Suggested Analyses

What types of analyses could be done with my air toxics data?

Motivation

- Ambient air toxics have been monitored since 2001/2002 as part of NATTS and even longer as part of other monitoring programs. While national-level analyses have been conducted, it is important that these data be investigated at a local, state, and regional level to better understand an area's air toxics issues.
- Regular data analysis may be conducted annually to identify potential problems with the data at the site level. Adjustments can then be made in collection or analysis to improve data quality before several years of potentially poor quality data have been collected.
- A list of suggested air toxics data analyses has been provided (*Introduction*). This list is a potential minimum set of analyses that each area could perform.
- Key areas of interest
  - Is the quality of data sufficient for analysis?
  - How would air toxics be characterized in the area?
  - What are local sources of air toxics?
  - Are there changes in toxics concentrations over time?

Suggested Analyses

What's Covered in This Section

A set of potential analyses using Arizona data has been used as an example.

- This section outlines a sample analysis of an urban data set from start to finish in order to provide a thorough example. These data were previously assessed and readily available.
- Note that this is an example analysis and is not intended to show the only way air toxics analyses should be performed. Deviations or additional analyses may be necessary depending on the data or the analyst's objectives.
- The following topics will be covered following the sequence of this workbook
  - **Background**
    - Introduction to the data
    - Understanding sources
  - **Data validation**
    (Workbook Section 3)
    - Determining data completeness
    - Assessing data below detection
    - Identifying censored data
    - Using quality-controlled data
    - Applying data validation techniques
  - **Data characterization**
    (Workbook Section 4)
    - Putting data in perspective
    - Spatial patterns
    - Temporal patterns
    - Model-to-monitor comparisons
    - Risk screening
  - **Trends**
    (Workbook Section 5)
  - **Advanced analyses**
    (Workbook Section 6)
    Source apportionment

Phoenix, AZ

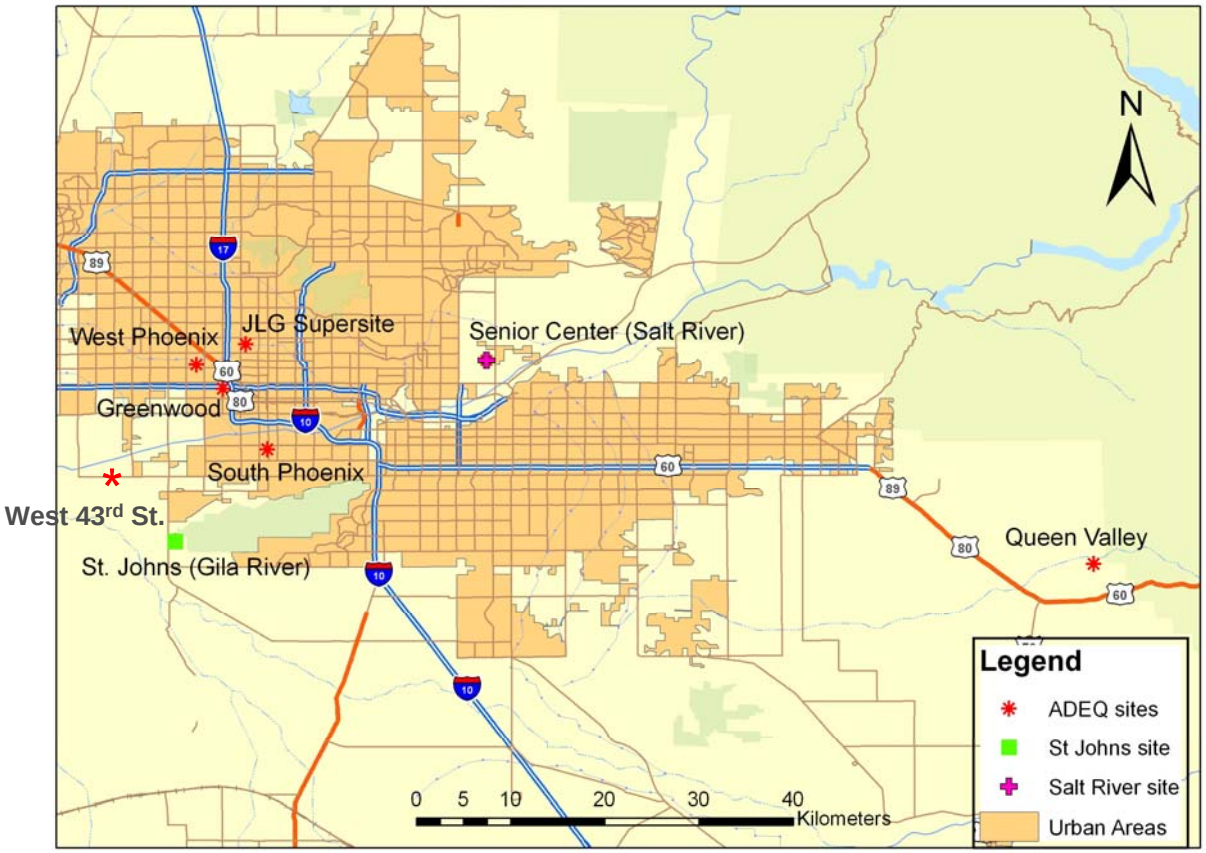
Introduction to the Data

Overview

- The sample data set used throughout this section is from an air toxics study performed in Arizona as part of the Joint Air Toxics Assessment Project (JATAP).
- The purpose of the study was to determine which air toxics are of most concern to the area and tribal communities.
- The study was conducted in two phases. (Analyses in this section focus primarily on Phase II data.)
  - Phase I: March 2003-March 2004
  - Phase II: February 2005 – March 2005
- Twenty-four-hour air toxics samples were collected every sixth day. On some days at some sites, two 12-hr samples were collected; for this analysis, these samples were 24-hr averaged. Only gaseous air toxics were collected and discussed here.
- A considerable quality assurance effort was made
  - Duplicate samples (collocated)
  - Replicate data (additional chemical analysis on canister)
  - Interlaboratory comparisons (more than one laboratory was involved)
  - Data validation
- For the trend assessment, we used historical data at two longer-term sites in the study area to illustrate air toxics concentrations over time in the area.

Introduction to the Data

Monitoring Site Locations

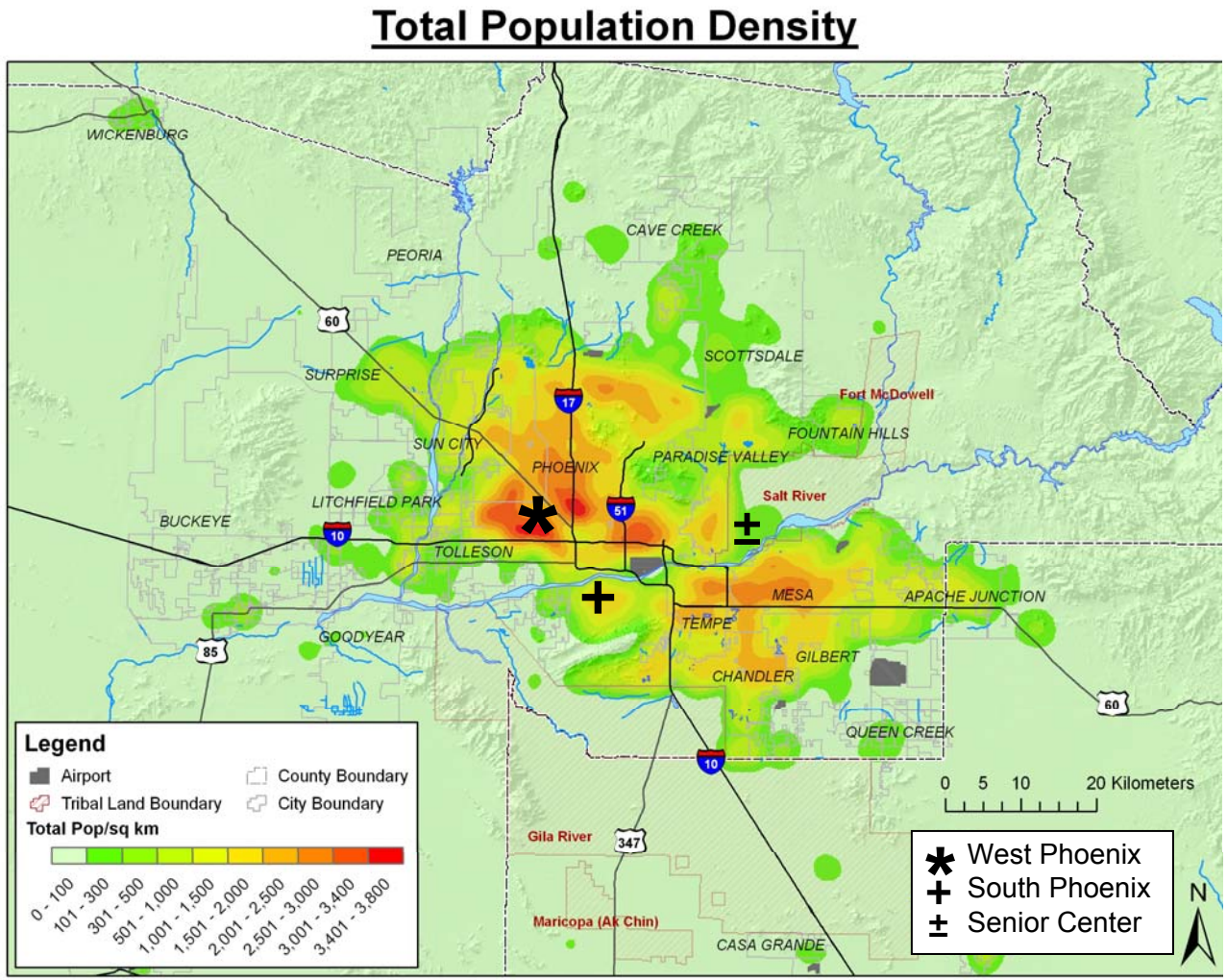


The map shows the eight monitoring sites in the study. The map was created with ArcMap. The West Phoenix, South Phoenix, and Senior Center sites are used most frequently in the sample analyses. The St. Johns site was operated by the Gila River Indian Community. The Senior Center site was operated by the Salt River Pima-Maricopa Indian Community.

Understanding Sources

Population Density

- The map shows population density in the study area. The three focus sites are indicated.
- Data from these sites help identify the most populated areas and potential air toxics source locations (e.g., high population density ≈ higher emissions).
- 2000 population density data were obtained from the U.S. Census Bureau.

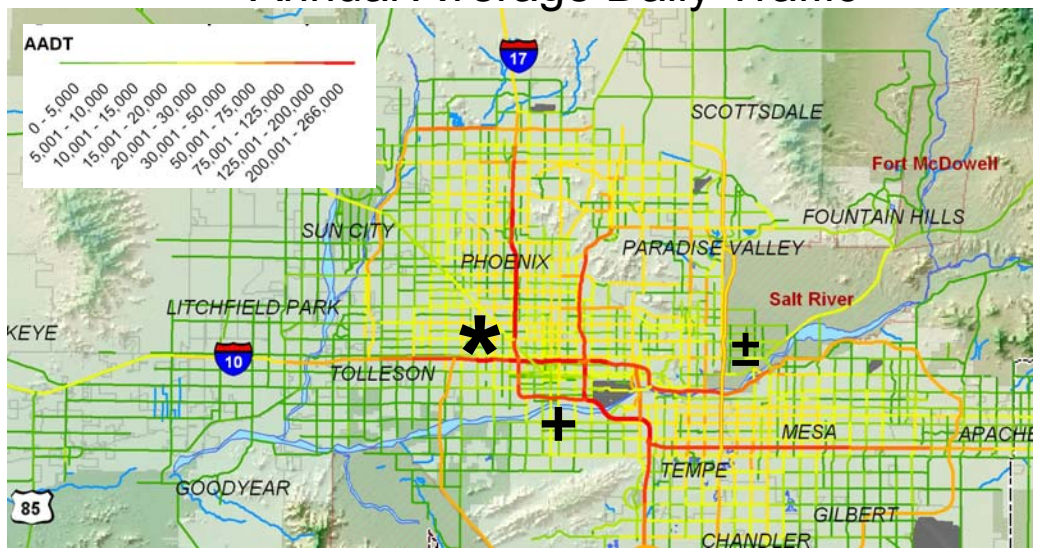


Understanding Sources

Mobile Sources

- The map shows annual average daily traffic (AADT) and heavy-duty vehicle (HDV) daily traffic for the study area (number of vehicles per day). The three sites of interest for this example are shown.
- AADT is an indicator of the relative on-road mobile source activity, and corresponding emissions levels, in the study area.
- Traffic data were obtained from the Arizona Department of Transportation (ADOT).

