

# A New Receptor Model Based on the Alternating Trilinear Decomposition Followed by a Score Matrix Reconstruction for Source Apportionment of Ambient Particulate Matter

Xiang Dong Qing<sup>1, 2, 3, \*</sup>, Lin Da Yin<sup>3</sup>, Xiao Hua Zhang<sup>4, \*</sup>, Yi Huang<sup>1, 2</sup>, Ling Xu<sup>1, 2</sup>, Min He<sup>5</sup>

<sup>1</sup>College of Materials and Chemical Engineering, Hunan City University, Yiyang, China

<sup>2</sup>Hunan Provincial Key Laboratory of Dark Tea and Jin-hua, Hunan City University, Yiyang, China

<sup>3</sup>School of Materials and Environmental Engineering, Hunan University of Humanities, Science and Technology, Loudi, China

<sup>4</sup>Food and Bioengineering College, Xuchang University, Xuchang, China

<sup>5</sup>School of Chemical Engineering, Xiangtan University, Xiangtan, China

## Email address:

xdqing123@hnu.edu.cn (Xiang Dong Qing), xhzhang2329@126.com (Xiao Hua Zhang)

\*Corresponding author

## To cite this article:

Xiang Dong Qing, Lin Da Yin, Xiao Hua Zhang, Yi Huang, Ling Xu, Min He. A New Receptor Model Based on the Alternating Trilinear Decomposition Followed by a Score Matrix Reconstruction for Source Apportionment of Ambient Particulate Matter. *Science Journal of Analytical Chemistry*. Vol. 8, No. 3, 2020, pp. 93-106. doi: 10.11648/j.sjac.20200803.12

Received: June 16, 2020; Accepted: July 3, 2020; Published: July 13, 2020

**Abstract:** A new receptor model based on the alternating trilinear decomposition followed by a score matrix reconstruction (ATLD-SMR) was developed for the source apportionment of urban PM<sub>10</sub> for the first time. First, simulated three-way data arrays of gas chromatography-mass spectrometry (GC-MS) were used to verify the feasibility of the ATLD-SMR method. Then, PM<sub>10</sub> samples (receptor) at five locations and TSP samples of ten pollution sources were collected during July and August, 2018 in Loudi City, China. The collected samples were measured by GC-MS. PAHs were used as tracers and their concentrations were accurately obtained by the ATLD-SMR analysis of GC-MS data of these samples after the problems of GC-MS including baseline drift, retention-time shift and unexpected peaks overlapping were successfully resolved. The highest concentrations of individual PAH in these samples were for phenanthrene and benzo [a] pyrene (40.76 ng m<sup>-3</sup> and 39.63 ng m<sup>-3</sup> in Liangang steel-making workshop, respectively). Last, a relative contribution matrix of the source to the receptor was estimated by the ATLD-SMR method. The proposed method was employed to apportion the source contributions to PM<sub>10</sub> particles at five locations and reasonable results were obtained, thus presenting a promising tool for source apportionment of complex ambient particulate matter.

**Keywords:** Source Apportionment, PM<sub>10</sub>, PAHs, GC-MS, ATLD-SMR

## 1. Introduction

Atmospheric pollution is an environmental concern worldwide leading to increasing morbidity due to cardiovascular disease and cancer as a result of exposure to polluted air [1-3]. Accurate source apportionment of ambient particulate matter has become among the hot topics in the atmospheric sciences. Until now, three approaches have been applied to estimate source contributions from source sample data and/or ambient sample data. The first, emission inventory is a list model of anthropogenic emissions based on their

history and heritage [4-5]. The second, diffusion model (also termed a source model), evaluates the influences of pollution sources on airborne particulates at a given site based on the discharge rates of the point source and meteorological data [6]. The third, known as a receptor model, focuses on studying the contribution of the pollution source to the receptor [7-9].

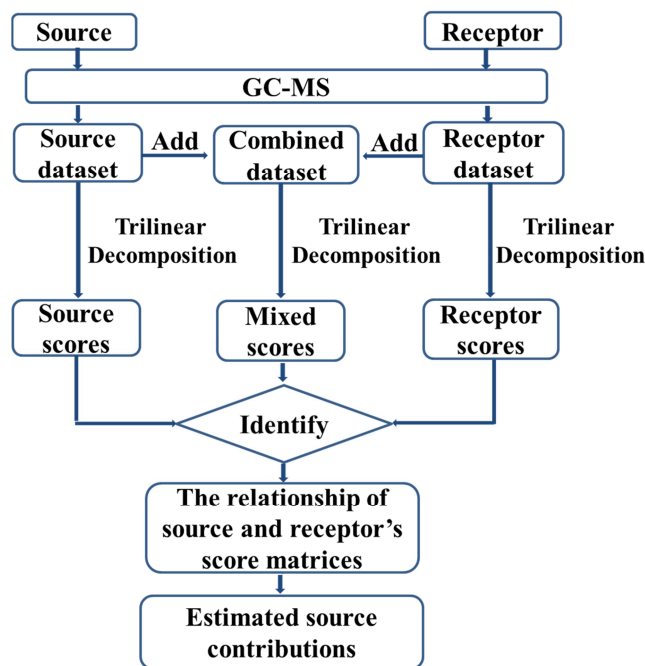
Receptor models include source unknown receptor models, such as principal component analysis (PCA) [10, 11], PCA/multiple linear regression (PCA/MLR) [12, 13] and positive matrix factorization (PMF) [14], and source known receptor models, for example, the classic chemical matter balance model (CMB) [15]. Receptor models have been

widely applied in the field of source apportionment of ambient particulate matter [7, 16-19]. In these methods, inorganic matter (metallic elements and inorganic ions) or organic compounds are commonly used as tracers to calculate the presence of the effluent from different sources within the ambient particulate matter [20, 21]. Therefore, there are two important problems that must be solved before source apportionment. The first is tracer identification in a complex environmental sample and the second is accurate quantification of tracers. However, ambient particulate samples contain the most chemically complicated components with low concentrations and severe mutual interference. This makes accurate tracer qualitative and quantitative determination in these samples even more difficult or impossible even with the aid of modern analytical instruments such as liquid or gas chromatography-mass spectrometry (LC/GC-MS). However, the accuracy of these methods is closely associated with tracer qualitative and quantitative results. Thus, there is a great need to develop one alternative strategy to carry out the nondestructive and accurate identification and quantification of tracers in source and receptor samples even in the presence of unknown interferences. Fortunately, the chemometrics second-order calibration method provides a good analytical strategy to conduct source apportionment of ambient particulate matter. It has the advantages of simplifying sample pretreatments, resolving overlapped chromatographic peaks and providing physicochemical information on components in an actual system because of its well-known “second-order advantage” [22-24].

According to the aforementioned discussion, there may be three methods which can be developed to construct source apportionment models for analysis of the three-way data array. A graphical representation of these ways was shown in Figure 1. The first method is based on the trilinear decomposition of only the source dataset with second-order calibration algorithm to obtain the relative contributions of pollution sources. The second is built on the iteration and decomposition of only the receptor dataset to extract the number and category of pollution sources. The third is based on first decomposing the source dataset and the receptor dataset with second-order calibration algorithm, respectively, and then estimating the relative contribution coefficients by combining the obtained source and receptor score matrices. These aforementioned methods may be found in some existing approaches elsewhere [25-27]. However, there are some deficiencies. For example, only the information from one side of the source and receptor can be utilized in the first two ways and the chemical compositions of sources may not be fully matched with those of the receptor in the third way.

Thus, in this work, a method based on another new way is proposed to construct a source apportionment model. The proposed method first decomposed the combined three-way GC-MS data array of source and receptor samples with alternating trilinear decomposition algorithm and then estimated the relative contribution of a pollution source to ambient particulate matter according to the reconstructed

score matrix of tracers extracted from the source and receptor samples. This new method, the alternating trilinear decomposition followed by a score matrix reconstruction (ATLD-SMR), has the advantages of fast, simple, accurate interference-free quantification of tracers and full utilization of the fingerprint information of samples to estimate the contributions of the source to the receptor. The developed method is elaborated in the following sections.



**Figure 1.** A graphical representation of different ways to construct a source apportionment model based on GC-MS data combined with trilinear decomposition method.

## 2. Method

### 2.1. Alternating Trilinear Decomposition

In this work, alternating trilinear decomposition (ATLD) was used to decompose the obtained three-way data array of the source and receptor samples. ATLD was proposed by Wu et al in 1998 [28]. The algorithm has fast convergence and is insensitive to excess factors, which are very useful properties in real environmental situations for three-way data decomposition. Details of ATLD can be found in the literature [28]. The matrices A, B and C can be provided by ATLD decomposition of three-way data array. The analyte (s) concentrations can be obtained by regression of the appropriate column of C corresponding to each analyte against its standard concentrations.

### 2.2. ATLD Followed by a Score Matrix Reconstruction

In the case of GC-MS, if the measured matrices of the  $k$  samples including the  $k_1$  calibration samples,  $k_2$  source samples and  $k_3$  receptor samples ( $k=k_1+k_2+k_3$ ), are stacked in a three-way data array (GC×MS×Sample), it can be decomposed via ATLD according to Equation (1) as follows:

$$x_{ijk} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} + e_{ijk}, i=1,2,\dots,I; j=1,2,\dots,J; k=1,2,\dots,K. \quad (1)$$

where,  $x_{ijk}$  is the element of the three-way data array  $X$  ( $I \times J \times K$ ), and  $a_{in}$ ,  $b_{jn}$ , and  $c_{kn}$  are the corresponding elements of the loading matrices  $A$  ( $I \times N$ ),  $B$  ( $J \times N$ ) and score matrix  $C$  ( $K \times N$ ) of  $X$ , respectively. The term  $e_{ijk}$  is the element of the

three-way residue data array  $E$  ( $I \times J \times K$ ).  $A$  and  $B$  collect the normalized elution-time and mass-to-charge ( $m/z$ ) profiles of the  $N$  factors, respectively.  $C$  collects the relative concentration or score of  $N$  factors in all samples.

