Sonoma Technology, Inc.
*Innovative Environmental Solutions*

# PAMS Data Validation and Analysis Training Material

Prepared by
Hilary R. Hafner
Bryan M. Penfold
Sonoma Technology, Inc.
Petaluma, CA

Prepared for
Kevin Cavender
Office of Air Quality Planning and Standards (OAQPS)
U.S. Environmental Protection Agency
Research Triangle Park, NC

January 4, 2018

# Contents

# Figures

# Tables

# 1. Overview

## 1.1 Goals of Training Material

To support the current ozone NAAQS, selected regions in the U.S. monitor volatile organic compounds (VOCs), nitrogen oxides, and other parameters under the umbrella of the Photochemical Assessment Monitoring Stations (PAMS) program. EPA sponsored the development of this training material to support state and local agencies in collecting, validating, analyzing, and visualizing the PAMS VOC data.[1] This workbook refers to the EPA's Data Analysis and Reporting Tool (DART) for most of the data validation and analysis tasks discussed. Although the workbook focuses on PAMS VOC data, DART can be used to validate, analyze, and report other types of data as well

The goals of this training material are to:

- Provide useful background information on data validation and analysis of PAMS VOC data.
- Guide state and local agencies in the use of validation and analysis methods, procedures, and tools for PAMS and similar data sets.
- Demonstrate how to use EPA's Data Analysis and Reporting Tool (DART) for selected validation and analysis steps for PAMS VOC data.

## 1.2 DART Basics

DART is a web-based data validation and analysis system that is integrated with AirNow-Tech. It not only provides a framework for validating and analyzing air quality data, but it will eventually enable access to complementary data sets from different sources and web services. DART can be used to validate and analyze any air quality data, including PAMS VOC measurements and other air quality measurements such as ozone, lead, air toxics, other gaseous pollutants, speciated particulate matter (PM), and meteorological measurements.

DART was launched in 2014, and updates are continuing. DART has a detailed user guide, which outlines how to bring data into DART via upload or via AQS, use the validation and plotting tools, and export data files.

[1] Note that the official term for PAMS VOCs in AQS is Total Nonmethane Organic Compounds, TNMOC. We use VOC and TNMOC interchangeably in this document.

## 1.3 How to Interact with Training Material within DART

This training material has a number of hyperlinks embedded in the text, which open up new browser windows for the subject in question. Users can keep the training material open in one browser window while using DART in another window; this allows users to easily switch between the two and facilitates training workshops.

# 2. PAMS Measurements

## 2.1 PAMS Network and Goals

On October 1, 2015, EPA made significant changes to the PAMS monitoring requirements and applicability (40 CFR part 58 Appendix D, section 5.0). Ambient concentrations of ozone and ozone precursors from the PAMS network will be used to make attainment/nonattainment decisions, aid in tracking VOC and NOx emission inventory reductions, better characterize the nature and extent of the ozone problem, and prepare air quality trends. In addition, data from the PAMS will provide an improved database for evaluating photochemical model performance, especially for future control strategy mid-course corrections as part of the continuing air quality management process.

Key changes in the network include:

- Requiring hourly VOC measurements – although there is a waiver to allow 3 8-hr canister samples in locations with low VOC concentrations and for "logistical and programmatic constraints"
- Requiring 3 8-hr carbonyls samples on a 1 in 3 day schedule – there is also an alternative to allow for continuous formaldehyde measurements
- Requiring "true $NO_2$ " in addition to existing NO and $NO_y$ measurements
- Requiring hourly mixing height measurement (replaces "upper air measurements") – There is a waiver option to allow measurements to be made at an alternative location (e.g., NOAA ASOS sites)
- Additional PAMS meteorology measurements that are not part of the NCore requirements include atmospheric pressure, precipitation, solar radiation, and UV radiation

A map of the PAMS network planned for 2019 deployment is shown in **Figure 1**.

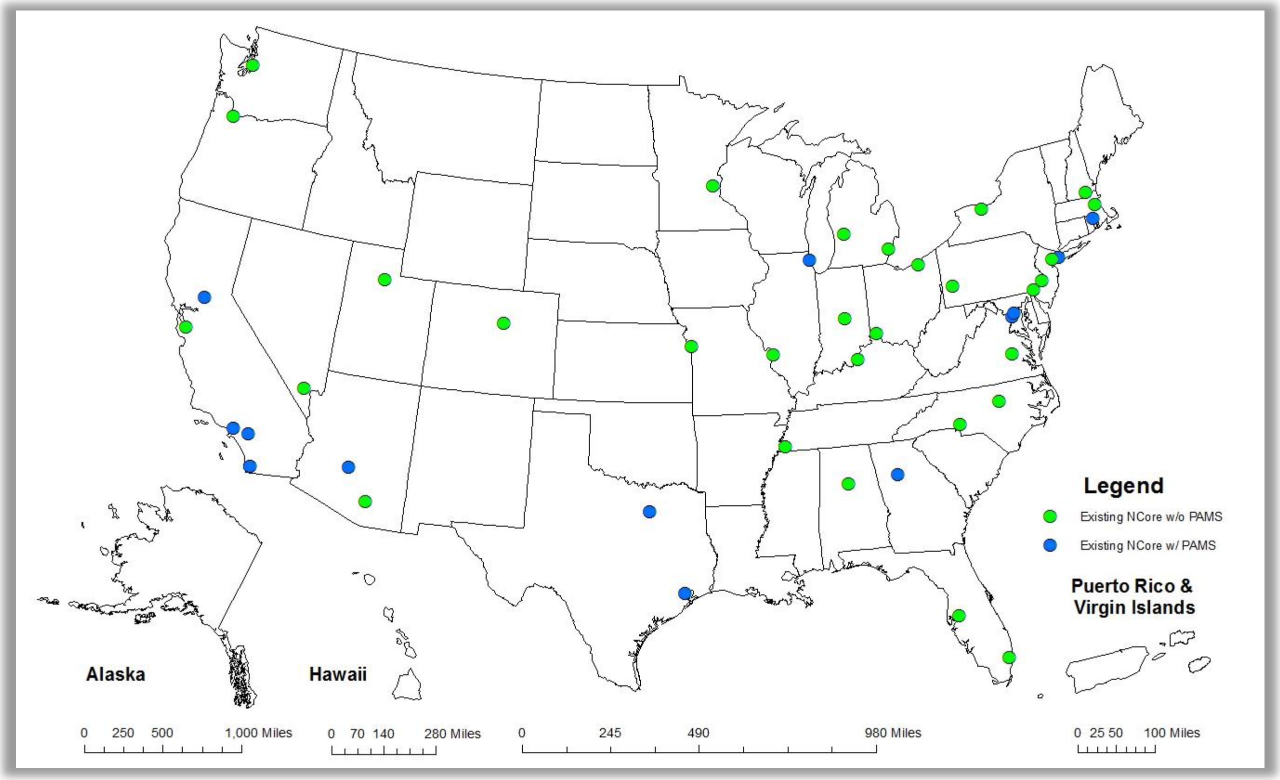


**Figure 1.** Map of PAMS sites for 2019 (from Cavender, 2016).

## 2.2 PAMS Measurements

Measurements of PAMS VOCs are typically made by an automatic gas chromatograph (auto-GC) or canisters with subsequent laboratory analysis. Auto-GCs typically provide hourly data, while canisters can be collected as 8-hour averages (i.e., three canisters per day). Table 1 shows the required PAMS VOC species, and Tables 2 and 3 list other species that often are, or have been, measured at PAMS monitoring sites. Table 4 summarizes the common emission sources of the target species.

**Table 1.** AQS codes and abbreviations of PAMS VOC species often or historically measured as part of PAMS..

| Parameter Name | AQS Code | Parameter Name | AQS Code | Parameter Name | AQS Code |
|---|---|---|---|---|---|
| 1,2,3-Trimethylbenzene | 45225 | Benzene | 45201 | n-Heptane | 43232 |
| 1,2,4-Trimethylbenzene | 45208 | cis-2-Butene | 43217 | n-Hexane | 43231 |
| 1,3,5-Trimethylbenzene | 45207 | cis-2-Pentene | 43227 | n-Nonane | 43235 |
| 1-Butene | 43280 | Cyclohexane | 43248 | n-Octane | 43233 |
| 1-Pentene | 43224 | Cyclopentane | 43242 | n-Pentane | 43220 |
| 2,2,4-Trimethylpentane | 43250 | Ethane | 43202 | n-Propylbenzene | 45209 |
| 2,2-Dimethylbutane | 43244 | Ethylbenzene | 45203 | n-Undecane | 43954 |
| 2,3,4-Trimethylpentane | 43252 | Ethylene | 43203 | o-Ethyltoluene | 45211 |
| 2,3-Dimethylbutane | 43284 | Formaldehyde | 43502 | o-Xylene | 45204 |
| 2,3-Dimethylpentane | 43291 | Isobutane | 43214 | p-Diethylbenzene | 45219 |
| 2,4-Dimethylpentane | 43247 | Isopentane | 43221 | p-Ethyltoluene | 45213 |
| 2-Methylheptane | 43960 | Isoprene | 43243 | Propane | 43204 |
| 2-Methylhexane | 43263 | Isopropylbenzene | 45210 | Propylene | 43205 |
| 2-Methylpentane | 43285 | m/p Xylene | 45109 | Styrene | 45220 |
| 3-Methylheptane | 43253 | m-Diethylbenzene | 45218 | Sum of PAMS target compounds | 43000 |
| 3-Methylhexane | 43249 | Methylcyclohexane | 43261 | Toluene | 45202 |
| 3-Methylpentane | 43230 | Methylcyclopentane | 43262 | Total NMOC (non-methane organic compound) | 43102 |
| Acetaldehyde | 43503 | m-Ethyltoluene | 45212 | trans-2-Butene | 43216 |
| Acetone | 43551 | n-Butane | 43212 | trans-2-Pentene | 43226 |
| Acetylene | 43206 | n-Decane | 43238 | | |

**Table 2.** AQS codes and abbreviations of meteorological parameters often or historically measured as part of PAMS..

| Parameter Name | AQS Code | Parameter Name | AQS Code | Parameter Name | AQS Code |
|---|---|---|---|---|---|
| Barometric pressure | 64101 | Solar radiation | 63301 | Wind direction - Resultant | 61104 |
| Dew point | 62103 | Ultraviolet radiation | 63302 | Wind direction - Scalar | 61102 |
| Outdoor temperature | 62101 | Ultraviolet radiation (type B) | 63304 | Wind speed - Resultant | 61103 |
| Rain/melt precipitation | 65102 | Vertical wind direction | 61112 | Wind speed - Scalar | 61101 |
| Relative humidity | 62201 | | | | |

**Table 3.** AQS codes and abbreviations of other air quality parameters often or historically measured as part of PAMS.

| Parameter Name | AQS Code |
|---|---|
| Nitric oxide (NO) | 42601 |
| Nitrogen dioxide ($NO_2$) | 42602 |
| Oxides of nitrogen ($NO_x$) | 42603 |
| Ozone | 44201 |
| Reactive oxides of nitrogen ($NO_y$) | 42600 |

Table 4. Key species, their sources, and comments relevant to data analysis.

| Species | Major Sources | Comments |
|---|---|---|
| Ethene (ethylene) | Mobile sources, petrochemical industry | Marker for vehicle exhaust |
| Acetylene | Mobile sources, combustion processes | Marker for vehicle exhaust. More abundant in gasoline exhaust than diesel exhaust |
| Ethane | Natural gas use | Non-reactive |
| Propene (propylene) | Refinery, chemical manufacturing, motor vehicle exhaust | More abundant in diesel exhaust than gasoline exhaust |
| Propane | LPG and natural gas use, oil and gas production | Relatively non-reactive, often underestimated in emission inventory. Also more abundant in diesel exhaust than gasoline exhaust |
| i-Butane | Consumer products, gasoline evaporative emissions, refining | Used as replacement of chlorofluorocarbons (CFCs) in consumer products |
| Butene | Motor vehicle exhaust | More abundant in gasoline exhaust than diesel exhaust. |
| n-Butane | Gasoline evaporative emission | Marker of gasoline use |
| t-2-Butene | Motor vehicle exhaust | Enriched in evaporated gasoline relative to exhaust |
| i-Pentane | Solvent use, refining, mobile sources | Among most abundant species in urban air. More abundant in gasoline exhaust than diesel exhaust |
| n-Pentane | Motor vehicle exhaust, gasoline evaporative emissions | Enriched in evaporative emissions relative to exhaust |
| Isoprene | Biogenics | Marker of biogenic emission; reactive |
| Internal olefins (e.g., t-2-pentene) | Gasoline evaporative emissions, plastics production | Reactive |
| 2,2-dimethylbutane | Motor vehicle exhaust | More abundant in diesel exhaust than gasoline exhaust |

| Species | Major Sources | Comments |
|---|---|---|
| Benzene | Motor vehicle exhaust, combustion processes, refining | Marker for vehicle exhaust |
| 2-Methylhexane | Motor vehicle exhaust | More abundant in gasoline exhaust than diesel exhaust |
| 2,2,4-Trimethylpentane | Gasoline evaporative emissions | Also in motor vehicle exhaust |
| n-Heptane | Surface coatings, degreasing | Also in motor vehicle exhaust |
| Toluene | Solvent use, refining, mobile sources | Among most abundant species in urban air |
| Styrene | Solvent use, chemical manufacturing | Also in motor vehicle exhaust |
| Heptane and octane isomers | Oil and gas production, asphalt, gasoline | Also in motor vehicle exhaust |
| n-Nonane | Dry cleaning, degreasing, motor vehicles | Also in motor vehicle exhaust |
| Xylenes | Solvent use, refining, mobile sources | Reactive |
| n-Decane, undecane | Fuel storage, surface coatings | More abundant in diesel exhaust than gasoline exhaust |
| Formaldehyde | Fuel combustion | Also a key photochemical reaction product (secondary source) |
| Acetaldehyde | Fuel combustion | Also a product of photochemistry |

### 2.2.1 Auto-GCs

Several automated measurement options for VOCs are being evaluated by EPA (https://www3.epa.gov/ttnamti1/pamsreeng.html), including automatic gas chromatographs. Auto-GCs can provide speciated hydrocarbon data on a 1-hr basis (typically referred to as continuous measurements). These instruments often use gas chromatography with mass spectrometry (GC/MS) or flame ionization (GC-FID). In order to capture both lower (C2-C5) and higher (C6+) carbon number species, many instruments have dual columns. Sometimes only one column will fail, so analysts must be alert to check for the presence of all expected species. Other common data validation items include peak misidentification and column contamination.

Continuous data provide a rich database from which diurnal variations, concentration responses to wind speed and direction, and comparisons to other pollutant concentrations can be explored.

### 2.2.2 Canister Sampling

When canisters are used for sample collection, the measurement process is to deploy canisters, collect samples according to a schedule (e.g., three 8-hr samples every third day), retrieve samples, and send them to a laboratory. At the laboratory, sample analysis follows standard compendium methods, such as TO-15, using GC/MS, GC-FID, or other methods. Potential errors found in canister data include contamination, peak misidentification, and poor recovery.

Canister data sets are useful for exploring source types through source apportionment techniques but lack sufficient time resolution to investigate diurnal changes, for example.

### 2.2.3 File Formats

Typically, auto-GCs provide raw text files that require processing and data aggregation. DART can process legacy instrument TX0 files (from Perkin Elmer auto-GCs) as well as more modern files. DART can ingest three file types: AQS, RD, crosstab, and Perkin Elmer, described below.

Currently, data from only one monitoring site can be included in the ingest file. A sample of each file format is provided below the description.

- U.S. EPA **AQS RD** (Hourly, Daily, and Sub Hourly Raw Data) pipe-delimited text file format.
- The **crosstab** format can be used to import data with a column for each measured parameter. DART currently supports three variations of the crosstab file format. In all three variations, the first column in the crosstab format must be a column labeled "Date." The three variations of the crosstab file format are as follows:
    - The date column is followed by a column for each parameter.
    - The date column is followed by a column for each parameter, and each parameter column is followed by a column for the method detection limit (MDL) of each parameter.
    - The date column is followed by a column for each parameter, and each parameter column is followed by a column for the units and then another column for the MDL of the parameter.

  Any time interval can be used (e.g., hourly, daily, etc.), but for a given data file, data must all be of the same time interval, with no duplicate date/time rows. In all three variations, the parameter columns must be labeled with their five-digit AQS parameter code, and the values in each column should be the concentration of the parameter at the date/time specified for that data row. Tables 5 and 6 show examples of crosstab formats for 24-hr and 1-hr data, respectively.
- **Perkin-Elmer** automatic gas chromatographs (auto-GCs) operated by monitoring agencies in the PAMS program are commonly used with Turbochrome or Total Chrome software. Typically, there are two TX0 files for each hour of data collection, representing the two channels of measurement.

For canister samples, most laboratories have their own lab-specific file formats, which can make data processing challenging. Additional details and examples are provided in the DART User Guide.

**Table 5.** Example of crosstab format for a 24-hr data file (this shows a date column followed by a column for each parameter).

| Date | 42153 | 43218 | 43502 | 43503 | 43819 |
|---|---|---|---|---|---|
| 6/14/2004 | 0.05 | 0.02 | 0.8 | 0.4 | 0.015 |
| 6/20/2004 | 0.05 | 0.02 | 0.5 | 0.2 | 0.015 |
| 6/26/2004 | 0.05 | 0.02 | 0.8 | 0.3 | 0.015 |
| 7/2/2004 | 0.05 | 0.02 | 0.9 | 0.4 | 0.015 |
| 7/8/2004 | 0.05 | 0.04 | 0.9 | 0.3 | 0.015 |
| 7/14/2004 | 0.1 | 0.02 | 0.7 | 0.2 | 0.015 |
| 7/20/2004 | 0.05 | 0.02 | 0.6 | 0.1 | 0.015 |

**Table 6.** Example of crosstab format for a 1-hr data file (this shows a date column followed by a column for each parameter).

| Date | 42153 | 43218 | 43502 | 43503 | 43819 |
|---|---|---|---|---|---|
| 6/14/2004 0:00 | 0.05 | 0.02 | 0.8 | 0.4 | 0.015 |
| 6/14/2004 1:00 | 0.05 | 0.02 | 0.5 | 0.2 | 0.015 |
| 6/14/2004 2:00 | 0.05 | 0.02 | 0.8 | 0.3 | 0.015 |
| 6/14/2004 3:00 | 0.05 | 0.02 | 0.9 | 0.4 | 0.015 |
| 6/14/2004 4:00 | 0.05 | 0.04 | 0.9 | 0.3 | 0.015 |
| 6/14/2004 5:00 | 0.1 | 0.02 | 0.7 | 0.2 | 0.015 |
| 6/14/2004 6:00 | 0.05 | 0.02 | 0.6 | 0.1 | 0.015 |

# 3. PAMS Data Validation

## 3.1 Data Validation Overview

*Validation* is confirmation through objective evidence that the requirements for a specific intended use are fulfilled—for example, that data have passed quality control checks, instruments have passed audits, and the data appear reasonable.

Data validation is vital because erroneous data values can cause serious errors in data analysis and modeling results. Monitoring agencies have the responsibility to prevent, identify, correct, and define the consequences of monitoring difficulties that might affect the precision and accuracy, and/or the validity, of their measurements.