Annual Average Daily Traffic



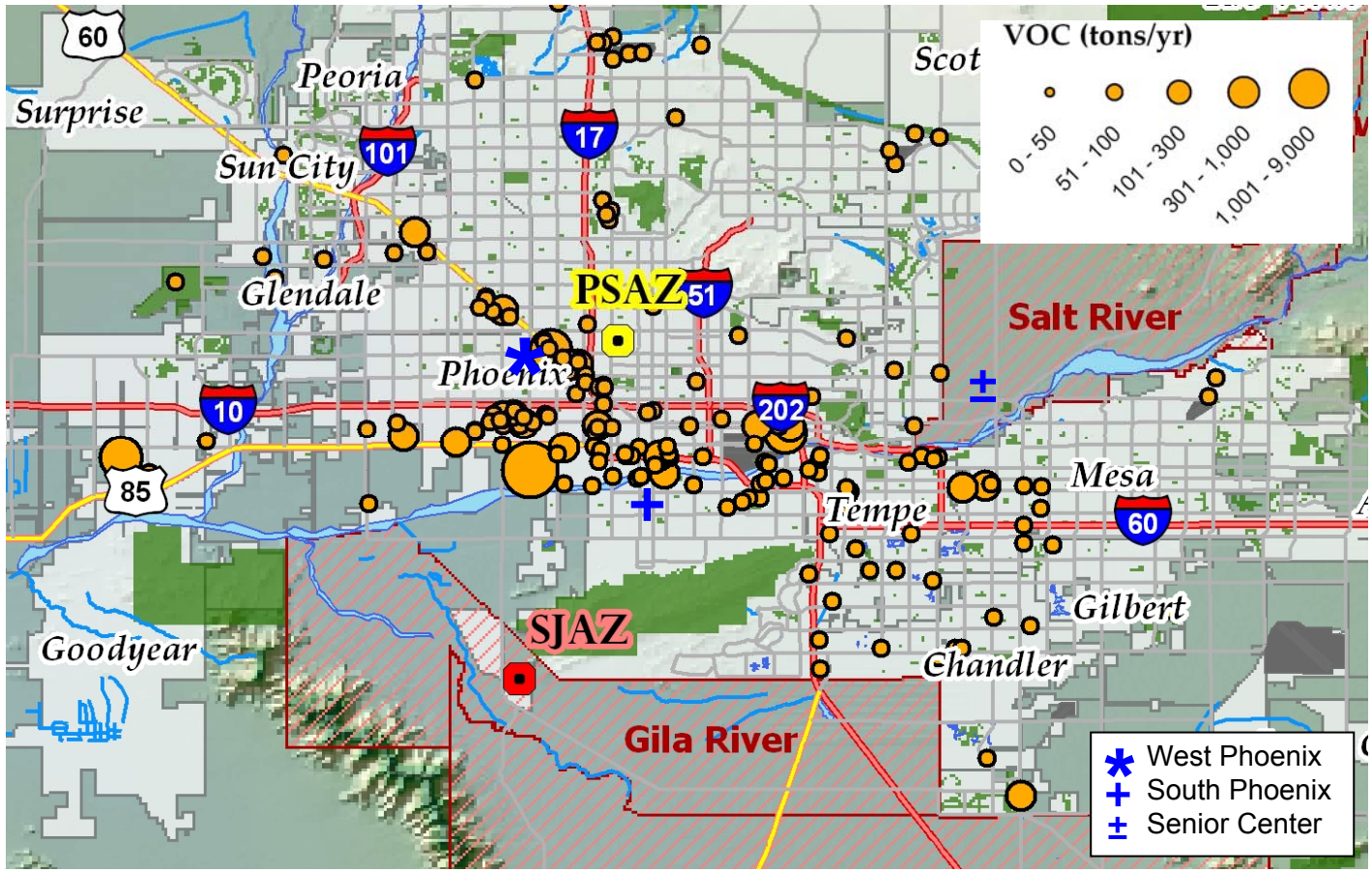
HDV Annual Average Daily Traffic

Understanding Sources

Point Sources

- The map shows point source emissions for total VOCs in the study area. The three sites of interest are shown on the map. Other sites in the area are also shown (Supersite [PSAZ] and St. Johns [SJAZ]).
- Note that mobile source emissions are not included in this data set (*see the average daily traffic maps on previous slide).*
- Emissions data were obtained from the 2002 NEI.

Point Source Emissions of VOCs



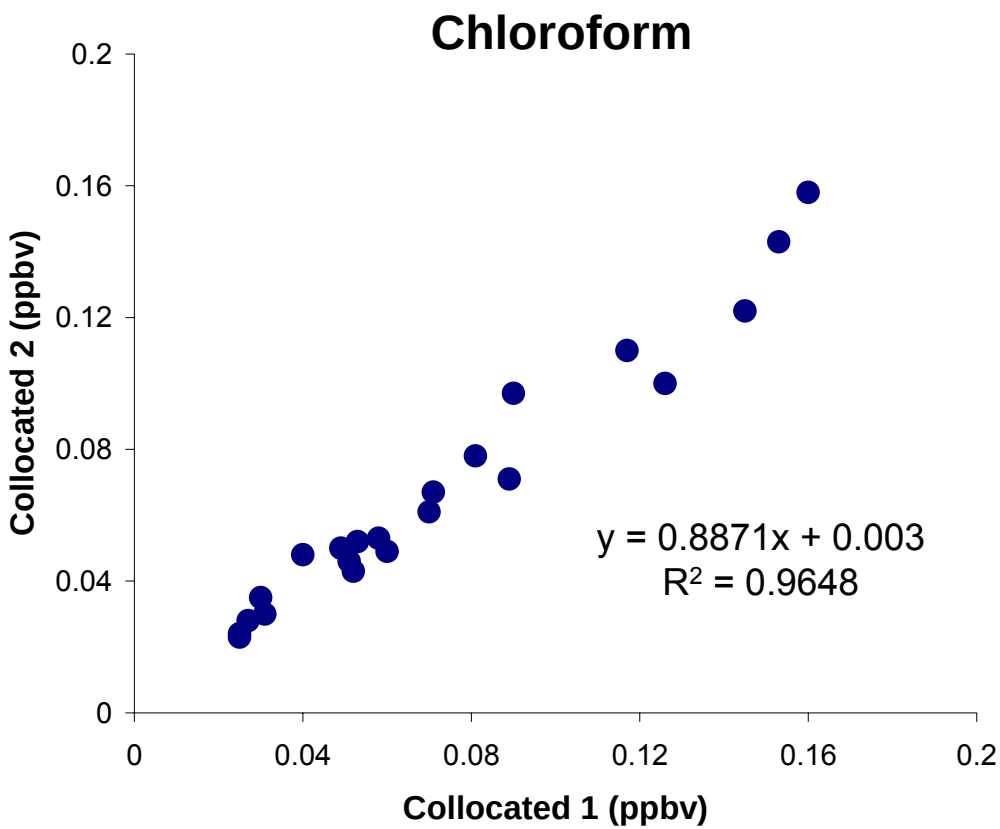
Using Quality Assurance Data
Overview

- Quality assurance (QA) is performed during sample collection and analysis to provide additional information about data quality and usefulness.
  - Collocated samples indicate agreement between *sample collection*
  - Replicate samples indicate agreement between *sample analysis*
- These data provide insight into biases and error that may occur in the process of collecting and analyzing samples.

Using Quality Assurance Data

Visual Inspection of Collocated Samples (1 of 2)

- Visual inspection of collocated samples is important to identify outliers and understand sampler performance.
- Collocated data for chloroform are plotted in the figure.
- The data indicate that chloroform is consistently measured; however Sampler 2 reported slightly lower values than Sampler 1 at higher concentrations.

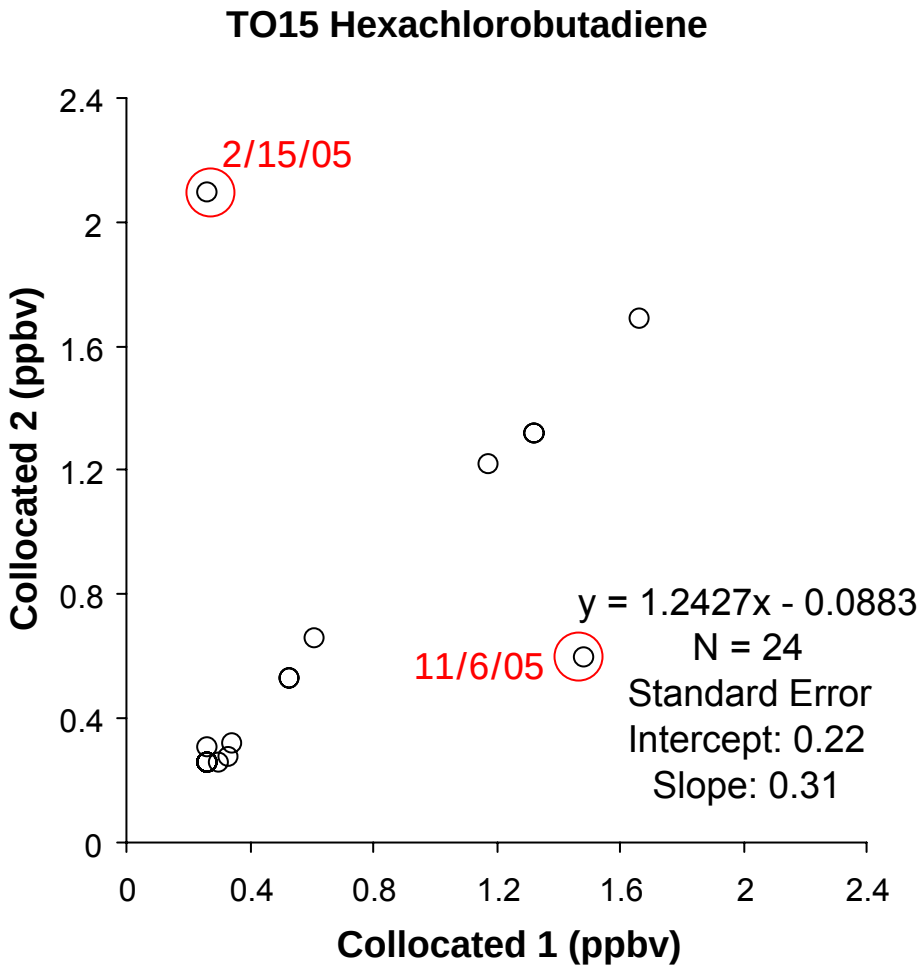


The figure shows collocated chloroform samples collected in the study. It was created with Microsoft Excel.

Using Quality Assurance Data

Visual Inspection of Collocated Samples (2 of 2)

- In this figure, collocated data for hexachlorobutadiene are plotted to the right; outliers are circled in red. Outliers identified from collocated samples should be excluded from further data analyses.
- The data indicate that hexachlorobutadiene is not consistently measured; Sampler 2 reported lower values than Sampler 1 at high concentrations. This is consistent with observations of collocated chloroform data.



Using Quality Assurance Data

Summarizing Sample Problems for Analysis

- The table shows an excerpt from the list of measurements, identifying problems in one of the study area site replicate comparisons.
- In site-level analyses, we typically exclude any of these failures. We flagged as suspect the pollutant identified as a problem in the indicated sample and did not use this pollutant/sample combination in subsequent analyses (e.g., toluene on 7/26/03).
- Flag 1 indicates that the percentage error was greater than 50%. Flag 2 indicates that the absolute difference in the two species was greater than three times MDL. Flag 3 indicates that the replicate or collocated average was suspect.

Date	Species Name	Flag 1	Flag 2	Flag 3	Suspect
7/26/2003	Toluene	x	x		x
7/26/2003	1,3,5-trimethylbenzene	x			x
7/26/2003	1,2,4-trimethylbenzene	x			x
8/25/2003	MTBE			x	x
8/25/2003	Methyl ethyl Ketone			x	x
8/25/2003	n-octane			x	x
8/25/2003	1,3,5-trimethylbenzene			x	x
8/25/2003	1,2,4-trimethylbenzene			x	x
9/24/2003	Methyl ethyl Ketone		x		x

Data Completeness
Overview

- For the site-level analysis, we summarized available data and calculated data completeness based on expected samples.
- This step included calculating the number of valid samples versus the expected number of samples based on collection frequency.
- In general, 75% data completeness is required to calculate valid aggregated values (e.g., monthly, quarterly, and annual averages).
- See *Preparing Data for Analysis,* Section 4, for a complete description of methods and rationale.

