

A source apportionment of U.S. fine particulate matter air pollution

George D. Thurston*, Kazuhiko Ito, Ramona Lall

Department of Environmental Medicine, School of Medicine, New York University, 57 Old Forge Rd, Tuxedo, NY 10987, USA

ARTICLE INFO

Article history:

Received 8 September 2010

Received in revised form

31 March 2011

Accepted 24 April 2011

Keywords:

Chemical Speciation Network (CSN)

Source apportionment

Fine particulate matter mass ($PM_{2.5}$)

Absolute Principal Component Analysis (APCA)

ABSTRACT

Using daily fine particulate matter ($PM_{2.5}$) composition data from the 2000–2005 U.S. EPA Chemical Speciation Network (CSN) for over 200 sites, we applied multivariate methods to identify and quantify the major fine particulate matter ($PM_{2.5}$) source components in the U.S. Novel aspects of this work were: (1) the application of factor analysis (FA) to multi-city daily data, drawing upon both spatial and temporal variations of chemical species; and, (2) the exclusion of secondary components (sulfates, nitrates and organic carbon) from the source identification FA to more clearly discern and apportion the $PM_{2.5}$ mass to primary emission source categories. For the quantification of source-related mass, we considered two approaches based upon the FA results: 1) using single key tracers for sources identified by FA in a mass regression; and, 2) applying Absolute Principal Component Analysis (APCA). In each case, we followed a two-stage mass regression approach, in which secondary components were first apportioned among the identified sources, and then mass was apportioned to the sources and to other secondary mass not explained by the individual sources. The major U.S. $PM_{2.5}$ source categories identified via FA (and their key tracers) were: Metals Industry (Pb, Zn); Crustal/Soil Particles (Ca, Si); Motor Vehicle Traffic (EC, NO_2); Steel Industry (Fe, Mn); Coal Combustion (As, Se); Oil Combustion (V, Ni); Salt Particles (Na, Cl) and Biomass Burning (K). Nationwide spatial plots of the source-related $PM_{2.5}$ impacts were confirmatory of the factor interpretations: ubiquitous sources, such as Traffic and Soil, were found to be spread across the nation, more unique sources (such as Steel and Metals Processing) being highest in select industrialized cities, Biomass Burning was highest in the U.S. Northwest, while Residual Oil combustion was highest in cities in the Northeastern U.S. and in cities with major seaports. The sum of these source contributions and the secondary $PM_{2.5}$ components agreed well with the U.S. $PM_{2.5}$ observed during the study period (mean = $14.3 \mu g m^{-3}$; $R^2 = 0.94$). Apportionment regression analyses using single-element tracers for each source category gave results consistent with the APCA estimates. Comparisons of nearby sites indicated that the $PM_{2.5}$ mass and the secondary aerosols were most homogenous spatially, while traffic $PM_{2.5}$ and its tracer, EC, were among the most spatially representative of the source-related components. Comparison of apportionment results to a previous analysis of the 1979–1982 IP Network revealed similar and correlated major U.S. source category factors, albeit at lower levels than in the earlier period, suggesting a consistency in the U.S. spatial patterns of these source-related exposures over time, as well. These results indicate that applying source-apportionment methods to the nationwide CSN can be an informative avenue for identifying and quantifying source components for the subsequent estimation of source-specific health effects, potentially contributing to more efficient regulation of $PM_{2.5}$.

© 2011 Elsevier Ltd. All rights reserved.

1. Background

Long-term exposure to fine particulate matter ($PM_{2.5}$) air pollution has been associated with increased risk of human mortality (e.g., Ozkaynak and Thurston, 1987; Dockery et al., 1993;

* Corresponding author. Department of Environmental Medicine, School of Medicine, New York University, Room 270, 57 Old Forge Rd, Tuxedo, NY 10987, USA. Tel.: +1 845 731 3564; fax: +1 845 351 5472.

E-mail address: george.thurston@nyu.edu (G.D. Thurston).

Pope et al., 2002). However, the composition of $PM_{2.5}$ mass can vary significantly with its origins, and it is likely that particles from different sources can have differing toxicities (NRC, 2004; WHO, 2007). The U.S. EPA decision to add a $PM_{2.5}$ fine particle National Ambient Air Quality Standard (NAAQS) in 1997 was an advancement in public health protection, as many toxic elements are most concentrated in that particle mass size fraction. However, it is possible that varying levels of protection are still provided by that standard, depending on variations in local $PM_{2.5}$ composition and associated sources. Future standard-setting and emissions controls implementation may consider these differences, but the challenge

remains as to which components within the particles are most damaging to health. To address that need, the analyses presented here were conducted as part of our continuing efforts, for the Health Effects Institute's National Particle Component Toxicity (NPACT) Initiative, to investigate the extent to which individual PM_{2.5} sources and/or components are responsible for the PM_{2.5}-mortality associations previously found in the American Cancer Society (ACS) cohort (Pope et al., 2002, 2004).

While numerous individual site source apportionments of PM_{2.5} have been published in the literature (e.g., Thurston and Spengler, 1985; Koutrakis and Spengler, 1987; Lewis et al., 2003; Kim et al., 2004; Qin et al., 2006; Chow et al., 2008), past investigations conducting national source-attributions of particulate matter (PM) mass are very limited in number. The first nationwide identification of sources of PM pollution using multivariate methods was conducted by Blifford and Meeker (1967), who applied rotated Factor Analysis (FA) to both particle elemental composition and mass data from 30 U.S. cities in order to determine the underlying groups of elements, and thereby, an indication of the sources that most strongly influence PM concentrations across the nation. On the basis of the sources' elemental emission characteristics, they identified independent components of particle variance as due to specific source classes (e.g., automotive exhaust). However, since Blifford and Meeker used traditional normalized factor analyses, they could not quantitatively attribute particle mass contributions to their respective sources. In 1987, Thurston and coworkers conducted a nationwide analysis of the Inhalable Particle (IP) Network data collected in some 36 Metropolitan Statistical Areas (MSAs) by the U.S. EPA in the late 1970's and early 1980's (Thurston et al., 1984; Ozkaynak and Thurston, 1987). This work relied on the Absolute Principal Component Analysis (APCA) source-apportionment method developed by Thurston and Spengler (1985). That analysis of the IP Network data represented the first published mortality analysis of long-term PM_{2.5}, as well as the first publication that used source-apportionment methods to assess the air pollution source categories most important to health (Ozkaynak and Thurston, 1987).

There is currently an opportunity to conduct a more extensive nationwide source-apportionment analysis for the U.S. using recent and more comprehensive PM_{2.5} chemical speciation and mass data. Beginning in 1999, the U.S. EPA instituted a nationwide PM_{2.5} mass and composition based Chemical Speciation Network (CSN) that provides the PM_{2.5} mass characteristics needed to characterize the compositional nature of PM_{2.5} exposures in the U.S., as well as to potentially conduct a statistical apportionment of air pollution impacts for cities across the U.S. (US EPA, 1999). While numerous studies have employed these CSN data in source-apportionment analyses of individual cities (e.g., for New York City: Ito et al., 2004; Qin et al., 2006), to date no study has yet attempted a source apportionment of the newly available data on a nationwide basis. Bell et al. (2007) has provided a descriptive assessment of this nationwide CSN dataset, which characterizes the variations in key PM_{2.5} trace constituents across cities in the U.S., but did not apply statistical models to these data to investigate source components. Thus, this study is the first to use CSN data in a comprehensive nationwide source-apportionment analysis.

For this study, a methodology was developed aimed at more specifically ascertaining mass contributions associated with their original primary source categories. While the total PM_{2.5} mass includes particles as a result of both primary (i.e., emitted directly from sources) and secondary (i.e., formed in the atmosphere) origins, the key tracers of sources identified in source-apportionment studies are typically based upon the elements that are usually highly enriched in the primary generation/direct emissions from that source class (e.g., Ni and V from residual fuel oil combustion). Therefore, it was decided to allow only these "primary" emission

trace elements to be used to define the underlying factors (i.e., source categories) in multivariate analyses, rather than also including secondary aerosols (such as sulfates) that are not specific to a particular source class. Most factor analysis-based source-apportionment studies that include PM constituents typically associated with secondary particle formation (e.g., sulfur) usually identify an ambiguous secondary aerosol source category (e.g., called "secondary sulfates"). However, as noted by Thurston and Lioy (1987): "Secondary and transported aerosol contributions complicate the apportionment of fine mass to sources". Thus, while the secondary aerosols explain much of the mass, they can confound the ability of a factor analysis to identify and quantify the original sources of PM_{2.5}. Therefore, in an effort to identify more distinct unambiguous sources, we have, in this new work, excluded predominantly secondary PM constituents (i.e., sulfates, nitrates and organic carbon) from the initial source identification factor analysis. The subsequent two-stage mass regression steps, set up to quantify the mass associated with each source category, then do incorporate both the secondary components and total PM mass, thereby apportioning both primary and secondary contributions to each source category.

Unlike earlier nationwide source-apportionment analyses (e.g., Blifford and Meeker, 1967; Thurston et al., 1984), this new work considers the daily values of measured variables (e.g., trace element concentrations), rather than annual average values at each site. Therefore, both the day-to-day variations in elemental concentrations, as well as the site-to-site variations, are employed to identify sources in these source-apportionment models. Additionally, factor analysis source-apportionment models can allow the inclusion of gaseous pollutants to aid the identification of sources. Since we are attempting to identify and quantify sources based on primary sources, being able to include gaseous pollutants, such as NO₂ as a tracer of local motor vehicle combustion, is clearly an added advantage of such a factor analysis model for better identifying source components.

The primary aim of this study is to apply source-apportionment methods to the newly available nationwide CSN data on a consistent nationwide basis in order to provide insights into the nation's principal sources of PM_{2.5} air pollution in the U.S. Such source-apportionment techniques can also provide an important input to epidemiology in order to help evaluate the relative toxicity of PM from various sources (Thurston et al., 2005). Therefore, the trace elemental concentrations, as well as the source-apportioned masses estimated by this work, can potentially be included in the further investigations of PM_{2.5} composition and health outcomes. These would provide an improved understanding of the role of composition and sources in past associations found between PM_{2.5} exposure and adverse morbidity and mortality outcomes (e.g., Pope et al., 2002).

2. Materials and methods

2.1. Data

The U.S. EPA Air Quality System provides routine air monitoring measurements for PM_{2.5} mass, PM_{2.5} anions (sulfate, nitrate) and cations (ammonium, sodium, and potassium), trace elements (Na through Pb on the periodic table), total carbon [including organic carbon (OC) and elemental carbon (EC)], and gaseous pollutant data (CO, NO₂, SO₂, O₃). These data have been compiled by the Health Effects Institute (HEI) and for this work PM_{2.5} chemical speciation data, EC, OC, Nitrates (NO₃⁻), SO₄, NH₄ and NO₂, SO₂ and CO gaseous pollutant data were download via HEI's Atmospheric and Environmental Research (AER) database (HEI, 2007). A detailed description of the original EPA data sources and quality assurance

steps conducted by HEI are provided at <https://hei.aer.com/aboutDatabase.php>.

Chemical speciation and NO₂ gaseous pollutant data were used in the factor analysis model used in this work. Sulfur (S), OC, NO₃[−] and total PM_{2.5} mass were considered in the mass regression models to quantify source contributions and estimate source profiles. In addition to the chemical speciation data considered directly in the factor analysis model, NH₄ and SO₄ were also considered in estimating source profiles. SO₂ and CO gaseous pollutants were only considered for the purpose of verification of source categories.

