



1360 Redwood Way, Suite C
Petaluma, CA 94954-1169
707/665-9900
FAX 707/665-9800
www.sonomatech.com

VOC DAT USER'S GUIDE

STI-902620-2015-UG

By:

Hilary R. Hafner

Jeffrey D. Prouty

Sonoma Technology, Inc.

1360 Redwood Way, Suite C

Petaluma, CA 94954-1169

Prepared for:

Sharon Nizich

**U.S. Environmental Protection Agency,
Office of Air Quality Planning and Standards
Monitoring and Quality Assurance Group**

MD C339-02

Research Triangle Park, NC 27711

September 30, 2004

ACKNOWLEDGMENTS

In support of the NARSTO-Northeast study with partial funding from EPRI Contract WO 9108-01, Sonoma Technology, Inc. (STI) developed a Windows-based, menu-driven program called VOCDat to display volatile organic compound (VOC) data and allow users to perform quality control tasks and begin data analysis. One of the goals behind the development of this software was to enable the Northeastern states to rapidly validate and release their 1995 and 1996 Photochemical Assessment Monitoring Stations (PAMS) data. Dr. Peter Mueller of EPRI, the NARSTO-Northeast Project Coordinator, was instrumental in the development and support of VOCDat for use in the Northeast. EPA's Office of Air Quality Planning and Standards (OAQPS) has funded modifications to the program and support since the late 1990s in support of the PAMS, PM_{2.5} and air toxics measurement programs. Hilary Hafner, Paul Roberts, and Jeff Prouty of STI developed VOCDat.

Many people have been involved in the development and modification of VOCDat:

- Mr. Alan Van Arsdale (EPA Region I) and Rich Poirot (Vermont) provided encouragement and support during the original development process.
- Al Leston and Pete Brunelli (Connecticut) performed extensive testing of the pre-release versions of the software. Their feedback greatly helped to improve the program.
- Similarly, Mark Schmidt, Joann Rice, and Sharon Nizich (EPA OAQPS) provided comments and feedback that have guided the program modifications.
- Lengthy discussions with Avraham Teitz, EPA Region II, were critical to understanding some of the problems faced by PAMS automatic gas chromatograph (auto-GC) operators.
- Discussions with Paul Nichols (Maine) and Tom McGrath (Massachusetts) led to several changes in the AIRS export capabilities.
- Kate Foster-Springer (Texas) provided screening ideas based on those used in Texas to review auto-GC data.
- Dirk Felton (New York), Tom Koehler (Illinois), Tracie Tenetylo (Pennsylvania), Charles Anderson (California), Atul Bhatt (Indiana), and many others served as excellent testers of the program during its development and refinement; their contributions to the program and its usefulness were invaluable.
- STI staff, Steve Brown, Mike McCarthy, and Theresa O'Brien have served as VOCDat users and trainers and have suggested and tested several modifications of the program.

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
LIST OF FIGURES	ix
LIST OF TABLES	xiii
EXECUTIVE SUMMARY	ES-1
1. DESCRIPTION OF THE SOFTWARE	1-1
1.1 Background.....	1-1
1.2 Overview of VOCDat Features	1-2
1.3 New Features in the Fall 2004 Release.....	1-2
1.4 Who Can Use This Program?	1-3
1.5 Customer Support	1-3
2. INSTALLING THE SOFTWARE.....	2-1
2.1 Obtaining the Software	2-1
2.2 Installing VOCDat	2-1
2.2.1 Initial Installation (Full)	2-1
2.2.2 Update Installation (Executable Only)	2-3
2.3 Common Problems Encountered During Installation.....	2-3
2.4 Getting Started with the Software.....	2-3
2.5 When VOCDat Times Out.....	2-4
3. VOCDAT FILE IMPORT, EXPORT, AND MANIPULATION.....	3-1
3.1 How VOCDat Works.....	3-1
3.1.1 The Species File	3-1
3.1.2 Species2.txt.....	3-2
3.1.3 Units.txt	3-3
3.1.4 Methods.txt.....	3-5
3.1.5 Nullcode.txt	3-5
3.1.6 Conversions.txt.....	3-5
3.2 VOCDat Files	3-5
3.3 Importing Files.....	3-10
3.3.1 Perkin-Elmer TX0 or TX1	3-11
3.3.2 Old AIRS Files	3-11
3.3.3 AQS R-2 Files	3-12
3.3.4 ASCII Files.....	3-13
3.3.5 Legacy Files	3-14
3.4 Common Import Problems	3-15
3.4.1 Missing Species.....	3-15
3.4.2 Other Errors.....	3-15
3.5 Exporting Files.....	3-16
3.5.1 Old AIRS AMP370 Format.....	3-16
3.5.2 AQS R-2 Format	3-18

TABLE OF CONTENTS (Continued)

<u>Section</u>	<u>Page</u>
3.5.3 Text Without QC Codes	3-20
3.5.4 Graph-Specific Data	3-21
3.5.5 Additional Export Notes.....	3-21
3.6 Appending Files.....	3-21
3.7 Customizing the Parameter List.....	3-22
3.8 Parameter Units	3-25
3.9 Changing the File Header Information	3-26
3.10 Closing Files	3-26
3.11 Handling Different Sampling Times	3-27
3.12 Species Abbreviations	3-28
3.13 Editing VOCDat Files in Excel	3-28
4. DISPLAYING AND FLAGGING DATA.....	4-1
4.1 Displaying Data in VOCDat.....	4-1
4.1.1 Time Series Graphs	4-1
4.1.2 Scatter Plots.....	4-3
4.1.3 Fingerprint Plots	4-4
4.1.4 Moving Around in the Plots	4-6
4.1.5 Saving Graph Configurations	4-7
4.1.6 Printing	4-7
4.2 Quality Checking Data	4-7
4.2.1 Performing Screening Checks.....	4-8
4.2.2 Calculating Summary Statistics.....	4-17
4.2.3 Enabling Editing.....	4-19
4.2.4 Changing QC Codes.....	4-20
4.2.5 Selecting Multiple Data Points.....	4-21
4.2.6 Changing the QC Code for a Data Point	4-21
4.2.7 Changing the QC Code for a Data Point	4-22
5. CHANGING FROM CONCENTRATION TO OTHER FORMATS.....	5-1
5.1 Weight Percent (PAMS only).....	5-1
5.2 Scaling Data Using Factors	5-2
5.2.1 Overview	5-2
5.2.2 Reactivity-weighted.....	5-2
5.2.3 Cancer Risk-weighted	5-4
5.2.4 User-defined Factors	5-5
5.3 Changing Units of Displayed Data.....	5-5
6. GENERAL DATA VALIDATION APPROACH AND EXAMPLES	6-1
6.1 Data Validation Objectives.....	6-1
6.2 Approach.....	6-1
6.3 Examples of Typical Problems Identified in PAMS VOC Data	6-3
6.3.1 Species Misidentification	6-3

TABLE OF CONTENTS (Concluded)

<u>Section</u>	<u>Page</u>
6.3.2 Calibration Data	6-5
6.3.3 Contamination	6-7
6.3.4 Miscellaneous Data Problems	6-9
7. ADDITIONAL FEATURES	7-1
7.1 Processing and Screening Aethalometer Data	7-1
7.1.1 Data Masher Features	7-1
7.1.2 Validating the Data	7-2
7.1.3 Obtaining the Aethalometer Data Masher	7-4
7.2 Precision Report Preparation	7-5
8. REFERENCES	8-1
APPENDIX A: IMPORT FILE FORMATS	A-1
APPENDIX B: EXPORT FILE FORMATS FOR DATA EXPLORATION	B-1
APPENDIX C: ANNOTATED CONFIGURATION FILE	C-1

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1-1. Opening screen for VOCDat.....	1-1
2-1. Screen that may be encountered during full installation of VOCDat software.	2-2
2-2. Primary screen for VOCDat (version 2.x).....	2-3
2-3. VOCDat “time out” error message indicating the need for the user to obtain a newer version of the program.	2-4
3-1. Example VOCDat file excerpt showing the header information and start of data.	3-9
3-2. Screen shot showing the options for “File - Import Data”.....	3-11
3-3. ASCII Import dialog box.	3-14
3-4. Incorrect header record error message.....	3-15
3-5. Incorrect species ID line error message.....	3-16
3-6. Incorrect date/time field designation message.....	3-16
3-7. Screen shot showing the options for “File - Export Data”.....	3-17
3-8. Example dialog box for export of a VOCDat data file to old AIRS format.	3-18
3-9. Example dialog box for export of a VOCDat data file to AQS format.	3-19
3-10. Dialog box for customizing the parameters list.	3-23
3-11. Dialog box for creating or editing a species definition showing the drop down menu for method code.	3-23
3-12. Dialog box for creating or editing a field definition.	3-24
3-13. Select Species File dialog box.	3-25
3-14. File header information screen.....	3-27
3-15. Species list window.....	3-29
4-1. “T” button on the VOCDat toolbar.....	4-1
4-2. Window for selecting species and sites to graph.	4-2
4-3. Value Axis scale pop-up window.	4-2

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
4-4. Time scale pop up window.	4-2
4-5. Example time series plot with highlighted data point information shown in the status bar.	4-3
4-6. Scatter plot button on the toolbar.....	4-4
4-7. Scatter plot x-scale and y-scale selection pop up windows.	4-4
4-8. Example scatter plot of propane versus ethane.....	4-5
4-9. Fingerprint button on the toolbar.	4-5
4-10. Example fingerprint plot with highlighted data point information shown in the status bar.	4-6
4-11. Scroll bar.....	4-7
4-12. Screening criteria available in VOCDat.	4-8
4-13. Example window for selecting the screening criteria for abundant species concentrations.	4-9
4-14. Example window for selecting the screening criteria for species concentration comparisons.	4-10
4-15. Example window for selecting the screening criteria for species concentration variability.	4-12
4-16. Window for applying the screening criteria.....	4-13
4-17. Excerpt from output of an example abundant species concentration screening check.	4-14
4-18. Example output from the screening test for concentration comparisons.....	4-15
4-19. Example output from the concentration comparison screening check.	4-16
4-20. Window providing selections for data to include in statistical calculations.	4-17
4-21. Example vertical output from statistics calculation.	4-18
4-22. Example horizontal output from statistics calculation.....	4-19
4-23. Enable editing window.	4-19

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
4-24. Initial configuration of the QC definitions.	4-20
4-25. Example user definitions for QC codes 1-6.....	4-21
4-26. Example of a highlighted species in the status bar for a scatter plot.	4-22
4-27. Window for adding a log file note with an example comment.....	4-23
4-28. Example LOG file showing changes to QC codes and example comments.....	4-23
5-1. Example time series plot of weight percent data.	5-1
5-2. Example time series plot of concentration data multiplied by Carter MIR factors.	5-3
5-3. Menu selection for converting units.	5-6
5-4. Example screen shots showing the same data in ppbC, ppb, and $\mu\text{g}/\text{m}^3$	5-7
6-1. Example of misidentification of toluene, a typically abundant species, in hourly data collected at a PAMS site during June 1995 using time series and scatter plots.....	6-4
6-2. Example of potentially misidentified toluene concentrations.....	6-5
6-3. Example of possible misidentification of 2-methylpentane and 3-methylpentane concentrations in a PAMS data set.	6-6
6-4. Example of typical and calibration fingerprints observed at a PAMS site.	6-7
6-5. Example of calibration gas carryover and its effect on n-decane and n-undecane concentrations for several hours after the calibration.	6-8
6-6. Concentrations of i-pentane and n-butane measured at a PAMS site during June 1995 showing high concentrations of i-pentane relative to n-butane.	6-9
6-7. Example of possible contamination of the column in the GC at the beginning of the month involving several species at a PAMS site during June 1996.	6-10
6-8. Example of high detection limit observed for 1,2,4-trimethylbenzene at a PAMS site during 1995; toluene concentrations are shown for contrast.....	6-10
6-9. Example of an analytical system change between two months that affected the relationship among three isomers.	6-11
6-10. Three years of 24-hr air toxics/PAMS concentrations with some problems noted.	6-12

LIST OF FIGURES (Concluded)

<u>Figure</u>	<u>Page</u>
6-11. Concentrations of carbon tetrachloride, dichlorodifluoromethane, and methyl chloride at an urban site in 2003 and 2004.	6-13
7-1. Entry screen for the Aethalometer Data Masher.....	7-1
7-2. Example Aethalometer Data Masher log file.....	7-3
7-3. Example hourly averaged PM _{2.5} standard temperature and pressure BC data.....	7-3
7-4. VOCDat ASCII file import pop-up windows.	7-4
7-5. Example time series plot of 1-hr average BC data.	7-5
7-6. Example of the data format required for preparation of an AQS Precision Report.....	7-6

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2-1. Directories and files installed by VOCDat by default.	2-2
3-1. Excerpt from a species file for air toxics containing the method TO-15 list of target compounds.	3-2
3-2. List of available species files.	3-3
3-3. Excerpt from the species2.txt file.	3-4
3-4. Units.txt file.	3-4
3-5. Excerpt from the Methods.txt file.	3-6
3-6. Nullcode.txt file.	3-7
3-7. Conversions.txt file excerpt.	3-8
3-8. Example of AIRS 80 character raw data conversion output.	3-12
3-9. Example of AQS R-2 data format for 1-hr VOC data.	3-12
3-10. Conversions made by VOCDat during import of AIRS files.	3-25
3-11. Example of how 1-hr and 3-hr data are listed in a VOCDat file.	3-28
5-1. Reactivity values (MIR and OH) for selected hydrocarbons.	5-4
6-1. List of steps for data validation.	6-2
6-2. Remote northern hemisphere background concentrations of selected air toxics taken from McCarthy et al. (2004) and other sources.	6-14

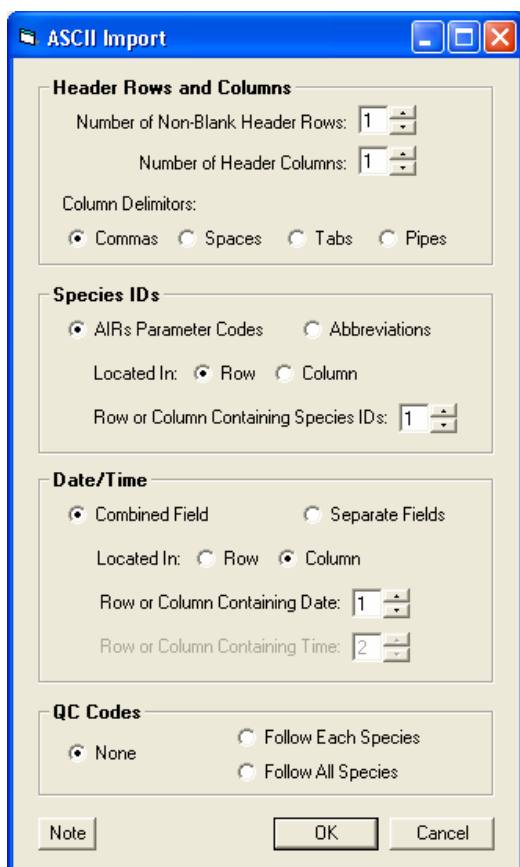
EXECUTIVE SUMMARY

VOCDAT: DATA VALIDATION AND AQS PREPARATION SOFTWARE

OVERVIEW

Since the mid-1990s, the U.S. Environmental Protection Agency has sponsored the development and enhancement of VOCDat software. VOCDat is a Windows-based program that imports and displays air pollutant data collected as part of the National Air Toxics Trends Sites (NATTS) and Photochemical Assessment Monitoring Station (PAMS) networks. Users can then perform quality control tasks on the data (e.g., apply screening criteria, visually inspect data), perform exploratory data analysis, and prepare data sets for import to the Air Quality System (AQS). One of the goals behind the development of this software was to enable the states to rapidly validate and release their data.

VOCDAT FEATURES



Imports and exports AQS R2 format. VOCDat can also import and export AMP370 (old AIRS) format data.

Imports flexible format ASCII files. This feature assists the user in handling data obtained from more than one laboratory and the various file formats in use.

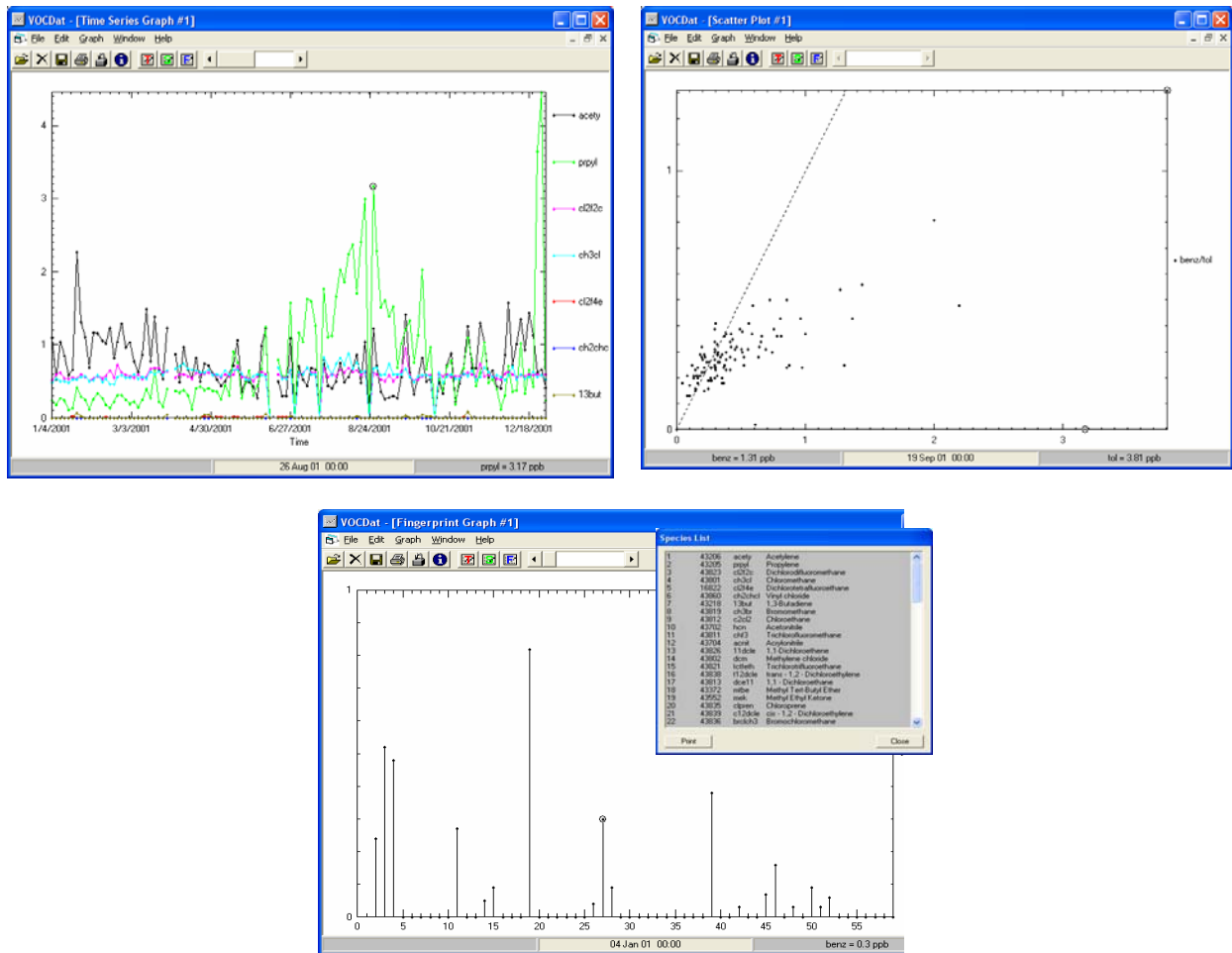
Imports other special types of data. For example, VOCDat imports pre-2003 versions of Perkin Elmer Turbochrome (automatic gas chromatograph) and more recent Totalchrome format data employed at some PAMS sites.

Provides export formats suitable for other software. To provide the user with more data analysis choices, VOCDat export options include formats suitable for use in spreadsheet and statistical software.

Allows on-screen editing of data quality control (QC) codes. QC codes are then translated into user-selected null data codes needed for AQS. The quality control (QC) codes may be changed for individual data points as well as for the entire sample on all plots. VOCDat keeps a log of changes including user-input comments.

Performs QC screening checks. The screening checks are designed to provide the user with samples that have potential validity problems. Screening checks are customizable to accommodate the differences in pollutant concentrations and characteristics among sites. VOCDat can also compute summary statistics of the pollutant concentrations.

Prepares graphical displays of time series, scatter, and fingerprint plots. The concentrations of the pollutants may be displayed using time series, scatter, and fingerprint plots. Time series plots show the concentrations of species over a specified time period and are useful in showing the temporal behavior. Using scatter plots, the relationship between species may be investigated. Fingerprint plots show the concentration of each species in a sample and help to identify unique characteristics of the samples.



Allows the user to convert units. The concentrations of the gaseous pollutants may be converted among ppb, ppbC, and $\mu\text{g}/\text{m}^3$ units.

OBTAINING VOCDAT

VOCdat software is available free at <http://vocdat.sonomatech.com>. To register for notification of future updates of the software and to obtain limited technical support, send an e-mail note to VOCdat@sonomatech.com. The VOCdat installation package includes species lists for the NATTS and PAMS networks and this user's guide. Training materials (presentation and training manual) are available by request.

1. DESCRIPTION OF THE SOFTWARE

1.1 BACKGROUND

Since the mid-1990s, the U.S. Environmental Protection Agency (EPA) has sponsored the development and enhancement of VOCDat software. VOCDat (**Figure 1-1**) is a Windows-based program that imports and displays air pollutant data collected as part of the National Air Toxics Trends Sites (NATTS), Photochemical Assessment Monitoring Station (PAMS), and particulate matter (PM) measurement networks. Users can then perform quality control (QC) tasks on the data (e.g., apply screening criteria, visually inspect data), perform exploratory data analysis, and prepare EPA Air Quality System (AQS)-ready data sets. VOCDat software was originally designed for the display, QC, and analysis of volatile organic compound (VOC) data; however, speciated air toxics and PM data can also be explored using this software. One of the goals behind the development of this software was to enable the states to rapidly validate and release their data.

