofuel use and waste disposal emissions as well as the VOC burden associated with solvent use and industrial sources may be overestimated by all emission inventories. Agricultural waste burning emissions of some of the detected compound groups (ketones, aldehydes and acids) are missing in the EDGARv4.3.2 inventory."

Reviewer comment: Methods:
Sec 2.3: Line 20, why 20%? Please explain and incorporate in the manuscript as well.
Author response: We chose to assign 20% uncertainty to all masses to avoid a situation where the difference in the assigned uncertainty drives the PMF to dedicate a disproportionate number of a factors towards minimizing Q of a few compounds at the expense of others which may be equally useful as tracers of specific activities. The lower reported uncertainty of some compounds (8-12%) in Sinha et al. 2014 can be primarily attributed to the fact that the instrument has been calibrated with more than one independently sourced calibration gas bottles for that particular compound and the fact that the respective m/z has a good signal to noise ratio and high signals. For some other compounds the sensitivity had to be derived from theory, because no calibration gas is available, hence they carry a larger error.

We have followed the advice of Paatero et al. 2014, Atmos. Meas. Tech., 7, 781–797 and performed sensitivity studies to better understand how errors and their handling can impact the PMF output in our specific case.

In our specific case the fact that toluene has one of the smallest reported measurement errors (8.6 % in Sinha et al. 2014) in combination with the fact that there is a genuine and abundant source with a normalized source profile that is dominated by toluene (tailpipe exhaust of petrol fuelled 2-wheelers) can result into a serious modelling error. This problematic behaviour is observed for this particular dataset, because the second most abundant compound in the same tailpipe exhaust source profile (the xylenes) carries a larger uncertainty (11.8 %) and can be accommodated in other source profiles with a smaller penalty on Q. Most real world traffic contains a mixture of 4-wheelers and 2-wheelers and the ratio of these two vehicle classes in the traffic varies as a function of air mass origin and time of the day. At the same time the benzene/toluene ratio of all aged plumes varies with the photochemical age of the air mass. When all these factors are combined the situation becomes such that while running the model with differential errors the lowest Q for the equation

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

is obtained by creating a separate toluene factor and removing toluene from the factor profile of all combustion source profiles. In other words, in this specific dataset assigning correct but different random errors to different m/z triggers a serious modelling error which appears already in a 4 factor solution and is retained through any higher number of factors. Assigning equal random errors to all m/z prevents this modelling error from occurring. Hence, we assigned an error of 20% to all masses even though in reality only few strong m/z ratios (formaldehyde, propyne, styrene and phenol) carry such a large error. The reviewer is, however, correct that this choice of using the largest error for all compounds is somewhat arbitrary and one could instead use the average uncertainty of all the strong compounds errors

(i.e. assigning ~10% uncertainty to all compounds would also prevent the modelling error). The magnitude of the chosen error will impact the magnitude of Q (which will increase by a factor of ~4 when 10% instead of 20% uncertainty is assigned) but will not change the model output as long as equal uncertainty is assigned to all masses. However, considering the accuracy and precision error while initializing the PMF may not actually be the right approach at all, considering that the software treats the errors as random. One could argue that only the precision error should be considered while assigning errors in the software. However, as long as equal uncertainty is assigned to all strong m/z the assigned uncertaintyt will not change the model output and conclusions.

Changes in the manuscript: We have incorporated this reason and the section now reads "All 32 species were assigned a fixed 20% in the uncertainty, which represents the largest uncertainty estimated for strong compounds, to avoid a situation where the difference in the assigned uncertainty drives the PMF to dedicate a separate factor towards minimizing Q of a single compound with low uncertainty (toluene) by taking it out of all other source profiles and opening a separate factor profile containing just a single compound."

Reviewer comment: Line 20, more than 50% of the measured species (18 of 32) are weak, isn't that going to influence the robustness & reliability of the PMF output?
Author response: This is definitely going to impact the robustness and reliability of the PMF output in a positive manner. A poor signal to noise ratio indicates that the measured values of a species throughout most of the time series are very close to the detection limit. All instruments have a higher precision error close to the detection limit. This is why the manual recommends assigning masses with low S/N ratio "weak" and we have followed this instruction for all compounds that have a poor S/N ratio and do not show any strong peaks. However, in our opinion S/N ratio should not be blindly used as a criterion to make masses weak. Let us consider the hypothetical scenario of a compound emitted only by a single source impacting the site. There can be a situation where such a source impacts the site only rarely (say less than 5% of the time) but when it does the plume brings a very high concentration of that tracer compound. In such a case that specific tracer could be extremely precious for constraining the rotational ambiguity of the PMF solution, even though its average S/N ratio would be very poor (because 95% of the values in the time series are noise around the detection limit). Hence one always needs to look at every species of the input dataset carefully to assess whether it should be made weak just because of its S/N ratio.

There can be reverse cases of masses with a high S/N ratio (for which the average concentration is always far above the detection limit throughout the time serious), which can negatively impact the PMF rotational ambiguity when not labelled as weak. This is the case for all masses with potential isobaric interferences. Let us consider an m/z where one of the compounds falling onto the mass to charge ratio is of pyrogenic origin and the other one a tracer for biogenic emissions or a product of daytime photochemistry and discuss how this will impact the PMF model output depending on whether the species is a strong or weak species. Any peak in the concentration observed can be due to either of the contributors i.e. due to a combustion source alone or due to biogenic emissions/daytime photochemistry alone or due to a mixture of both. The most serious impact of this on the model performance is that it can make resolving the rotational ambiguity difficult and can cause modelling errors. Resolving rotational ambiguity requires that the matrix contains a sufficient number of zero values where a source is totally absent. When two sources with different temporal profiles (night-time combustion and daytime biogenic emission or night-time combustion and daytime photochemistry) contribute different compounds to the same m/z ratio, zero values

are almost absent in that particular column of the matrix. When this column is made "weak" and given a higher uncertainty, other "strong" tracers with genuine zero values and strong peaks that can be attributed to a specific sources define source profiles and this reduces the rotational ambiguity of the model. The "weak" compounds with isobaric interferences are distributed among the source profiles available as per the solution that minimizes Q but they do not define any of the profiles. In our opinion, this is the most appropriate way to treat m/z ratios with potential isobaric interferences. As already described in the supplement and the main text we have made such masses weak in the PMF to improve the quality of the PMF output.

Changes in the manuscript: We have added a clarification that this makes the model more robust.

"Designating sources with isobaric interferences as weak is warranted because when two sources with different temporal profiles (night-time combustion and daytime biogenic emission or night-time combustion and daytime photochemistry) could potentially contribute different compounds to the same m/z ratio, zero values are almost absent in that particular column of the matrix and the tracer is affected by additional uncertainty not appropriately expressed by merely looking at the instrumental measurement error and the signal to noise ratio. When this column is made "weak" and given a higher uncertainty, other "strong" tracers, representing a single compound, define source profiles and this reduces the rotational ambiguity of the model. The "weak" compounds with isobaric interferences tend to be distributed among the source profiles available as per the solution that minimizes Q but they do not define any of the profiles."

Reviewer comment: Line 24, Why the authors chose to remove missing values instead of replacing them with some other values as mentioned in the literature? Is this the standard practice? what is the % of missing values in the total sampled points?

Author response: Replacing missing values with the median while assigning it a greater uncertainty in the PMF helps a lot when the PMF is run with different tracers measured with different sets of instruments, each of which has a different set of missing values. The default setting of the EPA PMF model described in the literature was developed for such a scenario. To illustrate let us consider using a dataset with data from 10 different instruments each of which individually has less than 10% downtime in a situation where unfortunately problems rotate. In that scenario for > 50% of the data points a few variables would be missing. Using the exclude missing value option in such a case would mean throwing out more than half of the dataset as the model removes all lines (=points in time) with a missing value, even if only a single column has a missing value. In such cases lines with missing values still have a lot of data (because one instrument is down the other instruments are running) and only a small subset of species is missing for each point to be potentially excluded. Hence the default model setup suggests filling in missing values with the median of the time series while assigning a greater uncertainty to that point.

However, we are dealing with measurements of a single instrument and <5% of missing values in a month. Filling missing values does not improve the quality of the model output in our case. When the PTR-MS is undergoing calibration or ion source cleaning, there is no ambient data at all available for the gap. Hence the gap filling is unnecessary. It serves no purpose and would hardly affect the model output as all parameters would be filled in with their respective median for that particular point in time.

Changes in the manuscript: added (<5%) after "missing values"

Reviewer comment: Sec 3.3

Line 1-3, R= 0.4 is not a good correlation, at best it can be termed as moderate. Please rewrite the explanation on why fire count is the best tracer for factor 2.
Author response: With best we simply meant that the R was better than that of other potential independent tracers such as NOx (which correlated more with transport sector emissions) and CO (which correlated best with the more regular open burning activities such as biofuel use and waste disposal). However, we understand now, that this could be misunderstood and have revised the sentence.
Changes in the manuscript: "Figure 3 shows that the factor profile correlates most strongly with flaming wheat residue burning (R=0.9). The average contribution of wheat residue burning to the total NMVOC mass at the receptor site and the daily fire counts over the NW-IGP show a moderate cross correlation of R=0.4 with a lag of 2 days (Figure 5)."

Reviewer comment: Sec 3.7
Line 7, in PM2.5, 2.5 should be subscript
Author response: Done this section has become Sec 3.1 in the restructured manuscript

Reviewer comment: Line 7, I don't think the way SOA being calculated enable the authors to make such strong quantitative assertion about the SOA contribution to PM2.5 in Mohali. At best, the adopted method can provide a qualitative and comparative assessment of SOA production efficiency among different PMF factors. I would suggest to remove or modify. line 6-8 to reflect this.
Author response: We understand that the SOA formation potential as calculated has its limitations and depends on the NO_X regime and may even show a non-linear dependence on VOC and NOx concentration for some compounds (Xu et al., 2015, Atmospheric Environment 101, 217-225). However, we believe that providing a boundary condition may be useful. We have modified lines 6-8. We now explicitly mention that we applied the "SOA yields for the low NOx regime" in the relevant section of the materials and methods, which the reviewer #1 asked to extend and in this section. We also now put the calculated SOA formation potential (i.e. the ~17 μg/m3) in brackets behind its first mention in the paragraph as we believe that despite all short comings this number provides an important perspective. In support, we have also qualified the estimate by citing more studies.
Changes in the manuscript:
"While the calculated SOA formation potential particularly from transport sector emissions (Ensberg et al., 2014) and aromatic compounds (Li et al., 2017, Li et al., 2018) is affected by large uncertainties and may depend in a non-linear fashion on NOx and VOC concentrations (Xu et al, 2015) our calculated SOA formation potential seem to indicate that SOA formation could contribute significantly to the overall PM2.5 burden (104 μgm-3).

Reviewer comment: Sec 3.8
I am not sure about the utility or purpose of this section. Are authors trying to use this comparison as another tool for PMF results validation? Or to suggest which inventory is better? Every emission inventory is developed based on some underlying assumptions and approximations. I would rather be very surprised if a single site based study can reproduce or match the emission inventory values. It is quite expected that differences will be there and even a perfect match doesn't necessarily validate emission inventories or the PMF results, especially in a complex source environment as in India. Several assumptive statements were made to explain the mismatch/less match between PMF results and emission inventory values. So, based on this comparison one can't really assert which inventory is better or more representative than others. Authors should remove or rephrase the section to reflect those concerns.

Author response: This section is meant to identify which of the currently used emission inventory represents the regional sources best. This is the major motivation behind any source-receptor modelling study. The anonymous reviewer is correct that every emission inventory is developed based on some underlying assumptions and approximations. Some of these assumptions and approximations can be awfully wrong and the purpose of source receptor modelling studies is to point out such discrepancies. For example, several PMF based source receptor modelling studies in Europe found that the solvent source could be overestimated in most emission inventories while the transport sector may be underestimated (Gaimoz et al. 2011 *Environ. Chem.* **2011**, *8*, 91–103., Niedojadlo et al. 2007 *Atmos. Environ.* **2007**, *41*, 7108., Lanz, et al. 2008 *Atmos. Chem. Phys.* **2008**, *8*, 2313.). Such discrepancies between inventories and source receptor modelling results which got replicated in several studies in different countries ultimately triggered a new series of road tunnel studies and on-road emission factor measurements to re-evaluate the assumptions and approximations used while building the transport sector emission inventories. These efforts not only resulted in a significant upward revision of transport sector emission estimates for NMVOCs while shifting from the EDGARv4.2 inventory for the year 2010 to EDGARv4.3.2 for the year 2012 but also exposed the diesel cheat software that switched off pollution control devices when the vehicles were driving on the road. Therefore, we believe that reality checks bases on source receptor modelling of ambient data perform an important role. Their potential significance is even larger in a complex environment where activity data for informal sector industries and activities that officially don't happen (e.g. open waste burning) is hard to obtain while at the same time proper emission factor measurements for many sources are lacking. We, therefore, insist that this section is important to retain.

The validation of PMF results in all prior studies has been performed by cross correlating one or several columns of one of the two matrices produced during the factor decomposition, namely the factor contribution matrix, with independent variable in the form of the time series of compounds that were not used to drive the model. We performed this cross verification step for all six identified factors using the species NOy (cars & 2-wheelers), SO_2 (industrial emissions), CO (domestic fuel usage and waste disposal), fire counts (wheat residue burning) and O_3 (mixed daytime factor). However, our study, to the best of our knowledge, is the first one to add an additional verification step in the form of grab samples collected at the source which were used to independently verify the factor profiles (i.e. the second matrix) that the PMF model created during the matrix decomposition. This validation was performed using samples collected at the source for five of the six factor profiles (wheat residue burning, domestic fuel usage and waste disposal, industrial emissions and solvent use, car tailpipe emissions, and 2-wheeler tailpipe emissions). It appears that this validation procedure was not described clearly enough, hence we have now added a section describing the procedure to the materials and methods section. We added a section 2.4 Validation of the PMF output. Some of the text in this section has been shifted from section 3.1 to this section. We also added a new reference since the source signature of brick kilns has recently been published and has now been included.
We have also switched to a new version of EDGAR (v4.3.2) which has recently become available and have added more depth to the comparison by looking at individual compound classes of the speciated emission inventory rather than just at the total VOC mass. We have also removed some of the quantitative statements.

Changes in the manuscript:
"2.4 Validation of the PMF output

The PMF generates two matrices from the intrinsic variability in the dataset. A factor contribution matrix and a factor profile matrix.
Traditionally the PMF output has been validated by cross-correlating the factor contribution matrix with independent tracers which were not used to initialize the model, but are considered useful tracers for the respective source (Brown et al. 2007, Leuchner et al. 2011, Bon et al. 2011, Gaimoz et al. 2011, Sarkar et al. 2017). We perform this validation step for all six source factors resolved by the PMF model. These were identified as "biofuel use and waste disposal", "wheat-residue burning", "four-wheelers", "two-wheelers", "industrial emissions and solvent use" and "mixed daytime sources", respectively. The factor contribution for 4-wheelers (R=0.7) and 2-wheelers (R=0.6) correlated best with the independent tracer NOy which is considered to be a vehicular exhaust marker (Ramanathan et al., 1985). The factor contribution of the domestic fuel usage and waste disposal factor correlated best with the independent tracer CO (R=0.9), a proxy for inefficient combustion, while the factor contribution of the industrial emission factor correlated best with the independent tracer SO_2. (R=0.6). The wheat residue burning factor days showed a moderate cross correlation with MODIS fire counts with an R=0.4 and a lag of 2 days. Ozone (R=0.8) was the best independent tracer for the mixed daytime factor.
However, our study goes one step further than all previous studies in validating the PMF output. For 5 out of 6 factors we validated the factor profiles generated by the PMF model (Figure 3) against grab samples collected at the source. Factor profiles were cross-correlated with the fingerprints of source samples collected from a number of potential sources including wheat residue fires (Chandra et al., 2017; Kumar et al., 2018), ambient air samples from a busy traffic junction (Chandra et al., 2017) and an industrial area (this study), tail-pipe exhaust of various vehicles (this study), waste burning (Sharma et al., 2019), leaf litter burning (this study) and domestic biofuel use (Stockwell et al., 2016) and brick kilns (Zhong et al., 2019) to identify the sources. Figure 3 shows the factor profiles obtained from the PMF run (in dark blue), the percentage of each species explained by the respective PMF factor (red squares) and the source profiles of those sources which best matched the factor profile (in various colors as indicated in the legend). The factor profile of residential fuel usage and waste disposal correlated most strongly with the measuredVOC source speciation profiles of domestic cooking (R=0.8), leaf-litter burning (R=0.7) and smoldering garbage fires (R=0.6), the wheat residue burning factor with flaming wheat residue burning (R=0.9), the 4-wheeler factor with petrol-fueled cars (R=0.5) and urban traffic junction grab samples (R=0.8) and the 2-wheeler factor with the tailpipe exhaust of petrol-fuelled 4-stroke two-wheelers (R=0.6). The industrial emissions correlated moderately with the source profile of brick kilns (R=0.5) and ambient air samples collected in an industrial area (0.6). For mixed daytime no source profile sampling is possible. "

The revised section 3.8 now reads:

"Figure 8 shows pie charts depicting the contribution of different sectors to the total VOC mass burden for the emission inventories and our PMF output. Biofuel use and waste disposal were responsible for 28.1% of the mass in our PMF but 39%, 44% and 42% of the mass in EDGARv4.3.2, GAINS and REASv2.1 respectively. The contribution of crop residue burning (27.1%) to the VOC mass in the month of May would be highly underestimated by both GAINS (7%) and EDGARv4.3.2 (4.76%) if the annual emissions are attributed equally to all months of the year. However, if both emission inventories would attribute their annual crop residue burning emissions over the region only to the 2.5 months when crop residue burning actually occurs (middle of October to end of November and May), these emission inventories could be reconciled with the PMF solution, as emissions in May would amount to

26.5% and 23% GAINS and EDGARv4.3.2, respectively as shown in Figure 8. At the same time the percentage share of domestic fuel use and waste disposal would drop to 32% and 35% in EDGARv4.3.2 and GAINS, respectively and the contribution of industrial emissions and solvent use would drop to 18% in GAINS and 30% in EDGAR, respectively. Our PMF solution indicates that industrial emissions and solvent usage (14.3%) are currently overestimated in all emission inventories but are closest to GAINS (540 Gg y-1, 18%) for industrial emissions and solvent use. For domestic biofuel use and waste disposal EDGARv4.3.2 (968 Gg y-1, 32%) appears to agree best with our PMF solution. For wheat residue burning GAINS agrees well with our PMF output, while the agricultural waste burning emissions of some of the detected compound groups (ketones, aldehydes and acids) appear to be missing in the EDGARv4.3.2 inventory. Our PMF solution for road transport sector emissions (30.5%) lies in between the estimates of GAINS (558 Gg y-1, 24%) and REAS (1230 Gg y-1, 36.2%), possibly, because not all pre-2000 super-emitters for which the 20-year vehicle lifetime has been exceeded have been retired as planned.
Overall it appears that none of the emission inventories is ideal at the present. Our PMF solution suggests that transport sector emissions may be are underestimated by GAINS and EDGARv4.3.2, while the combined effect of residential biofuel use and waste disposal emissions as well as the VOC burden associated with solvent use may be overestimated by all emission inventories. Similar results have been reported previously. Sarkar and co-workers (Sarkar et al., 2017) reported an underestimation of transport sector emissions for the REAS and EDGAR emission inventory for the Kathmandu valley in Nepal and an overestimation of the residential biofuel use and waste disposal source in all emission inventories, while Gaimoz and co-workers (Gaimoz et al., 2011) reported an overestimation of the VOC emissions from solvent use in Paris."

Reviewer comment: Figures:
I want to see Q/Qexp plot in SI.
done

Reviewer comment: Fig. 7: Why the evening peaks in Car & Two-wheeler contributions are significantly more pronounced than morning hours?
Author response: Because more traffic activity happens in shorter timespan in the evening and in addition the boundary layer is more shallow during that evening rush hour. In the mornings various activities associated with vehicular movement are spread out over a longer time period (e.g. schools tend to start earlier (8 am) than offices (9 am) and most markets open only at 11 am). In addition, the sun rises before 6 am and hence peak morning traffic occurs after the daytime boundary layer has been well established. This results in greater dilution and lower mixing ratios. We have included the explanation in the revised manuscript
Changes in the manuscript: Inserted
"Mass loadings during evening rush hour are higher than during morning rush hour, because peak morning traffic occurs after the breakup of the nocturnal boundary layer, while in the evening emissions accumulate in the shallow nocturnal boundary layer."

