



Evaluation of a Modified Receptor Model for Solving Multiple Time Resolution Equations: A Simulation Study

Ho-Tang Liao¹, Cheng-Pin Kuo¹, Philip K. Hopke², Chang-Fu Wu^{1,3,4*}

¹ Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University, Taipei, Taiwan

² Center for Air Resources Engineering and Science, and Department of Chemical Engineering, Clarkson University, Potsdam, NY, USA

³ Department of Public Health, National Taiwan University, Taipei, Taiwan

⁴ Institute of Environmental Health, National Taiwan University, Taipei, Taiwan

ABSTRACT

This study was conducted to evaluate the performance of an improved source apportionment model that is suitable for incorporating data with multiple time resolutions. This evaluation was achieved by using synthetic data sets that simulated environmental concentrations of volatile organic compounds (VOCs) and fine particulate matter (PM_{2.5}) from the five following sources: petroleum refinery, vehicle exhaust, industrial coating, coal combustion, and natural gas. Hourly VOCs and speciated PM_{2.5} data were simulated for a one-week period. The PM_{2.5} data were further averaged every twelve hours to generate data sets with mixed temporal resolutions. The Multilinear Engine program was applied to resolve the source profiles and contributions. A series of sensitivity analyses was conducted to examine how uncertainties in the profile variation, measurement error, and source collinearity affected the model performance. The resolved factor profiles closely matched the input profiles, and the measurement error had a larger impact on the modeling results than the profile variation. In the most comprehensive data set that contained all three types of uncertainty, the R² values between the input and retrieved source contributions were between 0.87 and 0.95. The estimated percentage contributions were also comparable with the input ones, demonstrating the applicability and validity of this improved model.

Keywords: Source apportionment; Receptor modeling; Multilinear engine; Multiple time resolution.

INTRODUCTION

Volatile organic compounds (VOCs) and fine particulate matter (PM_{2.5}) have been found to be associated with adverse health effects such as asthma induction and exacerbations, respiratory symptoms, and impaired lung function (Seagrave *et al.*, 2006; Wichmann *et al.*, 2009; Stanek *et al.*, 2011a, b). Thus, it is helpful to identify the contributions of air pollutants from different sources and evaluate the source-specific health impacts (Heal *et al.*, 2012; Lall *et al.*, 2012) in order to design effective control strategies. To achieve this analysis, source apportionment models, such as chemical mass balance (CMB), absolute principal component scores (APCS), and positive matrix factorization (PMF) models, can be used to quantify the contribution of a particular source at a receptor site (Cooper and Watson, 1980; Viana *et al.*, 2008a; Wang *et*

al., 2008; Gugamsetty *et al.*, 2012; Huang *et al.*, 2012). Based on the mass conservation concept, the receptor model is expressed by the following mass balance equation:

$$x_{ij} = \sum_{p=1}^P g_{ip} f_{pj} + e_{ij} \quad (1)$$

where x_{ij} represents the concentration of measured species j in sample i , g_{ip} is the contribution of source p to sample i , f_{pj} is the mass fraction of species j in source profile p , and e_{ij} is the residual of the model fit for the measured concentration. The multivariate models require less previous knowledge regarding the emission source profile than the CMB model. Therefore, when complete emission profiles are not available, the PMF model (one of the advanced multivariate models that simultaneously retrieve g and f from x) can be applied (Viana *et al.*, 2008a, b).

To date, few studies have considered combining both gaseous and particulate species into one source apportionment model. Previously, the Multilinear Engine (ME), a more flexible program for solving the PMF problem, was used to quantify the contributions of different sources to PM_{2.5} and VOC mixtures (Wu *et al.*, 2007). Nevertheless, different

* Corresponding author. Tel.: +886-2-3366-8096;
Fax: +886-2-3366-8096
E-mail address: changfu@ntu.edu.tw

sampling time schedules for VOCs and PM_{2.5} resulted in difficulties for conventional receptor modeling. For example, the PM_{2.5} composition data are commonly obtained by analyzing time-integrated filter samples. However, the hourly VOCs data can be monitored semicontinuously with automated monitors. For conventional receptor models, it is necessary to average the high-resolution data over longer sampling periods to match the low-resolution data. To prevent decreasing sample size and the loss of valuable information from the high-resolution data (Lioy et al., 1989), a more flexible model for solving multiple time resolution equations is needed. A previous study (Zhou et al., 2004) used field data to demonstrate that a modified ME program is suitable for this purpose.

This study was conducted to evaluate the performance of an improved source apportionment model that is suitable for incorporating data with multiple time resolutions. This assessment was achieved by using synthetic data as in several previous studies (Miller et al., 2002; Brinkman et al., 2006; Christensen et al., 2006; Lingwall and Christensen, 2007; Hemann et al., 2009; Vedal et al., 2009; Habre et al., 2011).

METHODS

Data Simulation

A primary data set that represented the hourly measurements in a week was created from five sources (i.e., petroleum refinery, vehicle exhaust, industrial coating, coal combustion, and natural gas) based on the assumption that the receptor site is located near an industrial complex (Buzcu and Fraser, 2006; Leuchner and Rappenglück, 2010). The hourly contribution matrix G (i.e., g_{ip} in Eq. (1)) was generated by using average contributions (Table S1 in the Supporting Information) that were available in the literature (Zheng et al., 2002; Lee et al., 2003; Kim et al., 2005; Marmur et al., 2005; Alastuey et al., 2006; Hopke et al., 2006; Kulkarni et al., 2007; Song et al., 2008; Yuan et al., 2009; Chan et al., 2011; Guo et al., 2011). Next, hourly fluctuations were added by assuming random variability and a lognormal distribution (Lingwall and Christensen, 2007) for each source, except for vehicle exhaust. Time series data of carbon monoxide (a classic tracer for traffic) collected at an air quality monitoring station in an urban area were used to approximate the diurnal and weekly vehicle exhaust variations. The factor profile matrix F (i.e.,

