Atmospheric Environment 45 (2011) 2811–2819

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Estimated contributions and uncertainties of PCA/MLR–CMB results: Source apportionment for synthetic and ambient datasets

Guo-Liang Shi [a], Fang Zeng [a], Xiang Li [a,b], Yin-Chang Feng [a,*], Yu-Qiu Wang [a], Guang-Xun Liu [a], Tan Zhu [a]

[a] State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Weijin Road 94, Tianjin 300071, China
[b] Department of Computer Science, University of Georgia, Athens, GA, USA

ARTICLE INFO

Article history:
Received 8 October 2010
Received in revised form
21 February 2011
Accepted 2 March 2011

Keywords:
PCA/MLR–CMB model
Source apportionment
Collinearity problem
Uncertainty
Nonnegative Constrained Principal Component Regression Chemical Mass Balance Model

ABSTRACT

The instability of a receptor model due to nearly collinear sources is often worsened by a large number of unknown sources, which usually results in unacceptable source contributions obtained by a single-stage model. To solve this problem, a combined (principal component analysis/multiple linear regression–chemical mass balance; PCA/MLR–CMB) model comprising two stages has been developed. In this study, synthetic datasets with a serious collinearity problem were generated to evaluate the performance of the combined model, and acceptable results were obtained in the presence of noise. Additional, the uncertainties of the estimated contributions were discussed. It was found that nearly collinear sources usually got higher uncertainties. Ambient data from Chengdu were then studied using both the PCA/MLR–CMB and NCPCRCMB models. For PCA/MLR–CMB model: vehicle exhaust emissions contributed 80.46 μg m^{-3} (28.71%) to the total PM_{10}; coal combustion got 68.52 μg m^{-3} (24.45%); resuspended dust, soil dust, secondary sulfate, secondary nitrate, cement dust and smelters got 53.91 μg m^{-3} (19.24%), 46.31 μg m^{-3} (16.53%), 33.35 μg m^{-3} (11.90%), 17.65 μg m^{-3} (6.30%), 12.47 μg m^{-3} (4.45%) and 1.81 μg m^{-3} (0.65%), respectively. For NCPCRCMB model, the results were: vehicle—86.70 μg m^{-3} (30.94%), cement—49.47 μg m^{-3} (17.65%), resuspended dust—41.22 μg m^{-3} (14.71%), coal—30.43 μg m^{-3} (10.86%), soil—27.45 μg m^{-3} (9.80%), secondary sulfate—16.41 μg m^{-3} (5.86%) and secondary nitrate—3.61 μg m^{-3} (1.29%).

1. Introduction

Understanding potential sources of the particulate matter in the urban atmosphere is important for air quality management (Mazzei et al., 2008). To identify sources, several receptor models have been developed (Watson et al., 2002). Chemical mass balance (CMB) model is a useful receptor model that uses the chemical characteristics of particles measured at the sources and receptor to quantify source contributions (Chow and Watson, 2002; Chow et al., 2007; Shi et al., 2009a; Stone et al., 2010). Principal component analysis/multiple linear regression (PCA/MLR) model, another important receptor model, does not require a priori knowledge of the number of sources or their compositions and instead uses sufficient concentration data from ambient samples (Thurston and Spengler, 1985; Cao et al., 2005; Hellebust et al., 2010). In addition, other important receptor models such as that based on positive matrix factorization (Paatero, 1997; Reff et al., 2007; Ulbrich et al., 2009; Chen et al., 2010; Henry and Christensen, 2010) and the Unmix model (Henry, 2003; Song et al., 2007; Henry and Christensen, 2010) have been used widely.

Each receptor has its strengths and weaknesses (Watson et al., 2008). The weakness of the CMB model is that a completely compatible source and receptor measurement is not commonly available (Watson et al., 2008). In addition, the collinearity problem may also disturb the modeled outcomes. Generally, the condition indexes (CI) and variance-decomposition propositions (VDP) values can be applied to identify the nearly collinear sources (Hopke, 1985). On another hand, the PCA/MLR model might fail to separate factors (sources). Some studies found that one factor (extracted by the PCA/MLR model) might relate to two or more source categories (Lioy et al., 1989; Okamoto et al., 1990; Harrison et al., 1996; Guo et al., 2004; Mamane et al., 2008).

In urban areas in China, some important sources have similar profiles, such as urban resuspended dust (URD), soil dust and coal combustion (Zhao et al., 2006). The similarity of these source profiles usually affects the results of the receptor models. If such

* Corresponding author. Tel./fax: +86 2223503397.
E-mail addresses: nksgl@hotmail.com (G.-L. Shi), fengyc@nankai.edu.cn (Y.-C. Feng).

1352-2310/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2011.03.007

data are apportioned by the CMB model, negative estimated contributions of these nearly collinear sources might be obtained. Or if they were apportioned by the PCA/MLR model, one factor containing these nearly collinear sources would be extracted.

We have presented a two-stage combined source apportionment technique, taking the advantages of the PCA/MLR and CMB models, and applied it to the study of contributions of nearly collinear sources to urban ambient particulate matter in our previous study (Shi et al., 2009b). In this study, synthetic datasets with a serious collinearity problem were developed and introduced into combined model. We want to test the accuracy of the results and their uncertainties of the combined receptor model, when serious collinearity problem exists. Next, the combined model was applied to study the contributions of nearly collinear sources in Chengdu.

2. Methodology of combined source apportionment

CMB model is based on the balance of chemical species masses among sources and receptors (US EPA, 2004). However, this balance might be disturbed. Firstly, besides the major sources (such as soil, coal and vehicles), there are thousands of insignificant sources that contribute to the ambient receptor measured. The contribution of each insignificant source to the ambient receptor is very small and usually ignored by the CMB model. However, the sum of them cannot be ignored and does disturb the balance. Secondly, uncertainties in the sampling and chemical analysis can disturb the balance. In this study, the sum of the disturbances to the balance due to these two factors is referred as noise.

According to our previous study (Shi et al., 2009c), with a low level of noise presenting at the receptor, CMB model can obtain acceptable results even for nearly collinear sources (which will be confirmed in the next section). Therefore, the PCA/MLR–CMB model aims at reducing the level of noise at the receptor and consequently weakens the collinearity problem. The method employed by the PCA/MLR–CMB model was described in our previous study (Shi et al., 2009b). The PCA/MLR–CMB model comprises three stages.

