Aerosol and Air Quality Research, 16: 2512–2522, 2016

ISSN: 1680-8584 print / 2071-1409 online
doi: 10.4209/aaqr.2016.05.0201

Accuracy of Advanced and Traditional Three-Way Factor Analysis Models for Determining Source Contributions to Particulate Matter

Ying-Ze Tian[1], Yan-Qi Huang-Fu[1], Guo-Liang Shi[1], Xu-Rong Shi[1], Bo Han[2], Yin-Chang Feng[1*]

[1] *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin, 300071, China*
[2] *Tianjin Key Laboratory for Air Traffic Operation Planning and Safety Technology, College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China*

ABSTRACT

Although three-way factor analysis models can take more information into account, their degree of accuracy must be investigated further. We simulated numerous synthetic datasets to evaluate traditional (PMF3) and advanced three-way models (AAB or ABB). On one hand, scenarios whereby sources share the same profiles but different emission patterns were constructed and introduced into PMF3 and advanced AAB models that can estimate the same source profile matrixes B but different emission pattern matrixes A. On the other hand, datasets with the same emission patterns but different profiles were set to simulate the variability of source profiles and were used to evaluate an advanced ABB model that can estimate the same emission pattern matrix A but different source profile matrixes B. The AAEs of PMF3 under two different conditions ranged from 2.95% to 90.22% and from 2.98% to 90.11%, respectively, while the results of the advanced three-way models were 2.88%–27.51% and 2.89%–29.89%, respectively. We observed that the PMF3 performed well when all sources showed strong emission patterns and source profiles were similar. The application of advanced AAB model was stable under various emission patterns; and ABB model was stable under different variability of source profiles. At the same time, several ambient datasets were estimated by the advanced three-way models. Our findings suggest that the performance of advanced three-way models make full use of spatial or size distribution information to enhance capacities to identify source categories.

Keywords: Three-way factor analysis model; PMF3; Receptor model; Simulation; Source apportionment; ME2.

INTRODUCTION

Atmospheric particulate matter (PM) is considered to be a major pollutant worldwide (Shen *et al.*, 2014; Srithawirat and Brimblecombe, 2015; Meena *et al.*, 2016). Numerous studies have described adverse effects of PM on public health, visibility, climate change and environmental quality (Zheng *et al.*, 2007). To improve urban quality levels, environmental management departments typically pay more attention to PM emitted sources, which have become a topic of scientific debate worldwide (Huang *et al.*, 2014; Kchih *et al.*, 2015; Malaguti *et al.*, 2015). However, individual contributions of each source category to PM cannot be measured directly (Shen *et al.*, 2014). Therefore, various source apportionment methods have been developed and applied (Zheng *et al.*, 2011; Shen *et al.*, 2014; Taiwo, 2016; Vu *et al.*, 2016).

Receptor-based source apportionment techniques have been developed to solve source estimation problems and to determine source profiles (B matrix) and contributions (A matrix) based on measurements collected from receptor sites (see Eq. (S1) in Supplementary Materials). Most receptor models address two-way datasets that are generally characterized by temporal variability and chemical composition (Eq. (S1)). However, with growing requirements to include more information in source apportionment methods, three-way blocks have become more popular. For example, with the growing number of monitoring sites for air quality monitoring networks, multi-site datasets that assess compositional, temporal and spatial variability have been produced (Escrig *et al.*, 2009; Beuck *et al.*, 2011; Mooibroek *et al.*, 2011; Larsen *et al.*, 2012; Minguillón *et al.*, 2012); another example pertains to multi-size datasets that include size, compositional and temporal information when size-segregated PM samples are synchronously collected (Chen *et al.*, 2015; Gaita *et al.*, 2016). To better explore inter-information in these three-way datasets, three-way factor analysis models come into play (Paatero, 1999; Faber

* Corresponding author.
Tel.: +862223503397; Fax: +862223503397
E-mail address: fengyc@nankai.edu.cn

et al., 2003; Abdollahi *et al.*, 2010). PMF3, one of the most commonly used three-way receptor models and a standard bilinear PMF was used to resolve the the source categories and contributions to PM (Mantas *et al.*, 2014), typically employs PARAFAC (Parallel Factor Analysis) strategies (Paatero, 1997).

As discussed in the related researches studies and in our previous works (Tian *et al.*, 2014), the PARAFAC approach could only extract a single source profiles matrix (matrix B) and the same source emission pattern matrix (matrix A) for different sites or for different sizes of PM, which might be inconsistent with the actual situation conditions as discussed above. While multi-size or multi-site three-way datasets with temporal variability, chemical species and size distributions or spatial variability levels as dimensions may provide more information for source apportionment (Pu *et al.*, 2015; Parthasarathy, *et al.*, 2016). Diverse three-way datasets generally present different characteristics. For multi-size three-way datasets, source compositions are dependent on particle sizes (Peré-Trepat *et al.*, 2007), meaning that source profiles may vary slightly for different sizes, though the time series of daily contributions for one source category should be highly consistent (Shi *et al.*, 2015). For multi-site three-way datasets, when sites fall within a certain geographical range (e.g., a city), the fundamental assumption is that source profiles for receptors are identical; however, the time series of daily contributions for one source category at different sites can be altered. When characteristics of various three-way datasets can be used fully, the accuracy of source apportionment results may be improved.

To take better advantage of three-way dataset information and to reflect real conditions, an advanced three-way factor analysis method based on ME2 (multilinear engine 2) was proposed and applied (Shi *et al.*, 2015; Peré-Trepat *et al.*, 2007). This advanced model can take size-composition variations into account, and namely it can obtain one emission pattern matrix (matrix A) but different source profile matrixes (matrix B) for multi-size three-way datasets. Thus, it is referred to as the ABB three-way factor analysis model. Another similar model can generate the same source profile matrix but different emission pattern matrices for multi-site three-way datasets, and it is thus referred to as the AAB three-way factor analysis model. The ABB model is suitable for multi-size datasets and the AAB model is appropriate for multi-site datasets to address real conditions and to take full advantage of three-way dataset information.

The accuracy of three-way factor analysis models is central to source apportionment and to further health-related studies. However, advanced three-way factor analysis models have not yet been systematically explored, and assessments of their accuracy are scarce. In fact, it may be nearly impossible to fully determine the "true" contributions of sources in real air ambient conditions. It is thus difficult to use real ambient datasets to explore the performance of models. Therefore, artificial dataset and simulation studies have typically been conducted by researchers (Shi *et al.*, 2011; Tian *et al.*, 2014).

In this work, extensive synthetic datasets for several scenarios involving two aspects (varying emission pattern and source profile variability levels) are developed and applied using advanced AAB and ABB three-way factor analysis models and the traditional PMF3 to evaluate the accuracy of three-way factory analysis models. The main purpose of this paper is to study the application potential of advanced three way models. We, in turn, propose protocols of model selection and application for use in environmental management departments.

METHODOLOGY

Advanced Three-Way Factor Analysis Models

The traditional PMF3 is based on the PARAFAC model, which can be expressed in the form of a matrix as follows (Hopke *et al.*, 1998):

$$X = ABC + E \tag{1}$$

where X is the three-way matrix of ambient concentrations ($\mu g\ m^{-3}$) for different types of PM (e.g., sizes or sites); A relates to the source contribution matrix, and namely the emission pattern matrix; B relates to the source profile matrix; and C is the fractional factor for different PM types. According to the equation, for PM of different types, the same source profile matrix (B) and the same emission pattern matrix (A) are extracted through traditional PARAFAC.