Data for all available CSN monitoring sites in the contiguous continental U.S. (i.e., excluding Hawaii, Alaska, and Puerto Rico) for 2/9/2000 through 12/31/2005 were considered in this work (for consistency with health data being analyzed within this time period), resulting in data for approximately 273 monitoring sites. Gaseous NO₂ 24-hour average concentrations were supplied by the HEI AER contractor for the nearest neighboring NO₂ sampler to each CSN site in a metropolitan area, as available. We only considered the use of NO₂ gaseous pollutant in the factor analysis source-apportionment model as it is considered a reliable marker of traffic sources, and to aid in the separation of this source versus other local combustion sources (e.g., biogenic burning). Other gases were investigated, but limitations in monitoring sites/data at or near CSN monitors would have resulted in additional “losses” of study sites if they were also included, so the only gaseous pollutant included in these factor analyses was NO₂.

2.1.1. Data quality control

Prior to the factor analysis source apportionment, quality control (QC) steps were applied to the PM_{2.5} CSN data to remove sites or observations from sites that did not meet study requirements, as described below. First, for a CSN site to be considered in the factor analysis source-apportionment model, a minimum of 60 observations was required (i.e., at least equivalent to 6 months of measurements at every 3rd day sampling). Two CSN sites situated in El Paso (Texas) were excluded from the analysis, as these sites are situated along the U.S.-Mexico border, and are dominated by unrepresentative cross-border pollution sources. Data outliers for the given set of sites were also identified for potential exclusion by computing the mean and standard deviations for each site, and flagging extreme outlier observations where the value was found to be greater than 12 standard deviations from the mean for that element for that site. Also, observations influenced by January 1st or July 4th fireworks (i.e., January 1st or the first week of July) were removed from the dataset to avoid the influence of firework events (associated with spikes in potassium, usually an important tracer of biogenic wood burning), which are relatively infrequent events, and not considered as being among the ubiquitous air pollution sources with major contributions on a year-round basis.

Once all QC steps were applied and sites without NO₂ excluded, 212 out of the original 273 U.S. CSN sites with sufficient data during the study period (2000–2005) were available for inclusion in the source-apportionment factor analysis. On average, the retained sites averaged approximately 220 observations per site.

Because of known Organic Carbon (OC) positive artifact (Chow et al., 2010), we applied the “regression intercept approach” described in Watson et al. (2009). We applied this approach on a region-season basis.

2.2. Methods

The fundamentals of source-apportionment receptor modeling techniques have been previously presented (e.g., Cooper and Watson, 1980; Henry et al., 1984). Currently, there are different

multivariate options available for apportioning mass among sources, such as, Positive Matrix Factorization, or PMF (Paatero, 1997), Multilinear Engine, or ME (Paatero, 1999), UNMIX (Henry and Kim, 1989) and Absolute Principal Component Analysis, or APCA (Thurston and Spengler, 1985). These techniques were part of the recent EPA Source Apportionment Workshop that found comparable results provided by the different techniques (Thurston et al., 2005; Hopke et al., 2006). While the application of alternate source-apportionment PMF model was considered as part of this work, a lack of the required trace concentration uncertainty values at the time of data download for our analysis limited the applicability of the PMF model in this dataset to a smaller subset of the available data, or required the estimation of uncertainties for sites and years of data lacking these uncertainty values. Therefore, for practical considerations, APCA was chosen for the work conducted here.

Traditional Factor Analysis (FA) and Principal Component Analysis (PCA) are useful methods for identifying source components contributing to the PM mass (e.g., see Cooper and Watson, 1980). The first step in this work was to identify the sources of PM nationwide using factor analysis. The second step was to apportion PM Mass to each of these sources. For the second step, two alternate mass regression models were attempted based on the factor analysis results, (a) using single key tracers for sources identified by FA, and (b) applying Absolute Principal Component Analysis (APCA) (Thurston and Spengler, 1985).

2.2.1. Factor analysis model for the identification of sources

In a manner analogous to traditional Chemical Mass Balance (CMB) source apportionment, the factor analysis approach to source apportionment assumes that the total concentration of each “observable” (e.g., elemental concentration) is made up of the sum of contributions from each of a smaller number of underlying pollution source factors (e.g., see Hopke et al., 1976). However, these source-related factors and their associated profiles are derived by the factor analysis from the correlation matrix of the input trace constituents, and are instead outputs of the analysis (not inputs to the model, as in the case for CMB). Seventeen input variables were considered for the factor analysis of the nationwide data, and these included daily values for: PM_{2.5} arsenic (As), calcium (Ca), copper (Cu), chlorine (Cl), iron (Fe), lead (Pb), manganese (Mn), nickel (Ni), selenium (Se), vanadium (V), silica (Si), zinc (Zn), potassium (K), sodium (Na), magnesium (Mg), elemental carbon (EC), and the gaseous pollutant nitrogen dioxide (NO₂). These specific trace elements were chosen for this analysis consistent with past source-apportionment studies (US EPA Report, 2003; DRI Report, 2005), as they are usually well measured by the x-ray fluorescence elemental analysis method (employed by the CSN), and because they were also deemed of key relevance as a potential tracer of a pollution source category.

The source identification factor analysis conducted here intentionally did not include PM constituents associated with secondary aerosols (i.e., sulfur (S), organic carbon (OC), and nitrates (NO₃)), as when included into a source-apportionment model these constituents typically align numerous source tracers on one ambiguous “secondary aerosols” factor, confounding the individual source category identification process. Instead, the secondary mass associated with each source were apportioned to the source-related factors in a subsequent two-stage mass regression procedure, as described below.

The source-apportionment factor analysis was conducted using statistical software SAS (version 9.1; SAS Institute Inc., Cary, NC). We considered factor analysis solutions for 6 to 9 factors, as well as both “Varimax” and “Oblimin” rotations that allow for orthogonal and oblique solutions, respectively, as described in Harman (1976).

2.2.2. Computation of absolute principal component scores for mass apportionment

While traditional Factor Analysis (FA) and Principal Component Analysis (PCA) are useful for identifying underlying source component “factors” contributing to the PM mass, they do not directly provide an apportionment of mass. However, the solutions can be employed to provide such a quantitative solution. In this work, the Absolute Principal Components Analysis (APCA) approach developed by Thurston and Spengler (1985) was used to produce such quantitative apportionments. Briefly, as provided in more detail in Thurston and Spengler (1985), this approach uses the scoring coefficients from the factor analysis to score a “zero” day pollution, which is then subtracted from observed factor scores to derive estimates of the absolute PC scores (APCS). These provide positive indices of source impact, upon which PM_{2.5} mass concentrations can be regressed to achieve a source apportionment. This approach has been found to give comparable apportionment results to other multivariate PM_{2.5} mass source-apportionment approaches (Thurston et al., 2005).

2.3. Quantification of sources impacts: two-stage mass regression models

Based on the factor analysis, two approaches for the apportionment of source-related mass were employed. First, a simple mass apportionment approach was applied to regress mass on single key tracers in a multivariate mass regression model, one tracer for each source class identified by FA (Kleinman et al., 1980). Second, a mass apportionment based on the factor analysis was also conducted using Absolute Principal Component Analysis (APCA) approach developed by Thurston and Spengler (1985).

In both cases, a linear mixed model was used for the mass apportionment regression analysis of nationwide tracer data or absolute factor scores. The linear mixed model was applied to the nationwide data, consisting of multiple daily measurement observations, i , from multiple monitoring sites, j . The tracer-mass and APCS-mass beta relationships were modeled as fixed terms, with a random intercept by monitoring site (I_j), thereby providing an overall beta (e.g., PM_{2.5} mass per absolute score increment) estimate across all sites.

Although the factor analyses steps did not explicitly include S, OC, or NO₃[−] in the model, we employed “secondary aerosol” mass regression steps prior to the PM_{2.5} mass apportionment regression models, in order to thereby account for the secondary mass “unexplained” ($I_j + e_{ij}$) by the absolute factors, APC _{k} , as follows:

$$\text{sulfur}_{ij} = \text{APC}_{1ij} + \dots + \text{APC}_{kij} + I_j + e_{ij} \quad (1a)$$

$$\text{OC}_{ij} = \text{APC}_{1ij} + \dots + \text{APC}_{kij} + I_j + e_{ij} \quad (1b)$$

$$\text{nitrate}_{ij} = \text{APC}_{1ij} + \dots + \text{APC}_{kij} + I_j + e_{ij} \quad (1c)$$

Therefore, for each secondary component, an “unexplained” remainder portion (not accounted for by the factors identified) could then be included in the final PM_{2.5} mass regression step to ascertain a complete mass apportionment for 197 sites. Analogous equations were also applied for the analysis using the eight key tracers, except that the tracers instead served as the dependent variables.

2.3.1. Single tracer-mass regression model

Based on the factor analysis results, this approach designated single key tracers (TR _{k}) for each of the k sources (e.g., Si for soil). Therefore, the PM_{2.5} mass regression step for the tracer-mass regression model, which included daily measurements i for each

site j , was set up for k source tracers and “unexplained” secondary mass components (S', OC' and NO₃'), such that:

$$\begin{aligned} \text{PM}_{ij} = & \beta_1 \times \text{TR}_{1ij} + \dots + \beta_k \times \text{TR}_{kij} + \beta_{(k+1)} \\ & \times S'_{ij} + \beta_{(k+2)} \times \text{OC}'_{ij} + \beta_{(k+3)} \times \text{NO}_3'_{ij} + I_j \end{aligned} \quad (2)$$

The beta coefficients from this model provided tracer-to-mass conversion factors, and the reciprocal of each of these coefficients yielded estimates of the fraction of each key tracer associated with source-related mass.

2.3.2. Absolute principal component mass regression model

In this mass regression approach using results from the factor analysis, the total PM_{2.5} mass was regressed onto the APCS and the three unexplained secondary mass terms. The model is set up for k PM source categories and “unexplained” secondary mass components (S', OC' and NO₃'), such that:

$$\begin{aligned} \text{PM}_{ij} = & \beta_1 \times \text{APC}_{1ij} + \dots + \beta_k \times \text{APC}_{kij} + \beta_{(k+1)} \\ & \times S'_{ij} + \beta_{(k+2)} \times \text{OC}'_{ij} + \beta_{(k+3)} \times \text{NO}_3'_{ij} + I_j \end{aligned} \quad (3)$$

By deriving “fixed” beta coefficients in this linear mixed regression model, absolute factor score to mass conversions were estimated. This analysis thereby provided an overall estimate of the source-to-mass relationships throughout the U.S. on a year-round basis.

As part of a sensitivity analysis, we also considered separate regression models for each region-season subset, using an analogous mass source-apportionment method (as shown above) to test whether the interpretation of results are influenced by U.S. regions and seasons. We considered ten separate region-season regression analyses that each considered 5 regions in the U.S. (Northeast, Southeast, Southwest, Northwest and California) and two seasons (winter = October–March, summer = April–September). Sites were designated one of the 5 regions based on their latitude and longitude. Latitude >39°N and longitude >−92°E are considered North and East, respectively, except for California, which was considered as a region separate from the Northwest or Southwest. Results from the region-season models were then compared to the nationwide model to evaluate any differences in the predicted PM_{2.5} contributions.

2.3.3. Computing spatial representativeness of source impacts

In order to assess the spatial representativeness of the CSN central-site monitors for the various PM_{2.5} source impacts and trace constituent concentrations, we compared all monitors located within 20 miles of another (mean separation distance = 9.9 miles). Using this subset of monitoring data, we calculated the difference in site concentration means for all pairs of sites within the 20-mile radius of each other ($n = 53$ unique site pairs nationwide). From these sites, we calculated the % coefficient of variation [(standard error/mean) × 100] of the mean difference for PM_{2.5}, each of the source category mass concentrations, and their respective key tracers.

