EPA's Data Analysis and Reporting Tool (DART)

Presented by Jennifer DeWinter
Sonoma Technology, Inc.
Petaluma, CA

on behalf of the
EPA OAQPS PAMS Program

National Air Toxics Workshop
October 26, 2015

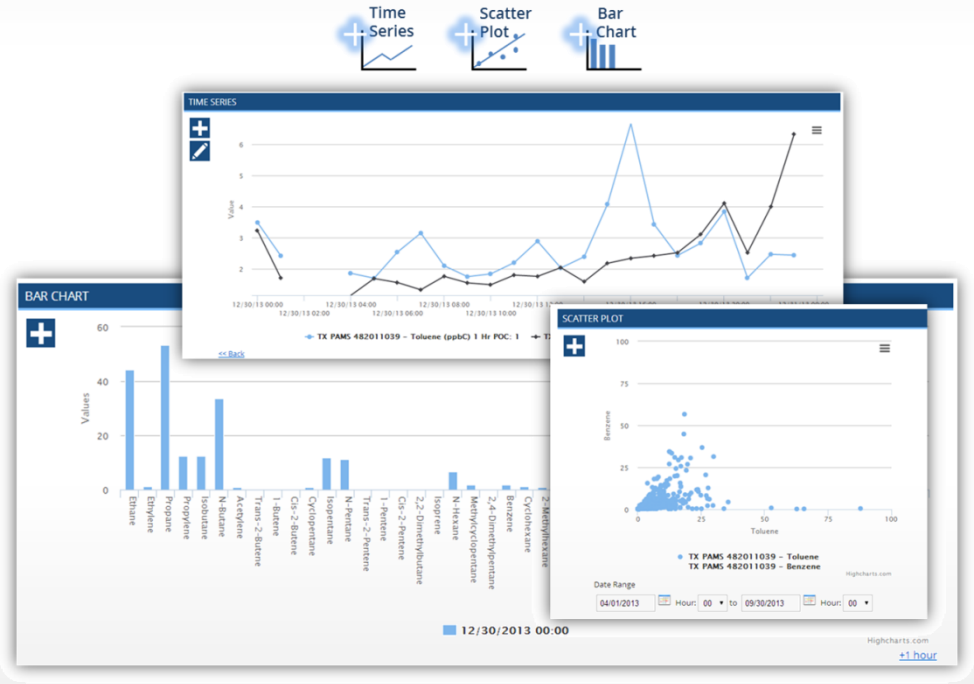
Sonoma Technology, Inc.

Glossary

Term	Definition
ANT	AirNow-Tech
AQS	Air Quality System
Auto-GC	Automated Gas Chromatograph
EPA	Environmental Protection Agency
NACAA	National Association of Clean Air Agencies
PAMS	Photochemical Assessment Monitoring Stations
$PM_{2.5}$	Particulate matter smaller than 2.5 micrometers in diameter
STI	Sonoma Technology, Inc.
VOC	Volatile organic compound

Agenda

- Introduction
- About DART
- How to Use DART
  - Live demo
- Upcoming Features
- Summary
- Q&A



