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 (75th percentile) for a particular wind direction sector thereby highlighting directional dependency of source factors and is defined as follows:

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

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:

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

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:

$$SOA\ potential = \left(\sum_i [VOC_i][SOAP_i]\right) \qquad (4)$$

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.

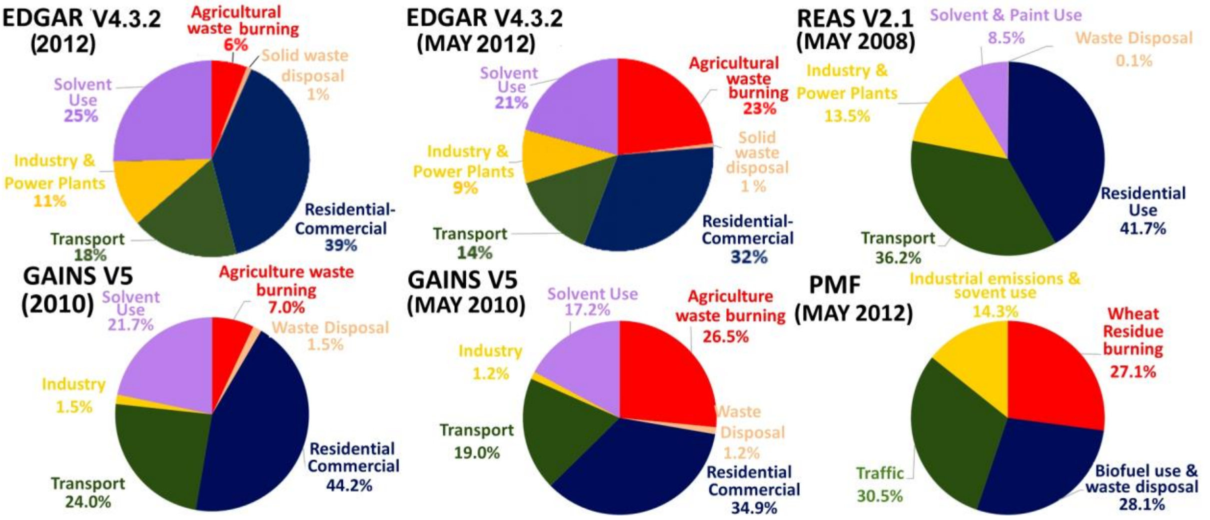


Figure 8 has been revised and now includes EDGAR v4.3.2 (2012) instead of v4.2 (2008) and have updated the discussion accordingly. The latest EDGAR represents a significant improvement over the EDGAR HTAP and v4.2.

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

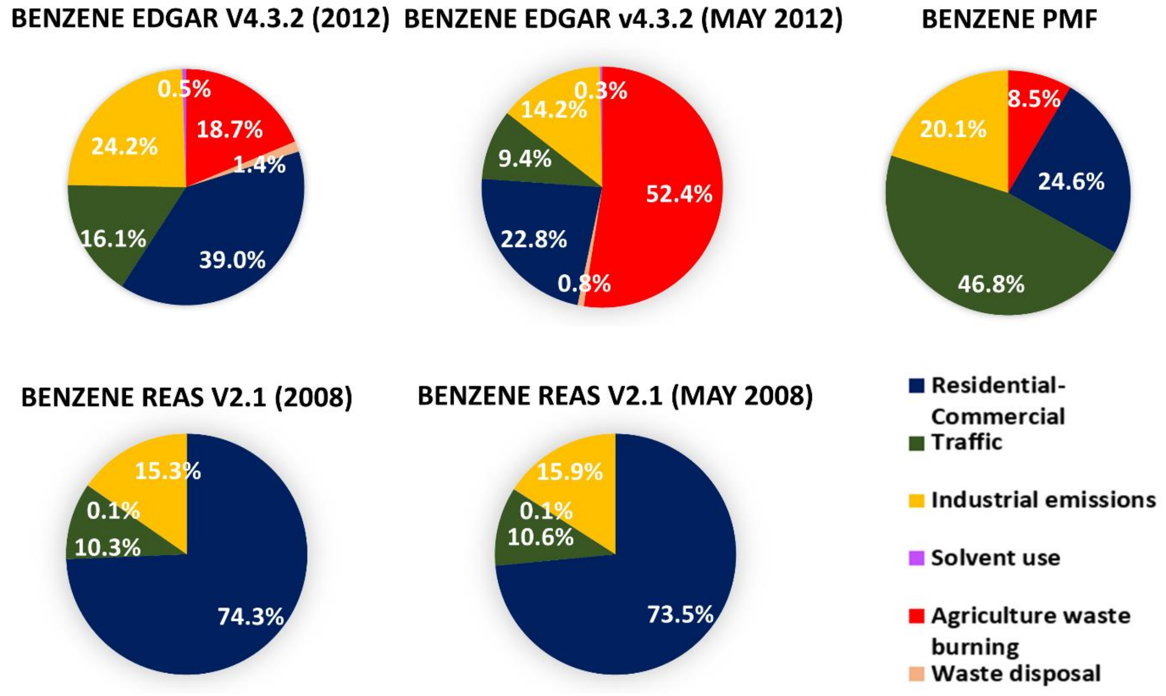


Figure S8a: Comparison of the PMF output with benzene emission inventories for the study region.