The main principle of advanced AAB three-way factor analysis models can be described as follows (Peré-Trepat *et al.*, 2007):

$$X_{ijk} = \sum_{p=1}^{p} a_{ipk} b_{jp} + e_{ijk} \tag{2}$$

where a_{ipk} is the estimated contribution of the pth source to the ith sample for the kth PM; b_{jp} is the jth species mass fraction in the pth source profile. The AAB model can estimate the same source profile matrixes B (the matrix of b_{jp}) but different emission pattern matrixes A (the matrix of a_{ipk}) for PMs of different types. Similarly, the ABB model can be described as follows:

$$X_{ijk} = \sum_{p=1}^{p} a_{ip} b_{jpk} + e_{ijk} \tag{3}$$

where x_{ijk} is the concentration value of the jth species of the ith sample for the kth PM; a_{ip} is the estimated contribution of the pth source to the ith sample; b_{jpk} is the jth species mass fraction of the pth source profile for the kth PM; e_{ijk} is the residual. From the ABB model, the same emission pattern matrix A (the matrix of a_{ip}) but different source profile matrixes B (the matrix of b_{jpk}) can be obtained for PMs of different types.

The advanced three-way factor analysis model was derived from multilinear engine 2 (ME2), which is a more flexible computer program. The ME2 was first developed by Paatero (1999). Its purpose is similar to that of the general factorization method in that it has typically been solved via

positive matrix factorization (PMF). Detailed descriptions of the ME2 are included in related literature (Paratero, 1999; Ramadan *et al.*, 2003; Amato and Hopke, 2012) and in our previous works (Tian *et al.*, 2014; Shi *et al.*, 2015). The advanced AAB and ABB models evolved from a logical two-way model, but they are organized as three-way arrays and input data should be three-way characterized by composition, time and PM type (Peré-Trepat *et al.*, 2007).

Synthetic Dataset Development and Data Analysis

To evaluate the accuracy of the advanced three-way factor analysis models, numerous synthetic datasets for different scenarios were developed. For the purposes of the present work, two types of synthetic datasets were developed: (1) datasets with different site-to-site correlations of emission patterns used to evaluate the performance of the AAB model and (2) datasets with differing degrees of source profile variability used to evaluate the performance of the ABB model.

The method used to construct the synthetic datasets is based on the published literatures (Habre *et al.*, 2011; Tian *et al.*, 2014; Shi *et al.*, 2015). Two PM types denoted as site I and site II (or size I and size II) were examined. For each PM type, the corresponding receptor concentration matrix $X_{m\times n}$ was developed according to:

$$X_{m\times n} = G_{m\times p} F_{m\times p} + E_{m\times n} \quad (4)$$

where $G_{m\times p}$ is the source contribution matrix; m is the number of daily samples; p is the number of sources; $F_{p\times n}$ is the source profile matrix; n is the number of species; and $E_{m\times n}$ is the random perturbation method used to simulate analytical and measurement errors.

Due to concerns related to feasibility and in accordance with previous related studies (Javitz *et al.*, 1988; Habre *et al.*, 2011), the collinearity of source profiles and contributions was not taken into account. For $F_{p\times n}$, the three source categories used in the present work are vehicular exhaust (denoted as VE), secondary sulfate (SS) and crustal dust (CD). Actual source profiles of vehicular exhaust and crustal dust measured in our previous work (Shi *et al.*, 2015) were applied. The profiles of secondary sulfate were largely expressed as pure $(NH_4)_2SO_4$. The source profiles used to simulate three-way datasets (referred to as actual source profiles) are shown in Table S1 Thus, p = 3 and n = 24. For the construction of $G_{m\times p}$, the Monte Carlo method was employed to randomly determine the contributions of each source category from a normal distribution with the set mean and standard deviations (Habre *et al.*, 2011). Three hundred daily samples were simulated (m = 300). The coefficient of variation (CV, which was defined as the standard deviation of the species concentration divided by its mean) was used to perturb receptor concentrations. Following from Javitz *et al.* (1988), a CV of 10% was used in all of the concentration matrices, and random uncertainties were generated from a log-normal distribution. The ambient concentration matrixes of two types of PMs were combined to develop each three-way input as a 300 × 24 × 2 array denoting temporal, compositional and type variability.

To systematically evaluate the performance of the AAB model and to examine the impacts of emission patterns on the advanced AAB and traditional PMF3 models, datasets with different site-to-site correlations of emission patterns were constructed based on four scenarios as follows (Table S2).

SCE 1: for each source category, site-to-site correlations of daily contributions between two types were set to zero to simulate emission patterns of different PM types that are completely inconsistent;

SCE 2: site-to-site correlations of daily contributions from secondary sulfate were set to 0.6 and the others were set to 0 to simulate one source exhibited moderate correlation between two types;

SCE 3: site-to-site correlations for three source categories were set to 0.6 to simulate three sources exhibiting moderate site-to-site correlations between two types;

SCE 4: site-to-site correlations for three source categories were set to 0.9 to simulate three sources exhibiting strong site-to-site correlations.

The site-to-site correlation of daily contributions between two sites (R) was denoted as:

$$R_j = CORREL(g_{ij}^{I},\ g_{ij}^{II}) \quad (5)$$

where CORREL denotes Pearson's correlation coefficients; $g_{ij}^{\text{Site I}}$ is the simulated contribution of the j^{th} source to the i^{th} sample for type Site I; and $g_{ij}^{\text{Site II}}$ is the simulated contribution of the j^{th} source to the i^{th} sample for type Site II.

To systematically evaluate the performance of the ABB model and the influence of source profile variability on advanced and traditional three-way factor analysis models, datasets with varying degrees of source profile variability were developed. CVs = 10%, 30% and 50% were used to simulate three degrees of source profile variability. Due to the presence of uniform emission patterns, site-to-site correlations for three source categories were set to 0.9.

To guarantee stability, representativeness and feasibility, ten synthetic datasets were generated for each scenario, creating a total of 70 synthetic datasets.

The 40 synthetic datasets with different site-to-site correlations of emission patterns were introduced into the advanced AAB and traditional PMF3 models. The 30 synthetic datasets presenting different degrees of source profile variability were computed from the ABB and PMF3 models.

Indicators used to estimate the performance of the three-way models included Pearson correlation coefficients (COR) and the average absolute error (AAE) between actual (simulated) and modeled daily contributions. The two indicators were defined as follows (Javitz *et al.*, 1988):

$$COR_j = CORREL(E_{ij}, T_{ij}) \quad (6)$$

$$AAE_j(\%) = 100 \times \frac{1}{n}\sum_{i=1}^{n}\frac{\left|E_{ij} - T_{ij}\right|}{T_{ij}} \quad (7)$$

where E_{ij} is the estimated contribution from the jth source

category to the ith sample for each PM type; T_{ij} is the simulated contribution (namely the true contribution) corresponding to E_{ij}; and n is the number of samples and is equal to 300 in this work. Some ambient datasets were estimated from the advanced 3-way model in our prior works to study multi-dimensional source apportionment patterns.

RESULTS AND DISCUSSION

Evaluation of the AAB Three-Way Model for Different Site-to-Site Emission Pattern Correlations

To investigate the performance of the advanced AAB model for different site-to-site correlations of source contributions, 40 synthetic datasets based on 4 scenarios (with different site-to-site correlations of source contributions) were introduced into the advanced AAB and traditional PMF3 models, and the results were then compared.

