# Comparison of SOC estimates and uncertainties from aerosol chemical composition and gas phase data in Atlanta

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

[a] Georgia Institute of Technology, School of Civil and Environmental Engineering, Atlanta, GA 30332, USA
[b] Georgia Institute of Technology, School of Earth and Atmospheric Sciences, Atlanta, GA 30332, USA
[c] Universidad de La Salle, Programa de Ingeniería Ambiental, Bogotá, Colombia

ABSTRACT

In the Southeastern US, organic carbon (OC) comprises about 30% of the PM$_{2.5}$ mass. A large fraction of OC is estimated to be of secondary origin. Long-term estimates of SOC and uncertainties are necessary in the evaluation of air quality policy effectiveness and epidemiologic studies. Four methods to estimate secondary organic carbon (SOC) and respective uncertainties are compared utilizing PM$_{2.5}$ chemical composition and gas phase data available in Atlanta from 1999 to 2007. The elemental carbon (EC) tracer and the regression methods, which rely on the use of tracer species of primary and secondary OC formation, provided intermediate estimates of SOC as 30% of OC. The other two methods, chemical mass balance (CMB) and positive matrix factorization (PMF) solve mass balance equations to estimate primary and secondary fractions based on source profiles and statistically-derived common factors, respectively. CMB had the highest estimate of SOC (46% of OC) while PMF led to the lowest (26% of OC). The comparison of SOC uncertainties, estimated based on propagation of errors, led to the regression method having the lowest uncertainty among the four methods. We compared the estimates with the water soluble fraction of the OC, which has been suggested as a surrogate of SOC when biomass burning is negligible, and found a similar trend with SOC estimates from the regression method. The regression method also showed the strongest correlation with daily SOC estimates from CMB using molecular markers. The regression method shows advantages over the other methods in the calculation of a long-term series of SOC estimates.

© 2010 Elsevier Ltd. All rights reserved.

## 1. Introduction

In the Southeastern US, OC comprises approximately 30% of the PM$_{2.5}$ mass. OC can be of both primary and secondary origin. Primary OC (POC) is mainly emitted from fossil fuel combustion in stationary, area and mobile sources, and biomass combustion (e.g., forest fires). In Atlanta, the major sources of POC are motor vehicles and biomass burning (Lee et al., 2007; Zheng et al., 2002). Secondary OC (SOC) is formed in the atmosphere by photochemical reactions of volatile organic compounds (VOCs) of biogenic and anthropogenic origin followed by the condensation of reaction products onto particles (Kroll and Seinfeld, 2008). At present, there is no

measurement approach that definitively differentiates between POC and SOC, though detailed speciation can identify specific components that would be dominantly primary or secondary. Epidemiologic studies suggest differences in health outcomes associated with POC attributed to mobile and biomass burning sources, versus other OC, presumably SOC (Sarnat et al., 2008).

Typically, as part of the Speciation Trends Network for example, OC in PM$_{2.5}$ is measured on 24-hour filter-based samples, although greater resolution is possible using semi-continuous *in situ* instruments (Solomon et al., 2000). The amount of OC on the filters is quantified using thermal-optical techniques (Chow et al., 1993; Turpin et al., 2000). These techniques are designed to measure the total OC fraction, and do not distinguish between primary and secondary components. Since the formation of SOC leads to oxygenated, polar compounds, it has been suggested that the water soluble fraction of the OC (WSOC) is a surrogate for the SOC when biomass burning impact is negligible (Hennigan et al., 2008; Weber et al., 2007). WSOC can be measured in the laboratory using

PM$_{2.5}$ filters and posteriori separation of the water soluble fraction or in-situ using a Particle Into Liquid Sampler (PILS) that captures particles in water from where the carbonaceous fraction is quantified using a Total Organic Carbon (TOC) analyzer (Sullivan et al., 2006). Summertime measurements in Atlanta find that WSOC is about 55–65% of total OC.

Different methods have been used to estimate SOC. Methods that rely on the use of tracer species of primary activity and secondary photochemistry formation include the EC tracer and regression methods. Receptor models have also been used to estimate primary and secondary fractions in the PM$_{2.5}$, notably Chemical Mass Balance (CMB) and Positive Matrix Factorization (PMF) methods. While estimates from chemical transport models (CTM) are available, simulated SOC values are viewed as highly uncertain, and likely biased (Eder and Yu, 2006; Tesche et al., 2006). Some studies have used organic molecular markers and specific compounds to separate the POC and SOC fractions (Zheng et al., 2006). Given that speciated organic compound concentrations are not widely available and that their measurement is resource intensive, methods that rely on typically available PM$_{2.5}$ speciation and gaseous data are preferable. Such methods are used in this study to construct multi-year time series of pollutants for epidemiologic analysis and air quality policy effectiveness studies.

Estimates of SOC in Atlanta vary between methods and have focused on different periods of time from one or two months during summer and winter to three years (Blanchard et al., 2008; Lee et al., 2008b; Lim and Turpin, 2002; Marmur et al., 2005; Zheng et al., 2007; Zheng et al., 2002). These studies have defined uncertainties in the SOC estimates as the standard deviation of the mean, with the exception of Blanchard et al. (2008) who estimated uncertainties as one-half the range from alternative regressions. The standard deviation represents a good measure of the variation in SOC estimates but does not consider the different types of uncertainties involved in the SOC calculation (e.g. ambient measurements, source profiles, regression coefficients, primary ratios, fitting methods). Here, we assess and compare the uncertainty in the SOC estimates from four different methods, considering uncertainties in input datasets and methods.

## 2. Methods

Nine-year time series of SOC concentrations and respective uncertainties are estimated using four methods: EC tracer (Turpin and Huntzicker, 1991), regression (Blanchard et al., 2008), CMB (Watson et al., 1984) and PMF (Paatero and Tapper, 1994). The results are then compared under the following metrics in order to choose the most accurate estimate: uncertainties (lowest uncertainty preferred) estimated by propagation of errors (Bevington and Robinson, 2003), seasonal estimates (summer SOC should exceed winter), day-to-day variability (smooth for a secondary pollutant), comparison with related work (i.e. molecular marker-based CMB) and comparison with WSOC measurements (as a surrogate of SOC).

### 2.1. EC tracer method

The EC tracer method consists of estimating a primary OC/EC ratio during periods when SOC is expected to be negligible (e.g. night, winter, overcast, clean background, minimal long range transport).

$$POC = (OC/EC)_p * EC + (OC)_{nc} \qquad (1)$$

$$SOC = OC - POC \qquad (2)$$

Here (OC)$_{nc}$ is the non-combustion contribution to the OC, from sources such as vegetative detritus, tire wear and industrial processes. Equation (2) can result in negative values of SOC, in which case SOC is set to zero.

