Supplementary of

Short-term Source Apportionment of Fine Particulate Matter with Time-dependent Profiles Using SoFi: Exploring the Reliability of Rolling Positive Matrix Factorization (PMF) Applied to Bihourly Molecular and Elemental Tracer Data

Q. Wang et al.

Correspondence to: Jian Zhen Yu (jian.yu@ust.hk)

Text S1. General description of the pollution at DSL site

Table S1 shows the average concentrations of measured $PM_{2.5}$ and its component during the campaign period, which separated into (1) before CNY period (29/12/2019-23/1/2020) and (2) CNY and post-CNY period (24/1-9/2/2020). During the second sub-period, anthropogenic sources related to human activities such as traffic and industrial activities were greatly restricted due to the Covid-19 pandamic. Figure S1 shows the percentage change in average concentration of $PM_{2.5}$ and the select tracer species in the before CNY period compared with the CNY and post-CNY period at DSL site. Due to the lockdown restriction, most of the pollutants showed decreased concentrations in the second sub-period. The $PM_{2.5}$ concentrations decreased by 23% from 62.0 ± 43.0 μg m^{-3} in the before CNY period to 47.8 ± 31.1 μg m^{-3} during CNY and post-CNY period. Large reduction (>70%) was observed for primary tracer species such as hopanes, Ca, Mn, and Zn, reflecting the reduction in associated traffic and industiral activities due to restriction. Levoglucosan decreased by 18%, much smaller than above-mentioned species, as biomass burning were not affected by the national lockdown policies. K and Ba showed obvious increase (77% and 184%) during the CNY and post-CNY period, a result of the increase in fireworks during holidays. For the secondary species, nitrate, ammonium, phthalic acid and DHOPA showed larger decrease (-44% to -58%), reflecting the reduction in the precursors. Sulfate and α-pinene SOA tracers showed smaller decrease (-18% and -30%). More details about the pollution characteristics at this site can be found in our previous paper (Wang et al., 2022b).

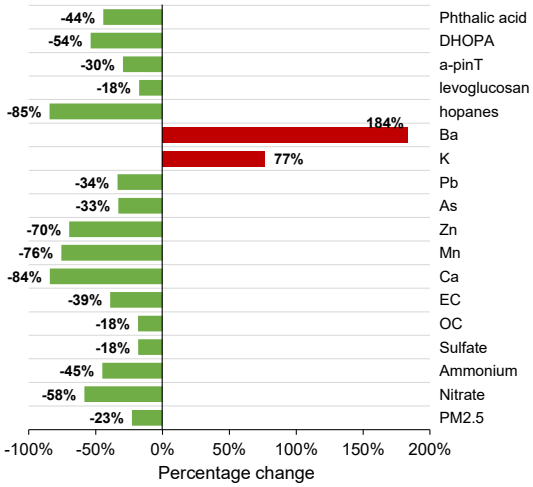


Figure S1. Percentage change in average concentrations of $PM_{2.5}$ and representative tracer species in CNY and post-CNY period compared with the before CNY period.

Table S1. Average concentration of $PM_{2.5}$ composition data during the before CNY period, CNY and post-CNY period and the whole campaign period at DSL site.

	Before CNY period		CNY and post-CNY period		Percentage change (%)
	Avg	Stdev	Avg	Stdev	
$PM_{2.5}$ and major components (μg m^{-3})					
$PM_{2.5}$	62.0	43.0	47.8	31.1	-22.8%
Sulfate	9.79	5.67	8.01	5.23	-18.2%
Nitrate	23.8	18.2	9.90	7.18	-58.4%
Ammonium	11.0	7.47	6.03	4.01	-45.1%
OC	5.67	2.98	4.63	2.60	-18.3%
EC	2.58	1.44	1.57	0.86	-39.1%
Ca	0.085	0.074	0.013	0.014	-84.2%
Mn	0.042	0.039	0.010	0.008	-75.7%
Fe	0.42	0.34	0.14	0.093	-67.2%
Zn	0.119	0.101	0.036	0.032	-69.8%
As	0.008	0.007	0.005	0.004	-33.1%
Pb	0.032	0.026	0.021	0.015	-33.7%
Cu	0.013	0.010	0.016	0.020	16.0%
K	0.55	0.39	0.97	1.00	76.9%
Ba	0.019	0.014	0.055	0.080	183.7%
Organic tracers (ng m^{-3})					
Hopanes*	0.55	0.54	0.09	0.07	-84.5%
Steranes#	0.19	0.20	0.04	0.03	-77.3%
Levoglucosan	37.6	19.8	31.0	15.3	-17.6%
Mannosan	5.47	5.42	3.16	2.36	-42.2%
α-pinene SOA tracers (α-pinT)^	52.7	53.1	37.1	23.9	-29.6%
β-caryophyllinic acid	13.3	10.0	7.55	5.10	-43.4%
2,3-Dihydroxy-4-oxopentanoic acid (DHOPA)	2.33	2.74	1.08	0.86	-53.7%
Phthalic acid	26.7	24.4	14.8	16.1	-44.3%
Gas pollutants (ppb)					
NO_x	39.1	25.3	16.0	3.9	-59%
SO_2	2.9	3.1	2.6	0.8	-12%
O_3	20	13	42	11	106%

