Sonoma Technology, Inc.

POLICY-RELEVANT LESSONS LEARNED FROM PHASE III AIR TOXICS ANALYSES

**WHITE PAPER
STI-903553-2585-WP**

**By:
Hilary R. Hafner
Michael C. McCarthy
Neil J. M. Wheeler
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169**

**Michael Koerber
Lake Michigan Air Directors Consortium
2250 East Devon Avenue, #250
Des Plaines, IL 60018-4511**

**Prepared for:
Lake Michigan Air Directors Consortium
2250 East Devon Avenue, #250
Des Plaines, IL 60018-4511**

August 10, 2004

ABSTRACT

Lessons Learned from Analyzing Air Toxics Data

The purpose of this document is to summarize what we have learned from three rounds of analyzing air toxics monitoring data (i.e., historical state and local data collected between 1990 and 2000, and the 10-city pilot data collected during 2001-2002). The first two rounds were intended to answer questions related to the design of the national air toxics monitoring network (Bortnick et al., 2001; Bortnick et al., 2003). The third round was intended to answer policy-relevant questions about data quality, usefulness, and applicability to policy decision-making (Hafner et al., 2004).

Based on this data analysis project, several important findings should be noted:

- A nationally consistent monitoring network is needed with common sampling and analysis procedures, a common set of compounds, and common quality assurance and data reporting. Monitoring objectives for this network include assessing trends and characterizing community-scale concentrations. To this end, the U.S. Environmental Protection Agency (EPA) has established a 22-site network, (known as the National Air Toxics Trends Sites), which includes 15 urban and 7 rural sites across the country, and is supporting a number of community-scale monitoring projects.

- Air toxics data are available from almost 800 sites in the historical data base and 37 sites in the pilot city data base, but the quality of these data is mixed. Confidence in the data varies by species, with high confidence for benzene, acetaldehyde, formaldehyde, manganese, and nickel; low confidence for acrolein, beryllium, and chromium VI, and improved confidence for others, such as lead and methylene chloride. (Improved measurement methods are needed to increase the data quality for these species.) Comparison of typical concentration values with cancer benchmarks indicates that the following species are the high risk drivers: benzene, acetaldehyde, carbon tetrachloride, formaldehyde, and arsenic.

- Concentrations for many species are comparable on a national scale, although some local variation may exist (e.g., benzene, formaldehyde, acetaldehyde, carbon tetrachloride). In many cities, typical regional profiles for air toxics are dominated by mobile source-related species. In some industrial areas, however, high local concentrations of individual species can occur near major point sources. Air toxics also vary temporally: by season (e.g., acetaldehyde and formaldehyde are higher in summer, and benzene is higher in winter); by day-of-week (e.g., diesel particulate-related compounds are lower on weekends); and by hour of day (e.g., formaldehyde is higher midday).

- Mobile and point source control programs have been successful in reducing air toxics concentrations. Examples include reductions in benzene and 1,3-butadiene concentrations due to implementation of reformulated gasoline, and reductions in tetrachloroethylene (and other compounds) due to implementation of MACT standards for various industries.

EXECUTIVE SUMMARY

Lessons Learned from Analyzing Air Toxics Data

Two data analysis efforts were funded in 2001-2003 to guide the design of a national air toxics monitoring program (Battelle Memorial Institute and Sonoma Technology, 2003; Bortnick et al., 2001; Bortnick et al., 2003). In Phase III, air quality analyses were performed using the historical (Phase I) and pilot city (Phase II) databases with a primary goal of addressing policy-relevant questions about data quality, usefulness, and applicability to policy decision-making. A secondary goal was to describe the process of answering these questions to national, regional, state, and local data analysts so that analyses will continue, particularly at the state level. This document summarizes the findings from the 2004 analyses in the context of policy questions developed for Phase III analyses (Hafner et al., 2004). Not all of the questions could be fully addressed within the scope of the project because of the data quality, data availability, and the type and location of data collection reflected in the historical and pilot city databases. However, useful data analysis approaches are provided to the stakeholders to further mine the data.

Data Quality and Availability

The current analyses focused on 18 hazardous air pollutants (HAPs) from almost 800 sites with at least one validated annual average in the historical data base (1960s–2000) and 37 sites in the pilot city data base (2001 and 2002). Data validation included automated data "cleaning", retaining as much data as possible, creating defensible average values, and applying relevant flags. The final database reflects considerable variation in the number of samples by species (e.g., lead total suspended particulate [TSP] the greatest number, acrolein the least) and in the site locations (i.e., most sites are urban and many regions of the country are poorly represented).

Confidence in the data varies by pollutant, with high confidence for some species, such as acetaldehyde, benzene, formaldehyde, manganese, and nickel (i.e., those with median concentrations well above MDLs); low confidence for others, such as acrolein, beryllium, chromium VI, and vinyl chloride (i.e., those with median concentrations close to or below MDLs); and improved confidence for others, such as lead and methylene chloride. To put the measured concentrations in perspective, the interquartile ranges (IQR, 25th–75th percentile) of the historical and pilot city data were determined and compared to the MDLs, background levels, and cancer benchmarks; the pilot city data, MDLs, background, and cancer benchmarks are provided in **Figure ES-1**.
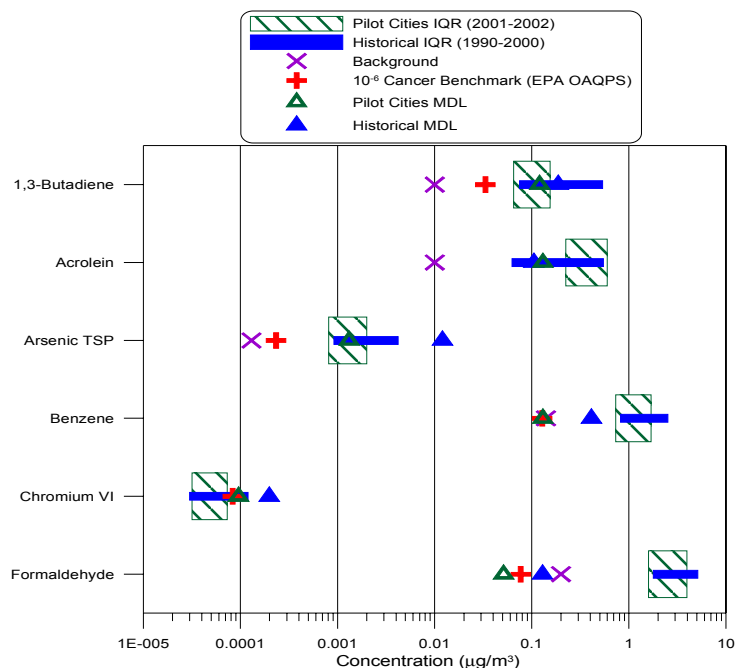
Figure ES-1. Typical air toxics levels: range of pilot city seasonal averages, MDLs, background, and cancer benchmarks for risk driver species.

The typical urban ambient data range exceeds the cancer benchmarks for acetaldehyde, formaldehyde, 1,3-butadiene, benzene, carbon tetrachloride, arsenic, and chromium. All species are well below noncancer reference concentrations except for acrolein (not shown in Figure S-1 for simplicity). The typical urban concentration ranges for most volatile organic compounds (VOC) and metals are about an order of magnitude higher (or more) than the background concentrations. Comparing the MDLs to cancer benchmarks shows that, for example, chromium measurement techniques need improvement in order to quantify cancer risk levels (i.e., the benchmark concentrations are at or below the current MDLs).

The available data are not sufficient to fully understand urban/rural differences for most pollutants. Other networks, such as Photochemical Assessment Monitoring Stations (PAMS), Interagency Monitoring of Protected Visual Environments (IMPROVE), and the Speciation Trends Network (STN), are needed to provide additional, primarily rural, concentrations. In addition, available supplemental information (e.g., site descriptions) in the historical database is not sufficient for the investigation of concentration spikes (or "hot spots"); more information is needed, including siting criteria and observations, monitoring objectives, emission source and activity information, and "local knowledge". The planned data collection and documentation process for the National Air Toxics Trends Site (NATTS) program needs to be more thorough than previous efforts.

National and Local Concentration Levels

Air toxics concentrations vary spatially. A number of case studies were performed to assess spatial variability. As an example, benzene concentrations show relatively little variation (factor of 2-3) across the United States as shown in **Figure ES-2** using 1999 data as an example.
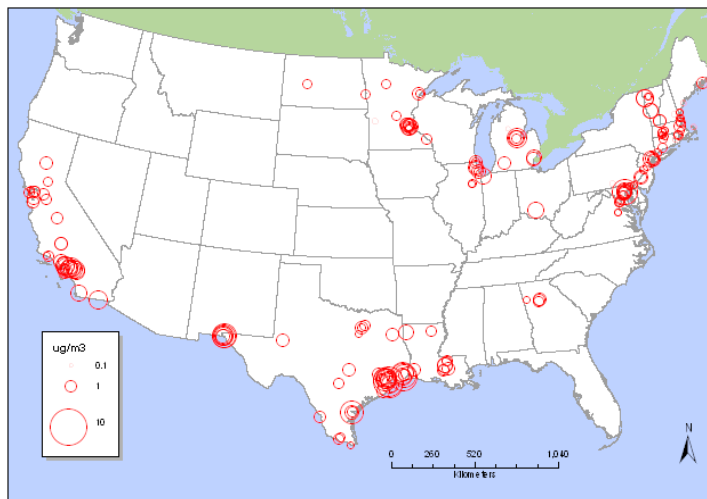

Figure ES-2.   Summer average benzene concentrations (1999). Note that not all available air toxics data are present in the historical database.

Over the period 1990–2000, a few spatial and temporal benzene concentration hot spots were identified. Examination of high concentrations in northwest Indiana, for example, showed the influence of a nearby coke oven facility (**Figure ES-3**).
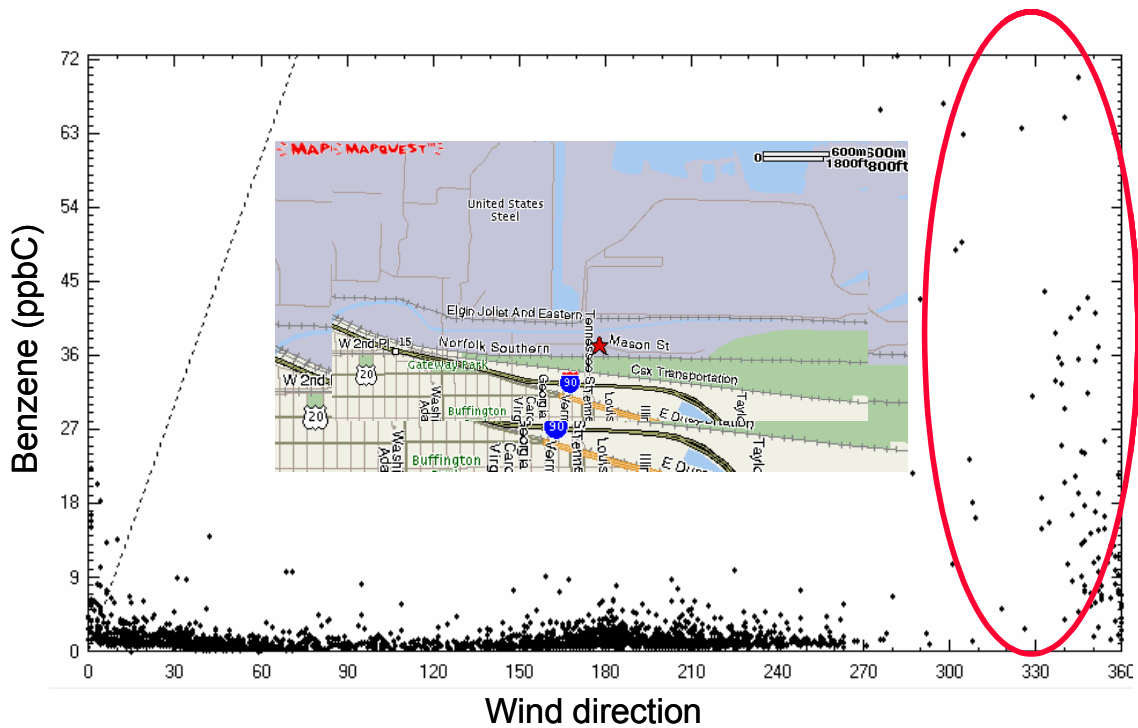
Figure ES-3. One-hour benzene concentrations in Gary, Indiana, by wind direction. The monitor location (★) relative to the coke oven facility (United States Steel) is shown in the insert.

Examination of urban-scale concentrations showed that average concentrations of species with lifetimes greater than a few hours or dominated by area source emissions vary by less than a factor of three on the urban scale, whereas those with shorter lifetimes or dominated by local point sources can vary by a factor of 10 or more. Also, fingerprints for cities and regions reflect similarities due to similar emission sources (e.g., motor vehicles) and some differences due to industrialization and monitor placement. To determine how many monitors are needed to capture citywide average air toxics concentrations, consideration should be given to the importance of point sources and species' residence times. To capture exposures near roadways, other studies have shown a significant spatial gradient suggesting the need for several monitors located at varying distances.