Then, Equation (1) can be written in terms of matrix  $X_k$ , where  $X_k$  is the  $k$ th ( $I \times J$ ) slice of  $X$ .

$$X_k = c_{k1} a_1 b_1' + c_{k2} a_2 b_2' + \dots + c_{kn} a_n b_n' + E_k, n=1,2,\dots,N; k=1,2,\dots,K. \quad (2)$$

where  $a_n$  and  $b_n$  are the  $n$ th column of the loading matrices  $A$  and  $B$ , respectively, and  $(a_n b_n')$  is the loading matrix of the  $n$ th factor. The score value ( $c_{kn}$ ) is the relative contribution of the loading factor ( $a_n b_n'$ ) to matrix  $X_k$ . However,  $N$  denotes the total number of tracers of interest, interference and background. Thus, we can extract the vectors of tracers from raw score matrix. By regression of the vector (corresponding to relative concentrations) of each tracer against its standard concentrations in calibration samples, the targeted concentrations of markers can be obtained, which is second-order calibration.

For the reconstruction of a score matrix of only markers to estimate the contribution of the source to the receptor, the following equation should first be calculated for the  $k_2$  source

samples:

$$\begin{bmatrix} X_{1,s} \\ X_{2,s} \\ \vdots \\ X_{k_2,s} \end{bmatrix} = \begin{bmatrix} c_{11,s} & c_{12,s} & \dots & c_{1n,s} \\ c_{21,s} & c_{22,s} & \dots & c_{2n,s} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k_2 1,s} & c_{k_2 2,s} & \dots & c_{k_2 n,s} \end{bmatrix} \begin{bmatrix} a_1 b_1' \\ a_2 b_2' \\ \vdots \\ a_n b_n' \end{bmatrix} + \begin{bmatrix} E_{1,s} \\ E_{2,s} \\ \vdots \\ E_{k_2,s} \end{bmatrix} \quad (3)$$

where the subscript “s” represents “source”.  $c_{k_2 n,s}$  is the targeted concentration of the marker in the source samples. Here,  $k_2$  is equal to  $n$ . Then, the measured matrices of the  $k_3$  receptor samples can also be obtained as follows:

$$\begin{bmatrix} X_{1,r} \\ X_{2,r} \\ \vdots \\ X_{k_3,r} \end{bmatrix} = \begin{bmatrix} c_{11,r} & c_{12,r} & \dots & c_{1n,r} \\ c_{21,r} & c_{22,r} & \dots & c_{2n,r} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k_3 1,r} & c_{k_3 2,r} & \dots & c_{k_3 n,r} \end{bmatrix} \begin{bmatrix} a_1 b_1' \\ a_2 b_2' \\ \vdots \\ a_n b_n' \end{bmatrix} + \begin{bmatrix} E_{1,r} \\ E_{2,r} \\ \vdots \\ E_{k_3,r} \end{bmatrix} \quad (4)$$

The subscript “r” represents “receptor”.  $c_{k_3 n,r}$  is the targeted concentration of the marker in the receptor samples. Thus, the Equation (5) can be obtained by combining Equations (3) and (4):

$$\begin{bmatrix} X_{1,r} \\ X_{2,r} \\ \vdots \\ X_{k_3,r} \end{bmatrix} = \begin{bmatrix} c_{11,r} & c_{12,r} & \dots & c_{1n,r} \\ c_{21,r} & c_{22,r} & \dots & c_{2n,r} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k_3 1,r} & c_{k_3 2,r} & \dots & c_{k_3 n,r} \end{bmatrix} \begin{bmatrix} c_{11,s} & c_{12,s} & \dots & c_{1n,s} \\ c_{21,s} & c_{22,s} & \dots & c_{2n,s} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k_2 1,s} & c_{k_2 2,s} & \dots & c_{k_2 n,s} \end{bmatrix}^{-1} \begin{bmatrix} X_{1,s} - E_{1,s} \\ X_{2,s} - E_{2,s} \\ \vdots \\ X_{k_2,s} - E_{k_2,s} \end{bmatrix} + \begin{bmatrix} E_{1,r} \\ E_{2,r} \\ \vdots \\ E_{k_3,r} \end{bmatrix} \quad (5)$$

where

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1k_2} \\ p_{21} & p_{22} & \dots & p_{2k_2} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k_3 1} & p_{k_3 2} & \dots & p_{k_3 k_2} \end{bmatrix} = \begin{bmatrix} c_{11,r} & c_{12,r} & \dots & c_{1n,r} \\ c_{21,r} & c_{22,r} & \dots & c_{2n,r} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k_3 1,r} & c_{k_3 2,r} & \dots & c_{k_3 n,r} \end{bmatrix} \begin{bmatrix} c_{11,s} & c_{12,s} & \dots & c_{1n,s} \\ c_{21,s} & c_{22,s} & \dots & c_{2n,s} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k_2 1,s} & c_{k_2 2,s} & \dots & c_{k_2 n,s} \end{bmatrix}^{-1} \quad (6)$$

$P$  is a reconstructed score matrix, which is also called a relative contribution matrix.  $p_{k_3 k_2}$  is the relative contribution of the source matrix  $X_{k_2,s}$  to the receptor matrix  $X_{k_3,r}$ . Finally, the measured matrix of any receptor sample can be expressed with respect to the  $k_2$  measured matrices of pollution sources as follows:

$$X_{k_3,r} = p_{k_3 1} X_{1,s} + p_{k_3 2} X_{2,s} + \dots + p_{k_3 k_2} X_{k_2,s} + \bar{E}_{k_3,r}, k_2 = 1,2,\dots,K; k_3 = 1,2,\dots,K. \quad (7)$$

$\bar{E}_{k_3,r}$  is the error matrix.

$$\begin{bmatrix} \bar{E}_{1,r} \\ \bar{E}_{2,r} \\ \vdots \\ \bar{E}_{k_3,r} \end{bmatrix} = \begin{bmatrix} E_{1,r} \\ E_{2,r} \\ \vdots \\ E_{k_3,r} \end{bmatrix} - \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k_2} \\ p_{21} & p_{22} & \cdots & p_{2k_2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k_31} & p_{k_32} & \cdots & p_{k_3k_2} \end{bmatrix} \begin{bmatrix} E_{1,s} \\ E_{2,s} \\ \vdots \\ E_{k_2,s} \end{bmatrix} \quad (8)$$

Thus, we can obtain the relative contribution of the pollution source to ambient particulate matter based on the alternating trilinear decomposition followed by a score matrix reconstruction (ATLD-SMR) of the three-way GC-MS data of source and receptor samples.

### 3. Experiment

#### 3.1. Simulated Chromatography-mass Spectrum-sample Data

A chromatography-mass spectrum data set was simulated for 15 samples with four components. The chromatographic profiles were generated as follows:

$$A(i, 1) = \text{Gaussian peak}(0.10, 4, 10, i),$$

$$A(i, 2) = \text{Gaussian peak}(0.20, 10, 25, i),$$

$$A(i, 3) = \text{Gaussian peak}(0.15, 7, 45, i),$$

$$A(i, 4) = \text{Gaussian peak}(0.55, 5, 55, i), \text{ all for } i=1, 2, \dots, 80$$

The Gaussian peak  $(x, a, b, i)$  represents the value at  $i$  of a Gaussian distribution with standard deviation  $a$ , where  $x$  is the height at the center of  $b$ .

The mass-spectrum profiles were reproduced by the following equation

$$B = \text{vsparse}(r, c, s, m, n)$$

where  $r$  and  $c$  are two index vectors which representing the positions of nonzero elements,  $s$  is a vector including all the nonzero elements corresponding to the index of  $(r, c)$ , and  $m$  and  $n$  are the row dimensions and column dimensions of matrix  $B$ , respectively. Herein,  $m$  and  $n$  are equal to 200 and 4, respectively.

The concentrations of four species in 11 samples including seven calibration samples and four source samples are distributed in the range of 0–15. Then, the concentrations of four species in four receptor samples are produced according to Equation (3), where the relative contribution matrix  $P$  was set as follows:

$$P = \begin{bmatrix} 0.40 & 0.35 & 0.15 & 0.10 \\ 0.15 & 0.45 & 0.30 & 0.10 \\ 0.15 & 0.20 & 0.57 & 0.08 \\ 0.17 & 0.23 & 0.15 & 0.45 \end{bmatrix}$$

Then, a three-way data array was obtained with the size of

$80 \times 200 \times 15$ . To simulate a real environmental measurement process, homoscedastic and heteroscedastic noises were added to the simulated data set as follows:

$$\text{Homo\_noise} = \sigma_{\text{homo}} \times \text{RANDN} \times \max(X_{..k})$$

$$\text{Heter\_noise} = \sigma_{\text{heter}} \times \text{RANDN} * X_{..k}$$

where  $\sigma_{\text{homo}}$  and  $\sigma_{\text{heter}}$  are the homoscedastic and heteroscedastic noise levels, taking values of 0.001, 0.002, 0.005 and 0.01.

#### 3.2. Experimental Data

##### 3.2.1. Sampling Procedure

PM10 (particulate matter with  $dp < 10 \mu\text{m}$ ) samples were collected for 12 hours once every four days during July and August of 2018 at five functional zones of Loudi City, China using a large-volume air sampler (YH-1000, Qingdao Jingcheng Instrument Co., Ltd, China) at a flow rate of  $1.0 \text{ m}^3 \text{ min}^{-1}$ . The sampler was mounted on the roof of a building approximately 15 m above the ground. PM10 was captured onto quartz-fiber filters  $200 \text{ mm} \times 240 \text{ mm}$  in size. TSP (total suspended particulates) samples of different sources were collected for 24 hours during the month by using four medium-volume samplers (JCH-120F (one) and JCH-6120 (three), Qingdao Juchuang Environmental Co., Ltd, China) at a flow rate of  $100.0 \text{ L min}^{-1}$ . TSP was captured onto 88-mm diameter quartz-fiber filters. Figure 2 shows the locations of the sampling sites of the PM10 and TSP samples. The information for these samples is provided in supplementary Table A1. These samples were separately wrapped in zip-top plastic bags and stored at  $-20.0^\circ\text{C}$  in a refrigerator.