Data Completeness

Site-Level Summary

Site	Sampling	Sampling Duration	Samples Expected	Samples Available	Valid Samples	Percent Valid
Greenwood	Cartridges[a]	24-hr	61	60	60	98
	Canisters	24-hr	61	61	59	97
JLG Supersite	Cartridges[a]	24-hr	61	61	49	80
	Canisters	24-hr	61	61	55	90
Queen Valley	Canisters	24-hr	31	31	30	97
St. Johns	Canisters	24-hr and 12-hr	30 (24-hr) 62 (12-hr)	37 (24-hr) 44 (12-hr)	79	95[b]
Senior Center	Canisters	24-hr and 12-hr	30 (24-hr) 62 (12-hr)	37 (24-hr) 46 (12-hr)	83	98[b]
South Phoenix	Cartridges[a]	24-hr	61	60	52	85
	Canisters	24-hr	61	60	59	97
West Phoenix	Canisters	24-hr	61	60	59	97

- The table shows data necessary to calculate the data completeness and the percent of valid data. The number of valid samples was computed after data validation steps but shown here for a complete summary.
- A high percentage of samples from all sites were valid.
- Additional samples may be marked as suspect during the process of data analysis.

[a] Carbonyls only.

[b] This percentage is based on 24-hr average sample days.

Assessing Data Above Detection

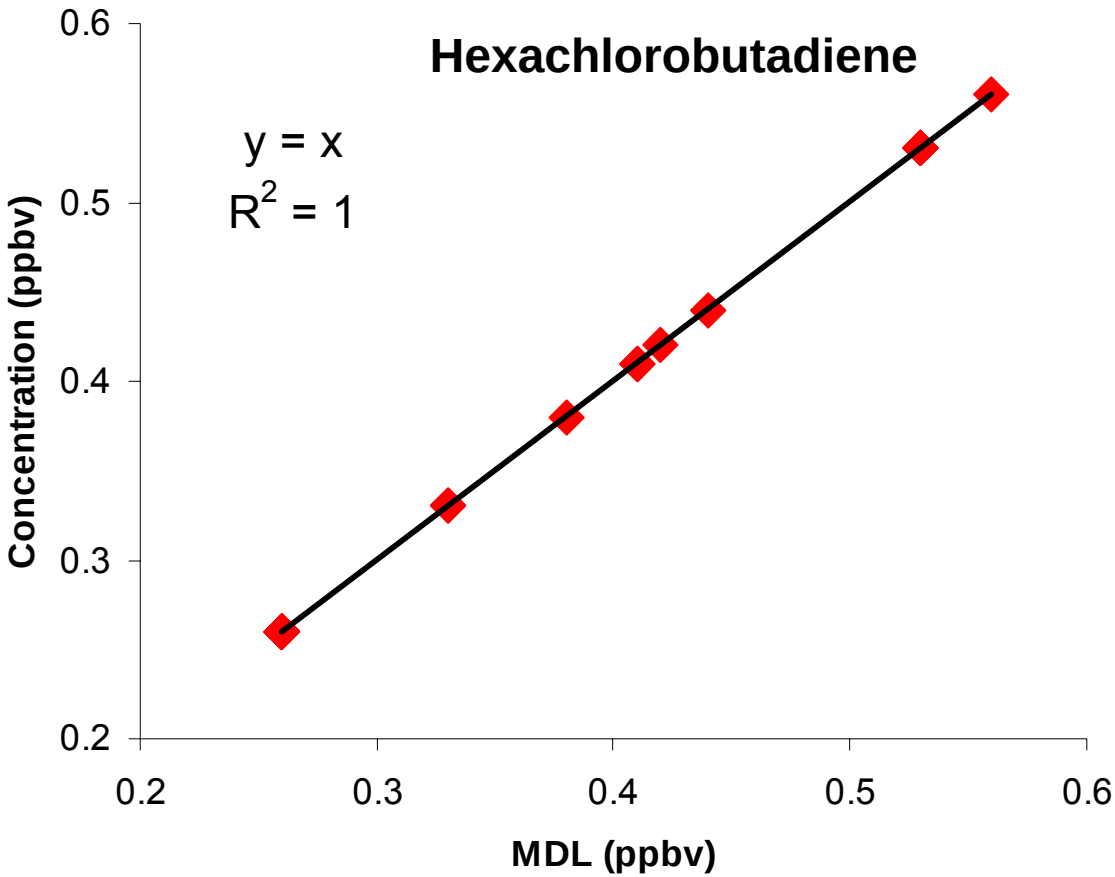
Species	2005 Percent Above MDL						
	St. Johns	Senior Center	South Phoenix	West Phoenix	Green-wood	JLG Supersite	Queen Valley
Benzene	100	99	100	100	100	100	100
Bromomethane	40	36	37	49	24	33	23
Carbon tetrachloride	89	89	89	83	100	100	100
Chloroform	43	90	77	83	98	100	53
Dichloromethane	76	94	97	98	100	100	97
Ethylbenzene	71	92	92	94	100	100	93
Hexachloro-butadiene	0	0	0	0	2	4	0

- The percent of data above detection should be calculated for each pollutant, site and year; additional calculations will be needed if monthly or seasonal aggregates are produced. The table shows an excerpt of the entire data set - the percent of data above detection for 2005. This example spans the range of data above detection observed in the data set.
- Data were color-coded in the table to illustrate potential patterns in data availability. More data were below detection at St. Johns and Queen Valley, consistent with their location away from sources. Hexachlorobutadiene was typically below MDL at all sites.

< 25% Above MDL
25% to 75% Above MDL
>= 75% Above MDL

Identifying Censored Data

- Alternate MDLs were included with the study data. Because alternate MDLs are often different for each sample, it is not always clear from the data that censoring (e.g., substitution with MDL or MDL/2) has occurred. We need to ensure that all samples are treated similarly when data are aggregated.
- Scatter plots are an easy way to identify whether data below detection are censored.
- Plot all data points that are less than or equal to the alternate MDL.
- The agreement between concentration and MDL indicates that the alternate MDL was substituted for values below detection. These samples were identified and MDL/2 substitution was subsequently applied for data aggregation.



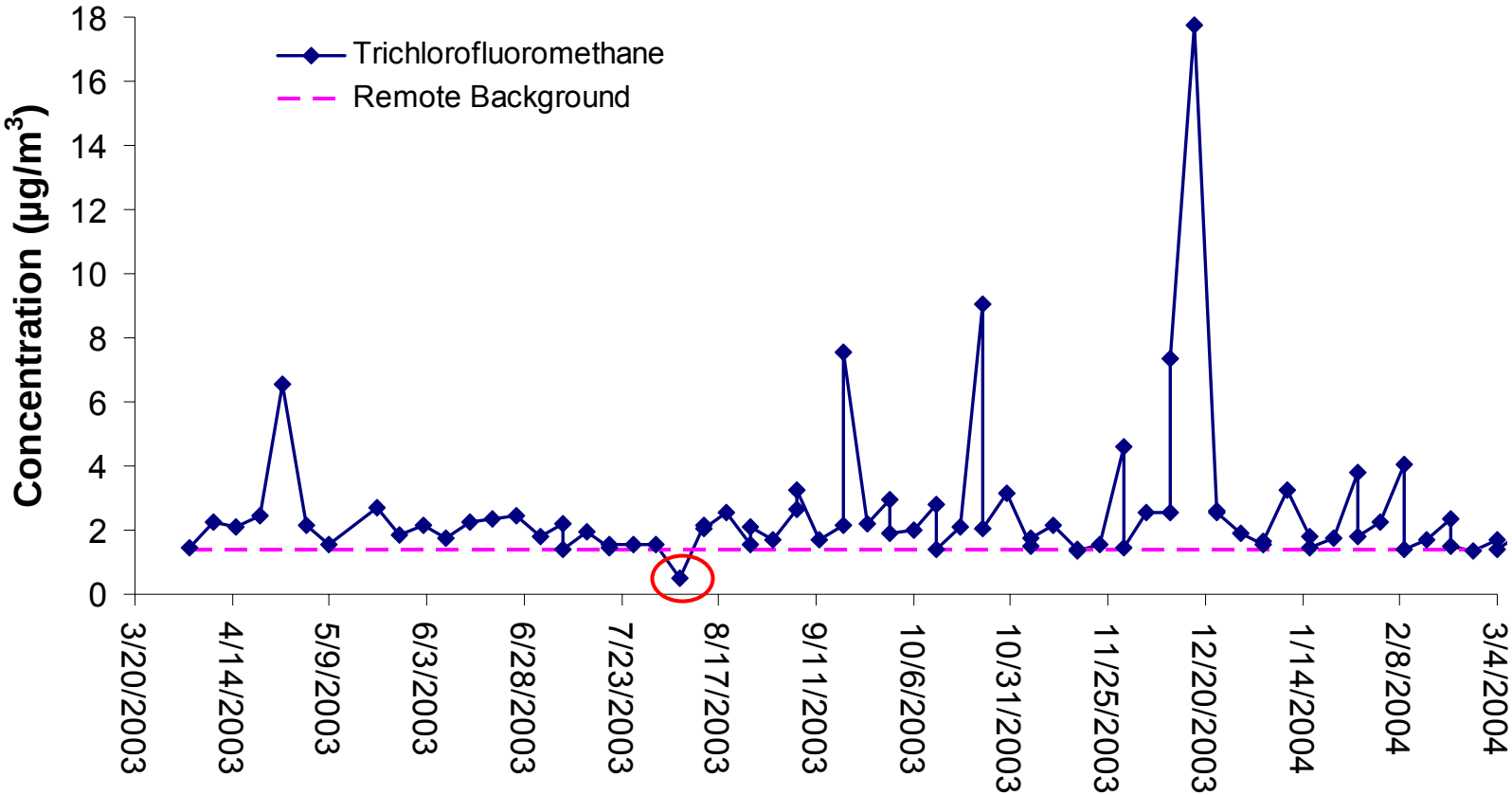
The graph shows the comparison of concentration values to their MDL for data at or below detection. It was created with Microsoft Excel.

Validation Techniques
Overview

- Once data are received from the laboratory, or a data repository such as AQS, it is useful to apply screening criteria during the early stages of data validation to identify suspect data that may not be representative of actual ambient concentrations.
- Basic visual analyses should be performed to identify potential problems in the data and to begin to understand data characteristics.
- Knowledge of similarity of sources, lifetime, and reactivity should be used to assist in data validation.
- The following screening checks are typically used
  - Comparison to remote background concentrations. Urban air toxics concentrations should not be lower than remote background concentrations.
  - Range checks. Check minimum and maximum concentrations for anomalous values.
  - Buddy site check. Compare concentrations at one site to nearby sites to look for anomalies.
  - Sticking check. Check data for consecutive equal data values which indicate the possibility of censored data not flagged appropriately.
  - Scatter plots. Investigate the relationship between species to identify sources and suspect data.
  - Fingerprint plots. Investigate the pattern of species concentrations and relationships among species to identify sources and suspect data.
- See the *Preparing Data for Analysis, S*ection 4, for a complete description of methods and rationale.

Validation Techniques

Remote Background Check

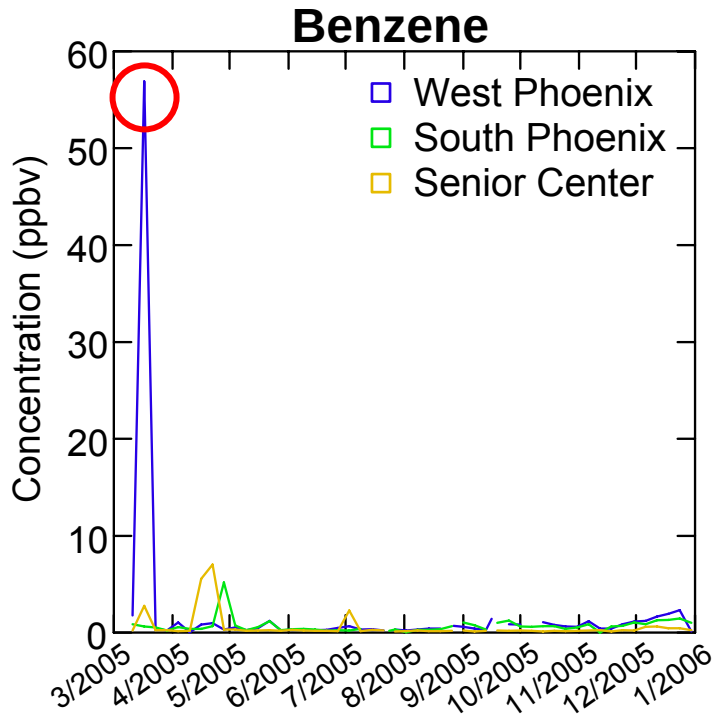


- The plot shows a time series of concentrations of trichlorofluoromethane compared to background concentrations measured at remote sites in the Northern Hemisphere.
- A significant dip in concentrations is circled in red. Concentrations at this monitor were typically equal to or greater than background concentrations, as expected for urban locations.
- The circled value was more than 20% below the background level and was identified as suspect for further review.