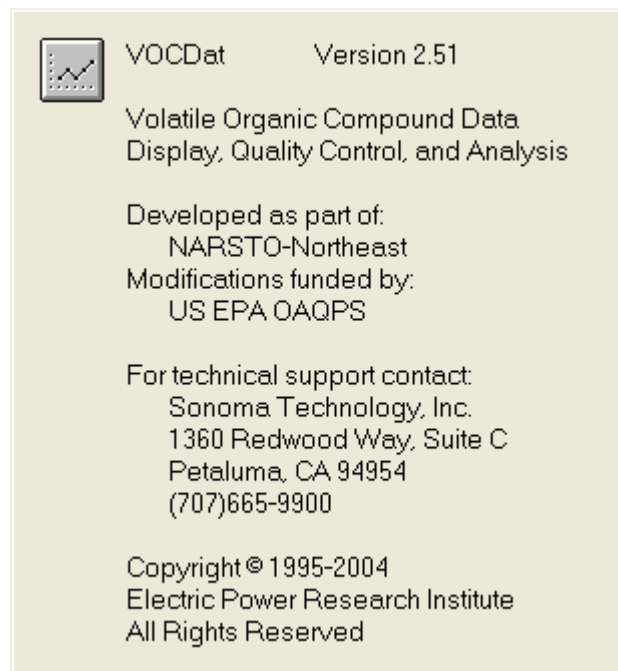


Figure 1-1. Opening screen for VOCDat.

In STI's experience with data analysis, much effort is often expended on validating the data prior to the analysis. VOCDat software provides a way for analysts to visualize their data and hopefully, fosters the validation of the data prior to submittal to AQS. 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. Data validation processes help to

- identify data with errors, biases, and physically unrealistic values before they are used for identification of exceedances, for analysis, or for modeling.

- produce a database with values that are validated and of a known quality.
- lead to the evaluation of the internal, spatial, temporal, and physical consistency of the data

1.2 OVERVIEW OF VOCDAT FEATURES

VOCDat allows the analyst to rapidly gain knowledge of a database. VOCDat can display pollutant concentrations using scatter, fingerprint, and time series plots. Using scatter plots, the relationship between species may be investigated. Fingerprint plots show the concentration of each species in a sample and help to identify unique characteristics of the samples. Time series plots show the concentrations of species in every sample over a specified time period and are useful in showing the diurnal, seasonal, or annual behavior of a species. The QC codes may be changed for individual and selected groups of data points as well as for the entire sample on all plots. Customizable screening criteria can be applied to the data to assist the user in identifying possible problems.

VOCDat is also equipped to calculate the weight percents of the PAMS data or multiply any pollutant concentration by a scaling factor, such as ozone formation potential or cancer risk. These altered data sets may then be explored using the graphical features of the software. For PAMS data, VOCDat calculates species groups, including the sum of the PAMS target species, total unidentified hydrocarbons, total aromatic hydrocarbons, total olefins, total paraffins, and total carbonyl compounds. In addition, other air quality and meteorological data (such as ozone, NO_x, wind direction, etc.) may be added to the data file, and new calculated data fields may be created (such as NMOC/NO_x ratios). Finally, the concentration data may be exported in a format suitable for submittal to the EPA AQS format or in a format easily importable into spreadsheet or statistical software packages.

VOCDat software has undergone several revisions since its first release to the NARSTO-Northeast participants in 1995 including Windows operating system and Visual Basic updates.

1.3 NEW FEATURES IN THE FALL 2004 RELEASE

The following items are new in the fall 2004 release of VOCDat:

- Enhanced import and export capabilities as follows:
 - Improved VOCDat capability to import data in a flexible ASCII delimited format including data stored in either rows or columns
 - Modified import and export options for AQS R-2 format
- Enhanced overall usability of the program:
 - Now contains several customized species lists including TO-15 and TO-11 air toxics, IMPROVE, and black carbon. Updates included adding unit risk estimates as factors for the air toxics files.
- Users can convert units of gaseous data among ppb, ppbC, and $\mu\text{g}/\text{m}^3$.

1.4 WHO CAN USE THIS PROGRAM?

Originally, VOCDat was developed for participants of NARSTO-Northeast in 1995 to enable the states to more rapidly validate and release their data. Through the 1995 study and a series of PAMS data analysis workshops (Main and Roberts, 2000), VOCDat has been employed by a wider audience including district, state, regional and federal air quality professionals across the United States. VOCDat is available to government employees free of charge. VOCDat is copyrighted by EPRI and all rights are reserved.

1.5 CUSTOMER SUPPORT

STI will try to assist users regarding operational questions or comments concerning the software based on the availability of funding for technical support. Please send an e-mail note to VOCDat@sonomatech.com or call STI for VOCDat assistance (707-665-9900); the receptionist will direct your call. For the most efficient assistance, please send an e-mail note with a description of your problem, an example file (including your species file), and the version of VOCDat and the computer operating system that you are using.

Users of the program are welcome to provide STI with their comments, suggestions, and questions so that efforts to improve the program and documentation can continue. We also need user e-mail addresses so that we can keep users informed of VOCDat updates.

2. INSTALLING THE SOFTWARE

2.1 OBTAINING THE SOFTWARE

The current version of VOCDat software is available from an Internet ftp site. As improvements and new features are added to the program, new versions will be placed at this site for distribution. The files of interest are compressed. The software may be obtained from the following web address:

<http://vocdat.sonomatech.com>

VOCDat Setup vx.xx.exe contains the latest full-install version of the VOCDat software, where x.xx refers to the current version number. Also included in the installation packet are copies of this user's manual, species files, and demonstration data. During installation, subdirectories are created automatically and these files are placed appropriately.

For assistance in obtaining or unzipping the files, please e-mail VOCDat@sonomatech.com. Once you begin using VOCDat, please register with STI by e-mail to the same address. By keeping a list of users of the program, we can send e-mail messages to all users when updates are available.

2.2 INSTALLING VOCDAT

2.2.1 Initial Installation (Full)

To install VOCDat on a Windows operating system, copy the file VOCDat Setup vx.xx.exe into any directory (e.g., C:\temp) and double-click on the file to execute it. Please note that the setup program may prompt you to reboot your computer; in this case, you will need to run setup again after rebooting. All temporary files created in the installation process are removed automatically during setup. The VOCDat program will be, by default, placed in a directory called C:\Program Files\VOCDat. **Table 2-1** lists the directories created and the files contained within the directories. During installation, please read pop-up windows carefully. The setup program will overwrite pre-existing versions of VOCDat, including species lists (i.e., species.txt and species2.txt) and configuration files (i.e., vocdat.cfg), so if you have made extensive changes to these files, be sure to back them up before you run the setup program.

Upon double clicking the setup executable, the window shown in **Figure 2-1** may appear. If you see this screen, choose Remove. Earlier versions of VOCDat will be removed. Once the uninstall process is complete, you will need to double click on the setup executable again to install the new version of VOCDat.

Table 2-1. Directories and files installed by VOCDat by default.

Folders	Files
C:\Program Files\VOCDat	executable VOCDat.exe; configuration file VOCDat.cfg; text files used by the program for methods, units, null codes, and conversions; default species files (species.txt and species2.txt)
C:\Program Files\VOCDat\Demo Files	Example Turbochrome files (*.TX1), ASCII toxics CSV file, PAMS VOCDat file, ASCII, NARSTO Northeast format file, AQS R2, old AIRS AMP370; template for precision reports
C:\Program Files\VOCDat\Manual	Most recent version of this user's guide
C:\Program Files\VOCDat\Species	Species files for PAMS, PM _{2.5} , PM, and air toxics (TO-11 and TO-15a lists)

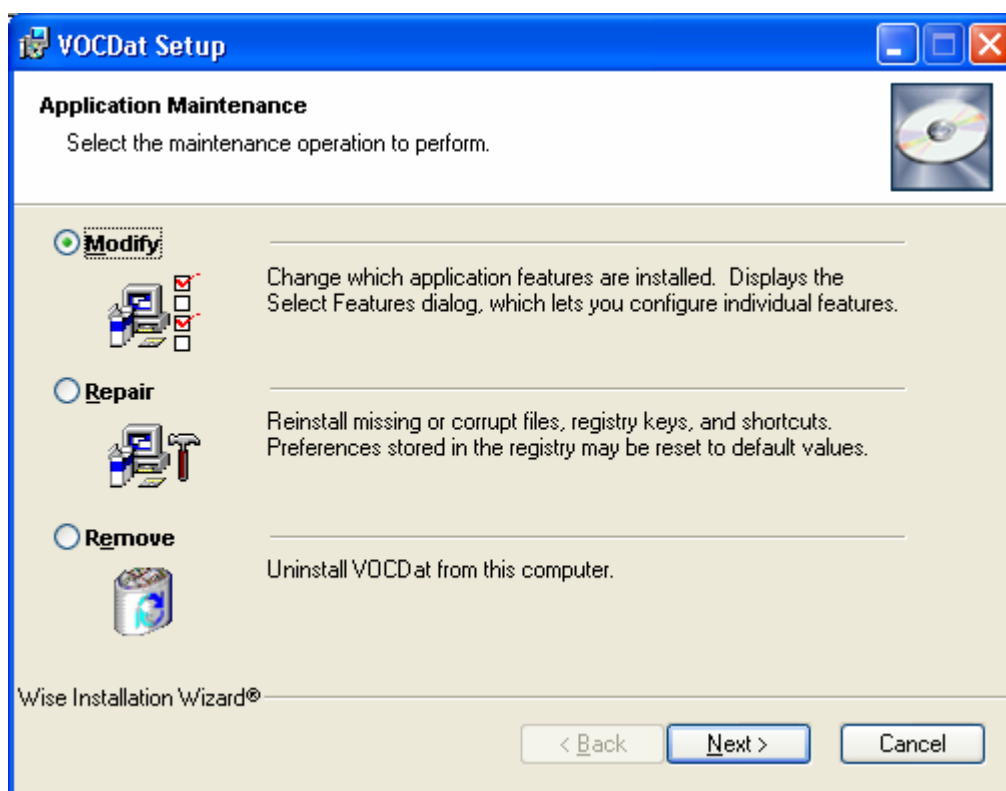


Figure 2-1. Screen that may be encountered during full installation of VOCDat software.

2.2.2 Update Installation (Executable Only)

If you already have VOCDat installed and are simply updating your system to use the most current executable file, all you need to download is the file labeled “VOCDat vx.xy Executable Only.zip”. Once it is on your computer, unzip this file and copy the latest executable file, “VOCDat.exe” over your older, existing file. No other changes should be necessary.

2.3 COMMON PROBLEMS ENCOUNTERED DURING INSTALLATION

Occasionally, a user will be prompted to reboot his/her computer during installation. If a reboot is necessary, the user must run setup again; no further reboots should be needed.

As with any other software installation, please close all open programs before you run setup. Some users have encountered problems installing VOCDat while the MS Office toolbar is running. Exit the installation process, close the toolbar, and try the installation again.

If you continue to have problems during installation, please first talk with your Information Technology (IT) support staff. If they are unable to solve your problem, please e-mail VOCDat@sonomatech.com with a description of your problem.

2.4 GETTING STARTED WITH THE SOFTWARE

After installation, select VOCDat from the Start button menu to start the program. The opening screen for VOCDat was shown in Figure 1-1. After loading, **Figure 2-2** shows the primary screen for VOCDat (version 2.x).

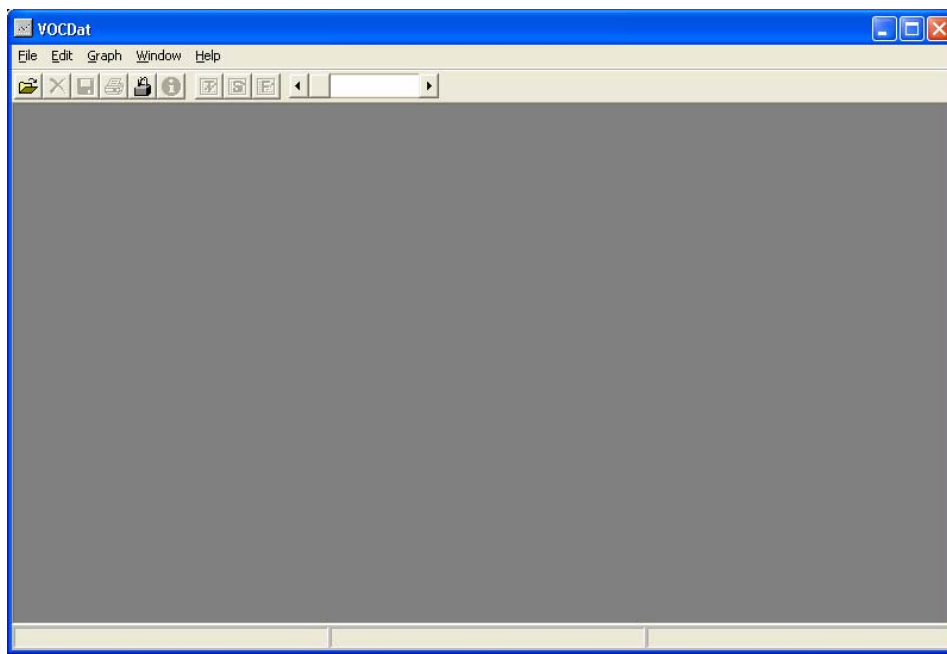


Figure 2-2. Primary screen for VOCDat (version 2.x).

To open an existing VOCDat file, select “File – Select Species List” from the menu at the top of the screen and choose “Species PAMS ppbC MIR.txt¹”. To find this file, select “browse”; the file is located by default in the “C:\Program Files\VOCDat\Species” folder. This selection tells VOCDat what species to expect in the data file to be opened. Next, select “File - Open” from the menu at the top of the screen or click the Open File icon. Open the example VOCDat file, “Example PAMS file.voc” (the file is located in the VOCDat software directory under the folder “Demo Files”); you are now ready to explore VOCDat features.

Section 3 discusses the procedure for creating new files and other detailed information on how to use the program.

2.5 WHEN VOCDAT TIMES OUT

VOCDat now contains a “time out” feature that reminds users to obtain the latest version of the software every year. When you receive the message shown in **Figure 2-3**, please obtain the new version of VOCDat from the ftp site discussed in Section 2.1. The old version will continue to operate; however, you will receive this message every time you open the program until the program has been updated. Updating the program periodically ensures that the most recent program adjustments and new features are being used.

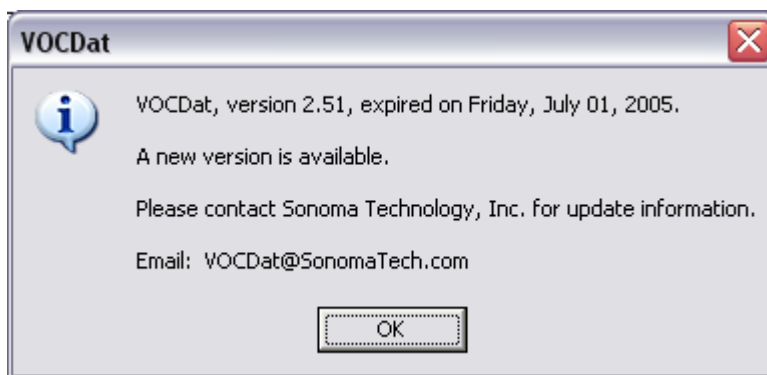


Figure 2-3. VOCDat “time out” error message indicating the need for the user to obtain a newer version of the program.

¹ This file contains a species list for PAMS target species in ppbC units with maximum incremental reactivity (MIR) as the scaling factor.

3. VOCDAT FILE IMPORT, EXPORT, AND MANIPULATION

This section discusses the VOCDat file formats including file types that may be imported and exported by the program. It also contains information on how to change file header information, append files, edit files in Microsoft Excel (or other spreadsheet software), close files, and customize the species list to fit your data.

3.1 HOW VOCDAT WORKS

VOCDat is very flexible in that the user can explore a wide range of data types. To support this flexibility, the user must specify the species contained in the files to be opened. VOCDat also uses several other supporting files to obtain information regarding units, method codes, and customized parameters.

3.1.1 The Species File

The species file is central to VOCDat (an example is shown in **Table 3-1**). This file tells VOCDat what species to expect upon import of new data or upon opening an existing VOCDat file. The file contains:

- AIRS parameter code, abbreviation, and full name
- Species type (used only for PAMS data) where the letters indicate O = olefin, P = paraffin, A = aromatic, C = carbonyl, and T = toxic. *These species groupings are not fully implemented for toxics and PM data and can be disregarded.*
- Weighting factor. The concentration data can be multiplied by user-defined weighting or scaling factors such as ozone formation potential (e.g., maximum incremental reactivity – MIR) or cancer risk.
- Units of the data. VOCDat assumes that the units in the data file are the ones listed in the species file – these are the units attached to the exported files.
- Whether the species are PAMS target species (TRUE) or not (FALSE). This field is used with PAMS data to identify which species to include in the computation of the sum of PAMS target compounds. *This field is not fully implemented for toxics and PM data and can be disregarded.*
- Analytical method code. VOCDat assumes that the method codes in the data file are the ones listed here – these are the units attached to the exported files.

Several species files have been developed and are available with the VOCDat software including those listed in **Table 3-2**. These files are found in the “Species” folder within the VOCDat software directory. The user may select which species file to use for each session to match the data files. If the same species file will be used for most data files, be sure to check the “Make this the default species file” checkbox. When the user customizes species files, we suggest storing them in the “Species” folder with a unique name.

Table 3-1. Excerpt from a species file for air toxics containing the method TO-15 list of target compounds.

AIRS Code	Abbrev.	Species	Type	Weighting Factor	Units	PAMS?	Method Code
43206	acety	Acetylene	T	0	ppb	FALSE	101
43205	prpyl	Propylene	T	0	ppb	FALSE	101
43823	cl2f2c	Dichlorodifluoromethane	T	0	ppb	FALSE	101
43801	ch3cl	Chloromethane	T	0	ppb	FALSE	101
16822	cl2f4e	Dichlorotetrafluoroethane	T	0	ppb	FALSE	101
43860	ch2chcl	Vinyl chloride	T	0	ppb	FALSE	101
43218	13but	1,3-Butadiene	T	0	ppb	FALSE	101
43819	ch3br	Bromomethane	T	0	ppb	FALSE	101
43812	c2cl2	Chloroethane	T	0	ppb	FALSE	101
43702	hcn	Acetonitrile	T	0	ppb	FALSE	101
43811	chf3	Trichlorofluoromethane	T	0	ppb	FALSE	101
43704	acnit	Acrylonitrile	T	0	ppb	FALSE	101
43826	11dcl	1,1-Dichloroethene	T	0	ppb	FALSE	101
43802	dcm	Methylene chloride	T	0	ppb	FALSE	101
43821	tctfeth	Trichlorotrifluoroethane	T	0	ppb	FALSE	101
43838	t12dcl	trans - 1,2 - Dichloroethylene	T	0	ppb	FALSE	101
43813	dce11	1,1 - Dichloroethane	T	0	ppb	FALSE	101
43372	mtbe	Methyl Tert-Butyl Ether	T	0	ppb	FALSE	101

The user should review weighting factors and units to make sure that they are up to date and appropriate for the data set in use. While we try to update these files periodically, the cancer risk factors often change. In the files available in August 2004, the most recent risk factors were only updated for the 18 core air toxics compounds explored by Hafner and McCarthy (2004). If available, the 2003 IRIS value was used. If this value was not available, the 2002 CAL EPA value was used. If this was not available, the value was left as it was in the previous file. If overall risk is being calculated, we recommend that the user check the online IRIS database at <http://www.epa.gov/iris/> and the CAL EPA web site at <http://www.arb.ca.gov/toxics/healthval/healthval.htm> for updated information.

All the species listed in the species file are exportable to AIRS or AQS formats.

3.1.2 Species2.txt

The species2.txt file also tells VOCDat which species to expect upon import or upon opening an existing VOCDat file. It differs from the species.txt file in that the species in species2.txt are not exportable to AIRS or AQS format. This file is intended to assist users in viewing complementary data or customized parameters with their concentration data without the possibility of accidentally (and inappropriately) exporting the data back into AIRS or AQS. The

Table 3-2. List of available species files.

Species File Name	Species Types	Units	Weighting Factor
species TO15 ppb risk_2004.txt	Air Toxics Method TO15	ppb	cancer risk ^a
Improve Species List.txt	IMPROVE PM _{2.5} species	µg/m ³	none
species TO15 ppb risk_2003.txt	Air Toxics Method TO15	ppb	cancer risk (older version)
Species PAMS ppbC MIR.txt	PAMS	ppbC	MIR
species pm 25 bc list.txt	PM _{2.5} Aethalometer Black Carbon	µg/m ³	none
species pm coarse tsp.txt	Coarse and TSP PM	µg/m ³	none
Species PM 25 10 STP ugm3 risk.txt	PM _{2.5} and PM ₁₀ at STP	µg/m ³	cancer risk
Species PM 25 10 LC ugm3 risk.txt	PM _{2.5} and PM ₁₀ at LC	µg/m ³	cancer risk
species TO11a risk.txt	Air Toxics Method TO11a	ppb	cancer risk
species TO11a MIR.txt	Air Toxics Method TO11a	ppb	MIR
Species PAMS ppbC OH.txt	PAMS	ppbC	OH reactivity

^a The risk factors in this version of VOCDat have been revised to reflect the most recent values from the IRIS and CAL EPA databases.

file contains the species type; criteria pollutants; meteorological parameters; and can contain user-specified ratios, sums, or other customized parameters (**Table 3-3**). This file is also customizable. The Type, Weighting Factor, PAMS?, and Method Code fields are placeholders and are not used by the program at this time. This file may be edited in Microsoft Excel or with a text editor that allows tab delimiters.

3.1.3 Units.txt

The Units.txt file is used by VOCDat to determine the units specified in the AIRS/AQS files and for export purposes. An excerpt of the file is provided in **Table 3-4**. This file may be edited in Microsoft Excel or with a text editor that allows tab delimiters.

