Using VOCDat to Validate and Process Air Toxics Data

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

U.S. Environmental Protection Agency
16th Annual National Air Quality Systems Conference
San Antonio, TX
June 6, 2006

Outline of Presentation

- Overview
- After obtaining data from laboratories
  - Using VOCDat for processing and validating data
  - Importing and exporting data
  - Validating examples and ideas
- Recent additions to VOCDat
- Hands-on training
- What other features would be helpful?

VOCDat: How Did We Get Here?

- Originally developed for Photochemical Assessment Monitoring Stations (PAMS) VOC data to facilitate
  - Importing data from AQS/AIRS format
  - Exporting data into AQS/AIRS format
  - Viewing and validating data before submitting to AQS!
- Applicable to other data sets, including air toxics, $PM_{2.5}$, and continuous data

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!

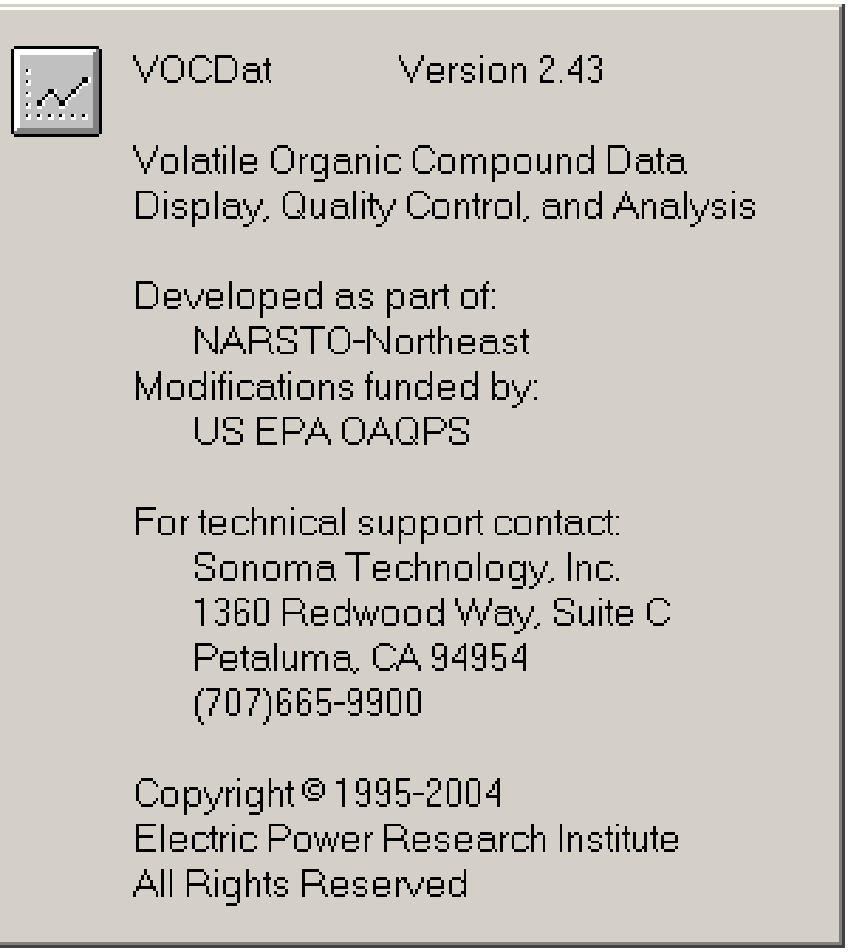
* U.S. Environmental Protection Agency (1984) Quality assurance handbook for air pollution measurement systems: Volume II. Ambient air specific methods. Sections 2.1, 2.2, 2.6, and 2.9. Report prepared by the Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-600/4-77-027a, July.

Example Approach Using VOCDat

- Obtain air toxics data from laboratory
- Import into VOCDat
- Review data quality and apply screening criteria
- Flag invalid data
- Export into AQS-ready format
- Prepare and report precision information

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)

- Imports and exports AQS AMP370 (old AIRS) format
- Imports and exports AQS R2 format
- Imports flexible format ASCII files
- Imports Turbochrome and TotalChrome (auto-GC) formats
- Exports formats suitable for other software (such as spreadsheets or databases)
- Edits data quality control (QC) codes on screen (keeps a log of changes)
- Prepares 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, and 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, and units
  - Several species lists are available
  - Only species in the list will be exported
- Once VOCDat knows what species to expect, you can input your data from AQS, text, etc.
- VOCDat also allows supplemental data handling
  - Species2.txt allows use of meteorological and criteria pollutant data in validation
  - These are not exported into AQS format

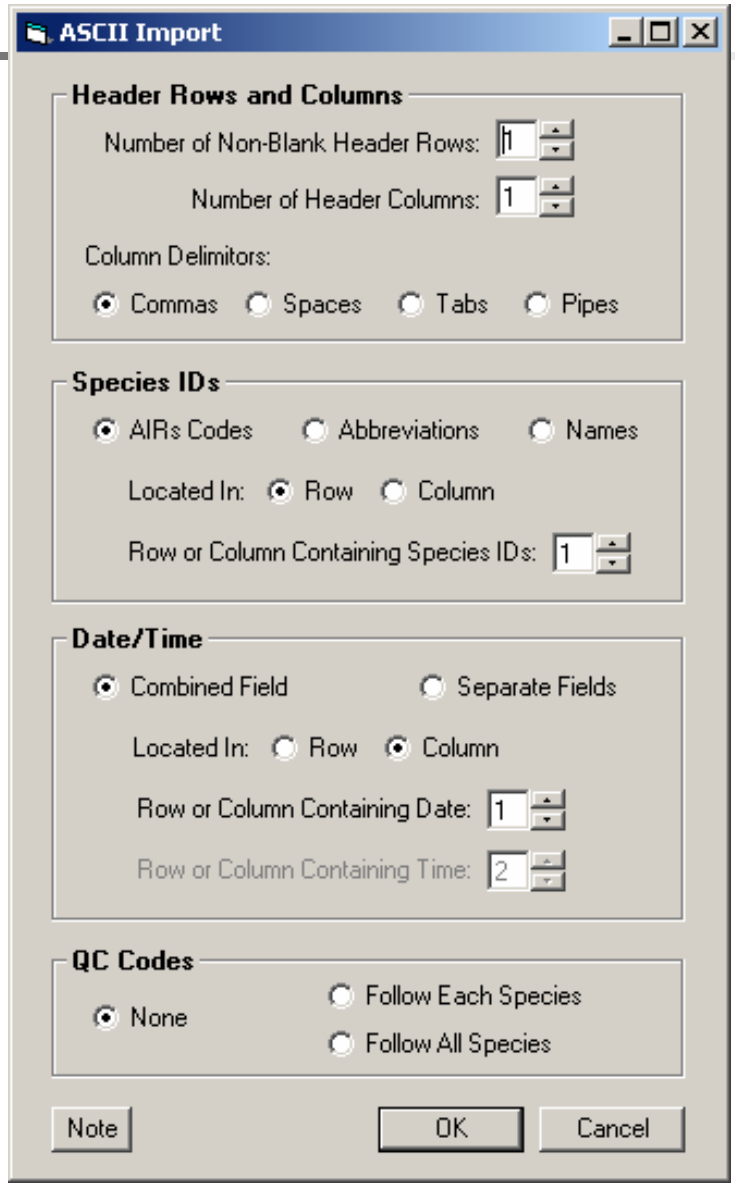
VOCDat Terminology

- Species list: list of species you tell VOCDat to expect; includes units and method code
- Import: bring data from various formats into VOCDat
- Open: open data in *.VOC file; these data have already been imported into VOCDat and saved
- Save: save current data as *.VOC file, including any validation flags
- Export: export current data into AQS or txt format, including any validation flags

From the Beginning

- Once you acquire air toxics data from the laboratory, then what?
- Select TO-15 method species list
  - List may be modified to accommodate other species
- Check format of data received
  - Are data in one of a number of specific formats?
  - If not, work with laboratories to get data in a more useful format — or, know anyone who likes Access? Use their knowledge for laboratory data!
  - Can now import database-style data (in ASCII format)
- Import into VOCDat

Importing ASCII Files



Validation Approach (1 of 2)

- 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 (2 of 2)

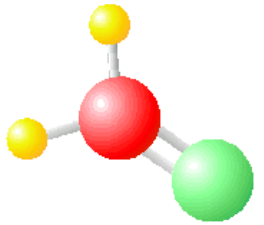
- Inspect all species
  - Time series plots
  - Scatter plots (internal consistency)
- Inspect every sample
  - Fingerprint plots (in elution order)
- Flag data and document modifications
- Prepare data for AQS

Making Validation Easier

- Once data are VOCDat-ready, validation is easy!
- With the EPA, STI has developed “cheat sheets” for 18 HAPs
  - Available at www.ladco.org/toxics.html
  - Useful in analysis!
- Easily generated simple graphs allow quick validation
- VOCDat user’s guide not as useful for examples and ideas