Typically, (OC/EC)$_p$ in Eq. (1) is determined from the linear regression between OC and EC (e.g. Demming regression, Saylor et al., 2006) over a long period of time, with the intercept determining (OC)$_{nc}$. Alternatively, (OC/EC)$_p$ can be estimated from days when primary or secondary activity is more pronounced (such as in Cabada et al., 2004). For this study, the (OC/EC)$_p$ ratio was estimated in three steps: i) we selected days from 1999–2007 with low photochemical activity, defined as days with O$_3$ (max 8 h average) concentration below the 25th percentile, O$_3$ < 41 ppb in summer and O$_3$ < 20 ppb in winter ii) we plotted time series of OC, EC, OC/EC, CO and O$_3$ and identified days when primary activity was more pronounced (an example of this selection is shown in Fig. 1) and iii) we computed averaged OC/EC ratios on those days, obtaining 1.7 for summer and 2.4 for winter. The application of a unique (OC/EC)$_p$ ratio for year-round estimates may not account for seasonal variation (Snyder et al., 2009). The larger winter value suggests an increased influence of biomass burning which has a higher OC/EC primary emissions ratio. Other studies in the area have found similar values for these ratios. Using time-resolved OC fractions in summer time for the estimation of SOC, Lim and Turpin (2002) found a ratio (OC/EC)$_p$ of 1.8 as reasonable and 2.1 as the upper limit. Using a multiscale air quality model over the United States, Yu et al. (2007) found (OC/EC)$_p$ ratios for Atlanta of 1.76 in summer and 2.76 in winter. For this study, the EC tracer refers exclusively to the application of the method using summer/winter ratios. Variation of the (OC/EC)$_p$ ratio on time scales less than half a year is beyond the scope of this study.

The initial estimate of the uncertainty ($\sigma$) is calculated using propagation of relative errors.

$$\sigma_{POC}^2 = \sigma_{EC}^2 \left(\frac{OC}{EC}\right)_p^2 + \sigma_{\left(\frac{OC}{EC}\right)_p}^2 EC^2 + \sigma_{OC_{nc}}^2 \qquad (3)$$

Here, the uncertainty in the EC and OC components was calculated using the procedure of Polissar et al. (1998). Briefly, the uncertainty
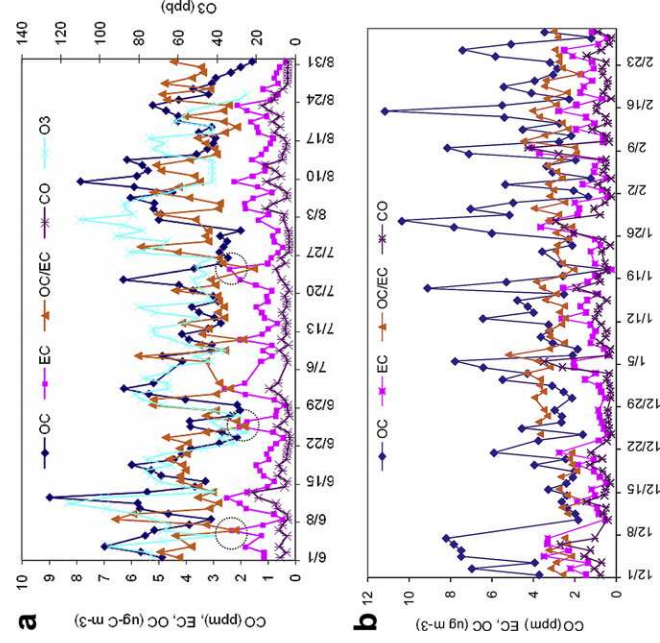


**Fig. 1.** Time series of primary and secondary species and OC/EC ratio during (a) summer 2002 (b) and winter 2002. During summer, the circled days have a decrease in O$_3$ concentrations, and high levels of OC, EC and CO, denoting a predominance of primary activity. For those days the average (OC/EC) ratio was 1.7. During winter, days with ozone concentrations below the 25th percentile had an average (OC/EC) ratio of 2.4.

in the observed concentrations was set as the sum of the analytical uncertainty times the concentration plus one-third of the detection limit (DL) value. The uncertainty in the primary (OC/EC) ratio was defined as one standard deviation of the estimated ratios. The uncertainty in the secondary organic fraction was calculated by propagating the uncertainties in the POC fraction and the measured OC.

$$\sigma_{SOC}^2 = \sigma_{OC}^2 + \sigma_{POC}^2 \qquad (4)$$

The root mean square average of the uncertainty for the POC and SOC estimates over the nine-year period of time is calculated as

$$\overline{\sigma_i} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\sigma_{ij}^2} \qquad (5)$$

where $\sigma_{ij}$ is the uncertainty in the $i$th parameter on the $j$th day, with a total of $N$ days.

### 2.2. Regression method

The regression method uses tracers of primary emissions (EC, 8-h average CO) as well as photochemical activity (8-h average $O_3$, sulfate $SO_4$, nitrate $NO_3$) to determine POC and SOC. We modified this approach by adding potassium (XRF K from the SEARCH data) to identify POC from biomass burning which accounts for a large part of the POC in the southeastern US (Kim et al., 2003, 2004):

$$OC = a + b*EC + c*CO + d*O_3 + e*lag(O_3) + f*SO_4 \\ + g*NO_3 + h*K \qquad (6)$$

$$POC_O = b*EC + c*CO + h*K \qquad (7)$$

$$SOC_O = d*O_3 + e*lag(O_3) + f*SO_4 + g*NO_3 \qquad (8)$$

The regression coefficients ($a$–$h$) are determined using least square fitting (LSF), and each coefficient is evaluated for its statistical significance. Here $POC_O$ and $SOC_O$ are initial estimates for each day. To guarantee that the sum of POC and SOC is equal to the observed OC, we distributed the initial estimates based on the mass fraction ratios.

$$POC = \left(\frac{POC_O}{POC_O + SOC_O}\right)OC \qquad (9)$$

$$SOC = \left(\frac{SOC_O}{POC_O + SOC_O}\right)OC \qquad (10)$$

On a year-round basis, multivariate regression of OC with EC, CO, K, $SO_4$, $NO_3$ and $O_3$ led to an $R^2 = 0.65$ ($n = 2921$), suggesting common sources between OC and primary and secondary pollutants. In summer, regression of OC with EC, CO, K, $SO_4$, $NO_3$ and $O_3$ results in a slightly stronger statistical fit ($R^2 = 0.68$, $n = 1476$). The regression coefficient for $NO_3$ was not statistically significant ($p > 0.05$) and the independent term 'a' (in Eq. (6)) had the lowest significance; therefore, the regression was performed with an intercept of zero. In this case, the significance of secondary tracers, such as $O_3$ ($t$-Stat = 22.7, $p < 0.01$), is comparable with primary tracers, such as EC ($t$-Stat = 21.4, $p < 0.01$). In winter, regression of OC with EC, CO, K, $NO_3$ and $O_3$ results in a stronger statistical fit ($R^2 = 0.78$, $n = 1427$) than the summer regression. The independent term 'a' and the $SO_4$ regression coefficient were not statistically significant ($p > 0.05$; EC ($t$-Stat = 26.8, $p < 0.01$) and K ($t$-Sat = 13.6, $p < 0.01$) were the most significant coefficients, suggesting a strong impact of mobile sources and biomass burning on OC. Hereafter the regression method will

refer to the application of the method using separate summer/winter regression results.