f_{pj} in Eq. (1)) was derived from the US EPA SPECIATE database (Lingwall and Christensen, 2007; Simon et al., 2010), which included 18 VOCs and 17 PM_{2.5} species (Fig. S1 in the Supporting Information). The species included in the simulated data set were determined according to source profile availability and proportion of values below the method detection limits (MDLs). The MDLs were developed based on the analytical techniques that were used to determine VOCs (Gas Chromatography) and elements (X-ray Fluorescence). For organic carbon, elemental carbon, nitrate, and sulfate, the MDLs were obtained from Hemann et al. (2009). After establishing G and F , the observation matrix X was calculated by using the following equation:

$$X = GF \quad (2)$$

The temporal resolution for PM_{2.5} data was intentionally reduced by averaging the associated source contributions at twelve-hour intervals to generate a mixed temporal resolution data set (D_{Pri}). If a simulated concentration value was below the MDL for species j , it was replaced by half of the MDL value. After the creation of the simulated data, the uncertainties of the input measurements were processed as follows: 1) the uncertainties for concentrations below the MDL were set to 5/6 of the MDL value and 2) the uncertainties for the other concentrations were determined by using the following equation (Miller et al., 2002):

$$u_{ij} = \sqrt{(0.5 \times MDL_{ij})^2 + (0.1 \times x_{ij})^2} \quad (3)$$

A series of sensitivity analyses was conducted to examine how the profile variation, the presence of measurement error, or the source collinearity affects the accuracy and precision of the modeling results (Table 1). Briefly, random noise was added to the source profiles (D_{FVar}) to reflect source profile variability by using Eq. (4) (Brinkman et al., 2006).

$$f'_{pj} = (1 + CV \times z_{pj}) f_{pj} \quad (4)$$

where CV (= 0.4) represents the coefficient of variation and z_{pj} is the random number that was sampled from the standard normal distribution for each species in each profile. In a previous study, the global CV average ranging

Table 1. Simulated data sets and sensitivity analysis scenarios.

Data set	Scenario
D_{Hour}	Hourly distributed primary data set
D_{Pri}	Primary data set with multiple time resolution
D_{FVar}	Add profile variation to D_{Pri}
D_{MErr}	Add measurement error to D_{Pri}
D_{Comb}	Add profile variation and measurement error to D_{Pri}
D_{Pri}^C	Primary data set with correlated source contribution
D_{FVar}^C	Add profile variation to D_{Pri}^C
D_{MErr}^C	Add measurement error to D_{Pri}^C
D_{Comb}^C	Add profile variation and measurement error to D_{Pri}^C

between 0.11 and 0.41 was used (Brinkman *et al.*, 2006). The measurement errors (D_{MErr}) were added in a similar manner with a CV of 0.1 (Miller *et al.*, 2002). Furthermore, a data set that combined profile variation and measurement error (D_{Comb}) was generated accordingly.

In the above data sets, the contributions of each source were assumed to be independent of each other (Table 2). We further simulated another series of data sets with correlated source contributions. The correlated primary data set ($D_{\text{Pri}}^{\text{C}}$) was simulated by increasing the cross-correlations of the temporal patterns between the industrial coating and coal combustion sources and between the vehicle exhaust and natural gas sources (Table 2). The associated data sets with profile variation ($D_{\text{FVar}}^{\text{C}}$), measurement error ($D_{\text{MErr}}^{\text{C}}$), or a combination of these two ($D_{\text{Comb}}^{\text{C}}$) were generated consequently by following the above mentioned procedures. Thirty simulation runs were conducted for each scenario in the sensitivity analyses and the total number of tested data sets was 270.

Model Implementation

The mass balance equation for the modified ME model that is suitable for multiple time resolution data is expressed as follows (Zhou *et al.*, 2004):

$$x_{sj} = \frac{1}{t_{s2} - t_{s1} + 1} \sum_{p=1}^P \left[\left(\sum_{i=t_{s1}}^{t_{s2}} g_{ip} \right) f_{pj} \right] + e_{sj} \quad (5)$$

where x_{sj} represents the concentration of measured species j in sample s and t_{s1} and t_{s2} are the start and end times, respectively, as described by the number of time units. In this study, the time unit is set to one hour based on the VOCs sampling interval. In Eq. (5), g_{ip} is the contribution of source p during the time units for sample s , f_{pj} is the mass fraction of species j in source profile p , and e_{sj} is the residual for species j in sample s .

In this study, the unmeasured mass of $\text{PM}_{2.5}$ was included and defined as the m th species as follows:

$$x_{s,j=m} = (\text{PM}_{2.5})_s - \sum_{j=1}^{m-1} x_{sj} \quad (6)$$

where $(\text{PM}_{2.5})_s$ is the measured $\text{PM}_{2.5}$ mass concentration of the s th sample. This inclusion eliminated the need for the post-hoc regression process (Larson *et al.*, 2006; Wu *et al.*, 2007). It also constrains the sum of all $\text{PM}_{2.5}$ species mass fractions in each source to equal 100% based on the following auxiliary equation:

$$\sum_{j=1}^m f_{pj} = 1.0 \pm \delta \quad (7)$$

where δ is set to 0.02 in this study.

Data Analysis

The agreement between the simulated and modeled contributions and profiles was evaluated for each data set by calculating the average absolute error (AAE) (Christensen and Gunst, 2004; Christensen *et al.*, 2006; Lingwall and Christensen, 2007). Unlike other model performance statistics, such as the root mean square error (RMSE), the AAE gives the amplitude of average error without being pulled by the large one (Javitz *et al.*, 1988; Willmott and Matsuura, 2005). The AAE of contribution (AAE_g) was defined as follows:

$$\text{AAE}_g = \frac{1}{n} \sum_{i=1}^n \sum_{p=1}^P |\hat{g}_{ip} - g_{ip}| \quad (8)$$

where n is the number of samples and \hat{g}_{ip} is the modeled contribution of source p to sample i . The AAE of profile (AAE_f) was obtained in a similar way. In addition, the coefficient of determination (R^2) of the contribution estimates for each source was calculated to evaluate the variance of the model fits (Brinkman *et al.*, 2006).

