Sonoma Technology, Inc.

1360 Redwood Way, Suite C
Petaluma, CA 94954-1169
707/665-9900
FAX 707/665-9800
www.sonomatech.com

APPORTIONMENT OF AIR TOXICS USING PRINCIPAL COMPONENT ANALYSIS

TECHNICAL MEMORANDUM
STI-903555-2590-TM

By:
Steven G. Brown
Hilary R. Hafner
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169

Prepared for:
Lake Michigan Air Directors Consortium
2250 East Devon Avenue, Suite 250
Des Plaines, IL 60018

August 6, 2004

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. BACKGROUND

The objectives of this investigation were to assess the feasibility of using the air toxics database to obtain meaningful source apportionment results, to investigate commonality of sources among sites, and to lay the foundation for future source apportionment of the air toxics data using more sophisticated tools. Specific policy relevant questions that were targeted included:

- What sources contribute to urban air toxics levels?
- What source apportionment methods are effective for assessing source contributions?
- Should other non-toxic species measurements be added to toxics monitoring sites to aid source apportionment?

2. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a simple multivariate analysis tool that is often used as a “first look” at the major components within a data set before using more robust source apportionment tools such as Positive Matrix Factorization (PMF). PCA extracts the principal components that explain the majority of the variance within a data set by grouping species together. These groups can then be qualitatively interpreted as individual sources or source types. A drawback to PCA is its foundation in least squares minimization, causing all data points to be weighted equally even though some species are known to have lower analytical and sampling uncertainties than others. PCA cannot utilize missing or below-detection data (a common occurrence with ambient measurements) in a meaningful manner either. Despite these limitations, PCA generally gives similar results on a qualitative level to more sophisticated source apportionment tools, and so is useful as an initial apportionment analysis.

3. DATA

Air toxics data from eight sites were obtained from a variety of sources, including the pilot cities database, the historical toxics database, and the Speciation Trends Network (STN). Data sources and sites are detailed in **Table 3-1**. The unique combination of sites selected met a variety of criteria:

- Sufficient data were available for a large number of species (i.e., at least one year of data available for more than 12 species),
- Different regions (Northwest, Midwest, Southeast, etc.) were represented,
- At least one pair of sites within the same urban area which allowed for analysis of intra-urban variability, and
- At least one pair of sites that allowed for analysis of urban-rural variability.

Table 3-1. Selected sites, data sources, types of species, temporal resolution, time frame, and number of samples.

Site	Data Source	Species Types	Sampling Interval	Years	Number of Samples
Detroit, MI – Allen Park	Pilot cities, STN	Gaseous, semi-volatile, $PM_{2.5}$	24-hr	2001-2002	65
Detroit, MI – 696/Lodge	Pilot cities, STN	Gaseous, semi-volatile, $PM_{2.5}$	24-hr	2001-2002	72
San Jose, CA	Historical	Gaseous, $PM_{2.5}$, PM coarse	24-hr	1996-2001	290
Minneapolis, MN	Historical	Gaseous, PM_{10}, TSP	24-hr	1996-2001	324
Wagner, MN	Historical	Gaseous, PM_{10}, TSP	24-hr	1996-1997	62
Phoenix, AZ	Historical	Gaseous	3-hr	1999-2000	341
Seattle, WA	Pilot cities, STN	Gaseous, $PM_{2.5}$	24-hr	2001-2002	70
Tampa Bay, FL – Lewis	Pilot cities, STN	Gaseous, $PM_{2.5}$, TSP	24-hr	2001	61

TSP = total suspended particulate

Since PCA cannot properly utilize below detection or missing data, only data with more than 50% of the measurements above MDL were used in this investigation. Data were normalized to minimize the influence of outliers and take into account differences in magnitude between species. Normalization was performed according to:

$$\frac{x_i - \bar{x}}{\sigma}$$

where x_i is the concentration in sample ‘i’, $\bar{x}$ is the average concentration for that parameter at that site, and σ is the standard deviation. A Varimax rotation was used to better help us interpret the results.

4. PCA RESULTS

Most of the factors identified with PCA make sense with the known or suspected emissions in a given area. For example, mobile sources were identified at each urban site, as expected. Secondary/regional, solvent use, and other industrial factors were also found. Overall, PCA provided a good “first look” into the contributing factors of air toxics, and more robust source apportionment and trajectory analysis will further identify and quantify individual source contributions.

One of the policy questions addressed in this overall project was whether nontoxic species measurements should be added to toxics monitoring sites to aid in source apportionment. We found in this preliminary assessment that supplemental collocated $PM_{2.5}$ data were vital in identifying factors. Results from each site are presented and discussed in this section.

4.1 SAN JOSE, CA

Historical data from 1996-2001 of gaseous, PM_{fine}, PM_{coarse}, and PM_{TSP} were used from San Jose. Nine factors were identified, accounting for 76% of the variance (see **Table 4-1**). A mobile source factor, identified by benzene, toluene, and 1,3-butadiene, accounted for the most variance at 25%. Fine manganese, which has been linked to diesel emissions, is also found in this factor. Road dust, characterized by TSP and coarse chromium, nickel, and manganese, accounted for 19% of the variance. Industrial factors included power plant emissions/manufacturing (mercury, arsenic), metal plating/oil combustion (fine manganese, chromium, nickel), microchip manufacturing/solvent use (methylene chloride, 1,1,1-trichloroethane) and rubber/alloy manufacturing (tin), which accounted for 6%, 8%, 5%, and 5% of the variance, respectively. Other factors include a dry cleaning/degreasing factor, a possible paint factor, and a chromium VI factor, which may be analytically biased. These factors are fairly consistent with emission inventories and the proximity of roads, point sources, and population as shown in **Figures 4-1 and 4-2**.

Table 4-1. PCA results for San Jose, CA (Jackson Street) using 24-hr samples from historical database (290 samples, 1996 to 2001).

Factor	% Variance	Species	Likely Source
1	24.7	1,3-Butadiene; Mn fine; Toluene; Tetrachloroethylene; o-Xylene; Styrene; Acetaldehyde; Ethylbenzene; 2-Butanone; Chloroform; Benzene; Formaldehyde	Mobile sources
2	18.5	Pb TSP; Ni TSP; Mn coarse; Cr TSP; Ni coarse; Cr coarse; Mn TSP	Road dust
3	6.0	Hg TSP; As TSP	Power plant/manufacturing
4	8.1	Mn $PM_{2.5}$; Cr $PM_{2.5}$; Ni $PM_{2.5}$; Se TSP	Metal plating/oil combustion
5	4.5	Methylene chloride; 1,1,1-Trichloroethane	Microchip manufacturing/solvent
6	4.6	Cr VI	Analytical bias?
7	4.5	Sb TSP	Rubber/alloy manufacturing
8	4.9	Pb $PM_{2.5}$; Pb coarse	Paint?
9	4.7	Trichloroethylene; p-Dichlorobenzene	Dry cleaning/degreasing

4.2 SEATTLE, WA

Pilot city and STN data at Seattle (Beacon Hill) of gaseous and $PM_{2.5}$ species from 2001-2002 were used. Eight factors were found, accounting for 77% of the variance (**Table 4-2**). A mobile source/road dust factor including benzene, organic carbon (OC), elemental carbon (EC), and 1,3-butadiene accounted for the most variance at 22%. A general industrial factor with sulfate, nickel, manganese, and chromium accounted for 13% of the variance. Soil, sea salt, and regional/secondary formation factors were also found. An unidentified factor with methylene chloride, acetaldehyde, and zinc accounted for 6% of the variance. Two factors attributed to industrial fuel composition accounted for 5% of the variance each and consisted of arsenic/vanadium and selenium/cadmium. The variety of industrial factors is consistent with the shipping and industry in the Seattle area (**Figures 4-3 and 4-4**).