* sum of five most abundant hopanes: 22,29,30-trisnorhopane, αβ-norhopane, αβ-hopane, αβ-22S-homohopane, and αβ-22R-homohopane.
# sum of five most abundant steranes: αββ 20R-cholestane, αα 20R-cholestane, αββ 20R-methylcholestane, αββ 20R-ethylcholestane, and ααα 20R-ethylcholestane.
^ sum of five α-pinene SOA tracers: pinonic acid, pinic acid, 3-acetylglutaric acid, 3-hydroxy-4,4-dimethylglutaric acid, 3-hydroxyglutaric acid.

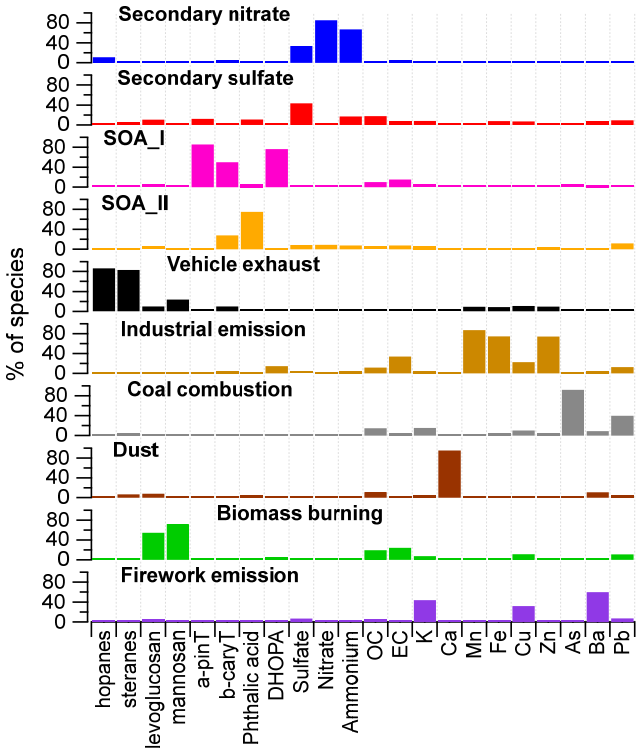


Figure S2. Source profiles resolved in the PMF_{ref} run using the campaign-wide data as input.

45

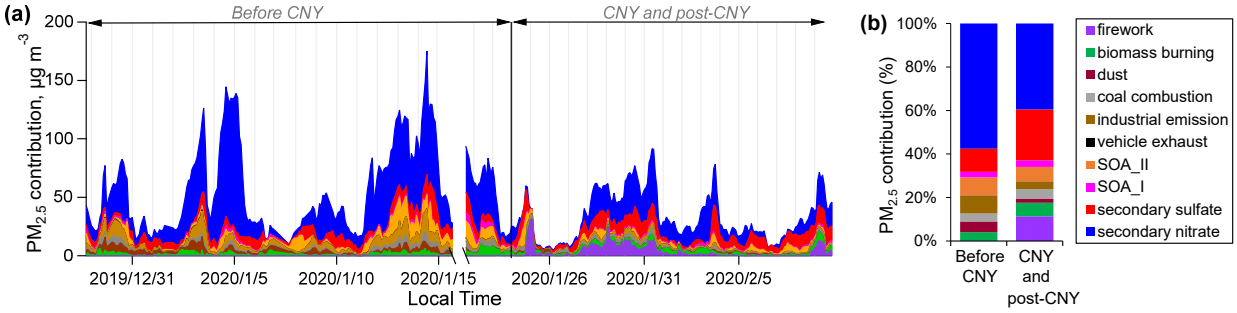


Figure S3. (a) Time series of $PM_{2.5}$ source contributions from individual factors resolved by PMF_{ref} and (b) percentage contributions to $PM_{2.5}$ from individual source factors resolved by PMF_{ref} during the before CNY and CNY and post-CNY period.

