APPORTIONMENT OF PM$_{2.5}$ AND AIR TOXICS IN DETROIT, MICHIGAN

FINAL REPORT
STI-906201.06-3103-FR

By:
Juli I. Rubin
Steven G. Brown
Katherine S. Wade
Hilary R. Hafner
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169

Prepared for:
Dennis Doll
Ellen Baldridge
U.S. Environmental Protection Agency
USEPA Mailroom, Mail Stop C304-04
Research Triangle Park, NC 27711

December 29, 2006

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Source apportionment of speciated $PM_{2.5}$ and air toxics data using positive matrix factorization (PMF) methods was applied to three Speciation Trends Network (STN) sites in Detroit, Michigan. The goals of the analysis were threefold: (1) to determine the effects of using recently updated STN uncertainties in PMF, (2) to determine the effects of using SANDWICH data on PMF, and (3) to explore PMF applications to combined $PM_{2.5}$ and air toxics data sets. The three selected STN sites—Luna Pier, Allen Park, and Dearborn—had 220, 565, and 190 samples dating from May 2002 to December 2005, December 2000 to December 2005, and May 2002 to December 2005, respectively. PMF was applied at each site with (1) STN data and recently updated uncertainties, (2) SANDWICH data, and (3) a combination of STN data and air toxics data. Additional data analysis techniques were applied to both STN and SANDWICH data sets including analysis of ambient data composition, yearly and seasonal trends, and species correlations to help evaluate PMF results.

The majority of the $PM_{2.5}$ mass was apportioned to ammonium sulfate, ammonium nitrate, and mobile sources at all three sites. The PMF results using recently updated uncertainties were compared with previous PMF efforts at Allen Park; results were similar across studies with the exception that previous PMF studies at Allen Park were able to separate out a diesel component from the mobile source factor. Differences are likely due to the larger relative uncertainties applied to metal species (i.e., the new uncertainties) without any changes to the uncertainties applied to the carbon species in the current study.

PMF runs conducted using STN and SANDWICH data sets produced similar factors with similar mass apportionment, on average. Major differences included larger mass apportioned to ammonium sulfate at the expense of ammonium nitrate when using the SANDWICH data. These results are as expected because the SANDWICH data set has higher sulfate concentrations and lower nitrate concentrations than the STN data set.

PMF runs conducted with combined $PM_{2.5}$ and air toxics data did not provide additional insight into mobile or other source contributions. This is likely due to the limited amount of collocated $PM_{2.5}$ and air toxics data, making it difficult to produce meaningful results; with a larger data set results may be improved.

1. INTRODUCTION

The U.S. Environmental Protection Agency (EPA) is interested in analyzing the relationships among pollutants and their precursor species. Increased awareness of interactions among secondary pollutants and the need to identify common culpable sources to assist with developing cost effective and efficient control strategies make it increasingly important that viable tools be available for attributing pollution sources to ambient measurement data.

Air pollution is composed of chemical species originating from natural and manmade emissions that can be transported from their original source areas. Sources typically emit a variety of pollutants, so efficient emission control strategies are needed to address multiple pollutants to bring air pollution concentrations below mandated health standards (e.g., 15 $\mu g/m^3$ annual average for $PM_{2.5}$ and 0.080 ppm for 8-hr ozone). These multiple pollutant control strategies depend on the ability to determine the relationships between emissions sources and elevated levels of air pollution observed at ambient monitoring sites. Source attribution needs to be performed across the range of pollutants to identify common sources among pollutants. Findings will support control strategy development.

The purpose of this work assignment (WA) was to perform source apportionment of multiple pollutants in selected cities using positive matrix factorization (PMF). This WA builds on past source apportionment efforts applied to speciated $PM_{2.5}$ data alone. Results are intended to help the EPA better understand the common sources of $PM_{2.5}$, ozone precursors, and air toxics. Results will also be useful to other ongoing investigations aimed at understanding the links among pollutants.

2. DATA AND METHODS

In this WA, PM$_{2.5}$ data from the Speciation Trends Network (STN) for sites in Detroit, Michigan, were combined with collocated air toxics data. This section summarizes the available data, uncertainty values and their quantification, the species selected for use in PMF, an overview of PMF, and discussion of a post-analysis technique called Potential Source Contribution Function (PSCF).

2.1   STN PM$_{2.5}$ DATA

2.1.1   Sampling and Analysis Details

Integrated 24-hr PM$_{2.5}$ samples were collected as part of the STN at three sites in the Detroit, Michigan, area: Allen Park (December 2000-December 2005, 565 samples, 1-in-3 day); Dearborn (May 2002-December 2005, 190, 1-in-6 day); and Luna Pier (May 2002-December 2005, 220 samples, 1-in-6 day). A map of the area is shown in **Figure 2-1**. Samples were collected by the STN using the MetOne Spiral Aerosol Speciation Samplers (SASS) and were taken on a 1-in-3 or 1-in-6 day schedule. Field blanks were collected for 1-in-10 routine samples, and trip blanks were collected for 1-in-30 routine samples (Research Triangle Institute, 2004). Blank correction for organic carbon (OC) is important for STN data (Kim et al., 2005c; Subramanian et al., 2004; Rice, 2004) and the average blank value over all blanks at each site was used to blank correct the OC concentrations. Only small seasonal variations in the blank values were observed, and with a small set of blanks, applying a seasonal blank correction would introduce additional bias and artificial trends in the ambient data; thus, only the average over the entire period was used. **Figure 2-2** shows the trends in OC blanks by site; the values used were 0.99, 1.32, and 1.16 μg/m^3 for Allen Park, Dearborn, and Luna Pier, respectively.

STN PM$_{2.5}$ samples were collected on Teflon, nylon, and quartz filters. The Teflon filter was used for mass concentrations and analyzed via x-ray fluorescence (XRF) for the elements. The nylon filter was analyzed for the ions sulfate, nitrate, ammonium, sodium, and potassium via ion chromatography (IC). The quartz filter was analyzed by the Research Triangle Institute (RTI) via the National Institute for Occupational Safety and Health/Thermal Optical Transmittance (NIOSH/TOT) protocol (NIOSH, 1999; Birch and Carey, 1996) for OC and elemental carbon (EC).

Figure 2-1. Detroit, Michigan, STN monitoring sites (2000 through 2005).
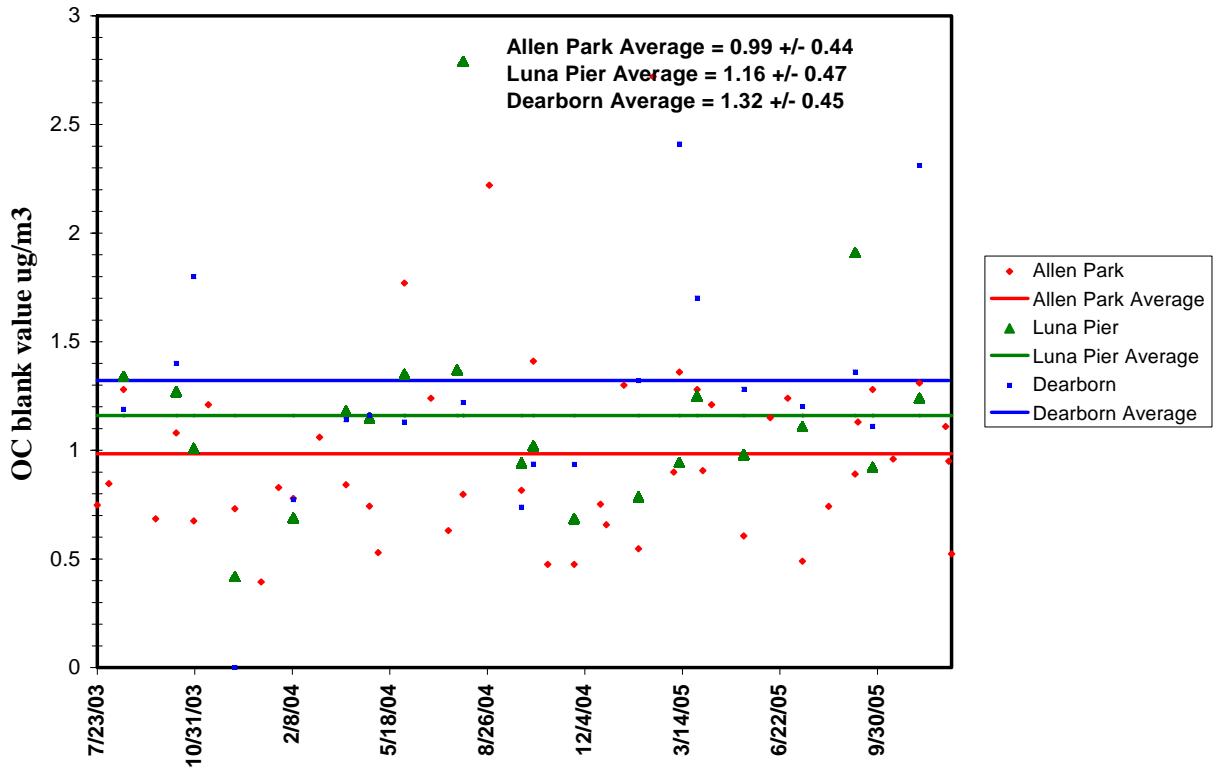
Figure 2-2. Time series of OC filter blank values at the three Detroit-area STN sites (Allen Park, Luna Pier, and Dearborn) and the average values used for blank correction.

2.1.2 Measurement Uncertainties

In addition to concentrations, analysts need to understand the associated uncertainties. Uncertainties reported by RTI for speciated PM$_{2.5}$ data are currently being updated to ensure consistency among the estimation methods used by the laboratories in STN. At present, the EPA's Air Quality System (AQS) only reports updated uncertainties for data measured between July 2003 and October 2005. It was, therefore, necessary to use the recently updated uncertainties to extrapolate uncertainties for the remaining samples without updated values.

For species with updated uncertainties, relationships were examined between concentration and uncertainty. Measurements from the three selected Detroit sites (Allen Park, Dearborn, and Luna Pier) were combined because relationships between concentration and uncertainty were found to overlap. The expected results of such plots are that a constant uncertainty for concentrations at or below the detection limit will be observed because the uncertainty is absolute. Above the detection limit, it is expected that an increase in uncertainty with concentration will be found because a relative component of the uncertainty is introduced. These patterns were observed for species such as potassium (**Figure 2-3**); however, some of the speciated metals did not follow the expected trend. No consistent relationship was found for selected speciated metals including titanium, vanadium, tin, and selenium (e.g., tin is shown in

Figure 2-4). Additionally, some metal species such as manganese, arsenic, lead, and bromine appeared to have two separate trend lines (e.g., sodium ion is shown in **Figure 2-5**). For species that showed two trends, the data were examined to determine if the trends were caused by different measurements, time periods, or other patterns. No identifiable patterns were found. In discussions with RTI, it was noted that samples were analyzed by different XRF instruments with different uncertainty relationships. However, the data are reported to AQS with one method code; AQS does not have identifiers at the instrument level.
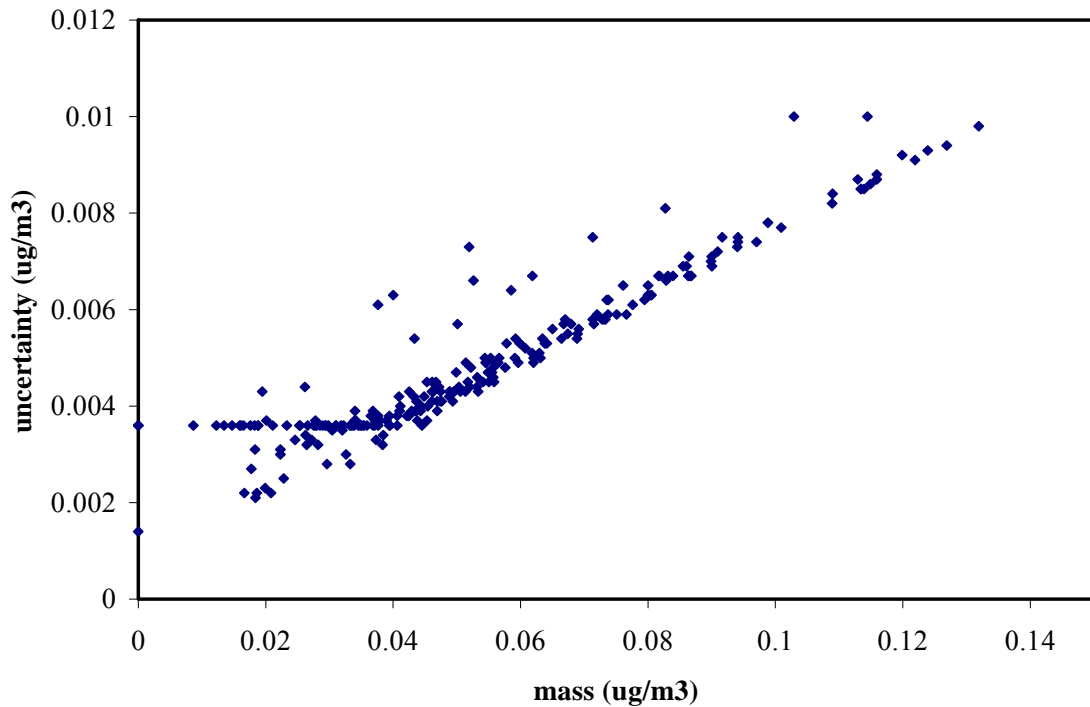
Figure 2-3. Potassium concentration versus uncertainty at all three Detroit STN sites (July 2003 to October 2005).
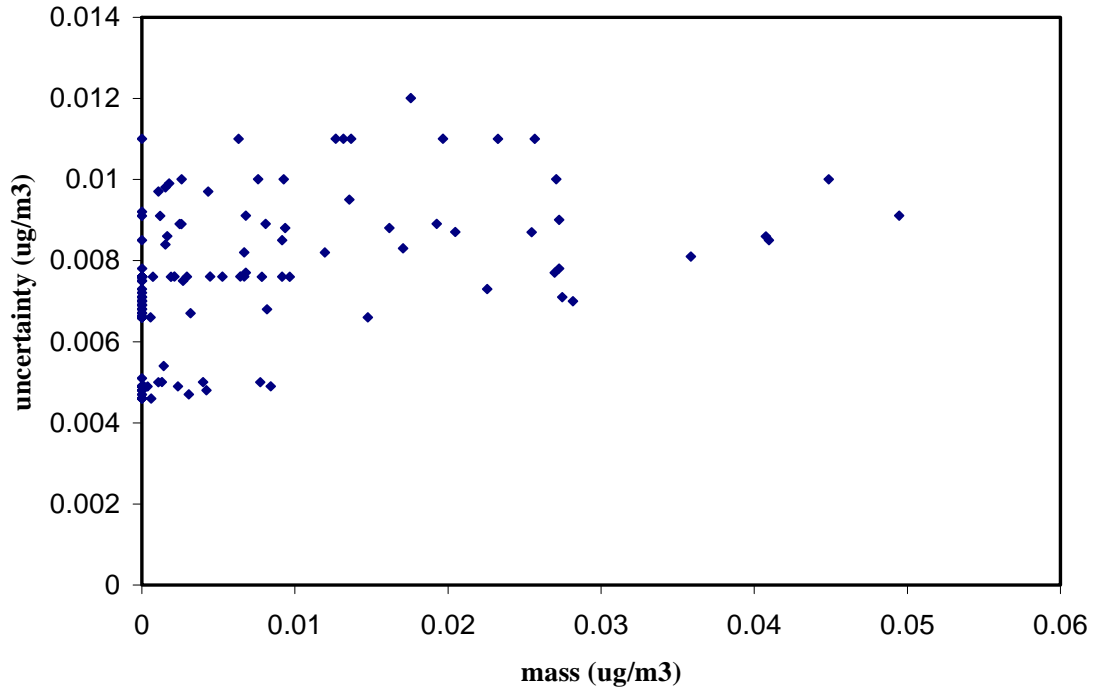
Figure 2-4. Tin concentration versus uncertainty at all three Detroit STN sites (July 2003 to October 2005).
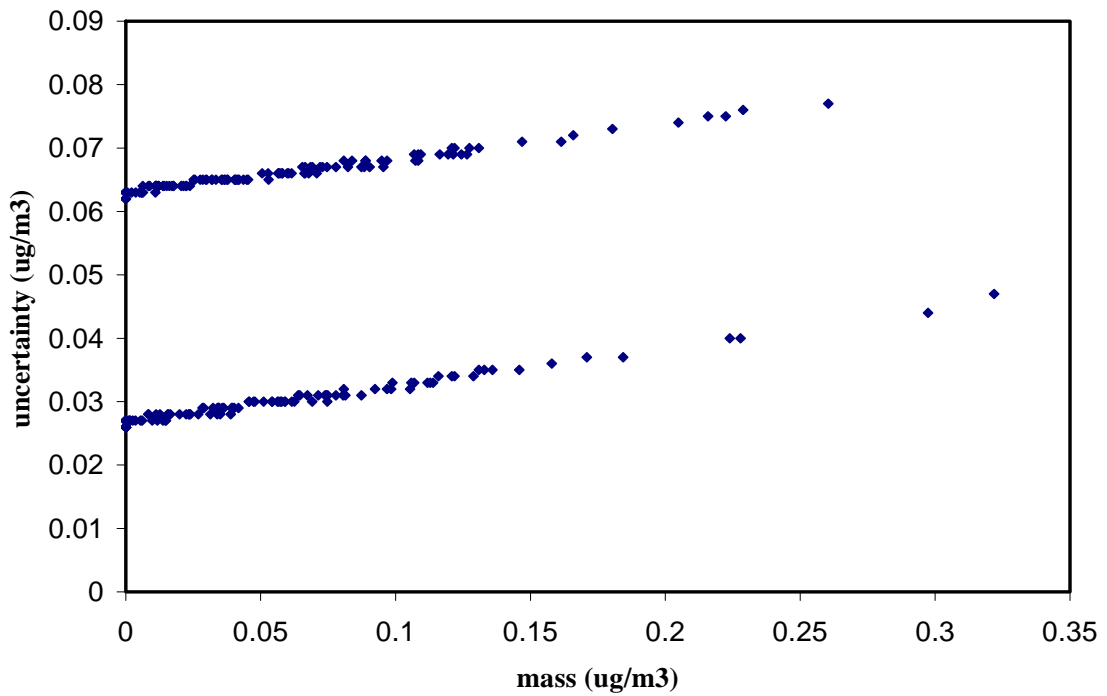
Figure 2-5. Sodium ion concentration versus uncertainty at all three Detroit STN sites (July 2003 to October 2005).

Different methods for estimating uncertainties were applied to address the different relationships found for the speciated data. For species that followed the expected curve, as shown for potassium (Figure 2-3), uncertainties were assumed to be constant below a determined cutoff point. Above the cutoff point, a linear regression was used to relate uncertainty to concentration. For the remaining species, the uncertainty was set to the maximum reported uncertainty (excluding outliers) because it is always best to assume a larger uncertainty for source apportionment purposes (Hafner, 2005). A species and uncertainty method summary is provided in **Table 2-1**.

The uncertainty development methods described could not be used for the carbon and ion species because the uncertainties for these species were not updated in AQS; only species measured by XRF were updated. As a result, the root median squared percent error (RMSPE) (Equation 2-1) was used to estimate the uncertainties using collocated measurements. This method is similar to the more common root mean squared error (RMSE) with two exceptions: (1) the error is a percent of the original value instead of an absolute error, and (2) the median of the squared errors is used instead of the mean. Using the percent enables the output to be applied directly to the concentrations to determine sample-specific uncertainties. Using the median excludes outlying values that heavily influence an RMSE but are not considered useful for PMF.

$$RMSPE = \sqrt{median\left(\left[\frac{2(x_1 - x_2)}{(x_1 + x_2)}\right]^2\right)}$$
(2-1)

where x and y are collocated measurements.