2.1. Stage 1: reducing noise from the original receptor by the PCA/MLR model

In stage 1, several factors identified as potential sources according to source markers (Hopke, 1985; Harrison et al., 1996; Hedberg et al., 2005) can be extracted from receptor (here is original receptor) using the PCA/MLR model. In this study, the actual source profiles measured in the real world are referred to as measured-source (M-source) profiles, while source profiles extracted by PCA/MLR model are referred to as extracted-source (E-source) profiles. If one factor can be identified as one source category (such as vehicle exhaust), it is referred to as an extracted simplex source (ES-source), whereas if it contains two or more source categories, it is referred to as an extracted complex source (EC-source). If a factor cannot be identified as any of the source category, it might be noise. In this process, the contributions of the EC-source, ES-source and noise can be described as follows.

$$\text{original receptor} = \text{EC}-\text{source}_1 + \text{EC}-\text{source}_2 + \ldots + \text{ES} - \text{source}_1 + \text{ES}-\text{source}_2 + \ldots + \text{noise} \quad (1)$$

2.2. Stage 2: applying the CMB model to analyze the secondary receptor

An EC-source obtained in stage 1 can be treated as a new receptor (here is secondary receptor). Compared with the original receptor, the number of source categories for a secondary receptor has been reduced. As discussed above, a secondary receptor normally contains several M-source categories (M-sources in a secondary receptor are referred to as Msub-sources). According to the loading values in secondary receptor, the emission inventory and the investigation of the studying area, the possible Msub-source categories can be identified. In this stage, the contributions of these Msub-sources to the secondary receptor can be estimated by CMB model.

2.3. Stage 3: combining the results of stages 1 and 2

Overall results of the combined source apportionment can be obtained from the results of stages 1 and 2 (i.e., the contributions of ES-sources in stage 1 and the contributions of Msub-sources in stage 2).

Detailed description of the PCA/MLR–CMB combined model can be found in our prior study (Shi et al., 2009b).

3. Experiments

3.1. Development of the synthetic receptor dataset

In this section, three synthetic receptor datasets were developed. The estimated source contributions by PCA/MLR–CMB model would be compared with the true contribution values.

Actual source profiles were used to generate the synthetic receptor dataset. For each synthetic receptor dataset, seven actual PM_{10} source categories were included: URD (urban resuspended dust), soil dust, coal combustion fly ash, cement dust, vehicle exhaust, secondary sulfate and secondary nitrate. These actual source profiles were obtained in three different cities in China: Taiyuan, Anyang and Tianjin (reported in our prior study (Bi et al., 2007)).

The method for developing the synthetic dataset was similar to that used in our previous study (Shi et al., 2009b). First, an m by n matrix (m is the number of samples and n is the number of chemical species) of concentrations ($\mu g\ m^{-3}$) C was generated:

$$c_{ij} = \sum_{k=1}^{7} s_{ik} f_{kj} \quad (2)$$

where f_{kj} is the fraction ($\mu g\ \mu g^{-1}$) of the jth species in kth source, s_{ik} is the synthetic total particulate matter mass contribution ($\mu g\ m^{-3}$) of the kth source to the ith sample, and 7 is the number of the actual source categories. The number of samples was 80, simulating the results of an 80-day sample campaign for the PCA/MLR analysis. The s_{kj} values were subjectively varied to reflect differences in the source emission pattern and the influence of metrological conditions. In this way, an 80×24 dataset can be obtained (80 samples and 24 chemical species). The synthetic source contributions to the synthetic receptor dataset (which are true values of source contributions) and their standard deviations are as follows: URD—$65.51 \pm 24.86\ \mu g\ m^{-3}$; soil—$25.24 \pm 10.66\ \mu g\ m^{-3}$; coal combustion—$31.96 \pm 13.00\ \mu g\ m^{-3}$; cement—$10.83 \pm 4.47\ \mu g\ m^{-3}$; secondary sulfate—$14.79 \pm 5.72\ \mu g\ m^{-3}$; secondary nitrate—$5.03 \pm 2.43\ \mu g\ m^{-3}$; vehicle—$35.98 \pm 14.28\ \mu g\ m^{-3}$.

For this synthetic receptor dataset, the seven sources and the receptor were highly compatible, for there was no noise. If such a synthetic receptor dataset and source profiles were fitted by the CMB model, good estimated source contributions can be obtained.

To disturb the balance, contributions of noise was added to the synthetic dataset. For each species in the source profiles, 80 random values were generated within a given range. Eighty noises were added into 80 synthetic receptor samples. The average

contributions and standard deviations of noises are 30.51 ± 11.53. These ranges for the noise were referred to our prior study (Shi et al., 2009b). Therefore, the synthetic receptor dataset was constructed from seven actual sources and noise as follows.

$$\text{synthetic original receptor} = \text{URD} + \text{soil} + \text{coal} + \text{cement} + \text{vehicle} + \text{sulfate} + \text{nitrate} + \text{noise} \quad (3)$$

Using the equation above, the original synthetic receptor dataset was generated.

Firstly, a synthetic receptor was developed by the source profiles from Taiyuan (called TY-receptor). The average concentrations of the species in TY-synthetic receptor (true values) are listed in Table 1.

3.2. Combined source apportionment for the TY-synthetic dataset

3.2.1. Stage 1: using the PCA/MLR model

PCA is a statistical technique that can be applied to a set of variables to reduce their dimensionality. The methods employed in the PCA/MLR model have been described in the literature (Hopke, 2003; Guo et al., 2004).

The normalized original TY-synthetic dataset was analyzed by PCA. After varimax rotation, seven factors (with eigenvalues of at least 1) were extracted and accounted for 80% of the variance.

These seven factors are identified as follows:

Factor 1 (31% of the variance): High loadings for Al, Si, K, Ca, Ti and Fe. This factor might be a complex source relating to two or more source categories. Factor 2 (11% of the variance): High loadings for total carbon (TC), moderate loading for Mn and etc. TC is the source marker of vehicle exhaust, which indicates that the factor might relate to vehicle exhaust. Factor 3 (10% of the variance): High loadings for NH_4^+, NO_3^- and SO_4^{2-}. This factor relates to secondary sulfate and nitrate sources. Factors 4–7 have high loadings for Pb, Co, Cl^-, Ba and other species. These factors mainly relate to noise.

Table 1
Average concentrations (μg m^{-3}) and standard deviations of species for the TY-synthetic receptor dataset (true values).

Species	Original receptor dataset		Secondary receptor dataset	
	Average	sd	Average	sd
Na	2.45	1.21	0.76	0.28
Mg	2.51	1.01	1.85	0.68
Al	12.79	4.08	10.25	3.81
Si	30.47	10.01	26.58	9.89
P	0.60	0.19	0.13	0.05
K	1.85	0.60	1.44	0.53
Ca	16.79	5.65	15.71	5.77
Ti	0.59	0.21	0.56	0.21
V	0.04	0.03	0.01	0.00
Cr	0.04	0.01	0.02	0.01
Mn	1.66	0.60	0.07	0.02
Fe	4.55	1.64	3.95	1.48
Co	0.07	0.05	0.00	0.00
Ni	0.04	0.02	0.01	0.00
Cu	0.71	0.56	0.01	0.00
Zn	0.31	0.26	0.05	0.02
Br	0.06	0.05	0.00	0.00
Ba	0.17	0.12	0.06	0.02
Pb	0.14	0.11	0.02	0.01
TC*	46.01	12.68	19.17	7.17
NH_4^+	6.72	1.95	0.04	0.01
Cl^-	1.02	0.58	0.20	0.07
NO_3^-	5.12	2.05	0.00	0.00
SO_4^{2-}	19.60	5.10	5.16	1.94
PM_{10}	219.85	54.58	133.54	49.33

*TC: total carbon.