2.3.4. Calculating source compositional profiles

In this work, source compositional profiles were estimated for both of the mass regression approaches applied. In the simple tracer-mass regression model, the reciprocal of the beta coefficients ($1/\beta_k$) provided estimates of the compositional profile percentages for each of the source tracers. For the APCS-mass regression model, source profiles were estimated by regressing each variable (e.g., daily trace element concentrations) onto the APCA's daily mass contributions for all eight sources in a linear model with an intercept term. We similarly assessed the source profile fractions of OC,

NH₄, NO₃ and SO₄, although these were not considered in the FA model. These source profiles yielded estimates of the compositional mix of the source mass, as well as providing an aid to the interpretation of their respective source category factors.

2.4. Computing average source-mass estimates for U.S. MSAs

Daily contributions of total PM_{2.5} and source-apportioned PM_{2.5} mass, as estimated using two-stage mass regression models, were employed to estimate source mass for U.S. MSAs. From this, summer (April–September) and winter (October–March) seasonal mean estimates for PM_{2.5} and PM_{2.5} sources were computed for each site, and the average of these two values provide an overall mean estimate for total and source-related PM_{2.5} mass at each site considered in this work. Due to missing data when applying the mass regression models, 30 sites with fewer than 60 observations for computing season average were excluded from MSA mean estimation, and overall means were restricted to non-negative values. Overall mean total and source-related PM_{2.5} were then computed for MSAs by averaging across sites in each MSA.

3. Results

The analysis dataset ultimately included 46,478 daily observations for 212 CSN monitoring sites distributed throughout the U.S. (Fig. 1). For the sites considered in this work, there was an average of 220 observations per site (or almost two years of data per site at the common network data collection rate of every-third-day sampling). Table 1 provides a nationwide summary of PM_{2.5} and its constituents considered in the factor model. Averages by season and by the five U.S. regions considered in the sensitivity analyses are also provided.

3.1. Factor analysis results: identification of major sources and their key tracers

An eight-factor solution using Varimax rotation was chosen as the “optimum” solution, based on an examination of both the factor eigenvalues (i.e., the data variance explained by the component) and the source-related interpretability of the factors. The oblique solution was found to provide similar results to the orthogonal solution, so the latter was chosen, as it requires the factor scores to be uncorrelated (i.e., independent of one another), which is desirable in the subsequent apportionment regression steps. Table 2 provides the “factor loadings” (i.e., the correlations between the factor scores and original input variables) that aid in the

interpretation of the factors. The physical identification of the factors is possible by comparing elements having the highest loading (correlation) with each component with elements emitted in large amounts and/or in high enrichment by known source types (relative to other elements or other sources), as previously documented for major U.S. source classes in the literature (e.g., see Cooper and Watson, 1980; Gordon, 1980; Thurston and Spengler, 1985; U.S. EPA, 2003). However, a potential limitation of this and all other source-apportionment analyses is that other sources of the key trace elements resulting from a source category not identified in the analysis can be statistically “picked up” and assigned to the most similar identified source component, potentially overstating the source-specific contribution estimates. Thus, each source category definition should be viewed as including the named source, plus any other sources similarly enriched in that source’s key trace elements.

On the basis of this information, the first factor in Table 2 is identifiable as Crustal/Soil particles because of its high loadings on the common earth crustal elements (e.g., Ca, Si) most commonly found together in soil-derived particles (e.g., wind-blown dust). Similarly, the second component is identifiable as associated with Metals Processing Industry because of its higher loadings on Pb and Zn. Historically, these elements were also related to refuse burning and automobile emissions, respectively, but the phasing out of open refuse burning and leaded gasoline has essentially eliminated these two alternative sources of these two elements in ambient air in the U.S., allowing for a better clarification of the interpretation of this factor as related to metals processing emissions. The third factor is identifiable as associated with Motor Vehicle Traffic from its loadings on both EC and NO₂. While each of these two pollutants individually has a variety of sources, the source most associated with both EC and NO₂ emissions is traffic. Without the inclusion of NO₂ to the analysis, it was not possible for the factor analysis to separate a distinct traffic component. Copper (Cu) was also found to be associated with Motor Vehicle Traffic source category factor, and this element has also been attributed to traffic-related brake wear by past studies (e.g., Schauer et al., 2006), consistent with our interpretation of the factor. Unfortunately, while EC is present in much higher quantities in diesel motor vehicle emissions (Schauer et al., 2006), a unique tracer for separating gasoline vs. diesel fuel vehicles was not available in this dataset, so a breakdown of traffic into these two sub-categories was not possible using the CSN database as it exists today. The fourth factor is identifiable as Salt Particles based upon its loadings on both Na and Cl, and would include contributions from both marine aerosols (e.g., along the coasts) and road salt (e.g., in the northern states). The fifth component is identified as Residual Oil Combustion based upon its loadings on both V and Ni, two elements known to be enriched in heavier fuel oils, such as residual fuel oil #6 (as burned by power plants) and cargo ship “Bunker” fuels. The sixth factor is identified as associated with the Steel Industry, based on its loadings on both Fe and Mn, elements known to be enriched in emissions from ferromanganese steel furnaces. The seventh factor is identifiable as Coal Combustion based on its loading on both As and Se. Although these two elements are also emitted by metals operations and smelters, respectively, their simultaneous loading on this factor, and the fact that coal emissions are highly enriched in and major U.S. sources of both of these elements (e.g., see Helble, 2000) uniquely identifies this factor as related to the combustion-of coal, some 94% of which is presently burned in coal-fired electric power plants in the U.S. (U.S. EIA, 2010). The last factor is identified as Biomass Burning (primarily as wood burning and forest fires in the U.S.), based on its association with potassium (K), an element emitted during vegetative burning (e.g., see U.S. EPA, 2003).

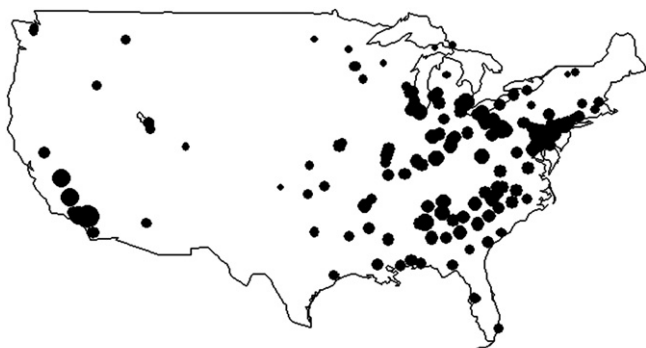


Fig. 1. CSN sites ($n = 212$) considered in the U.S. Source-Apportionment Analysis. Each circle radii are proportional to the overall mean PM_{2.5} concentration for each site. (relative to a maximum value = $26.5 \mu\text{g m}^{-3}$).

Table 1Observed PM_{2.5} Mass and Constituents, Means \pm Standard Errors of Site Average ($N = 212$), By Region and Season.

		US (Overall)	US (Winter)	US (Summer)	NE	SE	SW	NW	CA
$\mu\text{g m}^{-3}$	PM _{2.5}	14.2 \pm 0.2	13.7 \pm 0.2	14.6 \pm 0.3	14.7 \pm 0.2	15.0 \pm 0.1	11.2 \pm 0.2	10.8 \pm 0.2	18.9 \pm 0.3
	S	1.2 \pm 0.0	0.9 \pm 0.0	1.5 \pm 0.0	1.4 \pm 0.0	1.4 \pm 0.0	1.0 \pm 0.0	0.5 \pm 0.0	0.8 \pm 0.0
	NO ₃	1.7 \pm 0.1	2.5 \pm 0.1	1.0 \pm 0.1	2.0 \pm 0.0	1.1 \pm 0.0	0.9 \pm 0.0	1.8 \pm 0.1	5.6 \pm 0.2
	OC	4.2 \pm 0.1	4.2 \pm 0.1	4.2 \pm 0.1	3.9 \pm 0.1	4.6 \pm 0.1	3.6 \pm 0.1	4.3 \pm 0.1	6.4 \pm 0.1
ng m^{-3}	As	1.3 \pm 0.0	1.4 \pm 0.0	1.3 \pm 0.1	1.5 \pm 0.1	1.4 \pm 0.0	0.9 \pm 0.0	1.1 \pm 0.0	1.1 \pm 0.0
	Ca	54.9 \pm 2.5	47.6 \pm 2.4	62.6 \pm 2.8	45.4 \pm 1.6	50.8 \pm 2.9	73.4 \pm 2.8	69.7 \pm 2.6	74.1 \pm 3
	Cu	3.9 \pm 0.2	4.0 \pm 0.2	3.8 \pm 0.2	4.0 \pm 0.2	3.8 \pm 0.2	3.0 \pm 0.1	3.7 \pm 0.2	7.5 \pm 0.2
	Cl	29.0 \pm 3.3	41.0 \pm 4.4	18.6 \pm 2.9	30.6 \pm 3.2	16.6 \pm 2.2	42.9 \pm 5.6	32 \pm 2.3	50.6 \pm 1.4
	Fe	87.3 \pm 4.4	82.3 \pm 4.9	93.7 \pm 4.3	97.3 \pm 5.6	73.6 \pm 3.3	88.3 \pm 3.5	74.3 \pm 2.3	134.9 \pm 4.2
	Pb	4.7 \pm 0.3	4.8 \pm 0.3	4.6 \pm 0.4	6.0 \pm 0.3	4.5 \pm 0.4	2.6 \pm 0.1	3.5 \pm 0.1	3.9 \pm 0.1
	Mn	3.6 \pm 0.7	3.7 \pm 0.8	3.5 \pm 0.7	3.6 \pm 0.3	4.5 \pm 1.3	2 \pm 0.1	3.2 \pm 0.2	2.8 \pm 0.1
	Ni	1.7 \pm 0.2	1.8 \pm 0.2	1.5 \pm 0.1	2.4 \pm 0.2	1.0 \pm 0.0	0.9 \pm 0	1.2 \pm 0.1	2.5 \pm 0.1
	Se	1.3 \pm 0.1	1.4 \pm 0.1	1.3 \pm 0.1	1.8 \pm 0.1	1.4 \pm 0.0	0.6 \pm 0	0.7 \pm 0	0.9 \pm 0
	V	2.0 \pm 0.1	2.0 \pm 0.1	2.1 \pm 0.1	2.3 \pm 0.2	1.7 \pm 0.1	1.9 \pm 0.1	1.7 \pm 0.1	3.3 \pm 0.1
	Si	104.3 \pm 4.1	72.2 \pm 3.2	137.5 \pm 5.6	75.0 \pm 2.1	96.5 \pm 2.6	189.2 \pm 6.1	108.7 \pm 2.9	152.8 \pm 3.7
	Zn	16.2 \pm 1.7	18.3 \pm 1.7	14.4 \pm 1.8	19.6 \pm 1.1	17.1 \pm 2.8	9.7 \pm 0.4	11.5 \pm 0.4	12.5 \pm 0.6
	K	68.2 \pm 2.2	70.5 \pm 2.9	65.5 \pm 1.8	61.5 \pm 2.6	70.2 \pm 1.5	74.1 \pm 1.6	71.8 \pm 2.7	91.6 \pm 1.6
	Na	65.9 \pm 3.1	62.7 \pm 2.7	68.6 \pm 3.8	51.0 \pm 1.5	62.6 \pm 2.4	92.2 \pm 4.5	69.1 \pm 3.7	154.1 \pm 4.1
	Mg	9.0 \pm 0.4	6.3 \pm 0.3	11.8 \pm 0.5	7.5 \pm 0.3	7.3 \pm 0.3	15.3 \pm 0.6	9.1 \pm 0.3	15.6 \pm 0.3
	EC	646.0 \pm 22.2	724.4 \pm 26.0	568.5 \pm 20.2	678.0 \pm 24.5	623.5 \pm 19.2	521.3 \pm 20.3	637.1 \pm 17.6	1030.5 \pm 21.5
PPB	NO ₂	14.1 \pm 0.5	16.4 \pm 0.5	12.5 \pm 0.4	17.2 \pm 0.4	10.7 \pm 0.4	11.2 \pm 0.5	14.3 \pm 0.4	20.4 \pm 0.4

U.S. map of MSA-average estimated impacts confirm the factor interpretations (see Fig. 2): Motor Vehicle traffic-related impacts were elevated throughout the U.S. and highest in Southern California; Soil impacts were high in the desert Southwest; Steel impacts were highest in cities with major steel works (e.g., Birmingham, AL); Coal combustion impacts were highest in the Ohio Valley region (e.g., Pittsburgh, PA); Residual Oil burning impacts were highest in northeastern cities that use residual oil for heating and/or electricity generation (e.g., New York City, NY; Providence, RI; and, New Haven, CT), as well as in cities with major ports (e.g., Los Angeles, CA; Savannah, GA; Seattle, WA; and, Newark, NJ-NYC, NY), consistent with a major impact by cargo ships burning “bunker” fuel in those port cities.