The mean estimated contributions of the three source categories and the TC (total concentration of PM: the sum of contributions from three source categories) estimated from the advanced AAB and traditional PMF3 models are shown in Fig. 1. As discussed in our prior work (Shi *et al.*, 2015), PMF3 results are rather divergent and unstable with high standard deviations. Very PMF3 poor performance was observed, and especially for SCE 1 and 2. However, the performance of the advanced AAB model for different type-to-type correlations of source contributions was determined to be satisfactory. For SCE 1, the estimated contributions of VE, SS, CD and TC by AAB model were, respectively, 40.65 ± 4.10 μg m^{-3}, 35.79 ± 2.81 μg m^{-3}, 43.12 ± 4.40 μg m^{-3}, 119.55 ± 1.00 μg m^{-3} for Site I, respectively, 27.82 ± 3.47 μg m^{-3}, 35.35 ± 3.25 μg m^{-3}, 26.93 ± 3.06 μg m^{-3}, 90.11 ± 0.60 μg m^{-3} for Site II; while the results of PMF3 were 36.13 ± 10.21 μg m^{-3}, 62.58 ± 9.40 μg m^{-3}, 17.68 ± 15.66 μg m^{-3}, 116.39 ± 5.21.00 μg m^{-3} for Site I, 16.16 ± 24.84 μg m^{-3}, 50.73 ± 29.65 μg m^{-3}, 21.04 ± 14.09 μg m^{-3}, 87.93 ± 4.15 μg m^{-3} for Site II respectively. And the same conditions appeared in the other scenarios. We found the mean estimated contributions of the advanced AAB model to be very consistent with the real data, and the standard deviations were very small, denoting stable and accurate advanced AAB model results for different site-to-site correlations of source contributions.

To further evaluate whether the advanced AAB model and traditional PMF3 can function well for different site-to-site correlations of daily source contributions and to investigate the influence of site-to-site correlations on them, CORs and AAEs between estimated and predicted contributions were calculated. Fig. 2 presents the CORs of each source category and TC values for four scenarios for sites SI and Fi. S1 presents the conditions for site SII. It is evident that for all source categories and TC values, CORs were clearly unstable. For most fittings, and especially those in SCE1 and 2, estimated contributions of PMF3

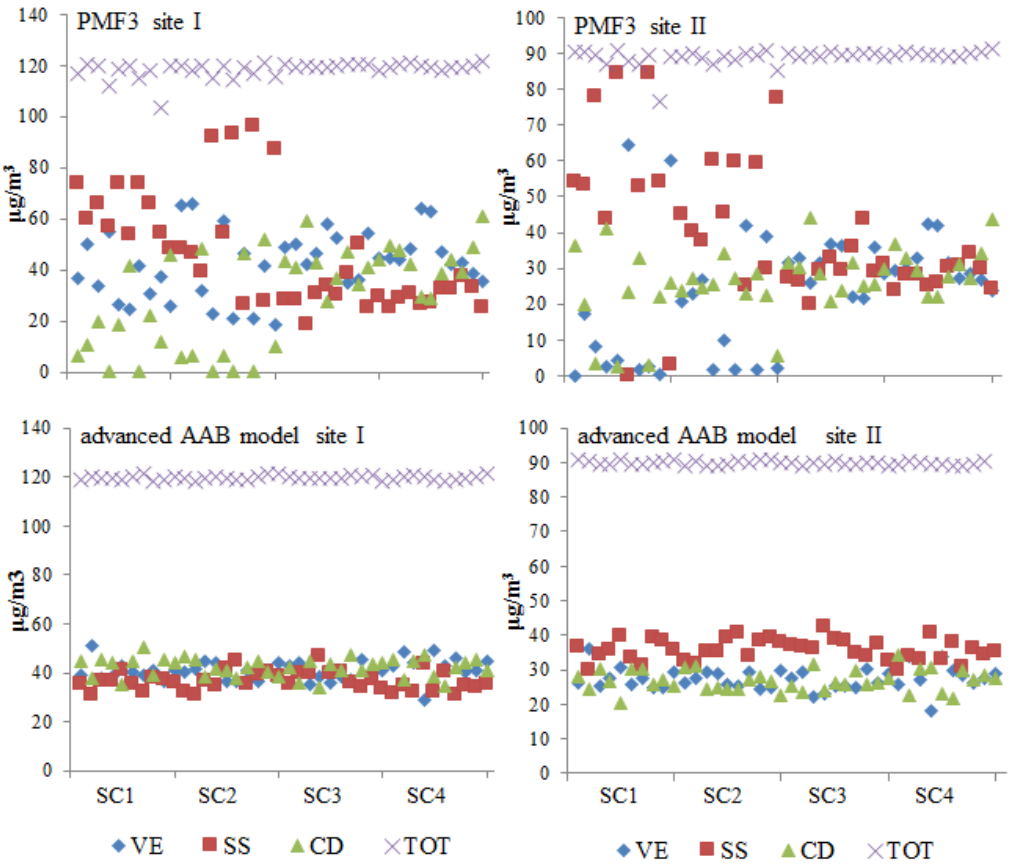

Fig. 1. The source contributions estimated by PMF3 and advanced AAB model in four scenarios with different site-to-site correlations of emission patterns.

