An ensemble-trained chemical mass balance (CMB) approach is developed for particulate matter (PM) source apportionment (SA) approaches for use in health studies. The approach uses results from a short-term emission-based chemical transport model (CTM) and multiple receptor-based approaches. Ensemble results have less day-to-day variation in source impacts and fewer biases between observed and estimated PM$_{2.5}$ mass compared to the original receptor model results. Ensemble results show increases in road dust, biomass burning, and coal impacts, but secondary organic carbon (SOC) impacts decrease. These results, along with observations, are then used to obtain new source profiles. Two sets of new source profiles based on ensemble results in summer (July 2001) and winter (January 2002) were developed, and used in separate CMB applications for a 12-month data set of daily PM$_{2.5}$ measurements at the Atlanta, GA, Jefferson Street site. Results show that ensemble-trained CMB approaches, using both summer profiles and winter profiles, effectively reduce day-to-day variability of source impact estimates by reducing fewer days of zero impact from sources known to be present, as compared to traditional receptor modeling, suggesting improved results.

1. Introduction

There are a number of reasons to develop direct relationships between emission sources, air pollutant concentrations, and health endpoints. In addition to establishing associations between sources and atmospheric pollutant concentrations (1–6), uncertainties in epidemiologic analyses might be reduced using particular matter (PM) source impacts rather than PM components (2, 5), and modeling approaches that link physically or chemically transformed products to sources may decrease classification error in epidemiologic analysis. However, there are issues associated with use of current source apportionment (SA) approaches in health analyses. As explored in Marmur et al. (2), the more commonly used receptor models, as well as those involving physically based models, have limitations that affect their usefulness. For example, commonly applied receptor models do not include all sources. Baek et al. (6) found that a typical receptor model application includes identifies sources/factors representing only 60–80% of inventoried primary emissions. Incomplete specification of the sources leads to both errors and bias. Receptor model SA approaches appear to introduce excessive day-to-day variability, as suggested by major sources with zero impact on some days and significant variation in relative impacts from one day to the next. A further limitation is that the information is representative of the location of the measurements. There are further issues specific to the method employed as well. For example, the chemical mass balance (CMB) (7) approaches depend on accurate characterization of source emissions and typically do not account for species transformations. Factor analytic approaches, such as Positive Matrix Factorization (PMF) (8, 9), on the other hand, result in factors which can include contributions from multiple sources and can be difficult to interpret. In addition, calculated impacts can change from subjective choices in the model application. Receptor modeling approaches are also limited in their ability to deal with species transformations, and hence in their ability to tie PM health impacts to sources of secondary pollutants.

These limitations of receptor modeling approaches suggested using detailed, physically based, chemical transport modeling (CTM), such as the Community Multiscale Air Quality (CMAQ) model (10). Marmur et al. applied CMAQ to the southeastern United States (2) with the intent of employing directly calculated source impacts on air quality in time-series health studies. Such an application is computationally demanding. Another identified limitation was that the fraction of total PM contributed by each source varied very little day by day. The lack of variation in source fractions was attributed to the fact that (1) the meteorological model used did not fully capture the fine scales (time and space) variations (11), (2) emission inputs had little day-to-day variation, except for weekends and changes due to meteorology, and (3) emissions are allocated based on land-use, not capturing actual spatial–temporal variations. On the other hand, the analysis suggested that the CMAQ results might be appropriate for use for longer term exposures (e.g., to study chronic health effects), and have the advantage of large continuous spatial coverage. A critical issue for both receptor-based and emission-based SA approaches is how to formally evaluate the results. Emission-based approaches can be indirectly evaluated by assessing how well the model correctly simulates the temporal and spatial concentrations of the observed air pollutants. Good agreement between model predictions and measurements, however, does not guarantee that the predicted source impacts are correct as there may be compensatory errors. Evaluating receptor model results is even more problematic. Models can be tested on artificial data sets, but that raises the question of how realistic such sets are in terms of capturing all of the complexities that impact PM composition at a receptor (e.g., source profile variation, source strength variation, physical and chemical transformations, and unidentified source contributions) (12). Alternatively, models can be tested on “event” days in which impacts from individual sources are known to dominate. Such events, however, are unusual, and can cause problems when applying factor analytic methods. Comparison of results from various modeling approaches can provide a false sense of comfort because the models may share the same assumptions or mathematical limitations. The more apparent
2. Approach

