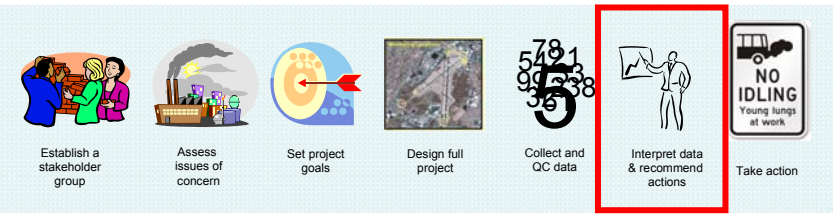
Data Analysis and Interpretation

- Spatial variability (e.g., exposure or environmental justice)
- Source characterization
- Supporting health effects assessments
- Methods evaluation
- Trends characterization
- Evaluating and improving air quality models



Data-Driven Analysis

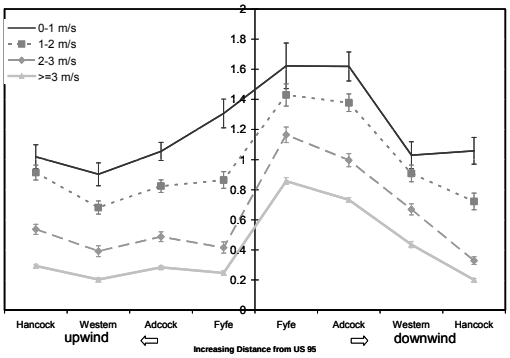
- Are data of sufficient quantity and quality to meet project objectives with statistical certainty?
  - Uncertainty – sampling, analytical, representativeness
  - $N^{0.5}$ – are there enough samples?
  - Data quality – contamination and other data issues
- Using the data
  - Evidence-based comparison to initial hypothesis
  - Accept, reject, or find inconclusive results regarding initial hypothesis

Project Objectives

- Community-scale monitoring
  - Spatial gradients near sources and environmental justice issues
  - Emissions source characterization
  - Support health effects assessments
  - Baseline concentrations for exposure assessment
  - Evaluate and improve air quality models
- Methods evaluation
  - Assess new methods for analysis of priority air toxics
  - Evaluate methods that may be operable on a routine basis to measure air toxics
- Analyze existing data
  - Same as community-scale monitoring and trends analysis

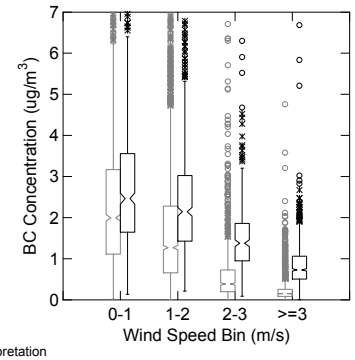
Spatial Comparisons

Means with 95% confidence intervals



US 95 MSAT Project Examples

Box plots

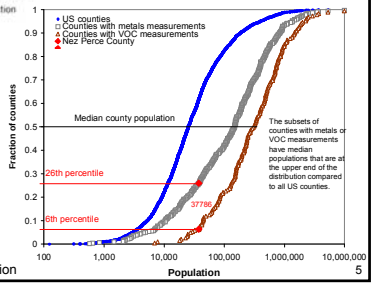


Spatial Comparisons (cont.)

Maps of average concentrations – Not quantitative

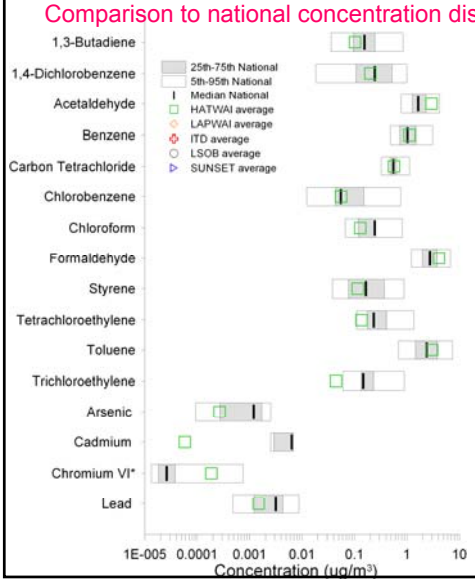
Nez Perce Example →

Cumulative distribution functions



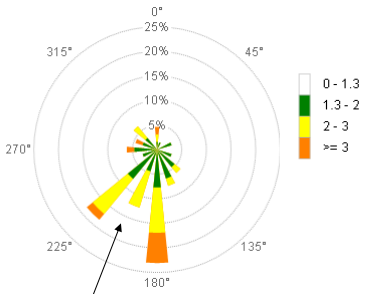
Spatial Comparisons (cont.)