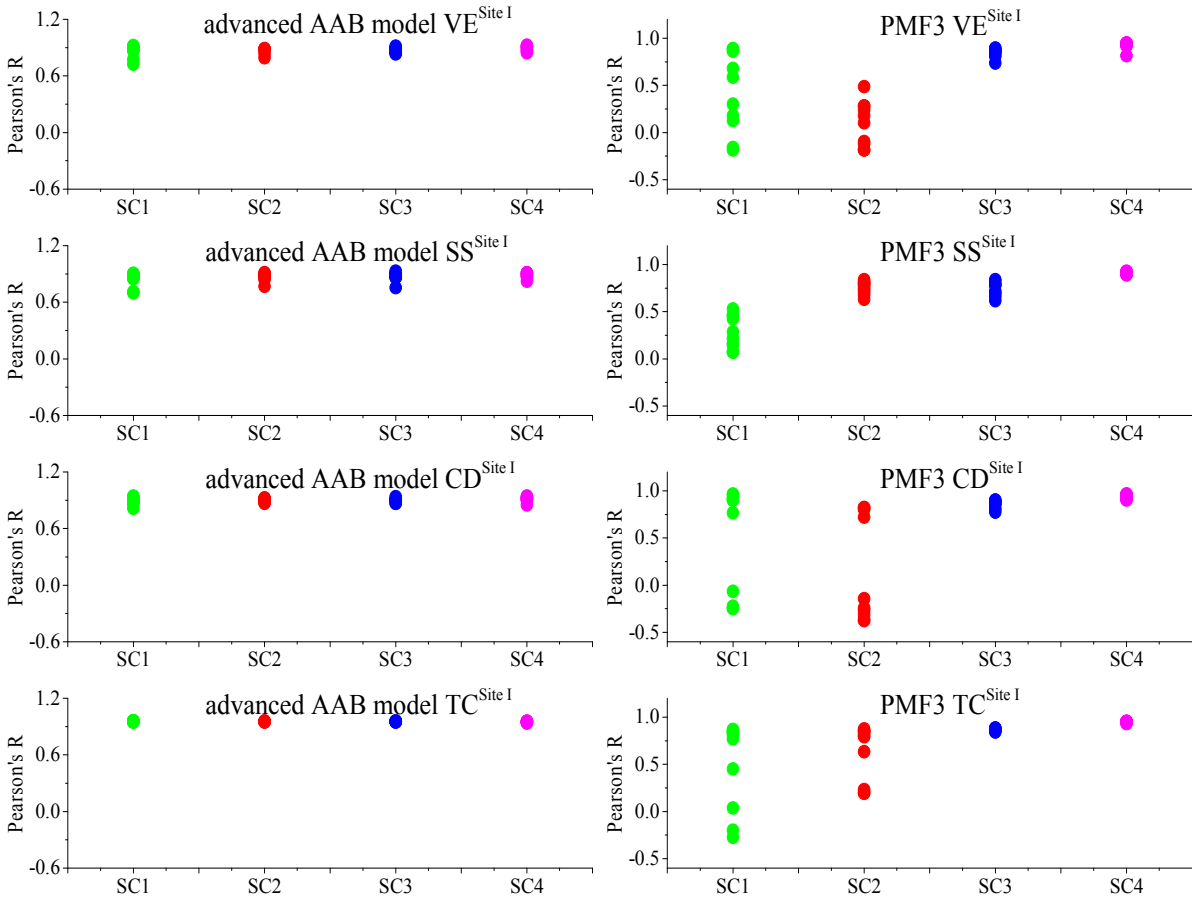

Fig. 2. The CORs of simulated and predicted factor contributions by advanced AAB model and PMF3 respectively for Site I.

diverged from true values while in SCE 3, the PMF3 performed better, and the results of PMF3 in SCE4 are the best. Thus, we found that the solutions of the traditional PMF3 are suitable only for high site-to-site correlations of source contributions. For the advanced AAB model, stable results with high CORs were found for all source categories at both sites: CORs ranged from 0.68–0.92 for vehicular exhaust, from 0.70–0.95 for secondary sulfate, and from 0.68–0.94 for crustal dust in SCE 1; from 0.79–0.89 for vehicular exhaust, from 0.77–0.94 for secondary sulfate, and from 0.77–0.92 for crustal dust in SCE 2; from 0.83–0.92 for vehicular exhaust, from 0.75–0.95 for secondary sulfate, and from 0.81–0.94 for crustal dust in SCE 3; and from 0.82–0.92 for vehicular exhaust, from 0.82–0.94 for secondary sulfate, and from 0.79–0.94 for crustal dust in SCE 4. The CORs of TC estimated from the advanced AAB model exceeded 0.94 in all cases, which means that the results of the model were stable even though the site-to-site correlations of fluctuated obviously.

Furthermore, the AAEs shown in Figs. 3 and S2 are in good agreement with the CORs. In SCE 1 and 2, the AAEs of the PMF3 presented larger ranges with high mean values, suggesting that PMF3 performs poorly when site-to-site correlations of source contributions are low. While site-to-site correlations of all source contributions were found to be higher (In SCE 3 and 4), the AAEs of PMF3 were relatively lower. All of these results show that the site-to-site correlations of source contributions should strongly influence the performance of PMF3 and could perform better when all sources show strong site-to-site associations. However, for the advanced AAB model, AAEs were stable. In SCE 1, the AAEs of three source categories ranged from 12.30%–39.49% with the average value of 21.89% for SI and from 12.12%–27.55% with an average value of 19.70% for SII. In SCE 2, the AAEs of three source categories ranged from 11.59%–34.65% with an average value of 19.36% for SI and from 12.28%–27.69% with an average value of 17.99% for SII. AAEs ranged from 10.48%–35.88% with an average of 19.95% for SI and from 10.29%–29.72% with an average of 18.81% for SII in SCE 3. In SCE 4, AAEs ranged from 9.97%–36.93% with an average value of 20.81% for SI and from 10.11%–40.65% with an average of 19.82% for SII. The AAEs of TC fell below 5% for all source categories at both sites.

These results can be explained by the different principles underlying PMF3 and the advanced AAB model. According to Eq. (2), the AAB model takes differences in source contributions fully into account in the surveyed region, while PMF3 extracts only one Matrix A (related to source contributions) for all of the sites, i.e., for each source, site-to-site correlations of contributions should be units. Therefore, PMF3 can perform well only when all sources show strong site-to-site correlations among all

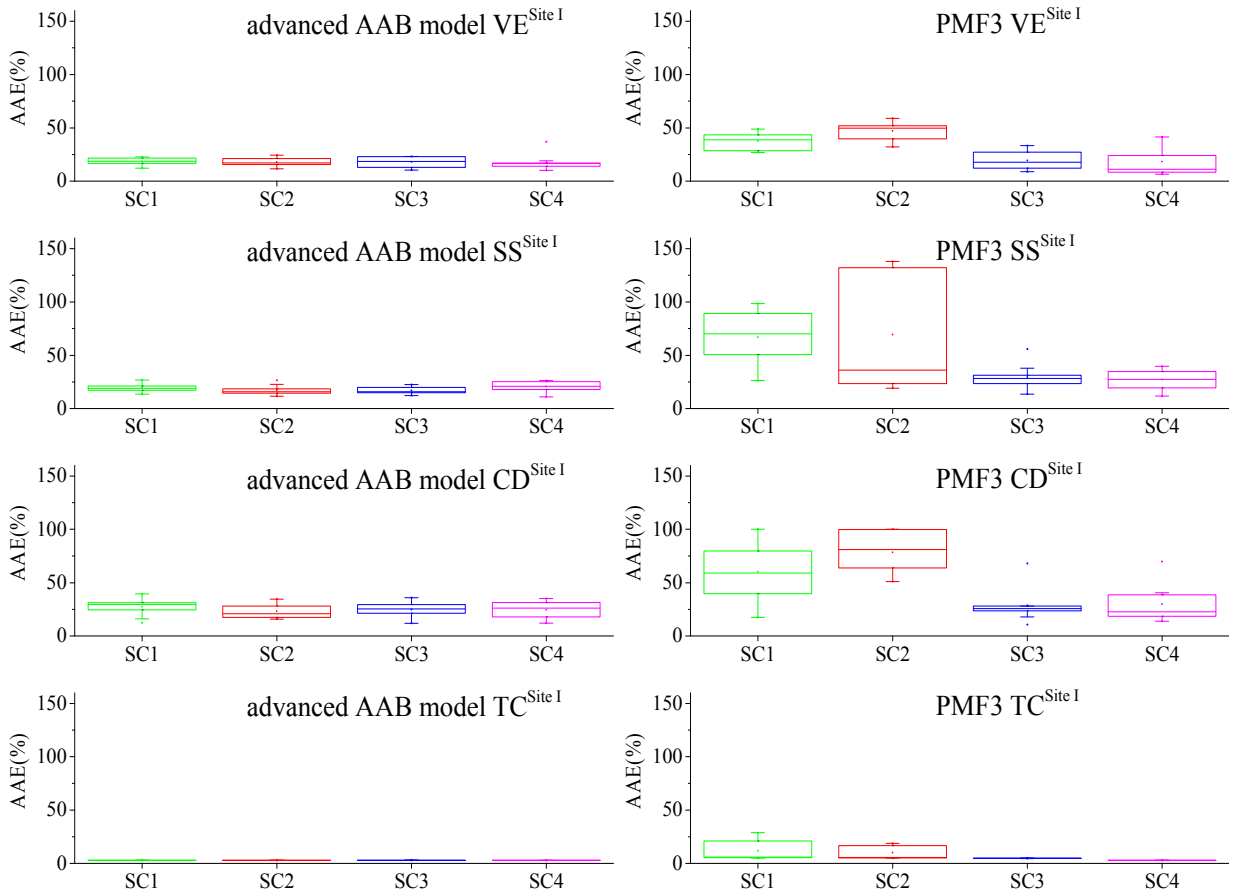

Fig. 3. The AEE of simulated and predicted factor contributions by advanced AAB model and PMF3 respectively for Site I.

investigated sites. In contrast, site-to-site correlations of source contributions do not influence the results of the advanced AAB model. Thus, both models can be used when all sources show high site-to-site correlations, as is the case of the period of regional pollution. When differences in source contributions increase, the performance of PMF3 worsens while the results of the advanced AAB model remain stable. When the datasets share the same source profile but different their emission pattern, AAB three-way model is appreciate to use.