Timely data validation is needed to minimize the amount of potentially invalid data that may be generated (and the corresponding level of effort required to address the problem) and thus maximize the recoverable data. As more time passes from data collection, more effort may be required to assess potentially invalid data.

For complex data sets such as PAMS data sets, data validation is challenging because each sample comprises multiple chemical species. This section provides definitions, steps to take, and examples.

## 3.2 Data Validation Definitions

There are typically four levels of data validation, from very basic validation of data collection and labeling (Level 0), to more advanced analyses that put data in historic and/or spatial context (Level III).[2]

- **Level 0:** Routine checks made during the initial data processing and generation of data, including proper data file identification, review of unusual events, review of field data sheets and result reports, and results from instrument performance checks, audits, and inter-laboratory comparisons.
- **Level I:** Internal consistency checks to identify data values that appear atypical when compared to values of the entire data set. Also includes review of data for gaps.
- **Level II:** Comparison of the current data set with historical data from the same site(s) to verify consistency over time. This level can be considered part of the data interpretation or analysis process.
- **Level III:** Tests for parallel consistency with data sets from the same population (e.g., region, period of time, or air mass) but from different sites to identify systematic bias. This level can also be considered part of the data interpretation or analysis process.

In practice, Level 0 validation is performed by the reporting agency prior to data submittal.

Most importantly, data should only be invalidated when there is compelling information supporting that there was a measurement error or contamination.

[2] From U.S. Environmental Protection Agency (1999) Particulate matter ($PM_{2.5}$) speciation guidance document. Available at http://www.epa.gov/ttnamti1/files/ambient/pm25/spec/specfinl.pdf.

## 3.3 Suggested Data Validation Process for PAMS Data

There are two ways to acquire or assemble data for validation in DART: (1) retrieve data from AQS through a DART query, or (2) upload your own data files. For the second option, it is important to regularly compile the raw data as data are collected; serious data issues or gaps can be caught during sampling rather than after sampling is completed. Level 0 validation should be performed by the reporting agency prior to the data being placed in AQS. For import to DART, the data need to be placed in a common data format with descriptive information concerning variables, validation level, QC codes, time standard, and standard units. Once data are brought into DART, the data are ready for Level 1 validation.

Before beginning data validation, it is helpful to think about potential impacts on pollutant concentrations and on data quality. Every sampling location and system has unique characteristics. Throughout all data validation and analysis steps, consider the following factors:

- Levels of other pollutants at the same site – was there a meteorological event or high pollution episode affecting measurements at the site?
- Levels of pollutants at other nearby, or similar, sites – are concentrations typical for the area or for a similar site?
- Time of day, day of week, and season – are concentration patterns consistent with known diurnal, day-of-week, or season patterns based on meteorology and emissions sources?
- Audit results (e.g., recurring problems with a particular compound) – are there known problems with some species but not others in the laboratory or instrument performance?
- Instrument performance history – are there consistent problems encountered with an instrument that affect data quality?
- Calibration or baseline drift – have concentrations drifted over time?
- Site characteristics and nearby emission sources – are spikes in concentrations consistent with nearby activities?
- Meteorology – are concentrations varying as expected with wind speeds, wind directions, temperature, or other meteorological phenomena?
- Exceptional events (e.g., holiday celebrations, fires, etc.) – are there nearby emissions impacts from unusual or holiday events?

After data are loaded into DART, data validation steps include the following:

1. Conduct Level I validation:
    a. Screen data using DART auto validation or customized auto validation checks. Review results visually, especially investigating hours surrounding data that fail screening. Screening helps analysts focus on the data needing the most attention.
    b. Review summary statistics for unrealistic maxima or minima.

   c. Review time series of each species group and then individual species; identify outliers, missing data, presence or absence of key species, “sticky” values, or other unusual data points (most of these are identifiable using automated screening). Verify these data, or, if there are reasons why data are incorrect, flag data with qualifier or null codes. Inspect every species, even to confirm the expectation that the species would normally be below the method detection limit. Other features to explore in time series include jumps and dips in data, data gaps, diurnal patterns, and baseline drift.
   d. Examine scatter plots to ensure that relationships among species are expected/typical; investigate samples with atypical results. Data lining up along two axes, also called an “open jaw” may indicate peak misidentification.
   e. Evaluate fingerprint plots to understand how concentrations change sample by sample; investigate radical, unexpected or unexplained changes.
   f. Apply flags to data and document changes. If you are using “DART Smarts,” some data will already have been flagged if they failed screening checks by a large margin.
   g. Adjust screening checks and customize them in light of experience with your site’s data.

2. Conduct Level II validation
   a. Compare your site’s data to data from similar sites, such as nearby sites or sites from similar sized urban areas (spatial).
   b. Compare data to data collected at the same site from previous years (temporal).

3. Conduct Level III validation
   a. Perform intercomparisons of the data (e.g., from two different instruments at the same site; this type of intercomparison is rarely available).

The subsequent sections provide more detail on how to use DART to accomplish data validation tasks.

## 3.4 Screening Checks

Given the volume and complexity of the PAMS data, automated screening can be conducted. These checks are useful to help analysts focus efforts on data that need the most attention. Checks include:

- Abundant species - are typically abundant hydrocarbons present in each sample?
- TNMOC – are TNMOC values available, does the unidentified portion (TNMOC minus the sum of PAMS target species) exceed 50% of TNMOC, or does the sum of PAMS species exceed TNMOC?
- Range – are concentrations higher or lower than typical for the pollutant and site? I
- Sticking – are there repeated values, above zero, for three or more consecutive hours?
- Chemical consistency – are typical chemical relationships as expected?

Screening checks built on these concepts are summarized in Table 7.

## 3.5 Level 1 Validation Checks

The Table 7 checks can be used during Level 1 (internal consistency) data validation. Screening checks can be run in DART from the Validate webpage using DART's built in "PAMS Basic" suite of checks or by configuring a custom suite of screening checks tailored to a particular monitoring site. DART Smarts can also be enabled to automatically apply null codes on data during the screening process.

Once you have the results of the screening checks, review results visually, especially investigating hours before and after data that fail screening and hours before and after already invalid or missing data. Sometimes, it may be necessary to invalidate data collected just before and just after invalid data because the data issue can be seen to have already started or is continuing. Examples from DART are provided for each check in the figures following Table 7.

**Table 7.** Data validation screening checks description, rationale, and criteria for failure.

| DART Check Name | Check Description | Rationale | Fails if... | DART Smarts Action |
|---|---|---|---|---|
| Abundant species | Are typically abundant hydrocarbons (e.g., benzene, propane, n-butane, isoprene, n-hexane, ethylbenzene) present in each sample? | Missing species that are expected to be present in nearly every sample may indicate a problem. | Any of the listed species are missing or 0 | If two or more species listed are missing or 0, flag sample with code "AQ" |
| TNMOC | Is the TNMOC provided with every sample and is it a reasonable value? | TNMOC is the sum of the sample mass and if it is missing or 0, there is a problem. The sum of PAMS is a subset of TNMOC, so this value should be less than or equal to TNMOC. | • TNMOC missing or 0; or<br>• the sum of PAMS exceeds TNMOC | • Flag TNMOC and unidentified with code "AN".<br>• Flag TNMOC and Sum of PAMS with code "DA" |
| Variability | Is the measurement an outlier? | Typically, a compound with a concentration 3 to 4 standard deviations above its mean can be considered an outlier and further investigation is needed to determine if the outlier is representative of real ambient conditions | Species concentration exceeds the mean plus 4 times the standard deviation. | None |

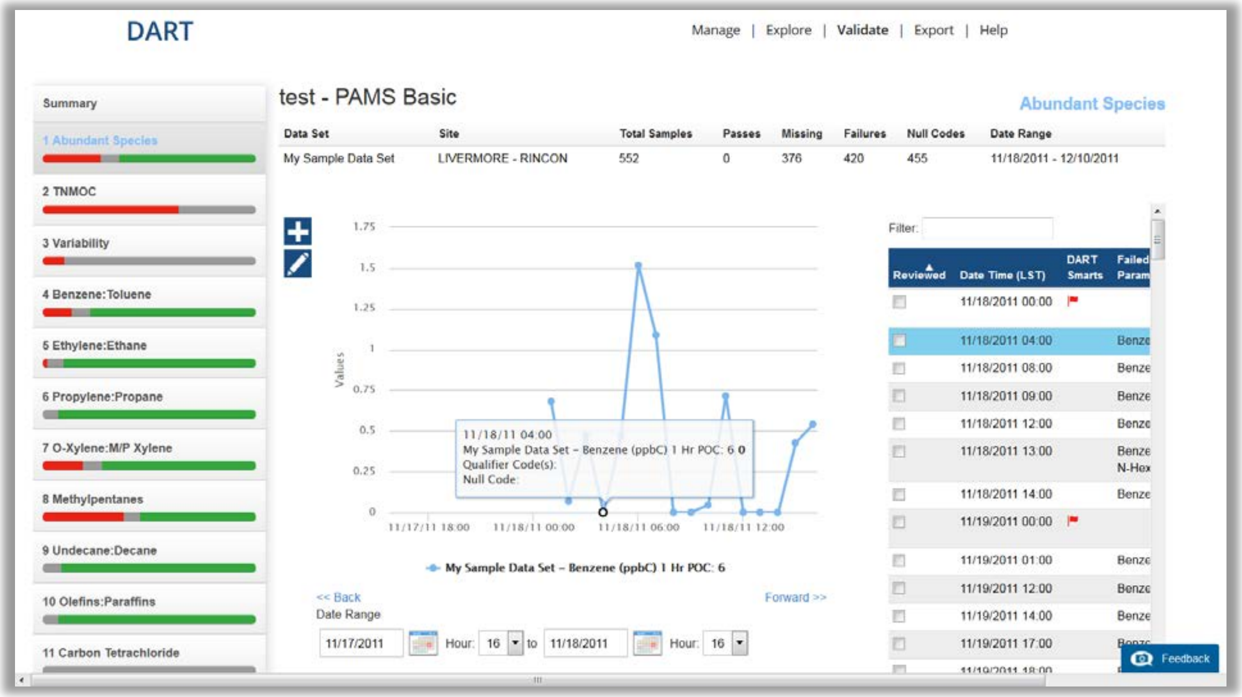
| DART Check Name | Check Description | Rationale | Fails if… | DART Smarts Action |
|---|---|---|---|---|
| Unidentified:TNMOC ratio | Does the total unidentified fraction of the sample exceed reasonable limits? Reasonable limits will be based on the site location. | A high unidentified fraction may indicate a problem. | Unidentified exceeds 50% of TNMOC | Flag unidentified with code "DA" |
| Sticking | Are consecutive identical measurements reported? | Several identical values may indicate an instrument issue. | Species has same non-zero value for 3 or more consecutive samples | Flag species with code "DA" |
| Benzene:toluene ratio | Are benzene concentrations greater than toluene? | Typically, unless there is a benzene source nearby, toluene concentrations exceed benzene. High benzene concentrations relative to toluene may indicate peak misidentification. | Benzene exceeds 0.2 ppbC and exceeds toluene | Flag both species with code "DA" |
| Ethylene:ethane ratio | Are ethylene concentrations great than ethane? | Typically ethane concentrations are greater than the much more reactive ethylene. High ethylene concentrations may indicate an instrument issue. | Ethylene exceeds 0.5 ppbC and exceeds ethane | Flag both species with code "DA" |
| Propylene:propane ratio | Are propylene concentrations greater than propane? | Typically propane concentrations are greater than the much more reactive propylene. High propylene concentrations may indicate an instrument issue. | Propylene exceeds 0.5 ppbC and exceeds propane | Flag both species with code "DA" |

| DART Check Name | Check Description | Rationale | Fails if… | DART Smarts Action |
|---|---|---|---|---|
| O-Xylene:M/P Xylene ratio | Are o-xylene concentrations greater than the sum of m- and p-xylenes? | These isomers typically correlate in ambient air. High o-xylene concentrations relative to the other two isomers may indicate an instrument issue. | o-xylene exceeds 0.5 ppbC and exceeds m-&p-xylenes | Flag xylenes with code "DA" |
| 2- and 3-Methylpentanes | Do these isomers correlate well? | Typically, these isomers correlate very well but because they may elute close together with each other and with other C6 isomers, they are sometimes misidentified. Lack of correlation may indicate peak misidentification. | 3-methylpentane exceeds 0.1 ppbC and exceeds 0.6 times 2-methylpentane | If 3-methylpentane exceeds 0.1 ppbC and exceeds 0.65 times 2-methylpentane, flag methylpentanes with code "BH" |
| Undecane:Decane | Are undecane concentrations greater than decane? Do concentrations of either or both of these species show high concentrations with slow decline over the subsequent hours? | Typically undecane, a possible indicator or diesel emissions, is present at very low concentrations. High concentrations or fall off in concentration may indicate sample contamination. These species are not always included in site target lists. | N-undecane exceeds 0.5 ppbC and exceeds n-decane | Flag both species with code "DA" |
| Olefins:Paraffins | Are olefin concentrations greater than paraffin concentrations? | Olefins are much more reactive than paraffins and are expected to be less abundant. High olefin concentrations may indicate an instrument problem. | Sum of olefins exceeds sum of paraffins | Flag both species sums with code "DA" |

| DART Check Name | Check Description | Rationale | Fails if… | DART Smarts Action |
|---|---|---|---|---|
| Carbon Tetrachloride | Are carbon tetrachloride concentrations significantly above or below global background levels? | If so, the entire sample may be suspect/invalid because this chemical compound should be at background levels within the precision of the measurement. | Carbon tetrachloride exceeds 0.16 ppb | Flag species with code "AQ" |
| Nighttime Isoprene | Are isoprene concentrations high overnight? | Isoprene is emitted by vegetation and concentrations typically correlate with temperature and sunlight. Isoprene is reactive as well. High concentrations at night have typically been identified as a result of instrument problems; however, in some cases, transport of high isoprene concentrations from upwind sites was confirmed. | Isoprene increases between 8 pm and 3 am local time | Flag isoprene with code "DA" |

Where DA = aberrant data, AQ = collection error, BH = Interference/co-elution/misidentification, AN = Machine Malfunction

Figures 2 through 13 show examples of output from automated screening and provide notes on interpretation.



test - PAMS Basic

| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| My Sample Data Set | LIVERMORE - RINCON | 552 | 0 | 376 | 420 | 455 | 11/18/2011 - 12/10/2011 |

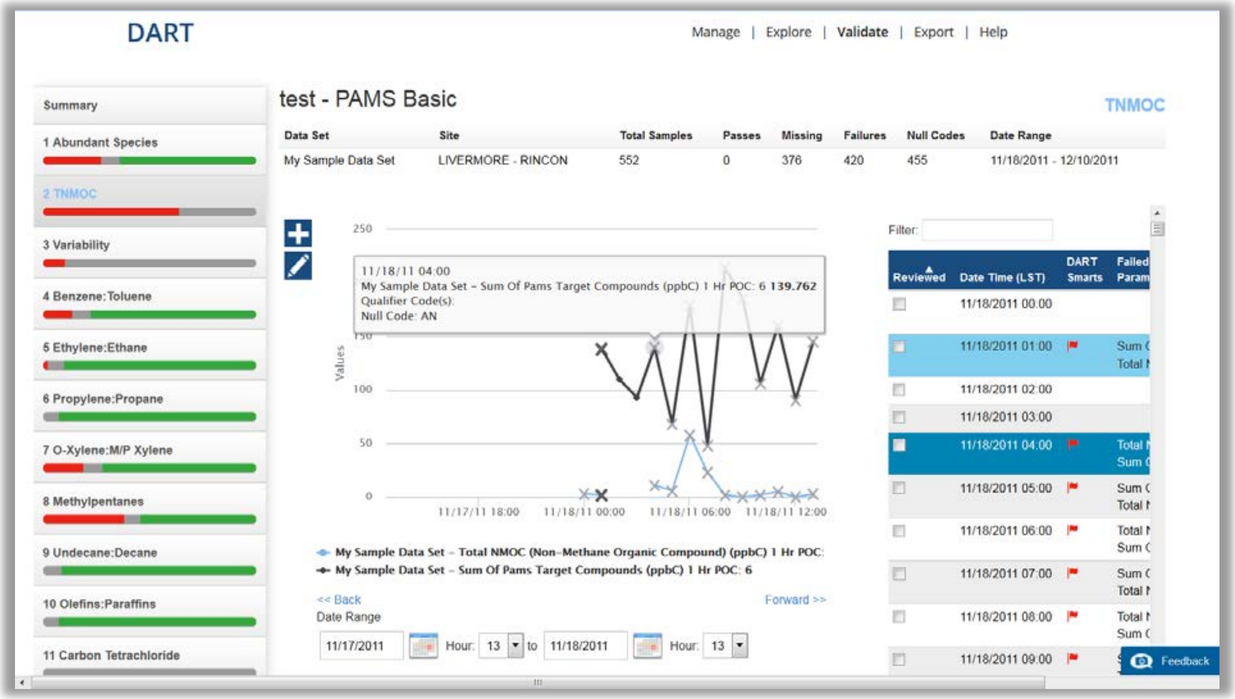
11/18/11 04:00
My Sample Data Set – Benzene (ppbC) 1 Hr POC: 6 **0**
Qualifier Code(s):
Null Code:

Date Range
11/17/2011 Hour: 16 to 11/18/2011 Hour: 16

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 11/18/2011 00:00 | 🚩 | |
| □ | 11/18/2011 04:00 | | Benze |
| □ | 11/18/2011 08:00 | | Benze |
| □ | 11/18/2011 09:00 | | Benze |
| □ | 11/18/2011 12:00 | | Benze |
| □ | 11/18/2011 13:00 | | Benze<br>N-Hex |
| □ | 11/18/2011 14:00 | | Benze |
| □ | 11/19/2011 00:00 | 🚩 | |
| □ | 11/19/2011 01:00 | | Benze |
| □ | 11/19/2011 12:00 | | Benze |
| □ | 11/19/2011 14:00 | | Benze |
| □ | 11/19/2011 17:00 | | Benze |
| □ | 11/19/2011 18:00 | | |

Figure 2. An abundant species check. This example shows benzene concentrations of zero in many samples. For this data set, from an urban site, this is unusual. If all species concentrations in these samples are low, the data are likely valid.



test - PAMS Basic

| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| My Sample Data Set | LIVERMORE - RINCON | 552 | 0 | 376 | 420 | 455 | 11/18/2011 - 12/10/2011 |