2.1.3   Data Used for Analysis and Source Apportionment

Data from the STN program are routinely screened and validated before being made publicly available. Additional quality control (QC) checks were performed on the data prior to source apportionment, including comparison of reconstructed fine mass to measured mass and comparison of XRF sulfur to IC sulfate. Approximately 10% of samples from each site did not pass these checks and were excluded from the PMF analysis. Those samples for which all species were missing were also excluded from the analysis.

Table 2-1. Summary of uncertainty methods by species and values used to estimate uncertainties. Values include maximum value reported (excluding outliers), the regression cutoff point, pre-cutoff uncertainty, slope and intercept for calculating post-cutoff uncertainties, and RMSPE values.

Species	Uncertainty Method	Maximum Value ($\mu g/m^3$)	Regression Cutoff ($\mu g/m^3$)	Precutoff Uncertainty ($\mu g/m^3$)	Slope	Intercept ($\mu g/m^3$)	RMSPE (%)
Aluminum PM$_{2.5}$	max value	0.0160	--	--	--	--	--
Antimony PM$_{2.5}$	max value	0.0325	--	--	--	--	--
Arsenic PM$_{2.5}$	max value	0.0028	--	--	--	--	--
Barium PM$_{2.5}$	max value	0.0733	--	--	--	--	--
Bromine PM$_{2.5}$	max value	0.0017	--	--	--	--	--
Cadmium PM$_{2.5}$	max value	0.0142	--	--	--	--	--
Cerium PM$_{2.5}$	max value	0.1083	--	--	--	--	--
Lead PM$_{2.5}$	max value	0.0044	--	--	--	--	--
Magnesium PM$_{2.5}$	max value	0.0375	--	--	--	--	--
Manganese PM$_{2.5}$	max value	0.0017	--	--	--	--	--
Samarium PM$_{2.5}$	max value	0.0077	--	--	--	--	--
Selenium PM$_{2.5}$	max value	0.0022	--	--	--	--	--
Strontium PM$_{2.5}$	max value	0.0032	--	--	--	--	--
Terbium PM$_{2.5}$	max value	0.0092	--	--	--	--	--
Tin PM$_{2.5}$	max value	0.0120	--	--	--	--	--
Titanium PM$_{2.5}$	max value	0.0026	--	--	--	--	--
Tungsten PM$_{2.5}$	max value	0.0183	--	--	--	--	--
Vanadium PM$_{2.5}$	max value	0.0018	--	--	--	--	--
Calcium PM$_{2.5}$	trend lines	--	0.0300	0.0026	0.0684	0.0009	--
Chlorine PM$_{2.5}$	trend lines	--	0.0060	0.0036	0.0555	0.0043	--
Chromium PM$_{2.5}$	trend lines	--	0.0040	0.0008	0.0629	0.0006	--
Copper PM$_{2.5}$	trend lines	--	0.0080	0.0008	0.0652	0.0004	--
Iron PM$_{2.5}$	trend lines	--	0.0066	0.0009	0.0690	0.0005	--
Nickel PM$_{2.5}$	trend lines	--	0.0045	0.0006	0.1699	-0.0003	--
Potassium PM$_{2.5}$	trend lines	--	0.0400	0.0036	0.0695	0.0010	--
Silicon PM$_{2.5}$	trend lines	--	0.0600	0.0030	0.0791	0.0021	--
Sulfur PM$_{2.5}$	trend lines	--	0	0	0.0711	0.0009	--
Zinc PM$_{2.5}$	trend lines	--	0.0095	0.0009	0.0691	0.0003	--
Ammonium Ion PM$_{2.5}$	RMSPE	--	--	--	--	--	7
EC PM$_{2.5}$	RMSPE	--	--	--	--	--	10
OC PM$_{2.5}$	RMSPE	--	--	--	--	--	33
Potassium Ion PM$_{2.5}$	RMSPE	--	--	--	--	--	0.7
Sulfate PM$_{2.5}$	RMSPE	--	--	--	--	--	10
Total Nitrate PM$_{2.5}$	RMSPE	--	--	--	--	--	8

RMSPE = root median squared percent error

To determine which species would be included as variables in the PMF analysis, two criteria were used: signal-to-noise ratio (SNR) and percent below detection limit (BDL). SNR can be used as a criterion to determine the "strong", "weak", and "bad" variables (Paatero et al., 2003). SNR is automatically calculated by EPA PMF 1.1 as

$$SNR_j = 1/2 \sqrt{\frac{\sum_{i=1}^{n} x_{ij}^2}{\sum_{i=1}^{n} s_{ij}^2}} \qquad \text{(2-2)}$$

where x_{ij} is the concentration for species j on day i and s_{ij} is the uncertainty. In general, the species j is defined to be "strong" if SNR > 2, "weak" if $0.2 \leq$ SNR ≤ 2, and "bad" if SNR < 0.2 (Paatero, Hopke 2003).

Table 2-2 summarizes the $PM_{2.5}$ measurements with the SNR and percent BDL values for the species with at least 10% of the data above detection for each site. In general, species having more than 70% BDL were discarded. Weak variables that have less than 70% BDL were included in the analysis, but those variables were down-weighted by a factor of three in PMF calculations. No sodium or chlorine data were used in this analysis because the impact of sea and road salt on $PM_{2.5}$ and air toxics was expected to be small and confidence (or SNR) in the sodium and chlorine data was often low. Because potassium ion (K^+) was mostly BDL (over 70% of the time), elemental potassium, which was mostly above detection, was used.

2.2 SANDWICH DATA

While STN measures $PM_{2.5}$ mass and the species that comprise the mass, the measurements are often slightly different than the Federal Reference Method (FRM) $PM_{2.5}$ mass measurements, which are the metric for regulations. To translate the STN measurements into "FRM equivalent" measurements, the Sulfate, Adjusted Nitrate, Derived Water, Inferred Carbonaceous mass and estimated aerosol acidity (H+) material balance approach (SANDWICH) was developed (Frank, 2006). This method assumes $PM_{2.5}$ on the filter is broken down as follows:

$$PM_{2.5} = NO_3(FRM) + SO_4 + NH_4 + H_2O + Crustal + TCM + Blank + Other \qquad \text{(2-3)}$$

The measurement of total carbonaceous material (EC and OC) has a higher analytical uncertainty than the other components of $PM_{2.5}$; therefore, the SANDWICH method uses the other components to calculate EC and OC. This method should eliminate blank corrections and artifacts on the filters. In addition, $NO_{3(FRM)}$ (nitrate measured by the FRM) is the retained NO_3 predicted using $NO_{3(STN)}$ (nitrate measured by the STN) and temperature and relative humidity. The result of these adjustments is usually a higher sulfate mass and a lower nitrate mass overall.

To compare analyses based on the STN measurements and ensure that results are applicable in a regulatory sense, SANDWICH data were also examined as part of this project. This is the first known application of PMF to SANDWICH data.

Table 2-2. SNRs and percent below detection for species measured at three STN sites (Allen Park, Dearborn, and Luna Pier). For SNR, values in bold are considered "strong"; the remaining values are considered "weak". Percent below detection values in bold have greater than 70% BDL.

Species	Signal-to-Noise Ratio (SNR)			% Below Detection Limit (BDL)		
	Allen Park	Dearborn	Luna Pier	Allen Park	Dearborn	Luna Pier
Arsenic PM$_{2.5}$	--	0.67	0.49	--	33.58	67.73
Aluminum PM$_{2.5}$	0.83	**2.32**	0.94	**71.63**	21.02	65.91
Antimony PM$_{2.5}$	--	0.28	--	--	**90.19**	--
Bromine PM$_{2.5}$	0.97	1.14	0.90	44.17	21.25	39.09
Calcium PM$_{2.5}$	**3.47**	**2.71**	**5.72**	1.44	7.72	3.18
Cadmium PM$_{2.5}$	--	0.29	--	--	55.23	--
Cerium PM$_{2.5}$	--	0.27	--	--	22.55	--
Chromium PM$_{2.5}$	1.57	1.84	1.44	68.94	14.53	67.27
Copper PM$_{2.5}$	**3.33**	1.58	1.59	28.01	11.12	54.09
Chlorine PM$_{2.5}$	**3.81**	**3.83**	1.50	64.99	9.06	67.27
Iron PM$_{2.5}$	**3.65**	**2.58**	**6.75**	0.72	4.50	0.91
Lead PM$_{2.5}$	0.59	1.42	0.48	68.94	18.09	61.82
Magnesium PM$_{2.5}$	--	1.23	--	--	22.42	--
Manganese PM$_{2.5}$	1.16	**3.93**	0.63	45.60	6.72	56.82
Nickel PM$_{2.5}$	0.98	1.01	0.86	**71.81**	19.03	**70.91**
Samarium PM$_{2.5}$	--	0.70	--	--	18.59	--
Selenium PM$_{2.5}$	0.50	0.71	0.86	**83.30**	30.34	67.73
Strontium PM$_{2.5}$	--	0.51	--	--	30.15	--
Terbium PM$_{2.5}$	--	0.77	--	--	50.99	--
Tin PM$_{2.5}$	0.36	0.32	--	**93.90**	64.16	--
Titanium PM$_{2.5}$	0.88	1.59	0.64	52.42	19.26	67.27
Tungsten PM$_{2.5}$	--	0.25	--	--	**77.89**	--
Vanadium PM$_{2.5}$	0.51	0.78	--	**84.56**	26.24	--
Silicon PM$_{2.5}$	**2.85**	**2.32**	**3.93**	10.23	9.40	15.45
Zinc PM$_{2.5}$	**4.07**	**3.19**	**5.55**	2.51	7.34	4.09
Sulfur PM$_{2.5}$	**4.36**	**2.52**	**6.98**	1.08	7.38	0.45
Potassium PM$_{2.5}$	1.46	**2.41**	1.13	1.62	7.59	2.27
Sodium PM$_{2.5}$	--	--	--	--	--	**75.00**
Ammonium Ion PM$_{2.5}$	**2.45**	**16.65**	**2.98**	1.62	3.10	0.45
Sodium Ion PM$_{2.5}$	**2.43**	1.18	1.15	16.34	41.24	18.18
Potassium Ion PM$_{2.5}$	1.14	**3.01**	1.59	57.81	10.54	63.64
OC PM$_{2.5}$	1.66	**5.16**	1.60	1.26	10.08	0.00
Total Nitrate PM$_{2.5}$	**3.00**	**12.50**	**3.64**	0.90	4.08	0.00
EC PM$_{2.5}$	1.54	**4.56**	1.71	3.95	11.05	13.64
Sulfate PM$_{2.5}$	**2.86**	**16.66**	**3.62**	0.18	3.10	0.00

2.3   PMF

2.3.1   Workings of PMF

PMF, described in detail elsewhere (Paatero, 1997; Paatero and Tapper, 1994) and briefly covered here, was used in this WA. PMF is an advanced multivariate receptor modeling technique that calculates site-specific source profiles with time variations of these sources based on correlations imbedded in ambient data. PMF has been successfully applied to PM data, air toxics data, and volatile organic compound (VOC) data in several studies (Anttila et al., 1995; Begum et al., 2005; Buzcu et al., 2003; Juntto and Paatero, 1994; Kim et al., 2003; Kim et al., 2004a; Kim et al., 2004b; Kim and Hopke, 2004a, b; Kim et al., 2005a; Kim et al., 2005b; Larsen and Baker, 2003; Lee et al., 1999; Poirot et al., 2001; Polissar et al., 2001; Ramadan et al., 2000; Zhou et al., 2004).

Given a data matrix X consisting of the concentration measurements of n chemical species in m samples and their corresponding uncertainties, the objective of PMF is to determine the number of factors p, the chemical composition profile of each factor, and the contribution of each factor to each sample. PMF factorizes the data matrix X into two matrices according to

$$X_{(m\ by\ n\ )}= G_{(m\ by\ p)}\ F_{(p\ by\ n)} + E_{(m\ by\ n)} \tag{2-4}$$

where G represents the contribution of each factor to each ambient sample, and describes the time variations of the factors because each ambient sample is an observation at different times. F is a matrix of chemical composition profiles of each factor. F and G are both forced to be non-negative in order to make physical sense (i.e., factors cannot have negative species concentrations and ambient samples cannot have a negative factor contribution). E is an m by n residual matrix of random errors. The elements of the residual matrix, e_{ij}, can be defined as

$$e_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj} \tag{2-5}$$

where $i = 1,\ldots,m$; $j = 1,\ldots,n$). In PMF, the sum of the squares of residuals, e_{ij}, weighted inversely by the variation of the data points, s_{ij}^2, is minimized according to the following constrained weighted least-squares model:

$$\text{minimize}\, Q = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{e_{ij}^2}{s_{ij}^2} \tag{2-6}$$

The objective is to determine the matrices G and F by minimizing Q. Equation 2-6 is solved using a unique iterative algorithm in which matrices G and F vary simultaneously at each iteration step (Paatero, 1997). Theoretically, if the uncertainties correctly characterize the data and every point is perfectly modeled, Q should be approximately the number of species multiplied by the number of observations, minus the number of factors multiplied by the number of species (i.e., the number of data points). In these analyses, Q from the modeling was required to be within 50% of the theoretical Q to ensure a reasonable fit for all observations.

The EPA's Office of Research and Development has recently developed a standalone version of PMF (EPA PMF) that has been freely distributed to the air quality management community (Eberly, 2005). EPA PMF version 1.1 is a graphical user interface that has been developed based on the PMF model and solved using the multi-linear engine as implemented in the program ME-2 (Paatero, 1999). EPA PMF operates in a robust mode, meaning that "outliers" are not allowed to overly influence the fitting of the contributions and profiles.

2.3.2 Final Data Set Development

In PMF, each data point is weighed individually, allowing the user to adjust the influence of each point depending on the confidence in the data. This feature is an advantage of PMF because samples with some species missing or below the minimum detection limit (MDL) can be used in the analysis, with associated uncertainty adjusted so that these data points are given less weight in the model solution. Data below MDL were substituted with the maximum MDL reported for the given species divided by two and missing data were substituted with median concentrations (Poirot et al., 2001; Polissar et al., 2001; Song et al., 2001). The maximum MDL was used for substitution because use of the sample-specific MDL could introduce a false source of variability in the data. Uncertainties for values below MDL were calculated as 5/6*max MDL and for missing values as 4*median concentration. For samples above detection, updated uncertainties were used, if reported, and the remaining uncertainties were estimated based on methods discussed in Section 2.1.2. Some species were given less or more weight by increasing or decreasing their uncertainty, which resulted in better modeling of individual species and the total mass.

2.3.3 Using PMF Output

Source contributions of each factor can be determined using PMF output and matrices G and F. Each element of the matrix G (g_{ik}) is a normalized source contribution of a factor k for a given sample i. Each element of F (f_{kj}) is a mass contribution of a chemical species j to a factor k. Using matrix F, the reconstructed mass (m_k) for an individual factor i can be calculated as

$$m_k = \sum_{j=1}^{n} f_{kj} \qquad \qquad \textbf{(2-7)}$$

The normalized source contributions of matrix G can then be converted to meaningful mass units by multiplying the individual elements of G and the corresponding factor mass m_k.

$$c_{ik} = g_{ik} * m_k \qquad \qquad \textbf{(2-8)}$$

Uncertainties in the EPA PMF solution are estimated using a bootstrapping technique, which is a re-sampling method in which "new" data sets are generated that are consistent with original data, each data set is decomposed into F and G matrices, and the resulting F and G matrices are compared with the base run (Eberly, 2005). Instead of inspecting point estimates, this method allows the analyst to review the confidence intervals for each species to obtain more

robust profiles. Output of the bootstrapping analysis consists of box whisker plots of species for each profile by both percent of species and concentration. The box shows where the middle 50% of the bootstrap values exist—the tighter the box, the more certainty in the profile, and the more consistent the results are across the bootstraps. In this study, 200 bootstrap runs were performed for the final analysis results.

2.4   WIND AND POTENTIAL SOURCE CONTRIBUTION FUNCTION ANALYSIS

A Potential Source Contribution Function (PSCF) (Draxler and Hess, 1997) was applied to help interpret the PMF results. The transport patterns on days with the highest 10% concentration of a given factor were compared with the climatological transport patterns. This comparison highlights the differences in transport and areas of influence between the general transport pattern (i.e., the climatology) and high concentration days of a given factor. Using the National Oceanic and Atmospheric Administration (NOAA) Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model, (Draxler and Hess, 1997) 72-hr backward trajectories were run for all sample dates at three ending heights (100, 300, and 1,000 meters). Ensemble backward trajectories were run every six hours to account for variability over a 24-hr sampling period. If a trajectory endpoint of the air parcel lies in the ij^{th} cell, the trajectory is assumed to collect $PM_{2.5}$ emitted in the cell. Once the $PM_{2.5}$ is incorporated into the air parcel, it is assumed to be transported along the trajectory to the monitoring site. $PSCF_{ij}$ is the conditional probability that an air parcel that passed through the ij^{th} cell had a high concentration upon arrival at the monitoring site defined as

$$PSCF_{ij} = \frac{m_{ij}}{n_{ij}} \qquad \textbf{(2-9)}$$

where n_{ij} is the total number of endpoints that fall in the ij^{th} cell and m_{ij} is the number of endpoints in the same cell that are associated with samples that exceeded the threshold criterion. In this study, the average contribution of each source was used as the threshold criterion. The sources are likely to be located in the areas that have high PSCF values (Draxler and Hess, 1997). Emissions data, including point sources and fire locations, were overlaid on the PSCF maps to identify specific emissions sources in likely source areas.