As a further verification of the sources, correlations of factor scores with variables not included in the analysis were also examined. For example, correlations were tested for daily observations of gaseous pollutants, such as SO₂ and CO, at nearby sites for the subsets of days for which each was available. Among the sources, SO₂ ($n = 42,093$) was most correlated over space and time in the daily data with oil combustion ($r = 0.32$, p -value < 0.001) and coal combustion ($r = 0.22$, p -value < 0.001). As also expected, CO ($n = 39,324$) was most correlated

with traffic ($r = 0.38$, p -value < 0.001). In addition, the coal combustion source factor had the highest factor score correlation with particulate sulfur, a secondary aerosol component that largely results from coal-burning emissions of SO₂, especially in the Eastern U.S. ($r = 0.25$, p -value < 0.001 , in the daily dataset). Similarly, we also investigated the source factor correlations with the CSN PM_{2.5} mercury (Hg), a pollutant primarily emitted into the air by coal combustion in the U.S. (U.S. EPA, 2005), and we found that the coal factor also had the highest source factor correlation with collocated Hg ($r = 0.19$, p -value < 0.001), further supporting the Se–As related source factor's interpretation as an index of PM_{2.5} impacts from coal combustion.

3.2. Source-related mass estimates

Summary mean \pm standard errors for source-related mass contributions and other secondary mass are shown in Tables 3 and 4. The overall nationwide average estimate for the 11 source categories presented in the results is based on averaging across all 197 U.S. sites considered in the regression models. Region-season averages were derived from the season-specific site averages for all sites in each region.

Table 2Factor “Loadings” for the 8-Factor Solution. Key signature tracers noted in bold. [* PM_{2.5} not included in Factor Analysis model].

	Crustal/Soil	Metals-Related	Traffic-Related	Salt Aerosols	Residual Oil	Steel Industry	Coal-Burning	Biomass Burning
As	0.11	0.58	0.07	−0.08	0.01	−0.08	0.49	0.05
Ca	0.75	0.06	0.15	0.01	0.05	0.04	0.01	0.16
Cu	0.10	0.27	0.73	0.05	−0.10	0.14	−0.10	−0.15
Cl	−0.12	0.00	0.09	0.58	0.02	0.01	0.34	0.44
Fe	0.46	0.14	0.29	0.02	0.12	0.64	0.10	0.16
Pb	0.04	0.87	0.09	0.02	0.01	0.06	0.11	−0.01
Mn	−0.01	0.13	−0.02	0.02	0.02	0.93	0.02	0.02
Ni	−0.01	0.02	0.15	0.02	0.82	0.08	0.03	−0.09
Se	0.00	0.04	0.09	0.01	0.01	0.07	0.87	−0.03
V	0.07	0.03	0.08	0.08	0.82	−0.01	−0.01	0.09
Si	0.85	−0.01	−0.03	0.12	0.00	0.06	0.02	0.07
Zn	−0.02	0.75	0.09	0.13	0.04	0.22	−0.12	0.09
K	0.33	0.09	0.11	0.03	−0.01	0.10	−0.06	0.78
Na	0.01	0.11	−0.05	0.86	0.10	0.02	−0.08	0.09
Mg	0.43	−0.02	0.00	0.72	−0.03	0.03	−0.04	−0.24
EC	0.14	0.09	0.68	0.02	0.22	0.03	0.29	0.30
NO ₂	−0.03	−0.05	0.71	−0.09	0.34	−0.04	0.10	0.13
*PM _{2.5}	0.17	0.13	0.34	0.01	0.13	0.06	0.25	0.26

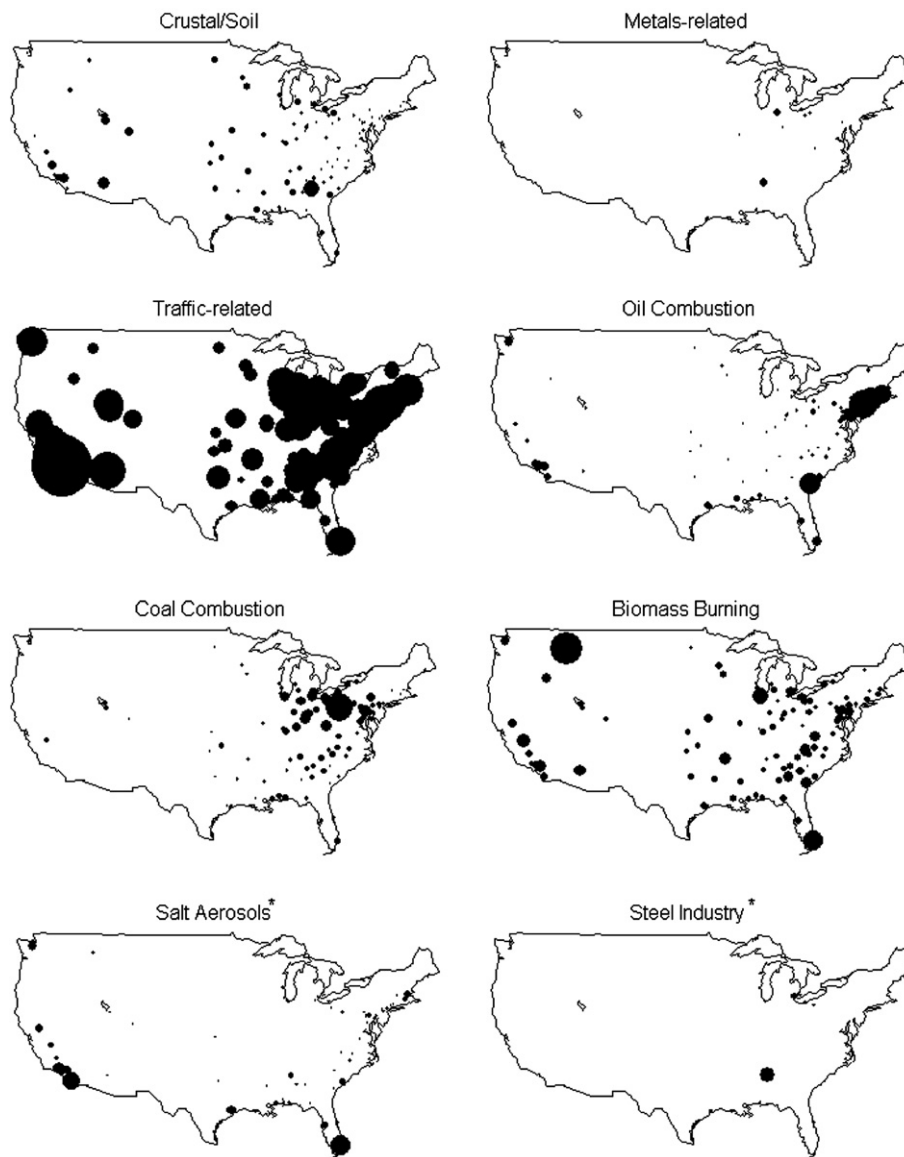


Fig. 2. Spatial plots of source-related mass impacts at 102 MSAs in the U.S., where the circle radii are proportional to the overall annual source contribution estimates for each MSA (relative to a maximum value = $12.4 \mu\text{g m}^{-3}$; *except for Salt Aerosols and Steel Industry, which are relative to a maximum value = $2.5 \mu\text{g m}^{-3}$).

Overall MSA averages for the source $\text{PM}_{2.5}$ contributions were also computed from the CSN site-specific averages. These allowed comparisons with historical MSA-based estimates for similar sources identified using data from the Inhalation Particle Network (IPN) that operated between 1979 and 1982.

3.2.1. Tracer-mass regression model

For the simpler single tracer-mass regression model, the single key tracers designated to represent each of the eight source categories identified were: Si (Soil), Se (Coal), Ni (Oil), K (Wood), Fe

(Steel), Cl (Salt), Zn (Metals) and EC (Traffic). These tracers were chosen based on their high correlations with each of the factor scores, respectively. Therefore, the mass apportionment regression was set up for the 8 key tracers and 3 “unexplained” or “other” secondary mass components, such that:

$$\begin{aligned} \text{PM}_{ij} = & \beta_1 \times \text{Si}_{ij} + \beta_2 \times \text{Se}_{ij} + \beta_3 \times \text{Ni}_{ij} + \beta_4 \times \text{K}_{ij} + \beta_5 \times \text{Fe}_{ij} \\ & + \beta_6 \times \text{Cl}_{ij} + \beta_7 \times \text{Zn}_{ij} + \beta_8 \times \text{EC}_{ij} + \beta_9 \times \text{S}'_{ij} + \beta_{10} \\ & \times \text{OC}'_{ij} + \beta_{11} \times \text{NO}_3'_{ij} + I_j \end{aligned}$$

Table 3
Results from the Single Tracer-Mass Apportionment Method. Nationwide Mean \pm Std. Error Estimate: a) $\text{PM}_{2.5}$ impact by source; and, b) %Key tracer associated with each source.

Source: (Tracer)	Crustal/ Soil (Si)	Metals-Related (Zn)	Traffic-Related (EC)	Salt Aerosols (Cl)	Oil Combustion (Ni)	Steel Industry (Fe)	Coal Combustion (Se)	Biomass Burning (K)	Other Sulfates (S)	Other Organic Carbon (OC)	Other Nitrates (NO_3)
$\mu\text{g m}^{-3}$	0.2 ± 0.1	0.1 ± 0.1	4.3 ± 2.1	0.1 ± 0.2	0.2 ± 0.3	0.1 ± 0.1	0.8 ± 0.6	1.5 ± 0.8	4.3 ± 2.0	0.9 ± 1.5	2.6 ± 1.0
%Tracer	42.3 ± 3.6	21.6 ± 1.8	15.3 ± 0.1	23.6 ± 1.9	0.9 ± 0.0	91.4 ± 19.3	0.2 ± 0.0	4.3 ± 0.1	20.3 ± 0.1	79.4 ± 1.0	79.4 ± 0.6

Table 4

Average Mass Contributions ($\mu\text{g m}^{-3}$) by Source Category for each region and season considered based on (a) U.S. overall Mass Regression model, and (b) separate Region/Season Mass regression models.