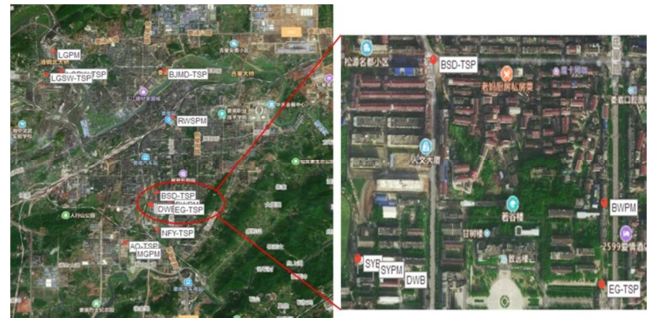


Figure 2. Locations of sampling sites in Loudi city, China.

##### 3.2.2. PAHs Extraction

The Soxhlet method was used to extract PAHs from the filter samples. A total of 300.0 mL of dichloromethane/n-hexane with a volume ratio of 3:2 was used as an extraction agent and added into a Soxhlet extractor to fully immerse the filter samples. The extraction temperature was set at  $78^\circ\text{C}$ . The extraction time was 20 min for each cycle and the filling/emptying cycles exceeded twelve. After the extraction, the solvent was concentrated using rotary evaporation at  $40^\circ\text{C}$  under low pressure to 2.0 mL and the solution was maintained on standby in a small beaker.

### 3.2.3. Analytical Procedures

#### 1. Reagents and chemicals

Standard substances of eight PAHs including acenaphthylene (ACN), fluorene (FLU), phenanthrene (PHE), pyrene (PYR), benzo [a] anthracene (BaA), chrysene (CHR), benzo [e] pyrene (BeP) and benzo [a] pyrene (BaP) were provided by Aladdin Chemical Co. Ltd. (Shanghai, China). The others including fluoranthene (FLO) and benzo [b] fluoranthene (BbF) were obtained from AccuStandard Inc. (Connecticut, USA). Methanol (HPLC-grade,  $\geq 99.9\%$ ) for dissolving all the analytical standards is obtained from Aladdin. Others solvents including dichloromethane (AR) and n-hexane (AR) were provided by Hunan Hui-hong Reagent Co. Ltd. (Changsha, China).

#### 2. Instrumentation and software

Retention time-mass spectral matrices were obtained on gas chromatography-mass spectrometry (GC/MS-QP2010, Shimadzu Corporation, Japan). The column for GC separation was an Rxi®-IMS (0.25  $\mu\text{m}$  and 0.25 mm  $\times$  30 m) nonpolar capillary column (Restek, USA). The temperature program of the GC was that it was initiated at 150°C for 2.0 min, increased to 300°C at a rate of 30°C min<sup>-1</sup>, then maintained at 300°C for 5.0 min. The mass spectrum detector was operated with electron impact (EI) source and under the following conditions: full scan mode with a range of 50-400 amu; injector temperature, 250°C; ion source temperature, 230°C; interface temperature, 250°C; and detection voltage, 0.94 kV. In addition, the sample injection volume was 1.0  $\mu\text{L}$  with a split ratio of 5:1 and helium was used as carrier gas at a flow rate of 2.7 mL/min. The data analysis was performed on Matlab (MATLAB, 7.9.0, Math Works, USA) software with home-written programs in the environment of a Microsoft Windows 8 operating system.

#### 3. Standard solutions

Each of analytical standards including BeP and BbF was weighed accurately, dissolved into a little dichloromethane, then transferred into a 100.0 mL volumetric flask and diluted to the mark with methanol to obtain the appropriate stock solutions. Stock solutions of the other eight PAHs were prepared by directly dissolving corresponding analytical standards into methanol. These stock solutions were stored at 4.0°C in a refrigerator.

#### 4. Sample sets

Seven calibration samples (C01-C07) were prepared by mixing the working solutions of 10 PAHs, and then diluting to volume with pure methanol in 10.0 mL volumetric flasks. Concentrations of 10 PAHs in these samples were summarized in supplementary Table A2. Samples of PM10 (n=5) and TSP (n=12) were directly extracted for analysis by GC-MS. All of the samples were filtered with 0.22- $\mu\text{m}$  nonsterile PTFE syringe filter (i-Quip®N2536) before auto-injection into the separation column. Note that human exposure to PAHs may cause fatal risk, so effective precautions have to be taken in the whole experiment. The sample preparation should be carried out in a fume hood with wearing suitable gloves, eye/face protection and other protective clothes.

## 4. Results and Discussions

### 4.1. Simulated Data Analysis

#### 4.1.1. A Noise-free Three-way Data Array

First, a three-way chromatography-mass spectrum data array without any noise was simulated to investigate the accuracy of the novel method. In the simulated data, four source samples were designed to reproduce four receptor samples according to Equation (7) and four species were used as markers in the source and receptor samples. A display of the normalized chromatogram, mass spectrogram and concentration profiles of simulated data is shown in Figure A1. From Figure A1 (A) and (B), we can see that it was complex system because there was serious overlap between the four components in the dimensions of both chromatography and mass spectrometry. However, ATLD-SMR can be firstly applied to decompose the simulated three-way data array to obtain the accurate chromatograms, mass spectrograms, and concentrations of the four components (The information is not shown here because the discussion on the resolution and quantification of analytes with second-order calibration methods can be found elsewhere). Then, the contributions of different sources to receptors can be estimated by the obtained concentrations of the four components with the ATLD-SMR method. The fourth column of Tables A3-A4 and the third column of Table 1 list the predicted contributions and statistical results of the four sources to four receptors in the simulated data. The accuracy of the developed method was 100% for the noise-free three-way data, which indicates that the proposed method is feasible.

#### 4.1.2. A Three-way Data Array with Different Homoscedastic Noise Levels

In real environmental analysis, noise often occurs in measured data arrays. Thus, different homoscedastic noise levels (0.001, 0.002, 0.005 and 0.01, respectively) were first added in the simulated data array to further check the performance of the proposed method. From Tables A3-A4 and Table 1, it can be seen that when the homoscedastic noise level was less than 0.005, the results of the developed method were satisfactory with an AD ( $\times 10^{-2}$ ) ranging from 0.2 to 3.8 and an RMSEP ( $\times 10^{-3}$ ) ranging from 0.3 to 9.5. When the noise level increased to 0.01, the results of the new method were still satisfactory with an ARC ranging from 0.9609 to 1.0417 and T-test values less than 1.5. These results of the developed method were acceptable for a homoscedastic noise level as high as 0.01.

#### 4.1.3. A Three-way Data Array with Different Heteroscedastic Noise Levels

Then, different heteroscedastic noise levels (0.001 to 0.01) were integrated into the simulated three-way data array to verify the feasibility of the novel method. The results are summarized in Table 2 and Supporting Information A5-A6. It was found that the results were similar to those of the homoscedastic noise. The AD and RMSEP values increased with an increase in the heteroscedastic noise level. When the

noise level was 0.01, the results of the proposed method remained acceptable, although there was one outlier value of related coefficient (see Supporting Information A6). During the real measurement process, the noise level was rarely

greater than 0.01. Thus, the new method is a promising tool for source apportionment of complex ambient particulate matter as demonstrated by real atmospheric data in the forthcoming sections.

**Table 1.** Statistical parameters obtained with the developed method in analysis of a simulated data set with different homoscedastic noise.

| Source | Parameter           | 0.000                 | 0.001  | 0.002  | 0.005                | 0.01                 |
|--------|---------------------|-----------------------|--------|--------|----------------------|----------------------|
| S1     | ARC <sup>a</sup>    | 1.0000                | 1.0001 | 0.9847 | 0.9945               | 1.0003               |
|        | AD <sup>b</sup>     | $6.4 \times 10^{-12}$ | 0.2    | 0.9    | 1.7                  | 4.1                  |
|        | T-test <sup>c</sup> | 0.4                   | 0.1    | 2.1    | 0.4                  | $1.0 \times 10^{-2}$ |
|        | RMSEP <sup>d</sup>  | $1.5 \times 10^{-11}$ | 0.3    | 3.8    | 4.3                  | 8.8                  |
| S2     | ARC                 | 1.0000                | 0.9966 | 1.0048 | 1.0264               | 0.9960               |
|        | AD                  | $8.3 \times 10^{-12}$ | 0.6    | 1.7    | 1.2                  | 1.7                  |
|        | T-test              | 0.6                   | 0.7    | 0.4    | 2.8                  | 0.3                  |
|        | RMSEP               | $4.2 \times 10^{-11}$ | 2.8    | 5.5    | 9.5                  | 6.7                  |
| S3     | ARC                 | 1.0000                | 1.0008 | 1.0038 | 0.9702               | 0.9609               |
|        | AD                  | $2.4 \times 10^{-11}$ | 1.2    | 1.9    | 2.2                  | 4.4                  |
|        | T-test              | 1.2                   | 0.1    | 0.3    | 1.9                  | 1.3                  |
|        | RMSEP               | $9.7 \times 10^{-11}$ | 2.8    | 6.7    | 9.3                  | 14.0                 |
| S4     | ARC                 | 1.0000                | 1.0144 | 1.0053 | 0.9988               | 1.0417               |
|        | AD                  | $1.8 \times 10^{-11}$ | 1.0    | 1.7    | 3.8                  | 3.4                  |
|        | T-test <sup>c</sup> | 1.6                   | 2.0    | 0.4    | $3.9 \times 10^{-2}$ | 1.5                  |
|        | RMSEP               | $3.9 \times 10^{-11}$ | 1.9    | 3.8    | 6.0                  | 12.0                 |

<sup>a</sup>ARC=Average related coefficients; the related coefficient is the ratio between the predicted and the true contribution.

<sup>b</sup>AD ( $\times 10^{-2}$ )= Average deviation.

<sup>c</sup>  $T = (\bar{X} - \mu_0) / (S / \sqrt{n})$ ; here,  $\bar{X}$  is the average related coefficient,  $\mu_0$  is 1.0000,  $n$  is the degree of freedom (where  $n+1$  is the number of evaluated levels), and the confidence level is 95%,  $T_{0.025}^4 = 3.18$ .

<sup>d</sup>The root mean square error of prediction (RMSEP,  $\times 10^{-3}$ ) can be calculated as follows:

$$RMSEP = \left[ \frac{1}{I-1} \sum (c_{\text{true}} - c_{\text{predicted}})^2 \right]^{1/2}$$

where  $c_{\text{true}}$  and  $c_{\text{predicted}}$  are the actual and predicted contribution, respectively, and  $I$  is the number of prediction samples.

