Supplement of Atmos. Chem. Phys., 18, 15601–15622, 2018
https://doi.org/10.5194/acp-18-15601-2018-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Source apportionment of fine particulate matter in Houston, Texas: insights to secondary organic aerosols

Ibrahim M. Al-Naiema et al.

Correspondence to: Elizabeth A. Stone (betsy-stone@uiowa.edu)

Supplemental Information

Selection of the number of AMS PMF factors

Figure S1 presents the ratio between the summation of scaled residuals (Q) and the expected value for this parameter ($Q_{expected}$) obtained for PMF models including 1 to 7 factors. The expected value of Q corresponds to the number of data points in the HR OA concentration matrix and represents a scenario in which the experimental errors equal the resulting model errors (Paatero, 2013). The $Q/Q_{expected}$ ratio under this ideal scenario would be 1, and thus, the relative value of this ratio can be used to evaluate deviations from this ideal condition. The summation of scaled residuals decreases as additional factors are incorporated into the model, and as illustrated in Figure S1, after a certain number of PMF factors, $Q/Q_{expected}$ exhibits only marginal reductions. According to Figure S1, major decreases in Q are observed when a second and third factor are considered in the model, while the retention of more than 5 factors does not lead to further evident reductions in the model residuals. Based on these observations, and considering the significant decrease in $Q/Q_{expected}$ after a third factor is incorporated, PMF models containing 3 to 5 factors were selected as potentially appropriate to explain the OA levels during the field campaign

The convergence of the 3 to 5-factor PMF models to a global minimum was examined by using twelve different initialization points (positions in a list of random values referred to as seeds; Paatero, 2013). The consistency of the model outputs utilizing twelve seeds was evaluated in terms of the variation in the $Q/Q_{expected}$ for each initialization point. As presented in Figure S2, at least two solutions with slightly differing $Q/Q_{expected}$ values were observed for the 4-factor model, while multiple outputs with close $Q/Q_{expected}$ levels resulted for the model containing 5 factors. In contrast, the 3-factor model exhibited $Q/Q_{expected}$ values resembling a single output likely corresponding to a global minimum.

The 3 to 5-factor PMF models were also evaluated in terms of their ability to reproduce the measured PM_1 OA mass concentrations during the sampling period. Figure S3 compares the experimental and modeled OA concentrations based on these models. As illustrated in Figure S3, the reconstructed OA mass concentrations based on PMF models including 3, 4 and 5 factors closely resemble the measured OA levels, indicating the suitability of these modeling approaches. According to Figure S3, the inclusion of a fourth/fifth factor did not lead to improved modeling of the OA mass concentration and the additional factors were insufficient to reproduce a large spike in OA observed on 5/23/2015.

The degree of similarity between the mass spectra of the factors in the 3 to 5-factor models was examined based on the spectral contrast angle (θ). As factor splitting may occur when additional factors are incorporated in the model, the θ between the factors provides insight on the presence of this phenomenon. As presented in Table S1, the factors in the 3-factor model closely resemble 3 of the factors in the models containing 4 and 5 factors. Similarly, the factors in the 4-factor model were highly similar (or identical) to 4 of the factors in the model including 5 factors. The fact that the additional factors in the 4 and 5-factor models were distinct from those in the 3 and 4-factor models, respectively, discards the occurrence of factor splitting in these PMF models.

As reported in the manuscript, the mass spectra of the factors contained in the 3-factor model resembled multiple factors previously reported in the literature (θ below ~17°). In contrast, the mass signature of the additional factors in the 4 and 5-factor models did not exhibit resemblance with any of the AMS factors reported in the UCB-AMS Spectral Database (Ulbrich et al. 2017), preventing better evaluation of their physical meaningfulness.

Based on the larger stability of the PMF model containing 3 factors (as reflected by its repeated convergence to a single output), the marginal reduction in Q when a number of factors exceeding 3 was included in the PMF model, and the comparable OA mass concentrations reconstructed using 3 to 5-factor models, the 3-factor PMF model was selected as the most appropriate approximation to represent the observed OA levels during the field campaign.

Figure S1: $Q/Q_{expected}$ for PMF models containing 1 and 7 factors.

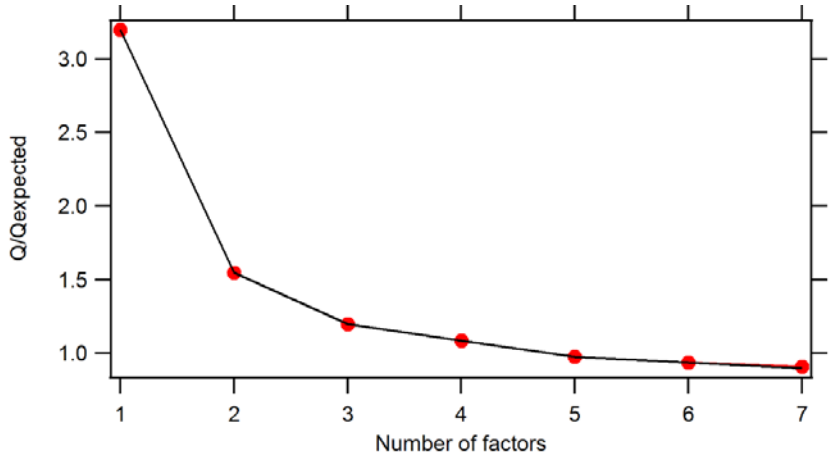

Figure S2: $Q/Q_{expected}$ for PMF models containing 3, 4 and 5 factors when different initialization points (seeds) are selected in PET v. 2.08D. Each seed number corresponds to a position in a list of randomly generated values (Paatero, 2013).

