Using VOCDat to Validate and Process Air Toxics Data

Steven G. Brown
Hilary R. Hafner
Sonoma Technology, Inc.
Petaluma, CA

14th Annual National Air Quality Systems Conference
Clearwater, FL
May 24-28, 2004

Outline of Presentation

- Overview: What and Why
- After Obtaining Data From Labs:
  - Using VOCDat for processing, validating data
  - How to import & export data
  - Validation examples and ideas
  - How to handle aethalometer data too!
- New: Precision Report
- Other New Additions to VOCDat
- What Other Features Would Be Helpful?

VOCDat: How Did We Get Here?

- Originally developed for Photochemical Assessment Monitoring Stations (PAMS) VOC data
- Later adapted to easily
  - Acquire data from AQS/AIRS
  - Process data into AQS/AIRS format
  - View and validate data before submitting to AQS!
- Easily applied to other data sets, including air toxics

Importance of Data Validation

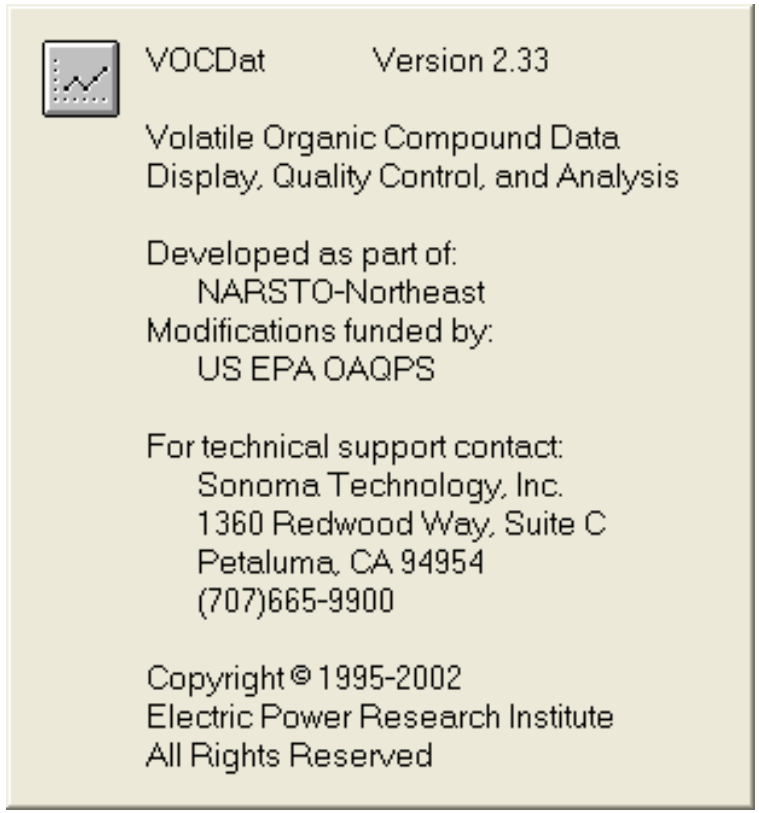
"The purpose of data validation is to detect and then verify any data values that may not represent actual air quality conditions at the sampling station."*

Without proper validation before analysis, erroneous conclusions may be drawn!

* U.S. EPA 1984, Sec. 2.0.3, p. 10

What is VOCDat?

- VOCDat is a Windows-based, menu-driven program used to
  - Import various data formats
  - Display VOC, toxics, and PM data
  - Validate the data
  - Begin data analyses
  - Export the data into EPA AQS format



VOCDat Features (1 of 2)

- Import and export AQS AMP370 (old AIRS) format
- Import and export AQS R2 format
- Import flexible format ASCII files
- Import Turbochrome and TotalChrome (auto-GC) formats
- Export formats suitable for other software (such as spreadsheets or databases)
- Edit data QC codes on screen (keeps a log of changes)
- Prepare graphical displays of time series, scatter, and fingerprint plots

VOCDat Features (2 of 2)

- Provides summary statistics
- Allows species list (selectable) customization
- Allows screening criteria customization
- Creates “weighted” data (using reactivity or risk)
- Calculates species group sums including paraffins, olefins, aromatics, unidentified, carbonyls
- Facilitates AQS precision report preparation and formatting

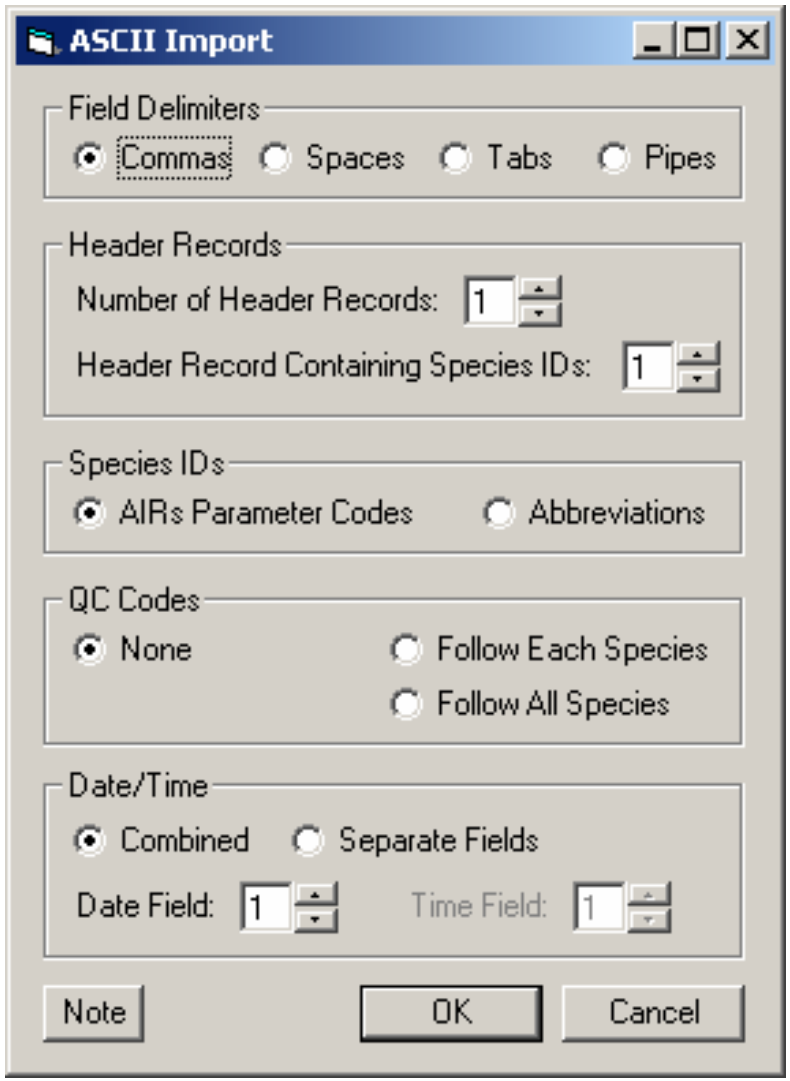
How VOCDat Works

- VOCDat needs to know what species to expect in the data file
  - Species.txt defines abbreviations, method codes, units
  - Several species lists available
  - Only species in the list will be exported
- VOCDat allows for supplemental data handling
  - Species2.txt allows use of meteorological and criteria pollutant data in validation
  - These are not exported into AQS

From the Beginning:

- Acquire toxics data from lab, then what?
- Select TO-15 method species list
  - List may be modified to accommodate other species
- Check format of data
  - Select one of a number of specific formats
  - Work with labs to get data in useful format
  - Know anyone who likes Access? Use them for lab data!
- Import into VOCDat

Importing ASCII Files



Validation Approach I

- Understand the pollutant sources, lifetimes, temporal behavior, etc. (use tools such as “cheat sheets”)
- Understand site location, sampling and analysis techniques
- Inspect summary statistics and apply screening criteria

Validation Approach II

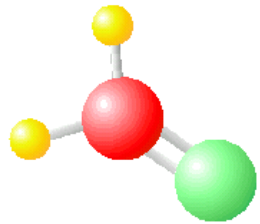
- Inspect all species
  - Time series plots
  - Scatter plots (internal consistency)
- Inspect every sample
  - Fingerprint plots (in elution order)
- Flag data and document modifications
- Prepare data for AQS

Making Validation Easier

- Once data is VOCDat ready, easy to validate!
- With EPA, STI has developed “cheat sheets” for the primary 18 HAPs
  - Available soon at www.ladco.org/toxics.html
  - Also useful in analysis!
- Easily generated simple graphs allow for quick validation
- User’s guide, data validation workshop useful for examples and ideas http://www.sonomatechdata.com/sti_workbooks

Sample HAPs Validation Sheet (1 of 2)

Formaldehyde



Synonyms: BFV; formalin; formalith; formic aldehyde; Formol; FYDE; HCHO; HOCH; karsan; lysoform; Methanal; methyl aldehyde; methylene glycol; methylene oxide; oxomethylene

Molecular formula: CH_2O ***CAS Registry #:*** 50-00-0
AIRS parameter code: 43502
Chemical group: Carbonyl

Summary: Formaldehyde is created by the oxidation of gas-phase hydrocarbons. Formaldehyde concentrations exhibit diurnal and seasonal variability because Formaldehyde photochemical production is a function of sunlight. Although it is also rapidly removed from the atmosphere via photochemical reaction, continuous production from the oxidation of methane and nonmethane hydrocarbons keep formaldehyde at concentrations above detection.

Reactivity	High	***Lifetime***	Hours
Background[1] conc.	0.2 ± 0.07 µg/m^3 (from 1980-2002)	***Spatial scale***	Local, subregional, regional
Molecular Weight	30.03	***Conversion factor[2]***	1 ppb = 1.35 µg/m^3
Sources	Key photochemical reaction product. Also produced by power plants, manufacturing facilities, incinerators, and automobile exhaust emissions. Used as a chemical intermediate, analytical reagent, concrete and plaster additive, used in cosmetics, disinfectants, fumigants, photography, and wood preservation and particle board products.		
Sinks	Photolysis; reaction with OH		
1 in 10^6 cancer risk:	0.08 µg/m^3 (IRIS, 2003); 0.17 µg/m^3 (CAL EPA, 2002)		
Non-cancer RfC[3]	3.6 µg/m^3 tier 2 (Axelrad *et al.*, 1998); 3 µg/m^3 (CAL EPA, 2000)		
Minimum Detection Limits (Pilot Cities)	Average: 0.05 µg/m^3 Range: 0.001-0.1 µg/m^3		

Regulations since 1990: Direct regulations include: wood furniture manufacturing in 1995 (33,000 tons/yr); off-site waste operations in 1996 (43,000 tons/yr); and pulp and paper production (155,000 tons/yr)[4]. Also, Maximum Achievable Control Technology (MACT) reductions on other VOCs will also indirectly reduce formaldehyde production (see Benzene, 1,3-butadiene).