Meet the Team

Kevin Cavender, EPA

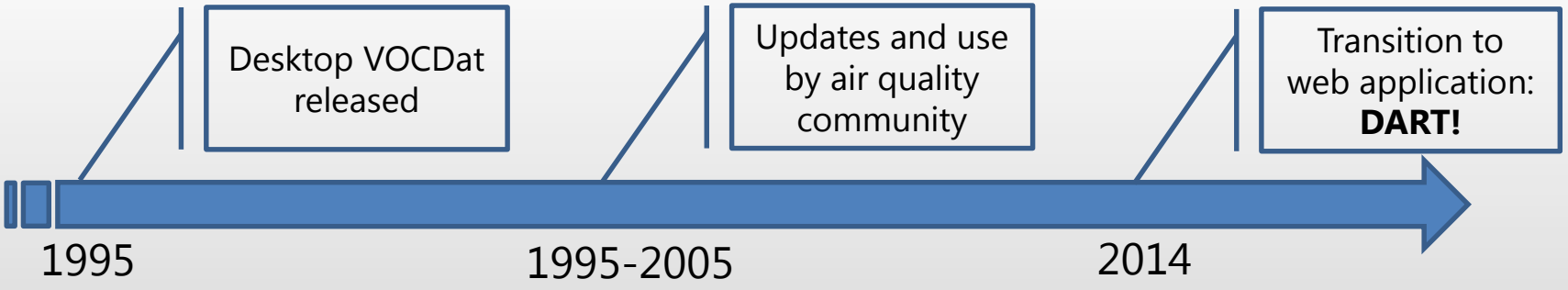
Jennifer DeWinter, STI

Steve Brown, STI

With excellent feedback and ideas from NACAA Steering Committee

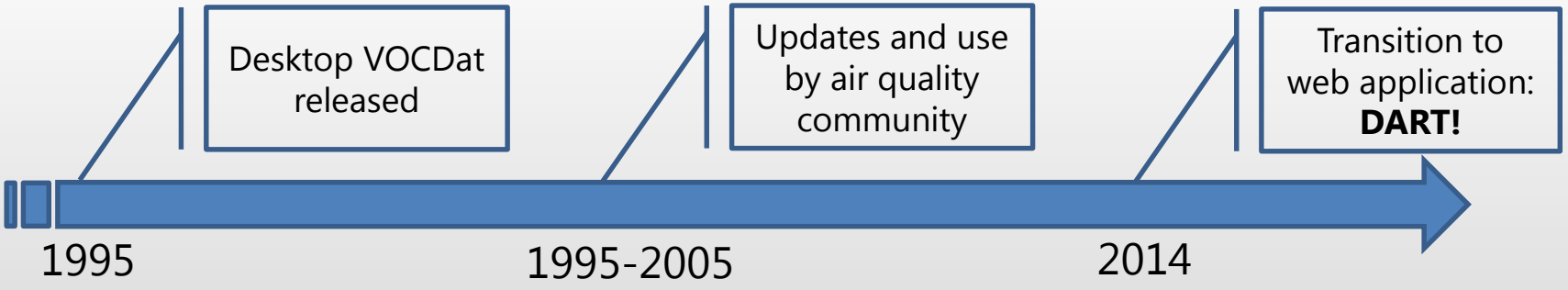
DART History

- Formerly VOCDat, a desktop software program
- Used by PAMS community to validate VOC data and prepare data for AQS submission
- Used to analyze carbonyls, air toxics, and speciated $PM_{2.5}$



What is DART Now?

- Web-based application for acquiring, analyzing and screening data
- Useful for all types of air quality data (criteria, VOCs, toxics, etc.)
- Available in AirNow-Tech



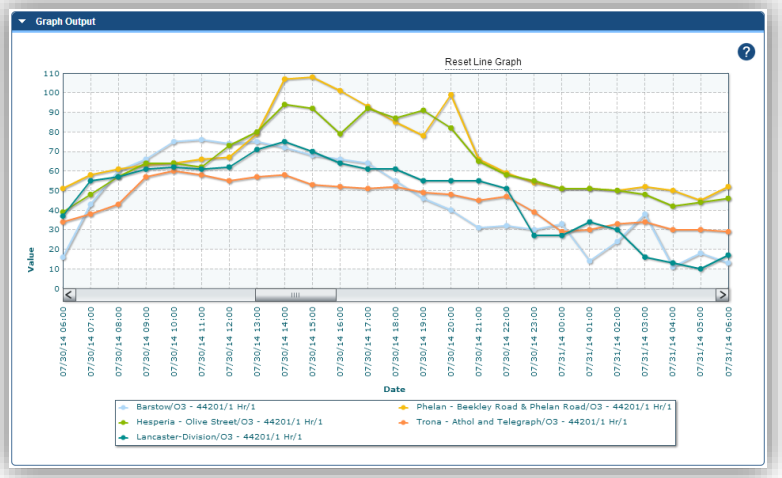
DART Version 2.0 – New Features

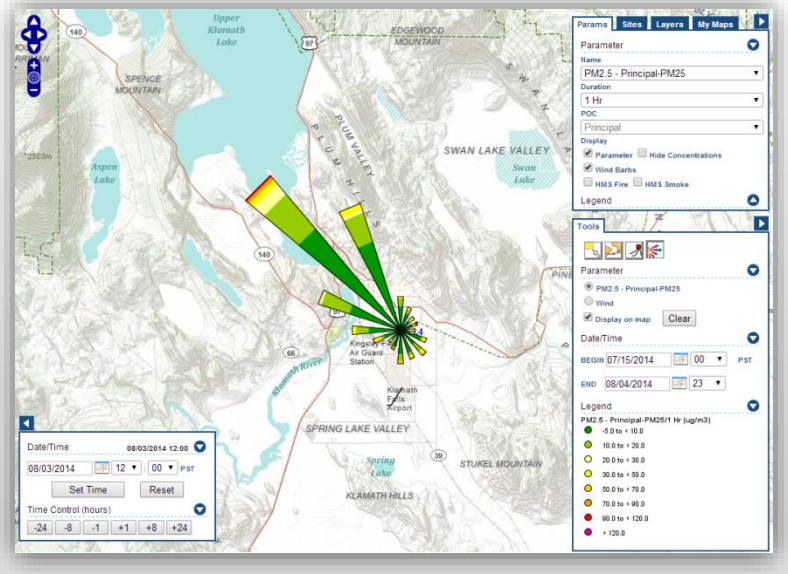
- **All-New Automated PAMS data screening**
  - "One-click" auto-screening provides interactive tables and plots to evaluate PAMS data
  - Screening checks based on recommended procedures in PAMS Data Analysis Workbook
- **Improved Data Imports**
  - Improved status information for data uploads and AQS requests
  - Unit conversions and customization of units on graphics