Text S2. Test of the window length for the short-term PMF

To find the appropriate window length for the short-term PMF, we tested the window length of 7 d, 10 d, 14 d and 18 d, and the statistical summary are showing in Table S2. PMF runs with window length less than 14 d showed mixed factor profiles of vehicle exhaust with other primary source factors, and the bootstrap error estimation failed, indicating non-robust solution. PMF run with window length of 14 d showed robust results, however, factor profile mixing between vehicle exhaust and biomass burning still occurred. The bootstrap resampling showed that SOA_I factor and secondary sulfate had less than 90% mapping. With increased window length of 18 d, the short-term PMF result showed increased bootstrap result with all factors showing mapping of >90% and less factor profile mixing between vehicle exhaust and biomass burning (Figure S4). Thus, window length of 18 d was chosen to conduct the short-term PMF analysis.

Table S2. Statistics of the different window length and the corresponding short-term PMF results.

PMF run No.	Window size (d)	Data coverage	PMF factors	Bootstrap results
1	7	00:00 29 Dec. 2019-22:00 4 Jan. 2020	Vehicle exhaust factor mixed with industrial emission	Failed
2	10	00:00 29 Dec. 2019-22:00 7 Jan. 2020	Vehicle exhaust factor mixed with biomass burning	Failed
3	14	00:00 29 Dec. 2019-22:00 11 Jan. 2020	Vehicle exhaust factor mixed with biomass burning	Passed, mapping of secondary sulfate and SOA_I factors < 90%
4	18	00:00 29 Dec. 2019-16:00 15 Jan. 2020	Less factor profile mixing	Passed, mapping of all factors > 90%

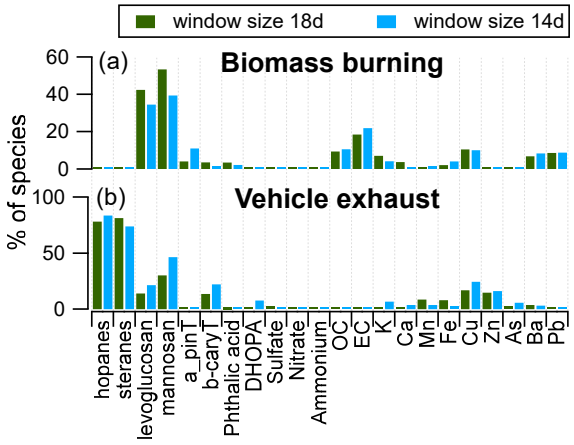


Figure S4. Comparison of factor profiles of biomass burning (a) and vehicle exhaust (b) from the base run PMF result for window length of 18 d and 14 d.

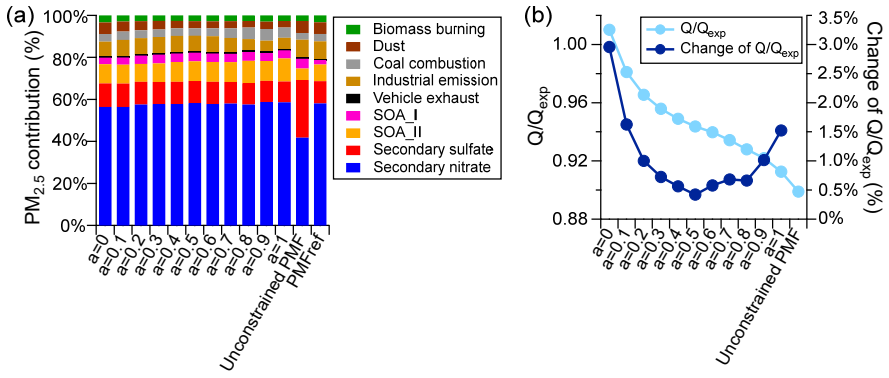


Figure S5. (a) Variation of the $PM_{2.5}$ contribution from individual factors for the *a*-value constrained runs (*a*=0-1, step 0.1), unconstrained PMF run and the reference result. (b) Change of the Q/Q_{exp} values for the *a*-value constrained runs (*a*=0-1, step 0.1) and unconstrained PMF run.