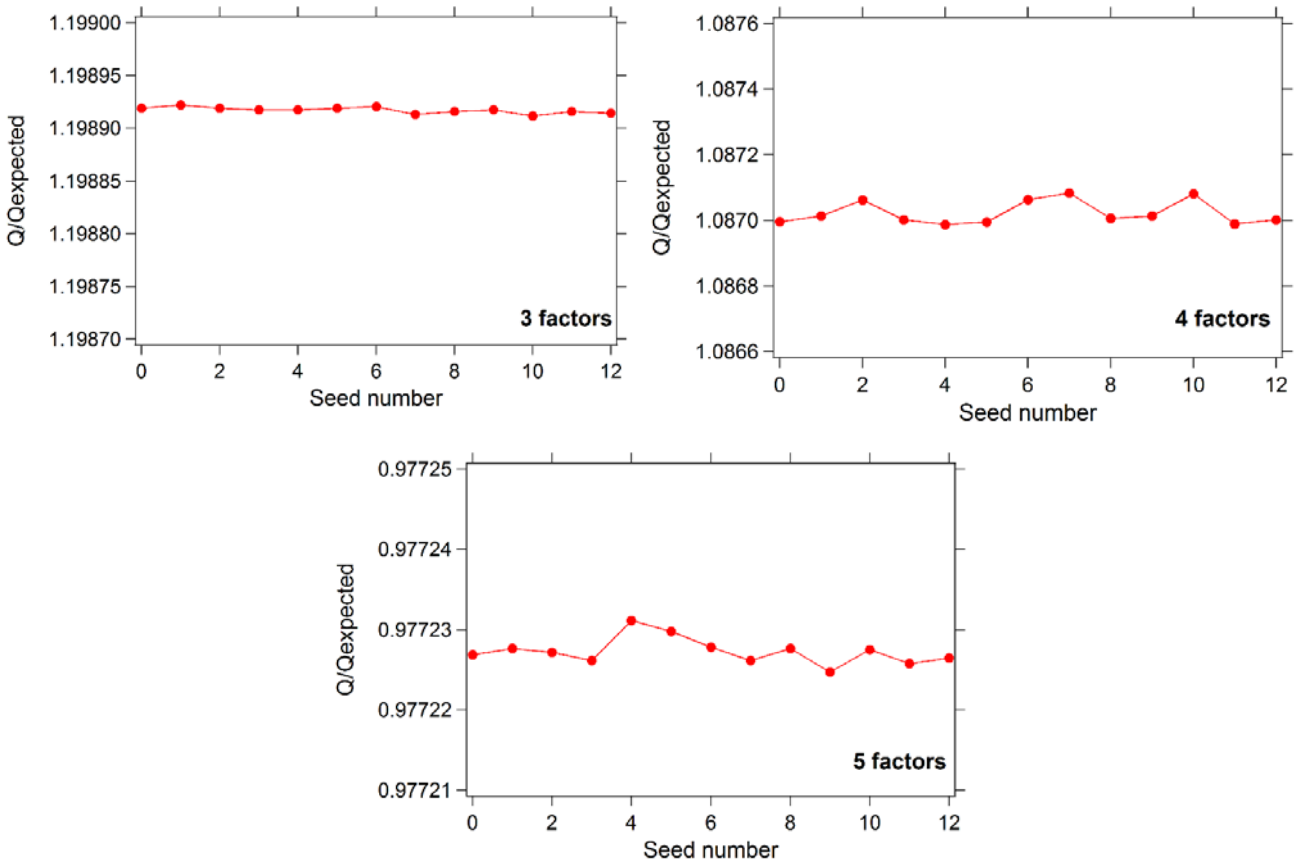

Figure S3: Measured and reconstructed OA mass concentration based on PMF models containing 3 (a), 4 (b) and 5 (c) factors.