Concentration data from small networks can be compared to historical and pilot city IQRs, MDLs, and background concentrations to understand how the data from small network sites fit into a broader urban context (e.g., Figure S-1). If concentrations differ significantly from typical urban ranges, analysts could follow case study examples to investigate the possible causes of high concentrations. Extrapolation for most air toxics beyond the urban scale is not recommended without a network of rural measurements capable of capturing gradients between urban and rural areas.

Effectiveness of Control Programs

Air quality improvements due to emission reductions have been measured. For example, benzene and 1,3-butadiene concentrations have declined due to the use of reformulated gasoline (RFG), but formaldehyde concentrations may have increased (**Figure ES-4**).
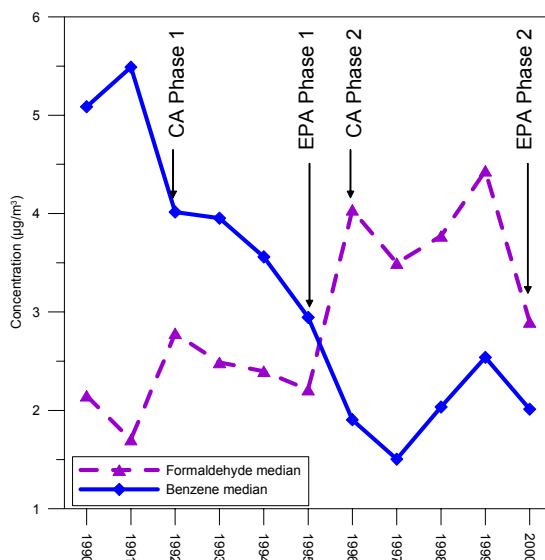

Figure ES-4.   Median seasonal average benzene and formaldehyde concentrations ($\mu g/m^3$) at all sites in California. RFG formulation changes are noted.

Other examples of decreasing trends include carbon tetrachloride, due to a phase-out in production by 1995 in accordance with Title VI of the Clean Air Act, and tetrachloroethylene, due to conformity with maximum achievable control technology [MACT] standards in the mid-1990s.

An initial statistical analysis was performed for several pilot cities (Seattle, Washington; Detroit, Michigan; and Tampa, Florida) and other cities (San Jose, California; Phoenix, Arizona; and Minneapolis and Wagner, Minnesota) to identify source contributions to air toxics concentrations. Some source types are common across these cities, including mobile sources, secondary/regional transport, solvent use, soil/road dust, combustion, and sea salt at coastal sites. To provide more definitive information about source contributions, a combination of other information (e.g., trajectories) and more quantitative apportionment methods should be considered.

Urban Air Toxics Modeling

A limited evaluation of urban-scale modeling techniques was conducted using pilot city data from Detroit, Michigan; Seattle, Washington; and Cedar Rapids, Iowa. Model inputs were prepared using National Weather Service meteorology and the U.S. Environmental Protection

Agency's (EPA) 1999 National Toxics Inventory. Comparisons of annual average modeled and monitored concentrations showed reasonable agreement for most VOCs, and generally poor agreement for most metals, which were often underpredicted (see **Figure ES-5**). When secondary formation was not considered, carbonyl compounds were underpredicted but at levels generally consistent with what was expected from primary emissions (see Figure ES-5c). However, carbonyl compound predictions for Seattle were better than average, which may indicate that the monitors were dominated by local emissions with little contribution due to secondary formation. Figure ES-5 also shows a tendency for underprediction for all species in Cedar Rapids.
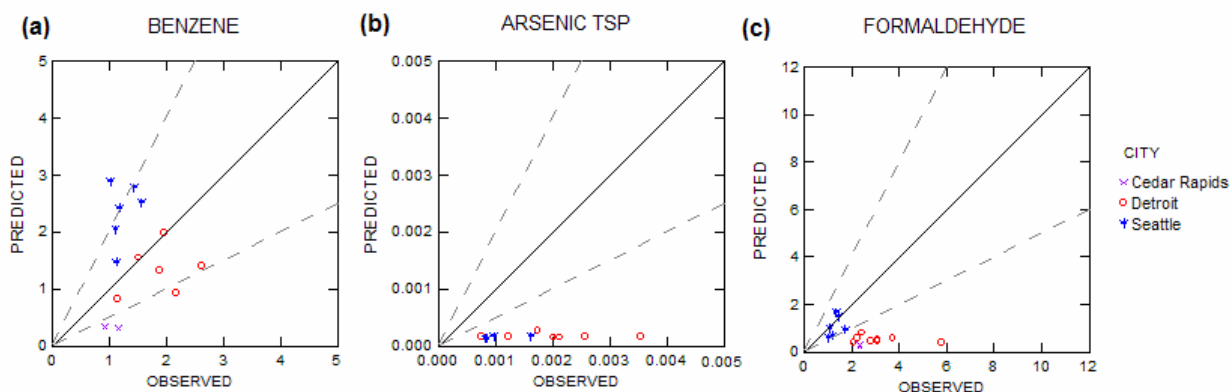

Figure ES-5.   Scatter plots of modeled v. measured concentrations ($\mu g/m^3$) for
(a) benzene, (b) arsenic TSP, and (c) formaldehyde.

When monitored values were compared to the best-fit modeled values from receptors within 4.5 km of the monitoring site, model performance improved for most species, which was to be expected in areas with significant spatial variations in emissions. Of the species considered risk drivers, arsenic had the poorest model performance, which did not improve significantly when the best-fit modeled values were used.

The simplified approach for estimating secondary production of HAPs using the Ozone Isopleth plotting Program for Research (OZIPR) model was found to be inadequate in some areas because it does not consider where a monitor is located relative to areas of precursor emissions. This can lead to significant overestimation of secondary HAPs at some locations. In Seattle, for example, the OZIPR estimated secondary formaldehyde and acetaldehyde concentrations alone were three and ten times the observed concentrations, respectively. Because these species are considered risk drivers, care should be taken to accurately model their secondary formation.

STI recommended that before modeling is used as a planning tool, efforts should be made to improve air toxics emissions inventories, better integrate physics and chemistry at different spatial scales, and improve meteorological inputs to air toxics models.

**POLICY-RELEVANT LESSONS LEARNED
FROM PHASE III AIR TOXICS ANALYSIS**

Two data analysis efforts were funded in 2001-2003 to guide the design of a national air toxics monitoring program (Battelle Memorial Institute and Sonoma Technology, 2003; Bortnick et al., 2001; Bortnick et al., 2003). While these analysis efforts used historical (Phase I) and pilot city (Phase II) air toxics databases to answer some policy-relevant questions, the analyses were not designed to answer the questions. In Phase III, air quality analyses were performed using the historical and pilot city databases with a primary goal of addressing policy-relevant questions about data quality, usefulness, and applicability to policy decision-making. The current analyses focused on 18 hazardous air pollutants (HAPs) from almost 800 sites with at least one validated annual average in the historical data base (1960s–2000) and 37 sites in the pilot city data base (2001). A secondary goal was to describe the process of answering these questions to national, regional, state, and local data analysts so that analyses will continue, particularly at the state level. The following discussion summarizes the findings from the 2004 analyses in the context of policy questions developed for Phase III analyses (Hafner et al., 2004). The questions are grouped into broad categories. Not all of the questions could be fully addressed within the scope of the project because of the data quality, data availability, and the type and location of data collection reflected in the historical and pilot city databases. However, we have attempted to provide useful data analysis approaches that can be used by other researchers to further mine the data. Broad policy-relevant questions in this white paper are followed by specific questions and the findings from Phase III analyses.

Can air toxics data be used to address policy-relevant questions?

Can policy questions be addressed with the available and planned data?

- Available data for a dozen air toxics provide a sufficiently long record (i.e., 10 years) at a large number of sites with which to investigate temporal and spatial (primarily urban) variability. For example, seasonal and annual averages from the historical database were used to identify declines in urban concentrations of benzene and 1,3-butadiene, and increases of formaldehyde from the introduction of reformulated gasoline (RFG).

- The available data are not sufficient to fully understand urban/rural differences for most pollutants. Other networks, such as Photochemical Assessment Monitoring Stations (PAMS), Interagency Monitoring of Protected Visual Environments (IMPROVE), and the Speciation Trends Network (STN), are needed to provide additional, primarily rural, concentrations. In addition, the available supplemental information (e.g., site descriptions) in the historical database is not sufficient for the investigation of concentration spikes (or hot spots); more information is needed, including siting criteria and observations, monitoring objectives, emission source and activity information, and "local knowledge". The planned data collection and documentation process for the National Air Toxics Trends Site (NATTS) program needs to be more thorough than previous efforts.

1

- Analyses discussed in this white paper indicate that planned data collection should include non-toxic species to facilitate data validation, transport assessment, and source apportionment efforts. Also, because diesel particulate matter (DPM) concerns policy makers, Phase I and II analyses indicated the need for measurements of potential DPM markers in addition to black carbon (BC) and the standard air toxics suite of species to provide sufficient information to start understanding the portion of particulate matter (PM) attributable to diesel emissions.

- In Phase III, the historical database has been further validated, and supplemental data added, to facilitate future analyses.

What is our confidence in the data?

- The historical database is sufficiently large that the central tendencies and overall patterns in air toxics concentrations can be confidently assessed for some species.   For the "big picture" of air toxics concentrations, anomalies, such as hot spots, spikes at individual sites, and differences in the minimum detection limit (MDL), are mitigated by the sheer magnitude of information. However, the historical database is missing information that is needed to improve our confidence in individual site data (e.g., site descriptions, sampling frequency, accurate proximate emission activity information, etc.).

- Confidence in air toxics data is a function of the pollutant, sampling and analysis methods, and analysis objectives. In general, there is more confidence in air toxics concentrations that are well above the MDL than those near or below the MDL. Our confidence in seasonal and annual averages can be summarized using comparisons of measured concentrations to MDLs and to cancer benchmark concentrations. **Table 1** shows an example approach for the applicability of the 18 target HAPs data to be used in risk assessment. Looking at the extremes in this approach, (1) if the median concentration and cancer benchmark of an air toxic are well above its MDL, then the health risk from that pollutant can be quantified, but (2) if both the median concentration and cancer benchmark are below the MDL, then all we really know is the upper limit of concentration; a lower MDL will be needed to quantify the annual average and to qualitatively assess risk. Note that even data not useful for risk assessment may be useful for trends (and potentially other) analysis. This approach can be used by analysts to address data quality and appropriate use of the data from their own cities.

- The information in Table 1 also suggests the need for better measurement methods for several compounds, including chromium VI and arsenic. Testing of new chromium VI sampling and analysis methods are underway at selected NATTS sites.

- The criteria for determining "good" need to be selected based on the data analysis objectives. For example, is the data precision for the historical air toxics data set good enough to assess risk? The MDL decision matrix (Table 1) shows that, for the pilot city data, only cadmium $PM_{2.5}$, cadmium TSP, chromium VI, tetrachloroethylene, chloroform, 1,3-butadiene, and arsenic $PM_{2.5}$ are not known well enough to confidently estimate the cancer risk associated with these species. For the other 18 core species, the typical cancer risk can be estimated with at least some confidence.

Table 1. MDL decision matrix using 2001 pilot city data and cancer benchmarks from Integrated Risk Information System (IRIS) (U.S. Environmental Protection Agency, 2003). Columns describe the ratio of the cancer benchmark (CB) to the MDL for a given species and rows classify the ratio of the median concentration to the MDL.

	CB/MDL<1	CB/MDL = 1 to 10	CB/MDL >10
Median/MDL<1	A lower MDL is needed to be able to quantify the annual average and qualitatively assess risk: *Cadmium PM$_{2.5}$, TSP* *Chromium (VI)* *Tetrachloroethylene* *Chloroform* *1,3-Butadiene* *Arsenic PM$_{2.5}$*	Upper limit of cancer risk is 10^{-6}. MDL is sufficient to estimate maximum risk: *Vinyl chloride* *Trichloroethylene*	Upper limit of risk is 10^{-7}. MDL is sufficient to determine that the upper limit of risk is small: *Beryllium PM$_{10}$*
Median/MDL= 1 to 10	Cancer risk is on the order 10^{-5}. MDL is sufficient to estimate risk: *Arsenic TSP* *Carbon tetrachloride* *Chromium PM$_{2.5}$, TSP*	Cancer risk is on the order of 10^{-6}. MDL is sufficient to estimate risk: *Beryllium TSP* *Methylene Chloride* *Chromium PM$_{10}$* *Benzene*	Cancer risk is $<10^{-6}$. MDL is sufficient to determine that upper limit of risk is small: *Cadmium PM$_{10}$* *Lead PM$_{2.5}$, TSP*
Median/MDL >10	Cancer risk is quantifiable and $>10^{-5}$. MDL is sufficient to quantify risk: *None*	Cancer risk is quantifiable and $>10^{-6}$. MDL is sufficient to quantify risk: *Acetaldehyde* *Formaldehyde*	Cancer risk is quantifiable and on the order of 10^{-6}. MDL is sufficient to quantify risk: *Lead PM$_{10}$* *Nickel PM$_{10}$*

Which toxics species are adequately represented in the database (e.g., national/regional or urban/rural coverage)?