Table 4-2. PCA results for Seattle, WA (Beacon Hill), using 24-hr samples from pilot cities and STN (70 samples, 2001 to 2002).

Factor	% Variance	Species	Likely Source
1	21.9	As $PM_{2.5}$; Cu $PM_{2.5}$; Benzene; Fe $PM_{2.5}$; OC $PM_{2.5}$; Pb $PM_{2.5}$; EC $PM_{2.5}$; Tetrachloroethylene; Trichloroethylene; Si $PM_{2.5}$; 1,3-Butadiene	Mobile sources/road dust + urban background
2	13.2	Fe $PM_{2.5}$; Cr $PM_{2.5}$; Sulfate $PM_{2.5}$; Ni $PM_{2.5}$; Mn $PM_{2.5}$	Industrial
3	9.1	Fe $PM_{2.5}$; Al $PM_{2.5}$;Total Nitrate $PM_{2.5}$; Ca $PM_{2.5}$	Soil
4	9.1	Na Ion $PM_{2.5}$; Mg $PM_{2.5}$; Cl $PM_{2.5}$	Sea Salt
5	7.2	NH_4 $PM_{2.5}$; Formaldehyde	Regional/secondary
6	6.4	Methylene chloride; Acetaldehyde; Zn $PM_{2.5}$	??
7	4.9	As $PM_{2.5}$; V $PM_{2.5}$	Industrial fuel combustion?
8	4.6	Se $PM_{2.5}$; Cd $PM_{2.5}$	Industrial fuel combustion?

4.3 MINNEAPOLIS, MN

Historical gaseous, PM_{10} and TSP data from 1996-2001 at Minneapolis were used. Six factors accounting for 71% of the variance were found (**Table 4-3**). A mobile source factor identified using benzene, toluene, and xylenes accounted for 21% of the variance. A secondary factor containing aldehydes accounted for 16%. Other factors included a possible solvent use factor, a lead source, a paint/enamel use factor, and a degreasing factor. Point source emissions of $PM_{2.5}$ and VOCs are shown in **Figures 4-5 and 4-6**. In general, these factors make sense given the species and potential sources in the area.

Table 4-3. PCA results for Minneapolis, MN (300 Nicollet Mall) using 24-hr samples from the historical database (324 samples, 1996 to 2001).

Factor	% Variance	Species	Likely Sources
1	20.6	Benzene; Ethylbenzene; Styrene; Toluene; o-Xylene	Mobile sources
2	15.5	Acetaldehyde; Propanal; Mn PM_{10}; Formaldehyde	Secondary
3	15.4	Tetrachloroethylene; Chlorobenzene; 1,1-Dichloroethane; 1,1,1-Trichloroethane; p-Dichlorobenzene; Chloroform	Solvent use?
4	7.1	Pb TSP; Pb PM_{10}	Lead source
5	6.6	Methylene chloride; Co PM_{10}	Paint/enamel use/production
6	6.1	Trichloroethylene	Degreasing

4.4 WAGNER, MN

Historical gaseous, PM_{10} and TSP data from 1996-2001 at Wagner, MN were used. This site is in rural Minnesota outside of Minneapolis. Eight factors were identified, accounting for 93% of the variance (**Table 4-4**). Two urban transport signatures, similar to factors seen in Minneapolis, were found, accounting for 26% and 12% of the variance, respectively. An additional secondary urban factor (aldehydes) accounting for 10% of the variance was also seen, demonstrating the impact of urban sources on this rural site. A mobile source factor accounted for an additional 11% of the variance. Other factors include paper production, fungicide/insecticide use, paint emissions, and global carbon tetrachloride background. Very few point sources are near Wagner (**Figures 4-7 and 4-8**), but these include polymer and paper production. Farm use of fungicides and insecticides is common nearby. The identified factors appear consistent with expected emissions.

Table 4-4. PCA results at Wagner Township (Sandstone), MN using 24-hr samples from historical database (62 samples, 1996 to 1997).

Factor	% Variance	Species	Likely Sources
1	25.8	o-Xylene; Styrene; Ethylbenzene; Bromomethane; p-Dichlorobenzene; Chlorobenzene; Ethylene dichloride	Urban transport?
2	12.4	Mn PM_{10}; Formaldehyde; Trichloroethylene	Similar to Minneapolis factor, transport?
3	11.2	o-Xylene; Toluene; Pb PM_{10}; Ethylbenzene; Benzene	Mobile sources
4	10.9	1,1,1-Trichloroethane; 1,1-Dichloroethane	Paper production?
5	9.7	Propanal; Acetaldehyde	Secondary urban formation
6	9.5	Methylene chloride; Tetrachloroethylene	Paint?
7	8.1	Bromomethane; Co PM_{10}	Fungicide/insecticide
8	5.5	Carbon tetrachloride	Global background

4.5 PHOENIX, AZ

Three-hour gaseous data from the historical database collected between 1999-2000 at Phoenix were used. Three factors, accounting for 63% of the variance, were found (**Table 4-5**). This small number of factors reflects the small number of species with sampling levels above the detection limit. A secondary aldehyde factor accounted for 27% of the variance and is consistent with enhanced photochemistry in the hot, arid desert area of Phoenix. A mobile source factor accounted for 23% of the variance and a solvent/degreaser factor (i.e., hexane, o-xylene) accounted for another 16% of the variance. These emissions are consistent with the species used and likely nearby emissions (**Figures 4-9 and 4-10**).

Table 4-5. PCA results at Phoenix Supersite, AZ, using 3-hr samples from historical database (341 samples, 1999 to 2000).

Factor	% Variance	Species	Likely Source
1	26.6	2-Butanone; Formaldehyde; Acetaldehyde; Propanal	Secondary
2	23.1	Ethylbenzene; Benzene; Toluene; 2,2,4-Trimethylpentane	Mobile sources
3	15.5	Hexane; o-Xylene	Solvent/degreaser

4.6 TAMPA BAY, FL

At Tampa Bay, gaseous, TSP and $PM_{2.5}$ data from the pilot cities and STN from 2001 were used. Eleven factors were identified, accounting for 81% of the variance (**Table 4-6**). A regional aerosol/transport factor and a regional mobile source dominated factor were identified, accounting for 11% and 16% of the variance, respectively. A mobile source factor contributed 10% of the variance. Other factors included solvent/degreaser emissions, coal combustion, oil combustion, incinerator emissions, soil, sea salt, and metal plating industrial emissions. This variety of sources is consistent with the wealth of emissions in the Tampa Bay area (**Figures 4-11 and 4-12**).

Table 4-6. PCA results at Tampa Bay (Lewis), FL using 24-hr samples from pilot cities and STN (61 samples, 2001 only).