The first factor, identified as a complex source (EC-source) in the analysis above, was treated as a secondary receptor, where the URD, soil dust, coal combustion and cement dust were the Msub-sources. They can be fitted by the CMB model in the second stage.

Next, the absolute principal component scores (APCS) were calculated according to the method described in detail by Thurston and Spengler (1985). The source contribution to the ith species of the receptor (c_i) can be calculated using multiple linear regression:

$$c_i = (b_0)i + \sum_{p=1}^{7} \text{APCS}_p{}^{*}b_{pi}, \quad (4)$$

where $(b_0)i$ is the constant term of multiple regression for ith species, b_{pi} is the coefficient of multiple regression for the pth factor and ith species of the receptor and APCS_p is the scaled value of the rotated pth factor. $\text{APCS}_p{}^{*}b_{pi}$ represents the contribution of the pth source to c_i. Therefore, the average of the product $\text{APCS}_p{}^{*}b_{pi}$ for all samples represents the average contribution of the pth source to c_i (Guo et al., 2004).

Similarly, the source contribution to the total PM_{10} mass can be described as

$$\text{total mass} = a_0 + \sum_{p=1}^{7} \text{APCS}_p{}^{*}a_p, \quad (5)$$

where the average of $\text{APCS}_p{}^{*}a_p$ for all samples represents the average contribution of the pth source to the total mass of the receptor.

The average contributions of the seven factors to the total mass and species of the TY-original synthetic receptor are listed in Table 2.

3.2.2. Stage 2: using the CMB model to fit the secondary receptor

The estimated average concentrations of the total mass (calculated according to Eq. (5)) and species (c_i) (calculated according to

Table 2
Factor profiles (μg m^{-3}) obtained by PCA/MLR stage for TY-original synthetic dataset.

Species	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Na	0.25	1.57	−0.75	−0.05	0.54	0.08	0.43
Mg	2.01	0.36	0.29	0.26	0.15	−0.05	0.07
Al	**9.76**	1.14	0.27	0.14	0.11	−0.05	0.42
Si	**24.68**	3.12	0.14	−0.01	0.19	−0.01	0.10
P	0.12	0.38	0.13	0.02	0.00	0.02	0.00
K	1.43	0.31	0.13	−0.03	0.03	0.00	0.04
Ca	**13.60**	0.65	−0.75	0.15	0.27	0.30	−0.01
Ti	0.52	0.02	−0.01	0.01	0.01	0.00	0.01
V	0.02	0.00	0.01	0.02	0.01	0.00	0.00
Cr	0.02	0.01	0.00	0.00	0.00	0.00	0.01
Mn	0.31	1.21	0.58	−0.07	−0.03	−0.01	−0.07
Fe	3.91	0.05	0.48	−0.07	0.05	0.03	0.06
Co	0.01	0.00	−0.01	0.00	0.04	0.00	−0.01
Ni	0.01	0.01	0.02	0.01	0.01	−0.01	0.01
Cu	−0.09	−0.04	0.26	0.34	0.16	0.18	0.19
Zn	0.17	−0.11	0.19	0.20	0.07	−0.04	0.03
Br	0.00	−0.01	−0.04	0.05	0.00	0.00	0.01
Ba	0.03	−0.01	0.01	0.01	−0.02	0.00	0.11
Pb	−0.02	0.05	0.01	0.11	−0.02	−0.01	−0.03
TC	21.12	**19.97**	7.17	−0.03	0.30	−0.03	−1.13
NH_4^+	0.24	0.64	**5.21**	−0.03	−0.12	0.17	0.06
Cl^-	0.09	0.24	0.18	−0.03	0.03	0.40	0.03
NO_3^-	−1.19	1.16	**2.94**	0.27	0.17	−0.18	−0.16
SO_4^{2-}	5.53	2.04	**11.08**	−0.66	−0.53	0.28	0.44
Total mass	127.93	33.09	28.54	6.34	3.60	2.09	4.59
	Complex source	Vehicle exhaust	Secondary sulfate and nitrate				

The species with bold value got high loading.

Eq. (4)) in the secondary receptor (factor 1) are listed in Table 2. Since the CMB model takes the uncertainties (standard deviations) of these average concentrations into account, the standard deviations of c_i in the secondary receptor should be first calculated.

According to our previous study (Shi et al., 2009b), the standard deviation can be estimated as

$$\sigma_i = \sqrt{\eta_i}\cdot\delta_i, \tag{6}$$

$$\eta_i = C_{i(\text{secondary})}/C_{i(\text{original})}, \tag{7}$$

where σ_i is the standard deviation of c_i in the EC-source, δ_i is the standard deviation of c_i in the original receptor, $C_{i(\text{secondary})}$ is the estimated average concentration ($\mu g\ m^{-3}$) of the ith species in the secondary receptor (Table 2) and $C_{i(\text{original})}$ is the average concentration ($\mu g\ m^{-3}$) of the ith species in the original receptor (Table 1).

For example, the estimated average concentration of Na in the secondary receptor (factor 1) was 0.25 $\mu g\ m^{-3}$ (see Table 2), while the average concentration of Na in the original receptor was 2.45 $\mu g\ m^{-3}$ with a standard deviation of 1.21 $\mu g\ m^{-3}$ (see Table 1). Therefore, the estimated standard deviation of the average concentration of Na in the secondary receptor was calculated as $\sigma_{Na} = \sqrt{0.25/2.45}\times 1.21 = 0.39$.

If the estimated average concentration of a species in the secondary receptor was negative (e.g., that for NO_3^- was −1.19 $\mu g\ m^{-3}$), it was replaced by half the detection limit.

All standard deviations of the average species concentrations of the secondary receptor were estimated according to Eqs. (6) and (7) and are listed in Table 3. To compare the estimated values of the standard deviation with the true values in the secondary receptor, we need to obtain the true average concentrations and their standard deviations. These true values can be calculated using to Eq. (2). If only the Msub-source profiles are included in the calculation, the secondary receptor can be generated:

$$\text{synthetic secondary receptor} = \text{URD} + \text{soil} + \text{coal} + \text{cement}. \tag{8}$$

Table 3
Estimated standard deviations of the average concentrations of all species of the secondary receptor ($\mu g\ m^{-3}$).

Species	Secondary receptor dataset		
	Estimated sd	Estimated sd/true sd	Estimated mean/true mean
Na	0.39	1.37	0.33
Mg	0.87	1.28	1.09
Al	3.59	0.94	0.95
Si	8.89	0.90	0.93
P	0.09	1.85	0.95
K	0.58	1.08	1.00
Ca	5.27	0.91	0.87
Ti	0.22	1.05	0.93
V	0.02	3.84	1.87
Cr	0.01	1.49	1.13
Mn	0.26	10.63	4.77
Fe	1.54	1.04	0.99
Co	0.02	32.89	5.46
Ni	0.01	3.51	1.68
Cu	0.00	0.00	0.00
Zn	0.19	11.41	3.74
Br	0.00	0.00	0.00
Ba	0.05	2.31	0.49
Pb	0.00	0.00	0.00
TC*	8.59	1.20	1.10
NH_4^+	0.37	26.83	6.49
Cl^-	0.17	2.32	0.46
NO_3^-	0.00	–	–
SO_4^{2-}	2.71	1.40	1.07
PM_{10}	42.09	0.85	0.96

sd: estimated standard deviations for secondary receptor.
*TC: total carbon.