This work relies on blending receptor-based modeling and detailed source apportionment results to develop an ensemble model of source contributions. The modeling approach is hybridized with conventional receptor modeling to provide a composite approach to source apportionment.

The Fractional Contributions of Source Categories

where \( C_i \) is the observed concentration of species \( i \), \( f_j \) is the fraction of species \( i \) contributed by source \( j \), and \( e_i \) is the error term. This equation is weighted by the composition of each source to account for the contribution of each source category to the receptor.

1. Chemical Mass Balance (CMB) Approaches Including CMB-MM, -LGO, -RG
2. PCA, PMF, CMB-TR
3. Source Apportionment Results

The Regional Planning Organization (RPO) consensus approach was based on weighted least-squares solutions solved to an equation that describes a mass balance between ambient concentrations of each chemical species and a linear sum of products of source profiles and source contributions:

where \( C_{ij} \) is the concentration of species \( j \) attributed to source \( i \), \( S_{ij} \) is the contribution of source \( j \) to the total concentration of species \( i \), and \( e_{ij} \) is the error term.

The model uses dayspecific emissions from major point sources, including to the Atlanta Jefferson Street (JST) SEARCH and BRAVO (15, 16), and requires emissions (e.g., SMOKE (17, 18)). Here, results from CMAQ-TR and model uncertainties, can be used together. This approach is constrained CMB to allow a constrained model-based approach (19).

In this study, we apportion PM sources to available factors/sources. Further, source impacts are classified as they are strongly affected by meteorology being collinearity of source factors can increase classification error in epidemiologic modeling.

Chemical composition and engine operating characteristics (including gasoline composition and engine operating characteristics) for the periods studied here. Details of PMF, and CMAQ-TR results from prior work (e.g., CMB-TR (20), PMF (21, 22), and CMAQ-TR (23)) are used to the CMB-LGO. CMB-LGO was developed as part of the 2002 VISTAS modeling, which included day-specific emissions from major point sources. The Regional Planning Organization (RPO) consensus approach was based on weighted least-squares solutions solved to an equation that describes a mass balance between ambient concentrations of each chemical species and a linear sum of products of source profiles and source contributions:

\[
C_i = \sum_{j=1}^{m} f_j C_{ij} + e_i
\]

where \( C_i \) is the observed concentration of species \( i \), \( f_j \) is the fraction of species \( i \) contributed by source \( j \), and \( e_i \) is the error term. This equation is weighted by the composition of each source to account for the contribution of each source category to the receptor.

The model uses dayspecific emissions from major point sources, including to the Atlanta Jefferson Street (JST) SEARCH and BRAVO (15, 16), and requires emissions (e.g., SMOKE (17, 18)). Here, results from CMAQ-TR and model uncertainties, can be used together. This approach is constrained CMB to allow a constrained model-based approach (19).

In this study, we apportion PM sources to available factors/sources. Further, source impacts are classified as they are strongly affected by meteorology being collinearity of source factors can increase classification error in epidemiologic modeling.
other SA approaches; the three nonprimary sources, ammonium sulfate, ammonium bisulfate, and ammonium nitrate, used in CMB-RG, CMB-LGO, and PMF, are divided into SULFATE, NITRATE, and AMMONIUM sources based on molecular weights. Out of 33 primary sources (Table S1) used in CMAQ, 17 were recategorized to five primary sources that were also used in CMB and PMF applications: (1) BURN impacts include agricultural burning, cigarette, fireplaces, prescribed burning, yard waste burning, wood combustion, and wild fires; (2) DUST includes construction, and paved and unpaved road dust; (3) GV includes gasoline-fuel engine sources such as gasoline vehicles, nonroad aircraft, and nonroad gasoline engines; (4) DV includes diesel vehicles and nonroad diesel engines; and (5) COAL includes just the primary PM from coal burning. However, the dominant precursor of sulfate PM in the eastern U.S. is the emission of SO\textsubscript{2} from the combustion of coal.

