Atmospheric Environment 61 (2012) 387–394

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

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

Ensemble-trained source apportionment of fine particulate matter and method uncertainty analysis

Sivaraman Balachandran [a,*], Jorge E. Pachon [a,b], Yongtao Hu [a], Dongho Lee [c], James A. Mulholland [a], Armistead G. Russell [a]

[a] Georgia Institute of Technology, School of Civil and Environmental Engineering, 311 Ferst Dr., Atlanta, GA 30332, USA
[b] Universidad de La Salle, Programa de Ingenieria Ambiental, Bogota, Colombia
[c] Gyeongnam Province Institute of Health and Environment, Changwon, Gyeongnam 641-702, South Korea

HIGHLIGHTS

- We ensemble averaged three receptor models and one chemical transport model.
- We develop a method to calculate new estimates of source impact uncertainties.
- The ensemble average had better performance measures than the individual methods.
- The ensemble has lower relative uncertainties as compared to the individual methods.

ARTICLE INFO

Article history:
Received 11 April 2012
Received in revised form
6 July 2012
Accepted 11 July 2012

Keywords:
$PM_{2.5}$
Source apportionment
Ensemble
Health
Air quality

ABSTRACT

An ensemble-based approach is applied to better estimate source impacts on fine particulate matter ($PM_{2.5}$) and quantify uncertainties in various source apportionment (SA) methods. The approach combines source impacts from applications of four individual SA methods: three receptor-based models and one chemical transport model (CTM). Receptor models used are the chemical mass balance methods CMB-LGO (Chemical Mass Balance-Lipschitz global optimizer) and CMB-MM (molecular markers) as well as a factor analytic method, Positive Matrix Factorization (PMF). The CTM used is the Community Multiscale Air Quality (CMAQ) model. New source impact estimates and uncertainties in these estimates are calculated in a two-step process. First, an ensemble average is calculated for each source category using results from applying the four individual SA methods. The root mean square error (RMSE) between each method with respect to the average is calculated for each source category; the RMSE is then taken to be the updated uncertainty for each individual SA method. Second, these new uncertainties are used to re-estimate ensemble source impacts and uncertainties. The approach is applied to data from daily $PM_{2.5}$ measurements at the Atlanta, GA, Jefferson Street (JST) site in July 2001 and January 2002. The procedure provides updated uncertainties for the individual SA methods that are calculated in a consistent way across methods. Overall, the ensemble has lower relative uncertainties as compared to the individual SA methods. Calculated CMB-LGO uncertainties tend to decrease from initial estimates, while PMF and CMB-MM uncertainties increase. Estimated CMAQ source impact uncertainties are comparable to other SA methods for gasoline vehicles and SOC but are larger than other methods for other sources. In addition to providing improved estimates of source impact uncertainties, the ensemble estimates do not have unrealistic extremes as compared to individual SA methods and avoids zero impact days.

1. Introduction

Controlling fine particulate matter poses unique challenges in developing strategies to improve public health and welfare (e.g., improved visibility). Unlike most other air pollutants, fine particulate matter (i.e. particles with an aerodynamic diameter less than 2.5 μm, or $PM_{2.5}$) is comprised of a heterogeneous mix of chemical species, some of which are emitted directly from a variety of sources and others that are formed via atmospheric processes which convert gaseous species into condensed-phase compounds. The health concern over $PM_{2.5}$ has grown as associations have been found between $PM_{2.5}$ mass and health outcomes (Dockery et al.,

* Corresponding author. Tel.: +1 206 250 6480; fax: +1 404 894 8266.
E-mail address: siv@gatech.edu (S. Balachandran).

1352-2310/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.atmosenv.2012.07.031

1993; U.S.EPA, 2009), and has led EPA to regulate $PM_{2.5}$ as a criteria pollutant as part of the US EPA's National Ambient Air Quality Standards (NAAQS).

Addressing $PM_{2.5}$ levels relies on quantifying source-to-air quality relationships, a process often termed source apportionment (SA). Historically, SA of $PM_{2.5}$ has been conducted using receptor-based modeling approaches such as chemical mass balance (CMB) modeling (Watson et al., 1984) or factor analytic (FA) approaches such as positive matrix factorization (PMF) and UNMIX (Henry, 1997, 2003; Paatero and Tapper, 1994). Receptor-based modeling approaches typically solve a mass balance equation that is used to reconstruct the mass of each measured species (Equation (1)):

$$C_i = f_{ij}S_j + e_i \quad (1)$$

where C_i, is the measured concentration of species i (μg species i m^{-3}), f_{ij} is the amount of species i emitted per unit amount from source j (μg of species i per μg of $PM_{2.5}$ emitted from j), S_j, is the impact of source j (μg $PM_{2.5}$ m^{-3}), and e_i is the error for the ith species between the measured concentration, C_i, and the calculated concentration, $f_{ij}S_j$. The most commonly used CMB approach, using more routinely available $PM_{2.5}$ observations (elemental and organic carbon: EC/OC, ionic and metal species), and EPA's CMB 8.2 software, is referred to here as CMB-regular, or **CMB-RG**, (U.S.EPA, 2004). CMB has also been applied using organic molecular markers, referred to here as **CMB-MM**, which allows identification of more primary organic sources than are typically quantified using CMB-RG (Cass, 1998; Zheng et al., 2007, 2002). Another CMB approach, called **CMB-LGO**, uses CMB and incorporates gaseous species measurements to constrain results (Marmur et al., 2005). Positive matrix factorization (PMF, version 3.0) (Paatero et al., 2003; Paatero and Tapper, 1994; U.S.EPA, 2008) is a commonly used factor analytic approach. Receptor models can be readily applied for long time periods for which observational data is available.

Recently, chemical transport models (CTMs), such as the Community Multiscale Air Quality (**CMAQ**), have been used to quantify source impacts on $PM_{2.5}$ (Baek et al., 2005; Byun et al., 1998; Cohan et al., 2005; Koo et al., 2009; Marmur et al., 2006b; Wang et al., 2009; Yang et al., 2000; Yarwood et al., 2007). CTMs utilize emissions inventories and meteorological information to model transport and atmospheric chemistry in a three dimensional grid, and calculate source impacts over a large spatial domain and over time scales that may not be available from observations. Another advantage of using chemical transport models is that they can directly link and quantify the impacts of gaseous emission sources on particulate matter, a weakness of receptor-based approaches.