To evaluate the performance of the advanced AAB three-way factor analysis model based on the ambient dataset, the results of the AAB model gleaned from our our other work were examined. Ambient $PM_{2.5}$ samples were collected from two sites in Tianjin, a megacity in China. The ambient datasets for the two sites (three-way datasets) were introduced into the AAB three-way model. The results of the AAB three-way model are summarized in Figs. S5 and S6. According to the source profiles estimated by the AAB three-way model, six factors (sea salt, secondary sulfate and secondary organic carbon, vehicular exhaust, crustal dust and cement dust, coal combustion and biomass burning and secondary nitrate) were identified. At the same time, the coastal and inland site datasets were apportioned by PMF (PMF2). Similar factors and source profiles were obtained in coastal site, but the sea salt factor can't be extracted in inland site Rather, the results of the advanced AAB three-way model are consistent with those of the PMF2, which is widely used. However, the AAB three-way model can make full use of spatial information to facilitate source category identification.

ABB Model Evaluation for Different Source Profile Variability Levels

As studied in our prior work, source profile variability should also affect three-way source apportionment. Thus, the advanced ABB model was employed to solve this problem. In this section, 30 synthetic datasets for 3 scenarios (with varying degrees of source profile variability: CV = 10%, 30% and 50%) were analyzed using the advanced ABB and traditional PMF3 models. The corresponding results were analyzed.

The contributions of vehicular exhaust, secondary sulfate, crustal dust and TC modeled by the advanced ABB model and the traditional PMF3 are shown in Fig. 4 The contributions estimated by the ABB model for vehicular exhaust were 41.56 ± 1.93 μg m^{-3} (Size I, CV = 10%), 41.17 ± 4.22 μg m^{-3} (Size I, CV = 30%), 42.03 ± 2.94 μg m^{-3} (Size I, CV = 50%), 27.83 ± 1.72 μg m^{-3} (Size II, CV = 10%), 27.96 ± 3.63 μg m^{-3} (Size II, CV = 30%) and 27.79 ± 1.89 μg m^{-3} (Size II, CV = 50%) while these results by PMF3 were 49.26 ± 6.17 μg m^{-3} (Size I, CV = 10%), 44.44 ± 22.30 μg m^{-3} (Size I, CV = 30%), 62.54 ± 17.61 μg m^{-3} (Size I, CV = 50%), 33.46 ± 3.98 μg m^{-3} (Size II, CV = 10%), 21.70

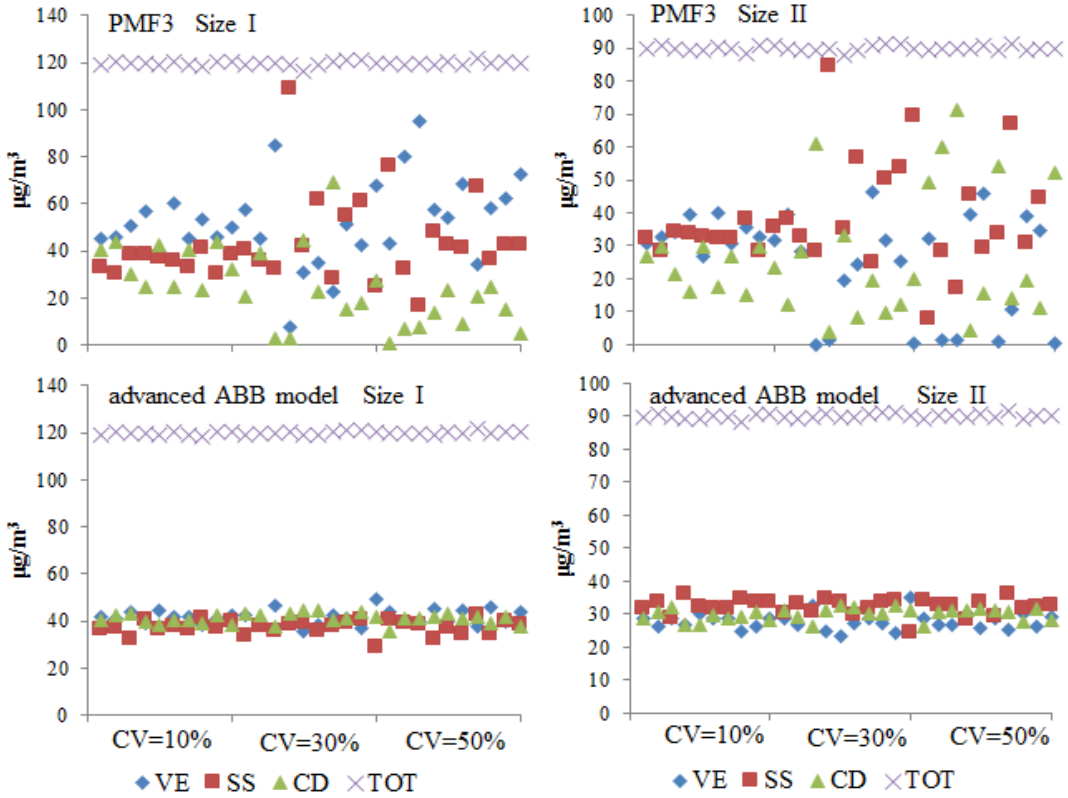

Fig. 4. The source contributions estimated by PMF3 and advanced ABB model in three CVs and two sizes of PM.

± 16.39 μg m^{-3} (Size II, CV = 30%) and 20.61 ± 19.10 μg m^{-3} (Size II, CV = 50%). The contributions estimated for other sources showed the same trends. It is no surprise that the variability of source profiles played a key role in the performance of the PMF3 model: the results of PMF3 were relatively stable in relation to the CV = 10% while the solutions diverged from high standard deviations. For the advanced ABB model, results were stable regardless of the different CVs of source profiles.

CORs between estimated and true contributions of the advanced ABB and traditional PMF3 models are summarized in Figs. 5 and S3 to evaluate modeling performance based on variations in daily contributions. According to Figs. 5 and S3, CORs of PMF3 for vehicular exhaust, secondary sulfate, crustal dust and TC exceeded 0.85 when source profile variability was weak (CV = 10%), but CORs remained rather poor with larger degrees of source profile variability (like CV = 30% and 50%). Fortunately, CORs of the advanced ABB model remained stable at approximately 1. When CV = 10%, the CORs of three source categories estimated by the ABB model ranged from 0.80–0.93 (Size I) and 0.79–0.94 (Size II); when CV = 30%, CORs ranged from 0.78–0.93 (Size I) and 0.80–0.94 (Size II); and when CV = 50%, CORs ranged from 0.80–0.92 (Size I) and 0.77–0.93 (Size II) for three source categories. The CORs of TC exceeded 0.94 in all cases.