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

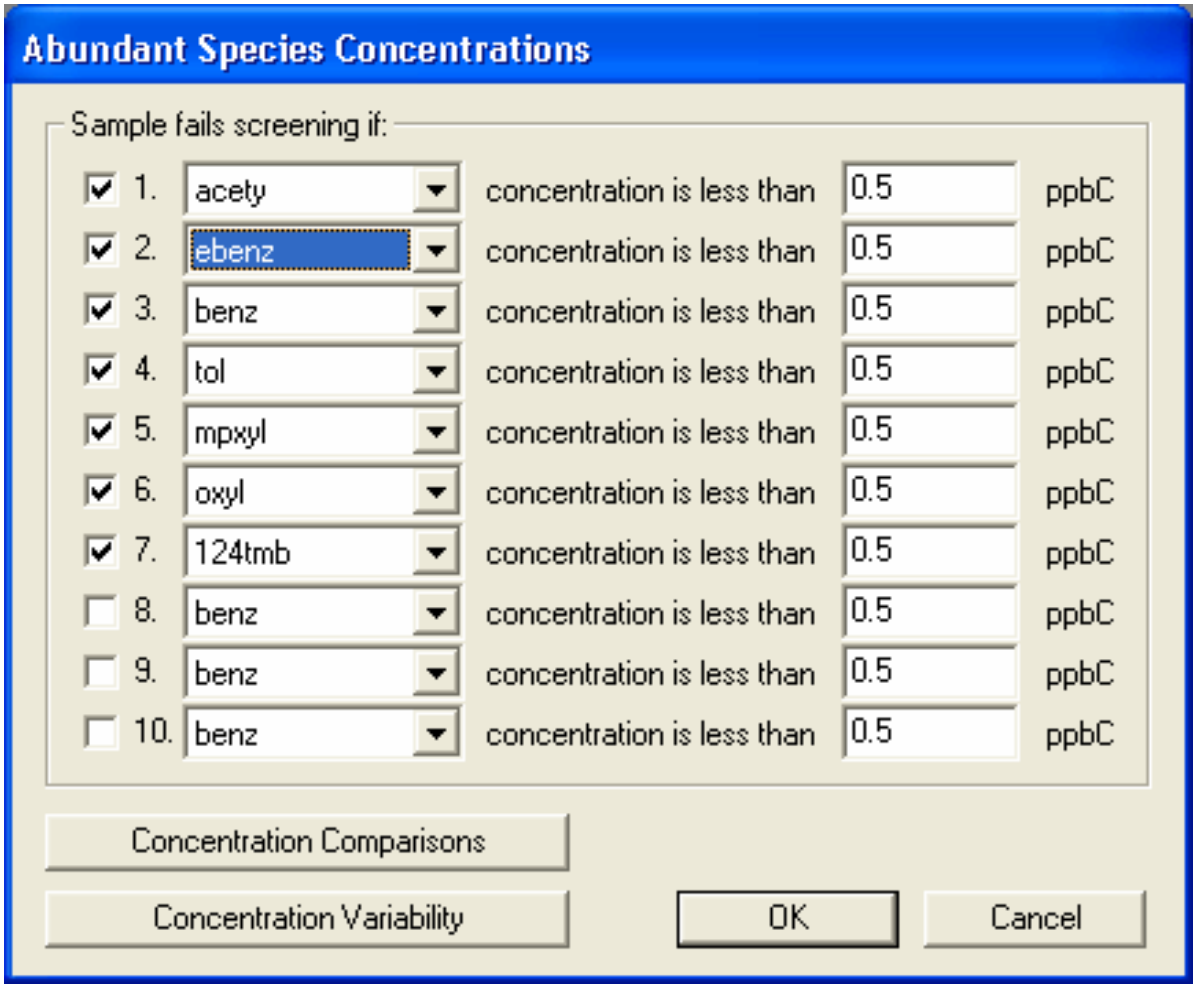
Temporal or Spatial Scale	*Expect Variability?*	*Behavior/explanation*	*Action*
Diurnal (time of day)	Urban-yes Rural-likely	Photochemical production from other VOCs is highest during the day	Check for midday-afternoon peak in concentrations
Day of week	Possible, but not likely	If VOC concentrations 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 $\mu g/m^3$

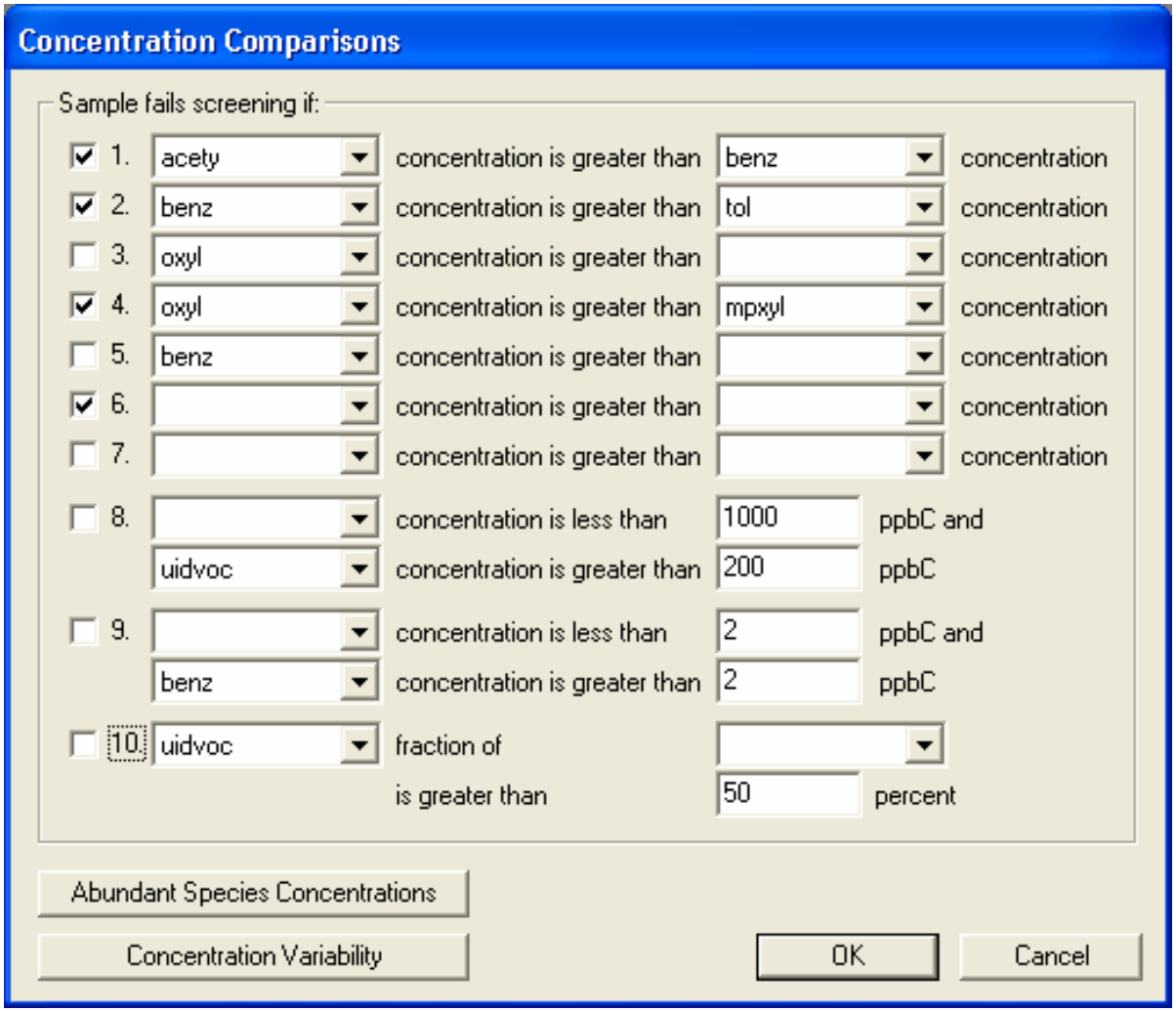
VOCDat Screening Checks (1 of 3)

Abundant Species Concentrations



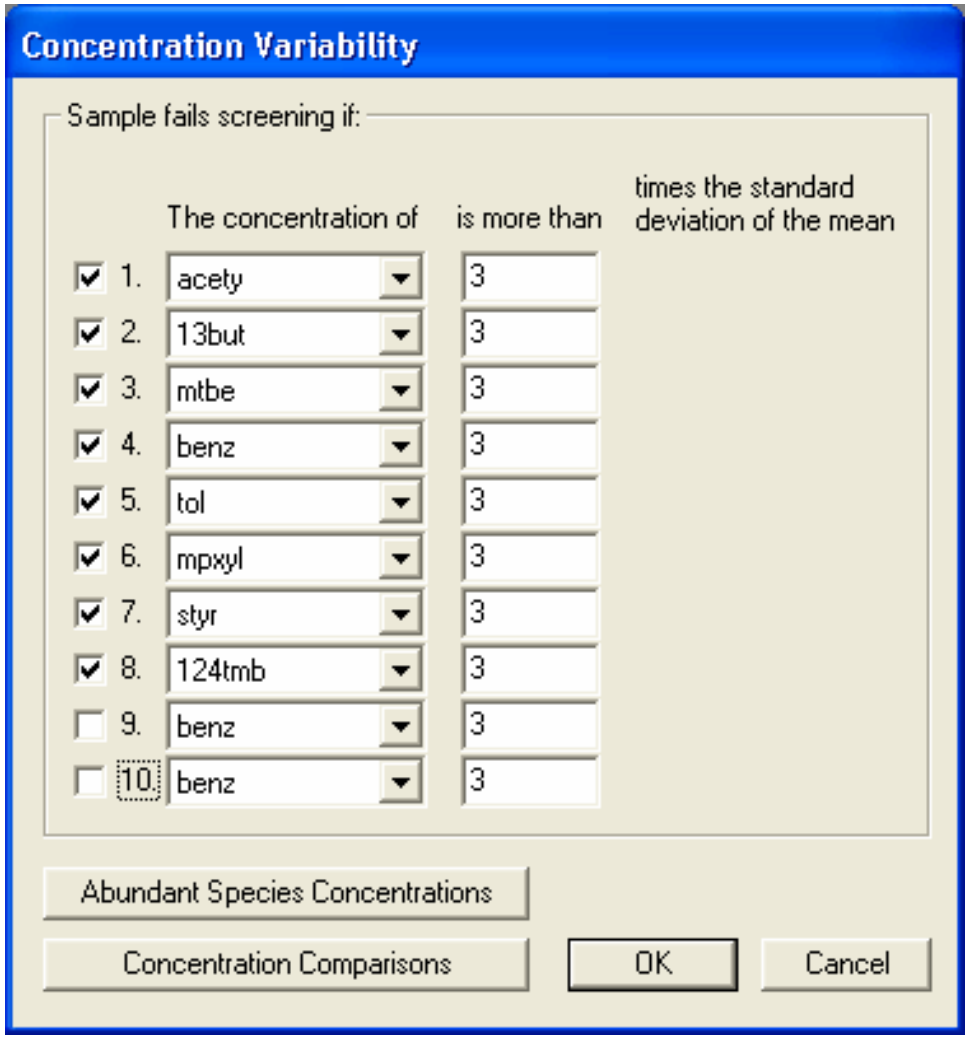
VOCDat Screening Checks (2 of 3)