There have been several efforts to determine relationships between sources of $PM_{2.5}$ and health outcomes (Laden et al., 2000; Mar et al., 2000; Marmur et al., 2006a; Sarnat et al., 2008; Stolzel et al., 2005; Thurston et al., 2005), though with different results. In Thurston et al. (2005) traffic sources were not significantly associated with both CVD and non-accidental mortality, and, as the authors note, the factor analytic approaches were limited in their ability to separate gasoline and diesel fractions. Subsequently, Sarnat et al. (2008) compared epidemiologic model results using a factor analytic SA method, PMF, and an optimized CMB method, CMB-LGO (Lipschitz global optimizer) (Marmur et al., 2005) to apportion sources for four years of speciated $PM_{2.5}$ data in Atlanta and using individual compounds that are viewed as reasonable tracers for various sources. They found good agreement in RRs for CVD and respiratory outcomes between using PMF, CMG-LGO and tracers, implying different SA methods yield similar results when incorporated into epidemiologic models. However, a positive association was shown between biomass burning and CVD outcomes but not respiratory outcomes, whereas a number of previous studies showed positive associations with respiratory but not CVD outcomes (Ito et al., 2006; Mar et al., 2006). As the authors note, several recent studies corroborate their findings, but there also may be confounding effects across source categories (Barregard et al., 2006; Barrett et al., 2006; Ostro et al., 2007; Sarnat et al., 2008). Thurston et al. (2005), who incorporated nine factor analytic SA results into epidemiologic models for Phoenix, AZ and Washington D.C., found that variability in SA results across investigators/methods increased 95% confidence intervals (CI) of relative risk ratio (RR) per inter-quartile range by approximately 15%. However, contributions from similar factors sometimes differed by an order of magnitude, making inter-comparisons between methods and their associations with health less clear (Grahame and Hidy, 2007).

Both receptor and emissions-based SA approaches have limitations that can affect their inclusion in health studies. Receptor-based SA results can vary substantially from method to method, and some approaches lead to bias and increased variability (Barregard et al., 2006; Barrett et al., 2006; Christensen et al., 2006; Christensen and Amemiya, 2003; Henry, 1987; Marmur et al., 2006a; Ostro et al., 2007; Sarnat et al., 2008). With a limited number of factors identified or source profiles available, these methods assign mass from other sources to available factors/sources, leading to bias. Typical receptor model applications use source profiles, or identify factors, associated with only about 80% of the estimated $PM_{2.5}$ emissions (Baek, 2009). The necessary resources required to apply CTMs over long periods inhibit their use, and they are subject to uncertainties in emission and meteorological inputs and model parameterizations.

A number of studies have evaluated SA results (Brinkman et al., 2006; Christensen and Gunst, 2004; Lee et al., 2009, 2008; Marmur et al., 2006a; Marmur et al., 2006b; Rizzo and Scheff, 2007; Tauler et al., 2009). Marmur et al. (2006a) showed that CMAQ had significantly less variability in fractional source impacts, than CMB-LGO, effectively precluding its use to provide source impact estimates that can be differentiated in terms of health impact associations in acute epidemiologic-based studies (Marmur et al., 2006a, 2006b). Christensen and Gunst (2004) evaluated the difference in CMB results for a simulated data set using four approaches to calculating source impacts and found that the weighted least squares approach performed better than the effective variance approach in most cases and was "slightly superior" in cases where the source profile variability is large. Christensen and Schauer (2008) showed that perturbations to species concentration uncertainties can lead to relatively large differences in PMF results. Lee and Russell (2007) found that source impact uncertainties in CMB-RG were more affected by source profile error contributed than measurement error.

Using an ensemble of air quality models has provided a means to evaluate air quality models (Delle Monache et al., 2006; Dennis et al., 2010; Rao et al., 2011; Wilczak et al., 2006). Ensemble averaging has been limited to CTMs and has often focused on uncertainties in modeling ozone concentrations. However, Lee et al. (2009) showed that using an ensemble average of SA results from four receptor models and one CTM resulted in improved fitting statistics, reduced variability (compared to individual receptor models) and reduced the number of days with no impact from sources that are known to be present. In this work, we build on the work of Lee et al. (2009) by ensemble averaging results from four SA methods and assessing SA uncertainties in the ensemble results. This work updates the approach by Lee et al. (2009) in three ways: this method uses a two step process to calculate the ensemble, uncertainties are calculated using propagation of errors that

includes covariance terms, and new estimates of uncertainties are calculated for the individual SA methods that are used in the ensemble. A compelling reason to quantify uncertainties is that they can be incorporated into epidemiologic studies, which can ultimately lead to improving our understanding of the relationships between $PM_{2.5}$ sources and health outcomes. Further, they can be used to inform policy makers of the effectiveness of control measures.

2. Methods

2.1. Ensemble source apportionment

Ensemble averaging of SA results is conducted in two steps. In the first step source impact estimates and the uncertainties from the SA methods described above (see SI for more on how uncertainties were calculated for each SA method) are averaged together. In the second step, the initial ensemble is used to re-estimate SA method uncertainties, which are then used as weights to calculate an updated average. Next, uncertainties for the updated ensemble source impact are re-calculated. In part, this can address concerns that the uncertainties provided by the traditional methods are biased. This process of re-estimating SA method uncertainties and re-updating the ensemble can be further iterated if desired.

The initial ensemble average, $\overline{S}_j(t_k)$, for source j at time t_k, is calculated using a weighted average:

$$\overline{S}_j(t_k) = \frac{\sum_{l=1}^{L} w_{jl}(t_k) \bullet S_{lj}(t_k)}{\sum_{l=1}^{L} w_{jl}(t_k)} \tag{2}$$

where $w_{jl}(t_k)$ is the weight for source j from method l, and $S_{lj}(t_k)$ is the source impact for source j from method l. The weights (Equation (3)) are based on each method's source impact uncertainties and the value of N determines if and how much source impact uncertainties weight the average:

$$w_{jl} = \frac{1}{\sigma_{S_{lj}}^{N}} \tag{3}$$

While there can be any choice for the weights, here we focus on using the inverse of the individual SA methods' uncertainty squared (i.e. $N = 2$) and equal weighting ($N = 0$). We also evaluate a mixed case, in which we use equal weighting for the initial ensemble and inverse square weighting for the updated ensemble. As discussed below, our focus is on the mixed case since we find that it provides the best results over both seasons. The initial and updated ensemble average uncertainty is calculated using weighted propagation of errors that includes covariance terms (Equation (4), see SI for derivation (Taylor and Kuyatt, 1994)):

$$\sigma^2_{\text{ensemble}} = \begin{bmatrix} \frac{\frac{1}{\sigma_{S_1}^N}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^N}} & \cdots & \frac{\frac{1}{\sigma_{S_L}^N}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^N}} \end{bmatrix} \begin{bmatrix} \sigma^2_{S_1S_1} & \cdots & \sigma^2_{S_1S_L} \\ \vdots & \ddots & \vdots \\ \sigma^2_{S_1S_L} & \cdots & \sigma^2_{S_LS_L} \end{bmatrix} \times \begin{bmatrix} \frac{\frac{1}{\sigma_{S_1}^N}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^N}} & \cdots & \frac{\frac{1}{\sigma_{S_L}^N}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^N}} \end{bmatrix}^T \tag{4}$$

where S_l is the $PM_{2.5}$ impact of source j (source index not shown for clarity) from method l. The middle matrix term in the right hand side of Equation (4) is a scaled uncertainty covariance matrix which takes into account the source impact uncertainties from the individual SA methods as well as the covariance of source impacts across methods; thus, each element $\sigma^2_{S_mS_n}$, where both m and n index the SA methods that range from 1 to L, is equal to (Equation (5)):

$$\sigma^2_{S_mS_n} = \frac{\sigma_{S_m}\sigma_{S_n} * \text{Cov}(m,n)}{\sqrt{\text{Cov}(m,m) * \text{Cov}(n,n)}} \tag{5}$$

Where σ_{S_m} and σ_{S_n} are source impact uncertainties from methods m and n, Cov (m,n) is the covariance of source impacts from methods m and n and Cov (m,m) and Cov (n,n) are the variances of source impacts from methods m and n, respectively.