Factor	% Variance	Species	Likely Source
1	15.8	NO_3 $PM_{2.5}$; Mn $PM_{2.5}$; Fe $PM_{2.5}$; Ca $PM_{2.5}$; Al $PM_{2.5}$; Mn TSP	Regional mobile?
2	11.3	K $PM_{2.5}$; SO_4 $PM_{2.5}$; Pb TSP; Cu $PM_{2.5}$; OC $PM_{2.5}$	Regional aerosol/transport
3	9.8	EC $PM_{2.5}$; OC $PM_{2.5}$; Pb TSP; Benzene; Ni $PM_{2.5}$; NH_4 $PM_{2.5}$	Mobile sources
4	7.1	Cd TSP; Be TSP	Industrial (aerospace, batteries)
5	6.6	Methylene chloride; Chloroform	Solvent/degreaser
6	6.0	Se $PM_{2.5}$; Si $PM_{2.5}$; As $PM_{2.5}$	Coal combustion
7	5.9	Formaldehyde; Acetaldehyde; V $PM_{2.5}$; EC $PM_{2.5}$	Secondary/Oil Combustion
8	5.5	Cl $PM_{2.5}$; Na $Pm_{2.5}$	Sea salt
9	4.5	Hg $PM_{2.5}$; Zn $PM_{2.5}$	Incinerator
10	4.2	Ti $PM_{2.5}$; Mg $PM_{2.5}$	Soil
11	4.0	Cr $PM_{2.5}$	Metal Plating

4.7 DETROIT, MI

Data from two sites in Detroit were analyzed, both with gaseous, semi-volatile and $PM_{2.5}$ data from 2001-2002 as part of the pilot cities and STN programs. At Allen Park, seven factors accounting for 86% of the variance were found (**Table 4-7**). A mobile source factor accounted for 39% of the variance. Other factors included smelter emissions, industrial emissions, road dust, regional aerosol, and smoke/fossil fuel usage. At 696/Lodge, eight factors accounted for 82% of the variance (**Table 4-8**). A mobile source factor accounted for the most variance at 23%. Other identified factors included auto manufacturing, oil combustion, soil/road dust, smelter, regional aerosol, and industrial emissions. The variety of sources is fairly consistent with the proximity of mobile and point sources (**Figure 4-13**).

Table 4-7. PCA results at Detroit (Allen Park), MI using 24-hr samples from pilot cities and STN (65 samples, 2001 to 2002).

Factor	% of Variance	Key Species	Likely Source
1	22.9	Benzene, aldehydes, toluene, naphthalenes, trimethylbenzenes, propene, 1,3-butadiene, As $PM_{2.5}$	Mobile sources
2	15.8	Xylenes, ethylbenzene, naphthalenes, chloromethane, OC $PM_{2.5}$	Mobile sources, industry?
3	13.8	Mn, Pb, Zn, and Fe $PM_{2.5}$; methyl ethyl ketone (MEK)	Industrial smelter
4	10.6	NH_4, NO_3, and SO_4 $PM_{2.5}$; MEK	Secondary background
5	8.2	n-octane, trimethylbenzenes, chloromethane	Industrial?
6	6.5	EC, OC, K $PM_{2.5}$	Combustion, wood smoke?
7	8.1	La, Si, Ca, Ni $PM_{2.5}$	Soil/road dust

Table 4-8. PCA results at Detroit (696/Lodge), MI using 24-hr samples from pilot cities and STN (72 samples, 2001 to 2002).

Factor	% of Variance	Key Species	Likely Source
1	23.3	Xylenes, benzene, toluene, acetylene, trimethylbenzenes, ethylbenzene, propene, 1,3-butadiene	Motor vehicle
2	12.4	Zn, NO_3, EC, K, Fe, OC, Cd $PM_{2.5}$; 2-methylnaphthalene	Regional/local aerosol
3	10.6	Aldehydes, dichlorobenzene, As $PM_{2.5}$	Industrial
4	6.0	Ni, Cr, La $PM_{2.5}$	Industrial
5	8.4	SO_4, NH_4 $PM_{2.5}$	Secondary aerosol
6	7.9	Pb, Mn, As, and Cd $PM_{2.5}$; carbon tetrachloride	Industrial – smelter? Coal?
7	5.3	Phenanthrene, Be $PM_{2.5}$	Industrial
8	8.2	Ca, Si, Mn $PM_{2.5}$	Soil/road dust

In previous work (Battelle Memorial Institute and Sonoma Technology, 2003), preliminary assessment of the feasibility of using PMF was performed using the Detroit data. **Tables 4-9 and 4-10** summarize the results of PCA and PMF. These tables illustrate that PCA does reasonably well in identifying likely factors but that in general, PMF provides more detail and more highly resolved factors.

Table 4-9. Comparison of PCA and PMF factors for Allen Park using the 2001 pilot city data.

Allen Park		
Factor	PCA	PMF
Mobile source	✓	✓
Secondary	(✓)	✓
Regional aerosol	(✓)	✓
Industrial	✓*	✓
Smelter/coal	✓	✓
Other	Soil Combustion	Diesel Road salt

(✓) = not separable

* = two or more factors

Table 4-10. Comparison of PCA and PMF factors for 696/Lodge using the 2001 pilot city data.

696/Lodge		
Factor	PCA	PMF
Mobile source	✓	✓
Secondary	✓	✓
Regional aerosol	✓	✓
Industrial	✓*	✓*
Other	Soil	Diesel Regional OC Regional Summer

* = two or more factors

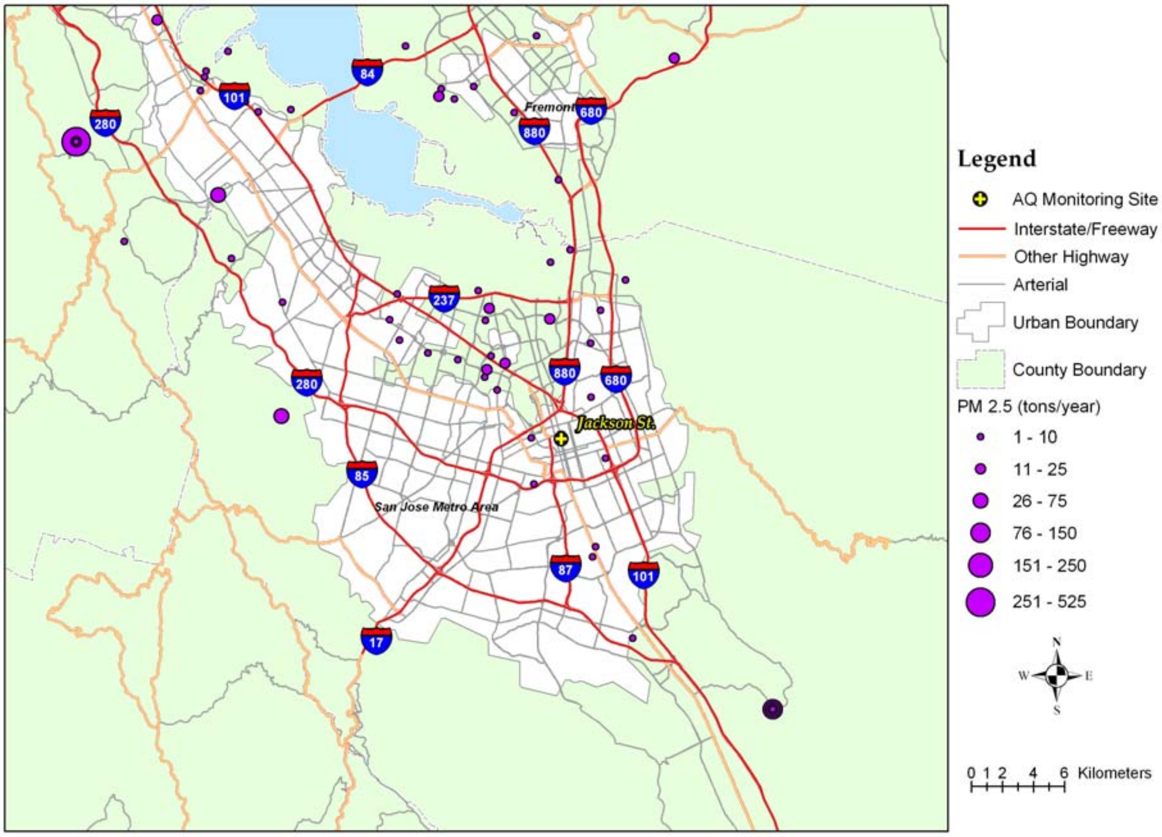

Figure 4-1. $PM_{2.5}$ emissions in the San Jose, CA area (U.S. Environmental Protection Agency, 1999).

