Characterizing Air Toxics

What are the diurnal, seasonal, and spatial characteristics of air toxics?

What do these characteristics tell us about emission sources, transport, and chemistry?

Characterizing Air Toxics

What's Covered in This Section

- Temporal Patterns
  - Diurnal
  - Day-of-week
  - Seasonal
- Spatial Patterns
  - Spatial characterization
    - National concentration plots for perspective
    - Maps
  - Variability within and between cities
  - Hot and cold spot analysis
  - Comparing urban and rural sites
- Risk screening

Characterizing Air Toxics

Overview

- Spatial and temporal characterizations of air toxics data are the basis for improving our understanding of emissions and the atmospheric processes that influence pollutant formation, distribution, and removal. Goals of these data analyses can include
  - Identifying possible important sources of air toxics.
  - Determining chemical and physical processes that lead to high air toxics concentrations.
- Characterization analyses help us develop a conceptual model of processes affecting air toxics concentrations and also provide an opportunity to compare data to existing conceptual models to identify interesting or problematic data. Following are some typical questions which may be addressed using these types of analyses:
  - Where are air toxics concentrations highest or lowest?
  - How do pollutant concentrations vary relative to each other – and what does this tell us about their sources?
  - What and where are the air toxics of concern?
  - How do urban and rural sites compare?
  - How do air toxics concentrations compare to criteria pollutants (e.g., ozone and $PM_{2.5}$)?
  - What local or regional sources influence a particular measurement site?

Quantifying Patterns

- When investigating temporal patterns, analysts should use statistical measures to understand if concentrations are statistically different.
- Testing statistical significance using T-test
  - The t-test is a very common method for assessing the difference in mean values of two groups of data (e.g., the difference in means of two years of data).
  - This test assumes that both data sets are normally distributed, a fact that is not true for many air toxics measurements. However, this is not a problem as long as there are sufficient data in each group (>~100). Each data set is also required to contain the same number of samples.
  - If there are fewer than 100 data points per group, a more advanced, non-parametric, test must be used. Some examples are
    - Kruskal-Wallis
    - Kolmogorov-Smirnov
    - Anderson-Darling (sample sizes of 10 to 40 only).
- Testing statistical significance using notched box plots
  - For the national analyses, SYSTAT notched box plots were used as a quick check of statistical significance between two groups. The notches on a box plot represent the range of the upper to lower 95th percentile confidence intervals surrounding the median (a full description of notched box plots can be found in *Preparing Data For Analysis*, Section 4, of this workbook). If the notches of two box plots do not overlap, the median concentrations are statistically significantly different.
  - Testing with notched box plots provides significance tests on the median concentration value, not the mean.
- Most of these statistical methods can be performed with Microsoft Excel or SYSTAT, as well as many other statistical programs.

StatSoft, Inc. (2005)

Characterizing Temporal Patterns

Motivation

- To more fully understand potential contributing air toxics sources, analysts may also wish to consider:
  - Diurnal patterns. How does the daily cycle of air toxics concentrations relate to emissions and meteorology? Are diurnal patterns properly reflected in exposure models?
  - Day-of-week patterns. Does the weekly cycle of air toxics concentrations tell us anything about emissions sources?
  - Seasonal patterns. Do air toxics concentrations show seasonal patterns and do these patterns make sense with respect to what we know about formation, transport, and removal processes?
- Understanding diurnal, day-of-week, and seasonal patterns may also help analysts understand potential biases in aggregated data, assess exposure, and evaluate models.

Diurnal Patterns

Overview

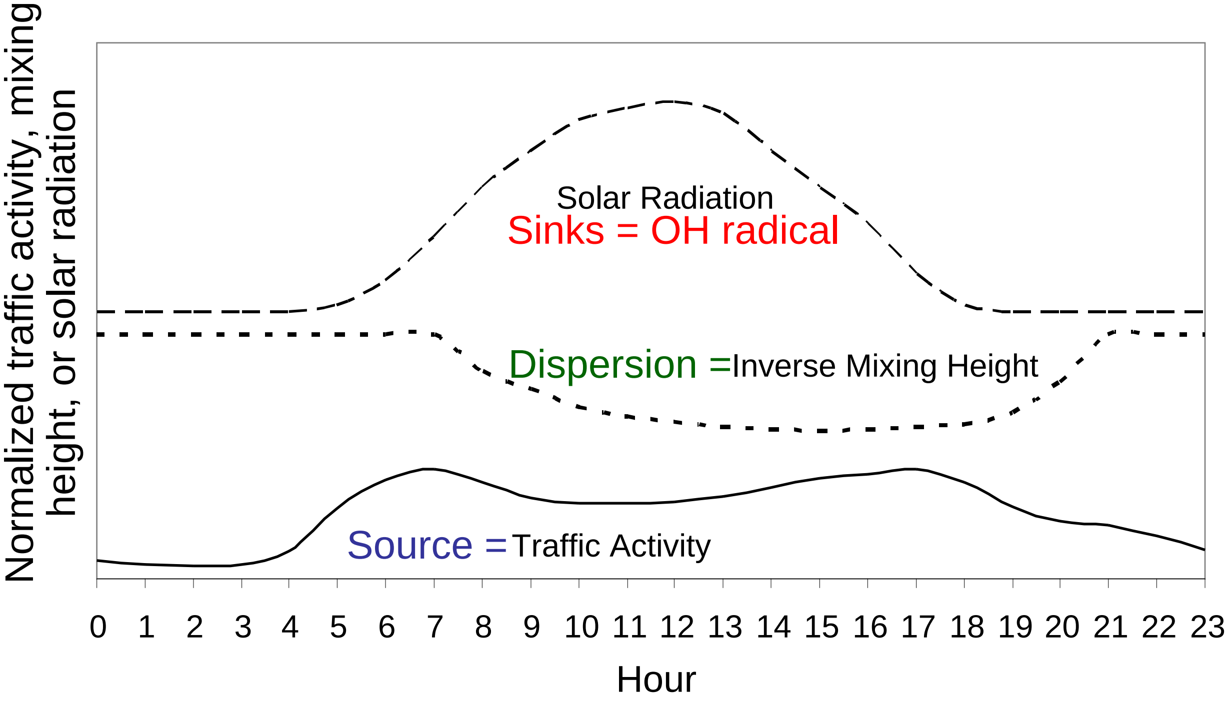
- Air toxics data are not routinely collected on a subdaily basis; most data are reported as 24-hr averages. However, the PAMS program provides subdaily measurements of nine air toxics: acetaldehyde, benzene, ethylbenzene, formaldehyde, hexane, toluene, styrene, xylenes (three isomers), and 2,2,4-trimethylpentane. The diurnal variation of some air toxics is unknown because of data limitations.
- Subdaily data allow us to:
  - Evaluate diurnal variation.
  - Understand general atmospheric processes (the physics, chemistry, and sources of air toxics).
  - Assess the performance of models that are attempting to capture diurnal cycles.
  - Provide input to receptor-based models.
- Reasons to understand diurnal patterns include
  - Assessing human exposure and health effects.
  - Identifying local sources vs. regional transport.
  - Contributing to an understanding of the physics and chemistry of air toxics.

Diurnal Patterns

Conceptual Model

- Daily concentrations are driven by dispersion (e.g., mixing height), sources (e.g., traffic patterns), sinks (e.g., oxidation by OH radical), and transport.
- Sources and transport from other areas increase concentrations at a monitor site, while sinks and dispersion reduce concentrations.
- The figure shows an example contribution of individual factors that commonly influence diurnal concentrations. The overall diurnal pattern may be driven by a combination of these factors and may be conceptually estimated in the following manner:

Concentrations = (Sources – Sinks + Transport)/Dispersion



Diurnal Patterns

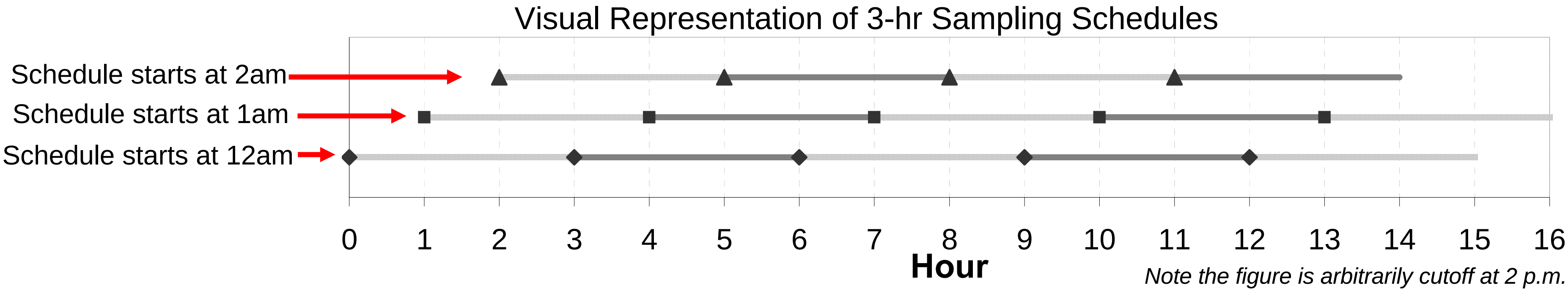
Approach *(1 of 3)*

- For the most valid diurnal patterns, the following data requirements are suggested:
  - 75% sampling completeness is recommended for each site, pollutant, and day (1) to ensure that data are representative of a full day and (2) to provide consistency with completeness requirements used to construct other aggregates (see *Preparing Data for Analysis*, Section 4).
  - Other completeness criteria (daily, monthly, yearly) may be necessary to aggregate data from multiple sites, depending on the length of time for which data are available and the objectives of the analysis.
  - The percent below detection should be tabulated for each pollutant and year. Initially, all data may be included regardless of the percent below detection.
  - To investigate diurnal patterns, there must be a sufficient number of measurements of each pollutant and sampling hour to accurately assess the value. In initial national level analyses, a minimum of 10 measurements for each air toxic and hour was set to try to include as many air toxics as possible in the analysis; more measurements are recommended if they are available.
  - Data should be inspected on both a concentration and normalized basis for each available duration. Normalization enables a comparison of diurnal patterns among sites and pollutants even if pollutant concentrations vary widely.
  - Data are normalized using the average concentration for each individual day, site, duration, and pollutant. To normalize data,
    - Calculate the average concentration by date, site, pollutant, and duration.
    - Divide the corresponding subdaily data by this average.
    - The resulting normalized values provide an indication of the magnitude of difference of the hourly concentration from the average concentration for that day. A value of 1 indicates that the hourly concentration value is the same as the daily average concentration. Values greater than one are greater than the average value (e.g., a value of 2 is 2 times greater than the average value) while values less than one are lower than the average value (e.g., a value of 0.5 is half as large as the average value).

Diurnal Patterns

Approach *(2 of 3)*

- Subdaily measurements may be made on different sampling schedules which must be taken into account when aggregating multi-site data.
  - Daily sampling schedules may differ between sites. For example, the sampling schedule for 3-hr measurements could begin at 12 a.m., 1 a.m., or 2 a.m., potentially creating three staggered hourly patterns among sites.
  - A visual representation of the possible 3-hr sampling schedules is shown in the figure below. The data points represent the sample start-time. The lines between points represent the duration of sample collection (3-hr). Subsequent sample lines are partitioned by shade for clarity.
  - Diurnal analyses can be obscured by the different sample schedules when aggregating multi-site data if the number of samples for each hour is not the same across all hours. This issue is typically not a problem within a single agency's network, but needs to be considered when data from different jurisdictions are used (such as at the national scale). Consider a hypothetical case in which Los Angeles sites used the 2 a.m. sample schedule and the rest of state used the 1 a.m. sample schedule.
  - If one considers the first three hours of the day—the sample that begins at 2 a.m. includes all three sampling schedules (i.e., all three samples overlap). For aggregating data with multiple sampling schedules, we calculated a weighted average of the hour representing the middle of staggered sampling schedules (i.e., 2 a.m. sampling schedule for 3-hr duration) from the raw data before completing the next steps.
  - A detailed example will be examined in following slides.



Note the figure is arbitrarily cutoff at 2 p.m. (14) and does not represent the whole day.

Diurnal Patterns

Approach *(3 of 3)*

- Summary statistics may be generated by pollutant and hour for the concentration and normalized data sets.
  - It is useful to inspect various parameterizations of the data (e.g., 10th, 50th, and 90th percentiles), especially when more than 50% of data is below detection.
  - Include the standard deviation or confidence interval as a measure of uncertainty in the data.
- Subdaily patterns can be visualizes the using line graphs of summary statistics with confidence intervals or notched box plots.

Diurnal Patterns

Effect of Sampling Schedule *(1 of 2)*

- The problem of staggered sampling schedules is discussed in Slides 8-10 of this section.
- This slide provides an example national 3-hr duration data set that exhibited all three possible sample schedules.
- The first table shows the raw measurements by begin-hour (i.e., the time that would be reported with the measurement).
- The second table provides the aggregated weighted averages. The formula used to calculate weighted averages, along with a calculation example, is also shown.

Table 1. Raw Measurements

Begin Hour Of Measurement	Number of Measurements	Median Concentration (µg/m^3)
0	66	0.777
1	66	0.708
2	64	0.729
3	66	0.665
4	66	0.697
5	65	0.857
6	70	0.947
7	71	0.995
8	68	0.836
9	66	0.692
10	64	0.554
11	64	0.490
12	78	0.500
13	70	0.463
14	67	0.479
15	67	0.479
16	66	0.495
17	64	0.511
18	66	0.585
19	66	0.692
20	64	0.793
21	64	0.852
22	64	0.852
23	64	0.814

Table 2. Aggregated Measurements

Aggregated Hour	Weighted Average Median Concentration (µg/m^3)
2	0.738
5	0.739
8	0.927
11	0.580
14	0.482
23	0.839

Weighted Average (WA) Formula:

$$\mathbf{WA = (1/\sum N_i)*\sum N_i C_i}$$

N = Number of Measurements

C = Concentration

Example calculation, aggregated to 2 a.m. sample schedule:

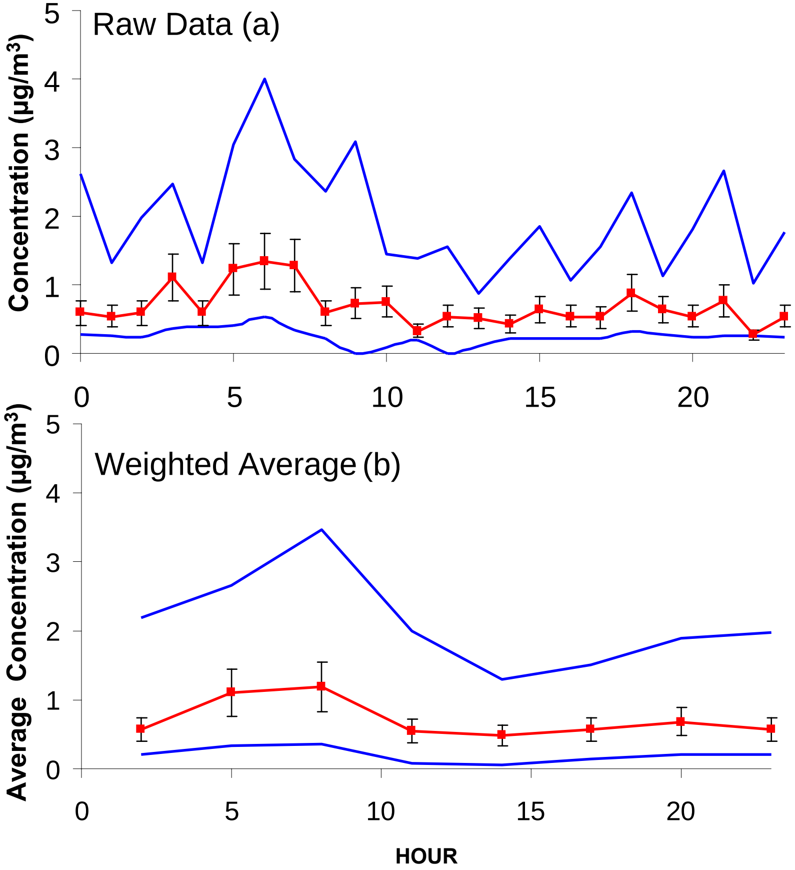
[1/(66+66+64)]*[66*0.777+66*0.708+64*0.729]
= 0.738

Diurnal Patterns

Effect of Sampling Schedule *(2 of 2)*

- The figures are a graphical representation of the calculations performed in the previous slide. (The data are not the same as those used in the previous slide.)
- Figure (a) shows the 10th, 50th, and 90th percentile of national 3-hr benzene data. The noise in this pattern is due to varying amounts of data available from three sampling schedules which begin at 12, 1, or 2 a.m. Sampling-schedule differences are typical when aggregating 3-hr or 4-hr measurements and can obscure diurnal patterns.
- Figure (b) shows the same data as a weighted average by the most representative hour. Averaging clarifies the diurnal pattern showing a morning peak trend as would be expected for benzene concentrations at most sites.
- This averaging method is recommended when aggregating multi-site data if multiple sampling schedules are used.

Benzene 3-hr Subdaily Data

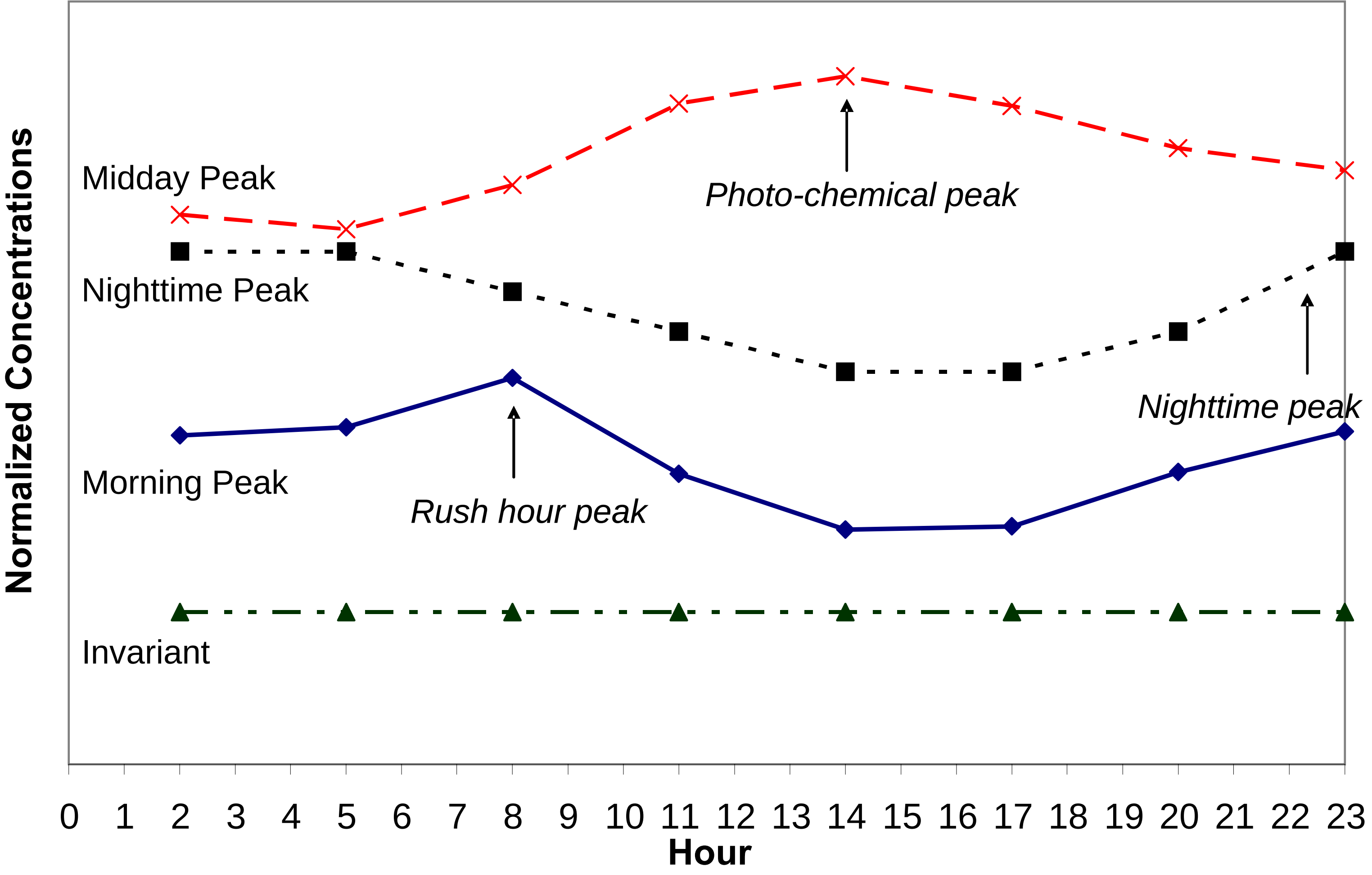


Figures show the 10th, 50th and 90th percentile of national 3-hr benzene data. They were created with SYSTAT11 and Microsoft Excel.

Diurnal Patterns

Commonly Observed Patterns

The figure shows a sample of four commonly observed diurnal patterns using national 3-hr duration data. The sources, sinks, transport, and dispersion leading to each pattern are discussed in this section. *Data were normalized as described in the approach to diurnal patterns.*



Diurnal Patterns

Morning Peak

- Morning peak patterns are observed from the combination of traffic emissions and mixing height dilution.
- The morning rush hour occurs while mixing heights are relatively low, causing a peak in concentration while emissions outweigh dilution.
- By mid-morning, mixing height dilution has outweighed traffic emissions, reducing concentrations below their nighttime value and obscuring the remaining traffic emission patterns.
- Evening concentration increases are a consequence of mixing height lowering.