The root mean square error (RMSE) for each method is determined by comparison against the ensemble average (Equations (6) and (7)):

$$\text{RMSE}_{jl} = \sqrt{\frac{\sum_{k=1}^{K}\left(S_{jlk} - \overline{S}_{jk}\right)^2}{K}} \tag{6}$$

where S_{jlk} is the source impact for source j, from method l, on day k and $\overline{S}_{jk}$ is the ensemble average for source j on day k, and K is the total number of days used in the ensemble. We then set the RMSE for each method as the updated uncertainty for each day (Equation (7)):

$$\sigma'_{jlk} = \text{RMSE}_{jl} \tag{7}$$

where σ'_{jlk} is the re-estimate of the source impact uncertainties for source j, from method l, on day k. A major consequence of using Equation (7) is that for a specific source the updated source impact uncertainties are the same for each day. We set new uncertainties in this way because regression analyses between SA method source impacts (S_{jkl}) and their errors $(S_{jkl} - \overline{S}_{jk})$ from the ensemble averages found little correlation. Next, new ensemble averages and uncertainties are calculated based on the weighted propagation of errors using the updated uncertainties for each SA method. The above procedure can be done using both the absolute and fractional source impacts and we focus here on results using absolute source impacts (both approaches were tested with similar results). Finally, we evaluate the individual SA methods and the ensemble by comparing the average source impact (by source category and season). To compare uncertainties between methods, we define the *overall method uncertainty*, $(\overline{\sigma}_{S_{jkl}})$, as the root mean square average of the daily updated source impact uncertainties (Pachon et al., 2010) (Equation (8)):

$$\overline{\sigma}_{S_{jl}} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\sigma^2_{S_{jkl}}} \tag{8}$$

As discussed previously, the base case was conducted using four SA techniques. SA impacts included previous results for CMB-MM (Zheng et al., 2007) and CMAQ (Baek et al., 2005) were used as inputs into the ensemble, and we applied CMB-LGO and PMF, for 1999–2004 using speciated $PM_{2.5}$ data from the SEARCH Jefferson St (JST) monitoring site (Edgerton et al., 2005, 2006; Hansen et al., 2003). The JST data set contains daily speciated concentrations of ions (sulfate, nitrate, and ammonium), organic carbon (OC), elemental carbon (EC), and trace metals. Data also includes speciated organic molecular markers for two one month periods (July 2001, January 2002) used as part of the CMB-MM work (Zheng et al., 2007). Further details on these methods can be found in Lee et al. (2009) and references therein.

Ensemble averaging was conducted for July 2001 to represent summer, and January 2002 to represent winter (SA results from CMAQ and CMB-MM were available for these months). Source

impacts from individual SA methods used in the ensemble were binned into nine source categories (Lee et al., 2009), and included five primary sources and four secondary sources. Primary sources include gasoline vehicles (GV), diesel vehicles (DV), dust (DUST), biomass burning (BURN), and coal combustion (COAL). Secondary categories include sulfate, nitrate, ammonium and other organic carbon (Other OC), which was treated as a surrogate for secondary organic carbon (SOC). CMAQ simulations tended to be biased high for sulfate, nitrate and ammonium in winter (Dennis et al., 2010). To account for this, we did not use CMAQ results for sulfate, nitrate and ammonium to calculate the ensemble impact in the equal weighting case. In addition, we performed a sensitivity analysis of the ensemble by replacing CMB-LGO with CMB-RG since this method is more widely used. We did not use both at the same time because they are very highly correlated, relying on similar data. A second sensitivity analysis was conducted by not including CMAQ results as such results may not be as readily available or for as long of a period. However, CTM-based source impact files are becoming increasingly available (Napelenok et al., 2006; Yarwood et al., 2007).

3. Results

3.1. Ensemble source impacts and uncertainties

Comparison of the four methods shows the relative biases of these methods across sources (Fig. 1, Fig. S1 and Tables S1 and S2). For example, CMB-LGO has significantly higher SOC impacts, especially in winter. PMF tends to have higher source impacts for DV and BURN with lower impacts for SOC. CMB-MM has higher estimates of SOC in summer and higher estimates of GV in winter. CMAQ has higher DUST impacts in both seasons and higher BURN and COAL impacts in winter. The three receptor models, as expected, have very similar results for ionic species while CMAQ estimates are higher. Ensemble averaging provides daily source apportionment that results in no zero-impact days, reduced variability (Fig. 2) and updated uncertainties to the daily source impacts in the five individual source apportionment methods. Ensemble averaging overcomes some limitations of the individual SA methods (e.g., when a particular method apportions $PM_{2.5}$ mass poorly for a given source, or does not resolve a set of sources for a given day). The ensemble avoids performing poorly for any particular source, a major limitation of traditional SA methods. The ensemble, for both seasons, has the lowest estimated relative uncertainty for all cases, when averaged across all sources (i.e. the average of the overall relative uncertainties for each source) (Table 1).

In summer, the ensemble, using inverse square weighting, has the lowest overall relative uncertainties (i.e. RMSE divided by average source impact) for BURN (49%), COAL (45%), and SOC (42%) and has the second lowest overall relative uncertainties for GV (77%), DV (36%) and DUST (62%). With equal weighting, the ensemble has the lowest overall relative uncertainties for DV (38%), DUST (48%) and BURN (35%), and has the second lowest uncertainties for GV (65%), COAL (39%) and SOC (40%). With