**Ensemble-Derived Source Impacts.** Daily PM source impact results using the different SA approaches at the JST site in Atlanta were ensembled using weighted averaging:

\[
\bar{S}_j(t) = \frac{\sum_{i=1}^n w_{ij}(t) \cdot S_{ij}(t)}{\sum_{i=1}^n w_{ij}(t)}
\]

where \(\bar{S}_j(t)\) is the ensemble-calculated impact of source \(j\) (in \(\mu g/m^3\)) at time \(t\), \(S_{ij}(t)\) is the impact developed by method \(l\) (e.g., CMAQ, CMB-RG, PMF, etc.), and \(w_{ij}(t)\) is the weight assigned to the calculated source impact as described below. Ensembling was applied for two different periods: July 2001 for summer and January 2002 for winter. There were 2–5 different \(S_{ij}(t)\) available for each day during July 2001 and January 2002. Uncertainties in each \(S_{ij}(t)\) were assessed and the inverse values were used for the weights.

\[
w_{ij} = \frac{1}{\sigma_{S_{ij}}}
\]

We explored using both \(1/(\sigma_{S_{ij}})\) and \(1/(\sigma_{S_{ij}}^2)\) and chose \(1/(\sigma_{S_{ij}})\) because we found a lower average reduced chi-square value (hereafter referred to as chi-square value) and fewer zero-impact days using \(1/(\sigma_{S_{ij}})\), particularly for DUST. In future work, we will explore the impact of using different weights and ways of estimating uncertainties to optimize performance based on average chi-square, zero-impact days, and day-to-day variability.

In CMB, uncertainties in the source contribution are estimated using an effective variance approach and an output of the model (17):

\[
\sigma_j = \left( \sum_{i=1}^n \frac{\sigma_{ij}^2}{\sigma_{ij}^2 + \sum_{j=1}^n \sigma_{ij}^2} \right)^{-1/2}
\]

where \(\sigma_j\) (in \(\mu g/m^3\)) is the uncertainty in source contribution, \(S_j\) (in \(\mu g/m^3\)), \(\sigma_{ij}\) is the uncertainty in the measurement of ambient species \(i\), and \(\sigma_{ij}\) is the uncertainty in the fraction of \(i\) species, in the \(j\) source profile. Uncertainties calculated using eq 4, available from prior studies, were used for CMB-RG (39) and CMB-MM (19), or calculated here for CMB-LGO and PMF. PMF factor profile uncertainties were developed using bootstrapping (28, 33). For CMAQ, uncertainties were adjusted based on the relations between CMAQ and CMB-RG models (derivation is in Supporting Information):

\[
4\sigma_{S_{ij}} = 4\sigma_{S_{ij,CMB}} + (S_{j,CMAQ} - S_{j,CMB})^2
\]

where \(S_{j,CMB}\) and \(S_{j,CMAQ}\) (in \(\mu g/m^3\)) are source contributions of source \(j\) in CMB and CMAQ, respectively. This approach directly accounts for differences observed and simulated concentrations.

**Ensemble-Trained CMB Source Apportionment.** While we view both CMB and FA approaches as attractive for application to long-term PM source apportionment, here we concentrate on using a CMB-based approach, similar to CMB-LGO. One reason to choose CMB-LGO is that it allows for constrained optimization of the source profiles to attain a best fit to ensemble results and can readily utilize gas-phase pollutant concentration data. In this case, CMB-LGO was applied to develop constrained ensemble-based source profiles (EBSPs) that best replicate the ensemble-derived \(S_{ij}\).