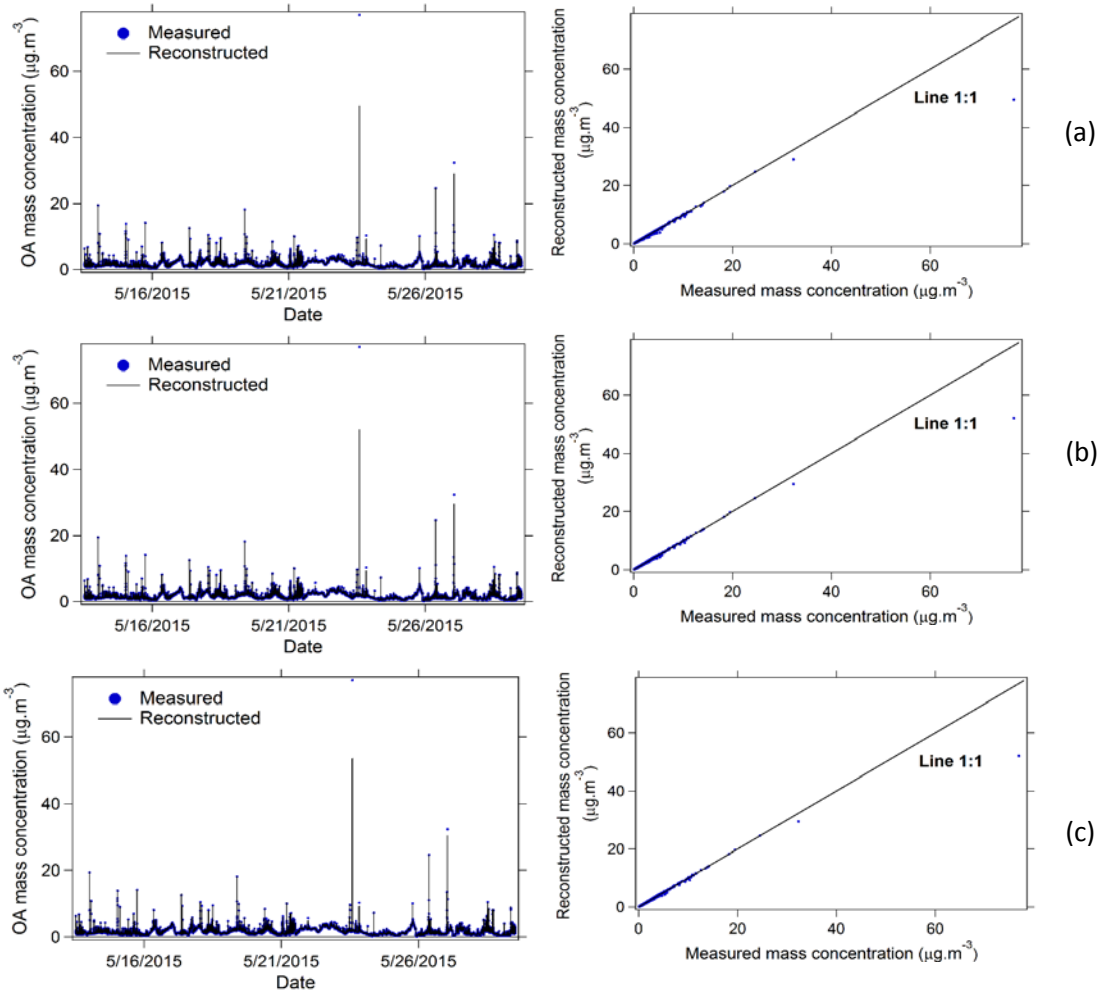

Figure S4: Box-and-whiskers plots showing the diurnal profiles of (a) $PM_{2.5}$, (b) nitrogen oxides, (c) ozone, (d) toluene, (e) isoprene, (f) UV radiation, (g) wind speed, and (h) ambient temperature at the Clinton Drive monitoring station in Houston, TX for 5-27 May 2015. The bottom whisker, bottom box line, top box line and top whisker indicate the 5th, 25th, 75th and 95th percentiles, respectively. Lines inside the boxes represent the hourly median and the continuous line represents the hourly mean. The hourly wind rose diagram (i) shows the frequency of occurrence (radial axis). All data were obtained from TCEQ (TCEQ, 2017). Daytime hours (7:00 –18:00 LT) are highlighted yellow and correspond to the times of the daytime $PM_{2.5}$ filter samples. Nighttime filter samples were collected 19:00- 6:00 LT.

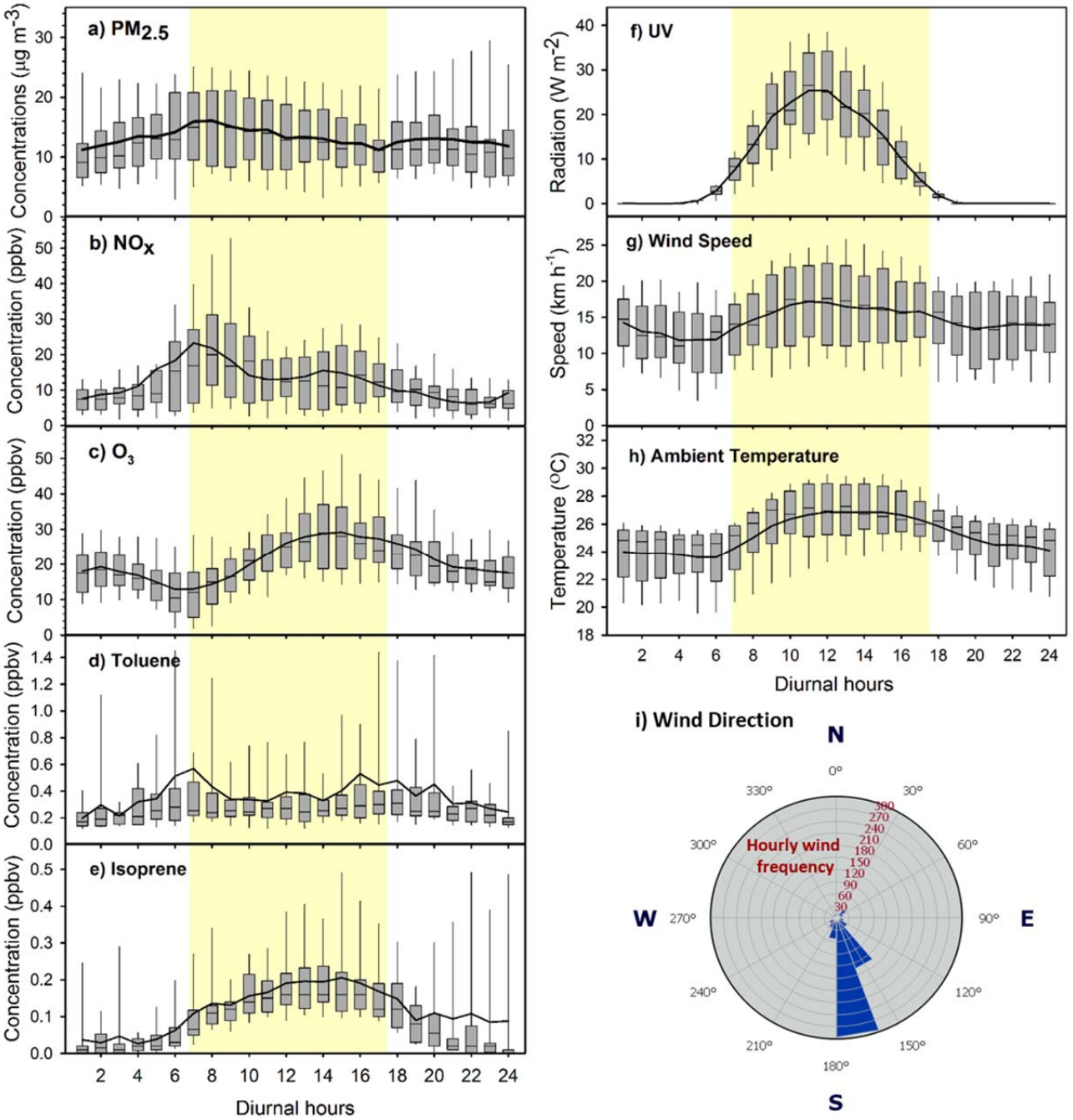

Figure S5: Hourly $PM_{2.5}$ concentrations (μg m^{-3}) and precipitation (cm) at the Clinton Drive monitoring station in Houston, TX, during May 2015.