Validation Techniques

Buddy Site Check

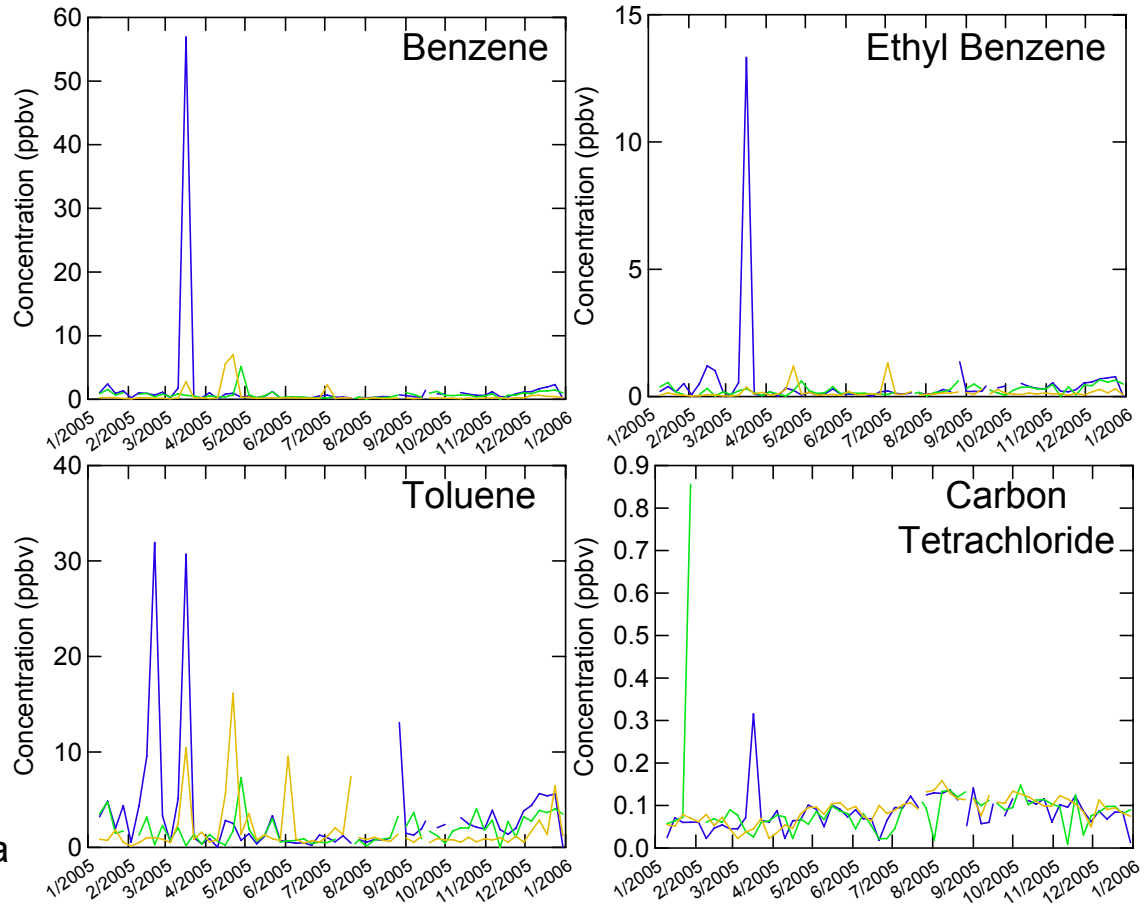
- Buddy site checks are useful in identifying suspect data.
- In the example, time series of benzene concentrations for three sites are plotted.
- There is clearly a suspect data point at the West Phoenix site in March 2005, which is not corroborated by the other sites. This indicates that the data point should be considered suspect because a concentration spike of that magnitude should register at nearby sites.
  - Investigation into these data showed that this event corresponds to a single data point significantly higher than the others.
  - Further investigation revealed that many species showed the same behavior at the West Phoenix site. The site may be impacted by a local source or sources.



Validation Techniques

Time Series

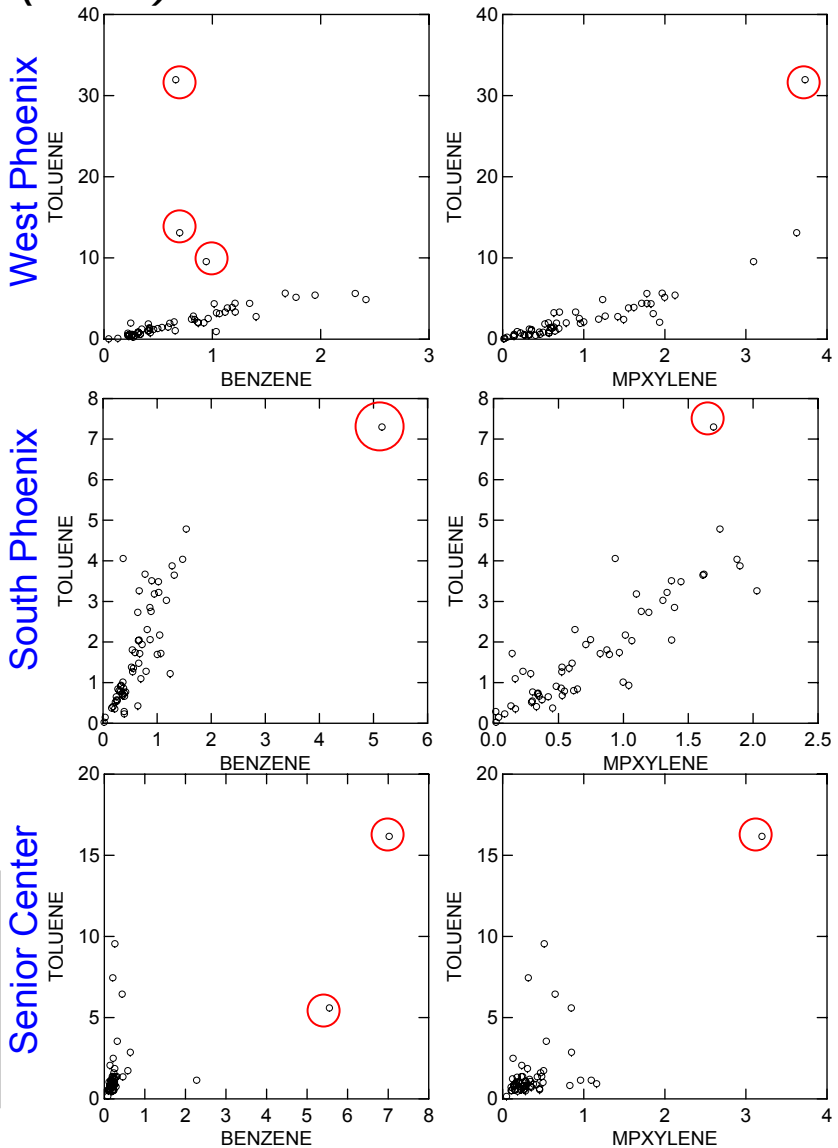
- The figures show the same benzene time series as the previous slide and matching time series for a variety of other compounds.
- Benzene, ethylbenzene, and toluene can all be emitted by mobile sources.
- The fact that these species peak at the same time is suspicious, because an increase of that magnitude from typical mobile source emissions is unlikely. However, an unusual event may have occurred, such as a gasoline spill very near the West Phoenix site that could have led to the high concentrations.
- Examining the time series of carbon tetrachloride helps confirm or reject this theory because there are no likely sources that would cause a spike of that magnitude. The time series of carbon tetrachloride shows a spike on the same day indicating that the event is in fact an instrument or analysis error. All data for that date and site should be flagged as suspect and not used in subsequent analyses.



Validation Techniques

Scatter Plots (1 of 2)

- The scatter plots show the relationship between toluene and benzene and toluene and m,p-xylene at three study sites. This method is another way to identify suspect data, which have been circled in red in the figures.
- At the West Phoenix site, the correlation between toluene, benzene, and m,p-xylene is strong, indicating that this site is highly mobile source-dominated.
- Outlier data points may point to data issues or other source influences. For toluene outliers, high toluene concentrations are often associated with solvent use or surface coatings; thus, the samples are likely valid.
- The correlations at the South Phoenix site are not quite as strong, but still indicate that the site is likely mobile source-dominated.
- The Senior Center site, on the other hand, shows a weak correlation between the three species as expected for a site farther from fresh emissions.

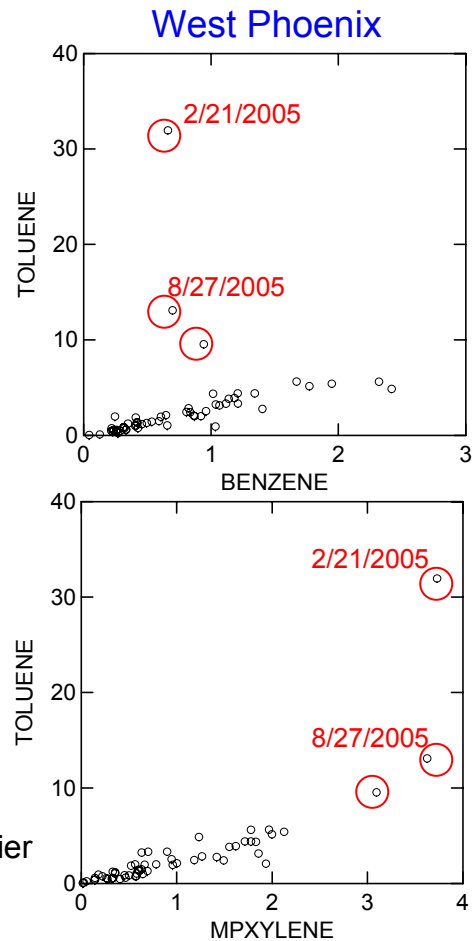


◯ = outlier

Validation Techniques

Scatter Plots (2 of 2)

- The figures show the same data as in the previous slide for the West Phoenix site only. The dates of the two highest outliers have been marked.
- The outlier values all correspond to the unusually high toluene concentrations. Significantly, the three toluene outliers correspond with the three highest m,p-xylene events.
- These correlations indicate that the high concentrations may not be due to collection or analysis errors, but may indicate solvent or surface-coating emissions impacting the site. Further exploration might include assessing the importance of these concentrations on the annual average and looking for possible sources of toluene in the emission inventory.
- The table below shows emission profiles for surface coating from EPA's SPECIATE. Xylenes and toluene account for almost one-third of this source profile supporting the hypothesis that the high concentration events are solvent-driven.



○ = outlier

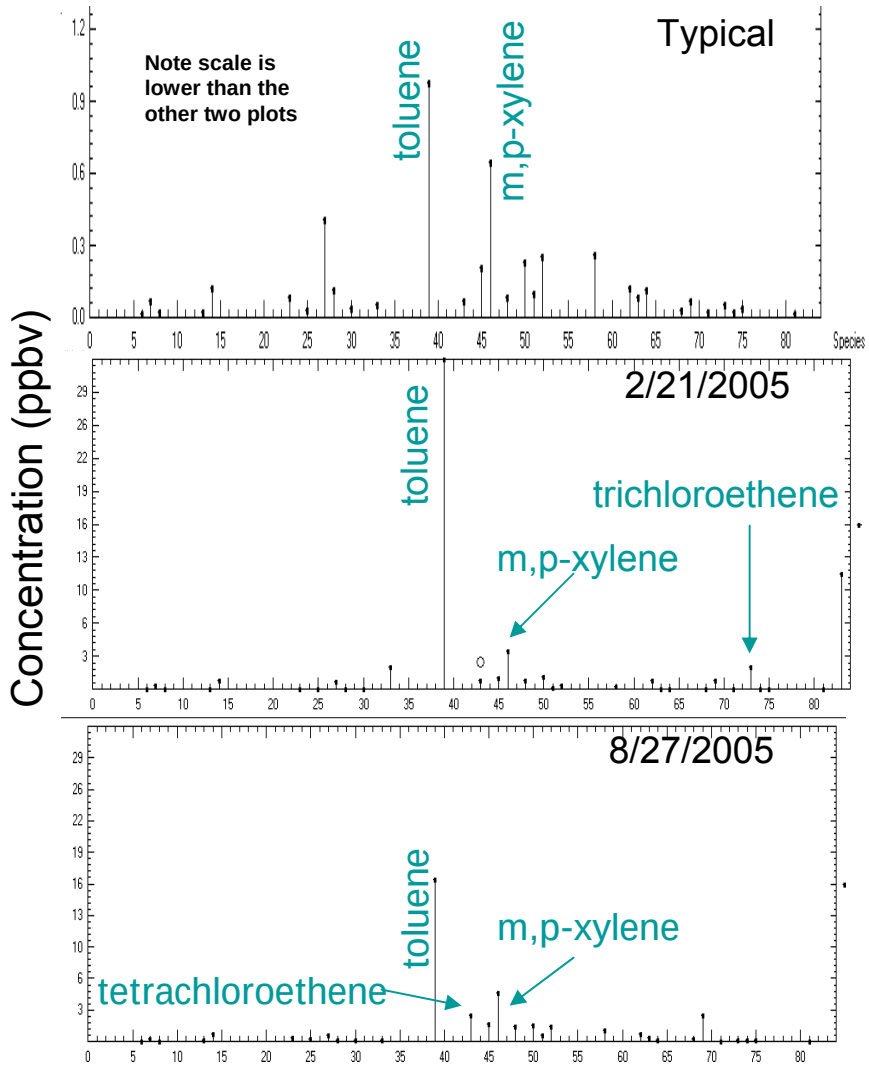
Profile Number: 6002
Profile Name: Surface Coating Operations (Industrial)
Percent Total: 100

POLLUTANT	CAS No.	Percent
ISOMERS OF XYLENE	1330207	15.800
TOLUENE	108883	14.700
METHYL ETHYL KETONE	78933	8.100
DIETHYLENE GLYCOL	111466	6.600
N-BUTYL ALCOHOL	71363	6.400

Validation Techniques

Fingerprint Plots

- Fingerprint plots represent concentrations of all species by date.
- They are useful for identifying relative pollutant concentrations on typical and unusual days.
- A typical fingerprint can be quantitatively determined (e.g., median sample composition) or qualitative (e.g., visual inspection of all fingerprints).
- The figures to the right show a typical fingerprint plot and fingerprint plots for 2/21/2005 and 8/27/2005 (the two dates of the highest outlier events in the previous slides).
- A review of fingerprints listed in EPA's SPECIATE shows that toluene and xylenes are prominent components of surface coatings.