Sample HAPs Validation Sheet (2 of 2)

Data Validation Guidelines for Formaldehyde

Table of expected temporal and spatial behavior

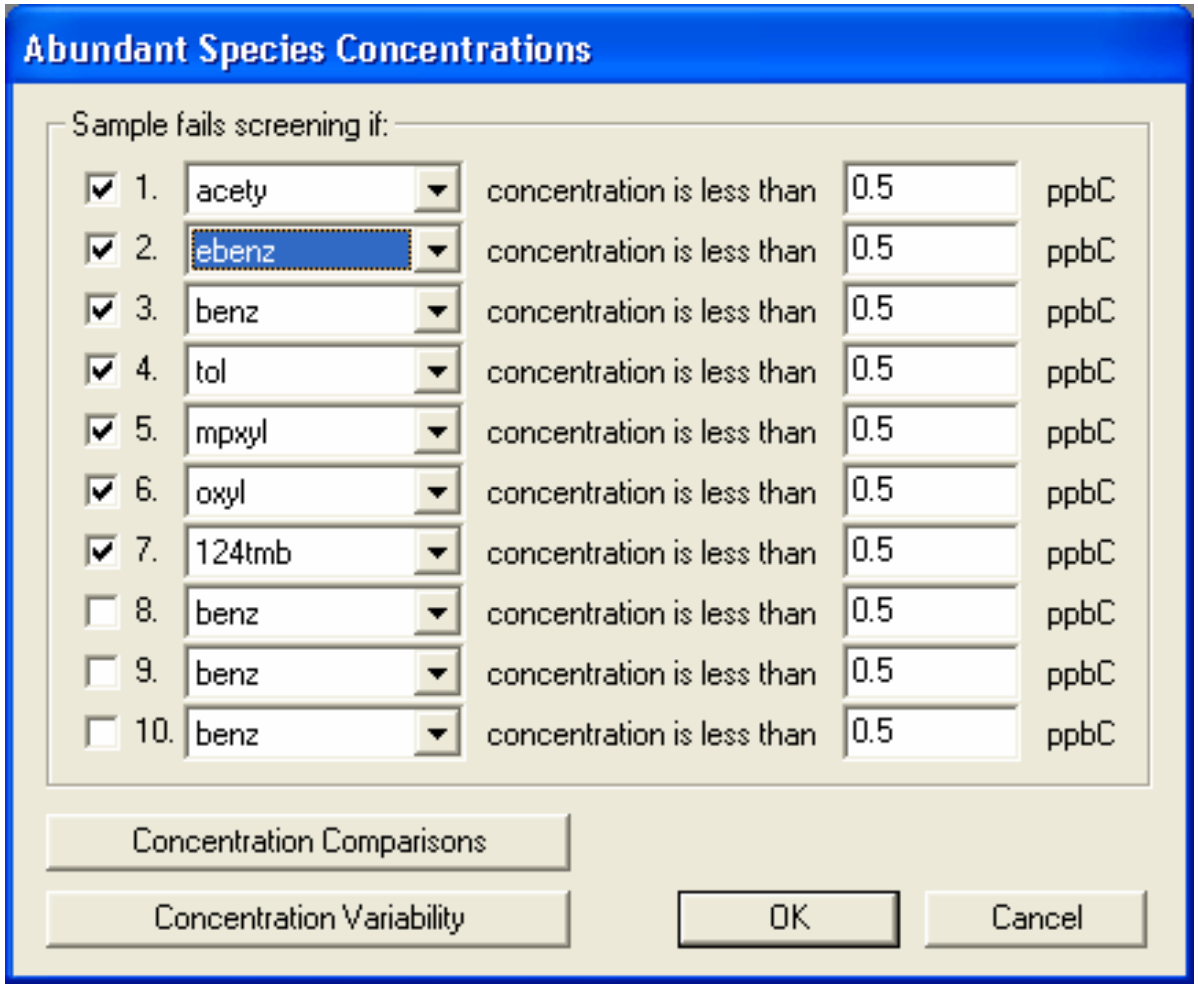
Temporal or Spatial Scale	*Expect Variability?*	*Behavior/explanation*	*Action*
Diurnal (time of day)	Urban-yes Rural-likely	Photochemical production from other VOCs is highest during the day	Check for midday-afternoon peak in concentrations
Day of week	Possible, but not likely	If VOC concentrations are significantly lower on weekends, production may be lowered	Check for lower concentrations of formaldehyde on weekends
Seasonal	Yes-highest during the summer	Increased production with increased sunlight	Check for differences in average concentrations as a function of season
Annual	Possible	May correlate with decreased concentrations of VOCs	Check for interannual differences, correlate with total non-methane organic carbon concentrations
Urban/Rural	Yes-likely higher urban	Urban has automobile emissions and industry; rural sites may be influenced by local forests	Check for urban/rural differences. Determine sources in each environment
Regional	Yes-highest in urban centers	Same as urban/rural	Same as urban/rural
Global	Yes	Same as urban/rural	Same as urban/rural

***Possible interspecies correlations*:** Acetaldehyde should show the strongest correlation with formaldehyde. May also correlate reasonable well with benzene, 1,3-butadiene, and total nonmethane organic carbon.

Typical Urban Concentrations: 1999 from historical database: 2.0-5.1 $\mu g/m^3$
2001-2002 from pilot city study: 1.8-3.8 $\mu g/m^3$

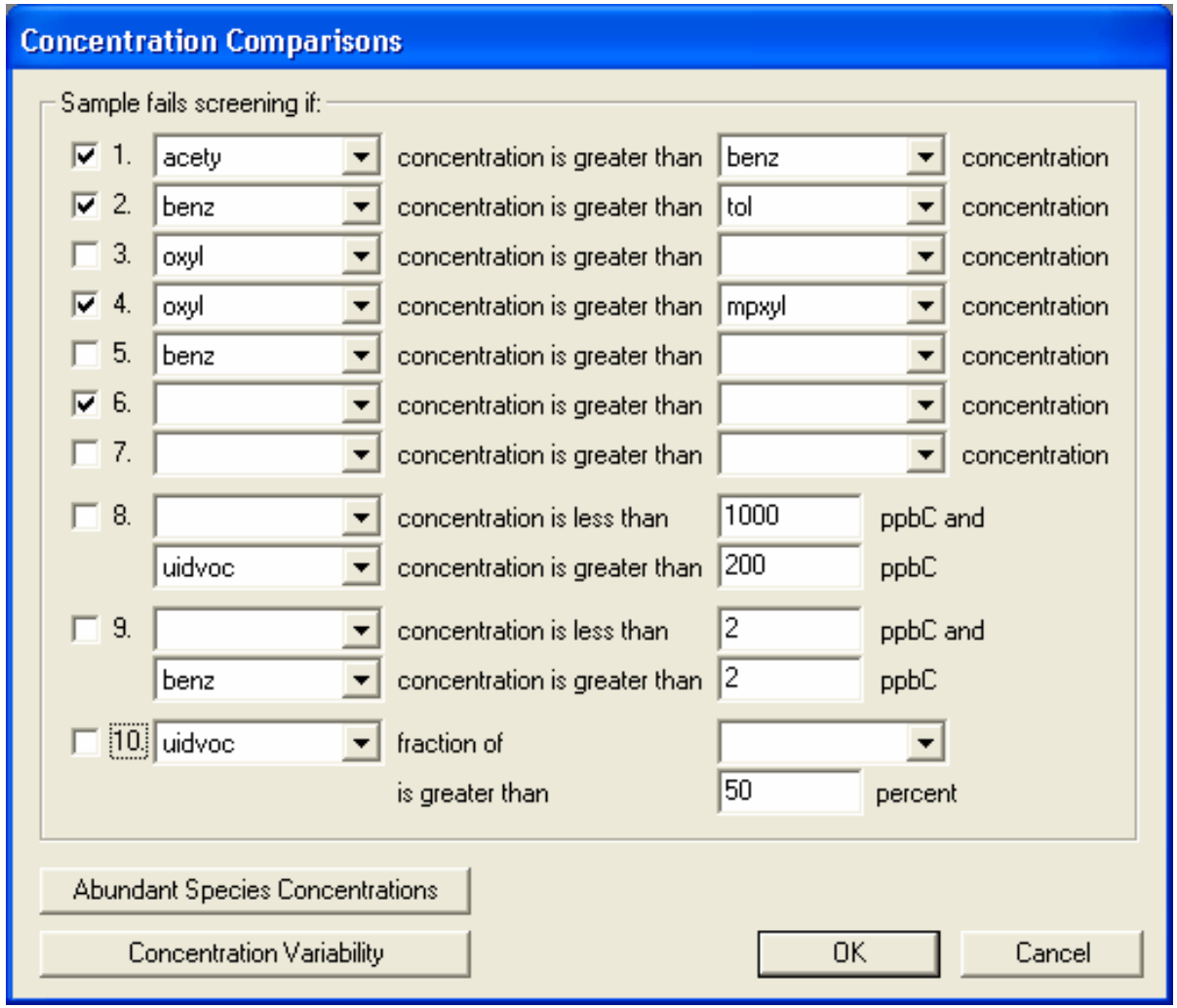
VOCDat Screening Checks (1 of 3)

Abundant Species Concentrations



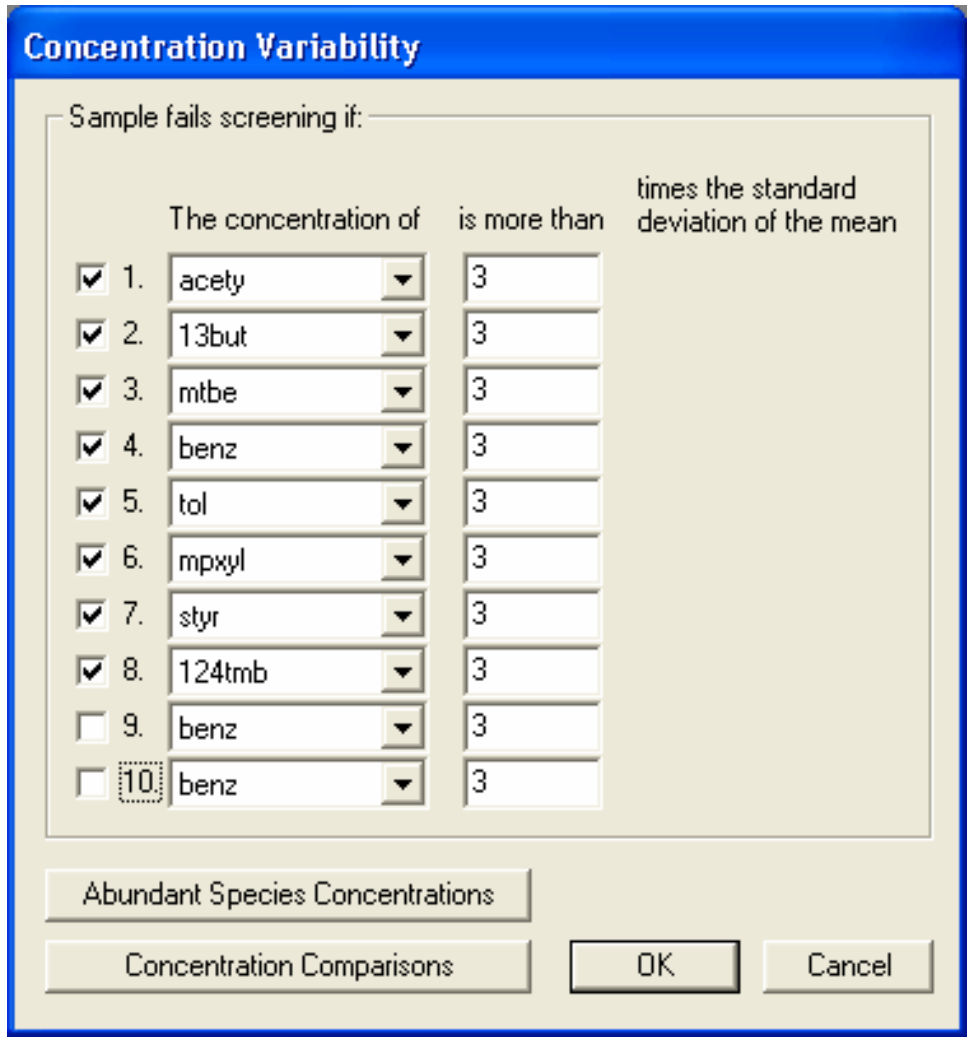
VOCDat Screening Checks (2 of 3)