Abundant Species Comparisons

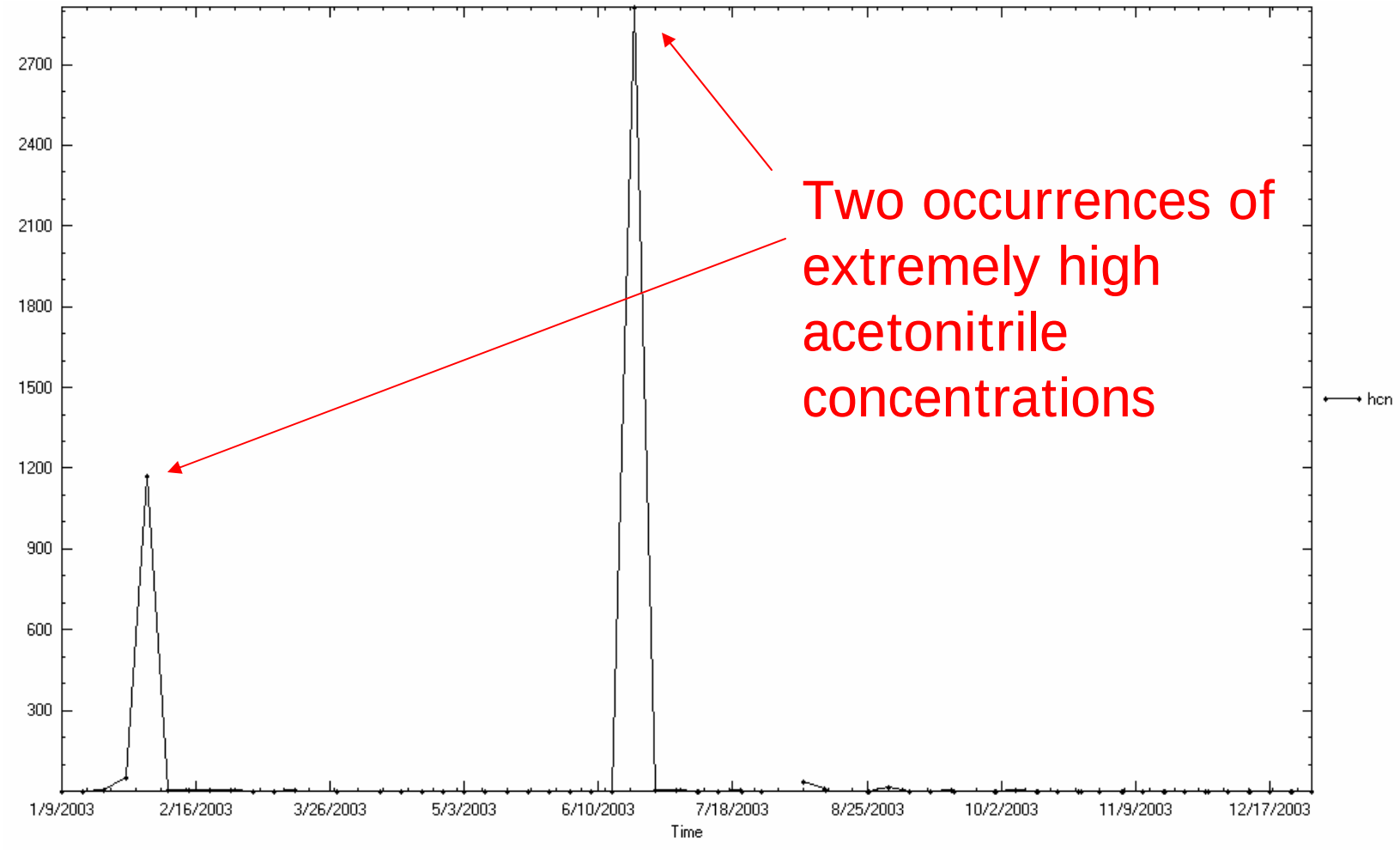


VOCDat Screening Checks (3 of 3)

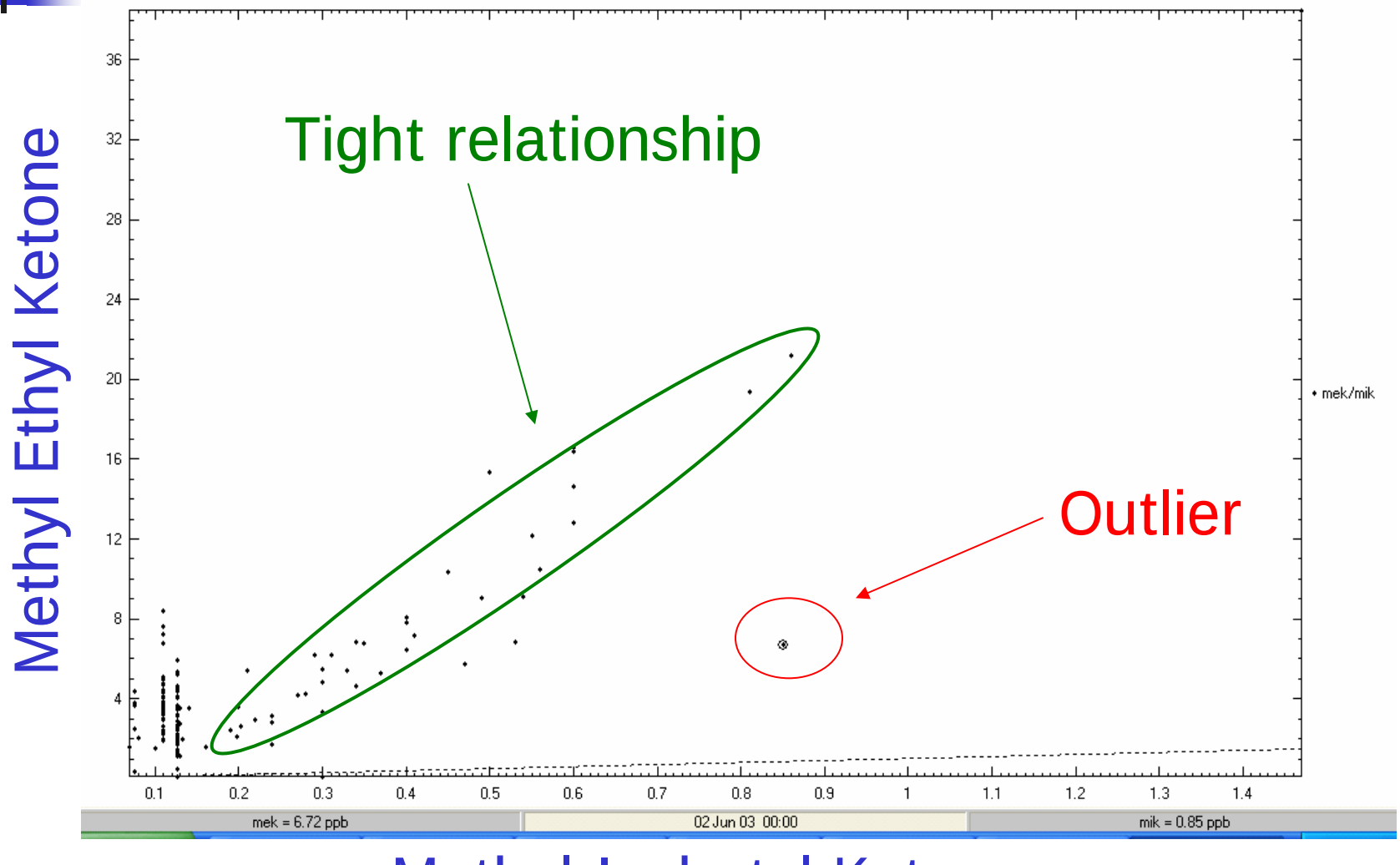
Abundant Species Concentration Variability