Comparison to national concentration distribution



← Nez Perce Example

Pollution rose



Concentrations are highest when wind blows from south and southwest.

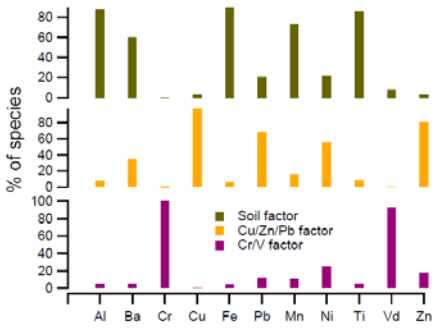
Spatial Comparisons (cont.)

- The site of interest has concentrations that
  - Are statistically significantly higher than other sites (mean, median, other metric)
  - Are higher when the wind is from a certain direction
  - Are higher than concentrations at other sites in the community, state, and/or nation
  - Are higher than expected given local population and emissions sources

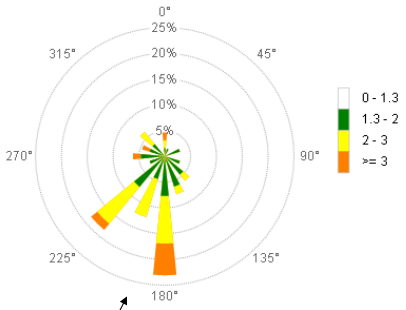
Note: Higher could be lower if you want to focus on clean sites.

Emissions Source Characterization

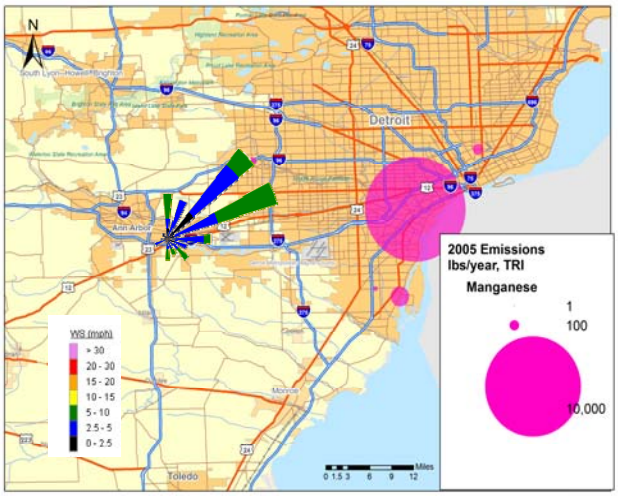
Chemical source profiles



Pollution rose



Concentrations are high when winds blow from the south



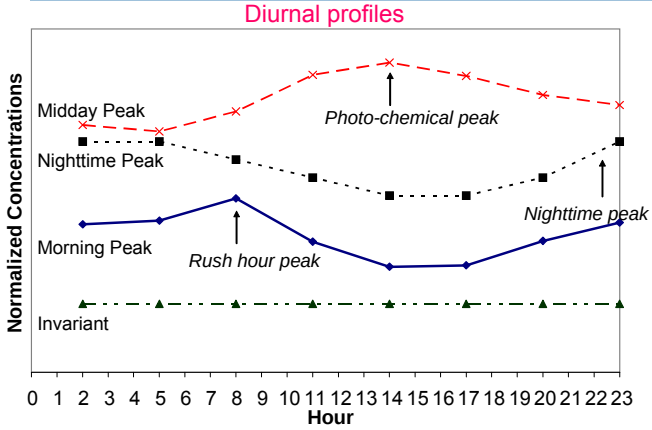
Emissions Source Characterization (cont.)

