

Supplement of Atmos. Chem. Phys., 20, 12047–12061, 2020  
<https://doi.org/10.5194/acp-20-12047-2020-supplement>  
© Author(s) 2020. This work is distributed under  
the Creative Commons Attribution 4.0 License.



*Supplement of*

## **Source apportionment of PM<sub>2.5</sub> in Shanghai based on hourly organic molecular markers and other source tracers**

**Rui Li et al.**

*Correspondence to:* Li Li (lily@shu.edu.cn) and Jian Zhen Yu (jian.yu@ust.hk)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

## **Section S1. The rationality of combining data from two sites**

In this manuscript, we combined the PM<sub>2.5</sub> major component data from the PD site and organic molecular markers data from the SAES site for the source apportionment analysis. The two sites are 12 km apart, both are characterized as a general urban location. We agree that it would be desirable to use chemical composition data from one single site to carry out the source analysis. Unfortunately, in this work, the SAES site had measurements of organic tracers, major ions, and elements, but did not have ECOC measurements. The PD site had data of major ions, elements, and OCEC, but no organic tracer data. As a result, we had to resort to “borrowing” certain composition data from a similar site nearby. The data combination provided a more comprehensive dataset to fully characterize the PM<sub>2.5</sub> pollution sources, for the general urban district in this region.

The detailed rationality of combining data from the two sites is explained in the following two sub-sections.

### **S1.1 The neighborhood characteristics of the two sites are similar**

Shanghai’s land area is part of the alluvial plain of the Yangtze Delta Region, with an average height of 2.19 meters above the sea. The average wind speed is around 3-4 m/s. The whole Shanghai area belongs to one air basin. The geographic distance between the two sites is 12.1 km. Figure S1 shows the surroundings of the two sites. Both sites are surrounded by large residential areas, in addition to scattered shopping malls and restaurants, indicating the similarities of mixed emissions influence for the two sites. The PD site is a designated urban monitoring station in the Pudong district (Zhao et al., 2013a). The SAES site is located in the Xuhui district and more pollution characteristics of this site can be found in Wang et al. (2018).

### **S1.2. The pollution characteristics at the two sites are similar**

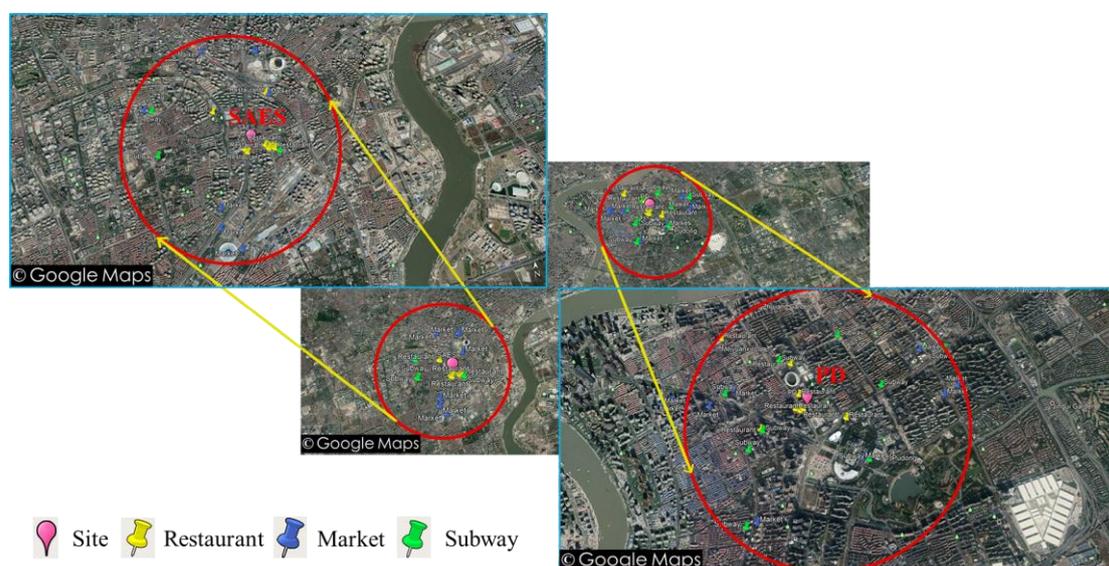
PM<sub>2.5</sub> mass concentrations, gaseous pollutants (CO, SO<sub>2</sub>, NO<sub>2</sub>) and PM<sub>2.5</sub> carbonaceous components (i.e., BC, OC and EC) concentrations between the two sites are examined. Table S1 lists the campaign-average concentrations and percentage difference of the average concentrations between the two sites. Figure S2 compares the time series and correlations of the PM<sub>2.5</sub> mass, and gaseous pollutants (CO, SO<sub>2</sub>, NO<sub>2</sub>) during the measurement period between the two sites.

As shown in Figure S2, the pollution levels of PM<sub>2.5</sub> and NO<sub>2</sub> at the two sites showed an excellent degree of agreement with each other, ( $R^2=0.92$ , slope=0.95) for PM<sub>2.5</sub> and ( $R^2=0.78$ , Slope=0.86) for NO<sub>2</sub>. The site-site difference in their average concentration was less than 6% (Table S1). CO also showed a high correlation ( $R^2=0.78$ ) and similar concentration levels at the two sites. SO<sub>2</sub> showed a moderate correlation between the two sites ( $R^2=0.45$ ), its temporal variations were broadly in synchrony.