Example – Extreme Events

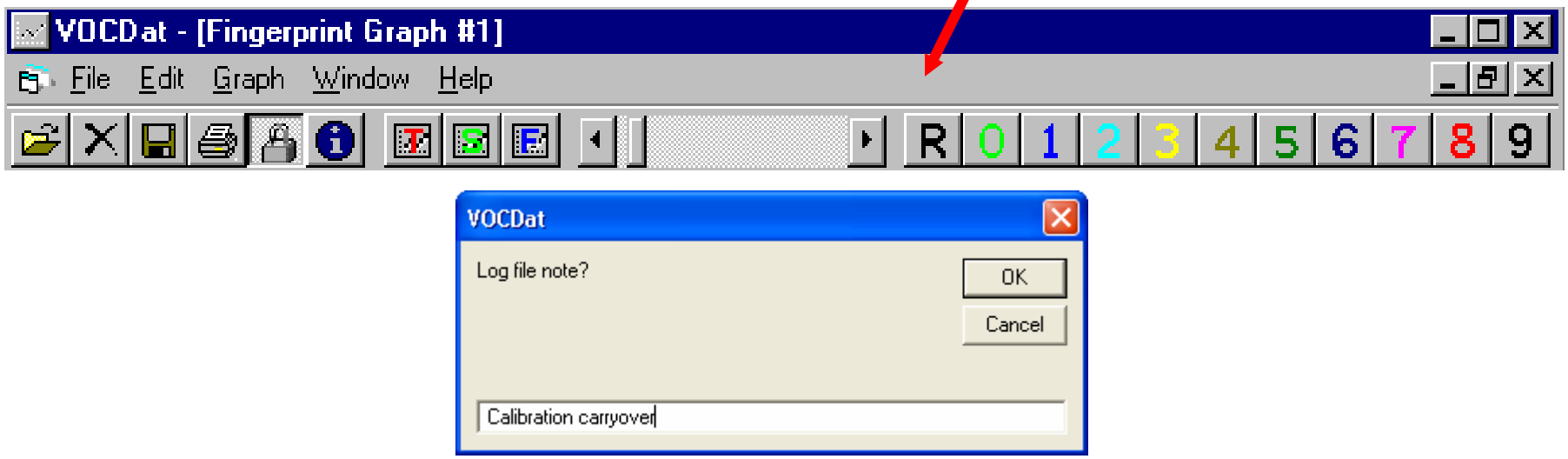


Example –Species Relationships



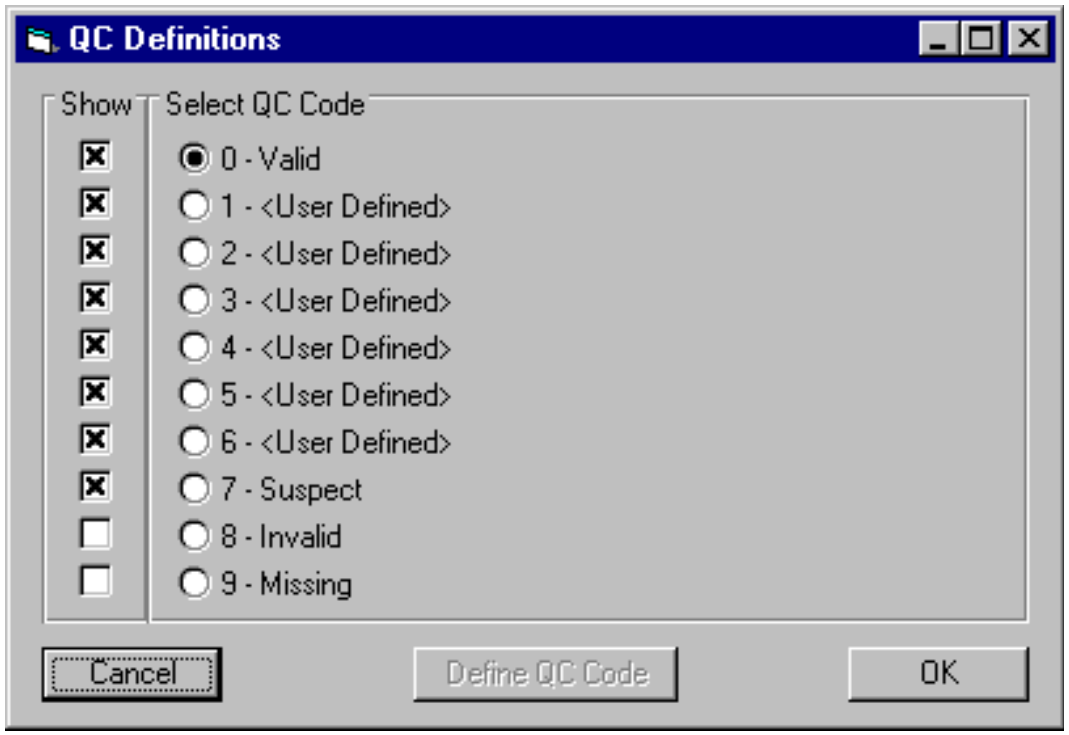
Flagging Suspect/Invalid Data (1 of 2)

Once erroneous data are identified, the VOCDat toolbar can be used to flag either the entire sample record or a single parameter



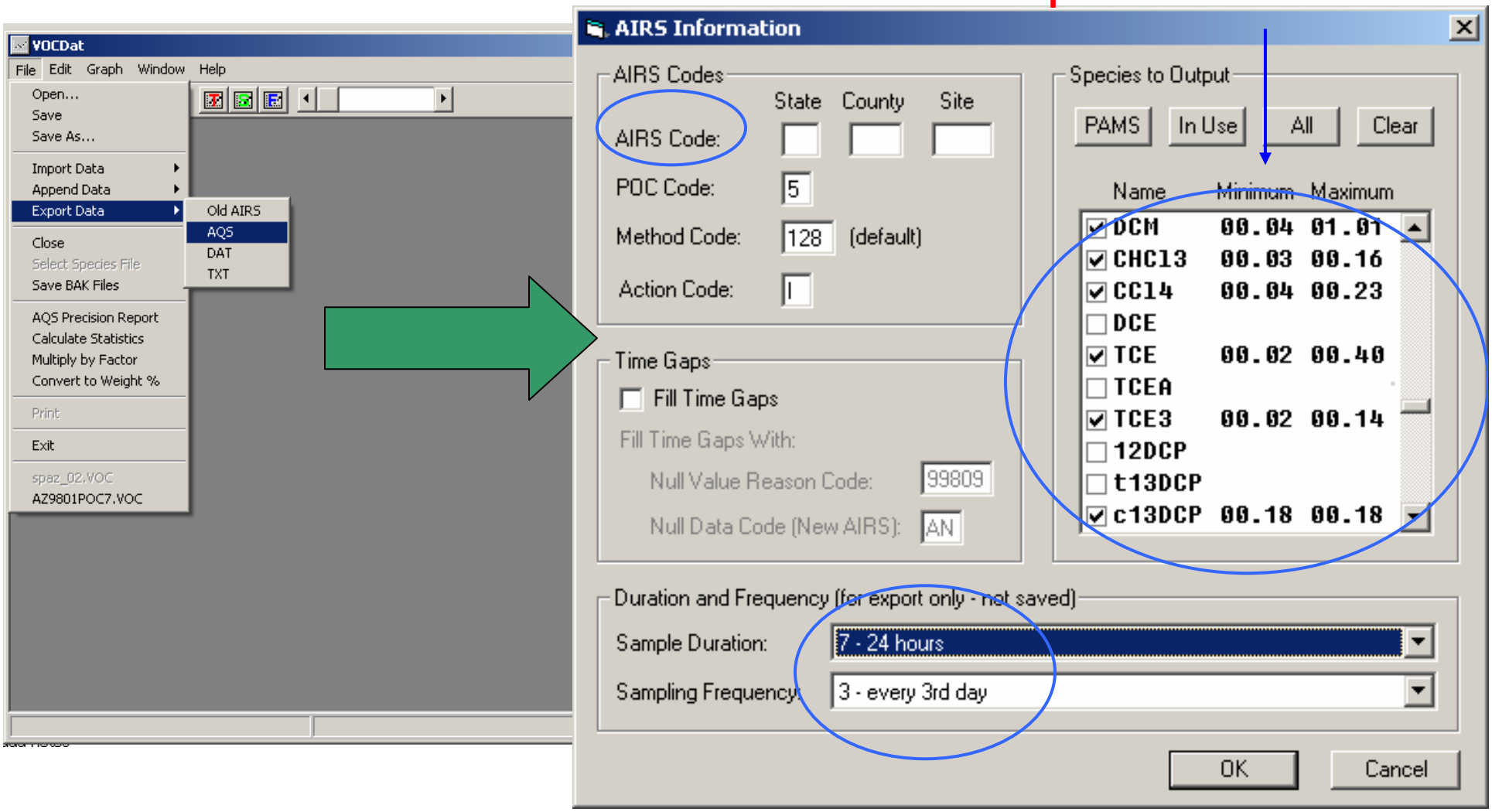
Flagging Suspect/Invalid Data (2 of 2)

Assign user-defined codes to indicate the problem

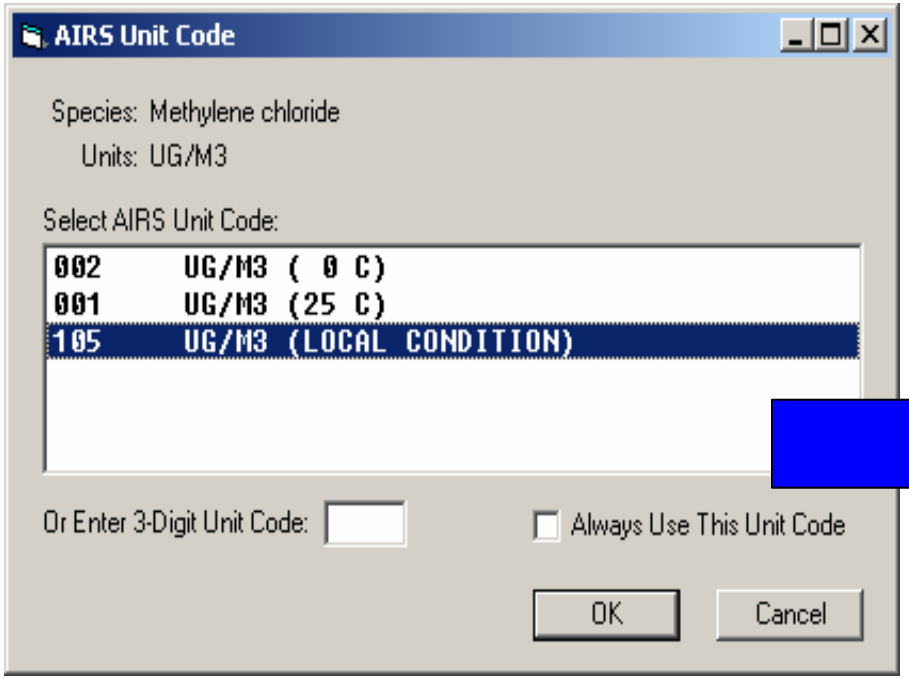


Exporting Into AQS Format I

Species Selected

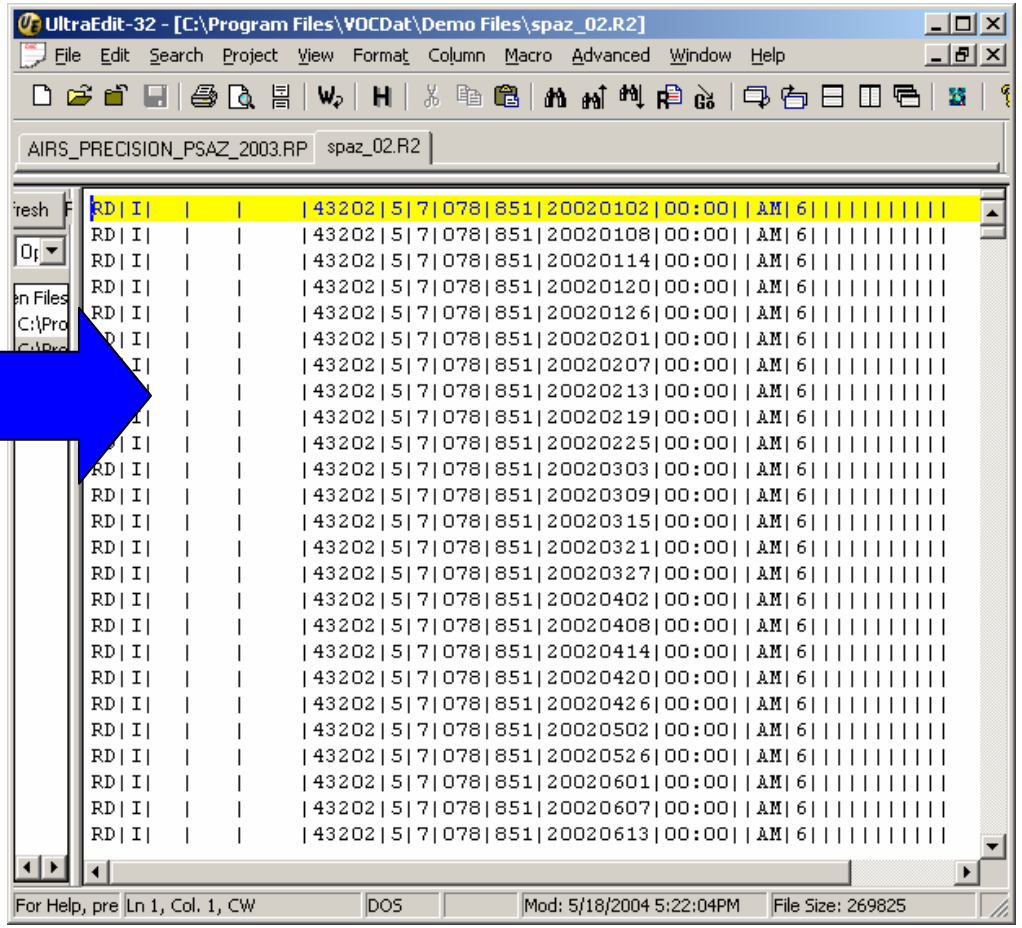


Exporting Into AQS Format II



Select units

Ready in AQS R2 format!