While trajectories provide useful information on regional transport, wind roses can provide insight into local transport. For the 10% of sample days with the highest concentrations, wind roses were examined for selected species. Hourly wind data were not available at the STN monitoring sites; as a result, wind data were used from nearby locations. A map of the study sites and the locations from which wind data were obtained is found in **Figure 2-6**. Site KDTW was used for Allen Park and Dearborn wind roses and site KDUH was used for Luna Pier. The wind roses were combined with point source emissions data to understand the link between PMF factors and specific sources.
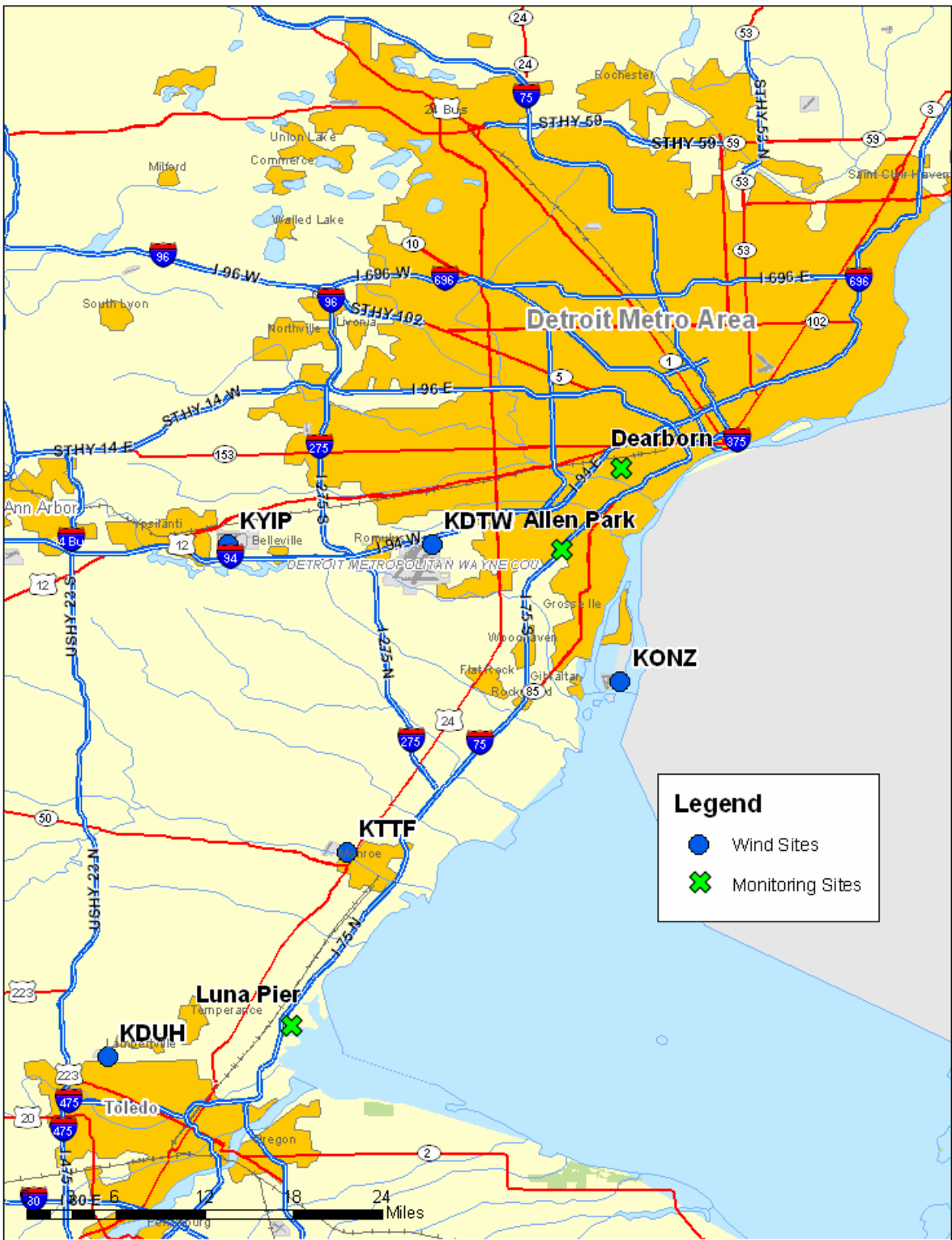
Figure 2-6. Meteorological monitoring sites located near the three STN sites.

3. RESULTS

This section includes summaries of the results of the ambient data analysis and the PMF runs for both STN and SANDWICH PM$_{2.5}$ data sets, and analysis of trajectories and wind as discussed in Section 2. Additional exploratory results are presented for PMF runs conducted with both STN PM$_{2.5}$ data and air toxics data.

3.1 DATA ANALYSIS

PMF should be considered one of a set of complementary analyses that help analysts understand and quantify source contributions to ambient concentrations. Prior to running PMF, it is essential to analyze ambient data to better understand the results of the source apportionment. These analyses can provide information on which factors are expected and can provide checks for determining whether results are sensible. Consequently, the composition of the ambient data and the correlations between various species were examined.

The average ambient PM$_{2.5}$ composition of data at Luna Pier, Allen Park, and Dearborn are displayed in pie charts in **Figures 3-1 through 3-3**. Each figure includes two pie charts, one with STN data and one with SANDWICH data. Allen Park and Dearborn have very similar ambient PM$_{2.5}$ compositions, with Dearborn having a slightly higher sulfate concentration and Allen Park having more organic mass (OM= blank corrected OC*1.8). At Dearborn, a larger fraction of the mass is unaccounted for (grey portion of the pie) and there is more soil mass. When comparing the STN and SANDWICH compositions, a larger fraction of the mass is associated with ammonium sulfate with SANDWICH data and there is a corresponding decrease in ammonium nitrate. There is also a decrease in the organic mass portion with SANDWICH data relative to STN.

Yearly and seasonal (quarterly) trends in the ambient data were examined for significant differences year to year or season to season. Yearly trends by season were also reviewed. Yearly trends in ambient composition for all three sites are shown in **Figure 3-4** and at all sites there is a clear decrease in PM$_{2.5}$ mass in 2004. Seasonal comparisons (**Figure 3-5**) showed a shift in composition from higher ammonium sulfate in the summer to higher ammonium nitrate in the winter, as expected. An increase in OM was also seen in the winter months, but overall, there is a decrease in total PM$_{2.5}$ mass. Yearly comparisons of winter season ambient data (**Figure 3-6**) again show decreased PM$_{2.5}$ mass in 2004, however, at Dearborn low mass is also observed in 2005. The higher 2003 mass has a larger organic mass compared with the two lower years. The winter data from 2002 were excluded for both Dearborn and Luna Pier because of insufficient data. Yearly comparisons of summer season ambient data (**Figure 3-7**) show decreased PM$_{2.5}$ mass in 2004 caused by a decrease in ammonium sulfate relative to other years.

Select scatter plots by season are shown in **Figures 3-8 and 3-9**. Zinc and manganese were found to be correlated at both Allen Park and Dearborn and nickel/chromium correlations were found at all three sites. As a result, we expect to find these species grouped in the same PMF factors. Another correlation found was between silicon and calcium, which is often associated with a soil source. However, as shown in **Figure 3-9**, the plot between silicon and

calcium is more scattered at Luna Pier than at the other two sites; the scatter indicates a second source of calcium at Luna Pier.
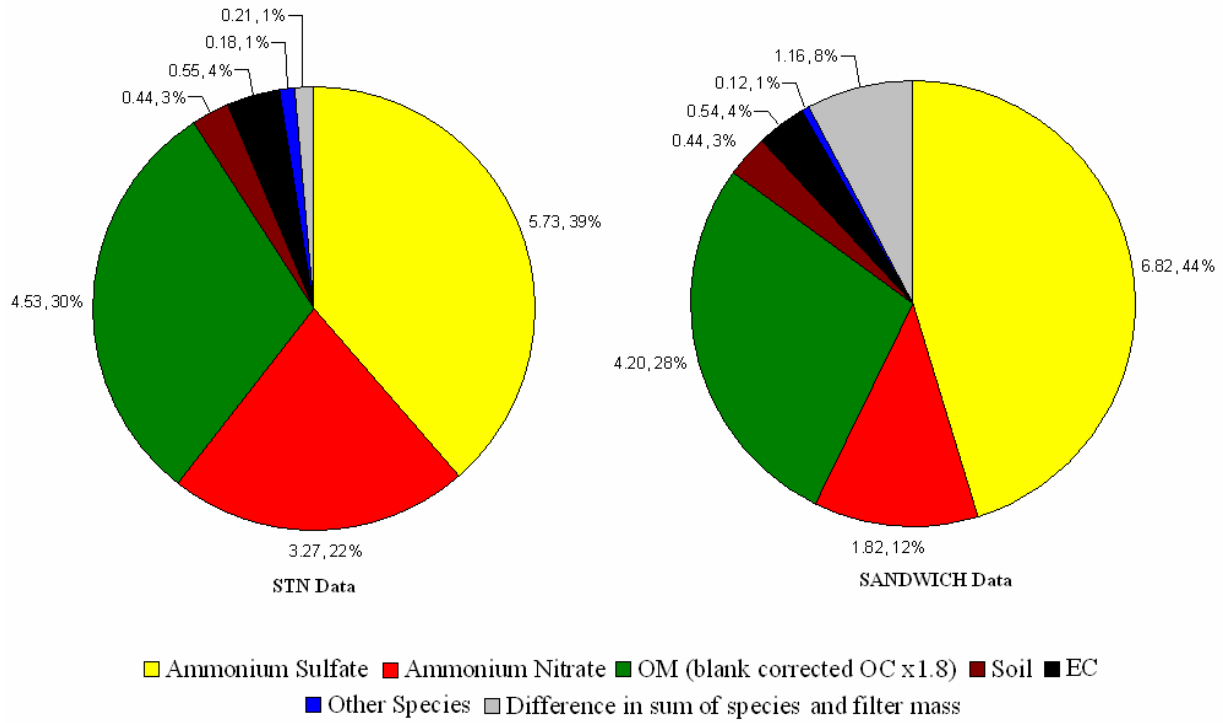

Figure 3-1. Luna Pier average ambient composition for STN and SANDWICH data (May 2002 through December 2005).
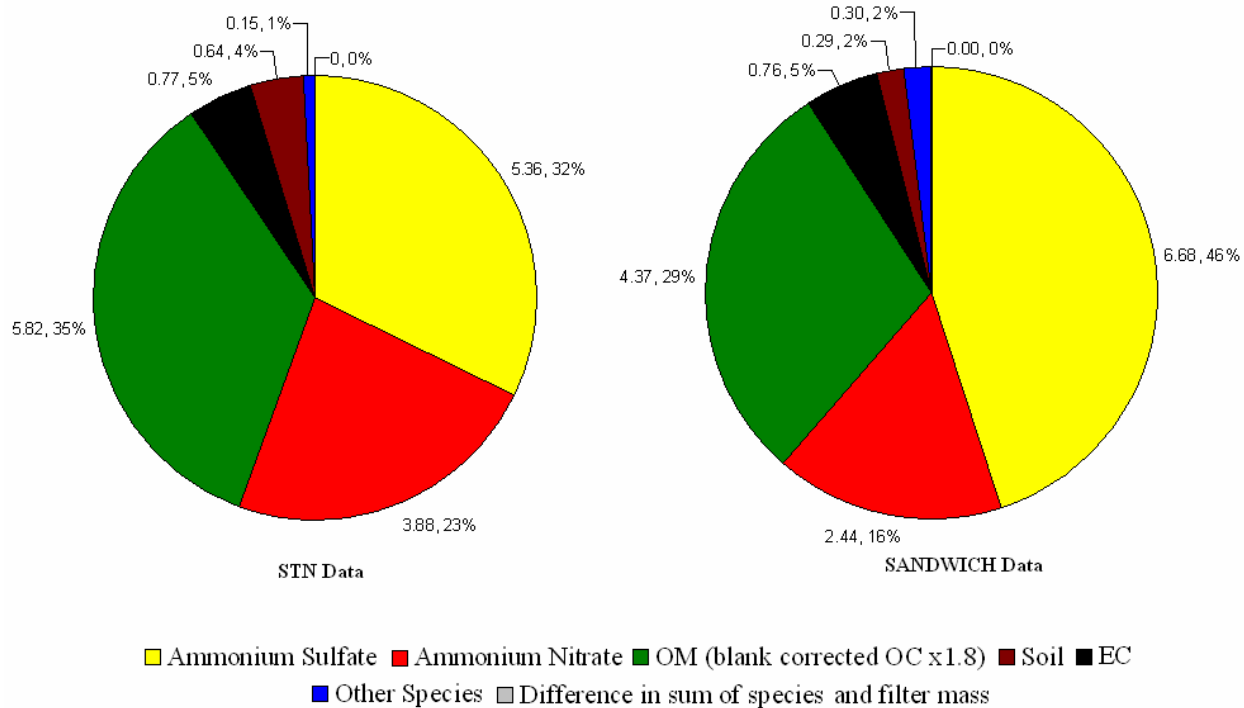
Figure 3-2. Allen Park average ambient composition for STN and SANDWICH data (December 2000 through December 2005).
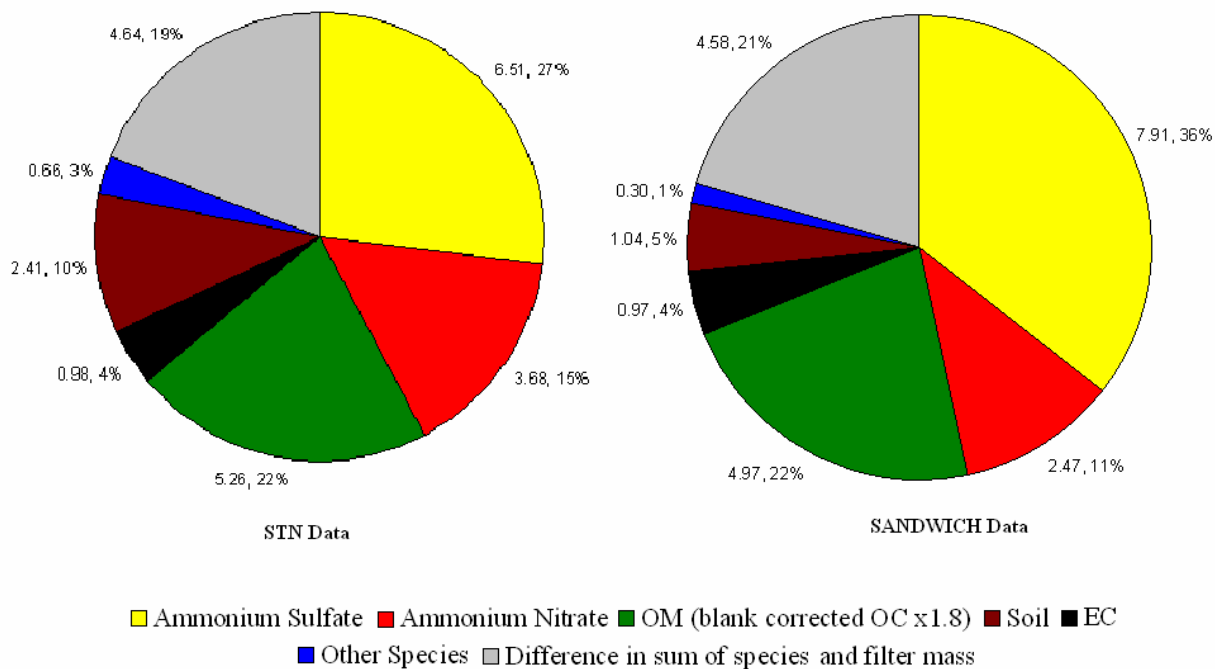