Table 3-3. Excerpt from the species2.txt file.

AIRS Code	Abbrev.	Name	Type	Weighting Factor	Units	PAMS?	Method Code
16114	olefin	Olefins			ppbC	FALSE	0
16115	parafn	Paraffins			ppbC	FALSE	0
45000	aromat	Aromatics			ppbC	FALSE	0
43131	uidvoc	Unidentified VOC			ppbC	FALSE	0
42401	SO2	Sulfur dioxide			ppb	FALSE	0
42600	NOY	Reactive nitrogen oxides			ppb	FALSE	0
42601	NO	NO			ppb	FALSE	0
42602	NO2	NO2			ppb	FALSE	0
42603	NOX	NOX			ppb	FALSE	0
44201	O3	O3			ppb	FALSE	0
61101	WS	Wind Speed			m/s	FALSE	0
61102	WD	Wind Direction			degrees	FALSE	0
61103	RS	Resultant wind speed			m/s	FALSE	0
61104	RD	Resultant wind direction			deg	FALSE	0

Table 3-4. Units.txt file.

AIRS Units Code	Description
53	# DEFECTS/7.7 SQ. IN./MO.
51	% LOSS IN REFLEC./MONTH
38	10,000 FIBERS/CU METER
49	100 PARTICLES/SQ. IN.
48	1000 PARTICLES/SQ. IN.
27	BETA SCATTER
36	CALORIES/SQ. CM./HOUR
88	CCAL/SQ. CM./MIN
9	COHS/1,000 LINEAR FEET
57	CONDEN. NUCLEI/CU. CM.
72	CUBIC FEET/MINUTE
65	CUBIC METER
83	CUBIC METERS/MINUTE
14	DEG
17	DEG C
15	DEG F
28	DEGREES CENTIGRADE/100M
37	DEGREES-KELVIN
26	DEGREES-RANKINE

3.1.4 Methods.txt

The Methods.txt file is used by VOCDat to determine the analytical method code specified in the AIRS/AQS files and for export purposes. An excerpt is provided in **Table 3-5**. This file is also customizable using Microsoft Excel or a text editor that allows tab delimiters and may need to be updated periodically to account for new method codes.

3.1.5 Nullcode.txt

The Nullcode.txt file is used by VOCDat to determine the null code specified substitution for invalid data in AIRS/AQS files (**Table 3-6**). This file is also customizable using Microsoft Excel or a text editor that allows tab delimiters and may need to be updated periodically to account for new null codes.

3.1.6 Conversions.txt

The Conversions.txt file is used by VOCDat to enable the user to convert among ppb, ppbC, and $\mu\text{g}/\text{m}^3$ units for gaseous species (**Table 3-7**).

3.2 VOCDAT FILES

After importing data, the user is prompted to save the files in VOCDat software format (*.VOC). (Note that the user need not understand the file format to use the program.) The VOCDat files are tab-delimited containing the following header information:

- Project information (*user input*)
- Site information such as site name (*user input*), AIRS site code (*from some import files*), site coordinates (*user input*), and elevation (*user input*)
- File information including begin and end dates of the data in the file, number of samples, VOCDat version used to create the file, and sample interval (*generated by VOCDat*)
- QC code definitions
- Data

The data (e.g., **Figure 3-1**) are arranged in chronological order with the sample time (e.g., date and sample begin time), species concentrations, and QC codes for each species in each sample. Each row in the data file represents a separate sample. In the data portion of the file, the first seven rows are as follows:

1. Code: AIRS parameter code
2. Group: Species group designation: O = olefin, P = paraffin, A = aromatic hydrocarbon, C = carbonyl compound. The “True” or “False” designation signifies whether or not the species are PAMS target species. *These fields should be ignored for PM and air toxics data.*

Table 3-5. Excerpt from the Methods.txt file.

PARAMETER CODE	METHOD CODE	COLLECTION DESC	ANALYSIS_DESC	REF METHOD SHORT DESC	EQUIVALENT METHOD DESC	MIN DET LIMIT	SUMMARY SCALE
11101	1	MANUAL	GENERIC GEMS METHOD			0.1	2
11101	75	LEAD-PLATE	TITRIMETRIC(HUEY)			0.1	1
11101	77	MANUAL	BRAZIL PARTICULATE SAMPLER			1	0
11101	78	MANUAL	THAILAND PARTICULATE SAMPLER			0.1	2
11101	79	INSTRUMENTAL R&P M1400A TSP HD	TEOM-GRAVIMETRIC			1	0
11101	91	HI-VOL	GRAVIMETRIC	RFP-1171-001	REFERENCE METHOD	1	0
11101	92	MEMBRANE-SAMPLER	GRAVIMETRIC			1	0
11101	93	MILLIPORE-FILTER	BETA ABSORPTION			1	1
11101	94	CASSETTE	GRAVIMETRIC			20	1
11101	95	HOFFMAN-HI-VOL	GRAVIMETRIC			1	0
11101	79	Instrument R&P M1400A TSP Head	TEOM-Gravimetric			1	0

Table 3-6. Nullcode.txt file.

AQS Code	AIRS Null Code	Definition
AA	9967	Sample Pressure Out Of Limits
AB	9968	Technician Unavailable
AC	9969	Construction/Repairs In Area
AD	9970	Shelter Storm Damage
AE	9971	Shelter Temperature Outside Limits
AF	9972	Scheduled But Not Collected
AG	9973	Sample Time Out Of Limits
AH	9974	Sample Flow Rate Out Of Limits
AI	9975	Insufficient Data (Can't Calculate)
AJ	9976	Filter Damage
AK	9977	Filter Leak
AL	9978	Voided By Operator
AM	9979	Miscellaneous Void
AN	9980	Machine Malfunction
AO	9981	Bad Weather
AP	9982	Vandalism
AQ	9983	Collection Error
AR	9984	Lab Error
AS	9985	Poor Quality Assurance Results
AT	9986	Calibration
AU	9987	Monitoring Waived
AV	9988	Power Failure (Powr)
AW	9989	Wildlife Damage
AX	9990	Precision Check (Prec)
AY	9991	Q C Control Points (Zero/Span)
AZ	9992	Q C Audit (Audt)
BA	9993	Maintenance/Routine Repairs
BB	9994	Unable To Reach Site
BC	9995	Multi-Point Calibration
BD	9996	Auto Calibration
BE	9997	Building/Site Repair
BF	9998	Precision/Zero/Span
BG	9966	Missing Ozone Data Not Likely To Exceed Level Of Standard

Table 3-7. Conversions.txt file excerpt.

Code	Abbreviation	Name	Molecular Weight	Number of Carbons
16822	c2cl2f4	Dichlorotetrafluoroethane	170.92	2
17141	NAPHTH	naphthalene	128.16	10
43127	BEABYL	1-Butene&iso-Butene	56.11	4
43141	N_DODE	n-dodecane	142.29	12
43145	OCT1E	octene-1	112.21	8
43172	mheps	2- and 3-methylheptanes	114.23	8
43201	METHAN	methane	16.04	1
43202	ETHANE	Ethane	30.07	2
43203	ETHENE	Ethylene	28.05	2
43204	N_PROP	Propane	44.1	3
43205	PROPE	Propene	42.08	3
43206	acety	Acetylene	26.04	2
43212	N_BUTA	n-Butane	58.12	4
43214	I_BUTA	Isobutane	58.12	4
43216	T2BUTE	trans-2-Butene	56.11	4
43217	C2BUTE	cis-2-Butene	56.11	4
43218	13but	1,3-Butadiene	54.09	4
43220	N_PENT	n-Pentane	72.15	5
43221	IPENTA	Isopentane	72.15	5

[Project Information]								
Demonstration File								
[Site Information]								
Name = Lynn, MA PAMS type II site								
Code = 2006								
AIRS Code = 250092006								
Latitude (ddmmss) =								
Longitude (dddmmss) =								
Latitude (deg) = 0								
Longitude (deg) = 0								
UTM Northing (km) = 4704.157								
UTM Easting (km) = 337.855								
UTM Grid Zone = 19								
Elevation (ft) = 170.6								
Elevation (m) = 52								
[File Information]								
Creator = VOCDat								
Version = 1.57								
Begin Date = 7/5/94 13:00:00								
End Date = 7/31/94 23:00:00								
Number of Days = 26								
Number of Records = 575								
Number of Data Fields = 66								
Time Zone = EST								
Sampling Interval (hr) = 1								
[QC Codes]								
0 = Valid								
1 = Estimated								
2 = 9985 POOR QUALITY ASSURANCE RESULTS								
3 = 9999 UNKNOWN								
6 = Failed QC								
7 = Suspect								
8 = Invalid								
9 = Missing								
[Data]								
				Code:	43206		43203	
				Group:	O	TRUE	O	TRUE
				Factor:	0.14		2.16	
				Desc:	Acetylene		Ethylene	
Val	Start	Start	Sample	Record	acety		ethyl	
Lvl	Time	Hour	Time	QC	ppbC	QC	ppbC	QC
0	7/5/94 13:00	7/5/94 13:00	0:40	0	2	0	2.47	0
0	7/5/94 14:00	7/5/94 14:00	0:40	3	2.49	3	2.92	3
0	7/5/94 15:00	7/5/94 15:00	0:40	3	-999	9	-999	9

Figure 3-1. Example VOCDat file excerpt showing the header information and start of data.

3. Factor: Scaling factor from the species file in use. Factors can include maximum incremental reactivity (MIR) factors in moles ozone per mole carbon from Carter, (2000) (see Section 5.2), cancer benchmark values, or other scaling factor needed by the user.
4. Desc.: Species name descriptions
5. Species abbreviations used in VOCDat plots (e.g., acety, ethyl)
6. Description of units (e.g., ppb, ppbC, $\mu\text{g}/\text{m}^3$)
7. First row of sample data

An important feature of the VOCDat program is that for PAMS data, the program computes the sums for the following species groups: total aromatic hydrocarbons, total olefins, total paraffins, PAMS target hydrocarbons, total unidentified hydrocarbons, and total carbonyl compounds (if available). The total unidentified hydrocarbon mass is the calculated difference between the total nonmethane organic compound (TNMOC) value and the sum of all the reported species marked as PAMS target species = TRUE in the species list. The sum of the PAMS target species is discussed in Section 3.4.

3.3 IMPORTING FILES

VOCDat can import data from the following formats (**Figure 3-2**):

- Specially formatted Perkin-Elmer auto-GC output (*.TX*)
- “Old” AIRS AMP370 card image files (80 character output)
- AQS R-2 pipe-delimited format
- ASCII, delimited file format
- NARSTO-Northeast (NNE) data bulletin board comma-separated variable format

In all cases, once you have imported a file, you are prompted by the program for file header information (see Section 3.7) and to save the file in the VOCDat format (*.VOC) so that on subsequent use of the data, you simply open the *.VOC file. Examples of these file formats are provided in Appendix A. If your data are in another format, the example files can be used as templates to create files suitable for importing to VOCDat.

In all files, missing periods of time are allowed. Missing data values for species (in a time record that contains one or more other values) are given values of -999 and a QC code of 9. If data values are present, then the QC code is set to 0 with the exception of ASCII import files in which existing QC codes (0-9) may be imported. No special codes in the concentration fields, such as “ND”, are allowed. If VOCDat encounters non-numeric data in the concentration fields, the non-numeric data are assigned a value of -999 and a QC code of 9.

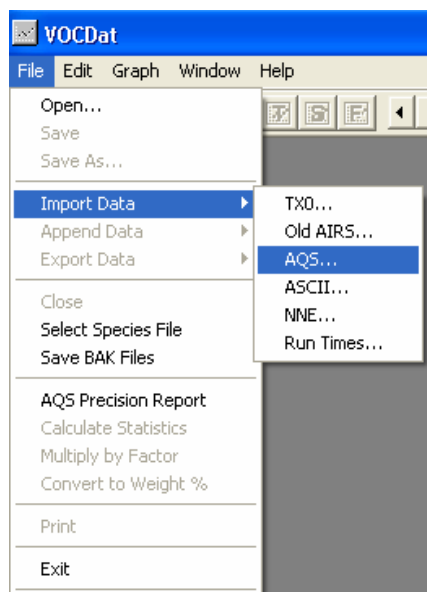


Figure 3-2. Screen shot showing the options for “File - Import Data”.

3.3.1 Perkin-Elmer TX0 or TX1

In the PAMS program, many states use Perkin-Elmer automatic gas chromatographs (auto-GCs) with Turbochrome or Total Chrome software. These systems have the capability to export data reports called *.TX0 or *.TX1 files. The user must follow the formatting instructions in Appendix A to create these files. VOCDat can only import the specially formatted files.

To import the correctly formatted *.TX* files, choose “File – Import Data - TX0”, find the directory containing your data, and highlight all the files you wish to import. This will create one *.VOC file that contains the concentration and QC code (initially set to 0 - valid) for each species in each sample. In addition to the concentration data available in the TX* files, retention time information is also available and can be imported into VOCDat. For this option, choose “File – Import Data – Run Times”, which will create a *.VRT file that contains the GC retention time (RT) and QC code for each species. The new files, *.VRT or *.VOC, must be saved using “File - Save” before exiting the program; the program will prompt you automatically once you have imported a file. Example Turbochrome files are provided (*.TX*) with the program. The retention time data may be explored using the same graphical features as the concentration data.

3.3.2 Old AIRS Files

Perhaps the most common data format used to import data into VOCDat in the past was an AIRS AMP370 raw data conversion format file. These files were originally submitted to or retrieved from the EPA AIRS database using the card image file format (e.g., 80 characters per line). An example old AIRS file is provided in **Table 3-8**.

Table 3-8. Example of AIRS 80 character raw data conversion output (report #370).

136005008342101110070549606010010009 0008 0008 0007 0008 0008 0009 0008 I
136005008342101110070549606010810008 0008 0008 0007 0005 0005 0005 0005 I
136005008342101110070549606011610005 0005 0006 0008 0009 0011 0012 0010 I
136005008342101110070549606020010008 0006 0005 0007 0007 0007 0008 0007 I
136005008342101110070549606020810007 0005 0005 0004 0004 0004 0004 0003 I

To import the AMP370 files, choose “File – Import Data – Old AIRS”, find the directory containing the data, and highlight the desired file. The program will read in the file and create a *.VOC file that contains the concentration and QC code (initially set to “0 - valid”) for each species. The user is prompted for the name of the new *.VOC file to save after import. We recommend that one site-month worth of hourly VOC data be imported at a time to VOCDat to keep file sizes manageable. Data from only one site per file may be imported into VOCDat. An example AIRS file is provided with the program (Example Old AIRS AMP370.air).

3.3.3 AQS R-2 Files

The re-engineered data warehouse at EPA is called the Air Quality System (AQS). An example AQS R-2 format is provided in **Table 3-9**. This format differs considerably from the original AIRS formats and contains more information. One of the notable changes is that there is only one sample provided per line and thus the R-2 files tend to be much larger than the older AIRS format discussed in Section 3.2.2. For a description of the reference information included in these files, visit http://www.epa.gov/aqspubl1/reference_tables.html (last accessed 7/28/04).

Table 3-9. Example of AQS R-2 data format for 1-hr VOC data.

RD I 25 009 2006 43202 5 1 078 125 19980601 00:00 6.52 A
RD I 25 009 2006 43202 5 1 078 125 19980601 01:00 4.13 A
RD I 25 009 2006 43202 5 1 078 125 19980601 02:00 5.1 A
RD I 25 009 2006 43202 5 1 078 125 19980601 03:00 5.01 A
RD I 25 009 2006 43202 5 1 078 125 19980601 04:00 5.47 A
RD I 25 009 2006 43202 5 1 078 125 19980601 05:00 4.97 A
RD I 25 009 2006 43202 5 1 078 125 19980601 06:00 5.24 A
RD I 25 009 2006 43202 5 1 078 125 19980601 07:00 3.93 A
RD I 25 009 2006 43202 5 1 078 125 19980601 08:00 3.06 A
RD I 25 009 2006 43202 5 1 078 125 19980601 09:00 AM A

To import the R-2 files, choose “File – Import Data - AQS”, find the directory containing the data, and highlight the desired file. The program will read in the file and create a *.VOC file that contains the concentration and QC code (initially set to “0 - valid”) for each species. An example file, “Example AQS R2.air”, is provided with the program. The user is prompted for the name of the new *.VOC file to save after import.

3.3.4 ASCII Files

Many users have air pollutant data stored in spreadsheets and requested some way in which to get the data into VOCDat. VOCDat import capabilities were extended by providing a method with which to import ASCII files with a variety of formats. The requirements on file formats are as follows:

- Data must be arranged with each row or column representing a separate sample time.
- Only data from one site at a time may be imported.
- Data should be sorted by date/time and should not contain any duplicate records.
- Acceptable delimiters are commas, spaces, tabs, or pipes.
- Any number of header records may be present. One of the header records must contain species identifications. VOCDat needs to know the number of non-blank header records.
- If the data already contain QC codes, VOCDat needs to know where the codes reside in the file (either after each species concentration value or after all the species concentrations).
- The species identifications may be AIRS parameter codes or abbreviations. Whichever one is used, the parameter codes or abbreviations in the species files must match exactly or the data will not be imported.
- The date and time of the samples must be specified. This information can be combined in a date/time field or provided in separate fields. VOCDat assumes the data are in standard time for exporting to AIRS format. VOCDat does not make any conversions of the time field. If your data are in daylight time, please change the data to standard time before importing to VOCDat.

To import ASCII files, choose “File – Import Data - ASCII”, find the directory containing the data, and highlight the desired file. The user will obtain the ASCII Import dialog box shown in **Figure 3-3**. This screen allows the user to input the information required to parse the data including the delimiter, number of header records, location of species identifications, type of identification, location of QC codes, and the location and format of the date and time fields. The program will read in the file and create a *.VOC file that contains the concentration and QC code (initially set to “0 - valid”) for each species. An example file, “Example ASCII Toxics File.csv”, is provided with the program. The user is prompted for the name of the new *.VOC file to save after import.

If incoming data values are non-numeric, such as “ND”, they are assigned a value of -999 and a QC code of 9. Previous versions of the software treated these values as zero. If you do not wish for these data to be identified as missing, we recommend replacing these fields with an appropriate value (such as MDL/2) prior to importing to VOCDat.

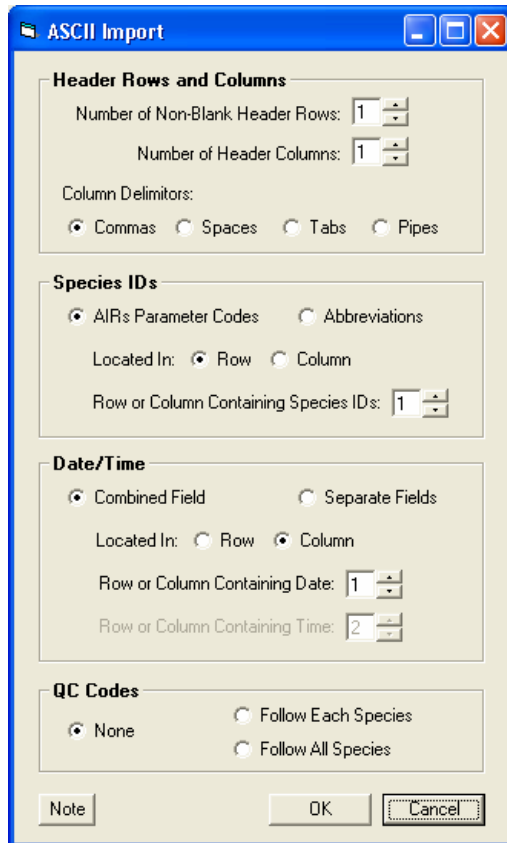


Figure 3-3. ASCII Import dialog box.

3.3.5 Legacy Files

NARSTO-Northeast CSV

VOCDat can import a special comma separated value (CSV) format that was designed to accommodate the VOC files that are posted on the NARSTO-Northeast (NNE) data bulletin board. The data are posted with one site-month worth of hourly or 3-hr VOC data. An example CSV file is provided with the program (Example NARSTO NE File.csv). Appendix A shows an example excerpt from a NNE format file. To read in the CSV format files, choose “File – Import Data - NNE”, find the directory containing the data, and highlight the desired file. This will create a *.VOC file that contains the concentration and QC code for each species. The user is prompted for the name of the new *.VOC file to save after import.

Experience with these files has shown that sometimes the file header may need to be edited to match the format of the test files. This is easily done using Microsoft Excel or a text editor that supports tab delimiters.

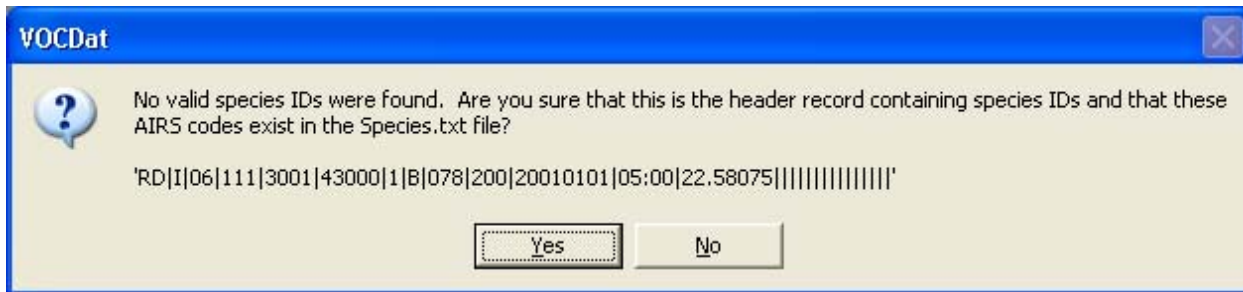


Figure 3-5. Incorrect species ID line error message.

Figure 3-6 shows an example error message received during import of an ASCII file indicating that there may be a problem with the date and time fields selected by the user. The user needs to check the import dialog box to see if the correct rows and columns are indicated. If correct, the user may need to reformat the columns using a spreadsheet or text editor. When all else fails, e-mail VOCDat@sonomatech.com for help.