Overall, the AAEs (as shown in Figs. 6 and S4) are in agreement with the COR results. The AAEs of three source categories obtained by PMF3 were relatively lower and more stable when CV = 10%, while AAEs were high with large standard deviations when source profile variability was strong. The AAEs of TC fitted by PMF3 were low in most cases with different degrees of source profile variability, but we found some outlier values. For the advanced ABB model, the AAEs of all source categories and TC were relatively low and stable under different degrees of source profile variability. When CV = 10%, AAEs modeled by the ABB model were: 17.30 ± 3.45% (vehicular exhaust, Size I) and 17.48 ± 3.77% (vehicular exhaust, Size II), 18.32 ± 2.42% (secondary sulfate, Size I) and 16.86 ± 2.85% (secondary sulfate, Size II), and 25.41 ± 5.32% (crustal dust, Size I) and 26.63 ± 6.38% (crustal dust, Size II) for CV = 10%; 19.81 ± 3.60% (vehicular exhaust, Size I) and 20.45 ± 3.70% (vehicular exhaust, Size II), 20.43 ± 4.59% (secondary sulfate, Size I) and 19.68 ± 5.79% (secondary sulfate, Size II), and 28.15 ± 4.06% (crustal dust, Size I) and 29.89 ± 5.14% (crustal dust, Size II) for CV = 30%; and 17.25 ± 2.76% (vehicular exhaust, Size I) and 17.54 ± 2.59% (vehicular exhaust, Size II), 19.81 ± 3.16% (secondary sulfate, Size I) and 18.23 ± 3.59% (secondary sulfate, Size II), and 25.20 ± 4.63% (crustal dust, Size I) and 28.35 ± 5.12% (crustal dust, Size II) for CV = 50%. The AAEs of TC ranged from 2.89%–3.19% with standard deviations of less than 0.2.

Due to differing assumptions and fundamental principles, the PMF3 and advanced ABB model performed differently. For the advanced ABB model, source profiles are individually established for each size; while PMF3 shares the same source profiles for all sizes. Therefore, the results can be explained. PMF3 performed well only when the source profiles of all sizes were strongly correlated. As receptor concentration

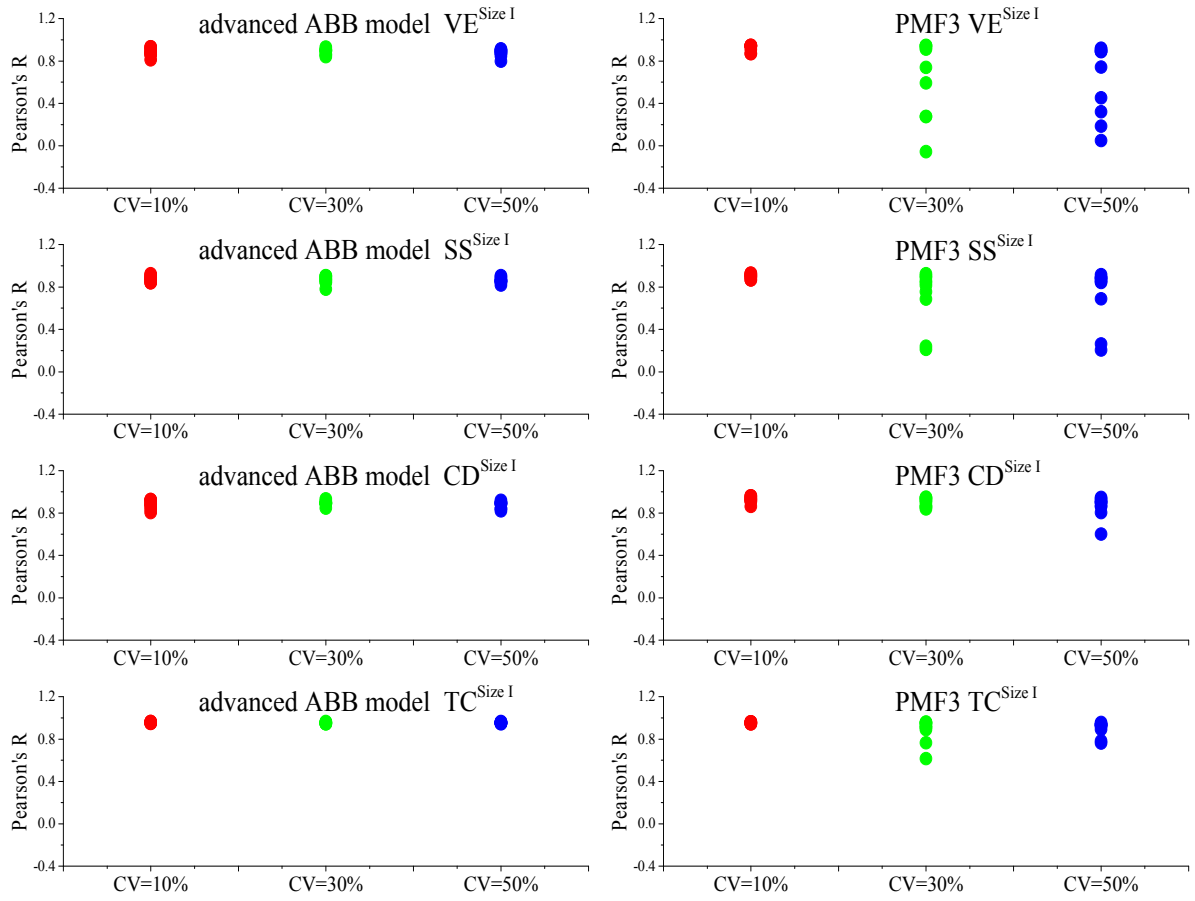

Fig. 5. The CORs of simulated and predicted factor contributions by advanced ABB model and PMF3 respectively in Size I.

disturbance increased, source profile similarities weakened, and PMF3 deteriorated considerably. However, the results of the advanced ABB model remained robust even though the correlation source profile varied. The ABB three-way model require the similar emission pattern but different source profile.

Furthermore, the ABB three-way model was applied to an ambient dataset reported in our previous works (Tian *et al.*, 2016). PM_{10} and $PM_{2.5}$ were synchronously sampled for Chengdu, a megacity in southern China. The ambient dataset for a single size was computed by PMF2 and the three-way dataset of two sizes was estimated by the ABB three-way model. The percentage source contributions evaluated by the two models are summarized in Fig. S7. For both PMF2 and ABB threes-way models, cement dust became enriched in the coarse area and vehicular exhaust, secondary nitrate and SOC became enriched in the fine area. It is interesting to note several differences between the size distributions determined by the two models. The contributions of secondary sulfate fitted by PMF2 were similar in fine and coarse PM while the secondary sulfate of the ABB three-way model clearly became enriched in PM to a lesser degree than it did in $PM_{2.5}$. According to related studies (Arimoto *et al.*, 1996; Srimuruganandam and Shiva Nagendra, 2012), it can be concluded that the size distribution of percentage source contributions (which is defined as the percentage of a certain source contribution accounting for corresponding PM concentration) estimated by the advanced ABB three-way model may be more reasonable than that of the PMF2. This may be explained by the fact that the three-way model can consider internal data on chemical species, temporal visibility levels and PM sizes while two-way models compute datasets of certain sizes independently, resulting in missing information between sizes.