Figure 3-3. Dearborn average ambient composition for STN and SANDWICH data (May 2002 through December 2005).
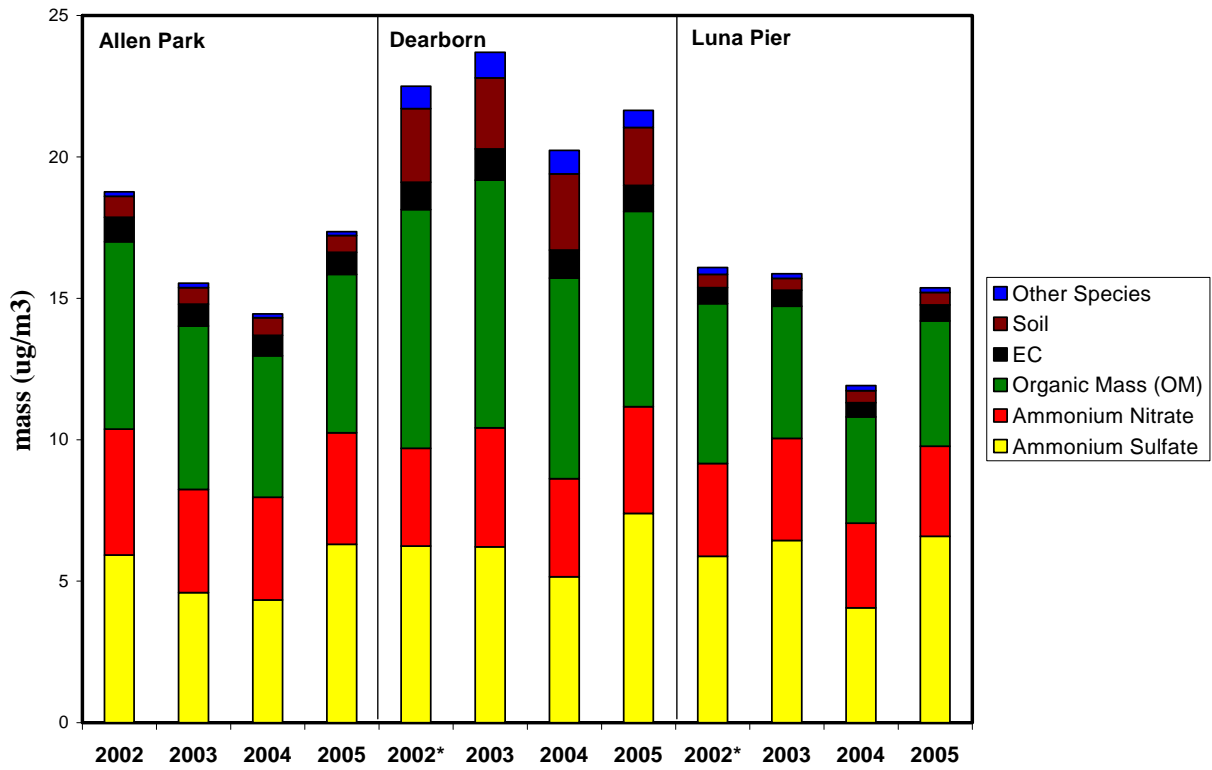
Figure 3-4. Yearly ambient PM$_{2.5}$ composition for STN data at Allen Park, Dearborn, and Luna Pier. An asterisk (*) indicates incomplete data collected in winter months.
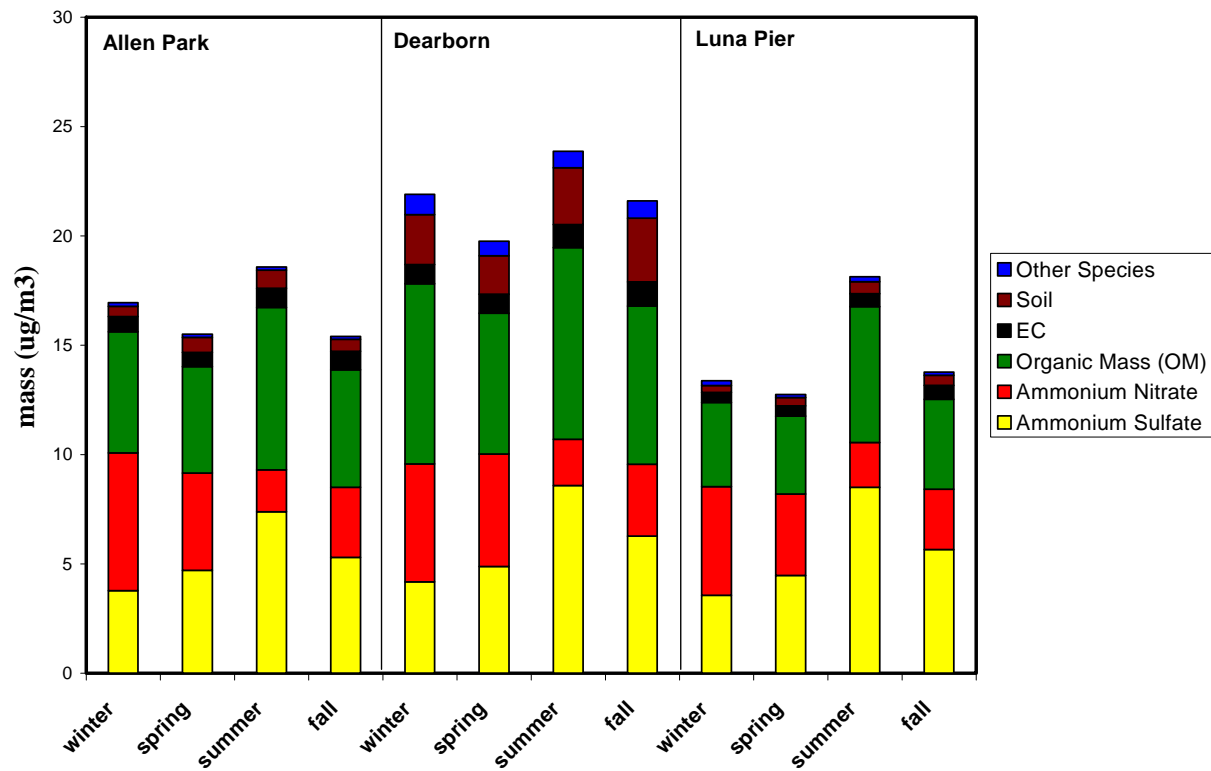
Figure 3-5. Seasonal ambient PM$_{2.5}$ composition for STN data at Allen Park, Dearborn, and Luna Pier (Allen Park: December 2000 through December 2005, Dearborn and Luna Pier: May 2002 through December 2005).
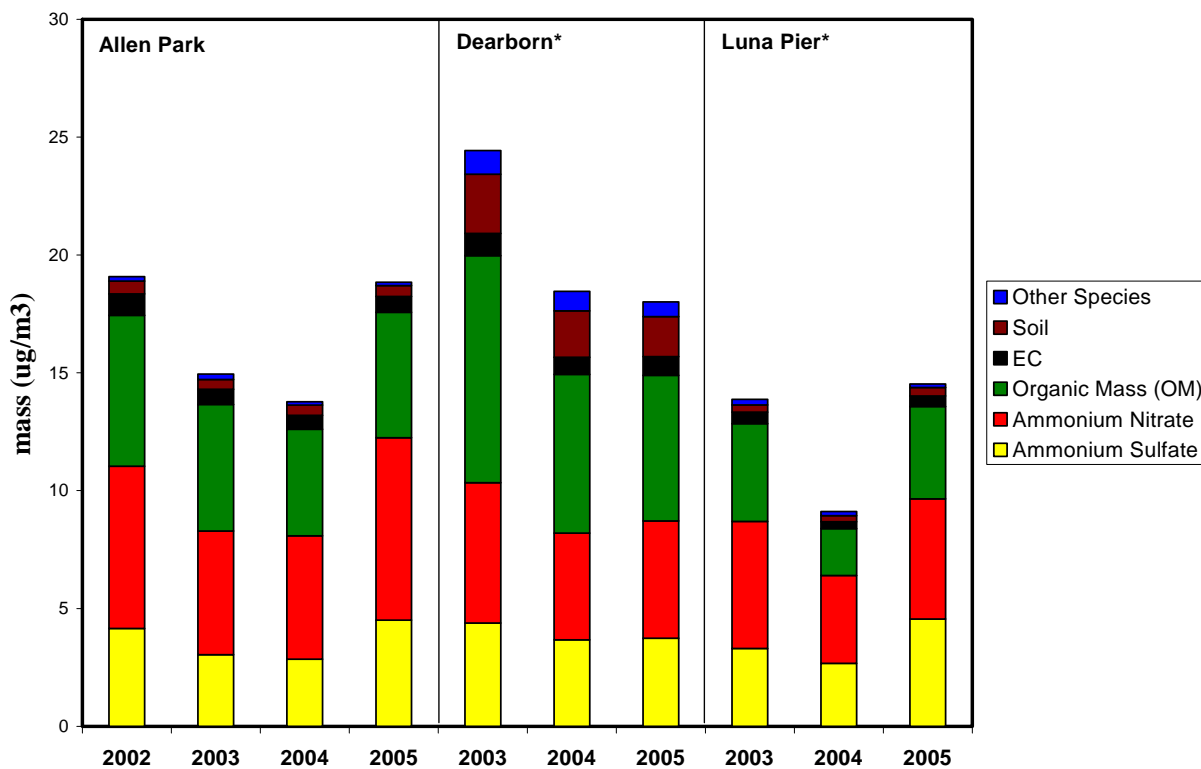
Figure 3-6. Yearly winter ambient PM$_{2.5}$ composition for STN data at Allen Park, Dearborn, and Luna Pier. An asterisk (*) indicates incomplete data were collected in winter months.
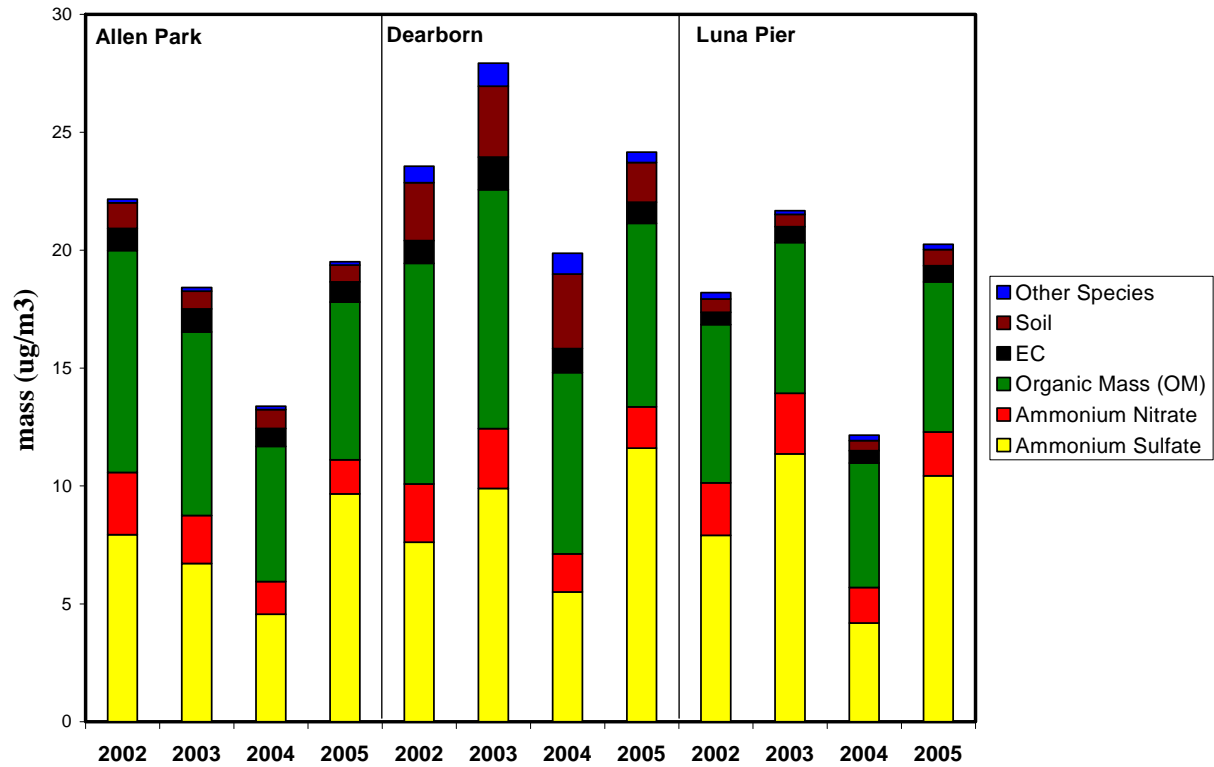
Figure 3-7. Yearly summer ambient PM$_{2.5}$ composition for STN data at Allen Park, Dearborn, and Luna Pier.
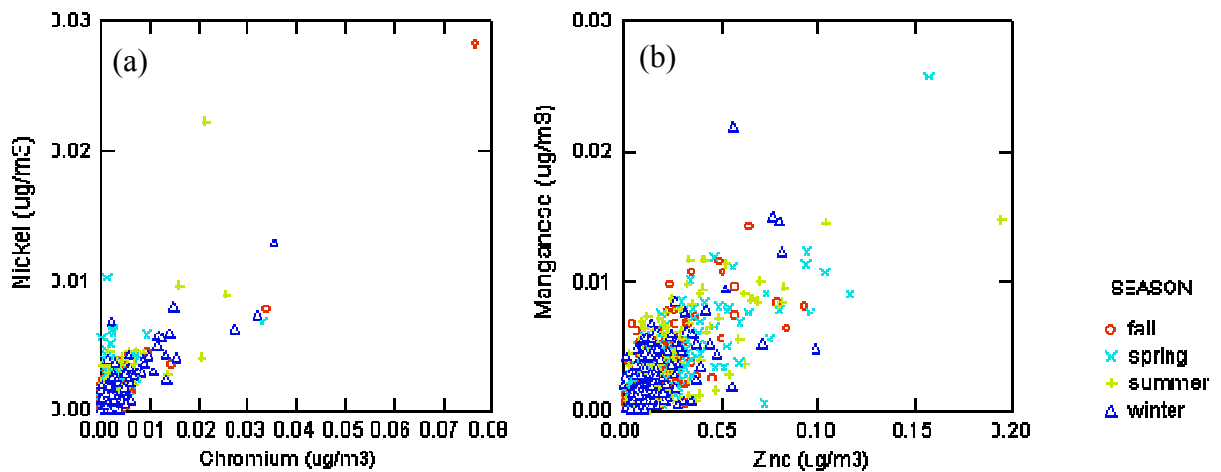

Figure 3-8. Scatter plot by season of (a) nickel and chromium and (b) manganese and zinc at Allen Park (December 2000 through December 2005).
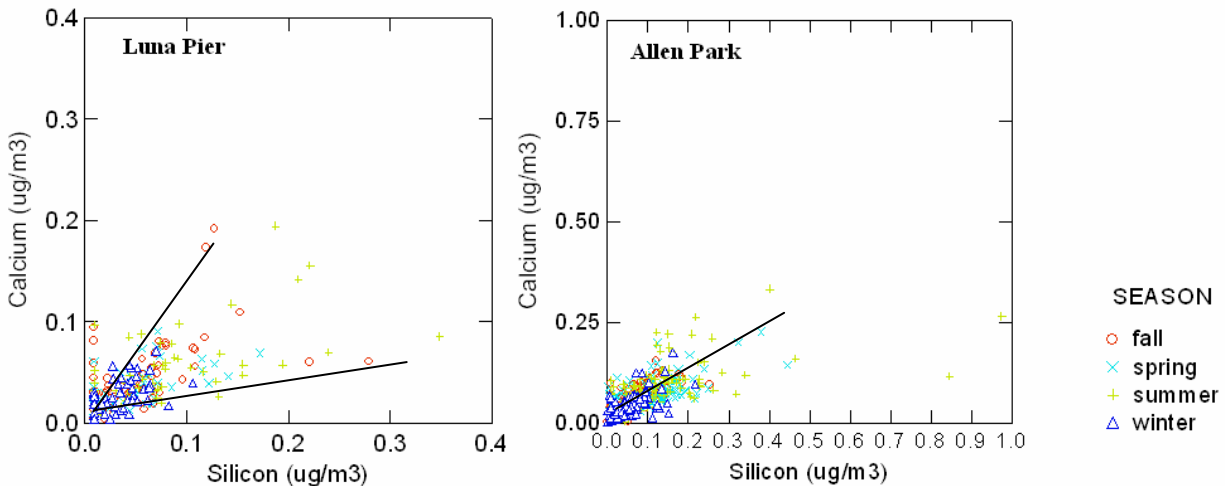
Figure 3-9. Scatter plot by season of silicon and calcium at Luna Pier and Allen Park (Luna Pier: May 2002 through December 2005, Allen Park: December 2000 through December 2005). Lines indicate edges or general trends.

3.2 SOURCE APPORTIONMENT OF STN PM$_{2.5}$ DATA

For all three sites, PMF was run using 7 to 11 factors and additional runs were conducted as sensitivity tests for inclusion of selected species. The sensitivity runs included species such as chlorine and various metals that had a large fraction of samples below detection. Species such as chlorine had very low R^2 values between measured and modeled results (0.11 at Allen Park) and had little effect on the overall results; they were therefore, excluded. Additional sensitivity runs were conducted to test the use of elemental potassium instead of potassium ion. Potassium ion, which is a typical tracer for wood burning, was mostly below detection whereas elemental potassium, not unique to wood burning, was mostly above detection. The results of the PMF run showed a strong correlation between potassium and potassium ion with both species coming out in the same factors, indicating that elemental potassium was useful as a wood smoke tracer for these sites.

Additional runs were conducted using data collected prior to July 2003 and using data collected from July 2003 to October 2005. Samples taken between July 2003 and October 2005 had recently updated uncertainties reported in AQS. For samples taken prior to July 2003, uncertainties were estimated based on methods previously discussed. The same factors and similar mass attributed to each factor were found using either set of data, showing that the estimated uncertainties do not change the solution.

The final number of runs and the species used at each site were based on model performance criteria including the Q-value, convergence, species correlations, and mass recovery. The number of factors for both STN and SANDWICH data is summarized in **Table 3-1**. For each site and data set, 10 random runs were conducted for the final number of factors to ensure robust results. Over the 10 runs, Q values were stable at all three sites and were within 50% of the theoretical Q values. Residuals of the PMF results were also within

recommended bounds (-3 < residual < 3) (Paatero et al., 2003) excluding a few outliers. At Allen Park, Luna Pier, and Dearborn, there were 11, 1, and 4 outliers, respectively.

Table 3-1. Number of factors used for final PMF runs at the three STN sites for both STN data and SANDWICH data.

Sites	STN Data	SANDWICH Data
Luna Pier	8	8
Allen Park	9	8
Dearborn	10	10

Ideally, tracers for PMF sources would be mostly above detection limits and unique to a particular source. Of the species within the STN data sets, iron and zinc are mostly above detection and are unique to industrial sources in the Detroit area. Sources of these species should, therefore, be easily quantifiable. Other metals such as nickel, chromium, and manganese have more than 50% of the samples below detection. These species are nearly unique tracers and can still be quantified. For sources such as wood burning, the only unique metal tracer is potassium. Potassium has other sources including soil; consequently, it will be hard to isolate a wood burning source.

The factor profiles from the PMF runs are presented in **Figures 3-10 through 3-12**. All three sites had some similar factors including ammonium sulfate, ammonium nitrate, soil, and mobile sources. The presence of sulfate, nitrate, and soil is consistent with the ambient data composition previously examined. In addition, we expect to find an influence of mobile sources as all three sites are surrounded by interstates and various highways as shown in Figure 2-1. Correlated species were grouped together in factors as indicated in Section 3-1. Silicon/calcium and nickel/chromium factors were identified in addition to a second calcium factor at Luna Pier. The identification of two industrial factors at Luna Pier may be attributed to the proximity of this site to both Toledo and Detroit (Figure 2-1). Of all three sites, Dearborn had the most factors resolved. This is expected due to the complexity of sources around the site.

The mass composition of the PMF results is presented in pie charts in **Figures 3-13 through 3-15** for STN runs. At all three sites, ammonium nitrate, ammonium sulfate, and organic carbon are the largest portions of the mass, which is in agreement with the ambient data. Yearly and seasonal trends were examined for the PMF results and are presented in **Figures 3-16 through 3-18**. Average yearly trends in the PMF results showed decreased mass in 2004, consistent with the ambient data. At all sites, there is in increase in apportioned mass in the summertime with increased ammonium sulfate and organic carbon, also consistent with the ambient data.
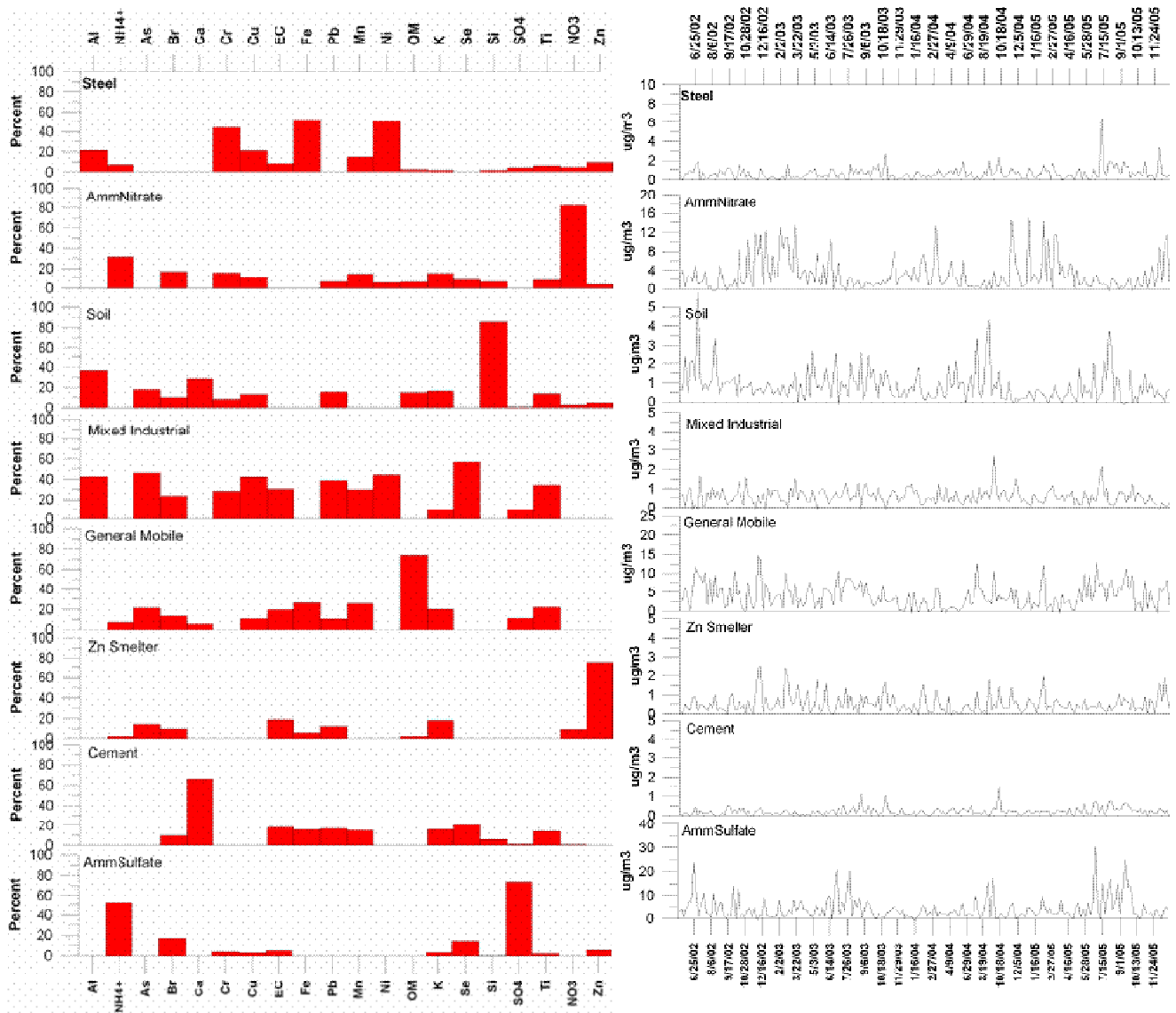
Figure 3-10. Luna Pier PMF factor profiles and time series for STN data
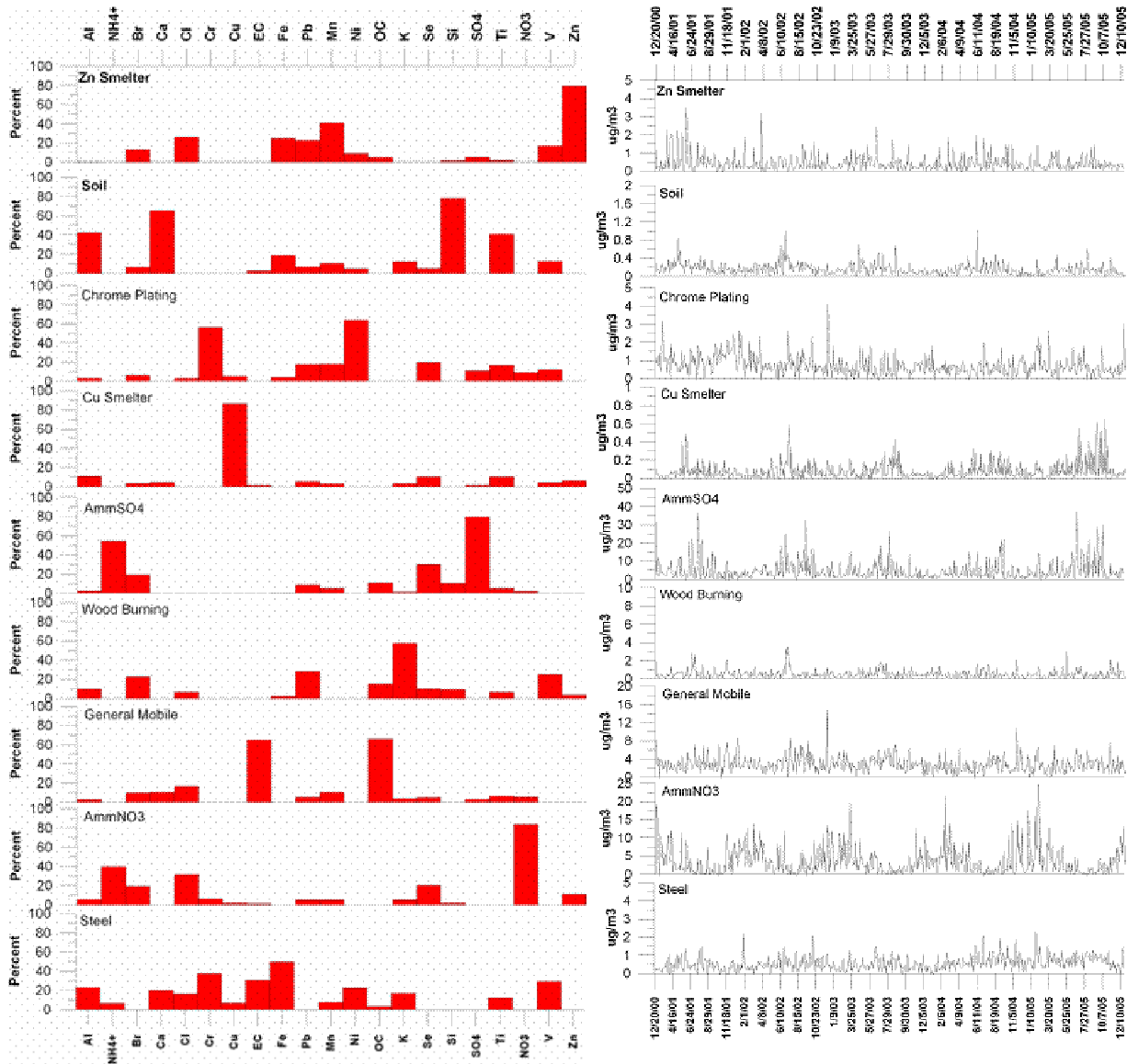(May 2002 through December 2005).