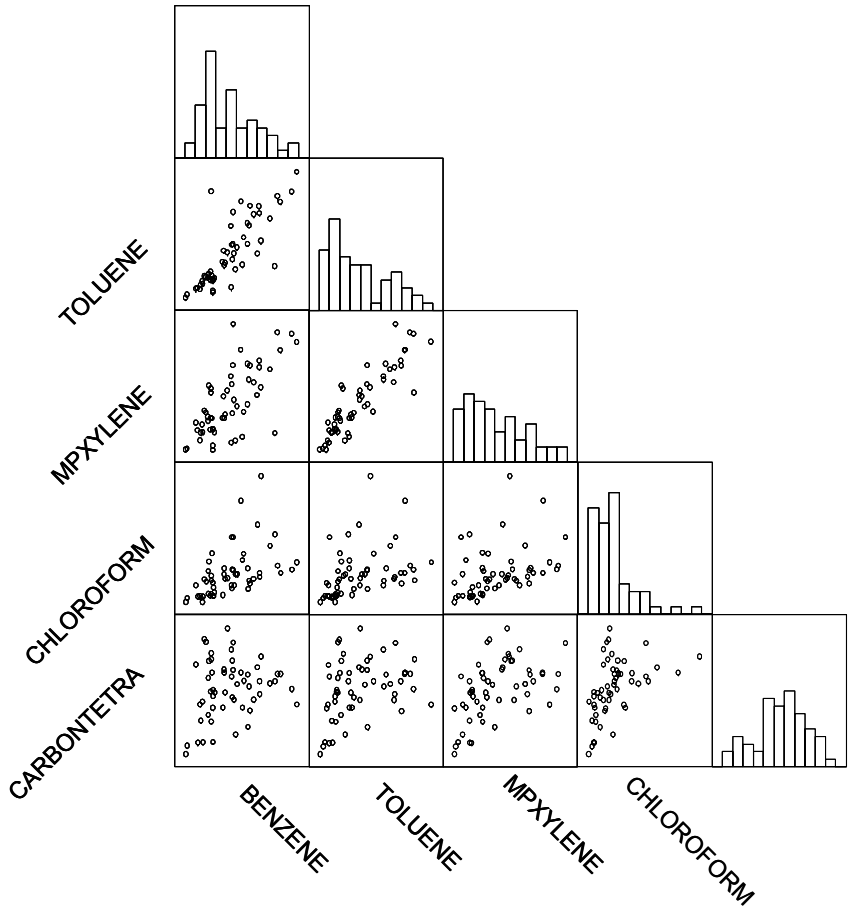
Validation Techniques
Summary

- What have we learned from applying these validation techniques?
  - Additional invalid and suspect data points were identified.
  - Data quality and limitations are better understood.
  - Spatial and temporal characteristics of the data are more thoroughly indicated.
  - Hypotheses about possible source influences for further investigation can be formed.
- These are a few examples of the data validation process that would be performed on the data set.
- Remember, data validation continues as part of data analysis.

Basic Understanding of Data

Scatter Plot Matrices

- Scatter plot matrices provide a quick and easy way to view correlations and outliers within a large amount of data.
- Scatter plot matrices are interpreted by matching the pollutant name on the row and column corresponding to the scatter plot. Histograms showing the distribution of measured values for each pollutant are included along the top diagonal.
- The graph to the right shows scatter plot relationships for five pollutants at the South Phoenix site. Note that previously identified outliers have been removed.
- The data show a clear correlation between toluene, m,p-xylene, and benzene, indicating that these pollutants are likely from mobile sources. Chloroform also shows a slight correlation with the mobile source pollutants (across the second row from the bottom) but the bifurcated relationship indicates a secondary source. Carbon tetrachloride shows little correlation with any species and shows a histogram that is roughly Gaussian, as expected for background pollutants.



Putting Data In Perspective

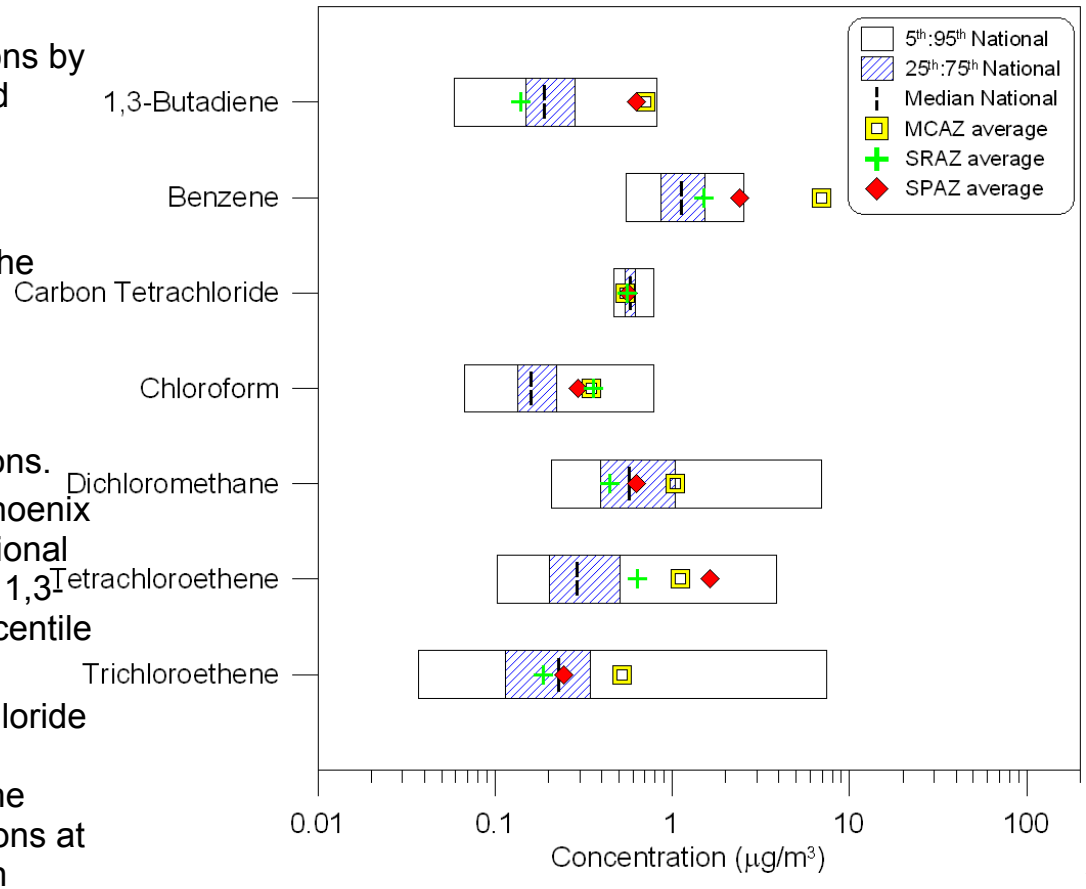
Overview

- Putting concentrations and MDLs into perspective provides a framework for comparing site-level concentrations to national levels and to other sites in the area.
- This information is useful in assessing whether concentrations are typical, low, or high and can help explain the impact of local source emissions on monitored concentrations.

Putting Data In Perspective

National Concentrations

- The figure shows the national 5th-95th, 25th-75th, and 50th percentile concentrations by species (bars) compared to site-averaged concentrations (symbols).
- Though Senior Center is the most rural (although within a few miles of urban emissions) of the other sites included in the figure, concentrations are typically higher than the national median and sometimes higher than the national 75th percentile concentration, showing that the site is impacted by urban emissions.
- Concentrations at the West and South Phoenix sites are also typically well above the national median. Concentrations of benzene and 1,3-butadiene are near or above the 95th percentile of national concentrations.
- National concentrations of carbon tetrachloride fall within a very small range due to its ubiquitous background concentration. The average carbon tetrachloride concentrations at all study sites are in good agreement with national levels, providing confidence that data collection in the study is representative of national data collection methods.

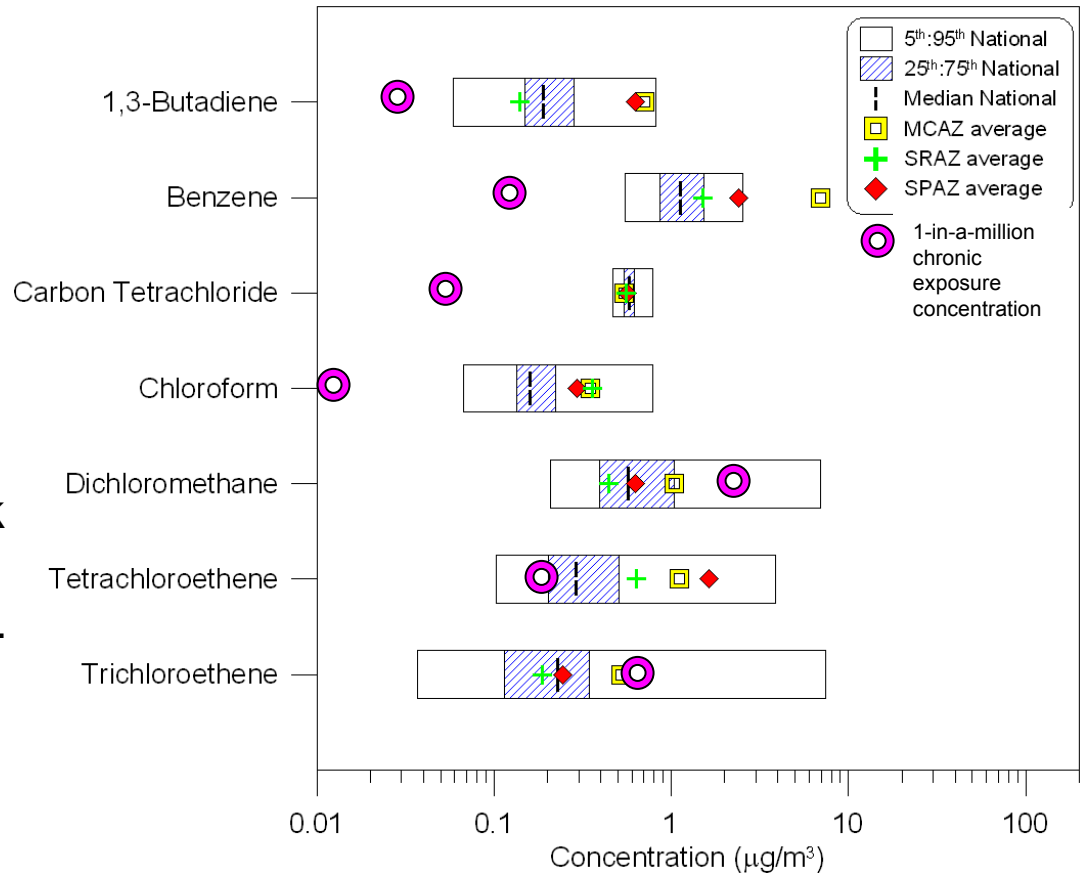


MCAZ = West Phoenix
SPAZ = South Phoenix
SRAZ = Senior Center

Putting Data In Perspective

Cancer Risk

- The figure shows the same data as the previous slide, with the addition of the chronic exposure concentration associated with a 1-in-a-million cancer risk to place health risks in perspective.
- Concentrations could be compared to other cancer risk levels: 0.1-in-a-million, 10-in-a-million, 100-in-a-million, etc.
- Concentrations are typically higher than the 1-in-a-million cancer risk level shown except for dichloromethane and sometimes trichloroethene.