- Some air toxics are data-rich while some are sparse. Spatial maps of annual average concentrations show that, for example, lead total suspended particulate (TSP) has the best

spatial and temporal coverage in the historical database, that formaldehyde has moderate coverage, and acrolein is poorly represented (**Figure 1**).
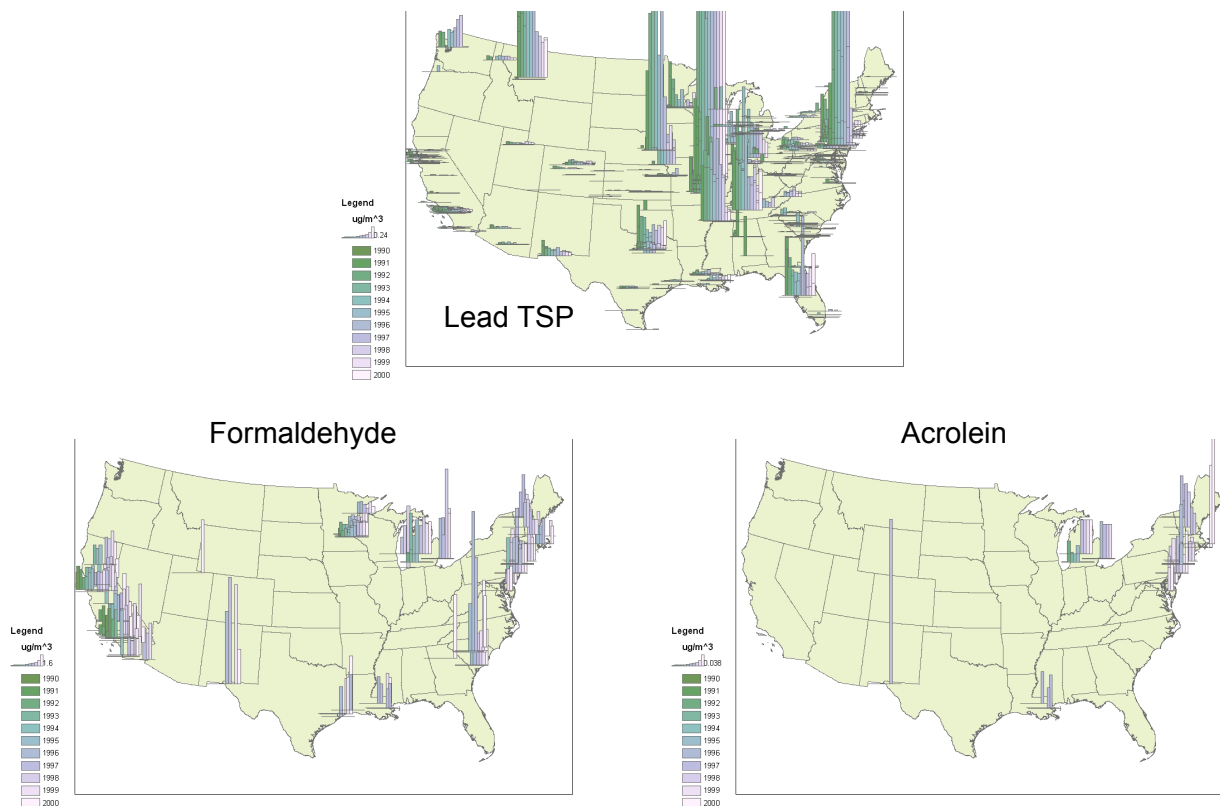

Figure 1.   Valid annual average concentrations ($\mu g/m^3$) from the historical database by year for 1990-2000 for the United States.

- At the national/regional level, only a few areas of the country are well-represented. Monitoring sites represented in the historical database are heavily concentrated in urban areas in California, Texas, Michigan, Minnesota, and the Northeast corridor (e.g., see **Figure 2**). The available data are likely adequate for capturing the concentrations for the bulk of the urban population in these states, but care needs to be taken in applying conclusions from analyses of data in these areas to other parts of the country because of differences among the states with respect to emissions sources, emissions controls, climate, meteorology, etc.

- Urban coverage for air toxics is good while rural coverage is very poor. Most of the data in the historical database during 1990-2000 are from urban monitors; there are few rural sites with data. Of the 4027 sites in the historical database with at least one valid sample, fewer than 250 sites could be considered rural (i.e., have a population density of <100 persons per square mile and are not within an urban city designation). Only a fraction of those sites have available data at any given time. Only 800 sites provided at least one valid annual average.

- Figure 2 illustrates that not all available air toxics data are present in the historical database. As noted in Phases I and II, not all the PAMS, air toxics, or special study/community study data were available in the Aerometric Information Retrieval System (AIRS) or the Air Quality System (AQS) at the time the data set was compiled. Missing data will likely be addressed in the next phase of analyses.
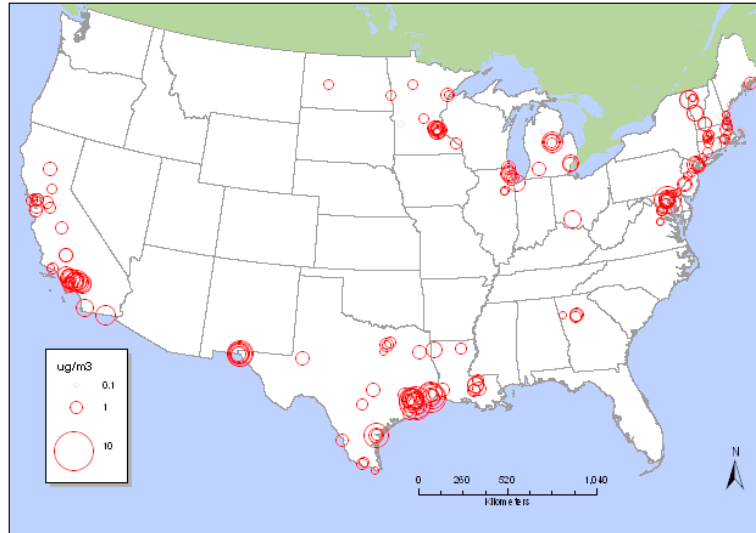

Figure 2.   Summer seasonal average benzene concentrations ($\mu g/m^3$) from 1999 available in the historical database.

How should missing data and data below detection levels be treated, and how do different data treatments affect data analysis and modeling results?

- Earlier work (Battelle Memorial Institute and Sonoma Technology, 2003; Bortnick et al., 2001; Bortnick et al., 2003) thoroughly treated the statistical veracity of MDL substitutions or treatment of missing data.

- In Phase III analyses, seasonal and annual averages were created with multiple flags to warn analysts of the quantity of data below the MDL. Data values below the MDL were retained; where null values for data below detection were provided, MDL/2 was used as the substitution. The flags (i.e., <25% of samples above the MDL, 25% to 50% of samples above the MDL, etc.) can then be used by analysts to decide which data meet their analysis objectives. An example of how the MDL flags apply to the data, and the effect of MDL substitution on our confidence in the data, is illustrated in **Figure 3**.

Species	%MDL1 >75% above MDL	%MDL2 50-75% above MDL	%MDL3 25-50% above MDL	%MDL4 0-25% above MDL	%MDL5 No MDL available
Formaldehyde	85.4	1.4	0.0	1.4	11.9
Benzene	83.6	6.6	1.8	3.7	4.4
Manganese (tsp)	51.1	1.6	5.4	41.7	0.2
Chromium (coarse)	41.5	47.4	2.2	8.9	0.0
Methylene chloride	39.2	11.4	14.6	33.4	1.4
1,3-Butadiene	31.8	22.9	24.5	20.6	0.2
Lead (tsp)	31.0	4.5	3.7	3.9	56.9
Cadmium (tsp)	25.3	6.3	10.9	56.8	0.7
Carbon tetrachloride	24.9	8.4	12.3	24.4	30.0

Figure 3. Example of how seasonal and annual average values in the historical database were flagged to indicate the amount of data below the MDL and how the flags relate to our confidence in the data.

- The effect of MDL substitution treatment on data analysis varies depending on the goal of the analysis. In assessing interannual trends in concentrations, for example, averages with >50% of measurements below MDL will be biased and unusable for quantification. However, these annual averages can be used to assess interannual trends qualitatively if the MDL has not significantly changed over the time period of interest.

How do we quantify uncertainty in the data analysis results?

- Table 1 provides guidance for interpreting and using data in relation to the MDL and cancer benchmarks. Depending on the analysis objective, quantification of uncertainty may not be necessary (or possible) for species with a large percentage of measurements below MDL. In those cases, only a qualitative answer may be possible. We know that data less than the MDL or near the MDL are more uncertain than concentrations well above the MDL.

- Precision and accuracy information was not investigated in this work. Additional analysis and discussion is needed to understand how to use precision, accuracy, and bias information to quantify uncertainty in air toxics.

Are data useful even if they do not meet quality assurance/quality control (QA/QC) and data completeness criteria?

- Data need to meet basic QA/QC measures to be useful. If contamination was known, instrument flow rates were out of specification, standard operating procedures were not met, etc., then data need to be invalidated. In the historical database, data were

6

invalidated when units were not specified, site coordinates were missing, etc. However, we retained as much data as possible to meet a range of analysis needs. For example, data for which the MDL was not documented may still be qualitatively useful and were retained; however, we do not know the uncertainty associated with these values.

- Data completeness could not be assessed with the historical database because sample frequency information was unavailable for >90% of measurements for the compilation of seasonal and annual averages. Using criteria developed for this project, seasonal averages appear to give reasonable and comparable results to those reported from the pilot city study where this information was available. Clearly, sample frequency information is a critical component of future data sets; consideration should be made to update the existing historical database with sample frequency.

- Although >75% data completeness is sufficient to ensure representativeness for a daily or seasonal average, this is not the only possible method to classify representative data. In the absence of frequency information, we prepared annual and seasonal averages by setting/checking sample spacing, monitoring campaign duration, and absolute sample numbers. For example, data completeness does not compare the absolute number of measurements used to compile a seasonal average. If a seasonal average with >75% completeness for 1-in-12-day sampling is representative, a seasonal average with >25% completeness for daily sampling can be more representative because it would have more measurements (i.e., 7 vs. 22 samples) as long as they are adequately spaced throughout the measurement period.

Is more detailed "standard metadata" needed to better define the specific, micro-scale characteristics of air toxics monitoring sites?

- The case study analyses showed that more detailed information, including local knowledge of monitoring sites near emission sources and control implementation (and schedule) actually applied to sources, is necessary and desirable. Proximity to local point sources, local fugitive emissions data, changes in local emission control strategies, and wind direction are all needed to understand spikes in concentrations. In the Phase III analysis, extensive investigation was required to track down the details of site characteristics and monitoring objectives to help explain data anomalies; these details were not readily available in the current data set. **Table 2** summarizes desirable metadata.

- In Phase III, some site metadata were added to the historical and pilot city database to begin to address the need for more detailed information including distance to roads, railroads, airports, and National Emission Inventory (NEI) volatile organic compound (VOC) sources; magnitude of traffic or emissions, number of tracks, or departures from each source; and emissions information from the 1996 National Toxics Inventory including the magnitude of emissions from major point sources, area sources, and off-road sources. However, sites did not always map correctly (i.e., incorrect coordinates), nor did the NEI databases always contain local sources of air toxics (e.g., **Figure 4**). These issues need to be addressed with local knowledge of individual sites, QC of site locations, and investigation into potential sources. Also information on plant closures, major changes in operation, or the addition of controls is needed to facilitate trends analyses.

- Monitor siting (and description) guidance akin to the national criteria pollutant guidance is desirable (e.g., definitions of microscale, neighborhood scale, etc.). The guidance will need to be pollutant group-specific (i.e., PM_{10}, TSP, reactive gaseous, long residence time gaseous, etc.).

- The addition of other species, either measured as part of the toxics program or measured as part of other programs collocated with toxics monitors, is useful to validation and necessary to source apportionment efforts.