Figure 3-11. Allen Park PMF factor profiles and time series for STN data
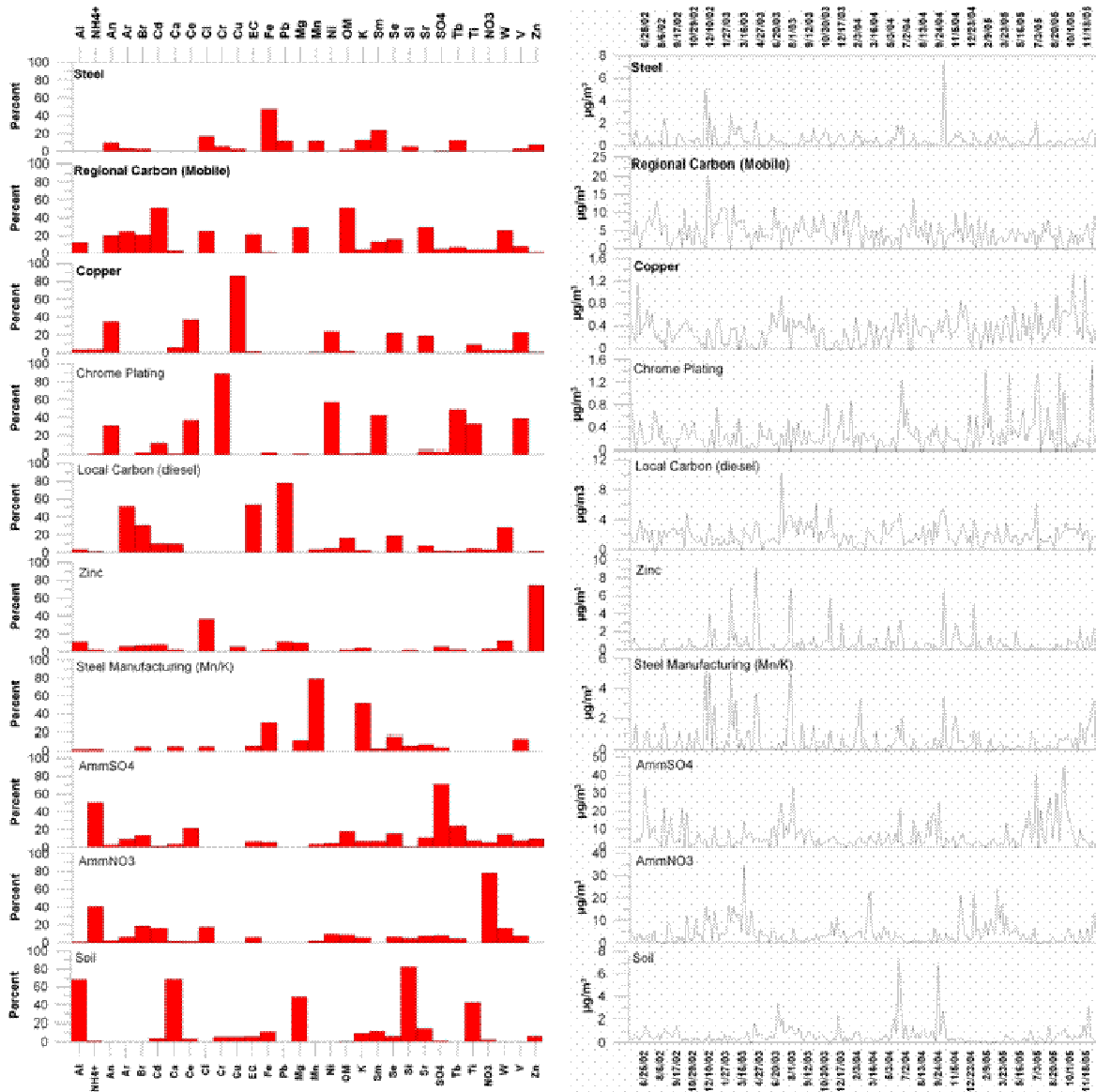(December 2000 through 2005).

Figure 3-12. Dearborn PMF factor profiles and time series for STN data
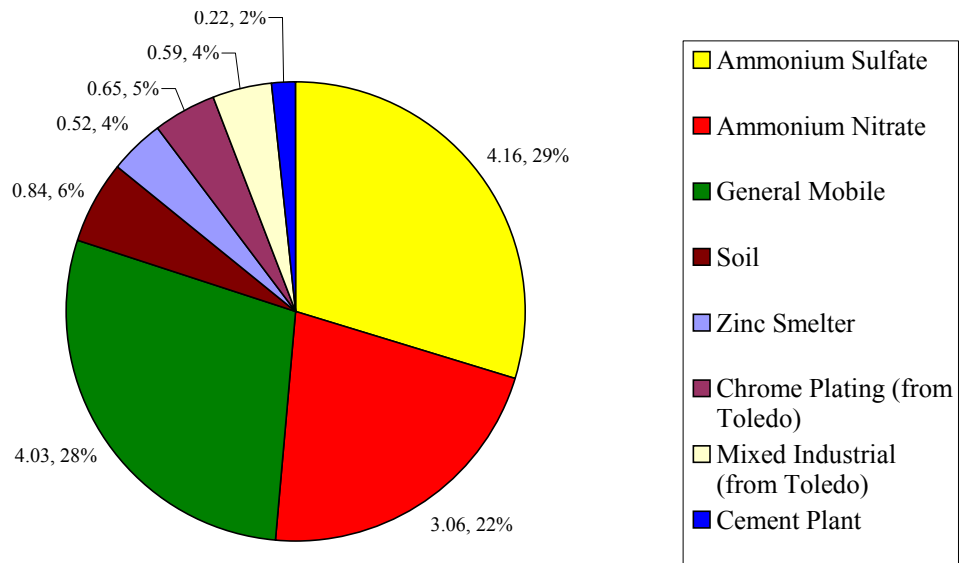(May 2002 through December 2005).

Figure 3-13. Average composition of 8-factor PMF results at Luna Pier for STN data (May 2002 through December 2005).
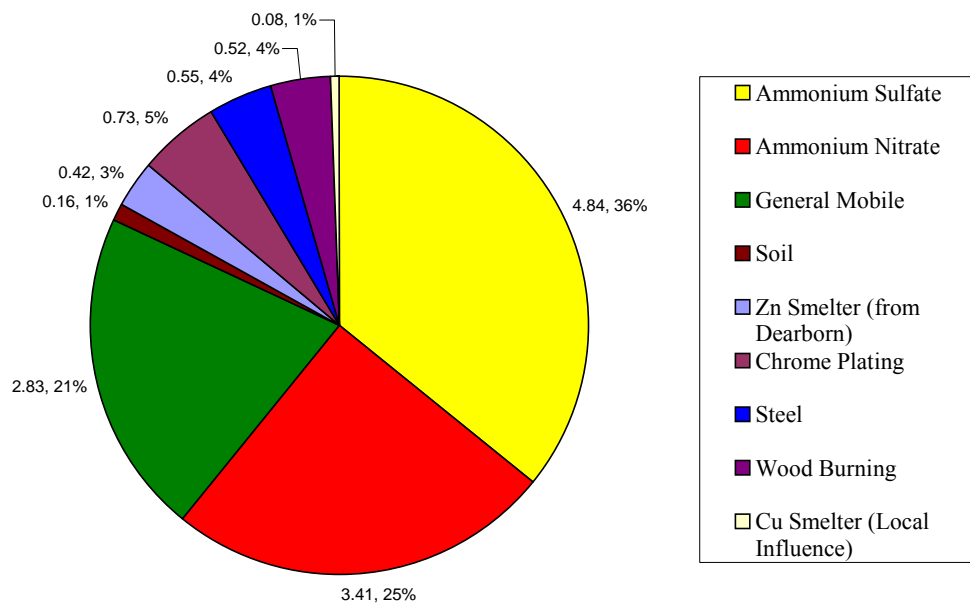

Figure 3-14. Average composition of 9-factor PMF results at Allen Park for STN data (2000 through 2005).
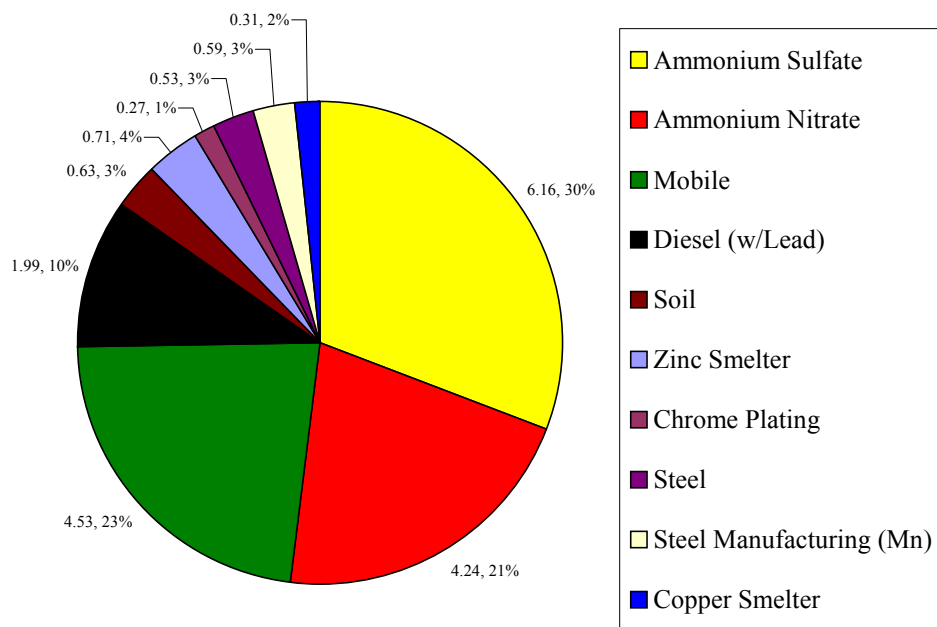
Figure 3-15. Average composition of 10-factor PMF results at Dearborn for STN data (May 2002 through December 2005).
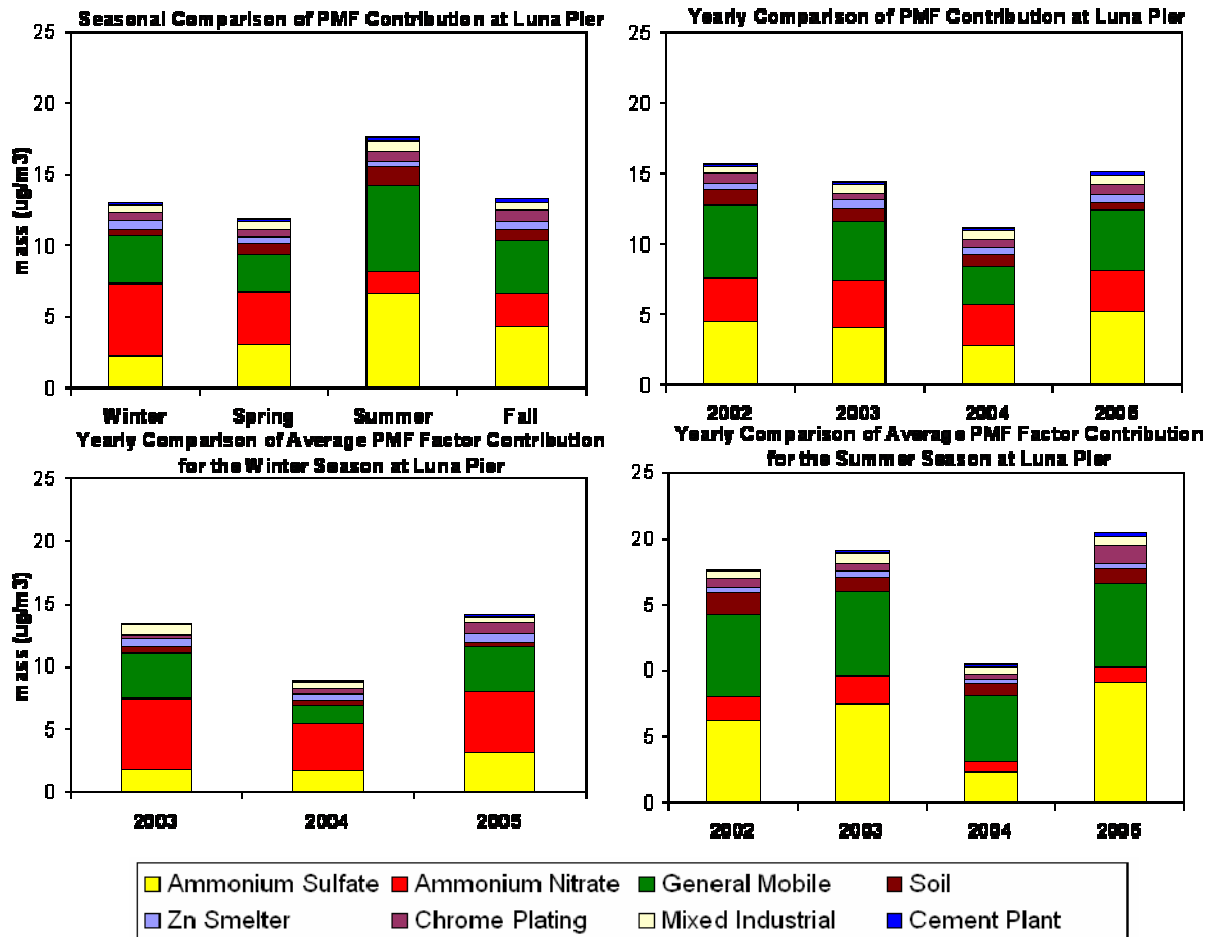
Figure 3-16. Yearly and seasonal trends in PMF mass composition at Luna Pier for STN data (May 2002 through December 2005).
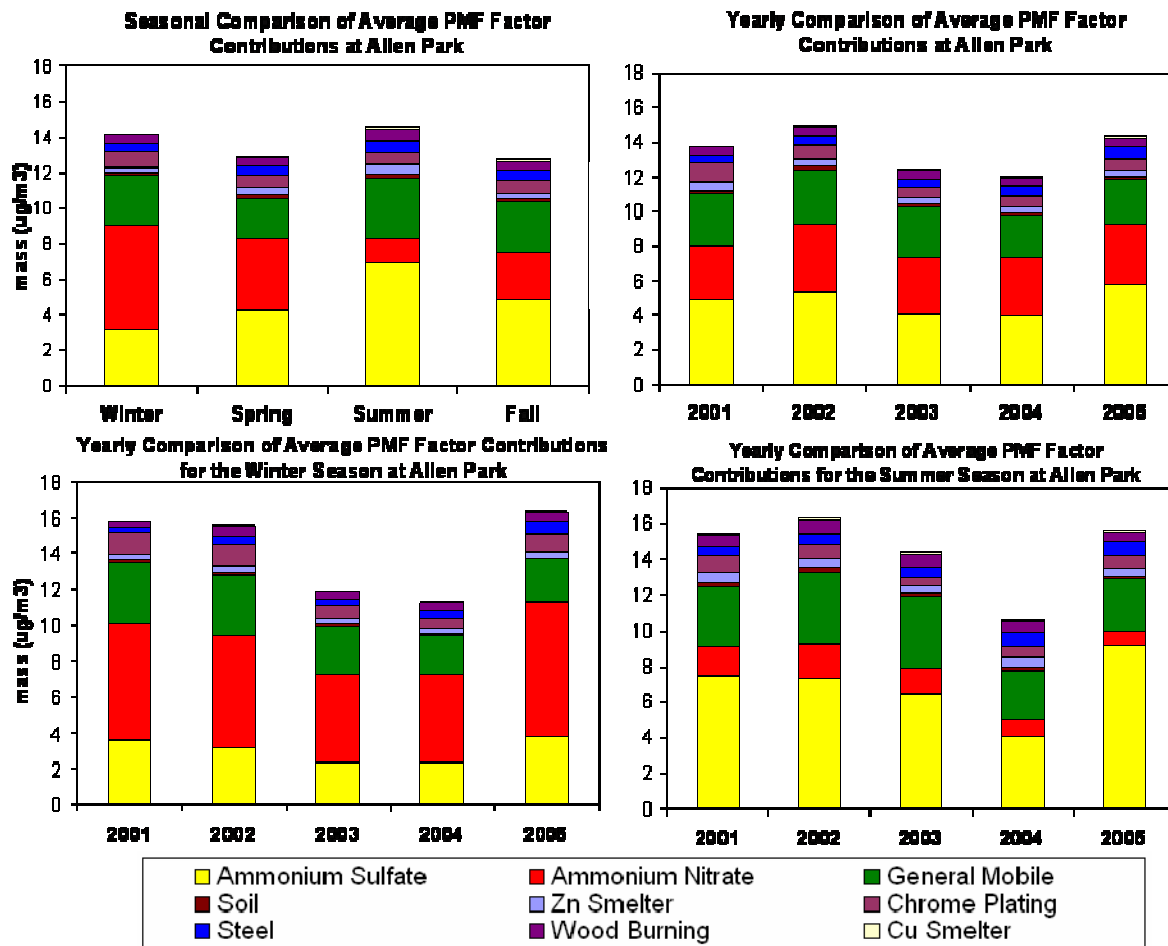
Figure 3-17. Yearly and seasonal trends in PMF mass composition at Allen Park for STN data (December 2000 through December 2005).
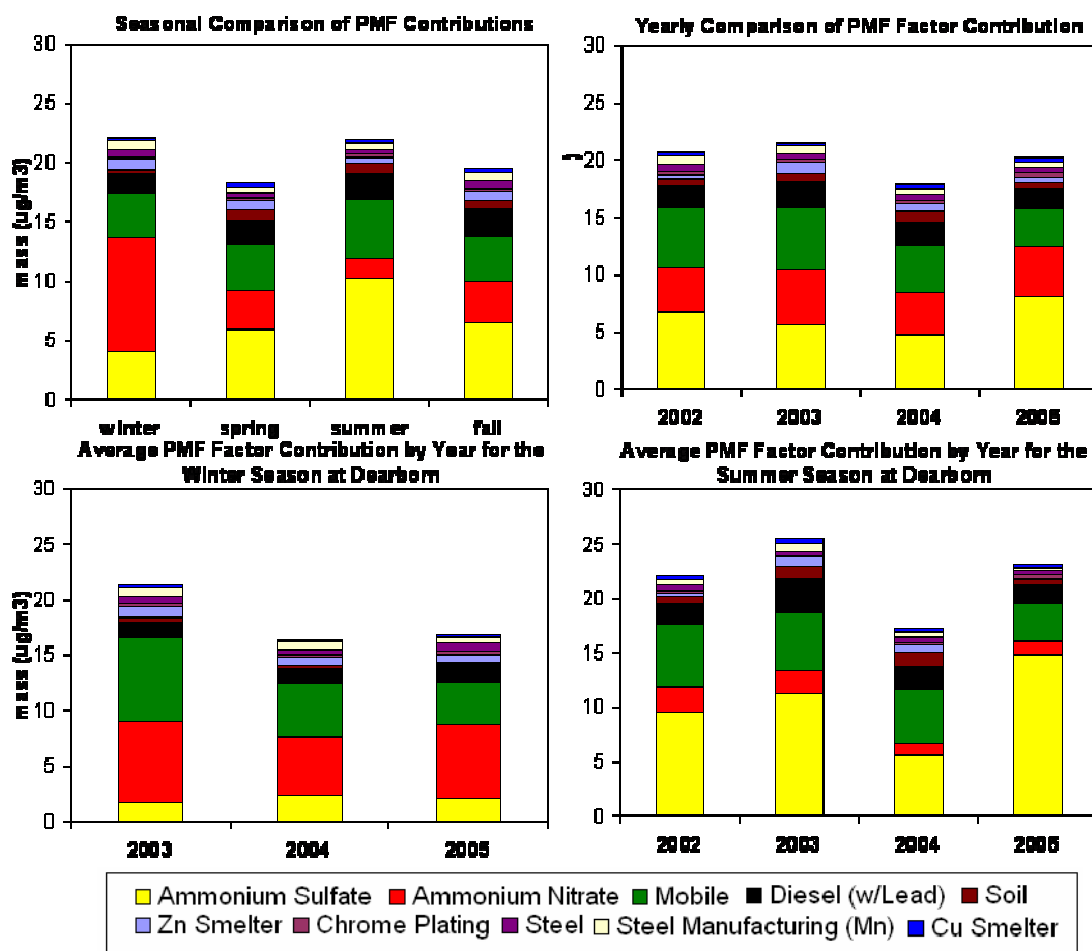
Figure 3-18. Yearly and seasonal trends in PMF mass composition at Dearborn for STN data (May 2002 through December 2005).