Abundant Species Comparisons

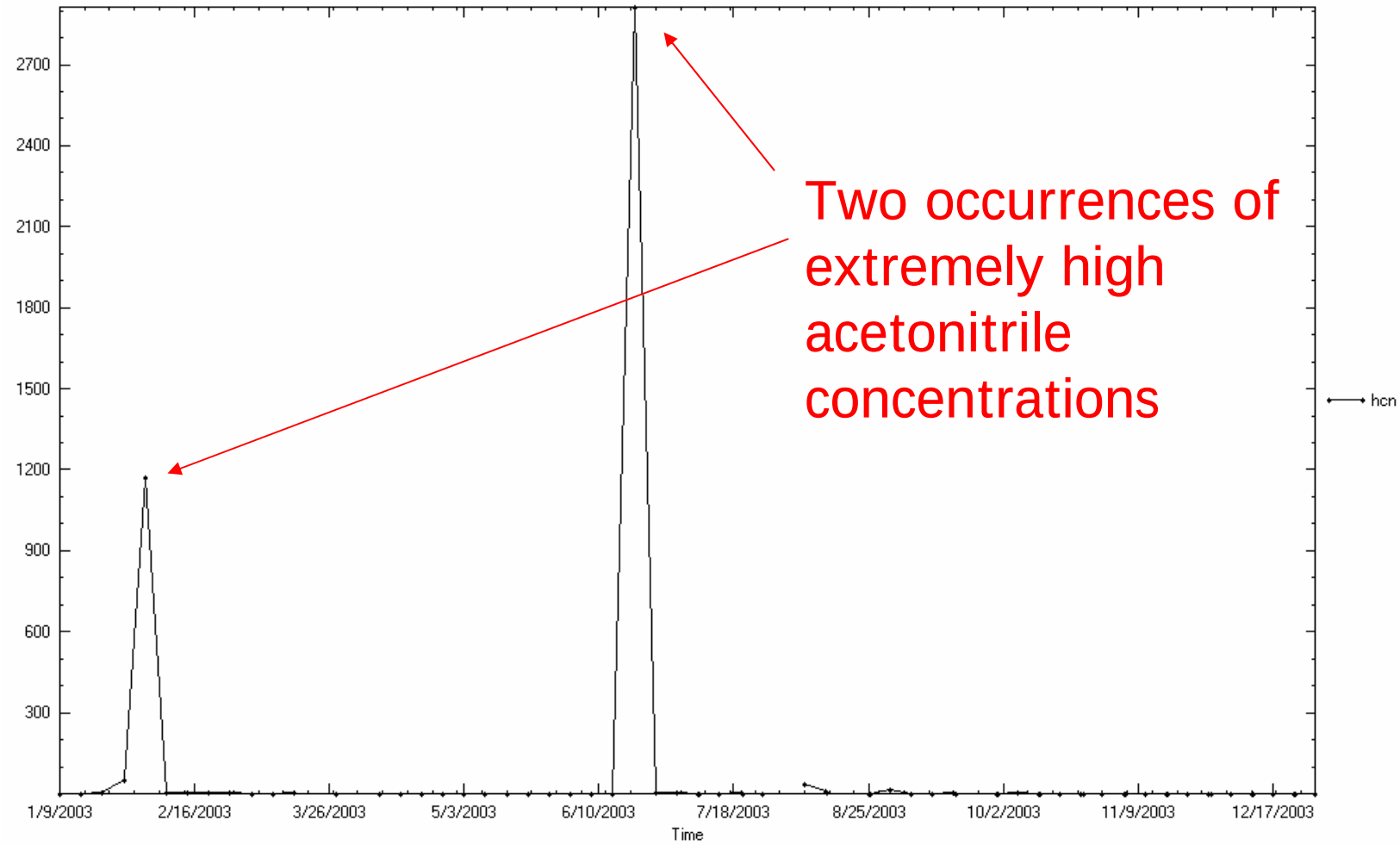


VOCDat Screening Checks (3 of 3)

Abundant Species Concentration Variability



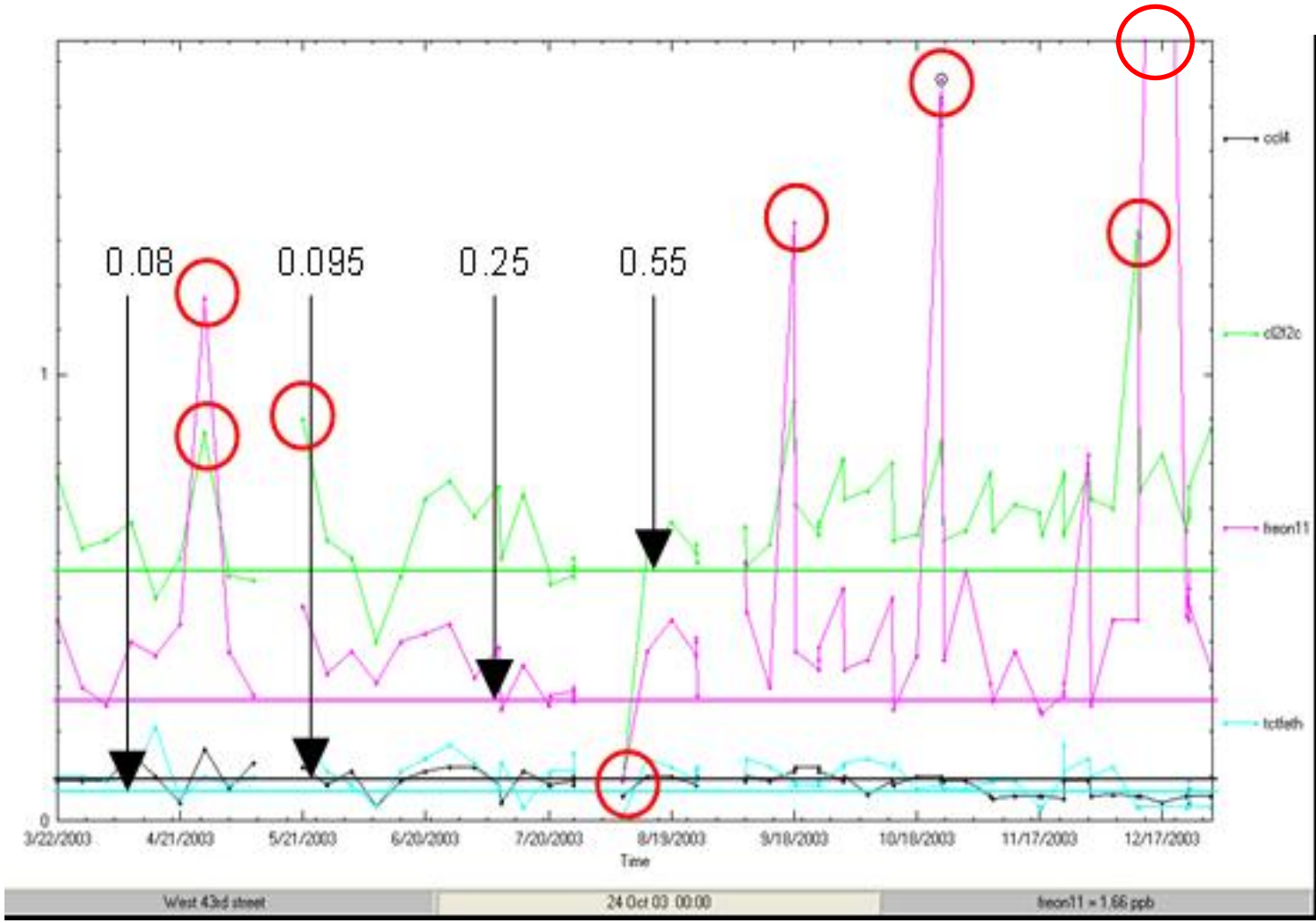
Example: Extreme Events



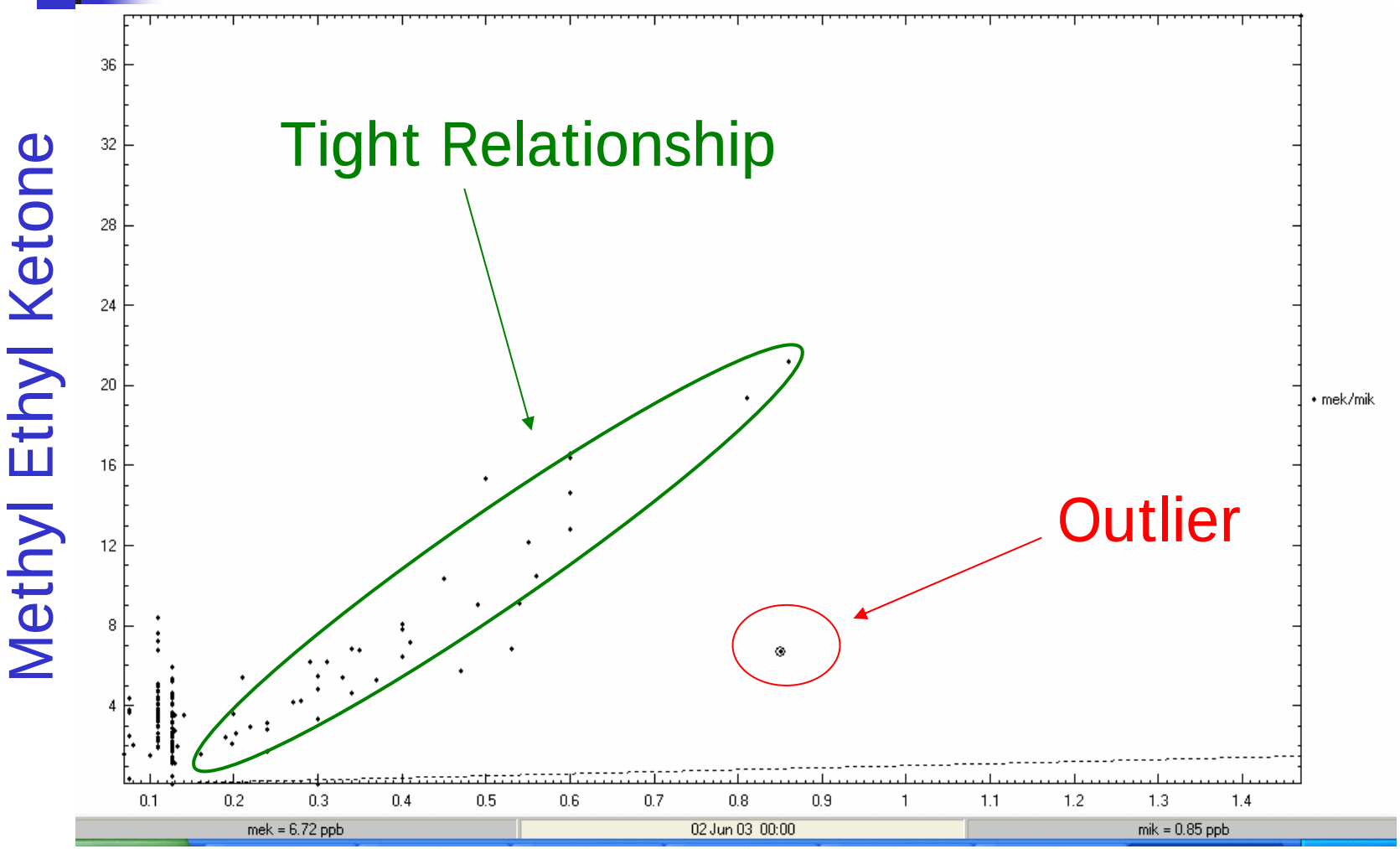
Example: Concentrations Exceeding Background

These CFCs are regional/global, and should not be significantly different than background.