	Crustal/Soil	Metals-Related	Traffic-Related	Salt Aerosols	Oil Combustion	Steel Industry	Coal Combustion	Biomass Burning	Other Sulfates	Other Organic Carbon	Other Nitrates	Predicted Mass	PM _{2.5}
a.													
Overall	0.8 ± 0.0	0.2 ± 0.0	4.6 ± 0.2	0.1 ± 0.0	0.9 ± 0.1	0.0 ± 0.0	1.1 ± 0.1	1.3 ± 0.1	4.3 ± 0.2	0.5 ± 0.1	0.6 ± 0.1	14.5 ± 0.3	14.3 ± 0.2
Winter													
NE	0.4 ± 0.0	0.2 ± 0.0	5.4 ± 0.2	0.1 ± 0.0	1.3 ± 0.2	0.0 ± 0.0	1.5 ± 0.2	1.5 ± 0.2	2.7 ± 0.1	0.0 ± 0.0	1.1 ± 0.1	14.4 ± 0.4	13.7 ± 0.3
SE	0.6 ± 0.1	0.2 ± 0.1	4.4 ± 0.2	0.1 ± 0.0	0.7 ± 0.1	0.0 ± 0.0	1.1 ± 0.1	1.5 ± 0.1	3.7 ± 0.1	1.4 ± 0.1	0.3 ± 0.1	14.0 ± 0.2	13.3 ± 0.2
SW	1.1 ± 0.2	0.1 ± 0.0	5.2 ± 1.0	0.1 ± 0.0	0.6 ± 0.1	0.0 ± 0.0	0.5 ± 0.1	1.7 ± 0.3	1.8 ± 0.4	0.2 ± 0.1	0.5 ± 0.2	11.7 ± 1.0	11.4 ± 0.6
NW	0.7 ± 0.1	0.2 ± 0.0	5.3 ± 0.6	0.1 ± 0.0	0.7 ± 0.2	0.0 ± 0.0	0.8 ± 0.1	2.2 ± 0.4	0.7 ± 0.2	0.4 ± 0.2	2.0 ± 0.4	13.1 ± 0.7	12.9 ± 0.6
CA	1.0 ± 0.2	0.1 ± 0.0	9.4 ± 1.2	0.2 ± 0.0	0.9 ± 0.2	0.0 ± 0.0	0.4 ± 0.2	2.8 ± 0.4	0.3 ± 0.1	2.1 ± 0.9	5.0 ± 1.4	22.1 ± 2.4	22.6 ± 2.6
Summer													
NE	0.9 ± 0.1	0.3 ± 0.0	4.8 ± 0.2	0.1 ± 0.0	1.0 ± 0.1	0.0 ± 0.0	1.4 ± 0.1	0.8 ± 0.1	6.5 ± 0.2	0.1 ± 0.0	0.1 ± 0.0	15.8 ± 0.4	15.7 ± 0.3
SE	1.1 ± 0.1	0.3 ± 0.1	3.3 ± 0.2	0.1 ± 0.0	0.6 ± 0.1	0.0 ± 0.0	0.9 ± 0.1	1.1 ± 0.1	7.5 ± 0.3	1.2 ± 0.1	0.0 ± 0.0	16.1 ± 0.3	16.6 ± 0.3
SW	1.8 ± 0.2	0.1 ± 0.0	3.0 ± 0.5	0.1 ± 0.0	0.6 ± 0.1	0.0 ± 0.0	0.5 ± 0.1	1.6 ± 0.2	3.8 ± 0.8	0.2 ± 0.1	0.0 ± 0.0	11.8 ± 0.7	12.1 ± 0.9
NW	1.4 ± 0.2	0.1 ± 0.0	3.2 ± 0.3	0.1 ± 0.0	0.8 ± 0.2	0.0 ± 0.0	0.5 ± 0.1	1.4 ± 0.2	1.1 ± 0.3	0.5 ± 0.2	0.1 ± 0.1	9.3 ± 0.6	8.7 ± 0.5
CA	1.3 ± 0.3	0.0 ± 0.0	5.8 ± 0.8	0.6 ± 0.1	1.6 ± 0.4	0.0 ± 0.0	0.2 ± 0.1	0.9 ± 0.2	2.6 ± 0.7	0.3 ± 0.1	2.6 ± 1.4	15.8 ± 2.7	15.4 ± 2.4
b.													
Overall	1.2 ± 0.1	0.2 ± 0.0	4.7 ± 0.2	0.2 ± 0.0	0.9 ± 0.1	0.0 ± 0.0	1.2 ± 0.1	1.4 ± 0.1	3.0 ± 0.1	0.4 ± 0.0	0.7 ± 0.1	13.9 ± 0.2	14.3 ± 0.2
Winter													
NE	0.5 ± 0.0	0.3 ± 0.0	5.6 ± 0.2	0.1 ± 0.0	1.1 ± 0.2	0.0 ± 0.0	1.3 ± 0.2	1.1 ± 0.1	2.5 ± 0.1	0.1 ± 0.0	1.0 ± 0.1	13.7 ± 0.4	13.7 ± 0.3
SE	0.7 ± 0.1	0.3 ± 0.1	3.2 ± 0.2	0.1 ± 0.0	0.5 ± 0.1	0.1 ± 0.1	1.1 ± 0.1	1.7 ± 0.1	4.5 ± 0.1	0.8 ± 0.1	0.7 ± 0.1	13.6 ± 0.2	13.3 ± 0.2
SW	1.1 ± 0.2	0.4 ± 0.1	3.7 ± 0.7	0.1 ± 0.0	0.7 ± 0.1	0.0 ± 0.0	0.4 ± 0.0	1.5 ± 0.2	2.4 ± 0.4	0.0 ± 0.0	0.8 ± 0.3	11.2 ± 0.7	11.4 ± 0.6
NW	0.5 ± 0.1	0.5 ± 0.1	6.3 ± 0.7	0.4 ± 0.1	0.8 ± 0.2	0.0 ± 0.0	1.2 ± 0.2	3.5 ± 0.6	1.0 ± 0.2	0.1 ± 0.0	0.9 ± 0.2	15.2 ± 1.0	12.9 ± 0.6
CA	0.3 ± 0.1	0.3 ± 0.1	13.9 ± 1.8	1.0 ± 0.1	2.2 ± 0.5	0.1 ± 0.1	1.1 ± 0.4	6.5 ± 1.0	0.8 ± 0.1	0.2 ± 0.1	0.9 ± 0.7	27.2 ± 2.3	22.6 ± 2.6
Summer													
NE	2.1 ± 0.1	0.1 ± 0.0	5.4 ± 0.3	0.1 ± 0.0	1.2 ± 0.1	0.0 ± 0.0	1.5 ± 0.2	0.8 ± 0.1	2.5 ± 0.2	0.2 ± 0.0	0.4 ± 0.0	14.4 ± 0.4	15.7 ± 0.3
SE	1.2 ± 0.1	0.3 ± 0.1	3.8 ± 0.3	0.1 ± 0.0	0.7 ± 0.1	0.0 ± 0.0	1.6 ± 0.1	1.2 ± 0.1	5.2 ± 0.2	1.1 ± 0.1	0.3 ± 0.0	15.6 ± 0.3	16.6 ± 0.3
SW	2.1 ± 0.3	0.2 ± 0.1	3.4 ± 0.6	0.1 ± 0.0	0.7 ± 0.1	0.0 ± 0.0	0.1 ± 0.0	1.3 ± 0.2	3.1 ± 0.6	0.3 ± 0.1	0.4 ± 0.1	11.7 ± 0.6	12.1 ± 0.9
NW	1.9 ± 0.2	0.1 ± 0.0	2.0 ± 0.2	0.1 ± 0.0	0.6 ± 0.2	0.0 ± 0.0	0.3 ± 0.0	1.2 ± 0.2	1.4 ± 0.4	0.4 ± 0.1	0.3 ± 0.1	8.3 ± 0.5	8.7 ± 0.5
CA	1.3 ± 0.3	0.0 ± 0.0	3.6 ± 0.5	0.7 ± 0.1	1.7 ± 0.4	0.0 ± 0.0	0.2 ± 0.1	0.8 ± 0.2	2.5 ± 0.4	0.1 ± 0.1	2.8 ± 1.5	13.8 ± 2.5	15.4 ± 2.4

Therefore, the beta coefficients β_k provide the “tracer-to-source mass” conversion factors. For example, β_1 is the “Si-to-Soil Mass” conversion factor. Table 3 provides the source-mass estimates and % key tracer associated with the source.

3.2.2. APCA mass regression model and results

The mass regression model using the APCA Scores (APCS) was set up for the 8 sources and 3 “unexplained” or “other” secondary mass components, such that:

$$\begin{aligned} \text{PM}_{ij} = & \beta_1 \times \text{SOIL}_{ij} + \beta_2 \times \text{COAL}_{ij} + \beta_3 \times \text{OIL}_{ij} + \beta_4 \times \text{WOOD}_{ij} \\ & + \beta_5 \times \text{STEEL}_{ij} + \beta_6 \times \text{SALT}_{ij} + \beta_7 \times \text{METALS}_{ij} + \beta_8 \\ & \times \text{TRAFFIC}_{ij} + \beta_9 \times \text{S}'_{ij} + \beta_{10} \times \text{OC}'_{ij} + \beta_{11} \times \text{NO}_3'_{ij} + I_j \end{aligned}$$

Therefore, the product of the β_k and the absolute scores provide the mass contributions associated with each source category.

The model beta coefficients β_9 , β_{10} , β_{11} for the secondary components, S, OC, and NO_3 , were 5.09, 1.22 and 1.25, respectively. These estimates compare favorably with known molar conversion factors for elemental sulfur to ammonium sulfate (4.125), organic carbon to organic mass (1.4) and nitrate to ammonium nitrate (1.29) (Malm et al., 1994), supporting the validity of this mass regression approach.

Table 4 provides the overall and the region-season average estimates \pm standard error for mass contributions by 8 source categories identified. For a better understanding of the variations across the multiple locations throughout the U.S., we used average source-related contributions for the 100 MSAs considered in this work. Fig. 3 provides the boxplots by source category that depicts the variation in source contributions ($\mu\text{g m}^{-3}$) across the MSAs. Sources, such as, motor vehicles, soil/crustal, and wood burning are found to be widespread throughout the U.S. (i.e., minimum MSA value > 0). On the other hand, certain sources, such as Iron/Steel

and metals-related, are only found for a very limited number of sites in the U.S.

3.2.3. Source contributions: single tracer vs. APCA approach

The overall predicted mass based on the single tracer approach is highly correlated with actual PM_{2.5} ($R^2 = 0.94$) (Fig. 4). As with the Single Tracer approach, the APCA approach also predicts PM_{2.5} mass well when considering the daily estimates or the computed overall MSA-average estimates (Figs. 4 and 5, respectively). Although the Single Tracer and APCA approach in this nationwide analysis provide similar results, the advantage of the APCA approach over the Single Tracer method is that it relies on multiple key tracers, yielding more clearly interpretable, independent

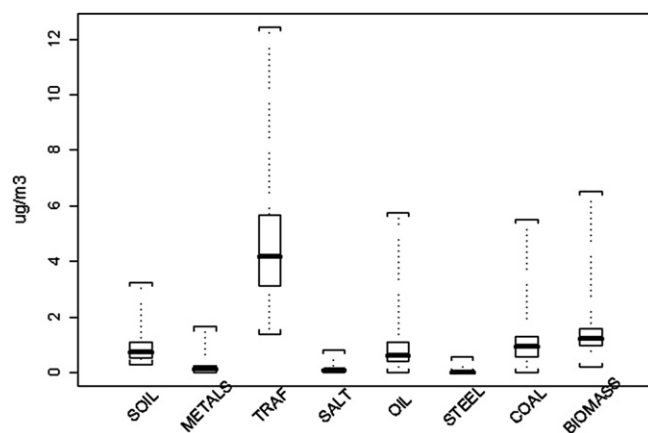


Fig. 3. Distribution of average source-mass estimates ($\mu\text{g m}^{-3}$) for 100 MSAs throughout the U.S.