We calculate the uncertainty by propagating errors for every term in the regression method. The uncertainty in each regression coefficient (i.e. $\sigma_b$) was obtained from the standard error in the regression analysis and the uncertainty in the species concentration (i.e. $\sigma_{EC}$) was estimated using the procedure of Polissar et al. (1998). The uncertainties were propagated to find daily uncertainties in POC and SOC:

$$(\sigma_{POC})^2 = (\sigma_{EC})^2*b^2 + (\sigma_{CO})^2*c^2 + (\sigma_K)^2*h^2 + (\sigma_b)^2*EC^2 \\ + (\sigma_c)^2*CO^2 + (\sigma_h)^2*K^2 \qquad (11)$$

$$(\sigma_{SOC})^2 = (\sigma_{O3})^2*d^2 + (\sigma_{SO4})^2*f^2 + (\sigma_{NO3})^2*g^2 + (\sigma_d)^2*O_3^2 \\ + (\sigma_f)^2*SO_4^2 + (\sigma_g)^2*NO_3^2 \qquad (12)$$

The average uncertainties for the POC and SOC estimates, over the nine-year period, are calculated using the root mean square average (Eq. (5)).

### 2.3. Chemical mass balance (CMB)

To estimate the SOC fraction in the CMB model, we include six primary source profiles and four profiles that represent secondary species formation (Marmur et al., 2005). $PM_{2.5}$ components $NO_3$, $SO_4$, $NH_4$, EC, OC, and metals Br, Al, Si, Ca, Fe, K, Mn, Pb, Cu, Se, Zn and Cr were used as fitting species. Primary source profiles used include gasoline vehicles (LDGV), diesel vehicles (HDDV), soil dust (SDUST), biomass burning (BURN), coal-fired power plants (CFPP) and cement production (CEM). Both BURN and LDGV have high fractions of OC in their source profiles (0.64 and 0.55 respectively). Profiles for components formed from atmospheric reactions are secondary ammonium sulfate (AMSULF), secondary ammonium bisulfate (AMBSLFT), secondary ammonium nitrate (AMNITR) and other OC (OTHROC). CMB reproduces 91% of the $PM_{2.5}$ mass ($R^2 = 0.90$, $n = 2698$, $\chi^2 = 3.39$), apportioning 15% of the $PM_{2.5}$ mass as 'other OC' which we take as the SOC fraction. It is recognized that there are potential non-secondary sources of OTHROC, including vegetative detritus, and unapportioned primary organic carbon in this source (e.g., Zheng et al., 2002) and therefore OTHROC may not include only SOC (Ding et al., 2008).

Uncertainties in CMB source contributions are given by the model and were calculated using a weighted variance approach:

$$\sigma_{g_{ik}} = \sum_{i=1}^{n}\left[\frac{f_{kj}^2}{\sigma_{C_{ij}}^2 + \sum_{k=1}^{N}\sigma_{f_{jk}}^2 g_k^2}\right]^{-1/2} \qquad (13)$$

where $f_{kj}$ is the source profile of species $j$ in source $k$, $\sigma_{f_{jk}}$ is the uncertainty in the profile, $g_k$ is the source contribution of source $k$, $\sigma_{g_{lk}}$ is the uncertainty in the contribution, and $\sigma_{c_{ij}}$ is the uncertainty in the measured concentration $c_{ij}$. The uncertainty in the POC fraction was estimated by propagating the uncertainties in the organic carbon fraction of the primary sources (SDUST, BURN, HDDV, LDGV, CFPP, CEM) and the uncertainty in the SOC fraction was estimated propagating the uncertainties in the POC and the measured OC (such as in Eq. (4)). The average uncertainties for the POC and SOC estimates, over the nine-year period, are calculated using the root mean square average (Eq. (5)).

### 2.4. Positive matrix factorization (PMF)

We used EPA-PMF 3.0 (Norris and Vedantham, 2008) for our simulations and classified species in the input model based on the

signal/noise ratio. Strong species for this study were $NO_3$, $SO_4$, $NH_4$, EC, OC, Br, Al, Si, Ca, Fe, and K. Weak species were Mn, Pb, Cu, Se, Zn and Cr. Since $PM_{2.5}$ was included and classified as a total variable, the model assigns it as a weak species in order to not double count its importance (Reff et al., 2007). We used 10 convergent runs and chose the run with the lowest error in the minimization of the mass balance equation. PMF reproduces 87% of the $PM_{2.5}$ ($R^2 = 0.91$, $n = 2931$). To identify the optimum number of factors, we ran PMF with five, six and seven factors and obtained the best fit with six factors (soil dust, biomass burning, secondary sulfate, secondary nitrate, cement and mobile sources). The SOC fraction in PMF is calculated by adding the OC fractions in the secondary factors and the unidentified OC fraction, defined as the difference between measured and fit OC (Lee et al., 2008b). The procedure of Polissar et al. (1998) was used in this study to calculate uncertainties in the species concentrations. Briefly, for data below DL, the concentrations were replaced with the value DL/2 and the uncertainty was set as (5/6)*DL. For missing data, concentrations were replaced by the geometric mean and the respective uncertainty was set at four times that of this mean concentrations. PMF provides uncertainties in factor profiles ($\sigma_{lkj}$), defined as the standard deviation of 100 bootstrapping runs. The uncertainty in factor contributions of species $\hat{j}(\sigma_{ij})$ is calculated as the product of the factor contribution ($g_{ik}$) times the uncertainty in the factor profiles (Eq. (14)).

$$\sigma_{ji}^2 = \sum_k \sigma_{f_{jk}}^2 g_{ki}^2 \quad (14)$$

Similar to CMB, POC uncertainty was propagated from the uncertainty in the OC fraction of primary factors (soil dust, biomass burning, cement and mobile sources). The uncertainty in the SOC estimate was propagated from the uncertainty in OC in the secondary factors (sulfate, nitrate) and the unidentified OC fraction. The average uncertainties for the POC and SOC estimates, over the nine-year period, are calculated using the root mean square average (Eq. (5)).