Manuscript in Track Changes mode

Source apportionment of volatile organic compounds in the north-west Indo-Gangetic Plain using positive matrix factorisation model

Pallavi [1], **Baerbel Sinha**[1], **and Vinayak Sinha**[1]

[1]Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Mohali, Sector 81, S.A.S Nagar, Manauli PO, Punjab, 140306, India.

Correspondence: Baerbel Sinha (bsinha@iisermohali.ac.in)

Abstract. In this study we undertook quantitative source apportionment for 32 volatile organic compounds (VOCs) measured at a suburban site in the densely populated North-West Indo-Gangetic Plain using the US EPA PMF 5.0 Model. Six sources were resolved by the PMF model. In descending order of their contribution to the total VOC burden these are ~~namely~~ "biofuel use and waste disposal"(23.2%), "wheat-residue burning"(22.4%), "cars"(16.2%), "mixed daytime sources"(15.7%) "industrial emissions and solvent use"(11.8%) and "two-wheelers" (8.6%) ~~and~~ .

Wheat residue burning is the largest contributor to the total ozone formation potential (32.4%). For the emerging contaminant isocyanic acid, photochemical formation from precursors (37%) and wheat residue burning (25%) were the largest contributors to human exposure. Wheat residue burning was also the single largest source of the photochemical precursors of isocyanic acid, namely, formamide, acetamide and propanamide, indicating that this source must be most urgently targeted to reduce human concentration exposure to isocyanic acid in the month of May. Our results highlight that for accurate air quality forecasting and modelling it is essential that emissions are attributed only to the months in which the activity actually occurs. This is important both for emissions from crop residue burning (which occur in May and from Mid-October to the end of November).

The SOA formation potential is dominated by cars (36.9%) and two-wheelers (21.1%), which also jointly account for 47% of the human class I carcinogen benzene in the PMF model. This stands in stark contrast to various emission inventories which estimate only a minor contribution of the transport sector to the benzene exposure (∼10%) and consider residential biofuel use, agricultural residue burning and industries to be more important benzene sources. Overall it appears that none of the emission inventories represent the regional emissions in an ideal manner. Our PMF solution suggests that transport sector emissions may be underestimated by GAINSv5.0 and EDGARv4.3.2 and overestimated by REASv2.1, while the combined effect of residential biofuel use and waste disposal emissions as well as the VOC burden associated with solvent use and industrial sources may be overestimated by all emission inventories. The agricultural waste burning emissions of some of the detected compound groups (ketones, aldehydes and acids) appear to be missing in the EDGARv4.3.2 inventory. ~~The biofuel and waste disposal, wheat residue burning, industrial emissions and solvent use, combined traffic sources, mixed daytime sources accounted for 23.2%, 22.4%, 11.8%, 25.1%, and 15.7% of the total VOC mass concentration respectively; 18.1%, 32.4%, 7.3%, 21.9%, and 20.3% of the total O_3 formation potential respectively; and 14.9%, 13.9%, 10.1%, 59.0%, and 2.2% of the SOA formation potential, respectively. Further the factors contributed 24.6%, 8.5%, 20.1%, 46.8%, and 0%, respectively, to the human class I carcinogen benzene and 18.4%, 25.4%, 5.9%, 13.3%, and 36.9%, respectively, to the toxic emerging contaminant isocyanic acid. Evaluation of emission inventories using the in-situ data derived PMF solution revealed that among EDGARv4.2, REASv2.1 and GAINSv5.0, the GAINSv5.0 emission inventory for year 2010, best agreed with the in-situ data derived PMF results for May 2012.~~

1 Introduction

Volatile organic compounds (VOCs) have diverse natural (760 $\mathrm{Tg(C)\,y^{-1}}$ (Sindelarova et al., 2014)) and anthropogenic sources (127 $\mathrm{Tg\,y^{-1}}$ average value (IPCC, 2013)).

Certain VOCs emitted primarily by anthropogenic sources such as benzene and isocyanic acid, have direct adverse impacts on human health even at low ppb concentration exposures (Chandra and Sinha, 2016). In densely populated regions like the Indo Gangetic Plain (IGP), reactive anthropogenic VOCs contribute significantly towards the formation of health relevant secondary pollutants such as ozone and secondary organic aerosol (Chandra and Sinha, 2016; Sarkar et al., 2016). At our study site, a representative suburban site in the NW-IGP, the 8 h average NAAQS (national ambient air quality standard) for ozone limit of 100 $\mu g\,m^{-3}$ was exceeded on 29 out of 31 days during May 2012 (Sinha et al., 2014), while the 24 h average NAAQS for $PM_{2.5}$ of 60 $\mu g\,m^{-3}$ was exceeded during 27 out of 31 days in the same period. It has been shown that wheat residue burning results in significant enhancement (by 19 ppb) of the daytime ozone mixing ratios in pre-monsoon season (Kumar et al., 2016) and long range transport in the form of dust storms from the Arabian Peninsula brings extremely high $PM_{2.5}$ mass loadings (with peak $PM_{2.5}$ mass loadings of 950 $\mu g\,m^{-3}$ on 17^{th} of May 2012) (Sinha et al., 2014; Pawar et al., 2015) and enhances the $PM_{2.5}$ mass by $\sim 30\,\%$.

However, ozone mixing ratios exceed the NAAQS even during the non-fire influenced days of the pre-monsoon season and the NAAQS for $PM_{2.5}$ is exceeded 60 % of the time for air masses with no history of long range transport (Kumar et al., 2016; Pawar et al., 2015). This indicates that local ozone and $PM_{2.5}$ precursor emissions deserve further study.

Previous source receptor modelling studies of VOC emission from India (Srivastava, 2004; Srivastava et al., 2005; Majumdar et al., 2009) produced results that conflicted strongly with the bottom up emission inventories, all of which ~~have~~ contain significant emissions from residential fuel usage even when filtered for the New Delhi National Capital Region (41-45 %), Greater Mumbai (32-36 %) and Greater Kolkatta (33-59 %). Transport sector emissions, according the bottom up emission inventories contribute only 15-35 %, 17-43 % and 6-14 % to the total VOC emissions in New Delhi National Capital Region, Greater Mumbai and Greater Kolkatta, respectively. ~~(43 % - 68 %) as their largest VOC source.~~ All previous studies employed a chemical mass balance (CMB) technique for ambient VOC source attribution and identified the transport sector as the main source in the form of evaporative emissions (40-87 %) in Mumbai (Srivastava, 2004)), diesel internal combustion engines (26-58 %) in Delhi (Srivastava et al., 2005) and roadway/refuelling exhaust (40 %) in Kolkata city (Majumdar et al., 2009). Except for the study performed in Kolkata which found a contribution of <10 % from wood combustion, residential fuel usage was not identified as a potential VOC source in those source receptor modelling studies. The observed discrepancy could be partially caused by the fact a CMB is not necessarily an ideal tool for conducting source receptor modelling study in understudied environments as the model needs to be initialized with locally measured source profiles of all potentially significant sources. However, it is unlikely that this is the only reason for the discrepancies between source receptor modelling outcomes and emission inventories. The only other source receptor modelling study in South Asia was conducted using a positive matrix factorisation model (EPA PMF5.0) with data collected in the Kathmandu valley, Nepal, as part of the SUSKAT campaign and attributed a negligible fraction of the anthropogenic VOC burden to residential biofuel usage (14 %). Instead different industrial sources including brick kilns (jointly 52 %) and the transport sector (21 %) were identified as the dominant VOC sources in the Kathmandu valley.

Different bottom up emission inventories have large discrepancies between each other when extracted for the NW-IGP ~~in this understudied region~~. For our study region (27.4-34.9 °N and 72-79.8 °E), EDGAR v4.~~2~~3.2 (Huang et al., 2017) estimates that the road transport sector contributes only ~~10.9~~18 % of the total anthropogenic VOC emissions (~~220~~440 $Gg\,y^{-1}$), while REAS v2.1 (Kurokawa et al., 2013) attribute~~s~~ 35.8 % of the total anthropogenic VOC emissions (1227 $Gg\,y^{-1}$) to this sector. For industrial emissions and solvent use, ~~EDGAR v4.2 again~~ GAINS (Amann et al., 2011) has the lowest (~~277~~540 $Gg\,y^{-1}$) and ~~REASv2.1~~EDGAR v4.3.2 the highest absolute emissions ~~of~~ (~~900~~736 $Gg\,y^{-1}$). Crop residue burning as VOC source is missing in REAS but accounted for a ~~4.7~~6 % (~~95~~145 $Gg\,y^{-1}$) and 7 % (163 $Gg\,y^{-1}$) share of the annual VOC emissions in EDGARv4.3.2 and GAINS, respectively. Considering the large discrepancies between bottom up inventories and different source receptor modelling studies, more source receptor modelling studies using robust statistical tools and better tracers for different biomass burning sources are necessary.

In the present study, we applied the US EPA's PMF 5.0 model in constrained mode for source apportionment of 32 VOCs measured at IISER Mohali Atmospheric Chemistry Facility in May 2012 with the objective of quantifying the most important sources of ozone and SOA precursors, the human class I carcinogen benzene and the emerging contaminant isocyanic acid (Chandra and Sinha, 2016), so that strategies for air pollution mitigation can benefit from quantitative evidence concerning the contribution of major sources. The month of May is of special interest, as it is affected by widespread wheat residue burning in the NW-IGP. In the present study, we quantify the contribution of this important area source to the VOC burden at a downwind site. Our analysis includes several rarely reported nitrogen containing compounds which appear to have strong pyrogenic sources in this particular study region. Compounds such as amines, amides and isocyanic acid are presently not included in global emission inventories and the default atmospheric chemistry mechanisms, despite their potential importance for secondary aerosol formation and human health. We compare our source-receptor modelling output with several emission inventories such as REAS v2.1, EDGAR v4.3.2 and GAINS

v5 to assess which emission inventory is most consistent with the results of our source receptor modelling study that employs in-situ observations.

2 Methods

2.1 Receptor site

The measurement facility is situated inside Indian Institute for Science Education and Research Mohali (IISER Mohali) campus (Figure 1a) which is a suburban site (30.667 °N, 76.729 °E, 310 m above mean sea level) in Mohali near Chandigarh in India (Figure 1b). Collectively the metropolitan of Chandigarh-Mohali-Panchkula forms a tri-city with a total population of 1,941,118 (Census, 2011). The main air transport toward the site was from the North West and the period studied was impacted by wheat residue burning, a dust storm and strong photochemistry (Sinha et al., 2014). Figure 1a shows 72 h HYSPLIT back trajectories arriving at the site. With average wind speeds of 5.6 $\mathrm{m\,s^{-1}}$ during the study period (range 1-20 $\mathrm{m\,s^{-1}}$) the meteorological conditions were conducive for capturing the contribution of regional emission sources. The measurement site, the meteorology and the primary dataset acquired during May 2012 have been discussed in detail elsewhere (Sinha et al., 2014).

2.2 VOCs and other Auxiliary measurements

We used hourly data of 32 measured organic ions which were assigned to volatile organic compounds (Supplementary Table S1) based on PTR-TOF-MS studies conducted by our group within the South Asian environment (Sarkar et al., 2016) to initialize the US EPA PMF 5.0 model and employed CO, SO_2, O_3 and NO_y as independent tracers to validate the results. ~~Since the technical details of the measurements and the QA/QC protocol have already been~~ As described in greater detail in (Sinha et al., 2014), ~~we provide only a quick summary here. A~~ ambient air sampling was performed continuously through a Teflon inlet line protected by an in-line Teflon filter. A high sensitivity proton transfer reaction quadrupole mass spectrometer PTR-QMS (HS Model 11-07HS-088, Ionicon Analytik Gesellschaft, Austria) was operated at drift tube pressure of 2.2 mbar, a drift tube temperature of 60 °C and a drift tube voltage of 600 V, which resulted in an operating E/N ratio of ∼ 135. Carbon monoxide (CO), Sulphur dioxide (SO_2), Ozone (O_3) and NO_y (NO, NO_2 and other nitrogen species converted to NO by a molybdenum converter such as nitric acid and PAN) were measured using Thermo Fischer Scientific 48i (IR filter correlation based spectroscopy), 43i (pulsed UV fluorescence), 49i (UV absorption photometry) and 42i trace level air quality analysers (chemiluminescence), respectively.

2.3 Positive Matrix Factorisation model

In the current study, US EPA PMF 5.0 receptor model (Norris et al., 2014) was applied to the ambient VOC dataset (in $\mu\mathrm{g\,m^{-3}}$) from May 2012 measured at the IISER-Mohali Atmospheric chemistry facility comprising of data matrix of 721 samples (rows) and 32 species (columns). The EPA PMF 5.0 receptor model (Paatero et al., 2014; Norris et al., 2014) is multivariate factor analysis tool (Paatero and Tapper, 1994; Paatero, 1997), which decomposes the data matrix x_{ij} with i number of samples and j number of measured VOCs into two matrices, the factor contribution matrix g_{ik} (which provides the mass g contributed by each factor to the individual sample) and the factor profiles matrix f_{kj} (which provides the source profile/fingerprint of each individual source). Both matrices are established for a user defined number of sources p from the existing intrinsic variability in the dataset leaving behind a matrix of residuals e_{ij}.

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

A detailed description of the model can be found elsewhere (Paatero and Tapper, 1994; Paatero, 1997; Paatero et al., 2014; Norris et al., 2014). The two primary advantages of the PMF over other source receptor modelling tools are its inherent non-negative constraints (Hopke, 2016) and its capability of optimally weighing individual data points and assigning uncertainties which makes it possible to include less robust species that can be useful for defining real sources. The EPAv5.0 model is superior when compared to other source receptor modelling tools as contains advanced rotational features (Paatero and Hopke, 2009) which allow to constrain the rotational ambiguity in a manner that pushes the PMF solution toward the real world space.

All 32 species were assigned a fixed 20 % in the uncertainty , which represents the largest uncertainty estimated for strong compounds, to avoid a situation where the difference in the assigned uncertainty drives the PMF to dedicate a separate factor towards minimizing Q of a single compound with low uncertainty (toluene) by taking it out of all other source profiles and opening a separate factor profile containing just a single compound. ~~and~~ 18 were identified as weak based on the signal to noise ratio and the presence of potential isobaric interferences as detailed in table S2. For weak species, the PMF model triples the stated uncertainty to reduce their impact on the models solution. Designating sources with isobaric interferences as weak is warranted, because when two sources with different temporal profiles (night-time combustion and daytime biogenic emission or night-time combustion and daytime photochemistry) could potentially contribute different compounds to the same m/z ratio, zero values are almost absent in that particular column of the matrix and the tracer is affected by additional uncertainty not

appropriately expressed by merely looking at the instrumental measurement error and the signal to noise ratio. When this column is made "weak" and given a higher uncertainty, other "strong" tracers, representing a single compound, define source profiles and this reduces the rotational ambiguity of the model. The "weak" compounds with isobaric interferences tend to be distributed among the source profiles available as per the solution that minimizes Q but they do not define any of the profiles. The extra modelling uncertainty was kept to zero and missing values ($< 5\%$) were excluded. For every base model run, we used 20 runs with random seeds. Stable Q-values were obtained for all runs. The model was run with 3 to 7 factors, to identify the appropriate number of factors as discussed in the supplementary text in greater detail. Figure 2 shows the percentage contribution of the identified sources to the VOC burden for these runs. Figures S4 a, b and c show how the factor profile, percentage of each VOC originating from a certain source, and the factor contribution change while increasing the number of factors in the model. Figure 2 shows that a 7 Factor solution provides little advantage over a 6 Factor solution while a 5 Factor solution does not resolve the wheat residue burning source which is independently verified by MODIS (Moderate Resolution Imaging Spectroradiometer) fire counts over the region. The residuals for all species for the 6 Factor solution depicted a normal curve and fall within -3.3 sigma and +3.3 sigma for all species indicating a good model fit. The constraints feature of the 5.0 version of the model was utilised to improve the performance of the model further as described in greater detail in the supplementary text. The constrained model operation of the PMF version 5.0 allows to reduce the rotational ambiguity of the model using external knowledge. For example, if a source is inactive for a particular period (as is photochemistry at night), then the source contribution (g_{ik}) due to that factor during that time period can be pulled to zero in the model to provide more robust output. Similarly, a compound that is known to be present only in primary emissions can be pulled down in the source composition (f_{kj}) matrix of the photochemistry factor. A list of the constraints applied is provided in the supplementary table S3. A detailed discussion of the use of constraints in a receptor model has been provided in previous studies (Paatero et al., 2002, 2014; Paatero and Hopke, 2009; Norris et al., 2014; Sarkar et al., 2016). Bootstrap model runs (Brown et al., 2015) were performed to assess the model uncertainty. Input parameters for the bootstrap runs constituted random seed, 100 number of bootstraps and default values for block size (10) and minimum correlation R-value (0.6) and there were no unmapped factors. Except for the car and two-wheeler factor (R=0.6) for which a certain degree of co-linearity is expected, none of the other factors showed cross correlation with each other (R<0.3) and the g-space plot even of this factor pair is well filled. The constraint mode was unable to force the PMF model to separate the wheat residue burning factor in a 5-factor solution without imposing a split between the car and 2-wheeler factor, indicating that these two indeed represent distinct source profiles.

2.4 Validation of the PMF output

The PMF generates two matrices from the intrinsic variability in the dataset. A factor contribution matrix and a factor profile matrix.

Traditionally the PMF output has been validated by cross-correlating the factor contribution matrix with independent tracers which were not used to initialize the model, but are considered useful tracers for the respective source (Brown et al., 2015; Leuchner and Rappenglück, 2010; Gaimoz et al., 2011; Bon et al., 2011; Sarkar et al., 2016). We perform this validation step for all six source factors resolved by the PMF model. These were identified as "biofuel use and waste disposal", "wheat-residue burning", "four-wheelers", "two-wheelers", "industrial emissions and solvent use" and "mixed daytime sources", respectively. The factor contribution for 4-wheelers (R=0.7) and 2-wheelers (R=0.6) correlated best with the independent tracer NO_y which is considered to be a vehicular exhaust marker (Ramanathan et al., 1985). The factor contribution of the domestic fuel usage and waste disposal factor correlated best with the independent tracer CO (R=0.9), a proxy for inefficient combustion, while the factor contribution of the industrial emission factor correlated best with the independent tracer SO_2 (R=0.6). The wheat residue burning factor days showed a moderate cross correlation with MODIS fire counts with an R=0.4 and a lag of 2 days. O_3 (R=0.8) was the best independent tracer for the mixed daytime factor.

However, our study goes one step further than all previous studies in validating the PMF output. For 5 out of 6 factors we validated the factor profiles generated by the PMF model against grab samples collected at the source. Factor profiles were cross-correlated with the fingerprints of source samples collected from a number of potential sources including wheat residue fires (Chandra et al., 2017; Kumar et al., 2018), a ambient air samples from a busy traffic junction (Chandra et al., 2017)and an industrial area (this study), tailpipes of various vehicles (this study), waste burning (Sharma et al., 2019), leaf litter burning (this study), domestic biofuel use (Stockwell et al., 2016) and brick kilns (Zhong et al., 2019) to identify the sources. Figure 3 shows the factor profiles obtained from the PMF run (in dark blue), the percentage of each species explained by the respective PMF factor (red squares) and the source profiles of those sources which best matched the factor profile (in various colors as indicated in the legend). The factor profile of residential fuel usage and waste disposal correlates most strongly with the measured VOC source speciation profiles of domestic cooking (R=0.8), leaf-litter burning (R=0.7) and smoldering garbage fires (R=0.6), the wheat residue burning factor with flaming wheat residue burning (R=0.9), the 4-wheeler factor with the tailpipe exhaust of petrol-fueled cars (R=0.5), gasoline

evaporation headspace for diesel (R=0.5) and urban traffic junction grab samples (R=0.8) and the 2-wheeler factor with the tailpipe exhaust of petrol-fuelled 4-stroke two-wheelers (R=0.6). The industrial emissions correlated best with the source profile of brick kilns (R=0.5) and ambient air samples collected in an industrial area (0.6). For mixed daytime sources no source profile sampling is possible.

