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}.

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

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

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

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. Figure S4a shows how the concentration of different species in different factor profiles changes with increasing number of factors, while Figure S4b shows how the percentage of each species explained by each of the different factors changes with increasing number of factors. Figure S4c 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.

Conditional Probability Function analysis The conditional probability function analysis aids in identifying physical locations of different PMF source factors without using back trajectories[3]. Conditional probability function (CPF) supplements the PMF model results and is computed using the factor contribution matrix generated by the PMF model and integrating it with general meteorological parameter of wind direction at the receptor site. It quantifies the probability of factor contributions surpassing a certain threshold criterion for a particular wind direction sector thereby highlighting directional dependency of source factors and is defined as follows:

$$CPF = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \tag{3}$$

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. Δθ was assigned a value of 30°. The threshold criterion value was set to 75th percentile of all the data to ensure only high mixing ratios are used to determine source location. A straight path of the air parcel is assumed for CPF analysis from source to the receptor site[4].

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)[5] and their corresponding relative percentage contributions are shown in Figure 11 (b). The following equation was used for the calculation:

$$Ozone\ production\ potential = \left(\sum_i k_{(VOC_i+OH)}\ [VOC_i]\right) \times\ [OH]\ \times n \tag{4}$$

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 shown in Figure 11 (c) was calculated for the PMF source factors using the literature SOA yields[6] 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:

$$SOA\ potential = \left(\sum_i [VOC_i][SOAP_i]\right) \tag{5}$$

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/m³)**	**Average mixing ratio (stddev) (μg/m³)**
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

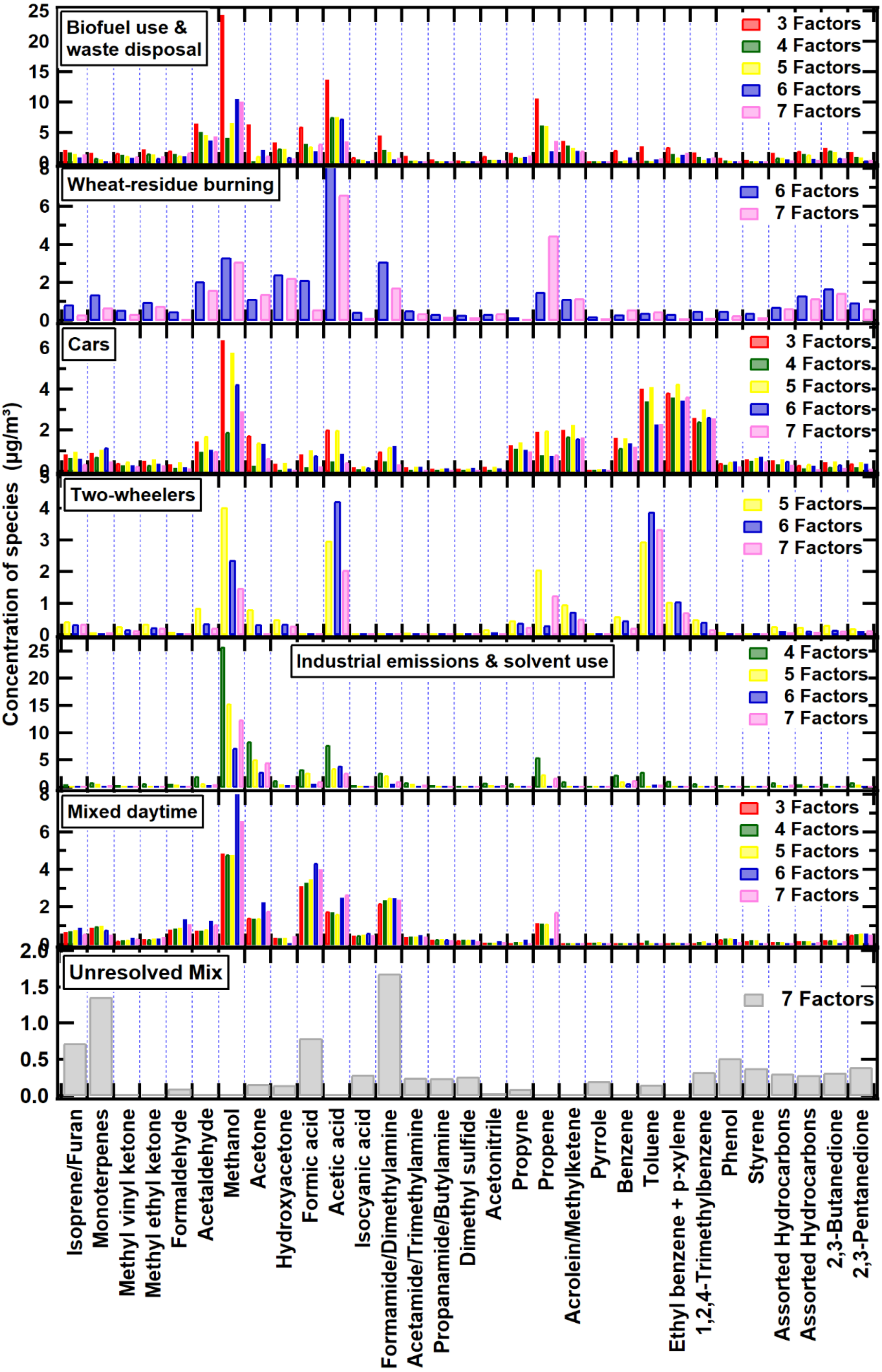

Figure S4a. Evolution of PMF factor profiles from 3 to 7 factor number solutions.