### 2.5. Air quality data

Aerosol chemical composition and gas phase data for this project were obtained for the Jefferson Street (JST) monitoring site, a mixed industrial-residential area near downtown Atlanta, GA (coordinates 33.7N, 84.4W and at an elevation of 275 m above sea level) during the period 1/2/1999–12/31/2007. Sampling at JST is part of a larger study called the Southeastern Aerosol Research and Characterization (SEARCH) network. Further information on this study and characteristics of the network are found elsewhere (Edgerton et al., 2005, 2006; Hansen et al., 2003). $PM_{2.5}$ monitoring includes daily 24-hour average measurements of ionic, carbonaceous and metal species concentrations. For the period, a total of 2937 days had valid data available. Data treatment of missing data and values below detection limits was performed as suggested by the network to ensure data quality (Hansen et al., 2003). A sample in which one or more major components were missing after the data treatment was discarded. Samples from the 4th of July, New Years (12/31) and adjacent days were removed from the analysis to avoid unusual noise in the concentrations due to fireworks (e.g. unusually high K concentrations). Measurements of WSOC in Atlanta were available for 120 days in the summer of 2007 (5/17–9/20). The WSOC fraction was measured semi-continuously using a PILS-TOC instrument at the roof of the Ford Environmental Science & Technology building at the Georgia Institute of Technology (GT). This site is approximately two miles away from the JST site. More information on the WSOC measurements can be found elsewhere (Hennigan et al., 2008; Sullivan and Weber, 2006). We found that OC measured with the continuous instruments at GT was higher than the OC measured at

JST (5.76 vs 3.97 ug-Cm⁻³). Explanation for this bias includes the loss of semi-volatile compounds from the filters (Edgerton et al., 2005; Turpin et al., 2000) and the positive artifact in the use of semi-continuous analyzers associated with the low air volume sampled and instrumental blanks (Offenberg et al., 2007; Peltier et al., 2007). To estimate the amount of WSOC at JST, we adjusted the WSOC at GT using the OC ratio between both sites.

$$WSOC_{JST} = WSOC_{GT} \left( \frac{OC_{JST}}{OC_{GT}} \right)_{avg} \quad (15)$$

For the summer of 2007, the ($OC_{JST}/OC_{GT}$) ratio was 0.69, giving an estimated averaged WSOC value of 2.29 ug-Cm⁻³ at JST (vs. 3.31 ug-Cm⁻³ at GT).

## 3. Results

During the nine-year period, the average OC concentration in Atlanta was 4.09±2.25 ug-Cm⁻³ (±one standard deviation), with a summer (April–September) mean of 3.90±1.80 ug-Cm⁻³ and a winter (October–March) mean of 4.25±2.63 ug-Cm⁻³. The higher OC value in winter in Atlanta is explained by an increase in mobile emissions and biomass burning activity (Lee et al., 2009; Zheng et al., 2002) accompanied by a decrease in the mixing layer.

### 3.1. EC tracer method

The EC tracer method estimates 1.51±1.36 ug-Cm⁻³ (±root mean square of the uncertainty as defined in Eq. (5)) of SOC in summer (39% of OC) and 0.77±1.96 ug-Cm⁻³ in winter (18% of OC). The lower amount of SOC in winter is consistent with the SOC formation mechanisms and fewer emissions of biogenic VOCs, which are responsible for a large portion of SOC in Atlanta (Weber et al., 2007). The greater SOC uncertainty in winter (>100% of the SOC) vs. summer (90% of the SOC) is explained by the higher uncertainties in the OC and EC species and the uncertainty in the primary (OC/EC) ratio during winter. The average of summer and winter estimates gives a SOC fraction of 1.19±1.71 ug-Cm⁻³ (29% of OC).

### 3.2. Regression method

The regression method estimates 1.70±0.80 ug-Cm⁻³ of SOC (44% of OC) in summer and 0.76±0.60 ug-Cm⁻³ of SOC (18% of OC)
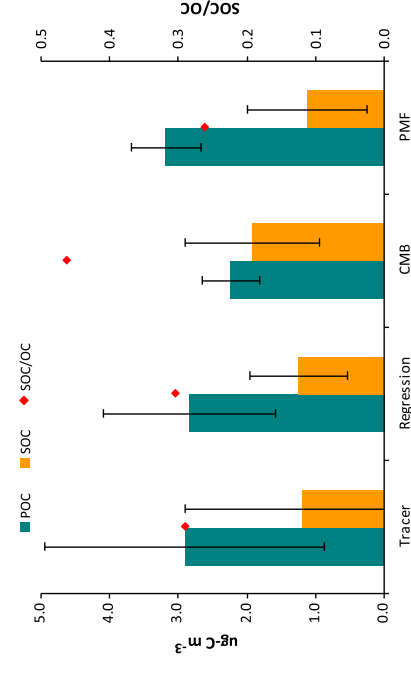


**Fig. 2.** Comparison of the four estimates from 1999-2007. The EC tracer ($n = 2932$) and regression ($n = 2932$) estimates include the use of summer/winter datasets with respective (OC/EC)$_p$ ratios and regression coefficients. For CMB ($n = 2698$) and PMF ($n = 2932$) the data was not separated by season. Error bars denote the root mean square of the uncertainty in POC and SOC fractions estimated by a propagation of errors.

**Fig. 4.** Day to day variability of SOC estimates for (a) June–July 2002 and (b) Dec. 2002–January 2003.

**Table 1**
Comparison of SOC Estimates using four methods.

| n (days) | EC Tracer | Regression | CMB | PMF |
|---|---|---|---|---|
| | 2931 | 2931 | 2698 | 2931 |
| POC (ug-C m$^{-3}$) | 2.90 (2.04)[a] | 2.84 (1.25)[b] | 2.24 (0.41)[c] | 3.18 (0.51)[d] |
| SOC (ug-C m$^{-3}$) | 1.19 (1.71) | 1.25 (0.71) | 1.92 (0.98) | 1.12 (0.87) |
| SOC/OC | 0.29 | 0.30 | 0.46 | 0.26 |
| σSOC/SOC | 1.44 | 0.57 | 0.51 | 0.78 |
| CV | 1.06 | 0.60 | 0.87 | 0.92 |
| Zero days of SOC | 478 | 0 | 114 | 0 |
| Summer SOC (ug-C m$^{-3}$) | 1.51 (1.36) | 1.70 (0.80) | 2.00 (0.93) | 1.37 (0.81) |
| Summer SOC/OC | 0.39 | 0.44 | 0.51 | 0.34 |
| σSOC/SOC | 0.90 | 0.47 | 0.46 | 0.60 |
| Winter SOC (ug-C m$^{-3}$) | 0.77 (1.96) | 0.76 (0.60) | 1.84 (1.03) | 0.86 (0.89) |
| Winter SOC/OC | 0.18 | 0.18 | 0.45 | 0.19 |
| σSOC/SOC | 2.56 | 0.80 | 0.56 | 1.03 |