AirNow-Tech

www.AirNowTech.org

- Decision support system for state, local, and federal agencies
- Password-protected, customizable, web-based software
- Contains a suite of tools to support analysis, forecasting, and custom reporting



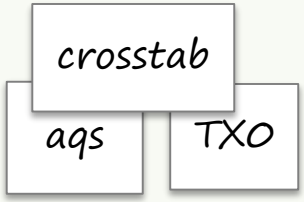


General Workflow in DART

1

Add data to DART

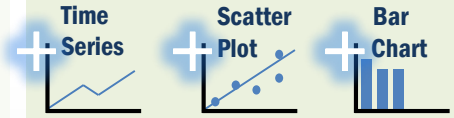
Manage



2

Explore data with graphs

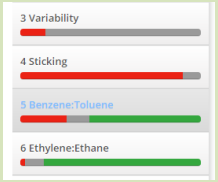
Explore



3

Perform validation

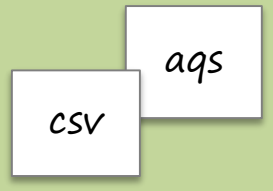
Validate



4

Export data and statistics

Export

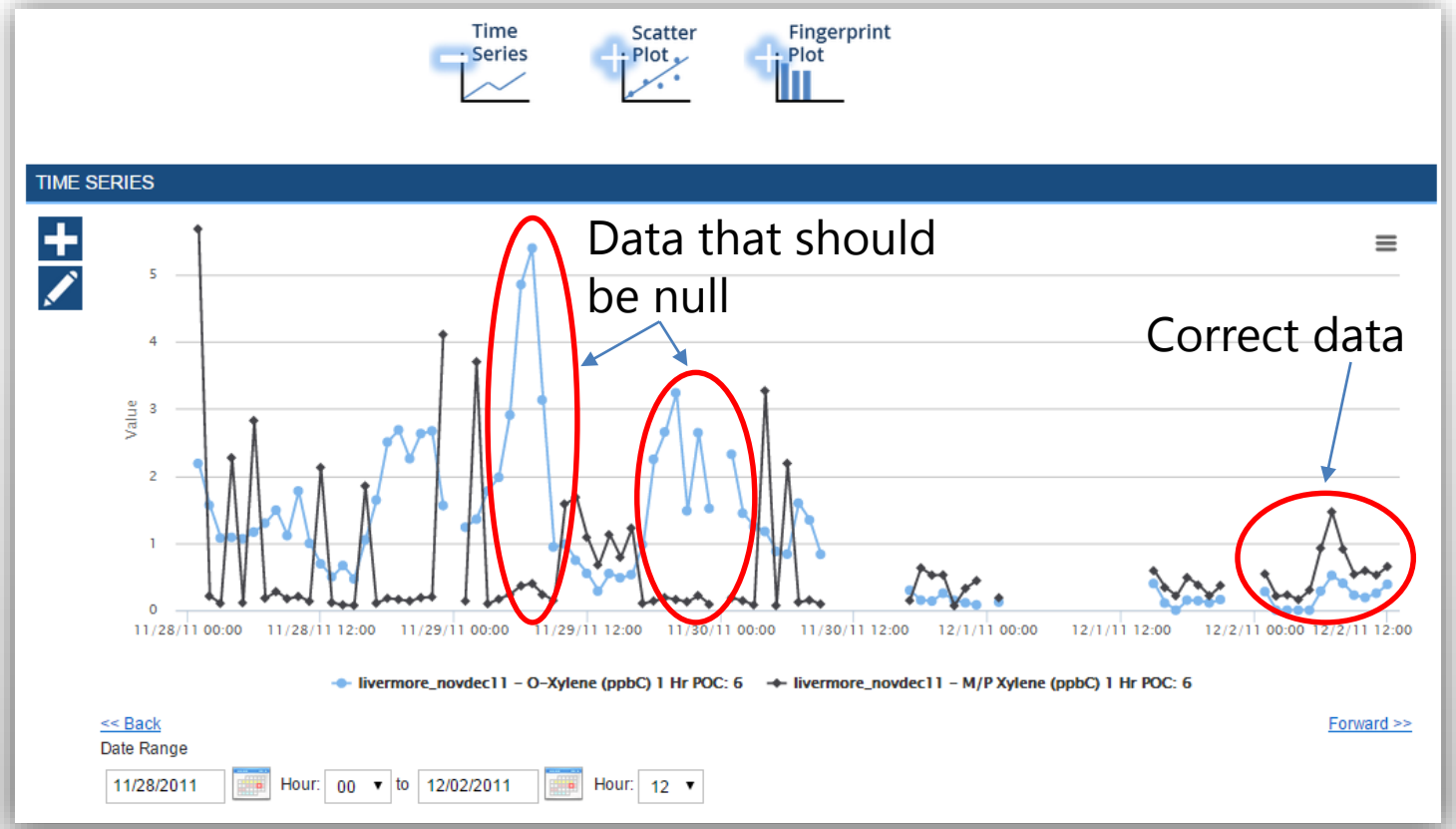


Demo

- Placeholder for demo of software
  http://airnowtech.org/

Time-Series Graphs

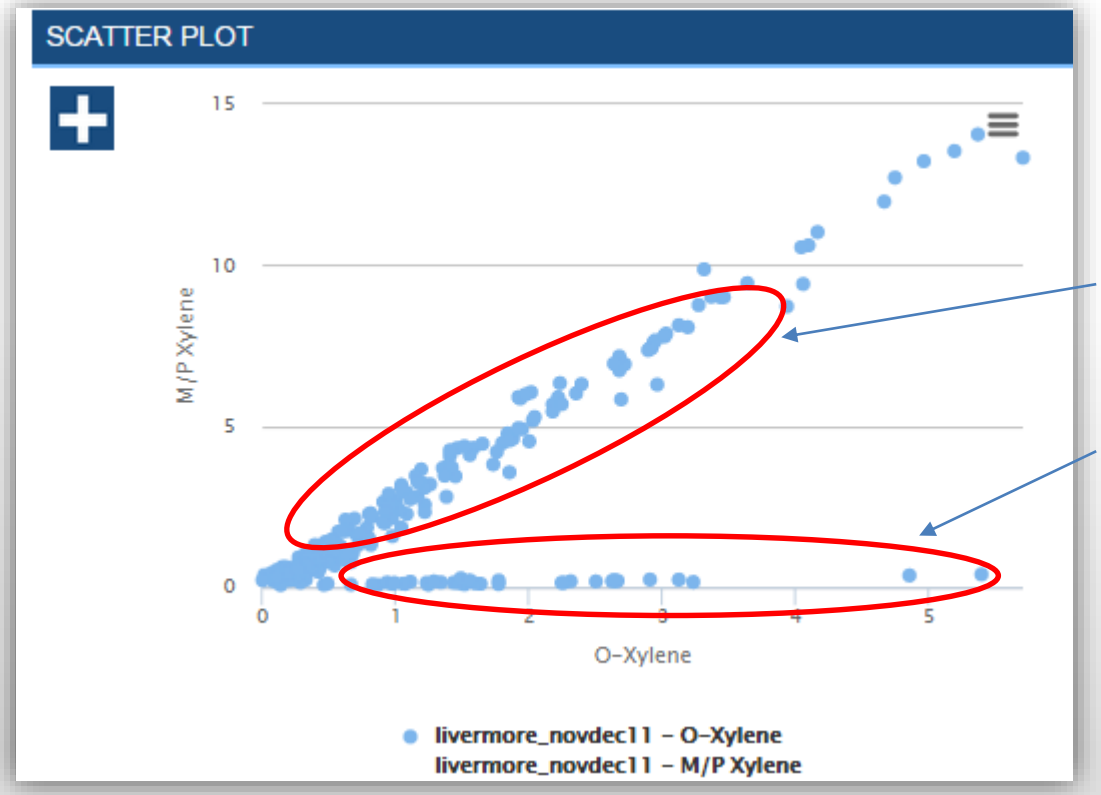
Time-series graphs are used to identify outliers, unusual data, trends, effects of meteorology in air quality, and changes in relationships among species



m/p-xylene is always > o-xylene in emissions and ambient air

Scatter Plots

Scatter plots are used to identify outliers outside of usual patterns in the data; some species have typical relationships based on meteorology and emissions