The average species concentrations and the standard deviations of the synthetic receptor data were obtained, and all values are listed in Table 1. Therefore, the values in Table 1 are the true values in this study. The ratios of estimated values to true values are also given in Table 3. From these ratio values, we see the estimated values of some species (such as the markers for factor 1: Mg, Al, Si, K, Ca, Ti, Fe and TC) were close to the true values (ratios approaches to 1). However, some species got high ratio values, such as Mn, Co, Zn and NH_4^+. It might be because that these species did not get high loadings in factor 1. PCA/MLR model estimates the high loading species more exactly than the low loading specie.

Next, the secondary receptor and the profiles of Msub-sources (URD, soil dust, coal combustion and cement dust) were fitted by the CMB model. The contributions of these sources were obtained: 65.51 $\mu g\ m^{-3}$ for URD, 25.25 $\mu g\ m^{-3}$ for soil dust, 31.13 $\mu g\ m^{-3}$ for coal combustion and 10.83 $\mu g\ m^{-3}$ for cement dust. The performance indices for CMB model (USEPA CMB8.2, 2004) were: $R^2 = 0.96$, $\chi^2 = 0.21$ and 100.96% of mass accounted for. These values all meet the requirements of the CMB model and indicate that acceptable results were obtained in stage 2.

On the other hand, factor 3 could also be a secondary receptor, and thus the contributions of secondary sulfate and secondary nitrate can be fitted by the CMB model. The results were: secondary sulfate—15.53 $\mu g\ m^{-3}$, secondary nitrate—3.81 $\mu g\ m^{-3}$.

3.2.3. Stage 3: final results

The final results for the PCA/MLR–CMB model are listed in Table 4. It is seen that the estimated source contributions were close to the true values.

In addition, the uncertainties of the estimated contributions were calculated.

(1) In stage 1, the contributions for each sample can be estimated by PCA/MLR, so the mean contributions and their standard deviations for simplex sources can be obtained accordingly.
(2) In stage 2, a balance between weighted receptor profile (C_w) and weighted source profiles (F_w) were established by CMB model:

$$C_w = F_w S \tag{9}$$

$$\text{where}\quad C_w = (V_e)^{-1/2}C,\quad F_w = (V_e)^{-1/2}F \tag{10}$$

$$V_e^{-1} = (V_e)^{-1/2}(V_e)^{-1/2} \tag{11}$$

$$(V_e)^{-1/2} = \left((V_e)^{-1/2}\right)' \tag{12}$$

where, C is the receptor profile vector; F is the source profile matrix; $V_e(n\times n)$ is the diagonal effective variance matrix (USEPA CMB8.2, 2004).

According to the reference (Belsley et al., 1980), for an ordinary least squares (OLS) regression process, the uncertainties of the regression coefficient β can be calculated as:

$$\sigma_{\beta_j}^2 = \left[\frac{1}{n-p}\sum_{k=1}^{n}\left(y_k - \sum_{i=1}^{p}x_{ki}\beta_i\right)^2\right]*(X'X)_{jj}^{-1} \tag{13}$$

Where, $y(n\times 1)$ is the dependent variable; $x(n\times p)$ is the independent variable; $\beta(p\times 1)$ is the regression coefficient.

Table 4
Source contributions (μg m^{-3}) for three synthetic receptor datasets estimated by the PCA/MLR–CMB model.

Sources	True values	Estimated contribution (mean ± sd)			
		TY-receptor	AY-receptor	TJ-receptor	
URD	65.51 ± 24.86	60.25 ± 17.45	61.84 ± 29.07	57.58 ± 14.85	Stage 2
Soil	25.24 ± 10.66	24.41 ± 16.69	23.72 ± 26.81	20.77 ± 12.82	Stage 2
Coal combustion	31.96 ± 13.00	34.57 ± 18.34	33.08 ± 16.75	29.31 ± 4.02	Stage 2
Cement	10.83 ± 4.47	9.92 ± 9.76	10.43 ± 8.60	8.23 ± 2.97	Stage 2
Secondary sulfate	14.79 ± 5.72	15.53 ± 0.44	15.68 ± 0.65	14.80 ± 0.12	Stage 2
Secondary nitrate	5.03 ± 2.43	3.81 ± 0.20	6.53 ± 0.36	4.59 ± 0.06	Stage 2
Vehicle	35.98 ± 14.28	33.09 ± 13.69	28.82 ± 15.82	35.20 ± 18.22	Stage 1
Noise	30.51 ± 11.53				
Total PM mass	219.85 ± 54.58				

TY-receptor: Taiyuan synthetic receptor, AY-receptor: Anyang synthetic receptor, TJ-receptor: Tianjin synthetic receptor, URD: urban resuspended dust.

So, for CMB model, the uncertainties of the estimated contribution s should be:

$$\sigma_{s_j}^2 = \left\{\frac{1}{n-p}\sum_{k=1}^{n}\left[(c_w)_k - \sum_{i=1}^{p}(f_w)_{ki}s_i\right]^2\right\} * \left(F_w' F_w\right)_{jj}^{-1} = \left\{\frac{1}{n-p}\sum_{k=1}^{n}\left[\left(c_k - \sum_{i=1}^{p} f_{ki}s_i\right)^2 \Big/ V_{e_{ii}}\right]\right\} * \left(F'(V_e)^{-1}F\right)_{jj}^{-1} \quad (14)$$

where, n is the number of the fitted species; p is the number of the source profiles. In this way, the receptor profile, source profiles and their uncertainties can be all considered in Eq. (14).

The uncertainties of estimated contribution for TY-receptor were listed in Table 4.

In addition, the collinearity for the source profiles can be discussed. The concepts of the condition indexes (CI) and the variance-decomposition propositions (VDP) (Hopke, 1985) were introduced to identify the collinearity problem in source profiles. According to Hopke's discussion, the collinearity problem arises when (1) there is a singular value with a high CI (>5) and (2) there are high VDP values (>0.5) for two or more sources associated with the same singular value (Hopke, 1985).

If the TY-original synthetic receptor dataset is apportioned using traditional CMB model, all the seven source profiles should be fitted by the CMB model. The results of CI and VDP analysis of the source profile matrix are presented in Table 5(a). In the table, there are two high CI values of 17.78 and 40.04. For the value of 17.78, there was only one VDP value greater than 0.5. And, for the value of 40.04, five source categories—URD, soil dust, cement dust, vehicle exhaust and secondary sulfate—had values greater than 0.5. It can be inferred that there is a serious collinearity problem in the source profiles if the apportionment is conducted in a traditional way.

However, when using the combined model, there were only four source profiles (URD, soil dust, coal combustion and cement dust) and the secondary receptor was fitted by the CMB model. The dimension of the source profile matrix was reduced. The CI and VDP values are listed in Table 5(b). In this instance, there was only one CI value (11.98) greater than 5, which was much lower than the CI value of 40.04. Therefore, the collinearity problem was greatly reduced in the combined model.