[a] Uncertainties in EC tracer method calculated with Eqs. (3)–(5).
[b] Uncertainties in the regression method calculated with Eqs. (5), (11),(12).
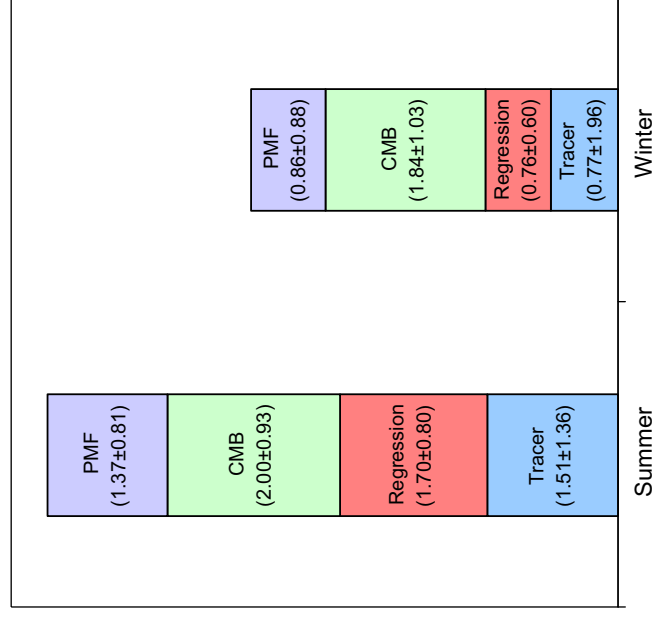[c] Uncertainties in CMB calculated with Eqs. (5), (13).
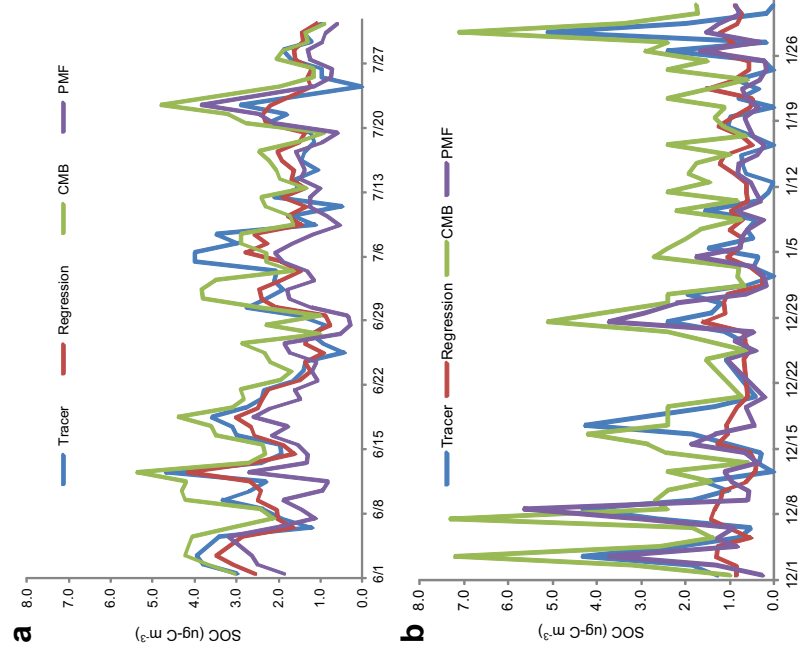[d] Uncertainties in PMF calculated with Eqs. (5), (14).

in winter. The SOC uncertainty is higher in summer given the larger concentrations and uncertainties in O$_3$ and SO$_4$ and the larger values of the regression coefficients. However, the amount of SOC is significantly lower in winter and the uncertainty represents 80% of the SOC value vs. 47% in the summer. The overall SOC uncertainty is driven by the estimate in winter, similar to the EC tracer method. The average of summer and winter estimates gives a SOC fraction of 1.25 ± 0.71 ug-C m$^{-3}$ (30% of OC).

### 3.3. Chemical mass balance and positive matrix factorization

We applied CMB and PMF with data from 1/2/1999 to 12/31/2007. The fit between measured and predicted OC was better in CMB ($R^2 = 0.99$, $n = 2698$) than PMF ($R^2 = 0.77$, $n = 2931$). The SOC

estimates are 1.92 ± 0.98 ug-C m$^{-3}$ (46% of OC) in CMB and 1.12 ± 0.87 ug-C m$^{-3}$ (26% of OC) in PMF. Summer SOC estimates are higher in both methods (2.00 ± 0.93 ug-C m$^{-3}$ in CMB and 1.37 ± 0.81 ug-C m$^{-3}$ in PMF) with lower uncertainties. In winter, the uncertainty in the SOC estimate is a significant fraction of the SOC concentration (56% in CMB & >100% in PMF). In CMB, it is known that uncertainties in source contributions are more influenced by uncertainties in the source profiles than ambient measurement data (Lee and Russell, 2007). Uncertainties in PMF are driven by the uncertainty in the measured OC species.



**Fig. 3.** Seasonal Estimates of SOC from 1999–2007. Units are ug-C m$^{-3}$, for concentrations and uncertainties (defined as the root mean square average).

**Table 2**
Comparison of SOC Estimates with related work, SOC (ug-C m$^{-3}$) or (%).

| | Year-round ug-C m$^{-3}$ (%) | Summer time ug-C m$^{-3}$ (%) |
|---|---|---|
| This study, EC tracer | 1.19 (30%) | 1.52 (40%) |
| This study, regression | 1.25 (33%) | 1.70 (44%) |
| This study, CMB | 1.92 (46%) | 2.00 (51%) |
| This study, PMF | 1.12 (26%) | 1.37 (34%) |
| EC tracer[a] | 34% | 34% |
| CO tracer[a] | 45% | 57% |
| Multiple regression[a] | 27% | 35% |
| Regular CMB[b] | 1.59 (39%) | – |
| CMB-LGO[c] | 2.59 (58%) | – |
| CMB-MM[d] | – | 2.43 (57%) |
| CMB-MM[e] | – | 3.18 (75%) |
| PMF[b] | 0.77 (19%) | – |
| Time resolved[f] | – | 3.9 ± 2.2 (46%) |

[a] EC tracer, CO tracer and Multiple regression from (Blanchard et al., 2008).
[b] Regular CMB and PMF from (Lee et al., 2008a).
[c] CMB-LGO from (Marmur et al., 2005).
[d] CMB-MM in 1999 from (Zheng et al., 2002).
[e] CMB-MM in summer 2001 and winter 2002 from (Zheng et al., 2007).
[f] Time resolved from (Lim and Turpin, 2002).