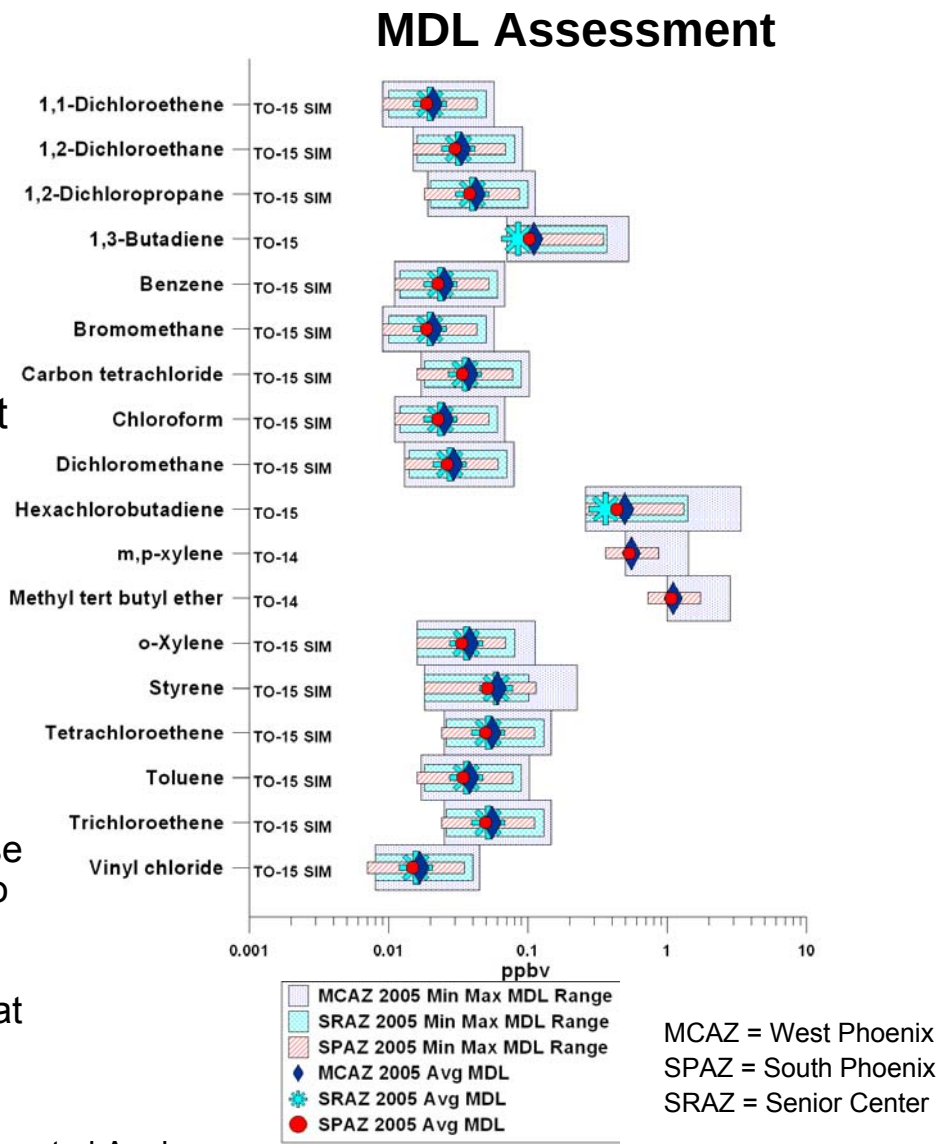


MCAZ = West Phoenix
SPAZ = South Phoenix
SRAZ = Senior Center

Putting Data In Perspective

MDLs

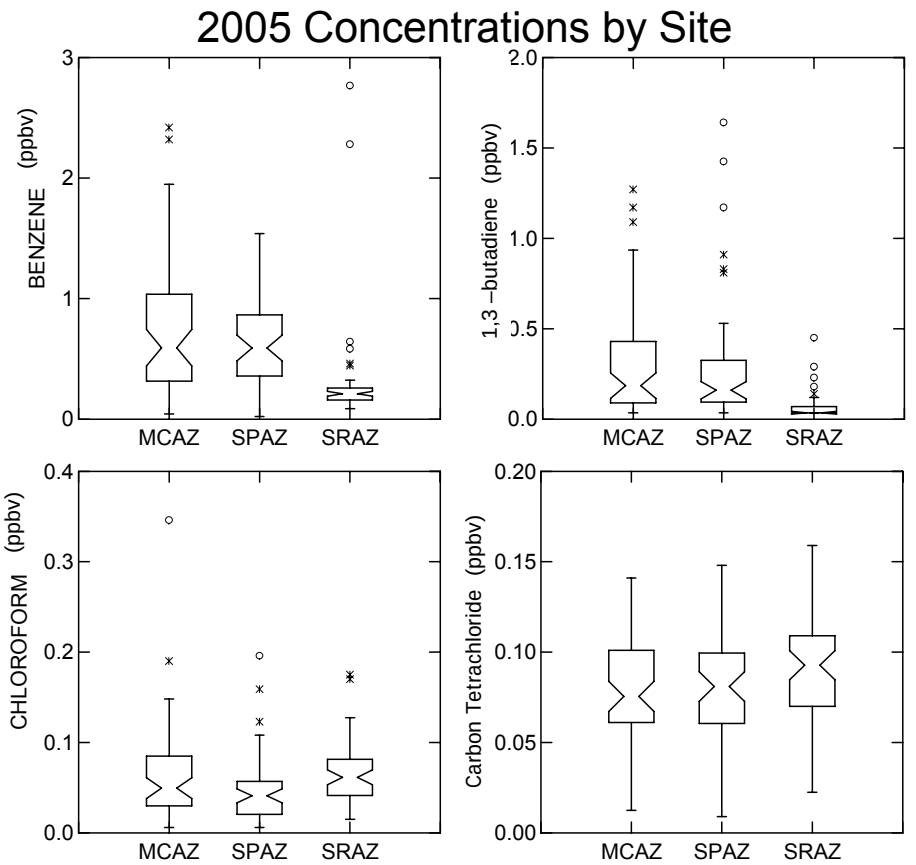
- Examining the relationship between MDLs at multiple sites is imperative to check that MDL/2 substitutions are not biasing the data differently at different sites.
- The graph shows the average MDL and minimum-to-maximum MDL range for three study sites.
- This graphical method allows the analyst to quickly confirm that MDLs are very similar between sites.
  - MDLs at the West Phoenix site (light purple bar) are sometimes higher than at other sites.
  - The difference is not enough to cause a major bias unless a high percentage of data is below the MDL. For example, hexachlorobutadiene is typically below detection so MDL/2 substitution may cause concentrations at the West Phoenix site to appear higher than at the other sites. However, hexachlorobutadiene, such a large portion of data is below detection that it cannot be reliably used for many analyses in the first place.



MCAZ = West Phoenix
SPAZ = South Phoenix
SRAZ = Senior Center

Spatial Patterns

- Understanding spatial patterns is important and can provide insight into
  - Improving monitoring networks
  - Verifying and improving emission inventories
  - Verifying and improving models
  - Identifying sources
- The box plots show 2005 concentrations of benzene, 1,3-butadiene, chloroform, and carbon tetrachloride at three study sites.
- Benzene and 1,3-butadiene concentrations are higher and more variable at the West and South Phoenix sites.
  - The lower concentrations and especially lower variability at the Senior Center site indicates that the site is removed from primary sources and is representative of the regional background.
- Chloroform and carbon tetrachloride are relatively consistent at all sites.
  - This behavior is expected for carbon tetrachloride which should be at background levels across the United States.
  - That chloroform does not follow the same pattern as benzene and 1,3-butadiene indicates the compounds probably have different sources. Benzene and 1,3-butadiene are primarily emitted by mobile sources while chloroform is emitted primarily from industrial operations.



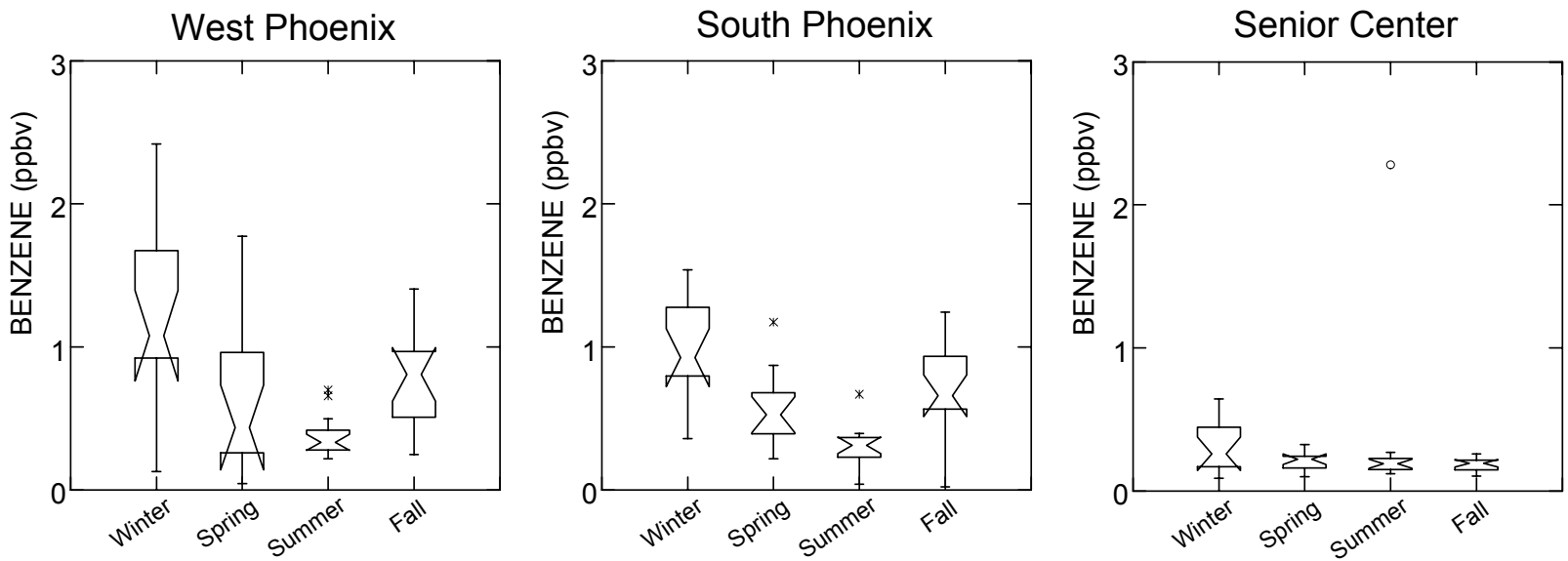
MCAZ = West Phoenix
SPAZ = South Phoenix
SRAZ = Senior Center

Temporal Patterns
Overview

- Characterization of temporal patterns can provide information on sources, physical or chemical processes affecting air toxics concentrations, and additional data validation.
- Before beginning temporal characterization, it is recommended to create valid aggregated data sets (examples in *Characterizing Air Toxics,* Section 5) to ensure the data are representative.
- There are sufficient data records in the example data set (i.e., one year of samples collected every sixth day) to characterize seasonal and weekday/weekend patterns.
- There are too few records in this data set to create day-of-week patterns (i.e., 95% confidence intervals on the means will overlap too much across the days because of the small sample size).
- 1- to 3-hr samples were not collected so diurnal patterns cannot be investigated.