Text S3. Sensitivity test of the reference profiles for the *a*-value approach

To test the sensitivity of the reference profiles, a list of new reference source profiles was synthesized based on the original one using the following equation:

$$f_{i,new} = f_{i,old}(1 + r \times b),\ r = \text{random number of 1 or -1};\ b = 0.1, 0.3, 0.5, 0.7 \quad (1)$$

where $f_{i,old}$ is the original reference profile, $f_{i,new}$ is the new reference profile, r is a random number of 1 or -1, and b represents the relative difference (RD) of the new reference profile from the original one. Four different b values of 0.1, 0.3, 0.5, and 0.7 were selected, representing the RD of 10%, 30%, 50% and 70% from the original reference profiles.

The short-term PMF run was conducted with the four new synthesized reference profiles using the same *a*-values obtained in Sec. 3.2 in the *a*-value approach in SoFi. The obtained source contributions to $PM_{2.5}$ from individual primary source factors from the four PMF runs were compared with that from original PMF result in Sec. 3.2 and the PMF_{ref} result (Figure S6). It can be seen that with the increasing deviation from the reference source profiles, the short-term PMF results showed larger RD of source contributions to $PM_{2.5}$. The PMF run with the original reference profiles showed RD of -14% to 5% for all primary factors. With the new deviated reference profiles, the RD clearly increased. For example, with reference profiles of RD of 30%, the RD of $PM_{2.5}$ contribution for coal combustion, biomass burning and dust increased to 20% and -19%; with RD of reference profiles increased to 70%, the RD for industrial emission and vehicle exhaust increased to -36% and -54%, respectively. The results suggested the effectiveness of using the source profiles from PMF_{ref} to do the *a*-value constraints,

which is closer to the true source profiles at this site. Larger deviation from the actual source profiles will lead to larger bias on the apportioned source apportionment results.

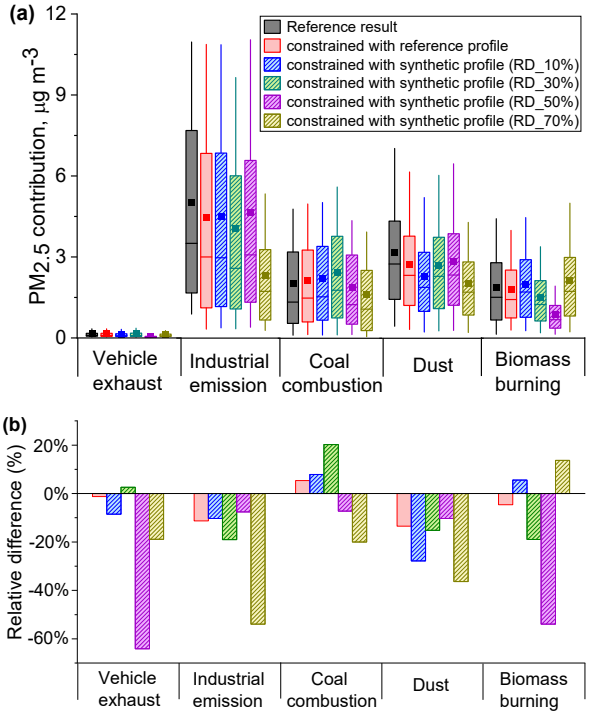


Figure S6. Comparison of the $PM_{2.5}$ source contributions from individual primary source factors resolved by the short-term PMF runs constrained by the a-value approach in SoFi using different synthesized reference profiles: (a) average $PM_{2.5}$ source contributions and (b) the relative difference (RD) of $PM_{2.5}$ source contributions compared with the reference result obtained in PMF_{ref}.

Table S3. Summary of the reference run (PMF_{ref}), short-term PMF testing run and rolling PMF runs (PMF_{roll}) using the remaining data performed in this study.