2.5 Conditional Probability Function analysis

We perform a conditional probability function (CPF) analysis (Leuchner and Rappenglück, 2010) which aids in identifying physical locations of different PMF source factors without using back trajectories (Xie and Berkowitz, 2006) The CPF is computed using the factor contribution of the PMF model in combination with the wind direction at the receptor site. It quantifies the probability of factor contributions surpassing a certain threshold ~~by calculating the probability of observing mass concentrations above the~~ (75^{th} percentile) for a particular wind direction sector thereby highlighting directional dependency of source factors and is defined as follows: ~~of a given factor contribution for every wind direction. This aids in identifying physical locations of different PMF source factors without using back trajectories.~~

$$CPF = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \tag{2}$$

Where $m_{\Delta\theta}$ represents the number of data points in the wind direction bin $\Delta\theta$ which exceeded the threshold criterion and $n_{\Delta\theta}$ represents the total number of data points from the same wind direction bin. $\Delta\theta$ was assigned a value of 30°.

2.6 Calculation of the ozone formation potential and SOA formation potential

Ozone production potential (O_3PP) for each of the PMF derived source factors was calculated based on the method used by Sinha and co-workers (Sinha et al., 2012) ~~as described in the supplementary text in greater detail~~ using the following equation:

$$O_3PP = (\sum_i k_{VOC_i+OH}[VOC_i]) \times [OH] \times n \tag{3}$$

using n = 2 and [OH] = 10^6 molecules cm^{-3}. The values were summed up for all the VOCs for obtaining the ozone production potential corresponding to each of the PMF derived factors for the daytime hours (07:00-18:00 LT).

Secondary organic aerosol (SOA) potential was calculated for the PMF source factors using the literature SOA yields (Derwent et al., 2010) under low NO_X conditions for benzene, toluene, ethylbenzene, trimethylbenzene, styrene, methanol, isoprene, formaldehyde, acetaldehyde, acetone, formic acid and acetic acid using the equation given below for 07:00-18:00 LT ~~as described in the supplementary text.~~

$$SOApotential = (\sum_i [VOC_i]) \times [SOA_i] \tag{4}$$

2.7 Methodology for the comparison of PMF source factors with existing emission inventories

Global Emission Database for Global Atmospheric Research (EDGARv4.3.2) inventory for the year 2012 (Huang et al., 2017) and two regional emission inventories: Regional Emission inventory in Asia (REAS v2.1) for the year 2008 (Kurokawa et al., 2013) and the Greenhouse Gas and Air Pollution Interactions and Synergies model (GAINS) (Amann et al., 2011) for the year 2010 (Stohl et al., 2015) were compared with our PMF output. The gridded inventory was filtered for Latitude: 27.4-34.9 °N and Longitude: 72-79.8 °E , i.e. the fetch region from which the air mass trajectories reach the receptor site within one day. This filtering is required because compounds with photochemical lifetimes of less than a day (e.g. styrene, C-8 and C-9 aromatics) feature prominently in several source profiles indicating that most of the transport sector emission were less than a day old when they reached the receptor site. Other compounds with longer lifetimes such as toluene (2 days), benzene (6 days) or acetonitrile (months) can reach the site from more distant sources. The wheat residue burning source shows the highest cross correlation with the regional fire counts for a lag time of 2 days indicating that emissions from distant sources can and do impact the site with a time lag. The chosen fetch region includes the areas where the maximum number of wheat residue burning fire counts are observed while avoiding a size that is too large to be consistent with the relatively unaltered signature of some of the other PMF source profiles.

Annual emissions were available for EDGAR (2012) and GAINS (2010), whereas, REAS provided monthly data (May 2008). However, Figure S6 shows that despite providing monthly data, the REAS emission inventory has very little seasonality for any of the sources.

To facilitate the comparison of the PMF output of the month of May which is affected by a strongly seasonal source (crop residue burning) with emission inventories that provide only annual data as of now, we calculate hypothetical pie charts which attribute annual crop residue burning emissions over the region only to the 2.5 months when crop residue burning actually occurs (middle of October to end of November and May)

3 Results and Discussion

~~3.1 Identification of PMF factors Six source factors were resolved by the PMF model. These were identified as "biofuel use and waste disposal", "wheat-residue burning", "four-wheelers", "two-wheelers", "industrial emissions and solvent use" and "mixed daytime sources", respectively. Factor profiles were cross-correlated with~~

~~the fingerprints of source samples collected from a number of potential sources including wheat residue fires~~ (Chandra et al., 2017; Kumar et al., 2018)~~, a busy traffic junction~~ (Chandra et al., 2017)~~, tail-pipes of various vehicles (this study), waste burning~~ (Sharma et al., 2019)~~, leaf litter burning (this study) and domestic biofuel use~~ (Stockwell et al., 2016) ~~to identify the sources. Figure 3 shows the factor profiles obtained from the PMF run (in dark blue), the percentage of each species explained by the respective PMF factor (red squares) and the source profiles of those sources which best matched the factor profile (in various colors as indicated in the legend). The identification of the factors is further supported by independent tracers such as the criteria air pollutants (NO_y, CO, SO_2, O_3) and MODIS (Moderate Resolution Imaging Spectroradiometer) fire counts as discussed in detail below.~~

3.1 Split up of VOC Emission Sources in Mohali and their contribution to Ozone and SOA Formation Potential

Figure 4 (a) shows the percent contribution of the different sectors to ambient VOC mass concentration loadings during May 2012 in Mohali, while Figure S7 shows a time series of the total VOC mass contributed by the individual factors to the overall mass. The two traffic factors combined together were found to be the strongest contributors to the total VOC mass concentration (25.1 %) followed by biofuel use and waste disposal factor (23.2 %), wheat-residue burning (22.4 %), the mixed daytime factor (15.7 %) and industrial emissions (11.8 %), with the residual not apportioned VOC mass only amounting to 1.7 % of the total. Early source receptor modelling studies from India attributed a slightly larger share 26-58 % of the total VOC mass to traffic related emissions (Srivastava, 2004; Srivastava et al., 2005), suggesting that the progression to the emission norms Bharat stage III & IV (which are equivalent to Euro 3 and Euro 4 norms, http://cpcb.nic.in/vehicular-exhaust/) may have brought down VOC emissions from the traffic sector.

Figure 4 (b) shows the contribution of the different sectors to the ozone formation potential during May 2012 in Mohali. Wheat residue burning factor was found to be the largest contributor to the ozone formation potential (32.4 %) and has been shown to enhance ambient tropospheric ozone mixing ratios by 19 ppb (Kumar et al., 2016). Both traffic sources combined, the mixed daytime sources, biofuel use & waste disposal, and industrial emissions and solvent use contributed 21.9 %, 20.3 %, 18.1 % and 7.3 %, respectively, to the ozone formation potential. It is clear that in order to bring ozone levels into compliance with the NAAQS, the wheat residue burning source of ozone precursors deserves the largest attention at this point, but the transport sector and biofuel use and waste disposal should not be neglected, either.

Figure 4 (c) shows the contribution of the different sectors to the SOA formation potential ($\sim 32\,\mu g\,m^{-3}$) under low NO_x conditions. Traffic is the single largest contributor and is responsible for contributing 59.0 % of the SOA formation potential followed by biofuel use and waste disposal (14.9 %), wheat residue burning (13.9 %), industrial emissions and solvent use (10.1 %) and the mixed daytime factor (2.2 %). While the calculated SOA formation potential particularly from transport sector emissions (Ensberg et al., 2014) and aromatic compounds (Li et al., 2017; Li and Cocker III, 2018) is affected by large uncertainties and may depend in a non-linear fashion on NOx and VOC concentrations (Xu et al., 2015) our calculated SOA formation potential seem to indicate that SOA formation could contribute significantly ~~Total SOA formation potential amounting to $\sim 17\,\mu g\,m^{-3}$, a resultant from all VOC source sectors indicates that at least 16% of~~ to the average $PM_{2.5}$ mass loading (104 $\mu g\,m^{-3}$) ~~for May2012 at IISER-Mohali could be secondary organic aerosols and that transport sector VOC emissions need to be targeted to reduce SOA formation~~.

3.2 Factor 1 - Biofuel use & waste disposal

~~Figure 4 shows that biofuel use & waste disposal contributes 23.2%, 18.1% and 14.9% of the total VOC mass, ozone formation potential and SOA formation potential, respectively. The factor profile correlates most strongly with the measured VOC source speciation profiles of domestic cooking (R=0.8), leaf litter burning (R=0.7) and smoldering garbage fires (R=0.6).~~ The biofuel use and waste disposal factor combines two sources with similar source profiles and high spatio-temporal overlap into one factor. As discussed previously for other South Asian atmospheric environments (Sarkar et al., 2017), the source contributions of domestic biofuel use and domestic waste burning are difficult to segregate ~~due to the high spatio-temporal overlap of the two activities~~. ~~As can be seen in~~ Figure 5~~, the factor~~ shows a weak bimodal behaviour with an early morning and late evening peak for this factor, as both domestic biofuel use and waste disposal fires peak in the early morning and in the evening hours (Nagpure et al., 2015). ~~CO serves as the best independent tracer (Figure 5) indicating that this factor represents a low temperature combustion with a low combustion efficiency. Figure 5 shows that~~ The highest conditional probability for this factor is from the North (>0.4), the direction of the Dadu Majra landfill in Chandigarh, followed by the wind direction NW where a large village (Mauli Baidwan) can be found within 1 km of the receptor and NE, the direction of Panchkula's garbage dump in Sector 23. This and the fact that the average contribution of this factor remains above ~~30~~56 $\mu g\,m^{-3}$ throughout the night indicates that garbage burning contributes significantly to the biofuel use & waste disposal factor.

Figure 3 and Figure 6 show that this factor explains a significant share of the mass of acetonitrile (a biomass burning tracer), aldehydes, ketones, acids, ~~explains 35%, 35%, 29%, 42%, 37%, 34% and 37% of the total acetonitrile, acrolein, methanol, acetaldehyde, methyl vinyl ketone,~~ propyne and propene ~~mass concentration, respectively,~~ in the PMF model. For propene (60%), aldehydes (85%) and ketones

(68%) the residential sector is the dominant source in the most recent speciated emission inventory EDGARv4.3.2. The percentage share for aldehydes and ketones in the inventory is higher than its share in the PMF because the agricultural residue burning source of these compounds is currently missing in the inventory. For acids, however, the residential fuel usage source in the inventory (0.5%) is dwarfed by solvent use associated emissions (96%), while in the PMF the two biomass burning sources (residential biofuel use and waste disposal and wheat residue burning) account for almost 69 % of the total acids in the model. ~~Most of the NMVOC mass in this factor was contributed by methanol (~ 10.6 $\mu g\,m^{-3}$), formic acid (1.9 $\mu g\,m^{-3}$) and acetic acid (7.4 $\mu g\,m^{-3}$).~~ High emission of oxygenated VOCs have been reported previously for source profiles of biofuel-stoves (Wang et al., 2009; Paulot et al., 2011; Stockwell et al., 2016) open waste burning (Sharma et al., 2019) and PMF factors' results of residential biofuel use and waste disposal factor in Kathmandu, Nepal (Sarkar et al., 2017).

It should be noted, that this factor is responsible for approximately 25 % of the total benzene emissions in our PMF model, while emission inventories attribute a larger share (39-74%) of this compound to this source. Since benzene is an identified Group-1 carcinogen (IARC, 1987) and emissions occur within the household itself (domestic cooking) or within close proximity of the house (waste disposal) this factor deserves special attention in programs targeted at emission reductions. However, the impact of such emission reductions in the residential and waste management sector on human benzene exposure are likely to be overestimated by modelling studies using present day emission inventories, as the inventories attribute 39-74% of the benzene emissions to residential fuel usage and waste disposal, while the PMF suggests the transport sector is the largest benzene source (Figure S8a). Direct emission of isocyanic acid, a highly toxic emerging contaminant and its photochemical precursors (Alkyl amines and Amides) was observed from this source and explained 18 % of the isocyanic acid mass concentration and 7-15 % of all the alkyl amines and amides in the PMF model, respectively.

3.3 Factor 2 - Wheat Residue burning

Wheat residue burning takes place every year in the NW-IGP in the post-harvest season and generally peaks in the month of May. It has been shown that wheat residue burning has a major impact on both ozone mixing ratios (Kumar et al., 2016) and VOC mixing ratios and hydroxyl radical reactivity (Kumar et al., 2018) ~~,~~ and result~~sing~~ in a large suite of unknown ($\sim 40\,\%$) and poorly quantified reactive gaseous emissions. ~~Figure 4 shows that wheat residue burning, contributes 22.4% of the total VOC mass concentration, 32.4% of the total ozone formation potential and 13.9% of the total SOA formation potential. Figure 3 shows that the factor profile correlates most strongly with flaming wheat residue burning (R=0.9) and Figure 5 illustrates that the best independent tracer for the average contribution of wheat residue burning to the total NMVOC mass are the daily fire counts with a cross correlation of R=0.4 and a lag of 2 days. Since wheat residue burning is an area source w~~Wheat residue burning ~~and~~ emissions are transported to the receptor site from a large fetch region and often with a significant lag time. Hence, there is no strong conditional probability for enhancements from any specific wind direction (Figure 5).

Figure 3 and Figure 6 show~~s~~ that the ~~four largest contributors to the total NMVOC mass in the~~ wheat residue burning factor explains a significant share of all acids, amines/amides, several ketones, and aldehydes, isoprene/furan, monoterpenes, acetonitrile, propene, styrene and phenol in the PMF model. ~~are acetic acid (11.4 $\mu g\,m^{-3}$), methanol (3.3 $\mu g\,m^{-3}$) acetaldehyde (2.1 $\mu g\,m^{-3}$) and acetone (1.1 $\mu g\,m^{-3}$). Figure 3 and Figure 6 demonstrate that more than 55% of the hydroxyacetone, 37% of the acetic acid, 32% of the total methyl ethyl ketone and 28-39% of the amides/amines as well as 28% of the isocyanic acid mass in the model can be explained by this factor.~~ This makes wheat residue burning the largest contributor to the human exposure to isocyanic acid in the month of May both through direct emissions of isocyanic acid and by virtue of being the largest source for its photochemical precursors.

In the EDGARv4.3.2 the agricultural residue burning source of ketones,aldehydes and acids is missing. On the other hand agricultural waste burning appears to be the dominant anthropogenic isoprene source (94%) in the EDGARv4.3.2 inventory while in our PMF model residential biofuel usage and the transport sector are equally important contributors to the isoprene/furan mass. The monoterpene emissions from agricultural residue burning (6%) in the EDGARv4.3.2 inventory are dwarfed by emissions from solvent use (90%), while in our PMF solution wheat residue burning and the transport sector appear to be the dominant anthropogenic sources of signals at m/z 81 and 137.

3.4 Factor 3 - Industrial emissions and solvent use

~~Figure 4 shows that~~ The source fingerprint of the industrial emissions and solvent use factor ~~jointly contribute 11.8% of the total VOC mass concentration, 7.3% of the total ozone formation potential and 10.1% of the total SOA formation potential.~~ is dominated by ~~M~~methanol (7.3 $\mu g\,m^{-3}$), acetic acid (3.9 $\mu g\,m^{-3}$) and acetone (2.9 $\mu g\,m^{-3}$) ~~are the largest contributors to this factor profile~~. This points towards solvent use (Gaimoz et al., 2011) and/or polymer manufacturing (Sarkar et al., 2017) contributing to the industrial emission and solvent use factor. In addition, Figure 3 and Figure 6 show that this factor explains a significant fraction of the benzene (20 %) and acetonitrile (17 %) mass in the PMF model. While both are known for their use as solvents (Brown et al., 2007), they can also be emitted from the combustion. ~~Figure 5 shows that the factor contribution~~ The EDGARv4.3.2 emission inventory has a strong industrial and solvent source of toluene, xylenes,

acids, formaldehyde and monoterpenes which is not reflected with equal strength in our PMF solution.

-The correlation of the industrial emissions and solvent use factor ~~correlated~~ with the SO_2 time series (R= 0.6), indicates that the emissions of coal or biofuel burning in industrial units and/or coal fired power plants may also be contributing to this factor profile. Figure 5 shows that the highest conditional probability of this factor is to the South East direction (120 ° -150 ° wind sector). The receptor site is downwind of a 600 MW coal fired power plant located in Jagadhri (80 km SE) as well as downwind of several industrial areas and brick kiln clusters located around Dera Bassi (15 km), Lalru (20 km) and Jagadhari (80 km) when the wind blows from this direction. In the Kathmandu valley, biofuel co-fired brick kilns explained a significant fraction of the benzene and acetonitrile mass (Sarkar et al., 2017) and the factor profile shows a moderate correlation with the source signature of brick kiln emissions (R=0.5), hence a combustion contribution from brick kilns to the factor profile cannot be ruled out. The diel profile broadly reflects boundary layer dynamics with factor contributions increasing continuously throughout the night indicating a buildup of constant emissions in the nocturnal boundary layer. Factor contributions peak in the early morning (32-49~~17-26~~ $\mu g\,m^{-3}$ between 5-9 am local time) and the factor contribution of this factor decreases from 9 am onwards after the breakup of the nocturnal boundary layer. This factor has higher average than the median factor contributions at night due to strong plumes (~~max~~ ~~~ 200~~ ~ 375 $\mu g\,m^{-3}$) reaching the receptor when it is downwind of the industrial sector but not during other nights when the wind direction is from rural Punjab (NW) or the urban sector (NE).

3.5 Factor 4 and 5 - cars and two-wheelers

~~Figure 4 shows that cars and two-wheelers contribute 16.2%, 8.9% of the total VOC mass concentration, 16.5%, 5.4% of the total ozone formation potential and 36.9%, 22.1% of the total SOA formation potential, respectively, at the receptor site.~~

~~As can be seen in Figure 3, factor 4 was identified as a factor dominated by car exhaust because it correlated best with the tailpipe exhaust of petrol fueled cars (R=0.5), urban traffic junction grab samples (R=0.8) and the independent tracer NO_y (R=0.7) which is considered to be a vehicular exhaust marker (Ramanathan et al. 1985).~~ The factor profile of the 4-wheeler factor explains a significant share of all aromatic compounds in the PMF model. ~~is characterized by elevated concentration levels of benzene (1.4 $\mu g\,m^{-3}$), toluene (2.3 $\mu g\,m^{-3}$), sum of C-8 aromatics (3.5 $\mu g\,m^{-3}$) and sum of C-9 aromatics (2.7 $\mu g\,m^{-3}$) and explained 35%, 30%, 53% and 58% of the total benzene, toluene, C-8 aromatics and C-9 aromatics mass in the PMF model, respectively.Features of car's factor profile also resemble gasoline evaporation headspace for diesel (R=0.5) collected at a petrol pump. This indicates that the factor profile consists~~ The factor represents a mixture of multiple components contributed by fuel exhaust and fuel evaporative running losses from vehicles and resembles ambient air samples from a busy traffic intersection. Similar profiles have been observed during field measurements in Beirut, Lebanon (Salameh et al., 2014, 2016) and Hong Kong (Ho et al., 2004). The ~~toluene to benzene ratio of this profile (1.4) is typical for traffic emissions (1.5-2.3) (Som et al., 2007; Hoque et al., 2008; Chandra et al., 2018) and the~~ highest conditional probability (Figure7) is observed for the Chandigarh wind sector (0-90 °). As reported previously from Mexico City during the Milagro campaign (Bon et al., 2011), a significant mass of methanol (4.3 $\mu g\,m^{-3}$) and other oxygenated VOCs were present in the traffic emissions factor. The fact that this factor explains 28 % of the total m/z 57 is consistent with the gasoline additive MTBE being detected at this m/z ratio as an interference to acrolein/methylketone (Karl et al., 2003; Warneke et al., 2003, 2005; Rogers et al., 2006). Signals at m/z 31, 47, 59, 61, 73, 87 in aged traffic plumes can be attributed to formaldehyde, formic acid, glyoxal, acetic acid, methylglyoxal and 2-butanedione which are products of the gas phase oxidation of toluene, C-8 and C-9 aromatic compounds (Bethel et al., 2000; Ervens et al., 2004). In addition, car exhaust also explained 34 % of the propyne mass in the model.