Temporal Patterns
Seasonal

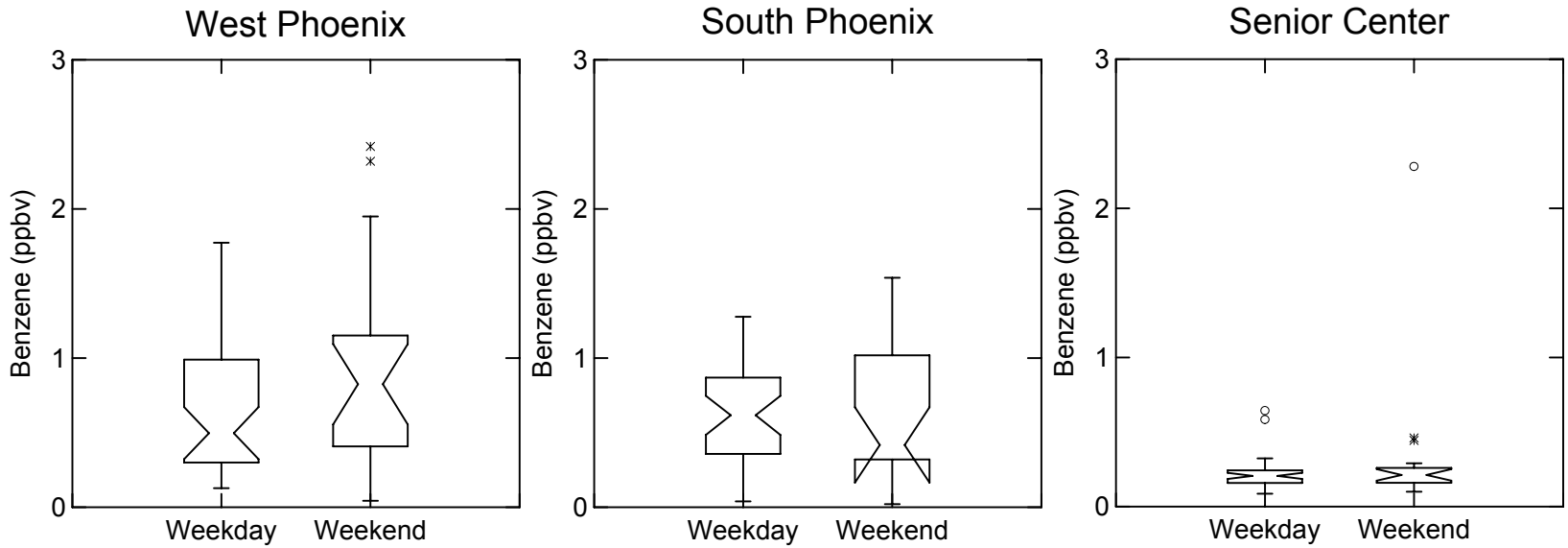
- The figures show seasonal patterns for benzene at three sites.
- The South and West Phoenix sites show typical benzene seasonal patterns (see *Characterizing Air Toxics*, Section 5) with lower concentrations during warm months and higher concentrations during cooler months. This is a result of mixing height differences and reactivity with season as opposed to changes in sources.
- At the Senior Center site, benzene shows an invariant seasonal pattern. While we expect higher concentrations in winter, note that the concentrations are generally lower during all seasons at this site. All samples are well-mixed upon arriving at the Senior Center and are similar to summer concentrations at the other sites.
- These data follow expectations for urban and downwind sites. The seasonal variability for these pollutants shows that for the urban data, computed annual averages without the winter quarter would be biased low and vice versa for a missing summer quarter.



Temporal Patterns

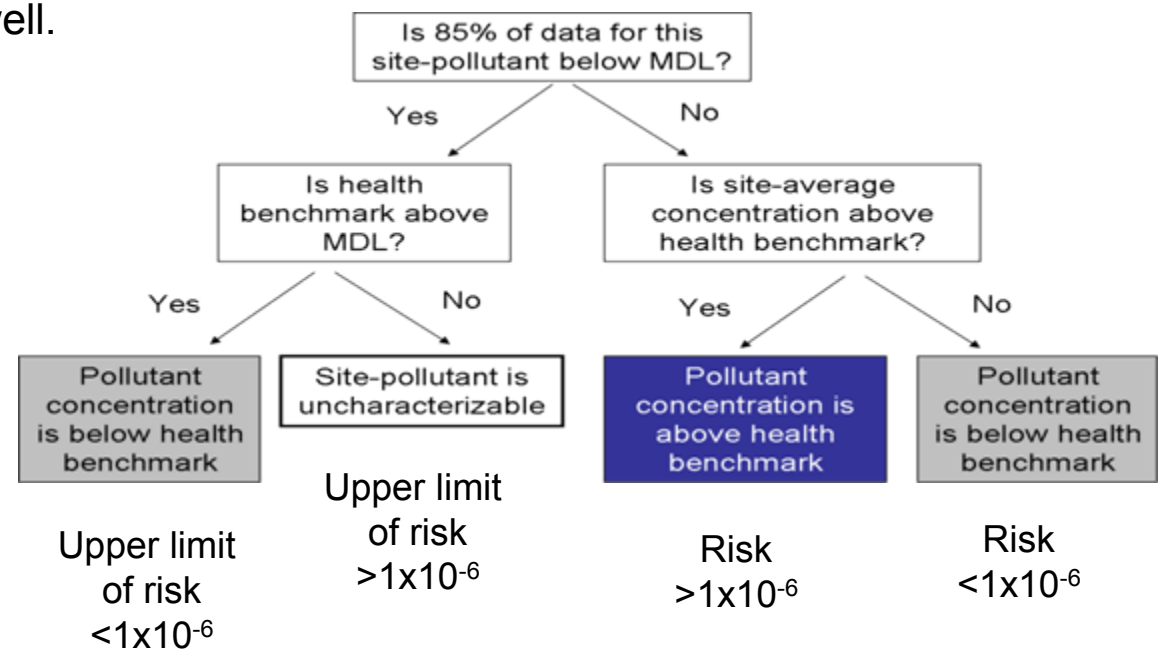
Weekday/Weekend

- The figures show weekday and weekend benzene concentrations at three study monitoring sites.
- Typically, we would expect lower MSAT concentrations on weekends, but in practice this is not always observed.
- The West Phoenix site shows higher weekend concentrations, but the difference is not statistically significant at 95% confidence. This difference may indicate that additional weekend events near the site are causing benzene emissions. For example, monitors placed near a facility with high use on weekends, such as a recreational facility, may cause this pattern. Additional investigation of the surrounding area may be warranted but was not done.
- The South Phoenix site shows slightly lower weekend concentrations (but not statistically significant). This pattern is more typical of urban sites at a national level.
- The Senior Center site shows invariant weekday/weekend patterns consistent with the well-mixed and aged nature of samples arriving at the site.



Risk Screening
Overview

- Risk screening may provide a summary of ambient concentrations of air toxics that may be of concern.
- To identify species which may indicate higher risk, follow the decision tree below for each pollutant.
- After risk species have been identified, you may wish to create risk-weighted annual averages.
- The screening here uses the 1-in-a-million cancer risk level – one could select a higher or lower risk level and define the level of concern depending on the purpose of the screening. Other health effects, such as non-cancer threshold values, could be used as well.



(ICF Consulting, 2004)

Risk Screening
West Phoenix Site

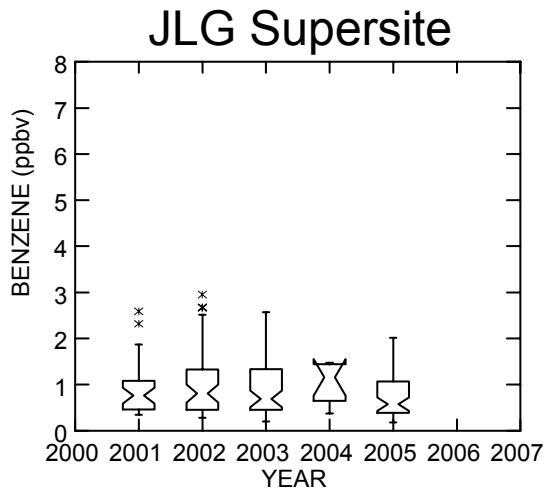
West Phoenix data necessary for risk screening

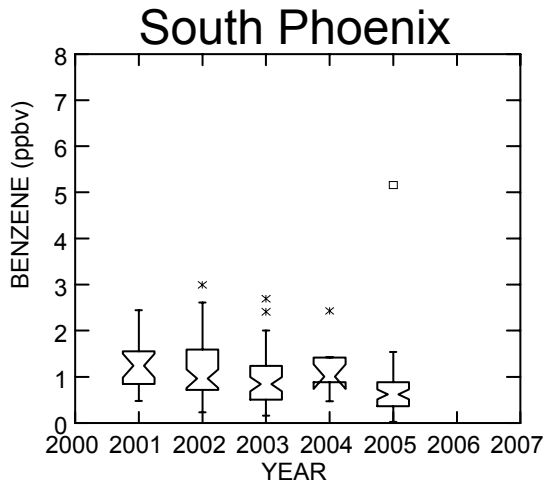
Pollutant	% Below Detection	1-in-a-million cancer risk (ppbv)	Average Method Detection Limit (ppbv)	West Phoenix Site Average Concentration (ppbv)
Benzene	0	0.040	0.50	1.7
Hexachlorobutadiene	100	0.0043	0.13	0.17

- Perform risk screening by applying all the data listed in the table to the risk-screening decision tree (see previous slide). Screening may be performed on a range of risk levels and also for non-cancer levels of concern.
- Benzene
  - More than 85% of data is above detection so there is high confidence in measured concentrations.
  - The site average concentration is above the chronic exposure concentration associated with a 1-in-a-million cancer risk.
- Hexachlorobutadiene
  - 100% of data is below detection so we have no confidence that the measured concentrations accurately reflect ambient concentrations. However, we know that concentrations are below the MDL (note that MDLs varied by sample and the average is shown).
  - The chronic exposure concentration associated with a 1-in-a-million cancer risk is below the MDL.
  - We know that both the data and the cancer risk level of 1-in-a-million are below the MDL-- improved data collection methods are necessary to more accurately characterize risk. The upper limit of risk is based on the MDL.

Trends
Five-Year Trends

- Inter-annual trends were investigated for all pollutants with sufficient data.
- The notched box plots show benzene concentrations at two sites with data available from 2001 to 2005.
- Benzene concentrations have remained relatively flat at the JLG Supersite and South Phoenix site. However, there is a statistically significant difference between the 2001 and 2005 concentrations at the South Phoenix site.
- Trends for other air toxics showed similarly consistent concentrations from year to year for this time period.
- Once six years of data are available, two 3-yr averages should be compared (i.e., average of 2001, 2002, and 2003 vs. 2004, 2005, and 2006; see *Quantifying Trends*, *S*ection 6).



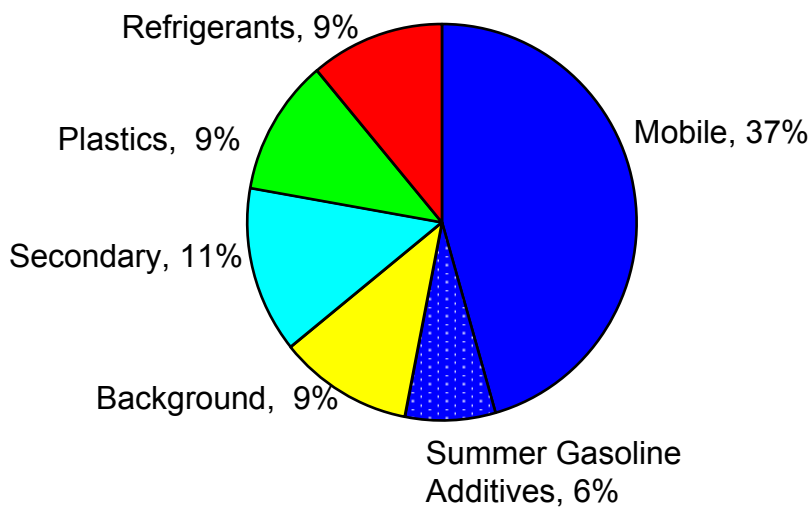


Source Apportionment

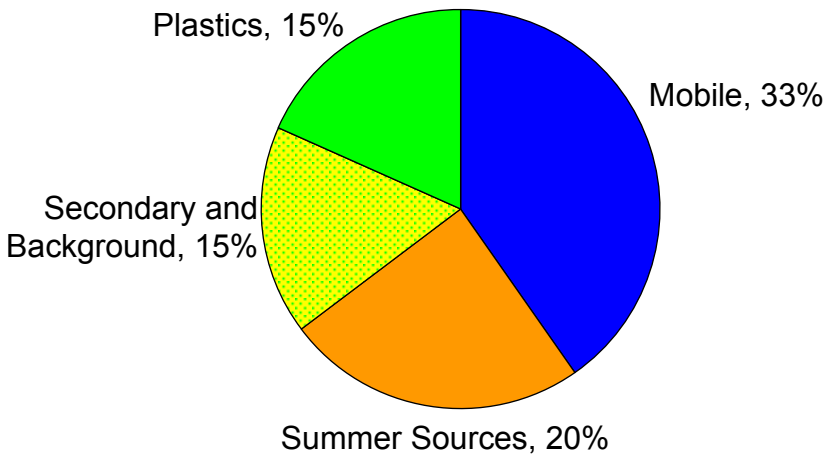
Example

- Principal component analysis (PCA) was applied to air toxics data from two sites, South Phoenix and West 43rd St., as part of an exploratory analysis. PCA uses correlation or covariance between each pair of variables to estimate relationships. PCA is relatively easy to perform with basic statistical packages; however, the analyst must infer source types from the factors.
- In South Phoenix, PCA resolved six factors, accounting for 81% of the variance. These data are illustrated in the top pie chart (note that the percentages are percent of variance explained in the data, not percent of the mass).
  - 37%: Mobile sources (benzene, 1,3-butadiene, xylenes, toluene, ethylbenzene)
  - 9%: Background (carbon tetrachloride, methyl ethyl ketone)
  - 11%: Secondary (formaldehyde, acetaldehyde)
  - 6%: Summer gasoline additives (MTBE)
  - 9%: Plastics (methylene chloride)
  - 9%: Refrigerants/AC (dichlorodifluoromethane, trichlorofluoromethane)
- PCA resolved four factors at the West 43rd Phoenix site, accounting for 82% of the variance; carbonyl compound data were not available at this site (so fewer factors were resolved).
  - 33%: Mobile sources (benzene, xylenes, toluene , ethylbenzene)
  - 20%: Summer sources, e.g., BBQs, air conditioning (trichlorofluoromethane, acetylene, propylene)
  - 14%: Secondary/background (MEK, MTBE, dichlorodifluoromethane)
  - 15%: Plastics (trimethylbenzenes)
- Next steps in this analysis may be to apply CMB or PMF to estimate source contributions.