Total PM$_{2.5}$ point source emissions in the Detroit area are shown in **Figure 3-19**. It is important to note that the Dearborn site is located close to many large point sources while the Luna Pier site is located between several large point sources to the northeast and many smaller sources from the Toledo area to the south. Point source emissions were also examined for representative species of the PMF factors identified including chromium, nickel, zinc, manganese, copper, and calcium. Point source emissions are shown in **Figures 3-20** (zinc and manganese), **3-21** (chromium and nickel), **and 3-22** (copper). The point source emissions combined with wind roses provide information on the sources associated with the PMF factors. Representative wind roses for a typical day at the three STN sites are shown in **Figure 3-23**. Typical wind roses provide a comparison for the high pollutant day wind roses. The typical wind roses show winds coming from all directions at Allen Park and Dearborn and winds coming from two directions at Luna Pier. Wind roses on high pollutant days of zinc/manganese, chromium/nickel, and copper are shown in **Figure 3-24**. High zinc days at Luna Pier occur when winds are from the Toledo area while high zinc/manganese days at Dearborn and Allen

3-17

Park occur when winds are from industrial sources between the two sites. On high nickel/chromium days at Dearborn, the wind roses point toward two small point sources directly by the site. Allen Park appears to be most affected by the Detroit Edison power plant southeast of the site for nickel and chromium emissions. For high copper days, both Allen Park and Dearborn are most affected by point sources to the southeast. Dearborn is also located near two much smaller copper sources. The source farther from Dearborn appears to have the most influence on high copper days. Calcium point sources are shown in **Figure 3-25** along with a wind rose for high calcium days at Luna Pier. The point sources show several cement, limestone, and gypsum facilities surrounding Luna Pier, likely explaining the non-soil calcium source found at this site. Wind roses for high emission days associated with the two industrial factors at Luna Pier (**Figure 3-26**) confirm previous suspicions that the two factors are coming from different directions. The mixed industrial factor at Luna Pier points to the large sources northeast of the site and the iron/chromium/nickel factor is influenced by winds from the Toledo area.

Figure 3-19. Proportional PM$_{2.5}$ point source emissions map (U.S. Environmental Protection Agency, 2002).

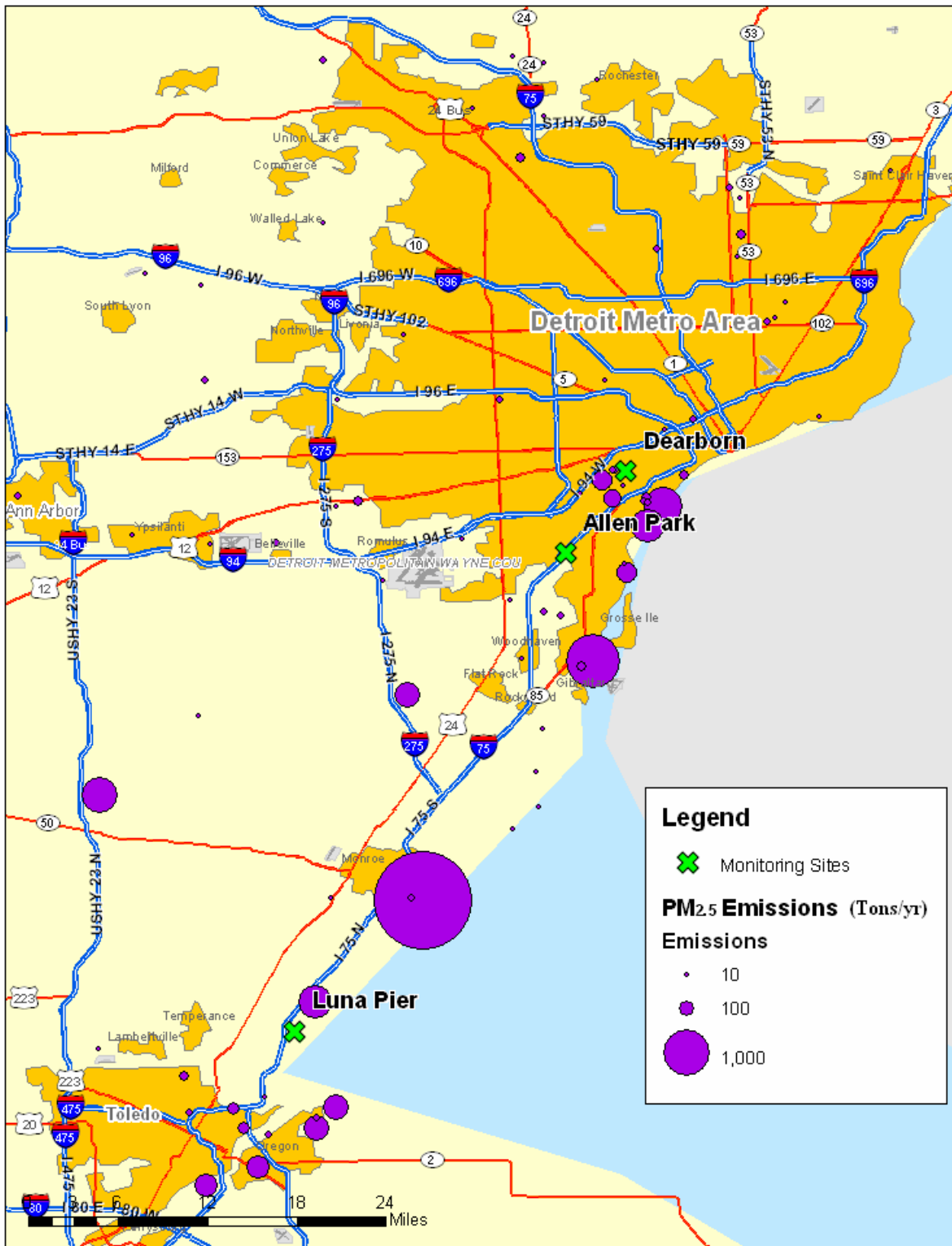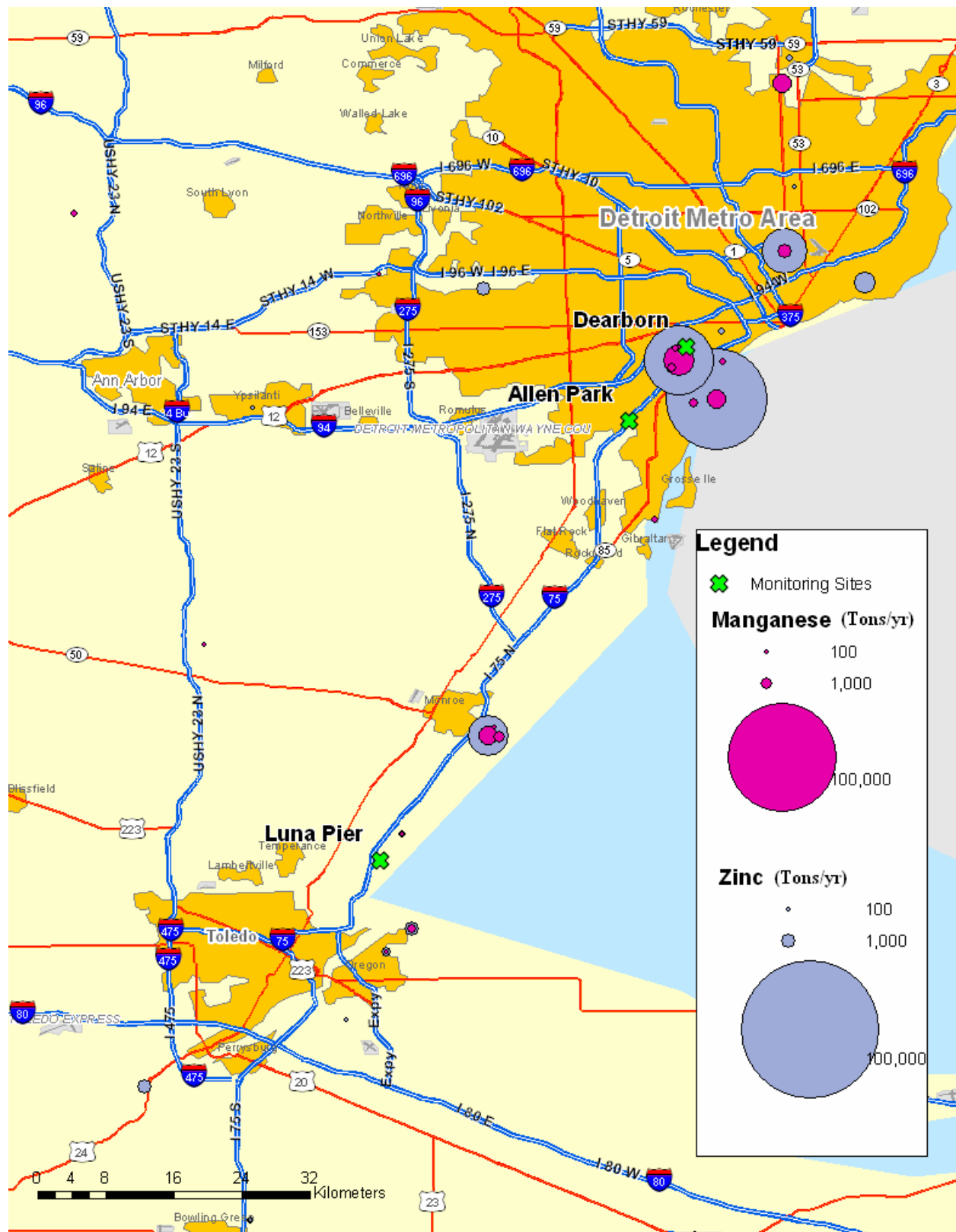Figure 3-20. Proportional zinc and manganese point source emissions map (U.S. Environmental Protection Agency, 2002).
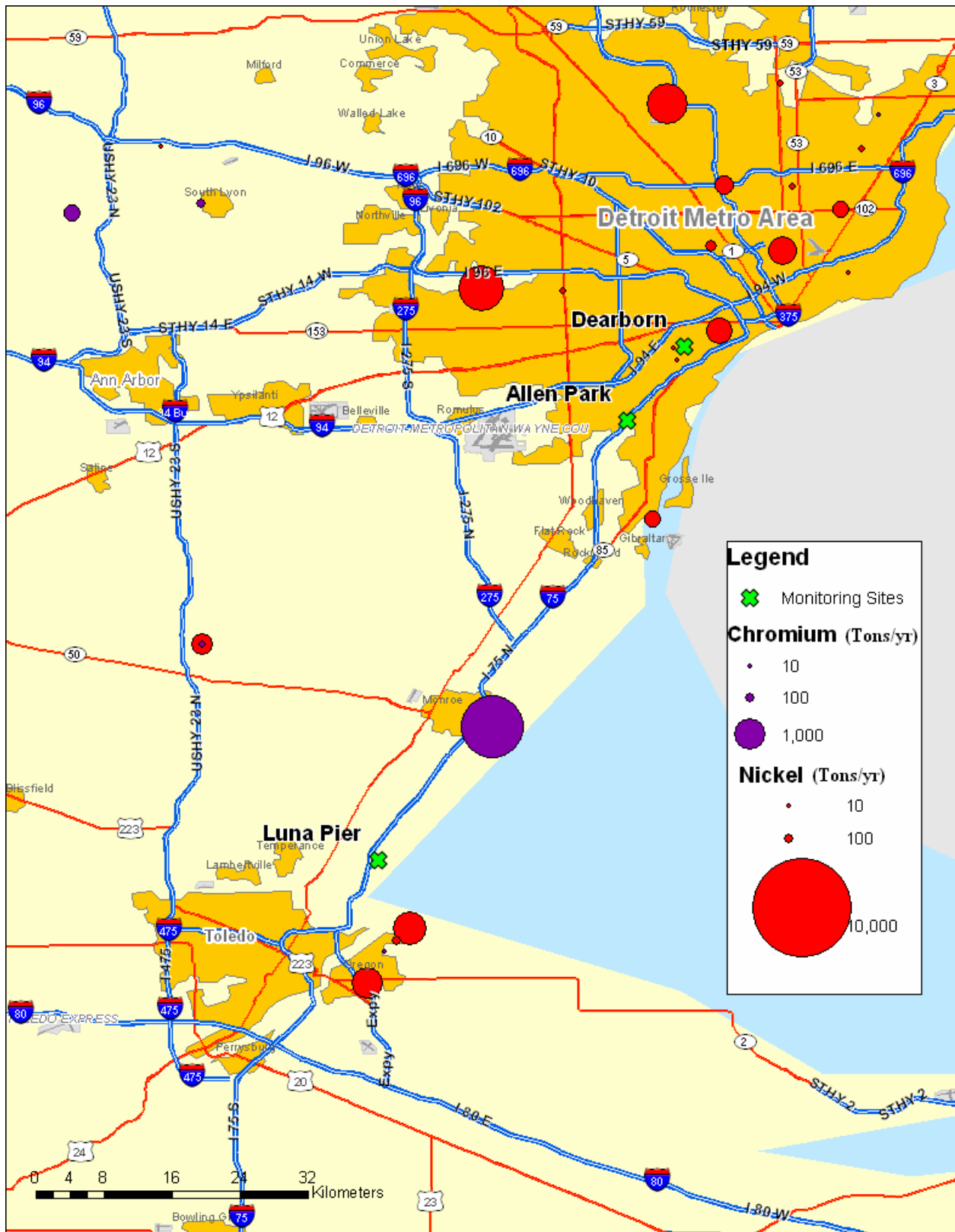
Figure 3-21. Proportional chromium and nickel point source emissions map (U.S. Environmental Protection Agency, 2002).
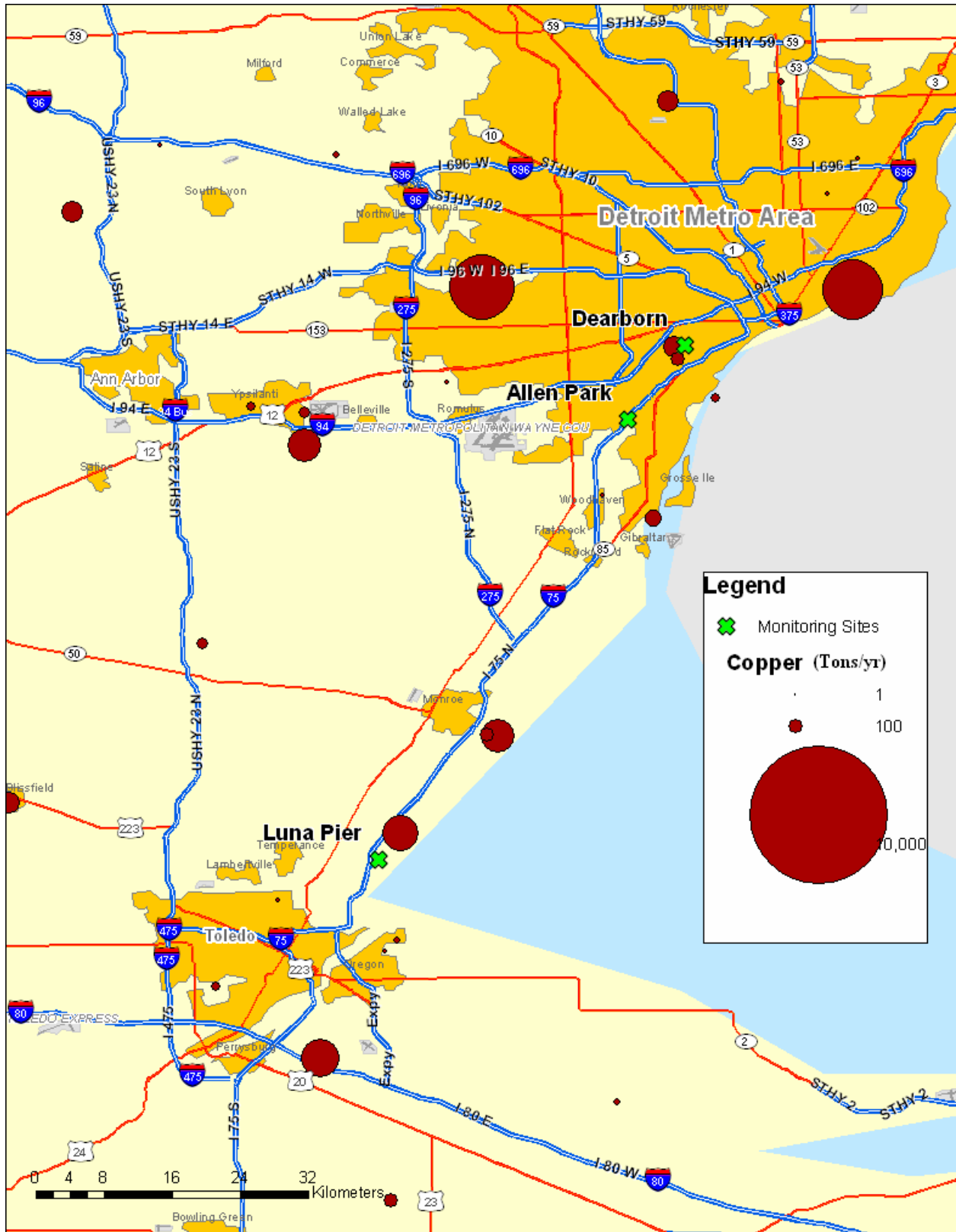
Figure 3-22. Proportional copper point source emissions map (U.S. Environmental Protection Agency, 2002).
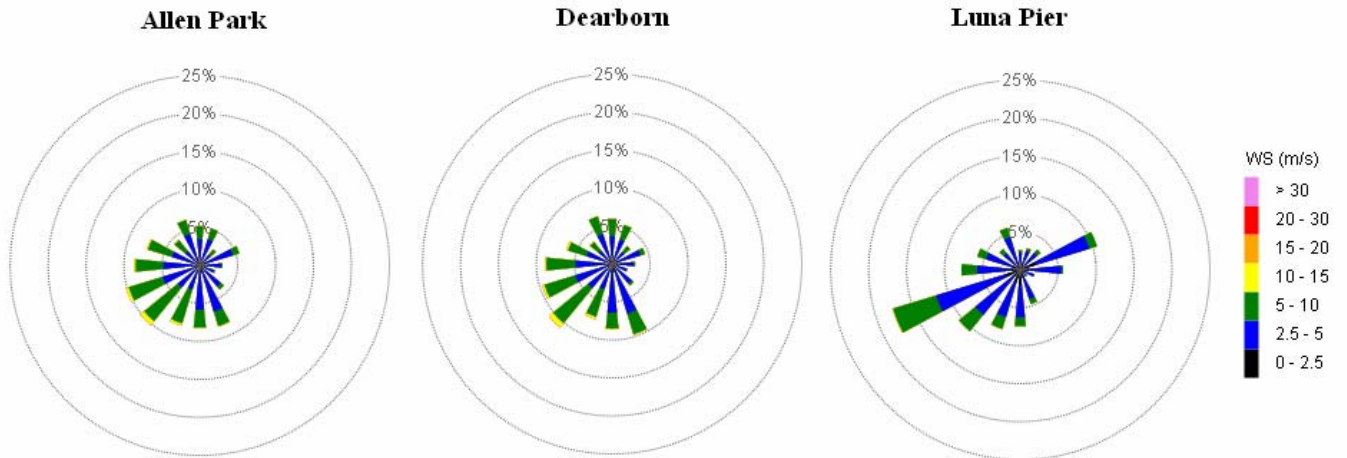
Figure 3-23. Wind roses for three STN sites: Allen Park, Dearborn, and Luna Pier (Allen Park: 2000 through 2005, Dearborn and Luna Pier: May 2002 through December 2005).
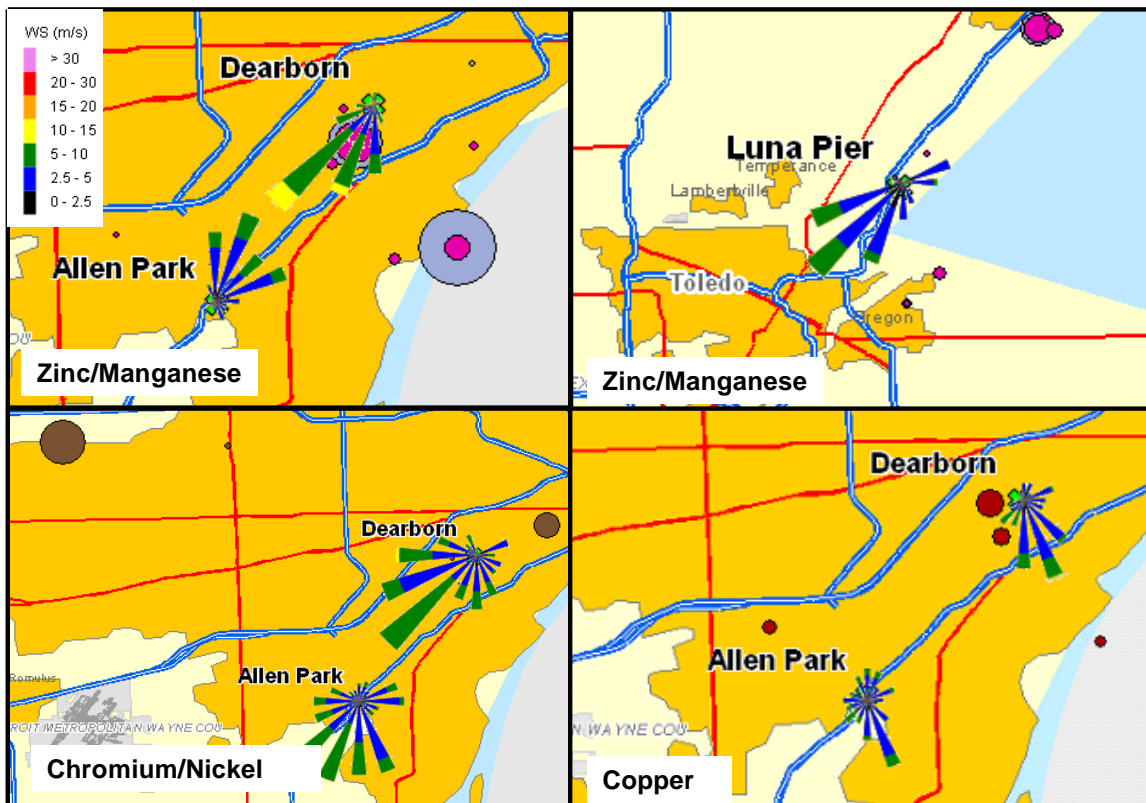
Figure 3-24. Wind roses of zinc/manganese, chromium/nickel, and copper on high pollutant days (U.S. Environmental Protection Agency, 2002). Units for emissions are the same as for Figures 3-20 through 3-22.
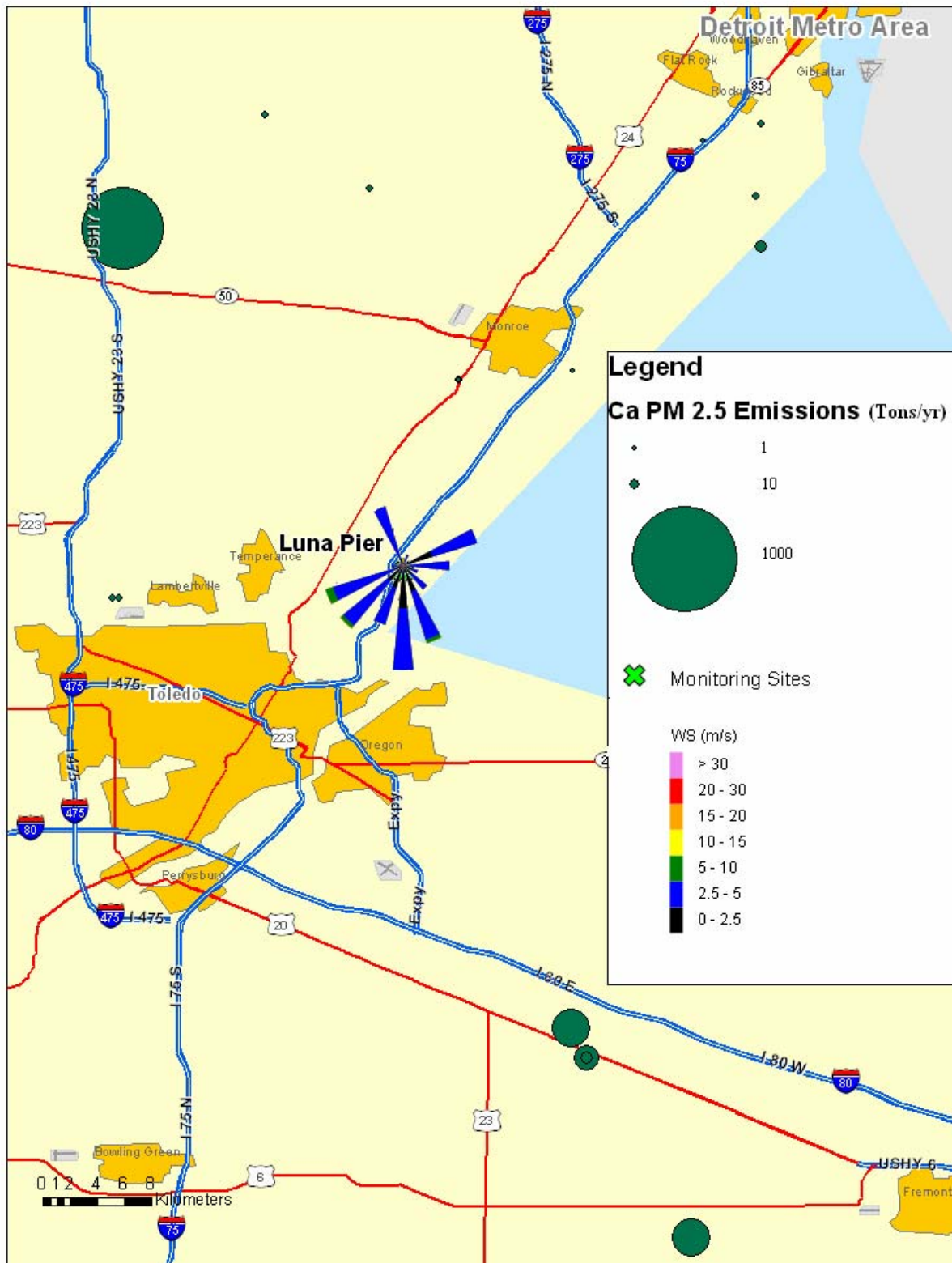
Figure 3-25. Wind roses of calcium on high concentration days with point source emissions at Luna Pier (U.S. Environmental Protection Agency, 2002).
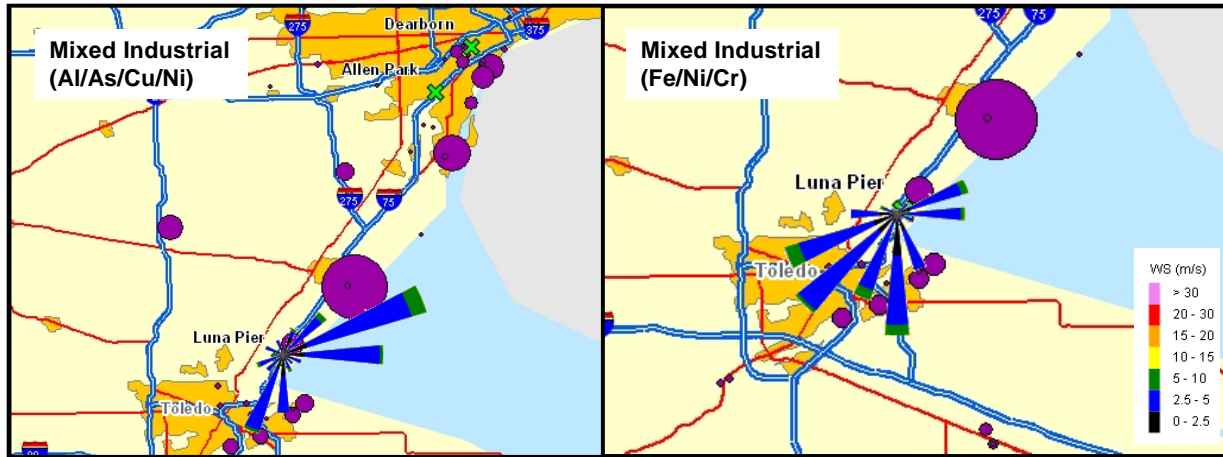
Figure 3-26. Wind roses on high mixed industrial factor days with PM$_{2.5}$ point source emissions at Luna Pier (U.S. Environmental Protection Agency, 2002). Units for emissions are the same as for Figures 3-19 through 3-22.