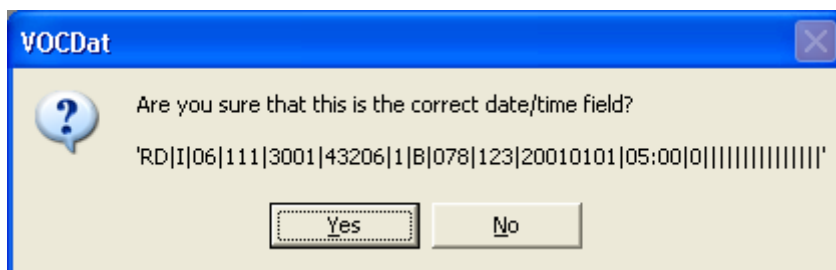


Figure 3-6. Incorrect date/time field designation message.

3.5 EXPORTING FILES

Data may be exported in the following formats: old AIRS format, AQS format, text without QC codes (DAT), and graph-specific (TXT) data (**Figure 3-7**). These formats are discussed in the following sections.

3.5.1 Old AIRS AMP370 Format

VOCDat can export data in the AIRS AMP370 format. This is a legacy format that was suitable for submittal to the EPA AIRS database; while still accepted by EPA at the time this manual was written, the preferred format is AQS R2. The old AIRS format is a card image file (e.g., 80 characters per line). Data with a QC code of 8 (invalid) are not exported to AIRS. Data with QC codes of 1 through 6 are exported as AIRS null codes rather than concentrations. Note that samples which have been flagged as suspect (QC code = 7) are exported to AIRS with no annotation. The concentration for each hydrocarbon and carbonyl species, the sum of the PAMS target species, and the total NMOC (AIRS code 43102) is exported for each sample. Other calculated sums and user-defined fields are not included in the output. Note also that only

concentration data are exportable in AIRS format; AIRS format files cannot be created with retention time, factor-scaled, or weight percent data.

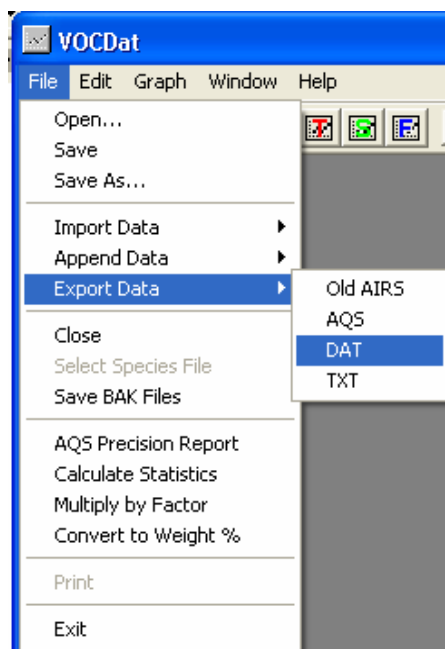


Figure 3-7. Screen shot showing the options for “File - Export Data”.

The old AIRS format files are given a *.AIR extension (e.g., filename.air). To export to old AIRS format, open a *.VOC file and choose “File - Export – old AIRS”. The user will obtain the AIRS Information dialog box as shown in **Figure 3-8**. The user-input information includes AIRS state, county, and site codes; POC code; method code; and action code. The option to fill time gaps in the data is provided, along with an AIRS null code for the gaps. Data gaps can only be filled between existing data points, i.e., if the last record in a file is missing, VOCDat cannot fill it in. The user can specify which species to include in the export. For reference, the minimum and maximum values for each species are provided in the window. Buttons to aid in selecting species for export are provided including:

- PAMS = selects the species identified as PAMS target species in the species file (labeled “TRUE”).
- In Use = selects species that are present in the current file.
- All = selects all species listed in the species file.
- Clear = clears all check marks so that the user can start over.

After providing the information in this screen, click ‘OK’. VOCDat will then create a *.AIR file that contains concentration (with units as specified in the species file) or an AIRS null code for each species. For PAMS data, VOCDat also exports the sum of PAMS target species (43000) and the total NMOC (43102). Note that VOCDat assumes all times are listed as standard time; the software does not make date/time conversions.

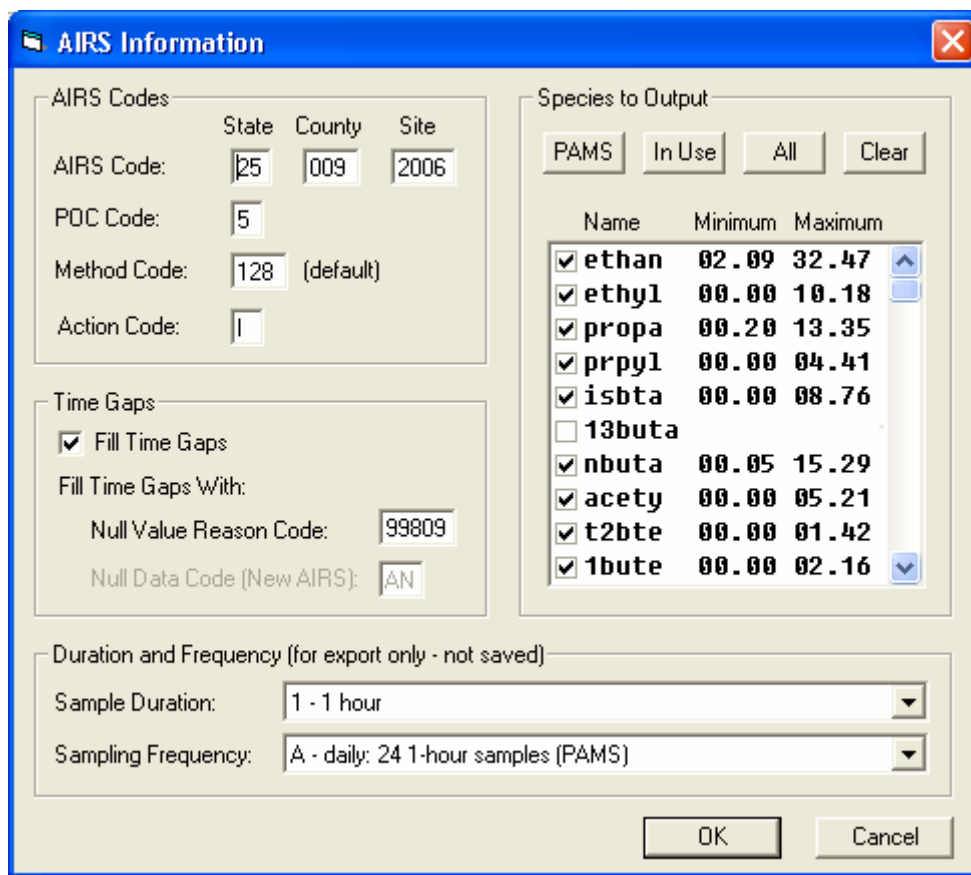


Figure 3-8. Example dialog box for export of a VOCDat data file to old AIRS format.

In exporting to old AIRS, the AIRS transaction record now includes a null data code of 9999 for PAMS target species that are missing from the sample. A decimal point (dp) indicator of 0 and a data flag of 9 are applied. Other user-selected null value reason codes are also given a dp of 0 and a QC flag of 9. If a species has a QC code of 0 or 7 (valid or suspect), the value is reported. If the QC code is 8, then 9978 (voided by operator) is reported. If the QC code is from 1 to 6, then the user-selected null value reason code is reported; these are defined in the QC Definitions and Define QC Code dialog boxes. If the QC code is 9, or no null value reason code has been assigned to QC codes 1-6, then 9999 is reported.

For PAMS data, a mixture of ppbC and ppmC units can be exported. This feature was added to accommodate some sites with very high concentrations.

3.5.2 AQS R-2 Format

VOCDat can export data in the AQS R-2 format suitable for submittal to the EPA's current data repository, AQS. Data with a QC code of 8 (invalid) are not exported to AQS. Data with QC codes of 1 through 6 are exported as AIRS null codes rather than concentrations (null codes for AQS are letters and letter combinations rather than numbers). Note that samples which

have been flagged as suspect (QC code = 7) are exported to AQS with no annotation. The concentration for each pollutant and a QC code are exported for each sample. For PAMS data, VOCDat also exports the sum of PAMS target species (43000) and the total NMOC (43102). Other calculated sums and user-defined fields are not included in the output. Note also that only concentration data are exportable in AQS R-2 format; AQS R-2 format files cannot be created with retention time, factor-scaled, or weight percent data.

The AQS format files are given an *.R2 extension (e.g., filename.R2). To export to the AQS format, open a *.VOC file and choose “File - Export Data – AQS”. The user will obtain the AIRS Information dialog box as shown in **Figure 3-9**. The user-input information includes AIRS state, county, and site codes; POC code; method code; and action code. The option to fill time gaps in the data is provided along with an AIRS null code (letters and letter combinations) for the gaps. The user can specify which species to include in the export. For reference, the minimum and maximum values for each specie are provided in the window. Buttons to aid in selecting species for export are provided:

- PAMS = selects the species identified as PAMS target species in the species file (labeled “TRUE”).
- In Use = selects species that are present in the current file.
- All = selects all species listed in the species file.
- Clear = clears all check marks so that the user can start over.

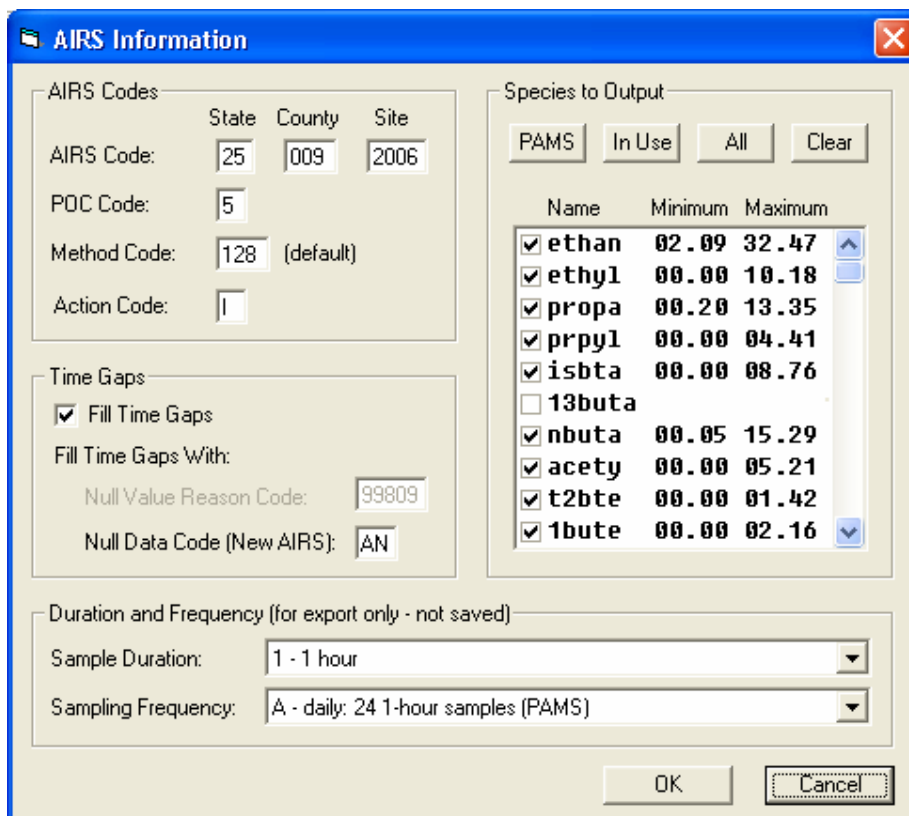


Figure 3-9. Example dialog box for export of a VOCDat data file to AQS format.

After providing the information in this screen, click 'OK'. VOCDat will then create a *.R2 file that contains concentration (with units as specified in the species file) or an AIRS null code for each species. For PAMS data, VOCDat also exports the sum of PAMS target species (43000) and the total NMOC (43102). Note that VOCDat assumes all times are listed as standard time; the software does not make date/time conversions.

The 2004 software version (v2.47 and above) no longer outputs the collection/sampling frequency code based on rejection of files by AQS that contained the code. The current documentation on the EPA web site asks for this code, apparently in error (see http://www.epa.gov/ttn/airs/airsaqs/manuals/aqstrans_format.pdf, page 9).

If a species has a QC code of 0 or 7 (valid or suspect), the value is reported. If the QC code is 8, then "AL" (voided by operator) is reported. If the QC code is 1-6, the user-selected null data reason code is reported as found in the QC Definitions and Define QC Code dialog boxes. If the QC code is 9 or no code has been selected for codes 1-6, then AM (miscellaneous void) is reported.

3.5.3 Text Without QC Codes

The VOCDat data files are set up to include a QC code in the column after every species concentration; this format can be cumbersome to use in other software packages. Rather than making the user manually delete each QC code column, VOCDat can export the data into one of two tab-delimited text file formats without QC codes; the QC codes which normally accompany each field are omitted for graphing convenience. There are two formats, *.DAT and *.TXT (see Appendix B). Both files contain the following:

- The data without QC codes for each individual species.
- Sample record QC code.
- Summary of the "worst" QC code for any individual species. Since the individual species QC codes are removed in this export, this feature was added so that analysts know that a particular sample had an individual specie that was flagged (such as missing or suspect). In our experience, an entire sample record is not always flagged.

Differences between the two file types are as follows:

- The TXT file has the same format as the VOCDat file (e.g., a header record, date/time field combined, several rows describing species names and units).
- The DAT file has only one header line that uses the species abbreviations. The date/time field is deconstructed into month, day, hour, day of week, and weekday/weekend. Species abbreviations may begin with a number.

To export a text file, open a *.VOC or *.VRT file and choose "File - Export Data -TXT" or "File - Export Data - DAT". These exported files contain the samples with QC codes of 0 through 7 (valid, user-defined, and suspect samples) including the data value for each species, the calculated sums, user-defined fields, and the total NMOC (for PAMS). The record QC code is retained so that records that have QC codes of 1 through 6 are identifiable and can be removed

by the user. Individual species QC codes are removed but a summary QC code is included for each record that indicates the “worst” QC given to any species in the record. All forms of the data may be exported using this format: concentration, factor-scaled, retention time, and weight percent data. These tab-delimited files open readily in Excel and are also importable into other software.

3.5.4 Graph-Specific Data

VOCdat also allows the user to export graph-specific data. First, create a graph (i.e., scatter plot, time series, or fingerprint). Then, select “Graph - Export CSV” (comma separated values) or “Graph - Export TXT”. These options export only the data shown on the selected graph. The format of the data varies depending on the type of graph exported, but all formats are easily imported into an Excel spreadsheet or other software for additional graphing capability. Please see Appendix B for examples of these output formats.

3.5.5 Additional Export Notes

For all export options (except *.DAT), the units in the exported files indicate the form of the data. If more than one data file is open, the user is prompted for the name of the file to export and for the name and location of the exported file.

3.6 APPENDING FILES

Once *.VOC and *.VRT files are created and saved, they can be read into VOCdat later by choosing “File - Open” or the Open File button on the tool bar. New data for the same site can be added to the *.VOC and *.VRT files by choosing “File - Append”. The file that is to be appended to must be open, and the *.TX0, *.AIR, or *.VOC files with the data to be added must be highlighted.

The software limits you (1) from appending files to more than four different sites open at one time (note that data may only be appended to the same site); (2) to 250 parameters (i.e., species and user-defined fields); and (3) to a number of records only limited by available computer memory and disk space. Also note that when two files are appended, while the QC codes (i.e., 0-9) in the second file (i.e., the one to be appended to the currently open file) are retained, the possibly unique QC code definitions for that file are ignored. This is only important if different QC definitions are used for different files; therefore, it is recommended that a standard set of QC definitions be used for all files.

Please take care with this feature when appending TX0 files. If you are adding data collected yesterday to a file containing the previous few days of data, only append yesterday’s data, not all the days. Errors can result if the same TX0 data are appended more than once.

3.7 CUSTOMIZING THE PARAMETER LIST

Upon importing data, the user may find that the species files supplied with the program do not contain all the parameters for which his/her data are available. In that case, the user needs to customize the parameter list to better suit the data. In this discussion, “species” is defined as a specific pollutant of interest (i.e., air toxic, PM, or VOC) and “fields” is defined as other parameters such as user-defined variables (e.g., TNMOC/NO_x ratios) and wind speed, criteria pollutant concentration, etc. The new field(s) (e.g., air quality parameters and user-defined parameters) must be added at the bottom of the field list and species must be added above PAMSHC in the list (the PAMSHC parameter is used by VOCDat as the final specie in the list). For VOCs, we recommend listing species in the order observed in a chromatograph to facilitate data validation. Note that “species”, which are listed above TNMOC, are stored in the species file in use. The “fields”, i.e., those parameters listed below TNMOC, are stored in the file called “species2.txt”. Only species, the TNMOC, and the sum of the PAMS target species are exported to AIRS or AQS formats.

The simplest way to edit the species lists is to use Microsoft Excel or a text editor such as Ultraedit. We recommend copying an existing line and replacing fields in the line in order to retain the proper spacing and delimiters.

If the user chooses to make the changes within VOCDat, the procedures for adding species or “fields” are as follows:

1. Open VOCDat and enable editing (“Edit - Enable Editing”, or click on the padlock icon). As delivered, the program does not require a password to enable editing, simply press the Enter key at the prompt, or click “OK”.
2. Close all open VOCDat files.
3. Select “Edit - Edit Field Definitions”. *All data files must be closed in order to access this option.* A window showing field definitions pops up (as shown in **Figure 3-10**). This window shows the order number, AIRS code, abbreviation, full name (description), species group (i.e., O, P, A, C), factor, units, and whether or not the parameter is a PAMS target species.
4. Highlight the row just below the selected insertion point. Select “Insert Field” (or “Insert Species”). Note that “species” refer to the species of interest and “fields” refer to other parameters as defined above. The “Edit Field Definition” (or “Edit Species Definition”) window appears (**Figures 3-11 and 3-12**).
5. Define the new species or field in this window. The “Species Code” or “Field Code” is usually the AIRS parameter code (may be left blank for user-defined field definitions). “Species Name” or “Field Name” is the value provided in plots; short names or acronyms are preferred. “Species Description” or “Field Description” is a more detailed description of your new field. For species, there is a “Species Group” drop down selection pertinent only to PAMS with A (aromatic), C (carbonyl), D (placeholder), O (olefin), or P (paraffin). “Species Units” or “Field Units” are the units of the new field; a drop down menu is available so that you can select the appropriate units. For species, there is “Species MIR” which is the field for a weighting factor. Also for species, “PAMS Target” is a true or false selection that defaults to FALSE. Only species that are labeled

TRUE in this field are summed for the PAMS target species group; this field is not pertinent to other sets of data (i.e., air toxics, PM). Finally, for species, “Method Code” allows the user to select the appropriate method code from a drop down window. If the method is not available from the list, the user can modify the Methods.txt file.

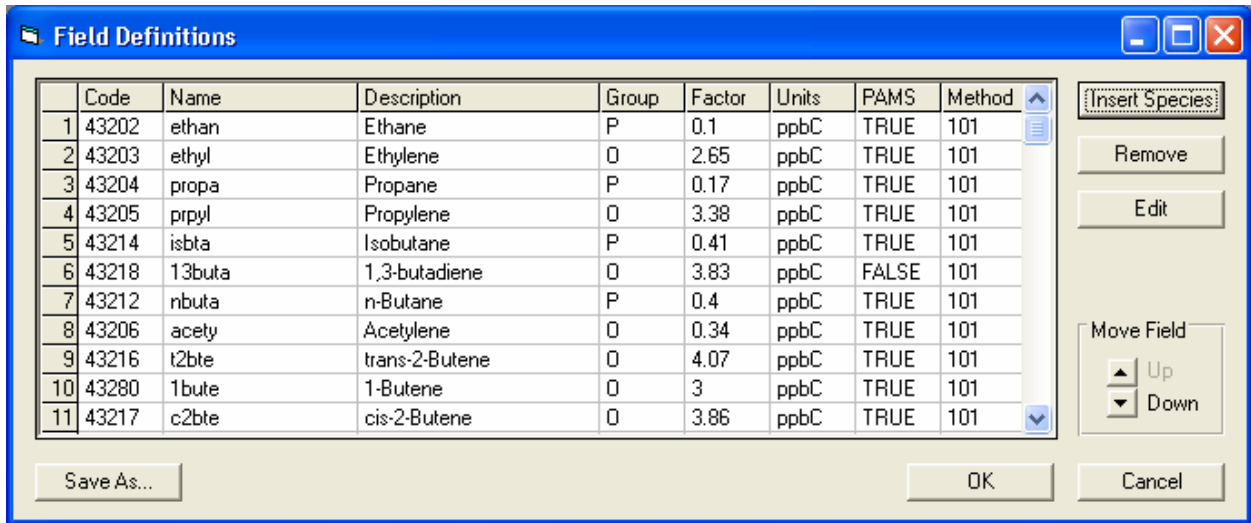


Figure 3-10. Dialog box for customizing the parameters list.

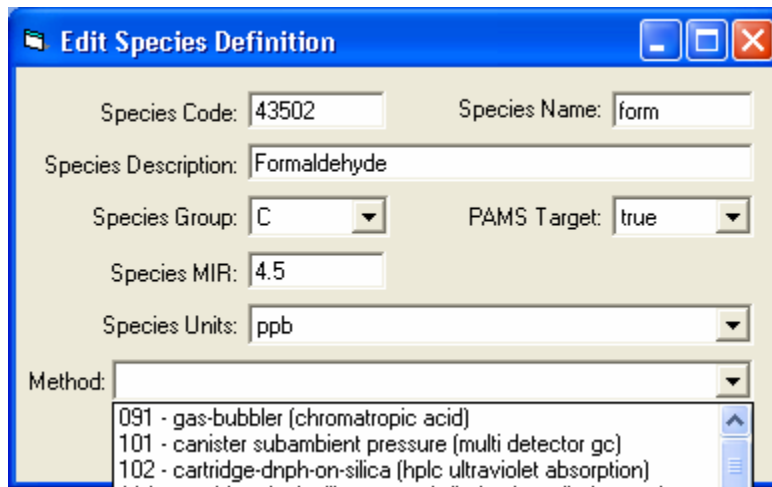


Figure 3-11. Dialog box for creating or editing a species definition showing the drop down menu for method code.