The general steps in the training are as follows: (1) use a best-fit to the ensemble SA series and develop optimized source profiles (i.e., EBSPs), (2) calculate the new source impacts and assess variability. The EBSPs should decrease the day-to-day variability found in the initial CMB applications, and account for location specific differences in relative source profile composition.

First, EBSPs which best replicate the ensemble results are calculated using the species mass balance equation, taking the observed concentrations, \(C_{ik}\) and ensemble source impacts, \(S_{ik}\), as known, leaving the source profiles, \(f_{ik}\), as unknown:

\[
C_{ik} = \sum_{j=1}^m \left( \frac{\bar{S}_{ij} \cdot \sigma_{S_{ij}}^2}{\sigma_{C_{ik}}^2} \right) f_{ik} + e_{ik}
\]

Equation 6 was solved for two periods using observed data and ensemble results at JST in July 2001 for summer and January 2002 for winter, by minimizing \(\chi^2\):

\[
\chi^2 = \sum_{i=1}^n \left( \frac{C_{ik} - \sum_{j=1}^m \bar{S}_{ij} \cdot f_{ik} \cdot \sigma_{C_{ik}}^2}{\sigma_{C_{ik}}^2} \right)^2
\]

where \(\bar{S}_{ij}\) are the ensemble-based source profiles. For all sources, the sum of relative source profile compositions, \(\Sigma S_{ik} = f_{ik}\) is limited to be less than 1 to consider the oxidized forms of metals and the relation between OC and OM (OC/OM > 1.4). Lower and upper bounds for the fractions of species, which were developed by Marmur et al. (22), are applied for the profile optimization, except the OC/EC ratio bounds for GV and DV sources, which are adopted from Cadle et al. (36), and the upper bounds for SO\textsubscript{2} and OC species in COAL source, which are adopted from the BRAVO study (Table S2) (24). Source impacts for each day in 2002, using the EBSPs in CMB-LGO, \(S_{ik}\), were then recalculated taking the \(C_{ik}\) and the \(f_{ik}\) as known. Gas phase pollutant concentrations of CO, SO\textsubscript{2}, and NO\textsubscript{x} provide constraints to the solutions as discussed in Marmur et al. (35).

There is no reason that the ensemble method cannot be easily employed at other sites, and it is not critical to have results from any one specific method. If measurement data are available, it is straightforward to apply CMB-RG, CMB-LGO, and PMF. Since the CMAQ domain includes all of the contiguous U.S., most receptor sites will be covered by a single CMAQ run as long as model runs are available for the corresponding time of the ensemble (e.g., Jan 2002 and July 2001, as done here, and many groups have applied CMAQ to annual and even multiyear periods). However, CMB-MM is the most resource-intensive method and therefore is not expected to be as widely utilized.

**Results**

**Ensemble.** Daily PM source apportionment results, developed by five different SA methods and the ensemble-
derived SA, show that the ensemble technique reduces the day-to-day variability of source impacts to varying degrees and decreases the number of zero-impact days for sources such as diesel exhaust and coal combustion (Figure 1). Daily variability is reduced when compared to CMB-LGO results for GV, COAL, and SOC sources, while little change is found for other sources. There were no zero-impact days with the ensemble averaging, while the original CMB-LGO results showed three zero-impact days in July 2001 (two days for COAL, one day for DV) and 11 zero-impact days in January 2002 (nine days for COAL, one day each for DV and SOC).

The number of source impacts available for each day varied for each source since receptor-based SA results were not conducted for those days missing the necessary ambient species measurements. However, in the original CMB-LGO, missing values were replaced by the geometric mean of the measured values, and their corresponding uncertainties were set equal to four times the geometric mean (35). All sources had all five source impacts on most days (Table S3) in summer and winter except for DUST in winter, which had CMB-MM results for only five days. It should be noted that for both COAL and BURN in the summer and for COAL in the winter, only four SA results were available since CMB-MM did not resolve those sources in these cases.