**Table 2.** Statistical parameters obtained with the developed method in analysis of a simulated data set with different heteroscedastic noise.

| Source | Parameter | 0.001  | 0.002  | 0.005  | 0.01   |
|--------|-----------|--------|--------|--------|--------|
| S1     | ARC       | 0.9952 | 1.0056 | 1.0171 | 1.0148 |
|        | AD        | 0.7    | 0.6    | 2.0    | 3.2    |
|        | T-test    | 0.8    | 1.2    | 1.1    | 0.7    |
|        | RMSEP     | 2.0    | 1.7    | 5.3    | 6.8    |
| S2     | ARC       | 0.9990 | 0.9882 | 0.9988 | 1.0161 |
|        | AD        | 0.2    | 1.4    | 2.7    | 5.0    |
|        | T-test    | 0.7    | 1.3    | 0.1    | 0.4    |
|        | RMSEP     | 1.2    | 6.2    | 8.8    | 25.0   |
| S3     | ARC       | 1.0007 | 1.0311 | 0.9632 | 0.9760 |
|        | AD        | 0.2    | 3.2    | 4.1    | 9.9    |
|        | T-test    | 0.4    | 1.3    | 1.1    | 0.3    |
|        | RMSEP     | 1.1    | 8.2    | 11.0   | 34.0   |
| S4     | ARC       | 1.0112 | 0.9874 | 1.0876 | 1.0737 |
|        | AD        | 1.6    | 1.2    | 6.4    | 8.5    |
|        | T-test    | 0.9    | 1.3    | 1.9    | 1.1    |
|        | RMSEP     | 2.1    | 1.9    | 11.0   | 12.0   |

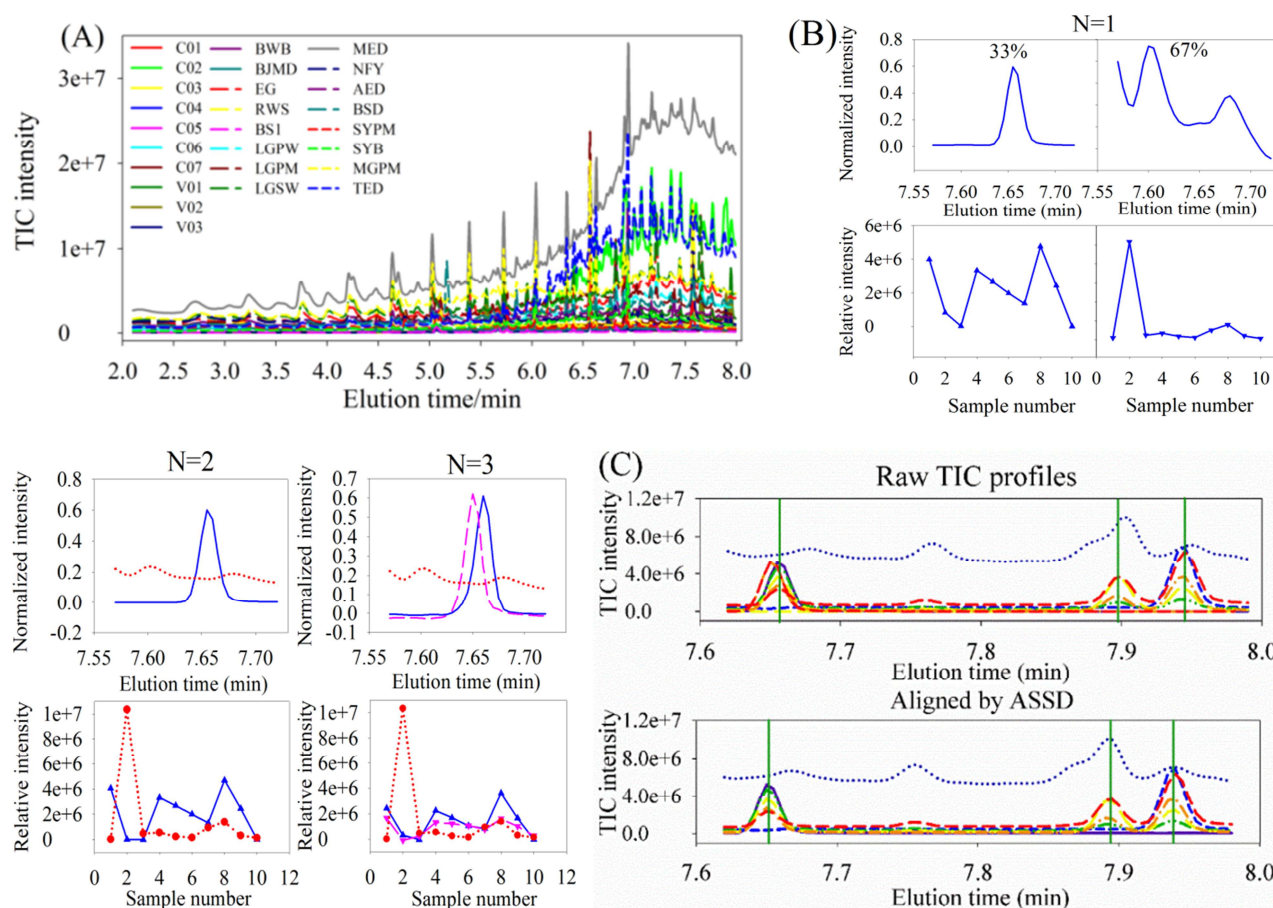
## 4.2. Real Data Analysis

### 4.2.1. The Problems of Baseline Drift and Retention-time Shift

Figure 3 (A) shows the total ion current (TIC) chromatograms of all of the samples studied in this work. As

seen from Figure 3 (A), there were serious baseline drifts in the studied retention domains for most of the samples. Another troublesome problem was that the chromatographic baseline drift of each sample differed from that of one another. Thus, it was clear that it was impossible to carry out an accurate identification and quantification of tracers of interest for such an actual complex situation with univariate regression. Usually, the baseline drift can be fitted as one or more extra component (s) in LC $\times$ LC-DAD or LC-MS data with second-order calibration method [29-30]. Herein, the ATLD method was firstly introduced to address the difficult problem of baseline drift in GC-MS data for the first time. As illustrated in Figure 3 (B), raw GC-MS data array of BbF in calibration samples (7.57-7.72 min, only one analyte of BbF) was decomposed by the ATLD method for 100 times with factor number  $N=1$ , the result was very poor and there was only 33% probability to get BbF with regression coefficient ( $R$ ) of only 0.9834. However, when  $N=2$ , the raw array was fitted very well by ATLD ( $R=0.9999$ ). The results demonstrated that baseline drift was fully extracted out as an extra component and “pure signal” of BbF can be exactly identified. Therefore, the problem of serious baseline drifts in GC-MS analysis can be easily resolved with the aid of the proposed strategy.





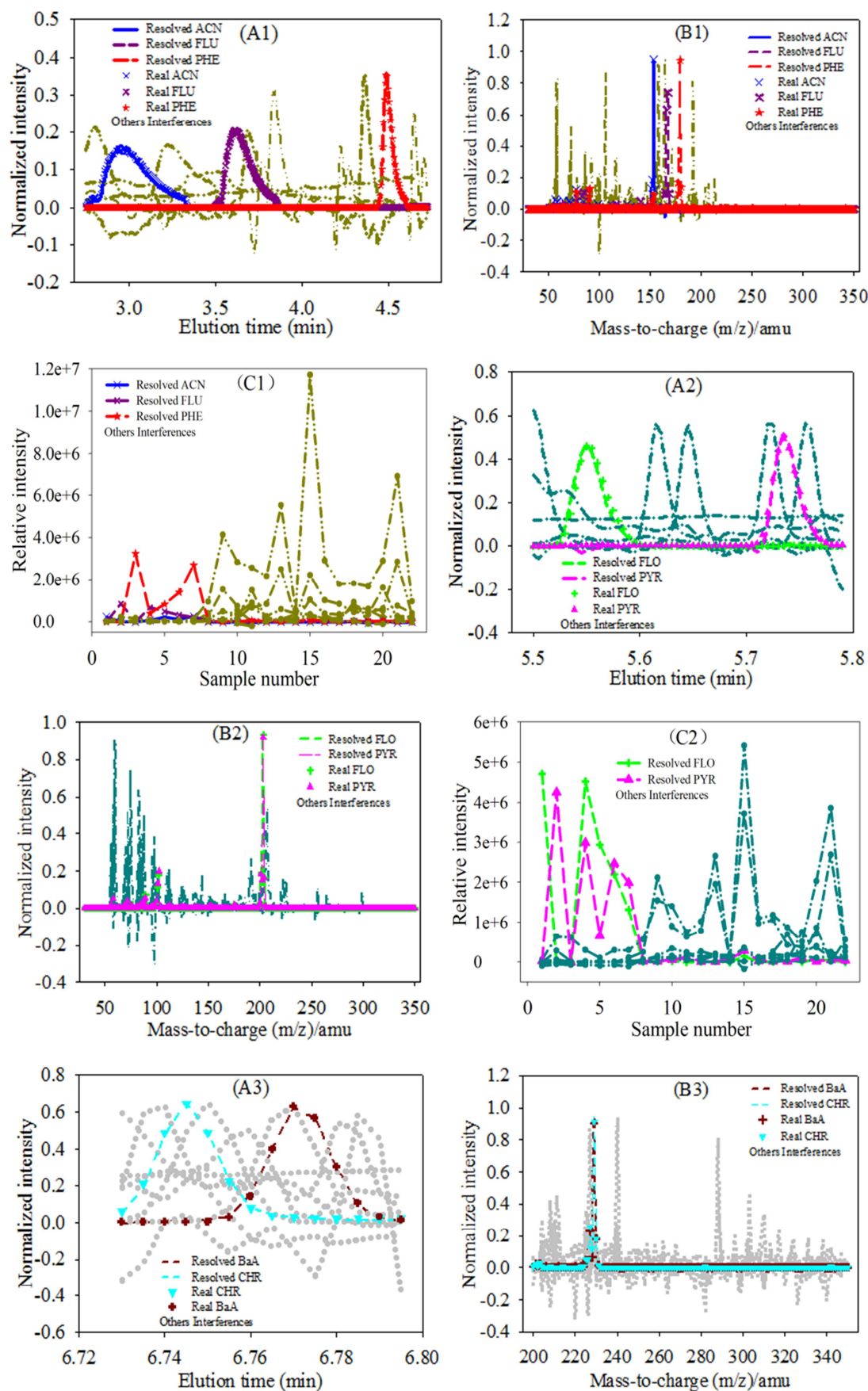
**Figure 3.** (A) Total ion current (TIC) chromatograms of all samples. (B) Raw GC-MS data of BbF in calibration samples was decomposed by the ATLD method with  $N=1$ , 2 and 3, respectively. (C) Raw TIC profiles of seven calibration samples, and their profiles aligned by ASSD method.