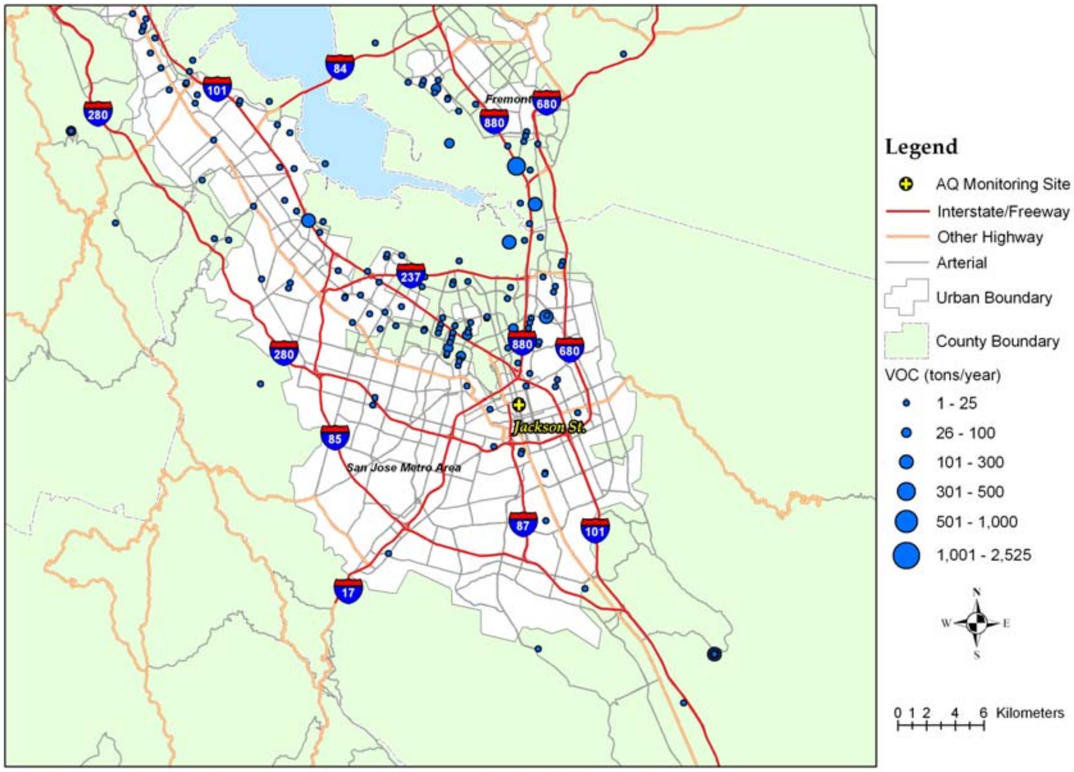

Figure 4-2. VOC emissions in the San Jose, CA area (U.S. EPA, 1999).

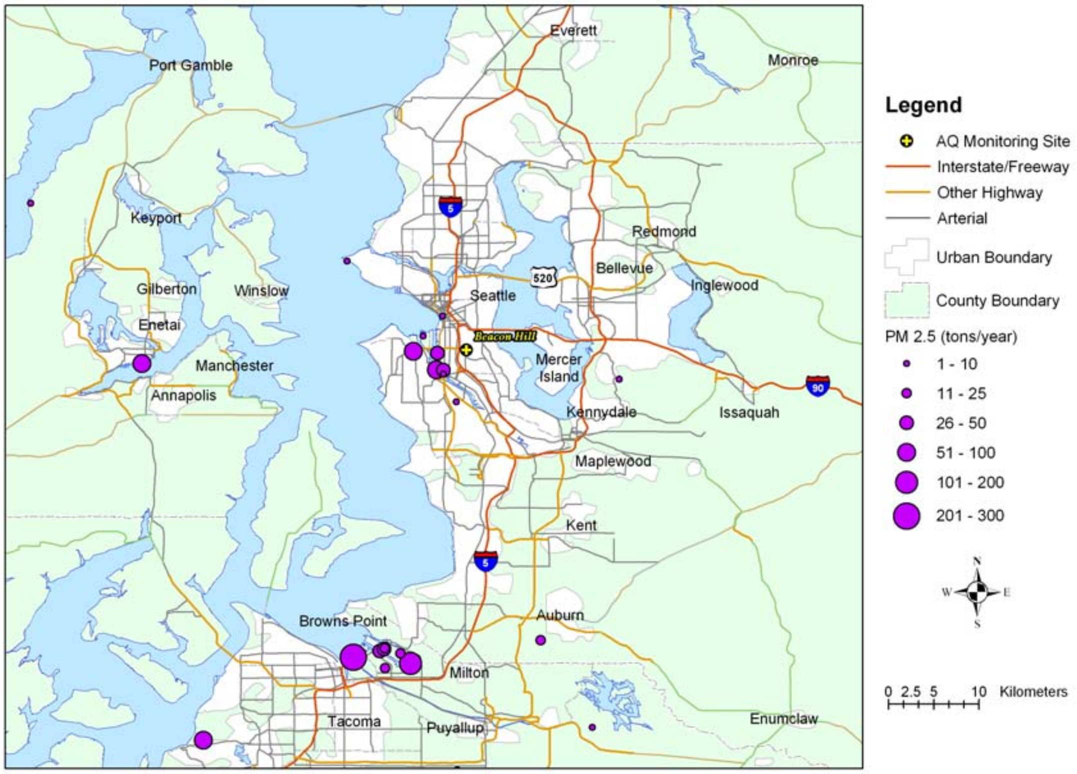

Figure 4-3. $PM_{2.5}$ emissions in the Seattle area near the Beacon Hill site (U.S. EPA, 1999).