RESULTS AND DISCUSSION

Primary Analysis

For the hourly distributed primary data set, all R^2 values were greater than 0.99, which indicated excellent agreement between the retrieved and simulated matrices (Table 3, D_{Hour}). In addition, the percentage contributions that were

Table 2. Correlation coefficients between source contributions in the primary data sets.

	Petroleum refinery	Vehicle exhaust	Industrial coating	Coal combustion	Natural gas
D_{Pri}					
Petroleum refinery	1	0.014	0.018	0.026	-0.020
Vehicle exhaust		1	-0.012	-0.029	-0.025
Industrial coating			1	0.005	0.007
Coal combustion				1	0.015
Natural gas					1
$D_{\text{Pri}}^{\text{C}}$					
Petroleum refinery	1	-0.018	-0.009	-0.029	0.002
Vehicle exhaust		1	-0.013	-0.002	0.664
Industrial coating			1	0.527	-0.008
Coal combustion				1	0.001
Natural gas					1

Table 3. R^2 values of the estimated source contributions and the modeled percentage contributions (PC, in %) in different data sets.

	Petroleum refinery	Vehicle exhaust	Industrial coating	Coal combustion	Natural gas
Inputted					
PC for VOC	31	32	21	6	10
PC for PM _{2.5}	28	33	13	21	5
D_{Hour}					
R^2	0.998	0.997	0.999	1.000	0.998
PC for VOC*	33 (31–33)	28 (27–30)	23 (22–24)	5 (5–5)	11 (10–11)
PC for PM _{2.5} *	29 (27–30)	30 (29–31)	15 (14–16)	20 (20–21)	6 (5–6)
D_{Pri}					
R^2	0.982	0.994	0.998	0.955	0.985
PC for VOC*	36 (34–38)	23 (22–25)	22 (21–22)	7 (6–8)	12 (12–13)
PC for PM _{2.5} *	27 (20–31)	28 (27–29)	13 (11–19)	23 (20–28)	8 (7–11)
D_{Comb}					
R^2	0.950	0.963	0.946	0.642	0.901
PC for VOC*	33 (29–39)	22 (18–26)	23 (20–27)	9 (6–12)	12 (11–15)
PC for PM _{2.5} *	28 (18–34)	26 (22–32)	13 (9–28)	23 (20–30)	9 (4–19)
D_{Comb}^C					
R^2	0.940	0.947	0.945	0.871	0.953
PC for VOC*	34 (29–36)	22 (18–26)	22 (20–26)	10 (7–14)	13 (11–15)
PC for PM _{2.5} *	28 (25–32)	25 (20–30)	13 (9–17)	26 (20–30)	9 (4–13)
D_{3S}^C					
R^2	0.943	0.915	0.944	0.784	0.953
PC for VOC*	35 (30–38)	19 (14–21)	22 (20–24)	10 (8–14)	14 (11–16)
PC for PM _{2.5} *	29 (23–34)	23 (16–30)	14 (9–17)	24 (14–29)	11 (6–19)
D_{FCor}					
R^2	0.949	0.937	0.942	0.594	(Diesel exhaust) 0.516
PC for VOC*	33 (28–37)	18 (12–24)	23 (21–28)	10 (6–22)	15 (12–19)
PC for PM _{2.5} *	28 (19–34)	24 (20–29)	14 (10–34)	24 (19–34)	11 (7–17)

* Median (range) of sensitivity analysis results (n = 30 runs).

reconstructed from the hourly distributed primary data set matched well with the input ones. After reducing the time resolution for the PM_{2.5} data, the R^2 values decreased slightly, but remained above 0.95 (Table 3, D_{Pri}). Nevertheless, the percentage contributions had higher deviation than those from the hourly distributed data set, especially for vehicle exhaust/VOC. This result potentially occurred because no major species had a mass fraction greater than 15% in the vehicle exhaust/VOC profile (Fig. S1). This relatively non-unique profile could lead to poor separation from other sources (Brinkman *et al.*, 2006). To test this hypothesis, the mass fractions of ethylene and isopentane in the VOC source profile of vehicle exhaust were increased to more than 20% and re-applied to the run with the lowest percentage contribution estimate (22% in Table 3) for vehicle exhaust/VOC. In this case, a more accurate value (29%) was obtained, which suggested this hypothesis was valid.

Sensitivity Analysis

Compared with the primary data set (D_{Pri}), the AAE did not change much after adding profile variation (D_{FVar}), although the variability was larger (Fig. 1). In contrast, the AAE significantly increased with the presence of measurement error (D_{MErr} and D_{Comb}). These results suggest

that the bias of the modeling results from measurement error is greater than that resulting from profile variation. The explanation is that the measurement error greatly influenced the residual term e_{sj} in Eq. (5), thus affecting the solutions of the minimization algorithm.

The R^2 values between the retrieved and simulated contributions in the primary data set are given in Fig. 2 (D_{Pri}) by source. Most R^2 values were greater than 0.8. Only minor influences resulted from profile variation (D_{FVar}). However, the R^2 values decreased for all sources when measurement error was considered (D_{MErr} and D_{Comb}), especially for the coal combustion source. In the input data, coal combustion was assumed to contribute only 6% to the VOC mass concentration. Because receptor models may not fully resolve sources with low mass contributions (Ulbrich *et al.*, 2009), this issue was investigated by doubling the average coal combustion contribution to VOC (6% to 12%). Increasing this percentage to 12% improved the median R^2 value from 0.525 to 0.778. This result indicated that the measurement error has a larger impact on modeling results when the percentage contribution of a source is smaller (Miller *et al.*, 2002; Brinkman *et al.*, 2006; Ulbrich *et al.*, 2009).