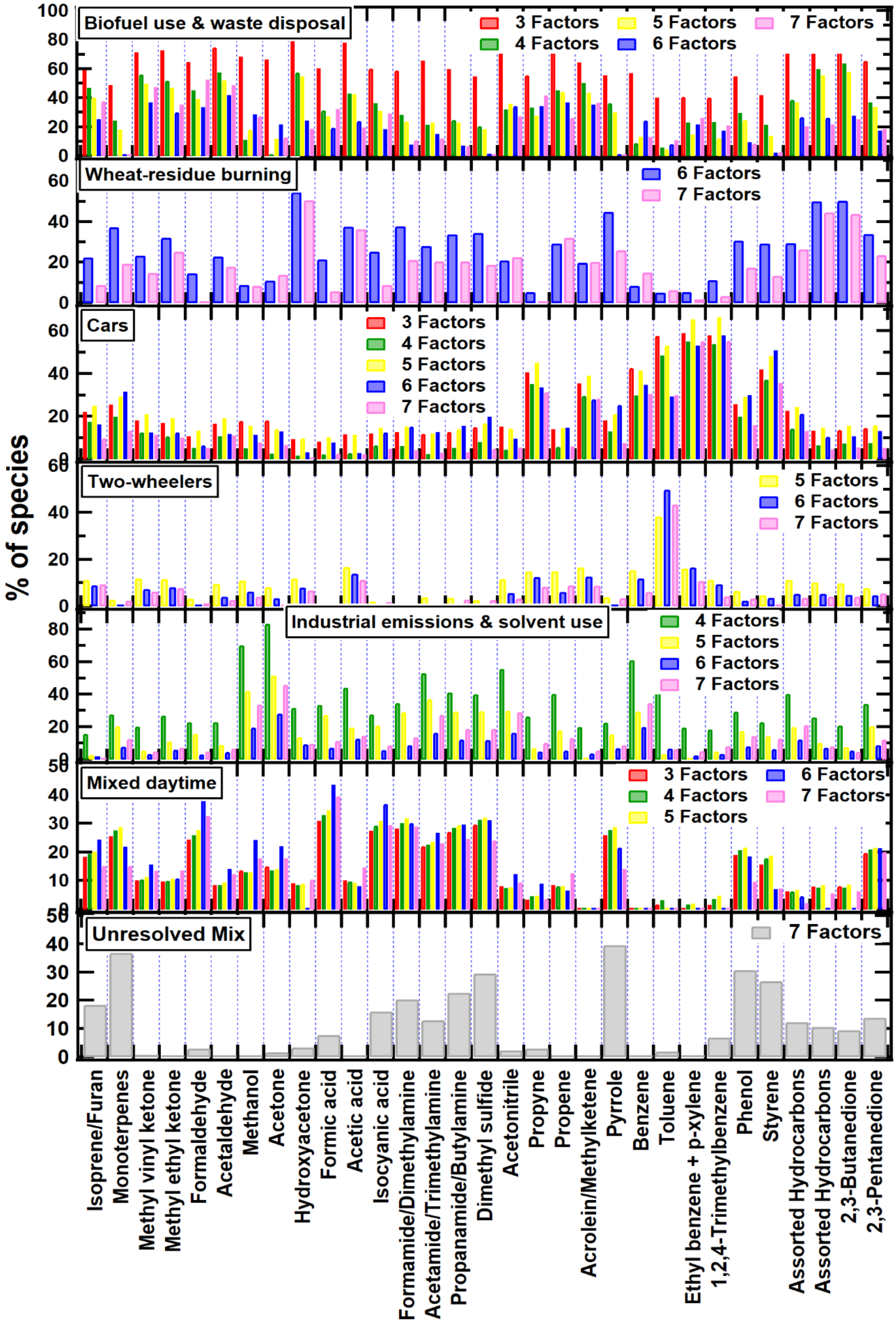

Figure S4b. Evolution of percentage contribution of different VOC species from 3 to 7 PMF factor solutions.