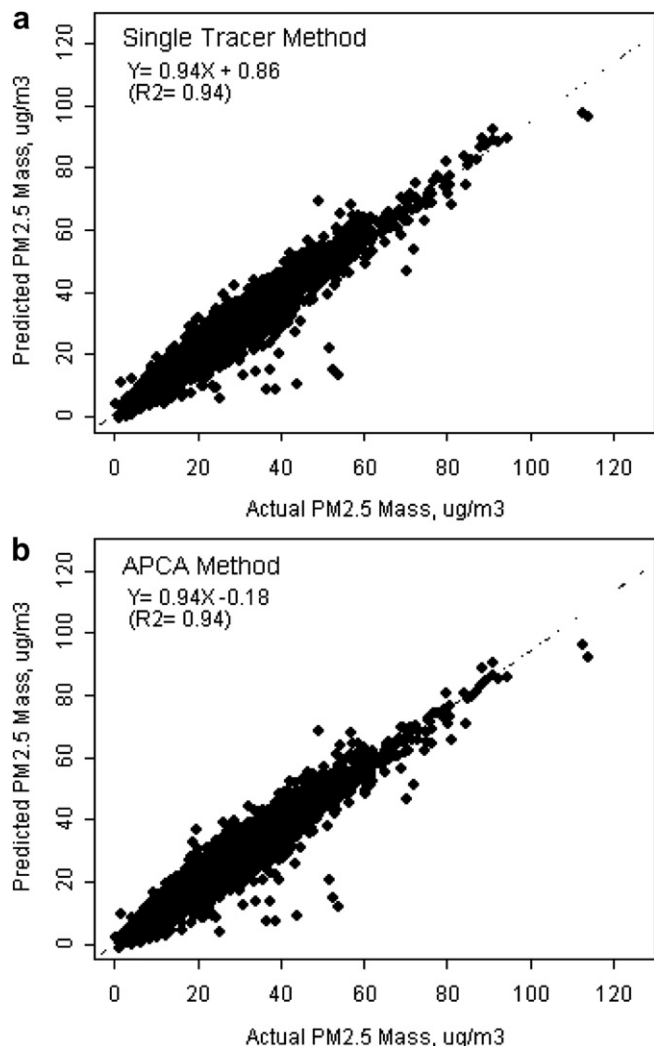


Fig. 4. Actual vs. Predicted Mass ($n = 46,478$) using (a) Single Tracer Method, and, (b) APCA approach (bottom) in a U.S. nationwide, year-round model.

(uncorrelated and not collinear), and unique source categories than the single tracer approach.

3.2.4. Sensitivity analysis results: source contributions based on region-season model

Table 4 shows the results of the sensitivity analyses from the APCA approach that compares the U.S. model versus the region-season models, indicating only minor differences in overall and by region-season average estimates. However, the plot of the region-season predicted mass vs. observed PM_{2.5} shown in Table 4, and reflected in Fig. 6, indicates that the nationwide estimates actually fit the observed data better than a regional analysis. Notably, the much higher California wintertime regional traffic-related impact estimate ($13.9 \mu\text{g m}^{-3}$ vs. $9.4 \mu\text{g m}^{-3}$ from the national model) contributes to an overestimate of PM_{2.5}: $27.2 \mu\text{g m}^{-3}$ PM_{2.5} in region-season model vs. observed $22.6 \mu\text{g m}^{-3}$ PM_{2.5} for that region-season in Table 4. In contrast, the national year-round model predicts a much closer-to-observed $22.1 \mu\text{g m}^{-3}$ PM_{2.5} for that same region and season. Thus, the simpler year-round U.S. model was apparently more stable, and was deemed to be preferable for the purpose of this nationwide analysis. The addition of more CSN data in the coming years may allow sufficient statistical power to determine regional profiles in future analyses.

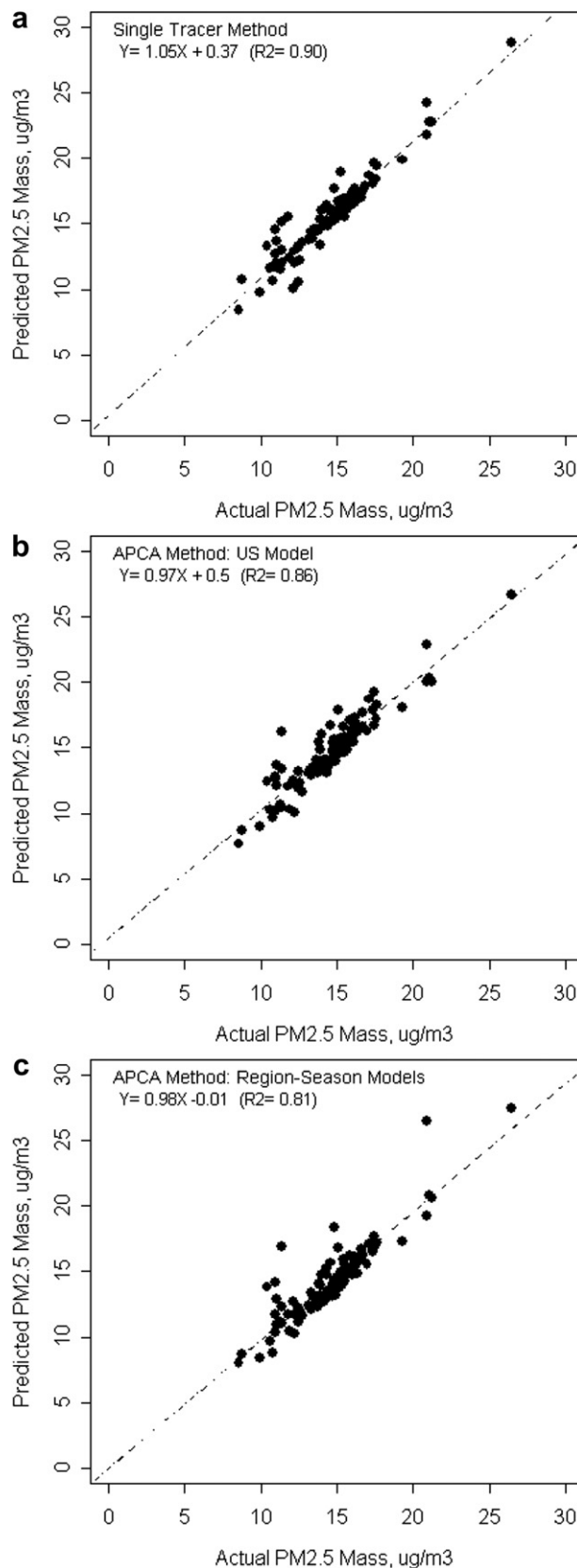


Fig. 5. Actual vs. Predicted Mass Estimates for 100 MSAs from the (a) Key Tracers, (b) APCA US and (c) APCA Region-Season Mass apportionment regression models.

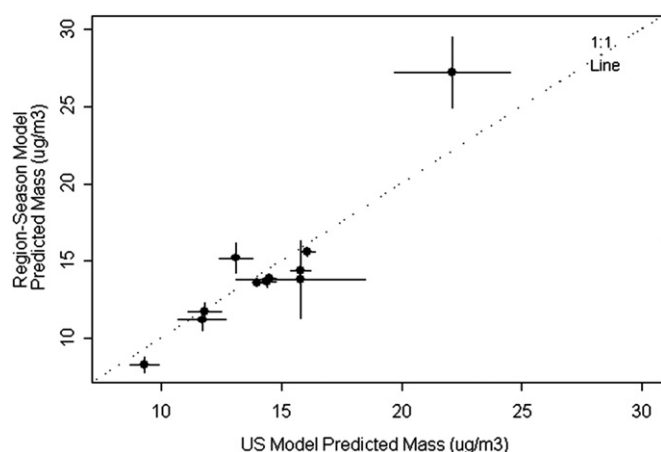


Fig. 6. Region-Season estimates of $PM_{2.5}$ mass, based on APCS, for estimates from Region-Season Models vs. U.S. Nationwide Model. Vertical and horizontal bars indicate the standard errors of the site-specific mean concentration estimates in a region-season for each model.

3.2.5. Spatial representativeness of source impacts

The % Coefficient of Variation of the mean between-site difference were calculated for all monitoring sites with 20 miles of each other for: 1) $PM_{2.5}$ mass, the mass associated with each of the source categories; and, 2) their respective key source tracers. The estimated mass CV's were, from lowest to highest: $PM_{2.5}$ = 11.1%; Traffic = 11.3%; Salt = 14.4%; Oil = 14.7%; Crustal/Soil = 16.8%; Wood = 21.5%; Metals-related = 25.5%; Coal = 30.9%; and, Steel = 54.6%. Among the key tracers, the CV values were, from lowest to highest: OC = 11.6%; S = 12.2%; Si = 13.9%; EC = 16.3%; Ni = 19%; Fe = 20.7%; Cl = 24.9%; K = 28.1%; Se = 28.6%; Zn = 32.2%, and Mn = 52.0%. These results indicate that the $PM_{2.5}$ mass and its secondary aerosol constituents (e.g., S and OC) were the most spatially homogeneous between sites, and therefore the most spatially representative central-site measurements, as expected. Among the source categories, traffic $PM_{2.5}$ and its tracer, EC, were the most spatially representative (i.e., lowest in %CV between paired sites). The metals and steel-related source $PM_{2.5}$ impact estimates and their associated trace elements (e.g., Zn, Mn) were generally among the least spatially representative of nearby monitors (i.e., having the largest %CVs).

3.3. Source compositional profiles

As shown in Fig. 7, the source composition profiles indicate the majority of the source-related mass is usually explained by the secondary components associated with the individual sources. However, the percentages of trace elements and metals associated with each source aids in the identification of the source, and provides perspective on the mass estimates.

For a comparison of these source profiles with other studies findings, we chose cities in different regions of the U.S. that are associated with major contributions from specific sources identified in this work, including: Phoenix (Soil) (Brown et al., 2007); Ohio (Coal) (Kim et al., 2007); New York City (Residual Oil) (Qin et al., 2006); Seattle (Wood) (Maykut et al., 2003; Wu et al., 2007); Detroit (Steel) (Gildemeister et al., 2007); Birmingham (Metals) (Baumann et al., 2008); and, Los Angeles (Traffic) (Kim and Hopke, 2007). In these various comparative cases, we find comparable estimates for key tracer percentages of mass, as follows: Soil (Si = 10.56% vs. 4%–75%); residual oil (Ni = 0.22% vs. 0.1%–3%); wood burning (K = 2.31% vs. 1%–20%); Steel (Fe = 22.64% vs. 7%–90%); Metals (Zn = 6.02% vs. 0.6%–10%) and, Salt (Cl = 15.66% vs.

0.1%–35%). In addition, our estimate for EC (10.34%) associated with the traffic source profile compares well with published estimates for “Gasoline Vehicle” (3%–20%) identified by these studies. Given that most past studies have not clearly identified a “Coal” factor, as they included secondary aerosols in their factor model, it was not feasible to compare the %Se associated with coal in this study vs. other apportionment studies.

3.4. Historical vs. current source impact estimates

Two past studies have conducted a similar nationwide analysis (Blifford and Meeker, 1967; Thurston et al., 1984; Ozkaynak and Thurston, 1987), with the more recent of these two using the Inhalable Particle Network (IPN). The IPN was set up by the EPA in the late 1970s to provide nationwide database of fine and coarse particle mass and elemental data. The network collected data from mid-1979 to the end of 1982. These data allow for a comparison between current vs. past concentrations for select 32 MSAs where data are available for both time periods. Results from this work were therefore also compared with source-apportionment APCA analysis of the nationwide IPN dataset (from Thurston et al., 1984), and these comparisons are provided here as Supplemental material to this analysis.