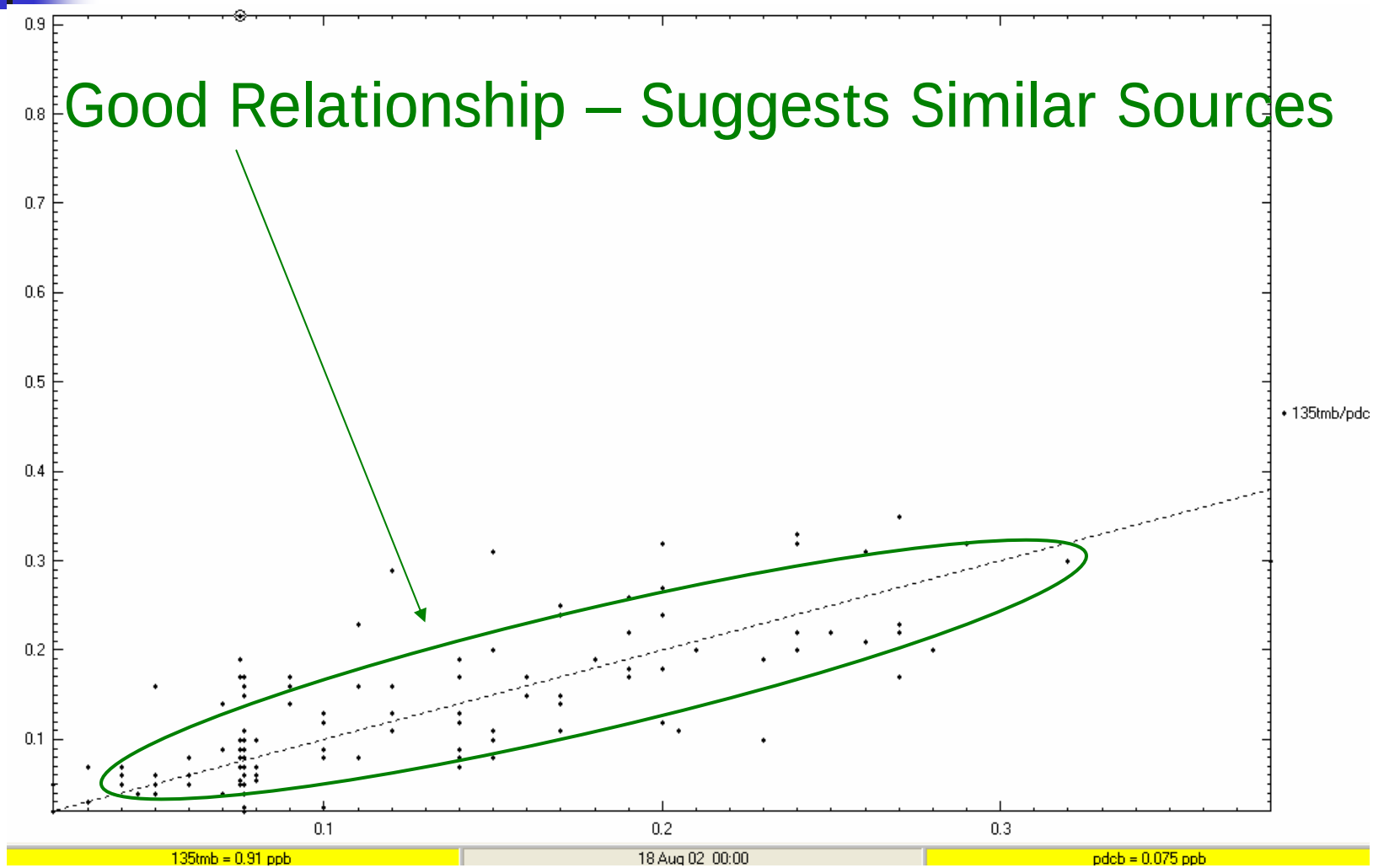
Spikes in concentrations may be from nearby sources or analytical error



Example: Species Relationships

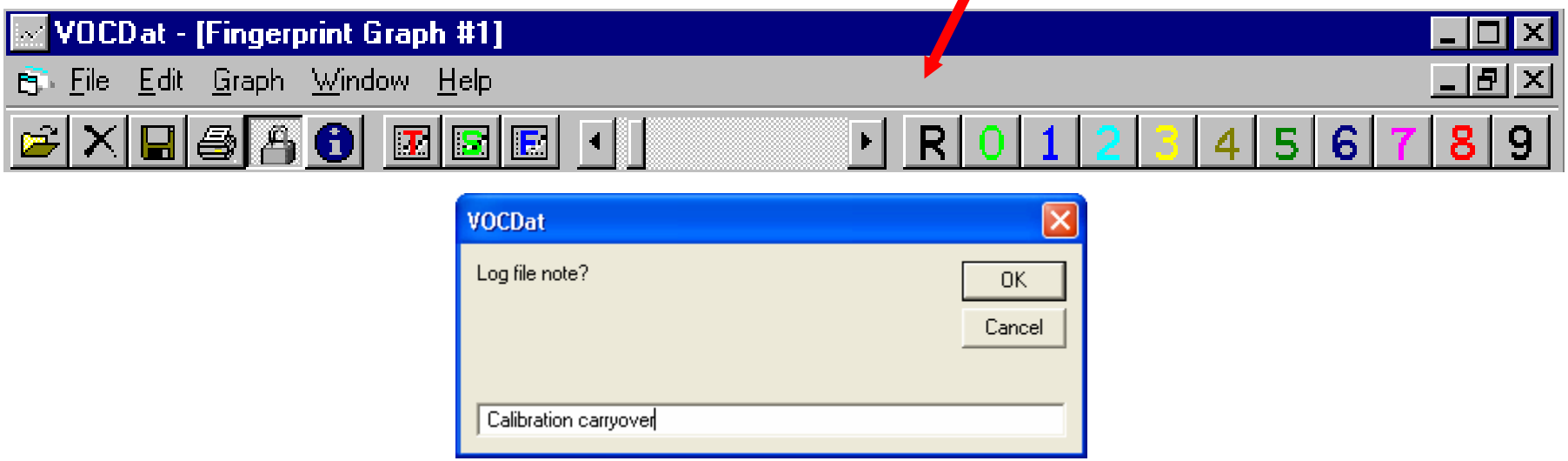


Example: Surprising Relationships



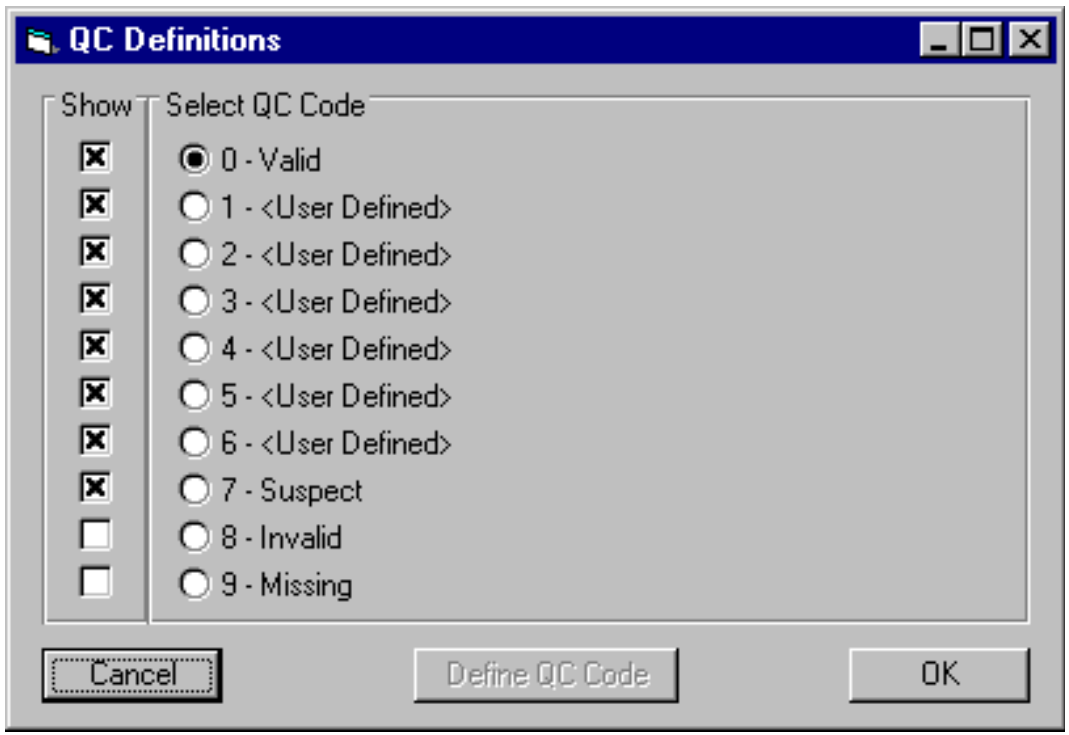
Flagging Suspect/Invalid Data (1 of 2)

Once you identify erroneous data, use the VOCDat toolbar to flag either the entire sample record or a single parameter

