# Understanding the sources and composition of the incremental excess of fine particles across multiple sampling locations in one air shed

Jerome E. McGinnis[1], Jongbae Heo[1], Michael R. Olson[1], Andrew P. Rutter[1], James J. Schauer[1,2,*]

1. University of Wisconsin-Madison, Environmental Chemistry and Technology Program, Madison, WI 53706, USA

2. Wisconsin State Laboratory of Hygiene, Madison, WI 53718, USA

A B S T R A C T

Well-designed health studies and the development of effective regulatory policies need to rely on an understanding of the incremental differences in particulate matter concentrations and their sources. Although only a limited number of studies have been conducted to examine spatial differences in sources to particulate matter within an air shed, routine monitoring data can be used to better understand these differences. Measurements from the US EPA Chemical Speciation Network (CSN) collected between 2002–2008 were analyzed to demonstrate the utility of regulatory data across three sites located within 100 km of each other. Trends in concentrations, source contribution, and incremental excesses across three sites were investigated using the Positive Matrix Factorization model. Similar yearly trends in chemical composition were observed across all sites, however, excesses of organic matter and elemental carbon were observed in the urban center that originated from local emissions of mobile sources and biomass burning. Secondary sulfate and secondary nitrate constituted over half of the PM$_{2.5}$ with no spatial differences observed across sites. For these components, the excess of emissions from industrial sources could be directly quantified. This study demonstrates that CSN data from multiple sites can be successfully used to derive consistent source profiles and source contributions for regional pollution, and that CSN data can be used to quantify incremental differences in source contributions of across these sites. The analysis strategy can be used in other regions of the world to take advantage of existing ambient particulate matter monitoring data to better the understanding of spatial differences in source contributions within a given air shed.

## Introduction

Exposure to ambient fine particulate matter with aerodynamic diameters less than 2.5 μm (PM$_{2.5}$) has been shown to cause adverse health effects (Laden et al., 2000; Pope et al. 2009), such as respiratory and cardiovascular disease (Dominici et al., 2006; Peng et al., 2009; Samet et al., 2000). PM$_{2.5}$ is a chemically complex pollutant emitted directly from a diverse range of sources, and formed from precursor gases by photochemical oxidation reactions. The composition of PM$_{2.5}$ varies spatially and temporally and

understanding this variability is a powerful tool regulation agencies can use to mitigate negative outcomes associated with exposure to PM.

The concentrations and composition of PM$_{2.5}$ vary greatly across local and regional scales. Although many countries are implementing mitigation techniques in an attempt to lower PM$_{2.5}$ concentrations, the levels are still high in many areas. The majority of world's population lives in urban areas, and this portion continues to grow (Bremner et al., 2009). This increase in urban population may lead to larger potential health problem related to PM$_{2.5}$ exposure. There have been a large number of studies elucidating physico-chemical characteristics and sources of PM$_{2.5}$ in major urban areas, including Pittsburgh Air

Quality Study (Pekney et al., 2006a), Southern California Air Quality Study (SCAQMD, 2008), Mexico City (Aiken et al., 2009) and Campaign of Air Quality Research in Beijing (van Pinxteren et al., 2009; Yue et al., 2009). These studies were performed using a limited number of fixed sites and uncertainties remain in the spatial variability of chemical components and sources of $PM_{2.5}$.