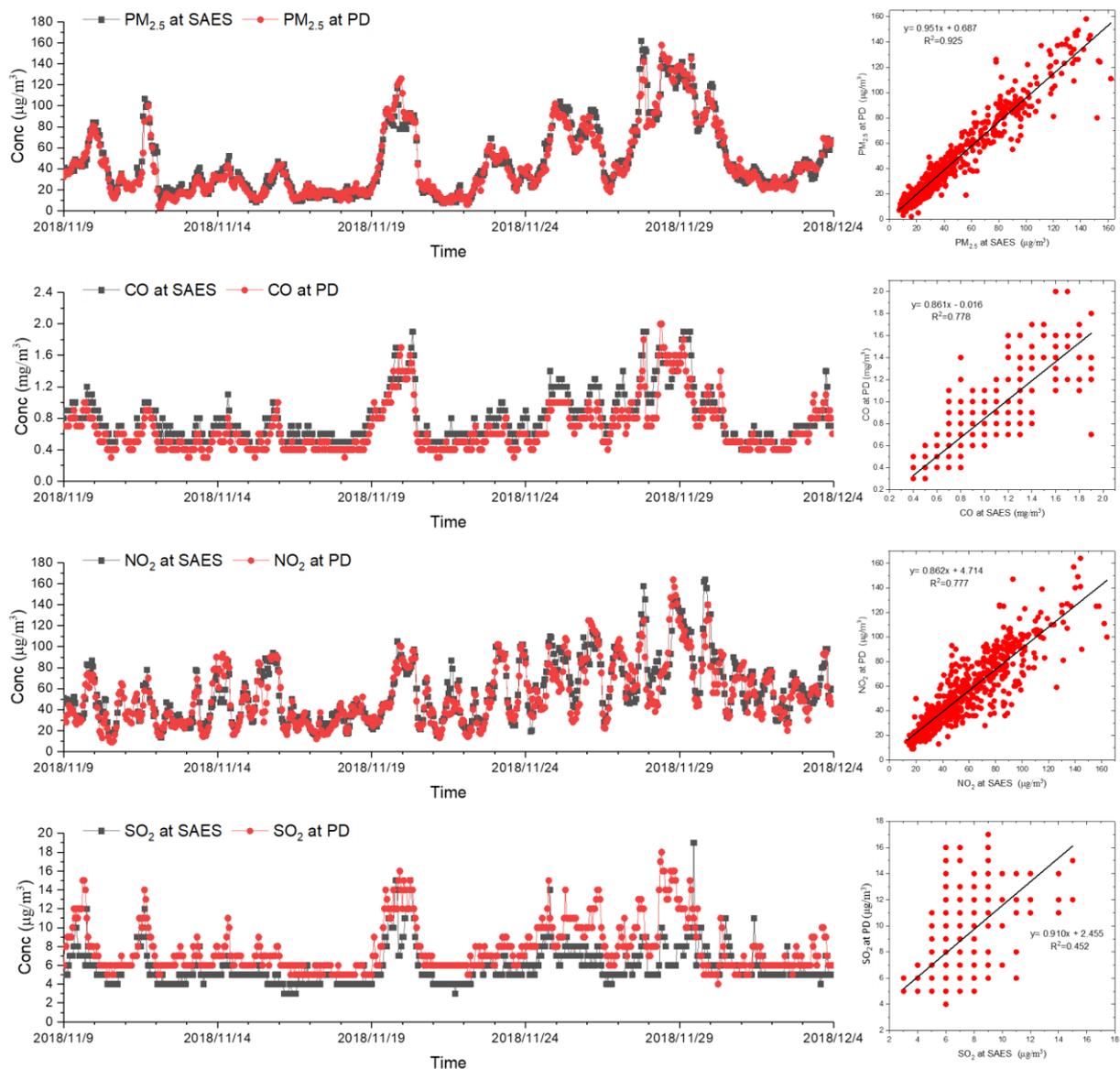
No OC and EC data were available for the SAES site during the sampling period. Thus, we next compare related carbonaceous measurements. Specifically, BC measured by aethalometer at SEAS is compared with EC measured at PD (Figure S3), showing synchronous variations, and their concentration levels were also similar, with a percentage difference of 33%. AMS-measured PM<sub>1.0</sub> OA at SAES is compared with

PM<sub>2.5</sub> OM mass estimated using OC at PD (Figure S3). The two quantities tracked very well throughout the measurement period, except for one OA episode lasting for a few hours on 29 Nov. Additionally, they showed comparable levels between the two sites, with a relative difference of 8-17%.