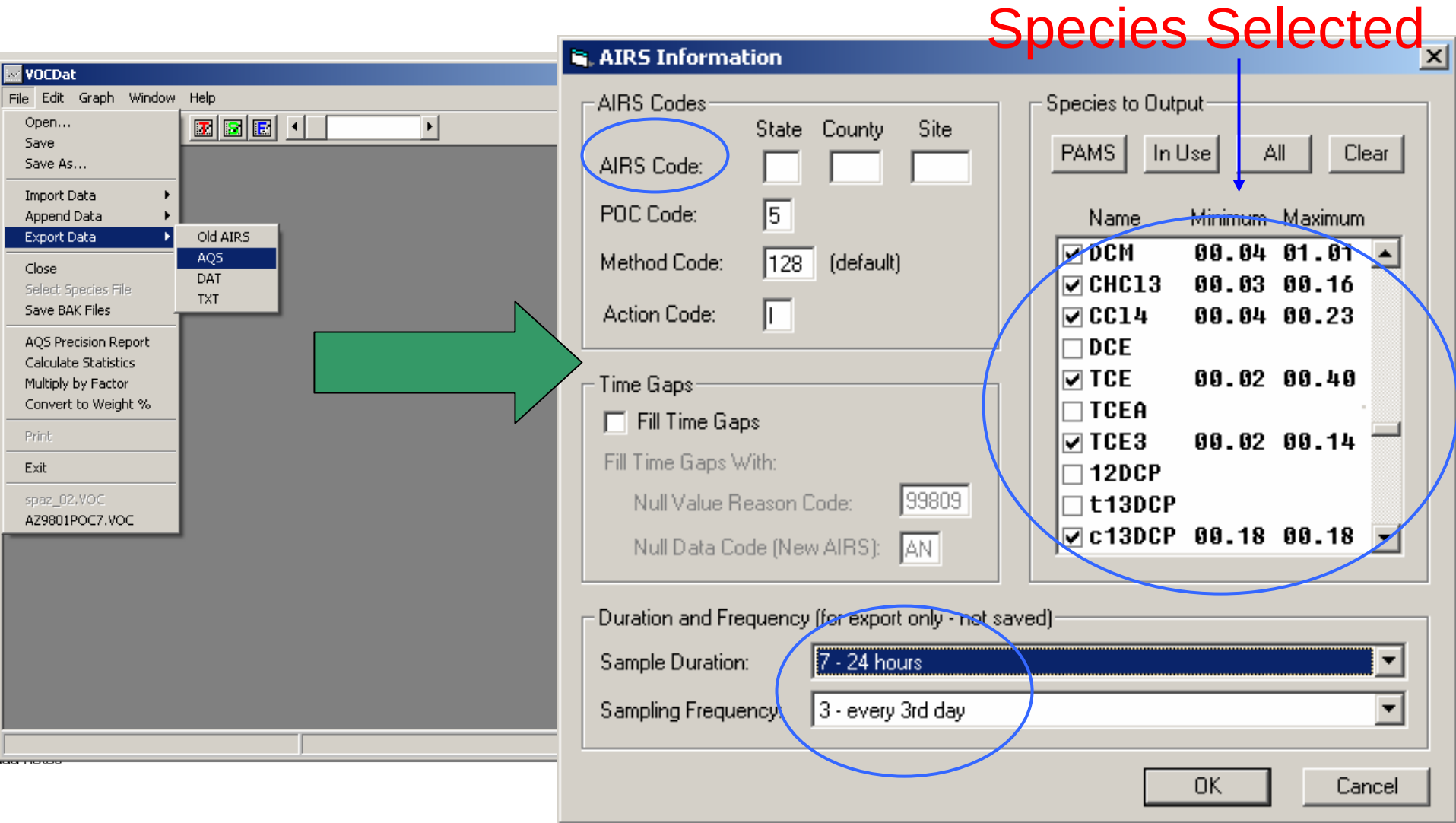


Flagging Suspect/Invalid Data (2 of 2)

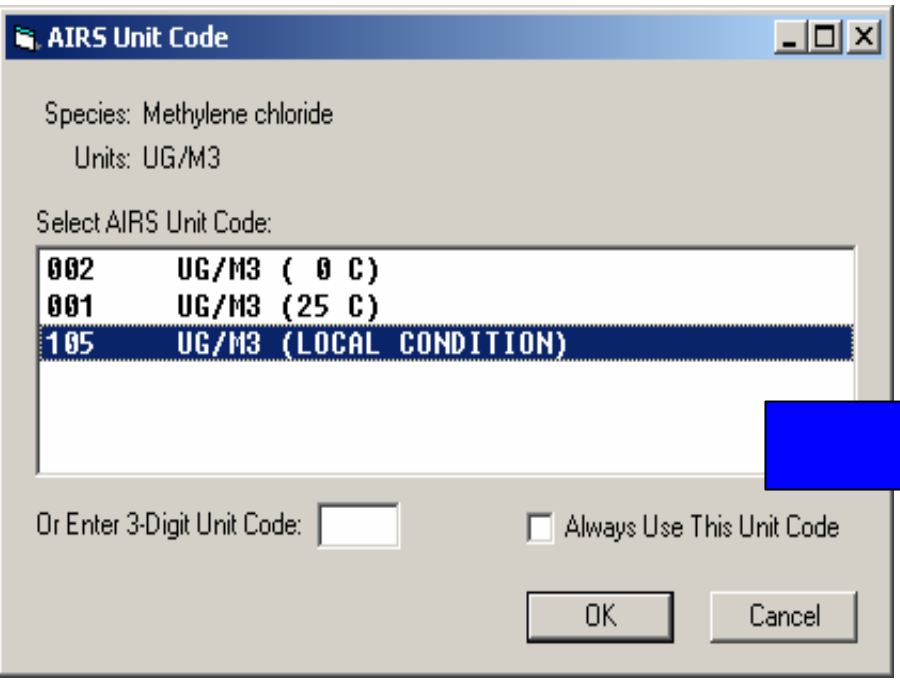
Assign user-defined codes to indicate the problem



Export into AQS Format I

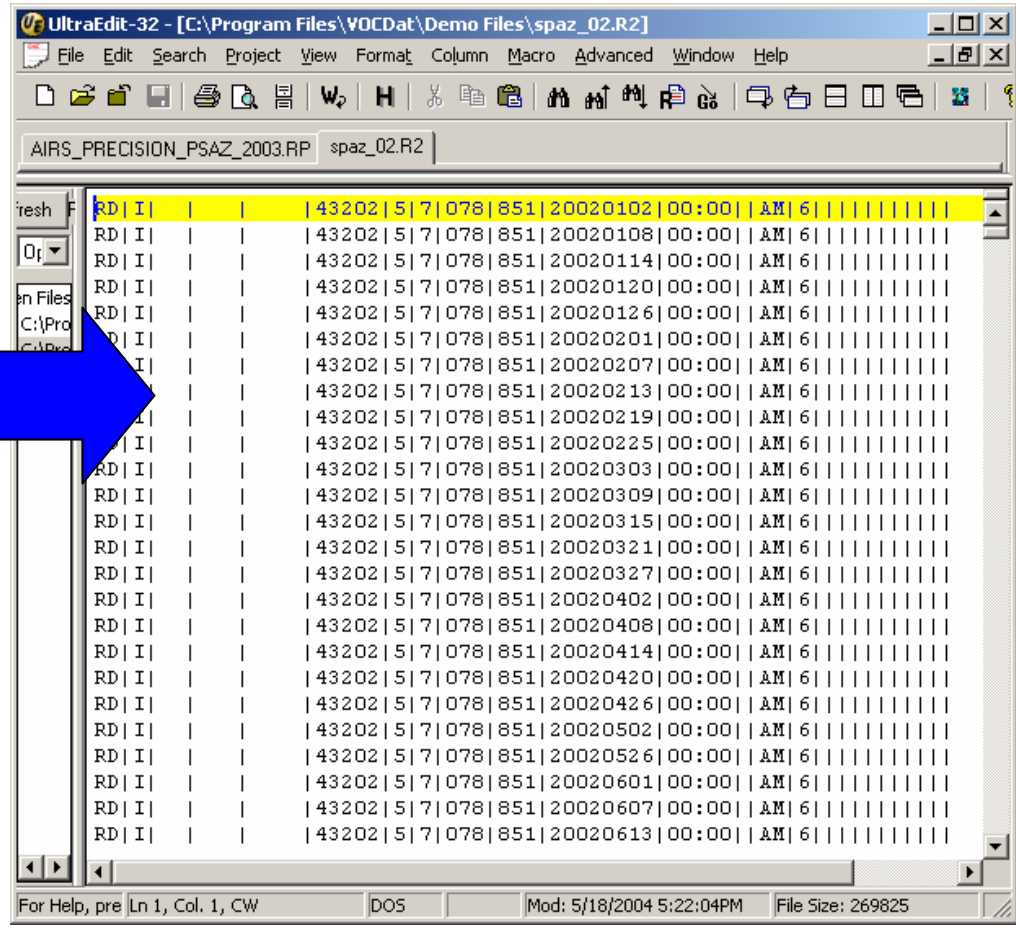


Exporting into AQS Format II



Select Units

Ready in R2 AQS format!

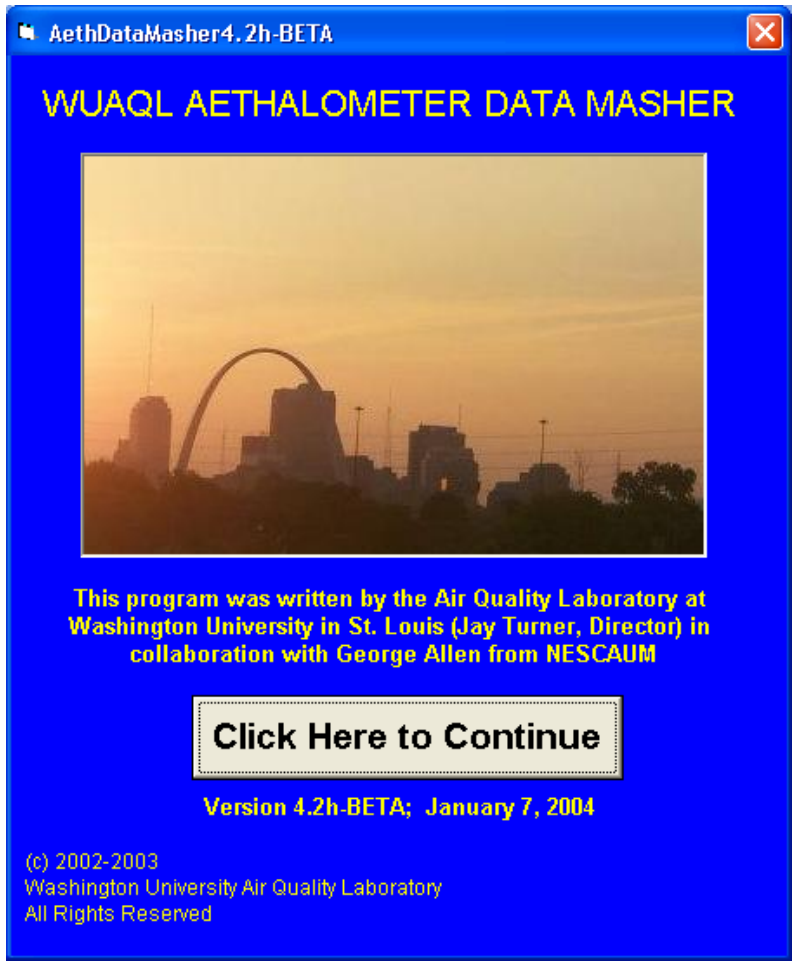


Aethalometer (Black Carbon)

- NATTS sites have aethalometers to monitor black carbon (BC), a marker (not unique!) for diesel particulate matter (DPM)
- What to do with the data?
- “Masher” developed by Jay Turner (Washington University, St Louis)
- Can bring the exported data from the aethalometer directly into the Masher
- Then into VOCDat! Ready for validation and export into AQS format.

