Advanced Analyses

What else can I do with my air toxics data?

Advanced Analyses

What's Covered in This Section?

- This section is an overview of selected advanced data analysis techniques that may be useful in further understanding air toxics data.
- Discussion of each of these topics could fill an entire workbook; a discussion is provided of the motivation behind using these techniques and the reader is referred to available documentation for further information.
- Not all of these analyses have yet been thoroughly applied to air toxics data, but approaches that have been applied to $PM_{2.5}$ and PAMS VOC data, for example, should be applicable to air toxics data sets.
- The following topics are covered
  - Source apportionment
  - Trajectory analysis
  - Emission inventory evaluation
  - Model evaluation
  - Monitoring network assessment

Advanced Analyses

Motivation

After basic data validation and "display and describe" analyses have been performed, more can be done with the data if sufficient resources (e.g., time, expertise) are available and more sophisticated analyses are needed because basic analyses did not sufficiently answer questions.

- **Source Apportionment**. Understanding the sources impacting your monitors can be explored with source apportionment techniques and tools.
- **Trajectory Analyses**. In addition to better understanding high and low concentrations, source apportionment results can be enhanced with trajectory analyses.
- **Evaluation of Emissions Inventories and Models**. A primary goal of national monitoring networks is to compare ambient data to emission inventories and model output. These evaluations can lead to improvements in the inventories and model performance.
- **Network Assessment**. The pollution sources impacting a site, nearby demographics, and monitoring purpose can change over time. EPA's air toxics monitoring plan includes regular network assessment.

Source Apportionment
Why Perform?

- Also known as receptor modeling, source apportionment is defined as a specified mathematical procedure for identifying and quantifying the sources of ambient air pollutants at a monitoring site (the receptor) primarily on the basis of concentration measurements at that site.
- Source apportionment relates source emissions to their quantitative impact on ambient air pollution.
- Receptor models can be used to address the following questions:
  - What emissions sources contribute to ambient air toxics concentrations?
  - How much does each source type contribute?
  - Which sources could be targeted with control measures to effect the highest reduction of air toxics concentrations (or risk)?
  - What are the discrepancies between emission inventories and sources identified by receptor models?
  - Are known control strategies affecting the source contributions to air toxics?
- When performing source apportionment, the analyst should be aware of uncertainties and limitations.
  - **Many emitters have similar species composition profiles.** The practical implication of this limitation is that one may not be able to discern the difference between benzene emitted from light-duty vehicles (LDV) versus benzene from gasoline stations or refineries. One solution to this problem is to add additional species to reduce collinearity. These profiles might help to qualitatively identify mobile sources.
  - **Species composition profiles change between source and receptor.** Most source-receptor models cannot currently account for changes due to photochemistry. Since carbonyl compounds such as formaldehyde and acetaldehyde have significant secondary sources, current methods cannot link these compounds to their primary emission sources.
  - **Receptor models cannot predict the consequences of emissions reductions.** However, source-receptor models can check if control plans achieve their desired reductions using historical data.

Source Apportionment

Single-Sample and Multivariate Models

Receptor models are classified into two types: single-sample or multivariate.

- In *single-sample* models, the analysis is performed independently on each available pollutant.
  - The simplest example of this is the “tracer element” method, in which a particular property (e.g., chemical species) is known to be uniquely associated with a single source. In this case, the impact of the source on the ambient sample is estimated by dividing the measured ambient concentration of the property by the property's known abundance in the source's emissions. This method is not often available because of the difficulties of finding unique tracers or knowing their abundances. However, even if a pollutant is not uniquely associated with a source of interest, knowledge of the abundance from that source can be used to provide an upper limit for the source's impact.
  - The best-known example of single-sample receptor modeling is the chemical mass balance model (CMB). CMB eliminates the need for unique tracers of sources but still requires the abundances of the chemical components of each source (source profiles) input.
- *Multivariate* receptor models use data from multiple pollutants and extract source apportionment results from all of the sample data simultaneously.
  - The reward for the extra complexity of these models is that they attempt to estimate not only the source contributions (i.e., mass from each source) but also the source compositions (i.e., profiles).
  - There are several tools to perform multivariate receptor modeling described in the literature; EPA has supported the development of two modeling platforms: Unmix and positive matrix factorization (PMF). These models are based on factor analysis, or the closely related principal component analysis.
  - Factor analysis is a statistical method used to describe variability among observed variables in terms of fewer unobserved variables called factors.
  - There is extensive literature available describing CMB and PMF applications to speciated PM data, less available literature describing applications to VOC data, and very little research on air toxics specifically.

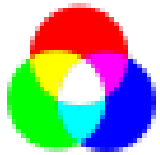
Source Apportionment

Positive Matrix Factorization

- PMF was originally developed by Paatero (1994, 1997) with additional development by Hopke et al. (1991, 2003). PMF can be used to determine source profiles based on the ambient data and associated uncertainties.
- PMF has been applied to many data sets to determine sources of $PM_{2.5}$, ozone precursors, and air toxics.
- PMF uses weighted least squares fits for data that are normally distributed and maximum likelihood estimates for data that are log normally distributed. Concentrations are weighted by their analytical uncertainties.
- PMF constrains factor loadings and factor scores to nonnegative values and thereby minimizes the ambiguity caused by rotating factors.
- Model input includes ambient monitoring data and associated analytical uncertainties (see Wade et al., 2007). A large (species and sample matrix) ambient data set is required.
- Model output includes
  - Factor loadings expressed in mass units which allows them to be used directly as source signatures.
  - Uncertainties in factor loadings and factor scores which makes the loadings and scores easier to use in quantitative procedures such as chemical mass balance.
- A free, standalone version of PMF was created by the EPA in 2005, available on the Internet at http://www.epa.gov/scram001/receptorindex.htm. Updates are underway.
- Data preparation and the interpretation of model diagnostics is covered in EPA's Multivariate Receptor Modeling Workbook (Brown et al., 2007b).

Source Apportionment

Unmix



- Unmix was developed by Ron Henry (1997) using a generalization of the self-modeling curve resolution method developed in the chemometric community.
- It originally used MATLAB computation routines. The EPA, along with Ron Henry, developed EPA Unmix and documentation that uses MATLAB features but is now a standalone model (i.e., MATLAB not needed).
- Unmix is a multivariate receptor modeling package that inputs ambient monitoring data and seeks to find the composition and contributions of influencing sources or source types. UNMIX also produces estimates of the uncertainties in the source compositions.
- Unmix requires many samples to extract potential sources, similar to PMF.
- It assumes that sources have unique species ratios, i.e., “edges” that can be observed in a scatter plot between species; uses these edges to constrain the results and identify factors; and does not need to weigh data points.
- Model input includes ambient monitoring data; uncertainty information and source profiles are not necessary.
- Model output includes source profiles with uncertainties.
- Unmix is available at http://www.epa.gov/scram001/receptorindex.htm.
- Data preparation and the interpretation of model diagnostics is covered in EPA’s Multivariate Receptor Modeling Workbook (Brown et al., 2007b).

Source Apportionment

Chemical Mass Balance

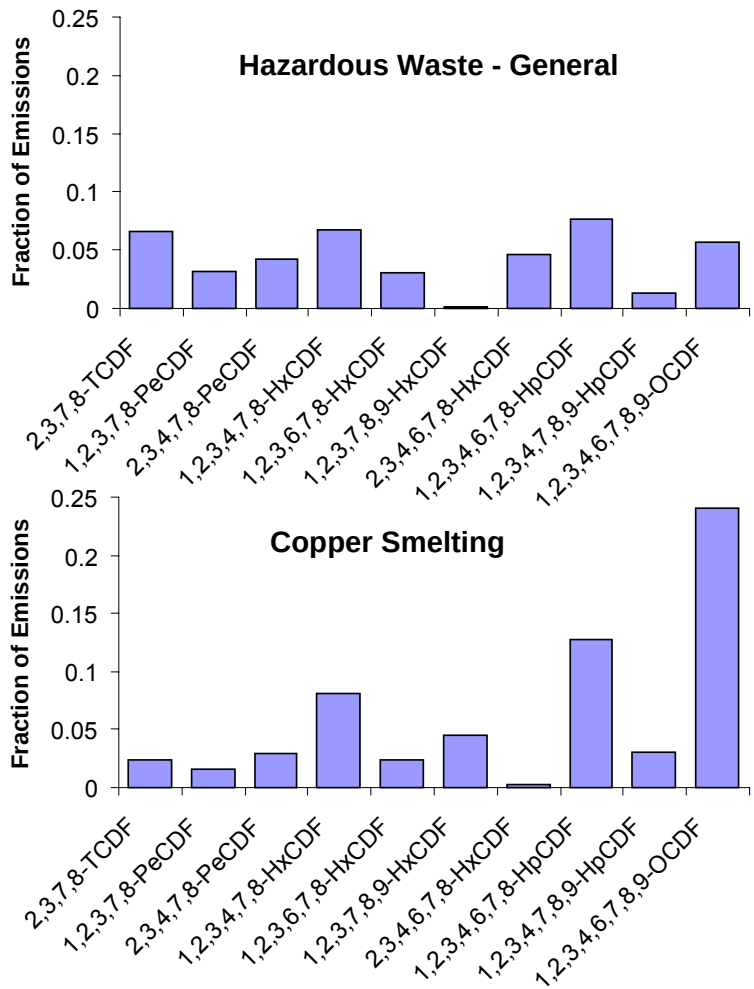
- The premise of chemical mass balance (CMB) is that source profiles from various classes of sources are different enough that their contributions can be identified by measuring concentrations of many species collected at the receptor site.
- To apportion sources, CMB uses an effective variance-weighted, least squares solution to a set of linear equations which expresses each receptor species concentration as a linear sum of the products of the source profiles and source contributions. This method can be applied to a single sample.
- Model input includes
  - Source profile species (fractional amount of species in emissions from each source type).
  - Receptor (ambient) concentrations.
  - Realistic uncertainties for source and receptor values. Input uncertainty is used to weigh the relative importance of input data to model solutions and to estimate uncertainty of the source contributions.
- Model output includes contributions from each source type and species to the total ambient concentration along with uncertainty.
- CMB has been used in a number of air pollution studies that examine particulate and VOC source apportionment, but few, if any, specific air toxics studies.
- CMB is available from EPA at http://www.epa.gov/scram001/receptorindex.htm.

Source Apportionment

Source Profiles

Accurate source profiles are the key to successful modeling.

- Source profiles provide information about the relative contribution of pollutants to emissions from a given source.
- Understanding source profiles is important because receptor modeling tools typically output source profile information that needs to be interpreted or requires user-input source profiles as a starting point for analysis.
- The figures to the right show example polychlorinated dibenzofuran (PBDF) source profiles for hazardous waste incinerators and copper smelting compiled by the EPA. Though the same compounds are present, the relative abundances are not the same, providing a mechanism for source identification.
- For CMB applications and for interpretation of PMF output, it is important to use source profiles that are representative of the study area during the period when ambient data were collected.
- In CMB, try available source profiles in sensitivity tests to determine the best ones for use (i.e., minimize collinearity).
- Source profiles can be obtained from
  - EPA SPECIATE, recently updated (version 4.0) and available at http://www.epa.gov/ttn/chief/software/speciate/index.html.
  - Literature review
    Source measurements made in your region during the period for which ambient data are available.
  - Local, state, and federal agencies.
  - Source profiles can also be procured via analysis of ambient data using tools such as PMF and UNMIX.