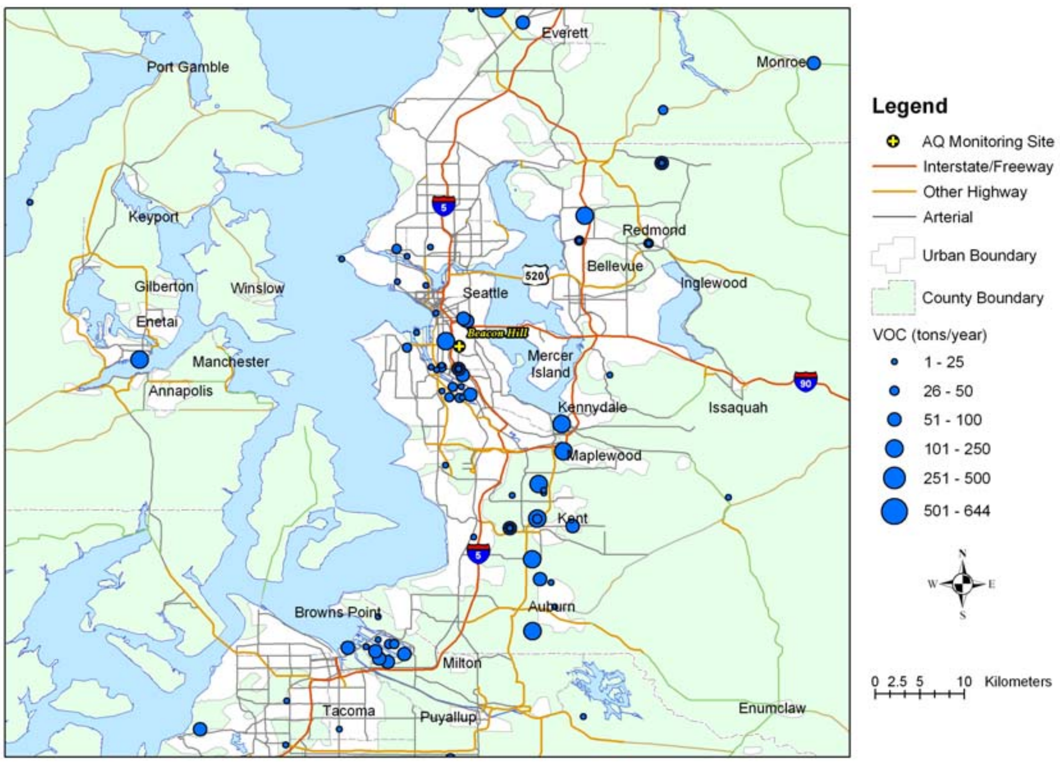

Figure 4-4. VOC emissions in the Seattle area near the Beacon Hill site (U.S. EPA, 1999).

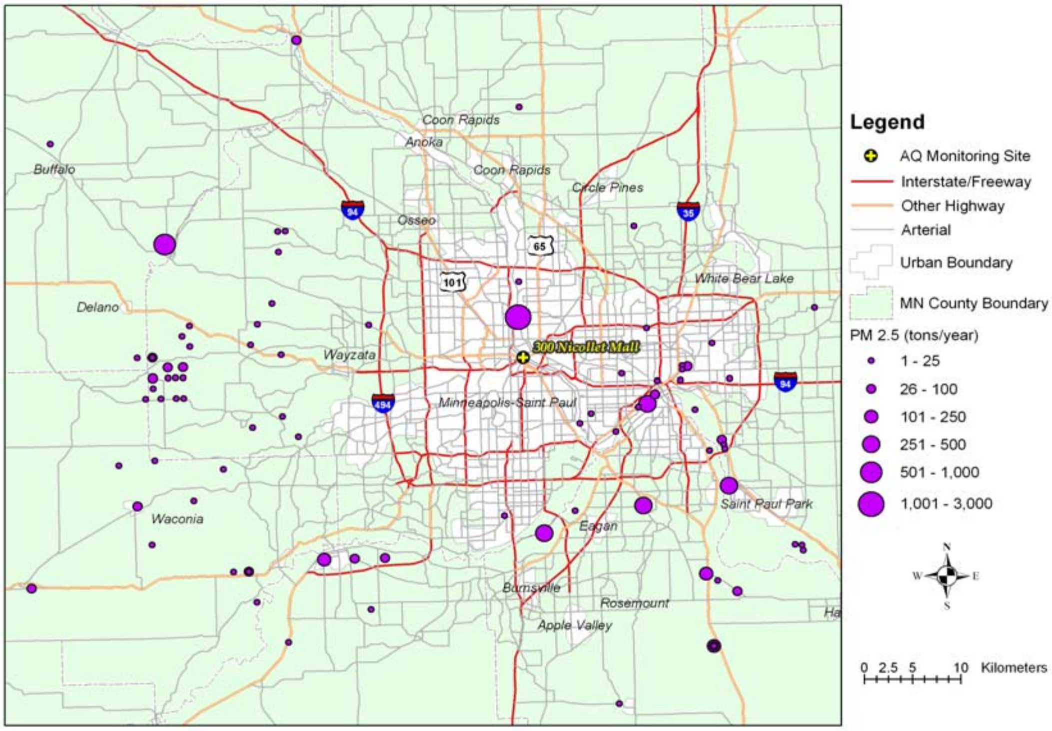

Figure 4-5. $PM_{2.5}$ emissions in the Minneapolis area (U.S. EPA, 1999).

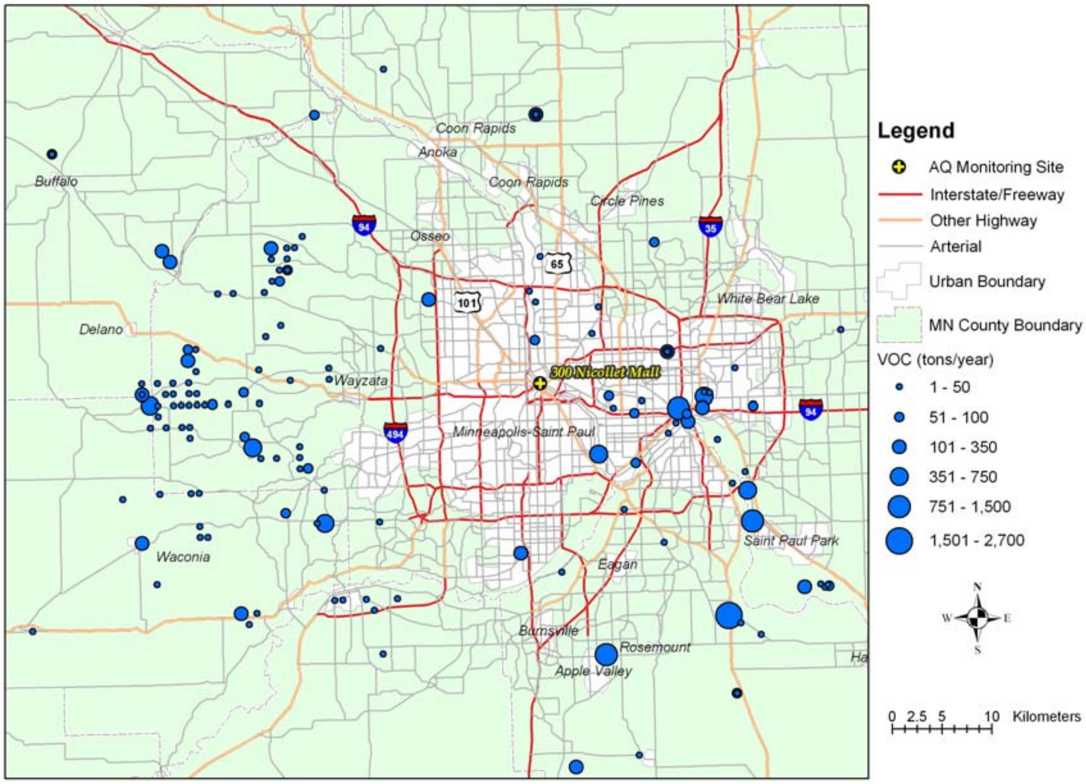

Figure 4-6. VOC emissions in the Minneapolis area (U.S. EPA, 1999).