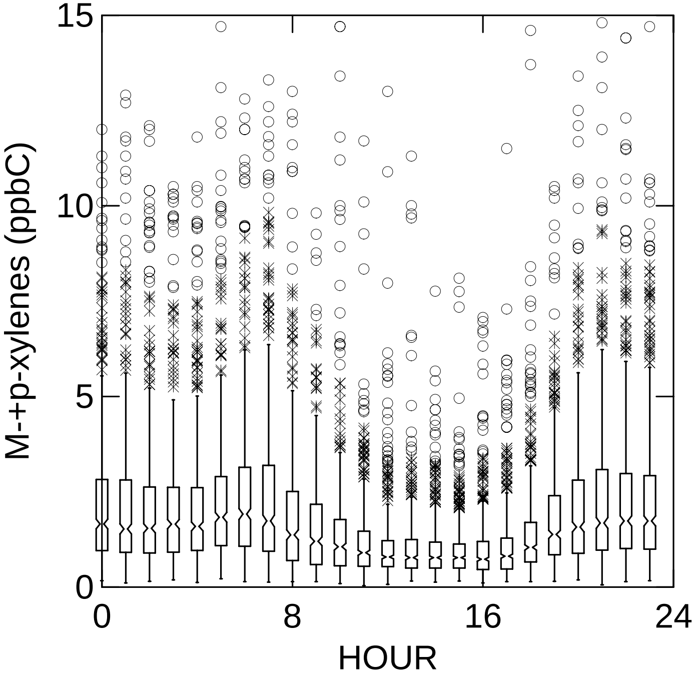
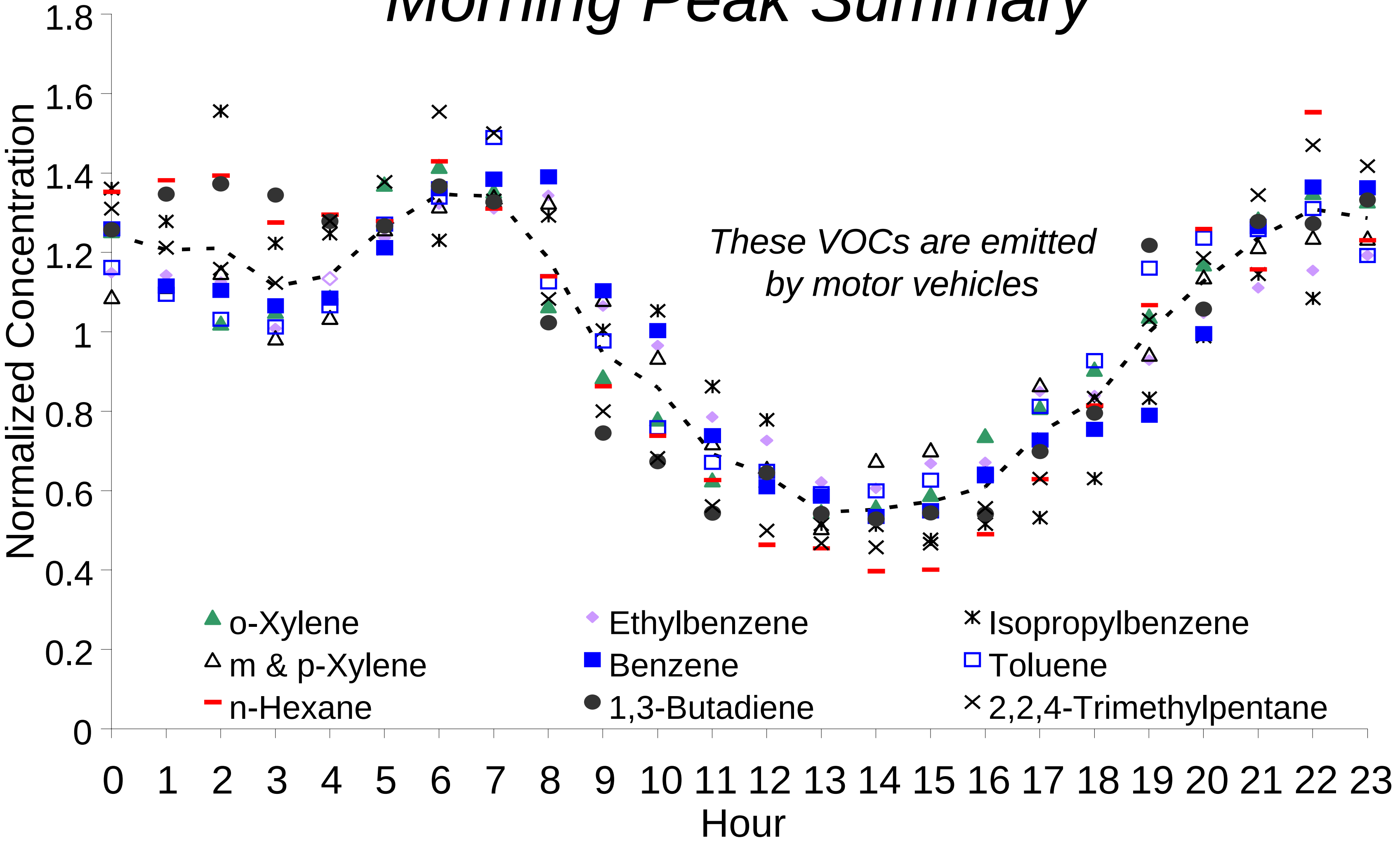


Figure shows notched box plot of m-&p-xylenes concentrations by hour at an urban site. Box plots are defined in *Preparing Data for Analysis,* Section 4. Several years of data are included. The plot was created with SYSTAT11.

Diurnal Patterns

Morning Peak Summary

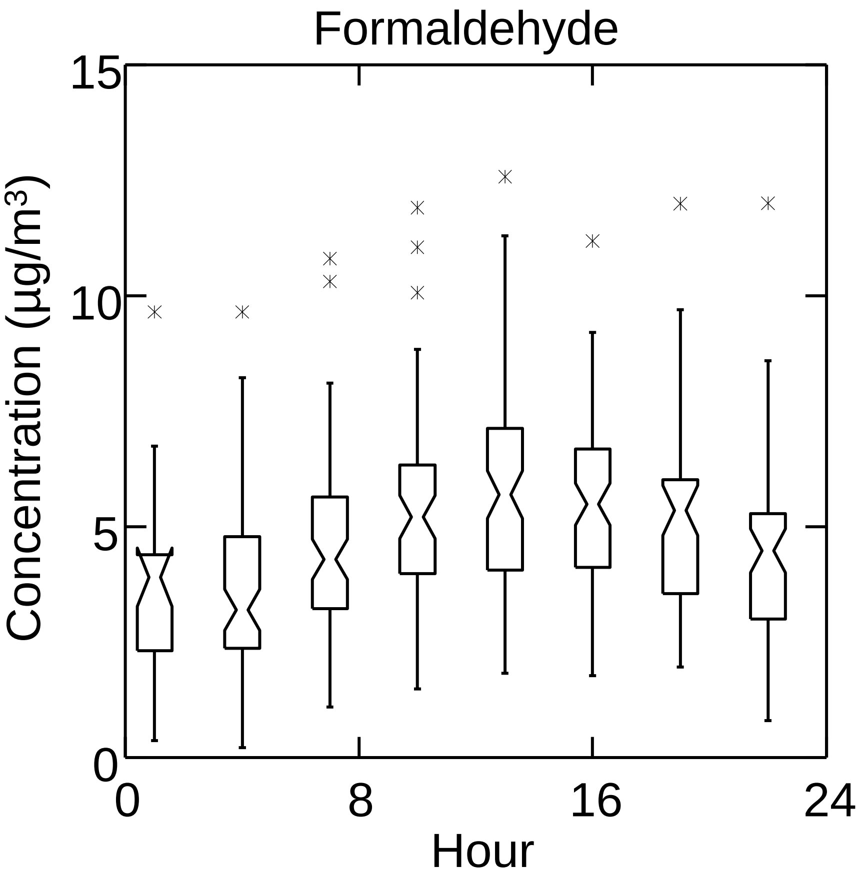


This figure shows 1990-2005 national hourly data normalized by site, pollutant, and day for all pollutants that exhibited a morning peak pattern on the national scale. Data were normalized as described in the approach to diurnal patterns.

Diurnal Patterns

Daytime Peak

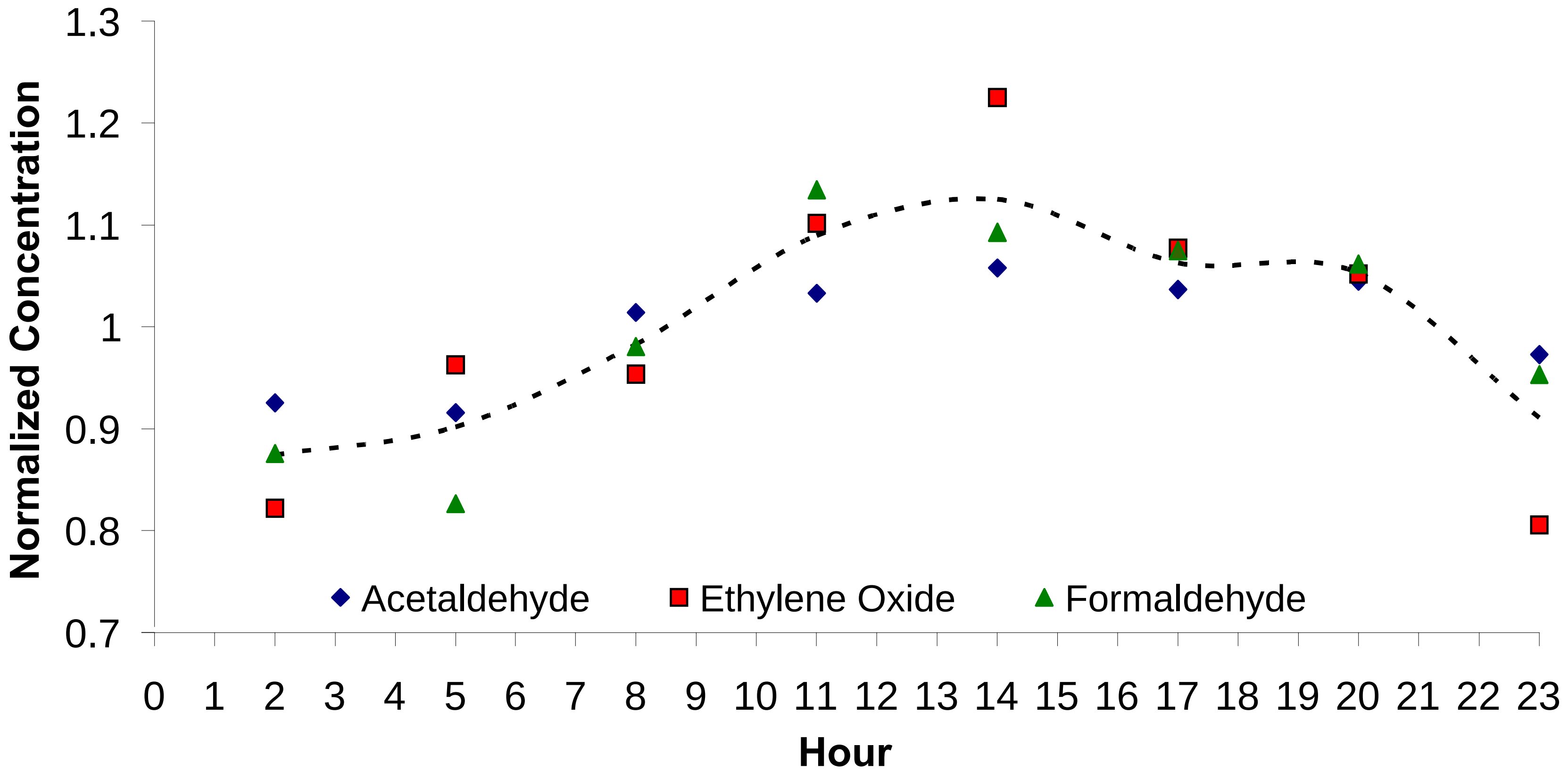
- The daytime pattern is driven by *in situ* secondary photochemical production mechanisms and mirrors the pattern of solar radiation.
  - Precursors of afternoon peak pollutants are typically emitted by motor vehicle sources and OH sinks. Afternoon peak pollutants experience daily dilution patterns in a manner similar to morning peak pollutants.
  - Secondary production of a pollutant (such as formaldehyde) must outweigh all these factors in order to create the observed pattern.



The figure shows notched box plots of national 3-hr formaldehyde concentrations by the middle sampling schedule (as discussed in Slides 8-10). The figure was created with SYSTAT11.

Diurnal Patterns

Daytime Peak Summary

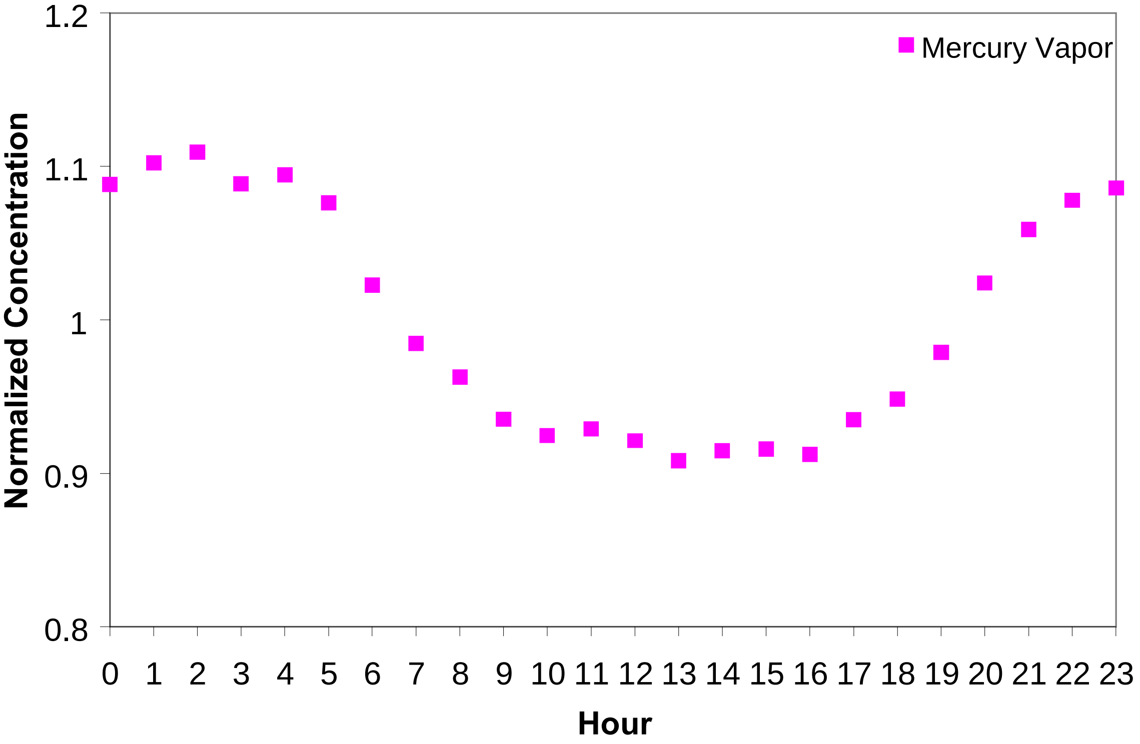


The figure shows 1990-2005 national-scale 3-hr duration data normalized by site, pollutant, and day for all pollutants that exhibit an afternoon peak pattern.

Diurnal Patterns

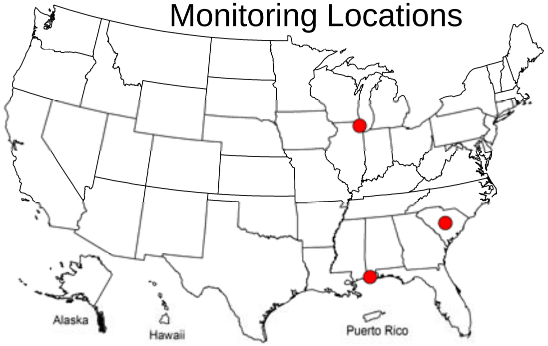
Evening Peak

- Mercury vapor is the only air toxic to exhibit a clear evening peak pattern in the air toxics investigated at the national level. However, data from only a few sites were available so this analysis may not be representative of a national pattern.
- Dilution appears to be the key factor affecting evening peak pollutants; emissions and sinks are likely invariant at the subdaily level.



1990-2005 national hourly mercury vapor data normalized by site, pollutant, and day. The figure was created with Microsoft Excel.

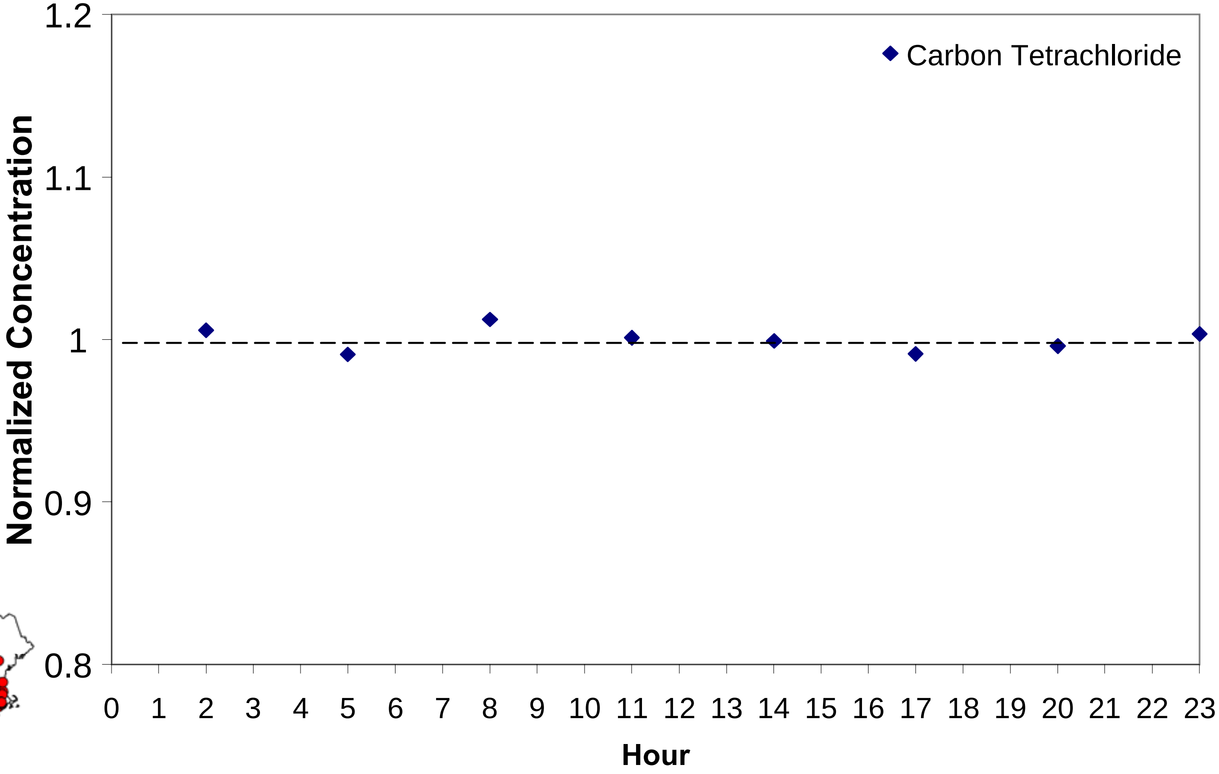
Mercury Vapor Monitoring Locations



Diurnal Patterns

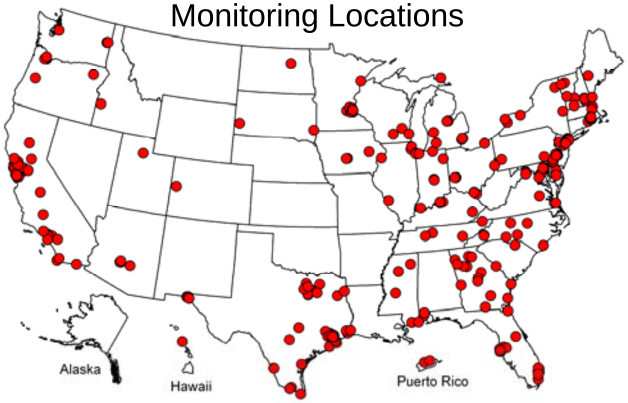
Invariant

- Invariant patterns are observed for global background pollutants (i.e., pollutant is no longer emitted).
- These pollutants show no sources or sinks and are evenly distributed worldwide so that transport and dilution have no effect on concentration.



The figure shows 1990-2005 national 3-hr carbon tetrachloride data normalized by site, pollutant, and day. Carbon tetrachloride is the only pollutant to exhibit an invariant diurnal pattern on the national scale. The figure was created with Microsoft Excel.