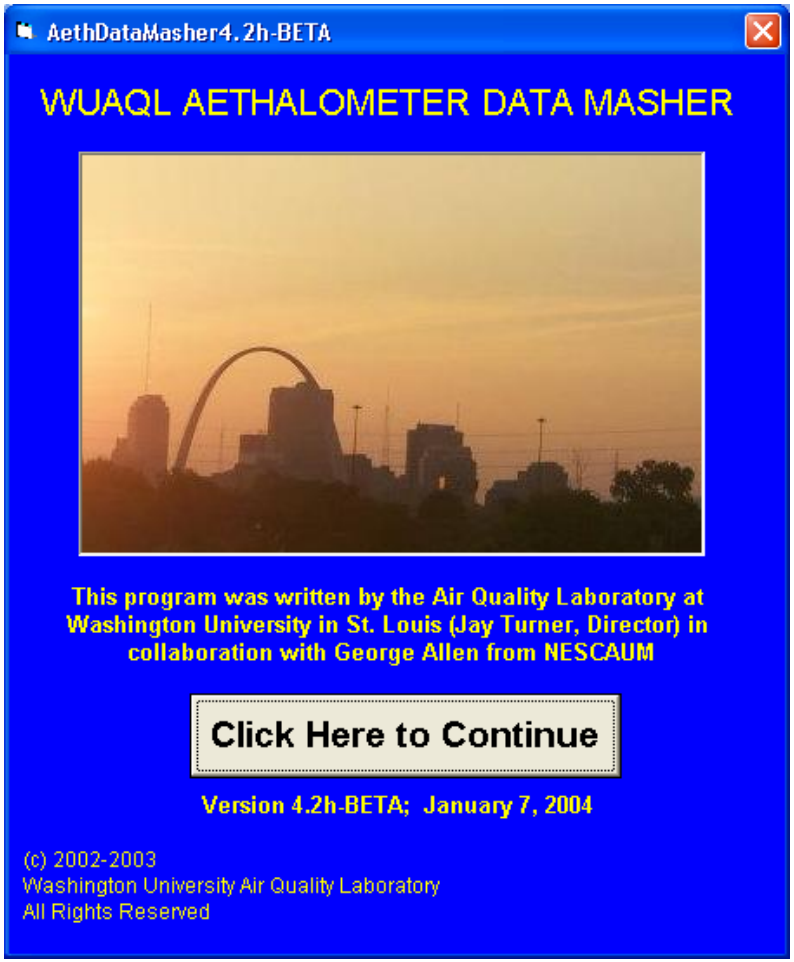


Aethalometer (Black Carbon)

- NATTS sites have Aethalometers™ to monitor black carbon (BC), a marker (thought not unique) for diesel particulate matter (DPM).
- What to do with the data?
  - Use “Data Masher” developed by Dr. Jay Turner (Washington University, St. Louis)
  - Bring the exported data from the Aethalometer™ directly into the Data Masher
  - Then into VOCDat – ready for validation and export into AQS format

Data Masher

- Five-minute or 1-hr averaged data can be input into the Data Masher
- Can be calculated by hourly averages
- Can be voided by hours with <75% completeness
- Available from Jay Turner: JRTurner@seas.wustl.edu (link also available from vocdat.sonomatech.com)



Example Import – Aethalometer Data (1 of 3)

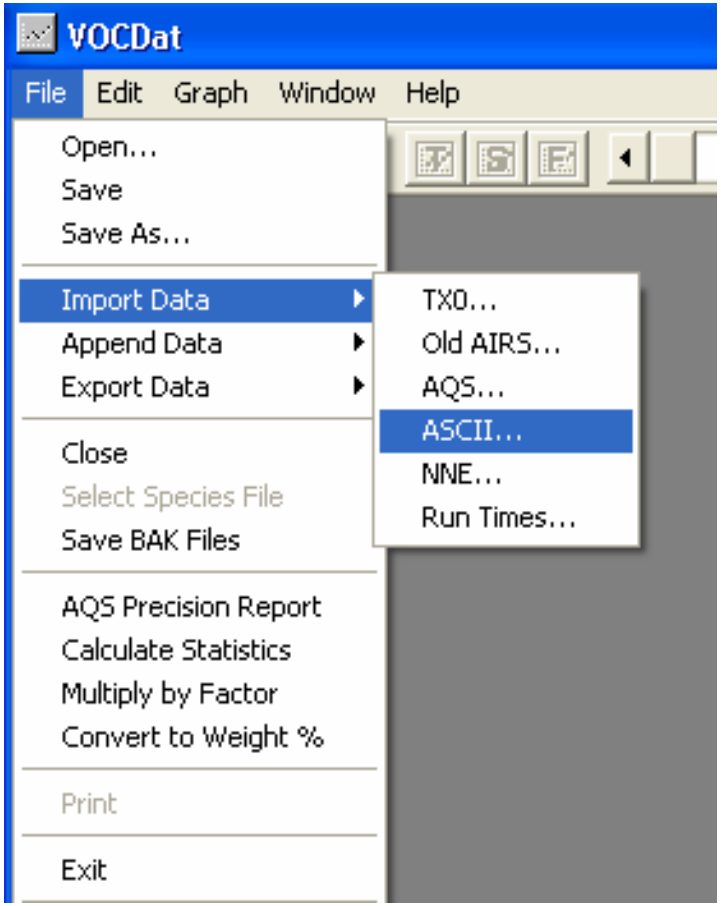
- Data are provided in a comma-separated text file with no header records.
- Save the file as tab-delimited *.txt (or *.CSV).

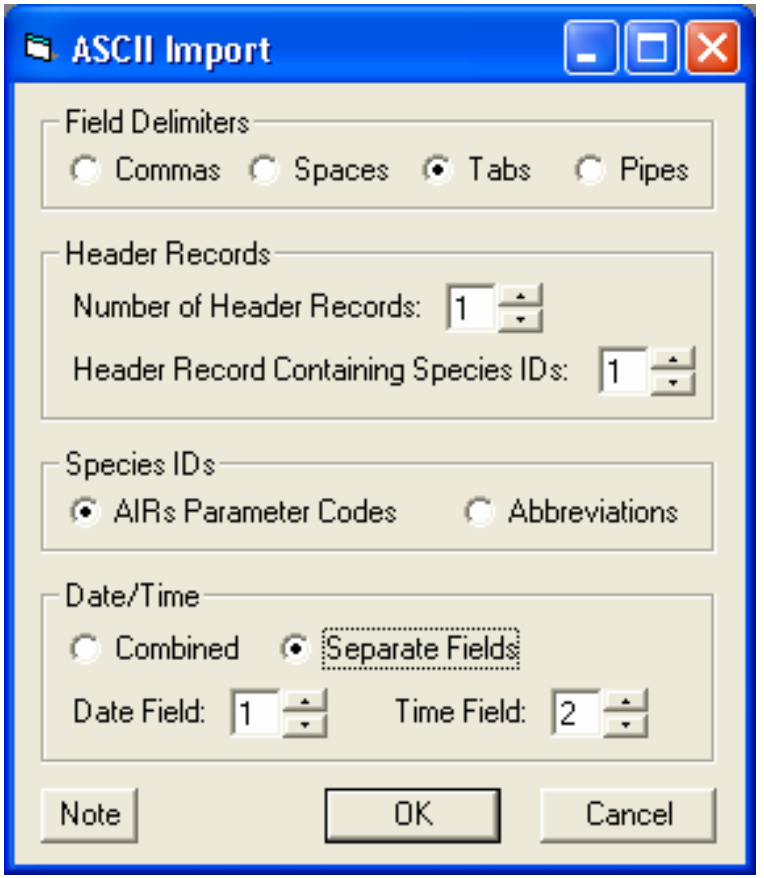
Date	time	84313
8/5/2003	21:00:00	0.251
8/5/2003	22:00:00	0.322
8/5/2003	23:00:00	0.19
8/6/2003	0:00:00	0.149
8/6/2003	1:00:00	0.289
8/6/2003	2:00:00	0.519
8/6/2003	3:00:00	0.316
8/6/2003	4:00:00	0.338
8/6/2003	5:00:00	0.511

In VOCDat Species file:

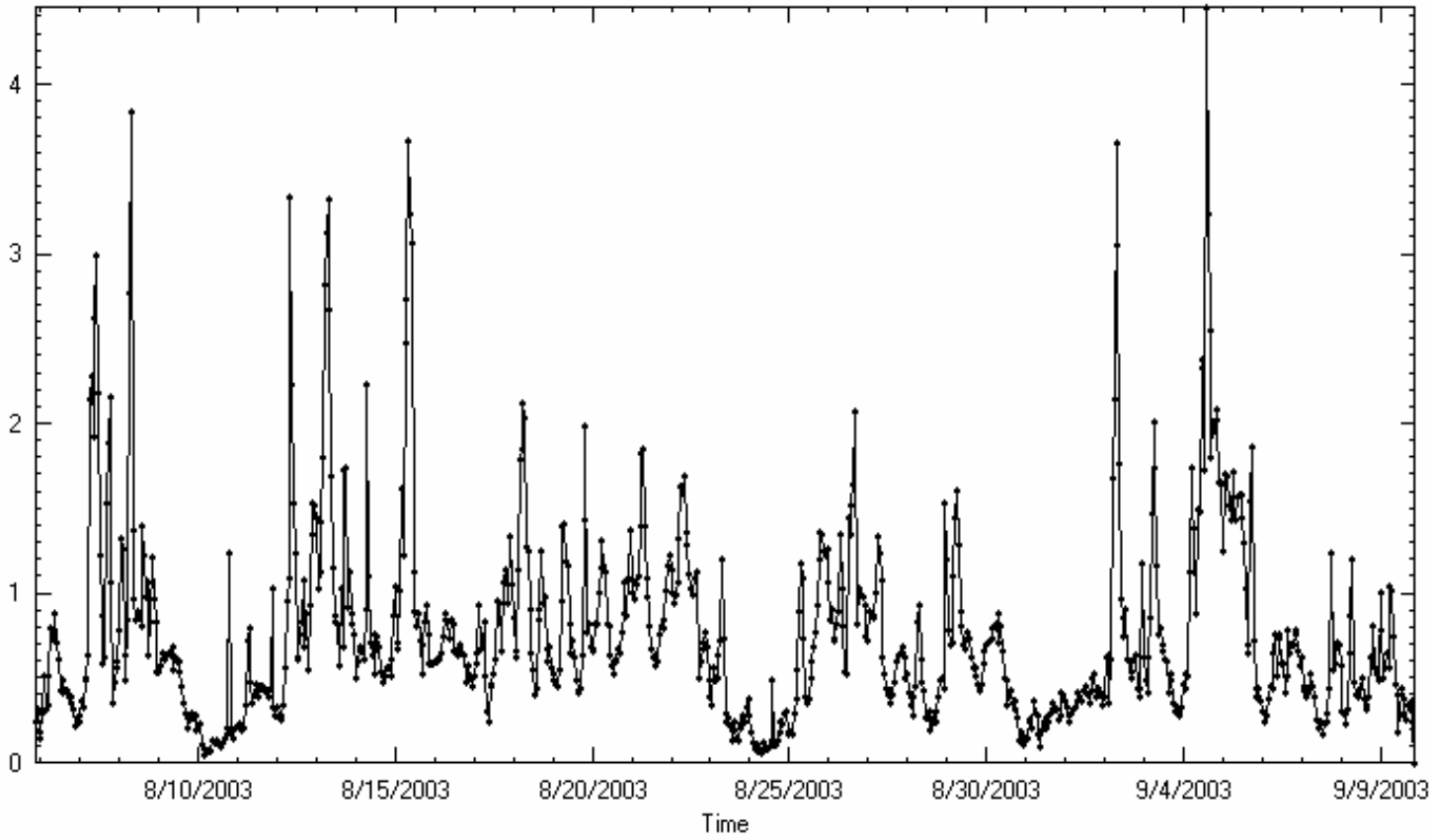
84313	Black carbon PM2.5 STP
88313	Black Carbon PM2.5 LC
84313	Black carbon PM2.5 STP
88313	Black Carbon PM2.5 LC

Example Import – Aethalometer Data (2 of 3)





Example Import – Aethalometer Data (3 of 3)



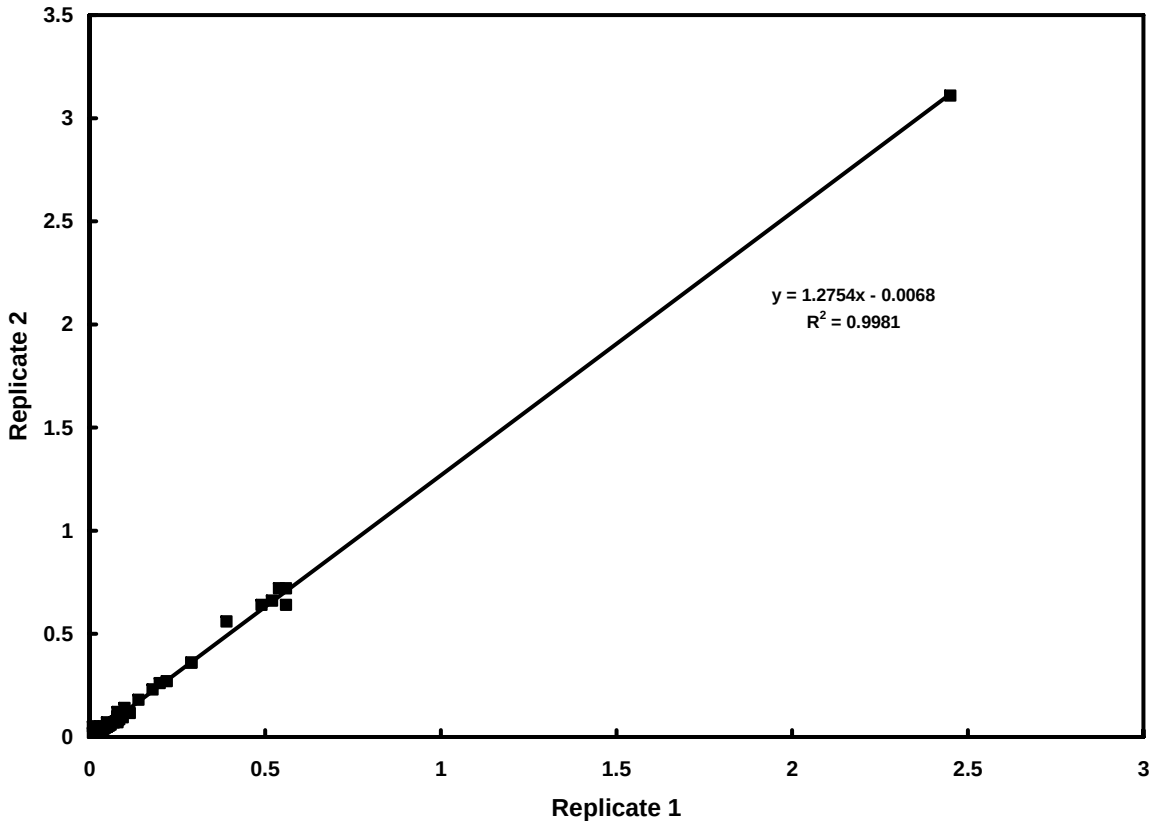
Ready for validation and export into AQS format!

Precision Report

- 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

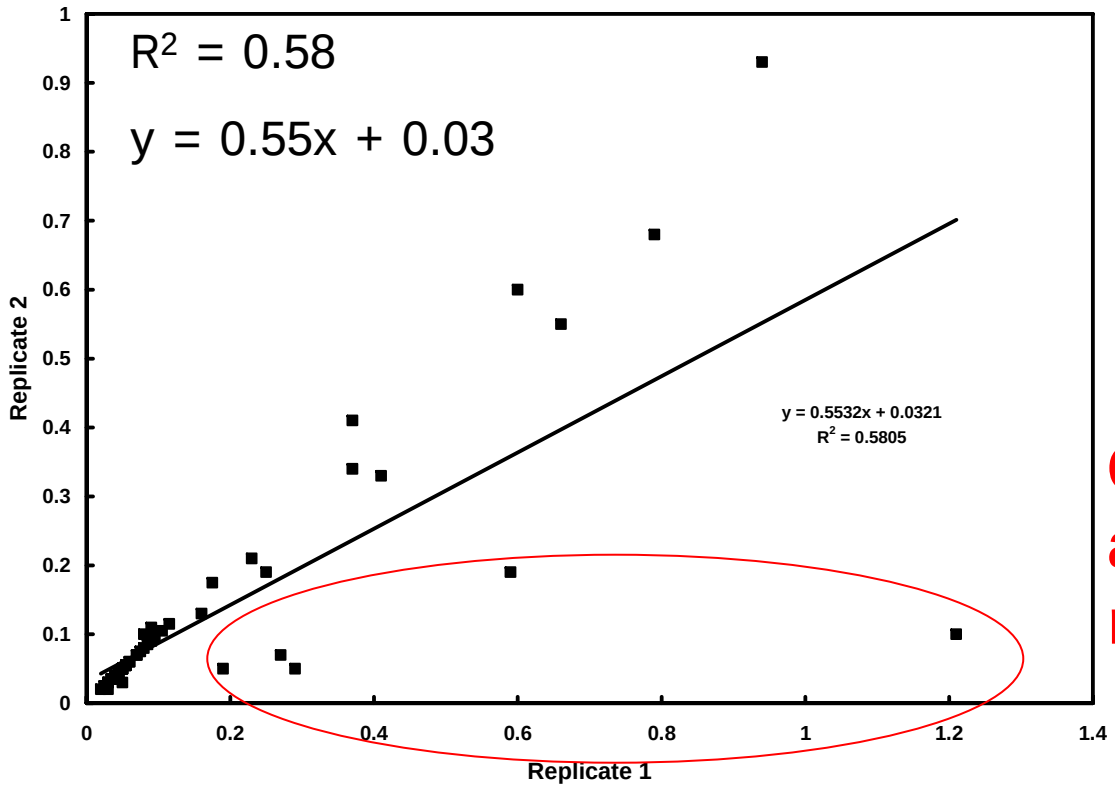
Usefulness of Precision Report

Compare replicate measurements (good example)