Source Apportionment
Approach

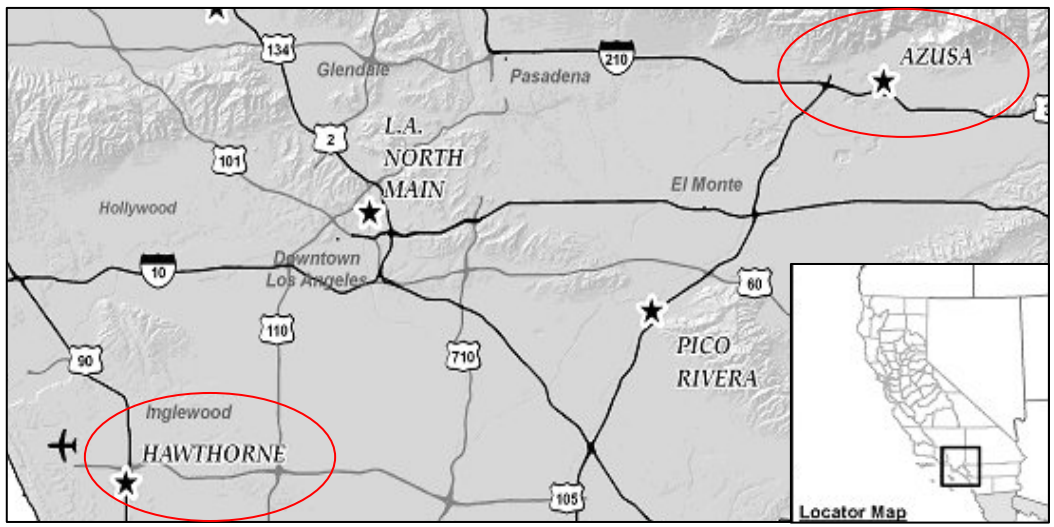
- Before beginning source apportionment, it is important to "know the data" in order to identify and assess the receptor model outputs. Understanding the data will be achieved in the process of data validation and analysis.
  - Understand airshed geography and topography using maps, photographs, site visits, etc.
  - Investigate the composition and location of emission sources.
  - Understand the typical meteorology of the site, including diurnal and seasonal variations.
  - Investigate the spatial and temporal characteristics of the data, including meteorological dependence.
  - Investigate the relationships among species using scatter plot matrices, correlation matrices, and other statistical tools.
- Apply cluster and factor analysis techniques using standard statistical packages to get an overall understanding of pollutant relationships and groupings by season, time of day, etc.
- If there are sufficient samples (e.g., more than two years of 1-in-6 day samples for more than 20 species and more than 50% of data above detection), Unmix and/or PMF may be applied to obtain "source" profiles with more species and further investigate data relationships.
- If samples are few and source profiles are available, CMB may be applied to obtain source contribution estimates.
- Compare source contributions estimates and source profiles from Unmix and PMF to the emission inventory.

Source Apportionment

Example

- PMF receptor modeling was performed for speciated VOC data collected at two PAMS sites, Hawthorne and Azusa, in the Los Angeles area during the summers of 2001-2003.
- Both toxic and non-toxic VOCs were investigated in order to provide as much data as possible for apportionment (Brown et al., 2007a).
- Air toxics included in the analysis were typically grouped as MSATs, though they have industrial sources as well.
- Data were collected as part of the PAMS network providing the advantage of subdaily data and speciated-versus-total mass measurements (total non-methane organic compounds, TNMOC).
- Uncertainty estimates were enhanced from the original analytical uncertainties by reducing the weighting of data below detection and missing data. Uncertainties for missing data were estimated with 4 times the median concentration, data below detection were given uncertainties of 1.5*MDL, and all other data were given the analytical uncertainty plus 2/3*MDL.

Site Map



Source Apportionment

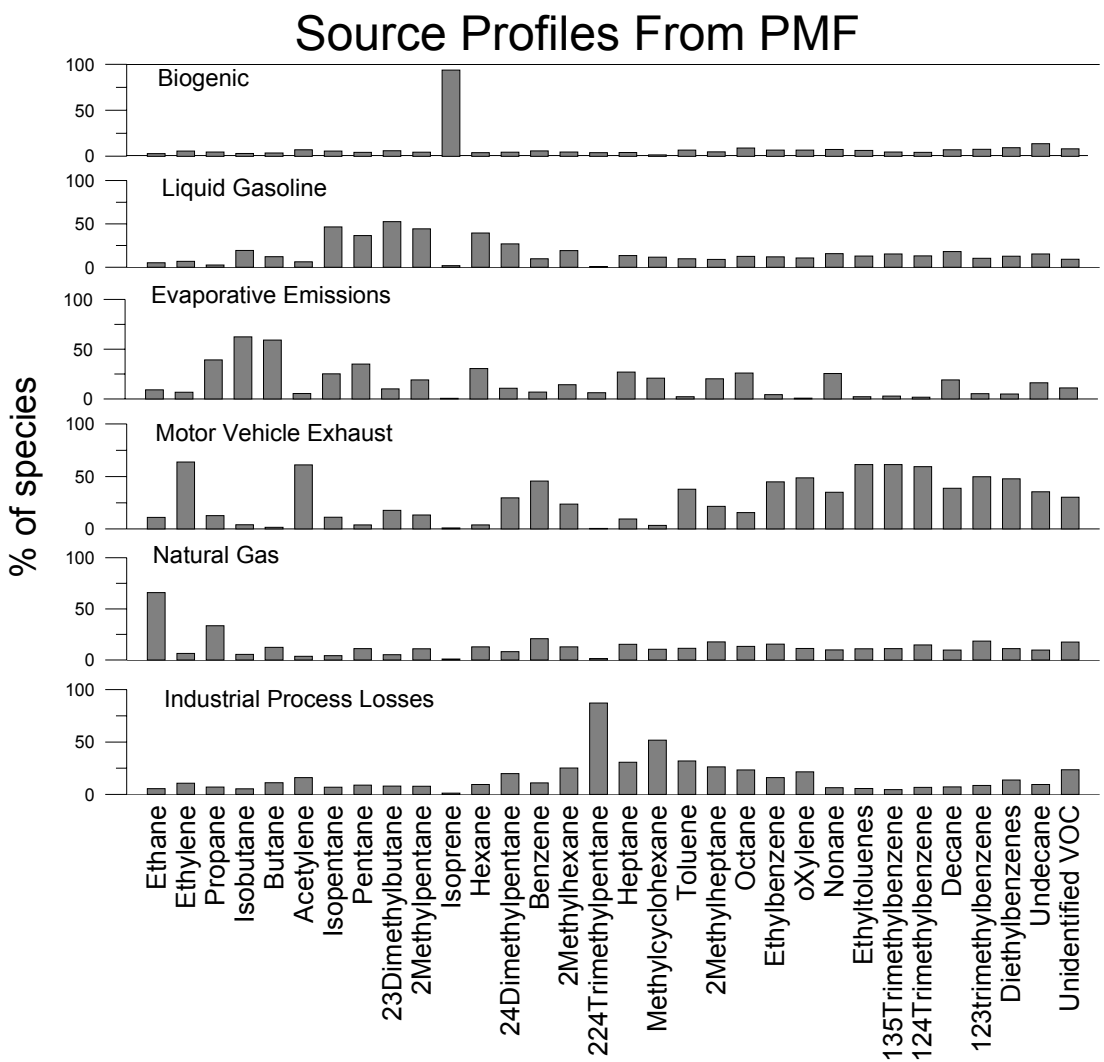
Example Preliminary Analyses

- Preliminary data analyses were performed including investigation into data quality, local emissions, species relationships, temporal patterns, etc.
- Findings
  - VOC concentrations were typically higher at Azusa compared to Hawthorne, a result consistent with site locations relative to the ocean.
  - The Azusa air mass was more aged, as indicated by loss of reactive species (except during rush hour); this is also consistent with the sites' locations in the air basin.
  - The Hawthorne site seemed to have constant, fresh emissions, with little change in the relative abundance of VOCs throughout the day, consistent with nearby industrial emissions.
  - Both sites are significantly influenced by mobile sources.

Source Apportionment

Example Hawthorne Site PMF Profiles

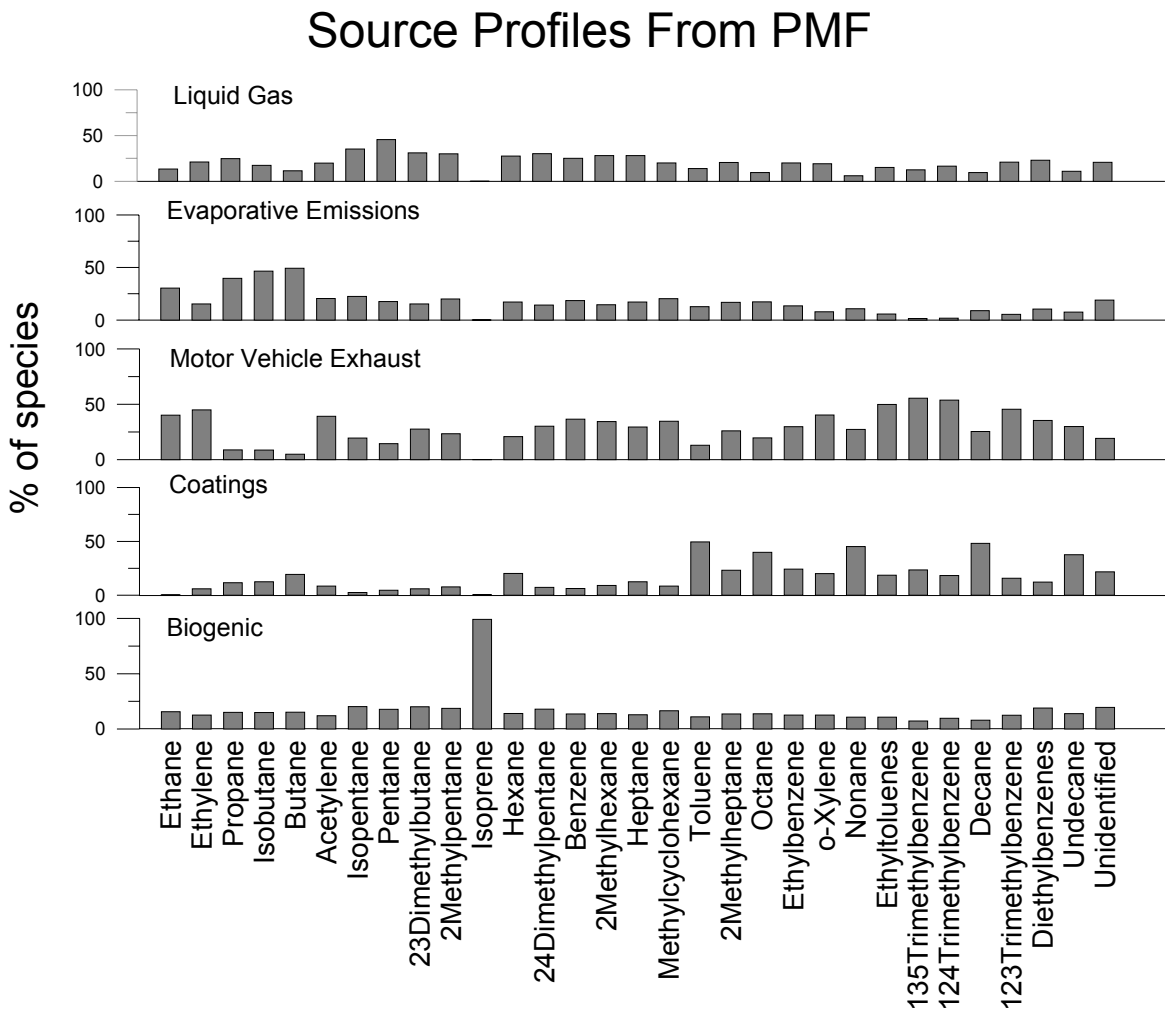
- Six factors were identified by PMF at the Hawthorne site following protocols discussed in the Multivariate Workbook (Brown and Hafner, 2005). The relative percent of species mass attributed to each profile is shown.
- Profile names indicate analyst-identified source types.
- Some of the rationale for source identification
  - Biogenic. Isoprene is the only marker for biogenic sources measured in this data set and anthropogenic sources of isoprene are insignificant; temporal patterns match expectations.
  - Liquid Gasoline. Abundance of C5 alkanes agrees with previous work; temporal patterns are consistent with mobile sources.
  - Evaporative Emissions. C3-C6 alkanes and temporal patterns are similar to diurnal temperature patterns.
  - Motor Vehicle Exhaust. Typical exhaust profile and temporal patterns are consistent with rush-hour traffic.
  - Natural Gas. Natural gas is mostly ethane and propane. These are also long-lived species that accumulate in the atmosphere.
  - Industrial Process. Losses. Consistent with nearby industrial emissions.