For the sake of providing accurate solutions and reducing data analysis and handling time, the elution region of 10 PAHs could be divided into four sub-segments, including 2.75-4.73, 5.50-5.79, 6.73-6.80 and 7.62-7.99 minutes, respectively. The analytes in these sub-segments overlapped with each other partially or fully on the retention time. In the fourth sub-segment, there were obvious retention-time shifts, which could be found in raw TIC profiles of Figure 3 (C). Additionally, when  $N=3$  in the example of BbF mentioned above, the shift component was also extracted out by the ATLD method (see the retention-time and sample-number plots of Figure 3 (B) ( $N=3$ )). Thus, the retention-time shift must be aligned before applying ATLD-SMR to analyze the three-way data array. Many chemometric strategies had been developed for aligning chromatographic peaks shifts among samples [31-33]. In the work, a named abstract subspace difference (ASSD) method was used to align the time shift occurring in the fourth sub-domain [34]. The aligned chromatographs of seven calibration and three test samples by ASSD method were shown in Figure 3 (C), which showed that the problem of retention-time shifts had been resolved successfully.

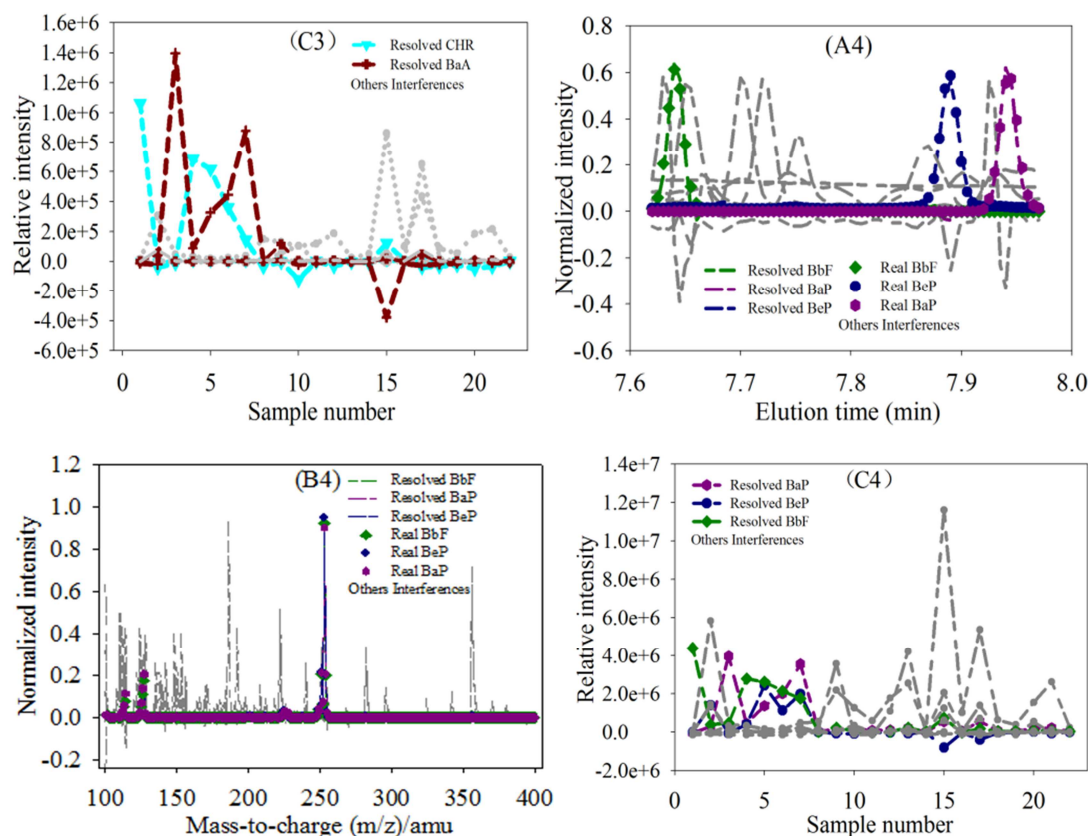
#### 4.2.2. Identification of Tracers in Source and Receptor Samples

During the progress of model construction, a three-way

dataset of twenty-two samples including seven calibration samples, ten source samples and five receptor samples was firstly divided into four sub-three-way data arrays according to four elution chromatographic segments mentioned above. A pictorial display of the identified results with the ATLD-SMR method was given in Figure 4. In the dimension of elution time, it was found that chromatograms of ten tracers are heavily overlapped with each other or unknown interference co-eluted. Even so, ATLD-SMR can still extract clear chromatographic profiles of ten tracers in all of samples. The resolved chromatographic profiles of ten tracers were in good agreement with their corresponding real ones, which were shown in Figure 4 (A1-A4). In the dimension of mass spectrum, besides overlapped with uncalibrated components, there were three pairs of tracers completely overlapped with each other, such as (1) FLO and PYR, (2) BaA and CHR, (3) BbF, BaP and BeP. However, the resolved mass spectra of ten tracers matched quite very well with the actual ones (shown in Figure 4 (B1-B4)). These results indicated the ATLD-SMR method had the accurate and reliable qualitative performance, which can be applied to identify tracers from complex environmental samples even in the presence of unknown and uncalibrated interference.







**Figure 4.** A pictorial display of the resolved results of 10 trackers in 22 samples by ATLD-SMR. A1-A4, normalized elution time fingerprints; B1-B4, normalized mass spectra fingerprints; C1-C4, relative concentration fingerprints.

#### 4.2.3. Quantification of Tracers in Source and Receptor Samples

Based on the above reliable qualitative results, a quantification of ten tracers in source and receptor samples can be obtained by the developed method. Table 3 listed the predicted concentrations of tracers in source and receptor samples. From the table we can see, the highest concentrations

of individual PAHs in these samples were for phenanthrene and benzo [a] pyrene ( $40.76 \text{ ng m}^{-3}$  and  $39.63 \text{ ng m}^{-3}$  in Liangang steel-making workshop, respectively). Then, the obtained concentrations of ten tracers can be applied to estimate the relative source contributions to different PM10 samples with the ATLD-SMR method in the following step.

**Table 3.** Predicted concentrations of PAHs in source and PM10 samples.

|                   | Samples | Predicted values/ng m <sup>-3</sup> |      |       |      |       |      |       |       |       |       |
|-------------------|---------|-------------------------------------|------|-------|------|-------|------|-------|-------|-------|-------|
|                   |         | ACN                                 | FLU  | PHE   | FLO  | PYR   | BaA  | CHR   | BbF   | BeP   | BaP   |
| Pollution Sources | SYB     | 15.51                               | 1.17 | 39.28 | 0.00 | 4.78  | 7.97 | 2.80  | 0.00  | 26.75 | 17.97 |
|                   | EG      | 0.00 <sup>a</sup>                   | 0.00 | 36.06 | 0.00 | 4.10  | 9.16 | 0.62  | 0.00  | 19.37 | 29.87 |
|                   | BSD     | 0.00                                | 0.28 | 6.90  | 0.00 | 1.88  | 1.57 | 0.88  | 0.00  | 2.59  | 5.87  |
|                   | NFY     | 0.00                                | 1.28 | 13.87 | 0.00 | 7.43  | 2.34 | 0.00  | 0.00  | 8.12  | 7.46  |
|                   | LGPW    | 0.00                                | 0.00 | 38.07 | 0.00 | 3.45  | 8.58 | 5.18  | 0.00  | 22.28 | 23.98 |
|                   | LGSW    | 0.00                                | 1.66 | 40.76 | 0.00 | 3.95  | 9.99 | 3.49  | 0.00  | 18.11 | 39.63 |
|                   | BJMD    | 0.00                                | 0.46 | 36.53 | 0.00 | 11.76 | 8.77 | 3.29  | 0.00  | 27.20 | 15.73 |
|                   | MED     | 0.00                                | 2.27 | 13.55 | 2.19 | 12.65 | 0.00 | 10.90 | 10.32 | 0.00  | 19.11 |
|                   | AED     | 0.00                                | 0.33 | 10.21 | 0.00 | 1.48  | 2.13 | 1.31  | 0.00  | 6.04  | 5.07  |
| PM10              | TED     | 0.00                                | 0.00 | 9.58  | 0.00 | 1.15  | 1.33 | 1.83  | 0.00  | 0.00  | 19.34 |
|                   | BW      | 14.05                               | 2.86 | 31.34 | 0.00 | 3.43  | 6.42 | 0.54  | 0.00  | 19.43 | 17.15 |
|                   | RWS     | 0.00                                | 0.19 | 4.06  | 0.00 | 1.85  | 0.92 | 0.12  | 0.00  | 2.64  | 1.99  |
|                   | LG      | 13.40                               | 0.93 | 4.26  | 0.00 | 0.52  | 0.99 | 0.00  | 0.00  | 3.14  | 3.29  |
|                   | MG      | 0.00                                | 0.77 | 38.00 | 0.00 | 9.54  | 8.01 | 3.38  | 0.00  | 19.20 | 37.17 |
|                   | SY      | 0.00                                | 0.02 | 1.20  | 0.00 | 0.27  | 0.28 | 0.12  | 0.00  | 0.87  | 0.57  |

<sup>a</sup> The predicted values which are less than or equal to zero are set to be equal to zero.

#### 4.2.4. The Estimation of the Source Contribution to PM10

According to the obtained concentrations of ten tracers in

the source and receptor samples, the relative contribution matrix of **P** can be calculated with the ATLD-SMR method.