**Table 3**
Comparison of SOC Estimates to SOC from CMB-MM and WSOC.

| | a. CMB-MM | | | b. WSOC | | |
|---|---|---|---|---|---|---|
| | $R^2$ | Bias[a] | Error[a] | $R^2$ | Bias | Error |
| Regression | 0.87 | −1.05 | 1.86 | 0.50 | −0.48 | 0.93 |
| EC tracer | 0.58 | −1.45 | 2.20 | 0.41 | −0.49 | 1.10 |
| CMB | 0.75 | −1.53 | 2.42 | 0.48 | −0.10 | 0.98 |
| PMF | 0.80 | −1.30 | 1.90 | 0.45 | −0.68 | 1.14 |

a. CMB-MM from Zheng et al., 2007, b. WSOC from Hennigan et al. 2008.
[a] Expressed in ug-Cm$^{-3}$, Bias expressed as $1/N \ \Sigma(SOC_i - WSOC)$ and Error expressed as $1/N \ \Sigma(SOC_i - WSOC)^2$, where $i$ denotes the method and $N$ the number of samples.

### 3.4. Comparison of SOC estimates and uncertainties

The four methods estimate SOC fractions between 1.12 ± 0.87 and 1.92 ± 0.98 ug-Cm$^{-3}$ representing 26–46% of the OC respectively (Fig. 2). CMB led to the highest estimate of SOC while the PMF led to the lowest. The EC tracer and the regression methods provided intermediate estimates of SOC. The higher SOC estimate in CMB is explained by the inclusion of all unapportioned OC into one secondary source. The other-OC source in CMB is correlated with both biomass burning ($R^2 = 0.57$) and mobile ($R^2 = 0.55$) factors in PMF. This correlation can be explained in part by: i) the other-OC includes primary OC from unidentified sources (such as meat cooking and natural gas combustion) that may correlate with biomass burning and mobile factors in PMF, ii) SOC may be included in the biomass burning factor in PMF since carbon emitted during biomass burning is in some cases oxygenated and water soluble (Lee et al., 2008a), or in the mobile factor since OC emissions from traffic can potentially evolve into SOC (Robinson et al., 2007).

The low estimate of SOC by PMF has been found in previous studies in the southeastern US (Lee et al., 2008b). Without use of detailed oxygenated species, PMF is not able to provide further information on SOC because of the colinearity of OC sources.

### 3.4.1. Uncertainties

The lowest uncertainty in the SOC estimate is found in the regression method and the highest is the EC tracer method (Table 1). The CMB uncertainties are comparable to the regression method, and if expressed as a fraction of the SOC concentrations they are even lower. The PMF uncertainties are significantly higher than the uncertainties in the CMB method.

### 3.4.2. Seasonal estimates

In summer, the proportion of SOC estimated by the four methods is similar, with CMB having the highest and PMF the lowest fractions (Fig. 3). In winter, CMB estimates are much higher than the other methods, indicating the likely inclusion of primary OC in this fraction and therefore, an overestimate of the SOC fraction.

### 3.4.3. Day-to-day variability

During the summer 2002, the four estimates exhibit similar day to day variability (Fig. 4). In winter 2002/2003, regression is the only method that yields smooth pattern, which would be expected for a secondary pollutant. The other estimates have significant variability typically more associated with primary pollutants. The lowest coefficient of variance, associated with this temporal trend, was for the regression method (Table 1). The EC tracer and the CMB methods had 478 and 114 days of zero estimated SOC, respectively, occurring when estimated POC is greater than measured OC.
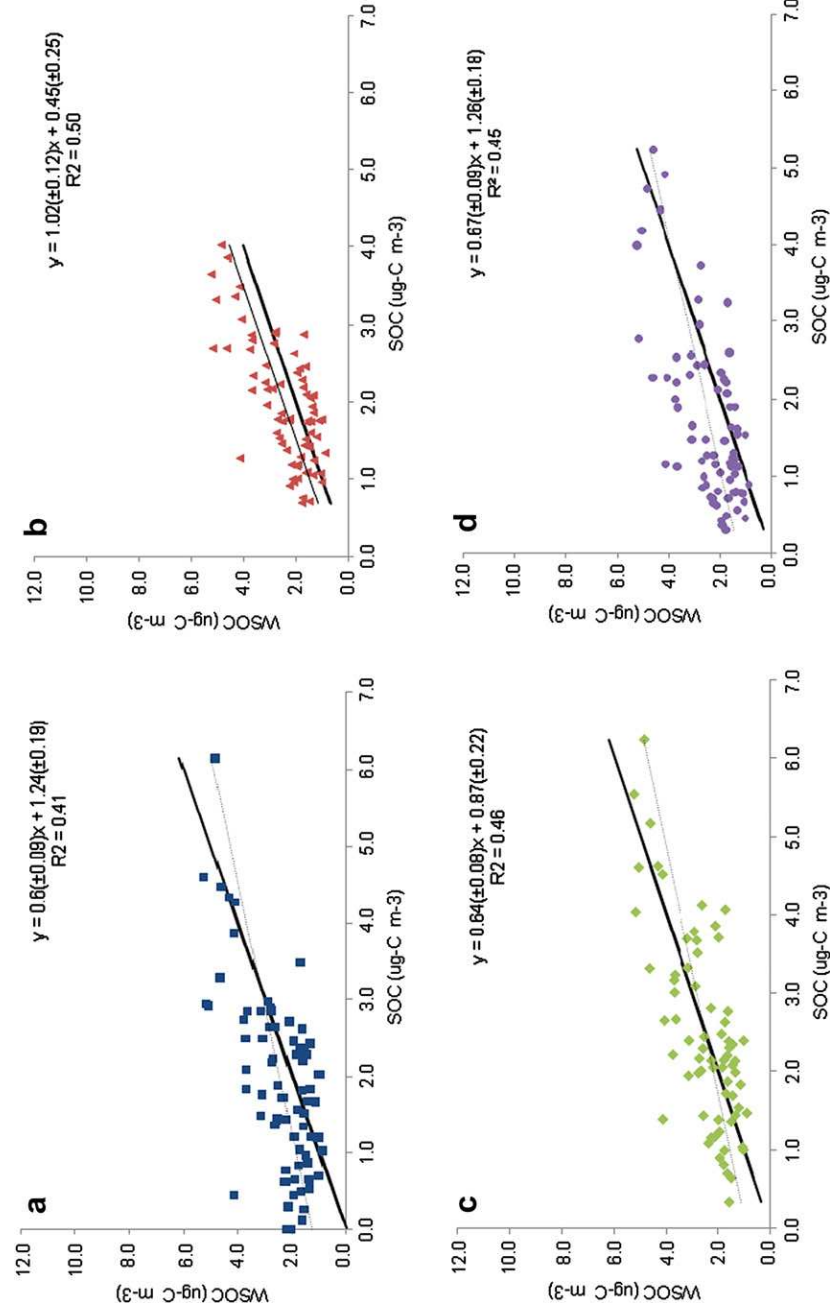


**Fig. 5.** Comparison of WSOC measurements with SOC estimates in 2007 by the (a) EC-tracer, (b) regression method, (c) CMB and (d) PMF. Solid line is the 1:1.