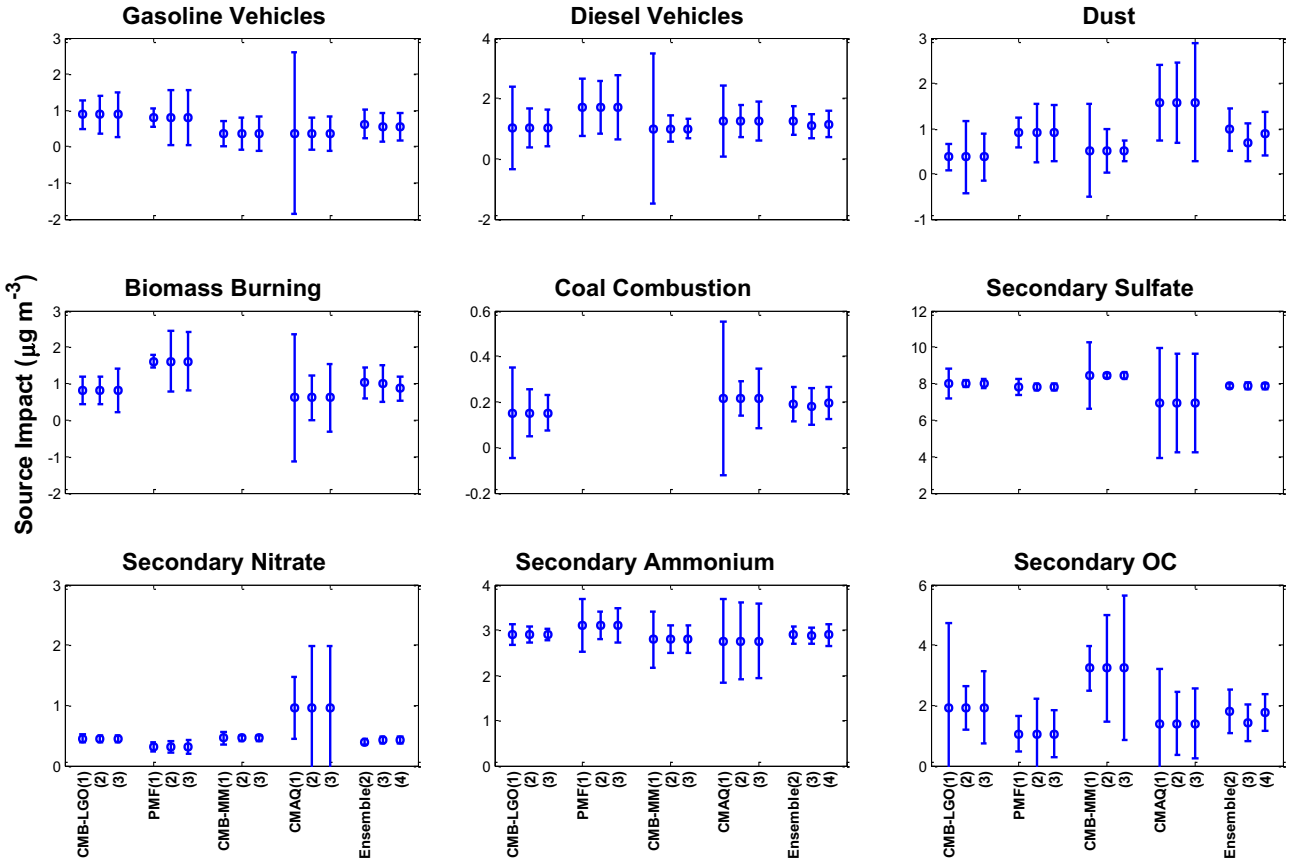

Fig. 1. Average source impacts and overall uncertainties (Equation. (8)) for the four SA methods and the ensemble (error bars represent one sigma) for July 2001 (note the changes in scales). For each method, the first data point (1) shows source impact and initial uncertainties. The second point (2) shows source impact and updated uncertainties using equal weighting (EW*). The third point (3) shows source impact and updated uncertainties using inverse square uncertainty weighting (ISW). The ensemble has three data point for the EW and ISW and a mixed case (4), respectively. The mixed case uses EW for the initial ensemble and ISW for the updated ensemble. *NOTE: the EW case does not include CMAQ results for secondary sulfate, secondary nitrate and secondary ammonium.

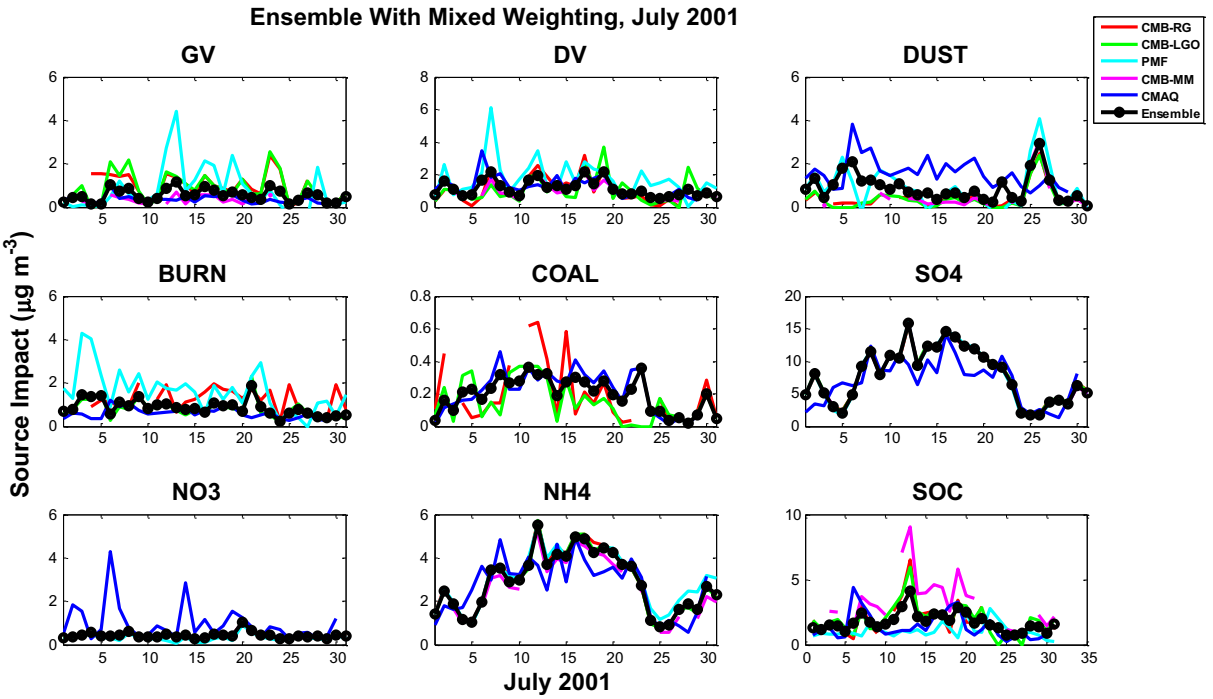

Fig. 2. Ensemble with mixed weighting for July 2001. NOTE: CMB-RG results shown here are not included in the base ensemble, but are used in the sensitivity analysis (Figs. S7 and S8).

mixed weighting, the ensemble has the lowest overall relative uncertainties for DV (36%), DUST (55%), BURN (33%), and SOC (29%). CMB-LGO has the lowest overall relative uncertainty for GV and CMAQ for COAL.

The ensemble overall relative uncertainties in winter are generally higher than in summer (Fig. S1). Also, source impacts in winter are more varied between methods than in summer leading to greater RMSEs between the SA methods and the ensemble.