South Phoenix



West 43rd St.



Model-to-Monitor Comparisons

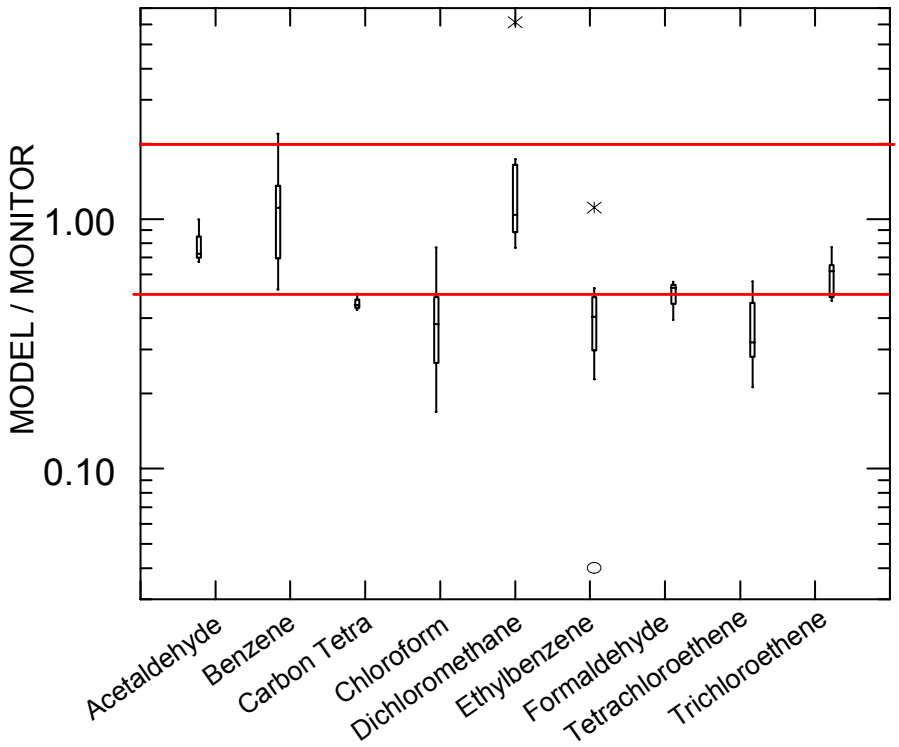
Overview

- EPA periodically performs national-scale air toxics assessment (NATA) to identify and prioritize air toxics emissions source types and locations which are of greatest potential concern in terms of contributing to population health risk. Modeled concentration estimates for 177 air toxics and DPM are provided by county. For more information on NATA see http://www.epa.gov/ttn/atw/natamain/.
- As part of an evaluation of how models used in NATA performed, EPA conducted a monitor-to-model evaluation to evaluate modeled values.
- A comparison of monitored and modeled data may help in checking the uncertainty of modeled values.

Model-to-Monitor Comparisons

Example

- The figure shows the ratio of NATA99 modeled data to annual averages computed from monitored data at the study area sites to indicate the accuracy of modeled data. This example is meant to illustrate a technique – note that the modeled and ambient data are from different years.
- When comparing modeled-to-monitored concentrations, results within a factor of 2 are considered reasonable agreement (U.S. Environmental Protection Agency, 2006b).
- Acetaldehyde, benzene, dichloromethane, and trichloroethene typically agreed within a factor of 2, consistent with national-level comparisons of modeled and monitored data.
- However, ethylbenzene, formaldehyde, carbon tetrachloride, chloroform, and tetrachloroethylene showed monitored concentrations more than a factor of 2 higher than model estimates at study area sites. There are many possible reasons for the differences. For example, the carbon tetrachloride model estimates have been shown to be low because of the use of background concentrations that were too low.



The graph shows the comparison of modeled to monitored annual averages at the study area sites. Boxes are described in *Section 4: Preparing Data for Analysis*.

Summary

What We Learned from this Data Analysis (1 of 2)

- **Data Validation – were data of sufficient quality for analysis?**
  - Overall data completeness was sufficient for analysis. (Slides 13 and 14)
  - For species data above detection were sufficient to perform most analysis, while a significant percent of some species' data were below detection. (Slide 15)
  - QA analyses showed agreement between collocated data were typical of what other studies have concluded. (Slides 10 and 11)
  - Data were validated using time series, buddy site checks, scatter plots, and fingerprint plots. Invalid data points were identified and removed. (Slides 17 to 25)
  - Data were determined to be of sufficient quality for most analyses.
- **Data Characterization – How would air toxics in the area be characterized?**
  - Air toxics concentrations in the study area were compared to national concentrations and chronic exposure concentrations associated with a 1-in-a-million cancer risk; concentrations of most air toxics are above the national median concentration at all study sites and are typically above the selected levels of risk. It is not clear why, and an evaluation/development of the air toxics emission inventory is planned (Slides 26 to 28)
  - MDLs at study sites were found to be similar across sites so that data are comparable. (Slide 29)
  - Spatial analyses showed concentrations were similar at the South and West Phoenix sites while significantly lower concentrations of MSATs at the Senior Center site were consistent with the sites' proximity to emissions. (Slide 30)

Summary

What We Learned from this Data Analysis (2 of 2)

- **Data Characterization – How would you characterize air toxics in the area? (Cont.)**
  - Temporal patterns were investigated. (Slides 31 to 33)
    - Seasonal patterns showed expected trends at the West and South Phoenix sites. Senior Center site benzene concentrations were low and showed no seasonal trend consistent with aged air impacting the site.
    - There were no significant weekend/weekday patterns, a typical result as truck traffic or weekday carryover often cause increased Saturday concentrations. There were not enough data points to reliably investigate trends by day-of-week.
  - Ambient annual average concentrations were compared to NATA 1999 modeled data. About half the species monitored at study area sites were more than two times above their modeled concentration values. Inspection of the emission inventory for the study area may be a next step. (Slides 38 and 39)
  - Risk screening was performed and the species of most concern were found to be benzene, 1,3-butadiene, acetaldehyde, carbon tetrachloride, chloroform, and tetrachloroethene. Hexachlorobutadiene may be a contributor to risk, but is not measured well enough to quantify the risk. (Slides 34 and 35)
- **Trends – Are there changes in air toxics concentrations over time?**
  - Five year trends (2001-2005) showed no significant change at the study sites (Slide 36)
- **Advanced Analyses – What are local sources of air toxics?**
  - PCA was performed for South Phoenix and West 43rd St. Mobile sources contributed to about one-third of the variance at both sites. Pollution related to plastics, background species, and secondary species contributed about another third. Both sites showed significant influence from "summer" pollutants related to BBQs, air-conditioning/refrigerants, and summer fuel additives. (Slide 37)
  - Mobile source influences were confirmed by other analyses.
    - Scatter plots showed strong correlation between mobile source air toxics. (Slide 21)
    - Spatial patterns revealed higher mobile source concentrations near busy roadways and much lower concentrations in remote areas (Slide 30)
  - Short-term solvent emissions events were identified during the process of data validation. (Slides 22 and 23)

References (1 of 2)

Arizona Department of Transportation (2005) Average Annual Daily Traffic (AADT). Available on the Internet at http://tpd.azdot.gov/datateam/aadt.php.

Brown S.G., Hafner H.R., and Shields E. (2004) Source apportionment of Detroit air toxics data with positive matrix factorization. Paper no. 41 presented at the *Air & Waste Management Association Symposium on Air Quality Measurement Methods and Technology, Research Triangle Park, NC, April 19-22* (STI-2450).

Brown S.G. and Hafner H.R. (2003) Source apportionment of Detroit pilot city air toxics data. Presented at the *National Workshop on Air Toxics Monitoring, Chicago, IL, May 13-14* (STI-902530-2371).

Brown S.G., Frankel A., and Hafner H.R. (2005) Principal component analysis and source apportionment of PAMS VOC data. Final report prepared for the South Coast Air Quality Management District, Diamond Bar, CA, by Sonoma Technology, Inc., Petaluma, CA, STI-904046-2723-FR, July.

Hafner H.R. and Brown S.G. (2005) 2005 JATAP monitoring project – gaseous air toxics data validation and analysis. Work plan prepared for the Arizona Department of Environmental Quality, Phoenix, AZ, by Sonoma Technology, Inc., Petaluma, CA, STI-905039.01-2814-WP, October.

Hafner H.R., O'Brien T.E., Frankel A.P., McCarthy M.C., and Brown S.G. (2006) 2005 JATAP monitoring project gaseous air toxics data validation and analysis. Presented at the *JATAP Workshop Meeting, Phoenix, AZ, March 6,* by Sonoma Technology, Inc., Petaluma, CA (905039.02-2921).

Hafner H.R. and O'Brien T.E. (2006) Analysis of air toxics collected as part of the Joint Air Toxics Assessment Project. Final report prepared for the Arizona Department of Environmental Quality, Phoenix, AZ, by Sonoma Technology, Inc., Petaluma, CA, STI-905039.03-3016-FR, December.

Henry R.C. (2000) Unmix Version 2 Manual. Available on the Internet at <http://www.epa.gov/ttnamti1/files/ambient/pm25/workshop/unmix2.pdf> last accessed September 9, 2005.

Hopke P.K. (2003) A guide to positive matrix factorization. Prepared for Positive Matrix Factorization Program, Potsdam, NY, by the Department of Chemistry, Clarkson University, Potsdam, NY.

ICF Consulting (2004) Air toxics risk assessment reference library, Volume 1. Prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, by ICF Consulting, Fairfax, VA, EPA-453-K-04-001A, April. Available on the Internet at <http://www.epa.gov/ttn/fera/risk_atra_main.html>.

References (2 of 2)

Lewis C.W., Norris G.A., Conner T.L., and Henry R.C. (2003) Source apportionment of Phoenix PM2.5 aerosol with the Unmix receptor model. *Journal of Air and Waste Management Association* **53** (3), 325-338.

McCarthy M.C., Brown S.G., Hafner H.R., Frankel A., and Broaders K.E. (2004) Data analyses for Phoenix, Arizona, air toxics data collected from 2001 to 2004. Final report prepared for Arizona Department of Environmental Quality, Phoenix, AZ, by Sonoma Technology, Inc., Petaluma, CA, STI-904236-2666-FR, December.

McCarthy M.C., Hafner H.R., and Montzka S.A. (2006) Background concentrations of 18 air toxics for North America. J. Air & Waste Manage. Assoc. 56, 3-11 (STI-903550-2589). Available on the Internet at http://www.awma.org/journal/ShowAbstract.asp?Year=&PaperID=1509.

Sundblom M., Armijo C., and Hafner H. (2006) Joint Air Toxics Assessment Project (JATAP) for the Maricopa/Pinal urban area, Arizona. Presentation for the *EPA National Air Monitoring Conference, Las Vegas, NV, November 8,* by the Arizona Department of Environmental Quality, the Salt River Pima Maricopa Indian Community, and Sonoma Technology, Inc., Petaluma, CA.

U.S. Environmental Protection Agency (1998) CMB8 application and validation protocol for PM2.5 and VOC. Report prepared by U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA 454/R-98-xxx, October.

U.S. Environmental Protection Agency (2005) Prioritization of Data Sources for Chronic Exposure. Available on the Internet at http://www.epa.gov/ttn/atw/toxsource/chronicpriority.html.

U.S. Environmental Protection Agency (2006) Technology Transfer Network, 1999 National-Scale Air Toxics Assessment, 1999 assessment results. Available on the Internet at <http://www.epa.gov/ttn/atw/nata1999/nsata99.html>.

U.S. Environmental Protection Agency (2006) A Preliminary risk-based screening approach for air toxics monitoring data sets. Available on the Internet at www.epa.gov/Region4/air/airtoxic/Screening-041106-KM.pdf.

U.S. Environmental Protection Agency (2006) Comparison of ASPEN Modeling System Results to Monitored Concentrations. Available on the Internet at http://www.epa.gov/ttn/atw/nata/mtom_pre.html.