Source impacts of several components increased compared to the original CMB-LGO results. The January fractional source impacts for DUST, BURN, and COAL sources increased to 3.3%, 20.1%, and 1.5% from 1.4%, 12.0%, and 0.5%, respectively. SOC decreased from 19.6% in the original CMB-LGO to 8.1% in the ensemble approach during January 2002 (Table 1). Similarly, the July fractional source impacts for...
DUST and COAL increased, whereas SOC decreased in ensemble results. Increases in primary sources were driven by CMAQ and PMF, which show greater impacts from primary sources than the other SA approaches. The decrease in SOC impacts in the ensemble approach is due to the negative bias in PMF-simulated SOC, which we find estimates less SOC than CMB. In addition to the impact changes between sources, the day-to-day fractional impacts of each source are more stable in the ensemble results (Figures 2 and 3).

The unusually high fraction of DUST for 10 days (2nd–6th, 8th–11th, and 13th of January 2002) (Figure 3) is caused by the absence of CMB-RG, CMB-MM, and PMF results during those days due to missing observations. Therefore, the SA results from only two methods, CMB-LGO and CMAQ, are used for ensemble weighting and, as a result, the higher dust impact in CMAQ dominates. The chi-square value averages, which use the errors between observed and estimated PM$_{2.5}$ mass for those two months, are decreased to 11.1 and 21.3 in January ensemble and July ensemble, respectively, from 14.6 and 23.4 in CMB-LGO. One concern appears to be successfully addressed in this approach as day-to-day variability in the receptor-based approach is reduced and modeling performance is improved. We also recalculated the ensemble trained source impacts by removing CMAQ and, separately, CMB-MM to see how sensitive the ensemble procedure was to these two SA methods. We found that overall, the ensemble source apportionments changed by less than 3% (contribution of any one source to the total) when these methods were removed (Figure S2). The major reason that removing CMAQ and CMB-MM had such little impact is that the ensemble averages are driven by CMB-RG, CMB-LGO and PMF, which typically have smaller uncertainties than CMAQ or CMB-MM. This characteristic of ensembling with weights based on uncertainty has a net effect of reducing variability.

**TABLE 1.** Comparison of Monthly Fractional Source Impact in Ensemble Results with CMB-LGO Results at the Atlanta Jefferson Street (JST) Site for July 2001 and January 2002

<table>
<thead>
<tr>
<th>Source</th>
<th>January 2002</th>
<th>July 2001</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CMB-LGO</td>
<td>ensemble</td>
</tr>
<tr>
<td>GV</td>
<td>0.121 (0.049)</td>
<td>0.127 (0.067)</td>
</tr>
<tr>
<td>DV</td>
<td>0.096 (0.050)</td>
<td>0.098 (0.050)</td>
</tr>
<tr>
<td>DUST</td>
<td>0.014 (0.110)</td>
<td>0.033 (0.032)</td>
</tr>
<tr>
<td>BURN</td>
<td>0.120 (0.073)</td>
<td>0.201 (0.058)</td>
</tr>
<tr>
<td>COAL</td>
<td>0.005 (0.006)</td>
<td>0.015 (0.008)</td>
</tr>
<tr>
<td>SO$_4^{2-}$</td>
<td>0.226 (0.090)</td>
<td>0.199 (0.080)</td>
</tr>
<tr>
<td>NO$_3^{-}$</td>
<td>0.137 (0.063)</td>
<td>0.139 (0.062)</td>
</tr>
<tr>
<td>NH$_4^{+}$</td>
<td>0.085 (0.028)</td>
<td>0.106 (0.034)</td>
</tr>
<tr>
<td>SOC</td>
<td>0.196 (0.082)</td>
<td>0.081 (0.031)</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>14.6 (14.0)</td>
<td>11.1 (7.9)</td>
</tr>
</tbody>
</table>

**FIGURE 2.** Daily fractional contributions to PM$_{2.5}$ in SA results developed from CMB-LGO method at JST during January 2002.