### 3.5. Comparison with related work

The range of SOC estimates in this study was 26–47% which is comparable with findings of other studies at Jefferson St in Atlanta (Table 2). The lowest SOC estimate (19% of OC) was obtained using PMF (Lee et al., 2008b) and the highest (58% of OC) using CMB-LGO (Marmur et al., 2005). For summer, our estimates vary from 34 to 51% as compared to results of other studies in Atlanta ranging from estimated SOC of 34% using the EC tracer method (Blanchard et al., 2008) to 75% using CMB-MM (Zheng et al., 2007). Since the time periods differ between studies, different SOC estimates are expected. Some studies (De Gouw and Jimenez, 2009; Robinson et al., 2007) suggest an underestimation of SOA in urban centers due to the rapid formation of SOA from semi-volatile and intermediate-volatile organic compounds emitted by traffic. Docherty et al. (2008) found ratios of SOA/OA between 70–90% on aged aerosols downwind of Los Angeles in a summer period with an ozone concentration of 86ppb. Our SOC estimate is equivalent to 35–57% being SOA using ratios of SOC/SOA = 1.8 and POC/POA = 1.2 (similar to Docherty et al., 2008) and for Atlanta the average 8 h-maximum $O_3$ concentration was 60 ppb, lower than the observed in the L.A. basin. While estimates of SOA formation using aerosol mass spectrometry have also being conducted (Jimenez et al., 2009), such data were unavailable in Atlanta for comparison here.

We compare our estimates with results from CMB using molecular markers during summer of 2001 (Zheng et al., 2007). Data were not available to conduct a long-term analysis of SOC estimated by CMB-MM. Here SOC is estimated the same way using regular CMB, as the difference between measured OC and the identified primary fraction, but using a greater number of fitting species from $PM_{2.5}$ organic speciation. The correlation was strongest with estimates from the regression method (Table 3a).

### 3.6. Comparison with WSOC measurements

In an effort to compare our estimates with new methods to quantify organic aerosols, we compared the four estimates with the WSOC fraction in Atlanta during the summer of 2007, when biomass burning contribution was negligible (Zhang et al., 2010) and therefore, we expect WSOC to be a good surrogate of SOC. The ratio of WSOC/OC observed was 0.52, slightly higher than our summer SOC/OC estimates (0.34–0.51). The strongest correlation and the lowest error were between WSOC and estimates from the regression method (Table 3b). The regression estimate had a slope close to 1.0 when plotted against WSOC (Fig. 5) indicating a good estimation of this secondary fraction.

### 4. Conclusion

Comparison of four methods to estimate the SOC fraction in the $PM_{2.5}$ suggests that between 26 and 47% of the OC in Atlanta is secondary in origin on a year-around basis. Uncertainties in the estimated SOC fraction range from 51% to more than 100% and are largely influenced by estimation of SOC in winter time. The SOC fraction estimated by the regression method has the lowest uncertainty, a greater value in summer than winter, shows less day-to-day variability and has a more similar trend to the WSOC measurements as compared to the other methods, suggesting the regression method is the most accurate method for developing multi-year SOC estimates, necessary in epidemiologic analysis and evaluation of air quality policy effectiveness. The regression method only requires readily measured speciated $PM_{2.5}$ components (i.e., EC, OC, K, sulfate and nitrate), ozone and CO data.

### Acknowledgements

### References

Bevington, P., Robinson, K., 2003. Data Reduction and Error Analysis for the Physical Sciences, third ed. McGraw Hill, New York.

Blanchard, C.L., Hidy, G.M., Tanenbaum, S., Edgerton, E., Hartsell, B., Jansen, J., 2008. Carbon in Southeastern US aerosol particles: empirical estimates of secondary organic aerosol formation. Atmospheric Environment 42, 6710–6720.

Cabada, J.C., Pandis, S.N., Subramanian, R., Robinson, A.L., Polidori, A., Turpin, B., 2004. Estimating the secondary organic aerosol contribution to PM2.5 using the EC tracer method. Aerosol Science and Technology 38, 140–155.

Chow, J.C., Watson, J.G., Pritchett, L.C., Pierson, W.R., Frazier, C.A., Purcell, R.G., 1993. The DRI thermal optical reflectance carbon analysis system – description, evaluation and applications in United-States air quality studies. Atmospheric Environment Part A General Topics 27, 1185–1201.

De Gouw, J., Jimenez, J.L., 2009. Organic aerosols in the Earth's atmosphere. Environmental Science & Technology 43, 7614–7618.

Ding, X., Zheng, M., Edgerton, E.S., Jansen, J.J., Wang, X.M., 2008. Contemporary or fossil origin: split of estimated secondary organic carbon in the Southeastern United States. Environmental Science & Technology 42, 9122–9128.

Docherty, K.S., Stone, E.A., Ulbrich, I.M., DeCarlo, P.F., Snyder, D.C., Schauer, J.J., Peltier, R.E., Weber, R.J., Murphy, S.M., Seinfeld, J.H., Grover, B.D., Eatough, D.J., Jimenez, J.L., 2008. Apportionment of primary and secondary organic aerosols in Southern California during the 2005 study of organic aerosols in Riverside (SOAR-1). Environmental Science & Technology 42, 7655–7662.

Eder, B., Yu, S., 2006. A performance evaluation of the 2004 release of Models-3 CMAQ. Atmospheric Environment 40, 4811–4824.

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.

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.

Hennigan, C.J., Bergin, M.H., Dibb, J.E., Weber, R.J., 2008. Enhanced secondary organic aerosol formation due to water uptake by fine particles. Geophysical Research Letters 35, L18801.

Jimenez, J.L., Canagaratna, M.R., Donahue, N.M., Prevot, A.S.H., Zhang, Q., Kroll, J.H., DeCarlo, P.F., Allan, J.D., Coe, H., Ng, N.L., Aiken, A.C., Docherty, K.S., Ulbrich, I.M., Grieshop, A.P., Robinson, A.L., Duplissy, J., Smith, J.D., Wilson, K.R., Lanz, V.A., Hueglin, C., Sun, Y.L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J.M., Collins, D.R., Cubison, M.J.E., Dunlea, J., Huffman, J.A., Onasch, T.B., Alfarra, M.R., Williams, P.I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J.Y., Zhang, Y.M., Dzepina, K., Kimmel, J.R., Sueper, D., Jayne, J.T., Herndon, S.C., Trimborn, A.M., Williams, L.R., Wood, E.C., Middlebrook, A.M., Kolb, C.E., Baltensperger, U., Worsnop, D.R., 2009. Evolution of organic aerosols in the atmosphere. Science 326, 1525–1529.