Source Apportionment

Example Azusa Site PMF Profiles

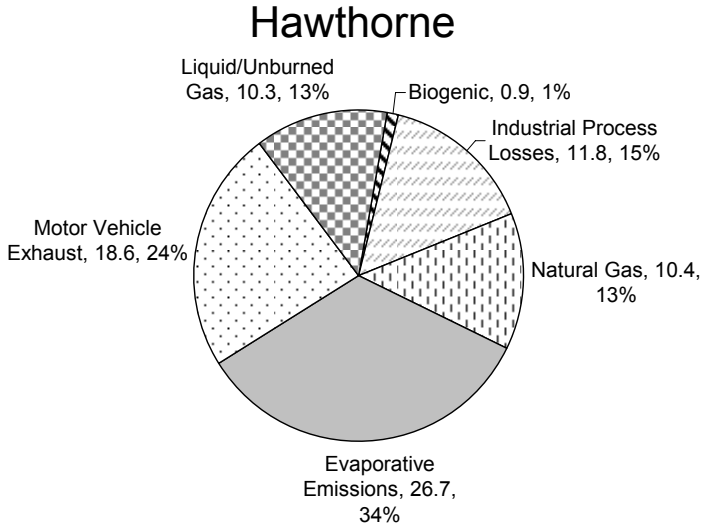
- Five factors were identified by PMF at the Azusa site. The relative percent of species mass is shown.
- Apportionment of these profiles to specific sources was performed by the analyst based on knowledge of source profiles and other investigations into the data.
- Some of the rationale for source identification
  - Coatings. Presence of C9-C11 alkanes is consistent with previous results; temporal pattern showed a daytime peak consistent with industrial operations.
  - Other profiles are similar to those observed at the Hawthorne site.

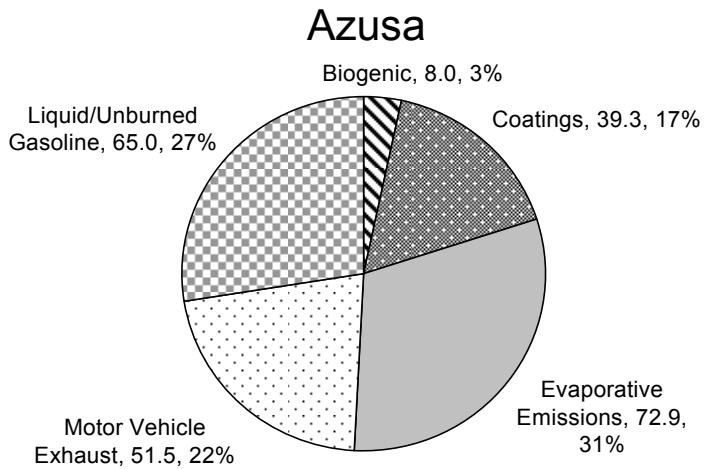


Source Apportionment

Example Percent of Total Mass

- The profiles in the previous slides indicate the relative fraction of VOCs within a profile.
- The pie charts to the right show the importance of each source profile by quantifying the amount of TNMOC mass represented by each profile. For example, in Hawthorne, evaporative emissions accounted for 34% of TNMOC mass during the summers of 2001-2003.
- Mobile source emissions are dominant contributors to TNMOC at both Hawthorne and Azusa with 71% and 80% of total mass, respectively (sum of liquid/unburned gasoline, motor vehicle exhaust, and evaporative emissions).
- The remaining VOC mass is attributed to coatings at the Azusa site and is split between industrial processes and natural gas at the Hawthorne site.

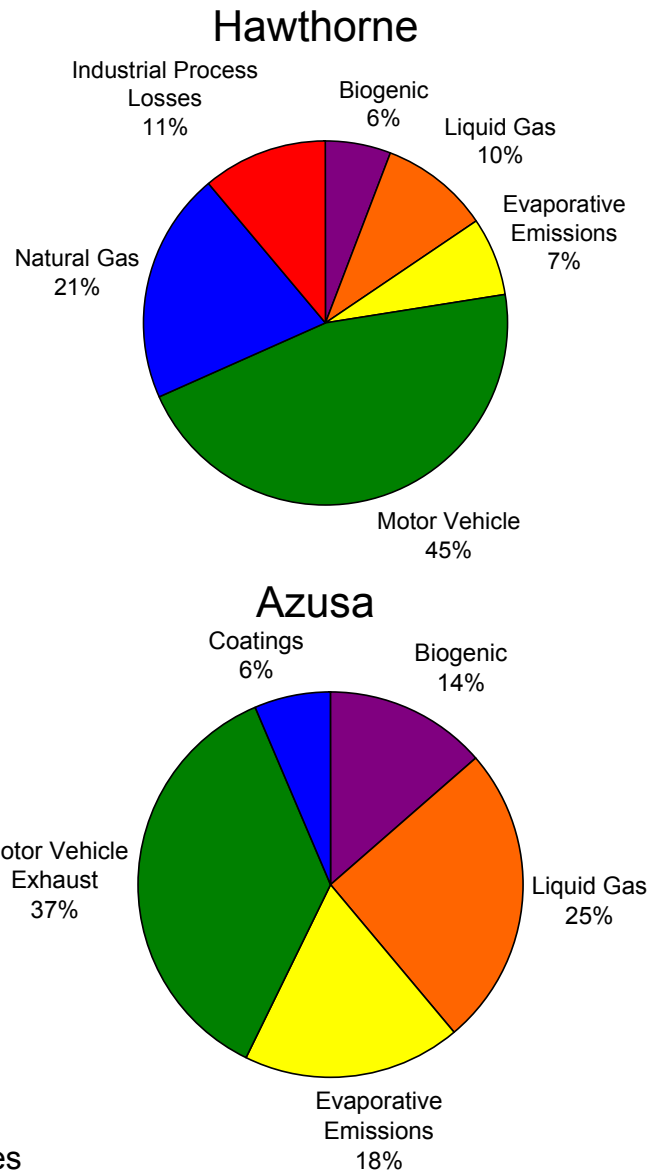




Source Apportionment

Example Apportionment of Benzene

- Apportionment of individual species between profiles can also provide interesting analyses.
- For example, benzene is a significant cancer risk driver at most sites in the United States. Source apportionment of benzene can help policy makers develop effective control regulations.
- The figures to the right show the percentage of benzene (by mass) attributed to each source profile identified by PMF at the Hawthorne and Azusa sites.
- As expected, both sites show a significant percentage of benzene mass attributed to mobile sources and gasoline evaporation. Interestingly, almost one-fourth of the benzene at the Hawthorne site is attributed to natural gas. Benzene is not emitted in natural gas (but may be emitted from combustion of natural gas); however, a significant fraction of ambient benzene is associated with air parcels containing ethane and propane (key components of natural gas). Since benzene is relatively long-lived, it is possible that benzene in this profile represents urban background. The same observation can be made for the benzene in the biogenic profile—biogenic benzene emissions are very small.



Source Apportionment
Summary

Source apportionment steps

- Review data quality and spatial/temporal characteristics.
- Prepare data for source apportionment.
  - Processing the necessary data differs among the tools, but typically the analyst needs to select pollutants with sufficient data above detection and understand/quantify uncertainty for each concentration. Guidance is provided in the EPA's Multivariate Receptor Modeling workbook (Brown et al., 2007b).
- Understand the air shed by assessing likely emissions sources and local meteorology. This helps set expectations for what the source apportionment results should show.
- With guidance from literature and workbooks, apply source apportionment tools. This is an iterative process!
- Evaluate results for reasonableness.
- Compare results to emission inventories.

With respect to toxics data, PMF and Unmix have been applied to a range of data sets while CMB applications have largely been focused on PM data.

Trajectory Analysis

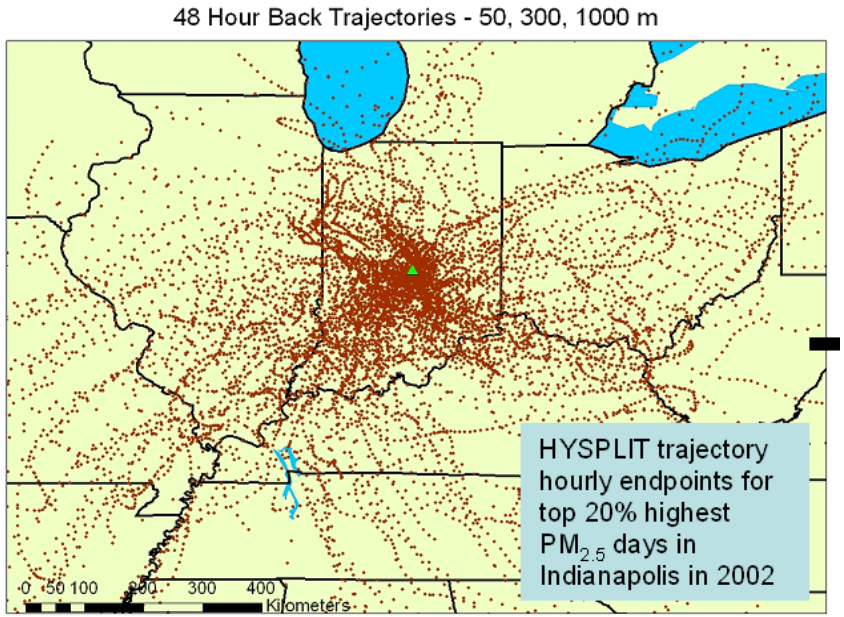
Introduction

- Trajectory analysis uses knowledge of air mass movement to trace the most likely areas of influence on high pollutant concentrations.
- The use of trajectory analysis after source apportionment helps analysts better understand, interpret, and verify source apportionment results.
- Analysis techniques
  - Backward trajectories
  - Trajectory densities
  - Potential Source Contribution Function (PSCF)
  - Conditional Probability Function (CPF)