A similar tendency was observed for each scenario in

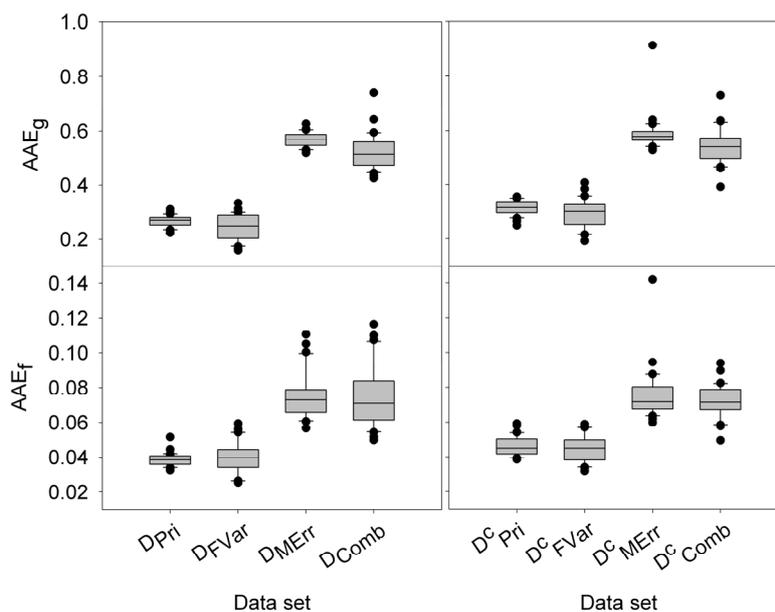


Fig. 1. Distribution of the AAE results for the source contributions and profiles in different data sets.

the correlated data sets when compared with the non-correlated data sets (Fig. 1). The minor deviation that resulted from the collinearity of the source contribution was unexpected. Fig. 2 shows that greater R^2 values were obtained for coal combustion in the correlated data sets than in the non-correlated data sets ($D_{\text{Comb}}^{\text{C}}$ vs. D_{Comb}). This result contradicts those of previous studies, in which the correlation between the simulated and predicted contributions decreased as collinearity increased (Brinkman *et al.*, 2006; Vedal *et al.*, 2009; Habre *et al.*, 2011). This contrasting result was explored by generating an additional data set ($D_{\text{Comb}}^{\text{C}2}$) with modified correlation structures among the simulated sources. The coal combustion in the $D_{\text{Comb}}^{\text{C}2}$ data set was simulated as cross correlating with natural gas, which also had a low source contribution (Table S2), instead of correlating with the high contributor (industrial coating) in the $D_{\text{Comb}}^{\text{C}}$ data set. It was assumed that the other three sources were independent of each other. A lower median (range) R^2 of 0.536 (0.004–0.891) was observed in the $D_{\text{Comb}}^{\text{C}2}$ data set relative to the D_{Comb} and $D_{\text{Comb}}^{\text{C}}$ data sets (R^2 of 0.642 (0.019–0.897) and 0.871 (0.644–0.951), respectively). These observations showed that better results indeed were obtained in scenarios that did not have source collinearity between two low contributors (R^2 : $D_{\text{Comb}} > D_{\text{Comb}}^{\text{C}2}$). On the other hand, source collinearity may reduce the disadvantage (R^2 : $D_{\text{Comb}}^{\text{C}} > D_{\text{Comb}}$) of a source having small percentage contribution (e.g., coal combustion) when it is correlated with another source having high percentage contribution (e.g., industrial coating in $D_{\text{Comb}}^{\text{C}}$). We further simulated a data set ($D_{3\text{S}}^{\text{C}}$) in which three sources (vehicle exhaust, coal combustion, and natural gas) were cross-correlated. For the high contributor of vehicle exhaust, decreased R^2 (0.915) and percentage contributions (19% for VOC and 23% for $\text{PM}_{2.5}$) were obtained, compared to the ones in the D_{Comb} data set (Table 3). On the other hand, for the low contributors of coal combustion and natural gas, their R^2 increased to 0.784 and 0.953, respectively, although

their percentage contribution estimates also increased.

Source profile collinearity was investigated by replacing natural gas in D_{Comb} with diesel exhaust which has a profile highly correlated with that of vehicle exhaust (Chin and Batterman, 2012). A lower R^2 of 0.516 and higher percentage contributions (15% for VOC and 11% for $\text{PM}_{2.5}$) were observed for diesel exhaust, compared to those of natural gas in the original D_{Comb} data set (Table 3, D_{FCor}). The problem of profile collinearity might be solved by including additional species to reduce similarity between profiles (Chin and Batterman, 2012). These above analyses indicated a trade-off between high and low contributors when they correlated with each other. The contribution estimates of the high contributors may be underestimated while those of the low contributors may be overestimated.

The percentage contributions in the most comprehensive data set, $D_{\text{Comb}}^{\text{C}}$, which combine profile variation, measurement error, and source collinearity, are given in Table 3. The modeled percentage contributions are comparable with the input contributions. A comparison between the simulated and retrieved source profiles is shown in Fig. 3. The error bars represent the standard deviations of each species for 30 runs. Good agreement was obtained between the input source profiles and the model estimates. It should be noted that the simulated data cannot completely illustrate the complexities of the real world. Nevertheless, results based on synthetic data are valuable for providing insights regarding the limitations and abilities of the improved receptor model.