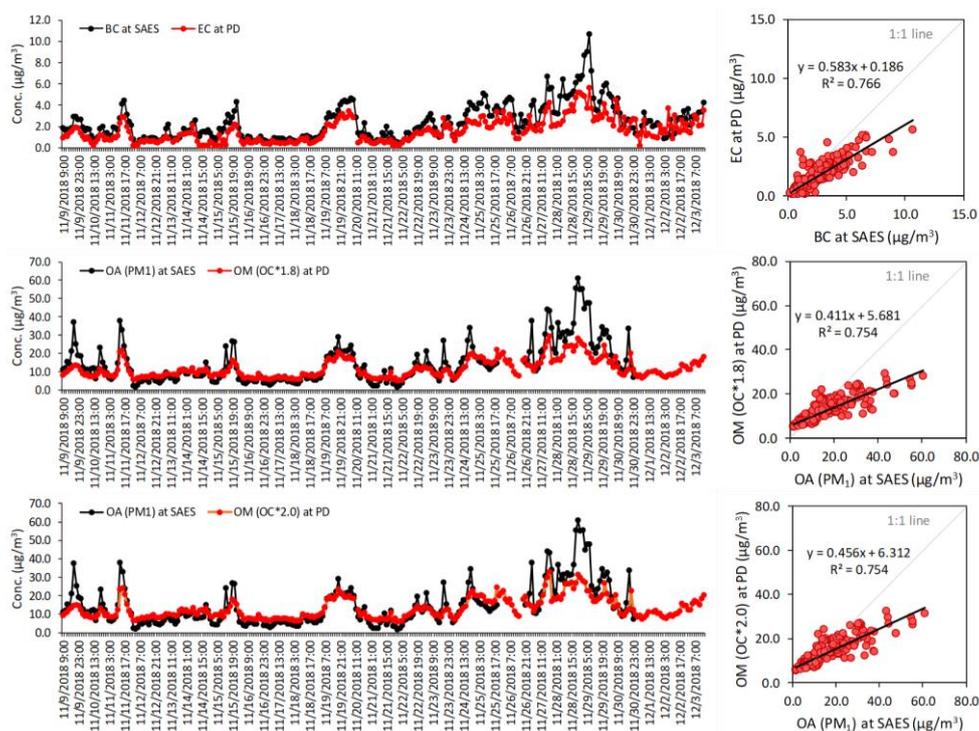
In summary, we have shown with measurement data that the major PM<sub>2.5</sub> components at the two urban sites are synchronous in temporal variations and highly similar in concentration levels. This provides data support to argue for the approach of using a combined data set to achieve a comprehensive source apportionment of PM<sub>2.5</sub> for the urban districts in this work. We interpret that the source apportionment results using the combined dataset are representative of the common and major sources at the PD and SAES sites. Otherwise it would not be possible to observe the highly similar level of PM<sub>2.5</sub>. On the other hand, we recognize that the data combination approach would be ineffective to extract a potential source specific at one single site. More specifically, as pointed out by the reviewer, the differences of organic markers at the two sites may be non-negligible, thus, the apportioned results, especially the sources apportioned by certain organic markers (i.e., cooking in this study) could be site-specific to the SAES site alone. Nevertheless, the major source factors should be consistent at the two sites and can be used to reflect the general urban pollution in our study location. Finally, a rigorous quantification of the uncertainties caused by the data combination at the two sites needs side-by-side online measurements of PM compositions, especially organic markers at the two sites, and currently this information is not available.



**Figure S1.** Surroundings of the two sampling sites (i.e., PD and SAES) in this study (© Google Maps).



**Figure S2.** Comparison of PM<sub>2.5</sub> and gaseous pollutants between the PD and SAES sites during the measurement period.



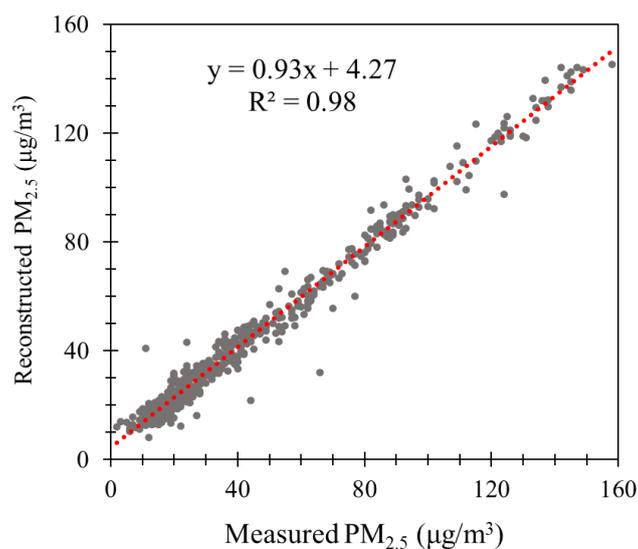
**Figure S3.** Comparison of PM<sub>2.5</sub> carbonaceous components between the PD and SAES sites during the measurement period. (OM/OC ratio from AMS measurement: 1.69±0.18 (1.4-2.0). 1.8→75<sup>th</sup> percentile of OM-to-OC ratio from AMS measurement; 2.0→ maximum value of OM-to-OC ratio from AMS measurement)

**Table S1.** Concentrations and relative difference of PM<sub>2.5</sub> mass, carbonaceous components, and gaseous pollutants (CO, NO<sub>2</sub> SO<sub>2</sub>) at the PD and SAES sites during the measurement period.

Species	SAES site (µg/m <sup>3</sup> )		PD site (µg/m <sup>3</sup> )		Relative Difference
	avg	stdev	avg	stdev	
PM <sub>2.5</sub>	47	34	46	33	-2.5%
BC at SAES vs. EC at PD	2.4	1.66	1.59	1.13	-33%
OC			6.5	2.8	
AMS-measured PM <sub>1</sub>			11.7 <sup>a</sup>	5.1	-17.4%
OA at SAES vs. PM <sub>2.5</sub>	14.1	11.1			
OM at PD estimated using OC			13 <sup>b</sup>	5.7	-8.2%
Gaseous pollutants					
CO (mg/m <sup>3</sup> )	0.79	0.33	0.66	0.32	-16.5%
NO <sub>2</sub>	57	30	54	29	-5.7%
SO <sub>2</sub>	5.9	1.96	7.8	2.7	32%

<sup>a</sup> Estimated PM<sub>2.5</sub> OM concentration assuming OM = 1.8×OC

<sup>b</sup> Estimated PM<sub>2.5</sub> OM concentration assuming OM = 2.0×OC



**Figure S4.** Comparison of reconstructed and measured  $\text{PM}_{2.5}$  mass for samples collected at PD during the measurement period. Reconstructed  $\text{PM}_{2.5}$  mass = sulfate + nitrate + ammonium + other water-soluble ions (sum of  $\text{Cl}^-$ ,  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Mg}^{2+}$ ) + EC + OM ( $1.8 \times \text{OC}$ ) + crustal material ( $2.49[\text{Si}] + 1.63[\text{Ca}] + 2.42[\text{Fe}]$ ) + other trace elements.