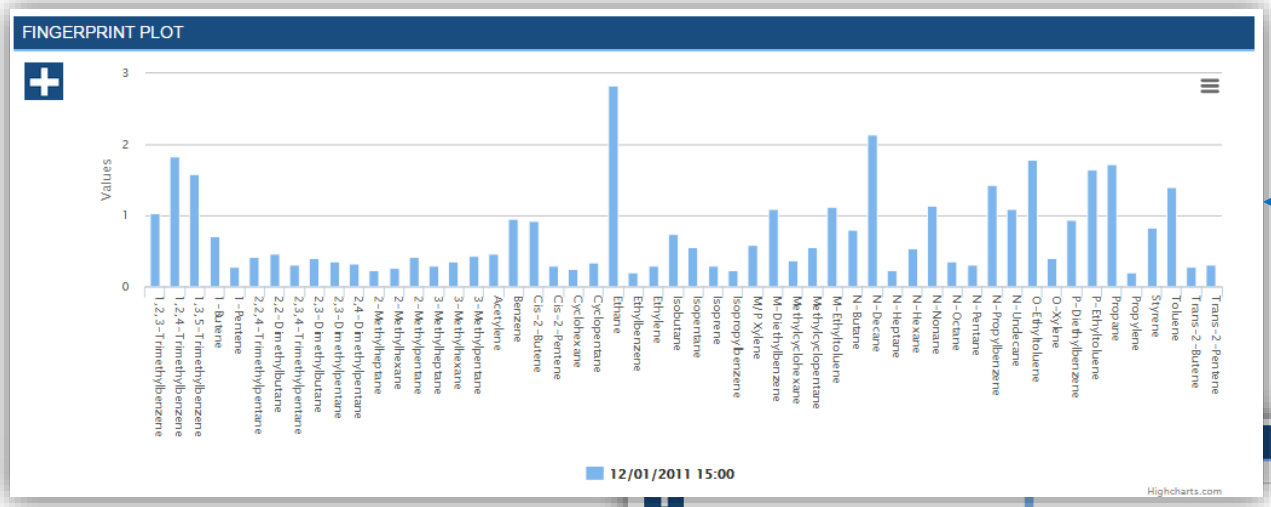
Correct data

Data that should be null

m/p-xylene is always > o-xylene in emissions and ambient air

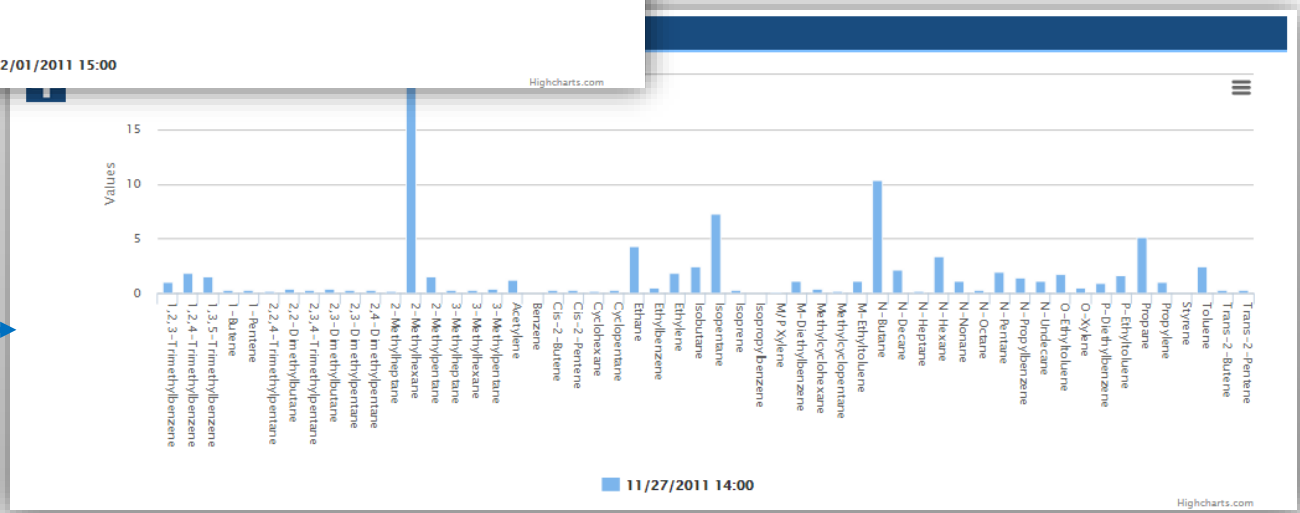
Fingerprint Plots

Fingerprint plots are used to scroll through data sample by sample to visually identify sudden changes in data



Good sample, typical of ambient VOCs

Unusual sample and likely bad data – examine this sample further in time series and scatter plots



Screening Checks (1)