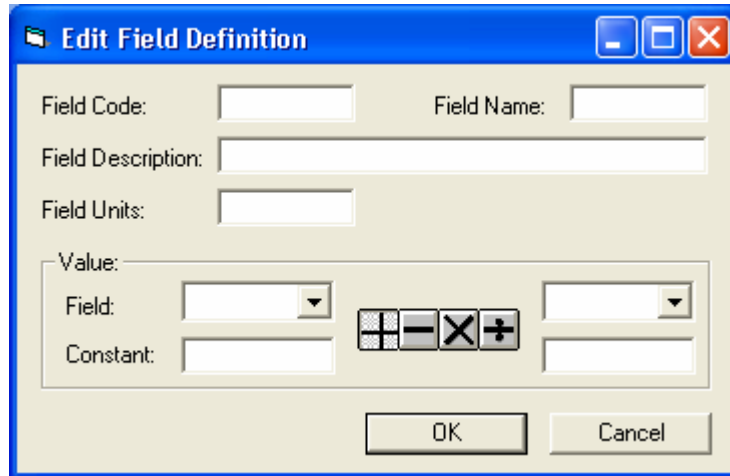


Figure 3-12. Dialog box for creating or editing a field definition.

6. In the “Edit Field Definition” window you can create new fields from existing fields and species using simple calculations (add, subtract, multiply, and divide). To create a new field requiring several mathematical manipulations, break the calculation down into simple steps. For example, to calculate the ratio of total xylenes to benzene, first create a field that sums m-and p-xylenes to o-xylene. Next, divide the new field (total xylenes) by benzene to get your ratio. It is recommended that the user retain all intermediate steps, in the order required, for recalculation purposes. In the case of the example, do not delete the total xylenes field.

The species list may need to be changed for the following reasons:

- The PAMS target list may have changed from year to year, or be different for different sites.
- Data delivered from different laboratories may have different method codes, abbreviations, or units.
- Reporting organizations using non-standard, but approved, parameter codes (such as the ones indicating coelution) can tailor the list to most accurately reflect their operations.
- The analyst wishes to create a “dummy” code to look at data not yet assigned a parameter code by EPA.

Note that the Summary field, including the sum of PAMS target species, will be updated to the current list, so be cautious when opening archived files. We recommend that a copy of the species list pertinent to a specific year be archived with the data for that year.

To facilitate moving among species lists, VOCDat allows the user to select a species file using “File – Select Species File”. The resulting dialog box is shown in **Figure 3-13**. The user can browse to find the selected file. Note that to select a species file all VOCDat data files must be closed. The default species file can be set using this feature.

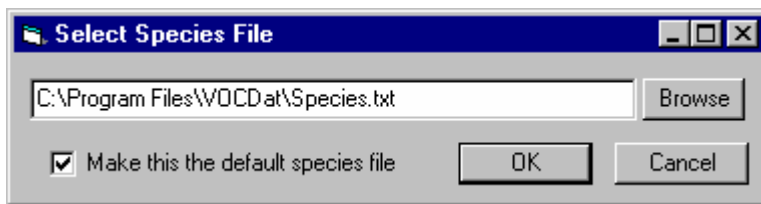


Figure 3-13. Select Species File dialog box.

3.8 PARAMETER UNITS

VOCDat currently performs a few conversions during import of an AIRS or AQS file based on the AIRS unit codes and VOCDat field descriptions (**Table 3-10**).

Table 3-10. Conversions made by VOCDat during import of AIRS files.

Unit	Converted to
ppmC	ppbC
pphmC	ppbC
ppm	ppb
pphm	ppb
mph	m/s
km/h	m/s
knots	m/s
deg F	deg C
deg K	deg C
in Hg	m
$\mu\text{g}/\text{m}^3$	ppb for NO, NO ₂ , NO _x , SO ₂ , O ₃ , CO

If the imported AIRS or AQS data are in the units on the left (e.g., mph), the data will be converted automatically by VOCDat to the units on the right (e.g., m/s). If the AIRS or AQS data are in the units on the right, no change is made. For ASCII files, no changes are made to the data on import because VOCDat has no way of knowing the units of the incoming data; the program relies on the units as designated in the selected species file. For ASCII files, VOCDat will import these data with no change and the data will be labeled (perhaps improperly) with whatever units are in the species file for that specie. Therefore, it is important to know the units of your data and to check these units against the species file you intend to use.

If the AIRS or AQS data are in other units not shown in this list, VOCDat will notify you that it does not recognize the units. VOCDat can be instructed to import the data (without change); however, VOCDat will retain the label shown in the right-hand column. However, new fields may be created to perform the desired conversion. Once the new fields are set up in your species2.txt, the AIRS or AQS file containing species that were ignored earlier may be appended

to your existing file or the file may be re-imported. The following parameters have been included in the current version of the species2.txt file: ozone, NO_x, NO, NO₂, NO_y, outdoor temperature, NMOC/NO_x ratio, wind speed, wind direction, resultant wind speed, resultant wind direction, relative humidity, and barometric pressure. Solar radiation and UV radiation are listed, but the user must add the actual units imported to the parameter list.

3.9 CHANGING THE FILE HEADER INFORMATION

During import, the user will be prompted for file header information including the project title, site name, site location in UTM coordinates (a utility is provided to facilitate this calculation from latitude/longitude data), sample duration (length of sample time, e.g., 1-hr, 24-hr), sample frequency (how often the samples are taken e.g., every third day), and the time zone of the data (see **Figure 3-14**). All this information is optional and is user-input. The “Site Name” field is used in the plots and screening criteria output to identify data. The “Site Code” field (an AIRS 4-digit code) is also used to identify the file in the plots. The file header can be changed later using the “File Info” button on the toolbar or “Edit - Edit File Header”. One source of the site information is the EPA AQS database. Plausible options for sample duration and sample frequency are provided in drop down menus. These selections are used to provide duration and frequency codes on export to AQS R-2 files.

3.10 CLOSING FILES

Multiple files may be open at the same time. During a QC or analysis session, the user may wish to close one or more files. Either “File - Close” or the Close File button may be used to close files. The user will be prompted for which file(s) to close. This window contains the file name, the site (if provided in the header record by the user), and the site code in parentheses. Be sure that you have highlighted the file to be closed (i.e., you must click on the desired file name).

If you have forgotten which files are open, the Close file function may also be used to display a list of open files. Choose “Cancel” to dismiss the dialog box without closing any files.

Figure 3-14. File header information screen.

3.11 HANDLING DIFFERENT SAMPLING TIMES

VOCDat can import 3-hr, 8-hr, or 24-hr average data, or data with other sampling intervals, in addition to 1-hr data. Note that two (or more) time intervals may be represented in a VOCDat file. For example, hourly GC data and 3-hr average carbonyl data may exist for the same site. All data, including the longer sampling intervals (e.g., 3-hr) are listed at their *begin hour*. Therefore, in a VOCDat file with 1-hr hydrocarbon and 3-hr average carbonyl data, the carbonyls would have data values at their begin hour and the two following hours would contain missing codes as shown in **Table 3-11**.

In graphs, the data for the longer time interval would only show a data value for the begin hour and nothing for the next hours. If the data in the file are all the same time interval, the data points will be connected by lines. Otherwise, only those data at or below the “sample duration” as listed in the file header will be connected.

Table 3-11. Example of how 1-hr and 3-hr data are listed in a VOCDat file.

Date	Hour	TNMOC	Formaldehyde
8/19/95	01:00	30.1	1.4
8/19/95	02:00	34.5	-999
8/19/95	03:00	32.3	-999
8/19/95	04:00	41.2	2.1
8/19/95	05:00	44.1	-999
8/19/95	06:00	40.2	-999

Note that if a file contains 3-hr TNMOC averaged values and 1-hr NO_x values, the TNMOC/NO_x ratios calculated by VOCDat are not correct because the program uses the two matching values in the ratio even though the intervals do not match. The user needs to separately calculate 3-hr average NO_x data (e.g., in a spreadsheet or database) and match it to the TNMOC data to obtain correct ratios.

3.12 SPECIES ABBREVIATIONS

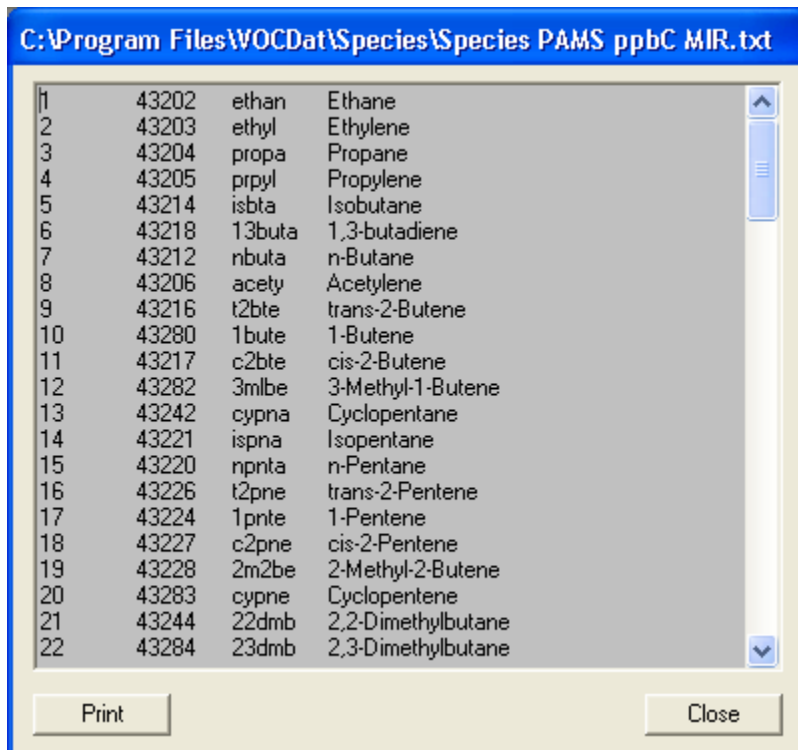
Each species and species group is assigned an abbreviation that is used in the legend of the graphs. Note that the abbreviations provided with the program are from the EPA PAMS manual (U.S. Environmental Protection Agency, 1994). On the fingerprint plots, the species are listed by number rather than by code because of space limitations. To view the species list at any time, select “Window - Show Species List” (see **Figure 3-15**). The species are displayed in the order shown in the list (e.g., 1 = ethane, 2 = ethylene, etc.). The user can print the list for reference purposes; this list contains the number shown in the fingerprint plot, the AIRS parameter code, the species abbreviation, and the full name as shown in the species file. The pop up window also shows the name and location of the species list currently in use.

The order of species may be changed using “Edit - Edit Field Definitions”. Select the species you wish to move by clicking on it and then clicking on the “move up” or “move down” arrows.

3.13 EDITING VOCDAT FILES IN EXCEL

To edit files, open the selected VOCDat file in Excel by selecting tab-delimited files and extensions of *.*. (You can also drag and drop VOCDat data and species files into Excel or on the Excel icon.) A file import window pops up; select “OK” to toggle through all the screens or select “Finished” to move right to the import (Excel will automatically recognize that the file is tab-delimited). Once imported, changes can be made to the file such as deleting some samples, changing the QC codes for one of the species across a time period, etc. Save the new file as a tab-delimited file, otherwise VOCDat will not recognize the file. After you save the tab-delimited file and exit the program or close the file, Excel displays a prompt that states that “the

current file format is not Microsoft Excel Workbook” and do you wish to “Save changes in ‘filename.voc’?”. Answer “No” to this prompt; otherwise the file will be saved as an Excel workbook and will not be readable by VOCDat.



The screenshot shows a window titled "C:\Program Files\VOCDat\Species\Species PAMS ppbC MIR.txt". The window contains a list of 22 species, each with a line number, a numerical ID, a three-letter code, and the full name of the species. The list is as follows:

Line	ID	Code	Species Name
1	43202	ethan	Ethane
2	43203	ethyl	Ethylene
3	43204	propa	Propane
4	43205	prpyl	Propylene
5	43214	isbta	Isobutane
6	43218	13buta	1,3-butadiene
7	43212	nbuta	n-Butane
8	43206	acety	Acetylene
9	43216	t2bte	trans-2-Butene
10	43280	1bute	1-Butene
11	43217	c2bte	cis-2-Butene
12	43282	3mlbe	3-Methyl-1-Butene
13	43242	cypna	Cyclopentane
14	43221	ispna	Isopentane
15	43220	npnta	n-Pentane
16	43226	t2pne	trans-2-Pentene
17	43224	1pnte	1-Pentene
18	43227	c2pne	cis-2-Pentene
19	43228	2m2be	2-Methyl-2-Butene
20	43283	cypne	Cyclopentene
21	43244	22dmb	2,2-Dimethylbutane
22	43284	23dmb	2,3-Dimethylbutane

At the bottom of the window, there are two buttons: "Print" and "Close".

Figure 3-15. Species list window.

The *.VOC files may also be appended using Excel. To append “file b” to “file a”, open both files in Excel as described above. Next, copy the portion you wish to append from “file b” to “file a”. The data should be kept in chronological order, all header information from “file b” should be omitted, and the species listed in both files should be the same. If a file is missing a particular species or data field, fill the field with -999 (VOCDat’s missing data code).

4. DISPLAYING AND FLAGGING DATA

Current users of VOCDat have found that they can become familiar with a fairly extensive database in a short period of time. The following sections present some of the ways in which analysts have been using VOCDat to display, explore, and QC their data.

4.1 DISPLAYING DATA IN VOCDAT

Pollutant concentrations can be displayed in three ways: time series, scatter, and fingerprint plots. Display options are available both under the Graph menu and by selecting buttons on the tool bar. Options are available for changing many aspects of a graph. Also, many features on a graph can be displayed with different colors. On each graph, the QC code of the entire sample (record) or for an individual species may be changed.

4.1.1 Time Series Graphs

Time series graphs are displayed by pressing the “T” button on the toolbar (**Figure 4-1**) or by selecting “Graph - Time Series - New”. The “Select Fields to Graph” window pops up to show the available sites (open files) and species or species groups (**Figure 4-2**). Multiple sites and species may be plotted. Hold the mouse button down and drag to select multiple contiguous species; use the Control key and mouse to add individual species selections; or use the Shift key and mouse to add all species between two already-selected species. Up to 12 species may be plotted in the same graph.



Figure 4-1. “T” button on the VOCDat toolbar.

The “Value Axis” Scale window sets the value of the y-scale parameters (**Figure 4-3**). The user can select a log scale, if desired. Accept the default parameter values by clicking the “OK” button or pressing the Enter key.

The “Time Scale” window provides the time scale options (**Figure 4-4**). The default is to show the data in 7-day increments (i.e., the “Last” box is checked). This can be altered by the user to include all the data in the graph (uncheck the “Last” box) or to include increments of 1 day, for example (change the 7 days to 1 day).

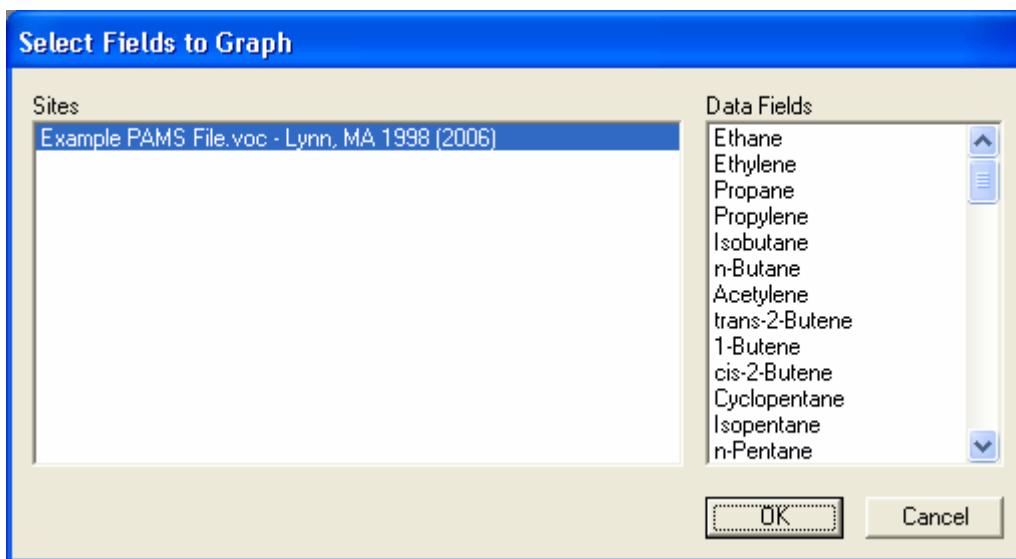


Figure 4-2. Window for selecting species and sites to graph.

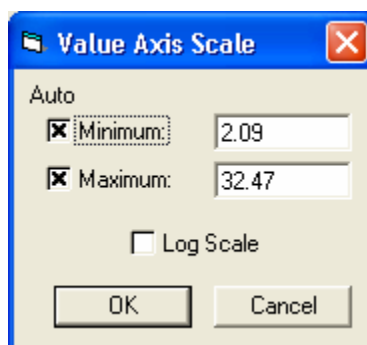


Figure 4-3. Value Axis scale pop-up window.

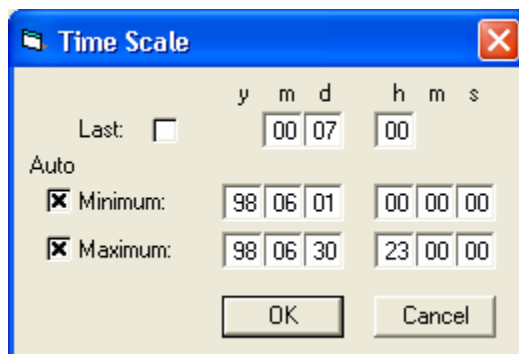


Figure 4-4. Time scale pop up window.

To later change a graph's features, select the graph, click the "T" button or select "Graph -Time Series", and change species, value axis (y-axis), or time axis features.

Figure 4-5 shows an example time series plot using all the defaults. When you click on a point on the plot, the site, date and time, and species name and concentration are shown in the status bar at the bottom of the screen. The QC codes for individual species, entire sample records, and/or samples from selected time periods may be edited in time series graphs.

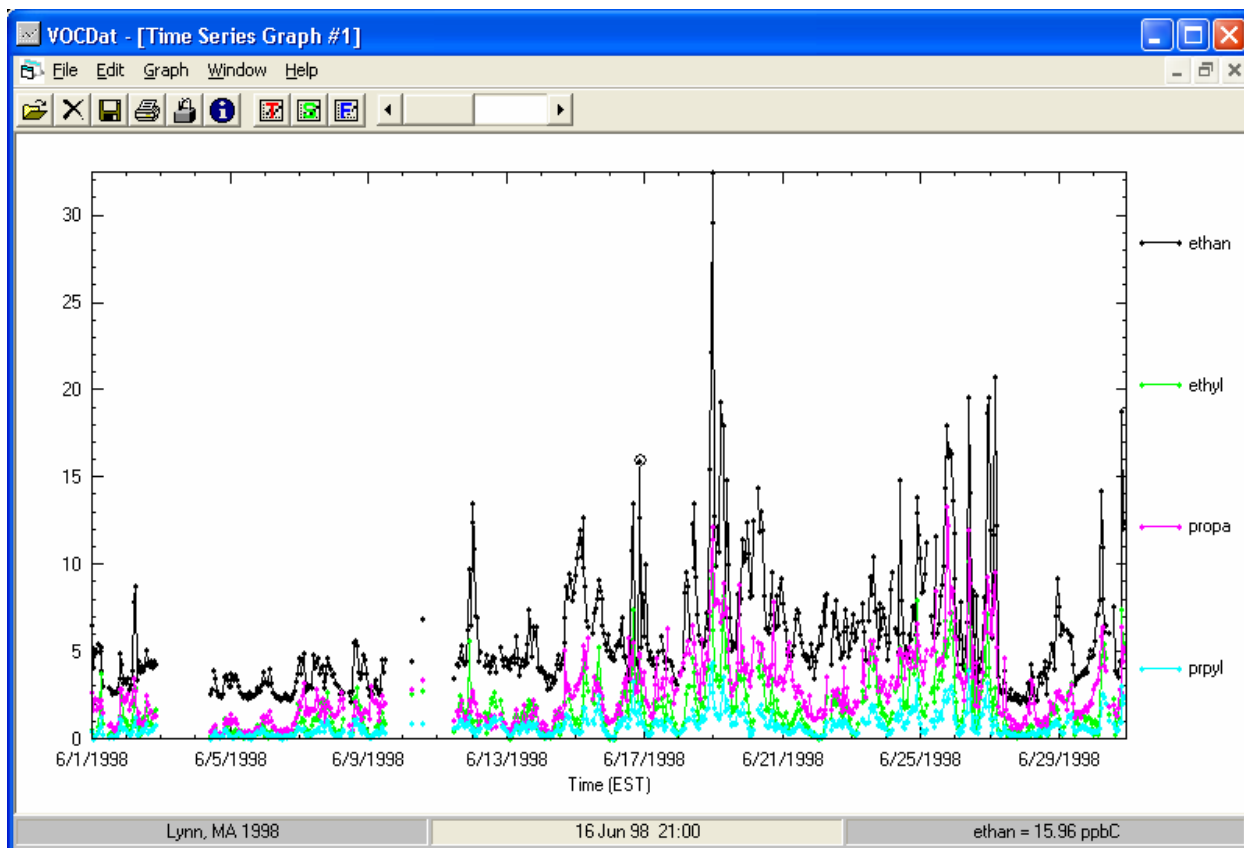


Figure 4-5. Example time series plot with highlighted data point information shown in the status bar.