CONCLUSION

To study the accuracy of advanced and traditional three-way factor analysis models, simulated three-way blocks were constructed and modeled by PMF3 and advanced AAB/ABB models. Based on the results and discussion listed above, it can be concluded that emission patterns and the variability of source profiles with different particle sizes should influence the performance of three-way models along with the accuracy of estimated results. For PMF3, the model performed well only when all sources showed strong site-to-site correlations among all investigated sites and when the source profiles of all sites showed strong correlations. In contrast, the advanced three-way factor analysis models were not influenced by emission pattern or source profile variability. These findings may serve as valuable information for environmental scientists as they carry out multi-site and multi-size source apportionment tasks. Ambient datasets

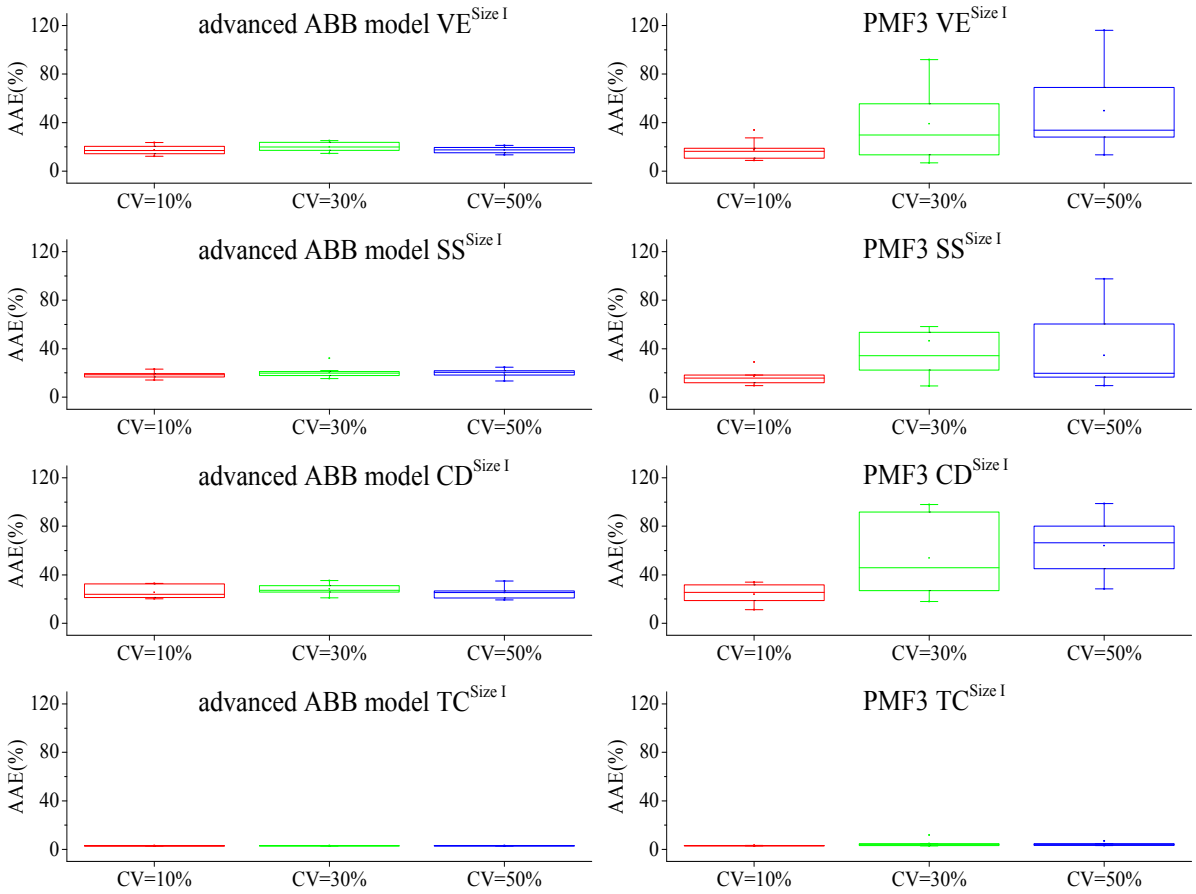

Fig. 6. The AEE of simulated and predicted factor contributions by advanced ABB model and PMF3 respectively in Size I.

were also used to evaluate the performance of the advanced three-way factor analysis model. From the results of our other work on Chengdu and Tianjin, we conclude that the results of the advanced three-way model are not only consistent with those of the PMF2 used internationally but also make full use of spatial information to facilitate source category identification.

The main focus of our analysis was to systematically investigate the accuracy of the models examined. Our findings are dependent on features of the simulated data analyzed. Though the ambient dataset was estimated, it may not reflect its performance perfectly when used to analyze real particle concentration data. More studies should be conducted on the advanced three-way model when run based on real particle concentration data. Furthermore, additional applications of the advanced three-way model for other characteristic three-way datasets should be explored.

ACKNOWLEDGMENTS

This study is supported by the Tianjin Natural Science Foundation (16JCQNJC08700), Fundamental Research Funds for the Central Universities, the National Key Research and Development Plan Project (2016YFC0208500), National Natural Science Foundation of China (21407174 and 21407081), Tianjin Research Program of Application Foundation and Advanced Technology (14JCQNJC08100), Special Funds for Research on Public Welfares of the Ministry of Environmental Protection of China (201309072) and Tianjin Significant Science and Technology Project (16YFZCSF00260, 14ZCDGSF00027 and 14ZCDGSF00029).

SUPPLEMENTARY MATERIAL

Supplementary data associated with this article can be found in the online version at http://www.aaqr.org.

REFERENCES

Abdollahi, H. and Sajjadi, S.M. (2010). On rotational ambiguity in parallel factor analysis. *Chemom. Intell. Lab. Syst.* 103: 144–151.

Amato, F. and Hopke, P.K. (2012). Source apportionment of the ambient $PM_{2.5}$ across St. Louis using constrained positive matrix factorization. *Atmos. Environ.* 46: 329–337.

Arimoto, R., Duce, R.A., Savoie, D.L., Prospero, J.M., Talbot, R., Cullen, J.D., Tomza, U., Lewis, N.F. and Ray, B.J. (1996). Relationships among aerosol constituents from Asia and the North Pacific during PEM-West A. *J. Geophys. Res.* 101: 2011–2023.

Beuck, H., Quass, U., Klemm, O. and Kuhlbusch, T.A.J. (2011). Assessment of sea salt and mineral dust

contributions to PM_{10} in NW Germany using tracer models and positive matrix factorization. *Atmos. Environ.* 45: 5813–5821.

Chen, P., Wang, T., Hu, X. and Xie, M. (2015). Chemical mass balance source apportionment of size-fractionated particulate matter in Nanjing, China. *Aerosol Air Qual. Res.* 15: 1855–1867.

Escrig, A., Monfort, E., Celades, I., Querol, X., Amato, F., Minguillón, M.C. and Hopke, P.K. (2009). Application of optimally scaled target factor analysis for assessing source contribution of ambient PM_{10}. *J. Air Waste Manage. Assoc.* 59: 1296–1307.

Faber, N.M., Bro, R. and Hopke, P.K. (2003). Recent developments in CANDECOMP/PARAFAC algorithms: A critical review. *Chemom. Intell. Lab. Syst.* 65: 119–137.

Gaita, S.M., Boman, J., Gatari, M.J., Wagner, A. and Jonsson, S.K. (2016). Characterization of size-fractionated particulate matter and deposition fractions in human respiratory system in a typical African city: Nairobi, Kenya. *Aerosol Air Qual. Res., in Press*.

Habre, R., Coull, B. and Koutrakis, P. (2011). Impact of source collinearity in simulated $PM_{2.5}$ data on the PMF receptor model solution. *Atmos. Environ.* 45: 6938–6946.