In addition, this synthetic dataset was also studied by traditional CMB model. The discussion of the results between traditional CMB and PCA/MLR–CMB models was discussed in the file of Supplementary material.

Table 5
Condition indexes (CI) and variance-decomposition propositions (VDP) analysis.

CI	VDP						
	URD	Soil	Coal	Cement	Vehicle	Secondary sulfate	Secondary nitrate
a. TY source matrix (seven source categories)							
17.78	0.92	0.02	0.02	0.05	0.41	0.12	0.01
40.04	**0.97**	**0.98**	0.09	**0.82**	**0.56**	**0.70**	0.13
b. TY source matrix (four source categories)							
11.98	**1.00**	**0.77**	**0.58**	0.49			
c. AY source matrix (four source categories)							
12.93	**0.99**	0.44	**0.73**	0.42			
d. TJ source matrix (four source categories)							
18.48	**0.87**	**0.79**	**1.00**	0.22			

The source with bold value got high VDP.

3.3. Combined source apportionment for other synthetic datasets

Besides the TY-synthetic dataset, two other synthetic datasets were generated according to Eq. (2): AY-synthetic receptor (source profiles obtained from Anyang) and TJ-synthetic receptor (source profiles obtained from Tianjin) datasets. To compare the estimated source contributions easily, the actual source categories for these two synthetic receptor datasets were the same as those for the TY-synthetic dataset; the true source contributions were the same as the values listed in Table 4.

The final results of PCA/MLR–CMB model for TY, AY and TJ synthetic datasets were shown in Table 4. For the three synthetic datasets, it can be found that the results of TJ-receptor datasets are less close to the true values, compared with those of TY and AY datasets. It might be because the different levels of collinearity among the source categories in different cities. Table 5(b–d) list the CI and VDP values for the source profile matrices from three cities. TJ source matrix got the highest collinearity. It might indicate that

Table 6
Summary statistics for the Chengdu dataset (μg m^{-3}).

Species	Average	Stdev	Max	Min
Na	4.69	2.87	11.47	0.13
Mg	1.77	0.89	4.94	0.06
Al	13.03	6.60	28.20	0.78
Si	25.02	16.25	77.21	1.02
K	5.14	3.27	13.39	0.01
Ca	14.06	6.72	32.29	0.37
Cr	0.13	0.23	1.53	0.00
Mn	0.18	0.23	1.65	0.00
Fe	2.76	1.10	5.58	0.03
Cu	0.24	0.22	1.03	0.01
Zn	1.36	0.94	6.69	0.04
Cd	0.00	0.00	0.00	0.00
TC	83.11	54.54	220.25	7.14
NH_4^+	6.40	3.24	13.73	0.27
Cl^-	0.53	0.73	3.54	0.00
NO_3^-	5.11	5.07	26.16	0.16
SO_4^{2-}	30.58	19.23	121.64	0.45
PM_{10}	280.21	138.19	608.71	16.67

TC: total carbon.

collinearity among the source profiles can affect the final results of the combined model.

Also, the uncertainties of the estimated source contributions were listed in Table 4. It can be found that the nearly collinear sources usually got higher values. It might be because that the source profile data can significantly affect CMB source apportionment results and uncertainties (Lee and Russell, 2007).

In next section, the ambient receptor from Chengdu is apportioned by the PCA/MLR–CMB model to test its practicability.

4. Sampling and chemical analysis of the Chengdu samples

4.1. Sampling site

The city of Chengdu (102°54′E–104°53′E, 30°05′N–31°26′N) is located in the western part of Sichuan basin, in the mid-river area of the Minjiang. Chengdu is one of the largest cities in China and the capital of Sichuan province, with an area of 12,390 km^2 and a population of 11.123 million people. Chengdu has a humid semitropical monsoon climate, with warm temperatures (16 °C yearly average) and plentiful rainfall (1000 mm yearly average). The climate of Chengdu has little sunshine, with the dominant weather being fog and clouds throughout the year (1071 sunshine hours per year on average).

4.2. Ambient data sampling

Ambient PM_{10} concentration data were obtained during sampling campaigns in 2007. A total of 135 samples were selected for analyzing. All samples were collected by filtration with a medium-volume air sampler situated about 5 m above the ground. The pump was set at 100 L min^{-1} and ran continuously for 24 h. Two parallel medium-volume air samplers were used for obtaining PM_{10} on polypropylene membrane filters and quartz fiber filters. The sampling process was referred to in our previous work (Bi et al., 2007).

4.3. Source sampling

Samples for five actual source categories were obtained in Chengdu. The samples of the soil dust source were swept from