Trajectory Analysis

Backward Trajectories

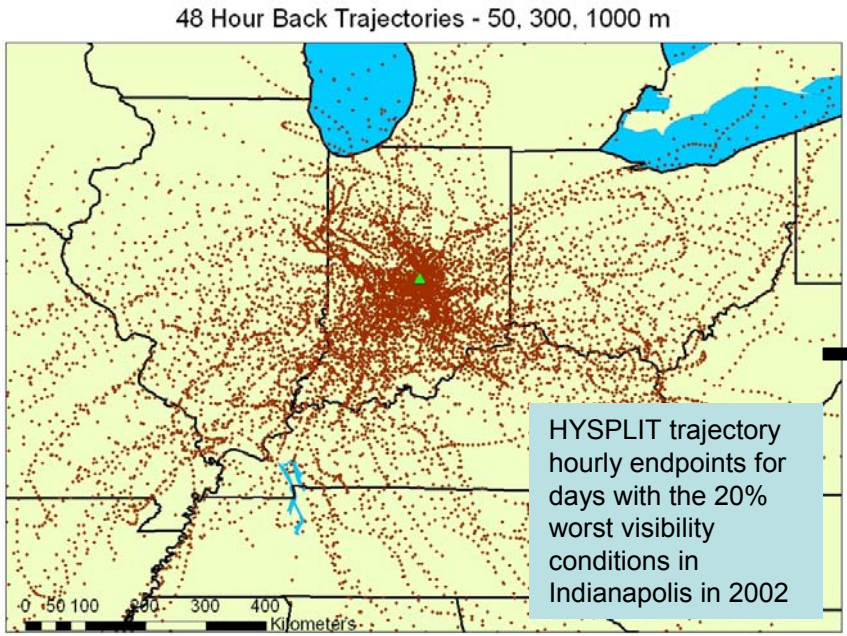
- Backward air mass trajectories estimate where air parcels were during previous hours.
- Air mass trajectories can be employed to investigate long-term, synoptic-scale meteorological conditions associated with high concentrations of individual factors.
- Estimates grow less certain as time elapses.
- The NOAA HYSPLIT model is one means to run trajectories. It is available at http://www.arl.noaa.gov/ready/hysplit4.html

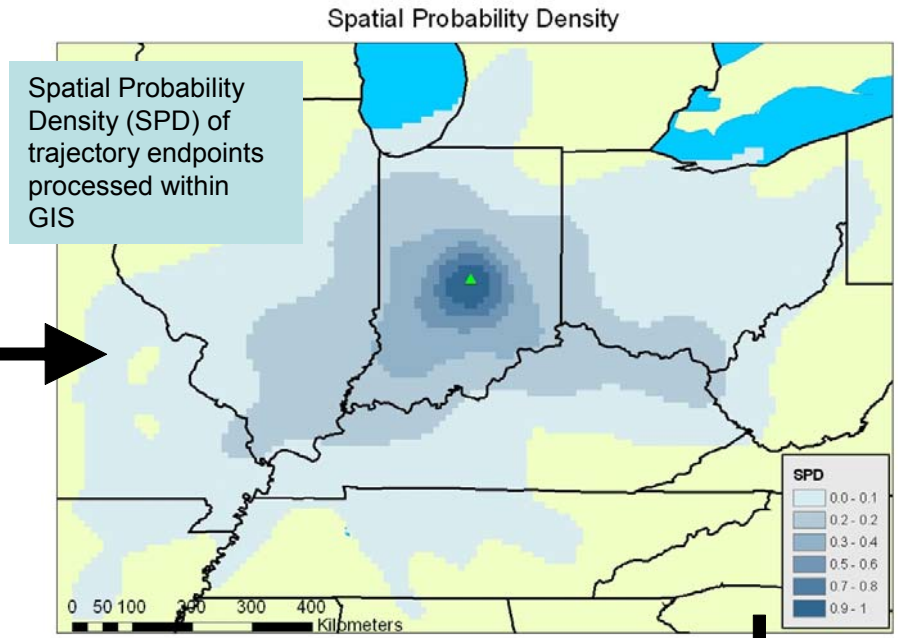


Trajectories are often plotted as single points for every hour backwards from the start point as shown here (also called a spaghetti plot). However, they should not be viewed as specific points, but rather as a small area around that point and with the last and next point.

Trajectory Analysis

Trajectory Densities





Trajectories are often processed into density, rather than “spaghetti”, plots. Higher density corresponds to more trajectories passing through that grid square. This plotting enables a number of useful analysis techniques, such as Potential Source Contribution Function (PSCF) analysis.

Trajectory Analysis

Potential Source Contribution Function (PSCF)

PSCF uses HYSPLIT backward trajectories to determine probable locations of emission sources.

$$PSCF = \frac{m_{ij}}{n_{ij}}$$

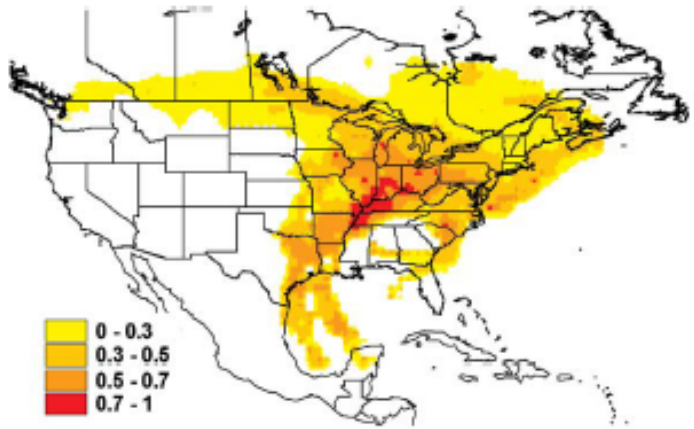
n_{ij} = number of times trajectory passed through cell (i,j).

m_{ij} = number of times source contribution peaked while trajectory passed through cell (i,j).

Top 10%-20% source contributions are used for m_{ij}.

In the example on the right, all five-day backward trajectories, for every two hours were applied to the corresponding 24-hr source contributions.

PSCF calculated for each cell sized 1°×1° and results displayed in the form of maps on which PSCF values ranging from 0 to 1 are displayed in a color scale.



PSCF function plot for sulfate affecting Philadelphia. Higher probability is associated with an area of high SO_2 emissions. Computations and graphics are made using ArcMap or other GIS tool.

(Source: Begum et al., 2005)

Trajectory Analysis

Conditional Probability Function (CPF)

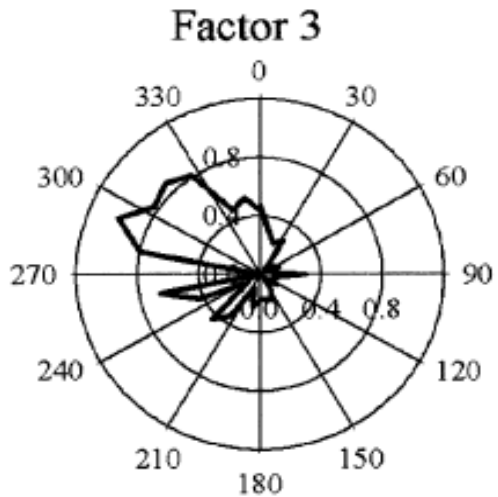
CPF uses wind direction, rather than trajectories, to determine the likely direction of sources. CPF compares days when concentrations were highest to the average transport pattern (i.e., the climatology).

$$CPF = \frac{m_{\Delta\Theta}}{n_{\Delta\Theta}}$$

$n_{\Delta\Theta}$= number of times wind direction is from sector $\Delta\Theta$.

$m_{\Delta\Theta}$= number of times source contributions are high while wind direction was from sector $\Delta\Theta$.

A CPF value close to 1.0 for a given sector ($\Delta\Theta$) indicates a high probability that a source is located in that direction.



Example CPF plot for the highest 25% contribution from a PMF factor pointing to the northwest of site as a possible source region. Computations can be programmed into Microsoft Excel or other statistical packages.

(Source: Kim et al., 2004)

Trajectory Analysis
Interpretation

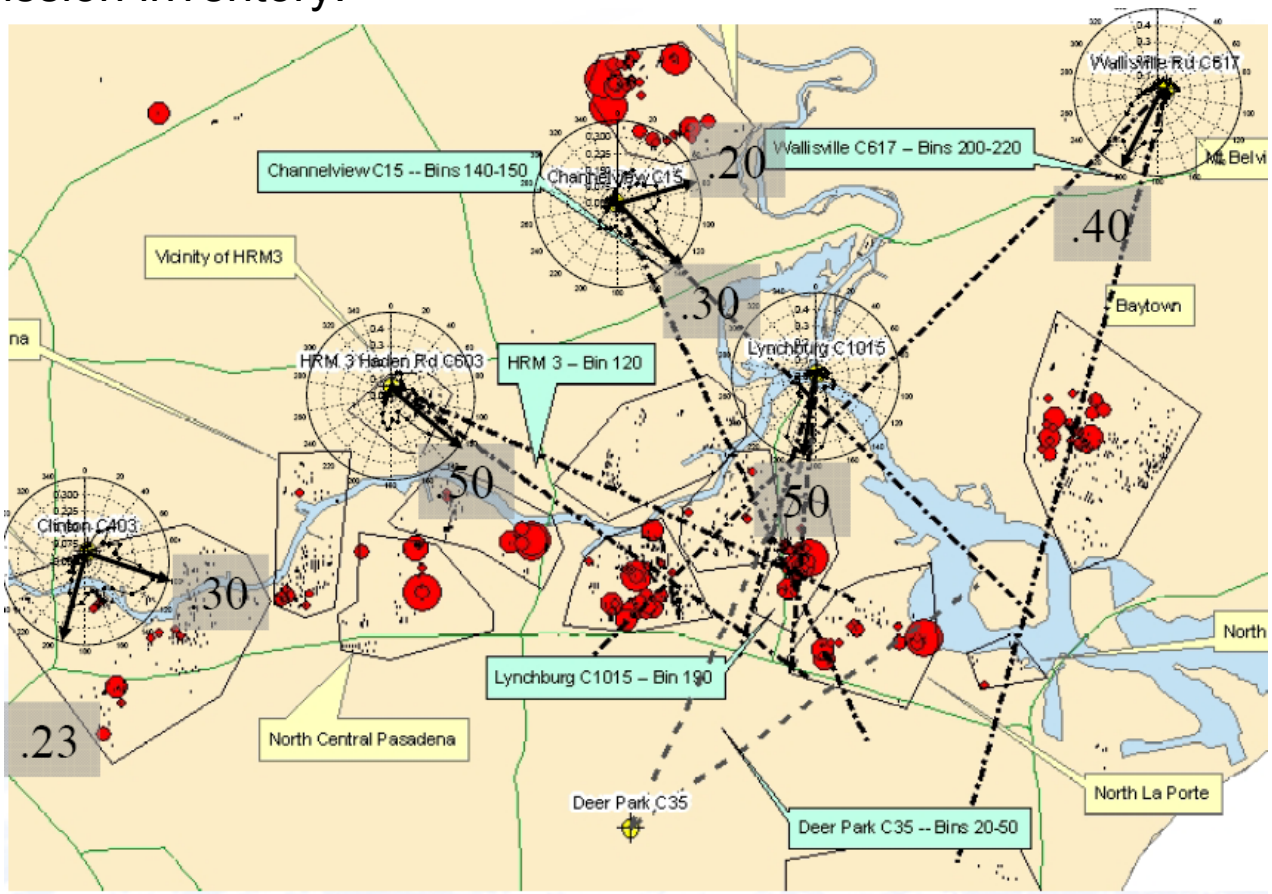
- No matter which trajectory analysis is used, interpretation of results is similar. These methods are all complementary to source apportionment or can be standalone to assess source regions. No one method shown is superior.
  - To investigate a number of days, ensemble methods are preferred (such as trajectory densities). These methods help identify source areas.
  - CPF also requires a number of days to be included, but helps point toward a particular direction.
  - Single trajectories are useful when investigating an individual sample.
- The following questions may be investigated for verification of results:
  - Do results meet the conceptual model of emissions and removal of air toxics?
  - Are these the areas from which emissions influence would be expected?
  - Does the transport pattern make sense with respect to the age/chemistry of a given factor (i.e., more transport and chemistry are associated with secondary pollutants such as formaldehyde)?