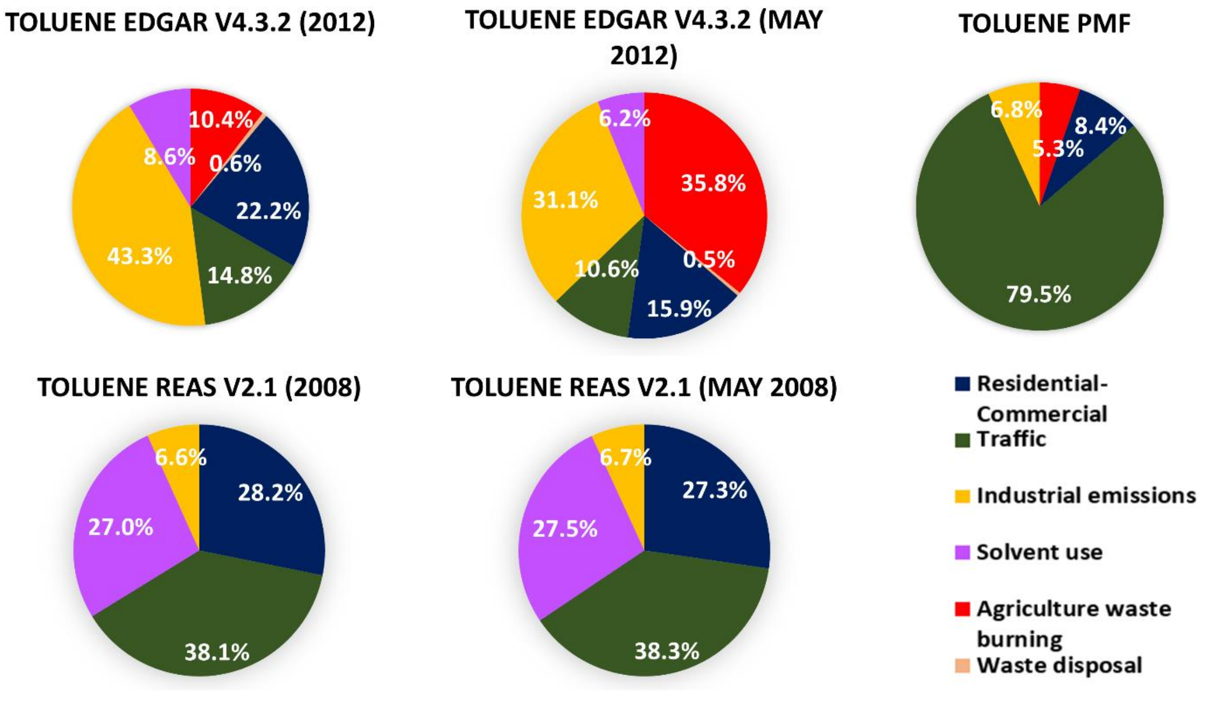


Figure S8b: Comparison of the PMF output with toluene emission inventories for the study region.