**FIGURE 3.** Daily fractional contributions to PM$_{2.5}$ in ensemble results at JST during January 2002.
FIGURE 4. Measured (MBSP) (23–25), LGO-derived (22), and ensemble-based (EBSP) source profiles: (a) GV, (b) DV, (c) DUST. Bars represent one standard deviation. The LGO profile was derived from a 447-day data set; the EBSP SUMMER and EBSP WINTER were derived from a 24-day data set in July 2001 and a 21-day data set in January 2002, respectively. (Note: BURN and COAL are shown in Figure S3).
TABLE 2. Performance Measures of Source Apportionments Using EBSPs and MBSP

<table>
<thead>
<tr>
<th>used source profile</th>
<th>EBSP</th>
<th>MBSP</th>
<th>EBSP</th>
<th>MBSP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>summer</td>
<td>winter</td>
<td>summer</td>
<td>winter</td>
</tr>
<tr>
<td>chi-square</td>
<td>7.2 ± 6.9</td>
<td>19.5 ± 9.5</td>
<td>13.2 ± 16.8</td>
<td>26.6 ± 19.6</td>
</tr>
<tr>
<td>PM mass ratio, %</td>
<td>88.8 ± 9.7</td>
<td>87.2 ± 9.6</td>
<td>98.3 ± 15.4</td>
<td>93.6 ± 15.1</td>
</tr>
<tr>
<td>correlation of species, R</td>
<td>0.994 ± 0.013</td>
<td>0.994 ± 0.013</td>
<td>0.999 ± 0.006</td>
<td>0.999 ± 0.008</td>
</tr>
<tr>
<td>frequency of zero source impact, day</td>
<td>8</td>
<td>8</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>gasoline vehicle</td>
<td>0.046 ± 0.026</td>
<td>0.060 ± 0.028</td>
<td>0.095 ± 0.047</td>
<td>0.098 ± 0.040</td>
</tr>
<tr>
<td>diesel vehicle</td>
<td>0.083 ± 0.041</td>
<td>0.080 ± 0.037</td>
<td>0.101 ± 0.056</td>
<td>0.099 ± 0.045</td>
</tr>
<tr>
<td>road dust</td>
<td>0.037 ± 0.038</td>
<td>0.027 ± 0.031</td>
<td>0.014 ± 0.009</td>
<td>0.010 ± 0.009</td>
</tr>
<tr>
<td>biomass burning</td>
<td>0.089 ± 0.053</td>
<td>0.075 ± 0.040</td>
<td>0.182 ± 0.085</td>
<td>0.090 ± 0.072</td>
</tr>
<tr>
<td>fractional source impacts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coal burning</td>
<td>0.008 ± 0.007</td>
<td>0.007 ± 0.007</td>
<td>0.015 ± 0.010</td>
<td>0.008 ± 0.008</td>
</tr>
<tr>
<td>SULFATE</td>
<td>0.359 ± 0.081</td>
<td>0.363 ± 0.082</td>
<td>0.213 ± 0.085</td>
<td>0.228 ± 0.088</td>
</tr>
<tr>
<td>NITRATE</td>
<td>0.057 ± 0.039</td>
<td>0.058 ± 0.040</td>
<td>0.123 ± 0.057</td>
<td>0.129 ± 0.059</td>
</tr>
<tr>
<td>AMMONIUM</td>
<td>0.146 ± 0.028</td>
<td>0.147 ± 0.029</td>
<td>0.107 ± 0.032</td>
<td>0.113 ± 0.032</td>
</tr>
<tr>
<td>SOC</td>
<td>0.174 ± 0.088</td>
<td>0.182 ± 0.068</td>
<td>0.190 ± 0.067</td>
<td>0.225 ± 0.082</td>
</tr>
</tbody>
</table>

* Estimated-to-observed PM mass ratio.  b Correlation of species between observed and estimated amount.