Trajectory Analysis

Using CPF Results

This approach is based on the assumption that wind direction and trajectory analysis results should be consistent with the spatial distribution of the sources in the emission inventory.

In the example at right, the directions of source regions from the CPF plots agree with the locations of propene sources in the area (red circles), giving more confidence to the source apportionment results. A similar approach can be employed for toxic species.



(Source: Berkowitz et al., 2004)

Emission Inventory Evaluation

Introduction

- **Why bother evaluating emissions data?**
  - Emission inventory development is an intricate process that involves estimating and compiling emissions activity data from hundreds of point, area, and mobile sources in a given region. Because of the complexities involved in developing emission inventories and the implications of errors in the inventory on air quality model performance and control strategy assessment, it is important to evaluate the accuracy and representativeness of any inventory that is intended for use in modeling. Furthermore, existing emission factor and activity data for sources of air toxics and their precursors are limited and the quality of the data is questionable. An emission inventory evaluation should be performed before the data are used in modeling.
- **What tools are available for assessing emissions data?**
  - Several techniques are used to evaluate emissions data including "common sense" review of the data; source-receptor methods such as PMF; bottom-up evaluations that begin with emissions activity data and estimate the corresponding emissions; and top-down evaluations that compare emission estimates to ambient air quality data. Each evaluation method has strengths and limitations.
  - Based on the results of an emissions evaluation, recommendations can be made to improve an emission inventory, if warranted. Local agencies responsible for developing an inventory can then make revisions to the inventory data prior to modeling.
  - $PM_{2.5}$ and PAMS data analysis workbooks provide some example analyses and approaches that are applicable to air toxics data (Main and Roberts, 2000; 2001).

Emission Inventory Evaluation

Using Ambient Data

- Ambient air quality data can be used to evaluate emission estimates ("top-down"); however, the following issues should be considered:
  - Proper spatial and temporal matching of emission estimates and ambient data is needed.
  - Ambient background levels of air toxics need to be considered.
  - Meteorological effects need to be considered.
  - Comparisons are only valid for primarily emitted air toxics.
  - To compare ambient concentrations to emissions estimates, a pollutant or total value (such as total VOC) is needed to create a ratio. Typically, NO_x or CO is used.

Emission Inventory Evaluation

Top-down Approach

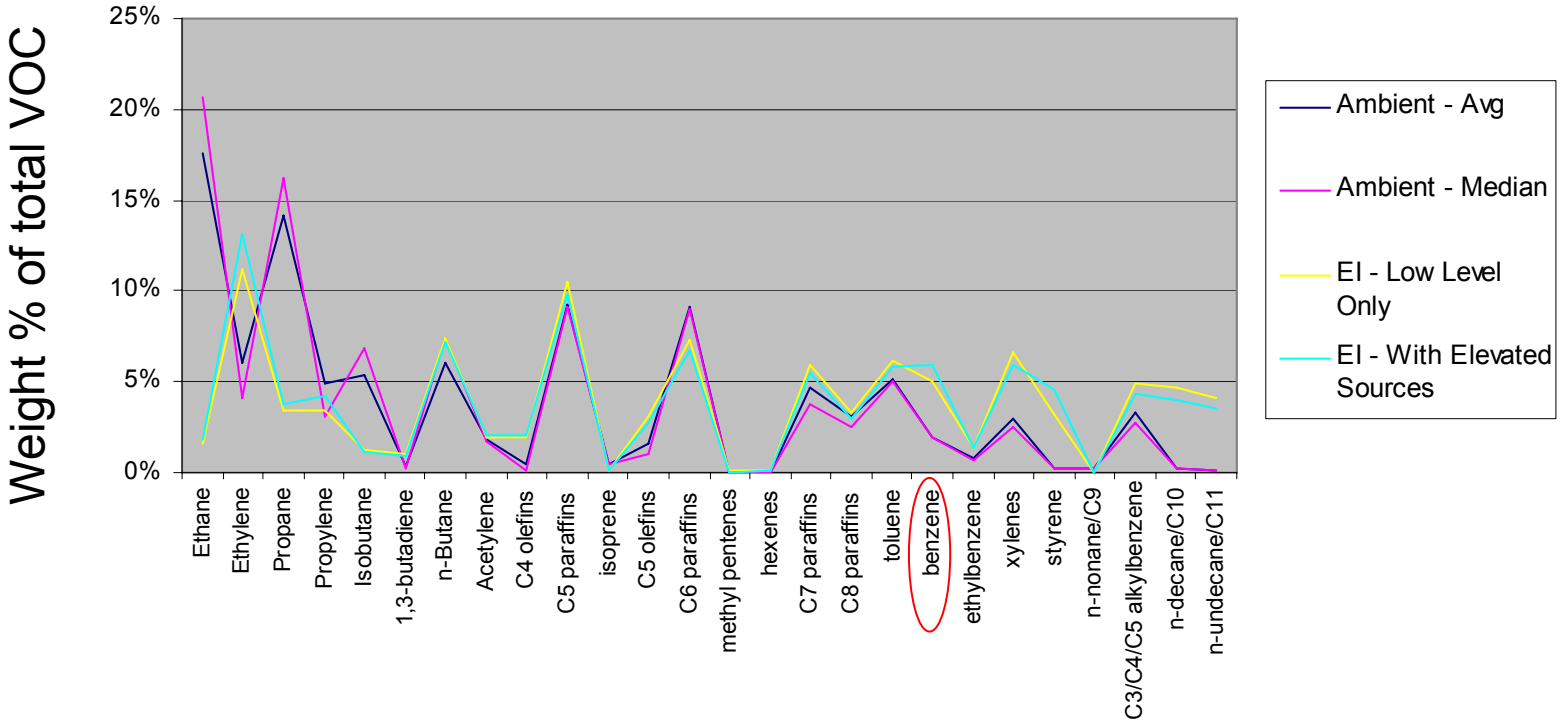
- **Top-down emissions evaluation** is a method of comparing emissions estimates with ambient air quality data. Ambient/emission inventory comparisons are useful for examining the relative composition of emission inventories; they are not useful for verifying absolute pollutant masses unless they are combined with bottom-up evaluations. The top-down method has demonstrated success at reconciling emission estimates of VOC and NO_x.
- **Top-down approach:**

  Compare ambient- and emissions-derived primary air toxic/NO_x, CO, or VOC ratios.

If early morning samples are available (such as with PAMS data), these sampling periods are the most appropriate to use because emissions are generally high, mixing depths are low, winds are light, and photochemical reactions are minimized.

Emission Inventory Evaluation

Example



- At this PAMS site, the EI-derived compositions of benzene are significantly higher than the ambient-derived compositions. Examination of point source records near the source indicates that the sources of these emissions are chemical manufacturing operations. It appears that the chemical speciation profiles used to speciate the point source inventory over-represent the relative amount of benzene (by about a factor of 2 to 5). Similarly, xylenes are overestimated.
- Toluene and 1,3-butadiene are only slightly overestimated in the EI at this site.

Evaluating Models

Introduction

- Air quality models have been used for decades to assess the potential impact of emission sources on ambient concentrations of criteria and toxic air pollutants.
- In the past decade, air quality models have also been used as planning tools for criteria pollutants, e.g., SIP development and attainment demonstration.
- However, until recently, air quality models have not been used as planning tools for air toxics, due to the lack of measurements with which to evaluate the models.
- The need to assess the usefulness of these models in air quality planning and to improve both modeling and evaluation methods has been identified – How well are we modeling air toxics?
- Reasonable agreement between model and monitor concentrations was set by EPA as "within a factor of 2".
- Example of model-to-monitor comparisons for NATA and methodology for comparisons are provided at:
  http://www.epa.gov/ttn/atw/natamain/index.html
  http://www.epa.gov/ttn/atw/nata1999/99compare.html

Evaluating Models

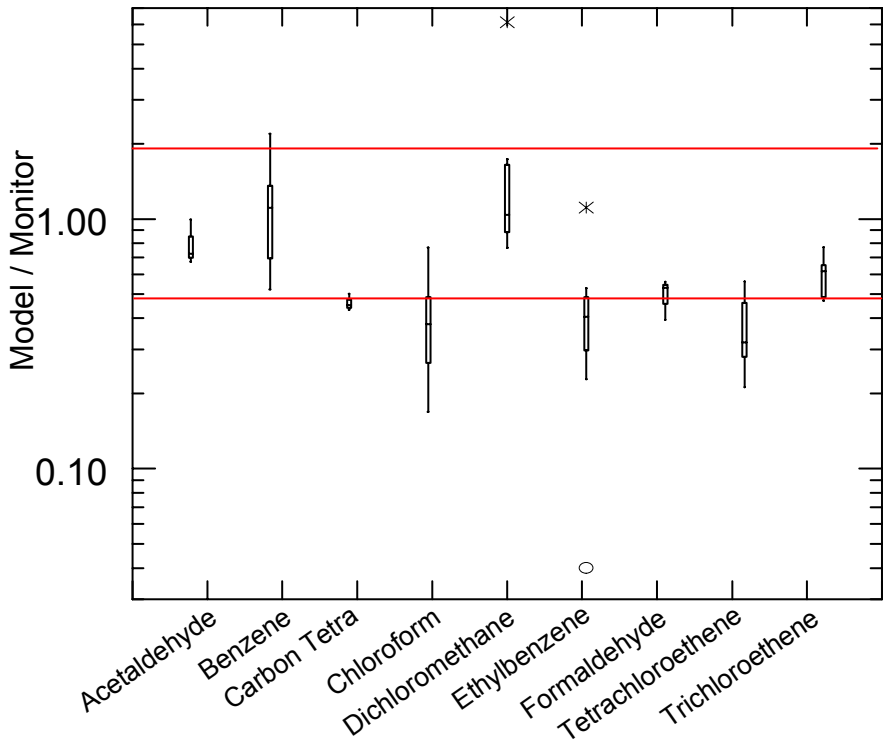
Methodology

- **Modeled Data**. Modeled data of interest for air toxics include publicly available and widely used NATA data. For this example, NATA99 model results were used.
- **Monitored Data**. In order to reduce perturbations from meteorology and other data biases in monitored data, the site average of 1998-2000 valid annual averages was used for comparison to model output.
- The lowest spatial resolution of NATA99 data is census tract level, so NATA99 modeled results should be related to ambient monitoring data at this level. If multiple sites fall into one census tract the sites should still be individually evaluated.
- **Analyses**. If data from many sites are available, box plots of modeled/monitored data can be examined; fewer sites lend themselves to a scatter plot approach of model-to-monitor data. Model-to-monitor ratios within a factor of 2 are considered to be within the acceptable limits of a good comparison; see http://www.epa.gov/ttn/atw/natamain/index.html.