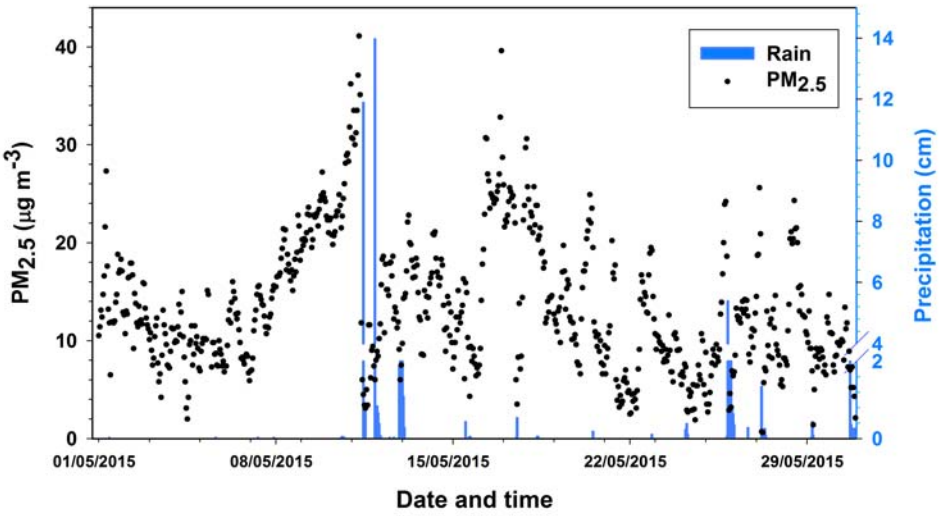

Figure S6: Hourly diurnal profiles (25^{th}, 50^{th}, and 75^{th} whiskers) of OM:OC, O:C, and H:C of NR-PM_1 OC at Clinton Drive measured by HR-ToF-AMS in May 2015.

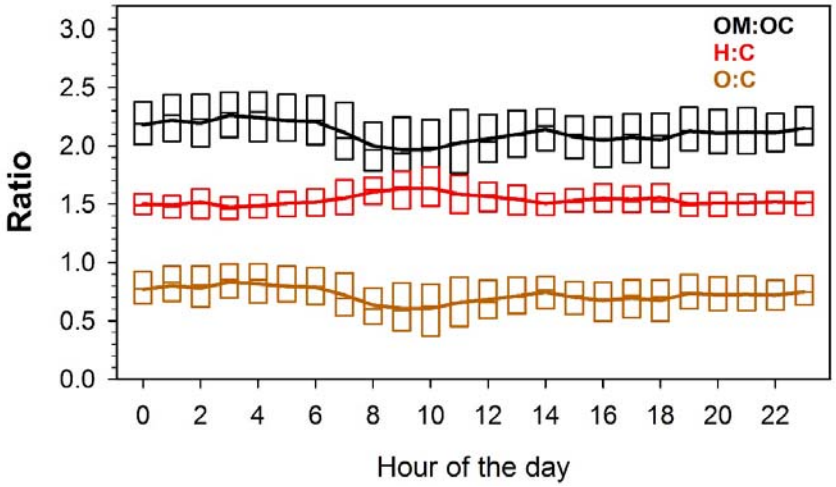

Figure S7: Active fires (1 month – Terra/MODIS) in May 2015; adapted from the maps created by Jesse Allen and Reto Stockli, NASA Earth Observatory, using data courtesy of the MODIS Land Science Team at NASA Goddard Space Flight Center (https://neo.sci.gsfc.nasa.gov/analysis/index.php).

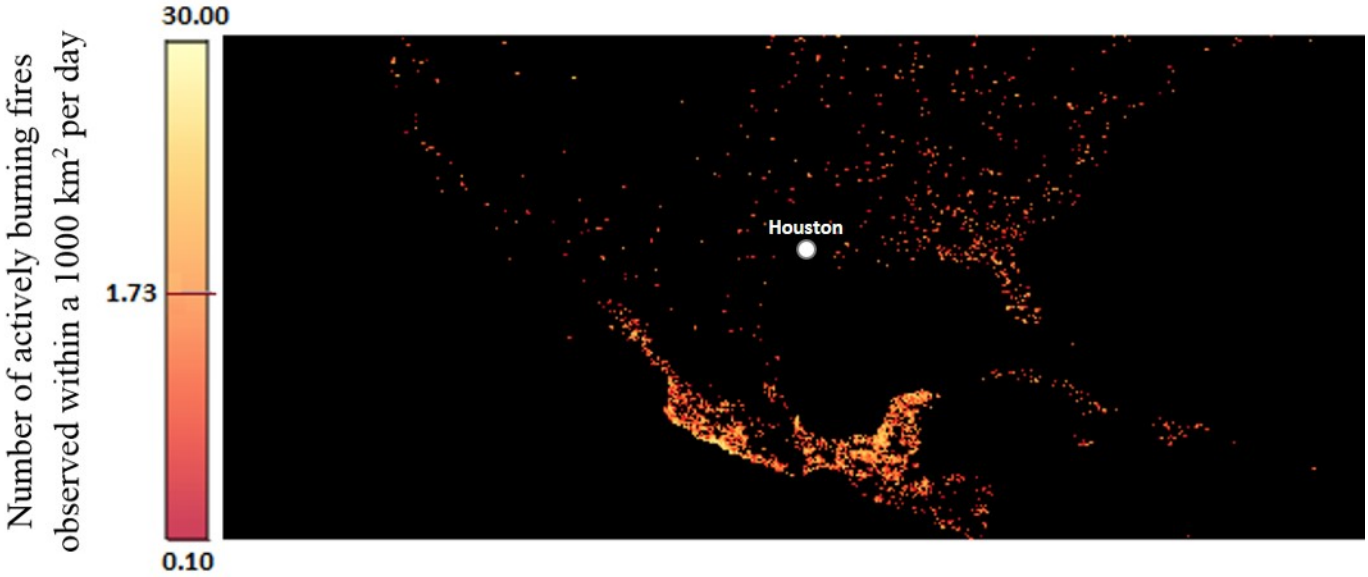

Figure S8: Three day back trajectories produced using the National Oceanic and Atmospheric Administration's (NOAA's) Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model. Data were collected for air masses ending 500 m above the ground level at the Clinton Drive monitoring site in Houston, TX (29.733943° N, 95.257684° W, indicated by star); a) from 05–17 May 2015 and b) from 17–29 May 2015 at 1200 UTC (07:00 LT). A new trajectory was started for every 24 hours. Symbols in each trajectory indicate the location of air mass for every 24 hours for up to 3 previous days.