The concentration of chemical species is generally greater in urban areas than the surrounding rural areas due to local source contributions. Regional sources tend to lead to fairly consistent concentration across multiple locations due to mixture during transport (Snyder et al., 2010). Many polluted regions around world are setting up similar monitoring networks, but the data from these networks are not fully analyzed and are largely used for looking at short term trends and annual averages. The United State Environmental Protection Agency (US EPA) started operating the Chemical Speciation Network (CSN), formerly called the Speciation Trends Network (STN), in 2001 to collect $PM_{2.5}$ compositional data at various locations throughout the United States. The data acquired from the various sampling sites can be analyzed to determine sources, trends, and define the difference between urban and rural sites. The Positive Matrix Factorization (PMF) has been an analysis method used in many studies to determine the source contribution of chemical constituents of ambient PM (Heo et al., 2013; Ito et al., 2004; Kim et al., 2005; Pekney et al., 2006b; Zhao and Hopke, 2006; Zhou et al., 2009). The PMF model uses the unique composition of $PM_{2.5}$ that is associated with various $PM_{2.5}$ source contributions. There have been a few studies that have combined multiple advanced receptor models such as Unmix and PMF to understand the sources of $PM_{2.5}$ (Lewis et al., 2003; Yuan et al., 2006). These studies have used sampling data from within urban areas, but rural surrounding areas were not factored in.

By using the CSN data along with the PMF model yearly trends were determined as well as the sources of $PM_{2.5}$. Comparing the sources across various locations can reveal whether sources are regional or local. This is a valuable tool to apply when developing $PM_{2.5}$ mitigation strategies.

# 1 Materials and methods

## 1.1 Chemical speciation network data

This study uses data from three sampling location located within 100 km of each other in Southeast Wisconsin in the Great Lakes Region of North America. The State of Wisconsin has a population of approximately 5.7 million, and the population of Southeast Wisconsin is approximately 1.5 million (Metropolitan Milwaukee Association of Commerce, 2012). Three sampling sites were investigated in the study: Milwaukee, WI (43.061111°N, 87.912500°W);

Mayville, WI (43.435000°N, 88.527778°W); and Waukesha, WI (43.003889°N, 88.231944°W), which all have sampling sites that are part of the US EPA CSN network (http://www.epa.gov/ttn/airs/airsaqs). Waukesha is approximately 27 km west of Milwaukee, and Mayville is approximately 54 km north of Waukesha and 65 km northeast of Milwaukee. Milwaukee and Waukesha Counties are both considered part of the metropolitan area, while Mayville, in Dodge County is a rural background location in greater Milwaukee (MMAC, 2012). A detailed map of the site locations can be seen in the supplemental materials (**Fig. S1**).

The CSN measurements were made every third day in Milwaukee and Mayville, and every sixth day in Waukesha. The data is made publically available through the US EPA's Air Quality Systems (AQS) (http://www.epa.gov/ttn/airs/airsaqs). The chemical speciation data from the networks includes elemental carbon, organic carbon, nitrate, sulfate, ammonium, sodium and potassium ions, and 48 elements. The CSN database also provides uncertainties and quality control flags on the dates the concentration was measured. X-Ray Fluorescence (XRF) Spectrometry is used on Teflon filters to determine the concentrations of 48 elements. Ion Chromatography is used on nylon filters for determining the concentrations of sulfate, nitrate, ammonium, sodium and potassium ions. The raw CSN data was not blank subtracted, but the CSN database reported available blank results. The average of the field blanks from the Milwaukee, Mayville, and Waukesha was found for each chemical species. The average field blank concentration over five years was then subtracted from each of the species concentration values. A description of the analytical methods and sampler configurations was given by Solomon et al. (2000).

Quality assurance/quality control was performed to validate data that was used in the analysis. Data was flagged on dates when an error with a monitoring device was reported including a machine malfunction, power failure, invalid flow rate, etc (https://aqs.epa.gov/aqsweb/codes/data/QualifierCodes.html). All flagged dates for specific locations were removed, with the exception of the flag termed "outlier". The "outlier" flag means that the data is significantly different from other dates but does not necessarily mean a device malfunction. After all the invalid data was removed, blank subtraction was performed using the average and standard deviation of the blanks from the AQS data. The method detection limit (MDL) was calculated using three times the standard deviation of the blanks. The data points above the MDLs were identified for each element at each location and are summarized in **Table S1**. The percent of data points above the MDL was calculated and used to determine what compounds and elements should be looked into further for complete analysis. Chemical constituents undetected in more than 90% of total collected samples