3.3   SOURCE APPORTIONMENT OF SANDWICH DATA

The PMF results using SANDWICH data are presented in **Figures 3-27 through 3-29**. PMF results using STN data and ambient data composition for both data sets are also shown. In the SANDWICH PMF results, a larger fraction of the mass is attributed to ammonium sulfate and less to ammonium nitrate, consistent with ambient data. Better mass recovery was achieved using the SANDWICH data set, mostly due to the difference in sulfate mass. With respect to the number of factors, Allen Park was the only site for which the SANDWICH and STN data sets did not agree. Using the SANDWICH data, PMF was able to split the carbon into a mobile and a diesel source, which was not achieved with the STN data. However, neither the wood burning nor the steel source was identified with the SANDWICH data. Overall, there is good agreement between SANDWICH and STN results. In general, results found using STN data are useful for FRM and SANDWICH applications. On a daily basis, the SANDWICH PMF results can be different than the STN PMF results, but these differences are nearly all due to the differences between SANDWICH and regular STN data (i.e., carbon, nitrate, and sulfate concentrations are already different in the two data sets).

Luna Pier
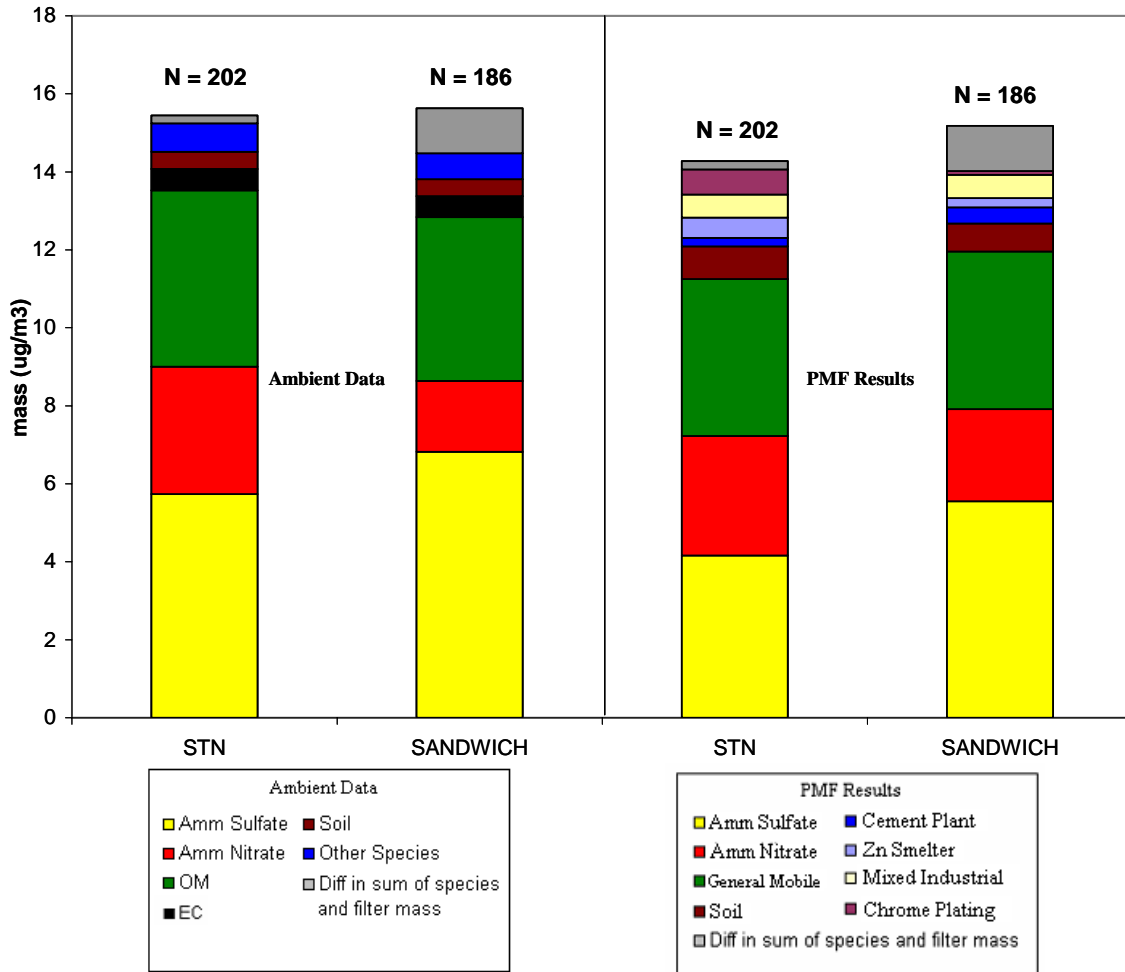
Figure 3-27. PMF results and ambient mass composition for both STN and SANDWICH data sets at Luna Pier (May 2002 through December 2005).
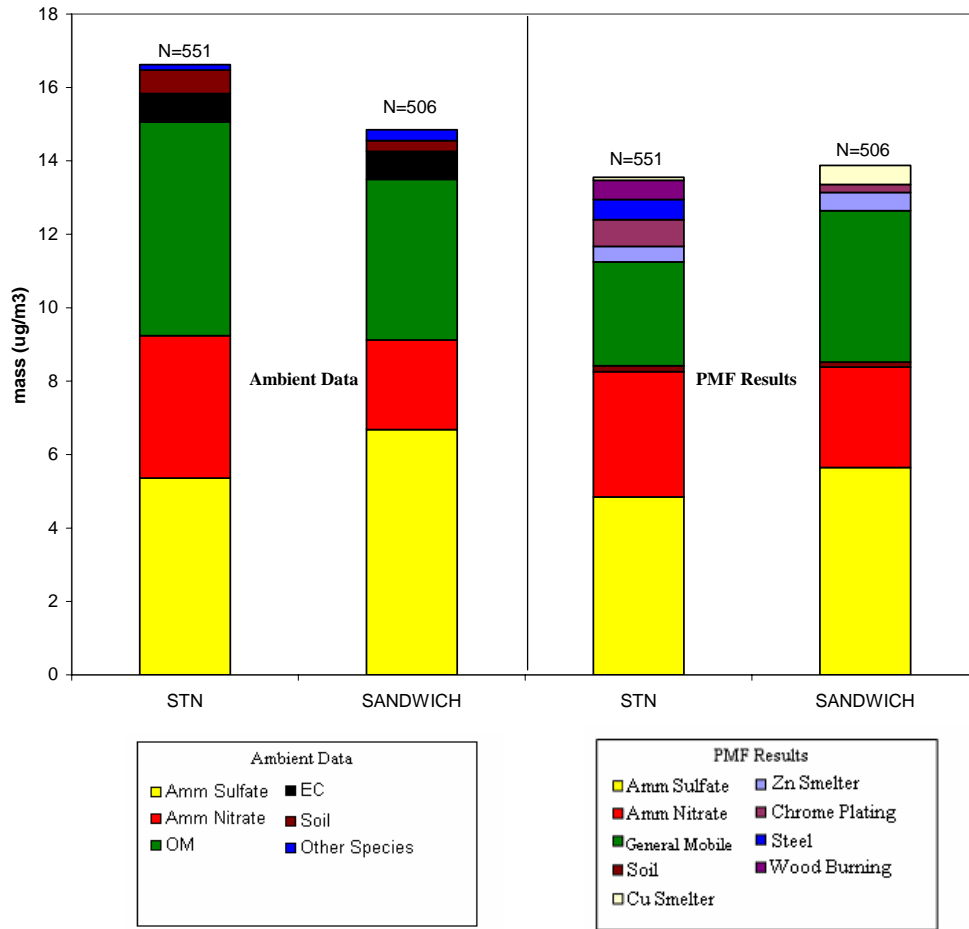
Figure 3-28. PMF results and ambient mass composition for both STN and SANDWICH data sets at Allen Park (December 2000 through December 2005).
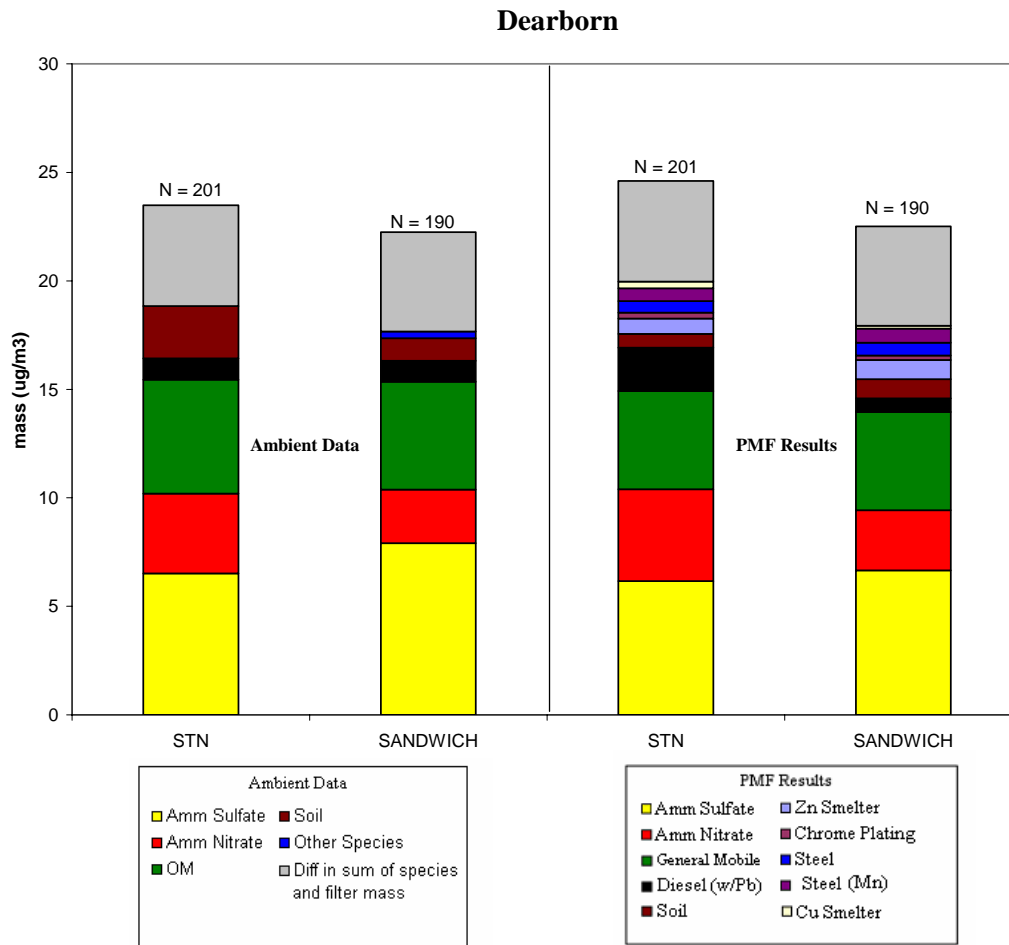
Dearborn