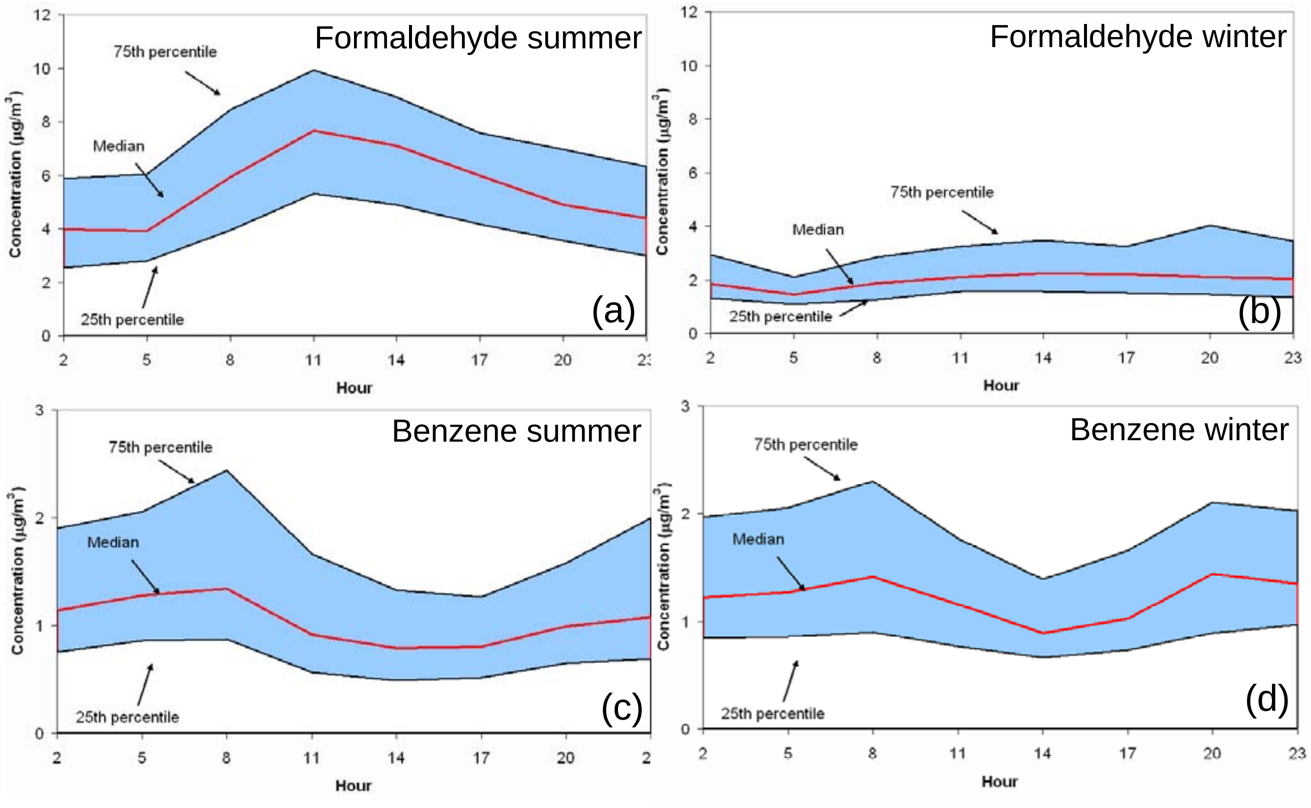
Carbon Tetrachloride Monitoring Locations



Diurnal Patterns

Seasonal Differences

- Seasonal differences may be observed in the diurnal patterns of some air toxics.
- For example, the diurnal pattern of formaldehyde on a national scale is highly affected by season, as seen in Figures a and b, because the main production of formaldehyde depends on sunlight which is less abundant in winter months; thus, midday production decreases significantly during these months.
- The diurnal pattern of benzene shows less seasonal dependence because it is driven by diurnal meteorology that is consistent throughout the year and benzene is less photochemically reactive (Figures c and d).



Figures show summary statistics of national diurnal patterns for formaldehyde and benzene partitioned into summer and winter patterns. Figures were created with Microsoft Excel.

Diurnal Patterns

Summary

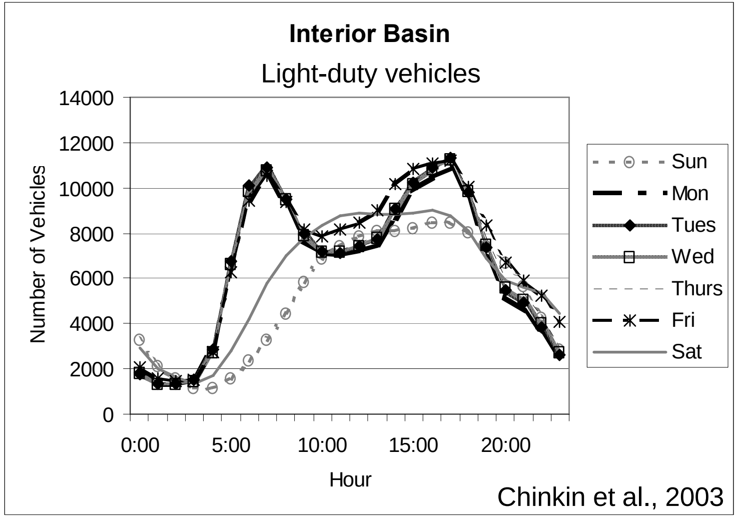
- Diurnal patterns of air toxics are influenced by sources, sinks, and dispersion processes that vary on a subdaily basis.
- Diurnal patterns are useful in classifying source type, transport, and reactivity of air toxics. These patterns can be used to improve exposure modeling, air quality modeling, and emissions inventories.
- Most air toxics data typically follow four diurnal patterns although many air toxics have not been characterized because of sampling and detection limitations.
    - <u>Morning peak</u>. Driven by mobile source emissions and mixing height dilution
    - <u>Afternoon peak</u>. Driven by secondary photochemical production
    - <u>Nighttime peak</u>. Driven by mixing height dilution
    - <u>Invariant</u>: Typical of global background pollutants that are not dependent on sources, sinks, transport, or dilution.
- If the diurnal pattern of a pollutant differs from the typical patterns shown at a national level, the analyst should explore possible reasons for the variation such as the presence of a nearby source.

Day-of-Week Patterns

Overview and Conceptual Model

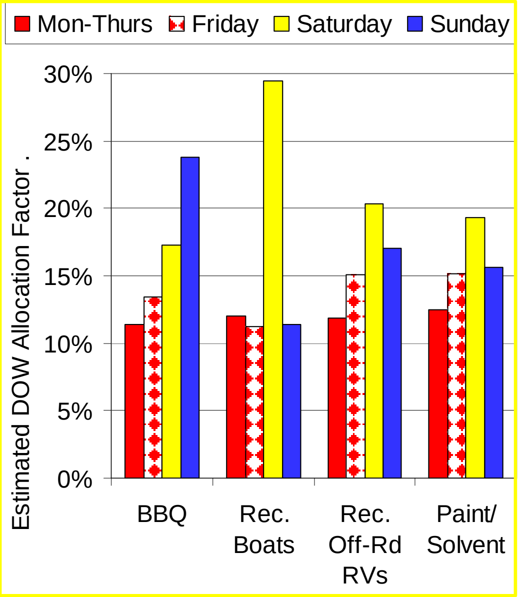
- Day-of-week patterns can be useful in identifying emissions sources.
- Expectations
  - Emission sources that operate every day, 24 hours per day (e.g., refineries) will not show a day-of-week pattern.
  - Emission sources with lower emissions on weekends should lead to lower ambient weekend concentrations of the emitted air toxics. Traffic studies (e.g., Chinkin et al., 2003) show that in many cities, light-duty vehicle activity is lower on Sunday compared to other days of the week (Figure a).
  - Emission sources with higher emissions on weekends should lead to high ambient weekend concentrations of the emitted air toxics. For example, studies in the Los Angeles area showed that recreational vehicle emissions may be higher on Saturdays (Figure b).

Los Angeles



Chinkin et al., 2003

Estimated allocation of residential emissions activity by day of week in Los Angeles (Coe et al., 2003)



Day-of-Week Patterns

Approach

- Day-of-week patterns are typically constructed from 24-hr averages. See *Preparing Data for Analysis,* Section 4, for a complete description of how to construct valid averages.
  - If subdaily data are available, it is sometimes useful to look at data subsets. For example, when creating day-of-week trends of an air toxic that exhibits morning peak diurnal patterns, the rush hour peak data subset (i.e., 6 to 9 a.m.) will provide more information about the mobile source signature than the 24-hr average. Mobile source signatures typically show day-of-week patterns, while mixing height dilution will occur on any day of the week. 24-hr averages will be more heavily weighted by mixing height dilution and may obscure mobile source day-of-week trends.
- A sufficient number of records for each day of the week is needed to create a representative day-of-week pattern. The actual data requirements will vary depending on the analysis types and variability of the data, among other factors.
  - Statistically, decreasing the sample size increases the confidence interval (CI). In general, if the 95% CIs of two data subsets (e.g., weekend vs. weekday concentrations) do not overlap, there is good evidence that the subset population means are different; therefore, it will be more difficult to discern statistically significant patterns with smaller sample sizes.
  - Quantify patterns using the statistical treatments described earlier in this section.
- Investigate the day-of-week pattern of multiple statistics (e.g., 10th, 50th, and, 90th percentile) with the standard deviation or confidence intervals as a measure of uncertainty.
- If data are insufficient for each day to determine a pattern, weekday vs. weekend patterns may be investigated.

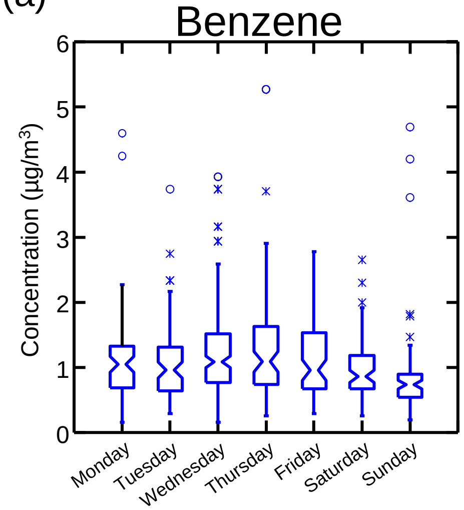
Day-of-Week Patterns (1 of 2)

Example

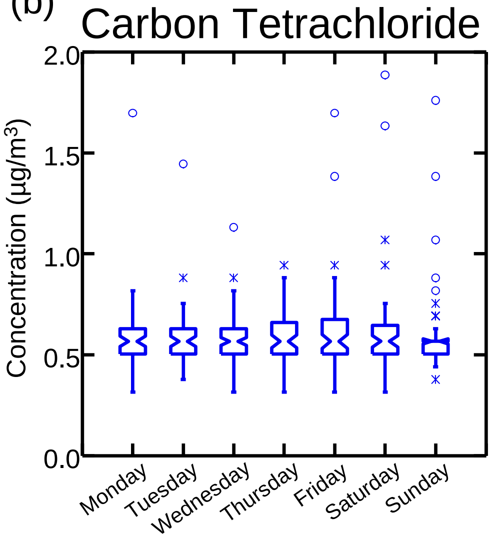
- In Figure (a), benzene concentrations at an urban site are statistically significantly lower on Sunday. The concentrations on Saturday seem slightly lower, but differences are not statistically significant. These results are consistent with our conceptual model of light-duty vehicle traffic.
- For carbon tetrachloride (Figure b), we expect concentrations to be the same every day. The central tendencies of the concentrations at the same site are consistent.

The figures show notched box plots of 24-hr concentrations by day of week at selected sites. They were created with SYSTAT11.

(a)



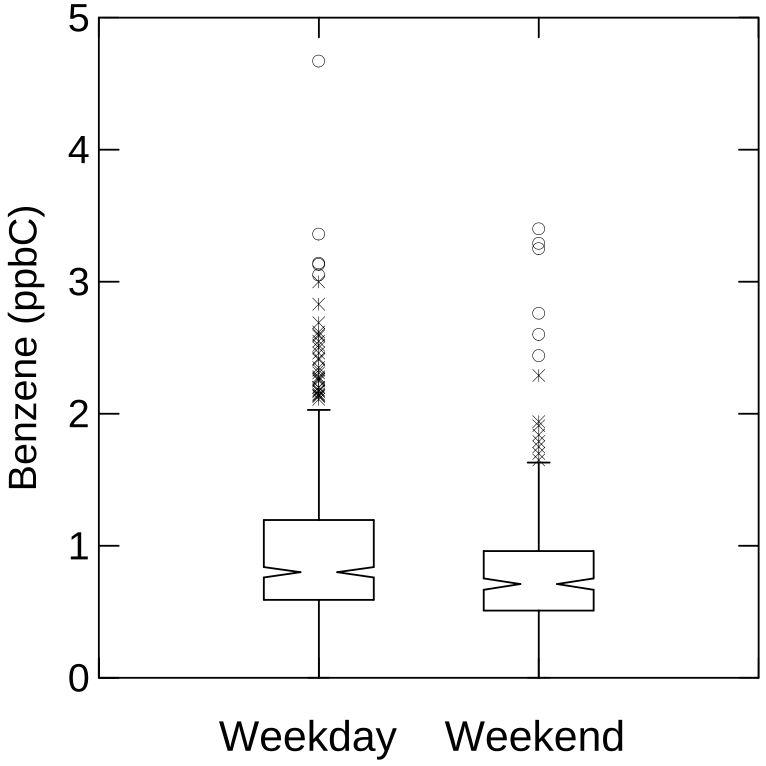
(b)



Day-of-Week Patterns (2 of 2)

Example

- Sometimes, not enough data are available to determine patterns by day of week—in some cases, the data can be combined into weekday vs. weekend groups.
- In the example, benzene concentrations at an urban site are lower on weekends than on weekdays (the difference in medians is statistically significant). These findings make sense because of the urban location of the monitor and lower motor vehicle emissions on the weekend compared to weekdays.
- The inspection of day-of-week patterns of all air toxics was not performed at a national level.



The figure shows a notched box plot of 1-hr benzene concentrations on weekdays vs. weekends at an urban site. All time periods were included—and weekend concentrations are statistically significantly lower than weekday concentrations. The figure was created with SYSTAT11.

Day-of-Week Patterns

Summary

- Typically, mobile source air toxics show the most obvious day-of-week pattern consistent with traffic patterns. Sunday concentrations were particularly low for most mobile source air toxics, a pattern consistent with reduced traffic.
- In general, day-of-week patterns can be difficult to discern due to interference from other sources, sinks, or meteorology.
- A low number of samples can obscure underlying patterns.
- In exploratory investigations of national-level data, few non-mobile source air toxics showed a clear day-of-week pattern.
- Note that day-of-week patterns are highly dependent on the proximity of the monitor's site to sources, the emission sources' schedule, and meteorology (e.g., wind direction); site-level examinations may provide a better explanation.

Seasonal Patterns

Overview

Understanding seasonal differences in air toxics concentrations helps analysts

- Formulate or evaluate a conceptual model of emissions, formation, removal, and transport of an air toxic.
- Better understand source types.
- Continue to validate data, i.e., do data meet expectations for seasonal variation?
- Construct and interpret annual averages when a season’s data are missing from the average (e.g., if the data for a winter quarter are missing, what biases in the annual average can be expected?).

Seasonal Patterns

Conceptual Model

- Cool season expectations
  - Mixing heights are lower in the cold months. Low mixing heights create less air available for pollutant dispersion which causes higher ambient concentrations.
  - Temperatures are lower and sunlight is reduced in cold months. This combination can lead to a reduction in evaporative emissions (e.g., gasoline) and reduced photochemistry. Reductions in temperature and sunlight also limit formation of hydroxyl radicals which efficiently oxidize many air toxics.
  - Typically more precipitation occurs during winter months and reduces dust emissions.
- Warm season expectations
  - Mixing heights are higher in warm months, allowing more dilution and transport of air toxics which, in turn, reduces ambient concentrations.
  - Higher temperatures and increased sunlight in warm months lead to an increase in evaporative emissions and photochemistry.
  - Conditions are typically drier, producing more dust.
  - Wildfire activity can also cause an increase in concentrations of pollutants emitted in smoke.

Seasonal Patterns

National Trends

- Seasonal patterns observed at a national level are shown in the table.
- These air toxics were selected because they were the ones with sufficient data for analyses.
  - Minimum of three valid seasonal averages by site and year
  - At least 20 monitoring sites meeting the above criteria
  - Additionally, limited to pollutants investigated in diurnal variability and annual analyses to focus on similar pollutants.
- Most of the VOCs, with the exceptions of styrene and isopropylbenzene, are cool season pollutants as expected.
- We are not sure why carbon tetrachloride shows a warm season peak—we expected it to be invariant. No obvious data issues suggested this pattern.

Pollutant Name	Pattern	Number of sites	Median CV	Median annual concentration (μg/m^3)
1,3-Butadiene	Cool	195	0.38	0.16
n-Hexane	Cool	159	0.30	0.88
2,2,4-Trimethylpentane	Cool	119	0.29	0.51
m- & p-Xylene	Cool	256	0.29	1.10
Tetrachloroethylene	Cool	137	0.29	0.26
Toluene	Cool	137	0.29	2.38
o-Xylene	Cool	261	0.28	0.46
Ethylbenzene	Cool	262	0.28	0.42
Benzene	Cool	306	0.27	1.03
Lead TSP	Cool	149	0.25	0.018
Dichloromethane	Cool	187	0.25	0.44
Styrene	Indeterminate	207	0.33	0.16
Isopropylbenzene	Indeterminate	91	0.31	0.068
Methyl Chloroform	Invariant	89	0.12	0.15
Chloromethane	Warm	245	0.09	1.20
Carbon Tetrachloride	Warm	240	0.09	0.56
Nickel TSP	Warm	44	0.20	0.0026
Manganese TSP	Warm	71	0.20	0.015
Chromium TSP	Warm	61	0.21	0.0039
Acetaldehyde	Warm	163	0.21	1.65
Propionaldehyde	Warm	112	0.27	0.28
Chloroform	Warm	102	0.29	0.123
1,4-Dichlorobenzene	Warm	97	0.32	0.19
Formaldehyde	Warm	163	0.36	2.75

McCarthy et. al, 2007

Seasonal Patterns

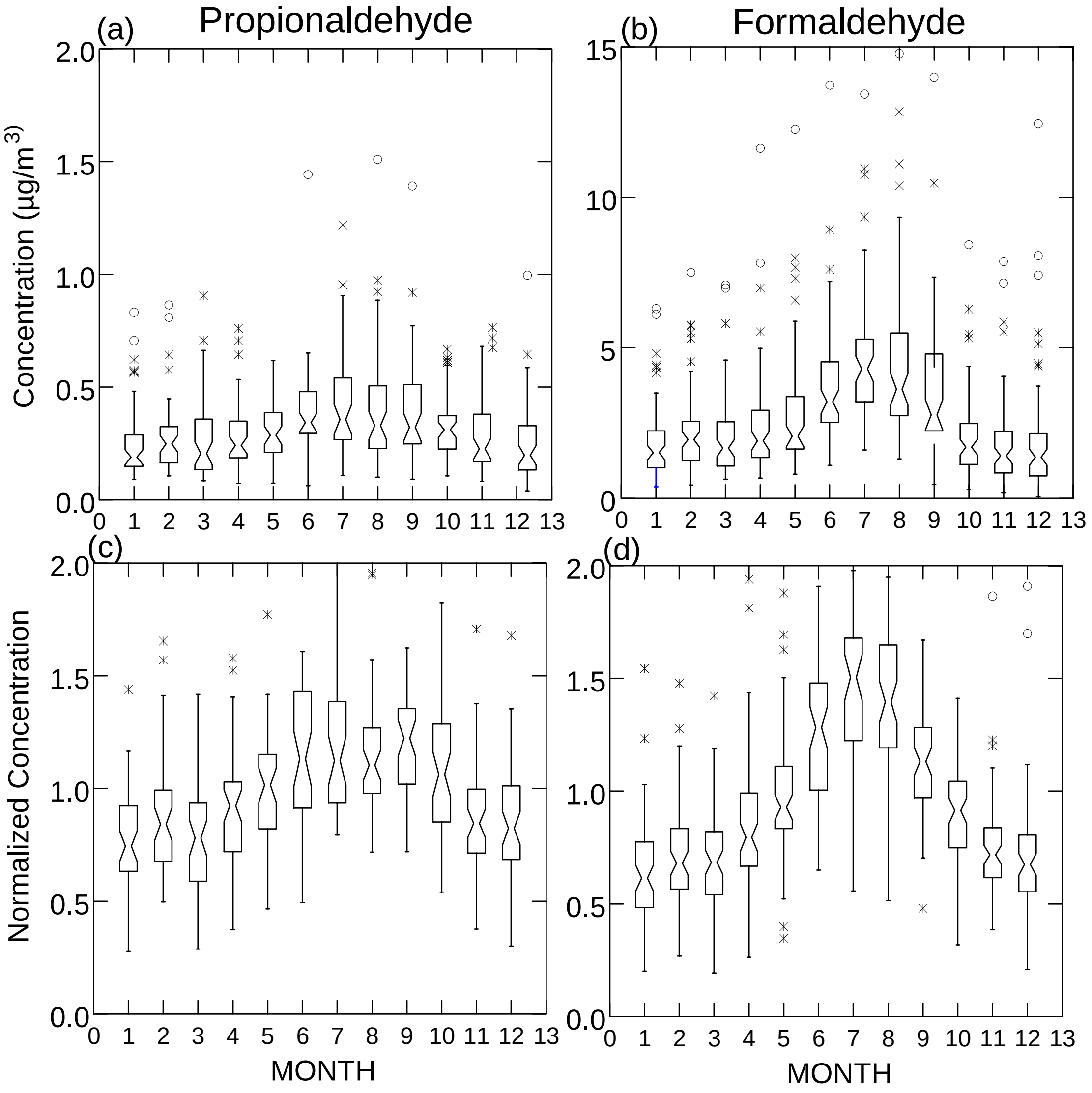
Approach

- Investigation of seasonal variability patterns using normalized monthly and/or quarterly averages.
    - See *Preparing Data for Analysis,* Section 4, for a complete description of how to construct valid monthly and quarterly averages.
    - Quarterly averages may be calendar quarters or seasonal quarters depending on the aim of analyses.
- Keep track of the percentage of data below detection; pollutants and years with >85% of data below detection result in too much bias to draw conclusions.
- Preferably, inspect monthly data for seasonal patterns if sufficient data are available.
    - Noise in monthly data may be high due to fewer measurements. For this reason, investigating quarterly (or specific monthly groupings relevant to the site) data in addition to monthly data can be useful.
    - Area-specific seasonal aggregations can be made.
- Normalize the data using the average value for each year, site, and pollutant.
    - Calculate an annual average for each year, site, and pollutant.
    - Divide the corresponding monthly or quarterly average by the annual average.
- Investigate seasonal patterns of normalized data using notched box plots or summary statistics with a measure of confidence (e.g., standard deviation or confidence intervals).