Factor 5, ~~was identified as~~ 2-wheeler exhaust, ~~as the factor profile showed the highest correlation with the tailpipe exhaust of petrol fuelled 4-stroke two-wheelers (R=0.6) and the independent tracer NO_y (R=0.6). Toluene (3.9 $\mu g\,m^{-3}$), acetic acid (4.2 $\mu g\,m^{-3}$) and methanol (2.4 $\mu g\,m^{-3}$) feature as the most abundant compounds in this factor profile, which~~ explains 50 % of the total toluene mass as well as 17 %, 12 % and 9 %, of the total C-8 aromatics, benzene and C-9 aromatics in the PMF model, respectively. The factor shows a ~~While part of the~~ signal at m/z 61 (acetic acid) which may partially be due to fragmentation of octane or ethyl acetate (Warneke et al., 2003; Rogers et al., 2006) which could be present in fuel.~~,~~ ~~t~~The mass has also been attributed to acetic acid in a previous study of diesel tailpipe emissions (Jobson et al., 2005). Nevertheless, it still seems that the 2-wheeler factor profile has a higher contribution from oxidised compounds compared to the car factor profile indicating that the plumes are typically more aged. Figure 7 shows that this factor displays higher conditional probability than the car factor towards the towns Kharar (8 km N), Dera Bassi (15 km SE) and Lalru (20 km SE), and a lower conditional probability than the car factor towards Chandigarh (NE) indicating 2-wheelers are more abundant in small towns, while cars dominate the traffic emissions in urban Chandigarh.

Figure 7 illustrates that both the traffic factors show bimodal peaks in morning (~~10.3~~19 $\mu g\,m^{-3}$ at 5-9 am local time) and evening (~~20~~38 $\mu g\,m^{-3}$ at 7-9 pm local time) during peak traffic hours. Mass loadings during evening rush hour are higher than during morning rush hour, because peak morning traffic occurs after the breakup of the nocturnal boundary layer, while in the evening emissions accumulate in the shallow nocturnal boundary layer. When the wind blows

from the urban sector (0-90 °) during peak traffic hour (7-9 pm) peak factor contributions of >~~140~~260 $\mu g\,m^{-3}$ for cars and >~~0~~150 $\mu g\,m^{-3}$ for 2-wheelers are observed.

As can be seen from Figure 6, the two traffic factors jointly explain 47 %, 80 %, 70 % and 67 % of the total benzene, toluene, C-8 and C-9 aromatic compounds in the model consistent with findings from the Kathmandu valley that traffic, not residential biofuel use and waste disposal is the more important source of aromatic compounds in South Asia. It is also clear that despite stringent regulations, the transport sector in the region is still the largest contributor to human benzene exposure. It can be seen from Figure S8a-d that at present, various emission inventories consider the transport sector to be a minor source of benzene (10-16%). The EDGAR v4.3.2 emission inventory also considers the transport sector to be only a minor source of, toluene (11-15%) and xylenes (17-22%). Residential fuel usage, industries and solvent use are considered to be the most significant year around source of benzene, toluene and xylenes in Edgar v4.3.2. Agricultural residue burning becomes the most significant source of all aromatic compounds in the EDGAR v4.3.2 emission inventory when crop residue burning emissions are treated as occurring during crop residue burning season only, which may imply that the annual emissions of aromatic compounds from the stubble burning may be overestimated. REAS v.2.1 appears to be overestimating the residential fuel burning contribution to benzene and toluene emissions and the solvent usage contribution to toluene emissions. However, it captures the contribution of the transport sector to xylenes and trimethylbenzenes emissions well. ~~and C-9 aromatic compounds.~~

3.6 Factor 6 - mixed daytime sources

Figures 4 and 6 shows that mixed daytime sources comprising of biogenic emissions and photochemically formed compounds ~~contribute 15.7% of the total VOC mass, 20.3% of the total ozone formation potential and 2.2% of the total SOA formation potential. Figure 7 illustrates that the mixed daytime factor correlates most strongly with the independent tracer O_3 (R= 0.8). It can be seen from Figure 6 that biogenic daytime emissions~~ explained 22 % of the monoterpenes and 25 % of the measured isoprene, respectively. Isoprene has a short chemical lifetime of 1.5 hours during the day and 16 % and 11 % of its first generation oxidation products MVK and MEK (Kesselmeier and Staudt, 1999) were also attributed to this factor . In addition, the mixed daytime factor explains 41 %, 44 %, 24 % and 22 % of the total formaldehyde, formic acid/ethanol, methanol and acetone mass, respectively. Photochemically formed isocyanic acid, formamide, acetamide and propanamide explain a slightly lower fraction (27-37 %) of the total mass concentration of these compounds compared to what has been reported from wintertime Kathmandu valley (36-41 %). ~~Methanol (8.9 $\mu g\,m^{-3}$), formic acid (4.4 $\mu g\,m^{-3}$), acetic acid (2.5 $\mu g\,m^{-3}$) were the most important contributors to the factor fingerprint.~~ Figure 7 illustrates that the mixed daytime factor peaks between 9 am and 4 pm and shows a slightly enhanced conditional probability for the 180 ° -330 ° rural wind sector (0.2-0.3) due to agroforestry plantations of poplar in the rural landscape.

3.7 Comparison of PMF source factors with existing Emission Inventories

~~Global Emission Database for Global Atmospheric Research (EDGARv4.2) inventory for the year 2008~~ (EDGARv4.2, 2011) ~~and two regional emission inventories: Regional Emission inventory in Asia (REAS v2.1) for the year 2008~~ (Kurokawa et al., 2013) ~~and the Greenhouse Gas and Air Pollution Interactions and Synergies model (GAINS)~~ (Amann et al., 2011) ~~for the year 2010~~ (Stohl et al., 2015) ~~were compared with our PMF output. The gridded inventory was filtered for Latitude: 27.4-34.9 °N and Longitude: 72-79.8 °E, i.e. the fetch region from which the air mass trajectories reach the receptor site within one day. Annual emissions were available for EDGAR (2008) and GAINS (2010), whereas, REAS provided monthly data (May 2008). However, Figure S5 shows that despite providing monthly data, the REAS emission inventory has very little seasonality for any of the sources.~~

Figure 8 shows pie charts depicting the contribution of different sectors to the total VOC mass burden for the emission inventories and our PMF output. Biofuel use and waste disposal were responsible for 28.1 % of the mass in our PMF but ~~67.9~~39 %, 44.2 % and 41.7 % of the mass in EDGARv4.3.2, GAINS and REASv2.1 respectively. The contribution of crop residue burning (27.1 %) to the VOC mass in the month of May would be highly underestimated by both GAINS (7 %) and EDGARv4.3.2 (~~4.7~~6 %) if the annual emissions are attributed equally to all months of the year. However, if both emission inventories would attribute their annual crop residue burning emissions over the region only to the 2.5 months when crop residue burning actually occurs (middle of October to end of November and May), these emission inventories could be reconciled with the PMF solution, as emissions in May would amount to 26.5 % and ~~19.2~~23 % of the monthly VOC emissions for the month of May for GAINS and EDGARv4.3.2, respectively as shown in Figure 8. At the same time the percentage share of domestic fuel use and waste disposal would drop to ~~54~~32 % and 35 % in EDGARv4.3.2 and GAINS, respectively and the contribution of industrial emissions and solvent use would drop~~s~~ to 18 % in GAINS and ~~11~~30 % in EDGAR, respectively. Our PMF (14.3 %) ~~lies in between the estimate of these two emission inventories~~ solution indicates that industrial emissions and solvent usage (14.3%) are currently overestimated in all emission inventories but are closest to GAINS (540 $Gg\,y^{-1}$, 18%) for industrial emissions and solvent use . For domestic biofuel use and waste disposal EDGARv4.3.2 (968 $Gg\,y^{-1}$, 32%) appears to agree best with our PMF solution. ~~but closer to GAINS~~ F~~f~~or wheat residue burning GAINS agrees well with our PMF output, while the agricultural waste burn-

ing emissions of some of the detected compound groups (ketones, aldehydes and acids) appear to be missing in the EDGARv4.3.2 inventory ~~and domestic biofuel use and waste disposal~~. ~~REAS overestimate the contribution of industrial activity and solvent use in the month of May (22%).~~ Our PMF solution for road transport sector emissions (30.5 %) lies in between the estimates of GAINS (558 $\mathrm{Gg\,y^{-1}}$, 24 %) and REAS (1230 $\mathrm{Gg\,y^{-1}}$, 36.2 %), possibly, because not all pre-2000 super-emitters for which the 20-year vehicle lifetime has been exceeded have been retired as planned.

Overall it appears that ~~GAINS, the emission inventory with the lowest absolute emissions from residential and commercial biofuel use shows the best agreement with our PMF solution~~none of the emission inventories is ideal at the present. Our PMF solution suggests that transport sector emissions may be are underestimated by ~~approximately a factor of 1.5 in~~ GAINS and EDGARv4.3.2, while the combined effect of residential biofuel use and waste disposal emissions as well as the VOC burden associated with solvent use may be overestimated by ~~a factor of 1.3 in the same~~ all emission inventor~~y~~ies. Similar results have been reported previously. Sarkar and co-workers (Sarkar et al., 2017) reported an underestimation of transport sector emissions for the REAS and EDGAR emission inventory for the Kathmandu valley in Nepal and an overestimation of the residential biofuel use and waste disposal source in all emission inventories, while Gaimoz and co-workers (Gaimoz et al., 2011) reported an overestimation of the VOC emissions from solvent use in Paris.

~~REAS and EDGAR overestimated residential bio fuel usage emissions even more than GAINS. EDGAR underestimated transport sector emissions and industrial emissions and solvent usage while REAS overestimates the importance of the same two sources. REAS also fails to include agricultural residue burning as a source.~~

~~Our results highlight that for accurate air quality forecasting and modelling it is essential that emissions are attributed only to the months in which the activity actually occurs. This is important both for emissions from crop residue burning (which occur in May and from Mid-October to the end of November) and emissions from wildfires (which are restricted to the dry season and peak in April and May). Annually averaged emissions are unlikely to yield accurate air quality forecast in regions affected by such seasonal events. At present, more specialized fire emission inventories such as FINN~~ (Wiedinmyer et al., 2011) ~~must be used to account for the full seasonality and day to day variations of open burning emissions. We also demonstrate, that the source profiles obtained as PMF output can be validated and matched against samples collected at the potential sources to validate the factor identification.~~

~~We find that the GAINSv5.0 emission inventory for the year 2010 agreed best with the in-situ data derived PMF solution for May 2012.~~

4 Conclusions

Our results highlight that for accurate air quality forecasting and modelling it is essential that emissions are attributed only to the months in which the activity actually occurs. This is important for emissions from crop residue burning (which occur in May and from Mid-October to the end of November). Annually averaged emissions are unlikely to yield accurate air quality forecast in regions affected by such seasonal events. At present, more specialized fire emission inventories such as FINN (Wiedinmyer et al., 2011) must be used to account for the full seasonality and day to day variations of open burning emissions. We also demonstrate, that the source profiles obtained as PMF output can be validated and matched against samples collected at the potential sources to validate the factor identification. ~~Six VOC emission sources were extracted via PMF simulations from the dataset comprising of 32 VOC species measured online at primary temporal resolution of 1 minute at a sub-urban site in Mohali in the summer of 2012. US EPA PMF 5.0 Model was used for source apportionment of VOCs and PMF-resolved factors included traffic exhaust, biofuel use and waste disposal, wheat residue burning and mixed daytime sources (comprising of biogenic emissions and photochemical formation), industrial emissions and solvent use, which along with the residuals,accounted for 25.1%, 23.2%, 22.4%, 15.7%, 11.8% and 1.7%, respectively, of the total VOC mass concentration.~~

For the human class I carcinogen benzene, the traffic factor alone contributed to 47 % of the total benzene mass at this receptor site followed by residential biofuel use and waste disposal (25 %) and industrial emissions and solvent use (20 %). This stands in stark contrast to various emission inventories which estimate the transport sector contribution to the benzene exposure as (10%) and consider residential biofuel use, agricultural residue burning and industries to be more important benzene sources. Since the annual NAAQS for benzene is exceeded at this receptor site (Chandra and Sinha, 2016), all three sectors must be targeted for emission reductions.

For the emerging contaminant isocyanic acid, photochemical formation from precursors (37 %), wheat residue burning (25 %) and biofuel usage and waste disposal (18 %) were the largest contributors to human exposure. The monthly average isocyanic mixing ratio of 1.4 ppb exceeds concentrations that can, after dissociation at blood pH, result in blood cyanate ion concentrations (Roberts et al., 2011) high enough to produce significant health effects in humans (Wang et al., 2007) such as atherosclerosis, cataracts and rheumatoid arthritis due to protein damage. Peak mixing ratios of this compound exceed 3 ppb in some night time wheat residue burning plumes. Wheat residue burning was also the single largest source of the photochemical precursors of isocyanic acid, namely, formamide, acetamide and propanamide, indicating that this source must be most ur-

gently targeted to reduce human concentration exposure to isocyanic acid.

~~Our results highlight that for accurate air quality forecasting and modelling it is essential that emissions that are both large in terms of their absolute contribution and display a significant seasonality in their occurrence are attributed only to the months in which the activity actually occurs. This is important both for emissions from crop residue burning (which occur in May and from Mid-October to the end of November) and emissions from wildfires (which are restricted to the dry season and peak in April and May). Annually averaged emissions are unlikely to yield accurate air quality forecast in regions affected by such seasonal events. We find that the GAINSv5.0 emission inventory for the year 2010 was best agreed with the in-situ data derived PMF solution for May 2012, as long as crop residue burning emissions were attributed to 2.5 months of the year only, and emissions from domestic biofuel use and solvent use were scaled down by a factor of 1.3 and transport sector emissions were scaled up by a factor of 1.5. The quantitative source apportionment results reported in this study for benzene, isocyanic acid and ozone and SOA precursors will provide much needed information for targeted mitigation efforts to improve the regional air quality.~~ Overall it appears that none of the emission inventories is ideal at the present. Our PMF solution suggests that transport sector emissions may be underestimated by GAINSv5.0 and EDGARv4.3.2, while the combined effect of residential biofuel use and waste disposal emissions as well as the VOC burden associated with solvent use may be overestimated by all emission inventories. Agricultural waste burning emissions of some of the detected compound groups (ketones, aldehydes and acids) are currently missing in the EDGARv4.3.2 inventory while aromatic emissions from the same source appear to be overestimated. Thus, large improvements are required in existing emission inventories for correct source attribution and inclusion of missing compounds over this densely populated region of the world.

Data availability. Data is available from the corresponding author upon request.

Author contributions. Pallavi performed the analysis and wrote the first draft of the paper. Dr. Baerbel Sinha conceived the analysis and revised the paper draft. Dr. Vinayak Sinha collected the data and commented on the paper draft.

Competing interests. The authors have no competing interests to declare.

Acknowledgements. We acknowledge the IISER Mohali Atmospheric Chemistry facility for data and the Ministry of Human Resource Development (MHRD), India for funding the facility. Pallavi acknowledges IISER Mohali for Institute PhD fellowship. This work was also partially supported through grant (SPLICE) DST/CCP/MRDP/100/2017(G) under the National Mission on Strategic knowledge for Climate Change (NMSKCC) MRDP Program of the Department of Science and Technology, India.

References

Amann, M., Bertok, I., Borken-Kleefeld, J., Cofala, J., Heyes, C., Hoeglund-Isaksson, L., Klimont, Z., Nguyen, B., Posch, M., Rafaj, P., Sandler, R., Schoepp, W., Wagner, F., and Winiwarter, W.: Cost-effective control of air quality and greenhouse gases in Europe: Modeling and policy applications, Environ. Model Softw., 26, 1489–1501., https://doi.org/10.1016/j.envsoft.2011.07.012, (http://www.iiasa.ac.at/web/home/research/researchPrograms/air/Asia.html), 2011.

Bethel, H. L., Atkinson, R., and Arey, J.: Products of the gas-phase reactions of OH radicals with p-xylene and 1, 2, 3- and 1, 2, 4-trimethylbenzene: effect of NO2 concentration, J. Phys. Chem. A, 104, 8922–8929, https://doi.org/10.1021/jp001161s, 2000.

Bon, D. M., Ulbrich, I. M., de Gouw, J. A., Warneke, C., Kuster, W. C., Alexander, M. L., Baker, A., Beyersdorf, A. J., Blake, D., Fall, R., Jimenez, J. L., Herndon, S. C., Huey, L. G., Knighton, W. B., Ortega, J., Springston, S., and Vargas, O.: Measurements of volatile organic compounds at a suburban ground site (T1) in Mexico City during the MILAGRO 2006 campaign: measurement comparison, emission ratios, and source attribution, Atmos. Chem. Phys., 11, 2399–2421, https://doi.org/10.5194/acp-11-2399-2011, 2011.

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, https://doi.org/10.1016/j.atmosenv.2006.08.021, 2007.

Brown, S. G., Eberly, S., Paatero, P., and Norris, G. A.: Methods for estimating uncertainty in PMF solutions: Examples with ambient air and water quality data and guidance on reporting PMF results, Sci. Total Environ., 518, 626–635, https://doi.org/10.1016/j.scitotenv.2015.01.022, 2015.

Census: Government of India, Ministry of Home Affairs, Office of the Registrar General & Census Commissioner, India. http://www.censusindia.gov.in/pca/Searchdata.aspx- accessed on 25july2018, 2011.

Chandra, B., Sinha, V., Hakkim, H., and Sinha, B.: Storage stability studies and field application of low cost glass flasks for analyses of thirteen ambient VOCs using proton transfer reaction mass spectrometry, Int. J. Mass Spectrom., 419, 11–19, https://doi.org/10.1016/j.ijms.2017.05.008, 2017.

Chandra, B., Sinha, V., Hakkim, H., Kumar, A., Pawar, H., Mishra, A., Sharma, G., Garg, S., Ghude, S. D., and Chate, D.: Odd-even traffic rule implementation during winter 2016 in Delhi did not reduce traffic emissions of VOCs, carbon dioxide, methane and carbon monoxide, Curr. Sci., 114, https://doi.org/10.18520/cs/v114/i06/1318-1325, 2018.

Chandra, B. P. and Sinha, V.: Contribution of post-harvest agricultural paddy residue fires in the NW Indo-Gangetic Plain to ambient carcinogenic benzenoids, toxic isocyanic acid and carbon monoxide, Environ. Int., 88, 187–197, https://doi.org/10.1016/j.envint.2015.12.025, 2016.

Derwent, R. G., Jenkin, M. E., Utembe, S. R., Shallcross, D. E., Murrells, T. P., and Passant, N. R.: Secondary or-

ganic aerosol formation from a large number of reactive man-made organic compounds, Sci. Total Environ., 408, 3374–3381, https://doi.org/10.1016/j.scitotenv.2010.04.013, 2010.

EDGARv4.2: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR), release version 4.2., 2011.

Ensberg, J. J., Hayes, P. L., L., J. J., Gilman, J. B., Kuster, W. C., de Gouw, J. A., Holloway, J. S., Gordon, T. D., Jathar, S., Robinson, A. L., and H., S. J.: Emission factor ratios, SOA mass yields, and the impact of vehicular emissions on SOA formation, Atmos. Chem. Phys., 14, 2383–239, https://doi.org/doi:10.5194/acp-14-2383-2014, 2014.

Ervens, B., Feingold, G., Frost, G. J., and Kreidenweis, S. M.: A modeling study of aqueous production of dicarboxylic acids: 1. Chemical pathways and speciated organic mass production, J. Geophys. Res. Atmos., 109, D15 205, https://doi.org/10.1029/2003JD004387, 2004.

Gaimoz, C., Sauvage, S., Gros, V., Herrmann, F., Williams, J., Locoge, N., Perrussel, O., Bonsang, B., d'Argouges, O., and Sarda-Estève, R.: Volatile organic compounds sources in Paris in spring 2007. Part II: source apportionment using positive matrix factorisation, Environ. Chem., 8, 91–103, https://doi.org/10.1071/EN10067, 2011.

Ho, K., Lee, S., Guo, H., and Tsai, W.: Seasonal and diurnal variations of volatile organic compounds (VOCs) in the atmosphere of Hong Kong, Sci. Total Environ., 322, 155–166, https://doi.org/10.1016/j.scitotenv.2003.10.004, 2004.