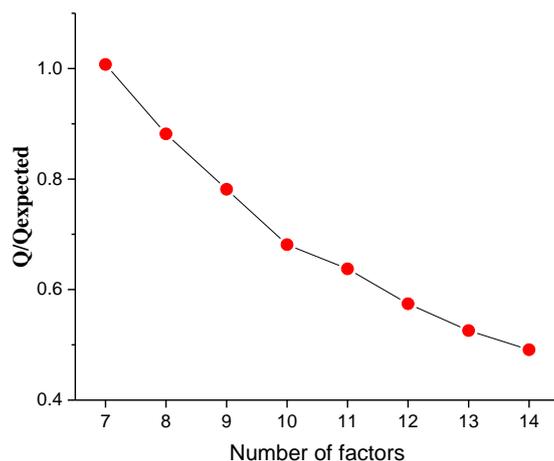
## Section S2. PMF factor number determination and model stability

### S2.1 PMF factor number determination

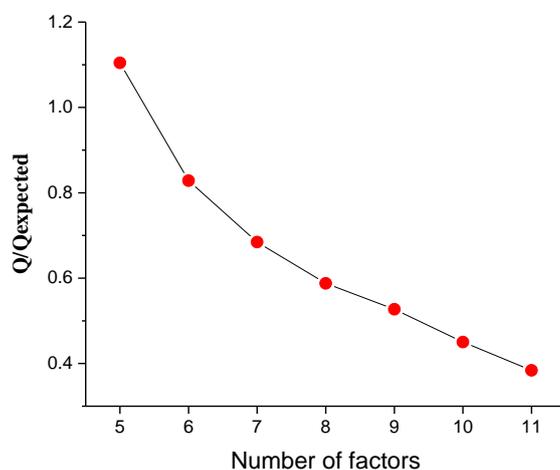
The change of  $Q/Q_{\text{exp}}$  values is often used as a reference to help in the factor number determination in PMF. When the number of factors increases to a certain value,  $Q/Q_{\text{exp}}$  will change less dramatically.

In MM-PMF,  $Q/Q_{\text{exp}}$  changed by 6.4% from the 10- to 11-factor model, less significant than the 11.4- 13.1% observed when the number of factors varied from 6 to 10, suggesting the factor number reaching to ten was needed to explain the input data (Fig. S5). However, when examining the factor profiles, the 11- factor solution provides the most reasonable source profiles by separating the vehicle exhaust and the secondary nitrate factor. The 11-factor solution was chosen as a final solution for MM-PMF.

In  $\text{PMF}_t$ ,  $Q/Q_{\text{exp}}$  changed by 10.3% from the 8- to 9-factor model, less significant than the 14.2- 25.0% observed when the number of factors varied from 5 to 8, suggesting the factor number reaching to eight or nine was needed for explaining the input data (Fig. S6). Increasing to 9 factors, Fe and Pb stood out as an unexplainable factor. Finally, the 8-factor result was chosen for  $\text{PMF}_t$ .



**Figure S5.** Change of  $Q/Q_{\text{exp}}$  values from 7 to 14 factors run for MM-PMF.



**Figure S6.** Change of  $Q/Q_{\text{exp}}$  values from 5 to 11 factors run for PMF<sub>t</sub>.

## S2.2 PMF model stability evaluation

Bootstrapping (BS), displacement (DISP) and bootstrap combined with displacement (BS-DISP) are three uncertainty estimation methods in the EPA PMF 5.0 software, to evaluate the model stability of the PMF solution (Norris et al., 2014). Tables S2 and S3 summarize the error estimation results for MM-PMF and PMF<sub>t</sub>, respectively. Tables S4 and S5 list the model performance of individual input species for MM-PMF and PMF<sub>t</sub>, respectively.

BS is used to identify whether there is a small set of observations that can disproportionately influence the solution. In BS, a new data set is constructed from the original input data with randomly sampling blocks. The BS factors from the resampled data matrices are mapped to the base run factors to provide the reproducibility of different base run factors due to the random errors. In the PMF user guide, it is suggested that factor mapping higher than 80% indicates robust factors. DISP mainly explores rotational ambiguity in the PMF results. In DISP, each element in the factor profiles is first adjusted up and down and then all other values are computed to achieve

the associated PMF within a pre-set change of  $dQ^{\max}$  values. In DISP, factor swap means that if factors change too much that they exchange identities. If factor swaps occur for the smallest  $dQ^{\max}$ , it indicates that there is significant rotational ambiguity and that the solution is not sufficiently robust to be used. BS-DISP is a hybrid approach with combination of the BS and DISP methods.