Wherein, the concentrations of ten tracers in the source samples were used as the score matrix of sources, and these in the receptor samples as the score matrix of receptors. Then, we obtained the relative contribution of each single source to the PM10 samples. They are listed in Table 4. At site BWPM, the main pollution sources were AED (73.35%), SYB (9.93%) and LGSW (6.19%). At site RWSPM, MGPM, and SYPM, the

contribution of AED was also the highest, accounting for 58.25%, 38.58% and 47.32%, respectively. Only for site LGPM, the main pollution sources were BSD (35.05%), SYB (21.29%) and AED (17.19%). These results showed that AED was the principal single pollution source to all measured PM10 samples except for LGPM.

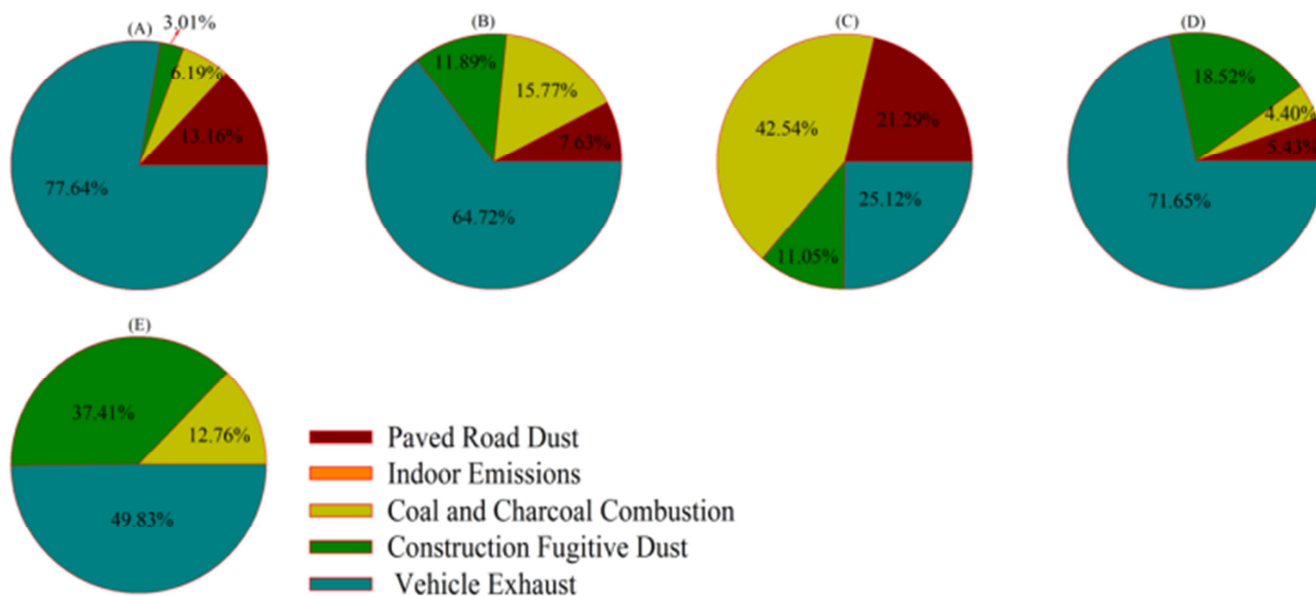
**Table 4.** Relative contributions of single source to different PM10 samples with ATLD-SMR.

| Samples | Relative contributions/% |      |       |      |      |       |       |      |       |       |
|---------|--------------------------|------|-------|------|------|-------|-------|------|-------|-------|
|         | SYB                      | EG   | BSD   | NFY  | LGPW | LGSW  | BJMD  | MED  | AED   | TED   |
| BWPM    | 9.93                     | 3.23 | 0.00  | 0.00 | 0.00 | 6.19  | 3.01  | 0.00 | 73.35 | 4.29  |
| RWSPM   | 0.00                     | 7.63 | 15.77 | 0.00 | 0.00 | 0.00  | 11.89 | 0.00 | 58.25 | 6.46  |
| LGPM    | 21.29                    | 0.00 | 35.05 | 0.00 | 0.00 | 7.49  | 11.05 | 0.00 | 17.19 | 7.93  |
| MGPM    | 0.00                     | 5.43 | 0.00  | 0.00 | 0.00 | 4.40  | 18.52 | 0.00 | 38.58 | 33.07 |
| SYPM    | 0.00                     | 0.00 | 0.00  | 0.00 | 1.14 | 11.62 | 37.41 | 0.00 | 47.32 | 2.51  |

In order to apportion the sources of PM10 better, 10 pollution sources were classified into five types, including paved road dust (SYB and EG), indoor emissions (NFY), coal and charcoal combustion (BSD, LGPW and LGSW), construction fugitive dust (BJMD) and vehicle exhaust (MED, AED and TED). A diagrammatic representation of the combined contributions is provided in Figure 5. As shown in Figure 5, vehicle exhaust was the most significant pollution source at the sampling point of the Biwu building, which was near a main urban road, contributing 77.64% of the PM10. Paved road dust provided the second highest contribution, accounting for 13.16%. Industrial coal combustion and construction fugitive dust accounted for 6.19% and 3.01%, respectively.

For the railway station point, vehicle exhaust was also the main contributor, accounting for 64.72%, which is easy to understand for a heavy traffic area, and then followed by the

contributions that coal and charcoal combustion (15.77%), construction fugitive dust (11.89%) and paved road dust (7.63%) gave (Figure 5 (B)). For the Liangang industrial area (Figure 5 (C)), coal and charcoal combustion, vehicle exhaust and paved road dust provided the main contributions to the PM10, accounting for 42.54%, 25.12% and 21.29%, respectively. For the municipal government and Shuyi building (Figure 5 (D) and 5(E)), the contribution of vehicle exhaust was the highest as 71.65% and 49.83%, respectively. These results showed that vehicle exhaust was the main contributor for all sampling points except for Liangang industrial area in the studied city. This can be easily explained as Loudi is a city with few coal-burning industries but the use of family car increases sharply in the city in recent years. Also, all the contributions estimated by the proposed method conformed to the actual situation in these five areas in the studied city.



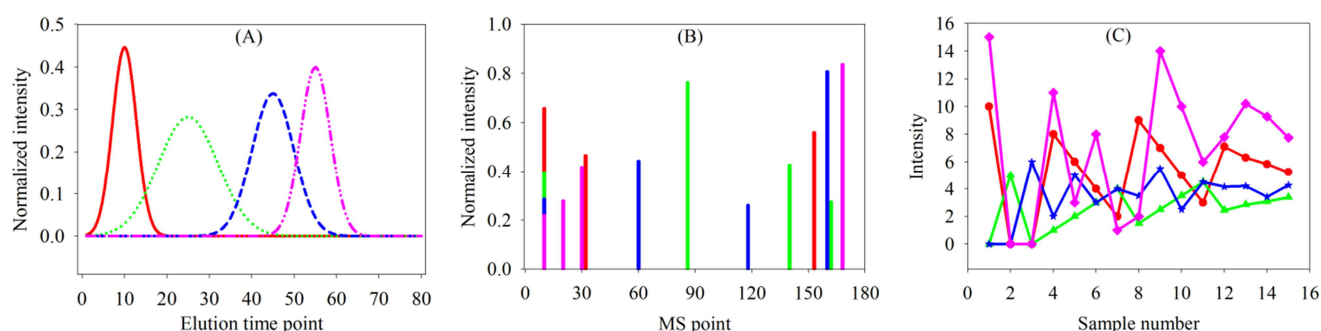
**Figure 5.** Relative source contributions to various PM10 samples in an urban area. (A) Biwu building, (B) railway station, (C) Liangang industrial area, (D) municipal government and (E) Shuyi building.

## 5. Conclusion

In the work, the procedure based on the alternating trilinear decomposition followed by a score matrix reconstruction (ATLD-SMR) of three-way GC-MS data of source and receptor samples allowed the source apportionment of PM<sub>10</sub> in different urban locations. The principal of the developed method was based on firstly decomposing the combined three-way dataset of source and receptor samples by ATLD algorithm, then calculating the relative contribution matrix according to the reconstructed score matrix of tracers extracted from source and receptor scores. The contributions of different sources as well as chromatographic and mass spectral profiles of ten tracers were obtained using the

procedure. From the obtained contributions, it was found that vehicle exhaust was the principal pollution source to all measured PM<sub>10</sub> mass except for Liangang. At Liangang, the main pollution source was industrial coal and charcoal combustion. The new ATLD-SMR method can also address other three-way data such as those produced by excitation-emission matrix fluorescence or phosphorescence, high-performance liquid chromatography-diode array detection, liquid chromatography-mass spectrometry, etc. It should be applied in other fields, such as the source apportionment of water and soil pollution, in the near future. Therefore, the newly developed method is a promising tool for source apportionment of environmental organic pollutants.

## Appendix



**Figure A1.** A plot of the normalized chromatogram (A), mass spectrogram (B) and concentration profiles (C) of the simulated data.