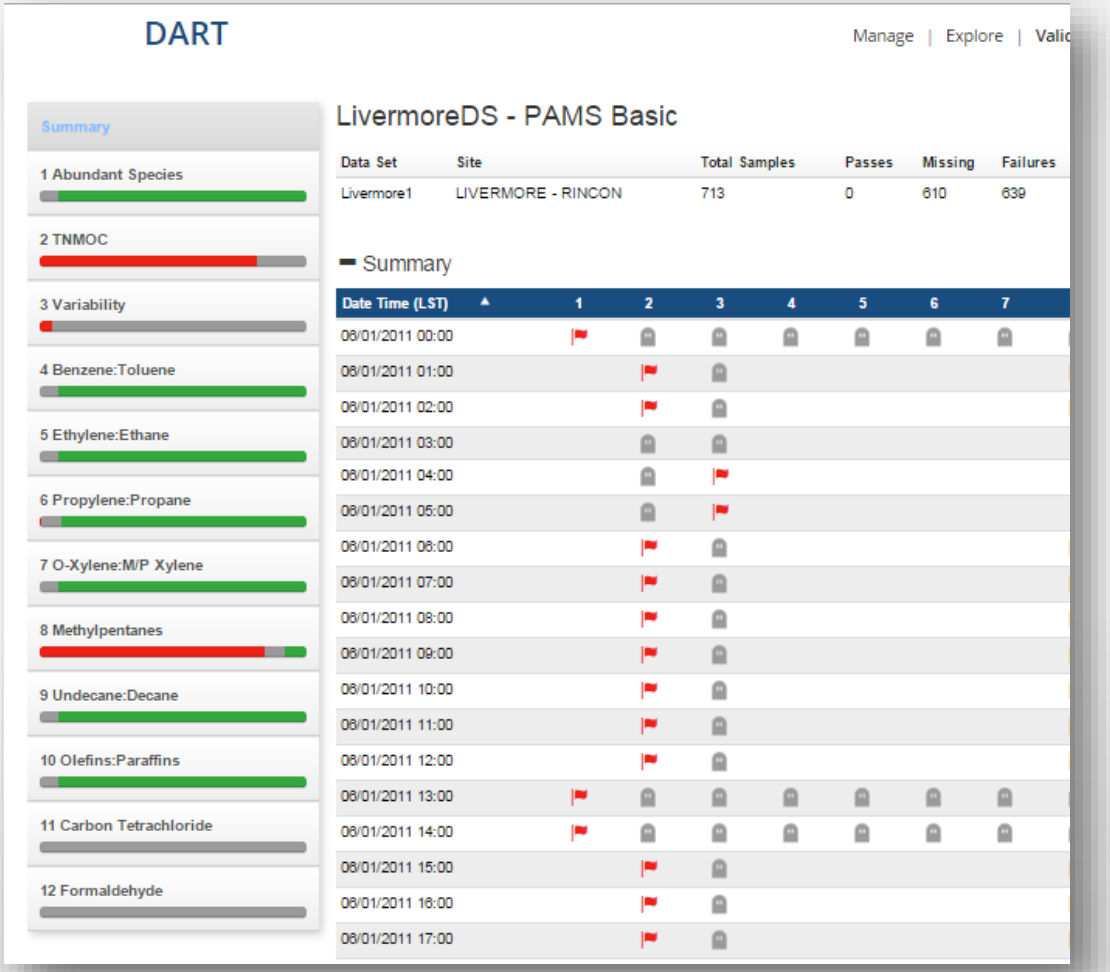
- Use screening checks to identify and export sample records that don't meet your conceptual model of ambient air quality or your sites, or that are physically unrealistic
- For example, look for samples with
  - High unidentified fraction (possible error in GC column, or data reporting error)
  - O-xylene>m/p-xylene (physically unreasonable, likely error in species identification)
  - Carbon tetrachloride below global background levels (physically unreasonable, likely error in sampling)