Data Masher

- 5-minute or 1-hour averaged data can be input into Masher
- Calculated hourly averages
- Voids hours with <75% completeness
- Available from Jay Turner: JRTurner@seas.wustl.edu



Example Import: Aethalometer Data (1 of 3)

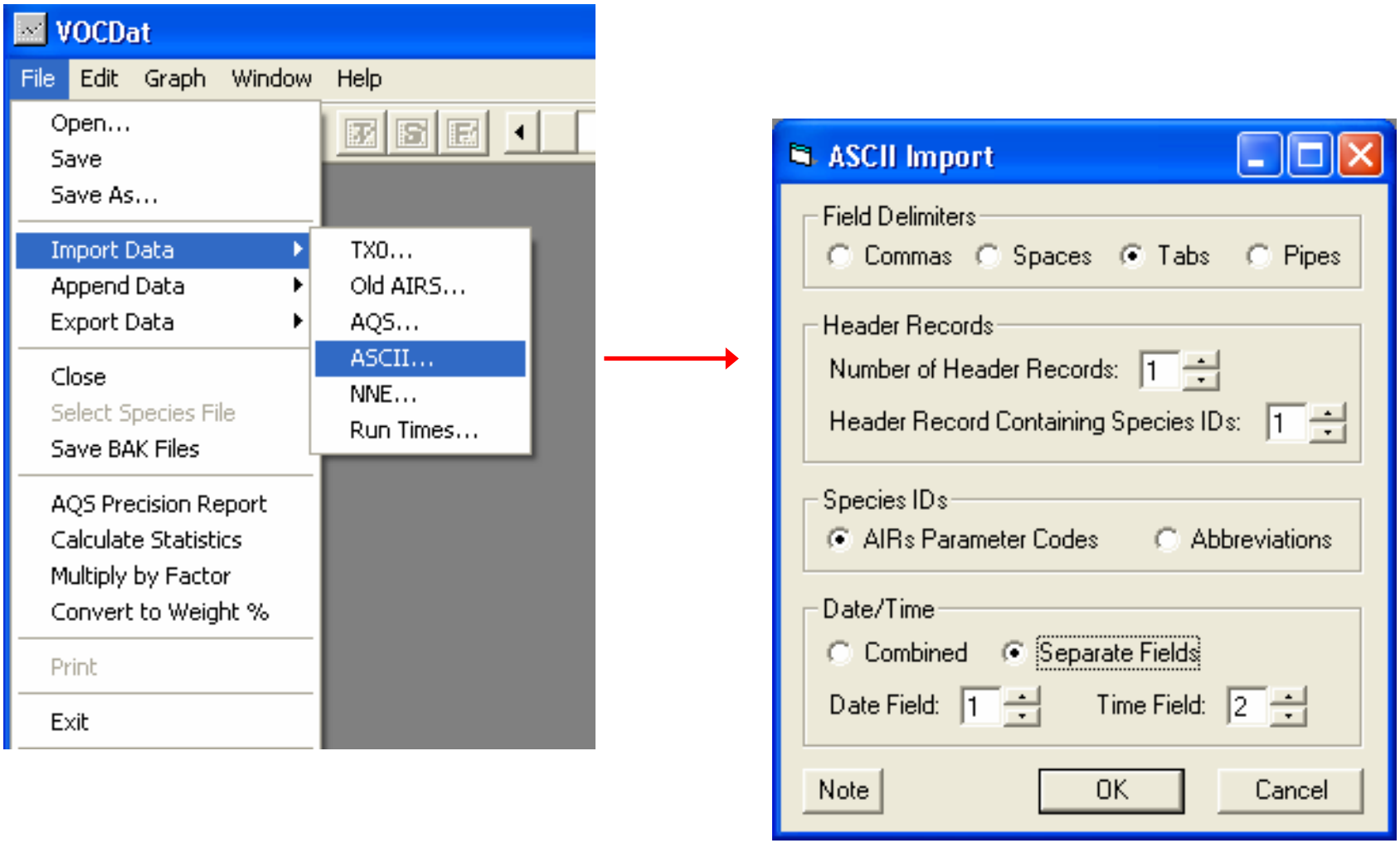
- Data are provided in a comma-separated text file with no header records.
- Save the file as tab-delimited txt (or CSV).

Date	time	84313
8/5/2003	21:00:00	0.251
8/5/2003	22:00:00	0.322
8/5/2003	23:00:00	0.19
8/6/2003	0:00:00	0.149
8/6/2003	1:00:00	0.289
8/6/2003	2:00:00	0.519
8/6/2003	3:00:00	0.316
8/6/2003	4:00:00	0.338
8/6/2003	5:00:00	0.511

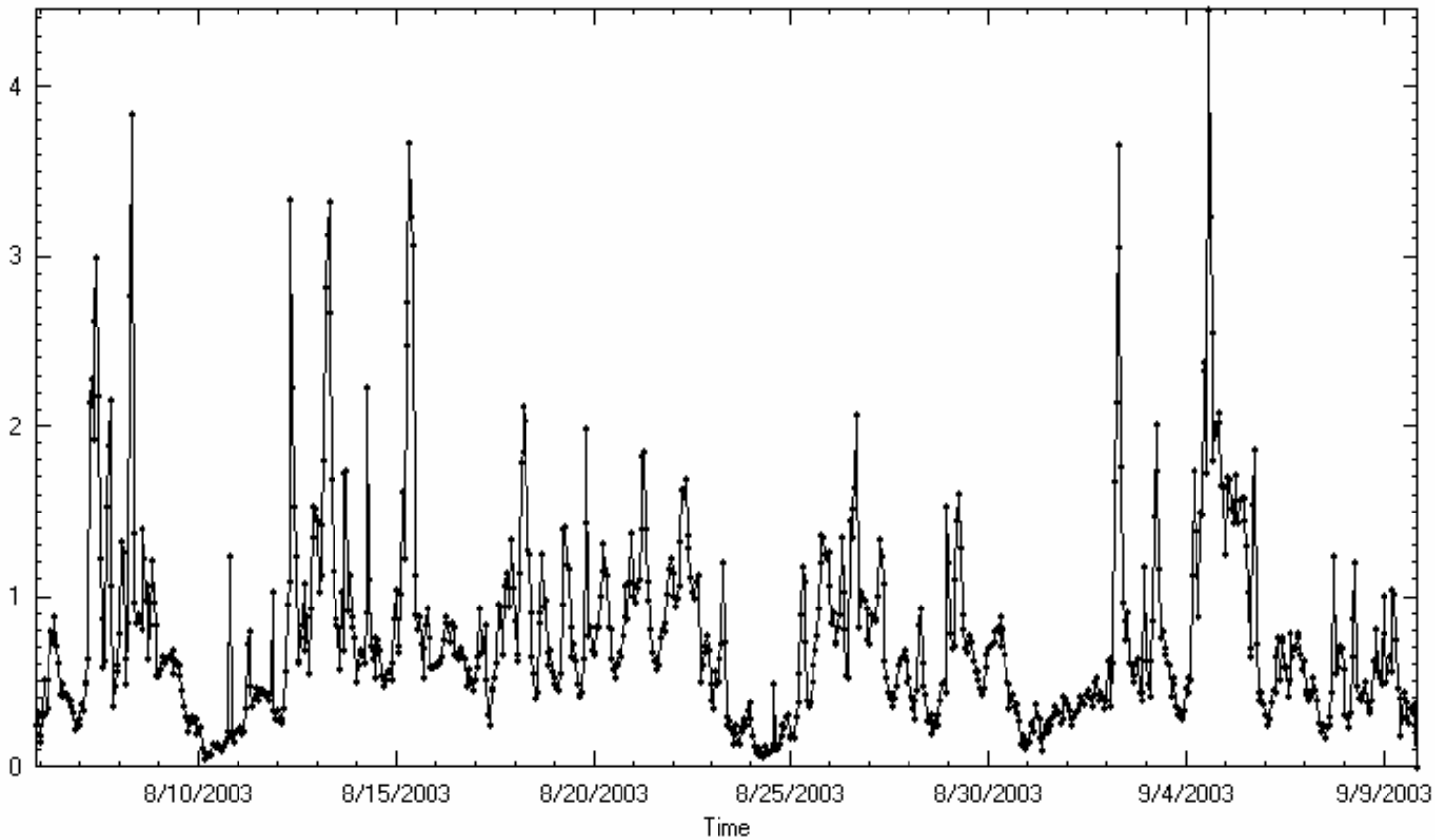
In VOCDat Species file:

84313	Black carbon PM2.5 STP
88313	Black Carbon PM2.5 LC
84313	Black carbon PM2.5 STP
88313	Black Carbon PM2.5 LC

Example Import: Aethalometer Data (2 of 3)



Example Import: Aethalometer Data (3 of 3)



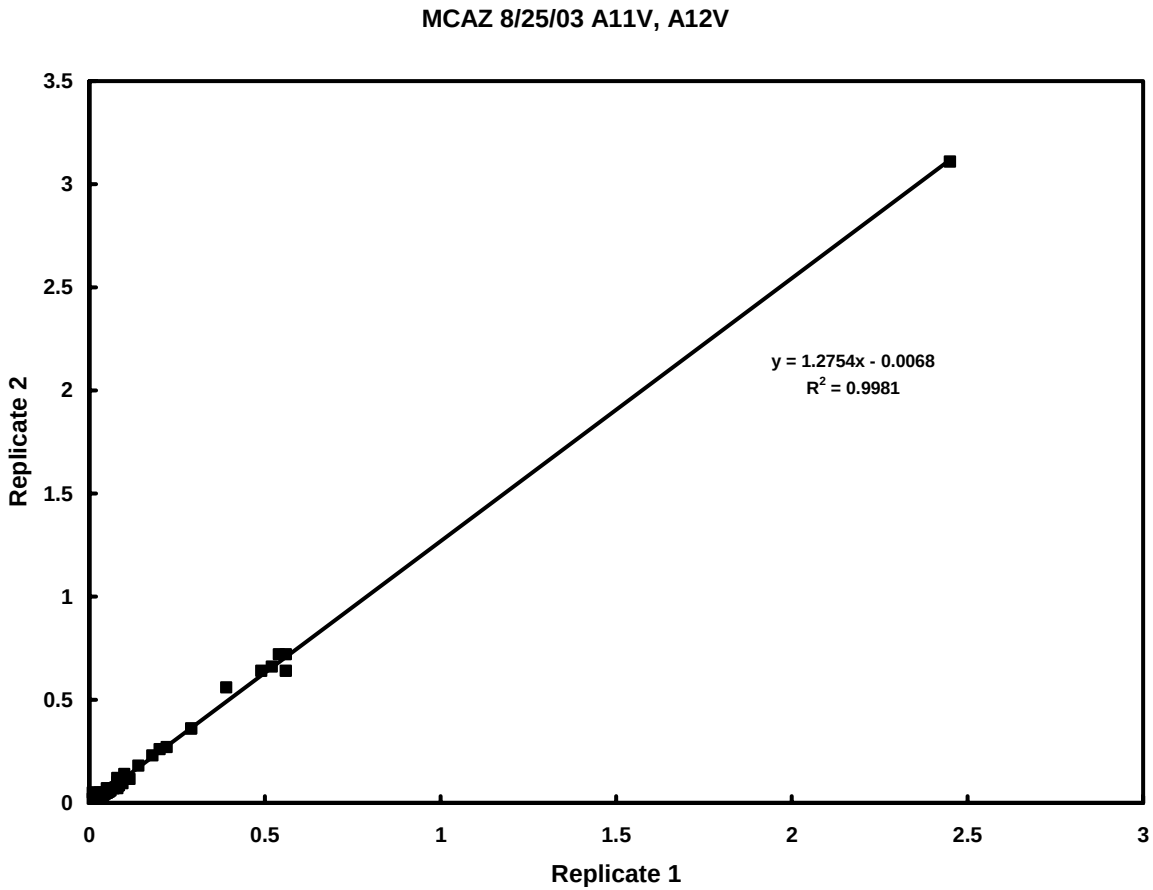
Ready for validation and export into AQS format!