Evaluating Models

Using Box Plots

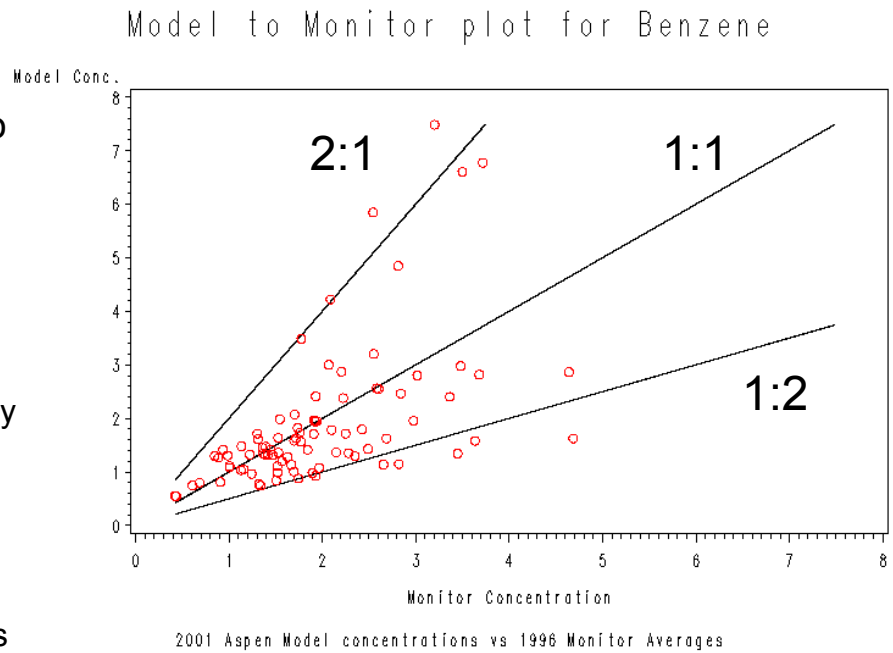
- The figure shows the ratio of NATA99 modeled data to monitored data at an urban area's sites to indicate the accuracy of modeled data.
- Red lines indicate the cutoff for modeled-to-monitored concentrations within a factor of 2.
- Acetaldehyde, benzene, dichloromethane, and trichloroethene typically agreed within a factor of 2, consistent with national level comparisons of model and monitor data.
- However, ethylbenzene, formaldehyde, carbon tetrachloride, chloroform and tetrachloroethylene showed monitored concentrations more than a factor of 2 higher than model estimates at these sites.



Evaluating Models

Using Scatter Plots

- Modeled and monitored concentrations can also be compared using scatter plots, plotting each data pair (ambient site-average, model output) separately. For NATA 1999, benzene data compared well to the modeled data.
- There are several reasons why we would expect good agreement between model prediction and monitor results for benzene.
  - It is a widely distributed pollutant which is emitted from point, area, and mobile sources. Thus, if the model is biased in the way it handles any one of these source categories, the bias will likely be dampened by one of the other sources.
  - An estimated background concentration was available for benzene in the modeling effort.
  - There is a large number (87) of monitoring sites for benzene for this comparison, resulting in an adequate sample size for the statistics in the comparison.
  - Monitoring technology for benzene has a long history, suggesting that the monitoring data reflects actual ambient concentrations.
  - Benzene emissions have been tracked for many years, so there is some confidence in emission estimates.



Model-to-monitor scatter plot for benzene. Most points fall within the factor of 2 wedge, and none are far outside the wedge. From http://www.epa.gov/ttn/atw/nata/draft6.html#secV

Network Assessment

Introduction

- Air quality agencies may choose to re-evaluate and reconfigure monitoring networks because
  - Air quality has changed;
  - Populations and behaviors have changed;
  - New air quality objectives have been established (e.g., air toxics reductions, $PM_{2.5}$, regional haze); and
  - Understanding of air quality issues and monitoring capabilities have improved.
- Network assessments may include
  - Re-evaluation of the objectives and budget for air monitoring;
  - Evaluation of a network's effectiveness and efficiency relative to its objectives and costs; and
  - Development of recommendations for network reconfigurations and improvements.
- Network assessment guidance is available from EPA at http://www.epa.gov/ttn/amtic/cpreldoc.html.

Network Assessment

Methodology

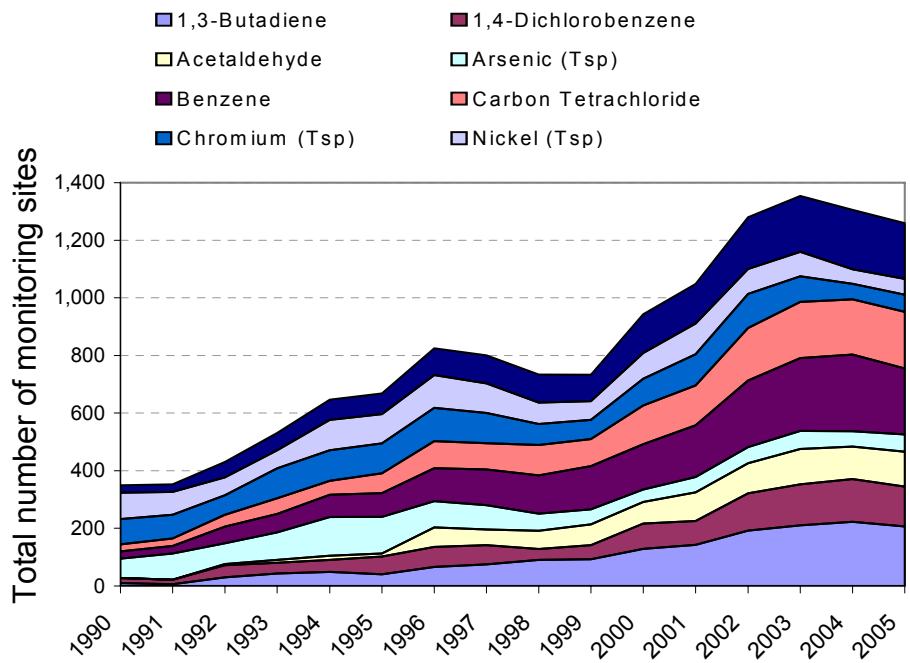
Some things to consider when performing a network assessment:

- Length of monitoring. Takes into account a site's monitoring history because long data records can be highly useful in trends and accountability analyses.
- Suitability analyses. Combines many data sets such as population or population change, meteorology, topography, and emissions to asses suitability of current or future monitoring locations.

Network Assessment

Period of Operation (1 of 2)

- Motivation
  - Monitors that have long historical trends are valuable for tracking trends.
  - This technique places the most importance on sites with the longest continuous trend record.
- Resources needed
  - Historical monitor data, typically valid annual averages.

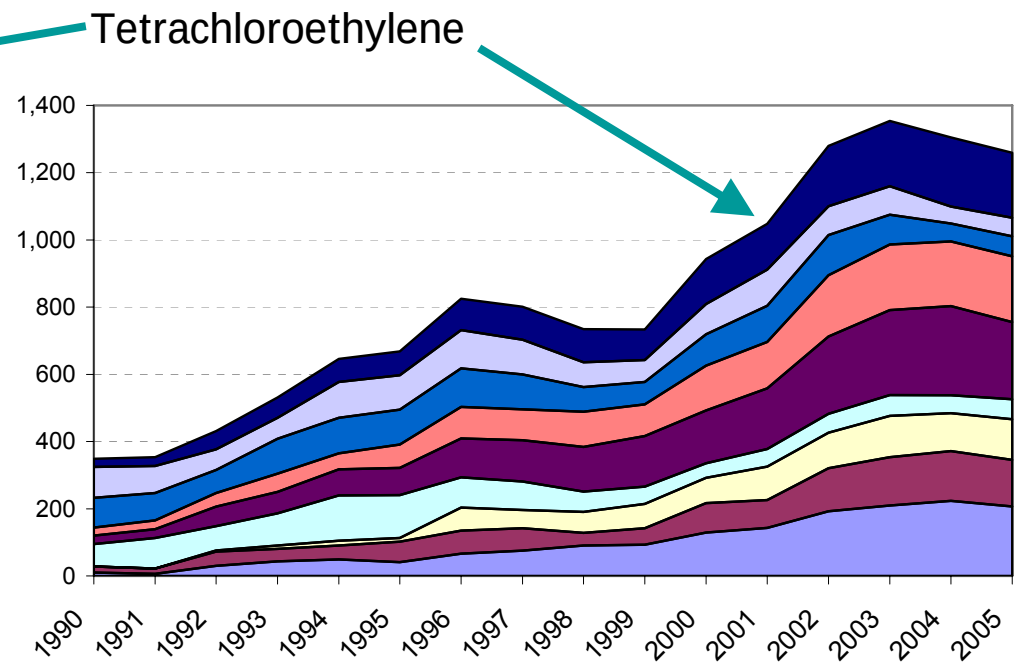


The figure shows the number of monitoring sites per year for a variety of air toxics. The number of air toxics monitoring sites has increased dramatically since 1990.

Network Assessment

Period of Operation (2 of 2)

City, State	AQS SiteID	Years
Stockton, CA	06-077-1002	13
Baltimore, MD	24-510-0040	12
Los Angeles, CA	06-037-1002	11
San Francisco, CA	06-001-1001	10
Fresno, CA	06-019-0008	10
Baltimore, MD	24-005-3001	10
Los Angeles, CA	06-037-1103	9
Los Angeles, CA	06-037-4002	9
San Diego, CA	06-073-0003	9
San Francisco, CA	06-075-0005	9
San Jose, CA	06-085-0004	9
Baltimore, MD	24-510-0006	9
Sacramento, CA	06-061-0006	8
San Diego, CA	06-073-0001	8
Oxnard, CA	06-111-2002	8
Chicago, IL-IN-WI	18-089-2008	8
Baltimore, MD	24-510-0035	8



The table lists the number of annual averages available for tetrachloroethylene at toxics monitoring sites from 1990 to 2003. For this analysis, sites with the longest record would be rated higher than those with shorter records.