Table 2.   List of desirable metadata to accompany air toxics.

| Metadata	Usefulness
QA/QC information from the field and laboratory: • standards • calibration technique and frequency • duplicate/replicate results • laboratory intercomparison results • example calculations • precision and accuracy	Allows analysts to better understand and quantify the uncertainty associated with the measurements
Monitor site information: • monitoring objectives • unusual events or changes near the site • presence of other monitors at the site • instrument model, inlets, dryers, denuders, sampler modifications	Used to interpret data and model results
Proximity information: • distance to nearby point sources, roads, railroads, airports • annual emissions, traffic counts, number of tracks, or departures from each source • population density • local fugitive emissions data	Used for source apportionment, outlier analysis, and understanding model performance
Control strategy implementation information: plant closures, major changes in operation, or the addition of controls changes	Assists analysts in quantifying and understanding trends, and understanding model performance.
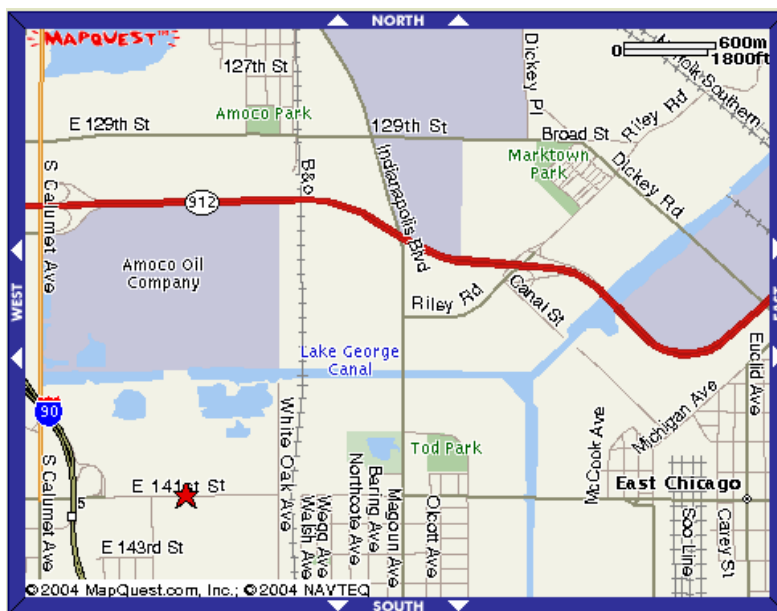
Figure 4. Example of useful metadata. A monitoring site is shown as a star and the nearest VOC emission source in the NEI is the Phillips Petroleum pipeline 2 km from the site. However, a quick scan of MapQuest also shows an Amoco Oil facility within 600 m of the site that was not in the NEI.

What are the air toxics concentration levels nationally and locally?

What does a broad national assessment say about air toxics concentrations across the country?

- The historical database (from 1990 to 2000) lacks data for the Pacific Northwest, the Rockies, the Southwest, the Southeast, and parts of the Midwest.   Urban monitoring sites located in urban centers in California, Texas, the Northeast, and the Great Lakes region are better represented.

- In California and the Northeast, typical regional profiles for air toxics were dominated by mobile source-related species. In contrast, monitoring sites in more industrial cities in the Midwest (e.g., Detroit) and Texas (e.g., El Paso, Houston areas) often showed marked differences in concentrations with high concentrations of individual species near major sources (e.g., benzene near a coke oven or metals near a smelter). Spatial maps at the national through local level were a useful method for visually identifying atypically high concentrations for a given species (e.g., Figures 1 and 2).

- Typical concentrations for urban sites were put in the context of cancer benchmarks, background concentrations, and chronic reference concentrations for inhalation (RfC) values in **Figures 5 and 6**. These figures show which species are risk drivers (i.e., the species representing the largest portion of total risk), typical urban ranges of air toxics (i.e., interquartile ranges – IQRs), and the contribution of remote background concentrations to urban concentrations. Of the 18 HAPs studied, risk drivers are arsenic,

9

acetaldehyde, formaldehyde (depending on the benchmark), benzene, carbon tetrachloride, and likely 1,3-butadiene.
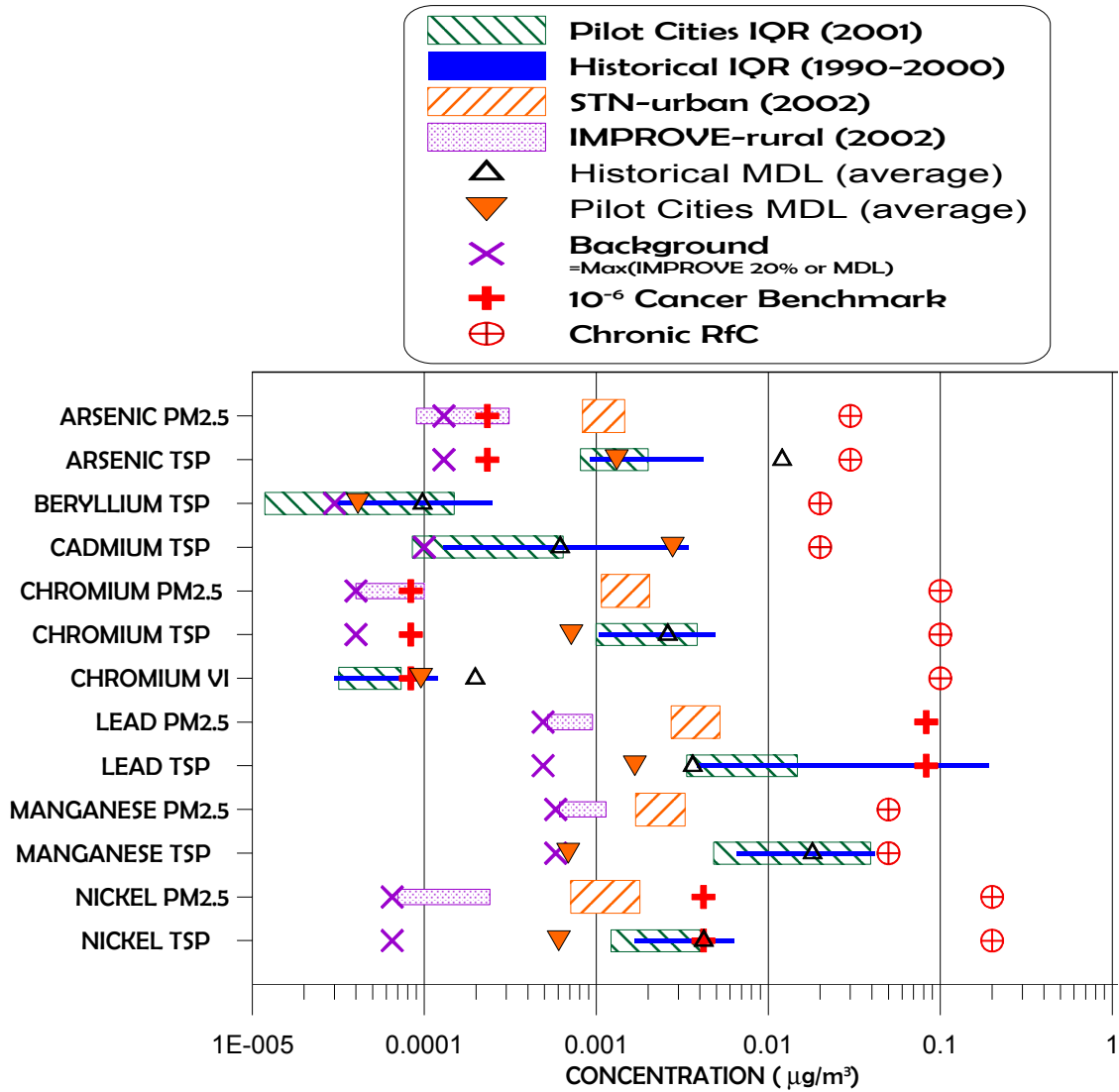

Figure 5. Historical and pilot city IQRs of concentrations for selected PM HAPs. The average MDLs, background concentrations, one-in-a-million cancer benchmark (from U.S. Environmental Protection Agency, 2003, 2004; California Air Resources Board, 2004), and RfC concentrations are shown for perspective.
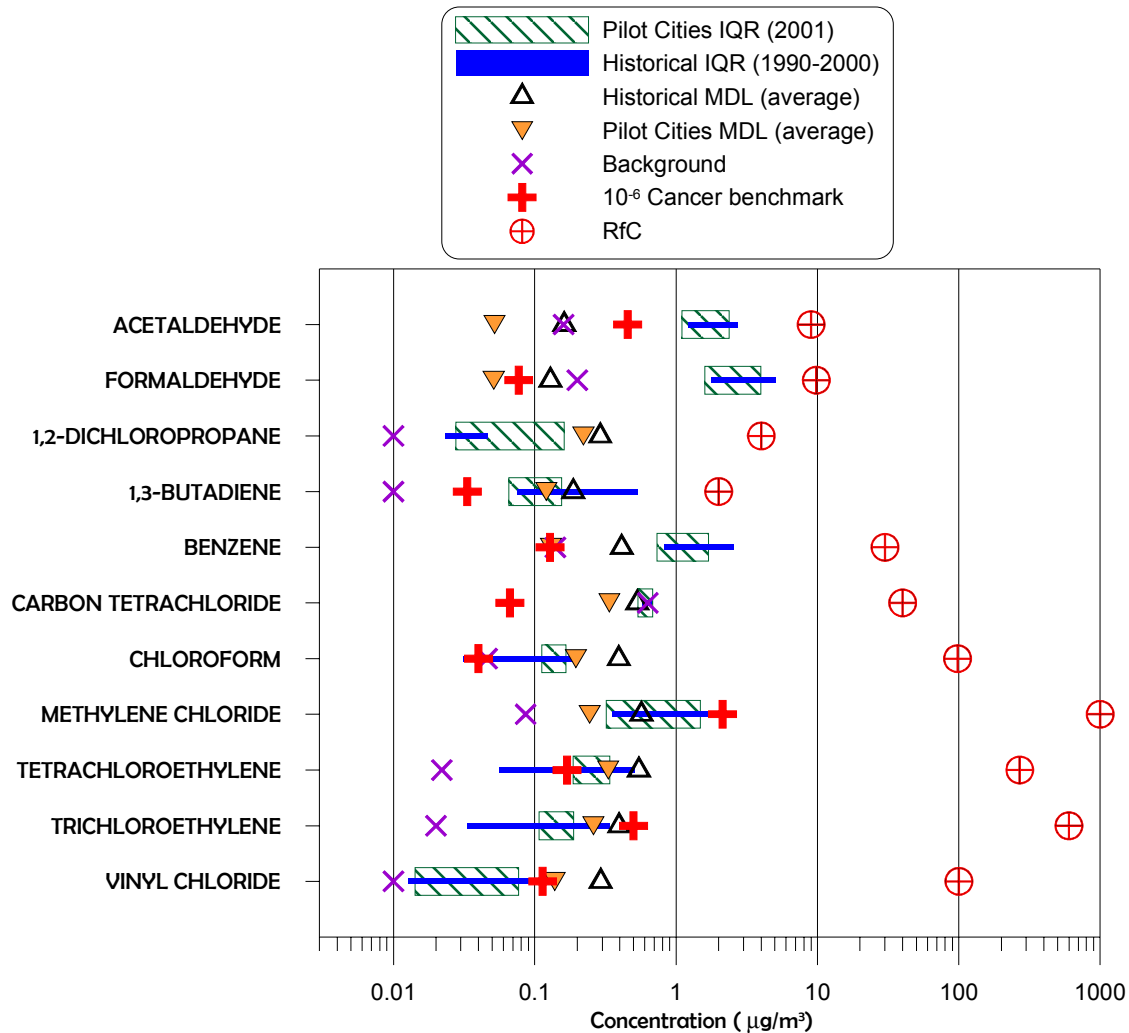
Figure 6. Historical and pilot city IQRs of concentrations for selected gaseous HAPs. The average MDLs, background concentrations, and one-in-a-million cancer benchmark concentrations (from U.S. Environmental Protection Agency, 2003, 2004; California Air Resources Board, 2004) are also shown.

Based on case studies, what can we say about air toxics concentrations on the urban scale?

- Analyses showed that where air toxics concentrations are not dominated by local emissions, urban concentrations of species with residence times of more than a few hours generally varied by less than a factor of 3 between seasonal or annual averages. For species with very short residence times (e.g., TSP metals or 1,3-butadiene) or with noticeably high point source emissions in the area, seasonal average concentrations varied by a factor of 5 to 10 or more (**Figure 7**).