MDEQ example

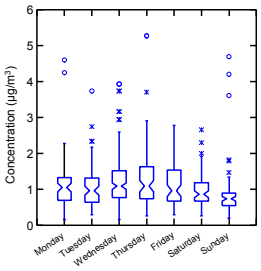
Wind roses for high Zn/Mn factor/source days at Detroit's Allen Park site point to large point sources in the industrial area of Detroit

Emissions Source Characterization (cont.)

Diurnal profiles



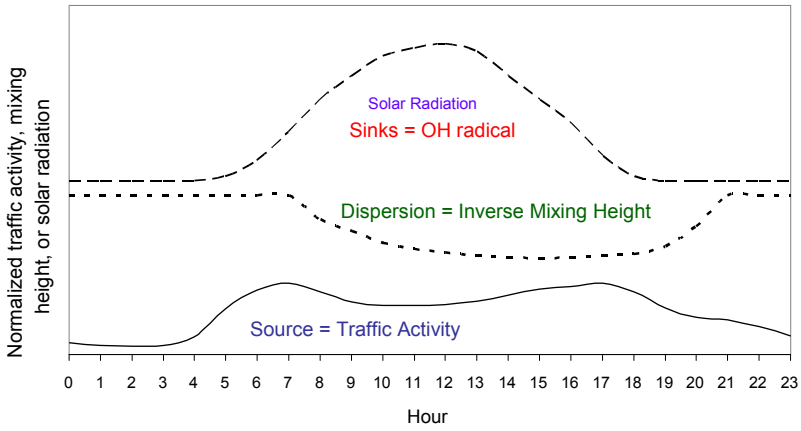
Benzene
Day-of-week profiles



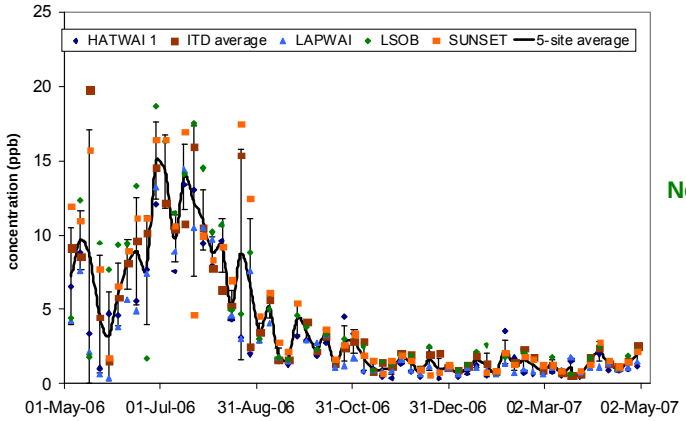
Diurnal and day-of-week profiles provide information on emissions temporal activity

Diurnal Patterns: Conceptual Model

Concentrations = (Sources – Sinks + Transport)/Dispersion



Time Series Analysis

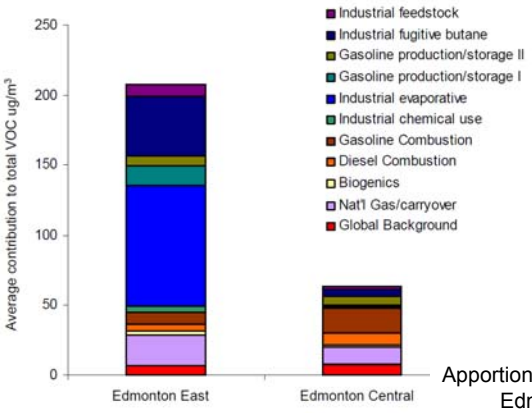


Nez Perce Example

Time series of formaldehyde concentrations (ppb) at each Lewiston area monitoring site and a five-site average. Error bars indicate the standard deviation in the five-site average.