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

Hoque, R. R., Khillare, P., Agarwal, T., Shridhar, V., and Balachandran, S.: Spatial and temporal variation of BTEX in the urban atmosphere of Delhi, India, Sci. Total Environ., 392, 30–40, https://doi.org/10.1016/j.scitotenv.2007.08.036, 2008.

Huang, G., Brook, R., Crippa, M., Janssens-Maenhout, G., Schieberle, C., Dore, C., Guizzardi, D., Muntean, M., Schaaf, E., and Friedrich, R.: Speciation of anthropogenic emissions of non-methane volatile organic compounds: a global gridded data set for 1970–2012, Atmos. Chem. Phys., 17, 7683–7701, https://doi.org/doi.org/10.5194/acp-17-7683-2017, 2017.

IARC: International Agency for Research on Cancer. Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs Volumes 1 to 42, Supplement 7, 1987.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, 10.1017/CBO9781107415324., https://doi.org/10.1017/CBO9781107415324, 2013.

Jobson, B., Alexander, M. L., Maupin, G. D., and Muntean, G. G.: On-line analysis of organic compounds in diesel exhaust using a proton transfer reaction mass spectrometer (PTR-MS), Int. J. Mass Spectrom., 245, 78–89, https://doi.org/10.1016/j.ijms.2005.05.009, 2005.

Karl, T., Jobson, T., Kuster, W. C., Williams, E., Stutz, J., Shetter, R., Hall, S. R., Goldan, P., Fehsenfeld, F., and Lindinger, W.: Use of protontransferreaction mass spectrometry to characterize volatile organic compound sources at the La Porte super site during the Texas Air Quality Study 2000, J. Geophys. Res. Atmos., 108, D16, https://doi.org/10.1029/2002JD003333, 2003.

Kesselmeier, J. and Staudt, M.: Biogenic volatile organic compounds (VOC): an overview on emission, physiology and ecology, J. Atmos. Chem., 33, 23–88, https://doi.org/10.1023/A:1006127516791, 1999.

Kumar, V., Sarkar, C., and Sinha, V.: Influence of postharvest crop residue fires on surface ozone mixing ratios in the NW IGP analyzed using 2 years of continuous in situ trace gas measurements, J. Geophys. Res. Atmos., 121, 3619–3633, https://doi.org/10.1002/2015JD024308, 2016.

Kumar, V., Chandra, B., and Sinha, V.: Large unexplained suite of chemically reactive compounds present in ambient air due to biomass fires, Sci. Rep., 8, 626, https://doi.org/10.1038/s41598-017-19139-3, 2018.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2, Atmos. Chem. Phys., 13, 11 019–11 058. (https://www.nies.go.jp/REAS/), https://doi.org/10.5194/acp-13-11019-2013, 2013.

Leuchner, M. and Rappenglück, B.: VOC source–receptor relationships in Houston during TexAQS-II, Atmos. Environ., 44, 4056–4067, https://doi.org/10.1016/j.atmosenv.2009.02.029, 2010.

Li, J., Zhang, M., Wu, F., Sun, Y., and Tang, G.: Assessment of the impacts of aromatic VOC emissions and yields of SOA on SOA concentrations with the air quality model RAMS-CMAQ, Atmos. Environ., 158, 105–115, https://doi.org/https://doi.org/10.1016/j.atmosenv.2017.03.035, 2017.

Li, W. and Cocker III, D. R.: Assessment of the impacts of aromatic VOC emissions and yields of SOA on SOA concentrations with the air quality model RAMS-CMAQ, Atmos. Environ., 184, 17–23, https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.03.059, 2018.

Majumdar, D., Mukherjee, A., and Sen, S.: Apportionment of Sources to Determine Vehicular Emission Factors of BTEX in Kolkata, India, Water Air Soil Pollut., 209, 379–388, https://doi.org/DOI 10.1007/s11270-008-9951-1, 2009.

Nagpure, A. S., Ramaswami, A., and Russell, A.: Characterizing the spatial and temporal patterns of open burning of municipal solid waste (MSW) in Indian cities, Environ. Sci. Technol., 49, 12 904–12 912, https://doi.org/10.1021/acs.est.5b03243, 2015.

Norris, G., Duvall, R., Brown, S., and Bai, S.: EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, EPA/600/R-14/108, https://www.epa.gov/sites/production/files/2015-03/epa$_p mf_5.0_s etup.exe, https: //www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user guide.pdf, 124, https: //doi.org/https: //www.epa.gov/sites/production/files/2015-03/epa_pmf_5.0_setup.exe, https:

//www.epa.gov/sites/production/files/2015 − 02/documents/pmf_5.0_user_guide.pdf, 2014.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemom. Intell. Lab. Syst., 37, 23–35, https://doi.org/10.1016/S0169-7439(96)00044-5, 1997.

Paatero, P. and Hopke, P. K.: Rotational tools for factor analytic models, J. Chemometr., 23, 91–100, https://doi.org/10.1002/cem.1197, 2009.

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

Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, Chemometr. Intell. Lab., 60, 253–264, https://doi.org/10.1016/S0169-7439(01)00200-3, 2002.

Paatero, P., Eberly, S., Brown, S. G., and Norris, G. A.: Methods for estimating uncertainty in factor analytic solutions, Atmos. Meas. Tech., 7, 23–35, https://doi.org/10.5194/amt-7-781-2014, 2014.

Paulot, F., Wunch, D., Crounse, J. D., Toon, G., Millet, D. B., DeCarlo, P. F., Vigouroux, C., Deutscher, N. M., González Abad, G., and Notholt, J.: Importance of secondary sources in the atmospheric budgets of formic and acetic acids, Atmos. Chem. Phys., 11, 1989–2013, https://doi.org/10.5194/acp-11-1989-2011, 2011.

Pawar, H., Garg, S., Kumar, V., Sachan, H., Arya, R., Sarkar, C., Chandra, B., and Sinha, B.: Quantifying the contribution of long-range transport to particulate matter (PM) mass loadings at a suburban site in the north-western Indo-Gangetic Plain (NW-IGP), Atmos. Chem. Phys., 15, 9501–9520, https://doi.org/10.5194/acp-15-9501-2015, 2015.

Ramanathan, V., Cicerone, R. J., Singh, H. B., and Kiehl, J. T.: Trace gas trends and their potential role in climate change, J. Geophys. Res. Atmos., 90, 5547–5566, https://doi.org/10.1029/JD090iD03p05547, 1985.

Roberts, J. M., Veres, P. R., Cochran, A. K., Warneke, C., Burling, I. R., Yokelson, R. J., Lerner, B., Gilman, J. B., Kuster, W. C., Fall, R., and de, G. J.: OIsocyanic acid in the atmosphere and its possible link to smokerelated health effects, Proc. Natl. Acad. Sci., 108, 8966–8971, 2011.

Rogers, T., Grimsrud, E., Herndon, S., Jayne, J., Kolb, C. E., Allwine, E., Westberg, H., Lamb, B., Zavala, M., and Molina, L.: On-road measurements of volatile organic compounds in the Mexico City metropolitan area using proton transfer reaction mass spectrometry, Int. J. Mass Spectrom., 252, 26–37, https://doi.org/10.1016/j.ijms.2006.01.027, 2006.

Salameh, T., Afif, C., Sauvage, S., Borbon, A., and Locoge, N.: Speciation of non-methane hydrocarbons (NMHCs) from anthropogenic sources in Beirut, Lebanon, Environ. Sci. Pollut. Res., 21, 10 867–10 877, https://doi.org/10.1007/s11356-014-2978-5, 2014.

Salameh, T., Sauvage, S., Afif, C., Borbon, A., and Locoge, N.: Source apportionment vs. emission inventories of non-methane hydrocarbons (NMHC) in an urban area of the Middle East: local and global perspectives, Atmos. Chem. Phys., 16, 3595–3607, https://doi.org/10.5194/acp-16-3595-2016, 2016.

Sarkar, C., Sinha, V., Kumar, V., Rupakheti, M., Panday, A., Mahata, K. S., Rupakheti, D., Kathayat, B., and Lawrence, M. G.: Overview of VOC emissions and chemistry from PTR-TOF-MS measurements during the SusKat-ABC campaign: high acetaldehyde, isoprene and isocyanic acid in wintertime air of the Kathmandu Valley, Atmos. Chem. Phys., 16, 3979–4003, https://doi.org/10.5194/acp-16-3979-2016, 2016.

Sarkar, C., Sinha, V., Sinha, B., Panday, A. K., Rupakheti, M., and Lawrence, M. G.: Source apportionment of NMVOCs in the Kathmandu Valley during the SusKat-ABC international field campaign using positive matrix factorization, Atmos. Chem. Phys., 17, 8129–8156, https://doi.org/10.5194/acp-17-8129-2017, 2017.

Sharma, G., Sinha, B., Jangra, P., Hakkim, H., Chandra, B. P., Kumar, A., and Sinha, V.: Gridded emissions of CO, NOx, SO2, CO2, NH3, HCl, CH4, PM2.5, PM10, BC and NMVOC from open municipal waste burning in India, Environ. Sci. Technol., 53, 4765–4774, 2019.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J.-F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmos. Chem. Phys., 14, 9317–9341, https://doi.org/10.5194/acp-14-9317-2014, 2014.

Sinha, V., Williams, J., Diesch, J., Drewnick, F., Martinez, M., Harder, H., Regelin, E., Kubistin, D., Bozem, H., and Hosaynali-Beygi, Z.: Constraints on instantaneous ozone production rates and regimes during DOMINO derived using in-situ OH reactivity measurements, Atmos. Chem. Phys., 12, 7269–7283, https://doi.org/10.5194/acp-12-7269-2012, 2012.

Sinha, V., Kumar, V., and Sarkar, C.: Chemical composition of pre-monsoon air in the Indo-Gangetic Plain measured using a new air quality facility and PTR-MS: high surface ozone and strong influence of biomass burning, Atmos. Chem. Phys., 14, 5921–5941, https://doi.org/10.5194/acp-14-5921-2014, https://www.atmos-chem-phys.net/14/5921/2014/, 2014.

Som, D., Dutta, C., Chatterjee, A., Mallick, D., Jana, T., and Sen, S.: Studies on commuters' exposure to BTEX in passenger cars in Kolkata, India, Sci. Total Environ., 372, 426–432, https://doi.org/10.1016/j.scitotenv.2006.09.025, 2007.

Srivastava, A.: Source apportionment of ambient VOCS in Mumbai city, Atmos. Environ., 38, 6829–6843, https://doi.org/10.1016/j.atmosenv.2004.09.009, 2004.

Srivastava, A., Sengupta, B., and Dutta, S.: Source apportionment of ambient VOCs in Delhi City, Sci. Total Environ., 343, 207–220, https://doi.org/10.1016/j.scitotenv.2004.10.008, 2005.

Stockwell, C. E., Christian, T. J., Goetz, J. D., Jayarathne, T., Bhave, P. V., Praveen, P. S., Adhikari, S., Maharjan, R., DeCarlo, P. F., and Stone, E. A.: Nepal Ambient Monitoring and Source Testing Experiment (NAMaSTE): emissions of trace gases and light-absorbing carbon from wood and dung cooking fires, garbage and crop residue burning, brick kilns, and other sources, Atmos. Chem. Phys., 16, 11 043–11 081, https://doi.org/10.5194/acp-16-11043-2016, 2016.

Stohl, A., Aamaas, B., Amann, M., Baker, L. H., Bellouin, N., Berntsen, T. K., Boucher, O., Cherian, R., Collins, W., and Daskalakis, N.: Evaluating the climate and air quality impacts of short-lived pollutants, Atmos. Chem. Phys., 15, 10 529–10 566, https://doi.org/10.5194/acp-15-10529-2015, 2015.

Wang, S., Wei, W., Du, L., Li, G., and Hao, J.: Characteristics of gaseous pollutants from biofuel-stoves in rural China, Atmos. Environ., 43, 4148–4154, https://doi.org/10.1016/j.atmosenv.2009.05.040, 2009.

Wang, Z., Nicholls, S. J., Rodriguez, E. R., Kummu, O., Hörkkö, S., Barnard, J., Reynolds, W. F., Topol, E. J., DiDonato, J. A., and L, H. S.: Protein carbamylation links inflammation, smoking, uremia and atherogenesis, Nat. Med., 13, 1176–1184, 2007.

Warneke, C., De Gouw, J. A., Kuster, W. C., Goldan, P. D., and Fall, R.: Validation of atmospheric VOC measurements by proton-transfer-reaction mass spectrometry using a gas-chromatographic preseparation method, Environ. Sci. Technol., 37, 2494–2501, https://doi.org/10.1021/es026266i, 2003.

Warneke, C., Kato, S., de Gouw, J. A., Goldan, P. D., Kuster, W. C., Shao, M., Lovejoy, E. R., Fall, R., and Fehsenfeld, F. C.: Online volatile organic compound measurements using a newly developed proton-transfer ion-trap mass spectrometry instrument during New England Air Quality Study Intercontinental Transport and Chemical Transformation 2004: Performance, intercomparison, and compound identification, Environ. Sci. Technol., 39, 5390–5397, https://doi.org/10.1021/es050602o, 2005.

Wiedinmyer, C., Akagi, S., Yokelson, R. J., Emmons, L., Al-Saadi, J., Orlando, J., and Soja, A.: The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning, Geosci. Model Dev., 4, 625, https://doi.org/10.5194/gmd-4-625-2011, 2011.

Xie, Y. and Berkowitz, C. M.: The use of positive matrix factorization with conditional probability functions in air quality studies: an application to hydrocarbon emissions in Houston, Texas, Atmos. Environ., 40, 3070–3091, https://doi.org/10.1016/j.atmosenv.2005.12.065, 2006.

Xu, J., Griffin, R. J., Liu, Y., Nakao, S., and Cocker III, D. R.: Simulated impact of NOx on SOA formation from oxidation of toluene and m-xylene, Atmos. Environ., 101, 217e225, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2014.11.008, 2015.

Zhong, M., Saikawa, E., Avramov, A., Chen, C., Sun, B., Ye, W., Keene, W. C., Yokelson, R. J., Jayarathne, T., Stone, E. A., Rupakheti, M., and Panday, A. K.: Nepal Ambient Monitoring and Source Testing Experiment (NAMaSTE): emissions of particulate matter and sulfur dioxide from vehicles and brick kilns and their impacts on air quality in the Kathmandu Valley, Nepal, Atmos. Chem. Phys., 19, 8209–8228, https://doi.org/doi.org/10.5194/acp-19-8209-2019, 2019.

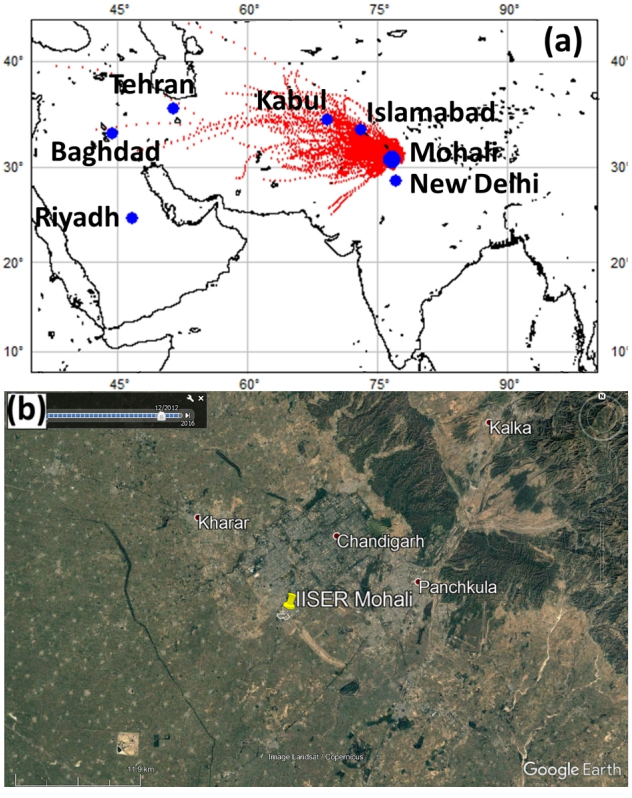

Figure 1. (a) Mohali located on Indian Subcontinent with the overlaid 72 h air mass back trajectories for May 2012 at 09:00 LT and 23:00 LT (UTC+5:30) (b) Precise location of IISER-Mohali Atmospheric chemistry facility (30.667 °N, 76.729 °E, 310 m above mean sea level) with nearby cities on Google Earth imagery. The campus of IISER Mohali is outlined in white.

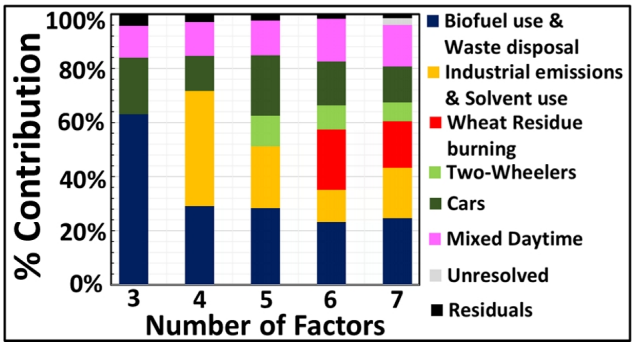

Figure 2. Percentage contribution assignment for various PMF factor number solutions (3-7) to the corresponding VOC emission sources.

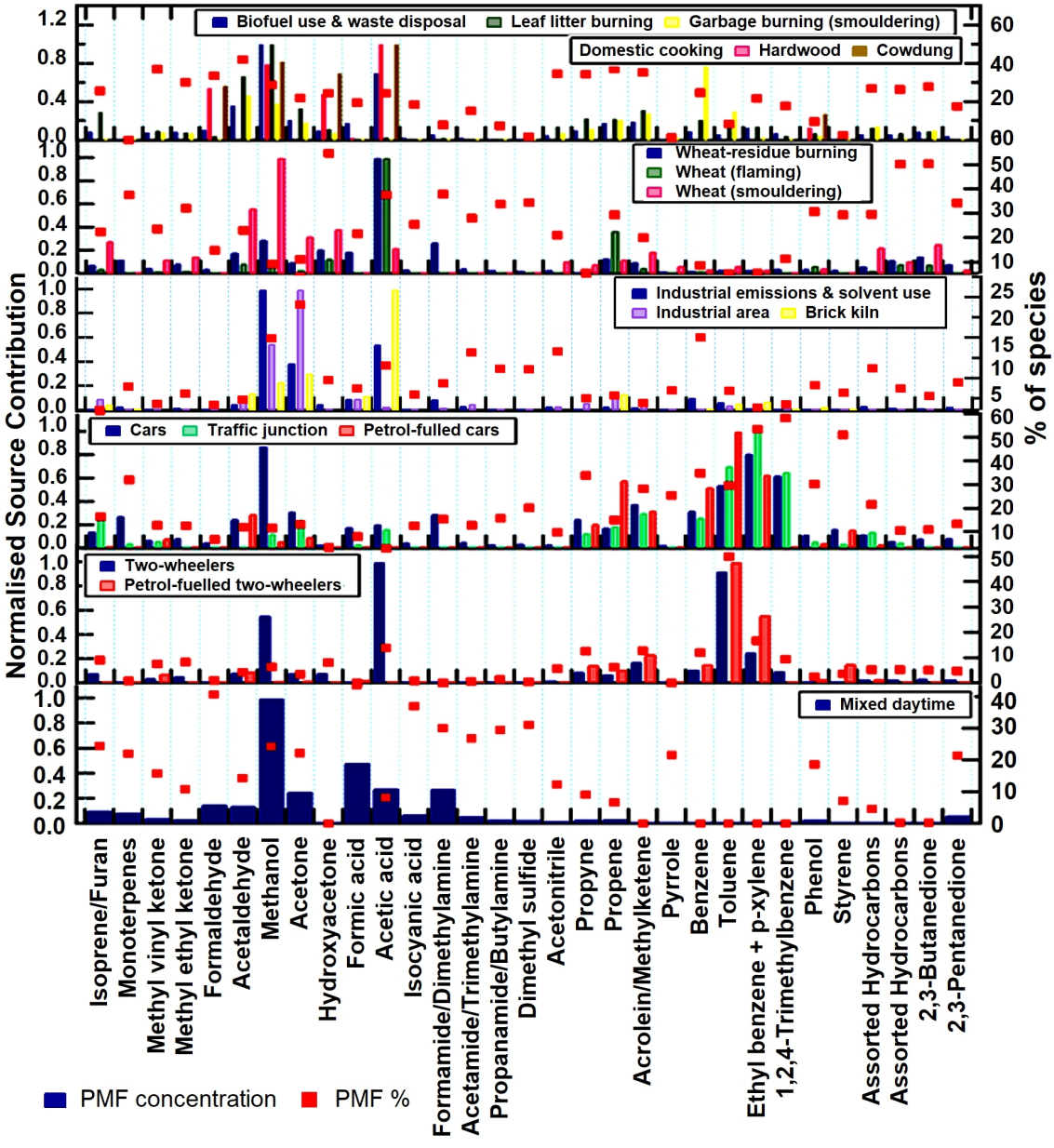

Figure 3. Factor profile composition for (6) PMF resolved factors at IISER-Mohali. It displays the normalized source fingerprints of the PMF factors (dark blue) and samples collected at source (in various colours) in bar-chart form. The value of the normalized species contribution is depicted on the left hand axis. The percentage of each species explained by each of the PMF factors is displayed in the form of a red square to be read from the right hand axis.