One additional issue to be discussed is determining the appropriate number of factors in the model. The above analysis assumed that the number of factors was known (i.e., 5 in this study). In the real world, the number of sources is generally unknown. Thus, we applied the model to the $D_{\text{Comb}}^{\text{C}}$ data set and varied the number of factors between 3 to 7. Next, we observed changes in the maximum individual column mean (IM) and the maximum individual column standard deviation (IS) of the scaled residual matrix

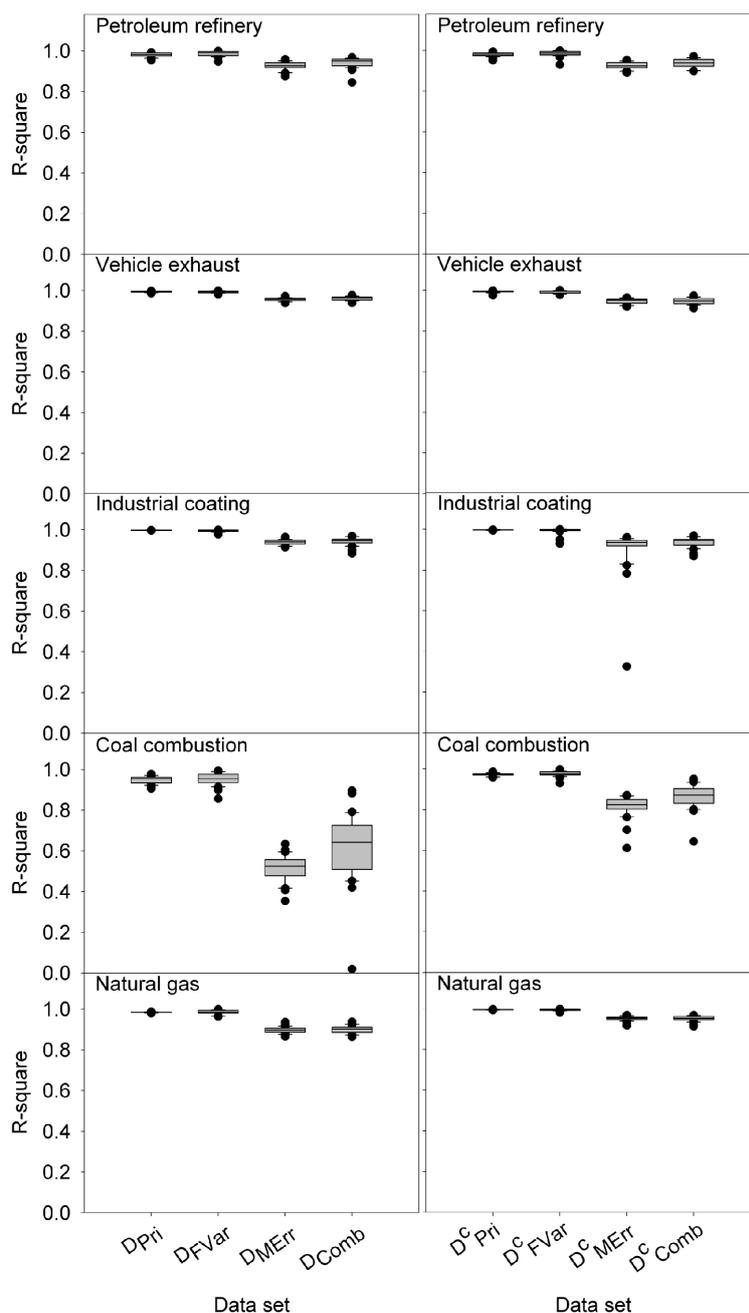


Fig. 2. Distribution of the coefficient of determination (R^2) for the source contributions in different data sets.

versus the number of factors in the model (Lee *et al.*, 1999; Brinkman *et al.*, 2006; Heo *et al.*, 2009). Because previous studies suggested that Q values depend greatly on assigned uncertainties, the common evaluation method of comparing the obtained and theoretical Q values was not used (Paatero and Tapper, 1993; Paatero, 2007). As shown in Fig. 4, the optimal number of factors should be between four and six based on the judging criterion of having an apparent decrease in IM and IS. This information usually provides initial guidance regarding the possible range of numbers of factors rather than pointing toward a single number of factors in the final solution (Lee *et al.*, 2003; Brinkman *et al.*, 2006). Previous studies showed that the optimal fit should be determined based on the most physically meaningful

solutions (Lee *et al.*, 2003; Kim *et al.*, 2005; Paatero, 2007; Ulbrich *et al.*, 2009; Yuan *et al.*, 2009). In the four-factor solution of the D^C_{Comb} data set, n-heptane, toluene, and m,p-xylene had large residuals that resulted from the combination of industrial coating and coal combustion sources. In the six-factor solution, two factors with profiles resembling vehicle exhaust were obtained. However, one of the factors was not interpretable because its profile lacked important i-pentane and organic carbon species. Consequently, a five-factor solution was the most reasonable.