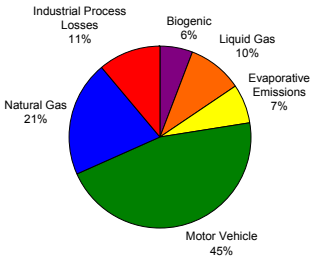
Emissions Source Characterization

Receptor Modeling



Apportionment of VOCs in Edmonton AB

Apportionment of benzene (in total VOC) at a Los Angeles site



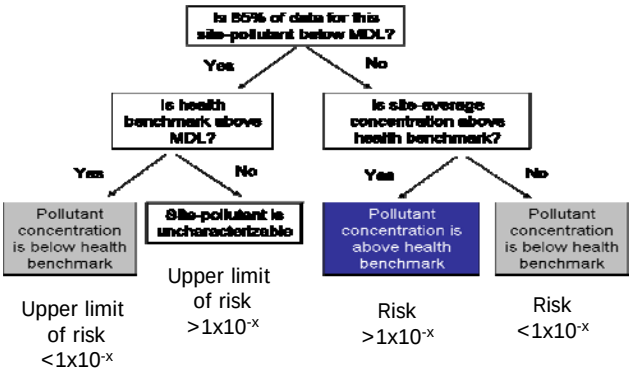
Emissions Source Characterization (cont.)

Analyses can demonstrate that

- Chemical fingerprint profiles are consistent with emissions source
- Concentrations of certain pollutants are higher when winds are from source direction
- Temporal variability is consistent with emissions activity
- Concentrations at nearby receptors are higher than at other sites
- Receptor modeling identifies and quantifies emissions source

Health Effects Assessments

Risk screening



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

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

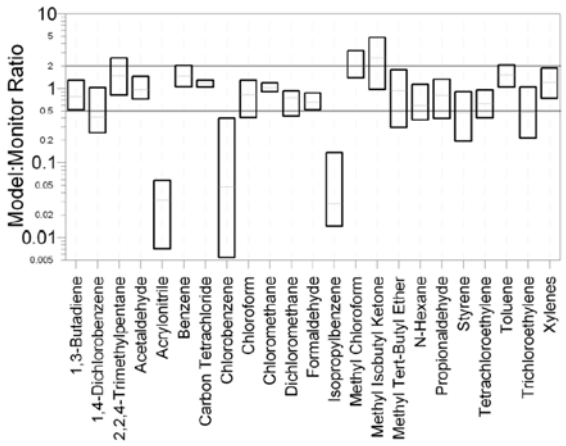
Risk $>1x10^{-x}$

Risk $<1x10^{-x}$

Where 10^{-x} is user defined

Health Effects Assessments (cont.)