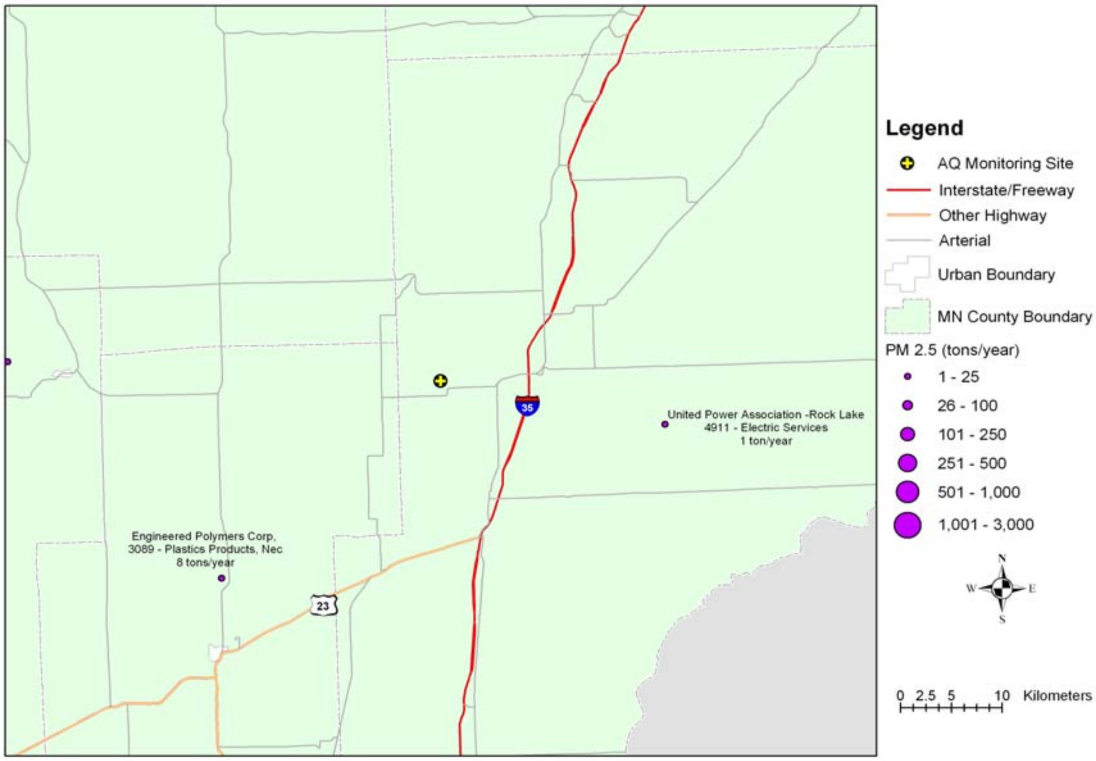

Figure 4-7. $PM_{2.5}$ emissions in the Wagner area (U.S. EPA, 1999).

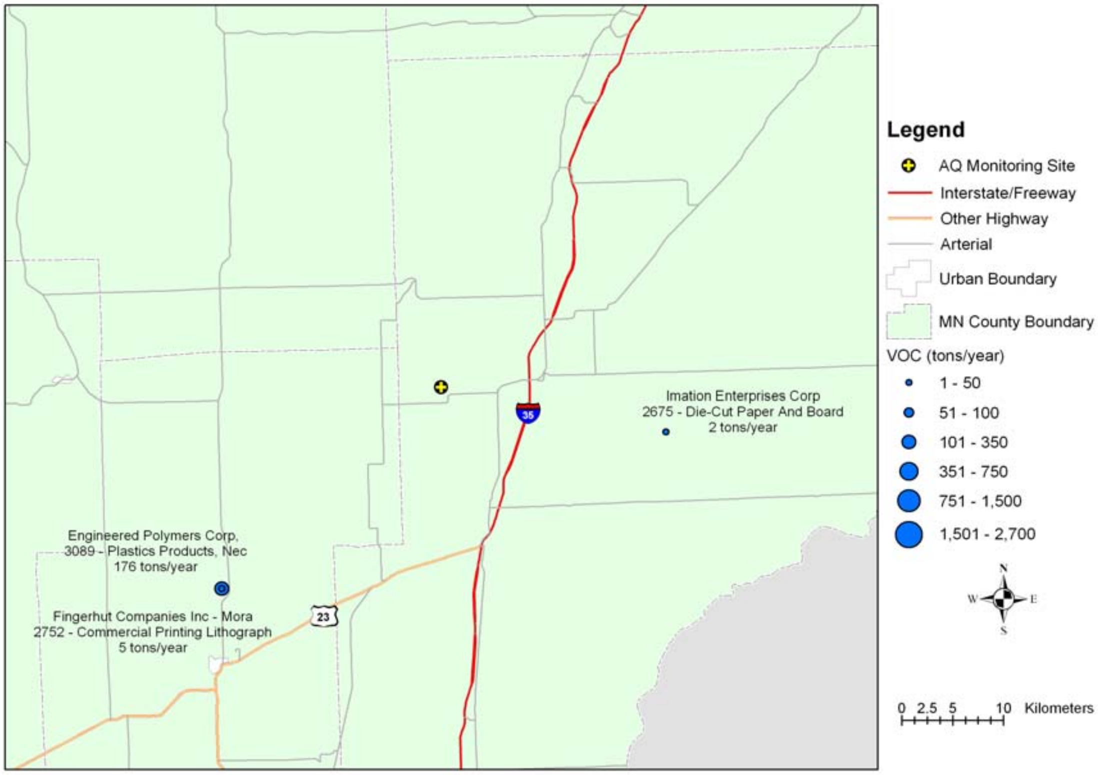

Figure 4-8. VOC emissions in the Wagner area (U.S. EPA, 1999).