CONCLUSIONS

In this study, a modified source apportionment model was

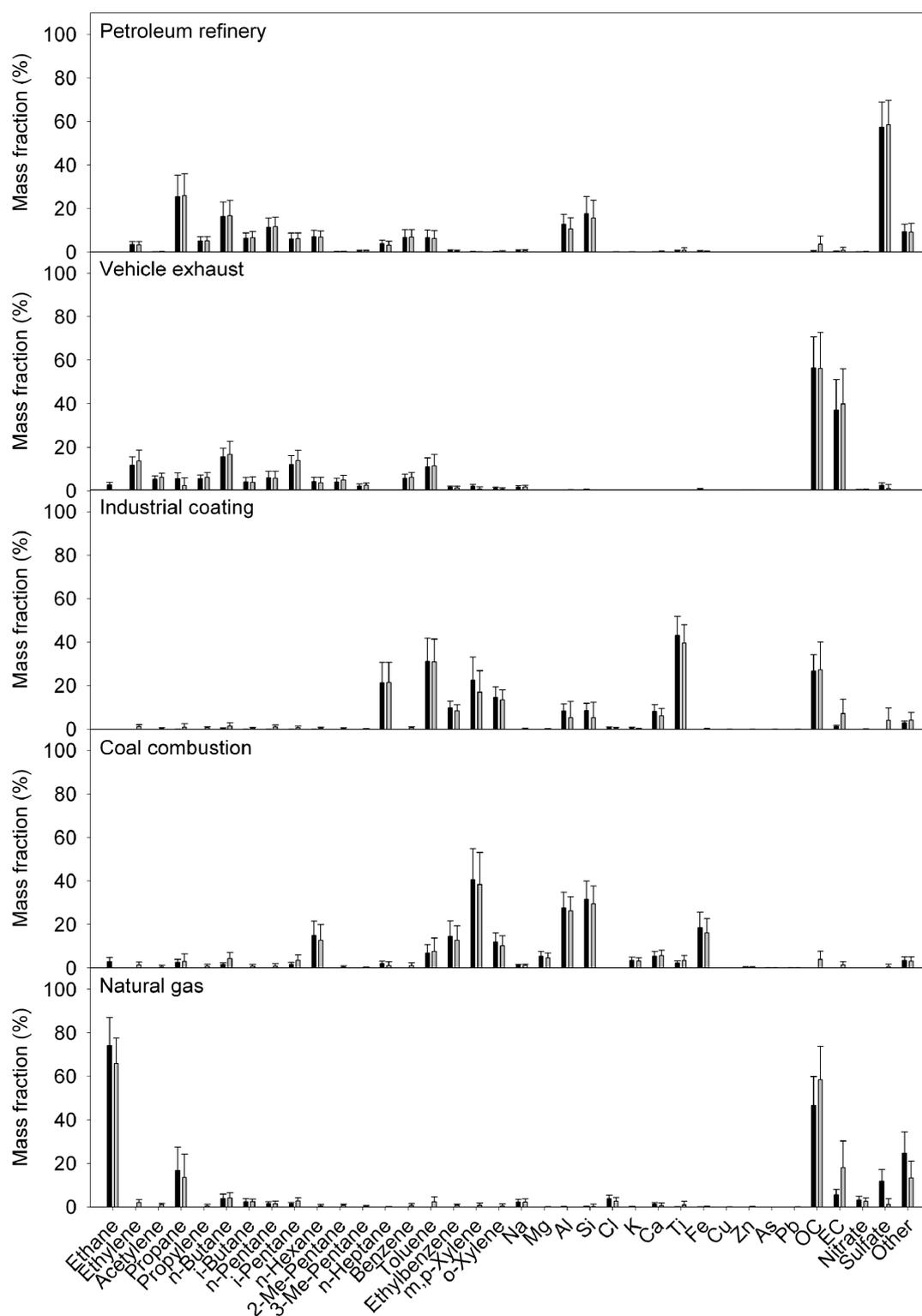


Fig. 3. Comparison of the simulated (black bar with error as the mean \pm SD) and retrieved (gray bar with error as the mean \pm SD) source profiles in the D^C_{Comb} data set.

evaluated for its performance with multiple time resolution data. This model was affected by measurement errors (similar to other models), especially when the contribution of certain sources was small. Nonetheless, the high correlations and the comparable percentage contributions

between the input data and the model results in most of the simulated scenarios demonstrated the applicability and validity of this improved model. Interpretation of model results should be particularly cautious when factors with small contributions are retrieved.

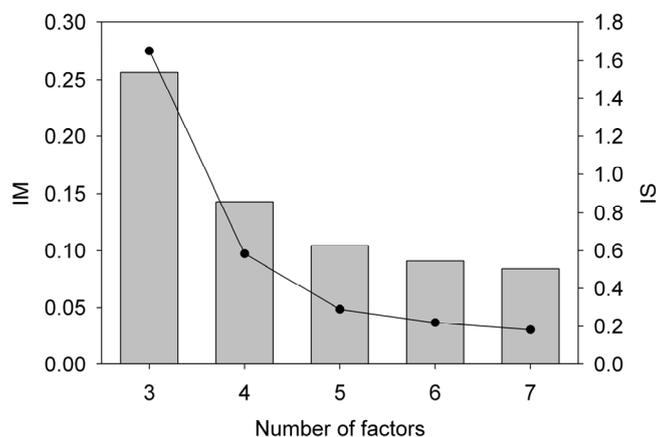


Fig. 4. Determination of the number of factors by using the maximum individual column mean (IM, filled circles) and the maximum individual column standard deviation (IS, gray bars) from scaled residuals.

ACKNOWLEDGMENTS

The authors thank the National Science Council of Taiwan for financially supporting this research under Contract NSC100-2221-E-002-038.