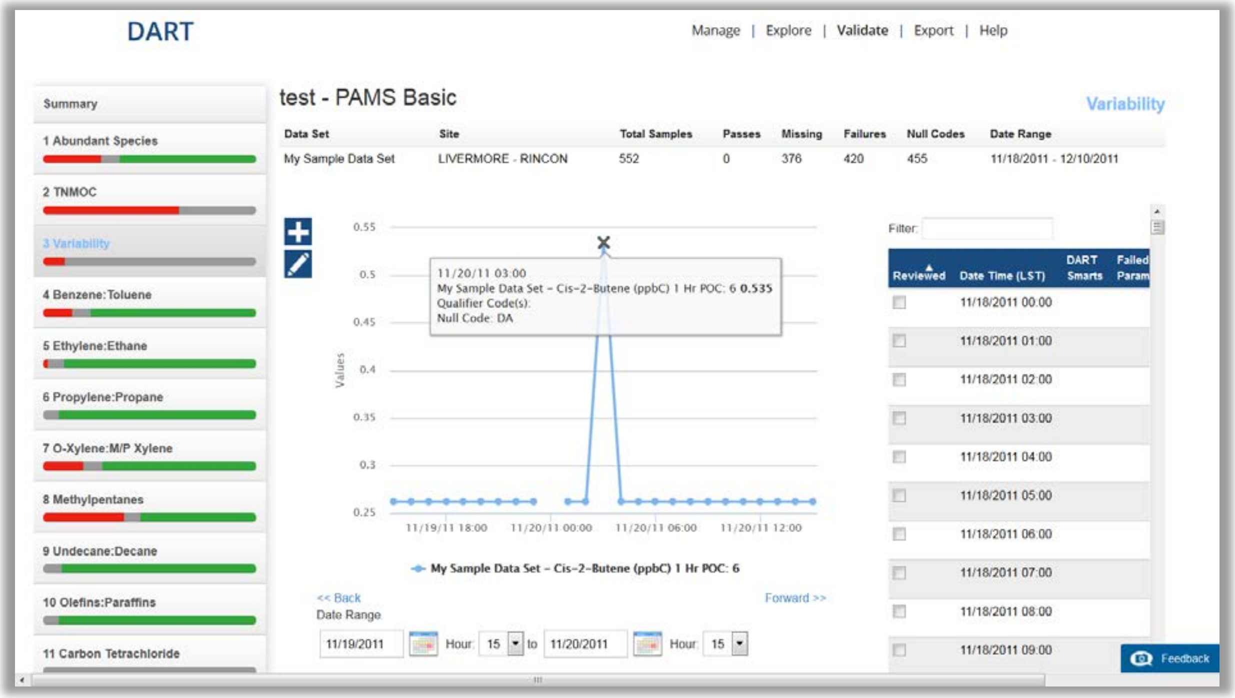
11/18/11 04:00
My Sample Data Set – Sum Of Pams Target Compounds (ppbC) 1 Hr POC: 6 **139.762**
Qualifier Code(s):
Null Code: AN

Date Range
11/17/2011 Hour: 13 to 11/18/2011 Hour: 13

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 11/18/2011 00:00 | | |
| □ | 11/18/2011 01:00 | 🟥 | Sum (<br>Total N |
| □ | 11/18/2011 02:00 | | |
| □ | 11/18/2011 03:00 | | |
| □ | 11/18/2011 04:00 | 🟥 | Total N<br>Sum ( |
| □ | 11/18/2011 05:00 | 🟥 | Sum (<br>Total N |
| □ | 11/18/2011 06:00 | 🟥 | Total N<br>Sum ( |
| □ | 11/18/2011 07:00 | 🟥 | Sum (<br>Total N |
| □ | 11/18/2011 08:00 | 🟥 | Total N<br>Sum ( |
| □ | 11/18/2011 09:00 | 🟥 | S |

**Figure 3.** A **TNMOC check**. In this example, the Sum of PAMS species concentrations is greater than TNMOC concentrations. Since TNMOC should be equal to the Sum of PAMS species plus unidentified mass, something is wrong with the TNMOC values. Sometimes this check shows failures because the units for TNMOC are incorrect and should be checked, or the TNMOC was not reported.



test - PAMS Basic

| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| My Sample Data Set | LIVERMORE - RINCON | 552 | 0 | 376 | 420 | 455 | 11/18/2011 - 12/10/2011 |

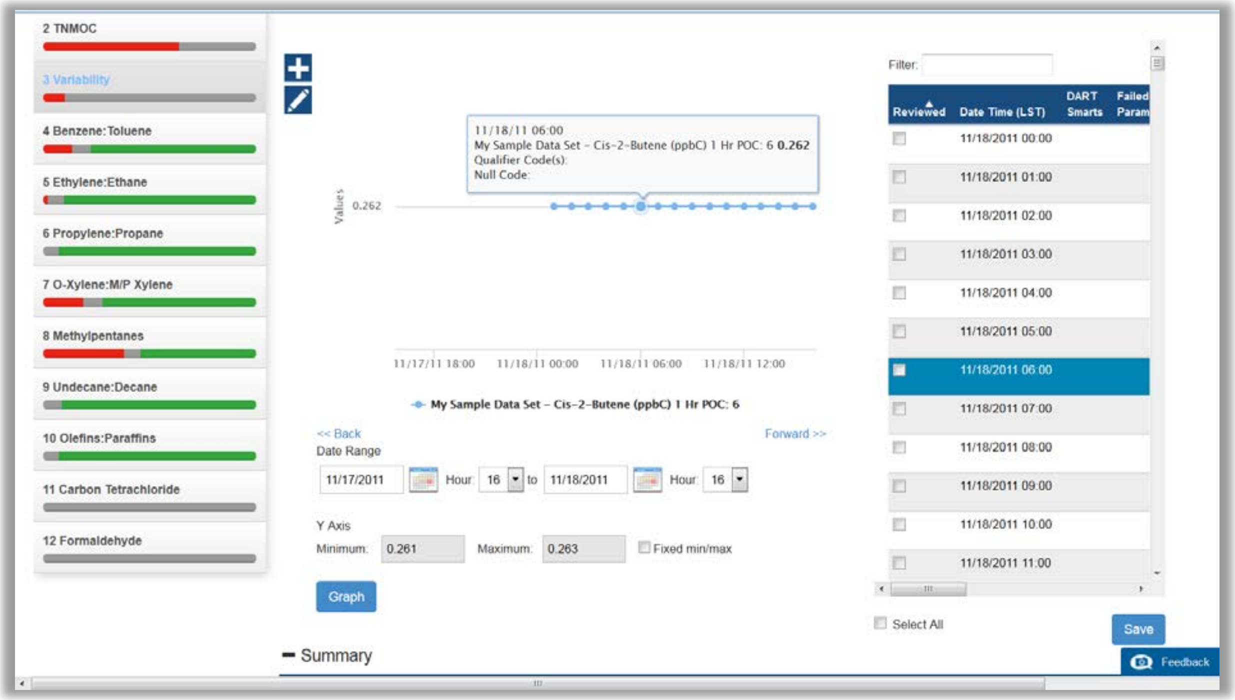
11/20/11 03:00
My Sample Data Set – Cis-2-Butene (ppbC) 1 Hr POC: 6 **0.535**
Qualifier Code(s):
Null Code: DA

Date Range
11/19/2011 Hour: 15 to 11/20/2011 Hour: 15

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 11/18/2011 00:00 | | |
| □ | 11/18/2011 01:00 | | |
| □ | 11/18/2011 02:00 | | |
| □ | 11/18/2011 03:00 | | |
| □ | 11/18/2011 04:00 | | |
| □ | 11/18/2011 05:00 | | |
| □ | 11/18/2011 06:00 | | |
| □ | 11/18/2011 07:00 | | |
| □ | 11/18/2011 08:00 | | |
| □ | 11/18/2011 09:00 | | |

**Figure 4.** A **variability check**. In this example, a concentration spike for cis-2-butene clearly stands out in this data set (all other values were at 0.26 ppbC). Concentrations are low for cis-2-butene in this data set. Investigating other olefins for this sample may help in identifying the reasonableness of the cis-2-butene.



11/18/11 06:00
My Sample Data Set - Cis-2-Butene (ppbC) 1 Hr POC: 6 **0.262**
Qualifier Code(s):
Null Code:

My Sample Data Set – Cis-2-Butene (ppbC) 1 Hr POC: 6

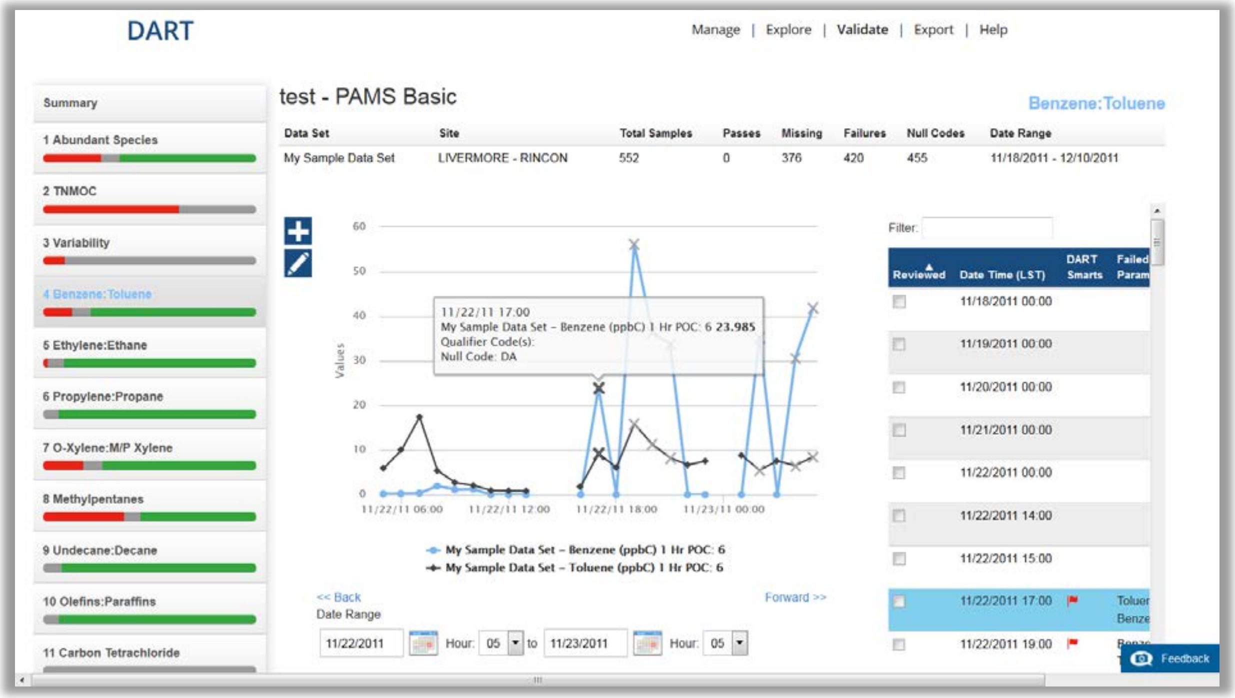
Date Range
11/17/2011 Hour: 16 to 11/18/2011 Hour: 16

Y Axis
Minimum: 0.261 Maximum: 0.263 □ Fixed min/max

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 11/18/2011 00:00 | | |
| □ | 11/18/2011 01:00 | | |
| □ | 11/18/2011 02:00 | | |
| □ | 11/18/2011 03:00 | | |
| □ | 11/18/2011 04:00 | | |
| □ | 11/18/2011 05:00 | | |
| □ | 11/18/2011 06:00 | | |
| □ | 11/18/2011 07:00 | | |
| □ | 11/18/2011 08:00 | | |
| □ | 11/18/2011 09:00 | | |
| □ | 11/18/2011 10:00 | | |
| □ | 11/18/2011 11:00 | | |

**Figure 5.** Example of **stuck values** for cis-2-butene. Looking at the data set may show that this is the instrument's detection limit for cis-2-butene.



test - PAMS Basic

| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| My Sample Data Set | LIVERMORE - RINCON | 552 | 0 | 376 | 420 | 455 | 11/18/2011 - 12/10/2011 |

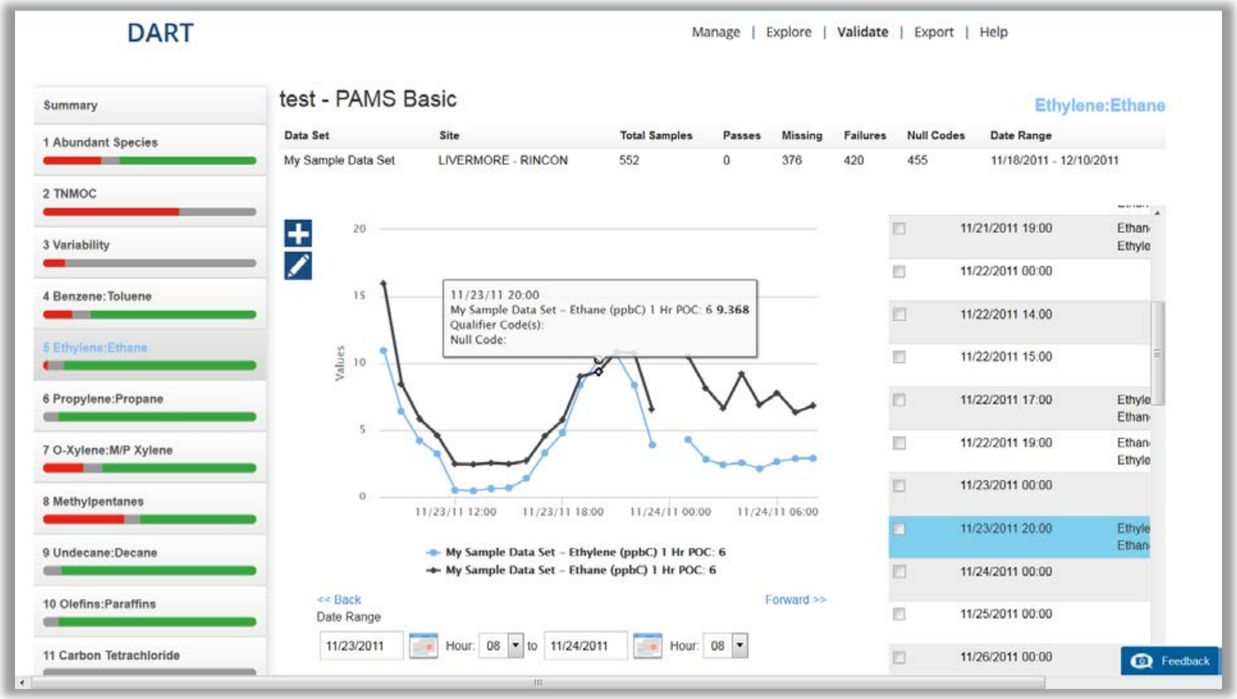
11/22/11 17:00
My Sample Data Set – Benzene (ppbC) 1 Hr POC: 6 **23.985**
Qualifier Code(s):
Null Code: DA

Date Range
11/22/2011 Hour: 05 to 11/23/2011 Hour: 05

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 11/18/2011 00:00 | | |
| □ | 11/19/2011 00:00 | | |
| □ | 11/20/2011 00:00 | | |
| □ | 11/21/2011 00:00 | | |
| □ | 11/22/2011 00:00 | | |
| □ | 11/22/2011 14:00 | | |
| □ | 11/22/2011 15:00 | | |
| □ | 11/22/2011 17:00 | 🚩 | Toluer<br>Benze |
| □ | 11/22/2011 19:00 | 🚩 | Benze |

Figure 6. A benzene:toluene ratio check. In this example, the benzene concentration is much higher than the toluene concentration. Unless there is a benzene source nearby, which is very rare, this is not a typical ambient relationship between these species, and misidentification is suspected. Note: concentrations are high in these cases, indicating that method detection limit (MDL) is likely not an issue.



test - PAMS Basic

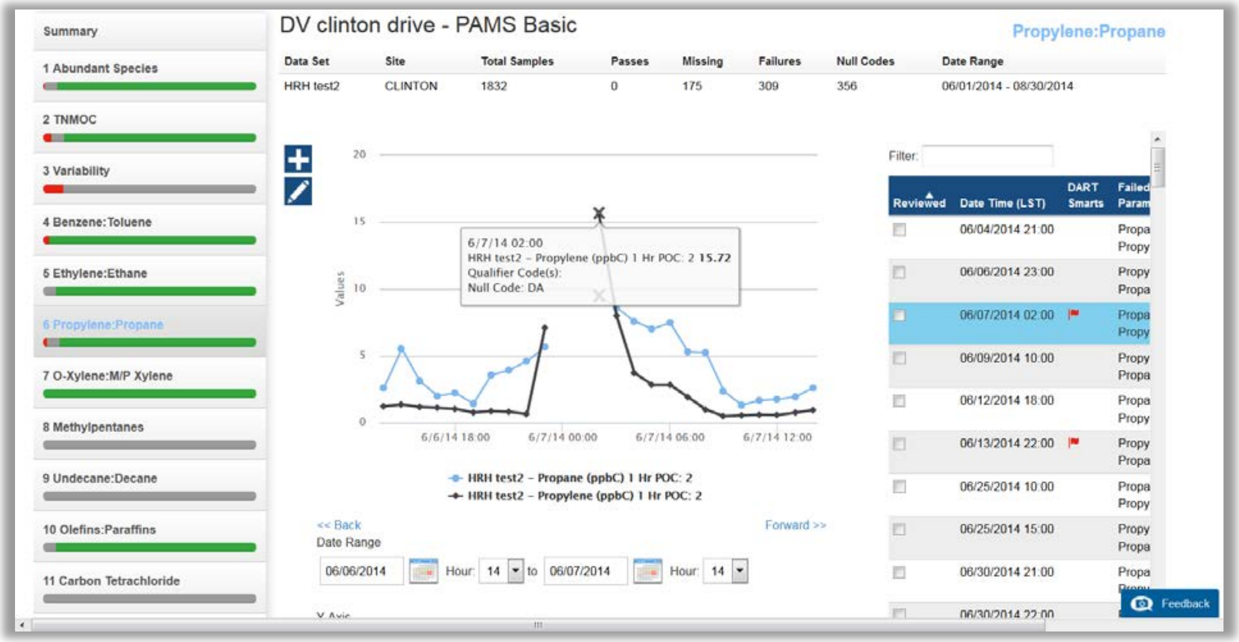
| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| My Sample Data Set | LIVERMORE - RINCON | 552 | 0 | 376 | 420 | 455 | 11/18/2011 - 12/10/2011 |

11/23/11 20:00
My Sample Data Set – Ethane (ppbC) 1 Hr POC: 6 **9.368**
Qualifier Code(s):
Null Code:

Date Range
11/23/2011 Hour: 08 to 11/24/2011 Hour: 08

| | | |
|---|---|---|
| □ | 11/21/2011 19:00 | Ethan<br>Ethyle |
| □ | 11/22/2011 00:00 | |
| □ | 11/22/2011 14:00 | |
| □ | 11/22/2011 15:00 | |
| □ | 11/22/2011 17:00 | Ethyle<br>Ethan |
| □ | 11/22/2011 19:00 | Ethan<br>Ethyle |
| □ | 11/23/2011 00:00 | |
| □ | 11/23/2011 20:00 | Ethyle<br>Ethan |
| □ | 11/24/2011 00:00 | |
| □ | 11/25/2011 00:00 | |
| □ | 11/26/2011 00:00 | |

**Figure 7.** An **ethylene:ethane ratio check**. In this example, the ethylene concentration is greater than ethane concentrations. However, the concentrations are very close, possibly within the precision of the instrument, and at user discretion the data may be deemed valid.