Choice of weighting does not result in large differences in the overall relative uncertainties in the ensemble averages for primary sources and SOC, though there can be large differences in the magnitude of source impacts (Fig. 1 and Fig. S1 and Tables S1 and S2). For example, the average GV source impact for the summer ensemble with inverse square weighting is 0.53 ± 0.21 (μg m^{-3}) and is driven by CMB-MM which has an average impact of 0.36 μg m^{-3} and an initial overall uncertainty of 0.35 μg m^{-3}. With equal weighting, the ensemble GV has an average of 0.62 ± 0.40 μg m^{-3}. With mixed weighting, the average source impact for the ensemble is 0.55 ± 0.38 μg m^{-3}. However, source impacts across the three cases are in general highly correlated, with low correlations only for SOC in the summer and DUST and SOC in winter (Table S3).

Table 1
Average overall relative uncertainties for equal weighting (EW), inverse square weighting (ISW) and a mixed case (MIX) using both EW and ISW for summer (July 2001) and winter (January 2002). The values shown are averaged over all source categories, excluding sulfate, nitrate and ammonium. Note: For MIX, the base SA methods have uncertainties based on EW.

	CMB-LGO	PMF	CMB-MM	CMAQ	ENS.
Summer					
Initial	97%	38%	143%	222%	–
EW	81%	76%	80%	72%	45%
ISW	76%	69%	72%	93%	52%
MIX	–	–	–	–	45%
Winter					
Initial	172%	53%	143%	388%	–
EW	219%	167%	202%	282%	59%
ISW	152%	124%	88%	409%	74%
MIX	–	–	–	–	62%

The ensemble results, as compared to measured $PM_{2.5}$, reconstruct $PM_{2.5}$ mass between 75% and 110% over all cases (Table 2). A somewhat low bias may be expected because the typical range of identified sources in receptor models account for only about 80% of the inventoried $PM_{2.5}$ emissions (Baek, 2009). In the work shown here, total mass from receptor models are biased slightly low in summer and slightly high in winter. There were no results for BURN or COAL in summer for CMB-MM, which may be why the predicted to observed $PM_{2.5}$ ratio is low for that method. CMAQ results for total $PM_{2.5}$ are biased low by about 20% in the summer and high by a factor of 2 in the winter. The ensemble, when using inverse square weighting, slightly under estimates $PM_{2.5}$ in both seasons. The ensemble results correlate more strongly with measured $PM_{2.5}$ in both seasons than other methods except PMF, regardless of weighting, having R^2 values between 0.84 and 0.96 (Table 2).

Sensitivity analyses were performed by re-running the ensemble in two different ways. First, the ensemble was run using CMB-RG results in lieu of CMB-LGO. In both seasons, using mixed weighting, the ensemble results change little because CMB-RG and CMB-LGO results are highly correlated for all source categories (Figs. S7–S8). Second, we also ran the ensemble without CMAQ results (i.e. ensemble with CMB-LGO, PMF and CMB-MM). In both seasons, changes are noted for GV, DUST, BURN and SOC since CMAQ is not always strongly correlated with receptor model results. Nevertheless, the changes are within the 67% confidence intervals of the full ensemble (Figs. S9 and S10).

4. Discussion

The ensemble gives insight into how well each SA method works, and provides improved estimates of source impacts and improved estimates of source impact uncertainties by SA method. The ensemble also overcomes poor or unrealistic performance

Table 2
Ratio of calculated to observed $PM_{2.5}$ for July 2001 and January 2002. Calculated $PM_{2.5}$ is defined as the sum of source impacts from the nine source categories. Observed $PM_{2.5}$ is from JST measurements, which use a gravimetric- based method similar to the Federal Reference Method (FRM). (*NOTE: Values are recalculated here because Lee et al. (2009) used a different protocol for calculating measurement uncertainties).

		CMB-LGO	PMF	CMB-MM	CMAQ	Ensemble (EW)	Ensemble (ISW)	Ensemble (MIX)	Lee et al. (2009)*
July 2001	Avg. calc./obs. PM2.5	0.78	0.84	0.71	0.77	0.77	0.81	0.79	0.74
	St. Dev. calc./obs. $PM_{2.5}$	0.10	0.13	0.09	0.29	0.10	0.10	0.10	0.08
	R^2	0.94	0.97	0.93	0.58	0.96	0.96	0.96	0.96
	Slope (Std. Error)	0.68 (0.03)	0.66 (0.02)	0.71 (0.04)	0.53 (0.09)	0.63 (0.02)	0.68 (0.03)	0.65 (0.03)	0.65 (0.02)
	Intercept (Std. Error)	1.77 (0.78)	2.95 (0.53)	−0.08 (1.10)	4.28 (2.05)	2.41 (0.57)	2.30 (0.62)	2.42 (0.61)	1.54 (0.59)
	Reduced Chi-Square	9	60	54	594	158	51	83	20
January 2002	Avg. calc./obs. $PM_{2.5}$	0.97	1.02	1.01	2.05	0.98	1.13	1.10	0.99
	St. Dev. calc./obs. $PM_{2.5}$	0.16	0.14	0.15	0.84	0.17	0.19	0.18	0.21
	R^2	0.83	0.88	0.84	0.34	0.84	0.85	0.84	0.76
	Slope (Std. Error)	0.74 (0.06)	0.90 (0.06)	0.76 (0.07)	1.21 (0.32)	0.65 (0.05)	0.78 (0.06)	0.76 (0.06)	0.65 (0.07)
	Intercept (Std. Error)	2.53 (0.83)	1.26 (0.84)	2.84 (0.91)	8.86 (4.28)	3.61 (0.70)	3.78 (0.81)	3.68 (0.83)	3.69 (0.92)
	Reduced Chi-Square	7	107	72	1661	212	71	124	805

(e.g., high day to day variability or days where source impacts are zero for sources known to present). The ensemble allows for comparison of uncertainties by calculating them in a consistent manner and avoids the need for bootstrapping methods or poorly characterized uncertainties in source profiles. For example, CMB-MM and PMF have very different GV impacts in winter (2.42 and 1.07 μg m^{-3}) with low overall uncertainties when calculated using traditional methods (0.44 and 0.33 μg m^{-3}). Thus, while the average source impacts are very different, the overall relative uncertainties are similar, 26 and 31%, respectively, making it difficult to determine which model provides more accurate estimates. The ensemble reconciles this inconsistency, suggesting uncertainties in both PMF and CMB-MM are larger. In another study using CMB-MM, it was shown that GV source impact uncertainties are sensitive to the percentage of high emitting vehicles for weekend traffic; when smoker vehicles are assumed to be 5% of the GV fleet, GV source impact uncertainties on Saturdays decrease from 51% to 25% while for other days they are below 17% (Lough and Schauer, 2007). Nevertheless, assumptions of fleet composition, vehicle type, driving conditions and driver behavior, all of which are significant sources of uncertainty, affect these types of analyses. Therefore, the uncertainties in Lough and Schauer (2007) should be viewed as tighter than achieved in general applications. In PMF, uncertainties are calculated by bootstrapping, which reflects how similar the bootstrapped data set's correlation structure is to the original data set, and may not reflect the actual factor contribution uncertainty.