Diurnal plots of sub hourly data may be prepared by selecting “Last 1 day” in the “Time Scale” window (Figure 4-4). The start hour for the graph may be changed by clicking on the ahead and back arrows on the scroll bar.

4.1.2 Scatter Plots

Scatter plots may be prepared by pressing the “S” button on the toolbar (**Figure 4-6**) or by selecting “Graph - Scatter - New”. A window pops up to show the available sites and species or species groups for the x-axis (see Figure 4-2). After selection, a window shows the available data for the y-axis (same as the list shown for the x-axis). The following windows provide the scales for the x- and y-axes (**Figure 4-7**). The next window provides the time scale (see Figure 4-4). The time scale default is to include all the data. However, the user may change this to include any “time window” of data.



Figure 4-6. Scatter plot button on the toolbar.

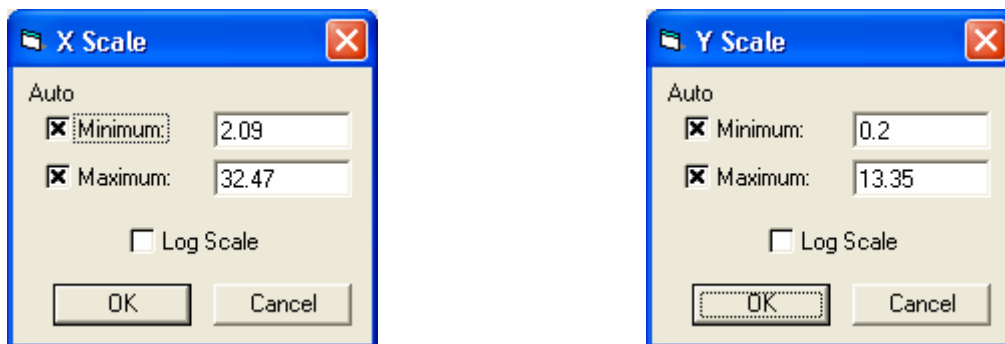


Figure 4-7. Scatter plot x-scale and y-scale selection pop up windows.

An example scatter plot is shown in **Figure 4-8**. The plots show a one-to-one line; this may be turned off by clicking on the “S” button (or selecting “Graph - Scatter”) and then the “one-to-one line” option. The data are shown as “y-axis/x-axis” in the legend. Multiple sites and species may be plotted. Click the “S” button (or select “Graph - Scatter”) to change species, x-axis, y-axis, or time span. When you click on a point on the plot, the site, date and time, and species names and concentrations are shown in the status bar at the bottom of the screen.

4.1.3 Fingerprint Plots

Fingerprint plots show the concentration of each species in a sample (in chromatographic order for PAMS) and help to identify unique characteristics of the samples. Fingerprint plots may be prepared by pressing the “F” button on the toolbar (**Figure 4-9**) or by selecting “Graph - Fingerprint - New”. If more than one file is open, a window pops up to show the available sites (see Figure 4-2). After selection of one or more sites, a window shows the scale for the value axis (y-axis); these values may be changed or left in auto-scale mode (see Figure 4-3). The legend contains the site names when more than one site is plotted. The next window provides the time span (see Figure 4-4). The time span default is to include all the data. However, the user may change this to include any “time window” of data. Click the “F” button (or select “Graph - Fingerprint”) to change sites, value axis (y-axis), or time span.

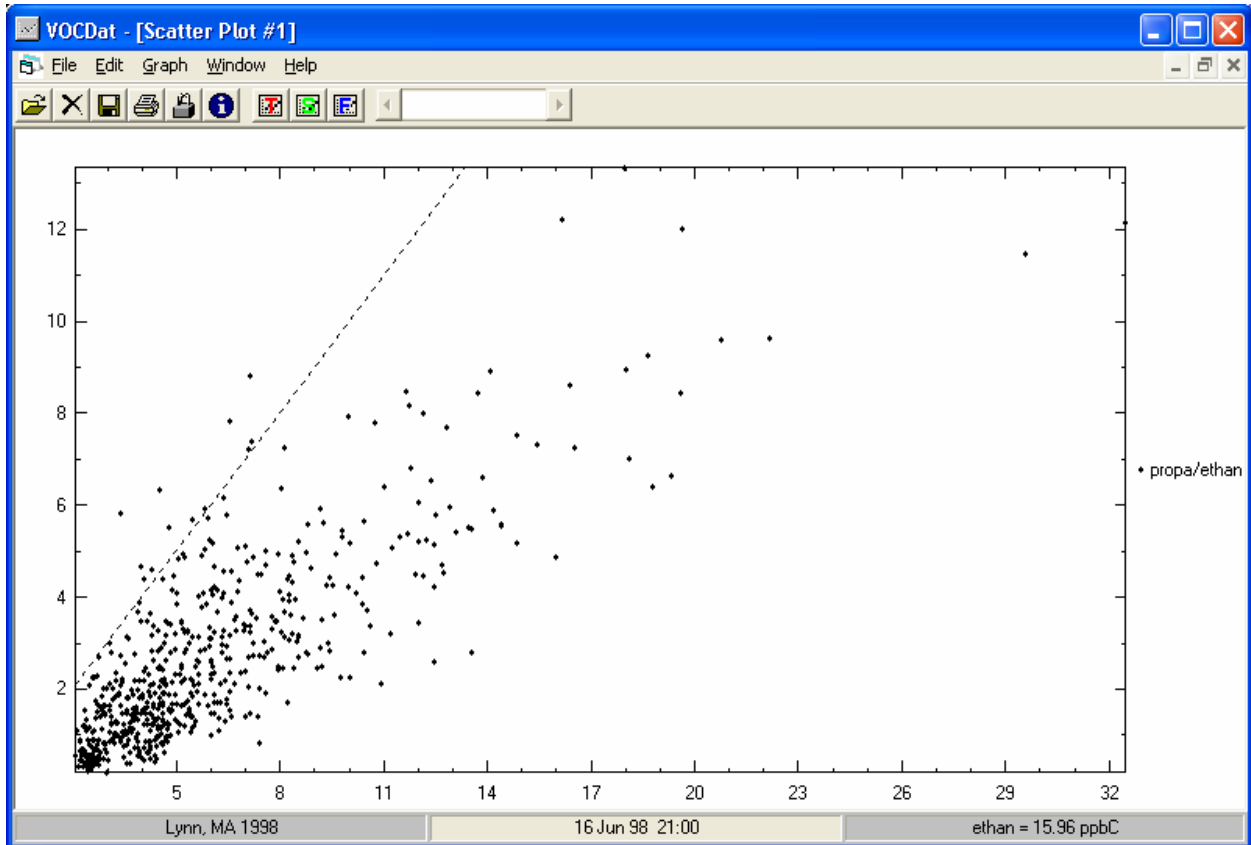


Figure 4-8. Example scatter plot of propane (propa) versus ethane (ethan).



Figure 4-9. Fingerprint button on the toolbar.

An example fingerprint plot is provided in **Figure 4-10**. When you click on a point on the plot, the site, date and time, and species names and concentrations are shown in the status bar at the bottom of the screen. The QC code of individual species and entire sample records may be edited using fingerprint plots. Note that only the species (listed in the species file in use) are plotted in fingerprint plots (i.e., calculated sums and user-defined fields are not plotted).

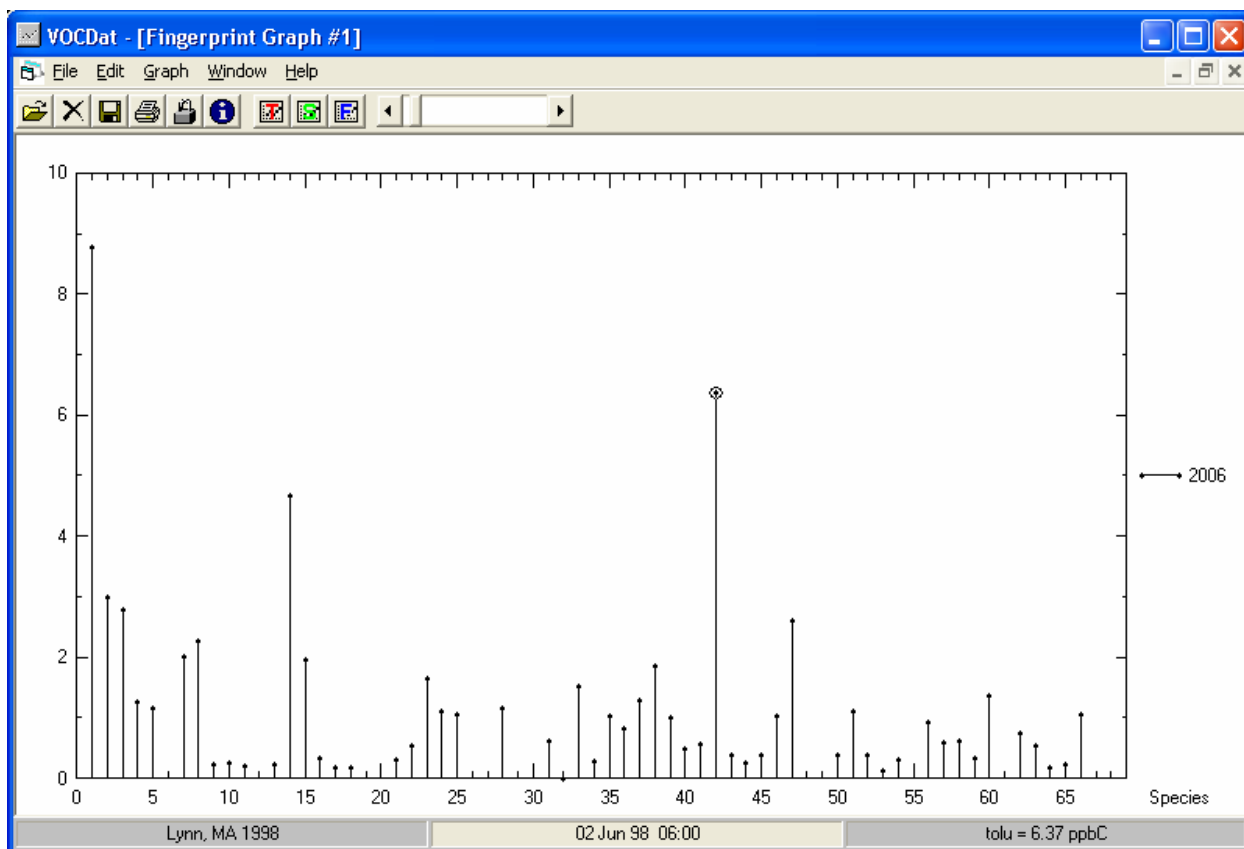


Figure 4-10. Example fingerprint plot with highlighted data point information shown in the status bar.

In the fingerprint plots, the species are plotted by number rather than name due to space considerations. Click on the data points to see the species name, concentration, date, and time of the sample. Also, a list of the species by species number may be reviewed by choosing "Window - Show species list". This list can be printed for easy reference.

4.1.4 Moving Around in the Plots

To move ahead or back in time within a time series or fingerprint plot, use the scroll bar provided at the top (**Figure 4-11**). Movement is made in the following ways:

1. Click on the left (back) or right (ahead) arrows to move one hour or sample interval at a time.
2. Click on the space to the left or right of the square located in the bar between the two arrows to move the data displayed ahead or back by the increment selected in the time axis (typically 7 days for the time series plots and about 24 hours for the fingerprints).
3. Drag the square located in the bar between the two arrows to move the data displayed ahead or back the desired amount.

4. For time series only, use the keyboard arrow keys: right and left arrow keys move the data display one graph width (default is 7 days) in time; and the up and down arrow keys change the species plotted to the next group in the species list. If three species are plotted, the down arrow key will select the next three species to be plotted over the same time period. This feature allows the user to quickly scroll through all the species for a set time period.



Figure 4-11. Scroll bar.

Portions of the time series and scatter graphs may be expanded by using the mouse to drag a box around the desired data points (using the left button of the mouse). This will create a graph showing the boxed area only. The graph may be reset to the previous scales by clicking the right button of the mouse anywhere on the graph. Multiple layers of zooming in and out on the data plots are now supported.

4.1.5 Saving Graph Configurations

All configurations, such as the screening criteria, graph colors, and the positions of graphs in your window, can be saved for future sessions by choosing "Windows - Save Configuration". You will then be prompted for a filename. A unique file name may be given to any number of configurations. Custom configurations can be restored in a future session (for different sets of graphs if desired) by choosing "Windows - Restore Configuration". An example file is provided in Appendix C.

4.1.6 Printing

The current highlighted graph can be printed simply by choosing "File - Print" or the Print button. The Windows default printer orientation (i.e., landscape or portrait), the number of copies, etc., can be selected before printing. Note that any printer configuration changes you make here are persistent and so can affect printed output from other programs. Note that the printed files are relatively crude and there is little control over options; we recommend using the print screen feature (e.g., using the keys Alt PrintScreen) to save a copy of the current window to the clipboard and pasting the image to a document.

4.2 QUALITY CHECKING DATA

One of the primary uses of VOCDat is in validating the data. "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” (U.S. Environmental Protection Agency, 1984, Sec. 2.0.3, p. 10). This section discusses how to identify potentially suspect or invalid data, flag data values, and change the QC codes associated with the values.

4.2.1 Performing Screening Checks

Through discussions with VOCDat users, it became clear that guidance was needed on the determination of which data should be flagged. Comparisons and statistical checks are useful in the determination of outliers. VOCDat provides modules to assist in this area. In consultation with other users, a few simple QC checks were prepared that can be applied to the hydrocarbon and carbonyl compound data. QC checks for PM and air toxics data need further development and implementation into VOCDat.

There are three screening tests currently available in VOCDat: checks of the abundant species concentrations, comparison of concentrations, and the variability in concentrations (**Figure 4-12**). A module for the computation of summary statistics is also available. Most screens and the statistics can be applied to the concentration, weight percent, or factor-weighted data. The output file will show a “ppbC”, “wt%”, or “factor” at the top of the file to indicate the format of the data.

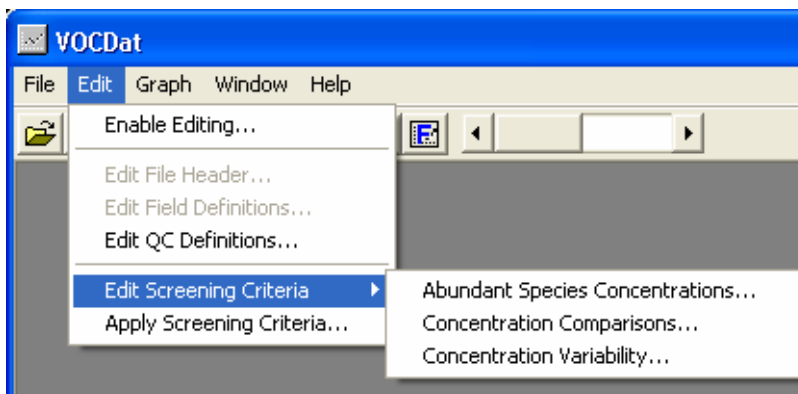


Figure 4-12. Screening criteria available in VOCDat.

Abundant Species Concentrations

At urban sites across the United States (as opposed to remote sites), there is a common set of hydrocarbon species that are typically abundant including acetylene, ethane, propane, n-butane, i-pentane, n-pentane, n-hexane, benzene, toluene, and the xylenes. Experience shows that if most of these species are present at relatively high concentrations, say above 1 ppbC, then all of these species should probably be present above the detection limit. VOCDat provides a way to test this:

1. Open a data file.
2. Select “Edit - Edit Screening Criteria - Abundant Species Concentrations” and the window shown in **Figure 4-13** appears.
3. In this window, the user can select the species and the cut-off concentrations as well as the number of species over which to apply the screening.
4. When the screening is executed by selecting “Edit - Apply Screening Criteria” or using the “Apply Screening Criteria” button, VOCDat checks each sample record for data that do not fit the specified criteria.

Item	Species	Comparison	Value
<input checked="" type="checkbox"/> 1	ethan	value is less than	0.5
<input checked="" type="checkbox"/> 2	propa	value is less than	0.5
<input checked="" type="checkbox"/> 3	isbta	value is less than	0.5
<input checked="" type="checkbox"/> 4	ispna	value is less than	0.5
<input checked="" type="checkbox"/> 5	npnta	value is less than	0.5
<input checked="" type="checkbox"/> 6	benz	value is less than	0.5
<input checked="" type="checkbox"/> 7	tolu	value is less than	0.5
<input checked="" type="checkbox"/> 8	m/pxy	value is less than	0.5
<input checked="" type="checkbox"/> 9	oxyl	value is less than	0.5
<input type="checkbox"/> 10	benz	value is less than	0.5

Buttons: Concentration Comparisons, Apply Screening Criteria, Concentration Variability, OK, Cancel

Figure 4-13. Example window for selecting the screening criteria for abundant species concentrations.

Users can start with default criteria and then as more experience with their own data is gained, the criteria can be customized to better represent the site under investigation. Screening concentrations should be set low enough to limit the potential number of data failing the screening criteria but high enough to be meaningful (e.g., 10 times the detection limit). The species list should reflect the most abundant species (or most problematic species!) at the site and thus, should be customized by the user.

Concentration Comparisons

In addition to commonly present species, there are also relationships among the hydrocarbons that are apparent at many sites. VOCDat provides a check of several expected relationships. (These checks are currently tailored to PAMS data but can be developed and implemented for air toxics and PM data.) For example, all three xylene isomers (ortho, meta, and para) tend to be present in about equal concentrations at ambient sites in and near urban areas. Since the m- and p-xylenes typically coelute in most GC systems, the concentration of the sum of these two species should exceed the concentration of o-xylene. Thus, one check of the data could be to see if o-xylene concentrations are greater than the sum of m- and p-xylenes. (Of course, there are always exceptions to this as there are sources of o-xylene independent of the other two isomers.) Under “Edit - Edit Screening Criteria - Concentration Comparisons” the window shown in **Figure 4-14** is provided. In this screen, the user can select the species to compare. In addition to a one-on-one check of species concentrations, the user can also check for species concentrations above a cut-off value or above a certain weight percent. When the screening is executed by selecting “Edit - Apply Screening Criteria”, VOCDat checks each sample record for data that do not fit the specified criteria.

Concentration Comparisons

Sample fails screening if:

<input checked="" type="checkbox"/>	1.	benz	value is greater than	tolu	value
<input checked="" type="checkbox"/>	2.	ethan	value is greater than	propa	value
<input checked="" type="checkbox"/>	3.	npnta	value is greater than	ispna	value
<input checked="" type="checkbox"/>	4.	oxyl	value is greater than	m/pxy	value
<input type="checkbox"/>	5.	benz	value is greater than		value
<input type="checkbox"/>	6.		value is greater than		value
<input type="checkbox"/>	7.		value is greater than		value
<input checked="" type="checkbox"/>	8.	tnmoc	value is less than	1000	ppbC and
		uidvoc	value is greater than	200	
<input checked="" type="checkbox"/>	9.	ethan	value is less than	2	ppbC and
		benz	value is greater than	2	
<input checked="" type="checkbox"/>	10.	uidvoc	fraction of	tnmoc	
			is greater than	50	percent

Abundant Species Concentrations Apply Screening Criteria

Concentration Variability OK Cancel

Figure 4-14. Example window for selecting the screening criteria for species concentration comparisons.

As with the other criteria, users can start with default criteria and then as more experience with their own data is gained, the criteria can be customized to better represent the site. Again, the species list should reflect the most abundant species (or most problematic species!) at the site. In the case of the weight percent cut-offs, some sites experience very high percentages of unidentified hydrocarbon (e.g., more than 70%) while other sites experience much lower (e.g., about 20%) contributions. The check of ethane concentrations less than 2 ppbC when benzene concentrations are greater than 2 ppbC derives from incidences using auto-GCs in which the cold trap failed causing low or zero concentrations of the C2 hydrocarbons while the C5+ species concentrations were normal. Ethane concentrations have been found at most sites to be above about 2 ppbC all the time.

Concentration Variability

A third set of quality checks on the data includes a check of the sample concentrations that lie outside the majority of the sample population. It is useful to determine a list of “outliers” using simple statistics in order to provide a check independent of the graphical checks on the data. For this check, the abundant species (or species groups, total NMOC) concentrations can be compared to the overall sample population using the standard deviation.

Under “Edit - Edit Screening Criteria - Concentration Variability” the window shown in **Figure 4-15** is provided. In this screen, the user can select the species to compare. When the screening is executed by selecting “Edit - Apply Screening Criteria”, VOCDat checks each sample record for data that do not fit the specified criteria. As with the other criteria, users can start with default criteria and then as more experience with their own data is gained, the criteria can be customized to better represent the site. The species list should reflect the most abundant species (or most problematic species!) at the site.

Screening Output

When the screening criteria are applied, the window shown in **Figure 4-16** is provided. The user may select the date range and QC codes over which the screening criteria are applied. Only vertical format for the output is allowed. Note that all open files are screened. Depending upon the size of the output, either Notepad or WordPad will open to display the results. From this window, the results may be edited, printed, and/or saved for future use.

Figure 4-17 shows example output for the abundant species concentration screening criteria applied to hydrocarbon data. First, the output shows the name of the data file that was screened (e.g., Lmajul94.voc), the sitename, units, time window selected, and the QC's included. Next, the criterion is listed followed by a list of samples (identified by date, time, and species concentration) that failed the screen. Finally, the total number of samples failing the particular criterion is provided. In the example provided, the screening criteria concentrations should probably be lower for a few of the abundant species because of the large number of “failures” (e.g., 124 acetylene values were below 1 ppbC). The user should check to see whether or not other abundant species were in low concentration for the two samples with low toluene concentrations. This check can be useful for finding peak misidentification or other problems for major species.