Seasonal Patterns

Using Normalized National-Scale Data

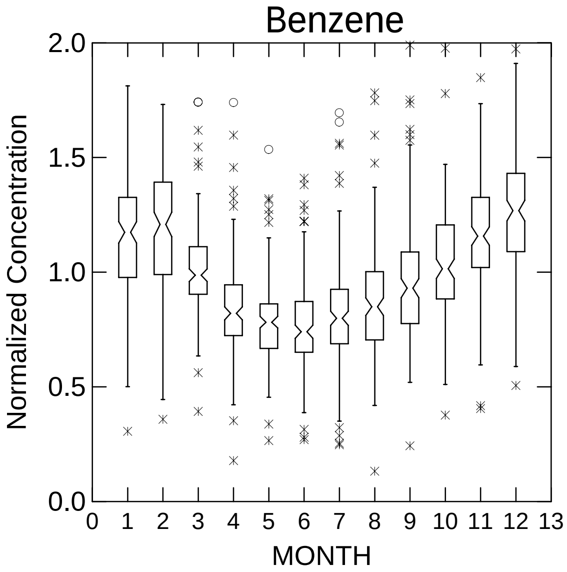
- To illustrate the use of normalized data, consider the monthly patterns of propionaldehyde and formaldehyde, both of which show concentrations that appear higher in summer (Figures a and b).
- However, normalized concentration patterns (Figures c and d) show that the monthly pattern of formaldehyde is more significant than that of propionaldehyde.
- On a relative basis, Figures c and d show that concentrations of formaldehyde are nearly three times higher in the summer than in winter.



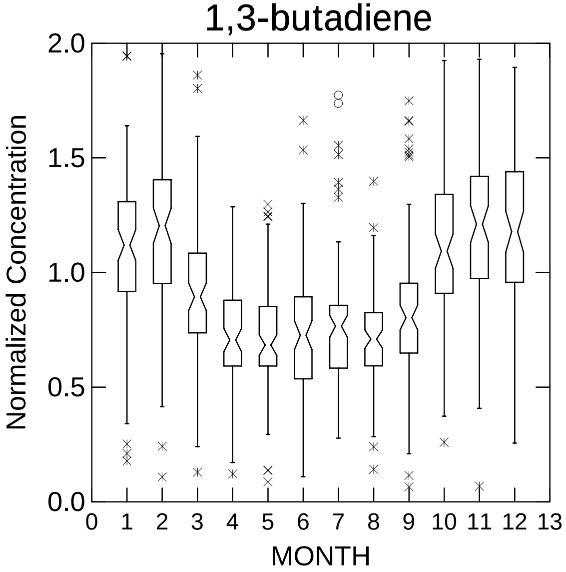
Seasonal Patterns

Cool Season Peak

- Cool seasonal patterns are generally observed because mixing heights are lower in winter and the enhanced removal by photooxidation observed during summer is absent.
- Heating-related emissions, such as wood burning, will typically be higher during winter months, contributing to increased concentrations of some air toxics.
- Benzene and 1,3-butadiene, two mobile source air toxics, show cool season peaks on the national scale.



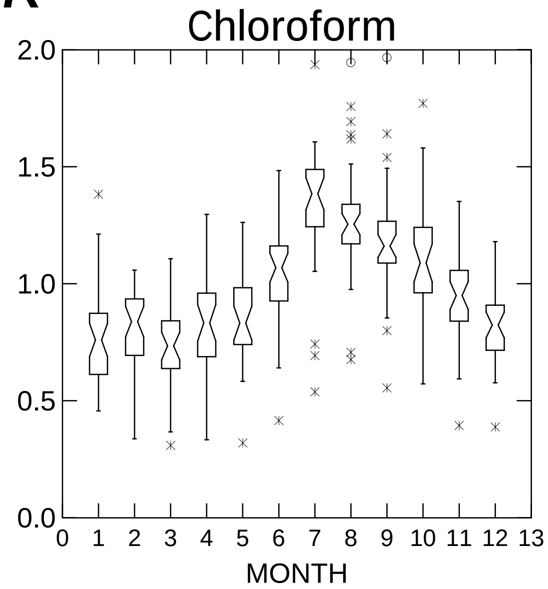
Figures show normalized monthly national concentration distributions for 2003-2005. Figures were created with SYSTAT11.

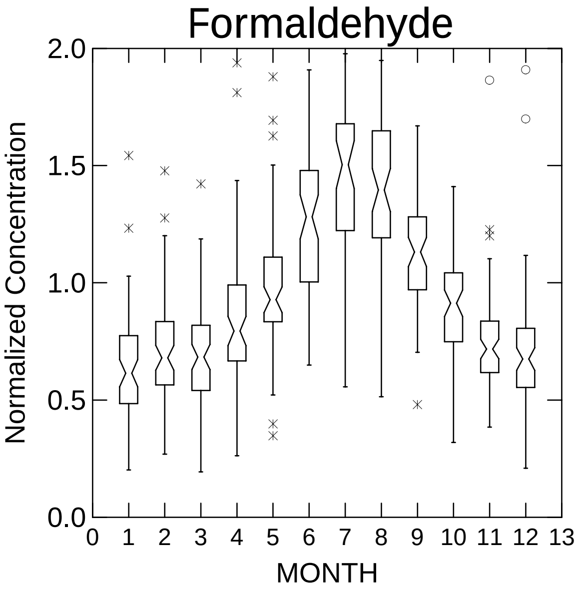


Seasonal Patterns

Warm Season Peak

- To display a warm peak pattern, summertime sources (emissions or secondary production) must significantly outweigh the higher mixing heights that occur during warm months.
- Chloroform emissions from water treatment processes and swimming pools may be enhanced during summer months, explaining the observed pattern.
- It has been estimated that 85-95% of formaldehyde concentrations originate from secondary photochemical production, which supports the observed warm season peak (Grosjean et al., 1983).





Figures show normalized monthly national concentration distributions for 2003-2005. Figures were created with SYSTAT11.

Seasonal Patterns

A National Perspective

- The figure shows the 10th, 50th, and 90th percentiles of national 2003-2005 normalized seasonal concentrations for selected pollutants by calendar quarter. Similar plots, such as regional summaries, can be prepared for any combination of sites.
- Parameters at the top of the figure show warm season peaks while those at the bottom show cool season peaks.
- Warm season peaks are likely due to secondary photochemical production and dust; it is unclear why carbon tetrachloride shows a warm season peak.
- Cool season peaks are primarily due to lower mixing heights in the winter.

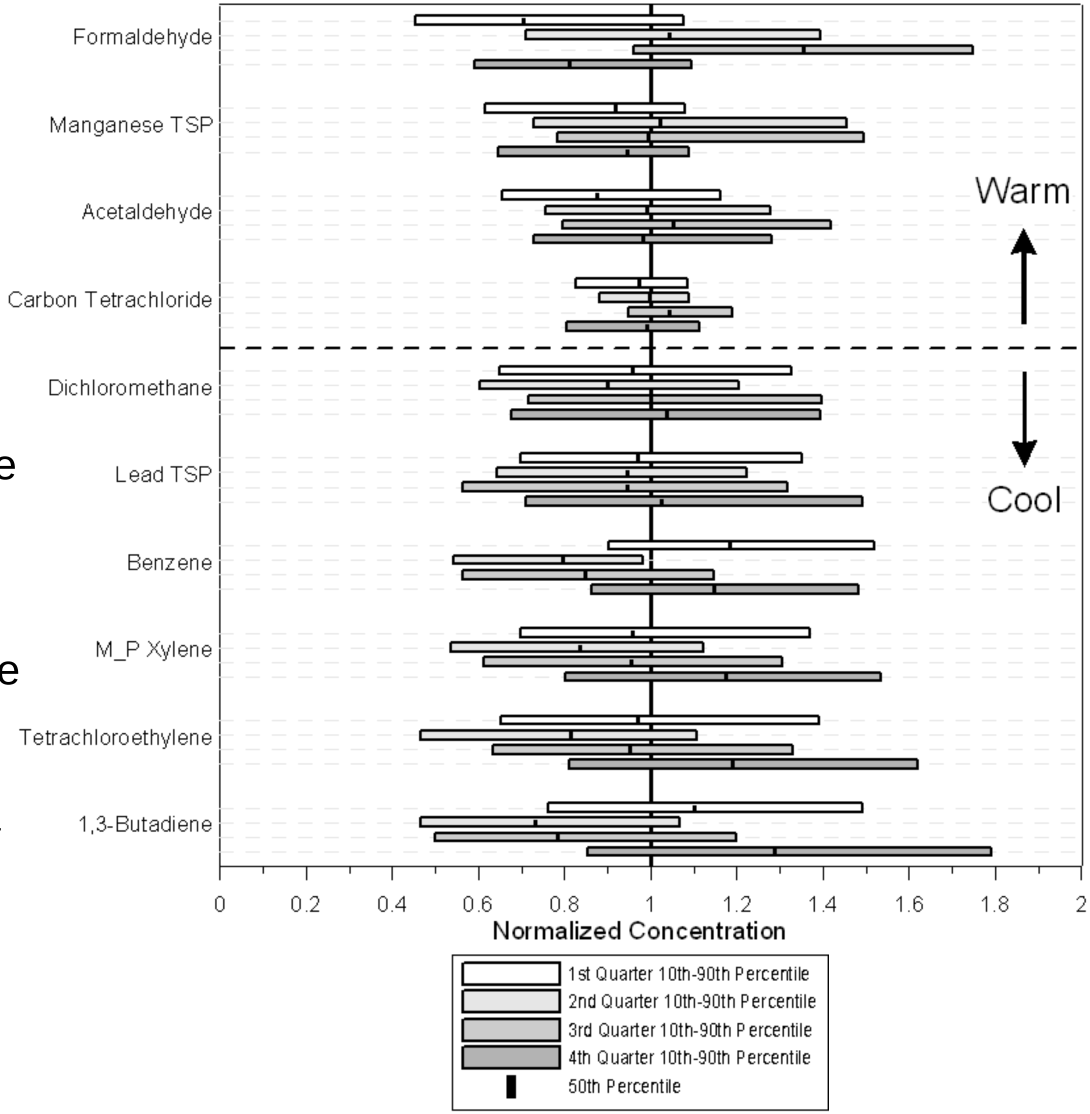


Figure created with Grapher.

Seasonal Patterns

Summary

- Three seasonal patterns were observed at a national level
  - Warm season peak. Photochemical production of secondary air toxics (e.g., formaldehyde and acetaldehyde) can be important at some sites. Concentrations (e.g., manganese) may also be high because of dust events and seasonally increased emissions (e.g., chloroform).
  - Cool season peak. Concentrations can be high because of lower inversions, changes in emissions through the use of wood-burning or fuel oil for home heating, and reduced photochemical reactivity.
  - Invariant. Invariant seasonal patterns are not commonly observed, but are typical of global background pollutants that are not affected by emissions changes or dilution which cause seasonal patterns of other air toxics.
- The quality of many air toxics data was low or seasonal patterns inconsistent at the national level; site level investigations may reveal additional seasonal patterns.
- Seasonal patterns assist in air toxics data analysis by providing insight into the chemistry, sources, and transport of air toxics. Deviation from expected seasonal patterns at a site may indicate additional sources of interest or transport.

Spatial Patterns

Overview

- Air toxics data are typically collected in urban locations. Given the large number of air toxics, their often disparate sources, and the wide range of chemical and physical properties, understanding spatial patterns and gradients is important.
- Understanding these gradients may help us
  - Improve monitoring networks, (Are we measuring in the right places to meet network objectives? Do we have the right number of monitors?)
  - Improve emission inventories. (How finely do emissions need to be spatially allocated?)
  - Improve models, including exposure models. (Are gradients in pollutants being properly represented in the model?)
  - Identify contributing sources. (Are concentrations higher when winds are predominantly from the direction of a source?)

Spatial Patterns
Conceptual Model

The concentration of a given species at any location is determined by local production, local sinks, and transport.

- *Production.* Local emissions—higher emissions lead to higher concentrations.
- *Loss*. Local removal (chemical or deposition)—reactive compounds and large particles are removed faster resulting in lower concentrations.
- *Transport*. Movement of species in the atmosphere—pollutants from sources are dispersed or diluted; local concentrations can either increase or decrease.

$$\frac{d(\text{Concentration})}{dt} = \text{Production} - \text{Loss} + \text{Transport}$$

Spatial Patterns

Methods

- To investigate spatial patterns, calculate one site average value for each air toxic for the time period of interest. This method removes temporal variability and focuses on spatial patterns.
  - The method is only valid if sites are temporally comparable. If not, results may be driven by a mixture of temporal and spatial patterns and will be difficult to interpret.
  - Averages should be constructed from valid aggregates. For example, if data are available for 2003-2005, you might first calculate the three valid annual averages then aggregate these averages to one site average. If data are not sufficient to create valid annual averages use valid seasonal or monthly averages. Note that site average values may be biased by temporal patterns if data are not representative of the full year. Relative spatial comparisons are still valid as long as data are available for all sites during the same time period.
  - If possible, multiple years of data should be used in order to mitigate meteorological effects.
  - Keep track of the percent of data below detection for each site average.
- Visualize concentration ranges by plotting summary statistics for each pollutant.
  - These plots give an overview of concentration values.
  - Supplementary data, such as levels of concern for increased cancer or noncancer risk (i.e., health levels of concern), remote background concentrations, and method detection limits (MDLs), are useful to put concentration data into perspective.
- Visualize site level concentrations using a mapping program to overlay supplementary data, such as the percent of data below detection, to enrich conclusions.
- The visualization methods may illuminate site-level data anomalies which become apparent upon comparison to other sites.

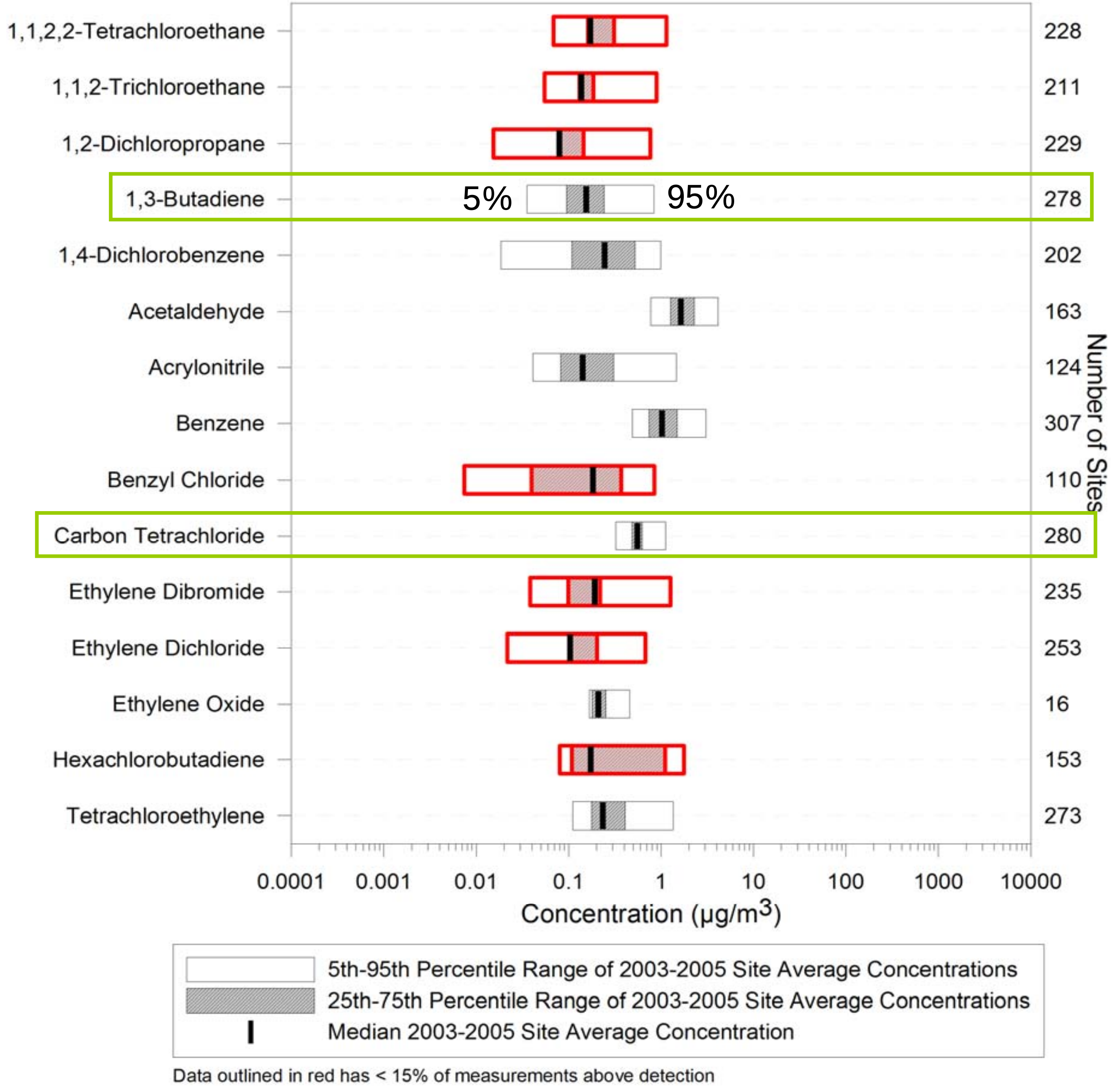
National Concentration Plots

Overview

- To put air toxics concentrations measured at a site or sites in perspective, a summary of the typical national concentration ranges is useful.
- The following national site average concentrations for 2003-2005 air toxics concentrations exemplify one way of visualizing summary statistics and supplementary data.
  - Are concentrations high, typical, or low?
  - How does this concentration compare to remote background? To MDL? To levels of concern?
- The following figures show the 5th, 25th, 50th (median), 75th, and 95th concentration ranges by pollutant; supplementary data are then overlaid as a progression. Wide ranges in concentration across sites indicate greater spatial variability of that pollutant.
- The number of sites included are shown on the right axis for each pollutant.
- Pollutants outlined in red represent <15% of samples nationally above their respective MDLs. The distribution of concentrations for these pollutants are mostly based on MDL/2 and should not be considered quantitative. Data used for these plots is included in *Preparing Data for Analysis,* Section 4.

All perspective plots were created in Grapher.

National Concentration Plots

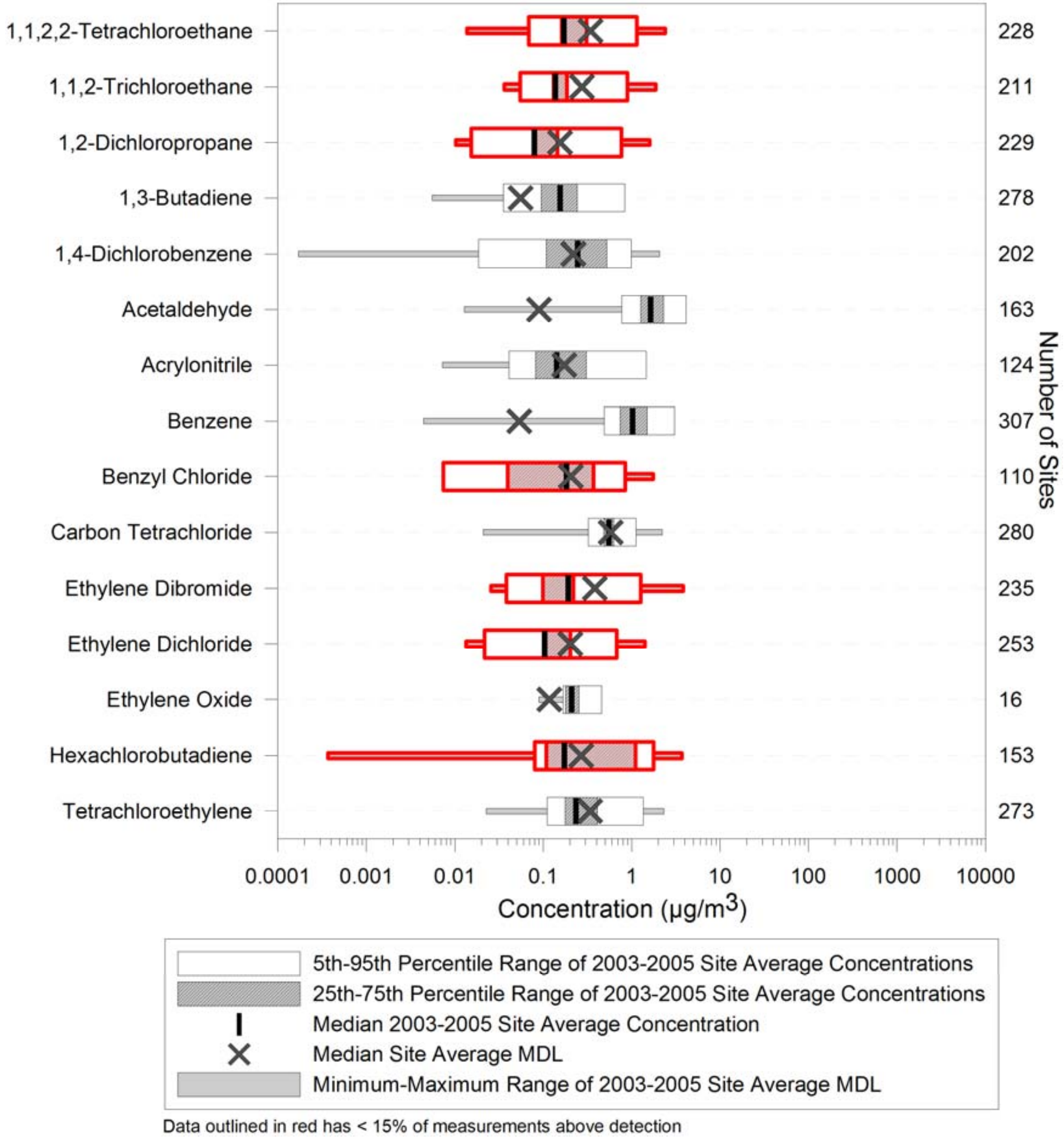


Data outlined in red has < 15% of measurements above detection

Interpretation

- Summary plots provide an overview of the spatial variability of, and a comparison within and between, air toxics. Spatial variability is represented by the width of the bar—nationally, air toxics concentrations typically varied by a factor of 3 to 10.
- The figure shows the high spatial variability of 1,3-butadiene. This variability is due to the relatively high reactivity of the compound.
- Conversely, carbon tetrachloride shows less spatial variability due to its low removal rate from the atmosphere and the absence of domestic emissions.
- A table of national concentration summary statistics can be found in the appendix to *Preparing Data for Analysis*, Section 4.