Optimized Source Profile Compositions. Two sets of source profile compositions were calculated using the two different seasonally based ensemble results. Only 21 days in January 2002 and 24 days in July 2001 had all the ambient data (all ions, carbon fractions, metals, and gases) desired, and are used for the profile optimization procedure (Tables S4 and S5). Source profiles of the five primary sources, GV, DV, DUST, BURN, and COAL are calculated, using CMB-LGO, to most closely reproduce the ensemble SA results.

Ensemble-based source profiles (EBSPs) are similar to those found applying CMB-LGO, though some profile species vary markedly. The GV winter EBSP has an OC/EC ratio of 4.0, which is higher than the LGO-generated profile of 1.6 and that from NFRAQS of 2.3. However, the OC/EC ratio of 1.5 in the summer EBSP is similar to the other two. The total carbon (TC) fractions in PM$_{2.5}$ from gasoline vehicle emission were 0.71 in winter and 0.74 in the summer EBSP, similar to both 0.78 in NFRAQS and 0.67 in the LGO generated profiles. The Zn fraction (Zn is a good marker for gasoline vehicle impact) is 0.005 for winter and 0.009 for summer compared to those from the two other procedures of 0.009 (NFRAQS) and 0.008 (LGO).

The OC/EC ratios in the DV EBSPs, 0.52 in the winter and 0.45 in the summer, differ from the NFRAQS ratio of 0.27, but are within a typical range of 0.3 to 0.9, as reviewed by Hodan, et al. (37). Their findings suggested that the OC fraction of 0.19 and EC fraction of 0.75 in the current HDDV profile, 1996 EPA PM$_{2.5}$ Split Factor Profile Number 35600, might underestimate the EC fraction and overestimate the OC fraction due to a failure to consider variable burn conditions, truck size, fuel type, and control scenarios. TC contents in the DV profiles are stable by season, 0.86 for the winter and 0.88 for the summer EBSP, being similar to 0.93 in the NFRAQS profiles.

Biomass burning source profiles are characterized by high OC/EC ratios and high K fractions. The OC/EC ratio of 3 in both summer and winter BURN EBSPs is lower than the measured profiles of 4.1 from BRAVO and 4.7 from the LGO profiles. Further, the K fractions, which drive the source contribution calculation of biomass burning, are also lower in both the winter EBSP (0.024) and summer EBSP (0.035), than the MBSP (0.056) and LGO (0.063) profiles. These differences may be due to burning conditions, wood type, grate feature, burning rate, etc. For instance, the estimated uncertainties in the vegetative burning profile from the BRAVO study (MBSP in Figure 4) are up to 97% for the fraction of EC and 98% for the K fraction due to the variability in burned materials. On the other hand, the OC and TC fractions in biomass burning EBSPs do not show much seasonal variation, with the same OC fractions of 0.55 in both summer and winter profiles and TC fractions of 0.88 in the summer and 0.86 in the winter. These values are similar to the OC fraction of 0.64 and TC fraction of 0.80 in the BRAVO profiles.

Fractions of major species in the DUST EBSPs are similar across seasons: 0.21 and 0.22 for OC, 0.08 and 0.07 for Fe, 0.03 and 0.02 for K, 0.16 and 0.17 for Si in the winter and summer EBSPs, respectively. However, the Ca fraction differed by season, being 0.07 in winter and 0.03 in the summer. OC and Fe fractions in the DUST EBSPs are slightly higher than the typical ranges of 0.04–0.19 for OC and 0.02–0.05 for Fe that have been reported in other studies (24, 37). On the other hand, both the MBSPs and the LGO source profiles are in the lower end of this typical range for OC. The higher OC levels in the DUST EBSPs are driven by OC. The higher OC levels in the DUST EBSPs are driven by sulfuric acid emissions from biomass burning and the PM$_2.5$ emissions from coal burning is theoretically stable by season. This bias comes from using PM$_2.5$ results that show a larger impact from coal burning, which can produce additional sulfate by oxidation of the concurrent SO$_2$ emissions.