Usefulness of Precision Report

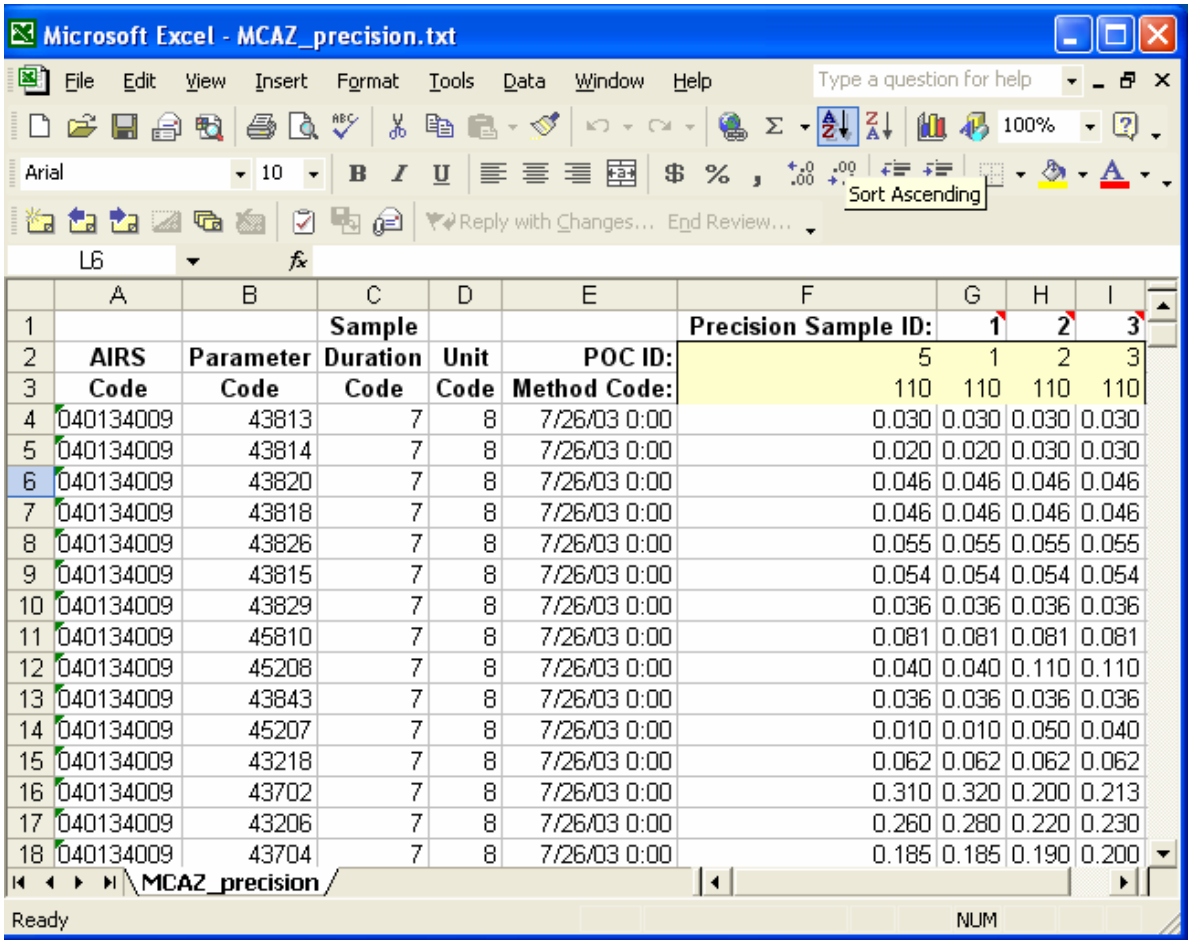
Compare replicate measurements (bad example)



Outliers flagged as suspect (poor reproducibility)

Precision Report Format

- Template is provided for data input
- Text-formatted cells were required to keep zeroes in front of AIRS Site Codes
- POC IDs indicate, from left to right, primary sample, duplicate, replicate, and duplicate-replicate samples

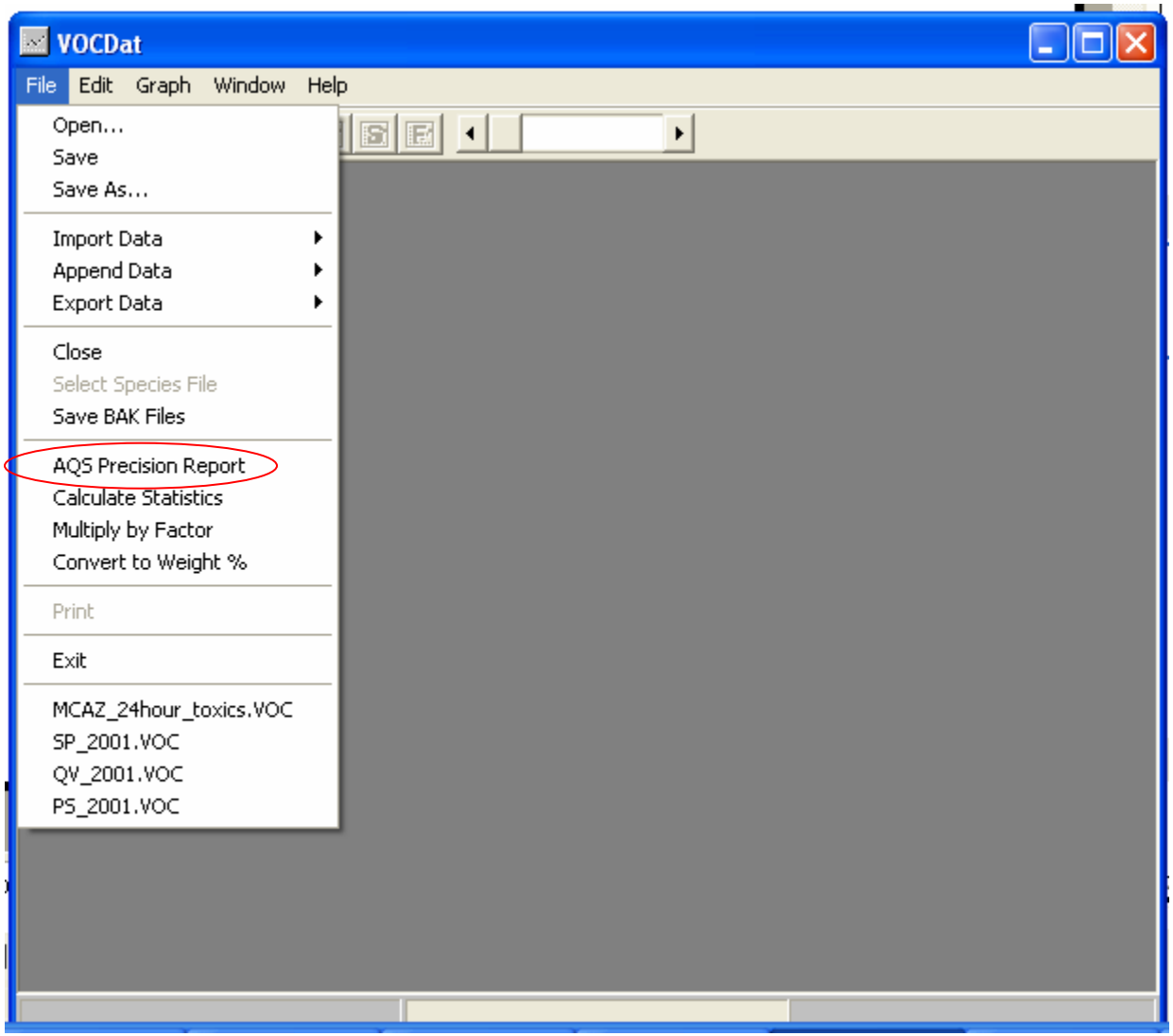


	A	B	C	D	E	F	G	H	I
1			Sample			Precision Sample ID:	1	2	3
2	AIRS	Parameter	Duration	Unit	POC ID:	5	1	2	3
3	Code	Code	Code	Code	Method Code:	110	110	110	110
4	040134009	43813	7	8	7/26/03 0:00	0.030	0.030	0.030	0.030
5	040134009	43814	7	8	7/26/03 0:00	0.020	0.020	0.030	0.030
6	040134009	43820	7	8	7/26/03 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