Precision Report

- New in 2003
- Allows reporting of duplicate (collocated) and replicate (second chemical analysis) samples into AQS
- Used to determine
  - Precision
  - Reproducibility of results
  - Confidence in measurements and sampling

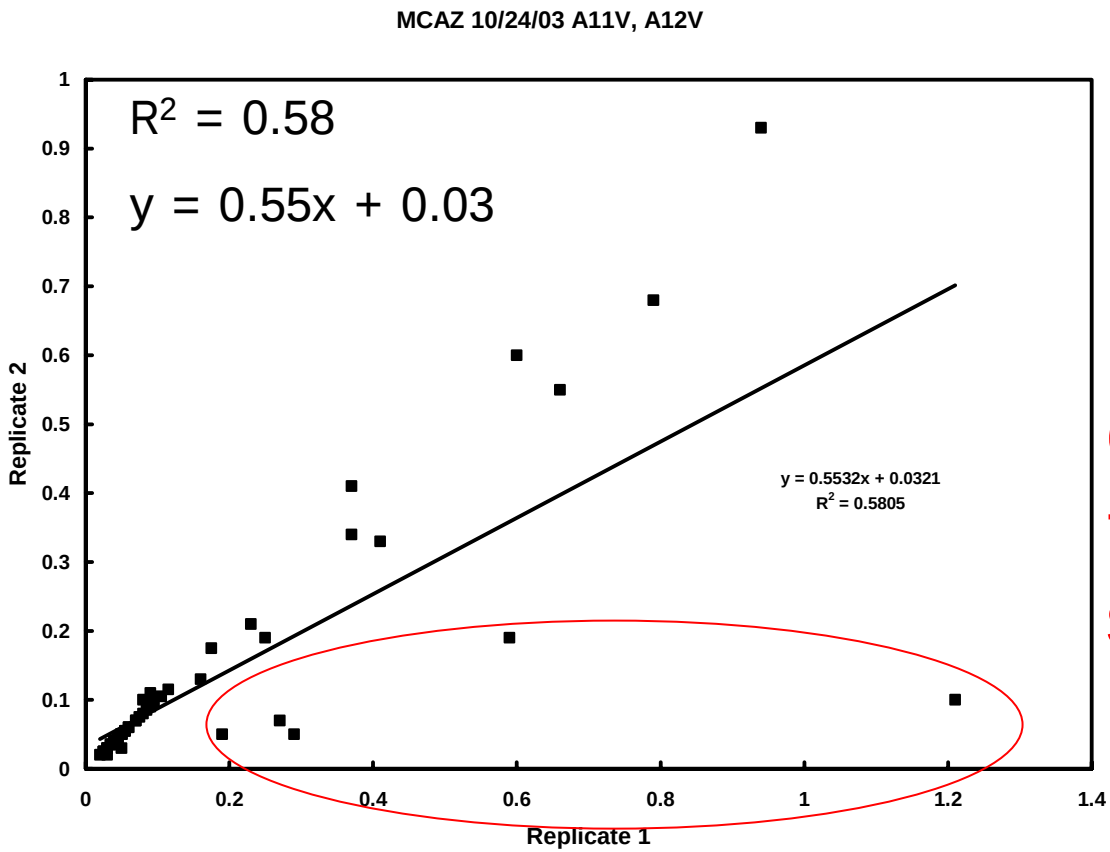
Usefulness of Precision Report

Compare Replicate Measurements (Good Example)



Usefulness of Precision Report

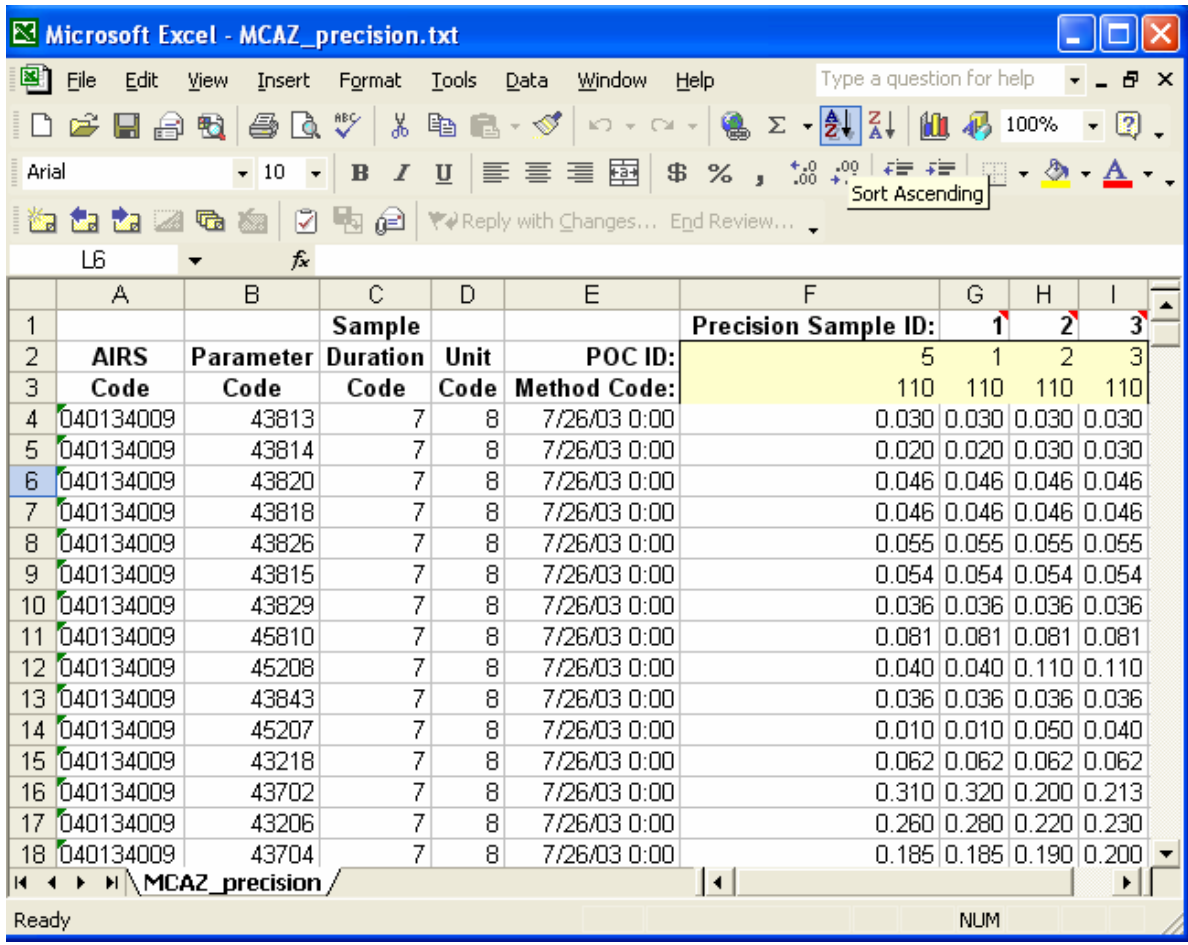
Compare Replicate Measurements (Bad Example)



Outliers flagged as suspect (poor reproducibility)

Precision Report Format

- Text formatted cells were required to keep zeroes in front of AIRS Site Code.
- POC IDs indicate, from left to right, primary sample, duplicate, replicate, and duplicate-replicate samples.



Microsoft Excel - MCAZ_precision.txt

File Edit View Insert Format Tools Data Window Help

Sort Ascending

L6

	A	B	C	D	E	F	G	H	I
1			Sample			Precision Sample ID:	1	2	3
2	AIRS	Parameter	Duration	Unit	POC ID:	5	1	2	3
3	Code	Code	Code	Code	Method Code:	110	110	110	110
4	040134009	43813	7	8	7/26/03 0:00	0.030	0.030	0.030	0.030
5	040134009	43814	7	8	7/26/03 0:00	0.020	0.020	0.030	0.030
6	040134009	43820	7	8	7/26/03 0:00	0.046	0.046	0.046	0.046
7	040134009	43818	7	8	7/26/03 0:00	0.046	0.046	0.046	0.046
8	040134009	43826	7	8	7/26/03 0:00	0.055	0.055	0.055	0.055
9	040134009	43815	7	8	7/26/03 0:00	0.054	0.054	0.054	0.054
10	040134009	43829	7	8	7/26/03 0:00	0.036	0.036	0.036	0.036
11	040134009	45810	7	8	7/26/03 0:00	0.081	0.081	0.081	0.081
12	040134009	45208	7	8	7/26/03 0:00	0.040	0.040	0.110	0.110
13	040134009	43843	7	8	7/26/03 0:00	0.036	0.036	0.036	0.036
14	040134009	45207	7	8	7/26/03 0:00	0.010	0.010	0.050	0.040
15	040134009	43218	7	8	7/26/03 0:00	0.062	0.062	0.062	0.062
16	040134009	43702	7	8	7/26/03 0:00	0.310	0.320	0.200	0.213
17	040134009	43206	7	8	7/26/03 0:00	0.260	0.280	0.220	0.230
18	040134009	43704	7	8	7/26/03 0:00	0.185	0.185	0.190	0.200