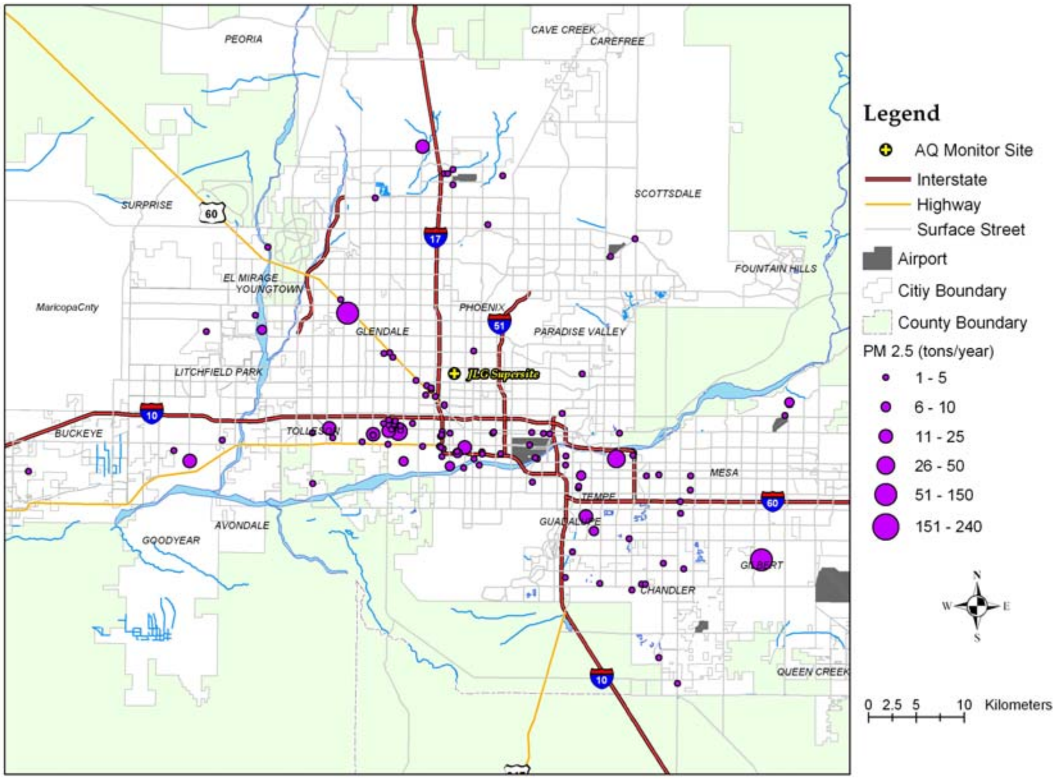

Figure 4-9. $PM_{2.5}$ emissions in the Phoenix area (U.S. EPA, 1999).

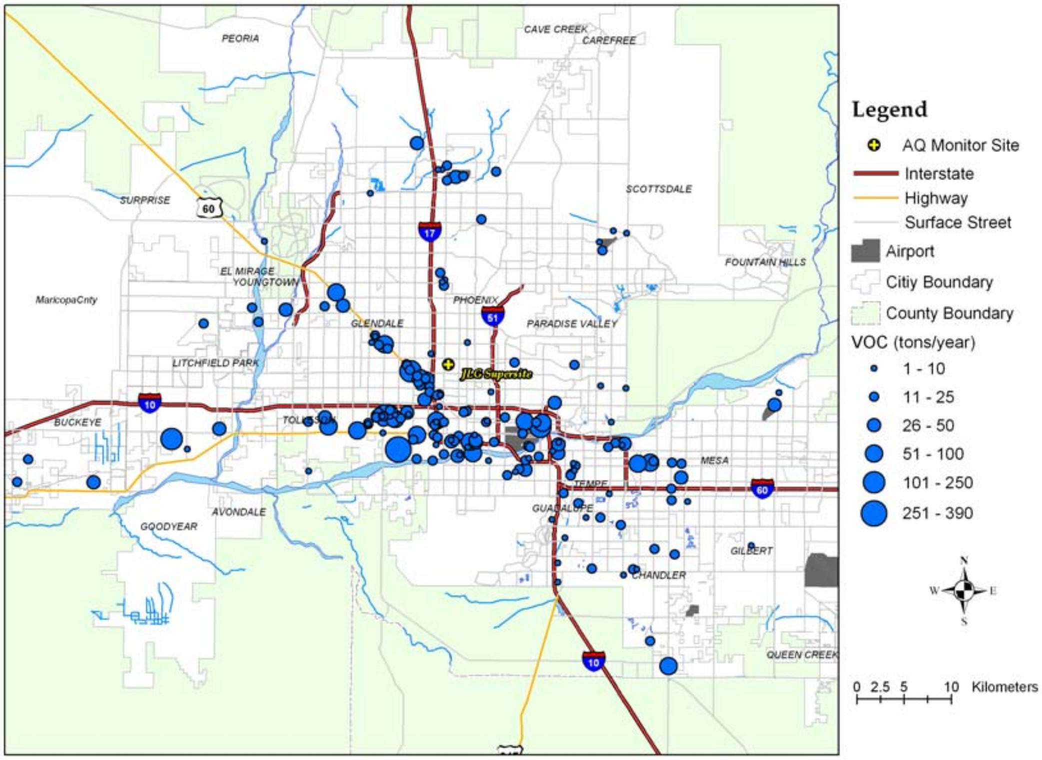

Figure 4-10. VOC emissions in the Phoenix area (U.S. EPA, 1999).

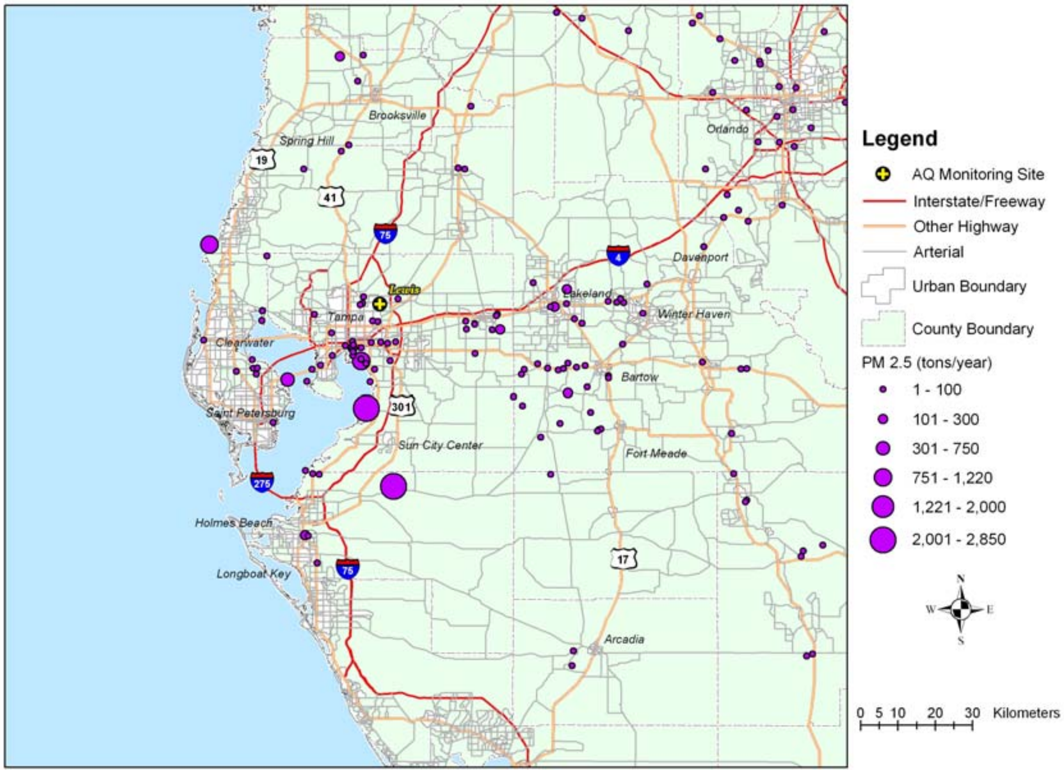

Figure 4-11. $PM_{2.5}$ emissions in the Tampa Bay area (U.S. EPA, 1999).