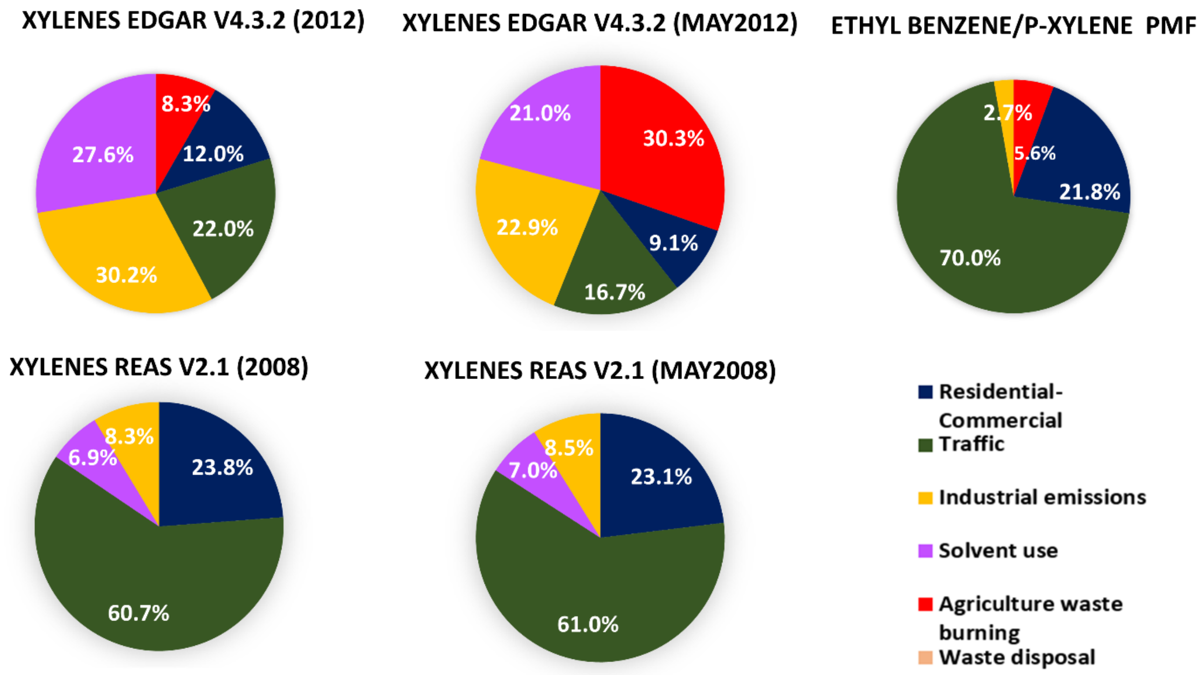


Figure S8c: Comparison of the PMF output with xylenes in the emission inventories for the study region.

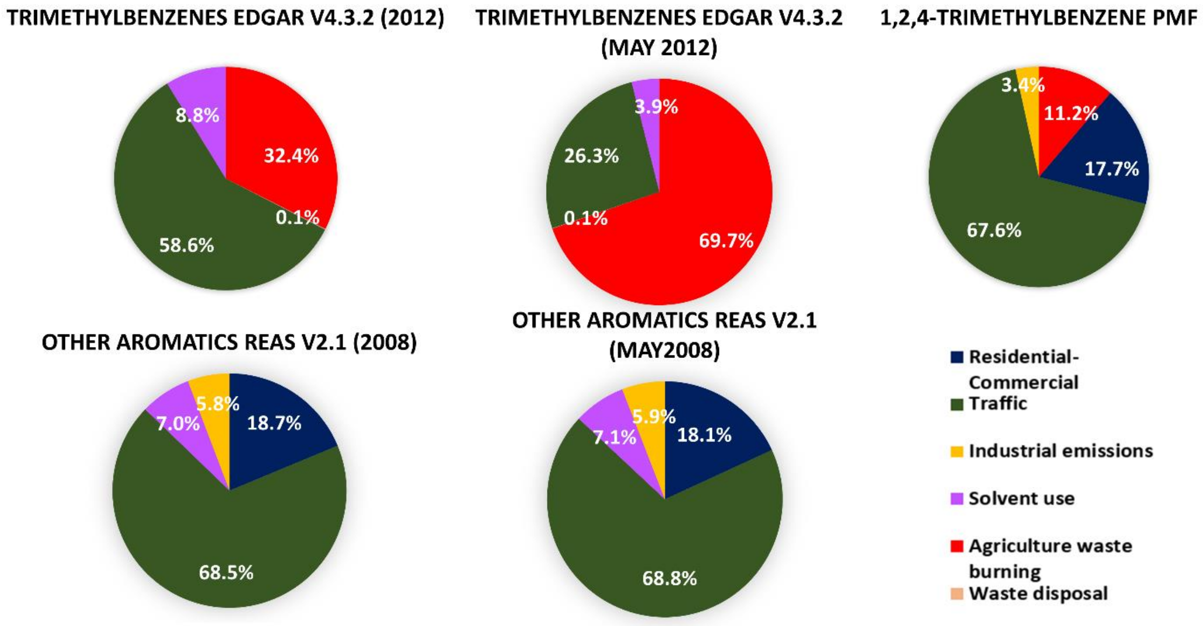


Figure S8d: Comparison of the PMF output of C-9 aromatic compounds with the class “other aromatic compounds” in the emission inventories for the study region.

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.

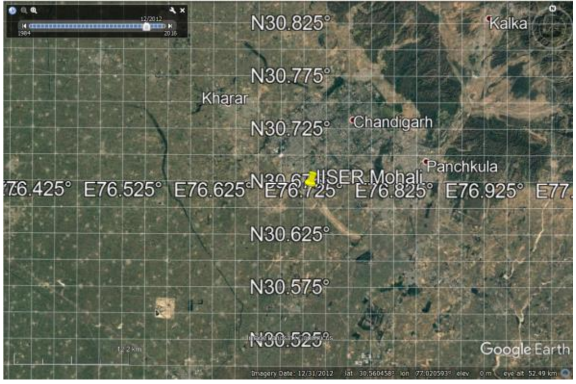
Author response: we have changed the numbering of several figures in the supplement as Reviewer #2 asked us to include an additional plot. The numbers are now consistent with the numbering in the manuscript.

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.