

**Supplementary material for Characterization of seasonal difference of HULIS-C sources from water soluble PM<sub>2.5</sub> in Seoul, Korea: probing secondary processes by Lee et al.**

Table S1. List of organic substances tested for retention by the SPE cartridge.

| Tested Substances | MW  | Recovery as HULIS-C (%) |
|-------------------|-----|-------------------------|
| Acetic acid       | 60  | 0±20                    |
| Oxalic acid       | 90  | 0±2                     |
| Succinic acid     | 118 | 0±4                     |
| Malic acid        | 134 | 11±4                    |
| Suberic acid      | 174 | 104±6                   |
| Azelaic acid      | 188 | 71±3                    |
| Phthalic acid     | 166 | 64±0                    |
| Levogluconan      | 162 | 3±4                     |
| Xylose            | 180 | 2±1                     |
| Sucrose           | 342 | 0±1                     |
| NAFA <sup>a</sup> |     | 92±10                   |
| SRFA <sup>b</sup> |     | 97±8                    |

<sup>a</sup>Nordic aquatic fulvic acid

<sup>b</sup>Suwannee river fulvic acid

Table S2. Contribution of target compounds-C in HULIS-C of collected samples

| Target compound | $\frac{\text{Target compound} - \text{C}}{\text{HULIS} - \text{C}}$ (%) |
|-----------------|-------------------------------------------------------------------------|
| Suberic acid    | 0.2±0.2                                                                 |
| Azelaic acid    | 0.5±0.3                                                                 |
| Phthalic acid   | 0.5±0.3                                                                 |

***PMF equations and factor number determination***

In Eq. (1), mass concentration  $x_{ij}$  of species  $j$  in sample  $i$  is decomposed by matrix multiplication to obtain the source contribution  $g_{ik}$  and the source profile  $f_{kj}$ . During the

modeling iteration, equilibrium will be achieved when the  $Q$  function reaches a minimum value as calculated in Eq. (2).

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

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

The input uncertainty  $u_{ij}$  is calculated by Eq. (3), where  $EF$  is regarded as error fraction and MDL is method detection limit. For data below their MDLs, the concentration was set to be  $0.5 \times \text{MDL}$  and the corresponding uncertainty was set as  $\frac{5}{6} \times \text{MDL}$  (Reff et al., 2007).

$$u_{ij} = x_{ij} \times EF + \frac{\text{MDL}}{3} \quad (3)$$

For major ions and elemental species, a constant  $EF$  was set to be 0.1. However, different OC components broken down by different separation procedures have their specific  $EF$  values calculated using Eq. (4). The advantage of using variable  $EF$  is to compensate the underestimation of high polluted samples by PMF model and the elaboration of which was established and illustrated elsewhere (Wang et al., 2017).

$$EF = EF_0 - \alpha \times \left( \frac{x_{ij}}{\text{MAX}[x_i]} \right) \quad (4)$$

$EF_0$  was set to 0.1 for WSOC, WISOC. Different  $EF_0$  was assigned to different OC components based on their analytical precision: 0.1 for levoglucosan; 0.3 for non-

polar organic compounds and 0.4 for polar organic compounds including SOA tracers (Wang et al., 2017). According to our past knowledge, the determination of  $\alpha$  value relied on the degree of variation between the “outliers” and the majority of the data points. We adopted  $\alpha = 0.2$  for non-polar organics and 0.3 for polar organics. There is no need to set the  $\alpha$  constraint on levoglucosan because it is infrequent to observe episodically high levoglucosan concentrations in the dataset.

Two methods were used to determine the optimum factor numbers: Q/Q<sub>exp</sub> ratio and the maximum individual column mean (IM) (Lee et al., 1999). First,  $Q_{exp} \approx nm - p(n + m)$  is used to determine the degree of freedom of the model solution by sample amount (n), species input (m) and numbers of factors (p) (Norris, 2014). Second, IM is calculated using Eq. (5) by considering residual ( $e_{ij}$ ) and input uncertainty ( $u_{ij}$ ). The determination is based on the principle that once the number of factors reaches a critical value, a sharp drop would be observed on the IM value. Thus, the principle of determining the minimum number of factors are based on the percentage change of the parameters (Q/Q<sub>exp</sub> or IM) reaching the maximum.

$$IM = \max_{f=1,2,\dots,m} \left( \frac{1}{n} \sum_{i=1}^n \frac{e_{ij}}{u_{ij}} \right) \quad (5)$$

The results in factor determination are shown in Fig. S1. From panel (a) first positive change in Q/Q<sub>exp</sub> drop was noticed from 5 to 6 factors. A similar drop trend could also be noticed from the change of IM in panel (b). Both results indicate that 6 is the minimum number of factors that should be considered. With the 6-factor solution, biomass burning, and other incomplete combustion are not fully separated

ed. When the number of factors increase from 6 to 7,  $Q/Q_{exp}$  and IM continue to decrease down to 0.7 and 0.15, respectively. Percentage in changes are around 3% for  $Q/Q_{exp}$  and over 20% for IM, indicating factor number of 7 could also give a mathematically stable profile. Thus, the combined consideration of mathematical criteria and source interpretability led us to select 7-factor solution as the optimal solution.