SUPPLEMENTARY MATERIALS

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

REFERENCES

- Alastuey, A., Querol, X., Plana, F., Viana, M., Ruiz, C.R., Sanchez de la Campa, A., de la Rosa, J., Mantilla, E. and Garcia dos Santos, S. (2006). Identification and Chemical Characterization of Industrial Particulate Matter Sources in Southwest Spain. *J. Air Waste Manage. Assoc.* 56: 993–1006.
- Brinkman, G., Vance, G., Hannigan, M.P. and Milford, J.B. (2006). Use of Synthetic Data to Evaluate Positive Matrix Factorization as a Source Apportionment Tool for PM_{2.5} Exposure Data. *Environ. Sci. Technol.* 40: 1892–901.
- Buzcu, B. and Fraser, M.P. (2006). Source Identification and Apportionment of Volatile Organic Compounds in Houston, TX. *Atmos. Environ.* 40: 2385–2400.
- Chan, Y.C., Hawas, O., Hawker, D., Vowles, P., Cohen, D.D., Stelcer, E., Simpson, R., Golding, G. and Christensen, E. (2011). Using Multiple Type Composition Data and Wind Data in PMF Analysis to Apportion and Locate Sources of Air Pollutants. *Atmos. Environ.* 45: 439–449.
- Chin, J.Y. and Batterman, S.A. (2012). VOC Composition of Current Motor Vehicle Fuels and Vapors, and Collinearity Analyses for Receptor Modeling. *Chemosphere* 86: 951–958.
- Christensen, W.F. and Gunst, R.F. (2004). Measurement Error Models in Chemical Mass Balance Analysis of Air Quality Data. *Atmos. Environ.* 38: 733–744.
- Christensen, W.F., Schauer, J.J. and Lingwall, J.W. (2006). Iterated Confirmatory Factor Analysis for Pollution Source Apportionment. *Environmetrics* 17: 663–681.
- Cooper, J.A. and Watson, J.G. (1980). Receptor Oriented Methods of Air Particulate Source Apportionment. *J. Air Pollut. Contr. Assoc.* 30: 1116–1125.
- Gugamsetty, B., Wei, H., Liu, C.N., Awasthi, A., Hsu, S.C., Tsai, C.J., Roam, G.D., Wu, Y.C. and Chen, C.F. (2012). Source Characterization and Apportionment of PM₁₀, PM_{2.5} and PM_{0.1} by Using Positive Matrix Factorization. *Aerosol Air Qual. Res.* 12: 476–491.
- Guo, H., Cheng, H.R., Ling, Z.H., Louie, P.K.K. and Ayoko, G.A. (2011). Which Emission Sources Are Responsible for the Volatile Organic Compounds in the Atmosphere of Pearl River Delta? *J. Hazard. Mater.* 188: 116–124.
- Habre, R., Coull, B. and Koutrakis, P. (2011). Impact of Source Collinearity in Simulated PM_{2.5} Data on the PMF Receptor Model Solution. *Atmos. Environ.* 45: 6938–6946.
- Heal, M.R., Kumar, P. and Harrison, R.M. (2012). Particles, Air Quality, Policy and Health. *Chem. Soc. Rev.* 41: 6606–6630.
- Hemann, J.G., Brinkman, G.L., Dutton, S.J., Hannigan, M.P., Milford, J.B. and Miller, S.L. (2009). Assessing Positive Matrix Factorization Model Fit: A New Method to Estimate Uncertainty and Bias in Factor Contributions at the Measurement Time Scale. *Atmos. Chem. Phys.* 9: 497–513.
- Heo, J.B., Hopke, P.K. and Yi, S.M. (2009). Source Apportionment of PM_{2.5} in Seoul, Korea. *Atmos. Chem. Phys.* 9: 4957–4971.
- Hopke, P.K., Ito, K., Mar, T., Christensen, W.F., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Larson, T.V., Liu, H., Neas, L., Pinto, J., Stolzel, M., Suh, H., Paatero, P. and Thurston, G.D. (2006). PM Source Apportionment and Health Effects: 1. Intercomparison of Source Apportionment Results. *J. Exposure Sci. Environ. Epidemiol.* 16: 275–286.
- Huang, C.H., Chen, K.S. and Wang, H.K. (2012). Measurements and PCA/APCS Analyses of Volatile Organic Compounds in Kaohsiung Municipal Sewer Systems, Southern Taiwan. *Aerosol Air Qual. Res.* 12:

- 1315–1326.
- Javitz, H.S., Watson, J.G. and Robinson, N. (1988). Performance of the Chemical Mass Balance Model with Simulated Local-Scale Aerosols. *Atmos. Environ.* 22: 2309–2322.
- Kim, E., Hopke, P.K., Pinto, J.P. and Wilson, W.E. (2005). Spatial Variability of Fine Particle Mass, Components, and Source Contributions during the Regional Air Pollution Study in St. Louis. *Environ. Sci. Technol.* 39: 4172–4179.
- Kulkarni, P., Chellam, S. and Fraser, M.P. (2007). Tracking Petroleum Refinery Emission Events Using Lanthanum and Lanthanides as Elemental Markers for PM_{2.5}. *Environ. Sci. Technol.* 41: 6748–6754.
- Lall, R., Ito, K. and Thurston, G.D. (2011). Distributed Lag Analyses of Daily Hospital Admissions and Source-Appportioned Fine Particle Air Pollution. *Environ. Health Perspect.* 119: 455–460
- Larson, T.V., Covert, D.S., Kim, E., Elleman, R., Schreuder, A.B. and Lumley, T. (2006). Combining Size Distribution and Chemical Species Measurements into A Multivariate Receptor Model of PM_{2.5}. *J. Geophys. Res.* 111: D10S09.
- Lee, E., Chan, C.K. and Paatero, P. (1999). Application of Positive Matrix Factorization in Source Apportionment of Particulate Pollutants in Hong Kong. *Atmos. Environ.* 33: 3201–3212.
- Lee, P.K.H., Brook, J.R., Dabek-Zlotorzynska, E. and Mabury, S.A. (2003). Identification of the Major Sources Contributing to PM_{2.5} Observed in Toronto. *Environ. Sci. Technol.* 37: 4831–4840.
- Leuchner, M. and Rappenglück, B. (2010). VOC Source-Receptor Relationships in Houston during TexAQS-II. *Atmos. Environ.* 44: 4056–4067.
- Lingwall, J.W. and Christensen, W.F. (2007). Pollution Source Apportionment using *A Priori* Information and Positive Matrix Factorization. *Chemom. Intell. Lab. Syst.* 87: 281–294.
- Lioy, P.J., Zelenka, M.P., Cheng M.D., Reiss, N.M. and Wilson, W.E. (1989). The Effect of Sampling Duration on the Ability to Resolve Source Types Using Factor Analysis. *Atmos. Environ.* 23: 239–254.
- Marmur, A., Unal, A., Mulholland, J.A. and Russell, A.G. (2005). Optimization-Based Source Apportionment of PM_{2.5} Incorporating Gas-to-Particle Ratios. *Environ. Sci. Technol.* 39: 3245–3254.
- Miller, S.L. anderson, M.J., Daly, E.P. and Milford, J.B. (2002). Source Apportionment of Exposures to Volatile Organic Compounds. I. Evaluation of Receptor Models using Simulated Exposure Data. *Atmos. Environ.* 36: 3629–3641.
- Paatero, P. and Tapper, U. (1993). Analysis of Different Modes of Factor Analysis as Least Squares Fit Problems. *Chemom. Intell. Lab. Syst.* 18: 183–194.
- Paatero, P. (1999). The Multilinear Engine - A Table-Driven, Least Squares Program for Solving Multilinear Problems, Including the *n*-Way Parallel Factor Analysis Model. *J. Comput. Graph. Statist.* 8: 854–888.
- Paatero P. (2007). End User's Guide to Multilinear Engine Applications. Available at <ftp://ftp.clarkson.edu/pub/hop/kep/pmf/>.
- Seagrave, J., McDonald, J.D., Bedrick, E., Edgerton, E.S., Gigliotti, A.P., Jansen, J.J., Ke, L., Naeher, L.P., Seilkop, S.K., Zheng, M. and Mauderly, J.L. (2006). Lung Toxicity of Ambient Particulate Matter from Southeastern U.S. Sites with Different Contributing Sources: Relationships between Composition and Effects. *Environ. Health Perspect.* 114: 1387–93.
- Simon, H., Beck, L., Bhave, P.V., Divita, F., Hsu, Y., Luecken, D., Mobley, J.D., Pouliot, G.A., Reff, A., Sarwar, G. and Strum, M. (2010). The Development and Uses of EPA's SPECIATE Database. *Atmos. Pollut. Res.* 1: 196–206.
- Song, Y., Dai, W., Shao, M., Liu, Y., Lu, S., Kuster, W. and Goldan, P. (2008). Comparison of Receptor Models for Source Apportionment of Volatile Organic Compounds in Beijing, China. *Environ. Pollut.* 156: 174–183.
- Stanek, L.W., Brown, J.S., Stanek, J., Gift, J. and Costa, D.L. (2011a). Air Pollution Toxicology-A Brief Review of the Role of the Science in Shaping the Current Understanding of Air Pollution Health Risks. *Toxicol. Sci.* 120 Suppl 1: S8–27.
- Stanek, L.W., Sacks, J.D., Dutton, S.J. and Dubois, J.J.B. (2011b). Attributing Health Effects to Apportioned Components and Sources of Particulate Matter: An Evaluation of Collective Results. *Atmos. Environ.* 45: 5655–5663.
- Ulbrich, I.M., Canagaratna, M.R., Zhang, Q., Worsnop, D.R. and Jimenez, J.L. (2009). Interpretation of Organic Components from Positive Matrix Factorization of Aerosol Mass Spectrometric Data. *Atmos. Chem. Phys.* 9: 2891–2918.
- Vedal, S., Hannigan, M.P., Dutton, S.J., Miller, S.L., Milford, J.B., Rabinovitch, N., Kim, S.Y. and Sheppard, L. (2009). The Denver Aerosol Sources and Health (DASH) Study: Overview and Early Findings. *Atmos. Environ.* 43: 1666–1673.
- Viana, M., Kuhlbusch, T.A.J., Querol, X., Alastuey, A., Harrison, R.M., Hopke, P.K., Winiwarter, W., Vallius, M., Szidat, S., Prévôt, A.S.H., Hueglin, C., Bloemen, H., Wählin, P., Vecchi, R., Miranda, A.I., Kasper-Giebl, A., Maenhaut, W. and Hitzenberger, R. (2008a). Source Apportionment of Particulate Matter in Europe: A Review of Methods and Results. *J. Aerosol Sci.* 39: 827–849.
- Viana, M., Pandolfi, M., Minguillón, M.C., Querol, X., Alastuey, A., Monfort, E. and Celades, I. (2008b). Inter-Comparison of Receptor Models for PM Source Apportionment: Case Study in an Industrial Area. *Atmos. Environ.* 42: 3820–3832.
- Wang, W.C., Chen, K.S., Chen, S.J., Lin, C.C., Tsai, J.H., Lai, C.H. and Wang, S.K. (2008). Characteristics and Receptor Modeling of Atmospheric PM_{2.5} at Urban and Rural Sites in Pingtung, Taiwan. *Aerosol Air Qual. Res.* 8: 112–129.
- Wichmann, F.A., Muller, A., Busi, L.E., Cianni, N., Massolo, L., Schlink, U., Porta, A. and Sly, P.D. (2009). Increased Asthma and Respiratory Symptoms in Children Exposed to Petrochemical Pollution. *J. Allergy Clin.*

- Immunol.* 123: 632–638.
- Willmott, C.J. and Matsuura, K. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Climate Res.* 30: 79–82.
- Wu, C.F., Larson, T.V., Wu, S.Y., Williamson, J., Westberg, H.H. and Liu, L.J.S. (2007). Source Apportionment of PM_{2.5} and Selected Hazardous Air Pollutants in Seattle. *Sci. Total Environ.* 386: 42–52.
- Yuan, Z., Lau, A.K.H., Shao, M., Louie, P.K.K., Liu, S.C. and Zhu, T. (2009). Source Analysis of Volatile Organic Compounds by Positive Matrix Factorization in Urban and Rural Environments in Beijing. *J. Geophys. Res.* 114: D00G15.
- Zheng, M., Cass, G.R., Schauer, J.J. and Edgerton, E.S. (2002). Source Apportionment of PM_{2.5} in the Southeastern United States Using Solvent-Extractable Organic Compounds as Tracers. *Environ. Sci. Technol.* 36: 2361–2371.
- Zhou, L., Hopke, P.K., Paatero, P., Ondov, J.M., Pancras, J.P., Pekney, N.J. and Davidson, C.I. (2004). Advanced Factor Analysis for Multiple Time Resolution Aerosol Composition Data. *Atmos. Environ.* 38: 4909–4920.

Received for review, November 21, 2012

Accepted, March 2, 2013