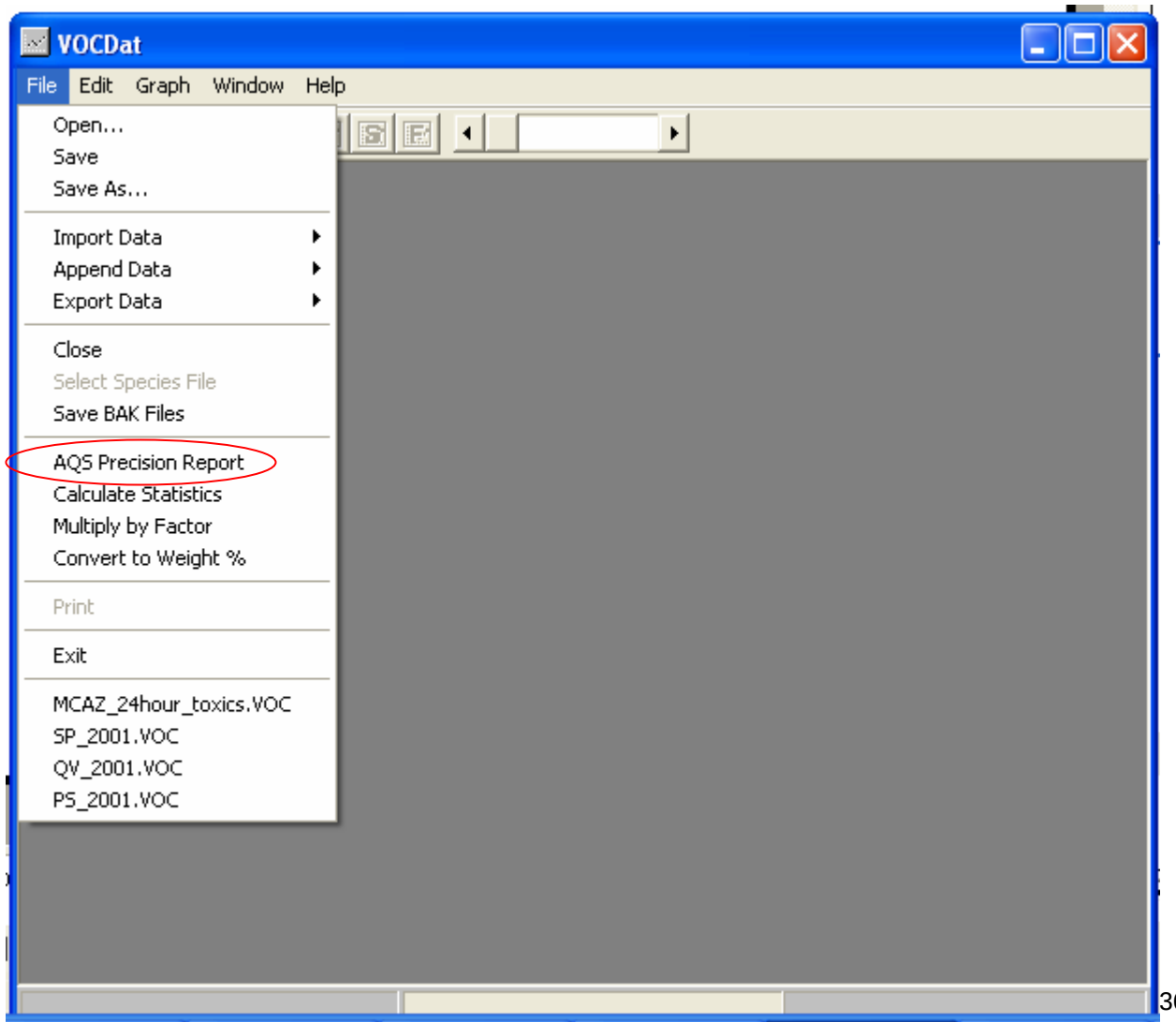
MCAZ_precision

Ready NUM

Example Setup File Available

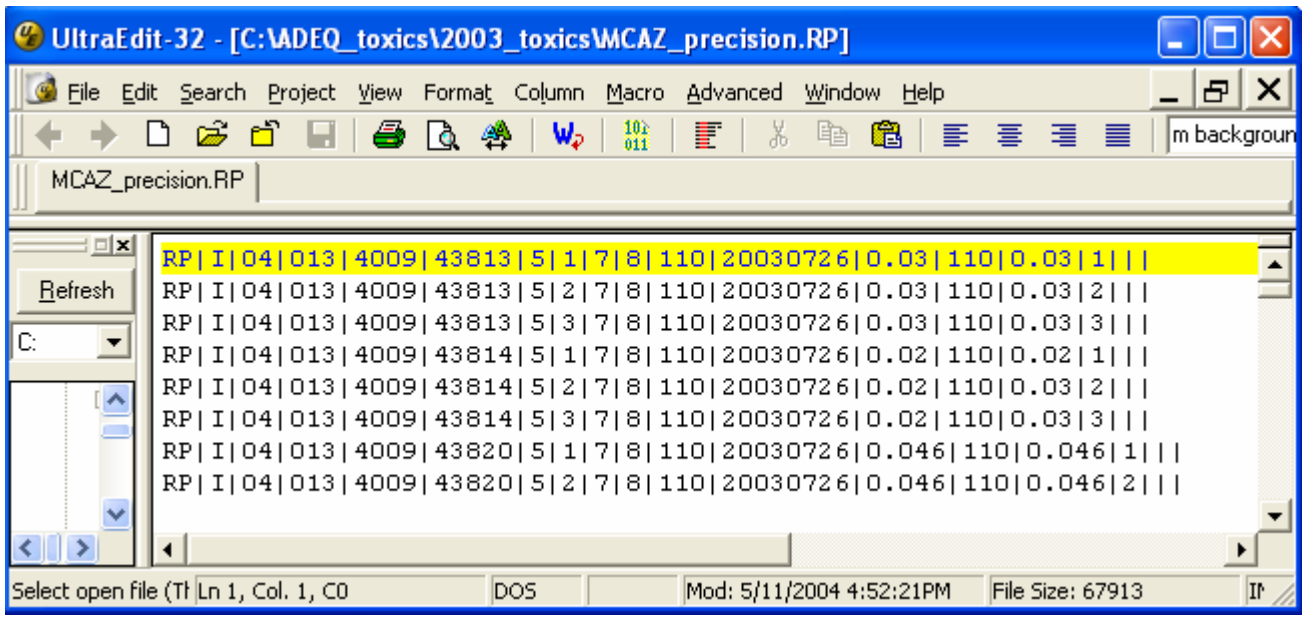
Creating Precision Report

- Import the text file (*.txt) into VOCDat by opening AQS Precision Report in the file menu.

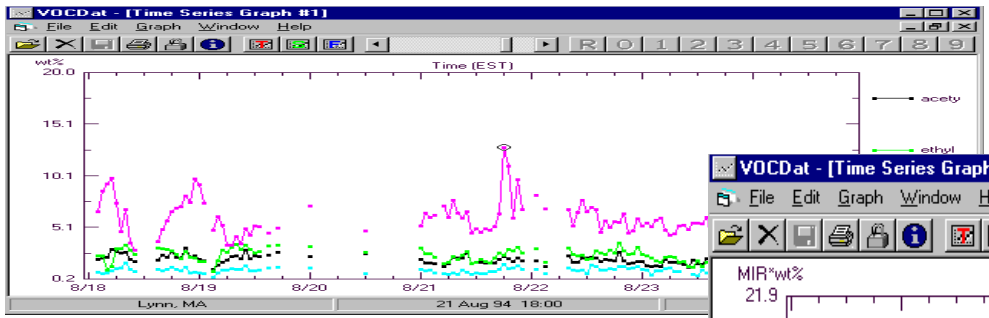


Resultant Precision Report

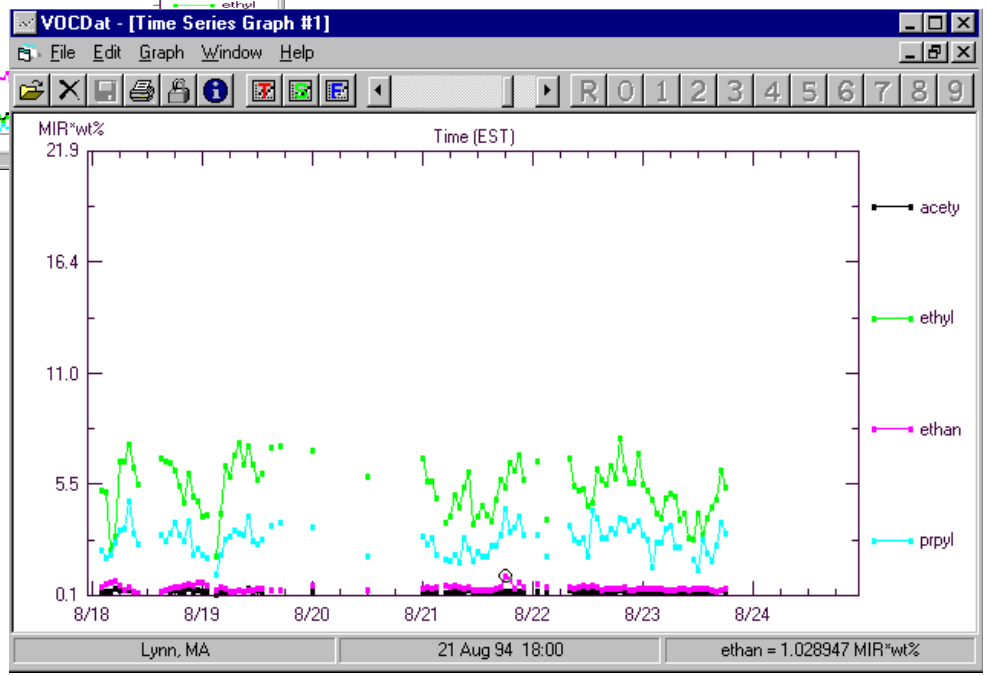
- Upon importing the file into VOCDat, the data is automatically exported as an *.RP file which is saved in the same folder as the imported text file.



Using Other Forms of the Data



Weight Percent



Reactivity-weighted or Risk-weighted

Toxicity/Risk Factors

- EPA's Integrated Risk Information System – IRIS (see http://www.epa.gov/iris/index.html).
- EPA Health Effects Notebook for Hazardous Air Pollutants (see http://www.epa.gov/ttn/atw/hapindex.html).
- EPA National Center for Environmental Assessment (NCEA) (see http://cfpub.epa.gov/ncea/cfm/nceahome.cfm).
- California EPA (see http://www.oehha.ca.gov/).
- Modeling Cumulative Outdoor Concentrations of Hazardous Air Pollutants by Systems Applications International, Inc. - Revised Final Report (see http://www.epa.gov/cumulativeexposure/resource/report.htm).
- California Air Resources Board Toxic Air Contaminant Identification List Summaries (see http://www.arb.ca.gov/toxics/tac/tac.htm).
- Application of Health Information to Hazardous Air Pollutants Modeled in EPA's Cumulative Exposure Project by Jane C. Caldwell, Tracey J. Woodruff, Rachel Morello-Frosch, and Daniel A. Axelrad (1998) (see http://www.epa.gov/cumulativeexposure/CEPpapers/paperCWMA.pdf).

New Additions – Unit Conversion

- Select units when bringing into VOCDat
- Convert among
  - ppbv (gaseous and semi-volatile HAPs)
  - ppbC (PAMS, auto-GC, VOC)
  - µg/m^3 (semi-volatile, particulate HAPs)
- µg/m^3 is focus for risk assessment
- Must assume pressure, temperature when converting, which can cause up to 20% variability/uncertainty

New Additions – Import Formats

- Currently working to be able to import data in additional file formats from labs:
  - ERG (HAPs)
  - RTI ($PM_{2.5}$)
- Some consistency will be needed in format
- Communicate with labs, EPA and STI – let us know what formats you want and could more easily use, with examples

Obtaining VOCDat

Available free at

ftp://ftp.sonomatech.com/public/vocdat/

To register and get technical support, call

Hilary Hafner or Steve Brown

at (707) 665-9900

or e-mail

VOCDat@sonomatech.com

Problems? Issues?

- User's guide has lots of good information (updates coming soon)
- Data validation and VOCDat workshop material available online:
  - http://www.sonomatechdata.com/sti_workbooks/
  - http://www.ladco.org/toxics.html
- Still can't get it to work? E-mail VOCDat@sonomatech.com
  - Specific problem file, species file, note version, screenshots of specific error

Summary

- Graphical techniques are useful to quickly and efficiently validate (and begin analysis of) aerometric data.
- VOCDat is extremely useful in processing and validating air toxics data
- New additions:
  - Precision report
  - New file formats
  - Unit conversions

Other Additions and Ideas?

- As always, a major VOCDat overhaul would be great...
- Other import file formats?
- Additional QC checks?
- Can we input comments into AQS?
- Is there a standard QC code list for comments?
- Web site for FAQs

Glossary

AIRS – Aerometric Information Retrieval System
AQS – Air Quality System
auto-GC – auto-Gas Chromatography
BC – black carbon
CFCs – chlorofluorocarbons
CSV – comma separated values
DPM – diesel particulate matter
EPA – U.S. Environmental Protection Agency
ERG – Eastern Research Group
HAPs – Hazardous Air Pollutants
NATTS – National Air Toxics Trends Sites
PAMS – Photochemical Assessment Monitoring Stations
PM – Particulate Matter
POC – parameter occurrence code
QC – Quality Control
RTI – Research Triangle Institute
VOC – Volatile Organic Compounds