DV clinton drive - PAMS Basic

| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| HRH test2 | CLINTON | 1832 | 0 | 175 | 309 | 356 | 06/01/2014 - 08/30/2014 |

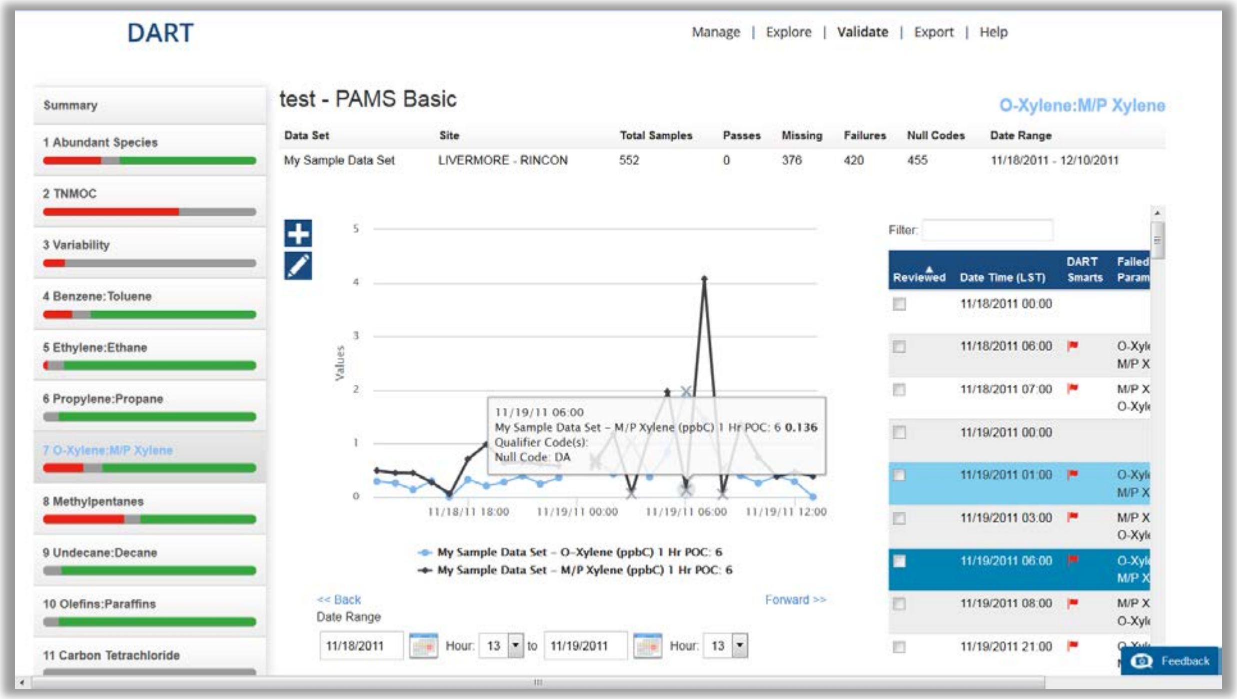
6/7/14 02:00
HRH test2 – Propylene (ppbC) 1 Hr POC: 2 **15.72**
Qualifier Code(s):
Null Code: DA

Date Range
06/06/2014 Hour: 14 to 06/07/2014 Hour: 14

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 06/04/2014 21:00 | | Propa<br>Propy |
| □ | 06/06/2014 23:00 | | Propy<br>Propa |
| □ | 06/07/2014 02:00 | 🟥 | Propa<br>Propy |
| □ | 06/09/2014 10:00 | | Propy<br>Propa |
| □ | 06/12/2014 18:00 | | Propa<br>Propy |
| □ | 06/13/2014 22:00 | 🟥 | Propy<br>Propa |
| □ | 06/25/2014 10:00 | | Propa<br>Propy |
| □ | 06/25/2014 15:00 | | Propy<br>Propa |
| □ | 06/30/2014 21:00 | | Propa<br>Propy |
| □ | 06/30/2014 22:00 | | |

**Figure 8.** A **propylene:propane ratio check.** Propylene is typically lower in concentration than propane because of its higher photochemical reactivity. The high propylene concentrations before and after a data gap may indicate that the samples before and after the gap are suspect. Inspecting other olefins may help determine whether these samples are valid or not.



test - PAMS Basic

O-Xylene:M/P Xylene

| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| My Sample Data Set | LIVERMORE - RINCON | 552 | 0 | 376 | 420 | 455 | 11/18/2011 - 12/10/2011 |

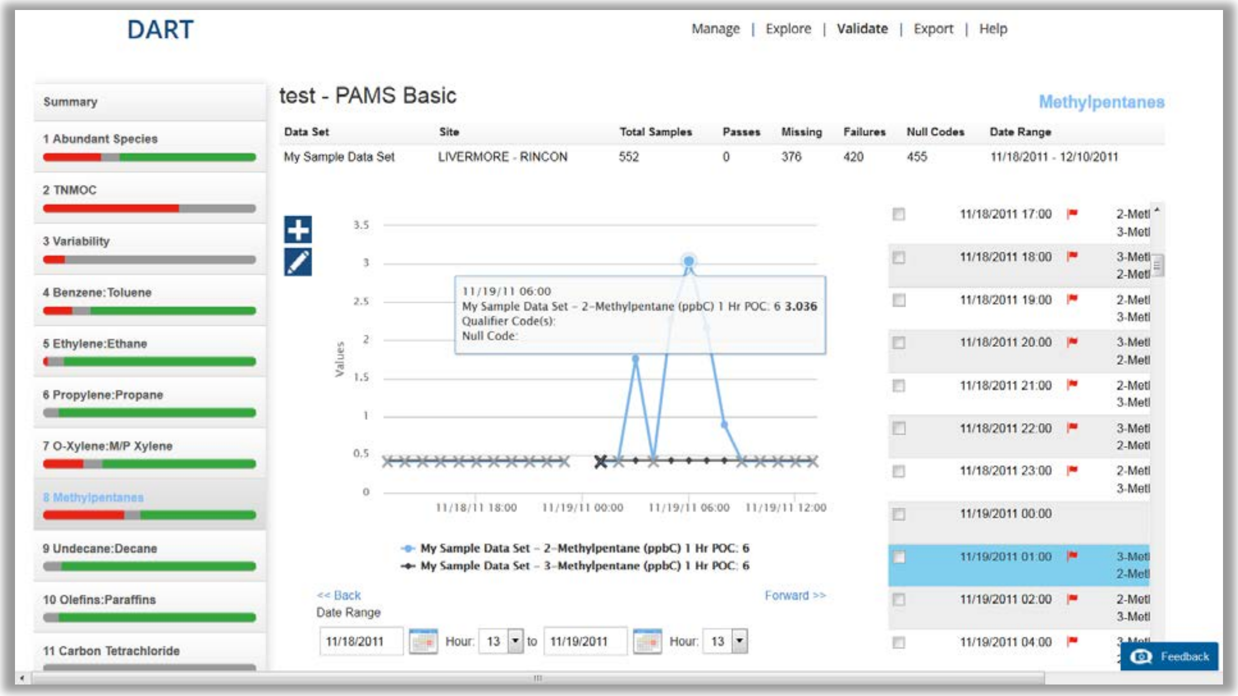
11/19/11 06:00
My Sample Data Set – M/P Xylene (ppbC) 1 Hr POC: 6 **0.136**
Qualifier Code(s):
Null Code: DA

Date Range
11/18/2011 Hour: 13 to 11/19/2011 Hour: 13

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 11/18/2011 00:00 | | |
| □ | 11/18/2011 06:00 | 🟥 | O-Xyl<br>M/P X |
| □ | 11/18/2011 07:00 | 🟥 | M/P X<br>O-Xyl |
| □ | 11/19/2011 00:00 | | |
| □ | 11/19/2011 01:00 | 🟥 | O-Xyl<br>M/P X |
| □ | 11/19/2011 03:00 | 🟥 | M/P X<br>O-Xyl |
| □ | 11/19/2011 06:00 | 🟥 | O-Xyl<br>M/P X |
| □ | 11/19/2011 08:00 | 🟥 | M/P X<br>O-Xyl |
| □ | 11/19/2011 21:00 | 🟥 | O-Xyl |

**Figure 9.** A **xylenes check**. In this example, the sum of m-&p-xylenes concentration is lower than o-xylene concentrations. This finding is unusual unless there is an o-xylene, but not m- and p-xylenes, source nearby (which is rare). These samples indicate a problem with the xylenes.



test - PAMS Basic

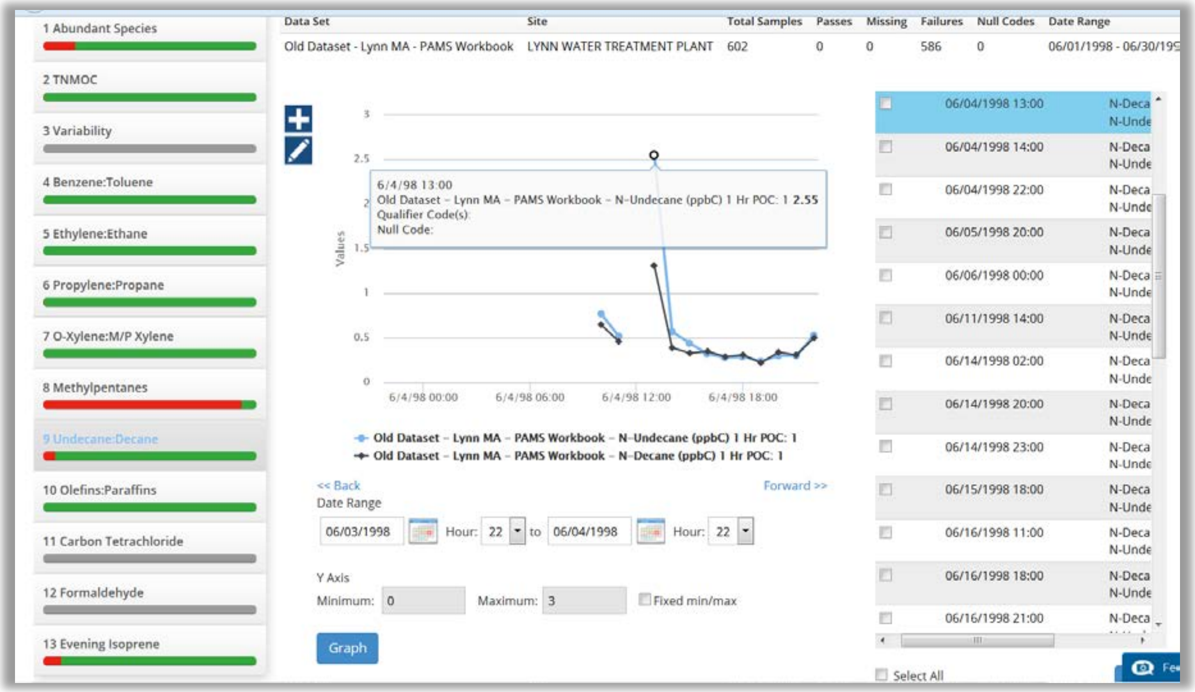
| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| My Sample Data Set | LIVERMORE - RINCON | 552 | 0 | 376 | 420 | 455 | 11/18/2011 - 12/10/2011 |

11/19/11 06:00
My Sample Data Set – 2-Methylpentane (ppbC) 1 Hr POC: 6 **3.036**
Qualifier Code(s):
Null Code:

Date Range
11/18/2011 Hour: 13 to 11/19/2011 Hour: 13

|  |  |  |  |
|---|---|---|---|
| □ | 11/18/2011 17:00 | 🟥 | 2-Metl<br>3-Metl |
| □ | 11/18/2011 18:00 | 🟥 | 3-Metl<br>2-Metl |
| □ | 11/18/2011 19:00 | 🟥 | 2-Metl<br>3-Metl |
| □ | 11/18/2011 20:00 | 🟥 | 3-Metl<br>2-Metl |
| □ | 11/18/2011 21:00 | 🟥 | 2-Metl<br>3-Metl |
| □ | 11/18/2011 22:00 | 🟥 | 3-Metl<br>2-Metl |
| □ | 11/18/2011 23:00 | 🟥 | 2-Metl<br>3-Metl |
| □ | 11/19/2011 00:00 |  |  |
| □ | 11/19/2011 01:00 | 🟥 | 3-Metl<br>2-Metl |
| □ | 11/19/2011 02:00 | 🟥 | 2-Metl<br>3-Metl |
| □ | 11/19/2011 04:00 | 🟥 | 3-Metl |

**Figure 10.** A **methylpentanes check**. 2-methylpentane and 3-methylpentane are typically abundant and correlate well with each other. In this example, the samples appear to have a method detection issue for several days and some spikes in 2-methylpentane, but not in 3-methylpentane, concentrations indicating an instrument problem.



| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| Old Dataset - Lynn MA - PAMS Workbook | LYNN WATER TREATMENT PLANT | 602 | 0 | 0 | 586 | 0 | 06/01/1998 - 06/30/199 |

6/4/98 13:00
Old Dataset – Lynn MA – PAMS Workbook – N-Undecane (ppbC) 1 Hr POC: 1 **2.55**
Qualifier Code(s):
Null Code:

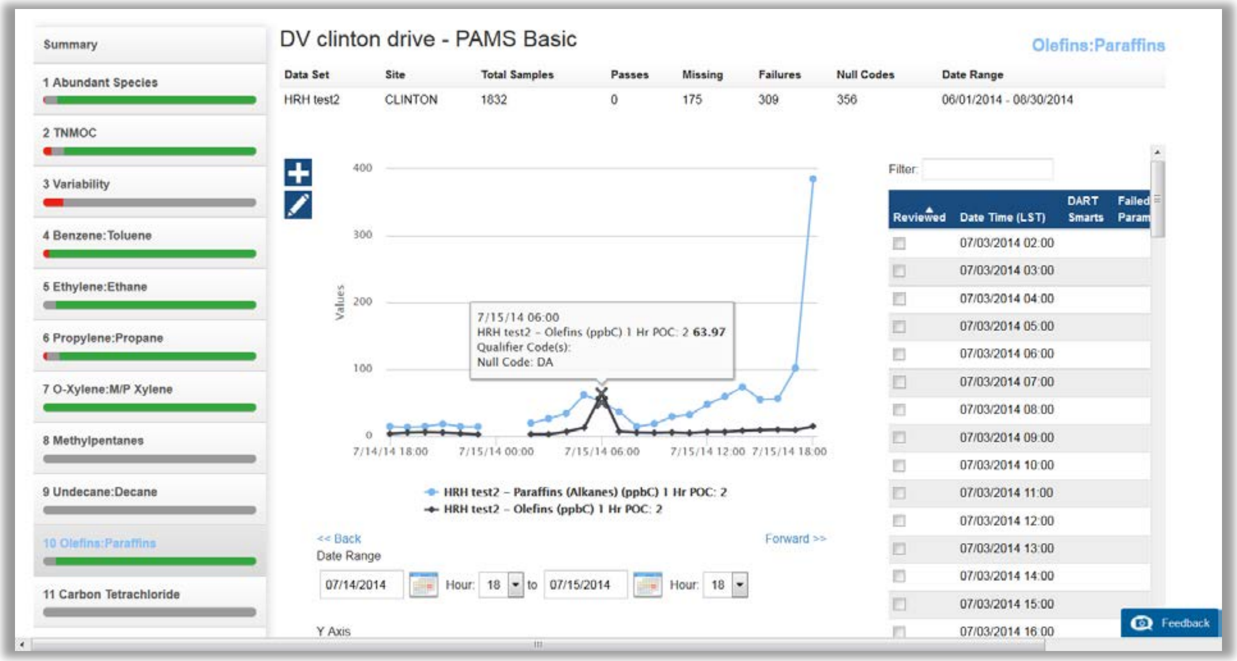
Date Range
06/03/1998 Hour: 22 to 06/04/1998 Hour: 22

Y Axis
Minimum: 0 Maximum: 3 □ Fixed min/max

| | | |
|---|---|---|
| □ | 06/04/1998 13:00 | N-Deca<br>N-Unde |
| □ | 06/04/1998 14:00 | N-Deca<br>N-Unde |
| □ | 06/04/1998 22:00 | N-Deca<br>N-Unde |
| □ | 06/05/1998 20:00 | N-Deca<br>N-Unde |
| □ | 06/06/1998 00:00 | N-Deca<br>N-Unde |
| □ | 06/11/1998 14:00 | N-Deca<br>N-Unde |
| □ | 06/14/1998 02:00 | N-Deca<br>N-Unde |
| □ | 06/14/1998 20:00 | N-Deca<br>N-Unde |
| □ | 06/14/1998 23:00 | N-Deca<br>N-Unde |
| □ | 06/15/1998 18:00 | N-Deca<br>N-Unde |
| □ | 06/16/1998 11:00 | N-Deca<br>N-Unde |
| □ | 06/16/1998 18:00 | N-Deca<br>N-Unde |
| □ | 06/16/1998 21:00 | N-Deca |

□ Select All

**Figure 11.** An **undecane:n-decane ratio check**. Undecane concentrations are typically lower than n-decane concentrations. In this example, there are two things to notice. First, the undecane is higher than decane for a few hours. Second, there is a gap in the data, followed by high concentrations and a decline in concentration over time. During this time period, the calibration gases had high concentrations and the heavier VOCs would "stick" in the GC and elute for the next few hours. This is an example of calibration gas carryover rather than real ambient data. Both the undecane and n-decane are invalid during the carryover period.



DV clinton drive - PAMS Basic

Olefins:Paraffins

| Data Set | Site | Total Samples | Passes | Missing | Failures | Null Codes | Date Range |
|---|---|---|---|---|---|---|---|
| HRH test2 | CLINTON | 1832 | 0 | 175 | 309 | 356 | 06/01/2014 - 08/30/2014 |

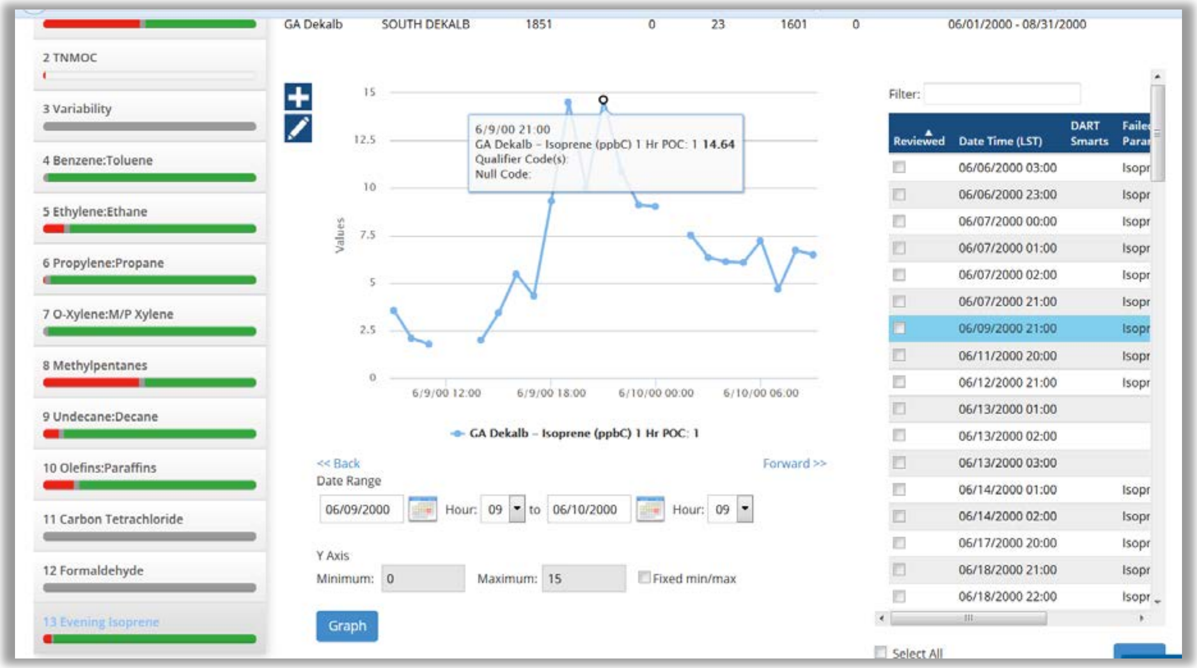
7/15/14 06:00
HRH test2 – Olefins (ppbC) 1 Hr POC: 2 **63.97**
Qualifier Code(s):
Null Code: DA

Date Range
07/14/2014 Hour: 18 to 07/15/2014 Hour: 18

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Failed Param |
|---|---|---|---|
| □ | 07/03/2014 02:00 | | |
| □ | 07/03/2014 03:00 | | |
| □ | 07/03/2014 04:00 | | |
| □ | 07/03/2014 05:00 | | |
| □ | 07/03/2014 06:00 | | |
| □ | 07/03/2014 07:00 | | |
| □ | 07/03/2014 08:00 | | |
| □ | 07/03/2014 09:00 | | |
| □ | 07/03/2014 10:00 | | |
| □ | 07/03/2014 11:00 | | |
| □ | 07/03/2014 12:00 | | |
| □ | 07/03/2014 13:00 | | |
| □ | 07/03/2014 14:00 | | |
| □ | 07/03/2014 15:00 | | |
| □ | 07/03/2014 16:00 | | |

Figure 12. An **olefins:paraffins ratio check**. The sum of olefins is typically lower than the sum of paraffins. In this sample, the next step would be to investigate whether a spike in a single olefin is causing the high total olefin concentration. This site is near industrial emissions, and a high olefin spike may be real.