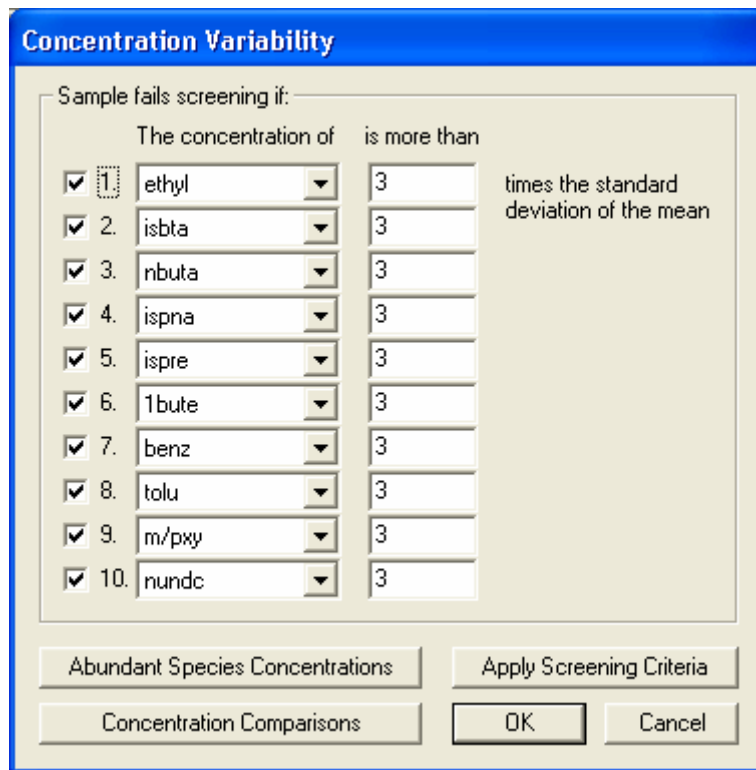


Figure 4-15. Example window for selecting the screening criteria for species concentration variability.

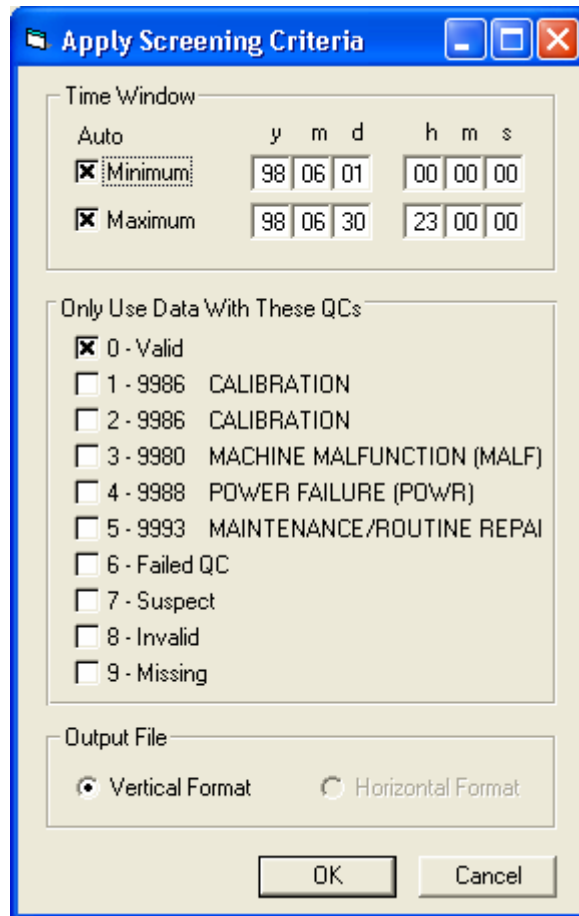


Figure 4-16. Window for applying the screening criteria.

```

Lmaaug94.voc - (ppbC)

Time Window: 8/1/94 to 8/31/94 11:00:00 PM
QC Codes Included:
  0 - Valid

acety concentration less than 1
  05 Aug 94 19:00 0.98
  05 Aug 94 20:00 0.87...
  124 samples failed out of 631 tested

ethan concentration less than 1
  0 samples failed out of 631 tested

propa concentration less than 1
  06 Aug 94 09:00 0.91
  06 Aug 94 10:00 0.82...
  29 samples failed out of 631 tested

nbuta concentration less than 1
  05 Aug 94 19:00 0.68
  05 Aug 94 21:00 0.84...
  72 samples failed out of 631 tested

npnta concentration less than 1
  05 Aug 94 19:00 0.94
  05 Aug 94 22:00 0.84...
  93 samples failed out of 631 tested

nhexa concentration less than 1
  02 Aug 94 00:00 0.90
  02 Aug 94 01:00 0.76...
  214 samples failed out of 631 tested

benz concentration less than 1
  05 Aug 94 22:00 0.81
  05 Aug 94 23:00 0.80...
  42 samples failed out of 631 tested

```

Figure 4-17. Excerpt from output of an example abundant species concentration screening check.

Figure 4-18 shows example output for the concentration comparison screening criteria. Each criterion, a list of samples that failed the screen, and the total number of samples failing are provided. In the example provided, one sample had a measurable concentration of o-xylene but was missing m- and p-xylene data (signified by -999 in the concentration field); this sample should be investigated. The check of ethane and benzene concentrations shows that there were probably no cold trap failures in the data set (e.g., missing ethane when benzene concentrations are well above the detection limit). This check also showed that only one sample had an unidentified fraction greater than 50 % of the NMHC.

Lmaaug94.voc - (ppbC)

Time Window: 8/1/94 to 8/31/94 11:00:00 PM

QC Codes Included:

0 - Valid

ethyl concentration greater than ethan concentration

06 Aug 94 19:00 4.41 4.03

06 Aug 94 21:00 5.71 4.25

22 samples failed out of 631 tested

prpyl concentration greater than propa concentration

0 samples failed out of 631 tested

npnta concentration greater than ispna concentration

03 Aug 94 06:00 25.80 22.00

05 Aug 94 21:00 3.04 1.93

20 Aug 94 06:00 49.40 -999.00

22 Aug 94 21:00 4.08 3.03

9 samples failed out of 631 tested

oxyl concentration greater than m/pxy concentration

01 Aug 94 09:00 1.31 -999.00

01 Aug 94 11:00 1.15 -999.00

10 Aug 94 16:00 0.83 0.72

3 samples failed out of 631 tested

TNMOC concentration less than 1000 and uidvoc concentration greater than 200

0 samples failed out of 631 tested

ethan concentration less than 2 and benz concentration greater than 2

06 Aug 94 12:00 1.96 3.57

1 samples failed out of 631 tested

uidvoc fraction of tnmoc is greater than 50%

06 Aug 94 02:00 23.17 41.60

30 Aug 94 12:00 58.91 97.00

2 samples failed out of 631 tested

Figure 4-18. Example output from the screening test for concentration comparisons.

The results from the third screening tool are shown in **Figure 4-19**. In this test, each species concentration was compared to the mean for all the data in the open file and species concentrations that were greater than two standard deviations from the mean were identified. For this example, 20 to 30 samples were identified as “outliers”. These samples should be further investigated in time series and fingerprint plots. In ambient hydrocarbon data, concentrations do fluctuate and simply being an “outlier” is not sufficient reason for invalidation.

```

Lmaaug94.voc - (ppbC)

Time Window: 8/1/94 to 8/31/94 11:00:00 PM
QC Codes Included:
    0 - Valid

acety concentration more than 2 standard deviations from the mean
    01 Aug 94 10:00 4.85
    02 Aug 94 20:00 9.11
    23 samples failed out of 631 tested

ethan concentration more than 2 standard deviations from the mean
    01 Aug 94 10:00 11.50
    19 samples failed out of 631 tested

propa concentration more than 2 standard deviations from the mean
    01 Aug 94 03:00 11.00
    02 Aug 94 22:00 12.00
    20 samples failed out of 631 tested

nbuta concentration more than 2 standard deviations from the mean
    01 Aug 94 21:00 18.20
    02 Aug 94 22:00 28.20
    16 samples failed out of 631 tested

ispna concentration more than 2 standard deviations from the mean
    01 Aug 94 10:00 30.40
    01 Aug 94 21:00 31.10
    21 samples failed out of 631 tested

npnta concentration more than 2 standard deviations from the mean
    02 Aug 94 20:00 13.50
    02 Aug 94 22:00 18.50
    17 samples failed out of 631 tested

benz concentration more than 2 standard deviations from the mean
    02 Aug 94 20:00 10.60
    02 Aug 94 21:00 7.73
    26 samples failed out of 631 tested

tolu concentration more than 2 standard deviations from the mean
    02 Aug 94 20:00 32.50
    02 Aug 94 21:00 26.80
    16 samples failed out of 631 tested

```

Figure 4-19. Example output from the concentration comparison screening check.

Using the Screening Output

Once the screening tests have been run, the analyst can examine the output for unusual samples, sampling periods, or other patterns. VOCDat does not automatically flag data, rather the screening output provides data that the analyst can use to perform further investigation of troublesome samples including inspection of fingerprint plots, scatter plots, and time series.

4.2.2 Calculating Summary Statistics

Summary statistics of a data set, including the minimum, maximum, median, mean, and standard deviation, are useful for assessing the variability in the data. While the VOCDat files can be imported to other software packages for the computation of these values, users expressed the desire to be able to perform these calculations in VOCDat. To do this, select “File - Calculate Statistics” and the window shown in **Figure 4-20** appears. In this window, the user can select the time period over which to perform the statistical calculation, the QC-coded data to include in the calculations, and the format of the statistical output.

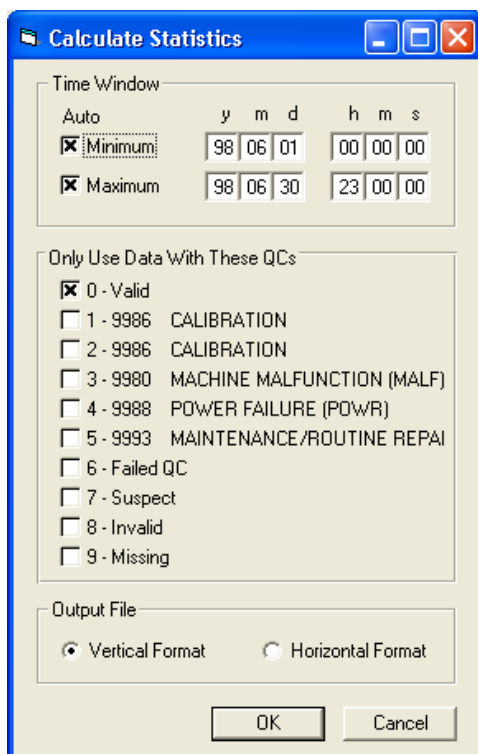


Figure 4-20. Window providing selections for data to include in statistical calculations.

Figures 4-21 and 4-22 provide example output in the vertical and horizontal formats, respectively. The output files include the file name, units of the file (e.g., ppbC for concentration file, wt% for the weight percent file, and MIR for reactivity-weighted data), and the following statistics: minimum, date/time of minimum, maximum, date/time of maximum, median, mean, standard deviation, and the number of records in the calculation (population).

Lmaaug94.voc - (ppbC)

Time Window: 8/1/94 to 8/31/94 11:00:00 PM

QC Codes Included:
0 - Valid

	Min	Date/Time of Min	Max	Date/Time of Max	Median	Mean	StdDev	Population
acety	0.4	8/15/94 1:00	9.54	8/31/94 11:00	1.7	2.03	1.36	537
ethyl	0.28	8/15/94 2:00	15.4	8/31/94 11:00	2.11	2.8	2.22	537
ethan	1.96	8/6/94 12:00	26.3	8/20/94 5:00	4.13	5.16	2.83	537
prpyl	0.17	8/10/94 11:00	6.03	8/2/94 20:00	0.81	1.08	0.84	532
propa	0.46	8/15/94 16:00	21.4	8/20/94 5:00	3.19	4.07	2.85	537
isbta	0.22	8/14/94 23:00	31.5	8/20/94 6:00	1.33	1.98	2.49	535
1bute	0.05	8/30/94 3:00	2.71	8/20/94 6:00	0.22	0.3	0.24	456
nbuta	0.39	8/30/94 14:00	77.4	8/20/94 6:00	2.4	4.51	6.27	537
t2bte	0.23	8/4/94 1:00	7.05	8/20/94 6:00	0.42	0.53	0.48	534
c2bte	0.06	8/30/94 12:00	5.95	8/20/94 6:00	0.23	0.34	0.44	477
3mlbe	0.08	8/11/94 2:00	15.1	8/20/94 6:00	0.57	0.83	1.1	526
ispna	0.99	8/30/94 14:00	98.5	8/20/94 5:00	28.7	9.33	10.18	536
1pnte	0.06	8/10/94 11:00	6.8	8/20/94 6:00	0.32	0.48	0.54	480
npnta	0.35	8/14/94 23:00	49.4	8/20/94 6:00	2.24	3.56	4.17	537
ispne	0.06	8/28/94 23:00	39.1	8/2/94 12:00	1.78	3.75	5	533
t2pne	0.05	8/30/94 15:00	10.4	8/20/94 6:00	0.35	0.57	0.8	484
c2pne	0.06	8/12/94 4:00	5.63	8/20/94 6:00	0.25	0.38	0.46	408
2m2be	0.04	8/22/94 10:00	2.85	8/20/94 6:00	0.19	0.27	0.27	356
22dmb	0.08	8/30/94 14:00	5.95	8/20/94 6:00	0.64	0.88	0.73	519
cypne	0.02	8/30/94 13:00	1.45	8/20/94 6:00	0.15	0.2	0.19	133
4mlpe	0.02	8/31/94 0:00	3.33	8/20/94 6:00	0.19	0.27	0.3	366
cypna	0.06	8/29/94 22:00	3.4	8/20/94 6:00	0.31	0.41	0.35	492
23dmb	0.12	8/30/94 13:00	7.22	8/20/94 6:00	0.57	0.74	0.67	528
2mpna	0.31	8/6/94 5:00	26.6	8/20/94 6:00	1.64	2.26	2.27	536
3mpna	0.13	8/3/94 18:00	15.6	8/20/94 6:00	1.1	1.5	1.46	491

Figure 4-21. Example vertical output from statistics calculation (excerpt).

The output files are opened automatically in Notepad or WordPad depending upon file size. From this window, the results may be edited, printed, and/or saved for future use. The data are provided in a tab-delimited format that makes it easy to import the data to Excel and other software packages.

Lmaa94.voc - (ppbC)						
Time Window: 8/1/94 to 8/31/94 11:00:00 PM						
QC Codes Included:						
0 - Valid						
	acety	ethyl	ethan	prpyl	propa	
Min:		0.4	0.28	1.96	0.17	0.46
Date/Time of Min:	8/15/94 1:00	8/15/94 2:00	8/6/94 12:00	8/10/94 11:00	8/15/94 16:00	
Max:		9.54	15.4	26.3	6.03	21.4
Date/Time of Max:	8/31/94 11:00	8/31/94 11:00	8/20/94 5:00	8/2/94 20:00	8/20/94 5:00	
Median:		1.7	2.11	4.13	0.81	3.19
Mean:		2.03	2.8	5.16	1.08	4.07
Standard Deviation:		1.36	2.22	2.83	0.84	2.85
Population:		537	537	537	532	537

Figure 4-22. Example horizontal output from statistics calculation (excerpt).

4.2.3 Enabling Editing

Once suspect or invalid data have been identified, the user needs to be able to change the QC codes associated with these data. To change the QC codes of the data, select “Enable Editing” under the Edit menu (see **Figure 4-23**). No password is provided with the program; as supplied, leave the password blank and click on “OK”. A password may be changed by editing the VOCDAT.cfg file with any text editing program. (Note that the password is not case sensitive.) Two decisions need to be made when enabling editing: Do you want to append changes to a log file and do you want to be able to add log notes? The default (and our recommendation) is yes to both questions. Once editing is enabled, the QC code buttons on the right side of the toolbar will then be visible.

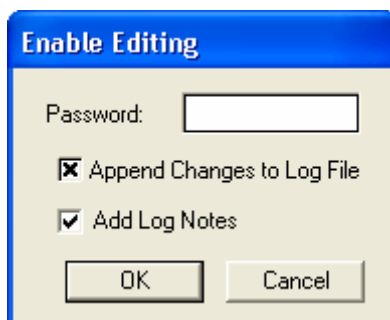


Figure 4-23. Enable editing window.

4.2.4 Changing QC Codes

Points on the graph can be selected with the mouse and the QC codes changed using the QC code buttons. The QC buttons are numbered R (for record) and 0 to 9. The QC definitions are provided in **Figure 4-24** and will appear when the cursor is held still over a numbered toolbar button. The Record (or R) button allows the user to change the QC code of an entire data record (sample) by selecting a data point on any of the three graphs, pressing Record (the button text will turn red or the R button will be depressed), and then pressing the appropriate QC button.

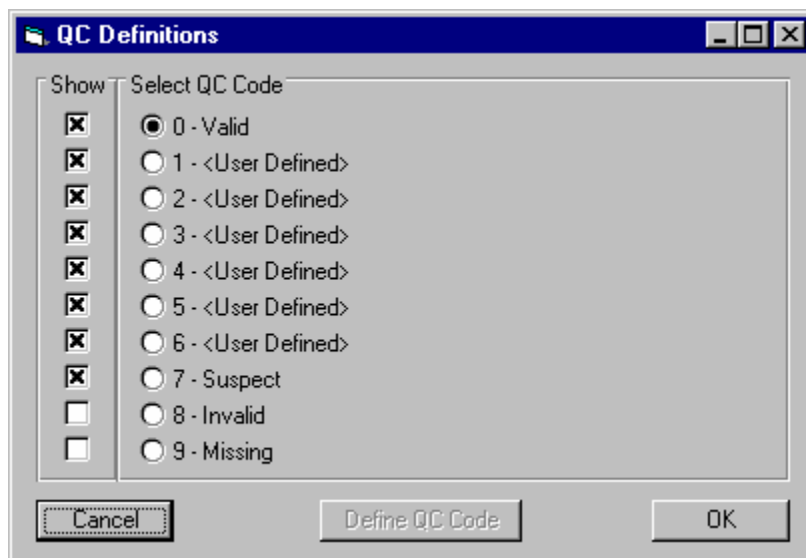


Figure 4-24. Initial configuration of the QC definitions.

Since many users output data to AIRS or AQS format, the reason for invalidating data needs to be provided. The QC codes can be customized to provide the appropriate output to AQS (see **Figure 4-25**). The null data codes are provided in NullCode.txt that may be edited in Microsoft Excel or with a text editor (that supports tab delimited files).

From the window shown in Figure 4-24, select the QC code you wish to apply to the data value. If you need to add a new definition, or alter an existing definition, select the “Define QC” button at the bottom of the window. The window shown in Figure 4-25 will be displayed. For any QC code, you may also choose whether or not you wish data with this QC code to be shown on graphs by toggling the x’ed boxes under “Show”; an x indicates the data marked with that code will show on the graphs. QC codes may also be edited via “Edit - Edit QC Definitions” on the menu. Individual species may be flagged as well as species groups or entire samples.

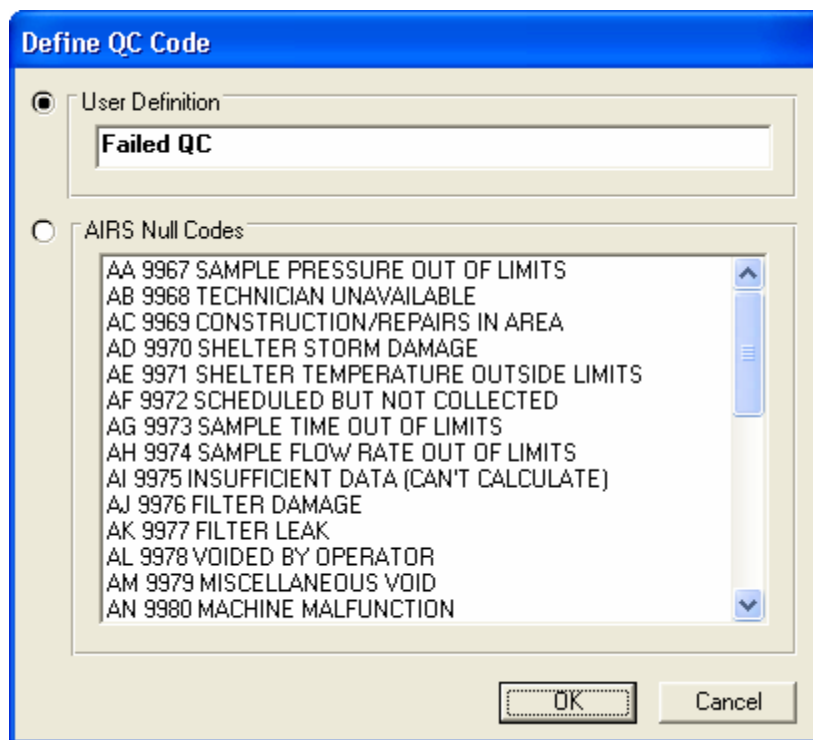


Figure 4-25. Example user definitions for QC codes 1-6. When data are exported in AIRS or AQS format, AIRS null codes are output instead of the concentrations for data with these designations.

4.2.5 Selecting Multiple Data Points

Multiple points may be selected by holding the “Ctrl” key while clicking on points. Another way to select multiple points is to hold down the “Ctrl” key while using the left mouse button to drag a box around the desired data points. (If you zoom in on data instead of selecting points, just right click and try again – remember to keep the Ctrl key depressed throughout the entire process of drawing the box.) The current limit for the number of data points in a multiple point selection is 600.

It may be useful during editing (but not necessary) to choose “Graph - Point Color = QC”. This will show all data as points with the color of the points corresponding to the QC code. Note that an invalid data point will not show on the graph after a redraw unless “Graph - Show all QC’s” is selected.

4.2.6 Changing the QC Code for a Data Point

To edit the QC code for a data point in a time series or fingerprint plot, click on the value (or several values) and select the QC code from the bar. On a scatter plot, highlight the species on the status bar which needs to be edited by clicking on the left or right hand side of the status bar (whichever side the species is on). That portion of the status bar will then change color (to yellow). Then click on the point that needs to be changed and select a QC code to apply to the

point. The user may select both species for changing QC codes by clicking on both sides of the status bar. An example of a scatter plot with one species highlighted is shown in **Figure 4-26**.