Save file as *.txt

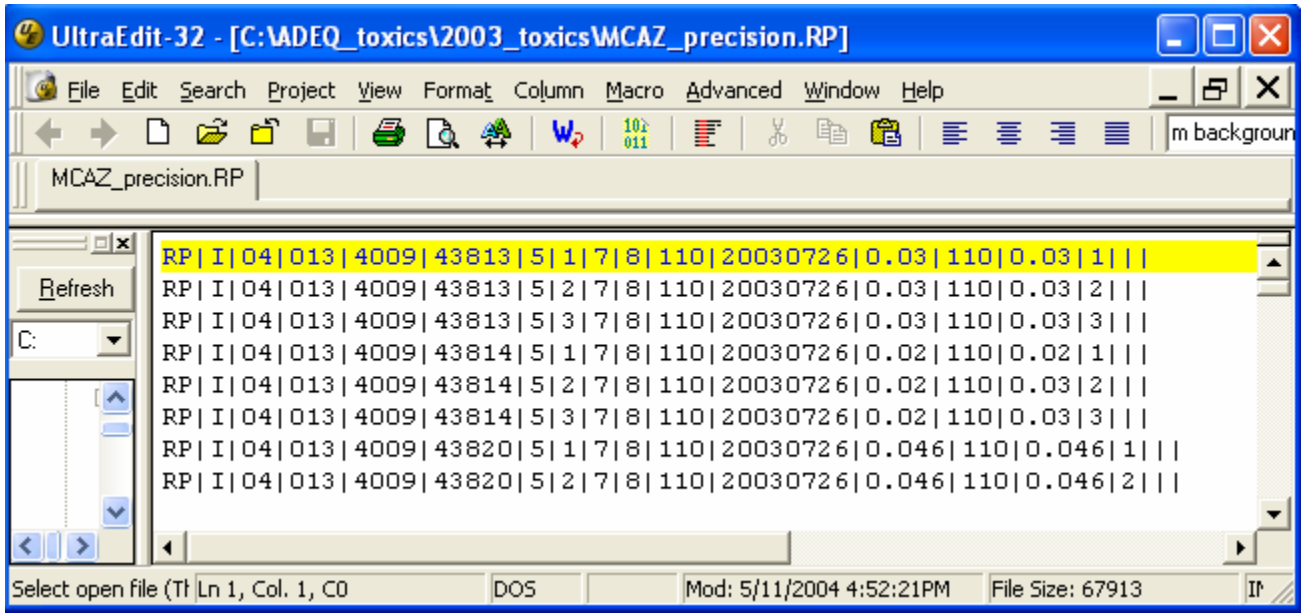
Creating Precision Report

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

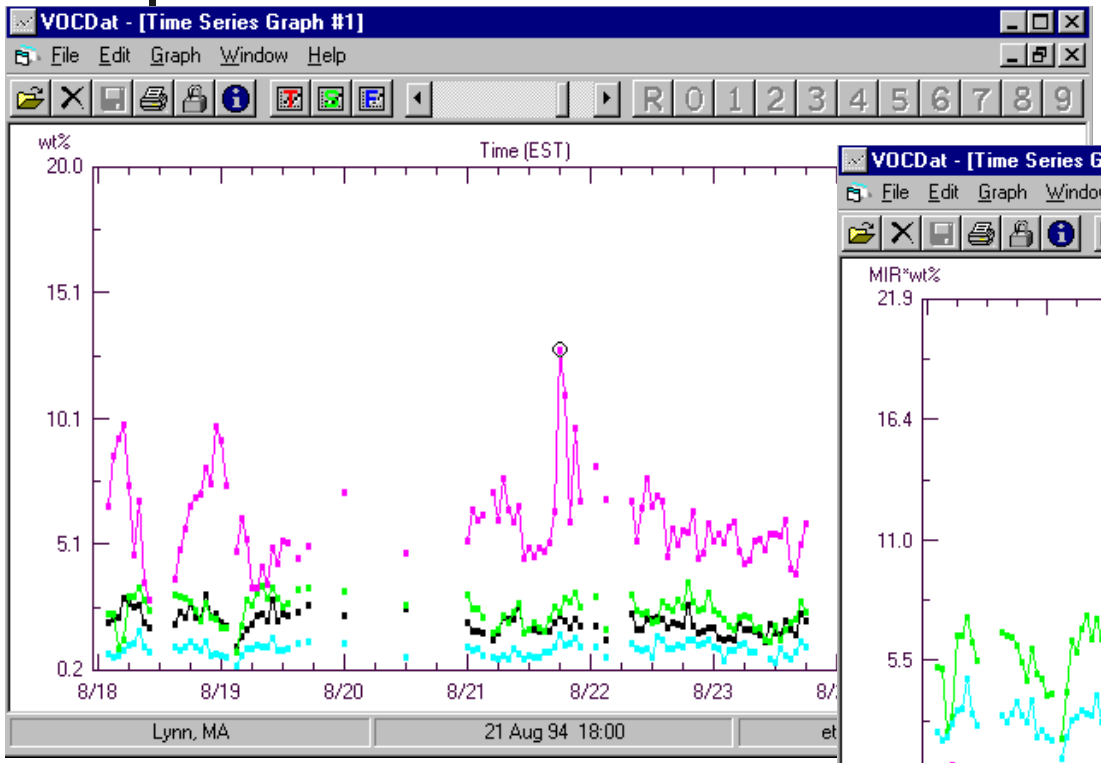


Resultant Precision Report

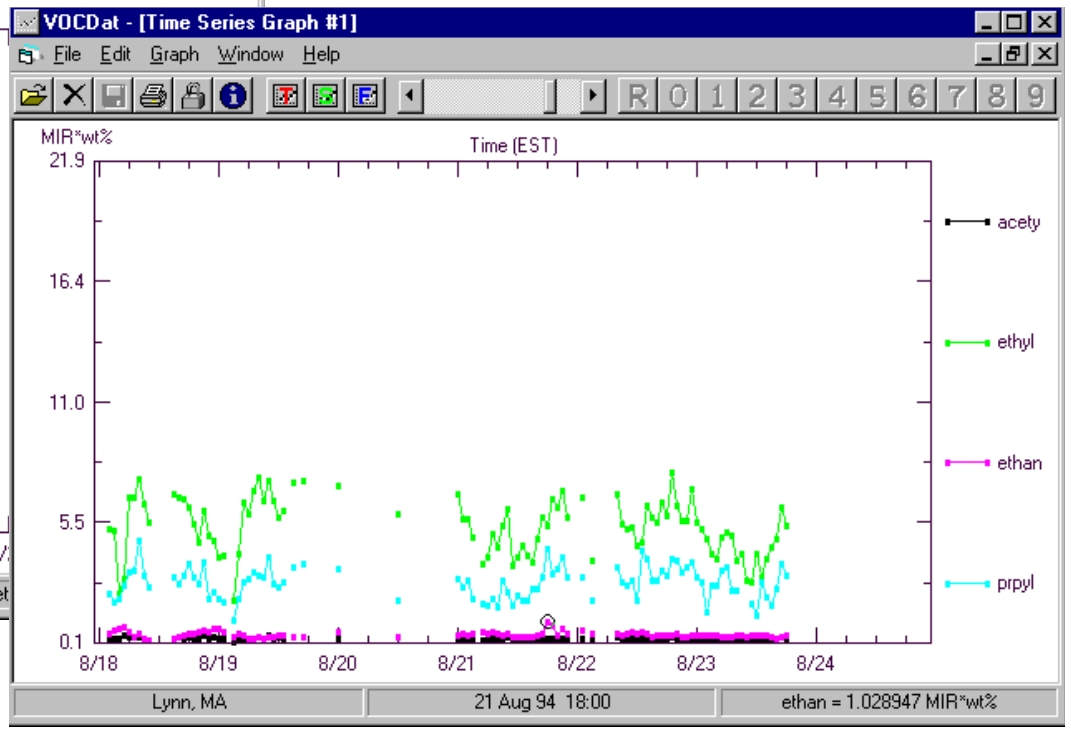
After the file is imported into VOCDat, the data are automatically exported as an *.RP file which is saved in the same folder as the imported text file



Using Other Forms of the Data



Weight percent



Reactivity-weighted or risk-weighted

Toxicity/Risk Factors

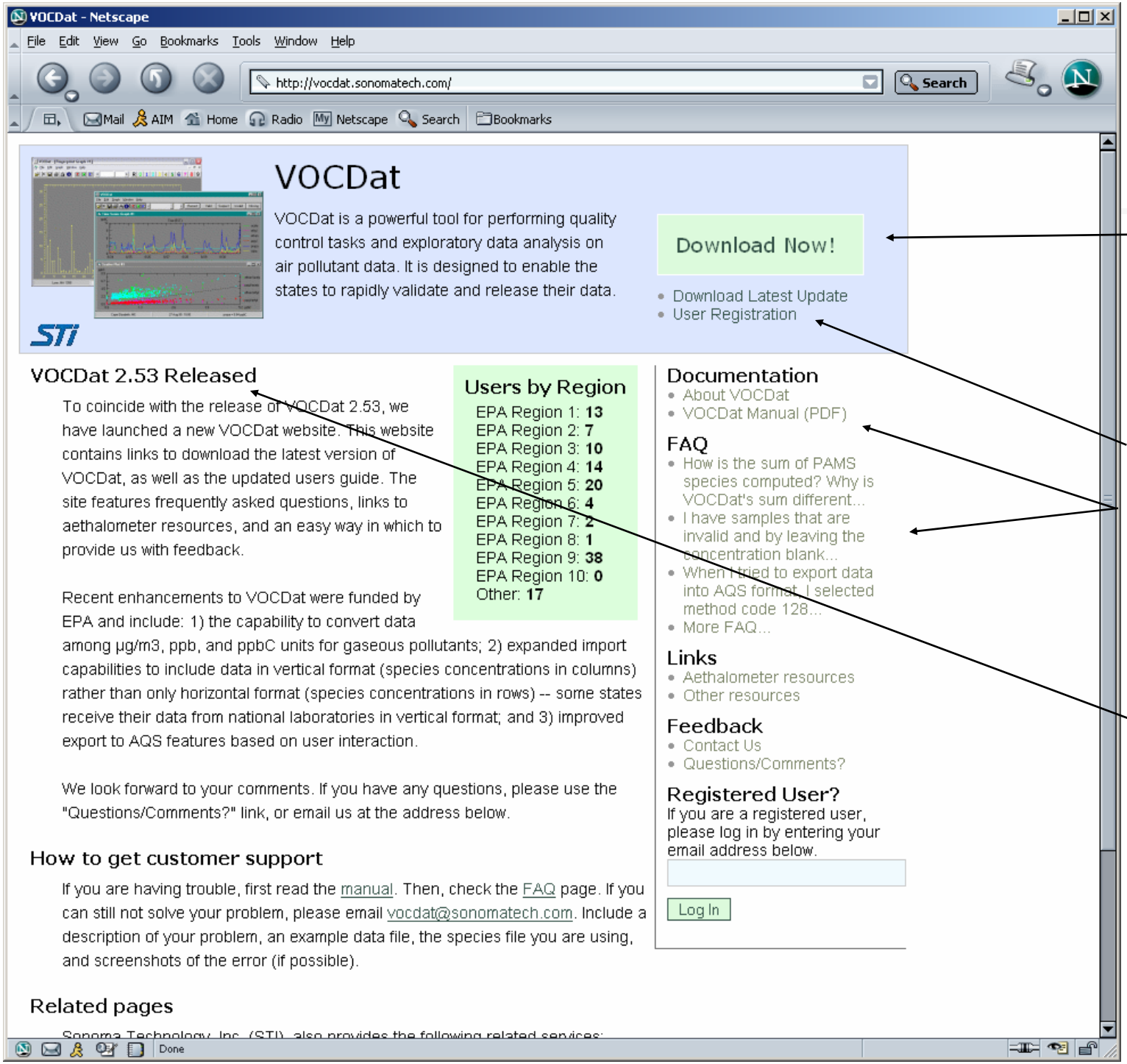
- EPA's Integrated Risk Information System – IRIS (see http://www.epa.gov/iris/index.html).
- EPA Health Effects Notebook for Hazardous Air Pollutants (see http://www.epa.gov/ttn/atw/hapindex.html).
- EPA National Center for Environmental Assessment (NCEA) (see http://cfpub.epa.gov/ncea/cfm/nceahome.cfm).
- 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).

Unit Conversion

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

Obtaining VOCDat

- Available free at
  http://vocdat.sonomatech.com/
- To register and get technical support, e-mail vocdat@sonomatech.com



Download VOCDat and species files, etc.

Please register!

Manual and FAQ

Check if you have most current version

Other Additions and Ideas?

- Other import file formats?
- Additional QC checks? More refined screening criteria?
- Can we input comments into AQS?
- Improve precision information transfer?
- Improve MDL information transfer?
- Is there a standard QC code list for comments?

Summary

- Graphical techniques are useful to quickly and efficiently validate (and begin analysis of) aerometric data
- VOCDat is useful in processing and validating air toxics data

Glossary

AIRS – Aerometric Information Retrieval System
AQS – Air Quality System
auto-GC – automatic-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