Figure 3-29. PMF results and ambient mass composition for both STN and SANDWICH data sets at Dearborn (May 2002 through December 2005).

3.4   SOURCE APPORTIONMENT OF STN AND AIR TOXICS DATA

Additional PMF runs were conducted that included air toxics data. The additional runs were conducted for Allen Park, because this site had the largest number of STN samples. The availability of air toxics data is much more limited than speciated PM$_{2.5}$ data; therefore, the number of samples used in PMF was drastically decreased to get overlapping sample days. The air toxics included in the analysis were benzene, o-xylene, ethylbenzene, toluene, formaldehyde, and acetaldehyde with a total of 158 samples from April 25, 2001, through November 6, 2005. PMF runs were first conducted using the 158 days of STN data only to ensure results were similar to those previously found with the 565 sample runs. The factors obtained from the 158 sample PMF runs were similar to those obtained with the full set of 565 samples with slight differences in average mass (**Figure 3-30**). A steel factor was not identified in the 158 sample run, instead a mixed industrial factor of various metals was obtained. The major components of

mass (ammonium sulfate, ammonium nitrate, and mobile sources), however, were apportioned similarly between the two data sets.
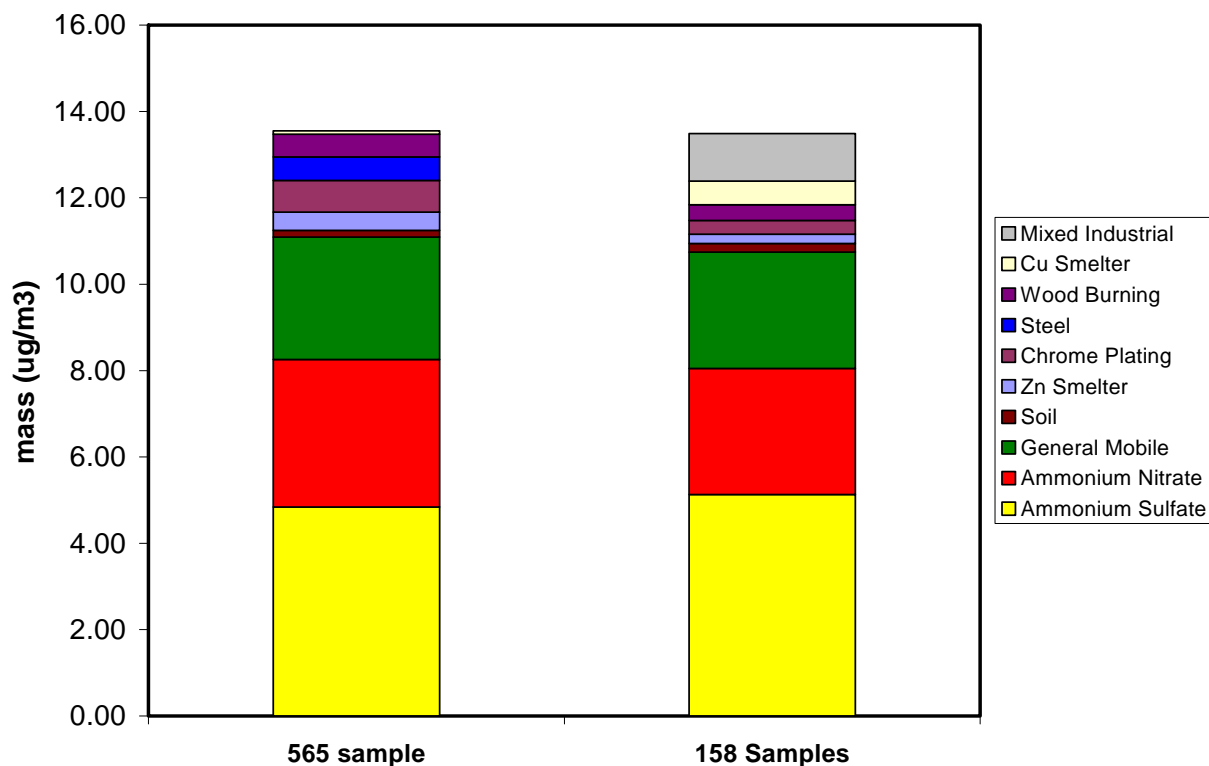
Figure 3-30. Comparison of average PMF mass contribution at Allen Park for a 9-factor 565 sample run and a 9-factor 158 sample run (April 25, 2001, through November 6, 2005).

PMF runs were conducted using 8 to 11 factors with air toxics species included. Over all runs conducted, OC and EC were not split into separate factors. Benzene, o-xylene, ethylbenzene, and toluene were grouped with the steel source (iron and chromium), while formaldehyde and acetaldehyde were grouped with the general mobile source (OC/EC) in all runs conducted. One of the expectations of using the air toxics data with STN data was that the additional species would help separate the mobile sources into gasoline and diesel factors. At Allen Park, though, no additional insight into the split of mobile sources was obtained.

4. SUMMARY AND CONCLUSIONS

Source apportionment using PMF was applied to both STN data sets and SANDWICH data sets at three Michigan STN sites. In addition, exploration of adding air toxics to $PM_{2.5}$ data for use in PMF was performed.

4.1   NEW STN UNCERTAINTIES

The application of PMF to STN data sets tested the use of recently updated uncertainties that are larger than those previously used, especially for metal species. **Table 4-1** shows a comparison of PMF results using recently updated uncertainties and previous results at Allen Park. The comparison shows good agreement in general with some differences including an increase in the number of industrial factors isolated with the new uncertainties. Additionally, separate mobile and diesel factors were no longer successfully isolated. As discussed in Section 2.1.2, uncertainties for carbon and ion species were not updated. OC and EC, therefore, have much smaller relative uncertainties than the other species used in the model. The increased uncertainty for metal species provides more flexibility in fitting these species, and may be the cause of the differences seen. Percent contributions were similar across studies; except for higher mass attributed to ammonium sulfate with the current work.

The recently updated STN uncertainties for $PM_{2.5}$ data were not available for the entire data set (only for July 2003-October 2005) and as a result, extrapolation methods were used for estimated the remaining uncertainties. PMF runs were conducted using samples collected prior to July 2003 and using samples collected from July 2003 to October 2005. The factors identified in the PMF runs, as well as the mass apportioned to each factor, were the same using both data sets. The uncertainty estimations used did not affect the solution. The methods used provide good estimates of uncertainties for PMF purposes.

Table 4-1. Comparison of PMF results using recently updated STN uncertainties and PMF results from previous studies at Allen Park. Values are the percentage of mass attributed to each source.

	Allen Park Previous Work (2002-2005)	Allen Park Current Work (2000-2005)
Ammonium Sulfate	26	36
Ammonium Nitrate	25	25
Soil	4	1
Mobile (OC)	21	21
Diesel (EC)	14	
Biomass Burning	2	4
Zinc Smelter	5	3
Chrome Plating	3	5
Copper		1
Steel		4

4.2   PMF ON STN AND SANDWICH DATA SETS

SANDWICH data were developed to adjust STN data to better match FRM data and SANDWICH has never been used for PMF purposes. PMF was run using STN data sets and SANDWICH data sets and results were compared. The major difference between the two data set results was that for the SANDWICH data, more mass was attributed to ammonium sulfate and less mass was attributed to ammonium nitrate, which is consistent with ambient data. Of the three sites considered, the same number of sources was identified using the two data sets at Luna Pier and Dearborn, while differences were found at Allen Park, due to larger changes in the carbon data between STN and SANDWICH data sets. SANDWICH data showed better mass closure overall, but there was more variability in the results on a day-to-day basis, mostly due to the differences in sulfate, nitrate, and carbon in the data sets. Overall, results found using STN data were similar to those found using SANDWICH data, and differences between the results were generally within the uncertainty of the application. Thus, at Detroit, the application of STN data is likely sufficient for most applications, though this may not hold true for all areas. A thorough comparison of STN versus SANDWICH data prior to source apportionment should be conducted in the future to understand if differences between the data sets are important enough to yield different source apportionment results.

4.3   LINKING PMF FACTORS TO SOURCES

Point source emissions of species that were representative of PMF factors found at the three STN sites were examined along with meteorological data to better understand the link between PMF factors and sources. Wind roses were developed for days on which high mass was apportioned to various PMF factors as well as for typical days at each of the three STN sites. Differences were noted between the wind roses for typical days at the sites and wind roses for

high mass days. At both Allen Park and Dearborn, typical day wind roses show winds coming almost equally from all directions. This is not found on high mass days. For example, on high zinc and manganese days, winds at Allen Park are from the northeast and winds at Dearborn are from the southwest. The winds at both sites come from the direction where industrial facilities that perform zinc smelting are located. The differences in typical and high factor mass wind roses indicate that high mass wind roses provide useful information for understanding which sources are associated with which factors. Wind roses indicate that the zinc/manganese factor at Allen Park and Dearborn is most likely associated with zinc smelting. The same analyses were performed for the other industrial factors at Allen Park, Dearborn, and Luna Pier. Understanding the meteorology of the sites provided useful information in understanding the appropriate sources to associate with PMF factors.

4.4 PMF RUNS WITH PM$_{2.5}$ DATA AND AIR TOXICS—EXPLORATORY ANALYSES

Additional PMF runs were conducted using both STN PM$_{2.5}$ data and air toxics data with the expectation of separating the mobile sources into gasoline and diesel factors. However, gasoline and diesel factors were not successfully isolated at Allen Park, even with the addition of several air toxic species. The inability to separate the mobile sources was likely caused by the limited data set. Air toxics data examined at the STN sites were mostly below detection limits and were measured with a lower frequency than the PM$_{2.5}$ data. As a result, the data set at Allen Park was reduced from 565 samples to 158 samples. Results from PMF runs with combined PM$_{2.5}$ and air toxics data may produce some insight into mobile sources at sites where more air toxics data are available.

5. REFERENCES

Anttila P., Paatero P., Tapper U., and Jarvinen O. (1995) Source identification of bulk wet deposition in Finland by positive matrix factorization. *Atmos. Environ.* (published without a date) **29** (14), 1705-1718, Pergamon Press, Ltd. (1352-2310/95).

Begum B.A., Kim E., Jeong C.H., Lee D.W., and Hopke P. (2005) Evaluation of the potential source contribution function using the 2002 Quebec forest fire episode. *Atmos. Environ.* **39**, 3719-3724.

Birch M.E. and Carey R.A. (1996) Elemental carbon-based method for monitoring occupational exposures to diesel exhaust. *Aerosol Sci. Technol.* **25**, 221-241.

Buzcu B., Fraser M.P., Kulkarni P., and Chellam S. (2003) Source identification and apportionment of fine particulate matter in Houston, TX, using positive matrix factorization. *Environmental Engineering Science* **20** (6), 533-545.

Draxler R.R. and Hess G.D. (1997) Description of the HYSPLIT 4 modeling system. Technical memorandum by NOAA, ERL ARL-224, December 24.

Eberly S. (2005) EPA PMF 1.1 user's guide. Prepared by the U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, June.

Frank N.H. (2006) Retained nitrate, hydrated sulfates, and carbonaceous mass in federal reference method $PM_{2.5}$ for six eastern U.S. cities. *J Air Waste Manag Assoc* **56** (4), 500-511.

Hafner H.R. (2005) Multivariate receptor modeling workbook. Work plan prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-905103-2736-WP, April.

Juntto S. and Paatero P. (1994) Analysis of daily precipitation data by positive matrix factorization. *Environmetrics* **5**, 127-144.

Kim E., Larson T.V., Hopke P.K., Slaughter C., Sheppard L.E., and Claiborn C. (2003) Source identification of $PM_{2.5}$ in an arid northwest U.S. city by positive matrix factorization. *Atmos. Res.* **66** (4), 291-305.

Kim E. and Hopke P.K. (2004a) Improving source identification of fine particles in a rural northeastern U.S. area utilizing temperature-resolved carbon fractions. *J. Geophys. Res.* **109** (D9).

Kim E. and Hopke P.K. (2004b) Source apportionment of fine particles at Washington, DC, utilizing temperature-resolved carbon fractions. *J. Air & Waste Manage. Assoc.* **54** (7), 773-785.

Kim E., Hopke P.K., and Edgerton E.S. (2004a) Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in positive matrix factorization. *Atmos. Environ.* **38**, 3349-3362.

Kim E., Hopke P.K., Larson T.V., Maykut N.N., and Lewtas J. (2004b) Factor analysis of Seattle fine particles. *Aerosol Sci. Technol.* **38** (7), 724-738.

Kim E., Brown S.G., Hafner H.R., and Hopke P.K. (2005a) Characterization of non-methane volatile organic compounds sources in Houston during 2001 using positive matrix factorization. *Atmos. Environ.* **39**, 5934-5946.

Kim E., Hopke P.K., Kenski D.M., and Koerber M. (2005b) Sources of fine particles in a rural midwestern U.S. area. *Environ. Sci. Technol.* (39), 4953-4960.

Kim E., Hopke P.K., and Qin Y. (2005c) Estimation of organic carbon blank values and error structures of the speciation trends network data for source apportionment. *J. Air & Waste Manag. Assoc.* **55**, 1190-1199.

Larsen R.K. and Baker J.E. (2003) Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: a comparison of three methods. *Environ. Sci. Technol.* **37** (9), 1873-1881.

Lee E., Chan C.K., and Paatero P. (1999) Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong. *Atmos. Environ.* **33**, 3201-3212.

NIOSH (1999) Method 5040: Elemental Carbon (Diesel Exhaust). In *NIOSH Manual of Analytical Methods*, 4th Ed., Issue 3, National Institute of Occupational Safety and Health, Cincinnati, OH.

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

Paatero P. (1997) Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems* **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. *Journal of Graphical Statistics* **8**, 854-888.

Paatero P., Hopke P.K., and Philip K. (2003) Discarding or downweighting high-noise variables in factor analytic models. *Anal. Chim. Acta* **490**, 277-289.

Poirot R.L., Wishinski P.R., Hopke P.K., and Polissar A.V. (2001) Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont. *Environ. Sci. Technol.* **35** (23), 4622-4636.

Polissar A.V., Hopke P.K., and Poirot R.L. (2001) Atmospheric aerosol over Vermont: chemical composition and sources. *Environ. Sci. Technol.* **35** (23), 4604-4621.

Ramadan Z., Song X.-H., and Hopke P.K. (2000) Identification of sources of Phoenix aerosol by positive matrix factorization. *J. Air & Waste Manage. Assoc.* **50**, 1308-1320.

Research Triangle Institute (2004) Extended semi-annual data summary report for the chemical speciation of $PM_{2.5}$ filter samples project. Report prepared by the Research Triangle Institute, Research Triangle Park, NC, RTI/07565/06DS. Available on the Internet at <http://www.epa.gov/ttn/amtic/files/ambient/pm25/spec/rtiqap.pdf>.

Rice J. (2004) Comparison of integrated filter and automated carbon aerosol measurements at Research Triangle Park, North Carolina. *Aerosol Science & Technology* (38), 23-36.

Song X.H., Polissar A.V., and Hopke P.K. (2001) Sources of fine particle composition in the northeastern U.S. *Atmos. Environ.* **35** (31), 5277-5286.

Subramanian R., Khlystov A.Y., Cabada J.C., and Robinson A.L. (2004) Positive and negative artifacts in particulate organic carbon measurements with denuded and undenuded sampler configurations. *Aerosol Sci. Technol.* **38**, 27-48.

U.S. Environmental Protection Agency (2002) National Emission Inventory (NEI), air pollutant emission trends, NEI emission trends data and estimation procedures, criteria pollutant data, current emission trends summaries. Available on the Internet at <http://www.epa.gov/airtrends> last accessed January 24, 2003.

Zhou L., Kim E., Hopke P.K., Stanier C.O., and Pandis S. (2004) Advanced factor analysis on Pittsburgh particle size-distribution data. *Aerosol Sci. Technol.* **38** (S1), 118-132.