**Table S2.** Summary of error estimation diagnostics from BS, DISP and BS-DISP for MM-PMF.

<b>BS Mapping(<math>R \geq 0.6</math>)</b>	Secondary Nitrate	Secondary Sulfate	Vehicle Exhaust	Industrial / Tire Wear	Industrial Emission II	Residual Oil Combustion	Dust	Coal combustion	Biomass Burning	Cooking	SOA	Unmapped
Secondary Nitrate	96	0	0	0	0	0	0	0	0	0	0	4
Secondary Sulfate	2	93	0	0	0	0	0	0	1	0	0	4
Vehicle Exhaust	13	0	61	3	0	0	1	1	3	1	2	15
Industrial / Tire Wear	0	0	0	96	0	0	0	0	0	0	0	4
Industrial Emission 2	0	0	0	0	96	0	0	0	0	0	0	4
Residual Oil Combustion	0	0	0	0	0	96	0	0	0	0	0	4
Dust	8	1	1	4	0	0	76	1	0	1	0	8
Coal combustion	0	0	0	1	0	0	0	93	0	0	0	6
Biomass Burning	0	0	0	0	0	0	0	0	96	0	0	4
Cooking	0	0	0	0	0	0	0	0	0	96	0	4
SOA	5	0	0	0	0	0	0	0	0	0	90	5
<b>DISP Diagnostics</b>		Error Code: 0					Largest Decrease in Q: 0 (0%)					
Factor Swaps	$dQ^{\max}=4$	0	0	0	0	0	0	0	0	0	0	0
	$dQ^{\max}=8$	0	0	0	0	0	0	0	0	0	0	0
	$dQ^{\max}=15$	0	0	0	0	0	0	0	0	0	0	0
	$dQ^{\max}=25$	0	0	0	0	0	0	0	0	0	0	0
<b>BS-DISP Diagnostics</b>		# of runs accepted: 33 out of 101					Largest Decrease in Q: -15.6 (-0.61%)					
# of Decreases in Q: 2			# of Swaps in Best Fit: 40					# of Swaps in DISP: 26				
Factor Swaps	$dQ^{\max}=0.5$	26	8	22	17	28	4	3	10	23	1	2
	$dQ^{\max}=1$	29	8	22	17	28	4	3	10	26	1	2
	$dQ^{\max}=2$	30	8	22	22	29	4	3	11	32	1	2
	$dQ^{\max}=4$	34	9	23	27	33	4	4	11	41	3	3

**Table S3.** Summary of error estimation diagnostics from BS, DISP and BS-DISP for PMF<sub>t</sub>.

<b>BS Mapping(<math>R \geq 0.6</math>)</b>	Secondary Nitrate	Secondary Sulfate	Vehicle Exhaust	Industrial /Tire Wear	Industrial Emission II	Residual Oil Combustion	Dust	Coal combustion	Unmapped
Secondary Nitrate	100	0	0	0	0	0	0	0	0
Secondary Sulfate	1	93	0	1	0	0	1	1	3
Vehicle Exhaust	11	0	73	5	0	0	0	3	8
Industrial / Tire Wear	0	0	0	100	0	0	0	0	0
Industrial Emission 2	0	0	0	0	100	0	0	0	0
Residual Oil Combustion	0	0	0	0	0	100	0	0	0
Dust	0	0	1	1	0	0	97	1	0
Coal combustion	2	1	0	1	0	0	0	96	0
<b>DISP Diagnostics</b>	Error Code: 0				Largest Decrease in Q: -17.6 (-1.3%)				
Factor Swaps	$dQ^{\max}=4$	0	0	0	0	0	0	0	0
	$dQ^{\max}=8$	0	0	0	0	0	0	0	0
	$dQ^{\max}=15$	0	0	0	0	0	0	0	0
	$dQ^{\max}=25$	0	0	0	0	0	0	0	0
<b>BS-DISP Diagnostics</b>	# of runs accepted: 52 out of 101				Largest Decrease in Q: -2.3 (-0.2%)				
# of Decreases in Q: 0		# of Swaps in Best Fit: 30				# of Swaps in DISP: 19			
Factor Swaps	$dQ^{\max}=0.5$	0	18	5	23	36	22	7	22
	$dQ^{\max}=1$	0	19	9	24	40	22	8	23
	$dQ^{\max}=2$	0	22	14	25	48	22	9	25
	$dQ^{\max}=4$	1	25	18	27	60	22	10	31

**Table S4.** Model performance of individual input species for the 11-factor solution of MM-PMF.

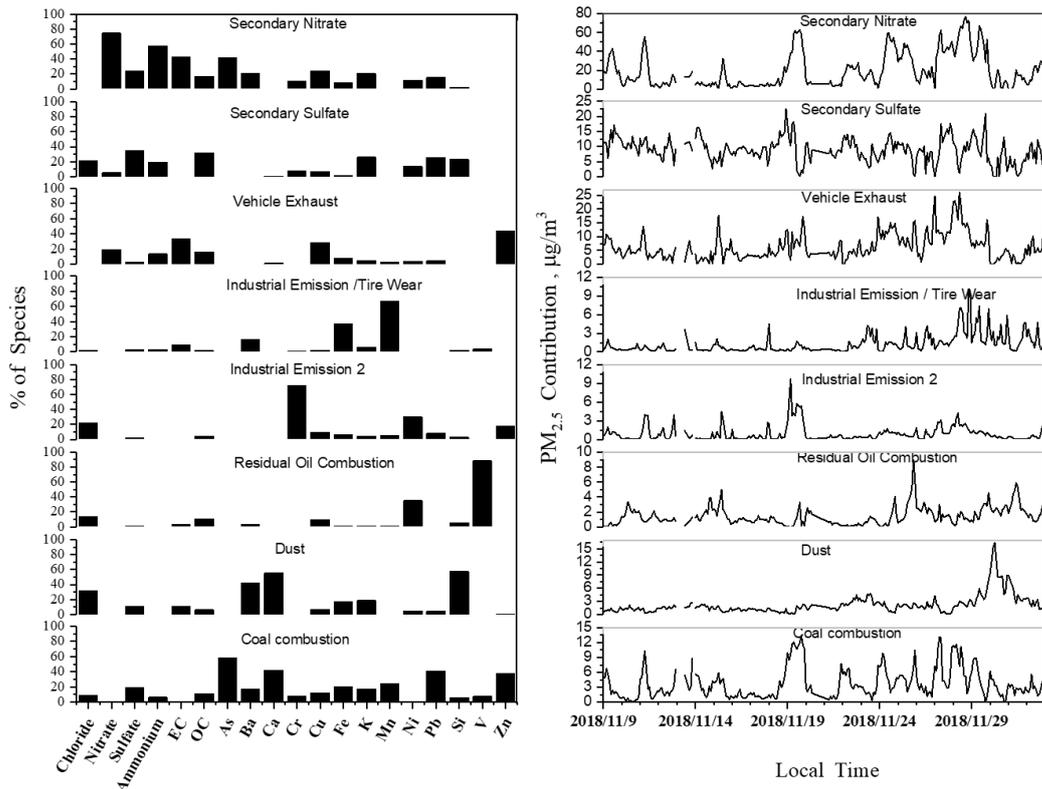
Species	R <sup>2</sup>	Intercept	Slope
Chloride	0.39	0.29	0.42
Nitrate	0.86	0.86	0.84
Sulfate	0.73	1.04	0.85
Ammonium	0.88	0.43	0.91
EC	0.86	-0.02	0.95
OC	0.86	-0.11	0.99
As	0.98	0.01	0.99
Ba	0.97	-0.01	0.99
Ca	0.95	0.01	0.93
Cr	0.95	0.01	0.92
Cu	0.88	0.01	0.86
Fe	0.42	0.22	0.32
K	0.95	0.01	0.99
Mn	0.98	0.01	0.97
Ni	0.91	0.01	0.85
Pb	0.54	0.02	0.42
Si	0.98	0.04	0.89
V	0.98	0.01	0.97
Zn	0.81	0.02	0.74
PAHs252	0.93	0.01	0.79
PAHs276	0.92	0.01	0.82
C <sub>6-8</sub> DICAs	0.78	0.02	0.65
OHBAAs	0.64	0.01	0.49
SFAs	0.81	0.02	0.62
phthalic acid	0.51	0.01	0.38
Mannosan	0.91	0.01	0.83
Levogluconan	0.88	0.01	0.93
α-pinT	0.72	-0.01	0.88
DHOPA	0.62	0.01	0.66
C <sub>9</sub> acids	0.84	0.02	0.82