The U.S. $PM_{2.5}$ source apportionment using IPN data indicated 5 major source groups as being most influential: ‘Soil’ (Si, Fe); ‘Motor Vehicles’ (Pb, Br); ‘Residual oil combustion’ (Ni, V); ‘Iron and Steel industry operations’ (Mn, Zn, Fe); and, ‘Coal combustion’ (Se, S) (see Supplemental Table S1). In addition to these five from Ozkaynak and Thurston (1987), the current analysis, using more comprehensive $PM_{2.5}$ composition data, also identifies ‘Biomass Burning’, a separate ‘Metals Industry’ and ‘Salt Aerosols’. The Harvard “Six Cities” Study also observed similar sources: ‘Mobile’ (Pb), ‘Coal’ (Se), ‘Crustal’ (Si), ‘Fuel oil’ (V), ‘Manganese’ (Mn), ‘Salt’ (Cl), ‘Metals’ (Zn and Ni), albeit in a much smaller set of cities (Laden et al., 2000).

While a set of major source categories similar to those identified in this work were also found for the past IPN analyses, current concentrations, across all the source mass and trace elements, are found to be lower than past levels, and this difference is found to be statistically significant for all sources and their key tracers (two-tailed paired *t*-test significance test at <0.05 probability) (see supplemental Fig. S1). The correlations are remarkably strong for past vs. current source-mass and trace elements across multiple MSAs throughout the U.S., given the number of years that have lapsed between sample collections. It is interesting to note that Pb (historically used as a marker for leaded gasoline fueled motor vehicles) is at extremely low concentrations today and poorly correlated with past levels ($r = 0.07$), as expected. However, the “automobile” and “traffic” factors reported by the two respective studies (i.e., based on the IPN and CSN databases, respectively) still correlate remarkably well ($r = 0.56$), despite using different tracers for that motor vehicle source class in different eras.

4. Discussion

This is the first study to conduct a nationwide factor analysis and source apportionment of the newly available daily CSN data. Additionally, this work takes a unique approach to the identification and quantification of PM sources: secondary aerosol constituents were not included in the source component identification factorization step. By only including tracers of primary emissions in the factor analysis (e.g., Se, As, Ni, V, EC), while excluding tracers of secondary formation (i.e., S, OC, and NO_3^-), this work was able to more clearly discriminate the original primary sources of $PM_{2.5}$ to which we aimed to attribute the $PM_{2.5}$ mass. This avoided the identification of ambiguous “secondary aerosol” factors as aerosol

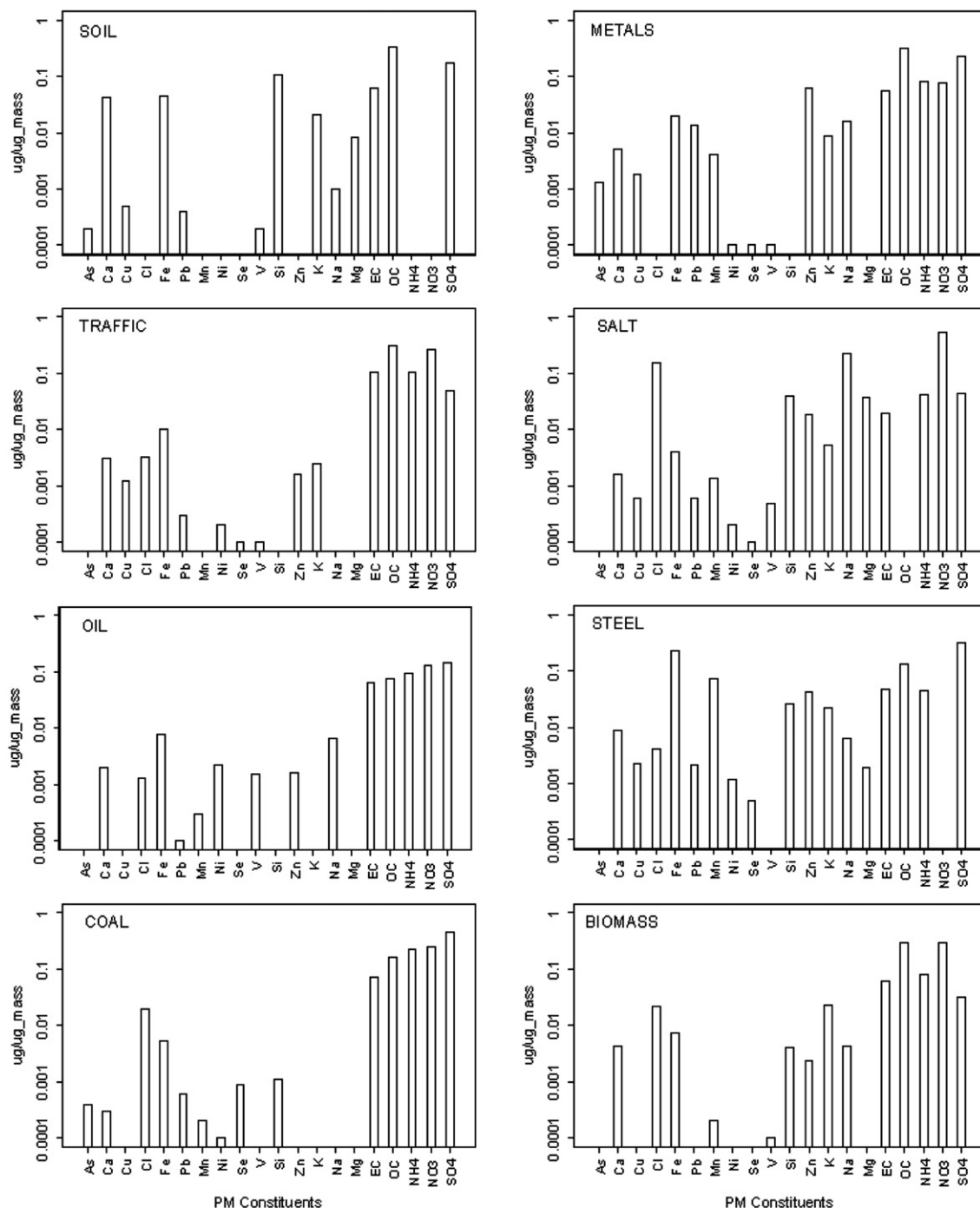


Fig. 7. Estimated fractional elemental source profiles of identified source categories.

“source” components, as reported in prior studies, to which primary sources may be subsumed and aggregated in the analysis. Especially notable, as a result, was the clearer identification of a major coal-related contribution to the mass based on the factor’s correlation with both As and Se, rather than the usual generic secondary sulfate aerosol component, to which coal was likely a major contributor, but was therefore, not separately identifiable in prior analyses. However, although we excluded the secondary PM components in the initial factor analysis model, our subsequent mass regression steps did include the secondary PM mass associated with each source component profile, as well as “other” secondary PM not associated with any of the specific sources

identified. As evident in the source profiles (Fig. 7), secondary aerosols (S, OC, NO₃) contribute a large portion of source-mass identified for the categories. Thus, a complete (i.e., including both primary and secondary aerosols) mass apportionment and source category elemental profile determination was ultimately achieved.

Sensitivity analyses of the apportionment results were also conducted. First, we considered the single key tracer approach for comparison with the APCA results. The single tracer approach results are consistent with the APCA results, but can be less definitive in their interpretation (i.e., single elements can come from multiple sources, whereas a component derived from multiple coincident elements is more source-specific in

interpretation). Further, there is the issue of co-linearity in the single tracer regression model, unlike the APCA method that employs orthogonal rotation that results in orthogonal (uncorrelated) factor scores. An overall U.S. year-round regression model was used to apportion mass, but, as part of sensitivity analysis, regional-seasonal models were also considered. The regional-seasonal models were not found to be significantly different from the overall US model, and in some cases (e.g., California) did not fit the actual $PM_{2.5}$ data as well. Apparently as the number of observations considered declined, the results became less robust. However, with more CSN data for each region with the passing of time, regional source profile and associated apportionment estimates may well prove more reliable and useful in the future.

We also conducted analyses of the site-to-site Coefficient of Variation (CV) of the key tracers and source-related mass impacts estimated in this work in order to investigate the spatial representativeness of the central-site source apportionment estimates provided in this work. This was conducted for all monitoring sites located within 20 miles of another monitor ($n = 53$ unique site pairs nationwide, mean inter-site distance = 9.9 miles). As expected, the secondary aerosols had the least spatial error (i.e., the smallest %CV values). Interestingly, of the source-related impacts, traffic and its key tracer (EC) had the lowest %CV value, indicating that central-site data for traffic $PM_{2.5}$ and its tracers are the most representative of impacts at other locations within a 20-mile radius (i.e., have the lowest spatial error), when compared to other source categories, such as oil and coal combustion, and especially vs. steel and metals-related operations $PM_{2.5}$ and their key tracers (which generally yielded the highest inter-site %CVs).

Numerous source-apportionment studies for individual sites have been conducted in the past decade, also providing useful comparisons with our results. Most of these consider CSN data in their analysis, but usually on a single-site, and never on a national basis. The U.S. EPA Technical Report “Compilation of $PM_{2.5}$ SA Studies” (2003) provides an assessment of over 40 studies conducted nationwide. This EPA Report lists a set of sources across these studies similar to the source categories identified in this work. In addition, the key “signature” tracers associated with each source in this work were also found to be consistent with these identified other studies. While contribution levels from each source can vary from year-to-year, making absolute $PM_{2.5}$ mass contribution comparisons with individual past studies difficult, we have compared our source trace element constituent profiles with other recently published source-apportionment studies’ source profiles for individual cities, finding consistency with the range of past studies’ source profile estimates, as well as with molar mass balance conversions to mass (e.g., for S to ammonium sulfate and NO_3 to ammonium nitrate).

We were also able to compare our results with an analogous past U.S. analysis of 1979–1982 IPN data for 32 MSAs (provided in Supplemental Results to this paper). In that comparison, we found current concentrations of trace elements and source-mass to be significantly lower than past levels, but with spatial correlations between past and current levels. This suggests that, while levels have declined, the relative ranks (i.e., high vs. low) among cities for these major sources have not changed appreciably across the U.S. over recent decades.

This work has provided a novel assessment of the major sources of $PM_{2.5}$ throughout the U.S., including source-mass estimates of the type that can potentially provide useful inputs into epidemiological studies investigating the human health effects of air pollution from specific source categories. However, the multi-site approach taken here to the source apportionment (i.e., analyzing all sites in the U.S. in one factor analysis, rather than each individual site independently) carries with it both strengths and weaknesses.

First and foremost, traditional single-site factor-based apportionment analyses have considered only day-to-day concentration variations that are highly inter-correlated with each other, due to their shared meteorology, limiting the ability of single-site analyses to separate impacts of sources. For example, different sources that are co-located, and therefore impact a single monitor from the same direction on each day, would be lumped together by a single-site analysis. In contrast, the inclusion of more than a single site at once in an apportionment factor analysis model adds the site-to-site variability in impacts to the variance analyzed, which can potentially provide more independent components than possible when considering only a single site. Another advantage of the multi-site approach is that it identifies shared source-related components for those source categories that are of the most overall importance in the group of sites, rather than those that may affect a specific site and not others. This thereby provides estimates of source category-related impacts that allow direct comparability of exposures across the entire spatial area encompassed by the factor analysis dataset. For example, an analysis of all New York City sites at once would identify shared source components that affect all of the sites, aided by differences between sites in day-to-day impacts of the various major sources.