  - Sulfate>3*sulfur (physically unreasonable, likely error on Nylon or Teflon filter)

Screening Checks (2)

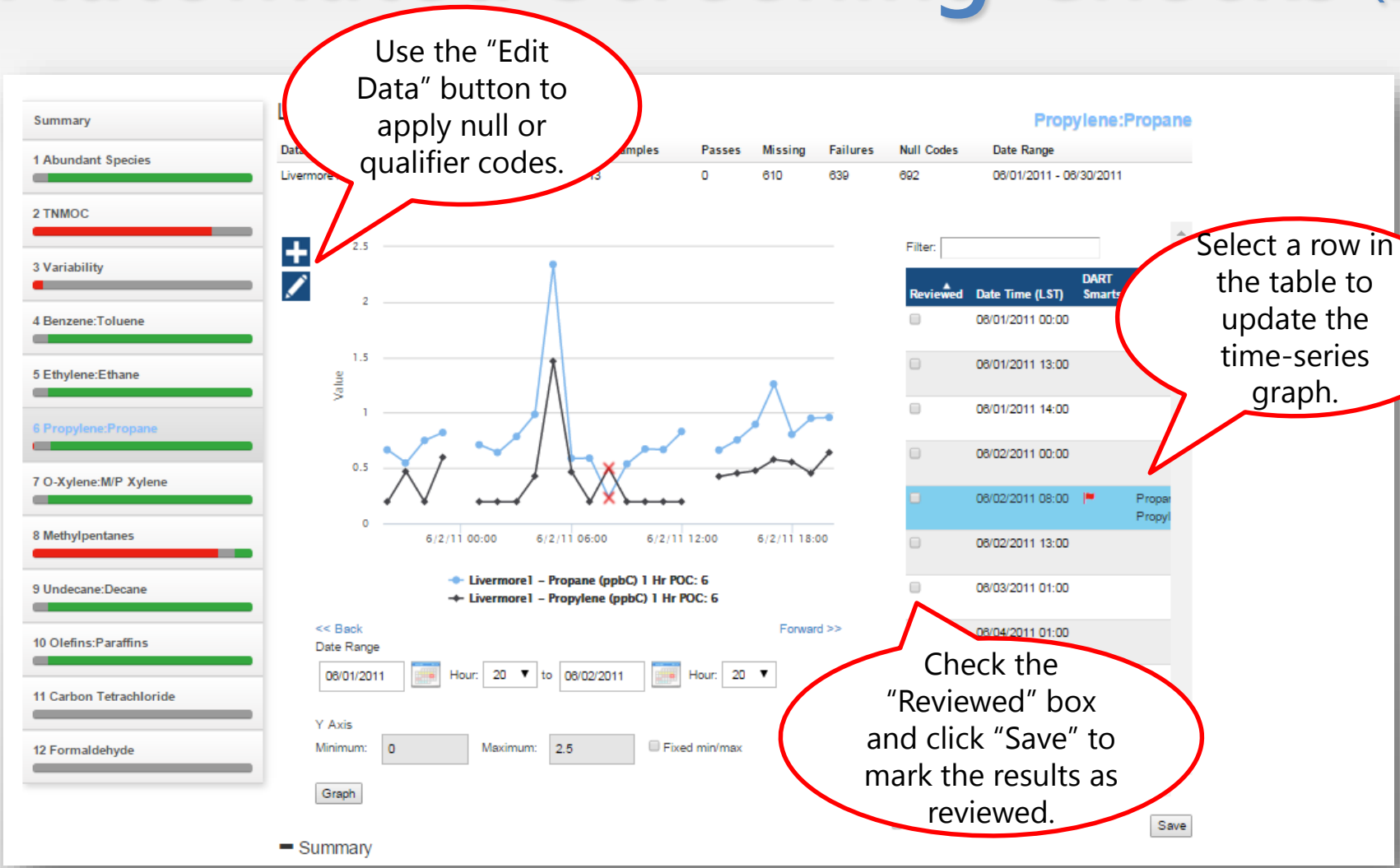
- **Species Threshold –** identify data values that exceed threshold criteria
- **Species Variability –** identify data within a specified variability range
- **Species Comparison –** compare data values between parameters according to defined criteria
- **Species Fraction –** identify data values that are within a specified fraction of another data parameter value
- **Multi-Condition –** create data screening queries that meet more than one condition