GA Dekalb SOUTH DEKALB 1851 0 23 1601 0 06/01/2000 - 08/31/2000

2 TNMOC

3 Variability

4 Benzene:Toluene

5 Ethylene:Ethane

6 Propylene:Propane

7 O-Xylene:M/P Xylene

8 Methylpentanes

9 Undecane:Decane

10 Olefins:Paraffins

11 Carbon Tetrachloride

12 Formaldehyde

13 Evening Isoprene

6/9/00 21:00
GA Dekalb – Isoprene (ppbC) 1 Hr POC: 1 **14.64**
Qualifier Code(s):
Null Code:

<< Back Forward >>

Date Range

06/09/2000 Hour: 09 to 06/10/2000 Hour: 09

Y Axis

Minimum: 0 Maximum: 15 □ Fixed min/max

Graph

Filter:

| Reviewed | Date Time (LST) | DART Smarts | Faile<br>Para |
|---|---|---|---|
| □ | 06/06/2000 03:00 | | Isopr |
| □ | 06/06/2000 23:00 | | Isopr |
| □ | 06/07/2000 00:00 | | Isopr |
| □ | 06/07/2000 01:00 | | Isopr |
| □ | 06/07/2000 02:00 | | Isopr |
| □ | 06/07/2000 21:00 | | Isopr |
| □ | 06/09/2000 21:00 | | Isopr |
| □ | 06/11/2000 20:00 | | Isopr |
| □ | 06/12/2000 21:00 | | Isopr |
| □ | 06/13/2000 01:00 | | |
| □ | 06/13/2000 02:00 | | |
| □ | 06/13/2000 03:00 | | |
| □ | 06/14/2000 01:00 | | Isopr |
| □ | 06/14/2000 02:00 | | Isopr |
| □ | 06/17/2000 20:00 | | Isopr |
| □ | 06/18/2000 21:00 | | Isopr |
| □ | 06/18/2000 22:00 | | Isopr |

□ Select All

**Figure 13.** Example of unusual **high nighttime isoprene** concentrations. Typically, isoprene concentrations are highest when solar radiation and temperature are high. Because solar radiation is zero at night, these high concentrations may need further investigation.

Additional exploration of your data can be made using time series, scatter, and fingerprint plots in the **Explore** menu. Figures 14 through 20 illustrate additional assessment of data issues identified in data screening.

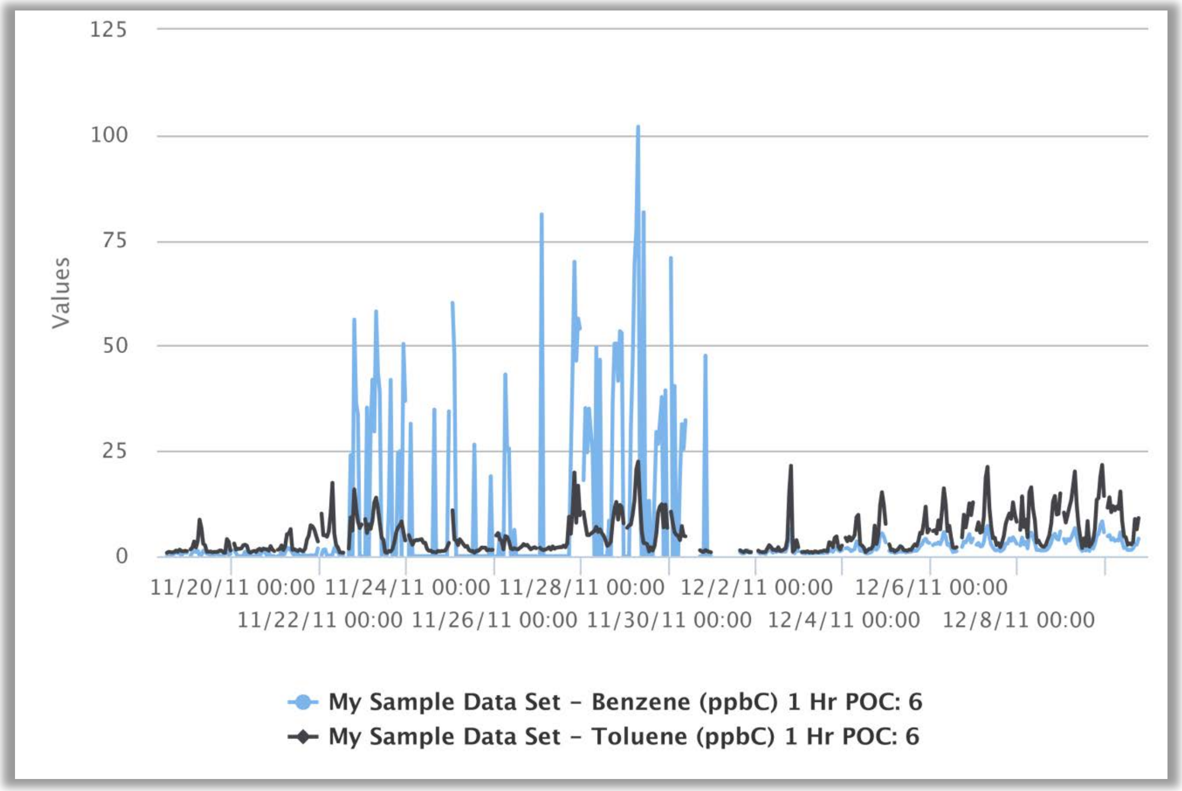


Figure 14. A time series plot used to further explore a period of high benzene concentrations relative to toluene. In this example, it appears that there may have been an instrument problem for a few days. In the periods before and after the benzene concentration spikes, the benzene concentrations are as expected relative to toluene. These are also very high benzene concentrations.

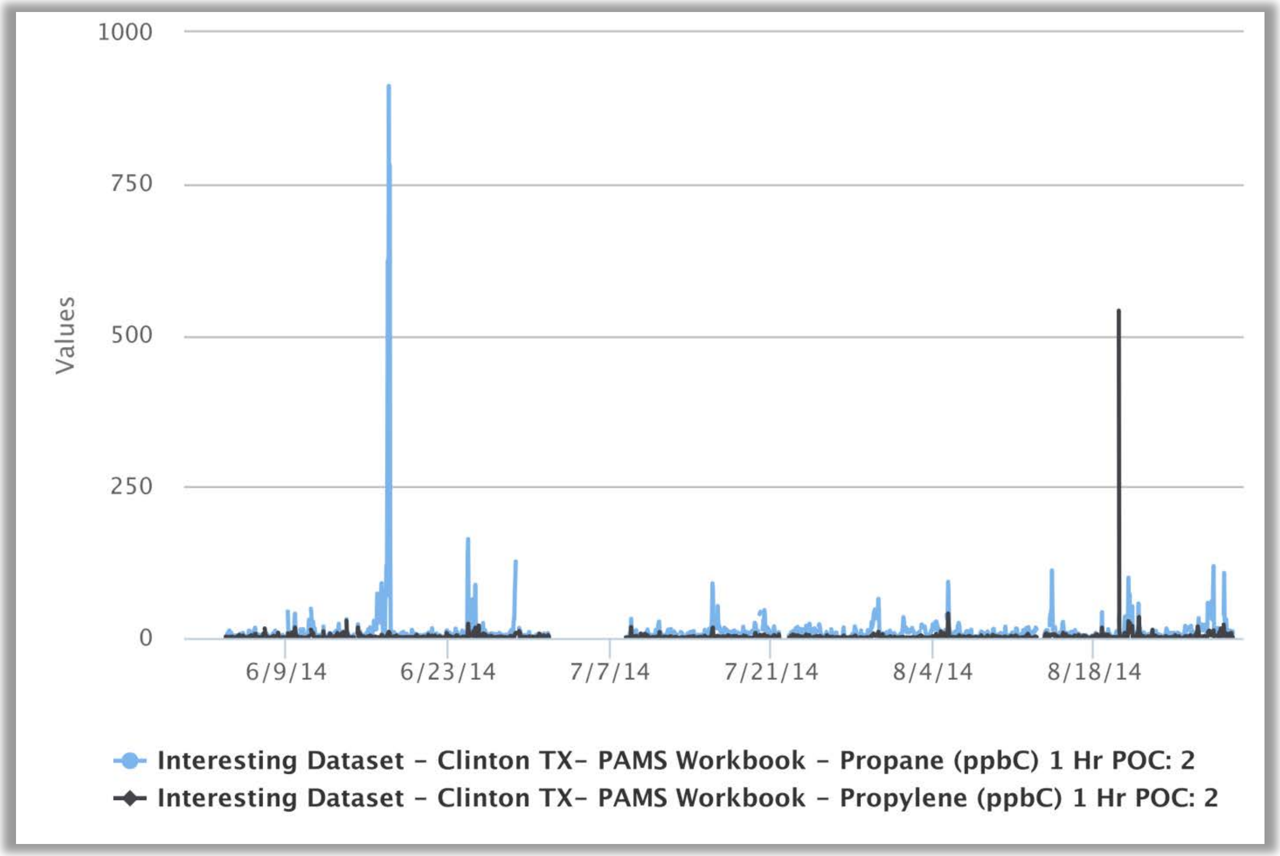


**Figure 15.** A time series plot used to further explore high propane and propylene concentrations. In this example, periodic spikes in concentration are seen for both VOCs. These spikes are consistent with fugitive emissions from nearby industrial sources.

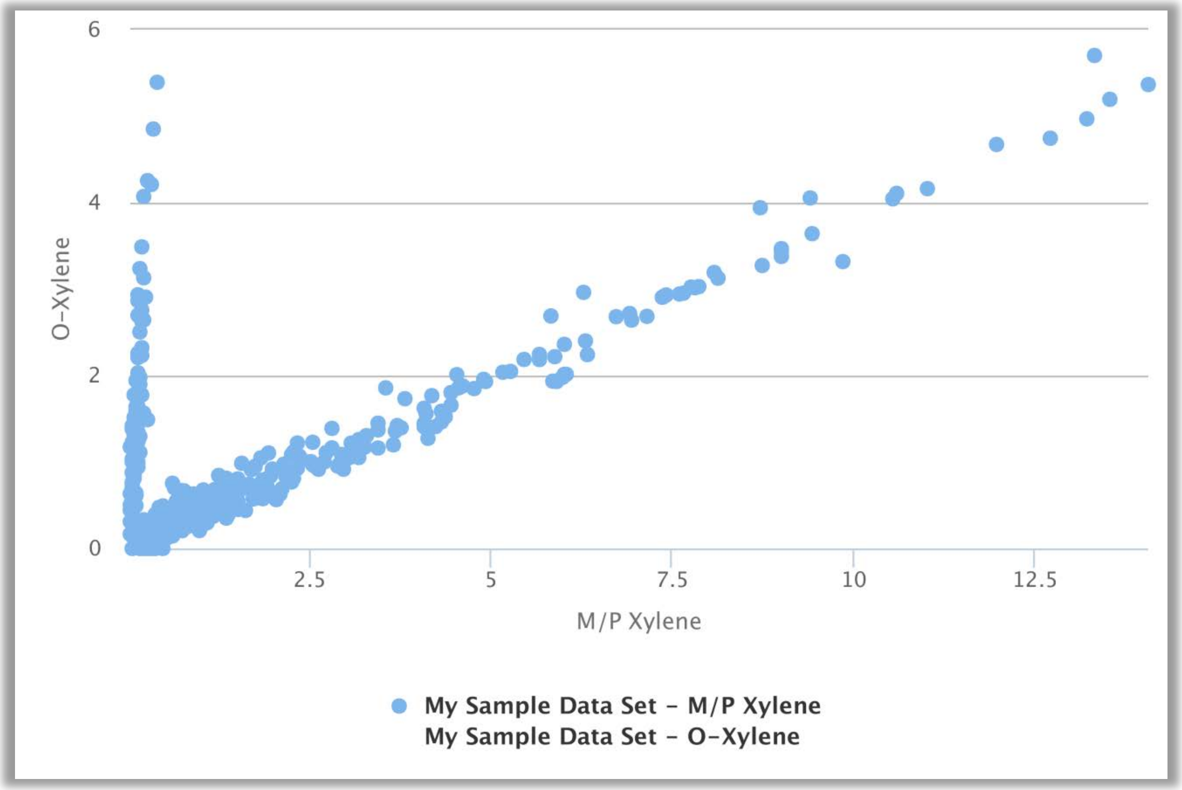


**Figure 16.** A scatter plot used to further explore xylene data problems. This plot is a classic example of peak misidentification. We expect a good correlation among the xylene isomers. In this case, the o-xylene data are consistently low in concentration relative to m- & p- xylenes, and the wrong peak in the chromatograph is likely being quantified as o-xylene. Further checking that the m- & p- xylenes correlate reasonably well with toluene would be a good next step to verify their validity.

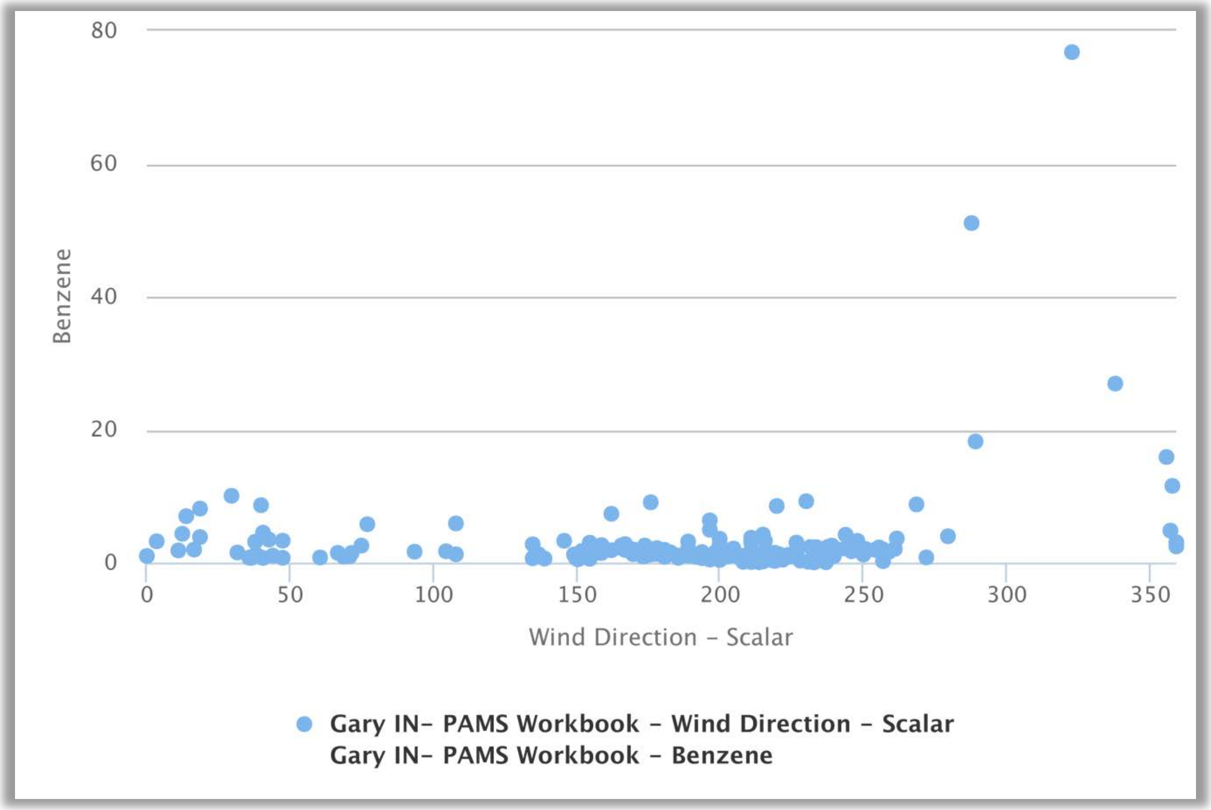
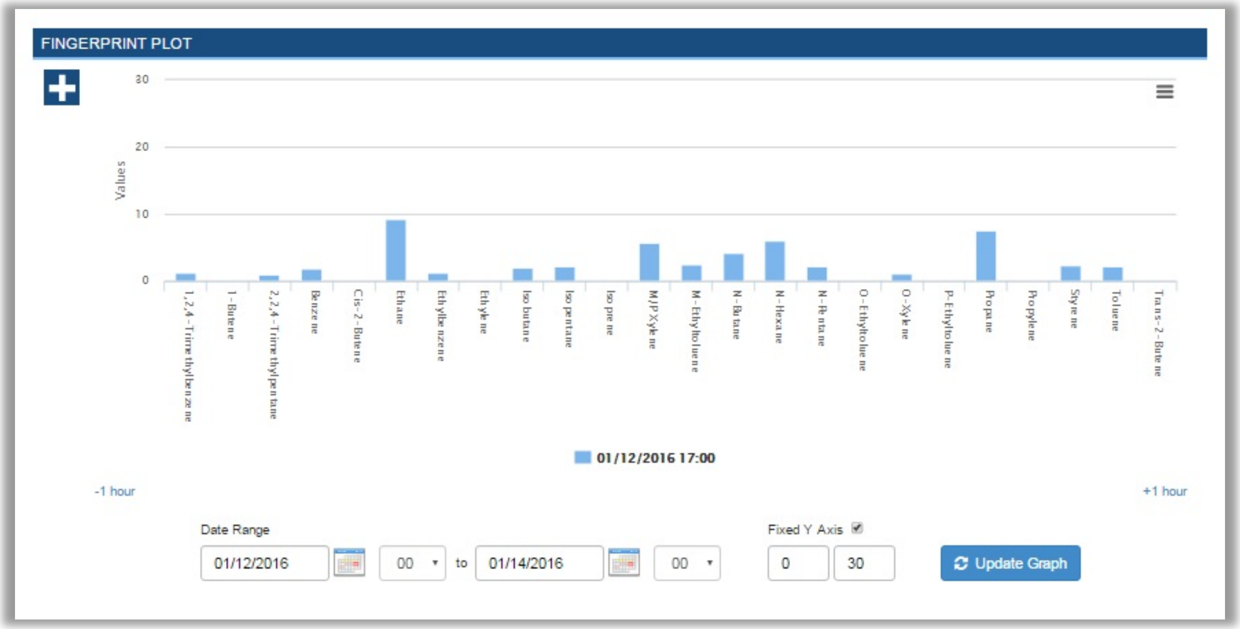