A limitation of this multi-site approach, however, is that it intrinsically assumes a single source profile (i.e., % trace element composition) for each source over the area encompassed by the sites considered, in this case over the entire nation. Before combining spatially disparate sites in such an analysis, there should be an evaluation as to whether this is appropriate, given potential spatial differences in emission characteristics (e.g., in soil composition). To evaluate the extent of this potential limitation in this work, we directly tested whether the profiles varied significantly from region-to-region across the U.S. by also conducting region-specific mass-PCA component regressions, but found that only the soil/crustal component varied significantly across the regions, while the anthropogenic sources did not differ significantly in their makeup. Moreover, this assumption of an overall spatial mean source category composition profile is no different, in practice, from the single-site derived profiles’ assumption that they do not vary over time, which they certainly do (e.g., from summer to winter). It should also be noted, however, that although a single source profile is assumed in an analysis of multiple sites at once, this source profile is data-driven, considering numerous sites in coming up with its composite central estimate of the source–mass relationship, and therefore can be considered as a representing a more robust “average” source profile that is not based on only limited data from a single site (as in most past studies). Thus, balancing these various strengths vs. weaknesses it was decided to use a nationwide multi-site approach for all sources in a consistent way. As time progresses, and more data become available at every site in the nation, it will be possible to more confidently define season and region-specific profiles. This may provide statistically significant improvements over the national estimates provided in this work, although this is yet to be proven, considering the underlying context of day-to-day variability in source composition.

Overall, the major $PM_{2.5}$ source categories (and their key tracers) identified in this nationwide analysis were: Soil (Si, Ca), Coal (Se, As), Oil (Ni, V), Biomass (K), Steel (Fe, Mn), Salt (Cl, Na), Metals (Zn, Pb) and Traffic (EC, NO_2). Among the eight primary sources identified by this work, certain sources (e.g., motor vehicular traffic and soil) were found to be ubiquitous across all 102 MSAs considered in this work, while some source categories (e.g., coal and residual oil) were found to affect certain areas (e.g., Ohio Valley region and northeast U.S.) and other source categories (e.g., Iron/Steel industry) were found to only significantly affect very few sites (e.g., Birmingham, AL and Detroit, MI).

Acknowledgements

Research supported by the Health Effects Institute (HEI) National Particle Component Toxicity (NPACT) Initiative and NYU's National Institute of Environmental Health Sciences (NIEHS) Center Grant (5P30ES000260).

Appendix. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.biochi.2011.04.017.

References

- Baumann, K., Jayanty, R.K., Flanagan, J.B., 2008 Jan. Fine particulate matter source apportionment for the chemical speciation trends network site at Birmingham, Alabama, using positive matrix factorization. *J. Air Waste Manag. Assoc.* 58 (1), 27–44.
- Bell, M.L., Dominici, F., Ebisu, K., Zeger, S.L., Samet, J.M., 2007 Jul. Spatial and temporal variation in PM_{2.5} chemical composition in the United States for health effects studies. *Environ. Health Perspect.* 115 (7), 989–995.
- Blifford, I.H., Meeker, G.O., 1967. A factor analysis model of large scale pollution. *Atmos. Environ.* 1 (2), 147–157.
- Brown, S.G., Frankel, A., Raffuse, S.M., Roberts, P.T., Hafner, H.R., Anderson, D.J., 2007 Jun. Source apportionment of fine particulate matter in Phoenix, AZ, using positive matrix factorization. *J. Air Waste Manag. Assoc.* 57 (6), 741–752.
- Chow, J.C., Watson, J.G., Chen, L.-W.A., Trimble, D.L., Ho, S.S.H., Veropoulos, K., 2008. Methods to Assess Carbonaceous Aerosol Sampling Artifacts for IMPROVE and Other Long-term Networks. Desert Research Institute, Reno, NV.
- Chow, J.C., Watson, J.G., Chen, L.-W.A., Rice, J., Frank, N.H., 2010. Quantification of PM_{2.5} organic carbon sampling artifacts in US networks. *Atmos. Chem. Phys.* 10, 5223–5239.
- Cooper, J.A., Watson Jr., J.G., 1980. Receptor oriented methods of air particulate source apportionment. *JAPCA* 30 (10), 1116–1125.
- Dockery, D.W., Pope 3rd, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris Jr., B.G., Speizer, F.E., 1993 Dec 9. An association between air pollution and mortality in six U.S. cities. *N. Engl. J. Med.* 329 (24), 1753–1759.
- Gildemeister, A.E., Hopke, P.K., Kim, E., 2007. Sources of fine urban particulate matter in Detroit, MI. *Chemosphere* 69 (7), 1064–1074.
- Gordon, G.E., 1980. Receptor models. *Environ. Sci. Technol.* 14, 792–800.
- Harman, H.H., 1976. *Modern Factor Analysis*, third ed. University of Chicago Press, Chicago.
- HEI, 2007. Air Quality Database for Health Effects Studies. <http://hei.aer.com> (accessed on: 05.17.07.).
- Helble, J.J., 2000. A model for the air emissions of trace metallic elements from coal combustors equipped with electrostatic precipitators. *Fuel Process. Technol.* 63, 125–147.
- Henry, R.C., Kim, B.M., 1989. A factor analysis model with explicit physical constraints. *Trans. Air Pollut. Control Assoc.* 14, 214–225.
- Henry, R.C., Lewis, C.W., Hopke, P.K., 1984. Review of receptor model fundamentals. *Atmos. Environ.* 18, 1507–1515.
- Hopke, P.K., Gladney, E.S., Gordon, G.E., Zoller, W.H., Jones, A., 1976. The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. *Atmos. Environ.* 10 (11), 1015–1025.
- Hopke, P.K., Ito, K., Mar, T., Christensen, W.F., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Larson, T.V., Liu, H., Neas, L., Pinto, J., Stolzel, M., Suh, H., Paatero, P., Thurston, G.D., 2006 May. PM source apportionment and health effects: 1. Intercomparison of source apportionment results. *J. Expo. Sci. Environ. Epidemiol.* 16 (3), 275–286.
- Ito, K., Xue, N., Thurston, G.D., 2004. Spatial variation of PM_{2.5} chemical species and source-apportioned mass concentrations in New York City. *Atmos. Environ.* 38, 5269–5282.
- Kim, E., Hopke, P.K., Larson, T.V., Maykut, N.N., Lewtas, J., 2004. Factor analysis of Seattle fine particles. *Aerosol Sci. Technol.* 38 (7), 724–738.
- Kim, E., Hopke, P.K., 2007. Source characterization of ambient fine particles in the Los Angeles Basin. *J. Environ. Eng. Sci.* 6, 343–353.
- Kim, M., Deshpande, S.R., Crist, K.C., 2007. Source apportionment of fine particulate matter (PM_{2.5}) at a rural Ohio River Valley site. *Atmos. Environ.* 41 (39), 9231–9243.
- Kleinman, M.T., Pasternack, B.S., Eisenbud, M., Kneip, T.J., 1980. Identifying and estimating the relative importance of sources of airborne-particulates. *Environ. Sci. Technol.* 14 (Issue 1), 62–65.
- Koutrakis, P., Spengler, J.D., 1987. Source apportionment of ambient particles in Steubenville, OH using specific rotation factor analysis. *Atmos. Environ.* 21, 1511–1519.
- Laden, F., Neas, L.M., Dockery, D.W., Schwartz, J., 2000 Oct. Association of fine particulate matter from different sources with daily mortality in six U.S. cities. *Environ. Health Perspect.* 108 (10), 941–947.
- Lewis, C.W., Norris, G.A., Conner, T.L., Henry, R.C., 2003 Mar. Source apportionment of Phoenix PM_{2.5} aerosol with the Unmix receptor model. *J. Air Waste Manag. Assoc.* 53 (3), 325–338.
- Malm, W.C., Sisler, J.F., Huffman, D., Eldred, R.A., Cahill, T.A., 1994. Spatial and seasonal trends in particle concentration and optical extinction in the United States. *J. Geophys. Res.* 99 (D1), 1347–1370.
- Maykut, N.N., Lewtas, J., Kim, E., Larson, T.V., 2003 Nov 15. Source apportionment of PM_{2.5} at an urban IMPROVE site in Seattle, Washington. *Environ. Sci. Technol.* 37 (22), 5135–5142.
- National Research Council (NRC), 2004. *Research Priorities for Airborne Particulate Matter: IV. Early Research Progress*. National Academy Press, Washington, DC, ISBN 0-309-09199-3.
- Ozkaynak, H., Thurston, G.D., 1987. Associations between 1980 U.S. mortality rates and alternative measures of airborne particle concentration. *Risk Anal.* 7, 449–460.
- Paatero, P., 1997. Least squares formulation of robust, non-negative factor analysis. *Chemometr. Intell. Lab. Syst.* 37, 23–35.
- Paatero, P., 1999. The multilinear engine – a table-driven least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Graph. Stat.* 8, 854–888.
- Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc. (JAMA)* 287 (9), 1132–1141.
- Pope, C.A., Burnett, R., Thurston, G.D., Thun, M., Calle, E., Krewski, D., Godleski, J., 2004 Jan 6. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. *Circulation* 109 (1), 71–77.
- Qin, Y., Kim, E., Hopke, P.K., 2006. The concentrations and sources of PM_{2.5} in metropolitan New York City. *Atmos. Environ.* 40, S312–S332.
- Schauer, J.J., Lough, G.C., Shafer, M.M., Christensen, W.F., Arndt, M.F., DeMinter, J.T., Park, J.S., 2006 Mar. Characterization of metals emitted from motor vehicles. *Res. Rep. Health Eff. Inst.* 133, 1–76.
- Thurston, G.D., Spengler, J.D., 1985. A quantitative assessment of source contributions to inhalable particulate matter in metropolitan Boston, Massachusetts. *Atmos. Environ.* 19, 9–25.
- Thurston, G.D., Ozkaynak, H., and Schatz, A., 1984. A chemical characterization and source apportionment of the IP network fine particle data. In: Paper 84-58.5. Presented at the 77th Annual Meeting of the Air Pollution Control Association, San Francisco, CA, June 24–29, 1984.
- Thurston, G.D., Liou, P.J., 1987. Receptor modeling and aerosol transport. *Atmos. Environ.* 21, 687–698.
- 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 Dec. Workgroup report: workshop on source apportionment of particulate matter health effects – intercomparison of results and implications. *Environ. Health Perspect.* 113 (12), 1768–1774.
- U.S. Energy Information Administration (EIA), 2010. U.S. Coal Supply and Demand 2009 Review. <http://www.eia.doe.gov/coal/page/special/consumption.html> (accessed 8.31.10.).
- U.S. Environmental Protection Agency (EPA), 1999. Particulate Matter (PM_{2.5}) Speciation Guidance Final Draft, first ed. US EPA Monitoring and Quality Assurance Group. Emission, Monitoring, and Analysis Division, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711.
- U.S. Environmental Protection Agency (EPA), 2003. *Compilation of Existing Studies on Source Apportionment for PM_{2.5}* Office of Air Quality Planning and Standards. Contract No. 68-D-02-061 Work Assignment 1-05. Washington, DC.
- U.S. Environmental Protection Agency (EPA), 2005. 2005 National Emissions Inventory. <http://www.epa.gov/ttn/chief/net/2005inventory.html> Washington, DC.
- Watson, J.G., Chow, J.C., Chen, L.-W.A., Frank, N., 2009. Methods to assess carbonaceous aerosol sampling artifacts for IMPROVE and other long-term networks. *J. Air Waste Manag. Assoc.* 59 (8), 898–911.
- World Health Organization (WHO), 2007. *Health Relevance of Particulate Matter from Various Sources: Report on a WHO Workshop*, Bonn, Germany, 26–27 March 2007. World Health Organization Regional Office for Europe, Scherfigsvej 8, DK-2100 Copenhagen Ø, Denmark.
- Wu, C.F., Larson, T.V., Wu, S.Y., Williamson, J., Westberg, H.H., Liu, L.J., 2007. Source apportionment of PM_{2.5} and selected hazardous air pollutants in Seattle. *Sci. Total Environ.* 386, 42–52.