Kim, E., Hopke, P.K., Edgerton, E.S., 2003. Source identification of Atlanta aerosol by positive matrix factorization. Journal of the Air & Waste Management Association 53, 731–739.

Kim, E., Hopke, P.K., Edgerton, E.S., 2004. Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in positive matrix factorization. Atmospheric Environment 38, 3349–3362.

Kroll, J.H., Seinfeld, J.H., 2008. Chemistry of secondary organic aerosol: formation and evolution of low-volatility organics in the atmosphere. Atmospheric Environment 42, 3593–3624.

Lee, D., Balachandran, S., Pachon, J., Shankaran, R., Lee, S., Mulholland, J.A., Russell, A.G., 2009. Ensemble-trained PM2.5 source apportionment approach for health studies. Environmental Science & Technology 43, 7023–7031.

Lee, S., Kim, H.K., Yan, B., Cobb, C.E., Hennigan, C., Nichols, S., Chamber, M., Edgerton, E.S., Jansen, J.J., Hu, Y., Zheng, M., Weber, R.J., Russell, A.G., 2008a. Diagnosis of aged prescribed burning plumes impacting an urban area. Environmental Science & Technology 42, 1438–1444.

Lee, S., Liu, W., Wang, Y.H., Russell, A.G., Edgerton, E.S., 2008b. Source apportionment of PM2.5: comparing PMF and CMB results for four ambient monitoring

sites in the Southeastern United States. Atmospheric Environment 42, 4126–4137.

Lee, S., Russell, A.G., 2007. Estimating uncertainties and uncertainty contributors of CMB PM2.5 source apportionment results. Atmospheric Environment 41, 9616–9624.

Lee, S., Russell, A.G., Baumann, K., 2007. Source apportionment of fine particulate matter in the Southeastern United States. Journal of the Air & Waste Management Association 57, 1123–1135.

Lim, H.-J., Turpin, B.J., 2002. Origins of primary and secondary organic aerosol in Atlanta: results of time-resolved measurements during the Atlanta supersite experiment. Environmental Science & Technology 36, 4489–4496.

Marmur, A., Unal, A., Mulholland, J.A., Russell, A.G., 2005. Optimization-based source apportionment of PM2.5 incorporating gas-to-particle ratios. Environmental Science & Technology 39, 3245–3254.

Norris, G., Vedantham, R., 2008. EPA Positive Matrix Factorization (PMF) 3.0. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Offenberg, J.H., Lewandowski, M., Edney, E.O., Kleindienst, T.E., 2007. Investigation of a systematic offset in the measurement of organic carbon with a semi-continuous analyzer. Journal of the Air & Waste Management Association 57, 596–599.

Paatero, P., Tapper, U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 111–126.

Peltier, R.E., Weber, R.J., Sullivan, A.P., 2007. Investigating a liquid-based method for online organic carbon detection in atmospheric particles. Aerosol Science and Technology 41, 1117–1127.

Polissar, A.V., Hopke, P.K., Paatero, P., 1998. Atmospheric aerosol over Alaska - 2. Elemental composition and sources. Journal of Geophysical Research-Atmospheres 103, 19045–19057.

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.

Robinson, A.L., Donahue, N.M., Shrivastava, M.K., Weitkamp, E.A., Sage, A.M., Grieshop, A.P., Lane, T.E., Pierce, J.R., Pandis, S.N., 2007. Rethinking organic aerosols: semivolatile emissions and photochemical aging. Science 315, 1259–1262.

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.

Snyder, D.C., Rutter, A.P., Collins, R., Worley, C., Schauer, J.J., 2009. Insights into the origin of water soluble organic carbon in atmospheric fine particulate matter. Aerosol Science and Technology 43, 1099–1107.

Solomon, P.A., Mitchell, W., Tolocka, M., Norris, G., Gemmill, D., Wiener, R., Vanderpool, R., Murdoch, R., Natarajan, S., Hardison, E., 2000. Evaluation of PM2.5 Chemical Speciation Samplers for Use in the EPA National PM2.5 Chemical Speciation Network. US Environmental Protection Agency, Research Triangle Park, NC.

Sullivan, A.P., Peltier, R.E., Brock, C.A., de Gouw, J.A., Holloway, J.S., Warneke, C., Wollny, A.G., Weber, R.J., 2006. Airborne measurements of carbonaceous aerosol soluble in water over northeastern United States: method development and an investigation into water-soluble organic carbon sources. Journal of Geophysical Research-Atmospheres 111, D23.

Sullivan, A.P., Weber, R.J., 2006. Chemical characterization of the ambient organic aerosol soluble in water: 1. Isolation of hydrophobic and hydrophilic fractions with a XAD-8 resin. Journal of Geophysical Research 111, D05314.

Tesche, T.W., Morris, R., Tonnesen, G., McNally, D., Boylan, J., Brewer, P., 2006. CMAQ/CAMx annual 2002 performance evaluation over the eastern US. Atmospheric Environment 40, 4906–4919.

Turpin, B.J., Huntzicker, J.J., 1991. Secondary formation of organic aerosols in the Los Angeles basin – a descriptive analysis of organic and elemental carbon concentrations. Atmospheric Environment Part A General Topics 25, 207–215.

Turpin, B.J., Saxena, P., Andrews, E., 2000. Measuring and simulating particulate organics in the atmosphere: problems and prospects. Atmospheric Environment 34, 2983–3013.

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.

Weber, R.J., Sullivan, A.P., Peltier, R.E., Russell, A., Yan, B., Zheng, M., de Gouw, J., Warneke, C., Brock, C., Holloway, J.S., Atlas, E.L., Edgerton, E., 2007. A study of secondary organic aerosol formation in the anthropogenic-influenced southeastern United States. Journal of Geophysical Research-Atmospheres 112, D13302.

Yu, S., Bhave, P.V., Dennis, R.L., Mathur, R., 2007. Seasonal and regional variations of primary and secondary organic aerosols over the continental United States: semi-empirical estimates and model evaluation. Environmental Science & Technology 41, 4690–4697.

Zhang, X., Hecobian, A., Zheng, M., Frank, N.H., Weber, R.J., 2010. Biomass burning impact on PM2.5 over the Southeastern US during 2007: integrating chemically speciated FRM filter measurements, MODIS fire counts and PMF analysis. Atmospheric Chemistry and Physics 10, 6839–6853.

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.

Zheng, M., Cass, G.R., Schauer, J.J., Edgerton, E.S., 2002. Source apportionment of PM2.5 in the Southeastern United States using solvent-extractable organic compounds as tracers. Environmental Science & Technology 36, 2361–2371.

Zheng, M., Ke, L., Edgerton, E.S., Schauer, J.J., Dong, M., Russell, A.G., 2006. Spatial distribution of carbonaceous aerosol in the southeastern United States using molecular markers and carbon isotope data. Journal of Geophysical Research 111, D10S06.