Automated Screening Checks (1)

- Run screening checks to identify problematic VOC data
- Review results in DART using interactive, linked tables and time-series graphs



Automated Screening Checks (2)



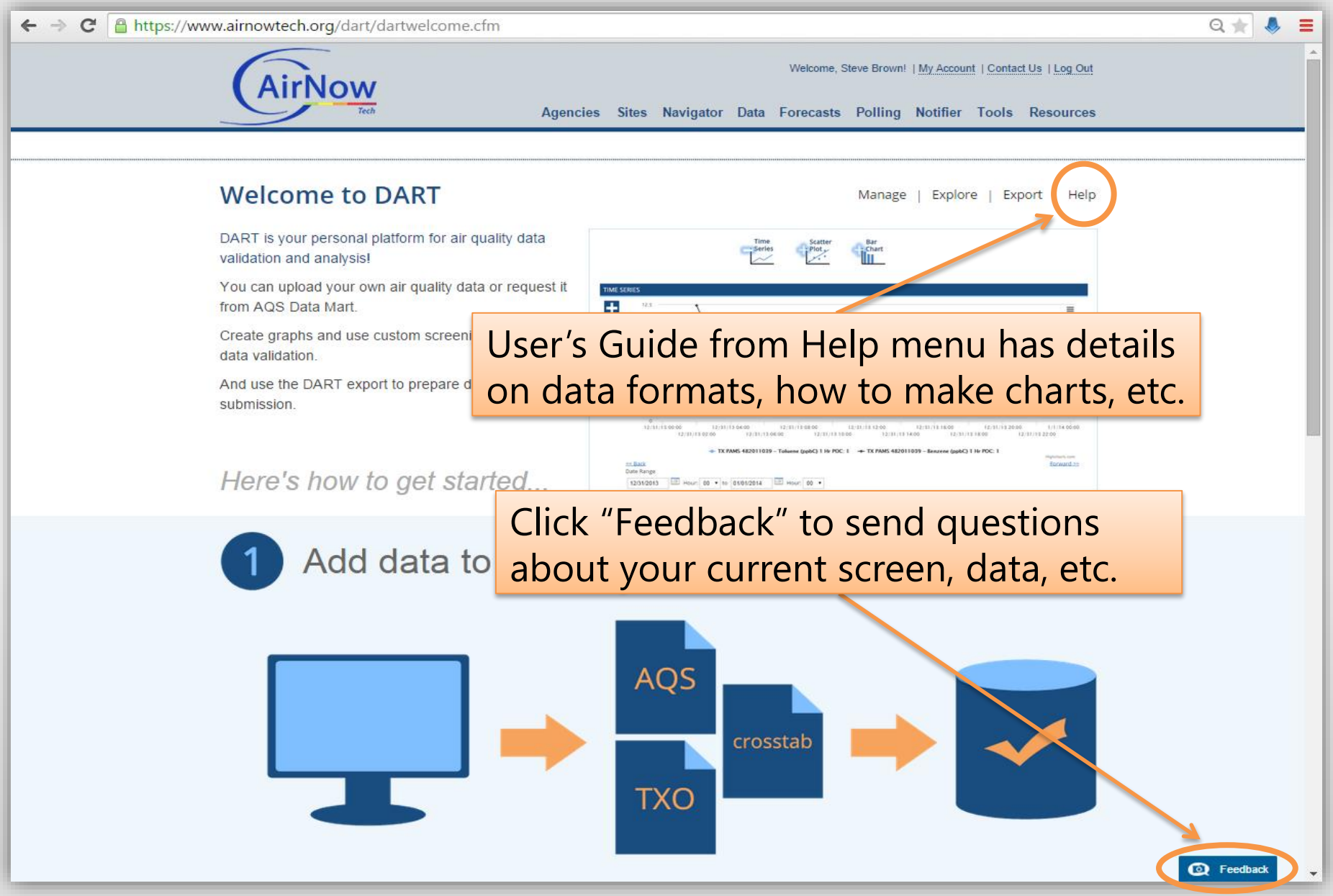
Upcoming Features

- **Customized automated data screening**
  - Setup your own “One-click” auto-screening checks
- **Secondary y-axis for time series**
- **Delete data sets**

Future Feature Ideas

- **Interactive map for Data Mart AQS requests**
- **Suite of automated screening checks for air toxics**
- **Compare site to national statistics**
- **New analyses and plot types**
  - Plot concentrations and MDL values
  - Plot concentrations and annual averages
- **Support for more import file formats**

How to Get Help



Summary

- DART is ready to use! Please let us know if you have questions or ideas for new features
- More new features to be deployed in November 2015
- After deployment, several webinars will be given
- Next phase of development to begin in 2016

Contact Us

Kevin Cavender
EPA PAMS Lead
cavender.kevin@epa.gov

Jennifer DeWinter
Atmospheric Scientist
jdewinter@sonomatech.com

Steve Brown
Senior Atmospheric Scientist
sbrown@sonomatech.com

Sonoma Technology, Inc.
Environmental Science and Innovative Solutions

707.665.9900 sonomatech.com @sonoma_tech

Extra Slides

Potential Features for CSN

- **Automatic file transfer –** allow use of an API to transfer laboratory analysis to DART and retrieve validated data from DART after states' approval
- **Import file formats –** support file format(s) useful for CSN data and expand parameter list to support key metadata (sample flow rates, etc.)
- **New plots/analyses –** stacked bar chart, display custom aggregates (e.g., calculate/display soil fraction)
- **Mass analysis** – calculate reconstructed mass on import and display in plots; compare to collocated mass available in AirNowTech
- **Automated screening for CSN** – suite of checks tailored for speciation data

More Potential Features

Feature Description
AQS Submission of Validated Data from DART
AQS Update of QC codes of data already in AQS from DART
Automated Data Screening & Summary Reports for CSN, Toxics, Criteria Pollutants
Downloadable AQS Data Availability Status Reports by site/agency
Semi-Automated "Level 3" Data Validation: compare to national averages, historical data, and nearby site information
AirNow-Tech Data Accessibility
Saving User Preferences for AQS Requests, Graphs, and customized validation procedures
Performance Improvements: schedule data exports
New Graphs and Graph Options: MDL's, stacked bar charts, wind & pollution analyses
File Upload Improvements: support auto-GC, CSN laboratory, aethalometer formats, etc.
AirNow-International Compatibility

Even More Potential Features

Feature	Description
Source Apportionment	Incorporate source apportionment tools (e.g., EPA PMF, Unmix and CMB) for user to run on DART datasets
Source Profile Comparisons	Incorporate SPECIATE or other source profile databases to enable CMB, ratio analyses, and source profile-oriented screening checks
Custom aggregates	Create/display custom sums (e.g., BTEX, unidentified) for user selected species; aggregate high time resolution data (e.g., 1-min to hourly).
Excel format	Support export to Microsoft Excel (*.xlsx) file format for crosstab/AQS
Update the POC information in file export	POC is assigned on file upload; this feature would allow the user to specify a parameter specific POC to use in the exported file. Some users reported that not all their compounds at a particular site have the same POC.
View data screening results	Display results in a table and link to plots
Weighting data	Enable cancer weighting, reactivity weighting, custom weights
Laboratory data	**Enable direct transfer from laboratory to DART**
Additional input formats	Support XML formatted files
Display box plot aggregates	Add option for box plots that would display aggregates of weekday/weekend, diurnal, day of week, season, etc.
Missing/zero data values	Provide information regarding missing/zero data values (e.g., data not found in input file)
QC codes	Enable applying standard (0-7) QC codes to data
QC log	Changes to null/qualifier codes during the validation process are logged and available for viewing/exporting
Reports	Provide precision/bias reporting useful for reviewing collocation data and replicate analyses
Drag and drop	Enable drag and drop option for file uploads
Apply filters	Use query results from data screening checks and apply filters to other plots
Data export	Enable export of data straight from plots