Inverse square weighting leads to the ensemble being heavily influenced by a particular method (e.g., CMB-MM for GV), having initial uncertainties that are apparently biased low. This indicates that, given no other information, all methods should be weighted equally, (i.e., using equal weighting). When using mixed weighting, the base case SA methods are also treated equally, but the updated ensemble is weighted by the new uncertainties to base case SA methods. We recommend mixed weighting because this incorporates the new uncertainties as weights to the updated ensemble average and performed well in the evaluation measures.

Ensemble averaging also allows uncertainties in CTM-based source impacts to be readily estimated. To our knowledge, this is the first work to estimate $PM_{2.5}$ source impact uncertainties in CMAQ. As new techniques are developed to estimate CTM uncertainties, ensemble averaging can provide a means to evaluate these estimates.

Another approach to evaluating the ensemble quantitatively is to compare our results with estimates of secondary organic carbon (SOC) impacts from other work (Table 3). Recently, Pachon et al. (2010) found that the regression method for estimating SOC had the lowest overall relative uncertainty, when compared to the EC Tracer Method, CMB-RG and PMF. They showed that both CMB-RG and PMF have high overall uncertainties that ranged from 47% to 56% for CMB-RG and 59% to 120% for PMF in summer and winter, respectively. The regression method estimated SOC to be 1.68 ± 0.14 μg m^{-3} and 0.80 ± 0.11 μg m^{-3} in July 2001 and January 2002, respectively. The ensemble estimates are comparable to the regression method's average impact and overall uncertainty for July 2001, but are higher for January 2002 (Table 3). The correlation of the ensemble-based SOC with the regression-based SOC is very encouraging since the regression method includes ozone concentrations, which are not used in any of the receptor models included in the ensemble. In addition, the regression method was more strongly correlated with measured water-soluble organic carbon (WSOC), which is hypothesized to be primarily from secondary reactions. This indicates a better fit with SOC than the other methods. WSOC is, likewise, not used in any of the ensemble methods. Further, it is interesting that the correlation

Table 3
Secondary Organic Carbon (SOC) Results for July 2001 and January 2002 (μg m^{-3}). NOTE: Ensemble with MIX uses EW uncertainties in base case SA methods.

	Summer					Winter				
			Uncertainty (±σ)					Uncertainty (±σ)		
	Average SOC		Ens. with EW	Ens. with ISW	Ens. with MIX	Average SOC		Ens. with EW	Ens. with ISW	Ens. with MIX
CMB-LGO	1.93	±	0.72	1.19	–	2.43	±	1.21	2.00	–
PMF	1.06	±	1.17	0.77	–	0.69	±	1.05	0.54	–
CMB-MM	3.23	±	1.73	2.39	–	1.89	±	0.89	1.77	–
CMAQ	1.40	±	1.06	1.15	–	0.97	±	0.71	0.76	–
Ensemble with EW	1.81	±	0.73	–	–	1.45	±	0.68	–	–
Ensemble with ISW	1.42	±	–	0.60	–	0.90	±	–	0.48	–
Ensemble with MIX	1.76	±	–	–	0.60	1.31.	±	–	–	0.63

between the ensemble SOC and the PMF SOC is very low ($R^2 = 0.01$ for July 2001).

To evaluate the choice of weighting, we conducted York regression (Saylor et al., 2006; York et al., 2004) between the ensemble and the regression method SOC impacts and found that mixed weighting reproduced regression method results better that equal or inverse square weighting ($R^2 = 0.82$ and slope = 0.87 for summer 2001) (Fig. S9). A similar analysis was performed for January 2002 (Fig. S10). It has been suggested that CMB based methods overestimate SOC because primary OC from some sources are not considered (Zheng et al., 2007; Zheng et al., 2002). Updated emissions information that include improved estimates of primary OC emissions in the winter, which suggest that gasoline vehicles emit more OC in cold weather than is captured in current inventories, can significantly alter how OC is apportioned (Donahue et al., 2009; Subramanian et al., 2006). It is expected that improved source profiles for CMB based methods and improved emissions processing in CTMs should lead to improved correlation of SOC estimates between the ensemble and the regression methods.

5. Conclusions

Commonly used methods to apportion sources of $PM_{2.5}$ have a number of issues that complicate their appropriate use. Results from the application of different SA methods can disagree substantially. Furthermore, calculation of source impact uncertainties varies from method to method, leading to very different uncertainty estimates and making inter-comparisons of source impacts and their associated uncertainties difficult. Here we average an ensemble of SA methods, which includes two CMB methods, PMF and CMAQ to estimate updated source impacts and uncertainties. Three weighting cases, equal weighting, inverse square weighting and a mixed case are evaluated.

Ensemble averaging results in source impact estimates that have reduced variability compared to individual SA methods, avoids zero impact days and resolves source impacts for all days. The choice of weighting impacts ensemble-based average source impacts and uncertainties, but in general ensemble source impact uncertainties are lower or very comparable with individual SA method uncertainties. Over both seasons, mixed weighting in the ensemble reproduces $PM_{2.5}$ better than equal or inverse square weighting and agrees better with SOC estimates from a separate approach (Pachon et al., 2010). In the absence of any prior information which would indicate otherwise, mixed weighting should be used.

The ensemble method provides updated uncertainties for the individual SA methods that are calculated in a consistent way across methods. In general, CMB-LGO and CMB-MM overall uncertainties, averaged over primary sources and SOC, decrease in summer and increase in winter as compared to those found using the traditional approach for these methods. The ensemble method also provides a way to estimate source apportionment uncertainties in CMAQ. CMAQ source impact uncertainties are comparable to other SA methods for GV and SOC and larger than other methods for DV, DUST and BURN.

Acknowledgments

This publication was made possible in part by USEPA STAR grants R833626, R833866, R834799 and RD83479901. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. We also acknowledge the Southern Company for their support and thank Eric Edgerton of ARA, Inc. for access to the SEARCH data. We also acknowledge the contribution of our colleagues at the Rollins School of Public Health at Emory University for helpful discussions.

Appendix A. Supplementary information

Supplementary information related to this article can be found online at http://dx.doi.org/10.1016/j.atmosenv.2012.07.031.

References

Baek, J., Park, S.K., Hu, Y., Russell, A.G., 2005. Source Apportionment of Fine Organic Aerosol Using CMAQ Tracers. In: The 4th Annual CMAS Models-3 Users' Conference. CMAS, Durham, NC.

Baek, J., 2009. Improving Aerosol Simulations: Assessing and Improving Emissions and Secondary Organic Aerosol Formation in Air Quality Modeling. Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA.