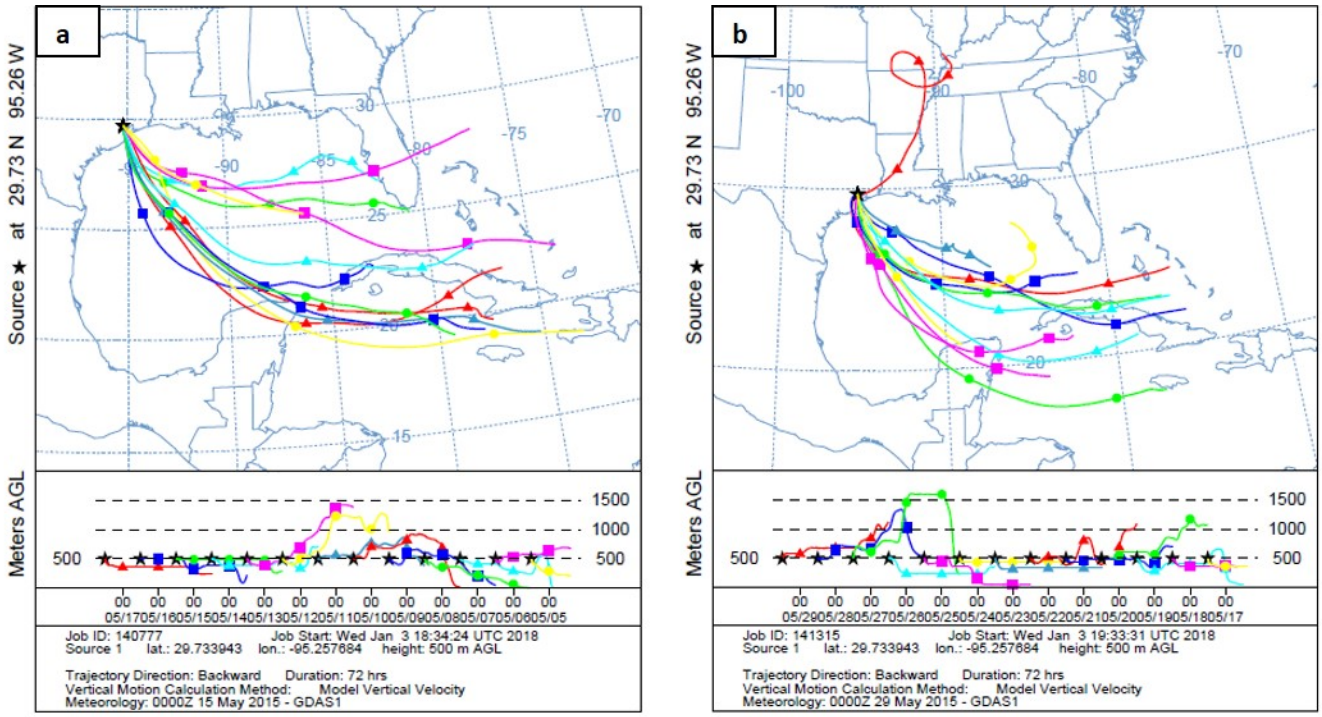

Figure S9: Isoprene emissions reported by industrial facilities in the Houston area in 2015 (https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-2016)

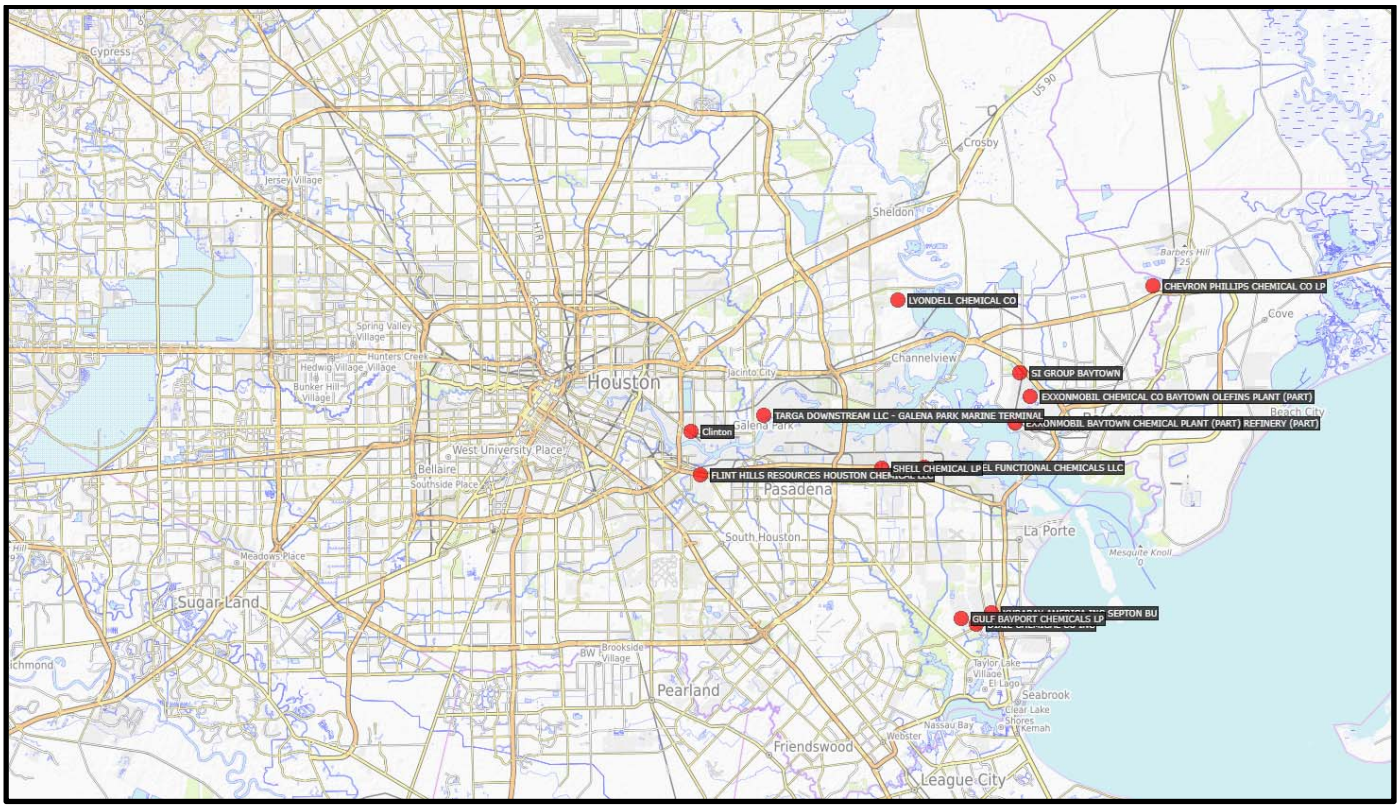

Figure S10: Average source contributions to $PM_{2.5}$ organic carbon (OC; µg m^{-3}, squares) in at Clinton Drive in May 2015 obtained for the 9 factor MM-PMF solution. The circles represent the median OC from bootstrapping (BS) and the average OC from displacement (DISP) and BS-DISP error estimation methods. The error bars represent the concentration range of OC from DISP, BS and BS-DISP. The lower and the upper whiskers represent the 5^{th} and the 95^{th} percentiles of the OC concentrations from BS and BS-DISP and the minimum and maximum OC concentrations from DISP.