Network Assessment

Suitability Modeling/Spatial Analysis (1 of 2)

- Motivation
  - This method may be used to identify suitable monitoring locations based on user-selected criteria.
  - Geographic map layers representing important criteria, such as emissions source influence, proximity to populated places, urban or rural land use, and site accessibility, can be compiled and merged to develop a composite map representing the combination of important criteria for a defined area.
  - The results indicate the best locations to site monitors based on the input criteria and may be used to guide new monitor siting or to understand how changes may impact the current monitoring network.
- Resources needed
  - GIS, site locations, population and other demographic/socioeconomic data, emission inventory data
  - Meteorology and concentration data may be helpful, but are not necessary
  - Skilled GIS analyst

Network Assessment

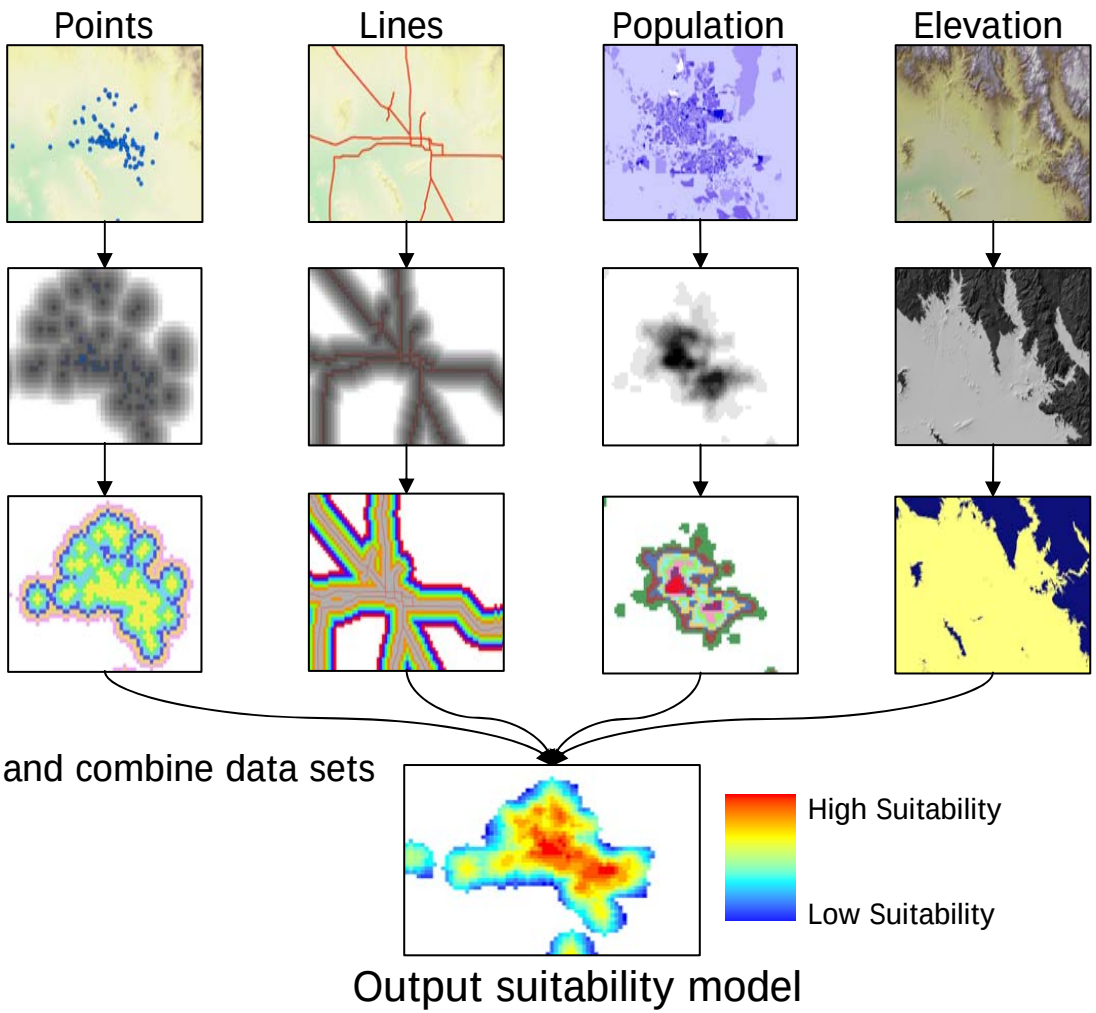
Suitability Modeling/Spatial Analysis (2 of 2)

A representation of the process of suitability modeling and spatial analysis

Input Data:
Point, line, or polygon geographic data

Gridded Data:
Create distance contours or density plots from the data sets

Reclassified Data:
Reclassify data to create a common scale



Network Assessment

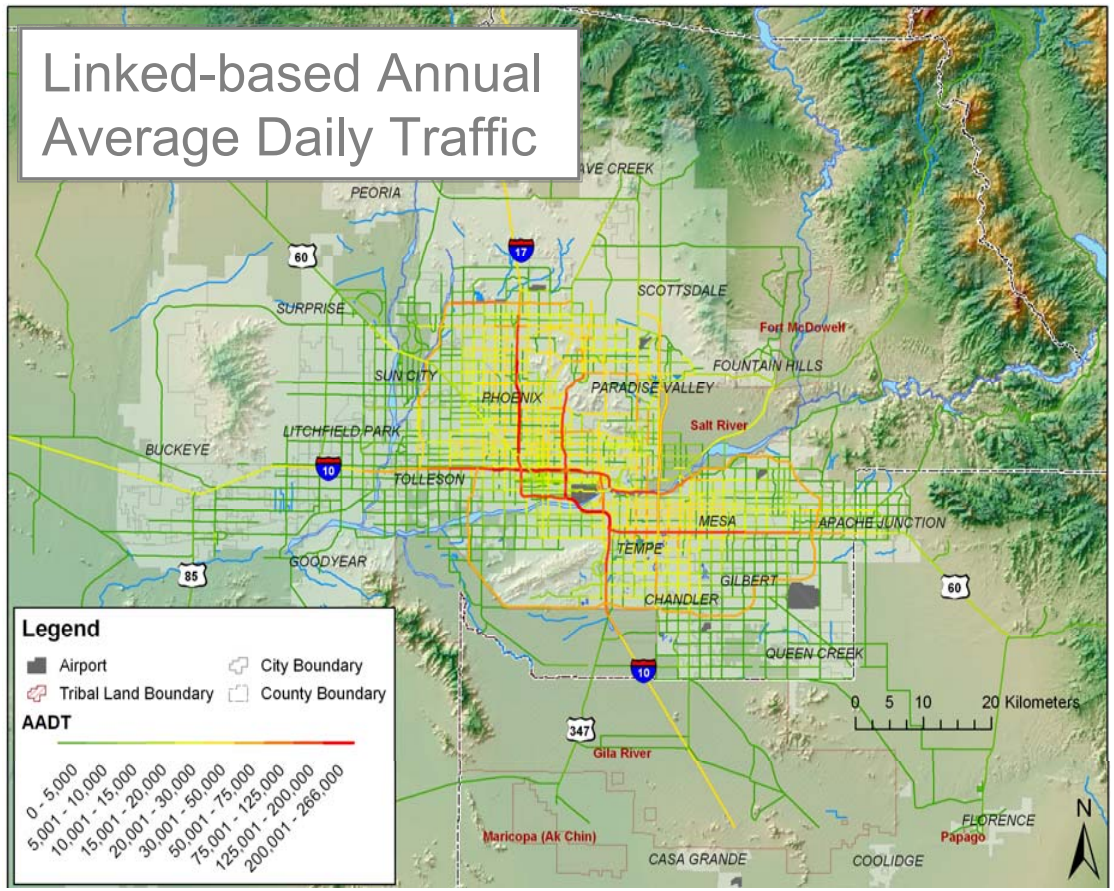
Suitability Modeling Example

- The goal of this analysis of the Phoenix area was to use GIS technology to identify locations within an area potentially suitable for placing air toxics and/or particulate monitors to better assess diesel particulate matter (DPM) emissions impacts on population.
- The emission inventory was assessed to determine
  - predominant sources of DPM; and
  - the best available geographic data to represent the spatial pattern of the identified emission sources in the region.
- The relative importance of each geographic data set was determined based on its potential DPM contribution.
- The input layers were weighted accordingly and combined to produce a suitability map using the Spatial Analyst GIS tool.

Network Assessment

Example Suitability Modeling Data Layers

1. Traffic volume (Annual Average Daily Traffic, AADT)
2. Heavy-duty truck volume (from AADT data)
3. Locations of railroads and transportation depots
4. Residential and commercial development areas
5. Golf courses and cemetery locations (lawn and garden equipment usage)
6. Airport locations
7. $PM_{2.5}$ point source locations (weight assigned to each source depends on the source's relative EC contribution)
8. Total population and sensitive population (e.g., under 5 and over 65 years of age) density
9. Annual average gridded wind fields representing predominant wind direction throughout the region



Network Assessment

Example Suitability Modeling Weighting

Weighting Scheme – two model scenarios were used:

1. Proximity to diesel emission sources (hot spot)
2. Proximity of population to diesel sources

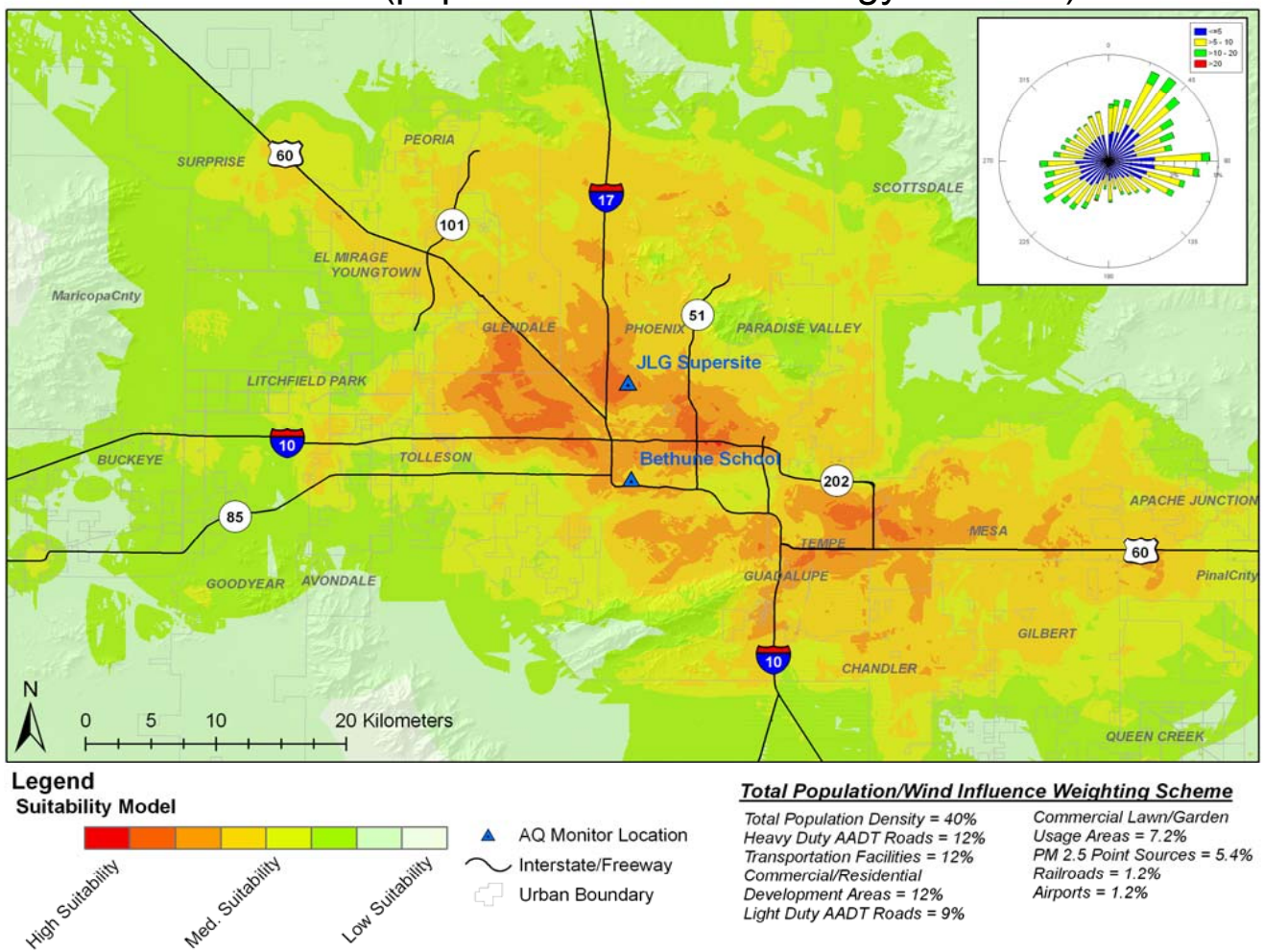
Layer	(1) Hot Spot	(2) Total Population	Weighting Criteria
Density of total population	–	40%	High population density = more suitable
Heavy-duty vehicle activity	20%	12%	High traffic density = more suitable
Light-duty vehicle activity	15%	9%	High traffic density = more suitable
Transportation distribution facility	20%	12%	Close to facility = more suitable
Lawn/garden activity areas	12%	7.2%	High activity density = more suitable
Commercial/residential construction activity areas	20%	12%	High activity density = more suitable
Distance to airports	2%	1.2%	Close to airport = more suitable
Distance to railroads	2%	1.2%	Close to railroad = more suitable
$PM_{2.5}$ point source activity	9%	5.4%	High non-EC $PM_{2.5}$ emissions density = less suitable