**Table A1.** Information of source and receptor samples.

| Number | Sample | Sampling Category | Sampling Site                             |
|--------|--------|-------------------|---|
| 1      | MED    | TSP               | Motorcycle Exhaust Dust                   |
| 2      | AED    | TSP               | Automobile Exhaust Dust                   |
| 3      | TED    | TSP               | Truck Exhaust Dust                        |
| 4      | SYB    | TSP               | Raise Dust near Shuyi Building            |
| 5      | EG     | TSP               | Raise Dust at the East Gate of school     |
| 6      | BSD    | TSP               | Raise Dust at BBQ Street                  |
| 7      | NFY    | TSP               | Newly decorated room at Nanfeng Community |
| 8      | LGPW   | TSP               | Liangang Power Workshop                   |
| 9      | LGSW   | TSP               | Liangang Steel-making Workshop            |
| 10     | BJMD   | TSP               | Construction site at Binjiang Mingdu      |
| 11     | BWPM   | PM <sub>10</sub>  | Biwu Building                             |
| 12     | RWSPM  | PM <sub>10</sub>  | Railway Station                           |
| 13     | LGPM   | PM <sub>10</sub>  | North Bridge at Liangang                  |
| 14     | MGPM   | PM <sub>10</sub>  | Municipal Government                      |
| 15     | SYPM   | PM <sub>10</sub>  | Shuyi Building                            |

**Table A2.** Concentrations of 10 PAHs in seven calibration samples (C01-C07).

| Sample No. | Analyte concentration (μg mL <sup>-1</sup> ) |      |      |      |      |      |      |      |      |      |
|------------|--|------|------|------|------|------|------|------|------|------|
|            | ACN  | FLU  | PHE  | FLO  | PYR  | BaA  | CHR  | BbF  | BeP  | BaP  |
| C01        | 6.70   | 0.00 | 0.00 | 5.60 | 0.00 | 0.00 | 3.42 | 0.00 | 0.00 | 6.70 |
| C02        | 0.00   | 2.69 | 0.00 | 0.00 | 6.40 | 0.00 | 0.00 | 3.60 | 0.00 | 0.00 |
| C03        | 0.00   | 0.00 | 4.90 | 0.00 | 0.00 | 0.68 | 0.00 | 0.00 | 4.20 | 0.00 |
| C04        | 1.34   | 2.15 | 0.98 | 4.48 | 4.99 | 0.14 | 2.85 | 0.72 | 0.84 | 1.34 |
| C05        | 5.36   | 1.61 | 1.96 | 3.36 | 1.24 | 0.27 | 2.28 | 2.88 | 1.68 | 5.36 |
| C06        | 2.68   | 1.08 | 2.94 | 2.24 | 3.84 | 0.41 | 1.71 | 1.44 | 2.52 | 2.68 |
| C07        | 4.02   | 0.54 | 3.92 | 1.12 | 2.56 | 0.54 | 1.14 | 2.16 | 3.36 | 4.02 |

**Table A3.** Contributions predicted by the developed method in analysis of a simulated data set with different homoscedastic noise.

| Receptor | Source | True Contribution (%) | Predicted contribution (%) |        |        |        |        |
|----------|--------|-----------------------|----------------------------|--------|--------|--------|--------|
|          |        |                       | 0.000                      | 0.001  | 0.002  | 0.005  | 0.01   |
| R1       | S1     | 0.4000                | 0.4000                     | 0.3996 | 0.3976 | 0.3952 | 0.3981 |
|          | S2     | 0.3500                | 0.3500                     | 0.3458 | 0.3465 | 0.3550 | 0.3432 |
|          | S3     | 0.1500                | 0.1500                     | 0.1536 | 0.1547 | 0.1477 | 0.1501 |
|          | S4     | 0.1000                | 0.1000                     | 0.1025 | 0.1014 | 0.1008 | 0.1073 |
| R2       | S1     | 0.1500                | 0.1500                     | 0.1503 | 0.1487 | 0.1452 | 0.1624 |
|          | S2     | 0.4500                | 0.4500                     | 0.4487 | 0.4458 | 0.4597 | 0.4580 |
|          | S3     | 0.3000                | 0.3000                     | 0.2998 | 0.3042 | 0.2870 | 0.2808 |
|          | S4     | 0.1000                | 0.1000                     | 0.1008 | 0.1023 | 0.1066 | 0.0974 |
| R3       | S1     | 0.1500                | 0.1500                     | 0.1502 | 0.1480 | 0.1531 | 0.1438 |
|          | S2     | 0.2000                | 0.2000                     | 0.1988 | 0.2078 | 0.2039 | 0.1956 |
|          | S3     | 0.5700                | 0.5700                     | 0.5701 | 0.5605 | 0.5698 | 0.5749 |
|          | S4     | 0.0800                | 0.0800                     | 0.0819 | 0.0777 | 0.0748 | 0.0863 |
| R4       | S1     | 0.1700                | 0.1700                     | 0.1697 | 0.1643 | 0.1703 | 0.1640 |
|          | S2     | 0.2300                | 0.2300                     | 0.2317 | 0.2298 | 0.2415 | 0.2318 |
|          | S3     | 0.1500                | 0.1500                     | 0.1469 | 0.1479 | 0.1409 | 0.1347 |
|          | S4     | 0.4500                | 0.4500                     | 0.4502 | 0.4556 | 0.4440 | 0.4683 |

**Table A4.** The related coefficients obtained by the developed method for the analysis of simulated data sets with different homoscedastic noise.

|    | contribution | 0.000               | 0.001  | 0.002  | 0.005  | 0.01   |
|----|--------------|---------------------|--------|--------|--------|--------|
| R1 | 0.4000       | 1.0000 <sup>a</sup> | 0.9991 | 0.9940 | 0.9880 | 0.9953 |
|    | 0.3500       | 1.0000              | 0.9880 | 0.9899 | 1.0144 | 0.9805 |
|    | 0.1500       | 1.0000              | 1.0242 | 1.0316 | 0.9849 | 1.0005 |
|    | 0.1000       | 1.0000              | 1.0252 | 1.0144 | 1.0075 | 1.0728 |
| R2 | 0.1500       | 1.0000              | 1.0019 | 0.9916 | 0.9679 | 1.0826 |
|    | 0.4500       | 1.0000              | 0.9972 | 0.9907 | 1.0216 | 1.0178 |
|    | 0.3000       | 1.0000              | 0.9994 | 1.0140 | 0.9566 | 0.9361 |
|    | 0.1000       | 1.0000              | 1.0076 | 1.0227 | 1.0660 | 0.9741 |
| R3 | 0.1500       | 1.0000              | 1.0013 | 0.9870 | 1.0203 | 0.9587 |
|    | 0.2000       | 1.0000              | 0.9941 | 1.0392 | 1.0196 | 0.9781 |
|    | 0.5700       | 1.0000              | 1.0001 | 0.9834 | 0.9997 | 1.0086 |
|    | 0.0800       | 1.0000              | 1.0240 | 0.9716 | 0.9349 | 1.0792 |
| R4 | 0.1700       | 1.0000              | 0.9982 | 0.9663 | 1.0019 | 0.9647 |
|    | 0.2300       | 1.0000              | 1.0073 | 0.9992 | 1.0501 | 1.0077 |
|    | 0.1500       | 1.0000              | 0.9793 | 0.9863 | 0.9396 | 0.8981 |
|    | 0.4500       | 1.0000              | 1.0005 | 1.0124 | 0.9866 | 1.0408 |

<sup>a</sup>1.0000 is the related coefficient between the predicted and the true value.**Table A5.** Contributions predicted by the developed method in analysis of a simulated data set with different heteroscedastic noise.

| Receptor | Source | True Contribution (%) | Predicted contribution (%) |        |        |        |
|----------|--------|-----------------------|----------------------------|--------|--------|--------|
|          |        |                       | 0.001                      | 0.002  | 0.005  | 0.01   |
| R1       | S1     | 0.4000                | 0.3984                     | 0.4012 | 0.4017 | 0.3982 |
|          | S2     | 0.3500                | 0.3504                     | 0.3409 | 0.3454 | 0.3538 |
|          | S3     | 0.1500                | 0.1498                     | 0.1626 | 0.1501 | 0.1405 |
|          | S4     | 0.1000                | 0.1002                     | 0.1000 | 0.1045 | 0.1053 |
| R2       | S1     | 0.1500                | 0.1513                     | 0.1493 | 0.1586 | 0.1567 |
|          | S2     | 0.4500                | 0.4480                     | 0.4501 | 0.4464 | 0.4145 |
|          | S3     | 0.3000                | 0.3016                     | 0.2991 | 0.2929 | 0.3509 |
|          | S4     | 0.1000                | 0.0997                     | 0.0988 | 0.1119 | 0.1006 |
| R3       | S1     | 0.1500                | 0.1474                     | 0.1520 | 0.1490 | 0.1456 |
|          | S2     | 0.2000                | 0.2000                     | 0.2008 | 0.1927 | 0.2075 |
|          | S3     | 0.5700                | 0.5692                     | 0.5703 | 0.5667 | 0.5591 |
|          | S4     | 0.0800                | 0.0835                     | 0.0770 | 0.0948 | 0.0994 |
| R4       | S1     | 0.1700                | 0.1689                     | 0.1718 | 0.1722 | 0.1783 |
|          | S2     | 0.2300                | 0.2299                     | 0.2242 | 0.2421 | 0.2518 |
|          | S3     | 0.1500                | 0.1500                     | 0.1564 | 0.1322 | 0.1226 |
|          | S4     | 0.4500                | 0.4512                     | 0.4494 | 0.4510 | 0.4466 |

**Table A6.** The related coefficients obtained by the developed method for the analysis of simulated data sets with different heteroscedastic noise.

|    | contribution | 0.001  | 0.002  | 0.005  | 0.01   |
|----|--------------|--------|--------|--------|--------|
| R1 | 0.4000       | 0.9959 | 1.0029 | 1.0042 | 0.9955 |
|    | 0.3500       | 1.0011 | 0.9740 | 0.9869 | 1.0109 |
|    | 0.1500       | 0.9985 | 1.0840 | 1.0010 | 0.9366 |
|    | 0.1000       | 1.0020 | 1.0005 | 1.0446 | 1.0530 |
| R2 | 0.1500       | 1.0085 | 0.9954 | 1.0574 | 1.0445 |
|    | 0.4500       | 0.9955 | 1.0003 | 0.9921 | 0.9211 |
|    | 0.3000       | 1.0054 | 0.9969 | 0.9765 | 1.1697 |
|    | 0.1000       | 0.9969 | 0.9877 | 1.1187 | 1.0063 |
| R3 | 0.1500       | 0.9830 | 1.0132 | 0.9936 | 0.9703 |
|    | 0.2000       | 0.9998 | 1.0038 | 0.9636 | 1.0377 |
|    | 0.5700       | 0.9986 | 1.0006 | 0.9942 | 0.9808 |
|    | 0.0800       | 1.0433 | 0.9626 | 1.1848 | 1.2430 |
| R4 | 0.1700       | 0.9933 | 1.0108 | 1.0131 | 1.0489 |
|    | 0.2300       | 0.9998 | 0.9749 | 1.0526 | 1.0949 |
|    | 0.1500       | 1.0002 | 1.0428 | 0.8812 | 0.8171 |
|    | 0.4500       | 1.0027 | 0.9986 | 1.0023 | 0.9925 |

## Acknowledgements

The authors gratefully acknowledge the National Natural Science Foundation of China (Grant No. 21707032, 31701689), the Research Foundation of Education Bureau of Hunan Province, China (Grant No. 16A109) and Hunan Provincial Key Lab of Dark Tea and Jin-hua (2016TP1022) for financial supports.