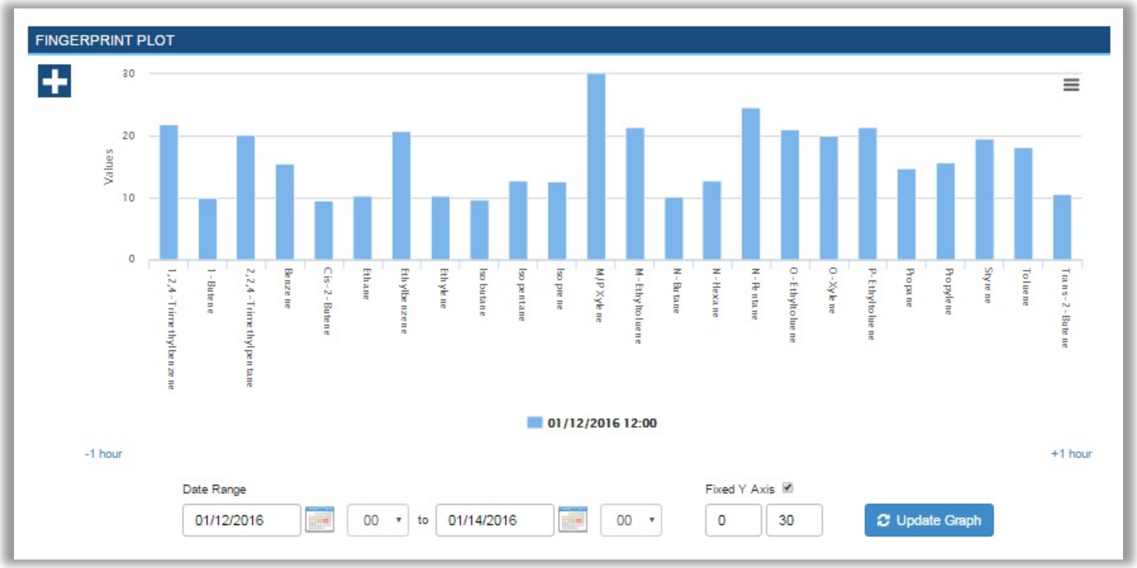
**Figure 17.** A scatter plot of benzene concentrations and wind direction used to further investigate high benzene concentrations relative to toluene. The highest benzene concentrations are typically coming from the north of the site. Emission inventory inspection showed a source of coke oven emissions to the north, which include benzene but not toluene, providing a reasonable explanation for these data (and helping prove their validity).



Date Range 01/12/2016 00 to 01/14/2016 00

Fixed Y Axis ☑ 0 30

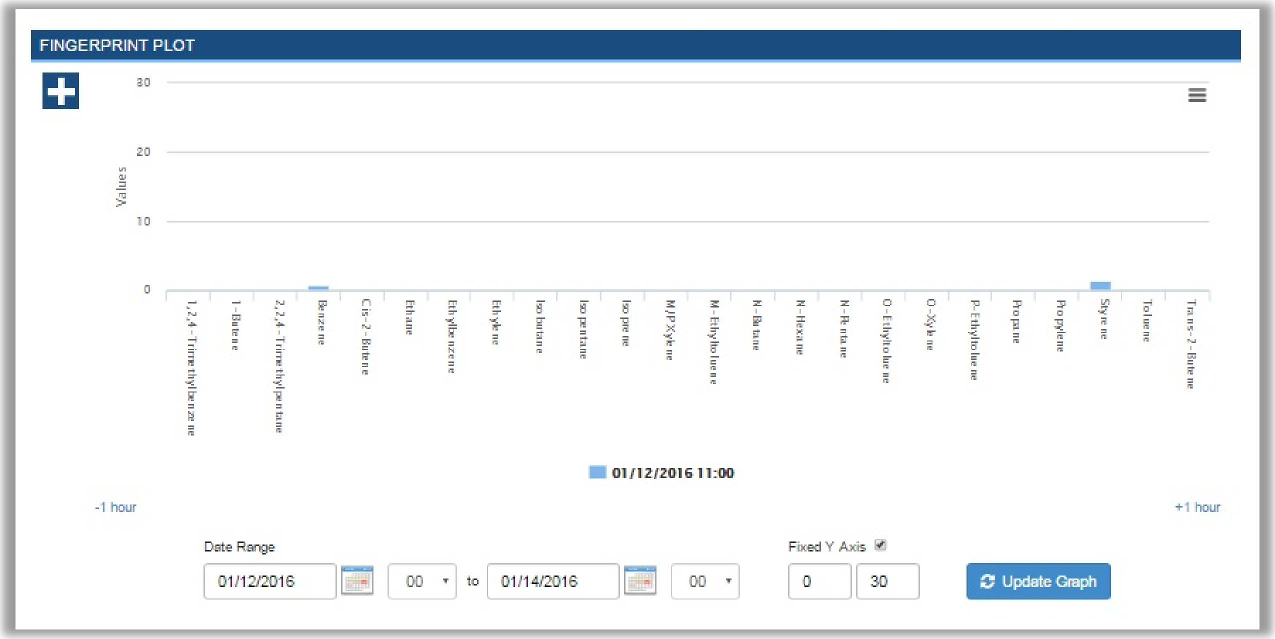
Figure 18. Typical fingerprint example showing a range of concentrations with typically abundant species present at higher concentrations than less abundant species.



Date Range 01/12/2016 00 to 01/14/2016 00

Fixed Y Axis ☑ 0 30

Figure 19. Precision check example showing that this fingerprint is quite different from a typical one. Higher concentrations are observed with all species present well above detection. This sample should be invalid and flagged as AX (precision check).



Date Range 01/12/2016 00 to 01/14/2016 00

Fixed Y Axis ☑ 0 30

Figure 20. Zero air example also showing that this sample has a very different fingerprint than either a typical sample or a precision check sample. This sample should be invalid and flagged as BF (Precision/Zero/Span).

## 3.6 Level II and III Validation

Level II and III validation actions help to put your data in perspective. Some Level II and III analyses can be performed in DART, or DART can produce statistical summaries to use for further exploration outside of DART.

In Level II, it is useful to compare your site's data to data from similar sites, such as nearby sites or sites from similar sized urban areas (spatial). Figure 21 shows an example of a toluene time series from two Texas sites, Clinton Drive in Houston near the Ship Channel (heavily industrial) and a Dallas urban site. As we might expect, concentrations are significantly higher at the Clinton Drive site.

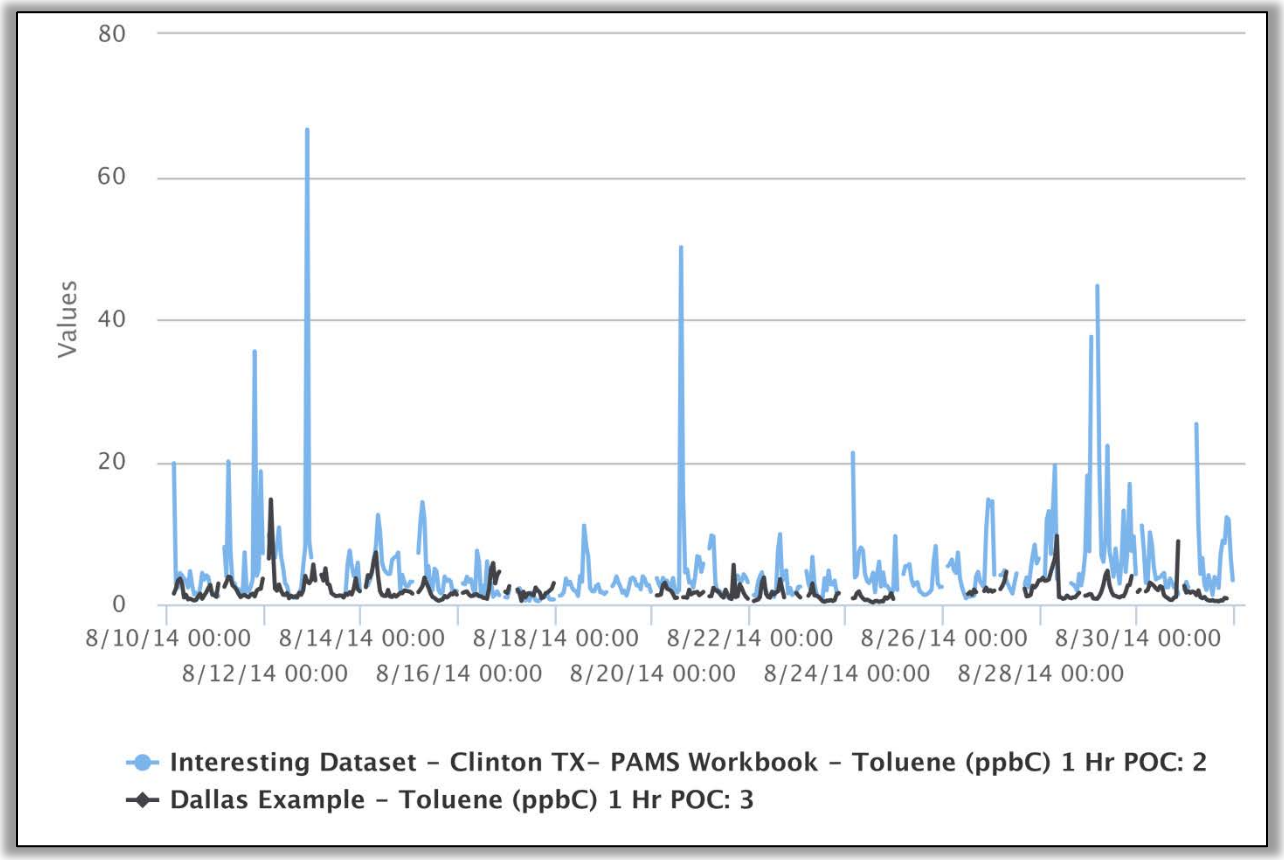


Figure 21. A toluene time series from an industrial site and an urban site in Texas.

Past analyses used data summaries compiled at a national level to show statistical ranges of concentrations at PAMS sites. Overlaying your own data on these graphics shows how your site concentrations compare to national percentile ranges (for example in Figure 22). Statistical summaries needed for this type of graphic are available within DART for your data set. However, an up-to-date national summary with which to compare to your data is not yet available.

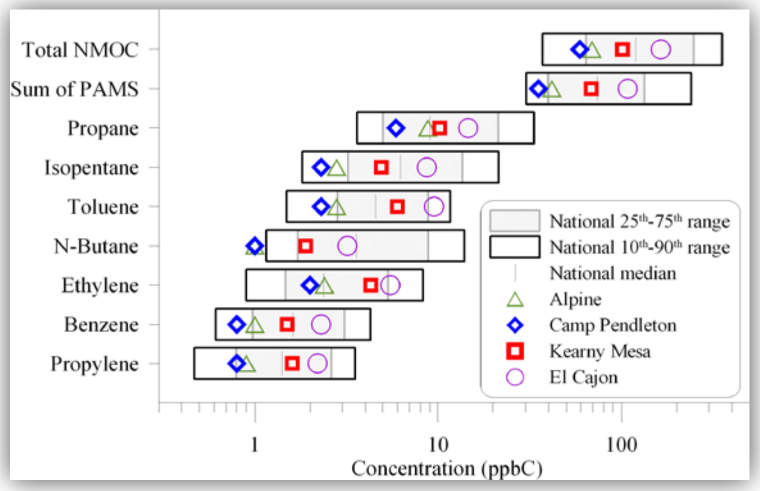


Figure 22. National concentration ranges for PAMS species compared to San Diego site data (2008). This graphic is not yet available in DART.

Another way to put your data in perspective is to compare data to data collected at the same site from previous years. First, you can use the summary statistics function within DART to create annual statistical files for further exploration outside of DART. Figure 23 shows annual statistics, annual average, and 95% confidence interval for formaldehyde at a site. This graphic and these metrics are not yet available in DART.

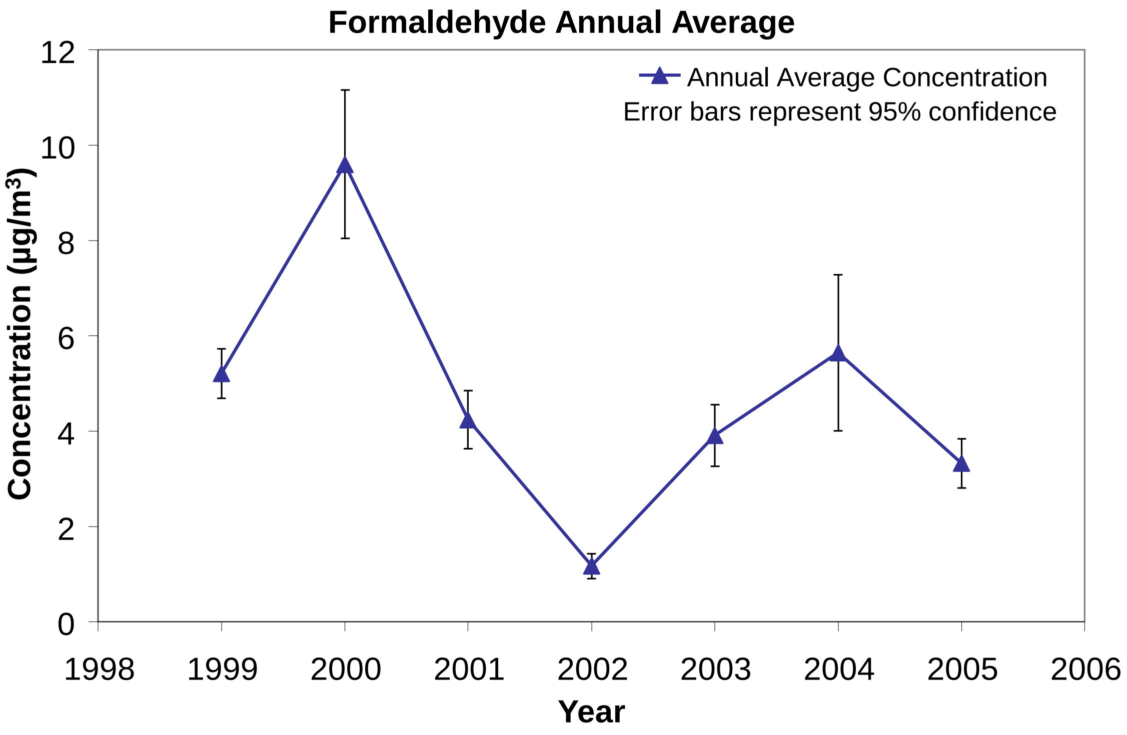


Figure 23. Annual formaldehyde concentrations represented as averages plus 95% confidence intervals. Concentrations in 2002 were statistically significantly lower than in other years because the confidence intervals do not overlap any other year. This graphic was made outside of DART.

A trends analysis is particularly useful to investigate whether ambient concentrations have shown a response to significant emissions changes in the area. A trends analysis, such as that shown in Figure 24, is not yet available in DART. This graphic was created MS Excel.

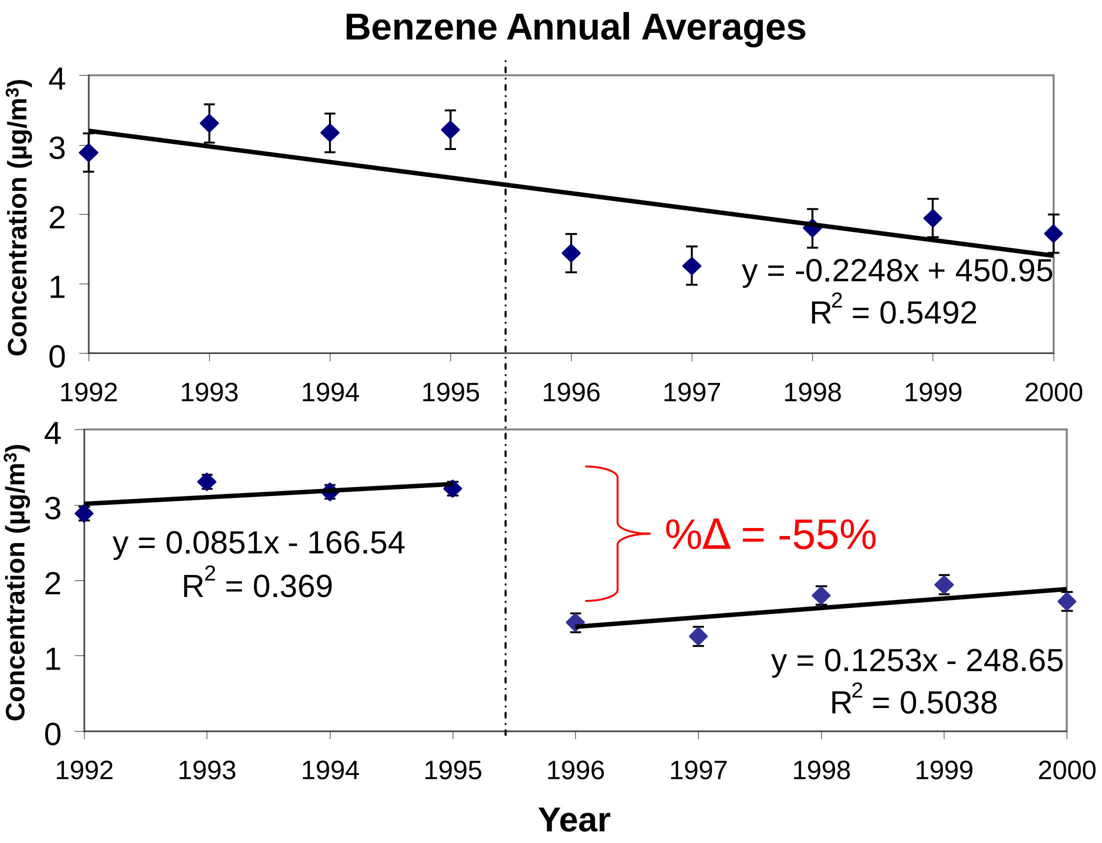


Figure 24. The same benzene annual averages (with 95% confidence intervals) fitted with regression lines in two ways. The first fits all data with one regression line and the second takes into account a large step change that occurred from regulations put into effect in 1995. The figure was created in Microsoft Excel.