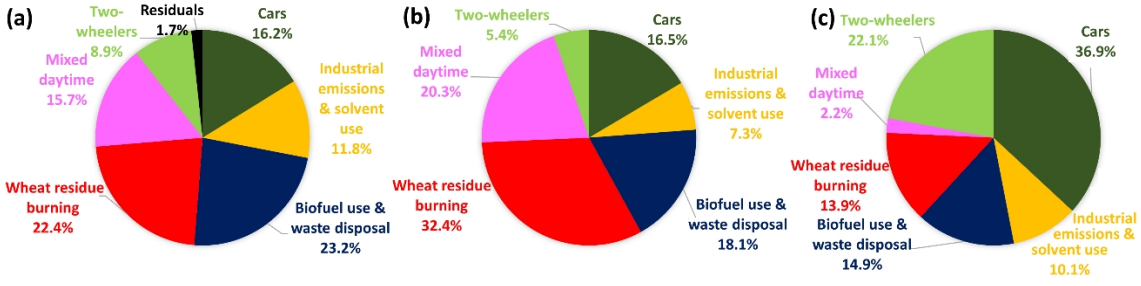

Figure 4. (a) Source contribution to the ambient VOC loading at the receptor site. (b) Ozone formation potential for PMF derived sources (c) SOA potential for PMF factors.

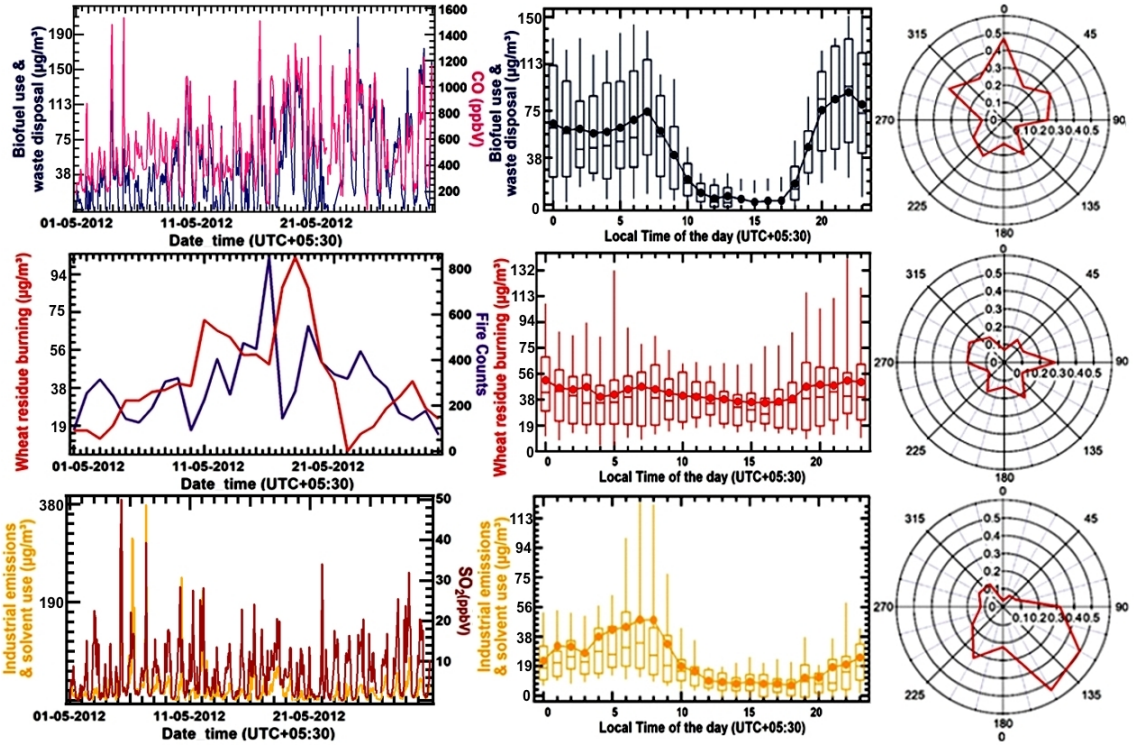

Figure 5. Factor contribution time series, factor diel variability and CPF plot for PMF Factor 1 (Biofuel use and waste disposal), PMF Factor2 (Wheat-residue burning) and PMF Factor3 (Industrial emissions and solvent use) for May2012. The time series of PMF factor's hourly mass in $\mu g\,m^{-3}$ is plotted against independent tracer species CO (in $ppbv$) for the biofuel use and waste disposal factor, daily fire counts for the wheat residue burning factor and SO_2 (in $ppbv$) for the industrial emission and solvent use factor. The Diel box and whisker plot shows the statistical parameters of factor's hourly mass contribution in $\mu g\,m^{-3}$ for every hour of the day plotted against the start time of the hour. The width of the box gives 25th and 75th percentiles, 50th percentile partitions the box; whiskers represent 10th and 90th percentiles of the dataset and average values are given by solid circles.

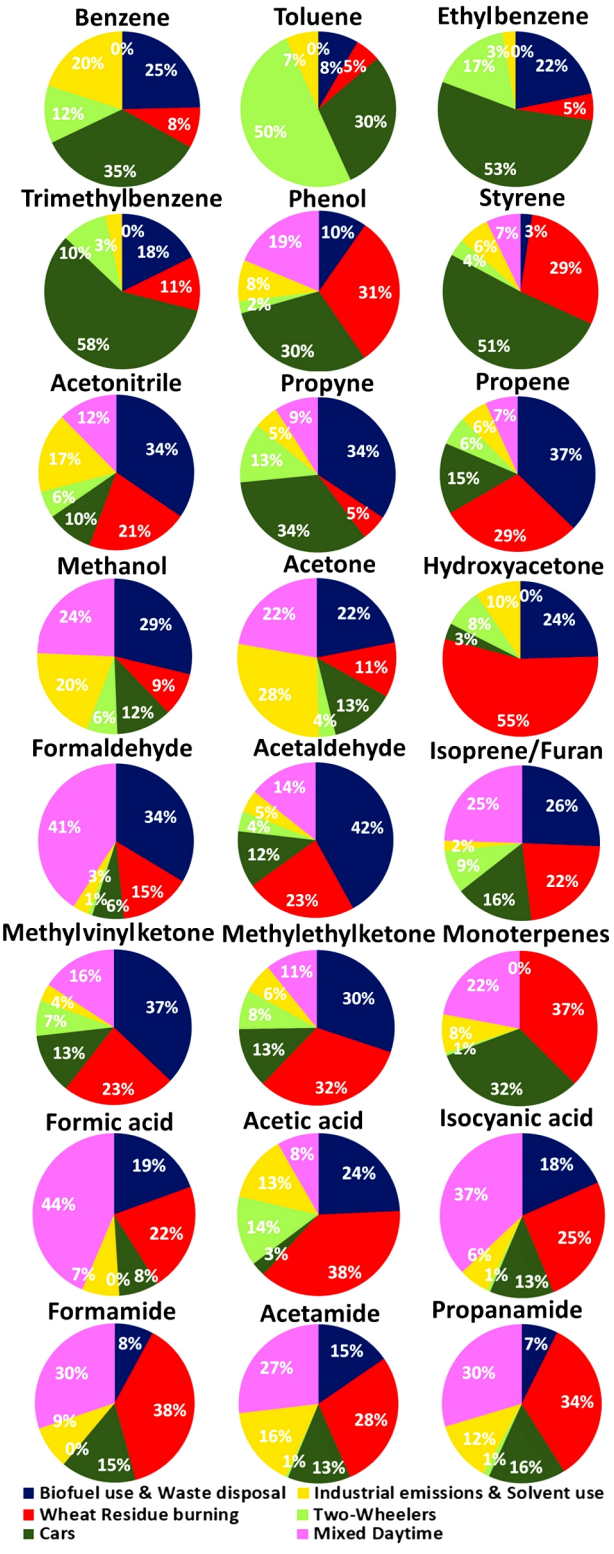

Figure 6. Contribution of individual PMF derived source factors to the total mass of different VOCs.

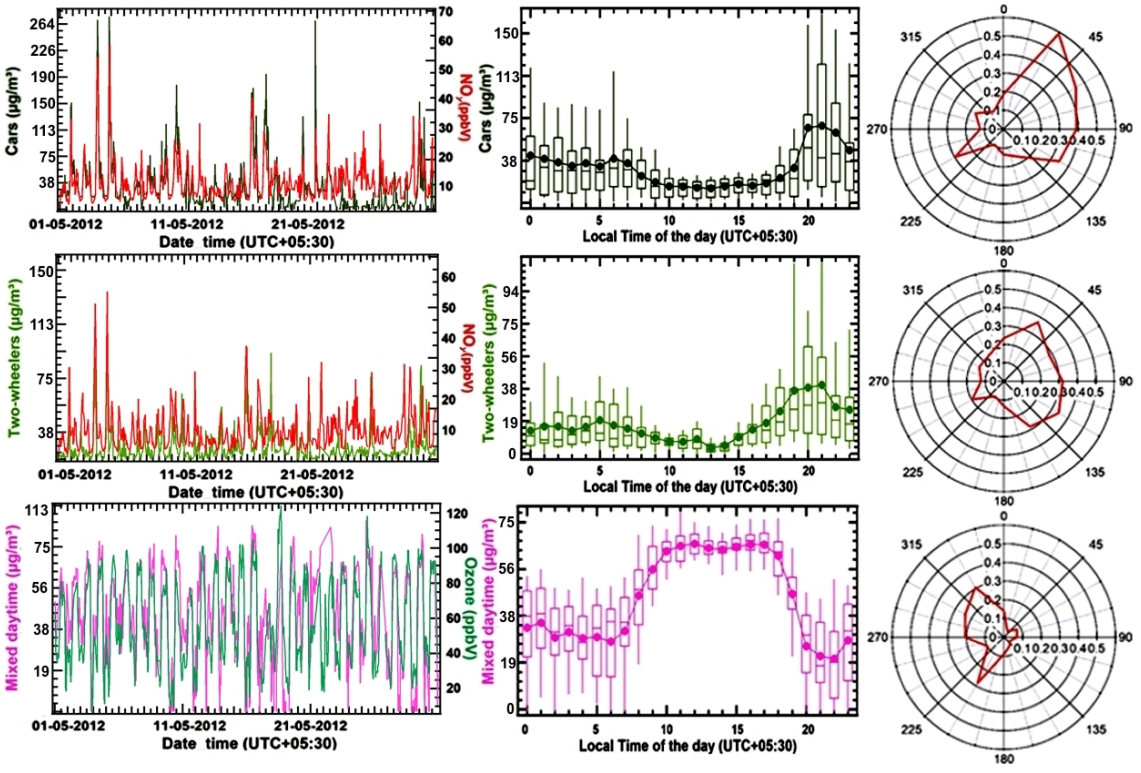

Figure 7. Factor contribution time series, factor diel variability and CPF plot for PMF Factor 4 and Factor 5 (Cars and two-wheelers) and PMF Factor 6 (Mixed daytime) for May2012. The time series of PMF factor's hourly mass in $\mu g\,m^{-3}$ is plotted against independent tracer species NO_y (in ppbv) for the car and two-wheeler factor and and O_3 (in ppbv) for the mixed daytime factor. The Diel box and whisker plot shows the statistical parameters of factor's hourly mass contribution in $\mu g\,m^{-3}$ for every hour of the day plotted against the start time of the hour. The width of the box gives 25th and 75th percentiles, 50th percentile partitions the box; whiskers represent 10th and 90th percentiles of the dataset and average values are given by solid circles.

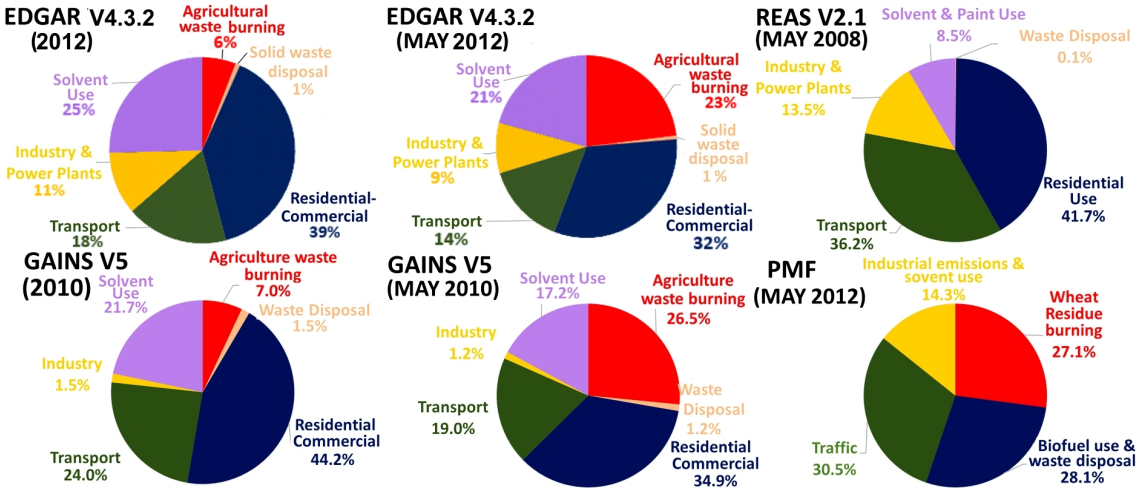

Figure 8. Comparison of PMF derived VOC source contribution to the EDGAR, REAS and GAINS Emission Inventory Database.

Supplement

Source apportionment of volatile organic compounds in the north-west Indo-Gangetic Plain using positive matrix factorisation model

Supplementary text:

Description of the PMF model: PMF is a multivariate factor analysis tool which decomposes the data matrix x_{ij} into two matrices, the factor contribution matrix g_{ik} and the factor profiles matrix f_{kj} both of which are established from the existing intrinsic variability in the dataset leaving behind a matrix of residuals e_{ij}.

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \quad (1)$$

The PMF aims at finding non-negative values of g_{ik} and f_{kj} for a given p that best reproduce x_{ij} while minimizing e_{ij}. The uncertainty weighted residuals are minimized using the parameter Q

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

wherein u_{ij} represents the matrix of measurement uncertainties for each data point and each species. Here, n and m represents the number of samples and number of species respectively. This method is described in more detail in Paatero & Tapper (1994)[1]; Paatero (1997)[2]. It is extensively employed in environmental air quality studies for source apportionment and air quality management.

Identification of the optimum number of factors: It is clear from Figure 2 in the main text and figure S4 that traffic emissions and photochemistry appear as separate sources even in a 3 Factor solution thanks to strong plumes from the urban sector and a distinct diurnal profile, respectively. Other combustion sources continue to be mixed till the model is run with a 6 Factor solution. A very distinct source, the wheat residue burning, which can be verified using MODIS fire counts appears first in the 6 Factor solution, indicating that at least 6 Factors are required to properly resolve the sources present. When the number of Factors is increased further to 7, the percentage contribution of all previously identified factors remains almost constant and an 'Unresolved Factor' accounting for only 2.5% of the total VOC mass appears. Since this factor could not be matched against any known VOC source, had no distinct diurnal

patterns and its contribution to the overall VOC burden was small we decided to retain the 6 Factor solution. The Q/Qexp plot Figure S4 also shows little improvement beyond 6 factors. Figure S5a shows how the concentration of different species in different factor profiles changes with increasing number of factors, while Figure S5b shows how the percentage of each species explained by each of the different factors changes with increasing number of factors. Figure S5c shows the evolution of the factor contribution time series with increasing number of factors.

Application of the constraint mode: Since wheat residue burning emits large quantities of oxygenated VOCs including methanol and acetic acid, clear separation of combustion derived and photochemically formed oxygenated compounds during daylight hours was an issue with some biomass burning emissions were attributed to the mixed daytime factor in the original solution. To improve the separation between photochemical formation and primary emissions, marked by compounds such as benzene, toluene, xylenes and trimethylbenzene, acetonitrile and styrene (which cannot be formed photochemically), these compounds were pulled down in the mixed daytime factor. In addition, the source contribution of the mixed daytime factor was pulled down between 2am and 4am at night. For a better separation of different combustion sources, strong plumes which represented the emission signature of the respective source were pulled up for the wheat residue burning, biofuel use and waste disposal, industrial and four wheeler source as detailed in supplementary table S3.

Normalization of factor and source profiles:

To facilitate the comparison, factor profiles comprising of the concentration of different species in µg/m³ and emission factors reported in g/kg of fuel were normalized using the following equation:

$$x' = \frac{x}{\max(x)}$$

Table S1. For each m/z used in PMF model, the table lists the major compound identifications and the references supporting such assignments from previous works along with detection limits and sensitivities.

Proto-nated mass (m/z)	Compound assignment (most likely)	Chemical formula	References	Sensitivity (ncps/ppb)	Detection limit (µg/m^3)	Average mixing ratio (stddev) (µg/m^3)
31	Formaldehyde	HCHO	[7-9]	16.3	0.472	3.414 (0.906)
33	Methanol	CH_3OH	[10, 11]	10.1	0.514	37.163 (16.049)
41	Propyne	C_3H_4	[9, 12]	16.5	0.630	3.270 (2.307)
42	Acetonitrile	CH_3CN	[11, 13]	20.7	0.065	1.745 (1.015)
43	Propene and fragment of acetic acid [1]	C_3H_6	[9, 12, 14, 15]	16.6	0.661	14.082 (7.236)
44	Isocyanic acid	HNCO	[9, 16]	16.6	0.677	1.839 (0.405)
45	Acetaldehyde	CH_3CHO	[11, 17]	20.2	0.125	9.123 (4.730)
46	Formamide/Dimethylamine	CH_3NO/ $(CH_3)_2NH$	[9, 17, 18]	16.6	0.708	8.626 (2.612)
47	Formic acid and ethanol	HCOOH	[9, 10, 19]	16.6	0.724	10.262 (2.243)
57	Acrolein/Methylketene	C_3H_4O	[9, 12]	16.5	0.881	5.990 (4.419)
59	Acetone	C_3H_6O	[20]	22.8	0.109	10.447 (5.603)
60	Acetamide/Trimethylamine	C_2H_5NO/ C_3H_9N	[12, 14]	16.5	0.929	1.962 (0.664)
61	Acetic acid	CH_3COOH	[11]	16.4	0.944	18.453 (9.551)
63	Dimethyl sulfide	C_2H_6S	[11]	16.4	0.976	0.920 (0.296)
68	Pyrrole	C_4H_5N	[19, 21]	16.2	1.055	0.528 (0.231)
69	Isoprene and Furan	C_5H_8	[11]	9.0	0.278	4.004 (1.710)
71	Methyl vinyl ketone	C_4H_6O	[11, 12]	16.0	1.102	2.577 (1.395)
73	Methyl ethyl ketone	C_4H_8O	[11, 12]	15.9	1.133	3.159 (1.578)
74	Propanamide/Butylamine	C_3H_7NO/ $C_4H_{11}N$	[22]	15.9	1.149	1.091 (0.331)
75	Hydroxyacetone	$C_3H_6O_2$	[9, 10, 21]	15.8	1.165	4.523 (2.791)
79	Benzene	C_6H_6	[11, 19]	13.5	0.196	4.105 (3.320)
83	Assorted Hydrocarbons	C_6H_{10}	[21]	15.3	1.291	2.531 (1.423)
85	Assorted Hydrocarbons	C_6H_{12}	[21]	15.1	1.322	2.686 (1.571)
87	2,3-Butanedione, 2-methyl-Butanal or pentanone	$C_4H_6O_2$ $C_5H_{10}O$	[21, 22]	15.0	1.354	3.407 (2.025)
93	Toluene	C_7H_8	[11, 19]	14.3	0.261	7.805 (6.977)
95	Phenol	C_6H_5OH	[9, 19]	14.2	1.480	1.766 (1.167)
101	2,3-Pentanedione, acetyl acetone , 2-butenoic acid methyl ester or hexanal,	$C_5H_8O_2$ $C_6H_{12}O$	[22]	13.5	1.574	2.935 (1.273)
105	Styrene	C_8H_8	[11, 12, 21]	13.1	1.637	1.477 (1.112)
107	Ethyl benzene + p-xylene	C_8H_{10}	[11, 19]	13.8	0.501	6.724 (6.381)
121	1,2,4-Trimethylbenzene	C_9H_{12}	[11, 19]	11.2	0.453	4.677 (4.102)
137	Sum of Monoterpenes	$C_{10}H_{16}$	[11, 13]	7.9	2.141	3.779 (1.577)

[1] Correction applied to the input concentration data for propene being the potential fragment of acetic acid with ~68% contribution.