## References

- [1] Q. Mu, M. Shiraiwa, M. Octaviani, N. Ma, A. Ding, H. Su, G. Lammel, U. Pöschl, Y. Cheng, Temperature effect on phase state and reactivity controls atmospheric multiphase chemistry and transport of PAHs, *Sci. Adv.* 4 (2018) 7314-7321.
- [2] Y. Zhang, H. Zheng, L. Zhang, Z. Z. Zhang, X. L. Xing, S. H. Qi, Fine particle-bound polycyclic aromatic hydrocarbons (PAHs) at an urban site of Wuhan, central China: Characteristics, potential sources and cancer risks apportionment, *Environ. Pollut.* 246 (2019): 319-327.
- [3] X. W. Li, X. Zhang, Z. Q. Zhang, L. Y. Han, D. P. Gong, J. Li, T. Wang, Y. H. Wang, S. Gao, H. W. Duan, F. L. Kong, Air pollution exposure and immunological and systemic inflammatory alterations among schoolchildren in China, *Sci. Total Environ.* 657 (2019) 1304-1310.
- [4] D. G. Streets, S. Gupta, S. T. Waldhoff, M. Q. Wang, T. C. Bond, Y. Bo, Black carbon emissions in China, *Atmos. Environ.* 35 (2001) 4281-4296.
- [5] S. Reis, R. W. Pinder, M. Zhang, G. Lijie, M. A. Sutton, Reactive nitrogen in atmospheric emission inventories, *J. Geophys. Res: Atmos.* 9 (2009) 7657-7677.
- [6] M. J. Kleeman, G. R. Cass, A. Eldering, Modeling the airborne particle complex as a source-oriented external mixture, *J. Geophys. Res: Atmos.* 102 (1997) 21355-21372.
- [7] P. K. Hopke, Recent developments in receptor modeling, *J. Chemom.* 17 (2003) 255-265.
- [8] C. A. Belis, D. Pernigotti, F. Karagulian, G. Pirovano, B. R. Larsen, M. Gerboles, P. K. Hopke, A new methodology to assess the performance and uncertainty of source apportionment models in intercomparison exercises, *Atmos. Environ.* 119 (2015) 35-44.
- [9] C. A. Belis, F. Karagulian, F. Amato, M. Almeida, P. Artaxo, D. C. S. Beddows, V. Bernardoni, M. C. Bove, S. Carbone, D. Cesari, D. Contini, E. Cuccia, E. Diapouli, K. Eleftheriadis, O. Favez, I. El Haddad, R. M. Harrison, S. Hellebust, J. Hovorka, E. Jang, H. Jorquera, T. Kammermeier, M. Karl, F. Lucarelli, D. Mooibroek, S. Nava, J. K. Nøjgaard, P. Paatero, M. Pandolfi, M. G. Perrone, J. E. Petit, A. Pietrodangelo, P. Pokorný, P. Prati, A. S. H. Prevot, U. Quass, X. Querol, D. Saraga, J. Sciare, A. Sfetsos, G. Valli, R. Vecchi, M. Vestenius, E. Yubero, P. K. Hopke, A new methodology to assess the performance and uncertainty of source apportionment models II: The results of two European intercomparison exercises, *Atmos. Environ.* 123 (2015) 240-250.
- [10] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37-52.
- [11] W. C. Ma, L. Y. Tai, Z. Qiao, L. Zhong, Z. Wang, K. X. Fu, G. Y. Chen, Contamination source apportionment and health risk assessment of heavy metals in soil around municipal solid waste incinerator: A case study in North China, *Sci. Total Environ.* 631-632 (2018) 348-351.
- [12] B. Wu, D. Zhao, Y. Zhang, X. Zhang, S. Cheng, Multivariate statistical study of organic pollutants in Nanjing reach of Yangtze River, *J. Hazard. Mater.* 169 (2009) 1093-1098.
- [13] I. Salim, R. U. Sajjad, M. C. Paule-Mercado, S. A. Memon, B. Y. Lee, C. Sukhbaatar, C. H. Lee, Comparison of two receptor models PCA-MLR and PMF for source identification and apportionment of pollution carried by runoff from catchment and sub-watershed areas with mixed land cover in South Korea, *Sci. Total Environ.* 663 (2019) 764-775.
- [14] I. Stanimirova, R. Tauler, B. Walczak, A comparison of positive matrix factorization and the weighted multivariate curve resolution method. Application to environmental data, *Environ. Sci. Tech.* 45 (2011) 10102-10110.
- [15] M. S. Miller, S. K. Friedlander, G. M. Hidy, A chemical element balance for the Pasadena aerosol, *J. Colloid Interf. Sci.* 39 (1972) 165-176.



- [16] J. A. Cooper, Review of the chemical receptor model of aerosol source apportionment, in: *Atmospheric Aerosol*, American Chemical Society, 1981, pp. 75-87.
- [17] H. Wallace, N. P. Sanchez, J. Flynn, M. H. Erickson, Source apportionment of particulate matter and trace gases near a major refinery near the Houston Ship Channel, *Atmos. Environ.* 173 (2018) 16-29.
- [18] S. M. Mudge, C. Bravo-Linares, L. Ovando-Fuentealba, J. P. Pinaud-Mendoza, A comparison between three unmixing models for source apportionment of PM<sub>2.5</sub> using alkanes in air from Southern Chile, *Environ. Forensics* 18 (2017) 226-240.
- [19] Z. P. Zhao, S. Lv, Y. H. Zhang, Q. B. Zhao, L. Shen, S. Xu, J. Q. Yu, J. W. Hou, C. Y. Jin, Characteristics and source apportionment of PM<sub>2.5</sub> in Jiaxing, China, *Environ. Sci. Pollut. Res.* 26 (2019) 7497-7511.
- [20] D. A. Olson, G. A. Norris, M. S. Landis, A. F. Vette, Chemical characterization of ambient particulate matter near the world trade center: elemental carbon, organic carbon, and mass reconstruction, *Environ. Sci. Tech.* 38 (2004) 4465.
- [21] G. R. Cass, Organic molecular tracers for particulate air pollution sources, *Trends Anal. Chem.* 17 (1998) 356-366.
- [22] M. R. Alcaraz, O. Monago-Maraña, H. C. Goicoechea, A. Muñoz de la Peña, Four- and five-way excitation-emission luminescence-based data acquisition and modeling for analytical applications. A review, *Anal. Chim. Acta* 1083 (2019) 41-57.
- [23] X. D. Qing, H. L. Wu, X. H. Zhang, Y. Li, H. W. Gu, R. Q. Yu, A novel fourth-order calibration method based on alternating quinquelinear decomposition algorithm for processing high performance liquid chromatography-diode array detection-kinetic-pH data of naptalam hydrolysis, *Anal. Chim. Acta* 861 (2015) 12-24.
- [24] S. M. Azcarate, A. de Araújo Gomes, A. Muñoz de la Peña, H. C. Goicoechea, Modeling second-order data for classification issues: Data characteristics, algorithms, processing procedures and applications, *Trends Anal. Chem.* 107 (2018) 151-168.
- [25] G. L. Shi, X. Peng, Y. Q. Huangfu, W. Wang, J. Xu, Y. Z. Tian, Y. C. Feng, C. E. Ivey, A. G. Russell, Quantification of source impact to PM using three-dimensional weighted factor model analysis on multi-site data, *Atmos. Environ.* 160 (2017) 89-96.
- [26] A. A. Karanasiou, P. A. Siskos, K. Eleftheriadis, Assessment of source apportionment by Positive Matrix Factorization analysis on fine and coarse urban aerosol size fractions, *Atmos. Environ.* 43 (2009) 3385-3395.
- [27] M. H. Nie, X. Huang, C. X. Yan, Y. Yang, J. L. Zhou, M. Liu, Fluorescence characterization of fractionated colloids in different sources of waters based on PARAFAC and SOM, *Acta Scien. Circum.* 38 (2018) 3672-3681.
- [28] H. L. Wu, M. Shibukawa, K. Oguma, An alternating trilinear decomposition algorithm with application to calibration of HPLC-DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons, *J. Chemom.* 12 (1998) 1-26.
- [29] Z. Liu, H. L. Wu, L. X. Xie, Y. Hu, H. Fang, X. D. Sun, T. Wang, R. Xiao, R. Q. Yu, Chemometrics-enhanced liquid chromatography-full scan-mass spectrometry for interference-free analysis of multi-class mycotoxins in complex cereal samples, *Chemom. Intell. Lab. Syst.* 160 (2017) 125-138.
- [30] Y. Zhang, H. L. Wu, A. L. Xia, L. H. Hu, H. F. Zou, R. Q. Yu, Trilinear decomposition method applied to removal of three-dimensional background drift in comprehensive two-dimensional separation data, *J. Chromatogr. A* 1167 (2007) 178-183.
- [31] M. Li, X. R. Wang, Peak alignment of gas chromatography-mass spectrometry data with deep learning, *J. Chromatogr. A* (2019).
- [32] T. B. Yang, P. Yan, M. He, L. Hong, R. Pei, Z. M. Zhang, L. Z. Yi, X. Y. Yuan, Application of Subwindow Factor Analysis and Mass Spectral information for accurate alignment of non-targeted metabolic profiling, *J. Chromatogr. A* 1563 (2018) 162-170.
- [33] S. Abou-el-karam, J. Ratel, N. Kondjoyan, C. Truan, E. Engel, Marker discovery in volatolomics based on systematic alignment of GC-MS signals: Application to food authentication, *Anal. Chim. Acta* 991 (2017) 58-67.
- [34] Y. J. Yu, H. L. Wu, J. F. Niu, J. Zhao, Y. N. Li, C. Kang, R. Q. Yu, A novel chromatographic peak alignment method coupled with trilinear decomposition for three dimensional chromatographic data analysis to obtain the second-order advantage, *Analyst*, 138 (2012) 627-634.