Model to monitor comparison



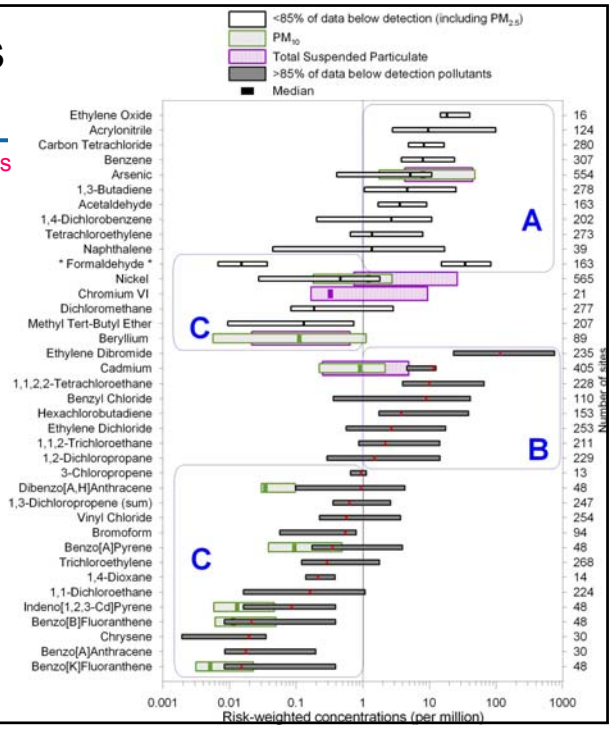
Health Effects Assessments

Risk-weighted concentrations

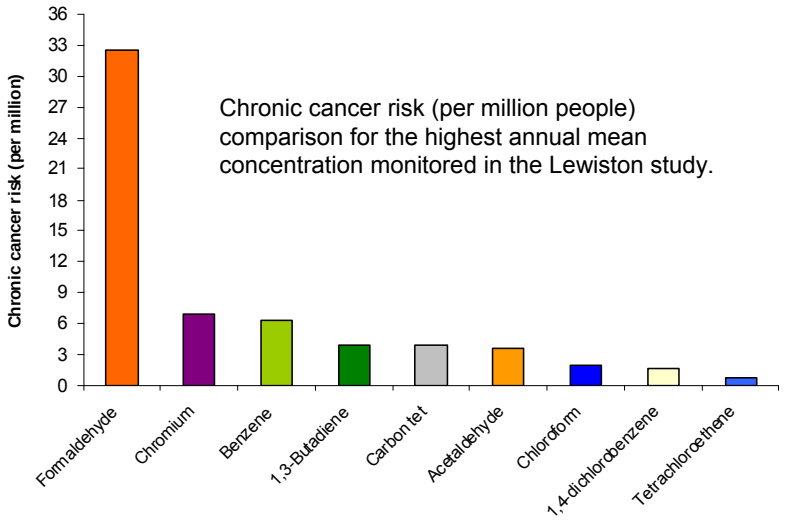
A. Pollutants with a majority of sites with risk estimates > 1-in-a-million risk level

B. Pollutants with most of the data < MDL, but detection limits above the 1-in-a-million risk level

C. Pollutants with the majority of monitoring sites reporting concentrations < the 1-in-a-million risk level including those usually above and below MDL



Health Effects Assessments (cont.)



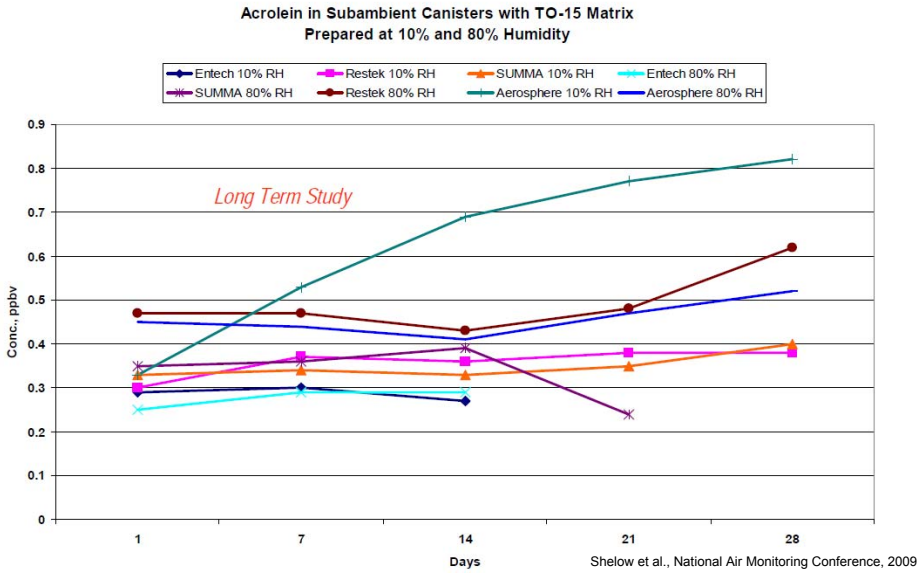
Chronic cancer risk (per million people) comparison for the highest annual mean concentration monitored in the Lewiston study.

Health Effects Assessments (cont.)