- In most of the case studies using the historical data, high spatial variability was caused by local point sources (and identified with source-oriented monitoring).
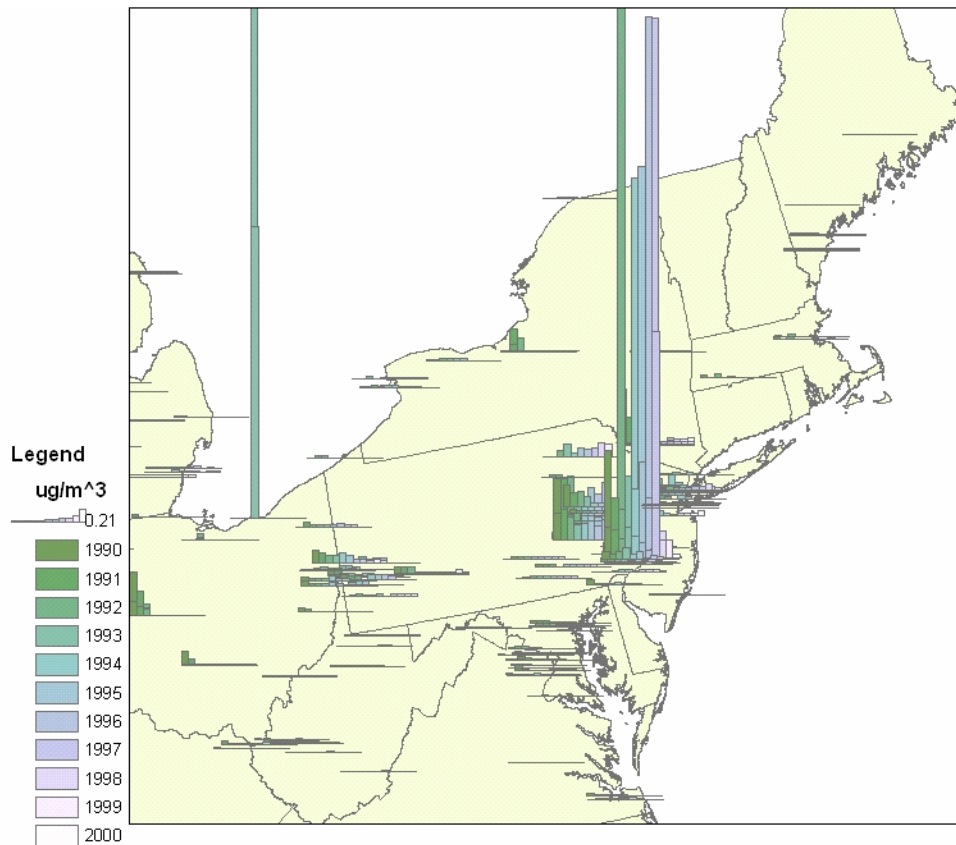
Figure 7. Summer seasonal averages of lead TSP concentrations ($\mu g/m^3$) for northeastern sites in the historical database. A hot spot in Philadelphia shows much higher lead TSP concentrations than at surrounding sites in the region.

How can concentrations from a small network be extrapolated to other areas?

- Small networks can compare their concentrations to historical and pilot city IQRs, MDLs, and background concentrations to understand how the data from their sites fit into a broader urban context (Figures 5 and 6). If concentrations differ significantly from typical urban ranges in these figures, analysts could follow case study examples to investigate possible causes of high concentrations.

- Extrapolation for most air toxics beyond the urban scale is not recommended without a network of rural measurements capable of capturing gradients between urban and rural areas. Modeling analyses suggest that concentrations gradients may be steep on upwind boundaries.

How representative are the existing monitoring sites (i.e., how broadly [spatially] can the annual average concentration at a particular site be applied)?

- Spatial representativeness depends on the species' residence time, whether the emissions are from area and mobile (i.e., well-distributed) or point (i.e., more localized) sources, and the locations of monitors with respect to the emission sources. Some air toxics have low spatial variability, such as carbon tetrachloride. Other air toxics with residence times

longer than a few hours can be estimated to within a factor of 2 or 3 in urban areas (see Figures 5 and 6) for most metropolitan statistical areas (MSAs). However, sites within 1 km of pollutant hot spots (i.e., source-oriented monitors) are likely to show elevated concentrations compared to typical urban ranges. Industrial cities exhibited steep concentration gradients that are unlikely to be captured without multiple monitors.

- Urban-scale models with realistic emissions, meteorology, and chemistry may be useful for addressing this question as well.

How do urban area concentrations compare with those in nearby rural areas?

- Unfortunately, there are few actual "nearby rural" sites with data with which to answer the question.

- Concentrations for many HAPs in rural areas are clearly lower than concentrations in city centers. However, because we have few measurements around the urban perimeter, it is difficult to assess the concentration gradient as we move from high population density to low.

- Leveraging the current urban network with special, regional, and national networks, like the National Oceanic and Atmospheric Administration (NOAA) Climate Monitoring and Diagnostics Laboratory (CMDL), PAMS, IMPROVE, and STN, is vital to better answer questions about urban and rural differences (e.g., **Figure 8**).
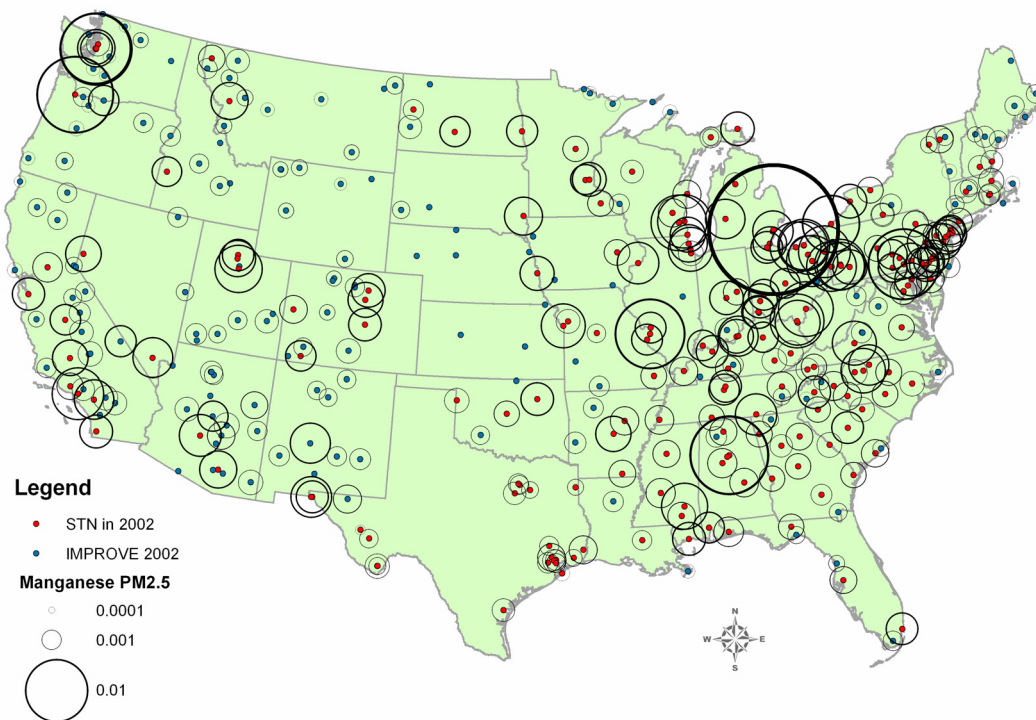

Figure 8.   Annual manganese PM$_{2.5}$ concentrations (black circles, in μg/m^3) using 2002 data from the STN and IMPROVE. Urban sites (STN) are in red, rural (IMPROVE) are in blue.

What is meant by "background" levels?

- Background concentrations can be defined as remote, regional, and operational:

  – The definition of remote background is the mean concentration of a compound measured at a site not impacted by local or regional emission sources. Examples of remote sites include islands in the middle of the ocean or Polar Regions. These data are useful to track global scale changes in long-lived air toxics, for example.

  – The definition of regional background is the mean concentration of a compound measured at a site not impacted by local sources. A regional background site might be one located in a coastal area with predominantly onshore flow, or a remote continental site uninfluenced by local sources. These data are useful to help understand what is transported into an area.

  – The definition of operational background is the mean concentration of a compound measured at an upwind site. This operational definition may include urban or rural sites upwind of other monitoring stations, since that air is the "background" for sites downwind. Operational background concentrations are useful for site-specific analyses in regions with high density of emissions (e.g., northeastern corridor of the United States).

- The definition of remote background indicates the lowest possible concentrations occurring in the atmosphere, whereas the definition of operational background indicates much higher concentrations depending on proximity to upwind sources and the nature of the pollutant. The operational definition of background concentration implies that backgrounds vary significantly by site location.

How can background levels be estimated, and what are reasonable estimates of background levels?

- Background levels can be estimated in a number of ways, depending on pollutant and definition used:

  – Remote background concentrations of many of the 18 target HAPs are routinely measured at monitoring stations operated by NOAA CMDL; data from other studies are available in the scientific literature.

  – Regional background concentrations may be estimated from rural monitoring stations for compounds with relatively long atmospheric residence times. The concentrations of these compounds should vary little across the United States (e.g., carbon tetrachloride and other chlorofluorocarbons [CFCs]).

  – Operational background concentrations consist of global background concentrations plus regional, natural, and upwind sources of pollutants. Use of upwind monitoring stations, chemical lifetimes, and ambient monitoring data can be used to help identify operational background concentrations on a site-by-site basis. This approach is not practical for application to all monitoring sites in the historical database.

- Remote background concentrations of several air toxics are available and were quantified for each of the 18 target HAPs in Phase III (e.g., Figures 5 and 6). Operational background concentrations need to be developed on a case-by-case basis.

- Remote background concentration data from other networks can be used to augment the NATTS measurements for some of the longer-lived species, but additional measurements in rural areas are necessary to better understand regional differences in background concentrations.

Are there typical urban and rural profiles (fingerprints)?

- This question can only be partially addressed using the historical database since it is composed primarily of urban sites and species on the HAPs target list. The results are limited in usefulness because

  – Only urban fingerprints were investigated.

  – Fingerprints were usually limited to one phase. Multi-phase pollutants (gaseous, semi-volatile, PM) were not always measured together and, therefore, data availability is inconsistent from site to site. However, fingerprint analysis of groups of species measured using similar techniques (e.g., chlorinated VOCs) reveals significant variability among urban areas. In contrast, within-region fingerprints are often consistent (e.g., **Figure 9**).

  – Generally, only the HAPs data are available. Other species would augment fingerprint analyses (e.g., 56 PAMS target hydrocarbons) and potentially help us understand the underlying sources of air toxics.

How does exposure to mobile source air toxics (MSATs) vary as a function of distance from roadways?

- Roadway proximity was added to the database; however, the monitoring sites were typically more than several hundred meters from roadways and may have multiple confounding variables. Recent research (Zhu et al., 2002; Hitchins et al., 2000; Kinney et al., 2000; Hewitt, 1991) has shown that spatial resolution down to tens of meters may be required to better characterize exposure to motor vehicle-related pollutants (especially those directly emitted from motor vehicles).
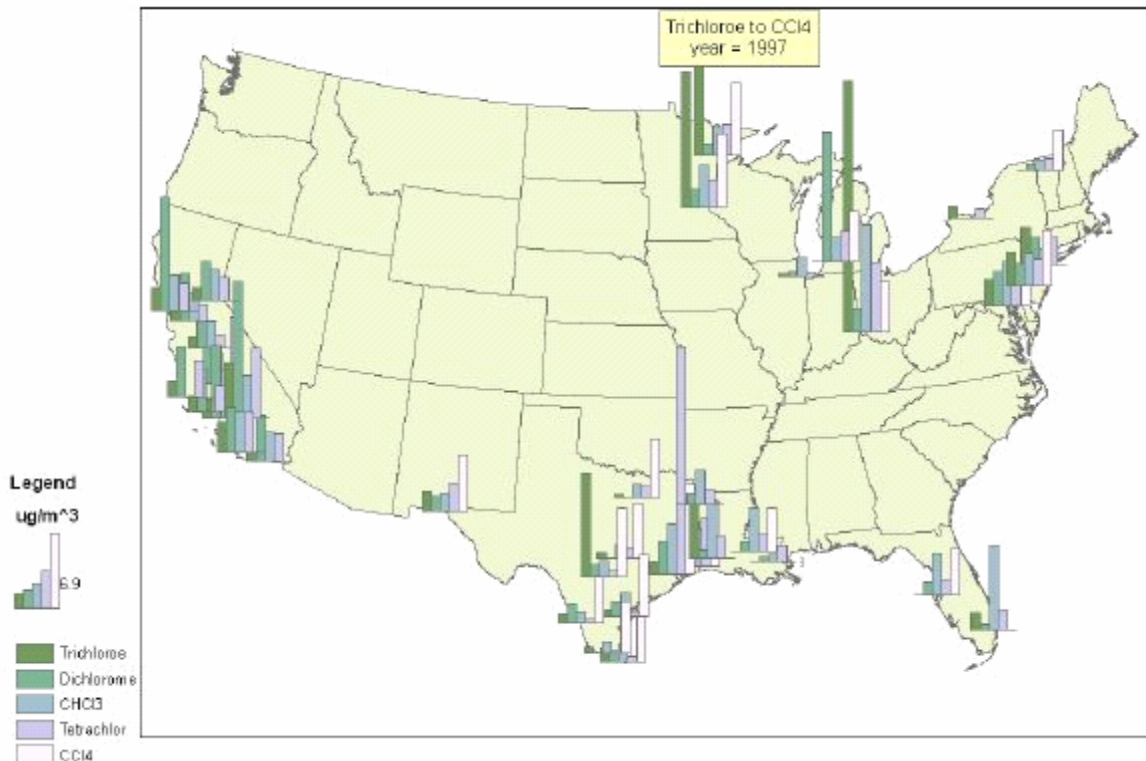
Figure 9. Normalized fingerprints of selected chlorinated VOC annual averages from 1997. The species are trichloroethylene (green), dichloromethane (aqua), chloroform (blue), tetrachloroethylene (purple), and carbon tetrachloride (pink).