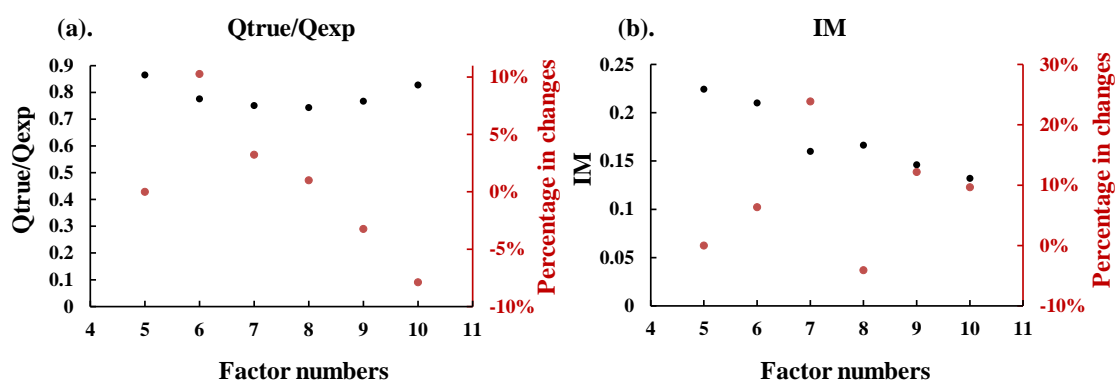


Fig. S1. Factor determination of PMF result: (a)  $Q_{true}/Q_{exp}$  (b) IM.

### ***Stability test of the PMF***

Both bootstrap (BS) and displacement (DISP) are used to evaluate the stability of the 7-factor solution. In BS, random sample selection from original dataset is carried out and source apportionment of this arbitrary sample set is performed. Number of times in BS is set to 100 with the default  $R > 0.6$ . The differences between the base run source profile  $f_{kj}$  and BS profiles archived during the 100 times iteration are interpreted as the reproducibility. For DISP, a series of preset profiles replaced the original source profile  $f_{kj}$ . In DISP, the F factor elements are displaced one by one. Each displacement is extended until the object function Q increases by a maximum allowed change in Q ( $dQ_{max}$ ). Each such extended displacement is interpreted as the upper and lower interval estimates for

each displaced factor element, representing the uncertainty of individual F factor (source profile).

In BS mapping results, all factors except for the sea salt factor could fully reproduce species' contribution in base run model while the sea salt in base run shows good similarity with BS by over 90 out of 100 times. Overall, this result indicates good reproducibility. In DISP, there was no factor swapping or un-defined factors indicating the stability of the result.

### *Time series of contribution of the factors to HULIS-C*

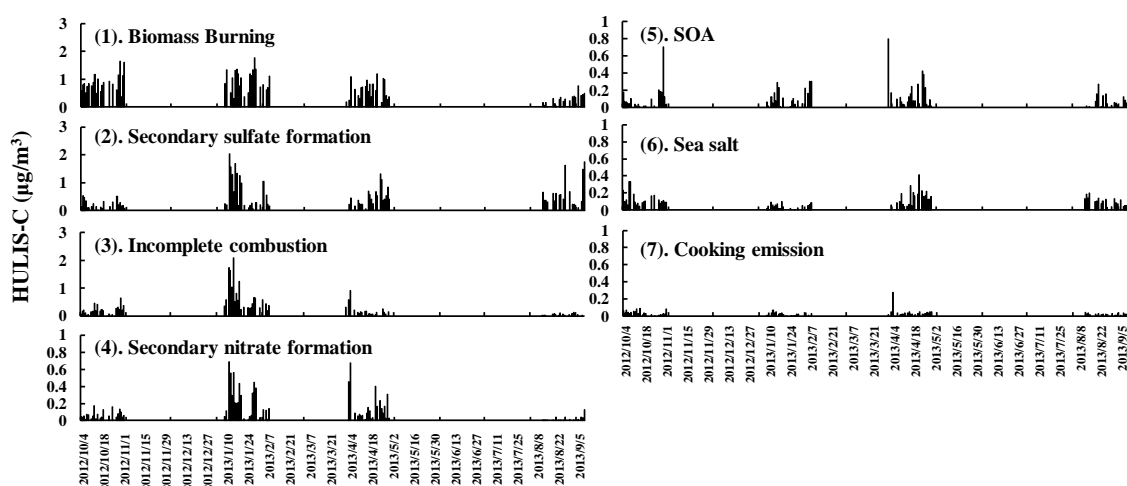


Fig. S2. Time series of contribution of the factors to HULIS-C.