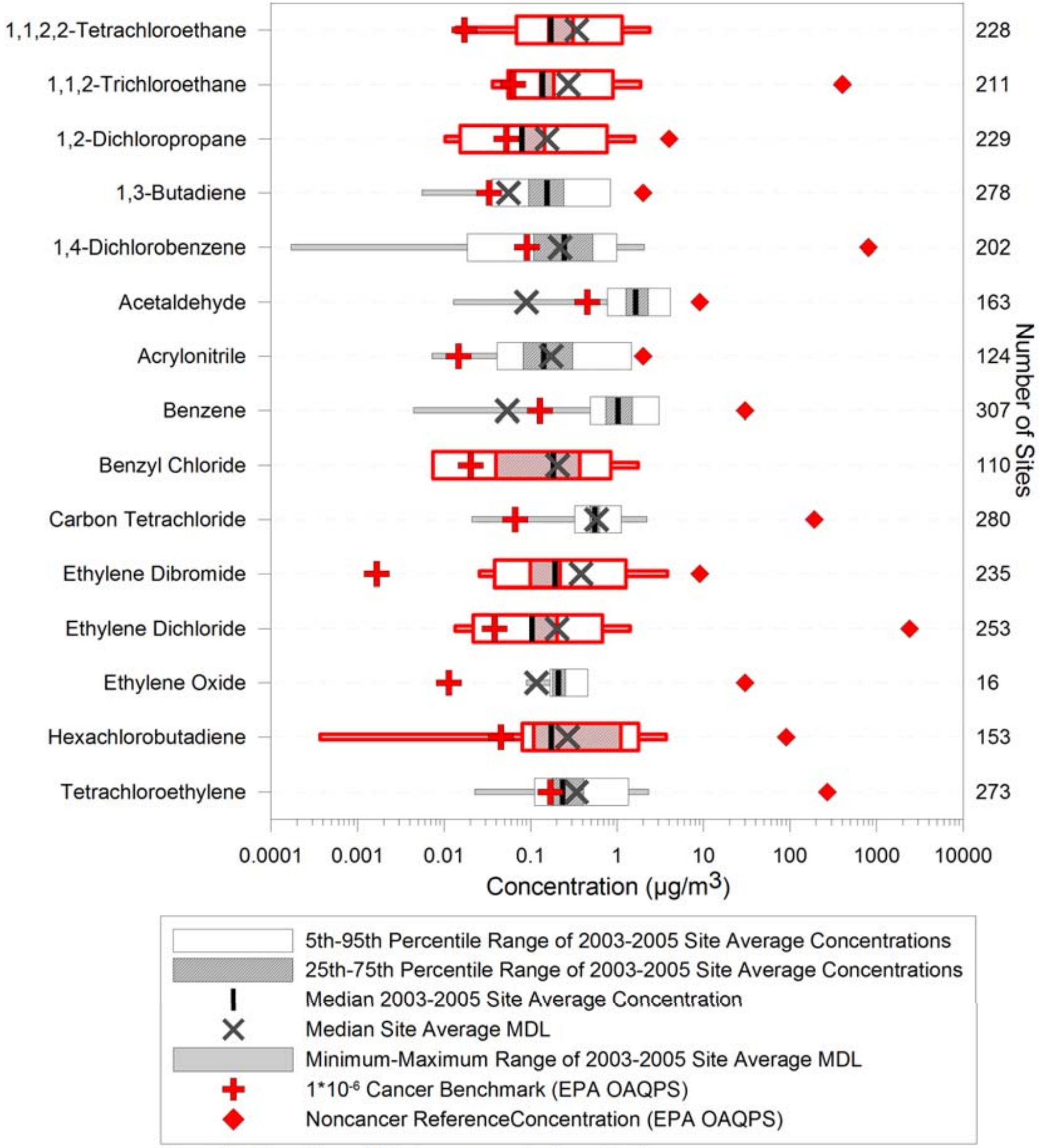
National Concentration Plots



Adding MDLs

- MDL ranges (thin lines) and median MDLs (X's) are added to the plot to illustrate how well pollutants are monitored.
- The minimum-maximum range of MDL concentrations and the median MDL concentration for a 2003-2005 site average are shown.
- The median concentration of the pollutants outlined in red are always below the median MDL. These pollutants are not adequately monitored in the national ambient monitoring networks (i.e., only a few sites have >15% of data above detection).

National Concentration Plots

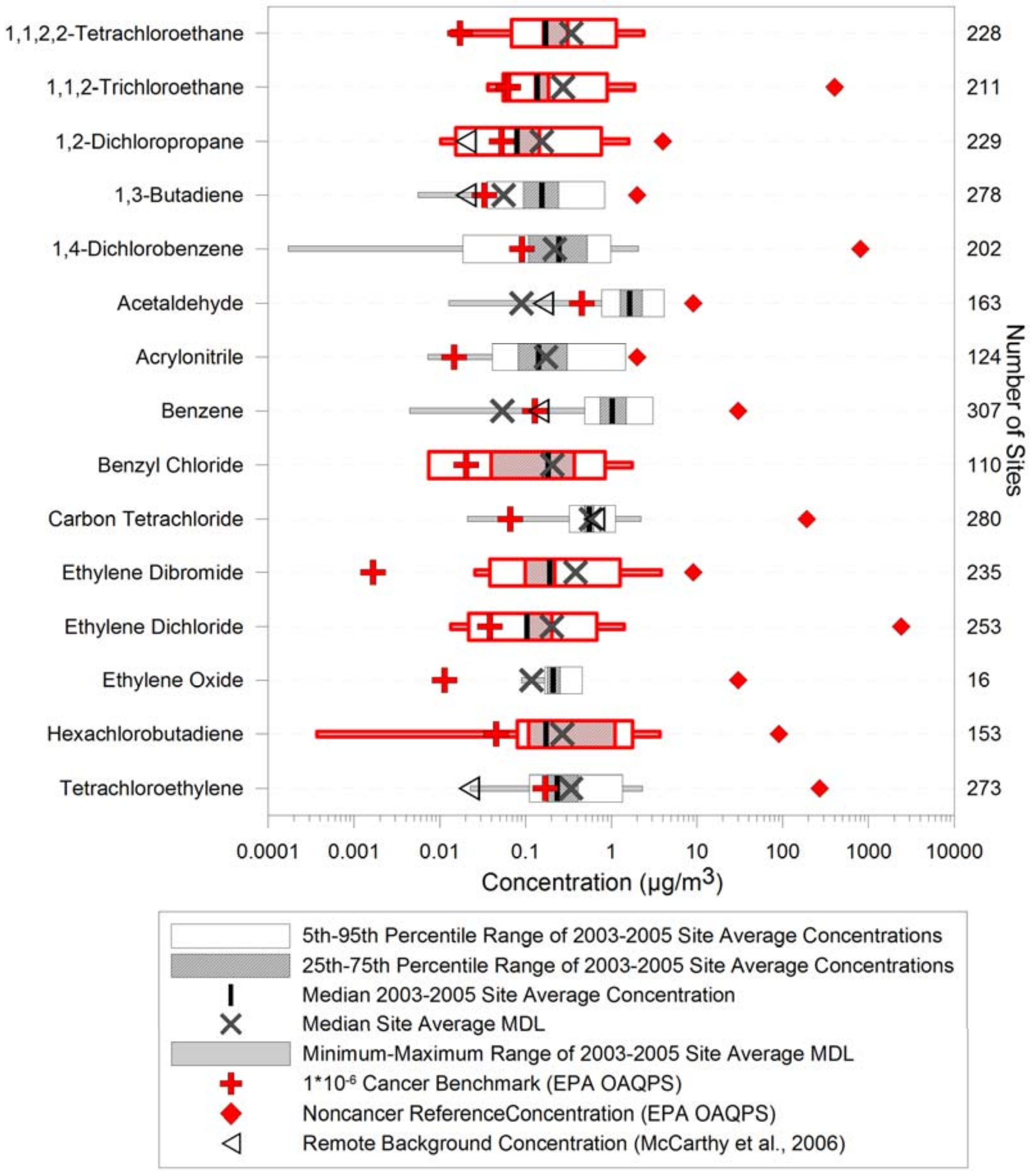


Data outlined in red has < 15% of measurements above detection

Risk Levels

- Chronic exposure concentration associated with a 1-in-a-million cancer risk (red crosses) and noncancer reference concentrations (red diamonds) are added to the plot to show a relationship to human health.
- National measured annual average air toxics concentrations are usually above the chronic exposure concentration associated with a 1-in-a-million cancer risk and below noncancer reference concentrations.
- Note that the pollutant concentration ranges outlined in red may actually be below levels of concern, but the data are not resolved well enough to characterize risk.

National Concentration Plots

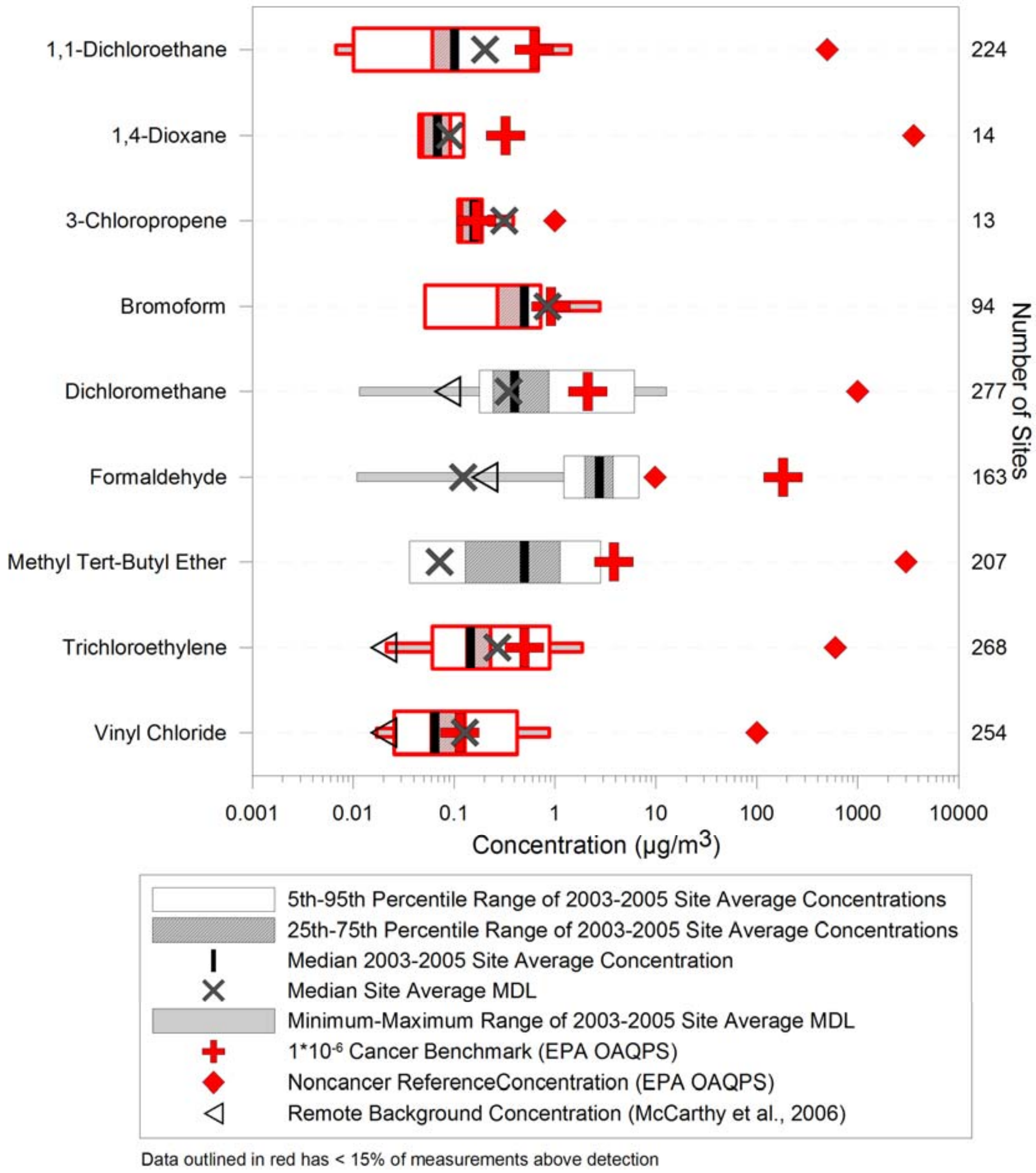


Data outlined in red has < 15% of measurements above detection

Remote Background

- Remote background concentrations (triangles) are added to the plot to show the lowest levels expected to be seen in the remote atmosphere; urban concentrations of most air toxics should not typically fall below this value.
- As expected, most air toxics are a factor of 5-10 above their remote background concentrations, with the exception of carbon tetrachloride – the only air toxic dominated by background concentrations.
- Background estimates are provided for about 40 air toxics (see *Preparing Data for Analysis,* Section 4).

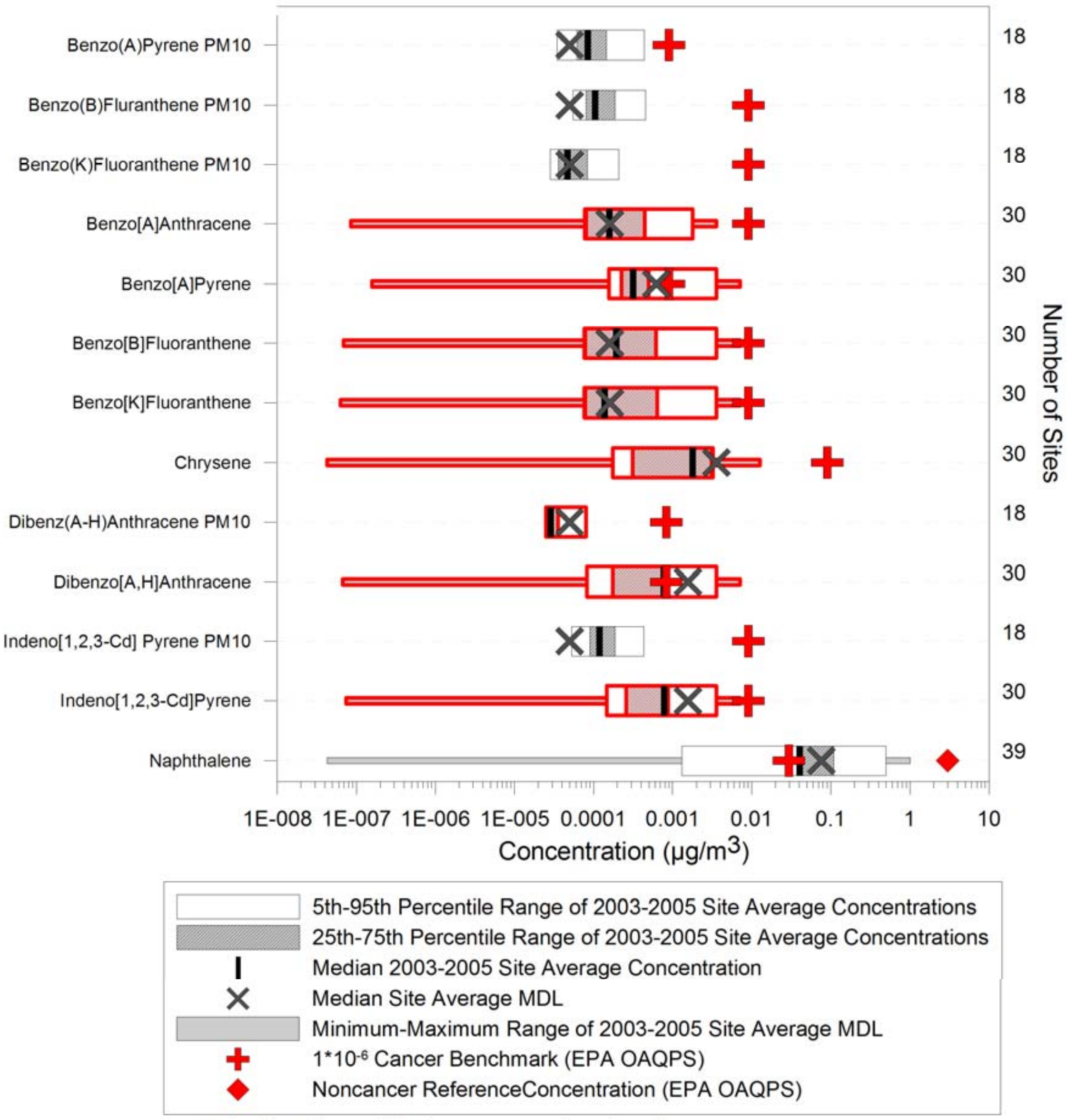
National Concentration Plots



Additional VOCs

- These VOCs are usually below their 1-in-a-million cancer risk level and noncancer reference concentrations.
- Note that the 1-in-a-million cancer risk level for formaldehyde was changed in 2004 from 0.08 to 182 µg/m³. 1-in-a-million cancer risk levels plotted are provided by EPA OAQPS.
- See the NATA website for more information regarding risk characterization, http://www.epa.gov/ttn/atw/nata1999/nsata99.html. For example, analysts can investigate the potential for health effects from air toxics by target organ/system.

National Concentration Plots



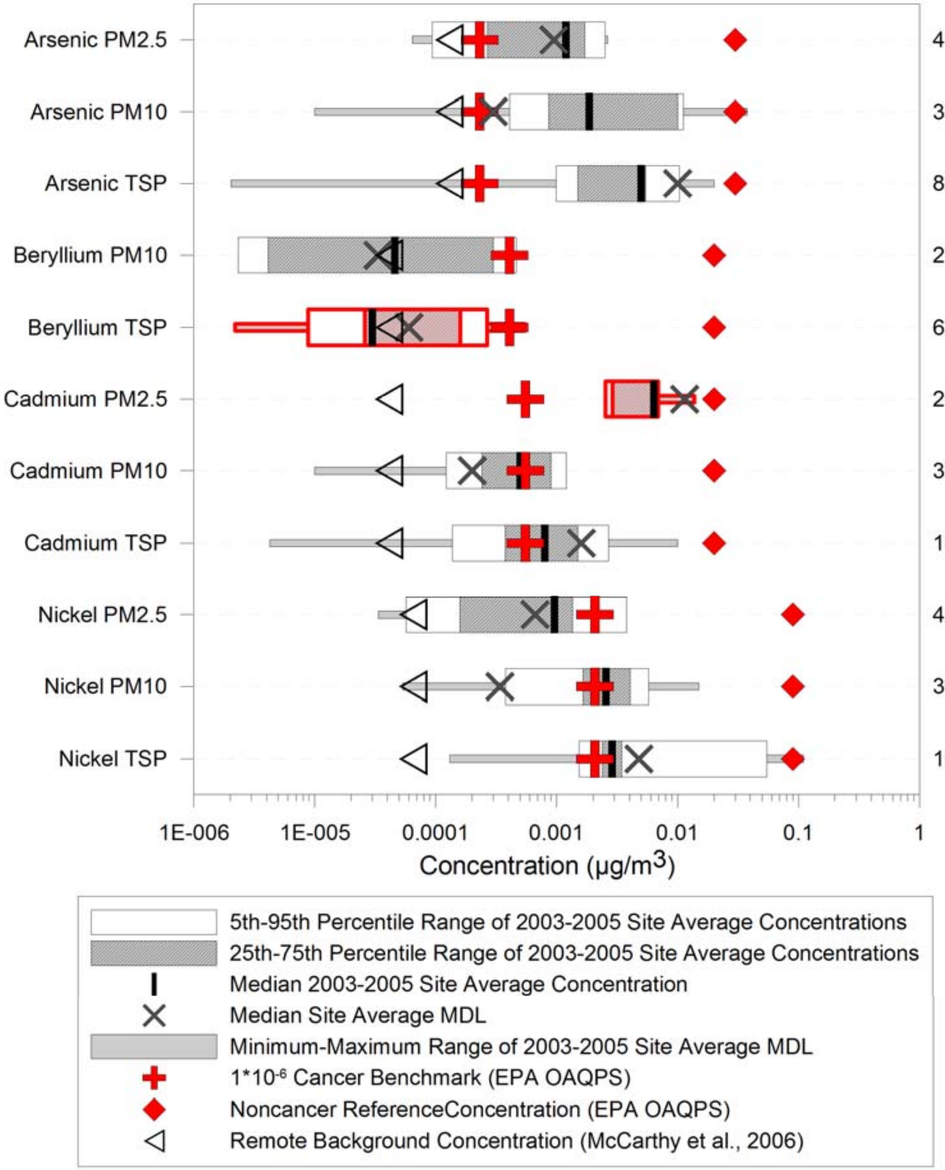
Data outlined in red has < 15% of measurements above detection

*SVOCs**

- The figure indicates that most SVOCs are below their 1-in-a-million cancer risk level. However, the data quality for many SVOCs is poor—less than 15% of measurements are above the detection limit.
- Only naphthalene is above its 1-in-a-million cancer risk level at most sites.
- Routine measurements of SVOCs are relatively rare across the United States.

* semi-volatile organic compounds

National Concentration Plots



Data outlined in red has < 15% of measurements above detection

Metals

- All metals are well below their noncancer reference concentrations.
- With respect to 1-in-a-million cancer risk level, arsenic is the most important of these metals, with more than 75% of sites measuring concentrations above the 1-in-a-million cancer risk level for $PM_{2.5}$.
- $PM_{2.5}$ metals are more commonly measured in rural and remote locations via the IMPROVE network; therefore, the lower range of $PM_{2.5}$ concentrations commonly overlaps remote background concentrations.
- Only four metals could clearly be shown in one figure (monitoring data are available for many more); ranges for other metals can be found in the appendix to *Preparing Data for Analysis,* Section 4.