How do air toxics concentrations vary temporally?

What can we say about the variation in air toxics concentrations on a yearly (trends), seasonal (or monthly), day-of-week, and hour-of-day basis?

- Several air toxics at national, regional, and local levels exhibited qualitatively decreasing concentrations from 1990 to 2000. The HAPs that show decreasing trends at the regional or national level over this time period were acetaldehyde, 1,3-butadiene (**Figure 10**), benzene, chloroform, lead (TSP and fine), tetrachloroethylene, and trichloroethylene. Additional investigation of these trends, including effects of meteorology, at national, regional, and local levels is needed to quantify these changes.
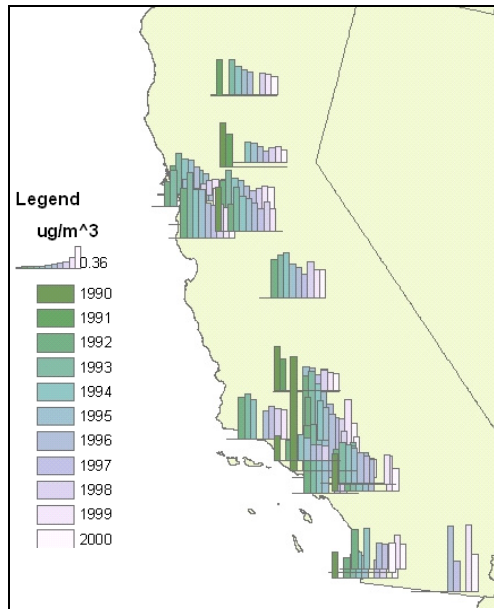
Figure 10.   Annual average concentrations (µg/m³) of 1,3-butadiene in California from 1990 to 2000.

- Remote background concentrations of carbon tetrachloride, chloroform, and tetrachloroethylene have shown a decreasing trend from 1995 to 2003.

- Seasonal, day-of-week, and hourly trends were addressed in Phase I. By season, acetaldehyde and formaldehyde concentrations were higher in the summer while benzene concentrations were typically higher in winter (**Figure 11**). By day of week, DPM-related compound concentrations were lower on weekends compared to weekdays (Bortnick et al., 2003). By hour of day, formaldehyde concentrations were generally higher midday because of photochemical production ((Bortnick et al., 2003). Sufficient samples are needed in a season or by day of week to investigate these trends; typically, one year of data with 1-in-12-day sampling is insufficient for these investigations. Hourly (or 3-hr) data from other networks (e.g., PAMS) are needed to assess diurnal variations; these investigations were documented in Phases I and II and in PAMS data analyses.
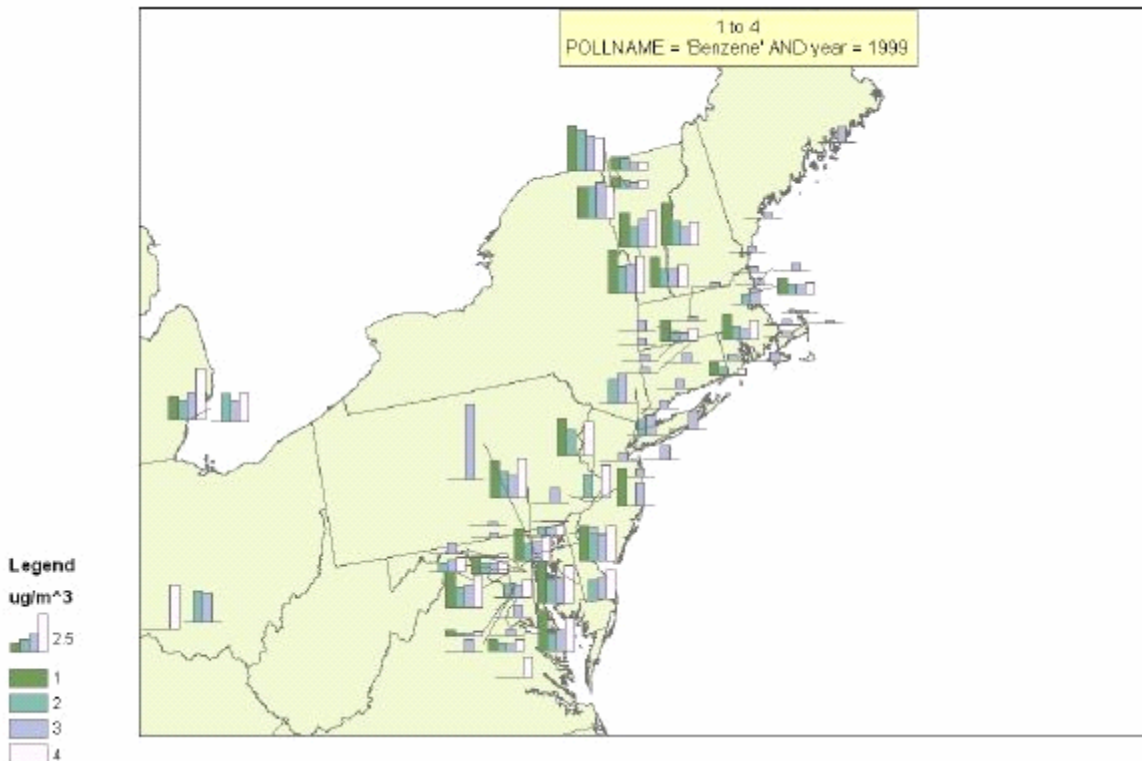
Figure 11.   Seasonal average benzene concentrations (μg/m^3) for 1999 in the Northeast region. The seasons are winter (green, 1), spring (aqua, 2), summer (blue, 3), and fall (pink, 4).

How do we characterize spikes in ambient concentrations?

- A spike is an annual, seasonal, or daily average concentration that is significantly higher than other averages. Aberrant concentrations should be compared to those at nearby sites to determine the spatial extent of the anomaly. In addition, co-measured species should be examined to assess whether a sampling or analytical error may have been responsible for high concentrations of all species for a given time period. Spatial maps facilitate the investigations of spikes (or hot spots) as shown in Figures 1, 2, and 7, for example.

- Another way to identify spikes is to note average concentrations above the IQRs shown in Figures 5 and 6; these data can then be investigated using case studies. In a case study, time series plots of concentrations (**Figure 12**), metadata (i.e., site maps such as Figure 4, satellite images, site photos, proximity of emission sources), pollution roses, and local knowledge of unusual conditions are needed to further understand and characterize the high concentrations. In the case study shown in Figure 12, only one measurement spiked at a Brazoria County, Texas, site in 1997 and may be suspect data, while three measurements were high in winter 1998 at the Beaumont site and may be attributable to a series of real emissions events.
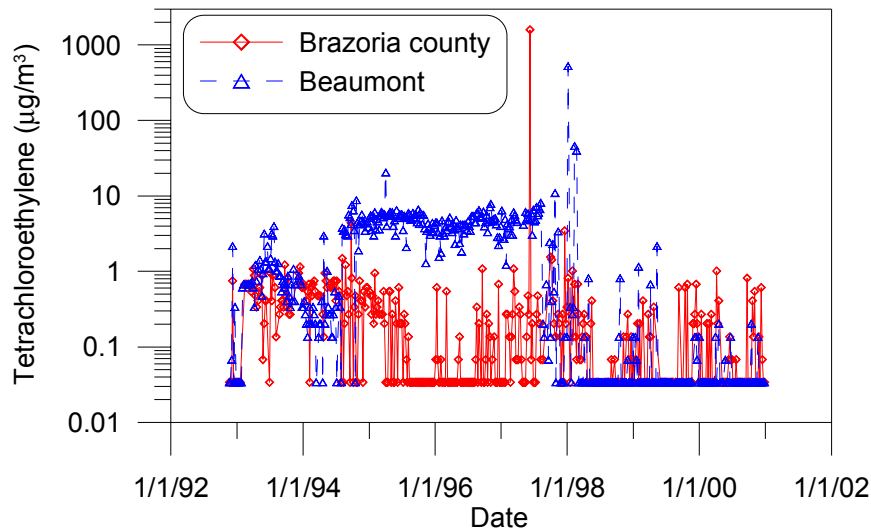
18

Figure 12. Time series of 24-hr tetrachloroethylene concentrations (µg/m³) collected every sixth day at two sites in Texas.

- In the case studies performed in Phases II and III, characterizing spikes in ambient concentrations required extensive investigation and local knowledge to discern whether the high concentrations were real or the result of sampling or analytical error.

What have we learned about the frequency of air toxics monitoring in order to capture the true story of exposure?

- Sampling frequency is a function of the length of time between sample collection (e.g., daily versus a sample every third day) and the duration of the sample (e.g., 1-hr versus 24-hr averages). For exposure assessment, one metric used is the annual average concentration. For several air toxics, archive data were used in Phase I to show that site-specific annual average concentrations can be estimated with 10% to 15% relative error using every third to every sixth day sampling (for data sets with concentrations mostly above the MDL). More frequent sampling is needed for higher concentration (source-oriented) sites to obtain similar relative error.

What do air toxics data say about the effectiveness of various control programs?

How effective have mobile source controls been in reducing exposure to toxics?

- An annual trend that can be associated with a control strategy change is the decrease in benzene concentrations due to the introduction of RFG in California (e.g., **Figure 13**). Other urban areas in the United States have experienced similar trends (Main, 2002). The historical data also show that 1,3-butadiene concentrations have declined over the same time period and likely may also be attributed to RFG. In contrast, formaldehyde concentrations may have increased from 1993 to 2000 (**Figure 14**) with introduction of RFG as predicted in model estimates (Stoeckenius et al., 1995). However, the change in formaldehyde concentrations may be in part or fully a result of different sampling

19

techniques (Ruiz, 1998) introduced between 1995 and 1996. Greater certainty in cancer benchmarks is needed to determine whether the net result is an increase or a decrease in total risk associated with these air toxics. In Figures 13 and 14, the box defines the IQR, the notch around the median is the 95% confidence interval, and the whiskers extend to data within 1.5*IQR.
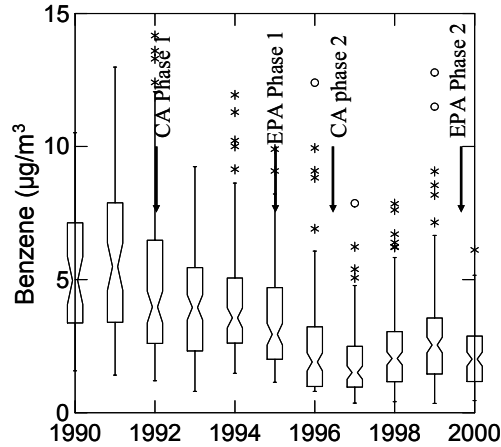
Figure 13. Notched box whisker plots of benzene concentrations ($\mu g/m^3$) from 1990 to 2000 for all sites in California (urban, motor vehicle-dominated). Introduction of different formulations of gasoline are noted.

Figure 14. Notched box whisker plot of formaldehyde concentration ($\mu g/m^3$) from 1990 to 2000 for all sites in California (urban, motor vehicle-dominated). Introduction of different formulations of gasoline are noted.
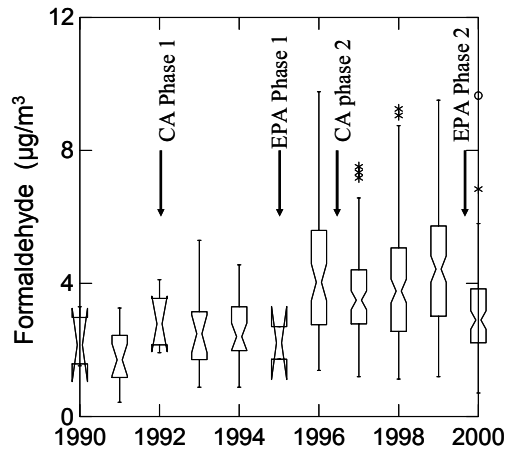
- The historical database lacks data with which to understand DPM contributions to ambient air. Investigation of pilot city data led to the addition of BC measurements in the NATTS; however, BC is not a unique marker for DPM. Additional analysis of speciated organic carbon data from special studies and source apportionment of the pilot city and

20

NATTS data are needed to further understand DPM contributions and appropriate DPM marker species.