Hopke, P.K., Paatero, P., Jia, H., Ross, R.T. and Harshman, R.A. (1998). Three-way (PARAFAC) factor analysis: Examination and comparison of alternative computational methods as applied to ill-conditioned data. *Chemom. Intell. Lab. Syst.* 43: 25–42.

Huang, R.J., Zhang, Y., Bozzetti, C., Ho, K.F, Cao, J.J., Han, Y., Daellenbach, K.R., Slowik, J.G., Platt, S.M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S.M., Bruns, E.A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann. R., An, Z., Szidat, S., Baltensperger, U., El Haddad, I. and Prévôt, A.S. (2014). High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* 514: 218–222.

Javitz, H.S., Watson, J.G. and Robinson, N. (1988). Performance of the chemical mass balance model with simulated local-scale aerosol. *Atmos. Environ.* 22: 2309–2322.

Kchih, H., Perrino, C. and Cherif, S. (2015). Investigation of desert dust contribution to source apportionment of PM_{10} and $PM_{2.5}$ from a Southern Mediterranean Coast. *Aerosol Air Qual. Res.* 15: 454–464.

Larsen, B.R., Gilardoni, S., Stenström, K., Niedzialek, J., Jimenez, J. and Belis, C.A. (2012). Sources for PM air pollution in the Po Plain, Italy: II. Probabilistic uncertainty characterization and sensitivity analysis of secondary and primary sources. *Atmos. Environ.* 50: 203–213.

Malaguti, A., Mircea, M., La Torretta, T.M., Telloli, C., Petralia, E., Stracquadanio, M. and Berico, M. (2015). Chemical composition of fine and coarse aerosol particles in the Central Mediterranean area during dust and non-dust conditions. *Aerosol Air Qual. Res.* 15: 410–425.

Mantas, E., Remoundaki, E., Halari, I., Kassomenos, P., Theodosi, C., Hatzikioseyian, A. and Mihalopoulosc, N. (2014). Mass closure and source apportionment of $PM_{2.5}$ by Positive Matrix Factorization analysis in urban Mediterranean environment. *Atmos. Environ.* 94: 154–163.

Meena, M., Meena, B.S., Chandrawat, U. and Rani, A. (2016). Seasonal variation of selected metals in particulate matter at an industrial city Kota, India. *Aerosol Air Qual. Res.* 16: 990–999.

Minguillón, M.C., Querol, X., Baltensperger, U. and Prévôt, A.S.H. (2012). Fine and coarse PM composition and sources in rural and urban sites in Switzerland: Local or regional pollution? *Sci. Total Environ.* 427–428: 191–202.

Mooibroek, D., Schaap, M., Weijers, E.P. and Hoogerbrugge, R. (2011). Source apportionment and spatial variability of $PM_{2.5}$ using measurements at five sites in the Netherlands. *Atmos. Environ.* 45: 4180–4191.

Paatero, P. (1997). A weighted non-negative least squares algorithm for three-way "PARAFAC' factor analysis. *Chemom. Intell. Lab. Syst.* 38: 223–242.

Paatero, P. (1999). The Multilinear Engine - A Table-driven least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Graph. Stat.* 8: 854–888.

Parthasarathy, K., Sahu, S.K. and Pandit, G.G. (2016). Comparison of two receptor model techniques for the size fractionated particulate matter source apportionment. *Aerosol Air Qual. Res.* 16: 1497–1508.

Peré-Trepat, E., Ginebreda, A. and Tauler, R. (2007). Comparison of different multiway methods for the analysis of geographical metal distributions in fish, sediments and river waters in Catalonia. *Chemom. Intell. Lab. Syst.* 88: 69–83.

Pu, W., Wang, X., Zhang, X., Ren, Y., Shi, J.S., Bi, J.R. and Zhang, B.D. (2015). Size distribution and optical properties of particulate matter (PM_{10}) and black carbon (BC) during Dust storms and local air pollution events across a Loess Plateau site. *Aerosol Air Qual. Res.* 15: 2212–2224.

Ramadan, Z., Eickhout, B., Song, X.H., Buydens, L.M.C. and Hopke, P.K. (2003). Comparison of positive matrix factorization and multilinear engine for the source apportionment of particulate pollutants. *Chemom. Intell. Lab. Syst.* 66: 15–28.

Shen, G.F., Xue, M., Yuan, S.Y., Zhang, J., Zaho, Q.Y., Li, B., Wu, H.S. and Ding, A.J. (2014). Chemical compositions and reconstructed light extinction coefficients of particulate matter in a mega-city in the western Yangtze River Delta, China. *Atmos. Environ.* 83: 14–20.

Shi, G.L., Tian, Y.Z., Ye, S., Peng, X., Xu, J., Wang, W., Han, B. and Feng, Y.C. (2015). Source apportionment of synchronously size segregated fine and coarse particulate matter, using an improved three-way factor analysis model. *Sci. Total Environ.* 505: 1182–1190.

Shi, G.L., Zeng, F., Li, X., Feng, Y.C., Wang, Y.Q., Liu, G.X. and Zhu, T. (2011). Estimated contributions and uncertainties of PCA/MLR-CMB results: Source apportionment for synthetic and ambient datasets. *Atmos. Environ.* 45: 2811–2819.

Srimuruganandam, B. and Nagendra, S.S. (2012). Application

of positive matrix factorization in characterization of PM_{10} and $PM_{2.5}$ emission sources at urban roadside. *Chemosphere* 88: 120–130.

Srithawirat, T. and Brimblecombe, P. (2015). Seasonal variation of saccharides and furfural in atmospheric aerosols at a semi-urban site. *Aerosol Air Qual. Res.* 15: 821–832.

Taiwo, A.M. (2016). Source apportionment of urban background particulate matter in birmingham, united kingdom using a mass closure model. *Aerosol Air Qual. Res.* 16: 1244–1252.

Tian, Y.Z., Shi, G.L., Han, B., Wang, W., Zhou X.Y., Wang, J., Li, X. and Feng, Y.C. (2014). The accuracy of two- and three-way positive matrix factorization models: Applying simulated multisite datasets. *J. Air Waste Manage. Assoc.* 64: 1122–1129.

Tian, Y.Z., Shi, G.L., Huang-Fu, Y.Q., Song, D.L., Liu, J.Y., Zhou, L.D., Feng, Y.C. (2016). Seasonal and regional variations of source contributions for PM_{10} and $PM_{2.5}$ in urban environment. *Sci. Total Environ.* 557–558: 697–704.

Vu, T.V., Beddows, D.C.S., Delgado-Saborit, J.M. and Harrison, R.M. (2016). Source apportionment of the lung dose of ambient submicrometre particulate matter. *Aerosol Air Qual. Res.* 16: 1548–1557.

Zheng, M., Cass, G.R., Ke, L., Wang, F., Schauer, J.J., Edgerton, E.S. and Russell, A.G. (2007). Source apportionment of daily fine particulate matter at Jefferson Street, Atlanta, GA, during summer and winter. *J. Air Waste Manage. Assoc.* 57: 228–242.

Zheng, M., Wang, F., Hagler, G.S.W., Hou, X., Bergin, M., Cheng, Y., Salmon, L.G., Schauer, J.J., Louie, P.K.K., Zeng, L. and Zhang, Y.H. (2011). Sources of excess urban carbonaceous aerosol in the Pearl River Delta Region, China. *Atmos. Environ.* 45: 1175–1182.

Received for review, May 19, 2016
Revised, July 28, 2016
Accepted, August 28, 2016