For Level III validation, an intercomparison of the data from two different instruments at the same site is needed. Examples include situations in which QA samples are collected (typically 24-hr canisters) along with a continuous sampling method, or in which 24-hr canisters are collected for the air toxics program in addition to the continuous PAMS measurements (these types of intercomparisons are rare). This intercomparison is not available in DART, but individual data sets could be obtained using the AQS request feature, explored in DART, and exported for further exploration outside of DART.

## 3.7 Handling Invalid Data

Once you have thoroughly reviewed the data identified by auto-screening using the displays provided in DART, you need to decide on the validity of the data and either keep the data as valid or mark the data as invalid using the appropriate AQS code. Commonly used AQS qualifier codes include AN (machine malfunction), AQ (collection error), BH (Interference/co-elution/misidentification), and DA (Aberrant Data [Corrupt Files, Aberrant Chromatography, Spikes, Shifts]).

With PAMS samples, you must determine whether flags apply to just the species you have identified as possibly invalid or if the flags apply to the entire sample. When determining whether to flag an individual VOC or an entire sample with a qualifier code, consider:

- When there is a problem with two or more of the most abundant species (e.g., toluene, pentanes, butanes, ethane, xylenes), flag the entire sample.
- When there are multiple validation screening check failures in the sample, flag the entire sample.
- For samples with one problematic VOC, flag the entire sample if that VOC represents a significant portion of the TNMOC – e.g., more than 20%.
- Flag individual species when there appear to be problems only with those species and concentrations are low relative to the rest of the sample.

In DART Smarts, if two or more of benzene, propane, n-butane, isoprene, n-hexane, or ethylbenzene are missing or zero, the entire sample is flagged with code AQ.

## 3.8 Post-Validation Steps

Once you are satisfied with data validation, it is time to put the data into AQS format. To export your data from DART, choose the **Export Data** option. Use the **Select data for export** button to specify the parameters to export. There are some decisions to be made upon data export:

- Include missing data and apply a Null code: Upon export, any gaps in the data set are identified and filled in; records have an empty value field, and a user-specified Null code is assigned. Gaps are automatically identified by determining the data sampling interval. This option is important to analysts using your data set.
- Perform MDL check or Data Substitution
  - Add qualifier code MD for values less than MDL: Upon export, the data set is screened for data values that are less than the MDL, and the Qualifier code MD is assigned. This option is useful for data analysts exploring the data at a later time.
  - Substitute MDL/2 for value and add MS Qualifier: Upon export, the data set is screened for data values that are less than the MDL, and those data values are replaced with the value MDL/2. This option is at the discretion of the agency. Substitution of data is typically discouraged because it complicates later data analysis.
- Apply Qualifier code ND (no value detected) to all records if the concentration is 0: Upon export, the data set is screened for data values that are equal to 0, and the Qualifier code ND is assigned. A value of zero for a species concentration would be below detection and ND is therefore recommended.

When you are ready to export your data, assign a file name and click on the **Export** button.

# 4. PAMS Data Analyses

## 4.1 Analysis Objectives

Ultimately, PAMS data are collected to help agencies understand important ozone precursors, to develop effective emissions controls for reducing ozone, and assess progress in reducing emissions. PAMS data are useful in reconciling emission inventories, evaluating models, assessing pollutant transport, and analyzing trends. Questions to ask of the data include:

- How do I ensure that the data I plan to use for analysis are of good quality? *(See Section 3)*
- How do concentrations change spatially and by time of day, day of week, and season?
- Which VOCs have similar patterns? Do these VOCs have common sources?
- What are the most important VOCs in terms of ozone formation potential?
- How do concentration levels for a given city/area compare to other cities?
- Have VOC concentrations declined over time in response to emission control programs?
- How do the most important VOCs compare with model output (e.g., are ambient concentrations high in locations not shown by the model)?

For data analysis, it is useful to apply several techniques and approaches. Obtaining consensus among results gives you more confidence in the findings. It is also useful to progress from simple display-and-describe analyses to more complicated analyses, and then to analyses that are used to interpret and integrate results. An example of a flow chart for data analysis is provided in Figure 25.

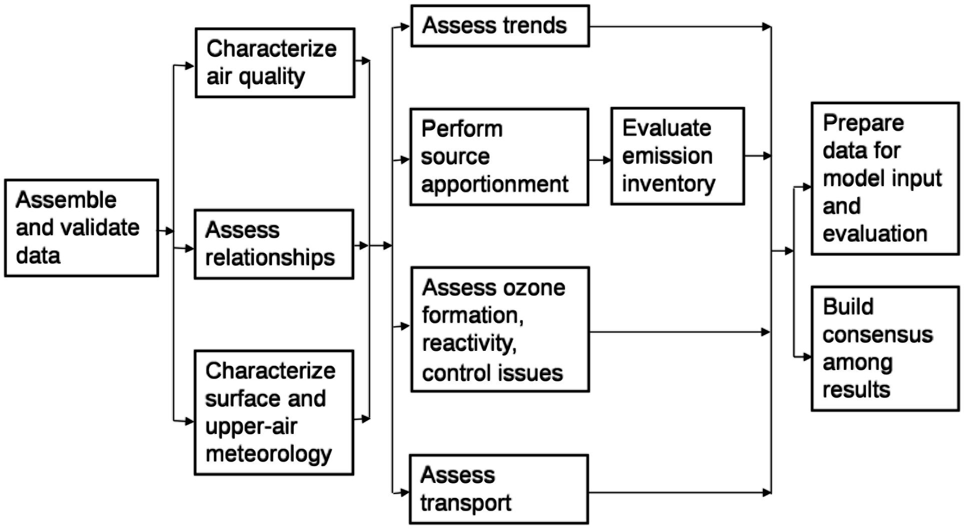

Assemble and validate data
Characterize air quality
Assess relationships
Characterize surface and upper-air meteorology
Assess trends
Perform source apportionment
Evaluate emission inventory
Assess ozone formation, reactivity, control issues
Assess transport
Prepare data for model input and evaluation
Build consensus among results

Figure 25. Example of a flow chart for PAMS data analysis.

## 4.2 Basic Analysis Examples

Data analysis starts with data validation—as an analyst progresses through data validation, much is learned about the data set. Basic "display and describe" analyses are first conducted to gain understanding of the VOCs' diurnal characteristics. Time series and box-whisker plots by time of day help analysts begin to understand how emissions sources, pollutant transport, and photochemistry affect VOC concentrations. Scatter plots help analysts inspect data for expected relationships between VOCs emitted by the same source type or from a particular direction. Stacked bar plots can show how the overall composition of ambient VOCs change from sample to sample, and whether there are sudden changes that may indicate problems in the data.  Many of these plots are provided in Section 3.

## 4.3 More Advanced Analyses

Many more analyses are useful to apply to the PAMS dataset. Most of these analyses need to be performed outside of DART at this time. However, they all rely on the validated dataset produced through DART.

### 4.3.1 Other "Display and Describe" Analyses

A spatial comparison of average concentrations, MDL values, or trends across many sites, such as a national map, can help to show jurisdictional differences attributable to sampling and analysis, national "hot spots" of certain VOCs, and regional differences in trends.

Figure 26 shows a map of average benzene concentrations (2003-2005) across the United States. Benzene concentrations have ambient measurements above detection across the country with only a few exceptions. Concentrations are consistent for areas dominated by mobile sources (e.g., the Northeast and California), while isolated high concentrations generally coincide with significant point source emissions of benzene such as refineries and coking operations. Sites that show unusually high concentrations with no clear emissions sources, or sites with concentrations that are very different from other sites (e.g., the yellow circles in Figure 26), might be further investigated to determine the cause.

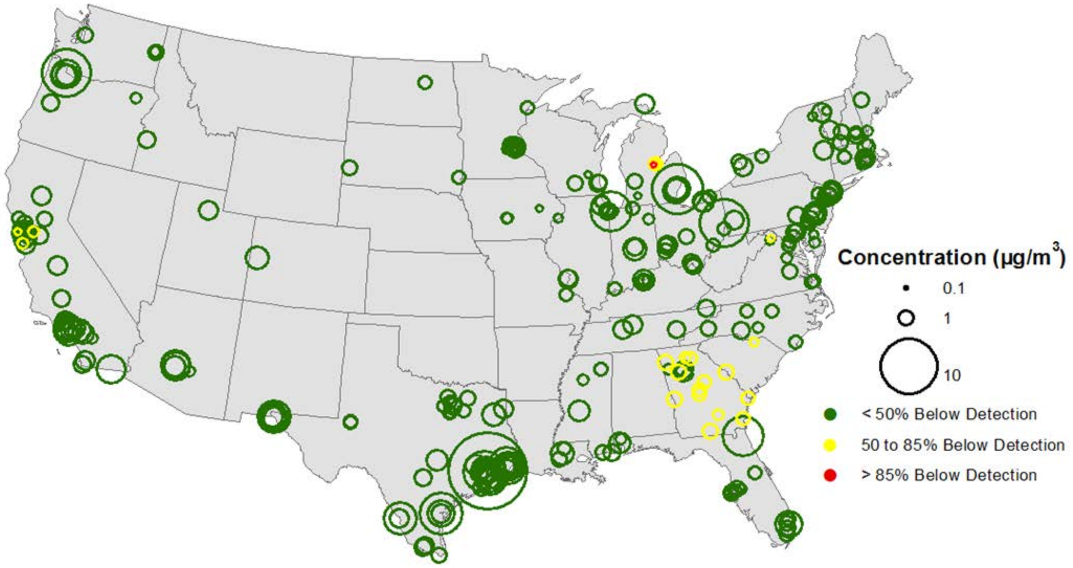


Figure 26. Average benzene concentrations for 2003-2005. The largest circle on the map corresponds to 17 µg/m$^{3}$.[3]

[3] Source: Hafner H.R., Charrier J.G., and McCarthy M.C. (2009) Air toxics data analysis workbook. Prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, STI-90830403-3224, January. Available at http://www.epa.gov/ttnamti1/files/ambient/airtox/workbook/AirToxicsWorkbook6-09.pdf.

A scatter plot matrix (SPLOM) allows analysts to compared relationships among multiple species. These plots can be generated using most statistical packages or R code. Figure 27 shows an example. To interpret a SPLOM, locate where a row and column intersect (e.g., ACETY-acetylene and MPXY-m-&p-xylenes on the bottom left hand corner). The intersection is the scatter plot of the row variable on the vertical axis against the column variable on the horizontal axis. Each column and row is scaled so that data points fill each frame. In this example, the isoprene (ISPRE) data do not appear to correlate well with the other hydrocarbons shown. In contrast, n-butane (NBUTA) and i-pentane (ISPNA) correlate very well, implying that the two hydrocarbons are from similar sources.

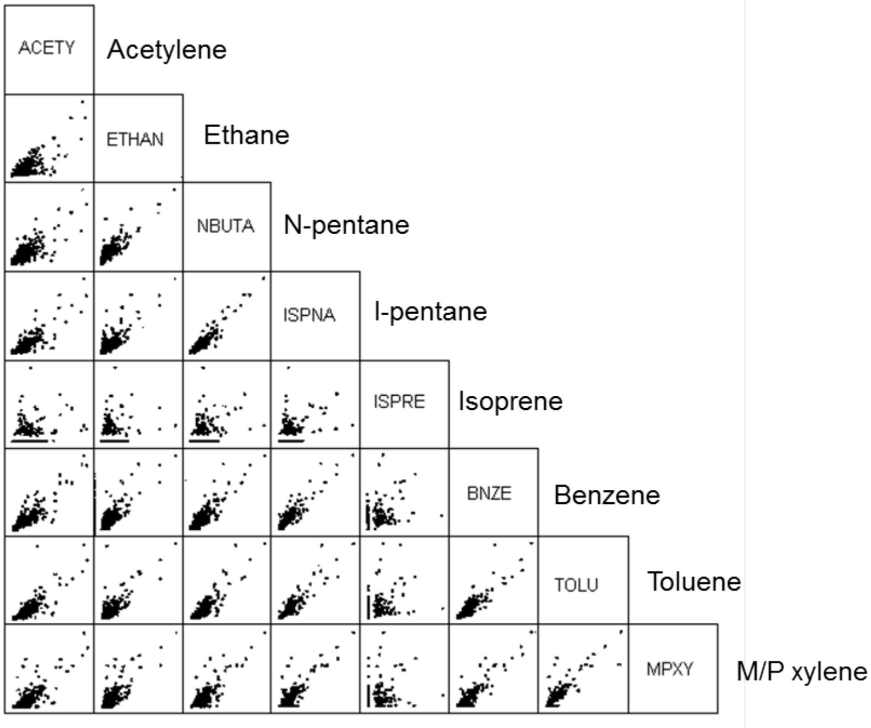


Figure 27. Example scatter plot matrix for abundant VOCs.

It is useful to assess the wind direction from which species concentrations are highest. A pollution rose shows the frequency of concentration bins by wind sector, as in Figure 28. These graphics are useful in assessing the sources that may be impacting the monitoring site. The example shows the frequency of pollutant concentrations in ppb with respect to wind direction.

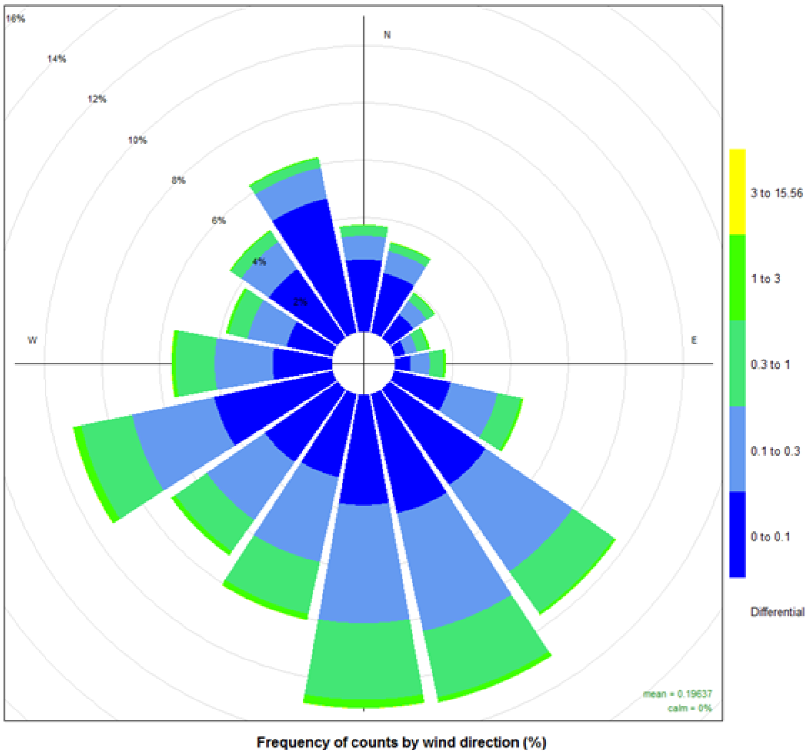


Figure 28. Example of a pollution rose plot (this example is for $PM_{2.5}$).

## 4.3.2 Reactivity-Weighted VOCs

The photochemical interaction of VOC and $NO_x$ forms ozone. Each VOC reacts at a different rate and with different reaction mechanisms. Therefore, VOCs can differ significantly in their influence on ozone formation. Incremental reactivity is the change in ozone caused by adding a small amount of test VOC to the emission in an episode, divided by the amount of test VOC added:

g ozone/g Carbon or moles ozone/mole Carbon

Incremental reactivity (measured by the maximum incremental reactivity [MIR] scale) is used to compare the ambient VOC mix among sites or episodes or to investigate VOCs important to ozone formation. Investigating the reactivity-weighted VOC data is very useful in a relative sense: Is an ambient sample more reactive than another? What are the most important VOCs with respect to ozone formation? Many less-abundant species (based on concentration) become important when reactivity is considered.

In the example shown in Figure 29, concentrations are overlaid with reactivity-weighted data. Note that ethane (ETHAN) and n-pentane (NPNTA) are quite abundant, but have low ozone formation potential, while isoprene (ISPRE) and xylenes (M_PXYL) have much higher ozone formation potential.

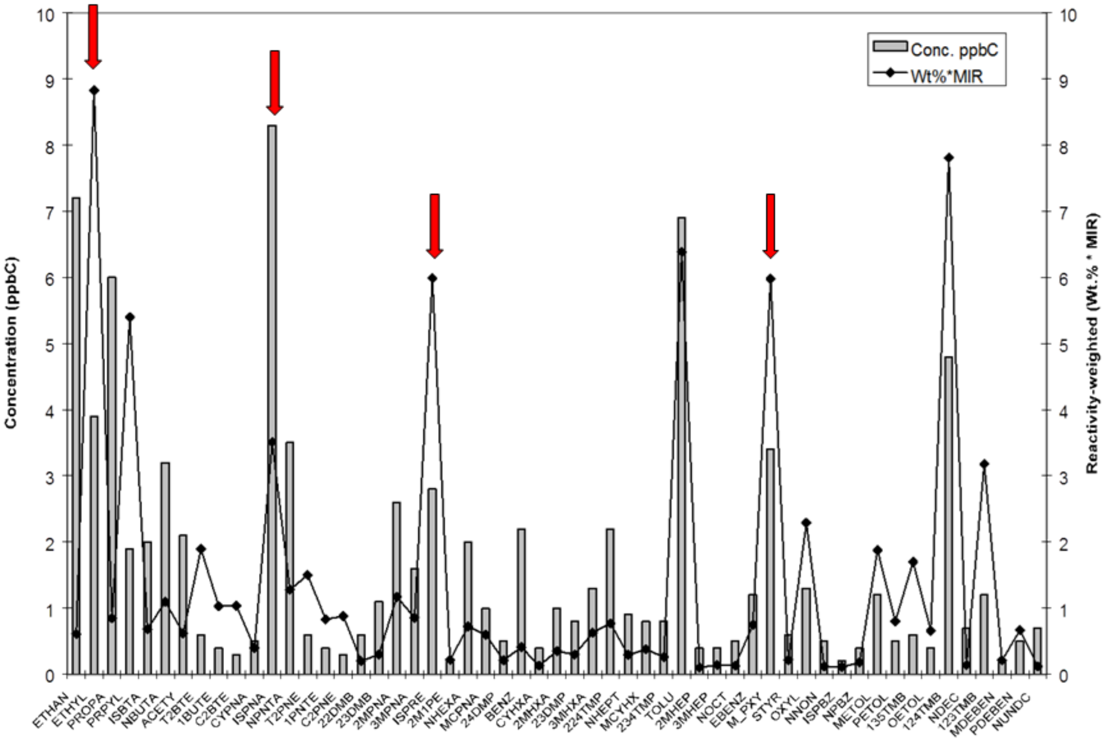


Figure 29. A typical morning, urban VOC fingerprint is superimposed on the same data weighted by MIR factors.[4] Arrows indicate pollutants discussed in the text.

[4] Source: 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 U.S. Environmental Protection Agency, Research Triangle Park, NC, STI-900243-1987-FWB, September.