	Run No.	Starting Time	Ending Time	Input sample size	Factor numbers
PMF_{ref} run		0:00 29 Dec. 2019	22:00 9 Feb. 2020	416	10
Short-term PMF run		0:00 29 Dec. 2019	16:00 15 Jan. 2020	190	9
PMF_{roll} with remaining dataset	1	0:00 30 Dec. 2019	22:00 22 Jan. 2020	195	9
	2	0:00 31 Dec. 2019	22:00 23 Jan. 2020	193	
	3*	0:00 1 Jan. 2020	22:00 24 Jan. 2020	193	10
	4	0:00 2 Jan. 2020	20:00 25 Jan. 2020	196	
	5	0:00 3 Jan. 2020	22:00 26 Jan. 2020	196	
	6	0:00 4 Jan. 2020	20:00 27 Jan. 2020	196	
	7	0:00 5 Jan. 2020	22:00 28 Jan. 2020	196	
	8	0:00 6 Jan. 2020	22:00 29 Jan. 2020	197	
	9	16:00 7 Jan. 2020	20:00 30 Jan. 2020	196	
	10	0:00 8 Jan. 2020	22:00 31 Jan. 2020	204	
	11	0:00 9 Jan. 2020	20:00 1 Feb. 2020	206	
	12	0:00 10 Jan. 2020	22:00 2 Feb. 2020	209	
	13	0:00 11 Jan. 2020	20:00 3 Feb. 2020	208	
	14	0:00 12 Jan. 2020	22:00 4 Feb. 2020	209	
	15	0:00 13 Jan. 2020	22:00 5 Feb. 2020	209	
	16	0:00 14 Jan. 2020	20:00 6 Feb. 2020	208	
	17	0:00 15 Jan. 2020	22:00 7 Feb. 2020	208	
	18	0:00 22 Jan. 2020	22:00 8 Feb. 2020	208	
	19	0:00 23 Jan. 2020	22:00 9 Feb. 2020	208	

*Run No. 3 was excluded due to very limited firework-influence data point in the input samples, leading to outlier results compared with other rolling PMF runs.

95

Text S4. Rolling PMF runs without the *a*-value approach

The rolling PMF runs without any constraints were also performed. A total of 19 runs were performed. The resolved factor contributions to $PM_{2.5}$ are shown in Figure S7. The scatter plot of the average factor contribution from the rolling PMF runs with the PMF_{ref} are shown in Figure S8.

The rolling PMF without *a*-value constraints showed large run-to-run variations, especially for vehicle exhaust and coal combustion. Besides, the secondary factors also showed larger variations when compared with the rolling PMF with the *a*-value constraints, although the secondary factors are not constrained in the *a*-value approach. The obtained average contributions from the rolling PMF without constraints showed poorer correlation with the PMF_{ref}, and the slope more deviated from unit. For example, vehicle exhaust and coal combustion were largely overestimated in the rolling PMF when compared with the PMF_{ref}, with slope of 3.4 and 1.8 respectively (Figure S8). The poor reproducibility of the result suggested that the short-term PMF runs, due to decreased data variability, showed high uncertainties when performed without any constraints. This highlighted the necessity of the source profile constraints to obtain robust source apportionment results when performing the PMF analysis across the short-term measurement data.