National Concentration Plots

Summary

- The national concentration plots provide perspective for local, state, regional, and tribal analysts to see how their data compare. A full list of the concentrations shown in the plots is provided in *Preparing Data for Analysis,* Section 4.
- Air toxics concentrations typically vary spatially by a factor of 3 to 10, depending on the pollutant.
- Almost all air toxics are below noncancer reference concentrations (except acrolein, not shown).
- At a national level, some air toxics are above their respective chronic exposure concentration associated with a 1-in-a-million cancer risk (http://www.epa.gov/ttn/atw/toxsource/table1.pdf).
- Most air toxics are well above their remote background concentrations.

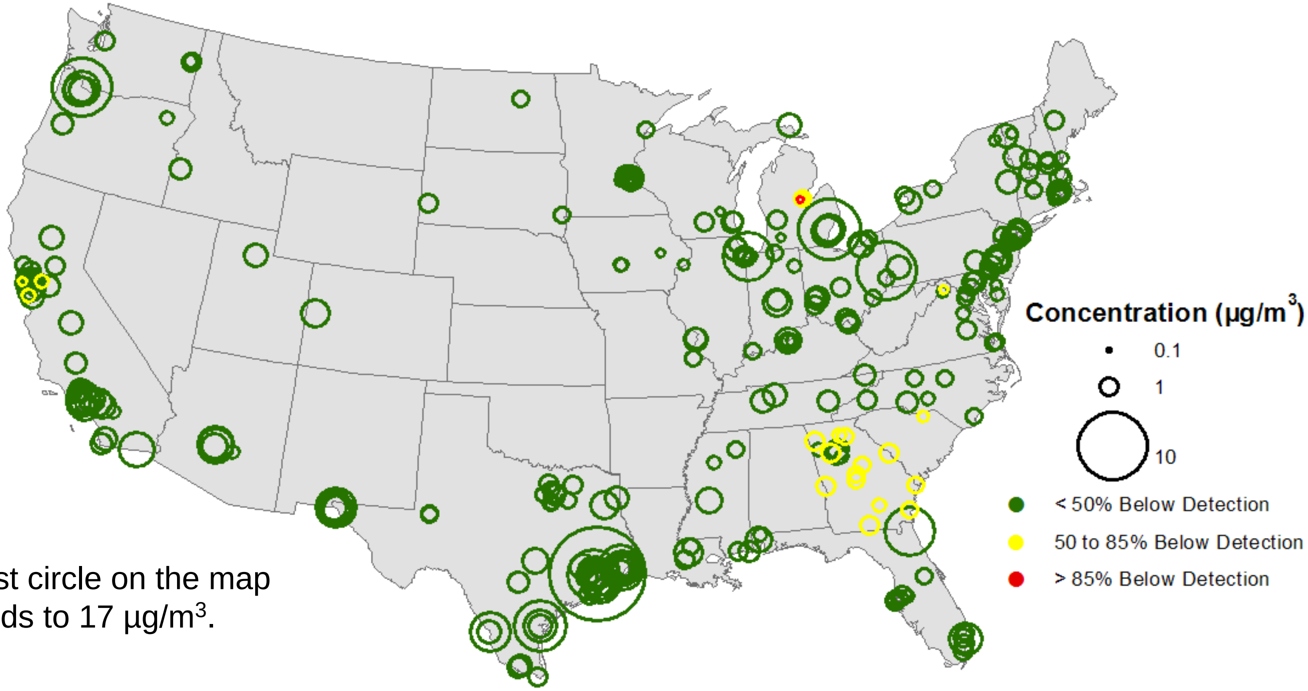
Spatial Patterns – Maps

Overview

- National concentration plots placing air toxics in a national context provide useful information for quantifying air toxics spatial variability. To view spatial patterns, though, it is also useful to plot site-level data on a map.
- Example maps of site average and risk-weighted concentrations (i.e., risk estimates based on ambient measurements) from 2003 through 2005 are shown in the following slides. These maps help analysts characterize the national picture of air toxics and are most useful in a qualitative sense to compare among sites, look for spatial patterns, and note data anomalies. The maps also illustrate a method of displaying data that can be applied to sites within a city, state, or region.
- In the examples, concentrations are displayed as proportional symbols which are color-coded to impart additional information.
- Maps are useful for communicating a range of information—similar depictions can be made using risk-weighted concentrations, percent change per year, or ratios—over a range of spatial dimensions (e.g., city, state, or region).
- The volume of concentrations is indicated on the maps by the diameter of the circle (the three sizes in the map legends) while the underlying percent of data below detection is signified by color. All maps were created with ESRI's ArcMap software.

Spatial Patterns – Maps

Benzene Concentrations 2003-2005

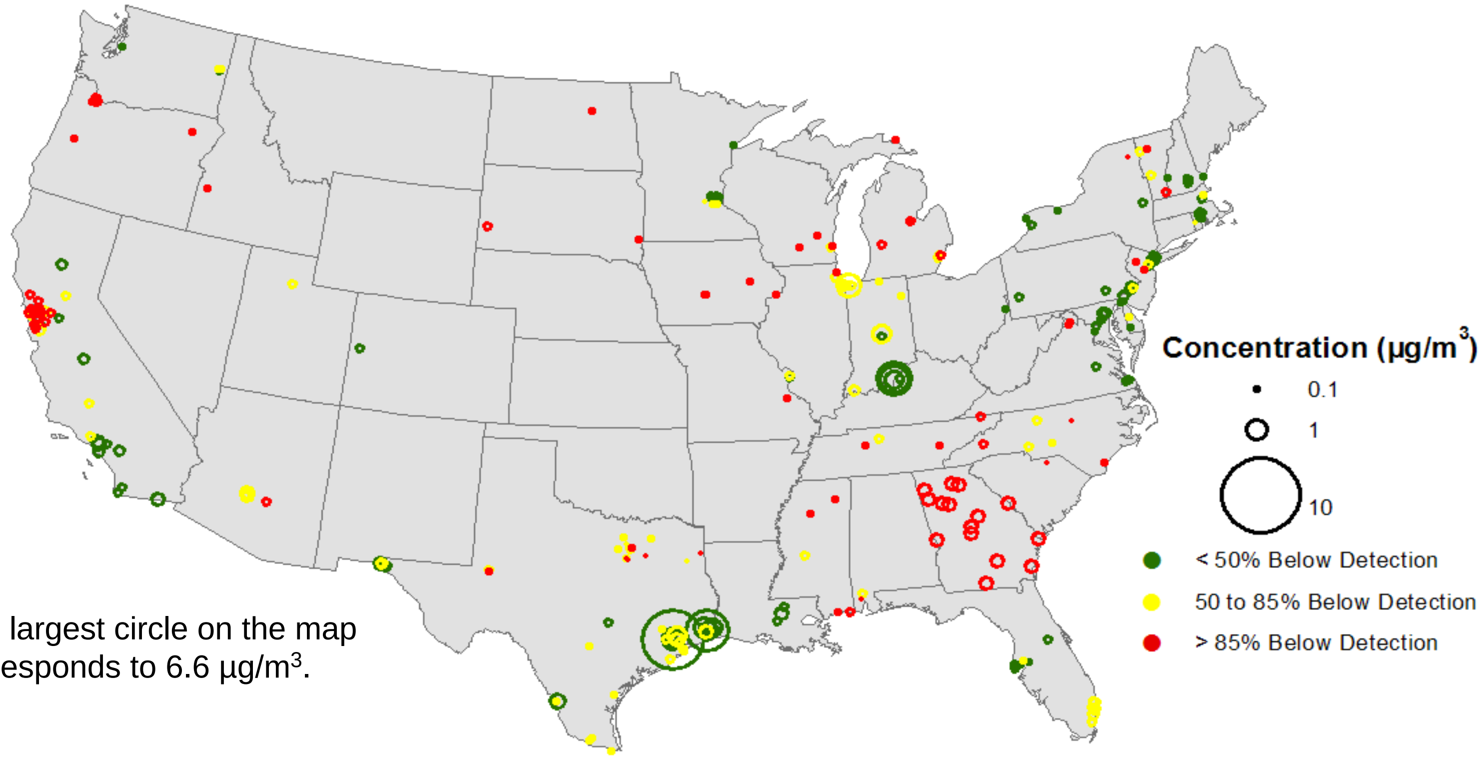


The largest circle on the map corresponds to 17 µg/m^3.

- The map shows that benzene concentrations have ambient measurements above detection across the country with only a few exceptions (i.e., 0-50% of the measurements at most sites are below detection).
- Concentrations are consistent for areas dominated by mobile sources (e.g., the Northeast and California) while isolated high concentrations generally coincide with significant point source emissions of benzene such as refineries and coking operations.
- Sites that show unusually high concentrations with no clear emissions sources, or sites with concentrations that are very different from other sites (e.g., the yellow circles in the map above), might be further investigated to determine the cause.

Spatial Patterns – Maps

1,3-Butadiene Concentrations 2003-2005

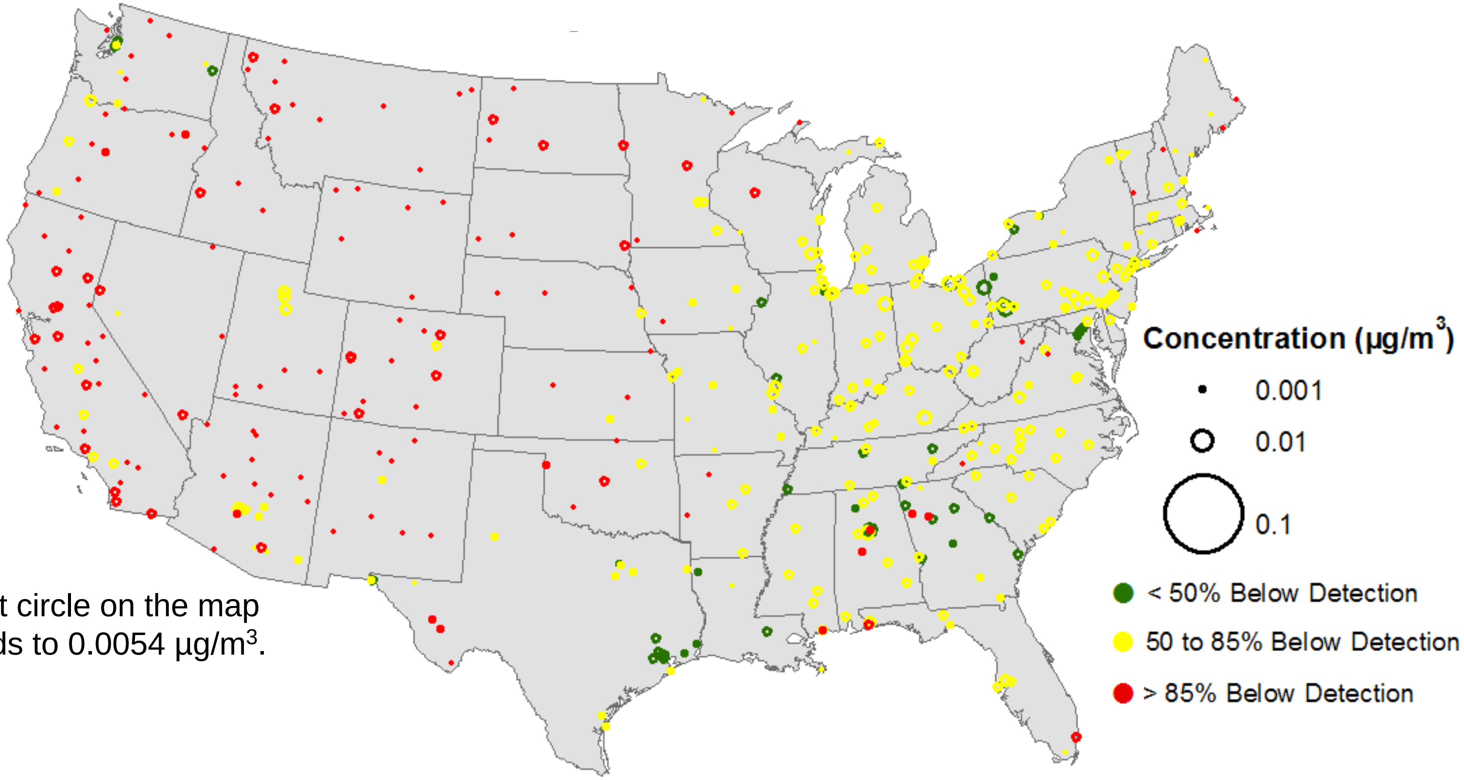


The largest circle on the map corresponds to 6.6 µg/m^3.

- The ability to obtain 1,3-butadiene concentration measurements above the MDL across the United States varies (note all the red circles and their varying sizes).
- Higher concentrations generally coincide with locations of known point source emissions.
- Differences in monitoring methods and methods application have resulted in large differences in reported MDLs across the United States.

Spatial Patterns – Maps

Arsenic $PM_{2.5}$ Concentrations 2003-2005

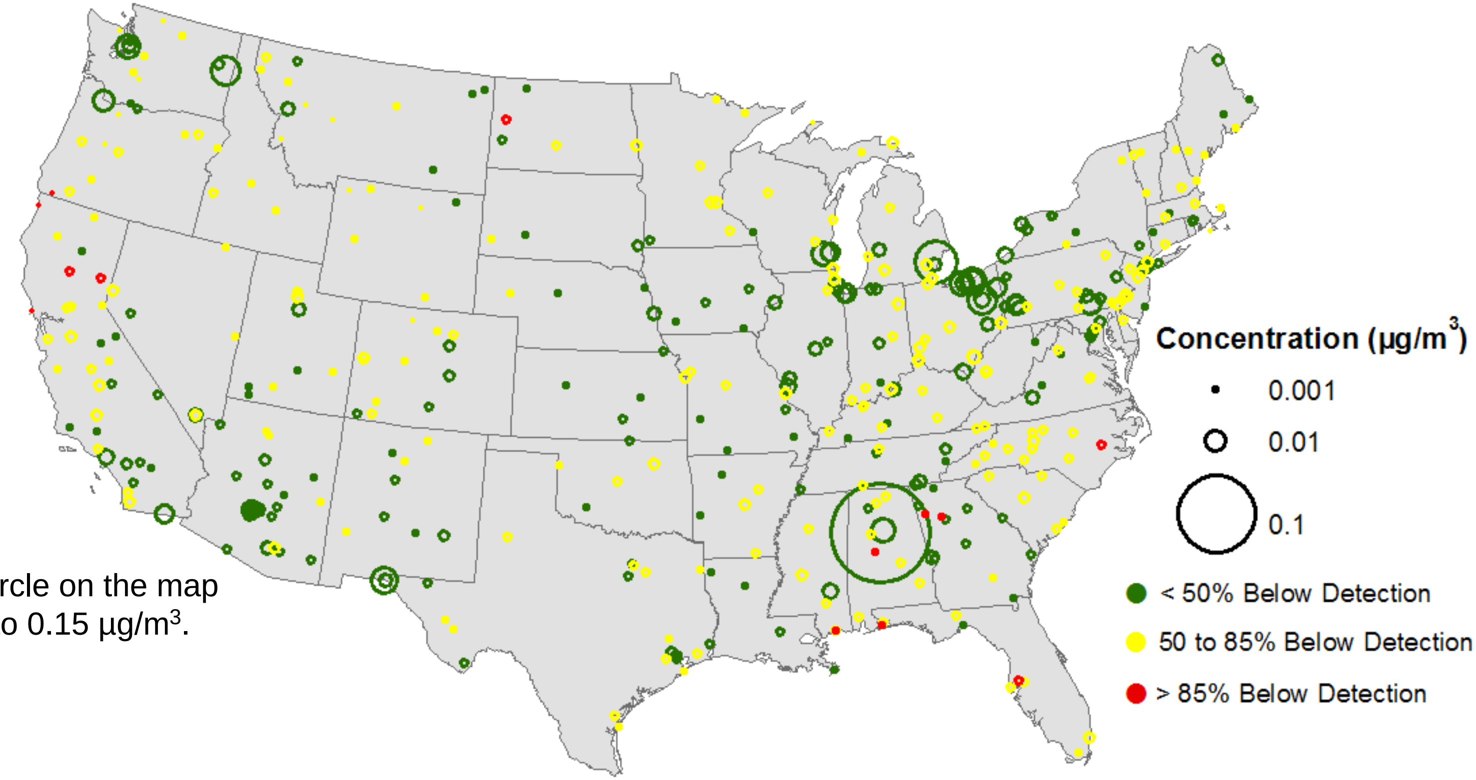


The largest circle on the map corresponds to 0.0054 µg/m^3.

- Arsenic concentrations are widely measured across the United States, and the entire range of data availability is observed from more than 50% of data above detection to less than 15% above detection.
- Significant MDL differences between networks make determining spatial patterns difficult.
- In general, concentrations are higher and more often above detection in the eastern half of the country.

Spatial Patterns – Maps

Manganese $PM_{2.5}$ Concentrations 2003-2005



The largest circle on the map corresponds to 0.15 µg/m^3.

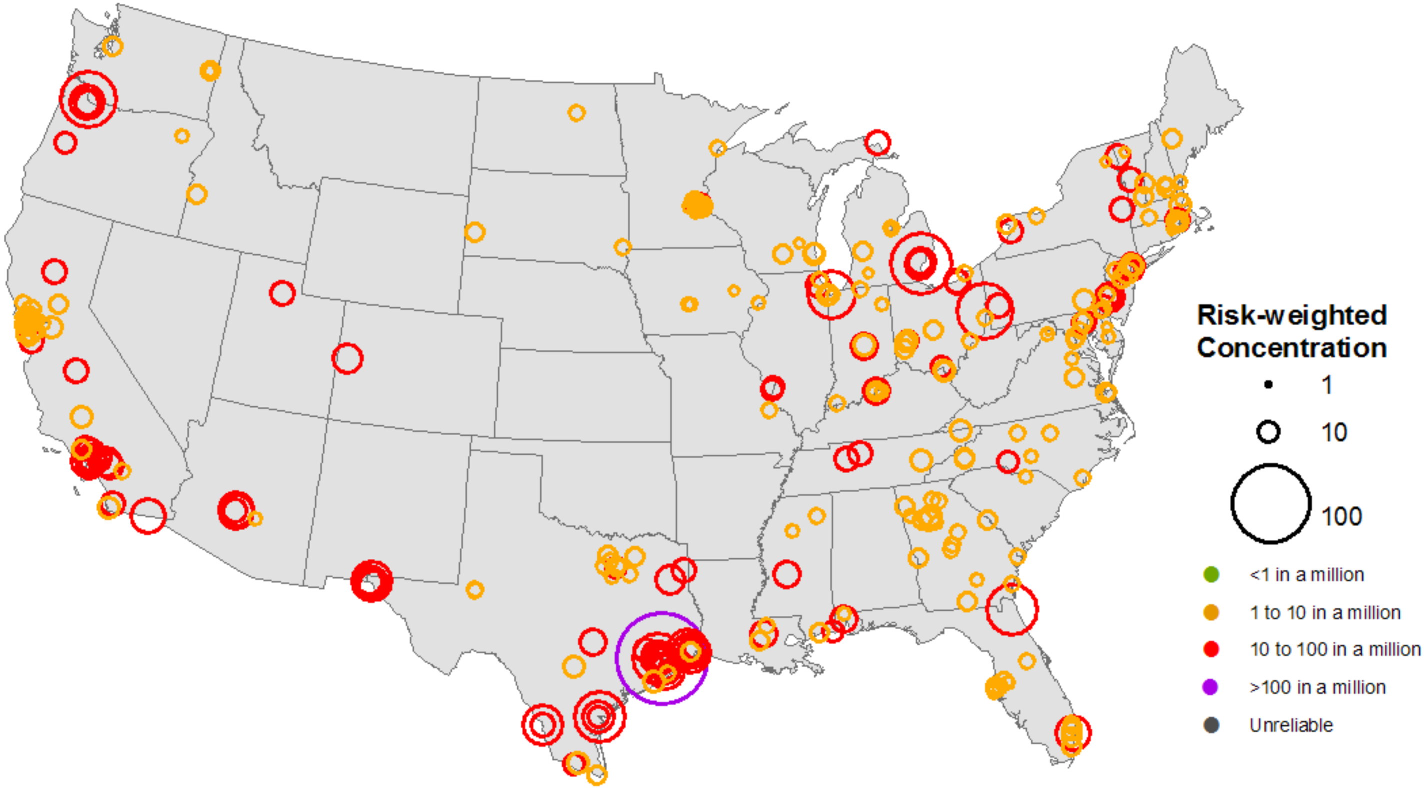
- In contrast to arsenic, manganese concentrations are widely measured across the country with most data recorded above the detection limit.
- Concentrations vary spatially and several “hot spots” can be identified that may lend themselves to additional investigation at a site level.