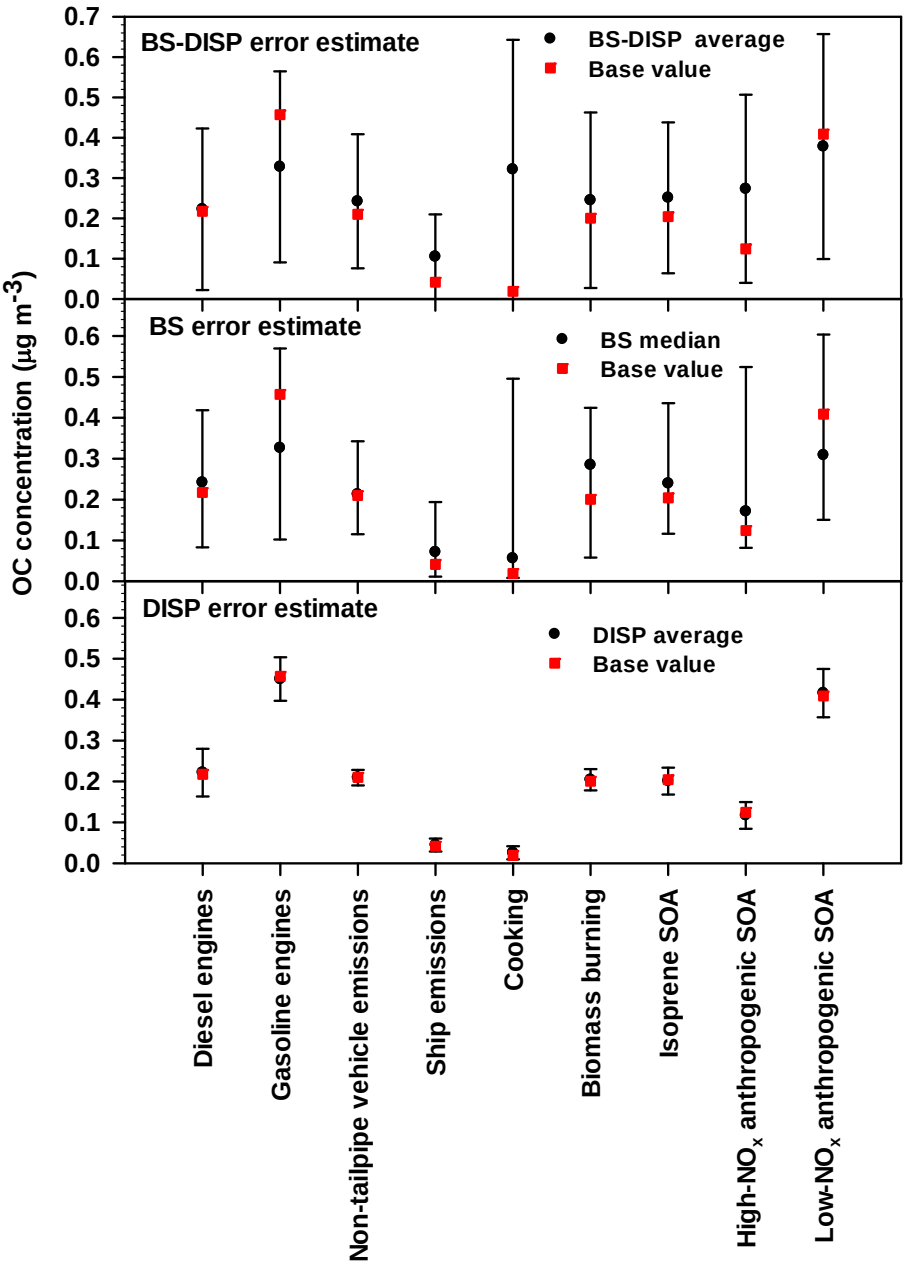

Table S1: Spectral contrast angle (θ °) between the mass spectrum of factors in the 3 to 5-factor AMS-PMF solutions. Numbers in bold indicate close resemblance between mass spectral signatures.

		4-factor solution				5-factor solution				
	Factor	1	2	3	4	1	2	3	4	5
3-factor model	1	**2.6**	49.1	32.1	75.3	44.0	**2.5**	31.5	51.0	76.2
	2	37.4	25.0	**2.7**	51.8	25.8	37.1	**7.3**	30.3	52.5
	3	76.8	45.8	52.6	**0**	54.2	76.7	55.9	35.7	**0**
4-factor model	1					46.6	**0**	33.7	53.0	78.0
	2					**12.0**	51.7	30.0	27.7	47.8
	3					26.9	34.3	**5.5**	31.6	54.0
	4					55.3	76.9	56.4	36.7	**0**

Table S2: The MM-PMF input data statistics for the modeled chemical compounds and base model diagnostics obtained for the 9-factor solution.