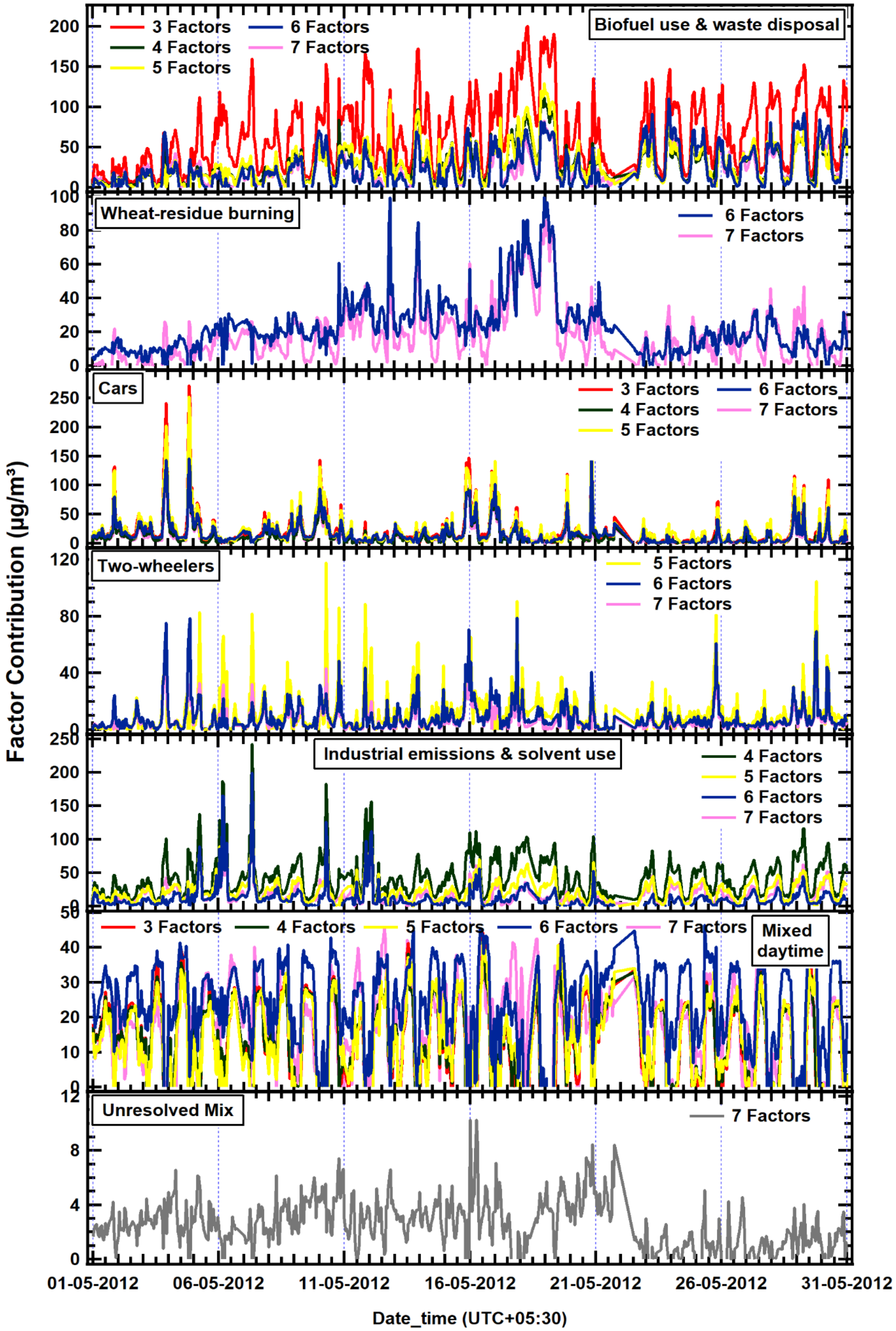

Figure S4c. Evolution of PMF factor contributions from 3 to 7 factor solutions.

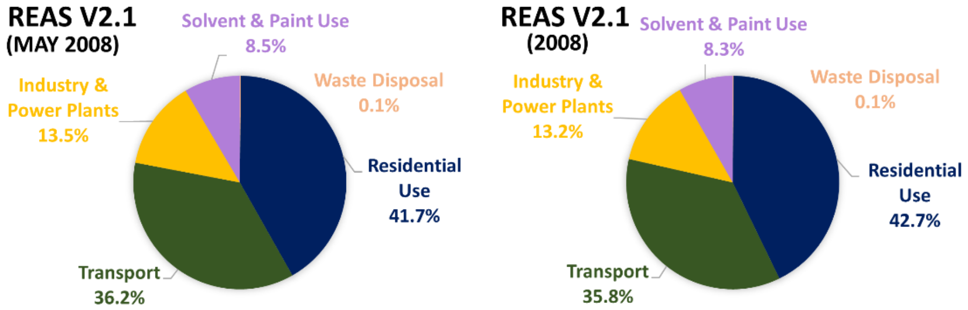

Figure S6. REAS database comparison to VOC source sectors on monthly and yearly resolution scales.

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.