Ambient data can be used to

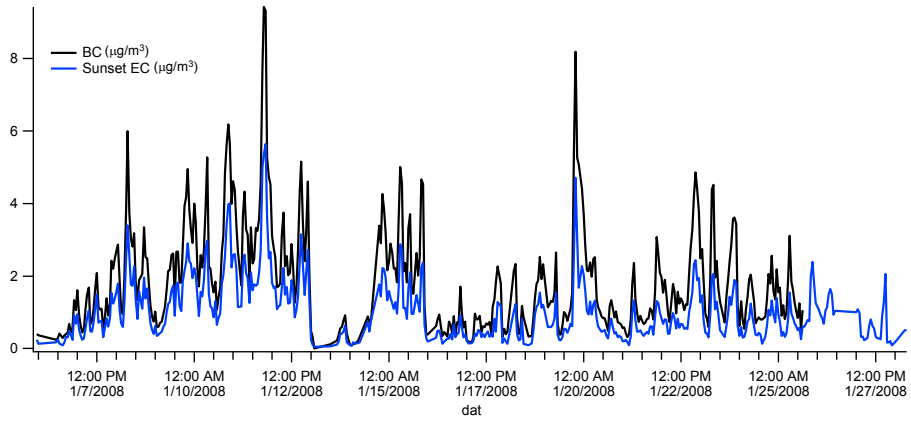
- Perform simple risk screening against levels of concern
- Calculate risk-weighted concentrations to estimate risk levels
- Validate and evaluate modeling efforts
- Identify temporal variability in concentrations for use in exposure modeling efforts

Methods Evaluation

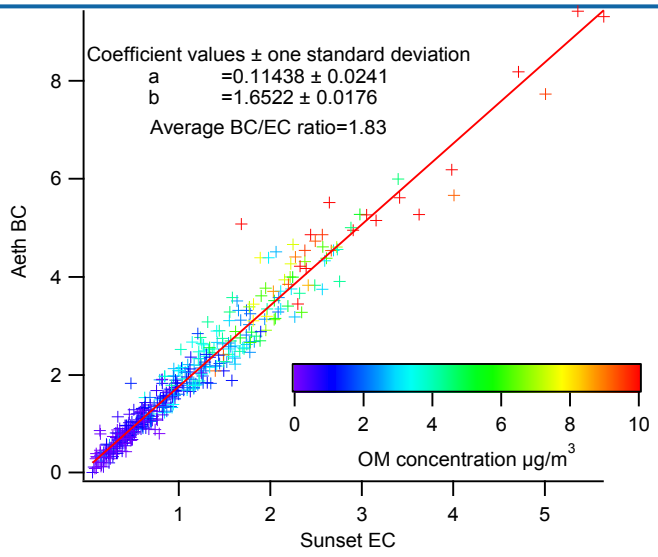


Shelow et al., National Air Monitoring Conference, 2009

Methods Evaluation (cont.)



Methods Evaluation (cont.)

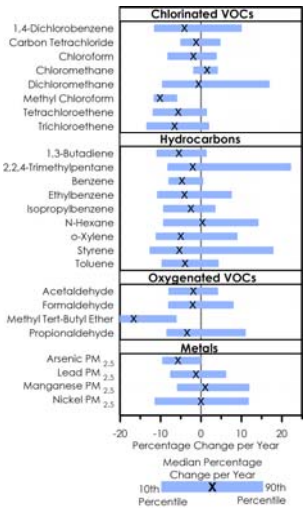


Methods Evaluation (cont.)

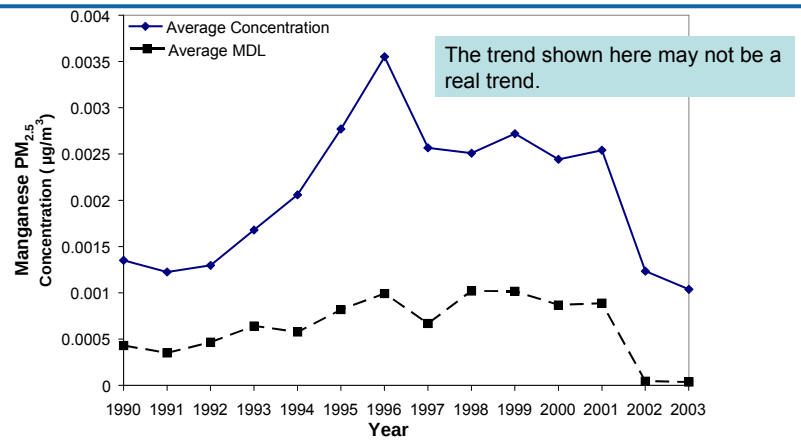
- Evaluation against standards can test accuracy and precision
- Evaluation against other existing methods can identify biases and real-world performance under ambient conditions
- Novel methods often provide surprising data that lead to better understanding of local emissions sources (e.g., local chrome facilities, solvent releases)

Trend Analyses

- Trends are useful for demonstrating progress (or lack thereof) in mitigating emissions of air toxics.
- Trend analysis can be complicated by data below MDL, changing methods, and step-changes in ambient concentrations.