**Table S5.** Model performance of individual input species for the 8-factor solution of PMF<sub>t</sub>.

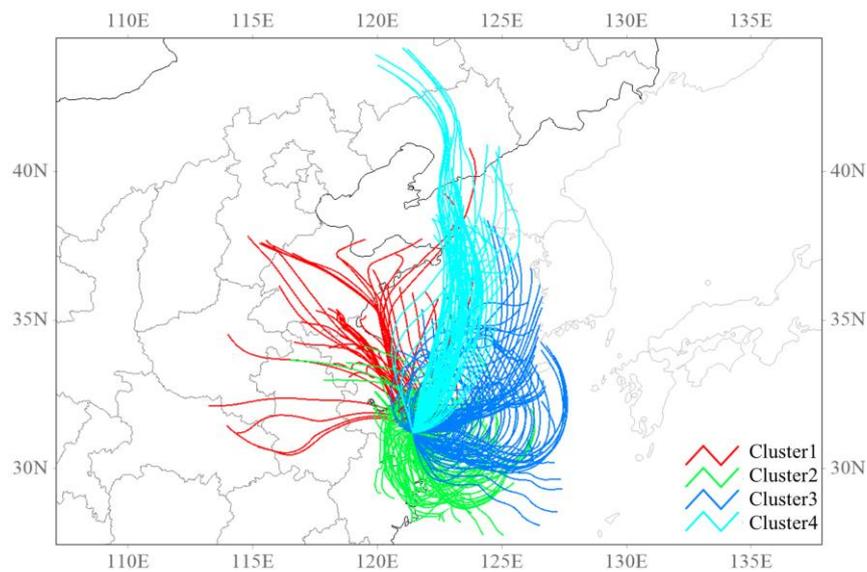
Species	R <sup>2</sup>	Intercept	Slope
Chloride	0.18	0.41	0.37
Nitrate	0.84	1.49	0.79
Sulfate	0.78	1.22	0.78
Ammonium	0.84	0.59	0.85
EC	0.85	-0.04	0.94
OC	0.84	0.22	0.93
As	0.98	-0.01	1.01
Ba	0.98	0	1.02
Ca	0.89	0.01	0.88
Cr	0.94	0.01	0.91
Cu	0.84	0.01	0.81
Fe	0.99	0.03	0.91
K	0.95	0.01	0.96
Mn	0.98	0.01	0.97
Ni	0.91	0.01	0.83
Pb	0.98	0	0.98
Si	0.98	0.04	0.88
V	0.99	0	0.99
Zn	0.81	0.01	0.75

**Table S6.** Pearson correlation (R) of MM-PMF factor contributions with meteorological parameters (wind speed (WS), temperature (T), and relative humidity (RH)) and gas pollutants (SO<sub>2</sub>, CO, and NO<sub>x</sub>). The highest correlation for each factor was highlighted in bold.