Barregard, L., Sallsten, G., Gustafson, P., Andersson, L., Johansson, L., Basu, S., Stigendal, L., 2006. Experimental exposure to wood-smoke particles in healthy humans: effects on markers of inflammation, coagulation, and lipid peroxidation. Inhalation Toxicology 18, 845–853.

Barrett, E.G., Henson, R.D., Seilkop, S.K., McDonald, J.D., Reed, M.D., 2006. Effects of hardwood smoke exposure on allergic airway inflammation in mice. Inhalation Toxicology 18, 33–43.

Brinkman, G., Vance, G., Hannigan, M.P., Milford, J.B., 2006. Use of synthetic data to evaluate positive matrix factorization as a source apportionment tool for $PM_{2.5}$ exposure data. Environmental Science & Technology 40, 1892–1901.

Byun, D.W., Young, J., Gipson, G., Godowitch, J., Binkowski, F., Roselle, S., Benjey, B., Pleim, J., Ching, J., Novak, J., Coats, C., Odman, T., Hanna, A., Alapaty, K., Mathur, R., McHenry, J., Shankar, U., Fine, S., Xiu, A.J., Jang, C., Amer Meteorol Soc, A.M.S., 1998. Description of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, 264–268 pp.

Cass, G.R., 1998. Organic molecular tracers for particulate air pollution sources. TracTrends in Analytical Chemistry 17, 356–366.

Christensen, W.F., Amemiya, Y., 2003. Modeling and prediction for multivariate spatial factor analysis. Journal of Statistical Planning and Inference 115, 543–564.

Christensen, W.F., Gunst, R.F., 2004. Measurement error models in chemical mass balance analysis of air quality data. Atmospheric Environment 38, 733–744.

Christensen, W.F., Schauer, J.J., 2008. Impact of species uncertainty perturbation on the solution stability of positive matrix factorization of atmospheric particulate matter data. Environmental Science & Technology 42, 6015–6021.

Christensen, W.E., Schauer, J.J., Lingwall, J.W., 2006. Iterated confirmatory factor analysis for pollution source apportionment. Environmetrics 17, 663–681.

Cohan, D.S., Hakami, A., Hu, Y.T., Russell, A.G., 2005. Nonlinear response of ozone to emissions: source apportionment and sensitivity analysis. Environmental Science & Technology 39, 6739–6748.

Delle Monache, L., Deng, X.X., Zhou, Y.M., Stull, R., 2006. Ozone ensemble forecasts: 1. A new ensemble design. Journal of Geophysical Research-Atmospheres 111.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S.T., Scheffe, R., Schere, K., Steyn, D., Venkatram, A., 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. Environmental Fluid Mechanics 10, 471–489.

Dockery, D.W., Pope, C.A., Xu, X.P., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G., Speizer, F.E., 1993. An association between air-pollution and mortality in 6 United-States cities. New England Journal of Medicine 329, 1753–1759.

Donahue, N.M., Robinson, A.L., Pandis, S.N., Kroll, J.H., Worsnop, D.L., 2009. Rethinking organic aerosols: semivolatile emissions and photochemical aging. Geochimica Et Cosmochimica Acta 73, A299.

Edgerton, E.S., Hartsell, B.E., Saylor, R.D., Jansen, J.J., Hansen, D.A., Hidy, G.M., 2005. The southeastern aerosol research and characterization study: part II. Filter-based measurements of fine and coarse particulate matter mass and composition. Journal of the Air & Waste Management Association 55, 1527–1542.

Edgerton, E.S., Hartsell, B.E., Saylor, R.D., Jansen, J.J., Hansen, D.A., Hidy, G.M., 2006. The southeastern aerosol research and characterization study, part 3: continuous measurements of fine particulate matter mass and composition. Journal of the Air & Waste Management Association 56, 1325–1341.

Grahame, T., Hidy, G.M., 2007. Pinnacles and pitfalls for source apportionment of potential health effects from airborne particle exposure. Inhalation Toxicology 19, 727–744.

Hansen, D.A., Edgerton, E.S., Hartsell, B.E., Jansen, J.J., Kandasamy, N., Hidy, G.M., Blanchard, C.L., 2003. The southeastern aerosol research and characterization study: part 1-overview. Journal of the Air & Waste Management Association 53, 1460–1471.

Henry, R.C., 1987. Current factor-analysis receptor models are Ill-posed. Atmospheric Environment 21, 1815–1820.

Henry, R.C., 1997. History and fundamentals of multivariate air duality receptor models. Chemometrics and Intelligent Laboratory Systems 37, 37–42.

Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. Chemometrics and Intelligent Laboratory Systems 65, 179–189.

Ito, K., Christensen, W.F., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Larson, T.V., Neas, L., Hopke, P.K., Thurston, G.D., 2006. PM source