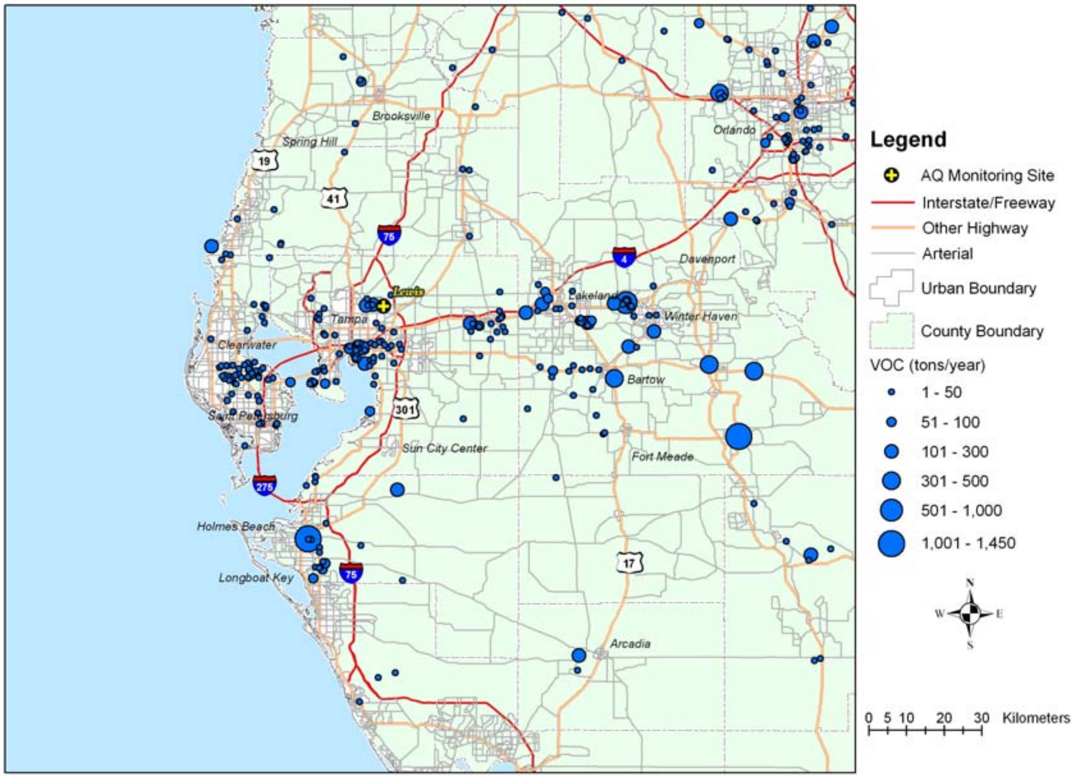

Figure 4-12. VOC emissions in the Tampa Bay area (U.S. EPA, 1999).

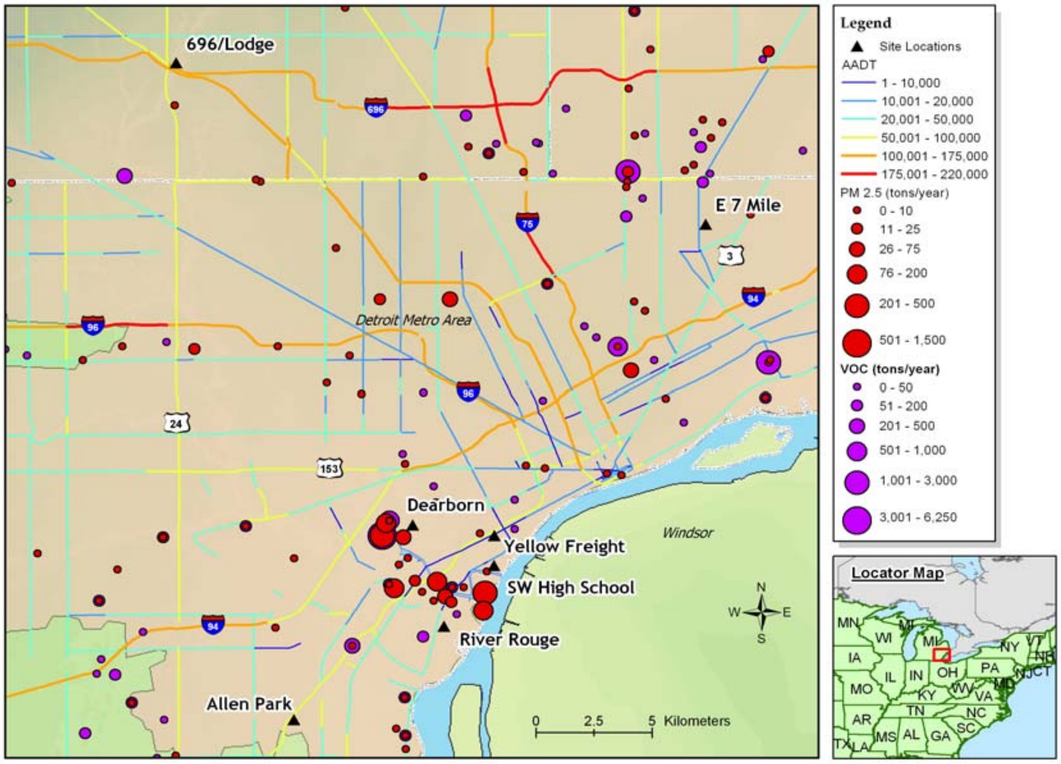

Figure 4-13. $PM_{2.5}$ and VOC emissions in the Detroit area (U.S. EPA, 1999).

5. CONCLUSIONS

PCA was successfully applied to air toxics and $PM_{2.5}$ data from eight sites. The factors observed at each site are summarized in **Table 5-1**. A mobile source factor was observed at each site, consistent with the species measured, proximity of mobile sources to each site, and prevalence of mobile source air toxics in the inventory. Regional secondary gaseous/aerosol factors were also found at all but one site, and industrial factors were found that were consistent with known emission sources.

Table 5-1. Factors identified at each site and the percent of variance explained by each factor.

Factors/Source Types	% of Variance						
	San Jose	Seattle	Minneapolis	Wagner	Detroit (Allen Park)	Tampa Bay	Phoenix (3-hr)
Mobile	25	22	21	11	23	10	23
Secondary/regional		7	16	47*	11	27	27
Misc. industrial	21*	13	7		24*	7	
Smelter or incinerator					14	5	
Combustion					7	12*	
Soil/road dust	19	9			8	4	
Solvent/dry clean	5		21*	11		7	16
Paint?	5		7	10			
Metal plating	8					4	
Sea salt		9				6	
Unknown		16*		8			
Global background				6			

* Two or more factors

This demonstration of PCA showed the following:

- Supplementing the air toxics data with additional collocated measurements such as speciated $PM_{2.5}$ data may improve source apportionment results.
- Leveraging toxics data with data from other networks or expanding the number of species analyzed for in the toxics methodology improves source apportionment efforts.
- Measurements of continuous data would further improve source apportionment analyses.
- Understanding of the range of industrial sources and signatures observed would be improved through acquisition of local knowledge and further investigation of the emission inventory.
- Combining data analysis (including further investigation of species relationships and meteorological influences, and trajectory analysis) and the application of a more

quantitative apportionment tool (e.g., PMF) is required to assess source contributions. PCA alone is insufficient to effectively apportion air toxics among sources.

6. REFERENCES

Battelle Memorial Institute and Sonoma Technology, Inc. (2003) Phase II air toxics monitoring data: analyses and network design recommendations. Final technical report prepared for Lake Michigan Air Directors Consortium, Des Plaines, IL by Battelle Memorial Institute, Columbus, OH, and Sonoma Technology, Inc., Petaluma, CA, December.

U.S. Environmental Protection Agency (1999) 1999 National Emission Inventory Documentation and Data. Available on the Internet at <http://www.epa.gov/ttn/chief/net/1999inventory.html> ; last accessed August 2004.