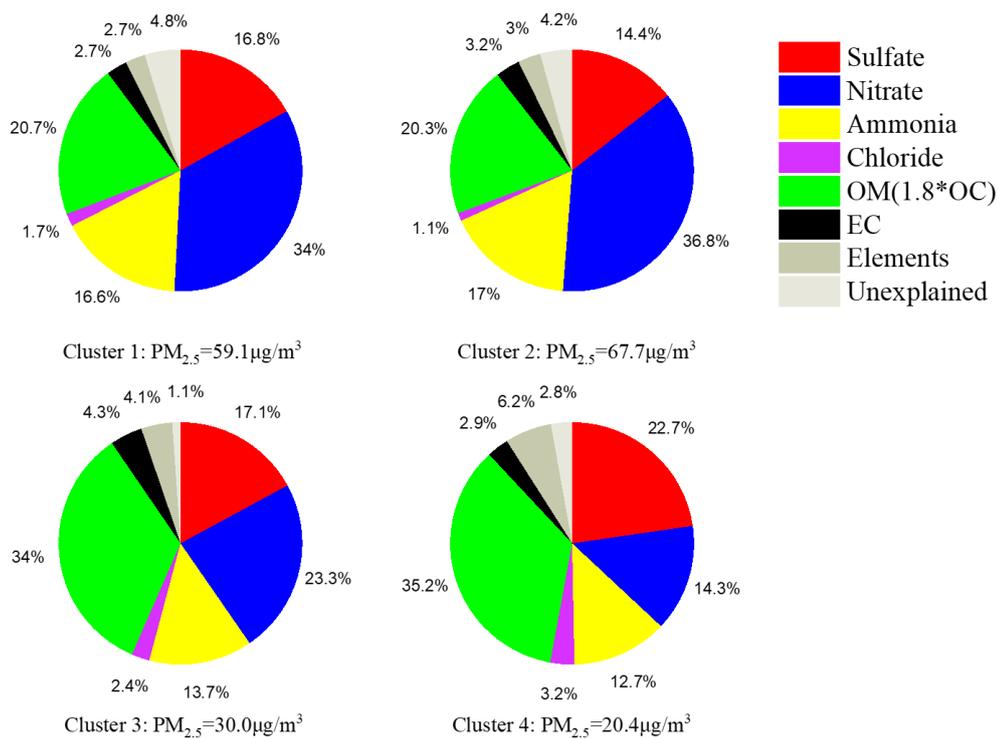
	Secondary Nitrate	Secondary Sulfate	Vehicle Exhaust	Industrial Emission /Tire Wear	Industrial Emission II	Residual Oil Combustion	Dust	Coal combustion	Biomass Burning	Cooking	SOA
WS	-0.50	0.41	-0.29	-0.32	-0.12	-0.08	-0.09	-0.28	-0.19	-0.10	-0.39
T	-0.13	-0.11	0.01	0.08	-0.21	0.37	0.22	-0.14	-0.19	-0.09	-0.28
RH	0.17	-0.30	-0.07	-0.01	0.07	0.20	-0.10	-0.18	0.16	0.27	0.05
SO <sub>2</sub>	0.47	0.11	0.53	0.25	<b>0.59</b>	0.01	-0.16	<b>0.68</b>	<b>0.58</b>	0.03	<b>0.69</b>
CO	<b>0.70</b>	0.17	0.48	0.32	<b>0.68</b>	0.01	-0.15	<b>0.68</b>	<b>0.70</b>	0.08	<b>0.79</b>
NO <sub>x</sub>	<b>0.50</b>	-0.34	<b>0.68</b>	<b>0.49</b>	0.38	0.31	0.12	0.39	0.49	0.21	0.46



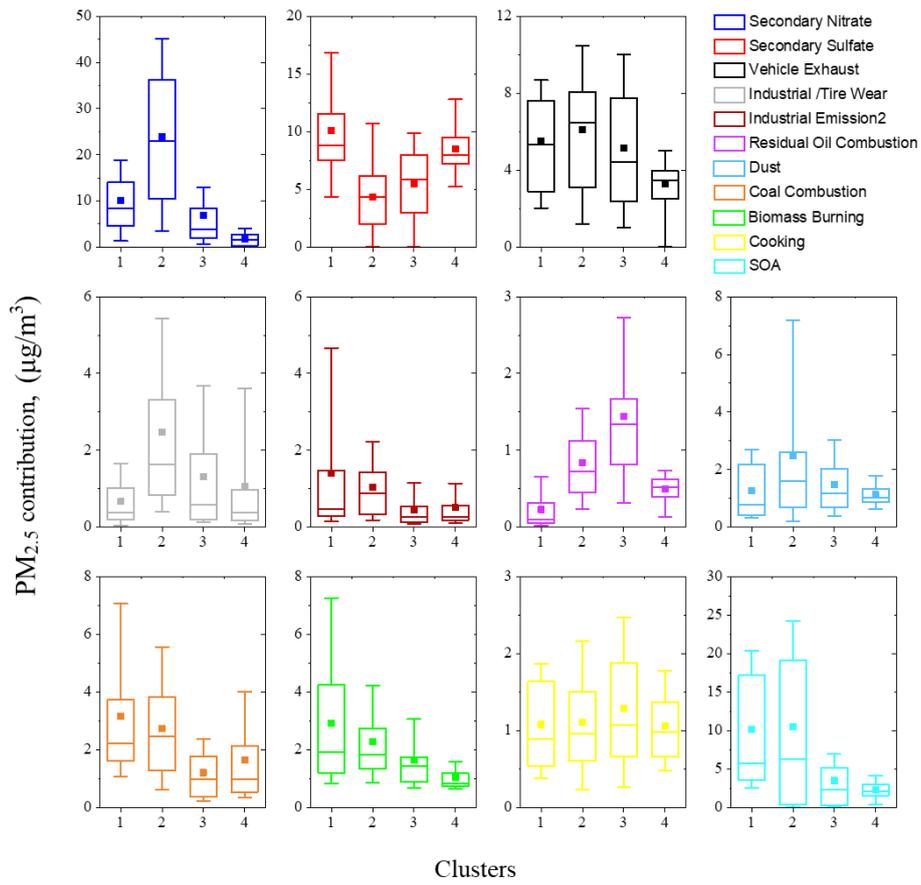
**Figure S7.** Resolved factor profiles (left) and factor contributions (right) in the eight-factor solution in PMF<sub>t</sub>.