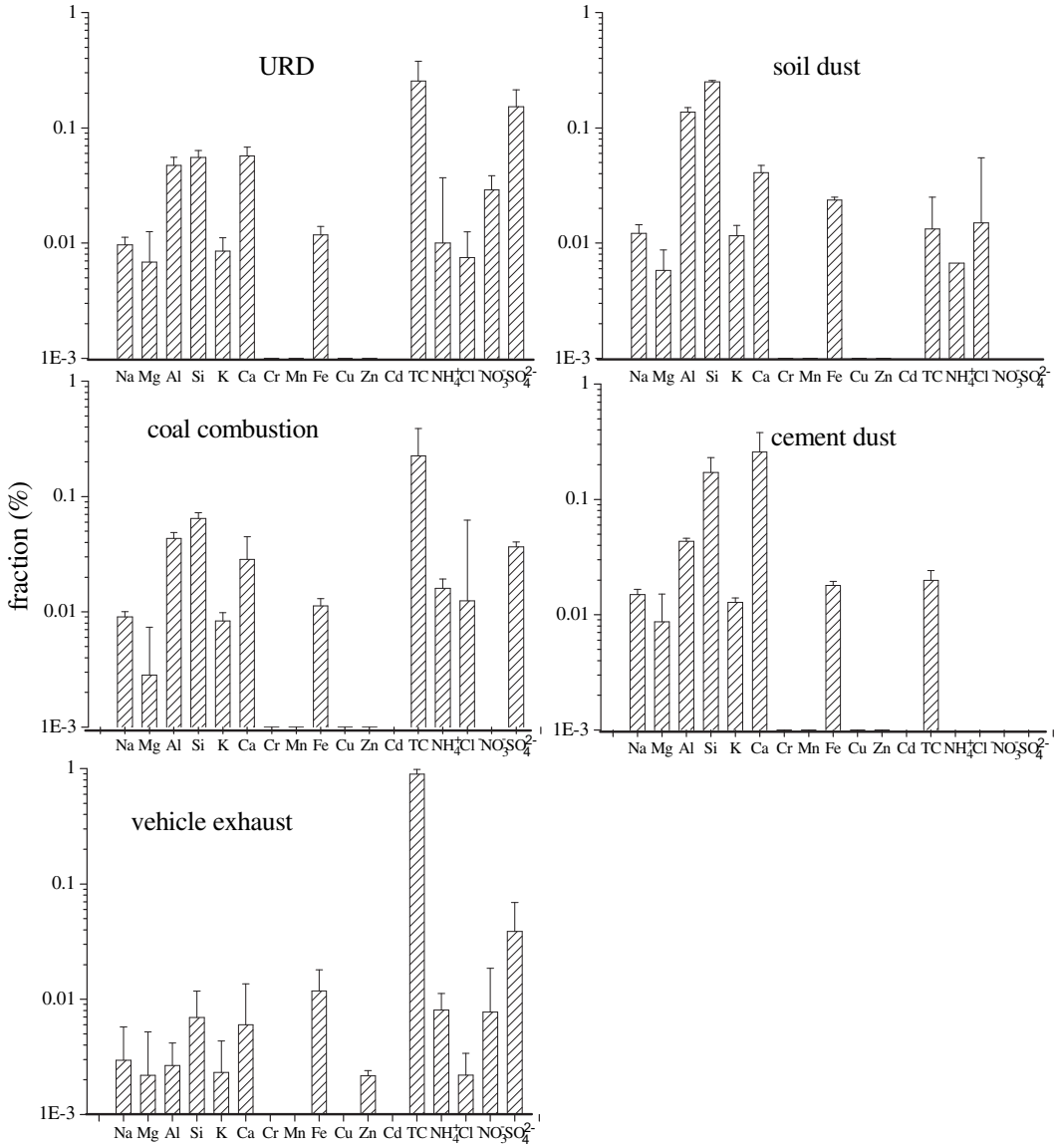

Fig. 1. Source profiles for Chengdu.

representative portions of the ground surface using a plastic brush and tray (Zhao et al., 2006). Coal combustion samples were collected from particulate pollution control devices (electrostatic precipitators, fabric filters or wet scrubbers) or sampled by a dilution stack sampler (Zhao et al., 2006). Cement dust was collected from roofs of residential buildings around the building site, or from production lines of nearby cement factories. The collected source samples were dried in a dark room with ventilation devices and sieved through 150 mesh sieves. The sieved material was then suspended in a chamber and sampled through size selective inlets onto filters to obtain the PM_{10} samples used for analyses (Chow et al., 1994). Vehicle exhaust dust was sampled from an exhaust pipes. The dilution stack and vehicle exhaust samplers could collect samples on their filters for PM_{10} measurements. All powder samples were sieved and suspended in a resuspension chamber or separated by a Bahco centrifugal machine for PM_{10} measurements (Zhao et al., 2006).

As secondary aerosol sources, ammonium sulfate and ammonium nitrate were expressed by "pure" secondary source profiles (Mazzera et al., 2001; Park and Kim, 2005).

4.4. Chemical analysis

Thirteen elements (Na, Mg, Al, Si, K, Ca, Cr, Mn, Fe, Cu, Zn and Cd) were analyzed by inductively coupled plasma (ICP) analysis (Zhao et al., 2006). These elements were extracted from polypropylene membrane filters using a laboratory system (ETHOS E, Milestone). Quartz fiber filters were cut into pieces for the analysis of ions and TC. Water soluble NH_4^+, Cl^-, NO_3^- and SO_4^{2-} were extracted by an ultrasonic extraction system (AS3120, AutoScience) and analyzed by ion chromatography (DX-120, Dionex). TC was determined using a carbon elemental analyzer (Vario EL, GmbH). The chemical analysis was detailed in our previous work (Zhao et al., 2006).

5. Results and discussion

The concentrations of analyzed species and the total mass are listed in Table 6. The average concentration of PM_{10} was 280.21 $\mu g\ m^{-3}$, which is higher than reported data for other cities in China, such as Shenzhen, Zhuhai (Cao et al., 2004), Tianjin and Taiyuan (Zhao et al., 2006). Some species such as Si, TC and SO_4^{2-} were at high levels in the ambient receptor data, while other elements such as Cr, Mn and Co were at low levels.

The source profiles are presented in Fig. 1.

The ambient receptor (shown in Table 6) (the original receptor data in this study) was analyzed using the PCA/MLR–CMB model.

5.1. Stage 1: PCA/MLR

PCA was performed on the original receptor data. After varimax rotation, four factors (with eigenvalues of at least 1) were extracted and they accounted for 80% of the variance.

Factor 1 (48% of the variance) is explained by crustal components such as Na, Mg, Al, Si and Ca. In addition, this factor is related to TC, the marker of vehicle exhaust. Therefore, this factor might be a complex source. According to the emission inventory and the investigation of the studying area (Feng et al., 2008), this factor might relate to five Msub-sources: URD, soil dust, coal combustion, cement dust and vehicle exhaust.

Factor 2 (12% of the variance) is related to NO_3^-, which is the marker of secondary nitrate. Note that there was also a high loading for Cl^- and moderate loading for some other elements (such as Mg, Ca, Fe and Zn).

Factor 3 (11% of the variance) had high loadings for NH_4^+ and SO_4^{2-}. Therefore, this factor relates to a secondary sulfate source.

Factor 4 (9% of the variance) had a high loading for Cr and moderate loadings for Mg, Mn, Zn and Cd. Hence, this factor might relate to a smelter source.

After MLR, the contributions to the total mass from each source and the species concentrations were obtained. The values are listed in Table 7. The complex source (factor 1) contributed 235.35 $\mu g\ m^{-3}$ to PM_{10}, the secondary sulfate contributed 33.35 $\mu g\ m^{-3}$, and the secondary nitrate and smelter sources contributed 17.65 and 1.81 $\mu g\ m^{-3}$ respectively.

5.2. Stage 2: CMB model

In stage 2, the complex source (factor 1) was treated as the secondary receptor. The standard deviations were calculated using Eqs. (6) and (7). Secondary receptor data were used for the ambient data input file, and information on the five Msub-source (URD, soil dust, coal combustion, cement dust and vehicle exhaust) profiles was used for the source profile input file. These input files were analyzed using the EPA CMB8.2 model.

The contributions of the five Msub-sources were estimated. The estimated contributions of vehicle exhaust, coal combustion, URD, soil dust and cement dust to the secondary receptor were 80.46 ± 67.99, 68.52 ± 152.23, 53.91 ± 43.61, 46.31 ± 70.80 and $12.47 \pm 20.38\ \mu g\ m^{-3}$, respectively. The performance indices meet the requirements of the CMB model: χ^2 was 0.29, R^2 was 0.94 and the percentage of mass accounted for was 111.18%. In addition, the diagnostic MPIN (modified pseudo-inverse normalized) matrix can be obtained by CMB model. The high MPIN value identifies the high influence of the fitting species on the source contribution estimated (Chow et al., 2007). For Chengdu dataset, the influential species were: Mg and NO_3^- for URD; Si for soil; Al and SO_4^{2-} for coal; Ca for cement; and TC for vehicle.

5.3. Stage 3: final results for the PCA/MLR–CMB model

The final results are listed in Table 8. Vehicle exhaust contributed most to ambient PM_{10}, and coal combustion was the second highest contributor.