Spatial Patterns – Maps

Benzene Risk-Weighted Concentrations 2003-2005

Note:

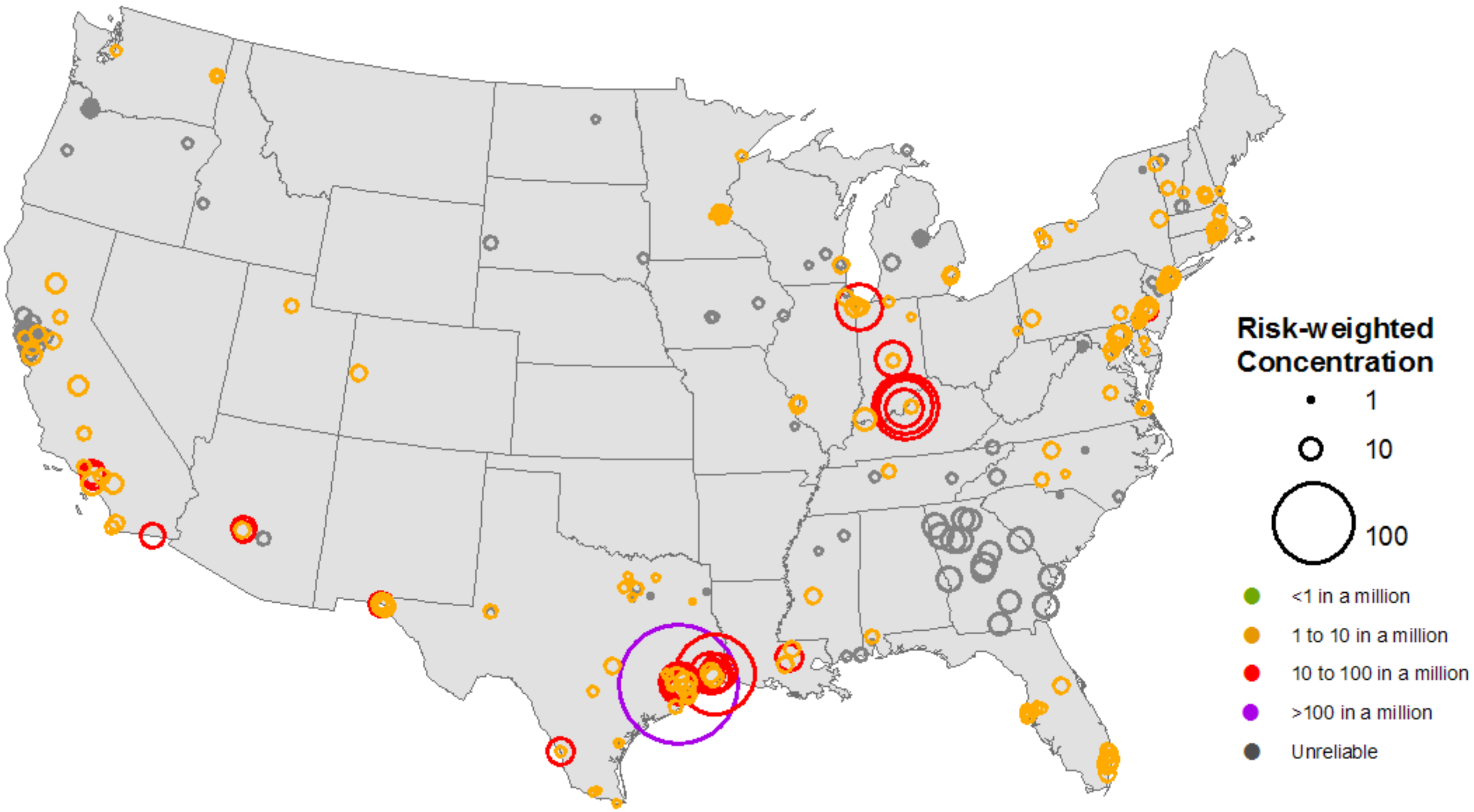
2003-2005 average concentrations are divided by the 1-in-a-million cancer risk concentration. Circle diameter represents this ratio while the chronic risk assessment is indicated by color. Sites at which >85% of data are below detection are considered unreliable (grey).



Benzene risk associated with measured ambient concentrations is almost always above the 1-in-a-million cancer risk level across the United States. Many areas are also above the 10-in-a-million cancer risk. These results are in good agreement with NATA 1999 results. The highest risk estimates are located in areas with significant point source benzene emissions.

Spatial Patterns – Maps

1,3-Butadiene Risk-Weighted Concentrations 2003-2005



Where measured reliably, 1,3-butadiene concentrations are almost always above the 1-in-a-million cancer risk level. Some areas do not measure concentrations well enough to evaluate risk (grey symbols). Highest concentrations are located in areas with known point source emissions (e.g., Houston and Louisville).

Variability Within and Between Cities
Overview

- A topic of interest for air toxics data analysis is assessing variability in concentration from site to site within a city. The aim of such analysis is to understand how representative a given site is with respect to air toxics concentrations in a city.
  - What is the variability of air toxics concentrations within cities and what are the implications for aggregating data at the city level?
  - Where do sites need to be located to accurately characterize variability within a city?
  - How many sites are needed to characterize spatial variability within a city?
  - How does within-city variability differ across cities?
- There may also be interest in assessing variability in air toxics from city to city.
  - What are the concentration distributions across all monitoring sites?
  - Do specific cities, states, or regions have demonstrably higher or lower concentrations?
  - Do demonstrably lower concentrations occur at rural and remote sites?
  - Are concentration differences associated with monitoring agency differences?

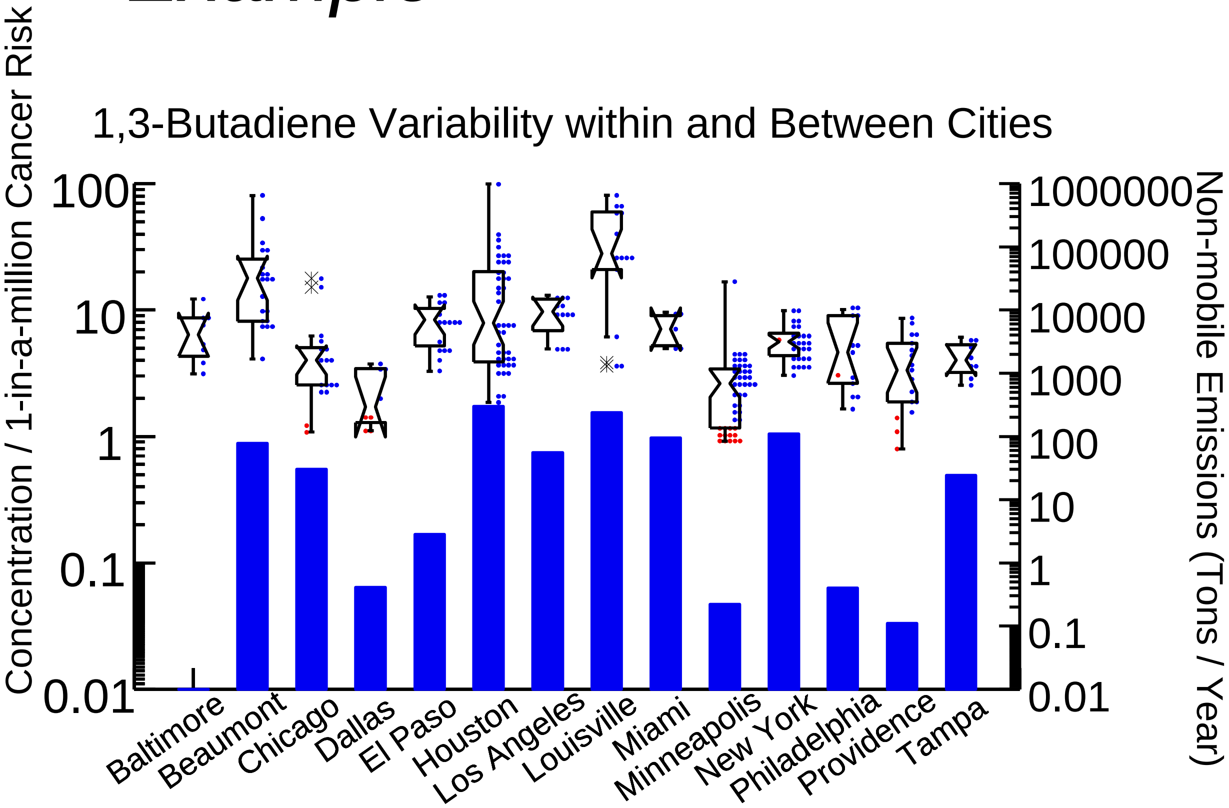
Variability Within and Between Cities
Approach

- To investigate within-city variation, a city of interest should have multiple monitors. For example, for a national trend analysis, EPA required a city to have at least four monitors to be included in analysis.
- Valid annual averages are calculated for each monitor in a city. To reduce noise from year-to-year changes (e.g., the effect of meteorology), it is best to use multiple years of data when available. The national study used 2003-2005 data.
- Data can be visualized using notched box plots by air toxic, city, and year. If variation between years at a given city is minor, notched box plots by air toxic and city only can be constructed to increase the amount of data.
- Advanced Plotting Techniques
  - Include a color-coded measure of the percent of data below detection to understand the reliability of the data.
  - Divide annual averages by the chronic exposure concentration associated with a 1-in-a-million cancer risk (or other risk level) to show variation in risk estimates within and between cities.
  - Include a measure of relevant emissions by city to explain possible reasons for high or low concentrations.

Variability Within and Between Cities

Example

- In the example, risk estimates have been used to provide a secondary layer of information.
- A single box in the figure contains one annual average for each monitor within the city; thus, each box represents intra-city concentration variation.
- The variability between cities is also represented by including multiple cities on the same plot.
- The within-city spatial variability of 1,3-butadiene is usually less than a factor of 8 for the cities in the figure.
- 1,3-butadiene variability between cities, however, can be greater than an order of magnitude.
- Emissions from major sources at a county level are generally higher for the cities with greater within-city variability and higher concentrations, but there are exceptions that could be explored.



1,3-Butadiene Variability within and Between Cities

The figure shows benzene risk-weighted (1-in-a-million) annual average variation for 2003-2005 for selected U.S. cities along with non-mobile emissions. Notched boxes include annual averages for each monitor within a city, providing within-city variation. Dots over the notched boxes show the individual data points and whether they are above (blue) or below (red) the average MDL. Bars show county-level non-mobile emissions of 1,3-butadiene from EPA's AirData. The figure was created with SYSTAT11.

Variability Within and Between Cities

National Perspective

- At a national level, spatial variability within cities was found to be pollutant- (or pollutant group-) specific.
- Most toxic measurements are highly variable within cities; risk values span an order of magnitude within some cities.
- The spatial variability between cities is a good metric to estimate the variability within cities *a priori*. Spatial variability analysis helps set expectations for sampling in a new city.
- Cities with point source emissions (e.g., Houston) showed higher within-city variability than those dominated by area/mobile sources (e.g., Los Angeles).
- Some of the observed variability is due to differences in sampling/analysis method and method detection limit.

Hot and Cold Spot Analysis

Overview

- Hot and cold spot analysis is an investigation of sites with the highest and lowest concentrations.
- The objective of this analysis includes:
  - Data validation. The highest and lowest values may be due to some type of error, possibly reporting.
  - Comparison to the spatial conceptual model. Are the highest concentrations consistent with known sources, transport, and dispersion?
  - Risk screening. Where are the toxic concentrations highest?

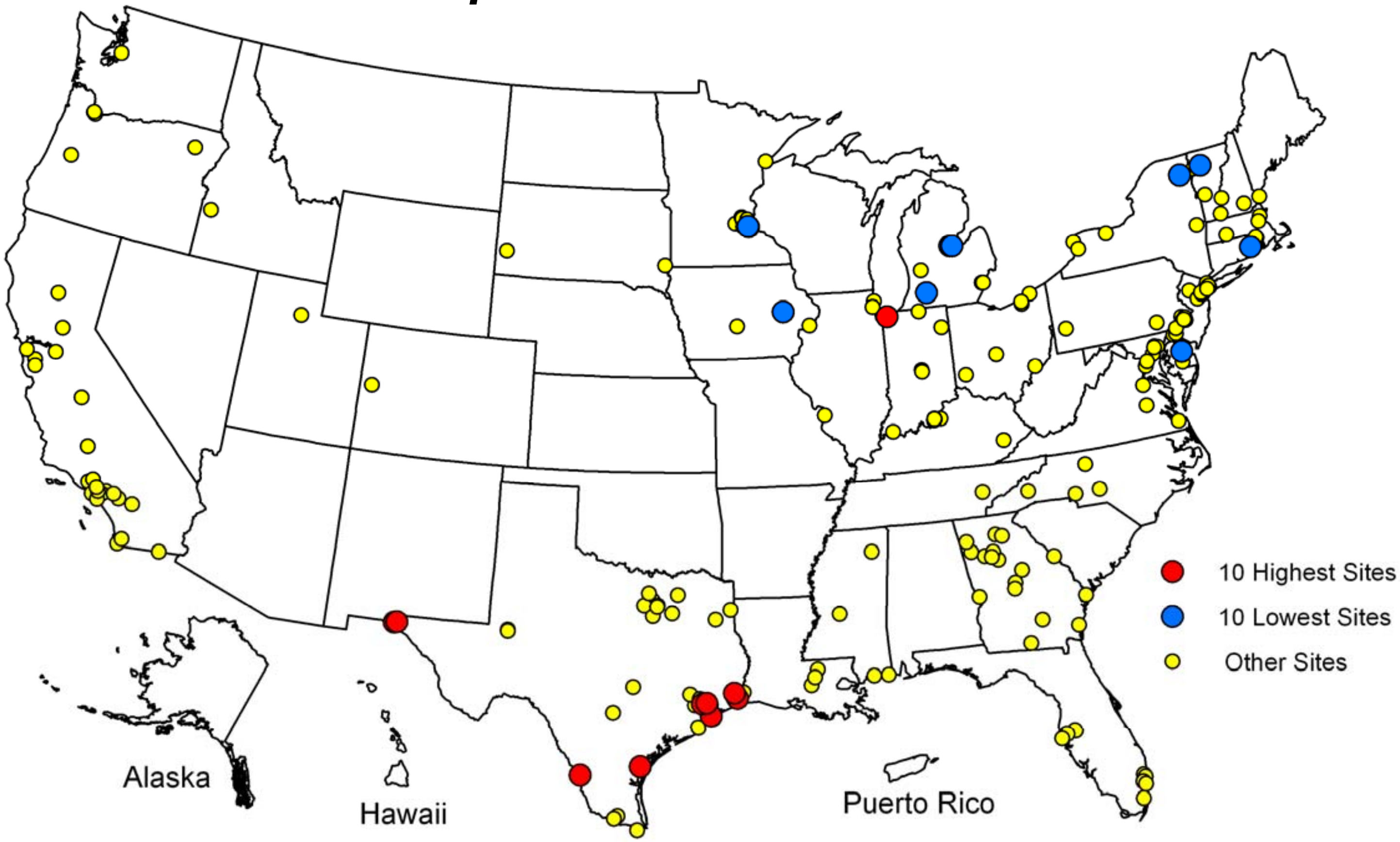
Hot and Cold Spot Analysis

Approach

- Create valid annual averages (see *Preparing Data*, Section 4) for each site and pollutant and rank each site by its concentration (highest to lowest). The number of high- and low-ranked concentration sites investigated depends on the number of available sites. At a national level, the 10 highest and 10 lowest ranking sites were investigated to illustrate the approach.
- Map all sites, marking the highest and lowest ranked sites to investigate spatial variation.
- Identify why high or low concentrations occur at those sites and whether the occurrence of those concentrations meets expectations.
  - Review metadata about the sites (e.g., Google Earth images, local emissions, and meteorology). Do concentrations meet spatial conceptual models with respect to scale, sources, transport, and dispersion?
  - Inspect time series of concentration and MDL (e.g., is the value stuck, are data outliers driving the average, is the MDL higher than the concentrations at an average site?).

Hot and Cold Spot Analysis

Example – Benzene *(1 of 2)*



The figure shows sites with the 10 highest and 10 lowest benzene concentrations based on 2003-2005 annual averages. Other monitoring sites are shown in yellow. The sites ranked lowest were either a result of data reporting or siting issues or were located in rural areas, consistent with our conceptual model of low concentrations.

Hot and Cold Spot Analysis

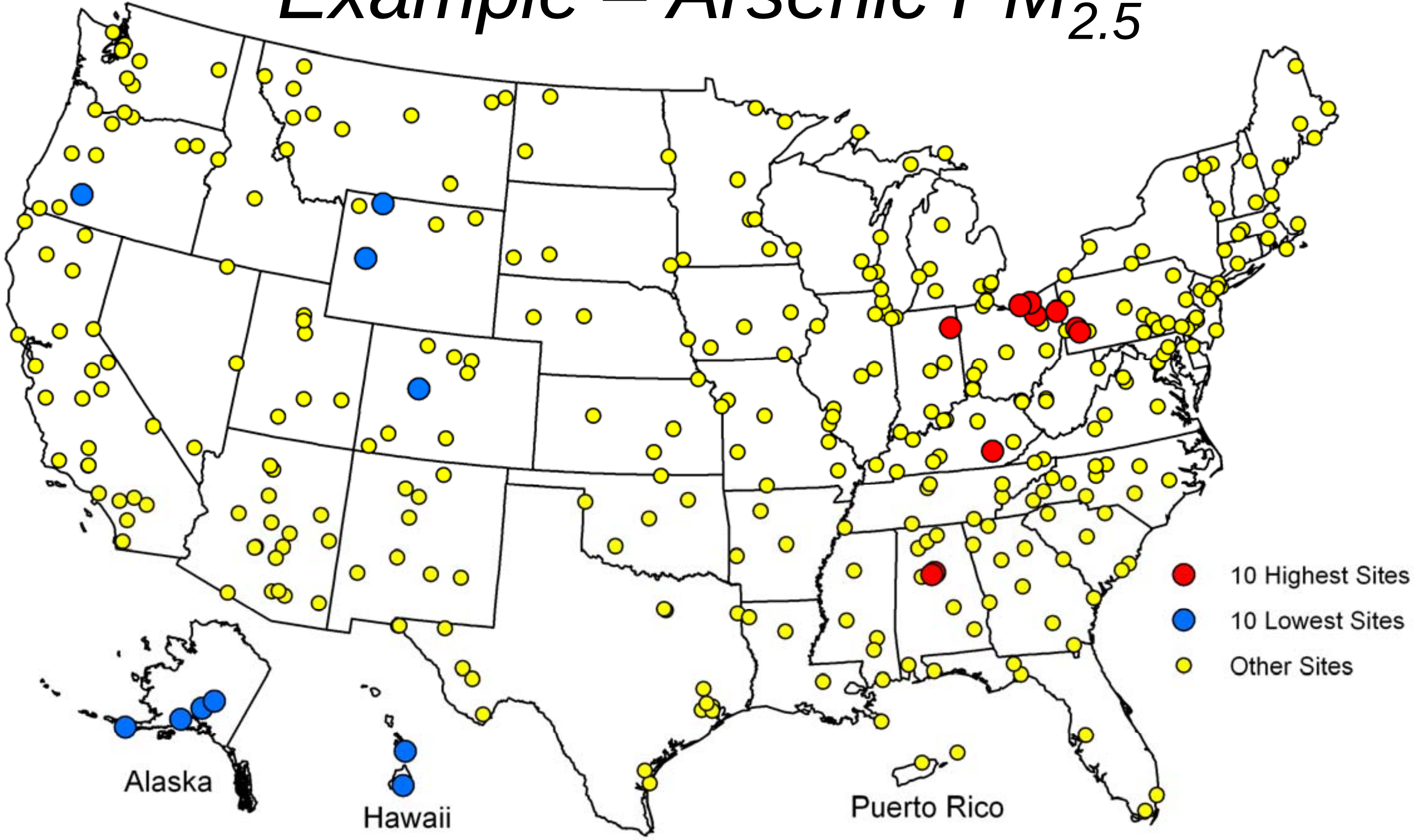
Example – Benzene (2 of 2)

- The sites measuring the highest concentrations in the nation were dominated by nearby point source emissions; the site identified in the figure measured the second highest benzene concentration in the nation.
- This site is very close to two refineries that emit a significant amount of benzene each year according to the NEI.

Google Earth image of the site with the second highest benzene concentrations in the United States. Refineries to the right and left emitted 84,000 and 44,000 lbs of benzene in 2004 (NEI).

Hot and Cold Spot Analysis

Example – Arsenic $PM_{2.5}$



The figure shows sites with the 10 highest and 10 lowest arsenic $PM_{2.5}$ concentrations based on 2003-2005 annual averages. Other monitoring sites are shown in yellow. Conceptually, we would expect Arsenic $PM_{2.5}$ concentrations to be highest in locations dominated by point source emissions, especially smelting and coal combustion. The highest sites are consistent with this conceptual model. The lowest sites are located in extremely remote locations such as Alaska and US national parks which is reasonable for the lowest arsenic $PM_{2.5}$ concentrations.

Urban vs. Rural Analysis

Overview

- Measured concentrations can be highly dependent on individual monitor locations, geography, emissions sources, and meteorological conditions (e.g., prevailing winds).
- **Urban areas – conceptual model**
  - Urban areas contain sources of air toxics that result in increased concentrations and, in some cases, "hot spots" (areas with disproportionately higher concentrations) in the spatial pattern.
  - Urban concentrations vary greatly from day to day due to the mix of local sources and meteorology.
- **Rural areas – conceptual model**
  - Rural areas typically have fewer sources of air toxics. Air toxics concentrations that are transported from urban locations are typically near background levels when they reach rural areas (a function of source strength, distance, and the lifetime of the pollutant).
  - Concentrations do not vary consistently day to day. Daily and seasonal patterns that are dependent on meteorological conditions may still be observed.
- Urban and rural sites that do not meet the expectations of conceptual models may indicate monitoring location effects or data errors or problems with the conceptual model.