Variable	PMF input data statistics		Base model diagnostics obtained for the 9 factor solution		
			Measured (X) vs. modeled (Y) scatter plots		Residual analysis
	S/N[a]	Category[b]	Slope	R^2	Q_{true}[d]/Q_{exp}[e]
Organic carbon[c]	7.4	Weak	0.853	0.650	0.8
Elemental carbon	6.9	Strong	0.718	0.706	16.0
Cyclopenta(cd)pyrene	1.7	Strong	0.391	0.646	12.1
Benz(a)anthracene	4.6	Strong	0.745	0.792	4.1
Chrysene	5.2	Strong	0.855	0.893	6.7
Benzo(b)fluoranthene	2.2	Strong	0.501	0.492	4.3
Benzo(k)fluoranthene	2.3	Strong	0.810	0.907	4.7
Benzo(e)pyrene	2.0	Strong	0.431	0.464	5.0
Benzo(a)pyrene	2.6	Strong	0.799	0.831	6.5
Indeno(1,2,3-cd)pyrene	1.1	Strong	0.381	0.400	4.5
Benzo(ghi)perylene	2.6	Strong	0.731	0.800	4.9
Pristane	0.7	Weak	0.496	0.521	0.4
Octadecane	1.0	Strong	0.562	0.551	2.4
Nonadecane	0.8	Weak	0.558	0.649	0.5
Docosane	0.9	Weak	0.692	0.979	0.2
Tricosane	4.1	Strong	0.814	0.990	3.8
Tetracosane	3.5	Strong	0.877	0.993	1.2
Pentacosane	3.5	Strong	0.932	0.994	1.0
Hexacosane	2.4	Strong	0.986	0.987	1.1
Heptacosane	2.3	Strong	0.982	0.970	2.1
Octacosane	4.5	Strong	0.761	0.725	8.4
Nonacosane	5.7	Strong	0.671	0.633	8.9
17α(H)-21β(H)-30-norhopane	5.8	Strong	0.844	0.810	10.6
17α(H)-21β(H)-Hopane	3.2	Strong	0.958	0.986	1.3
Cholesterol	2.6	Strong	1.265	0.951	6.2
Levoglucosan	8.5	Strong	0.664	0.706	6.6
5-Nitrosalicylic acid	0.9	Weak	0.273	0.311	0.4
4-Nitrophenol	1.5	Strong	0.060	0.056	9.3
2-Methyl-4-nitrophenol	0.5	Weak	0.166	0.096	0.3
4-Methyl-3-nitrophenol	1.7	Strong	0.373	0.972	11.7
4-Methyl-2-nitrophenol	1.9	Strong	0.001	0.006	19.8
2,3-dihydroxy-4-oxopentanoic acid	4.5	Strong	0.811	0.809	10.4
Phthalic acid	9.0	Strong	0.262	0.298	10.8
4-Methylphthalic acid	8.4	Strong	0.744	0.912	6.9
Terephthalic acid	4.1	Strong	0.057	0.099	14.7
Isophthalic acid	6.0	Strong	0.099	0.301	23.5
2-Methylglyceric acid	8.8	Strong	0.789	0.970	2.0
2-Methylthreitol	9.2	Strong	1.011	0.989	1.4
2-Methylerythritol	9.9	Strong	0.987	0.990	2.1
cis-Pinonic acid	3.0	Strong	0.089	0.141	13.9

[a]signal to noise ratio calculated by PMF (EPA, version 5); [b]weak, if S/N=0.5-1, strong, if S/N>1 and the uncertainties of the compounds categorized as weak are automatically increased by a factor of 3; [c] total variable, which by default categorized as weak; [d]goodness-of-fit parameter calculated including all points by PMF; [e]goodness-of-fit parameter calculated using the difference between the number of non-weak data points in the input data matrix and the number of elements in the two matrices, source profiles and source contributions, taken together

Table S3: Summary of MM-PMF settings for base runs and error estimations

Parameter	Description
Data type; sample collection/averaging time frame	$PM_{2.5}$; 05-27 May 2015 based on daytime (07:00-18:00 LT) and nighttime (19:00 – 06:00 LT) schedule
# of species	40
Total variable	Organic carbon
# of samples	46
# of factors	3 to 11
Treatment of missing data	No missing data
Treatment of data below detection limit (BDL)	No modifications or censoring of BDL data
Treatment of concentrations equal to or less than zero	No modifications or censoring of data ≤ 0
Lower limit for normalized factor contributions g_{ik}	-0.2
Robust mode	Yes
Constraints	None
Seed value	Random
# of base runs	20 When developing a solution (3-11 factors) and 100 (5-9 factors) when determining a final solution
# of bootstraps in BS	100
R^2 for BS	0.6
BS block size	1
DISP dQ_{max}	4, 8, 16, 32
# of DISP active species	34 (only the species categorized as strong)
# of bootstraps and r^2 for BS in BS-DISP	100 and 0.6
BS-DISP active species	Elemental carbon, chrysene, benzo(b)fluoranthene, benzo(GHI)perylene, 17α(H)-21β(H)-hopane, cholesterol, levoglucosan, methylphthalic acid, 2-methylglyceric acid, 2-methylthreitol, 2-methylerythritol, 2,3-dihydroxy-4-oxopentanoic acid
BS-DISP dQ_{max}	0.5. 1, 2, 4

Table S4. Elemental ratios, carbon oxidation state, m/z 44 and m/z 43 fractions of PMF factors identified in the PM_1 OA during the field campaign.

PMF Factor	O:C	H:C	$\overline{OS}_C$	f_{43}	f_{44}
HOA	0.06	2.03	-1.91	0.11	0.02
CI-LO-OOA	0.61	1.57	-0.35	0.10	0.11
MO-OOA	1.24	1.21	1.27	0.03	0.26

Table S5: Correlation between CI-LO-OOA and mass fragments previously reported as tracers of food cooking activities

Mass fragment	Coefficient of correlation (R)	Reference(s)
$C_3H_3O^+$	0.89	(Mohr et al., 2012) (Sun et al., 2016) (Wallace et al., 2018)
$C_2H_3O^+$	0.88	(Mohr et al., 2009) (Liu et al., 2017)
$C_5H_8O^+$	0.73	(Sun et al., 2016) (Sun et al., 2011)
$C_2H_4O_2^+$	0.70	(Mohr et al., 2009)
$C_6H_6O^+$	0.75	(Wallace et al., 2018)
$C_6H_{10}O^+$	0.51	(Elser et al., 2016) (Cao et al., 2018) (Sun et al., 2016) (Sun et al., 2011)

Table S6: CMB model performance metrics. The R^2 values indicate the fit of the profile to the ambient data, with values greater than 0.8 indicating a good model fit. The χ^2 values are the weighted sum of squares of the differences between the calculated and measured fitting species concentrations, with a value of 0 indicating a perfect model fit, <1 indicating a very good fit, 1-2 indicating an acceptable fit, and >4 indicating that one or more species concentrations are not well explained by the model. In one sample (21 May, nighttime) this value was greater than 4, with 2.85 the next-highest value. The calculated-to-measured concentration ratios of the fitting species indicate the extent to which individual tracers were fit by the model. SOA tracers, which behaved ideally coming from only one source, had calculated-to-measured concentrations of 1.