5.4. Study by NCPCRCMB

In addition, the receptor dataset from Chengdu was also studied by Nonnegative Constrained Principal Component Regression

Table 7
Factor profiles ($\mu g\ m^{-3}$) obtained by PCA/MLR stage for Chengdu original receptor dataset.

Species	Factor 1	Factor 2	Factor 3	Factor 4
Na	4.79	0.42	0.58	0.21
Mg	1.24	0.20	−0.05	0.32
Al	**11.15**	0.63	0.96	−0.03
Si	**26.94**	0.79	1.88	−1.00
K	5.52	0.42	0.57	0.08
Ca	**10.23**	1.73	−0.63	1.09
Cr	−0.01	−0.01	0.07	0.19
Mn	0.27	−0.01	−0.04	0.10
Fe	1.54	0.31	0.49	−0.07
Cu	0.23	−0.05	−0.04	0.01
Zn	1.00	0.23	−0.12	0.28
Cd	0.00	0.00	0.00	0.00
TC	**91.91**	0.58	3.21	1.31
NH_4^+	−0.26	−0.06	4.05	0.23
Cl^-	0.24	0.43	−0.36	0.05
NO_3^-	1.29	**3.20**	1.83	−0.70
SO_4^{2-}	20.16	0.80	**17.53**	4.95
Total mass	235.35 Complex source	17.65 Secondary nitrate	33.35 Secondary sulfate	1.81 Smelter

The species with bold value got high loading.

Table 8
Source contributions (μg m^{-3}) to Chengdu ambient receptor data estimated by the PCA/MLR–CMB and NCPCRCMB models.

Sources	Estimated contribution: μg m^{-3} (100%)			
	PCA/MLR–CMB		NCPCRCMB	Traditional CMB
URD	53.91 ± 43.61 (19.24%)	Stage 2	41.22 ± 107.97 (14.71%)	161.91 ± 110.52
Soil	46.31 ± 70.80 (16.53%)	Stage 2	27.45 ± 58.53 (9.80%)	59.10 ± 58.11
Coal combustion	68.52 ± 152.23 (24.45%)	Stage 2	30.43 ± 122.78 (10.86%)	−99.95 ± 159.90
Cement	12.47 ± 20.38 (4.45)	Stage 2	49.47 ± 34.70 (17.65%)	23.51 ± 29.02
Vehicle	80.46 ± 67.99 (28.71%)	Stage 2	86.70 ± 57.80 (30.94%)	75.39 ± 59.69
Secondary sulfate	33.35 ± 24.63 (11.90%)	Stage 1	16.41 ± 11.78 (5.86%)	16.59 ± 14.66
Secondary nitrate	17.65 ± 23.98 (6.30%)	Stage 1	3.61 ± 6.28 (1.29%)	0.14 ± 5.91
Smelter	1.81 ± 1.92 (0.65%)	Stage 1	–	–

Chemical Mass Balance Model (NCPCRCMB). NCPCRCMB is a new receptor which was developed in our prior work (Shi et al., 2009a). The Nonnegative Constrained Principal Component Regression route is added into the traditional CMB model iteration, to solve the near collinearity problem for source apportionment. The detailed description of this model can be found in the reference (Shi et al., 2009a).

For NCPCRCMB model, profiles of seven source categories (the source categories were listed in Table 8 and source profiles were descript in Fig. 1) and receptor (shown in Table 6) were introduced into the model. The results of NCPCRCMB model for Chengdu dataset were also shown in Table 8 (the uncertainties of the estimated contribution were also calculated according to Eq. (14)). The performance indices are: $\chi^2 = 0.72$, $R^2 = 0.81$ and PM = 91.11%.

The results show that vehicle got the highest contribution (86.70 μg m^{-3}), which was similar to the results of PCA/MLR–CMB. However, the estimated contribution of coal by NCPCRCMB was relatively lower than that by PCA/MLR–CMB; while the cement was the opposite case. Secondary sulfate and nitrate got higher estimated contributions by PCA/MLR–CMB than those by NCPCRCMB model. It is because that the factors (factor 2 and 3 in Table 7) extracted by PCA/MLR–CMB included unknown sources. See Table 7, the estimated concentration of NO_3^- in factor 2 was 3.20 μg m^{-3} and that of SO_4^{2-} in factor 3 was 17.53 μg m^{-3}. These values are close to the estimated contributions of secondary sulfate and nitrate by NCPCRCMB, respectively. It can be found that the sum of estimated contributions by NCPCRCMB was less than that by PCA/MLR–CMB model. The differences are reasonable. In several studies, different results might be obtained by two receptor models. NCPCRCMB and PCA/MLR–CMB model are based on traditional CMB model. For CMB model, the diagnostic PM within 80%–120% meets the requirement. As presented above, the diagnostic indices of results of PCA/MLR–CMB and NCPCACMB all meet the CMB requirement. So the results of both models can be acceptable. In addition, for uncertainty, the nearly collinear sources got relative higher values for both models.

The Chengdu ambient dataset was analyzed by traditional CMB model. The results are shown in Table 8. Among the sources, soil and coal combustion are the near collinear sources. For traditional CMB results, coal combustion got negative results.

6. Conclusions

According to the discussion above, the combined model can be used for source apportionment. The noise in the receptor can be reduced in PCA/MLR stage; and some similar source categories might be contained in one factor, this factor is a complex source; then in CMB-stage, the complex source can be treated as a secondary receptor, then the secondary receptor and its contained sources can be more compatible. In this way, a more accurate result (source contributions) of apportionment can be obtained.

The experiment in this study evaluated this technique by applying synthetic datasets. Results for each stage of combined model were compared with the true values, the results were acceptable. In addition, the synthetic tests show that nearly collinear sources usually got higher uncertainties. In addition, the ambient receptor dataset was studied by PCA/MLR–CMB and NCPCRCMB models, and acceptable results were obtained.

Acknowledgments

This study is supported by SDP of Tianjin (No. 09ZCGYSF02400) and the combined laboratory of the Tianjin Meteorological Bureau and Nankai University.

Appendix. Supplementary material

Supplementary material related to this article can be found online, at doi:10.1016/j.atmosenv.2011.03.007.