Network Assessment

Example Results of Suitability Modeling

- The map shows the results of combining all data layers in Scenario 1 (table on previous slide).
- The map indicates that the Glendale area is a hot spot for both diesel influence and population, as well as the area around the Phoenix Supersite.
- The area between Guadalupe and Mesa is also suitable for monitoring to better understand DPM impacts.

Scenario 1 (population and meteorology included)



Network Assessment

Suitability Analysis Summary

- Results of this analysis assisted decision makers in
  - Assessing the utility of current monitors;
  - Selecting locations for new monitors;
  - Setting monitoring priorities; and
  - Investigating a range of monitoring objectives and considerations.
- Suitability analysis can improve the effectiveness of monitoring decisions

Resources

- PMF, Unmix, and CMB:
  http://www.epa.gov/scram001/receptorindex.htm
- EPA's Multivariate Receptor Modeling Workbook:
  http://www.sonomatechdata.com/sti_workbooks/#MVRMWB
- NOAA HYSPLIT model:
  http://www.arl.noaa.gov/ready/hysplit4.html
- EPA SPECIATE, recently updated (version 4.0): http://www.epa.gov/ttn/chief/software/speciate/index.html.
- Network assessment guidance:
  http://www.epa.gov/ttn/amtic/cpreldoc.html

References

Begum B.A., Kim E., Jeong C.H., Lee D.W., and Hopke P. (2005) Evaluation of the potential source contribution function using the 2002 Quebec forest fire episode. *Atmos. Environ.* **39**, 3719-3724.

Berkowitz C.M., Xie Y.-L., Jolly J., and Estes M. (2004) Receptor modeling and analysis: early first results from the 2003 enhanced Houston Auto-GC network. Presented at the *TERC Science Advisory Committee (SAC) Meeting, October 13*. Available on the Internet at <http://www.harc.edu/harc/Projects/AirQuality/Documents/Meetings/20041013/Presentation09.pdf>.

Brown S.G., Frankel A., and Hafner H.R. (2007a) Source apportionment of VOCs in the Los Angeles area using positive matrix factorization. *Atmos. Environ.* **41**, 227–237 (STI-2725).

Brown S.G., Wade K.S., and Hafner H.R. (2007b) Multivariate receptor modeling workbook. Prepared for the U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-906207.01-3216, August.

Chinkin L.R., Coe D.L., Hafner H.R., and Tamura T.M. (2003) Air toxics emission inventory training workshop. Sponsored by the U.S. Environmental Protection Agency, Region IX, Richmond, CA. Prepared by Sonoma Technology, Inc., Petaluma, CA, STI-903320-2398, July 15-16.

Friedlander S.K. (1973) Chemical element balances and identification of air pollution sources. *Environ. Sci. Technol.* **7**, 235-240.

Fujita E.M., Croes B.E., Bennett C.L., Lawson D.R., Lurmann F.W., and Main H.H. (1992) Comparison of emission inventory and ambient concentration ratios of CO, NMOG, and NOx in California's South Coast Air Basin. *J. Air & Waste Manag. Assoc.* **42**, 264-276.

Fujita E.M., Watson J.G., Chow J.C., and Lu Z. (1994) Validation of the chemical mass balance receptor model applied to hydrocarbon source apportionment in the Southern California Air Study. *Environ. Sci. Technol.* **28**, 1633-1649.

Gordon G.E. (1988) Receptor models. *Environ. Sci. & Technol*. **22(10)**, 1132-1142.

Hafner H.R., Penfold B.M., and Brown S.G. (2005) Using GIS tools to select suitable DPM monitoring locations: Phoenix, Arizona. Presented at the *2005 Air Toxics Summit, Seeking Solutions for our Rural and Urban Communities, Portland, OR, October 18-19*, by Sonoma Technology, Inc., Petaluma, CA (STI-904234-2755).

References

Henry R. C. (1997) History and fundamentals of multivariate air quality receptor models. *Chemometrics and Intelligent Laboratory Systems* **37**, 525-530.

Henry R.C., Lewis C.W., Hopke P.K., and Williamson H.J. (1984) Review of receptor model fundamentals. *Atmos. Environ.* **18** (8), 1507-1515.

Henry R.C. (1997) History and fundamentals of multivariate air quality receptor models. *Chemometrics and Intelligent Laboratory Systems* **37**, 525-530.

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

Henry R.C. (2002) Receptor modeling. In *Encyclopedia of Environmetrics*, A.H. El-Shaarawi and W.W. Piegorsch eds., John Wiley & Sons, Ltd, Chichester, 1706-1721.

Hidy G.M. and Friedlander S.K. (1971) The nature of the Los Angeles aerosol. In proceedings from the *Second International Clean Air Congress*, 391-404, Academic Press, New York.

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.

Hopke P.K., Ramadan Z., Paatero P., Norris G., Landis M., Williams R., and Lewis C.W. (2003) Receptor Modeling of Ambient and Personal Exposure Samples: 1998 Baltimore Particulate Matter Epidemiology-Exposure Study. *Atmos. Environ.* **37**, 3289-3302.

Kim E., Hopke P.K., Larson T.V., and Covert D.S. (2004a) Analysis of ambient particle size distributions using UNMIX and positive matrix factorization. *Environ. Sci. Technol*. **38** (1), 202-209.

Kim E., Hopke P.K., Larson T.V., Maykut N.N., and Lewtas J. (2004b) Factor analysis of Seattle fine particles. *Aerosol Sci. Technol.* **38** (7), 724-738.

Kim E., Hopke P.K., Kenski D.M., and Koerber M. (2005b) Sources of fine particles in a rural Midwestern U.S. area. *Environ. Sci. Technol,* **39** (13), 4953-4960.

Kim E. and Hopke P.K. (2004) Improving source identification of fine particles in a rural northeastern U.S. area utilizing temperature-resolved carbon fractions. *J. Geophys. Res.* **109** (D9), D09204, doi: 09210.01029/02003JD004199.

References

Kim E., Hopke P.K., and Qin Y. (2005c) Estimation of organic carbon blank values and error structures of the speciation trends network data for source apportionment. *J. Air & Waste Manag. Assoc.* **55**, 1190-1199.

Lindsey C.G., Chen J., Dye T.S., Richards L.W., and Blumenthal D.L. (1999) Meteorological processes affecting the transport of emissions from the Navajo Generating Station to Grand Canyon National Park. *J. Appl. Meteorol.* **38** (No. 8), 1031-1048.

Main H.H. and Roberts P.T. (2001) PM2.5 data analysis workbook. Draft workbook prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-900242-1988-DWB, February.

Main H.H. and Roberts P.T. (2000) PAMS data analysis workbook: illustrating the use of PAMS data to support ozone control programs. Prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-900243-1987-FWB, September.

Larsen R.K. and Baker J.E. (2003) Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: a comparison of three methods. *Environ. Sci. Technol.* **37** (9), 1873-1881.

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. *J. Air & Waste Manag. Assoc.* **53** (3), 325-338

Paatero P. and Tapper U. (1994) Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **5**, 111-126.

Paatero P. (1997) Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems* **37**, 23-35.

Paatero P., Hopke P.K., and Philip K. (2003) Discarding or downweighting high-noise variables in factor analytic models. *Anal. Chim. Acta* **490**, 277-289.

Poirot R.L., Wishinski P.R., Hopke P.K., and Polissar A.V. (2001) Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont. *Environ. Sci. Technol.* **35** (23), 4622-4636.

Raffuse S.M., Sullivan D.C., McCarthy M.C., Penfold B.M., and Hafner H.R. (2006) Analytical techniques for technical assessments of ambient air monitoring networks. Guidance document prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-905212.02-2805-GD, September.

References

Raffuse S.M., Brown S.G., Sullivan D.C., and Chinkin L.R. (2005) Estimating regional contributions to atmospheric haze. Presented at the *2005 ESRI International User Conference, San Diego, CA, July 26* (STI-2649).

Raffuse S.M., Sullivan D.C., and Chinkin L.R. (2005) Emission impact potential - a method for relating upwind emissions to ambient pollutant concentrations. Presented at *the U.S. Environmental Protection Agency 14th International Emission Inventory Conference, Las Vegas, NV, April 11-14* by Sonoma Technology, Inc., Petaluma, CA (STI-2715, STI-2722). Available on the Internet at <http://www.epa.gov/ttn/chief/conference/ei14/session2/raffuse.pdf>.

Rosenbaum A.S., Ligocki M.P., and Wei Y.H. (1999) Modeling cumulative outdoor concentrations of hazardous air pollutants. Revised final report prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC (SYSAPP-99-96/33R2), February.

Seigneur C., Pun B., Lohman K., and Wu S.-Y. (2002) Air toxics modeling. Report prepared for Coordinating Research Council, Inc., Alpharetta, GA and U.S. Department of Energy's Office of Heavy Vehicle Technologies through the National Renewable Energy Laboratory, Golden, CO by Atmospheric & Environmental Research, Inc., San Ramon, CA, CRC Project Number A-42-1, Document Number CP079-02-3, August. Available on the Internet at <http://www.crcao.com/reports/recentstudies00-02/A-42-1%20Final%20Report.pdf>;

U.S. Environmental Protection Agency (1999) Air dispersion modeling of toxic pollutants in urban areas: Guidance, methodology and example applications. Prepared by the Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/R-99-021, July

Watson J.G. (1979) Chemical element balance receptor model methodology for assessing the source of fine and total particulate matter. Ph.D. Dissertation, Oregon Graduate Center, Portland, OR, University Microfilms International, Ann Arbor, MI.

Watson J.G. (1984) Overview of receptor model principles. *J. Air Poll. Cont. Assoc.* **34** (6), 619-623.

Watson J.G., Fujita E.M., Chow J.C., Zielinska B., Richards L.W., Neff W., and Dietrich D. (1998) Northern front range air quality study. Final report prepared for Colorado State University, Cooperative Institute for Research in the Atmosphere, Fort Collins, CO, by Desert Research Institute, Reno, NV, STI-996410-1772-FR, June.