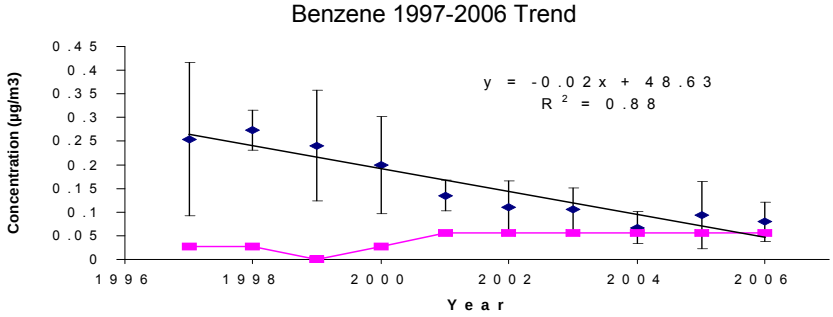


Effect of Changes in MDL on Trends Assessment



In the national-level investigation of manganese (Mn) trends, MDL trends were similar to concentration trends, making us suspicious of the reliability of the overall ambient trend. This example shows average Mn $PM_{2.5}$ concentrations and MDLs from 1990 to 2003. For this data set, Hyslop and White (2007) showed that reported MDLs are much lower than actual detection limits. Current recommendations are to be cautious with data within a factor of 6 to 10 of the reported MDL.

Effect of Changes in MDL on Trends Assessment (cont.)

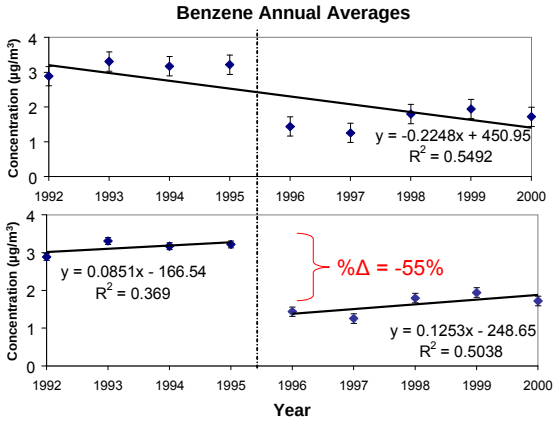


In contrast to the previous Mn $PM_{2.5}$ trend, this benzene trend does not show influence from a change in MDL (i.e., the trends in concentration and MDL show different patterns).

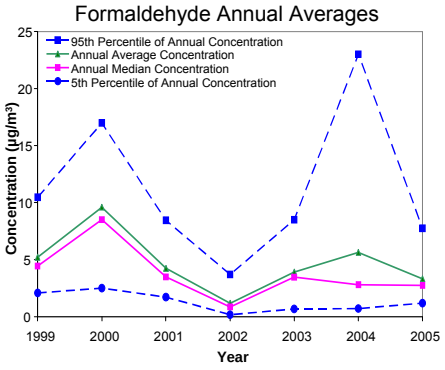
How to Quantify Trends

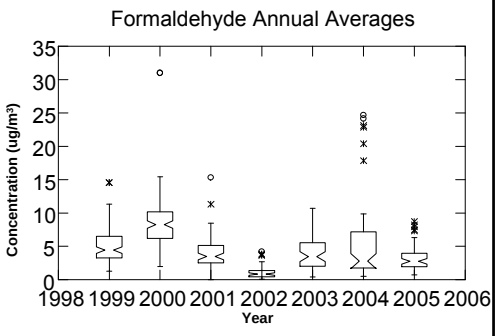
- Initial investigation of trends
  - Inspect first and last year of the trend period or two multi-year averages for change.
  - Use simple linear regression to determine the magnitude of a trend over the trend period.
- Quantifying trends
  - The percent difference between the first and last year of the trend period provides a rough sense of the change.
  - The difference between two multi-year averages provides another measure of change and helps smooth out possible influences of meteorology.
  - The percent change per year is provided by the slope of the regression line. This "normalized" value allows the analyst to compare changes across varying lengths of time (i.e., sites with different trend periods).
- Test for significance (F-test or others)

Visualizing Trends



Visualizing Trends (cont.)