Urban vs. Rural Analysis

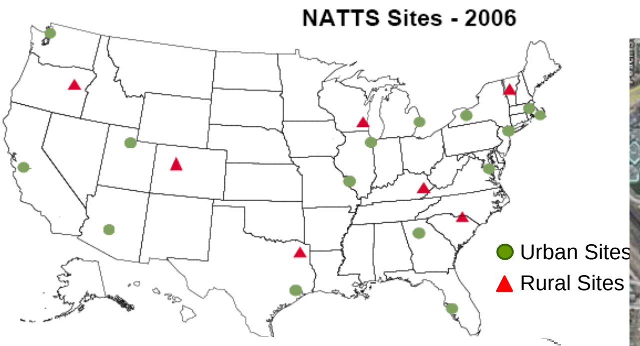
Approach

- Characterize each site as urban or rural.
  - If available, start with EPA urban/rural designations as listed in AQS (note that these designations are not always up to date)
  - Verify the designations using Google Earth—they may be outdated or incorrect
  - Be wary of defining a site using population density, total county population, or other metrics—local knowledge of the site appears to be the best way to identify site characteristics.
- Identify pollutant availability and time period for each site.
  - The goal is to have a spatially representative mix of urban and rural sites measuring a pollutant over the same time period. This mix can be a challenge since toxics are more commonly measured in urban locations.
- Choose pollutant/site combinations that are spatially and temporally representative.
  - Pollutant-specific monitoring time periods need to be the same for site comparison; otherwise differences in observed concentrations could be biased by seasonal or inter-annual patterns.
- Estimate valid 24-hr averages for the sites, pollutants, and time periods of interest.
  - Characterize all concentration averages that are below the associated average MDL
- Visualize the data by site by preparing plots of data distributions, including some measure of the data below detection. Look for differences in concentrations.
- Identify statistically significant differences in urban vs. rural site concentrations.
- Summarize the results with a focus on neighboring urban vs. rural sites.
  - Which urban and rural sites measured significantly higher or significantly lower concentrations, if either? Which showed no difference?
- Investigate data that do not meet expectations (e.g., concentrations as a rural may be significantly higher than those at a nearby urban site).
  - Are the sites representative of the area (i.e., compare to other urban or rural sites)?
  - Are there monitor location abnormalities (e.g. local terrain, prevailing winds)?
  - Are there measurement methods or MDL differences between the sites?
  - Is there a significant rural emissions source?
  - Are possible data errors or outliers driving the trend?

Urban vs. Rural Analysis

Example – Investigating Urban vs. Rural Sites (1 of 2)

- When beginning an urban vs. rural analysis, it is important to verify that sites are properly designated “urban” or “rural”. This example is qualitative.
- The pictures below show a map of urban and rural NATTS sites across the United States along with Google Earth pictures of two of the rural sites—Grand Junction, Colorado, and La Grande Oregon.
- Both sites are designated as rural in AQS, but the Colorado site appears quite urban in character, and it is likely that air toxics concentrations will not conform to the model for a rural site.
- The Oregon site, on the other hand, is rural-based on the observation that the surrounding area is mainly farmland.



•Urban Sites		•Rural
•E. Providence, RI •Boston (Roxbury), MA •New York, NY •Rochester, NY •Washington, DC •Decatur, GA •Tampa, FL •Detroit, MI	•Chicago, IL •Houston (Deer Park), TX •St. Louis, MO •Bountiful, UT •San Jose. CA •Phoenix, AZ •Seattle WA	•Underhill, VT •Hazard, KY •Chesterfield, SC •Mayville, WI •Grand Junction, CO •La Grande, OR •Harrison County, TX


Grand Junction, CO

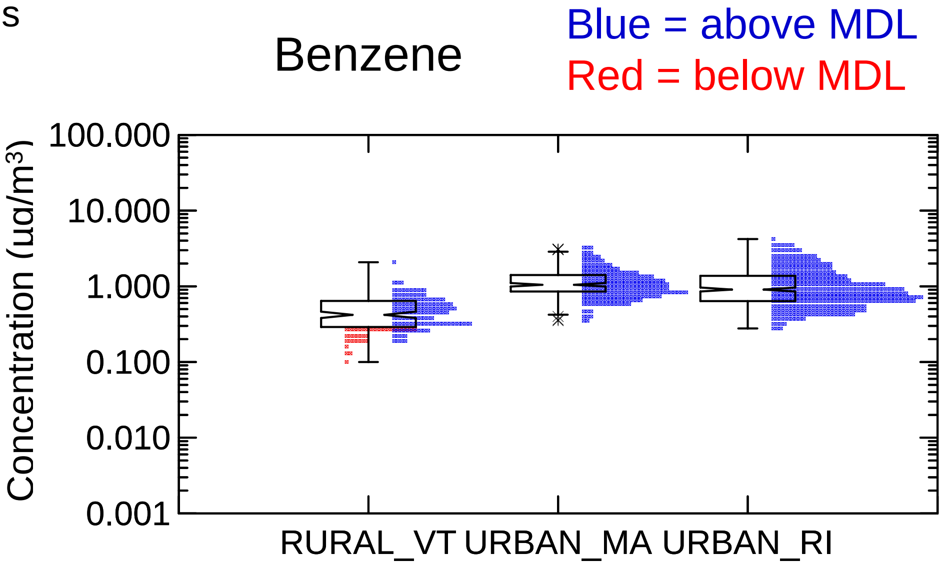
La Grande, OR

Two rural sites in the NATTS network. Images obtained from Google Earth.

Urban vs. Rural Analysis

Example – Investigating Urban vs. Rural Sites (2 of 2)

- The figure shows benzene concentrations at a rural Vermont site compared to concentrations at two urban northeastern sites.
- The rural site shows statistically significantly lower concentrations.
- If a site does not fit an urban or rural definition as expected, check for
  - Measurement method or MDL differences
  - Local emissions sources
  - Time series comparing the two sites with color-coded data below detection.
  - Evaluate data subsets when both sites have measurements above detection. Does this tell a different story?



The example figure is from an analysis of NATTS sites using 2003-2005, 24-hr average, benzene data. The box plots encompass all data while the overlaid dot density shows each data point and whether it is above or below detection (blue vs. red). It was produced in SYSTAT11.

Spatial Patterns

Summary

- Analyses described in this section provide information about a variety of aspects of air toxics spatial variability and help analysts evaluate multiple conceptual models.
- Spatial patterns can provide information about sources, sinks, transport, and dispersion which are of interest for air toxics analyses.
- At a national level, the following spatial patterns were observed for air toxics.
  - Benzene, 1,3-butadiene
    - Concentrations vary around the United States and are high in urban areas. The highest concentrations of these two air toxics, however, are found in areas influenced by point source emissions in addition to mobile sources.
    - Within- and between-city variability is generally near a factor of 5.
  - Carbonyl compounds
    - Carbonyl compounds are measured widely and show very consistent concentrations across the nation. This is due to the dominant secondary formation mechanism.
    - Within and between-city variability is relatively low with few exceptions.
  - $PM_{2.5}$ metals
    - The spatial character of $PM_{2.5}$ metals is difficult to determine due to differences in measurement methods and MDLs among monitoring networks.
    - Overall it seems that concentrations are slightly higher in the eastern half of the United States.

Risk Screening

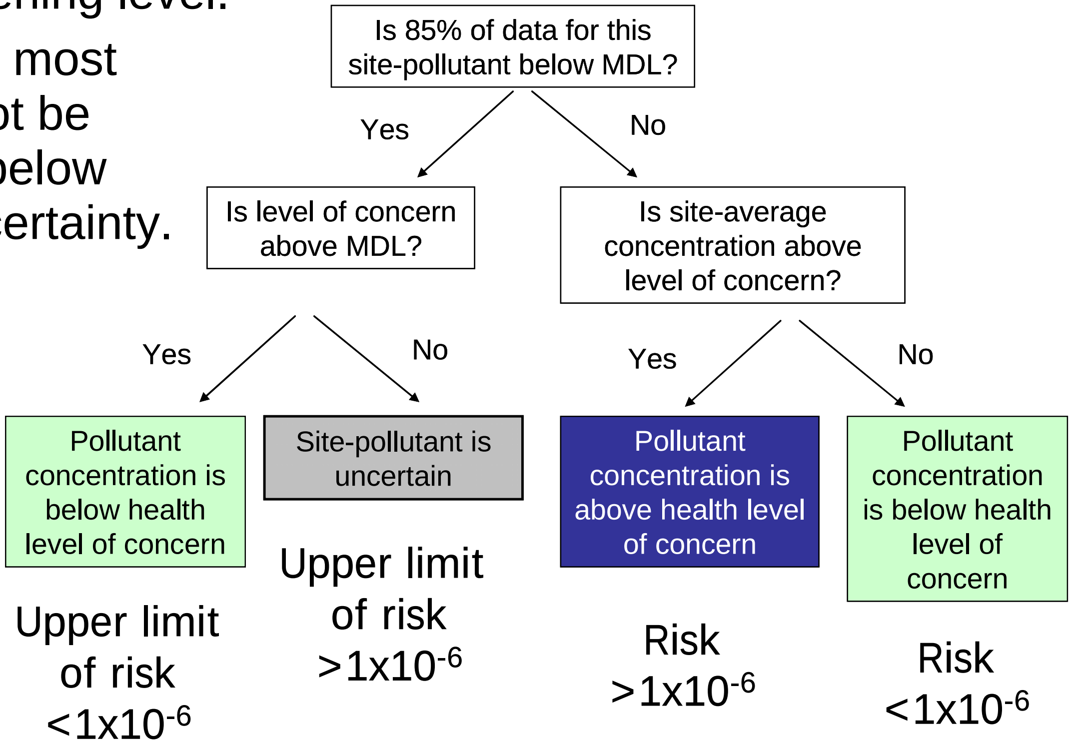
Overview

- A key use of air toxics data is to compare annual average concentrations to health thresholds to put ambient levels into context.
- Risk screening can help identify air toxics of concern.
- Information to consider in conducting a risk screening is available, for example, in "A Preliminary Risk-Based Screening Approach for Air Toxics Monitoring Data Sets", http://www.epa.gov/region4/air/airtoxic/Screening-041106-KM.pdf.
- For information on a more thorough air toxics risk assessment, see the Air Toxics Risk Assessment Library: http://www.epa.gov/ttn/fera/risk_atra_main.html.

Risk Screening

Approach

- For this first level of screening, site average concentration data from the most recent year (s) (e.g., 2003-2005) were used to identify the number of sites at which a pollutant was definitively above or below the relevant EPA OAQPS chronic exposure concentration associated with a 1-in-a-million cancer risk as found at: http://www.epa.gov/ttn/atw/toxsource/summary.html. Results are ranked by screening level.
- Air toxics were also noted if most site concentrations could not be characterized as above or below the relevant risk level with certainty.
- The figure shows steps through a decision tree for performing risk screening.



Is 85% of data for this site-pollutant below MDL?

Yes

No

Is level of concern above MDL?

Is site-average concentration above level of concern?

Yes

No

Yes

No

Pollutant concentration is below health level of concern

Site-pollutant is uncertain

Pollutant concentration is above health level of concern

Pollutant concentration is below health level of concern

Upper limit of risk $<1x10^{-6}$

Upper limit of risk $>1x10^{-6}$

Risk $>1x10^{-6}$

Risk $<1x10^{-6}$

The % of data below MDL listed in the first box may need to be stricter or less strict to meet your DQOs.

Risk Screening

Example

Decreasing risk →

(Red = Notes)

Concentrations above 1-in-100,000 cancer risk level at >25% of sites	**Concentrations above 1-in-1,000,000 cancer risk level at >50% of sites**	**Concentrations above 1-in-1,000,000 cancer risk level at 10-50% of sites**
Benzene	Arsenic ($PM_{2.5}$ and PM_{10})	Tetrachloroethylene
Acrylonitrile[1]	Acetaldehyde[2]	Cadmium (PM_{10} and TSP)
	Carbon tetrachloride	Naphthalene
	1,3-Butadiene	1,4-Dichlorobenzene
	Nickel (PM_{10} only)	Benzyl Chloride
	Chromium (estimated Cr VI from Cr $PM_{2.5}$)	

- This table displays only pollutants whose concentrations were monitored well enough to support a conclusion that they were above the relevant health levels of concern for pollutants for which at least 20 monitoring sites existed in the United States from 2003-2005.
- We are confident these cancer-risk pollutants are at or exceed the categories of cancer risk (i.e., may be higher, but are not lower)

[1] May have sampling issues biasing concentrations high, magnitude unknown

[2] May have sampling issue biasing concentrations low by a factor of 2 (Herrington et al., 2007)

Risk Screening

Summary

- Risk screening results at a national level are provided in the following table.
- At a regional, state, or local level, results may differ. This table provides a context for comparing local results.

Higher confidence – chronic cancer risk (ordered by importance)	Lower confidence – chronic cancer risk (ordered by importance)	High confidence – chronic and acute noncancer hazard
Benzene	Ethylene dibromide	Acrolein
Acrylonitrile[1]	1,1,2,2-tetrachloroethane	
Arsenic	1,2-dibromo-3-chloropropane	<u>Local chronic hazard</u>
Acetaldehyde[2]	Ethylene oxide	Formaldehyde
Carbon tetrachloride	Ethylene dichloride	Manganese
1,3-Butadiene	Hexachlorobutadiene	Acrylonitrile[1]
Nickel[3]	1,2-dichloropropane	1,3-Butadiene
Chromium[3]	1,1,2-trichloroethane	Nickel
Tetrachloroethene	Vinyl chloride	
Naphthalene	Trichloroethylene	
Cadmium	Benzo[A]pyrene	
1,4-Dichlorobenzene	Dibenzo[A,H]anthracene	
Benzyl chloride	3-Chloropropene	

[1] May have sampling issues biasing concentrations high, magnitude unknown

[2] May have sampling issue biasing concentrations low by a factor of 2 (Herrington et al., 2007)

[3] Concentrations adjusted to estimate toxicity based on subset expected to be in either Cr VI or Nickel subsulfide.

Summary (1 of 2)

Check List for Ways to Characterize Air Toxics

Temporal Characterization

- ☐ The general procedure for investigating temporal patterns is the same for all aggregates.
  - – Prepare valid concentration and normalized temporal aggregates and summary statistics.
    - Normalization allows comparison between sites and pollutants even if absolute concentration values vary widely.
    - Keep track of the amount of data below detection.
  - – Plot data with notched box plots or line graphs of multiple statistics (e.g., mean vs. 90th and 10th percentiles) with confidence intervals.
  - – Characterize patterns by pollutant
    - Do patterns fit your conceptual model?
    - Are they statistically significant?
  - – Investigate unexpected results
- ☐ Diurnal patterns – If alternate sampling schedules are used, calculate the weighted average by the most representative sampling hour; otherwise, diurnal patterns may be obscured.
- ☐ Day-of-week patterns – Examine data availability by day-of-week.
  - – If sufficient data exist for each day of the week, examine day-of-week patterns.
  - – If insufficient data exist, weekday vs. weekend groupings can be used.
- ☐ Seasonal patterns – Aggregate to the monthly level if sufficient data exist. Use quarterly averages if data are not sufficient or monthly patterns are too noisy.
- ☐ Compare what you have learned from the different temporal aggregates. Do conclusions make sense in the larger temporal picture?

  For example, the diurnal pattern of formaldehyde suggests that concentrations are highly dependant on sunlight. This dependency is confirmed by the seasonal pattern, which shows higher concentrations in summer (i.e., more sunlight.

Spatial Characterization

- ☐ General spatial patterns
  - – Create site level average values by pollutant for the time period of interest. Make sure data are temporally comparable at all sites.
  - – Investigate spatial variability by calculating and graphing summary statistics of the site averages. The results provide overview information about the magnitude of spatial variation.
  - – Visualize spatial variability by creating maps of the site-level average concentrations.
    - Results will provide more specific information about the spatial gradients of air toxics.
    - Including supplementary data such as MDLs, remote background concentrations, and cancer and noncancer risk levels provides a framework for the observed concentrations.
- ☐ Within- and between-city variation
  - – Calculate valid annual averages for each site within a city that has more than one monitor.
  - – Create notched box plots of annual averages by city.
    - Each box will contain one point for each monitor, so the box will indicate within-city variability.
    - Including multiple cities on one plot will provide a comparison of between city variability.
- ☐ Hot and cold spot analysis
  - – Calculate valid annual averages for each site.
  - – Rank the averages in order of concentration.
  - – Using maps, compare sites with highest and lowest concentrations to all sites.
  - – Investigate data and metadata for the sites with highest and lowest concentrations. Do concentrations make sense based on the metadata and conceptual models?
- ☐ Urban vs. rural site analysis
  - – Verify the EPA urban/rural designation of each site using Google Earth.
  - – Identify pollutant data availability and time period.
  - – Create a data set of pollutant/site combinations that are spatially and temporally representative.
  - – Plot valid 24-hr average data as a notched box plots for neighboring urban and rural sites.
  - – Summarize the results and investigate sites that do not meet the conceptual model of an urban or rural site.

Summary (2 of 2)

Check List for Characterizing Air Toxics

Risk Screening

- Create valid site average concentration data for the most recent years.
- Calculate the percent of sites above the selected risk level and the percent of data below detection.
- Follow the risk screening decision tree to identify the exposure risk for each pollutant.
- More advanced risk analyses should be performed by risk assessment professionals.

A Final Note on Data Below Detection

- Most air toxics have enough data below detection to cause uncertainties and/or biases in aggregated data if not handled properly.
- Note, however, that it is not valid to remove these data because they are representative of true values on the lower end of the concentration spectrum; removal would cause even more significant positive biases.
- It is always important to know the amount of data below detection when looking at any data set. The effects of data below detection should be considered in all analyses.
- In national analyses, we did not draw conclusions when more than 85% of the measurements of a pollutant was below detection.

Resources

- Statistical
  - StatSoft: Background on a variety of statistics http://www.statsoft.com/textbook/stathome.html
  - NIST Engineering Statistics: Background on a variety of statistics http://www.itl.nist.gov/div898/handbook/index.htm
  - SYSTAT: A graphical and statistical tool http://www.systat.com/
  - Minitab: A graphical and statistical tool http://www.minitab.com/Emissions
- Emissions
  - EPA AirData: Air toxics emissions reports to the county level http://www.epa.gov/air/data/reports.html
  - National Emissions Inventory 2002: Emissions inventory for the United States; some Canada and Mexico data also available. http://www.epa.gov/ttn/chief/net/2002inventory.html
  - EPA Toxics Release Inventory (TRI): A variety of emissions data sets http://www.epa.gov/triexplorer/chemical.htm.

References

Bortnick S.M. and Stetzer S.L. (2002) Sources of variability in ambient air toxics monitoring data. *Atmos. Environ.* **36**, 1783-1791 (11).

Demerjian K.L. (2000) A review of national monitoring networks in North America. *Atmos. Environ.* **34**, 1861-1884.

Fortin T.J., Howard B.J., Parrish D.D., Goldan P.D., Kuster W.C., Atlas E.L., and Harley R.A. (2005) Temporal changes in U.S. benzene emissions inferred from atmospheric measurements. *Environ. Sci. Technol.* **39**, 1403-1408 (6).

Grosjean D., Swanson R.D., and Ellis C. (1983) Carbonyls in Los Angeles air-contribution of direct emissions and photochemistry. *Sci. Total Environ.* **29**, 65-85 (1-2).

Grosjean D. (1982) Formaldehyde and other carbonyls in Los-Angeles ambient air. *Environ. Sci. Technol.* **16**, 254-262 (5).

Hafner H.R. and McCarthy M.C. (2004) Phase III air toxics data analysis workbook. Workbook prepared for the Lake Michigan Air Directors Consortium, Des Plaines, IL, by Sonoma Technology, Inc., Petaluma, CA, STI-903553-2592-WB, August.

Herrington, J.S., Fan, Z., Lioy P.J., and Zhang, J. (2007) Low acetaldehyde collection efficiencies for 24-hour sampling with 2,4-dinitrophenylhydrazine (DNPH)-coated solid sorbents. *Environ. Sci. Technol*. **41** (2), 580 -585.

Kao A.S. (1994) Formation and removal reactions of hazardous air-pollutants. *J. Air & Waste Manag. Assoc.* **44**, 683-696 (5).

Main H.H., Roberts P.T., and Reiss R. (1998) Analysis of photochemical assessment monitoring station (PAMS) data to evaluate a reformulated gasoline (RFG) effect. Final report prepared for the U.S. Environmental Protection Agency, Office of Mobile Sources, Fuels and Energy Division, Washington, DC, by Sonoma Technology, Inc., Santa Rosa, CA, STI-997350-1774-FR2, April. Available on the Internet at <http://www.epa.gov/oar/oaqps/pams/rfg_oms.pdf>.

Main H.H. and Bortnick S. (2002) Temporal variability in ambient air toxics: implications to monitoring network design. Presentation at the *Coordinating Research Council (CRC) Air Toxics Modeling Workshop, Houston, TX, February 26-27* (STI-2153).

McCarthy M.C., Hafner H.R., Chinkin L.R., Touma J.S., and Cox W.M. (2005) Temporal variability of selected air toxics: a national perspective. Prepared for the United States Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, and Sonoma Technology, Inc., Petaluma, CA. Available on the Internet at <http://www.epa.gov/ttn/amtic/toxdat.html> last accessed September 2, 2005.

McCarthy M.C., Hafner H.R., Chinkin L.R., and Charrier J.G. (2007) Temporal variability of selected air toxics in the United States. *Atmos. Environ.*, doi:10.1016/j.atmosenv.2007.1005.1037 (STI-2894).

McCulloch A. (2003) Chloroform in the environment: occurrence, sources, sinks and effects. *Chemosphere* **50**, 1291-1308 (10).

Seinfeld J.H. and Pandis S.N. (1998) *Atmospheric chemistry and physics: from air pollution to global change*, J. Wiley and Sons, Inc., New York, New York.

Singh H.B., Salas L., Viezee W., Sitton B., and Ferek R. (1992) Measurement of volatile organic-chemicals at selected sites in California. *Atmos. Environ. Part a-General Topics* **26**, 2929-2946 (16).

Spicer C.W., Buxton B.E., Holdren M.W., Smith D.L., Kelly T.J., Rust S.W., Pate A.D., Sverdrup G.M., and Chuang J.C. (1996) Variability of hazardous air pollutants in an urban area. *Atmos. Environ.* **30**, 3443-3456 (20).

US EPA Toxics Release Inventory (TRI) Explorer. Available on the internet at http://www.epa.gov/triexplorer/chemical.htm.

US EPA National Air Toxics Trends Stations. Available on the internet at http://www.epa.gov/ttnamti1/natts.html.

Vardoulakis S., Gonzalez-Flesca N., Fisher B.E.A., and Pericleous K. (2005) Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris. *Atmos. Environ.* **39**, 2725-2736 (DOI:10.1016/j.atmosenv.2004.05.067).

StatSoft, Inc. (2005) The Statistics Homepage. Available in the Internet at http://www.statsoft.com/textbook/stathome.html.