Performance Metric	Range			Mean	Median
R^2	0.833	-	0.999	0.979	0.987
χ^2	0.06	-	4.35	0.93	0.71
Calculated-to-measured concentration ratios					
elemental carbon	0.99	-	1.01	1.00	1.00
levoglucosan	0.61	-	1.32	0.98	0.99
17α(H)-21β(H)-hopane	0.00	-	0.42	0.21	0.20
17α(H)-21β (H)-30-norhopane	0.75	-	1.07	0.99	1.00
17α(H)-22,29,30-trisnorhopane	0.67	-	1.33	0.99	1.00
benzo(b)fluoranthene	0.67	-	2.00	1.05	1.00
benzo(ghi)perylene	0.00	-	1.78	0.92	1.00
indeno(1,2,3-cd)pyrene	0.00	-	2.00	1.06	1.00

Table S7: Summary of MM-PMF diagnostics and error estimation statistics obtained for 5 to 9 factor solutions.

Diagnostic	5 Factors	6 Factors	7 Factors	8 Factors	9 Factors
$Q_{expected}$	1164	1084	1004	924	844
Q_{robust} (minimum)	8361	7290	6406	5568	4841
Q_{true} (corresponding to min Q_{robust})	11117	9569	7961	6770	5844
$Q_{robust}/Q_{expected}$	7.2	6.7	6.4	6.0	5.7
$\Delta(Q_{robust}/Q_{expected})$	-	0.46	0.34	0.35	0.29
DISP %dQ	-9.57E-05	0	0	0	0
DISP swaps at dQ_{max} 4, 8, 16, and 32	0	0	0	0	0
Factors with BS mapping < 80 %	Mixed biomass burning, non-tailpipe vehicle emissions, and ship emissions (34%)	Ship emissions (58%) and mixed biomass burning and non-tailpipe vehicle emissions (58%)	Gasoline engines (73%) and ship emissions (76%)	High-NO_x anthropogenic SOA (56%) and diesel engines (72%)	High-NO_x anthropogenic SOA (67%) and biomass burning (54%)

Note: When changing from 7 to 8 factor solution high-NO_x anthropogenic SOA resolved from low-NO_x anthropogenic SOA. When changing from 8 to 9 factor solution non-tailpipe vehicle emissions resolved from biomass burning. Also, for the 9-factor solution levoglucosan is completely resolved from isoprene SOA.

Works Cited

Cao, L.-M., Huang, X.-F., Li, Y.-Y., Hu, M., He, L.-Y., Volatility measurement of atmospheric submicron aerosols in an urban atmosphere in southern China. Atmos. Chem. Phys. 18, 1729-1743. 10.5194/acp-18-1729-2018, 2018.

Elser, M., Huang, R.J., Wolf, R., Slowik, J.G., Wang, Q., Canonaco, F., Li, G., Bozzetti, C., Daellenbach, K.R., Huang, Y., Zhang, R., Li, Z., Cao, J., Baltensperger, U., El-Haddad, I., Prévôt, A.S.H., New insights into PM2.5 chemical composition and sources in two major cities in China during extreme haze events using aerosol mass spectrometry. Atmos. Chem. Phys. 16, 3207-3225. 10.5194/acp-16-3207-2016, 2016.

Liu, T., Li, Z., Chan, M., Chan, C.K., Formation of secondary organic aerosols from gas-phase emissions of heated cooking oils. Atmos. Chem. Phys. 17. 10.5194/acp-17-7333-2017, 2017.

Mohr, C., DeCarlo, P.F., Heringa, M.F., Chirico, R., Slowik, J.G., Richter, R., Reche, C., Alastuey, A., Querol, X., Seco, R., Peñuelas, J., Jiménez, J.L., Crippa, M., Zimmermann, R., Baltensperger, U., Prévôt, A.S.H., Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data. Atmos. Chem. Phys. 12, 1649-1665. 10.5194/acp-12-1649-2012, 2012.

Mohr, C., Huffman, J.A., Cubison, M.J., Aiken, A.C., Docherty, K.S., Kimmel, J.R., Ulbrich, I.M., Hannigan, M., Jimenez, J.L., Characterization of Primary Organic Aerosol Emissions from Meat Cooking, Trash Burning, and Motor Vehicles with High-Resolution Aerosol Mass Spectrometry and Comparison with Ambient and Chamber Observations. Environmental Science & Technology 43, 2443-2449. 10.1021/es8011518, 2009.

Paatero, P., User's Guide for Positive Matrix Factorization Programs PMF2 and PMF3, Part 1: Tutorial. University of Helsinki, Finland. http://helsinki.fi/~paatero/PMF/pmf2.zip, 2013.

Sun, Y., Du, W., Fu, P., Wang, Q., Li, J., Ge, X., Zhang, Q., Zhu, C., Ren, L., Xu, W., Zhao, J., Han, T., Worsnop, D.R., Wang, Z., Primary and secondary aerosols in Beijing in winter: sources, variations and processes. Atmos. Chem. Phys. 16. 10.5194/acp-16-8309-2016, 2016.

Sun, Y.L., Zhang, Q., Schwab, J.J., Demerjian, K.L., Chen, W.N., Bae, M.S., Hung, H.M., Hogrefe, O., Frank, B., Rattigan, O.V., Lin, Y.C., Characterization of the sources and processes of organic and inorganic aerosols in New York city with a high-resolution time-of-flight aerosol mass apectrometer. Atmos. Chem. Phys. 11, 1581-1602. 10.5194/acp-11-1581-2011, 2011.

TCEQ: Texas Commission on Environmental Quality monitoring site information. www.tceq.texas.gov/cgi-bin/compliance/monops/agc_monthly_summary.pl?user_site=48_201_1035, (accessed May 09, 2017), 2017.

Ulbrich, I. M., Handschy, A., Lechner, M., and Jimenez, J. L.: High-Resolution AMS Spectral Database. URL: http://cires.colorado.edu/jimenez-group/HRAMSsd/, Accessed August 10, 2017.

Wallace, H.W., Sanchez, N.P., Flynn, J.H., Erickson, M.H., Lefer, B.L., Griffin, R.J., Source apportionment of particulate matter and trace gases near a major refinery near the Houston Ship Channel. Atmospheric Environment 173, 16-29. https://doi.org/10.1016/j.atmosenv.2017.10.049, 2018.