References

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. A John Wiley & Sons, INC., Hoboken, New Jersey.
Bi, X.H., Feng, Y.C., Wu, J.H., Wang, Y.Q., Zhu, T., 2007. Source apportionment of PM_{10} in six cities of northern China. Atmospheric Environment 41, 903–912.
Cao, J.J., Lee, S.C., Ho, K.F., Zou, S.C., Fung, K., Li, Y., Watson, J.G., Chow, J.C., 2004. Spatial and seasonal variations of atmospheric organic carbon and elemental carbon in Pearl River Delta region, China. Atmospheric Environment 38, 4447–4456.
Cao, J.J., Wu, F., Chow, J.C., Lee, S.C., Li, Y.S., Chen, W., An, Z.S., Fung, K.K., Watson, J.G., Zhu, C.S., Liu, S.X., 2005. Characterization and source apportionment of atmospheric organic and elemental carbon during fall and winter of 2003 in Xi'an, China. Atmospheric Chemistry and Physics 5, 3127–3137.
Chen, L.-W.A., Lowenthal, D.H., Watson, J.G., Koracin, D., Kumar, N., Knipping, E.M., Wheeler, N., Craig, K., Reid, S., 2010. Toward effective source apportionment using positive matrix factorization: experiments with simulated PM2.5 data. Journal of the Air & Waste Management Association 60, 43–54.
Chow, J.C., Watson, J.G., Houck, J.E., Pritchett, L.C., Rogers, C.F., Frazier, C.A., Egami, R.T., Ball, B.M., 1994. A laboratory resuspension chamber to measure fugitive dust size distribution and chemical composition. Atmospheric Environment 28, 3463–3481.
Chow, J.C., Watson, J.G., 2002. Review of $PM_{2.5}$ and PM_{10} apportionment for fossil fuel combustion and other sources by the chemical mass balance receptor model. Energy & Fuels 16, 222–260.
Chow, J.C., Watson, J.G., Lowenthal, D.H., Chen, L.W.A., Zielinska, B., Mazzoleni, L.R., Magliano, K.L., 2007. Valuation of organic markers for chemical mass balance source apportionment at the Fresno Supersite. Atmospheric Chemistry and Physics 7, 1741–1754.
Feng, Y.C., Zhou, L.D., Zhang, P., Wu, J.H., Deng, X.Y., 2008. Source Apportionment for Air Particulate Matter in Chengdu. Chengdu Research Academy of Environmental Sciences and Nankai University.
Guo, H., Wang, T., Simpson, I.J., Blake, D.R., Yu, X.M., Kwok, Y.H., Li, Y.S., 2004. Source contributions to ambient VOCs and CO at a rural site in eastern China. Atmospheric Environment 38, 4551–4560.
Harrison, R.M., Smith, D.J.T., Luhana, L., 1996. Source apportionment of atmospheric polycyclic aromatic hydrocarbons collected from an urban location in Birmingham, UK. Environmental Science & Technology 30, 825–832.
Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. Chemometrics and Intelligent Laboratory Systems 65, 179–189.

Hedberg, E., Gidhagen, L., Johansson, C., 2005. Source contributions to PM_{10} and arsenic concentrations in Central Chile using positive matrix factorization. Atmospheric Environment 39, 549–561.
Hellebust, S., Allanic, A., O'Connor, I.P., Wenger, J.C., Sodeau, J.R., 2010. The use of real-time monitoring data to evaluate major sources of airborne particulate matter. Atmospheric Environment 44, 1116–1125.
Henry, R.C., Christensen, E.R., 2010. Selecting an appropriate multivariate source apportionment model result. Environmental Science & Technology 44, 2474–2481.
Hopke, P.K., 1985. Receptor Modeling in Environmental Chemistry. John Wiley & Sons, Inc.
Hopke, P.K., 2003. Recent developments in receptor modeling. Journal of Chemometrics 17, 255–265.
Lee, S., Russell, A.G., 2007. Estimating uncertainties and uncertainty contributors of CMB $PM_{2.5}$ source apportionment results. Atmospheric Environment 41, 9619–9624.
Lioy, P.J., Zelenka, M.P., Cheng, M.D., Reiss, N.M., Wilson, W., 1989. The effect of sampling duration on the ability to resolve source types using factor analysis. Atmospheric Environment 23, 239–254.
Mamane, Y., Perrino, C., Yossef, O., Catrambone, M., 2008. Source characterization of fine and coarse particles at the East Mediterranean coast. Atmospheric Environment 24A, 2089–2097.
Mazzei, F., D'Alessandro, A., Lucarelli, F., Nava, S., Prati, P., Valli, G., Vecchi, R., 2008. Characterization of particulate matter sources in an urban environment. Science of the Total Environment 401, 81–89.
Mazzera, D.M., Lowenthal, D.H., Chow, J.C., Watson, J.G., 2001. Sources of PM10 and sulfate aerosol at McMurdo station, Antarctica. Chemosphere 45, 347–356.
Okamoto, S., Hayashi, M., Nakajima, M., Kainuma, Y., Shiozawa, K., 1990. A factor analysis-multiple regression model for source apportionment of suspended particulate matter. Atmospheric Environment 42, 6114–6130.
Paatero, P., 1997. Least squares formulation of robust non-negative factor analysis. Chemometrics and Intelligent Laboratory Systems 37, 23–35.
Park, S.S., Kim, Y.J., 2005. Source contributions to fine particulate matter in an urban atmosphere. Chemosphere 59, 217–226.
Reff, A., Eberly, S.I., Bhave, P.V., 2007. Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. Journal of the Air & Waste Management Association 57, 146–154.
Shi, G.L., Feng, Y.C., Zeng, F., Li, X., Zhang, Y.F., Wang, Y.Q., Zhu, T., 2009a. Use of a Nonnegative Constrained Principal Component Regression Chemical Mass Balance Model to study the contributions of nearly collinear sources. Environmental Science & Technology 43, 8867–8873.
Shi, G.L., Li, X., Feng, Y.C., Wang, Y.Q., Wu, J.H., Li, J., Zhu, T., 2009b. Combined source apportionment, using positive matrix factorization–chemical mass balance and principal component analysis/multiple linear regression–chemical mass balance models. Atmospheric Environment 43, 2929–2937.
Shi, G.L., Li, X., Wu, T., Feng, Y.C., Bi, X.H., Wu, J.H., Zhu, T., 2009c. Effects of Unknown Source on the Collinearity Problem in CMB Model, The 237th ACS National Meeting, Salt Lake City, UT, March 22–26.
Song, Y., Xie, S.D., Zhang, Y.H., Zeng, L.M., Salmon, L.G., Zheng, M., 2007. Source apportionment of PM2.5 in Beijing using principal component analysis/absolute principal component scores and UNMIX. Science of the Total Environment 372, 278–286.
Stone, E., Schauer, J., Quraishi, T.A., Mahmood, A., 2010. Chemical characterization and source apportionment of fine and coarse particulate matter in Lahore, Pakistan. Atmospheric Environment 44, 1062–1070.
Thurston, G.D., Spengler, J.D., 1985. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. Atmospheric Environment 19, 9–25.
Ulbrich, I.M., Canagaratna, M.R., Zhang, Q., Worsnop, D.R., Jimenez, J.L., 2009. Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data. Atmospheric Chemistry and Physics 9, 2891–2918.
US Environmental Protection Agency, 2004. EPA CMB8 2 User's Manual. Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711.
Watson, J.G., Zhu, T., Chow, J.C., Engelbrecht, J., Fujita, E.M., Wilson, W.E., 2002. Receptor modeling application framework for particle source apportionment. Chemosphere 49, 1093–1136.
Watson, J.G., Chen, L.W.A., Chow, J.C., Doraiswamy, P., Lowenthal, D.H., 2008. Source apportionment, findings from the US Supersites Program. Journal of the Air & Waste Management Association 58, 265–288.
Zhao, P.S., Feng, Y.C., Zhu, T., Wu, J.H., 2006. Characterizations of resuspended dust in six cities of North China. Atmospheric Environment 40, 5807–5814.