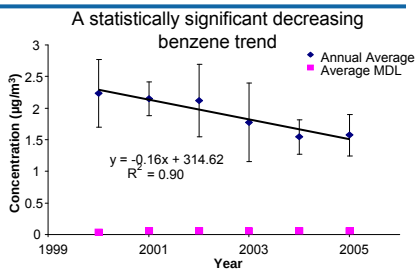


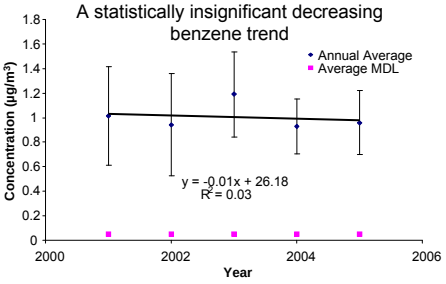
29

Summarizing Trends

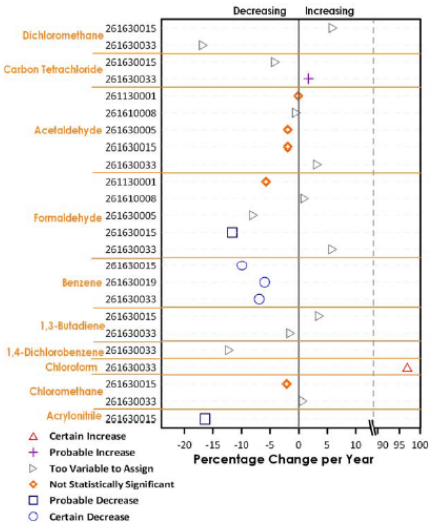
Comparing Trends

- Site-level trends for benzene from two U.S. sites.
- Confidence in these results is high. The data are mostly above detection, MDLs are consistent for the whole trend period, and no outliers appear to influence the trend.





Summarizing Trends (cont.)



Example from MDEQ trends report

Interpretation of Results

- Were data of sufficient quantity and quality to meet project objectives with statistical certainty?
  - Uncertainty – sampling, analytical, representativeness
  - $N^{0.5}$ – are there enough samples?
  - Data quality – contamination and other data issues
- Using the data to test project hypotheses
  - Emissions characterization
    - Chemical source profile comparison (ambient vs. emissions)
    - Emissions activity matches expected temporal patterns after adjusting for meteorology
    - Wind analysis to corroborate impact of emitter on monitor
    - Source apportionment

Interpretation of Results (cont.)

- Using the data to test project hypotheses
  - Health effects assessment
    - Comparison of concentrations to health benchmarks
    - Comparison of concentrations to other sites/cities/states/nation
    - Identifying pollutants above health benchmarks
  - Community baseline
    - Characterizing annual averages, seasonal variability
    - Quantifying toxics concentrations likely to be targeted by emissions reductions measures
    - Characterizing spatial variability
  - Methods evaluation
    - Is method more accurate, precise, sensitive?
    - Does it have better time resolution?
    - How much does it cost versus routine method?

Lessons Learned from Data Analysis

- Plan for data analysis in your study.
- Reserve adequate funds and time (schedule) to conduct data analysis.
- Start looking at your data early in the project as it is first collected – don't wait until the end.
- Isolating a particular source impact on pollutant concentrations is tricky (and local met data are vital).
- Many studies noted issues with differing MDLs across labs, too much data below MDL, etc.

Lessons Learned from Data Analysis (cont.)

- Analyses do not always lead to the answer you anticipated.
- Getting a similar result using different analysis approaches gives you more confidence in your results.
- Visualization of data is key (…a picture tells a thousand words).
- Show uncertainty in results to demonstrate statistical significance of findings.