### 4.3.3 VOC/NO$_x$ Ratios

Emission control strategies are based on assessments of whether an area is "VOC-limited" or "NO$_x$-limited." Assessing VOC/NO$_x$ ratios is one method to determine whether NO$_x$ and/or VOC controls would be effective to reduce ozone.

The ratio of VOC to NO$_x$ in the morning is an important parameter for photochemical systems. The ratio characterizes the efficiency of ozone formation in VOC-NO$_x$-air mixtures. At low ratios (<5 ppbC/ppb), ozone formation is slow and inefficient (i.e., VOC-limited or VOC-sensitive chemistry). Ozone formation is limited by VOC availability—reducing VOCs can reduce ozone. Decreasing NO$_x$ levels may result in increased ozone formation. At high ratios (>15 to 20 ppbC/ppb), ozone formation is limited by the availability of NO$_x$ rather than VOCs (i.e., NO$_x$-limited or NO$_x$-sensitive chemistry). Thus, reducing NO$_x$ reduces ozone. Ratios between 5 and 15 are considered transitional, and both NO$_x$ and VOC controls may be effective. Note that the range of ratios used to define VOC and NO$_x$ limitations varies among researchers.

Analyses of VOC/NO$_x$ ratios include creating frequency distributions of ratios by site and by time of day; scatter plots of VOC and NO$_x$ to assess relationships; spatial and temporal variations in ratios; and ratios as a function of time of day or along a trajectory. Figure 30 shows a histogram of VOC/NO$_x$ ratios for from an urban PAMS site circa 1998. Data were screened to include only NO$_x$ concentrations ≥5 ppb and VOC concentrations ≥100 ppbC to reduce outliers. This location was typically VOC-limited.

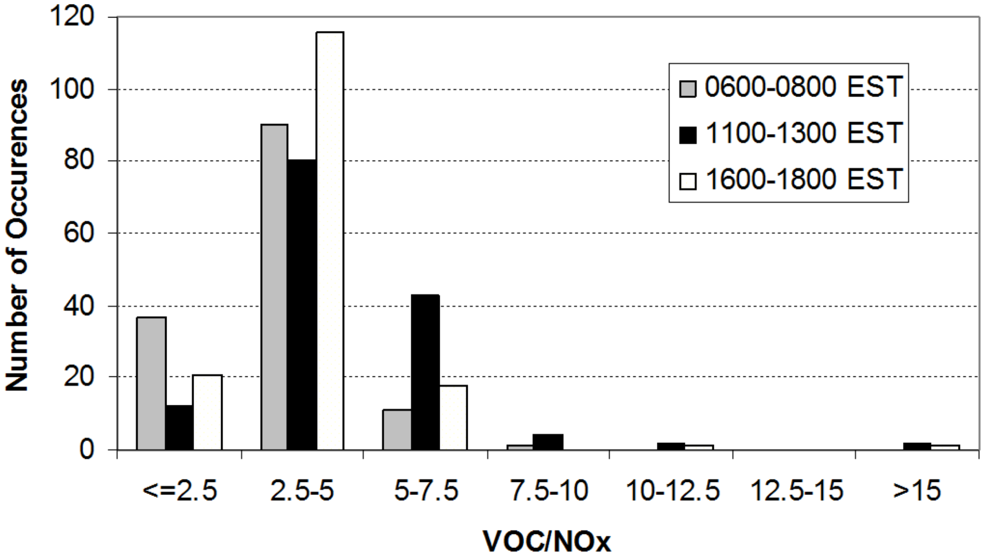


Figure 30. A VOC/NO$_x$ histogram at an urban PAMS site.[5]

[5] Source: 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 U.S. Environmental Protection Agency, Research Triangle Park, NC, STI-900243-1987-FWB, September.

### 4.3.4 Weekday/Weekend Differences

Typically, urban vehicle traffic is different on weekends relative to weekdays because of differences in commuter and business operation patterns. Comparing VOC concentrations, average diurnal profiles of concentrations, and VOC/$NO_x$ ratios on weekdays versus weekends helps to show the impact of motor vehicle emissions on VOCs and $NO_x$. Findings help with understanding possible impacts of controls on motor vehicle emissions to ambient VOC and $NO_x$ levels and thus, ozone. Statistical summaries and graphical depictions of data organized by day of week or weekday/weekend are the basis for this type of analysis. Figure 31 shows a box whisker plot of morning hour TNMOC/$NO_x$ ratios and $NO_x$ concentrations by day of week. The TNMOC/$NO_x$ ratio is higher and $NO_x$ concentrations are lower on Sunday when traffic is reduced at this urban site.[6]

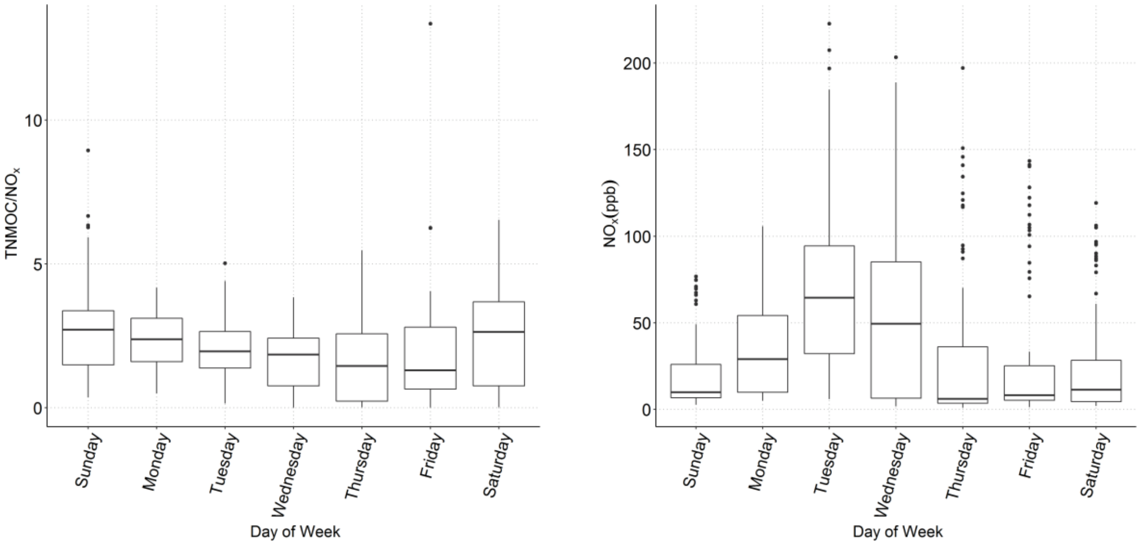


**Figure 31.** Box whisker plots of TNMOC/$NO_x$ ratios and $NO_x$ concentrations (ppb) at an urban site by day of week (DOW).

[6] A notched box-whisker plot shows the entire distribution of concentrations for each year. In box-whisker plots, each box shows the 25th, 50th (median), and 75th percentiles. The boxes are notched (narrowed) at the median and return to full width at the 95% lower and upper confidence interval values. These plots indicate that we are 95% confident that the median falls within the notch. If the 95% confidence interval is beyond the 25th or 75th percentile, then the notches extend beyond the box (hence a "folded" appearance).

### 4.3.5 Trend Analyses

It is important to track ambient concentrations and their changes over time to see if progress is being made to reduce emissions (i.e., are emission control programs working?). Starting with valid data, data preparation for trend analysis includes setting data completeness criteria (typically 75%), handling data below detection (many options), and determining statistical metrics for the analysis.

Quantifying trends can include assessing

- The percent difference between the first and last year of the trend period (rough, "first cut" sense of the change).
- The difference between two multi-year averages (helps account for changes in meteorology or abrupt change in emissions).
- The percent change per year (the slope of the regression line). This approach allows comparison of changes across varying trend lengths and between sites.

Testing the significance of observed trends includes:

- Calculating the significance of the slope of the regression line using the F-test, which is a statistical measure of confidence that the regression line does not have a slope of zero.
- Using other methods such as t-tests, nonparametric tests (such as Spearman's rho test of trend or Kendall's tau test of trend, which test for and estimate a trend without making distributional assumptions), and analysis of variance.

Visual inspection of any trend is necessary to ensure the results make sense (or not). Showing confidence intervals, concentration ranges (such as box plots), average method detection limits, or multiple metrics is useful in assessing trends. Obtaining consensus among results (such as similar trends from a range of statistical metrics) increases certainty in the observed trends.

Figure 32 shows an example of trends for three VOCs across three trend periods: 1990-2005, 1995-2005, and 2000-2005. The number of sites with data over the three periods varied significantly. The metric plotted is the percent change per year. Variability for shorter trend periods is much higher. Concentrations have generally declined by about 5% per year.

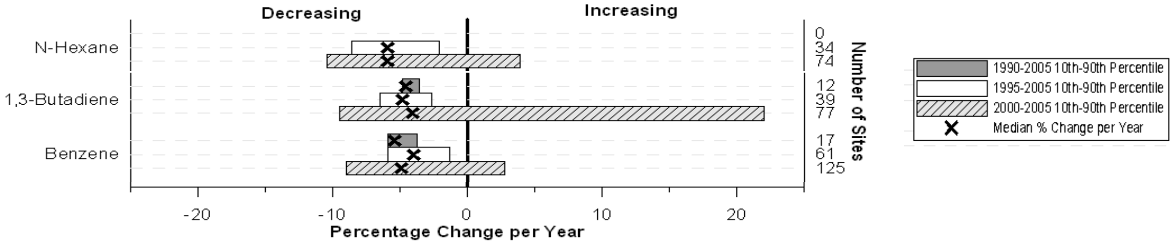


Figure 32. Percent change per year for n-hexane, 1,3-butadiene, and benzene at air toxics sites across the U.S. for three time periods.[7]

[7] Source: Hafner H.R., Charrier J.G., and McCarthy M.C. (2009) Air toxics data analysis workbook. Prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, STI-90830403-3224, January. Available at http://www.epa.gov/ttnamti1/files/ambient/airtox/workbook/AirToxicsWorkbook6-09.pdf.

### 4.3.6 Comparing Ambient Data to an Emissions Inventory

Emissions inventories are routinely used for planning purposes and as input to comprehensive photochemical air quality models. Significant biases in either VOC or $NO_x$ emission estimates can lead to poor baseline photochemical model performance and erroneous estimates of the effects of control strategies. The basic approach is to compare early morning (e.g., 0700-0900 LT) ambient- and emissions-derived data:

- NMOC/$NO_x$ ratios.
- Relative compositions of individual chemical species and species groups.
- Relative reactivities of individual chemical species and species groups.

Early morning sampling periods are the most appropriate for these evaluations because they have the best potential to minimize the effects of upwind transport and photochemistry. During the morning, emissions are generally high, mixing depths are low, winds are light, and photochemical reactions are minimized.

A difficult part of this analysis is processing the emission inventory to put it on the same basis as the collected ambient data. For example, the emissions data need to be converted from a mass basis to a molar basis, and species in the inventory that are not measurable by the ambient data system (such as halogenated species) must be excluded.

Figure 33 shows an example of a comparison of reactivity weighted composition of the emission inventory and ambient data. In this case, the emission inventory has more reactive olefins than the ambient data. This analysis is complementary to a ratio analysis.

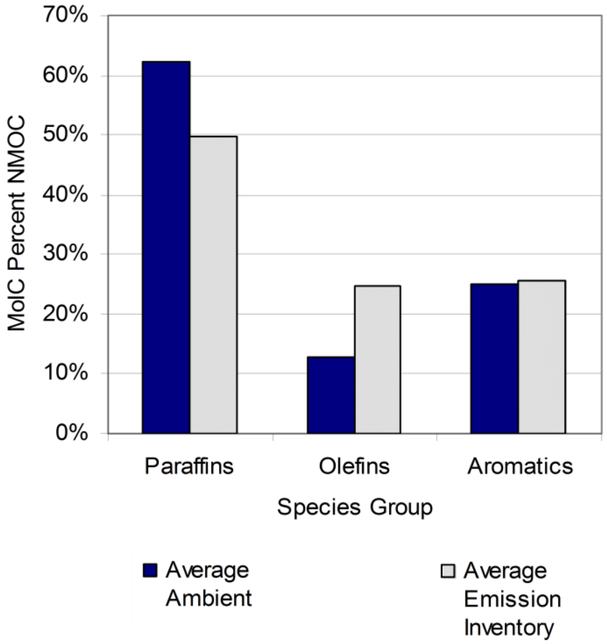


**Figure 33.** VOC group composition and reactivity-weighted comparisons for ambient and emission inventory data at a site.[8]

[8] Source: 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 U.S. Environmental Protection Agency, Research Triangle Park, NC, STI-900243-1987-FWB, September.

### 4.3.7 Source Apportionment

Source apportionment modeling uses data collected at a monitoring site(s) to reconstruct the impacts of emissions from various sources of pollutants. Common approaches for source apportionment modeling include the following:

- Numerical evaluation of data to identify sources: correlating pollutants associated with specific sources (e.g., scatter plots); correlating wind speed or wind direction with specific source markers (e.g., scatter plots); and subtracting urban-regional concentrations of a specific pollutant.
- Dispersion models, photochemical models, and/or emissions inventories: CALPUFF, AERMOD, CALINE3, CMAQ. These models are freely available at http://www3.epa.gov/scram001/aqmindex.htm.
- Statistical algorithms requiring data from a receptor site(s) (commonly called receptor modeling): Principal Component Analysis (PCA), Positive Matrix Factorization (PMF), Chemical Mass Balance (CMB), and Multilinear Engine (ME/ME-2). CMB, PMF, and Unmix models use this approach. Some of these are available at http://www.epa.gov/scram001/receptorindex.htm.

Figure 34 shows an example of the distribution of source types contributing to VOC in the Los Angeles area. This PMF analysis used PAMS data from Azusa for the period 2001-2003.

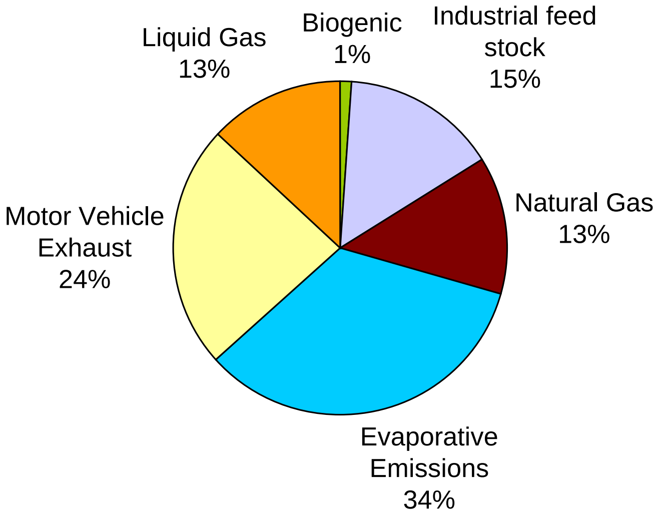


Figure 34. Source apportionment of ambient VOCs in Hawthorne, California, using data from 2003-2005.[9] (Source: Brown et al., 2007).

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

### 4.3.8 Transport Analyses

As the ozone standard has become more stringent, background concentrations and pollutant transport are of increasing importance. Nonattainment areas need to understand ozone and ozone precursors being transported into the domain. Typical investigations include exploring the relationship between surface meteorology and air quality data (diurnal plots, maps, pollution and wind roses) and performing case studies of periods of high ozone concentrations (time series, HYSPLIT trajectory analysis). An analysis of mixing height evolution during case studies is also useful.

Figure 35 shows an example of trajectory cluster analysis output. In this example, trajectories were run four times per day at a selected height for the entire study period. Trajectories were then grouped by direction. This analysis is useful to trace the frequency of high or low concentrations of a pollutant during different transport regimes.

Transport analysis pairs well with source apportionment analyses, allowing analysts to gain further information about potential source regions.

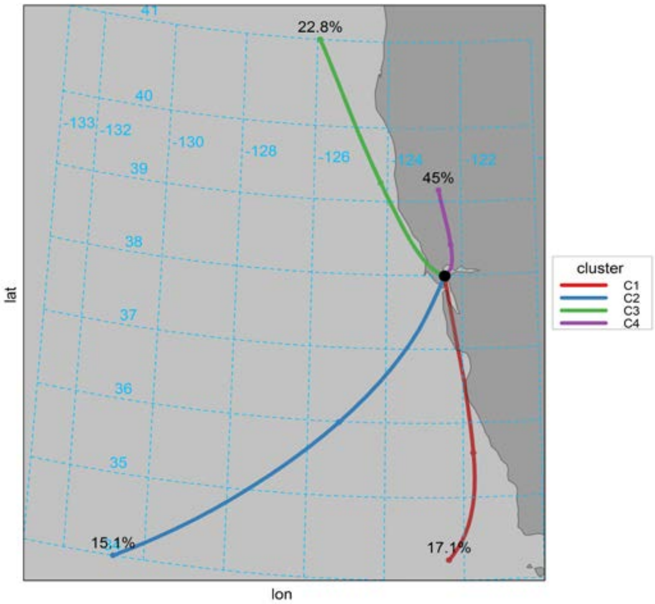


Figure 35. Trajectory cluster output showing hourly trajectory clusters for the period of interest. Clusters are labeled with the proportion of hourly trajectories falling in each cluster.

### 4.3.9 Accountability Analysis

Tying changes in ambient concentrations to changes in control programs is difficult. In addition to accounting for meteorological impacts on ambient concentrations, analysts need to understand the spatial scale of the control's influence: did emission changes affect several states or were the changes more localized? Over what period of time were the changes made? Previous investigations of ambient air quality changes found confounding influences from multiple controls applied within similar time frames and at different spatial scales. The most straightforward assessments of control effectiveness are possible when a significant change happens abruptly (such as introduction of reformulated gasoline, Figure 36) and when ambient data are available both before and after the change.

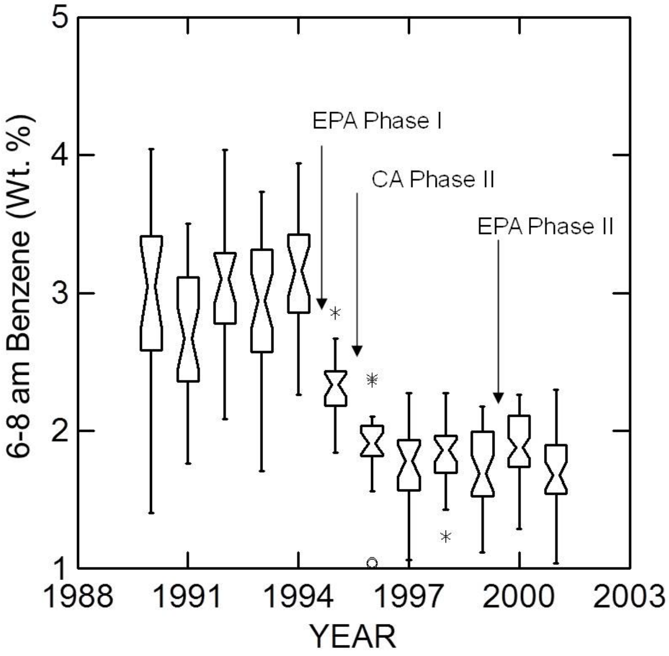


Figure 36. Illustrating the impact of the introduction of reformulated gasoline in California on ambient benzene concentrations.