How effective have maximum achievable control technology (MACT) standards been in reducing exposure to toxics?

- Due to the relatively long phase-in periods of MACT standards, it is difficult to unambiguously ascribe changes in ambient concentrations to MACT implementation. However, species that have been targeted for reduction by MACT standards often show decreasing concentrations over the time period of interest:

  - Chloroform concentrations show marked decreases at a number of hot spot areas throughout the decade. For example, decreases occurred in Houston, Texas; the Great Lakes; Miami, Florida; and the Northeast as shown in **Figure 15**. These reductions may be due to a number of MACT regulations, including those for halogenated solvents in 1994, off-site waste operations in 1996, or pulp and paper mills in 1998.
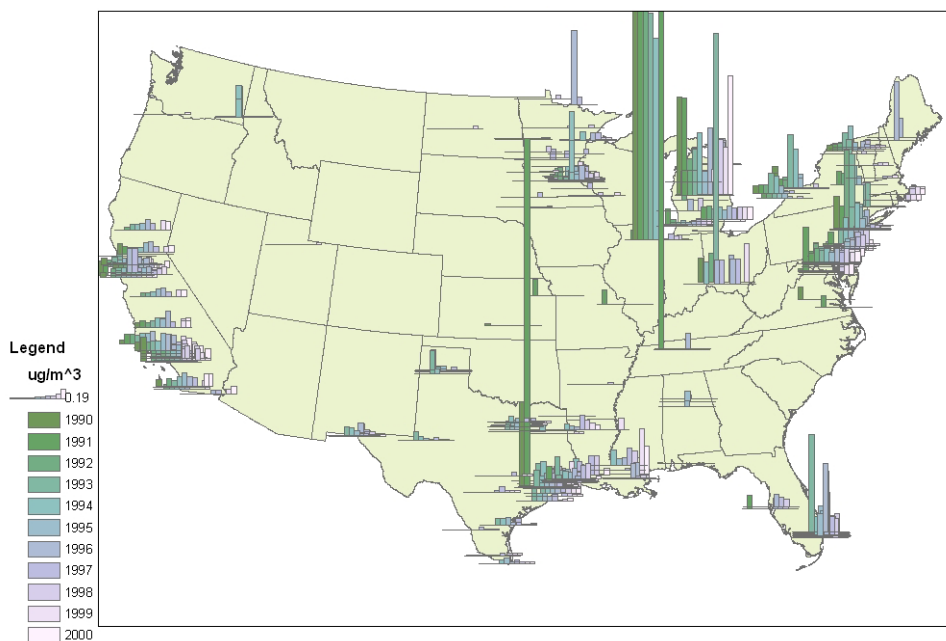

Figure 15. Summer seasonal average chloroform concentrations ($\mu g/m^3$) for the United States from 1990 to 2000.

  - Dichloromethane concentrations decreased significantly between 1994 and 1995. Regulations aimed at reducing dichloromethane concentrations targeted halogenated solvents in 1994, aerospace manufacturing in 1995, pharmaceuticals in 1998, and polyurethane foam production in 1998. In addition, dichloromethane was a key intermediate in the production of CFCs, whose production was largely phased out in 1995 due to the Montreal Protocol. In 1994 there were spikes at dozens of sites around the country (e.g., El Paso and Midland, Texas; Minneapolis, Minnesota;

Miami, Florida; Vermont; San Diego and Santa Barbara, California; and New York City, New York). It is unclear why a large spike in concentrations occurred in 1994 relative to the surrounding years, but impending regulations may have led to an increase in emissions in 1994. Further analysis of this phenomenon through case studies may yield additional insight into these multiple spikes.

− Tetrachloroethylene concentrations showed a markedly decreasing trend from 1992 to 1999 (**Figure 16**). Analyses using spatial maps indicate that tetrachloroethylene concentrations have decreased noticeably at many urban sites throughout the nation. Regulations aimed at reducing tetrachloroethylene were directed at dry cleaning in 1993 and halogenated solvents in 1994.

− Trichloroethylene concentrations require additional analyses to determine nationwide and regional trends because of the high variability among sites and the large number of concentration spikes.

• For all assessments of MACT implementation and the effects on air toxics concentrations, we have found that local knowledge is required to understand how and when the controls were implemented and the likely result and magnitude of the controls.
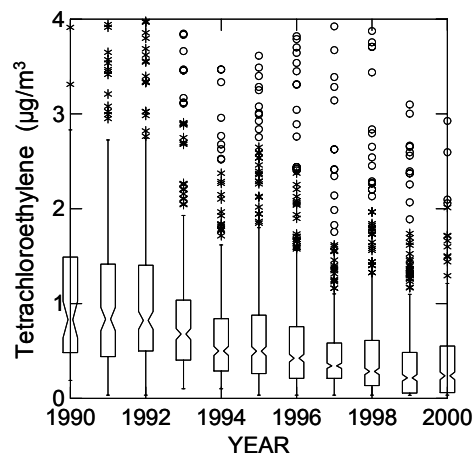

Figure 16. Notched box whisker plot of tetrachloroethylene seasonal average concentrations ($\mu g/m^3$) from all sites across the nation from 1990 to 2000.

Have emissions controls reduced background concentrations?

• For a few air toxics, background concentrations have been reduced. Carbon tetrachloride and tetrachloroethylene show decreasing trends from 1995 to 2003, likely as a result of the Montreal Protocol. Background concentrations and trends are further addressed by McCarthy et al. (2004).

How has implementation of ozone and PM controls reduced air toxics levels (and vice versa)?

• RFG was implemented primarily to reduce ozone formation; the result of this control strategy on air toxics was discussed previously in this paper.

- NATTS data will be useful to understand the effects of forthcoming diesel fuel regulations.

Can current models predict the range, variability, and gradients of measured toxics?

How good was the modeling approach used in this study?

- **Figure 17** provides an overview of model performance in terms of the ratio of model-predicted to observed concentrations for each city and species modeled. A ratio of 1.0 indicates perfect agreement between the model and observations. Values less than 1.0 indicate underprediction by the model and ratios grater than 1.0 indicate overprediction. The following is a general description of those results:
    - TSP metals (Seattle and Detroit only) were generally underpredicted except for cadmium.
    - Predictions of arsenic, beryllium, and manganese were the worst, with underpredictions of 85-94% on average at the monitoring sites.
    - Lead was undepredicted by 66% on average in Detroit but overpredicted by 77% in Seattle, which was dominated by an overprediction at the Seattle-Tacoma airport (SeaTac) monitor of 473%. However, this overprediction can be explained by the misallocation of general aviation fuel emissions to the SeaTac airport, which is discussed further in the following question concerning emission inventory quality.
    - Without considering secondary formation, acetaldehyde concentrations were underpredicted by 92% in Cedar Rapids, 80% in Detroit, and 53% in Seattle.
    - Without considering secondary formation, formaldehyde concentrations were underpredicted by 86% in Cedar Rapids and 78% in Detroit. On average, Seattle showed no under- or overprediction.
    - EPA estimates of secondary acetaldehyde and formaldehyde production from OZIPR model runs for 10 cities (U.S. Environmental Protection Agency, 1999) indicated that 87% to 92% of acetaldehyde and formaldehyde were the result of secondary formation. Based on these results, EPA suggested that if none of the cities in this report typifies the area of interest, the secondary component of the HAPs concentrations may be estimated from values for the primary component. For formaldehyde and acetaldehyde, EPA recommended multiplying the spatial average primary concentrations by 9.0 (90%/10%) to estimate the secondary contribution to these species and then add that concentration to the site-specific primary concentrations. Applying this methodology to the cities in this study resulted in large overpredictions (some exceeding 1000%) at most sites in Detroit and Seattle. For Cedar Rapids, using this method, acetaldehyde was underpredicted by 35% and formaldehyde was overpredicted by 13%.
    - 1,3-Butadiene, benzene, and tetrachloroethylene are overpredicted in Detroit and Seattle, and underpredicted in Cedar Rapids.

23

–   Methylene Chloride is generally underpredicted, but unlike the other VOCs modeled, it is best predicted in Cedar Rapids.
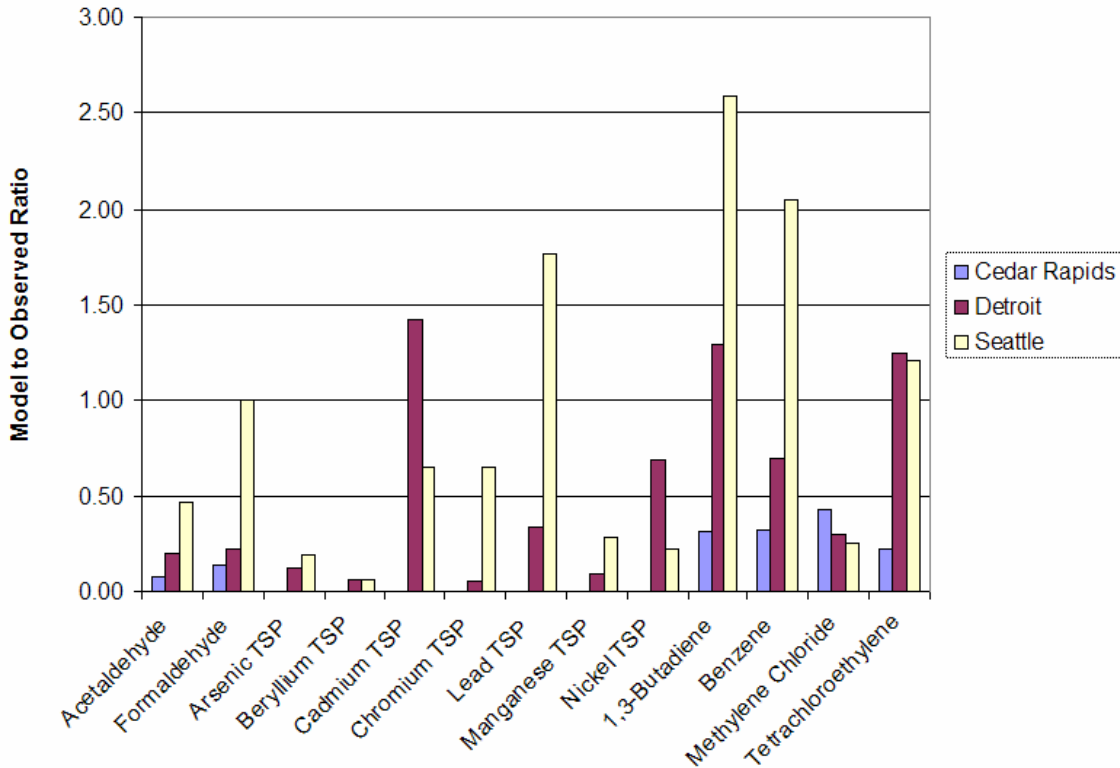

Figure 17.   Ratios of annual average site-specific model predictions to observed concentrations of HAPs.

- The model-predicted concentrations were generally less variable than the observations. This is consistent with distributing non-point source emissions to 1-km areas for modeling.

- When the best-fit model predictions (i.e., the closest to the observed value) for receptors within 4.5 km of the monitoring sites were used, there were significant improvements in model performance. **Figure 18** provides an overview of model performance in terms of the ratio of model-predicted to observed concentrations using the best-fit model predictions.

- Of the species considered risk drivers, arsenic had the poorest model performance, which did not significantly improve when the best-fit model predictions were used.
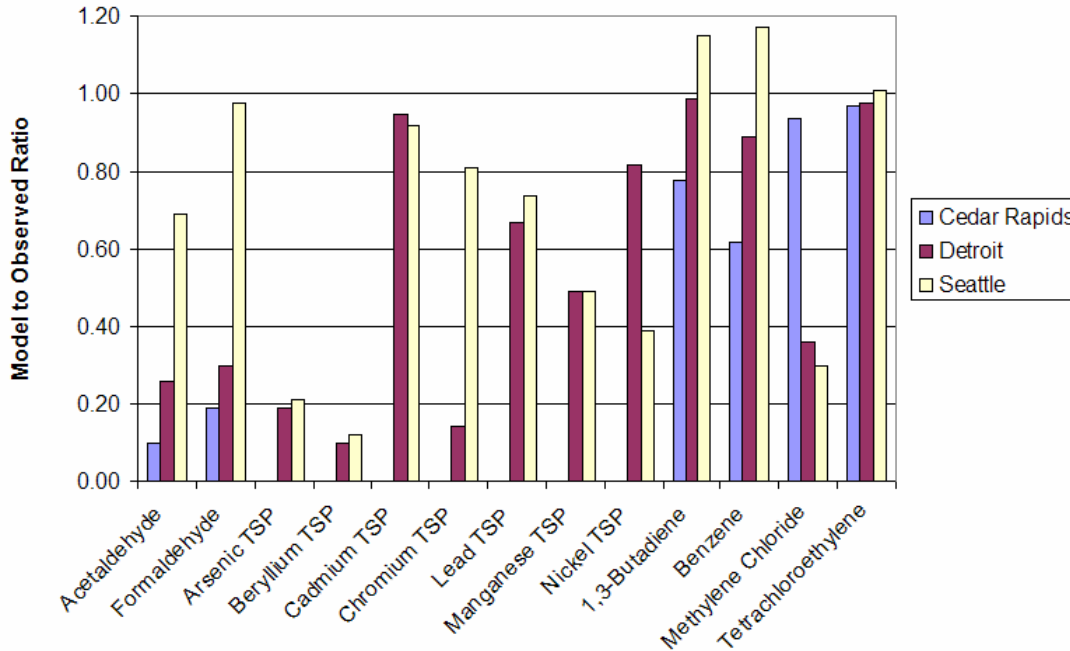
Figure 18. Ratios of annual average best-fit (within 4.5 km of the observation site) model predictions to observed concentrations of HAPs.