Species fractions in the EBSPs are optimized from the relations between fixed source impacts (ensemble source contribution) and PM$_{2.5}$ compositions at the receptor. Thus, theoretically, the EBSPs are expected to best replicate the ensemble source impact, but there is no direct method to estimate the accuracy of the optimized profiles. Instead, performance measures from applying CMB-LGO source apportionment using EBSPs are compared to the original CMB-LGO using MBSP (Table 2).

Source Apportionment Using EBSPs. The two seasonal EBSPs are applied to estimate source impacts of each day in 2002 (over 320 days for which all species of ambient data are available) at JSF. Source apportionments for the winter season (January, February, November, and December) use winter EBSPs while source apportionments for the other eight months (from March to October) use summer EBSPs. Source impacts developed from ensemble-trained CMB (using...
EBSs show a decreased number of zero-impact days in winter, source apportionment introduces the relative importance changes between primary sources and secondary sources. During the winter period, source apportionment shows a high percentage of biomass burning in the area, while in the summer, the relative source importance changes between sources are not as significant as in the winter. This can be attributed to the decreased importance of biomass burning in the summer period.

Source apportionment results suggest that the CVAPs and CMB-LGO approaches are complementary in terms of source attribution. CVAPs are effective in identifying the relative importance of sources, while CMB-LGO can be used to identify the relative contributions of sources to specific pollutants. The ensemble approach used in this study allows for the development of more robust and reliable source apportionment results.

The results of this study are relevant for policy makers and health professionals as they provide insights into the sources of particulate matter and their potential health impacts. The findings can be used to inform strategies for reducing particulate matter concentrations and improving air quality.

**Acknowledgments**

This study was supported by the U.S. EPA under Grants RD-83215901, RD83369601, RD83369801, RD836876902, and RD836876903. This information is available free of charge via the Internet at http://pubs.acs.org.

**Supporting Information Available**

This information is available free of charge via the Internet at http://pubs.acs.org.
(26) Paatero, P.; Tapper, U. Positive matrix factorization - a non-
negative factor model with optimal utilization of error-estimates

apportionment of Phoenix PM$_{2.5}$ aerosol with the UNMIX receptor

Foley, C. *EPA Positive Matrix Factorization (PMF) 3.0 Funda-
mentals & User Guide*; EPA 600/R-08/108; U.S. Environmental

(29) Dennis, R.; Roselle, S.; Gilliam, R.; Arnold, J., High time resolved
comparisons for in-depth probing of CMAQ fine particle and
gas predictions. In *2004 Models 3 Workshop, Research Triangle

Models-3 Community Multiscale Air Quality (CMAQ) Modeling
System*; EPA/600/R-99/030; U.S. Environmental Protection

(31) Houyoux, M. R.; Vukovich, J. M. Updates to the Sparse Matrix
Operator Kernel Emissions (SMOKE) Modeling System and
Integration with Models-3. The Emission Inventory: Regional
Strategies for the Future, Raleigh, NC, October 26–28, 1999;
Raleigh, NC, 1999.

(32) Seaman, N. L. Meteorological modeling for air-quality assess-

(33) Lee, S.; Liu, W.; Wang, Y. H.; Russell, A. G.; Edgerton, E. S. Source
apportionment of PM$_{2.5}$: Comparing PMF and CMB results for
four ambient monitoring sites in the southeastern United States.

(34) Hansen, D. A.; Edgerton, E. S.; Hartsell, B. E.; Jansen, I. J.;
Kandasamy, N.; Hidy, G. M.; Blanchard, C. L. The southeastern
aerosol research and characterization study: Part 1-overview.

(35) Marmur, A.; Unal, A.; Mulholland, J. A.; Russell, A. G. Optimiza-
tion-based source apportionment of PM$_{2.5}$ incorporating gas-
3254.

Snow, R. Composition of light-duty motor vehicle exhaust

(37) Hodan, W. *Recommendations for the Update and Improvement
of Existing PM$_{2.5}$ Split Factors*; Report to EPA by Pacific
Environmental Services, Inc.: Research Triangle Park, NC,