apportionment and health effects: 2. An investigation of intermethod variability in associations between source-apportioned fine particle mass and daily mortality in Washington, DC. Journal of Exposure Science and Environmental Epidemiology 16, 300–310.
Koo, B., Wilson, G.M., Morris, R.E., Dunker, A.M., Yarwood, G., 2009. Comparison of source apportionment and sensitivity analysis in a particulate matter air quality model. Environmental Science & Technology 43, 6669–6675.
Laden, F., Neas, L.M., Dockery, D.W., Schwartz, J., 2000. Association of fine particulate matter from different sources with daily mortality in six US cities. Environmental Health Perspectives 108, 941–947.
Lee, S., Russell, A.G., 2007. Estimating uncertainties and uncertainty contributors of CMB $PM_{2.5}$ source apportionment results. Atmospheric Environment 41, 9616–9624.
Lee, S., Liu, W., Wang, Y.H., Russell, A.G., Edgerton, E.S., 2008. Source apportionment of $PM_{2.5}$: comparing PMF and CMB results for four ambient monitoring sites in the southeastern United States. Atmospheric Environment 42, 4126–4137.
Lee, D., Balachandran, S., Pachon, J., Shankaran, R., Lee, S., Mulholland, J.A., Russell, A.G., 2009. Ensemble-trained $PM_{2.5}$ source apportionment approach for health studies. Environmental Science & Technology 43, 7023–7031.
Lough, G.C., Schauer, J.J., 2007. Sensitivity of source apportionment of urban particulate matter to uncertainty in motor vehicle emissions profiles. Journal of the Air & Waste Management Association 57, 1200–1213.
Mar, T.F., Norris, G.A., Koenig, J.Q., Larson, T.V., 2000. Associations between air pollution and mortality in Phoenix, 1995–1997. Environmental Health Perspectives 108, 347–353.
Mar, T.F., Ito, K., Koenig, J.Q., Larson, T.V., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Neas, L., Stolzel, M., Paatero, P., Hopke, P.K., Thurston, G.D., 2006. PM source apportionment and health effects. 3. Investigation of inter-method variations in associations between estimated source contributions of $PM_{2.5}$ and daily mortality in Phoenix, AZ. Journal of Exposure Science and Environmental Epidemiology 16, 311–320.
Marmur, A., Unal, A., Mulholland, J.A., Russell, A.G., 2005. Optimization-based source apportionment of $PM_{2.5}$ incorporating gas-to-particle ratios. Environmental Science & Technology 39, 3245–3254.
Marmur, A., Mulholland, J., Kim, E., Hopke, P., Sarnat, J., Klein, M., Tolbert, P., Russell, A., 2006a. Comparing results from several $PM_{2.5}$ source-apportionment methods for use in a time-series health study. Epidemiology 17, S200.
Marmur, A., Park, S.K., Mulholland, J.A., Tolbert, P.E., Russell, A.G., 2006b. Source apportionment of $PM_{2.5}$ in the southeastern United States using receptor and emissions-based models: conceptual differences and implications for time-series health studies. Atmospheric Environment 40, 2533–2551.
Napelenok, S.L., Cohan, D.S., Hu, Y.T., Russell, A.G., 2006. Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3D/PM). Atmospheric Environment 40, 6112–6121.
Ostro, B., Feng, W.Y., Broadwin, R., Green, S., Lipsett, M., 2007. The effects of components of fine particulate air pollution on mortality in California: results from CALFINE. Environmental Health Perspectives 115, 13–19.
Paatero, P., Tapper, U., 1994. Positive matrix factorization – a nonnegative factor model with optimal utilization of error-estimates of data values. Environmetrics 5, 111–126.
Paatero, P., Hopke, P.K., Hoppenstock, J., Eberly, S.I., 2003. Advanced factor analysis of spatial distributions of $PM_{2.5}$ in the eastern United States. Environmental Science & Technology 37, 2460–2476.
Pachon, J.E., Balachandran, S., Hu, Y.T., Weber, R.J., Mulholland, J.A., Russell, A.G., 2010. Comparison of SOC estimates and uncertainties from aerosol chemical composition and gas phase data in Atlanta. Atmospheric Environment 44, 3907–3914.
Rao, S.T., Galmarini, S., Puckett, K., 2011. Air Quality Model Evaluation International Initiative (AQMEII) advancing the state of the science in regional photochemical modeling and its applications. Bulletin of the American Meteorological Society 92, 23–30.
Rizzo, M.J., Scheff, P.A., 2007. Fine particulate source apportionment using data from the USEPA speciation trends network in Chicago, Illinois: comparison of two source apportionment models. Atmospheric Environment 41, 6276–6288.
Sarnat, J.A., Marmur, A., Klein, M., Kim, E., Russell, A.G., Sarnat, S.E., Mulholland, J.A., Hopke, P.K., Tolbert, P.E., 2008. Fine particle sources and cardiorespiratory morbidity: an application of chemical mass balance and factor analytical source-apportionment methods. Environmental Health Perspectives 116, 459–466.
Saylor, R.D., Edgerton, E.S., Hartsell, B.E., 2006. Linear regression techniques for use in the EC tracer method of secondary organic aerosol estimation. Atmospheric Environment 40, 7546–7556.
Stolzel, M., Laden, F., Dockery, D.W., Schwartz, J., Kim, E., Hopke, P.K., Neas, L.M., 2005. Source apportionment of fine and coarse particulate matter and daily mortality in two US cities – a comparison of different methods. Epidemiology 16, S95.
Subramanian, R., Donahue, N.M., Bernardo-Bricker, A., Rogge, W.F., Robinson, A.L., 2006. Contribution of motor vehicle emissions to organic carbon and fine particle mass in Pittsburgh, Pennsylvania: effects of varying source profiles and seasonal trends in ambient marker concentrations. Atmospheric Environment 40, 8002–8019.
Tauler, R., Viana, M., Querol, X., Alastuey, A., Flight, R.M., Wentzell, P.D., Hopke, P.K., 2009. Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies. Atmospheric Environment 43, 3989–3997.
Taylor, B.N., Kuyatt, C.E., 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. National Institute for Standards and Technology, Gaithersburg, MD.
Thurston, G.D., 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., Hopke, P.K., 2005. Workgroup report: workshop on source apportionment of particulate matter health effects – Intercomparison of results and implications. Environmental Health Perspectives 113, 1768–1774.
U.S.EPA, 2004. In: EPA-CMB8.2 User's Manual. U.S.E.P.A. Office of Air Quality & Standards, Research Triangle Park, NC. Publication No. EPA-452/R-04–011.
U.S.EPA, 2008. In: EPA-PMF 3.0 User's Manual. U.S.E.P.A. National Exposure Research Laboratory, Research Triangle Park, NC.
U.S.EPA, 2009. Integrated Science Assessment for Particulate Matter (Final Report).
Wang, Z.S., Chien, C.J., Tonnesen, G.S., 2009. Development of a tagged species source apportionment algorithm to characterize three-dimensional transport and transformation of precursors and secondary pollutants. Journal of Geophysical Research-Atmospheres 114.
Watson, J.G., Cooper, J.A., Huntzicker, J.J., 1984. The effective variance weighting for least-squares calculations applied to the mass balance receptor model. Atmospheric Environment 18, 1347–1355.
Wilczak, J., McKeen, S., Djalalova, I., Grell, G., Peckham, S., Gong, W., Bouchet, V., Moffet, R., McHenry, J., McQueen, J., Lee, P., Tang, Y., Carmichael, G.R., 2006. Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during the summer of 2004. Journal of Geophysical Research-Atmospheres 111, 15.
Yang, Y.J., Wilkinson, J.G., Odman, M.T., Russell, A.G., 2000. Ozone sensitivity and uncertainty analysis using DDM-3D in a photochemical air quality model. In: Gryning, S.E., Batchvarova, E. (Eds.), Air Pollution Modeling and Its Application, vol. Xiii, pp. 183–194.
Yarwood, G., Morris, R.E., Wilson, G.M., 2007. Particulate matter source apportionment technology (PSAT) in the CAMx photochemical grid model. In: Borrego, C., Norman, A.L. (Eds.), Air Pollution Modeling and Its Applications, vol. XVII, pp. 478–492.
York, D., Evensen, N.M., Martinez, M.L., Delgado, J.D., 2004. Unified equations for the slope, intercept, and standard errors of the best straight line. American Journal of Physics 72, 367–375.
Zheng, M., Cass, G.R., Schauer, J.J., Edgerton, E.S., 2002. Source apportionment of $PM_{2.5}$ in the southeastern United States using solvent-extractable organic compounds as tracers. Environmental Science & Technology 36, 2361–2371.
Zheng, M., Cass, G.R., Ke, L., Wang, F., Schauer, J.J., Edgerton, E.S., Russell, A.G., 2007. Source apportionment of daily fine particulate matter at Jefferson street, Atlanta, GA, during summer and winter. Journal of the Air & Waste Management Association 57, 228–242.