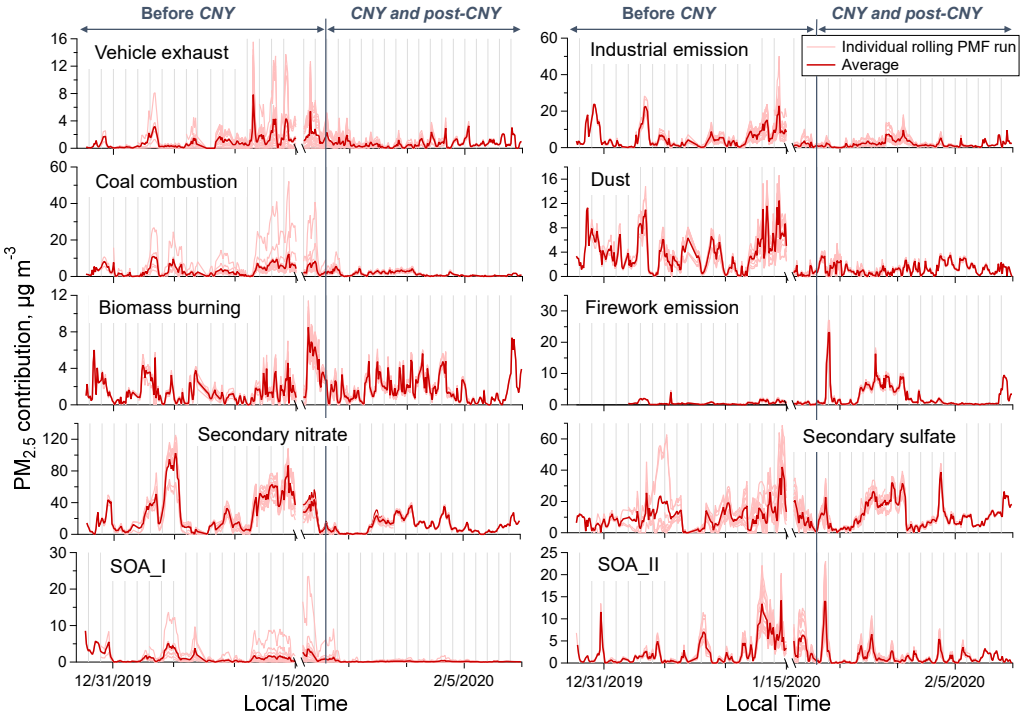
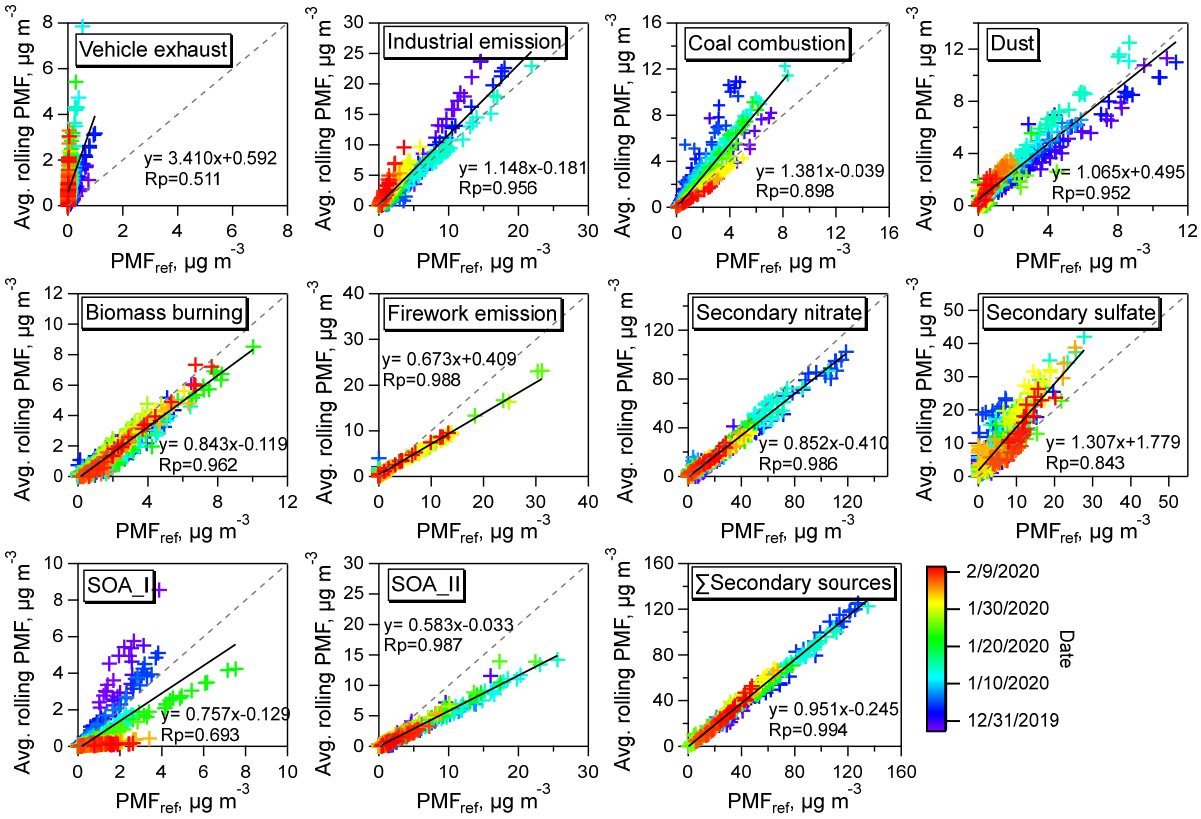


Figure S7. Time series of individual factor contributions to $PM_{2.5}$ for individual rolling PMF runs without a-value constraints and the average source contributions. The individual rolling PMF run is shown in light red line and the average rolling PMF result is shown in dark red line.



115 **Figure S8. Comparison between the average source contribution from the rolling PMF runs without *a*-value constraints and the reference result from PMF_{ref} for individual source factors and the sum of the four secondary sources.**