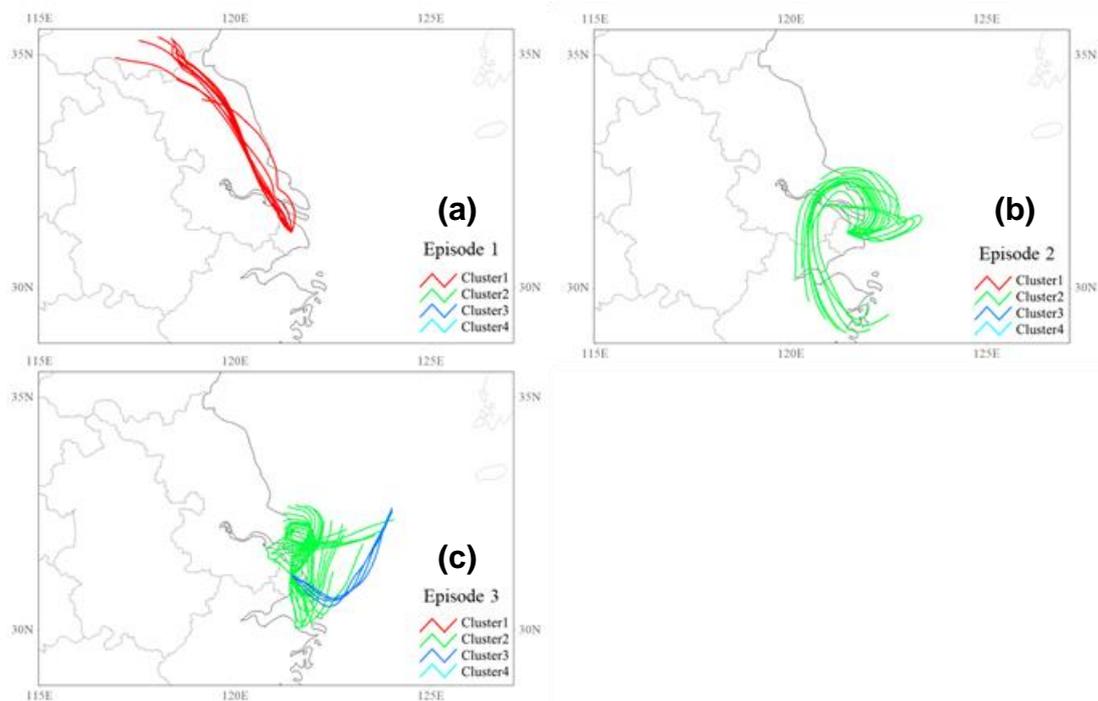
**Figure S8.** Distributions of all air mass trajectories during the whole sampling period, color coded by the corresponding clusters.



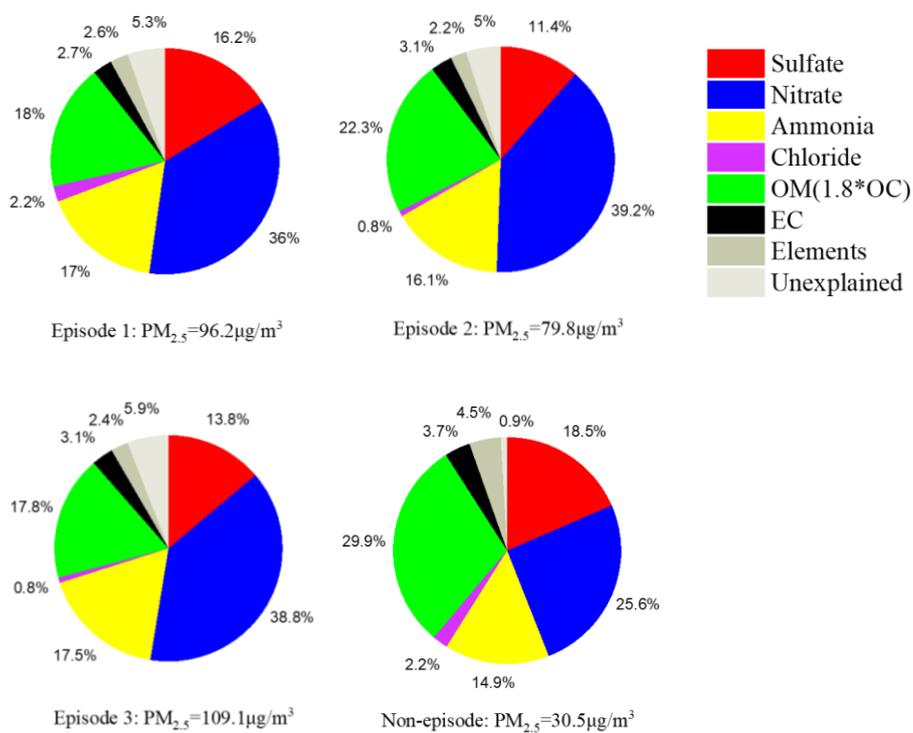
**Figure S9.** Average concentration of PM<sub>2.5</sub> and its major components for different air mass clusters.



**Figure S10.** Box plot of individual MM-PMF resolved source contributions ( $\mu\text{g}/\text{m}^3$ ) to  $\text{PM}_{2.5}$  for different air mass clusters (25<sup>th</sup> and 75<sup>th</sup> percentile boxes, 10<sup>th</sup> and 90<sup>th</sup> percentile whiskers; lines inside the boxes represent the hourly median and the red points represent the hourly mean).



**Figure S11.** Distributions of individual backward trajectories during the three episodes: (a) Episode 1, (b) Episode 2, and (c) Episode 3. The trajectories are color-coded by their corresponding clusters.



**Figure S12.** Average concentrations of PM<sub>2.5</sub> and its chemical compositions under each episode. The PM compositions for non-episodic hours are also shown for comparison.