Table S2. Input data statistics for PMF Model runs.

VOC Species	Category	S/N	Min	25th	Median	75th	Max
Isoprene/Furan	Weak	3.89	1.25	2.78	3.48	4.82	11.48
Benzene	Strong	3.87	0.57	1.83	3.21	5.15	24.64
Toluene	Strong	3.93	0.82	3.48	5.54	9.07	49.95
Ethyl benzene + p-xylene	Strong	3.72	0.80	2.99	4.72	8.07	62.51
1,2,4-Trimethylbenzene	Strong	3.61	0.54	2.31	3.38	5.53	31.45
Methyl vinyl ketone/hydrocarbon fragments	Weak	2.10	0.56	1.53	2.13	3.44	8.01
Methyl ethyl ketone/butanal	Weak	2.43	0.88	1.88	2.77	4.02	9.54
Acetic acid	Strong	3.97	5.54	18.64	28.18	38.53	107.72
Dimethyl sulfide	Weak	0.74	0.32	0.67	0.95	1.10	2.41
Pyrrole	Weak	0.14	0.09	0.37	0.54	0.66	2.50
Propanamide/Butylamine	Weak	0.75	0.37	0.86	1.07	1.24	2.32
Hydroxyacetone	Strong	2.81	0.98	2.50	3.91	5.75	16.24
Assorted Hydrocarbons	Weak	1.79	0.66	1.38	2.20	3.19	8.82
Assorted Hydrocarbons	Weak	1.85	0.55	1.50	2.32	3.37	9.36
$C_4H_6O_2$	Weak	2.18	0.77	1.90	2.95	4.34	11.88
Phenol	Strong	1.02	0.54	1.21	1.54	2.06	16.50
$C_5H_8O_2$	Weak	1.83	1.13	2.08	2.64	3.39	11.64
Styrene	Strong	0.62	0.20	0.92	1.21	1.65	9.73
Methanol	Strong	4.00	14.98	24.28	33.86	45.96	129.91
Acetonitrile	Strong	3.95	0.51	0.97	1.50	2.27	7.23
Acetaldehyde	Strong	3.99	2.30	5.13	7.83	12.16	24.69
Acetone	Strong	4.00	3.58	6.74	9.14	12.54	54.48
Monoterpenes	Weak	1.73	0.72	2.63	3.88	4.62	10.35
Formaldehyde	Strong	3.68	1.68	2.69	3.21	4.07	6.28
Propyne	Strong	3.11	0.70	1.63	2.63	3.98	16.63
Propene	Weak	3.56	0.94	2.92	4.71	6.79	17.32
Isocyanic acid	Weak	2.61	0.98	1.57	1.79	2.04	3.21
Formamide/Dimethylamine	Weak	3.87	3.01	6.91	8.42	9.90	21.52
Formic acid	Weak	3.91	5.22	8.74	9.99	11.58	18.55
Acrolein/Methylketene	Weak	3.29	1.07	2.63	4.90	7.63	32.31
Acetamide/Trimethylamine	Weak	2.10	0.66	1.52	1.82	2.25	4.78
Total VOC	Weak	4.00	64.00	124.77	171.23	230.51	515.52

Table S3. List of constraints applied to the wheat residue burning, biofuel use and waste disposal, industrial and four wheeler source.

PMF SOURCE FACTOR	DATE (May2012)	TIME	CONSTRAINT APPLIED
Wheat residue burning	12	19:00-20:59	Pull up
	13	22:00-23:59	Pull up
	18	3:00-3:59, 6:00-7:59	Pull up
	19	2:00-2:59	Pull up
Cars	1	20:00-21:59	Pull up
	4	19:00-21:59	Pull up
Industrial emissions and solvent use	6	4:00-5:59	Pull up
	7	8:00-8:59	Pull up
Biofuel use and waste disposal	22	22:00-22:59	Pull up
	23	12:00-12:59, 6:00-6:59, 22:00-23:59	Pull up
	24	2:00-4:59, 21:00-22:59	Pull up
	25	7:00-7:59	Pull up

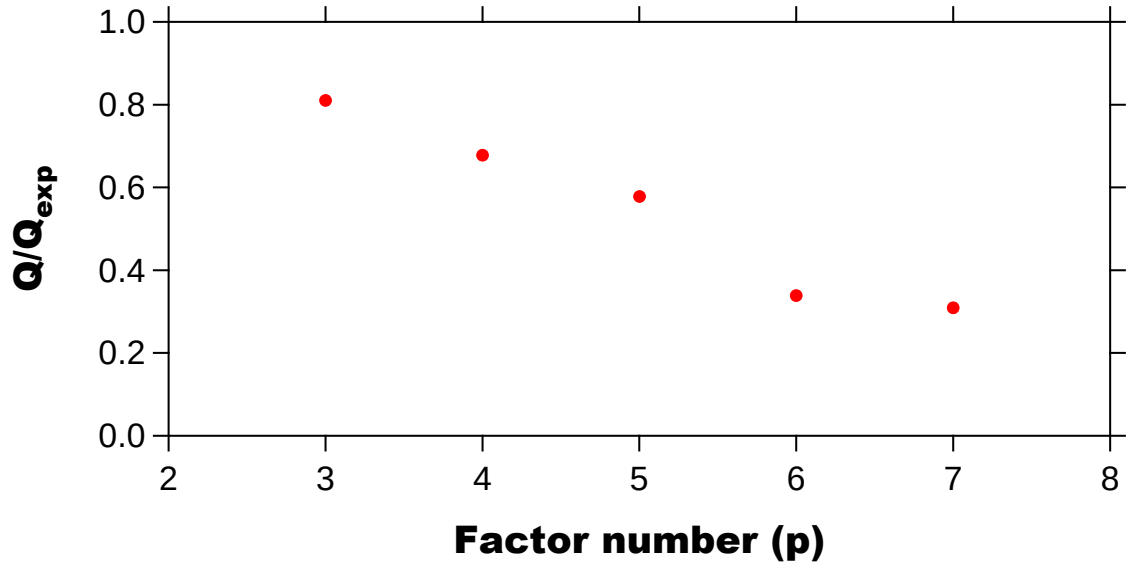

Figure S4 Q/Q_{exp} plot with increasing number of factor. The absolute Q is relatively low indicating that it may be more appropriate to only consider the 10% precision error of the PTR-MS instead of including the accuracy error while specifying the uncertainty in the PMF input. However, since equal uncertainty was applied to all strong m/z this only affects the absolute Q value and not the model output.

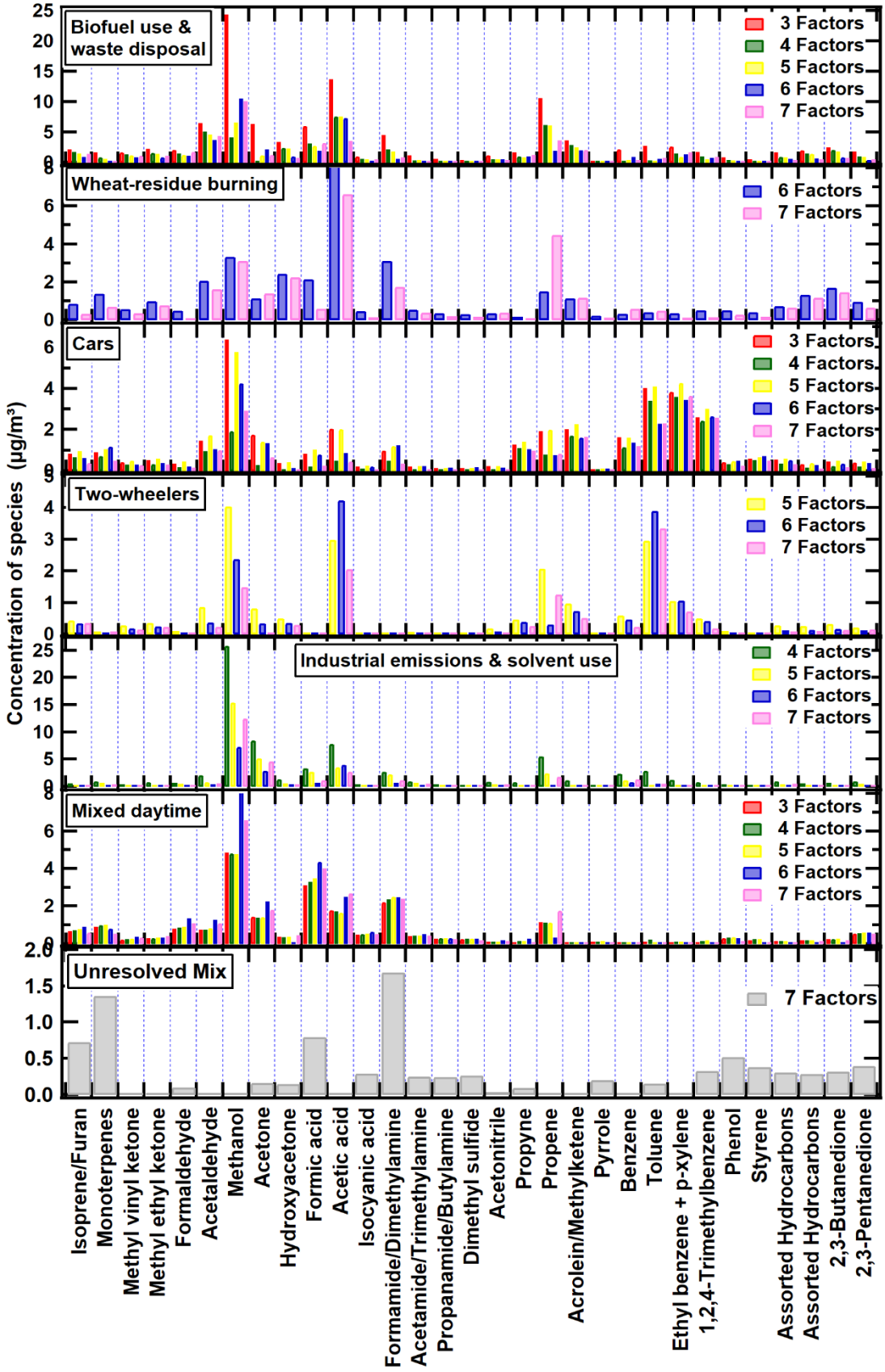

Figure S5a. Evolution of PMF factor profiles from 3 to 7 factor number solutions.

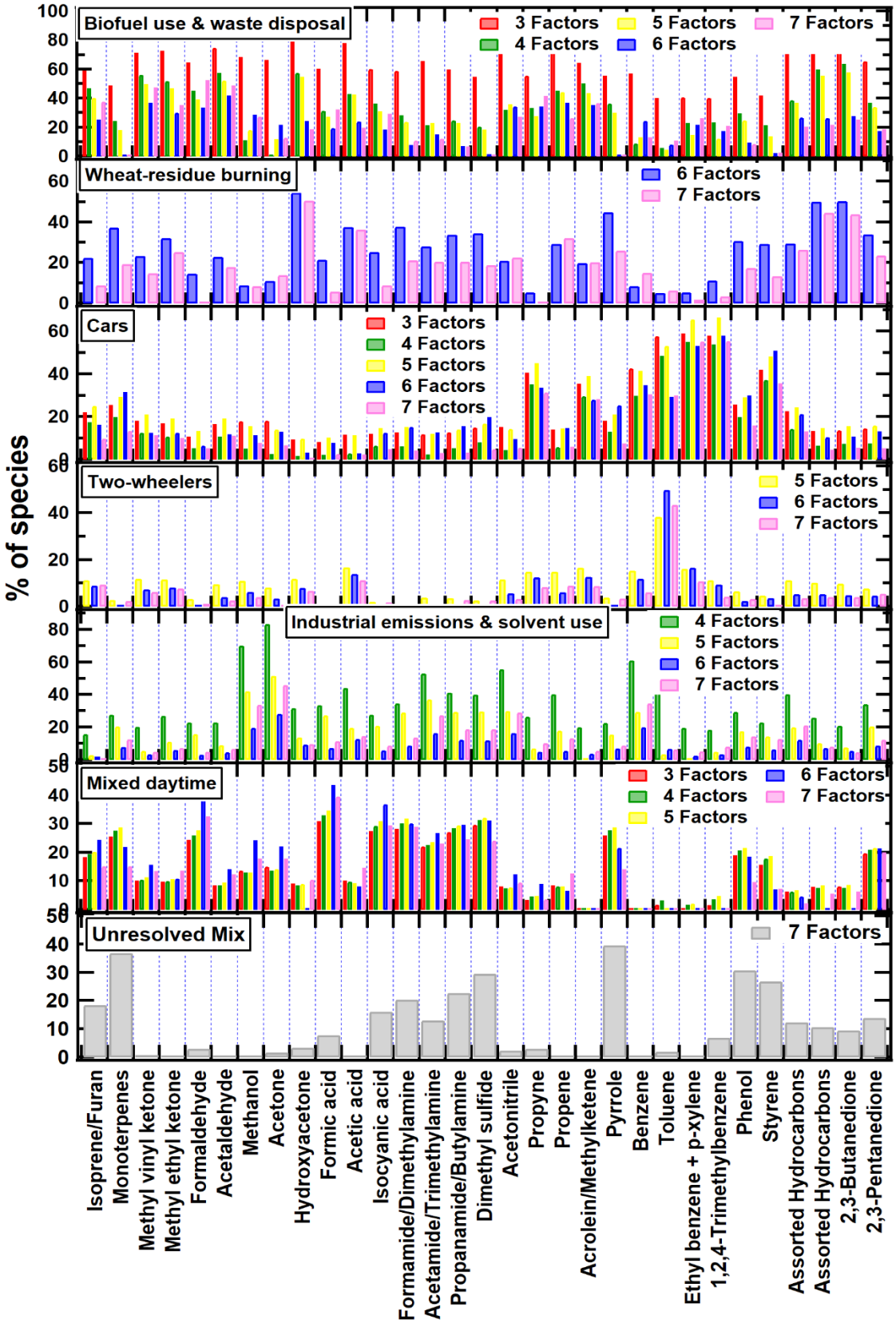

Figure S5b. Evolution of percentage contribution of different VOC species from 3 to 7 PMF factor solutions.

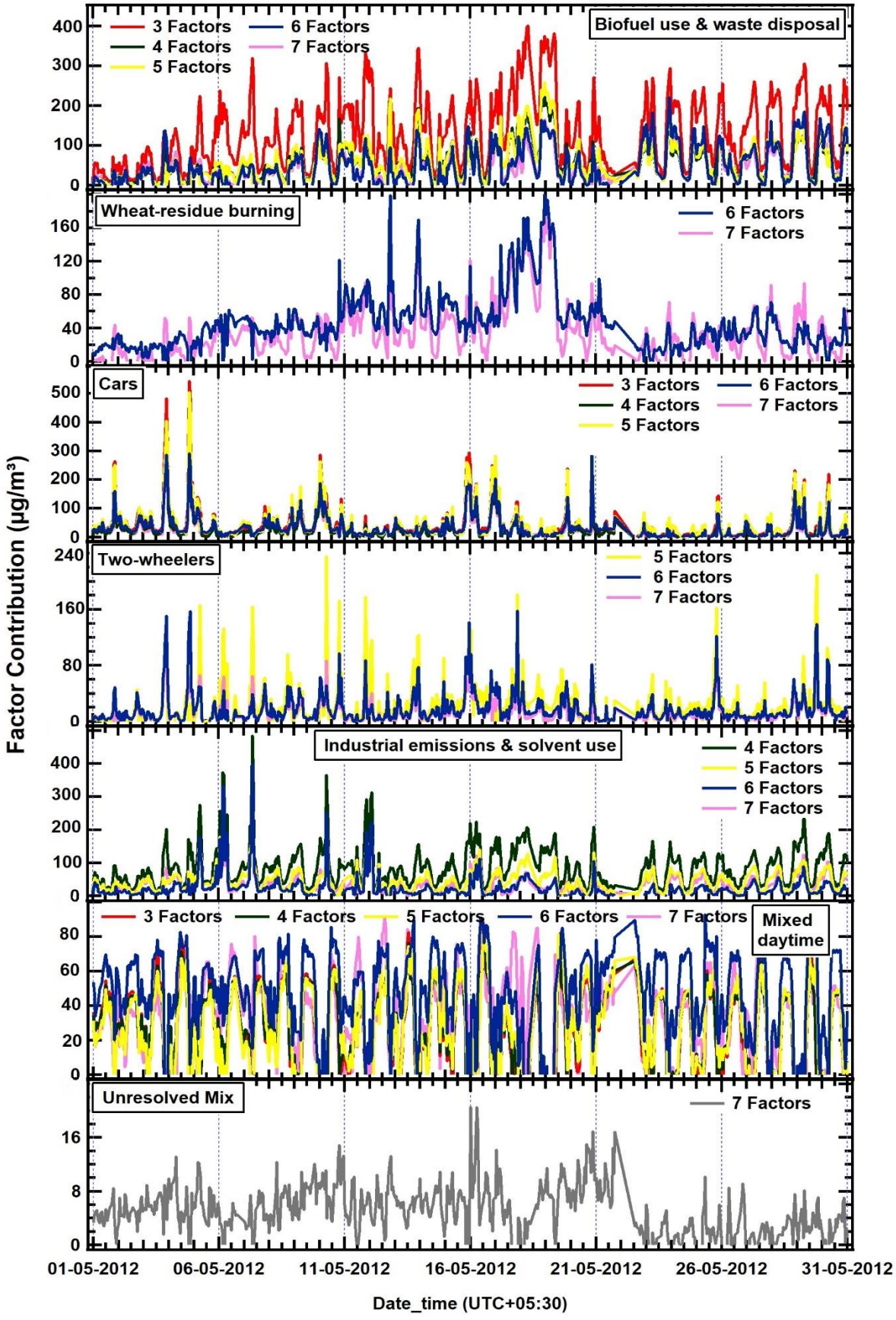

Figure S5c. Evolution of PMF factor contributions from 3 to 7 factor solutions.

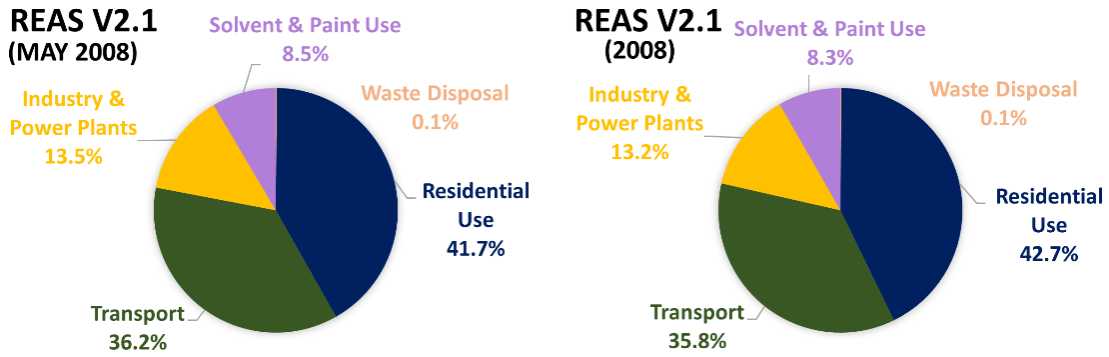

Figure S6. REAS database comparison to VOC source sectors on monthly and yearly resolution scales.