How does the performance of the current modeling compare to that done for the 1996 National Air Toxics Assessment (NATA)?

- Model evaluation for the NATA involved a much large number of cities and sites. Therefore, comparisons between the two modeling efforts are not necessarily commensurate.

- Overall model performance for benzene was about the same as in the NATA but with a tendency to be overpredicted in the current modeling and underpredicted in the NATA.

- While formaldehyde and acetaldehyde were underpredicted by 35-40% in the NATA modeling, the use of the OZIPR model to estimate secondary formation for these species resulted in overpredictions.

- Model performance for lead, chromium, and especially cadmium was much better in the current modeling than in the NATA.

How good are the toxics emission inventories and what is the impact of inventory quality on predicted concentrations?

- Based on model performance, the NEI appears to be of reasonable quality for the gas-phase species, except for methylene chloride. This is consistent with many of the gaseous species being dominated by mobile sources emissions, which have undergone continuous research and improvement over the past decade. The consistent underprediction of methylene chloride across all sites indicates that a significant amount of this solvent may be missing from the emissions inventory.

25

- Most metals have underpredicted in past and current modeling studies. Because model predictions for some metals (e.g., cadmium) and some low-reactivity VOCs (i.e., benzene) were not underpredicted, it is not believed that these are not a result of meteorological biases. These results infer that the emission inventory for most of the metals is underestimated. Improvements to inventories of toxic metals should focus initially on risk drivers such as arsenic.

- Missing sources can significantly affect model performance. Extremely high concentrations (i.e., greater than 10,000 $\mu g/m^3$) of methylene chloride were measured at the Allen Park site in Detroit, which dominated the annual average concentration. The emissions inventory contained only known sources of these emissions and the model predicted only 1% of the observed annual average.

- Care must be taken in the selection of spatial surrogates. The overprediction of lead TSP at the SeaTac site in Seattle was found to be a result of lead emission from general aviation fuel being allocated to all airports based on their spatial size. Instead of allocating these emission, which make up 98% of the lead emissions for the area, to dozens of small general aviation airports, most of the emissions were allocated to the SeaTac airport, which is upwind of the SeaTac monitor, and to Boeing Field, which is downwind of the monitor. It was noted that general aviation fuels currently contain lead but that low-lead fuels (approximately 25% of the original lead content) have been introduced in the past decade. Future modeling studies should verify that current speciation profiles for general aviation fuels are used.

- Allocation of mobile source emissions to 1-km areas for modeling was not sufficient to resolve concentration gradients near roadways. For example:

  o Mobile-source-related species were significantly overpredicted at the SeaTac monitoring site in Seattle. While the SeaTac monitor is upwind of major roadways, emissions from those roadways were allocated to the 1-km area in which the monitor was located.

  o Mobile-source-related species were significantly underpredicted at the two monitoring sites in Cedar Rapids. Both sites are relatively near interstate highways that appear to have a larger impact on the monitors than is estimated by the model when those emissions are spread over a 1-km area.

- These results indicate that it may be more difficult to model mobile sources in smaller cities where emissions are less ubiquitous, and in areas where sites are located at the upwind edge of the city, than in larger cities.

How can background concentrations for toxics be defined for use in near-source modeling?

Are there sufficient measurements to establish operational background concentrations?

- Models require operational background concentrations, which consist of global background concentrations plus regional, natural, and upwind sources of pollutants. For some locations and species, there are measurements sufficient to establish operational

background concentrations; however, for most cities and species, there are insufficient measurements.

- Regional background concentrations may be estimated from rural monitoring stations for compounds with relatively long atmospheric residence times. However, air toxics monitoring has historically focused on areas where there is likely to be higher levels of human exposure. This focus has not provided adequate measurements to establish regional background concentrations of air toxics for use in modeling assessments.

- Upwind monitoring stations, chemical lifetimes, and ambient monitoring data can be used to help identify operational background concentrations on a site-by-site basis. However, few monitoring stations are always upwind of a site.

- The remote background concentrations routinely measured at NOAA CMDL monitoring stations provide reasonable estimates of global background concentrations and can provide reasonable estimates of regional background for species that currently have no significant sources of emissions in the United States (e.g., carbon tetrachloride) or that are removed rapidly from the atmosphere (e.g., 1,3-butadiene).

What alternatives to using measurements can be used for establishing background concentrations?

- Establishing boundary conditions is complicated because some important toxics (e.g., benzene, formaldehyde, 1,3-butadiene, and acetaldehyde) undergo chemical transformation in the atmosphere. Further complications exist because formaldehyde and acetaldehyde can be formed from every VOC in the atmosphere. Major contributors to aldehyde formation are toluene, xylenes, other hydrocarbons present in automobile exhaust, and biogenic hydrocarbons. It is estimated that more than 80% of these aldehydes may be due to atmospheric formation, not emissions (Luecken, 2002).

- Photochemical grid models can be used to establish regional and urban-scale operational background concentrations for both reactive and non-reactive species. While photochemical grid models require considerable resources to setup, run, and evaluate, properly representing the secondary formation of risk drivers, such as formaldehyde and acetaldehyde, will be necessary before air quality models can become useful planning tools.

- Rural measurements will still be needed to validate regional photochemical models.

What are the impacts of complex meteorology, terrain, and emissions patterns on predicted concentrations?

- While these impacts were not explicitly investigated in this study, some information about these impacts was inferred from the differences between sites and cities.

- Many cities have meteorology that is influenced by terrain or water bodies resulting in spatial variations of wind direction and speed, temperature, moisture, and vertical mixing across the city. For those cities, the use of single-station meteorology, as used in this study, cannot adequately represent the transport, diffusion, transformation, and removal

processes affecting concentrations of HAPs. For example, winds are typically from the south-southwest in southern Seattle, southerly in central Seattle, and southeasterly in northern Seattle. Therefore, using the winds from any one area will be incorrect for the others.

- The gradients in modeled concentrations near monitoring sites indicate that while a model may accurately predict near-site concentrations, it may not predict the concentrations in the correct location. This may be a result of meteorological complexity (the meteorology used in the model is not representative of all locations) or emissions complexity (cases with many industrial point sources or spatial allocation of non-point sources by spatial surrogates). While this may not be an issue for criteria pollutants where maximum concentrations are of concern, HAPs need to be evaluated in terms of population exposure. Failure to get concentrations in the correct locations may result in any exposure estimates being useless for planning purposes.

Can current modeling tools be used in air toxics planning?

- To use modeling tools in air toxics planning means to use their predictive capabilities to assess the effectiveness of potential control measures in reducing exposure to HAPs.

- For cases where species are non-reactive, emissions are well-known, and meteorology does not spatially vary significantly, the tools used in this study may be appropriate for planning purposes. However, in such cases, simpler approaches such as liner rollback may be as effective for planning purposes.

- For cases where meteorological complexity is an issue, the tools used in this study have limited applicability for planning purposes because of their inability to use varying meteorology needed to properly estimate the exposure to HAPs.

- For cases with significant contributions of HAPs from sources beyond the 50-km modeling domain (e.g. the transport of benzene from major upwind cities), the tools used in this study are insufficient for use in air toxics planning.

- For cases where secondary formation of HAPs is being addressed, the tools used in this study may have limited usefulness. The simplified approach for estimating secondary production of HAPs using the OZIPR model was found to be inadequate in some areas because it does not take into account monitor location relative to areas of precursor emissions. This can lead to significant overestimation of secondary HAPs at some locations. In Seattle, for example, the OZIPR estimated secondary formaldehyde and acetaldehyde concentrations alone were three and ten times the observed concentrations, respectively. Because these species are considered risk drivers, care should be taken to accurately model their secondary formation.

- While the modeling methods used in the NATA and this study have many limitations when used for assessing current exposure or predicting future exposure to HAPs, there are currently available meteorological and air quality modeling tools capable of addressing those limitations. These other models may be more complex but, that level of complexity will be required to predict the effect of emission controls on HAPs.

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.

Bortnick S., Coutant B., Holdren M., Stetzer S., Holdcraft J., House L., Pivetz T., and Main H. (2001) Air toxics monitoring data: analyses and network design recommendations. Report prepared for Lake Michigan Air Directors Consortium, Des Plaines, IL, by Sonoma Technology, Inc., Petaluma, CA and Battelle Memorial Institute, Columbus, OH, October.

Bortnick S.M., Coutant B.W., and Biddle B.M. (2003) Estimate background concentrations for the national-scale air toxics assessment. Final technical report prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, by Battelle, Columbus, OH, Contract No. 68-D-02-061, Work Assignment 1-03, June.

California Air Resources Board (2004) Consolidated table of OEHHA/ARB approved risk assessment health values. Available on the Internet at <http://www.arb.ca.gov/toxics/healthval/healthval.htm> ; last accessed August 3, 2004.

Hafner H.R., Wheeler N.J., and Roberts P.T. (2004) Analysis of air toxics monitoring data. Work plan prepared for Lake Michigan Air Directors Consortium, Des Plaines, IL, by Sonoma Technology, Inc., Petaluma, CA, STI-903555-2442-WP2, January.

Hewitt C.N. (1991) Spatial variations in nitrogen dioxide concentrations in an urban area. *Atmos. Environ.* **25B**, 429–434.

Hitchins J., Morawska L., Wolff L., and Gilbert D. (2000) Concentration of submicrometer particles from vehicle emissions near a major road. *Atmos. Environ.* **34**, 51-59.

Kinney P.L., Aggarwal M., Northridge M.E., Janssen N.A., and Shepard P. (2000) Airborne concentrations of $PM_{2.5}$ and diesel exhaust particles on Harlem sidewalks: a community-based pilot study. *Environ. Health Perspect.* **108**, 213-218.

Luecken D. (2002) Atmospheric formation and decay of air toxics - implications for exposure assessments. Presented at the *Air Toxics Exposure Assessment Workshop, San Francisco, CA, June 25-27*.

Main H.H. (2002) Evaluation of control program effectiveness. Technical memorandum prepared for the Northeast States for Coordinated Air Use Management, Boston, MA, by Sonoma Technology, Inc., Petaluma, CA, STI-902071-2253-TM, September.

McCarthy M.C., Hafner H.R., and Montzka S.A. (2004) Background Concentrations of 18 Core Air Toxics in the Northern Hemisphere. *J. Air & Waste Manage. Assoc.* (in progress) (STI-2589).

Ruiz G. (1998) Report on a parallel carbonyl monitoring study using C18 and silica gel sampling media. Final Report prepared for California Air Resources Board. Available on the Internet at <http://www.arb.ca.gov/aaqm/qmosprog/papers_studies/carbonylstudy.pdf> , last accessed August 30, 2004.

Stoeckenius T.E., Yarwood G., Ligocki M.P., Cohen J.P., Shepard S.B., Looker R.E., Fujita E.M., Main H.H., and Roberts P.T. (1995) Feasibility study for a 1995-1996 Southern California air quality monitoring program. Final report prepared for Coordinating Research Council, Atlanta, GA by Systems Applications International, San Rafael, CA, Desert Research Institute, Reno, NV, and Sonoma Technology, Inc., Santa Rosa, CA, SYSAPP94-94/065, January.

U.S. Environmental Protection Agency (1999) A simplified approach for estimating secondary production of hazardous air pollutants (HAPs) using the OZIPR model. Office of Air Planning and Standards, Research Triangle Park, NC, EPA-454/R-99-054.

U.S. Environmental Protection Agency (2003) Integrated risk information system (IRIS). Available on the Internet at <http://www.epa.gov/iris/> ; last accessed April 2, 2004.

U.S. Environmental Protection Agency (2004) Toxics sources. Technology Transfer Network Air Toxics Website. Available on the Internet at <http://www.epa.gov/ttn/atw/toxsource/table1.pdf> ; last accessed August 3, 2004.

Zhu Y., Hinds W.C., Kim S., Shen S., and Sioutas C. (2002) Study of ultrafine particles near a major highway with heavy-duty diesel traffic. *Atmos. Environ.* **36**, 4323-4335.