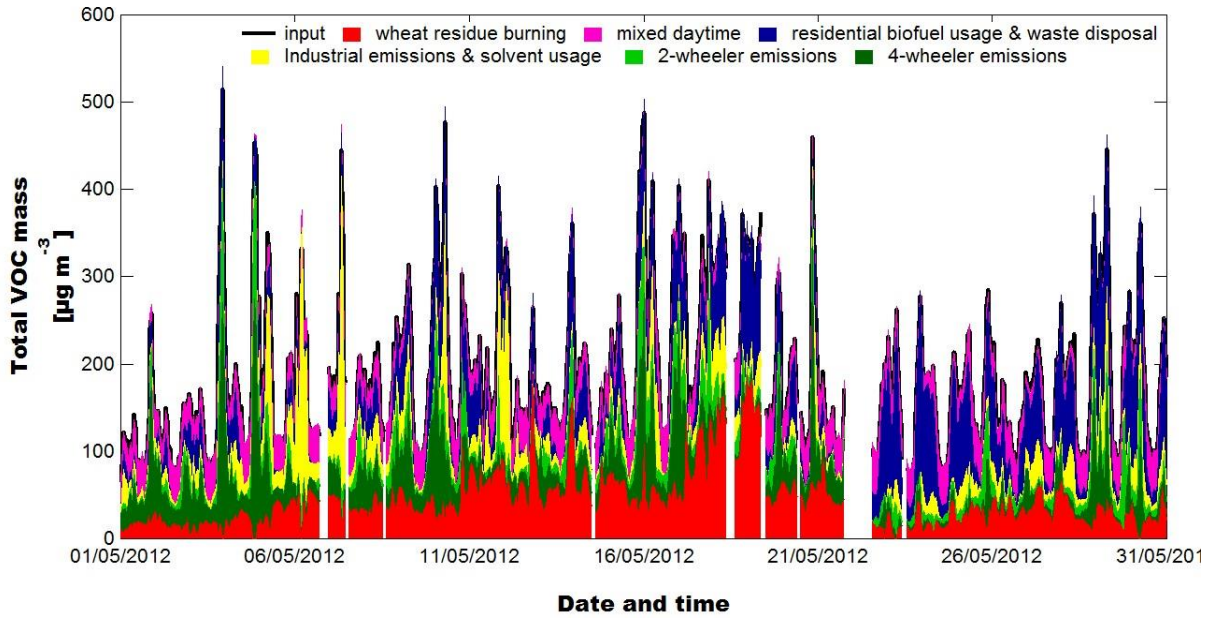

Figure S7. Time series of the total mass contributed by the different sources to the overall VOC mass

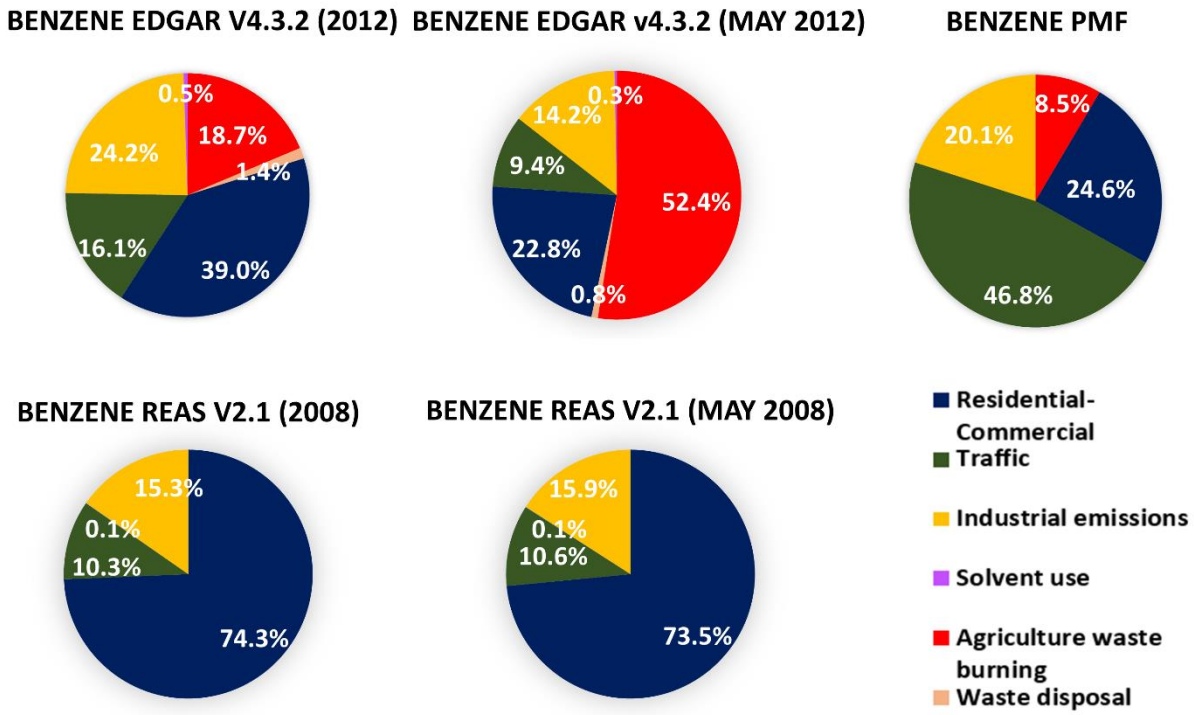

Figure S8a: Comparison of the PMF output with benzene emission inventories for the study region.

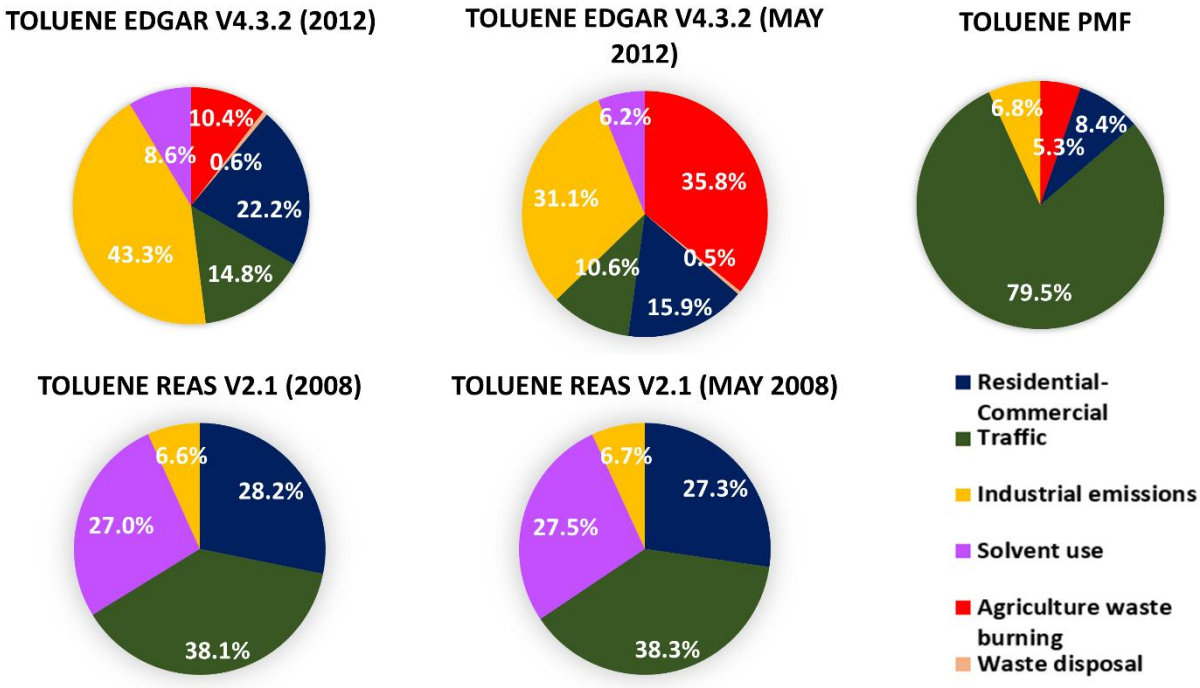

Figure S8b: Comparison of the PMF output with toluene emission inventories for the study region.

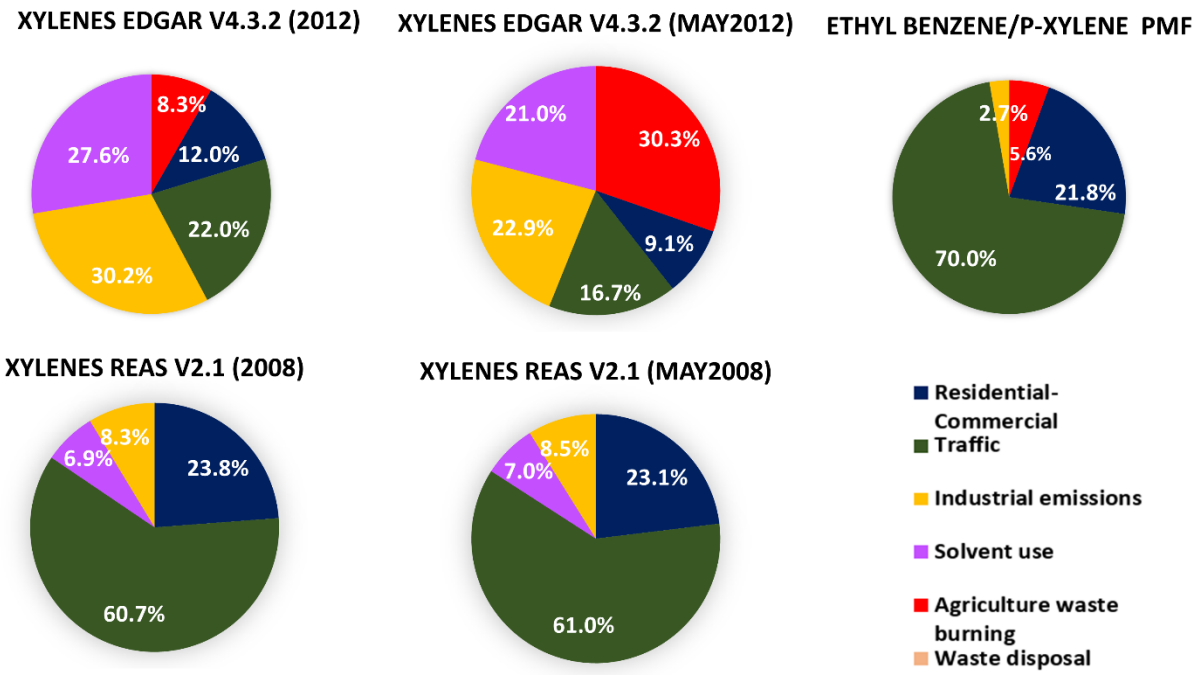

Figure S8c: Comparison of the PMF output with xylenes in the emission inventories for the study region.

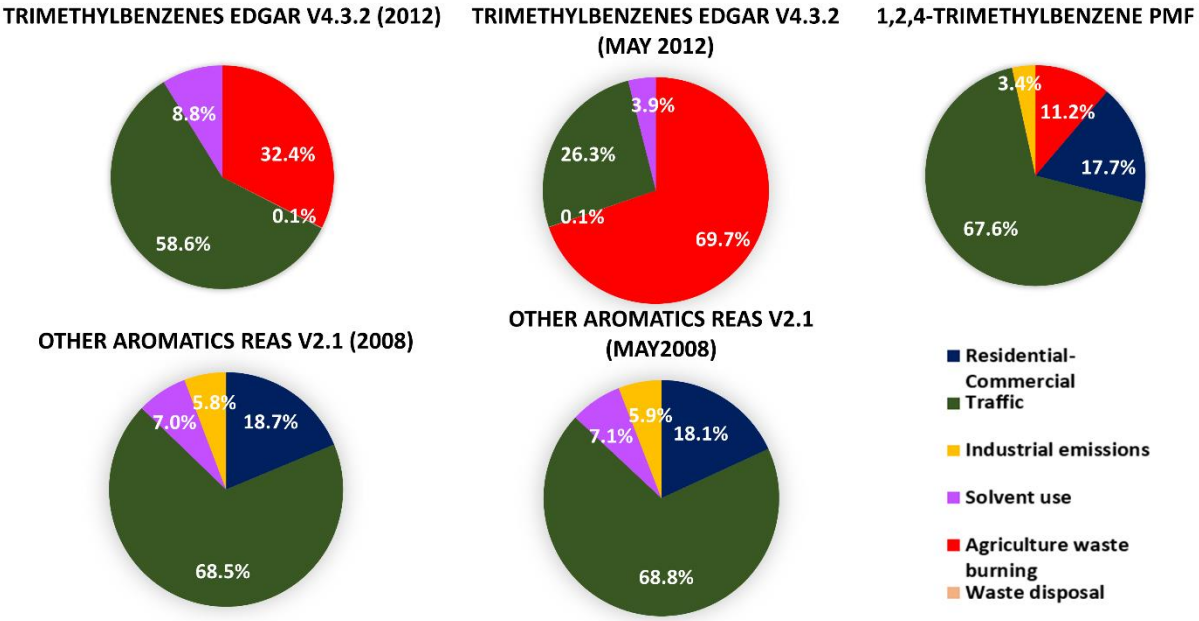

Figure S8d: Comparison of the PMF output of C-9 aromatic compounds with the class "other aromatic compounds" in the emission inventories for the study region.

References

1. Paatero, P.; Tapper, U., Positive matrix factorization: A non- negative factor model with optimal utilization of error estimates of data values. *Environmetrics.* **1994,** *5*, (2), 111-126.

2. Paatero, P., Least squares formulation of robust non-negative factor analysis. *Chemom. Intell. Lab. Syst.* **1997,** *37*, (1), 23-35.

3. Xie, Y.; Berkowitz, C. M., The use of positive matrix factorization with conditional probability functions in air quality studies: an application to hydrocarbon emissions in Houston, Texas. *Atmos. Environ.* **2006,** *40*, (17), 3070-3091.

4. Leuchner, M.; Rappenglück, B., VOC source–receptor relationships in Houston during TexAQS-II. *Atmos. Environ.* **2010,** *44*, (33), 4056-4067.

5. Sinha, V.; Williams, J.; Diesch, J.; Drewnick, F.; Martinez, M.; Harder, H.; Regelin, E.; Kubistin, D.; Bozem, H.; Hosaynali-Beygi, Z., Constraints on instantaneous ozone production rates and regimes during DOMINO derived using in-situ OH reactivity measurements. *Atmos. Chem. Phys.* **2012,** *12*, (15), 7269-7283.

6. Derwent, R. G.; Jenkin, M. E.; Utembe, S. R.; Shallcross, D. E.; Murrells, T. P.; Passant, N. R., Secondary organic aerosol formation from a large number of reactive man-made organic compounds. *Sci. Total Environ.* **2010,** *408*, (16), 3374-3381.

7. Jobson, B.; Alexander, M. L.; Maupin, G. D.; Muntean, G. G., On-line analysis of organic compounds in diesel exhaust using a proton transfer reaction mass spectrometer (PTR-MS). *Int. J. Mass Spectrom.* **2005,** *245*, (1-3), 78-89.

8. Warneke, C.; Roberts, J.; Veres, P.; Gilman, J.; Kuster, W.; Burling, I.; Yokelson, R.; De Gouw, J., VOC identification and inter-comparison from laboratory biomass burning using PTR-MS and PIT-MS. *Int. J. Mass Spectrom.* **2011,** *303*, (1), 6-14.

9. Sarkar, C.; Sinha, V.; Sinha, B.; Panday, A. K.; Rupakheti, M.; Lawrence, M. G., Source apportionment of NMVOCs in the Kathmandu Valley during the SusKat-ABC international field campaign using positive matrix factorization. *Atmos. Chem. Phys.* **2017,** *17*, (13), 8129-8156.

10. Williams, J.; Pöschl, U.; Crutzen, P.; Hansel, A.; Holzinger, R.; Warneke, C.; Lindinger, W.; Lelieveld, J., An atmospheric chemistry interpretation of mass scans obtained from a proton transfer mass spectrometer flown over the tropical rainforest of Surinam. *J. Atmos. Chem.* **2001,** *38*, (2), 133-166.

11. de Gouw, J.; Warneke, C., Measurements of volatile organic compounds in the earth's atmosphere using proton- transfer- reaction mass spectrometry. *Mass Spectrom. Rev.* **2007,** *26*, (2), 223-257.

12. Stockwell, C.; Veres, P.; Williams, J.; Yokelson, R., Characterization of biomass burning emissions from cooking fires, peat, crop residue, and other fuels with high-resolution proton-transfer-reaction time-of-flight mass spectrometry. *Atmos. Chem. Phys.* **2015,** *15*, (2), 845-865.

13. Yuan, B.; Koss, A. R.; Warneke, C.; Coggon, M.; Sekimoto, K.; de Gouw, J. A., Proton-transfer-reaction mass spectrometry: Applications in atmospheric sciences. *Chem. Rev.* **2017,** *117*, (21), 13187-13229.

14. Sarkar, C.; Sinha, V.; Kumar, V.; Rupakheti, M.; Panday, A.; Mahata, K. S.; Rupakheti, D.; Kathayat, B.; Lawrence, M. G., Overview of VOC emissions and chemistry from PTR-TOF-MS measurements during the SusKat-ABC campaign: high acetaldehyde, isoprene and isocyanic acid in wintertime air of the Kathmandu Valley. *Atmos. Chem. Phys.* **2016,** *16*, (6), 3979-4003.

15. Bruns, E. A.; Slowik, J. G.; Haddad, I. E.; Kilic, D.; Klein, F.; Dommen, J.; Temime-Roussel, B.; Marchand, N.; Baltensperger, U.; Prévôt, A. S., Characterization of gas-phase organics using proton transfer reaction time-of-flight mass spectrometry: fresh and aged residential wood combustion emissions. *Atmos. Chem. Phys.* **2017,** *17*, (1), 705-720.

16. Gylestam, D.; Karlsson, D.; Dalene, M.; Skarping, G., Determination of gas phase isocyanates using proton transfer reaction mass spectrometry. *Anal. Chem. Lett.* **2011,** *1*, (4), 261-271.

17. Janda, M.; Morvova, M.; Machala, Z.; Morva, I., Study of plasma induced chemistry by DC discharges in CO_2 /N_2 /H_2O mixtures above a water surface. *Orig. Life Evol. Biospheres* **2008,** *38*, (1), 23-35.

18. Ge, X.; Wexler, A. S.; Clegg, S. L., Atmospheric amines–Part I. A review. *Atmos. Environ.* **2011,** *45*, (3), 524-546.

19. Karl, T.; Christian, T. J.; Yokelson, R. J.; Artaxo, P.; Hao, W. M.; Guenther, A., The Tropical Forest and Fire Emissions Experiment: method evaluation of volatile organic compound emissions measured by PTR-MS, FTIR, and GC from tropical biomass burning. *Atmos. Chem. Phys.* **2007,** *7*, (22), 5883-5897.

20. de Gouw, J.; Warneke, C.; Holzinger, R.; Klüpfel, T.; Williams, J., Inter-comparison between airborne measurements of methanol, acetonitrile and acetone using two differently configured PTR-MS instruments. *Int. J. Mass Spectrom.* **2004,** *239*, (2-3), 129-137.

21. Yokelson, R. J.; Burling, I.; Gilman, J.; Warneke, C.; Stockwell, C.; Gouw, J. d.; Akagi, S.; Urbanski, S.; Veres, P.; Roberts, J., Coupling field and laboratory measurements to estimate the emission factors of identified and unidentified trace gases for prescribed fires. *Atmos. Chem. Phys.* **2013,** *13*, (1), 89-116.

22. Brilli, F.; Gioli, B.; Ciccioli, P.; Zona, D.; Loreto, F.; Janssens, I. A.; Ceulemans, R., Proton Transfer Reaction Time-of-Flight Mass Spectrometric (PTR-TOF-MS) determination of volatile organic compounds (VOCs) emitted from a biomass fire developed under stable nocturnal conditions. *Atmos. Environ.* **2014,** *97*, 54-67.