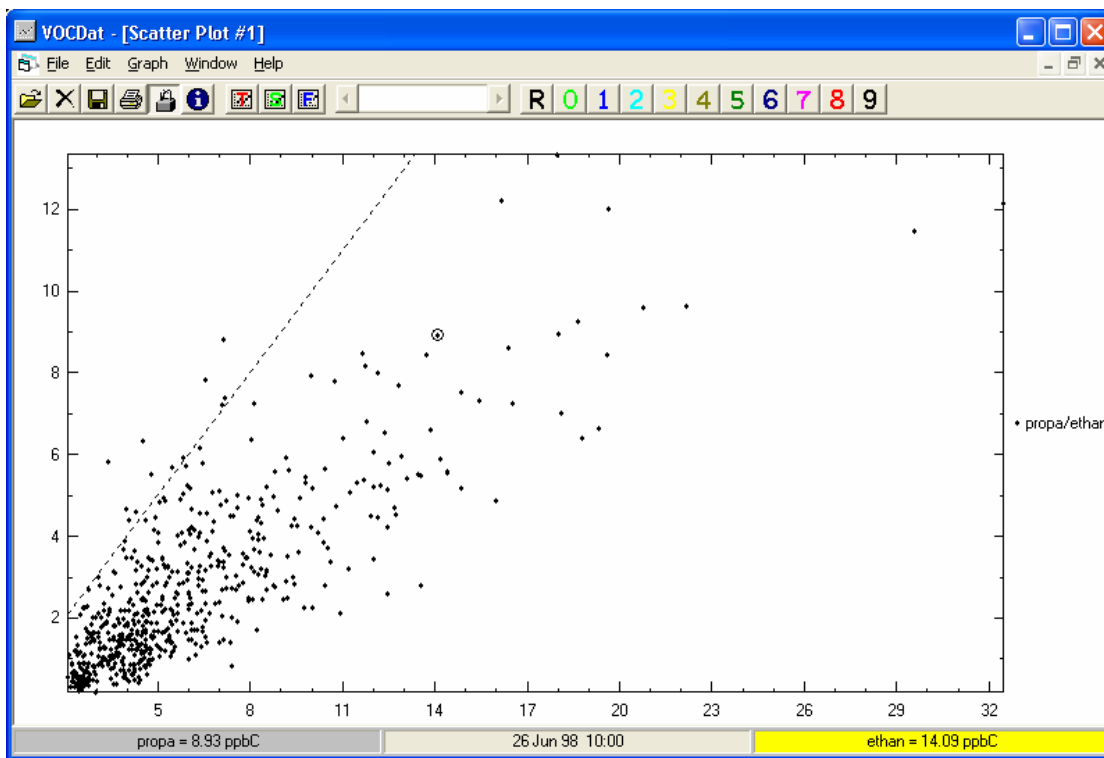


Figure 4-26. Example of a highlighted species in the status bar for a scatter plot.

4.2.7 Changing the QC Code for a Data Point

If you decided to “Add Log Notes” when editing was enabled, after you have selected the QC code, a window will appear as shown in **Figure 4-27**. You can add a comment (not required) to the log file as a reminder of why you flagged the data. If more than one data point is highlighted, the same log file note will apply to all the points, but separate entries in the log will be made. these log notes are very useful in later data analyses and in following the data validation steps that have been made.

Note that the summary totals of the organic groups for the PAMS data are not automatically updated to reflect QC changes; the user must update the QC code of the group summaries separately. This feature is user-performed because not all changes to the QC code of an individual species warrant changes to the QC code of the species group. Missing and invalid species concentrations (set to -999) are not included in summaries; if changes are made that invalidate species, group summaries need to be updated by selecting “Edit - Update Summary Fields”. Also, when the file is saved the summary fields will be updated automatically.

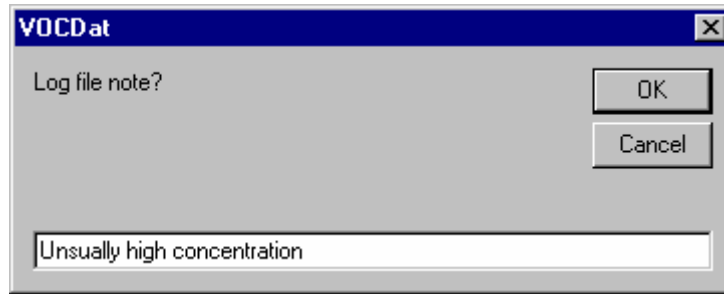


Figure 4-27. Window for adding a log file note with an example comment.

Changes to the file are recorded in a log file with the same filename as the VOCDat file but with a *.LOG extension. This file is saved when the VOCDat file is saved. If you are making frequent changes to a file, we recommend saving your file often to avoid any lost data. An example LOG file is shown in **Figure 4-28**. The files are tab-delimited.

09-25-2000	15:13:05	temp.VOC	
28 Jun 98	22:00	Record QC code set to 7	Unusually high concentration
11 Jun 98	22:00	ethyl QC code set to 7	Odd fingerprint
15 Jun 98	15:00	nbuta QC code set to 8	sample invalidated by operator
14 Jun 98	17:00	Record QC code set to 1	calibration
14 Jun 98	18:00	Record QC code set to 1	calibration
14 Jun 98	19:00	Record QC code set to 1	calibration

Figure 4-28. Example LOG file showing changes to QC codes and example comments. In this data set, the QC code of 7 = suspect, 8 = invalid, and 1 = calibration.

5. CHANGING FROM CONCENTRATION TO OTHER FORMATS

In data validation and analysis it is sometimes useful to look at data in different ways. Since concentrations may fluctuate widely in ambient data, it may be useful to normalize the data on the total concentration and inspect the relative composition. Also, some species that are not very abundant in concentration can still be important when ozone formation potential or cancer risk is considered. Thus, VOCDat provides the options to convert concentration data to weight percent or to multiply the data by a weighting factor (such as reactivity or cancer risk). These options are discussed in this section.

5.1 WEIGHT PERCENT (PAMS ONLY)

PAMS data may be converted from concentration to weight percent by selecting “File - Convert to Weight %”. The conversion is made by dividing the species' concentrations by the TNMOC and multiplying by 100. The units on graphs and in exported data indicate “wt.%” (see **Figure 5-1**). Note that all open files will be converted - thus, you will be prompted by the program before performing this step to allow you to save the files before proceeding. The scales of the graphs are automatically updated. Only individual hydrocarbon species are affected by these conversions; carbonyl compounds, TNMOC, summary fields, and user-defined fields are not changed.

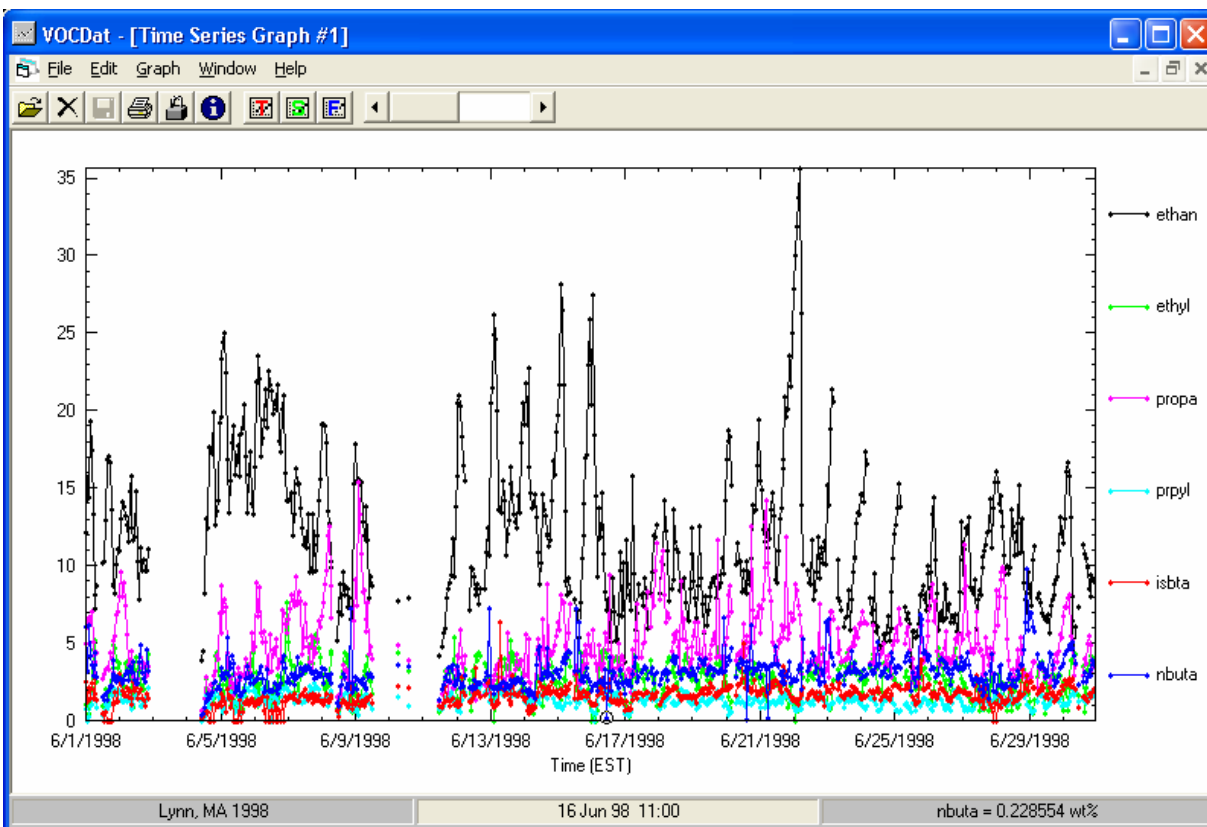


Figure 5-1. Example time series plot of weight percent data.

Once the file has been converted to weight percent, the program will not allow another conversion to weight percent and the file cannot be output to AIRS/AQS or saved as a VOC file. The data can be exported as *.TXT or *.DAT files using “File - Export - TXT” and “File – Export – DAT” options.

5.2 SCALING DATA USING FACTORS

5.2.1 Overview

Data in the open files may also be multiplied by scaling factors including ozone formation potential and cancer risk. This feature is included in VOCDat for exploratory data analysis purposes. Concentration or weight percent data may be multiplied by a factor by selecting “File – Multiply by Factor”. The units on graphs and in exported data indicate “Factor*ppbC” or “Factor*wt%” for factor-scaled concentration (for units in ppbC) or weight percent data, respectively. Note the following:

- The factors applied are those in the selected species file. It is up to the user to check the factors prior to use.
- All open files will be converted.
- The scales of the graphs are automatically updated if auto-scaling is in use.
- Factors are assumed to be zero if unknown.
- Some species that do not appear important when concentrations are considered may prove to be very important when they are scaled by reactivity, toxicity, or cancer risk.
- Only individual species are converted; TNMOC, group summaries, and user-defined fields are not changed.
- The files cannot be saved as *.VOC or to AQS/AIRS once the data have been multiplied by a factor.
- The data can be exported as *.TXT or *.DAT files using “File - Export - TXT” and “File – Export – DAT” options if the user wishes to further explore the scaled data.

5.2.2 Reactivity-weighted

The ozone formation potential of the hydrocarbons and carbonyl compounds is of interest to analysts. The degradation of VOCs by photochemistry and the resulting conversion of NO to NO₂ and formation of ozone do not occur at the same rate for all VOCs. The ozone formation potential of a specific hydrocarbon depends on its concentration, structure, and removal pathways. If a reactive compound is low in concentration, it will generally not have a high ozone formation potential while a somewhat unreactive compound with a high concentration may have a larger ozone formation potential. Two sets of reactivity scaling factors have been provided in the species files delivered with the software: hydroxyl reactivity scale and maximum incremental reactivity (MIR).

The hydroxyl reactivity scale (OH) (Atkinson, 1989, 1994) utilizes the reaction coefficient of an individual hydrocarbon with hydroxyl radical. This is strictly the rate at which the hydrocarbon is oxidized by hydroxyl radical only and does not consider competing removal mechanisms for either the VOC or hydroxyl radical or the influence from the overall composition of VOCs in an air mass. The species files that contain these factors are indicated by an “OH” in the name, such as “Species PAMS ppbC OH.txt”.

Incremental reactivity (Carter, 1994, 2001) is the change in ozone caused by adding a small amount of test VOC to the emissions in an episode, divided by the amount of test VOC added: g ozone/g C or moles ozone/moles C. Incremental reactivity may be used to assess the effect of changing emissions of a given VOC on ozone formation, to compare the ambient VOC mix among sites or episodes, or to investigate VOCs important to ozone formation. This scale considers NO_x sinks as well as the generation and loss of hydroxyl radicals, all of which affect the rate of reaction for VOCs. The MIR scale was developed by W.P.L. Carter (1994) and used in “low emission vehicles and clean fuels” regulations in California. The MIR list was recently expanded to include more VOCs and MIR values were updated (Carter, 2001). An example time series plot showing reactivity-weighted data is provided in **Figure 5-2**.

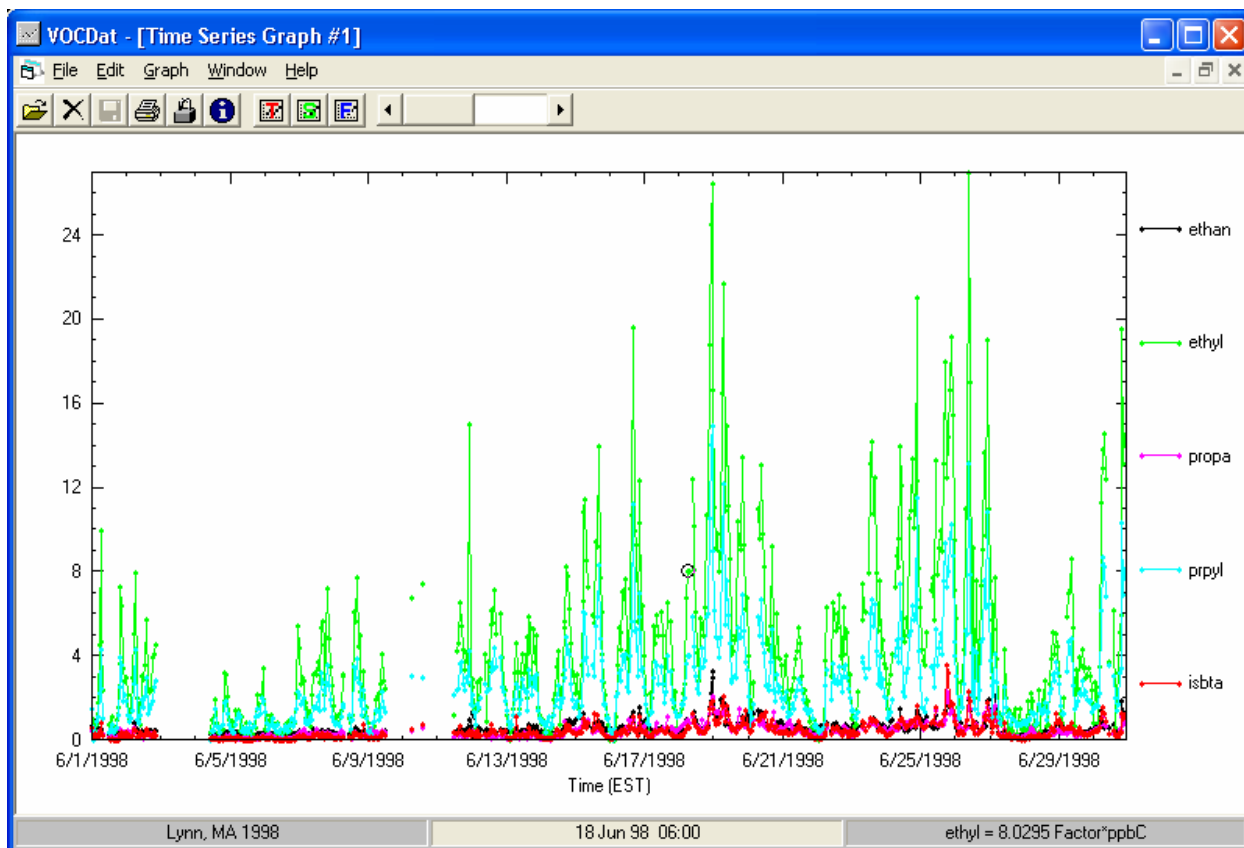


Figure 5-2. Example time series plot of concentration data multiplied by Carter MIR factors.

In assessing VOC data, analysts have found that the MIR scale is most useful in a relative manner (i.e., is an ambient sample more reactive than another?) rather than in an absolute manner (i.e., how much ozone can be generated with this air parcel?). Furthermore, the uncertainty associated with MIR scale values and the notion that total reactivity equals the sum of incremental reactivities from individual species is unverified. The analyst needs a low unidentified fraction of TNMOC to best assess the potential reactivity of a hydrocarbon mixture. If high unidentified fractions exist, this analysis is less useful. When comparing samples, the weight percent of each hydrocarbon multiplied by its reactivity is often used. Scaling by a sample's TNMOC allows for differences of the entire sample to be assessed on a relative basis, instead of on a per species basis.

There are a number of differences between the two reactivity scales. One is that carbonyl compounds are much more reactive on the MIR scale than on the OH scale. Another is that isoprene is much more reactive on the OH scale, so that even small amounts become significantly amplified. Propene is much more reactive than ethene on the OH scale, but less so on the MIR scale. Lastly, styrene is much more reactive on the OH scale than on the MIR scale, so that low amounts appear more significant on the OH scale than on the MIR scale. Values for a number of species on the OH and MIR reactivity scales are given in **Table 5-1**.

Table 5-1. Reactivity values (MIR and OH) for selected hydrocarbons.

Compound	MIR Reactivity (mol O ₃ /mol C)	OH Reactivity (rate constant with OH * 10 ¹²) (cm ³ molecule ⁻¹ s ⁻¹)
Ethene	2.65	8.5
Propene	3.38	26.3
n-butane	0.4	2.4
Trans-2-butene	4.07	64.0
Isopentane	0.51	3.7
Cis-2-pentene	2.99	67.0
m/p-xylene	2.06	23.6
Toluene	1.09	5.95
1,3,5-trimethylbenzene	3.12	57.5
Isoprene	3.03	101.0

5.2.3 Cancer Risk-weighted

One way to scale air toxics data is to multiply the data by a unit risk estimate (URE). The species files that contain "risk" in the titles indicate that URE's have been included as the scaling factor. The unit risk estimate is the upper-bound excess lifetime cancer risk estimated to result from continuous exposure to an agent at a concentration of 1 µg/m³ in air. The interpretation of the URE would be as follows: if the URE = 1.5 x 10⁻⁶ per µg/m³, 1.5 excess tumors are expected to develop per 1,000,000 people if exposed daily for a lifetime to 1 µg of the chemical in 1 cubic

meter of air. Unit risk estimates are considered upper bound estimates, meaning they represent a plausible upper limit to the true value.

The risk factors in the 2004 version of the species files have been revised to reflect the most recent values from the IRIS and CAL EPA databases. The most recent risk factors have only been updated (if necessary) for the 18 core air toxics compounds. If available, the 2003 IRIS value was used. If this value was not available, the 2002 CAL EPA value was used. If this was not available, the value was left as it was in the previous file. We recommend you check the online IRIS database at <http://www.epa.gov/iris/> and the CAL EPA web site at <http://www.arb.ca.gov/toxics/healthval/healthval.htm> for updated information prior to using these values. The risk factors should only be used in a relative, rather than quantitative, manner.

5.2.4 User-defined Factors

The user can select other factors by creating a custom species file.

5.3 CHANGING UNITS OF DISPLAYED DATA

VOCDat allows the user to change the units of the gaseous data among ppb, ppbC, and $\mu\text{g}/\text{m}^3$. Units conversion information for the PAMS and air toxics species are provided in a file called Conversions.txt provided during VOCDat installation. The file contains the AIRS parameter code, species abbreviation and name, molecular weight, and number of carbon atoms in the chemical formula for the compound. This file can be modified by the user using Microsoft Excel or a text editor that supports tab delimiters.

Gaseous data may be converted from concentration to weight percent by selecting “File - Convert Units to” where the analyst selects ppb, ppbC, or $\mu\text{g}/\text{m}^3$ (**Figure 5-3**). The units on graphs and in exported data indicate current units (see **Figure 5-4**). Note that all open files will be converted - thus, the first time you select units conversion, you will be prompted by the program before performing this step to allow you to save the files before proceeding. The scales of the graphs are automatically updated if auto-scaling is in use. Only the gaseous species in Conversions.txt are affected by these conversions; TNMOC, sum of PAMS species, summary fields, user-defined fields, or any other fields in species2.txt are not changed. If a parameter is not included in the Conversions.txt list, VOCDat will display a message indicating which species were not converted. The user can convert among all three units as many times as necessary.

For the present version of VOCDat (v2.51, September 2004), the units conversion feature is for exploratory data analysis. Therefore, once the file has been converted among units, the file cannot be output to AIRS/AQS or saved as a VOC file. However, the data can be exported as *.TXT or *.DAT files using “File - Export - TXT” and “File - Export - DAT” options for additional exploration or delivery to other analysts.

Units conversions are not allowed for files that have been converted to weight percent or multiplied by a factor.

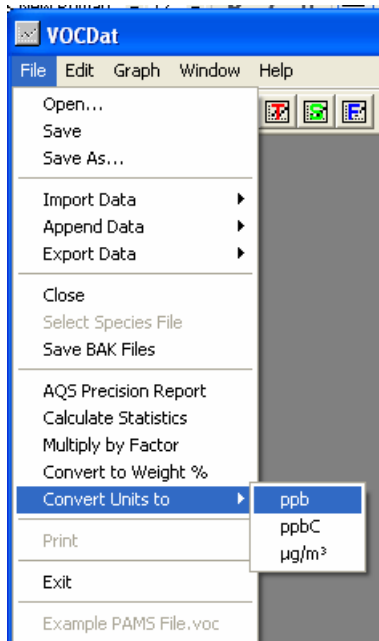


Figure 5-3. Menu selection for converting units.

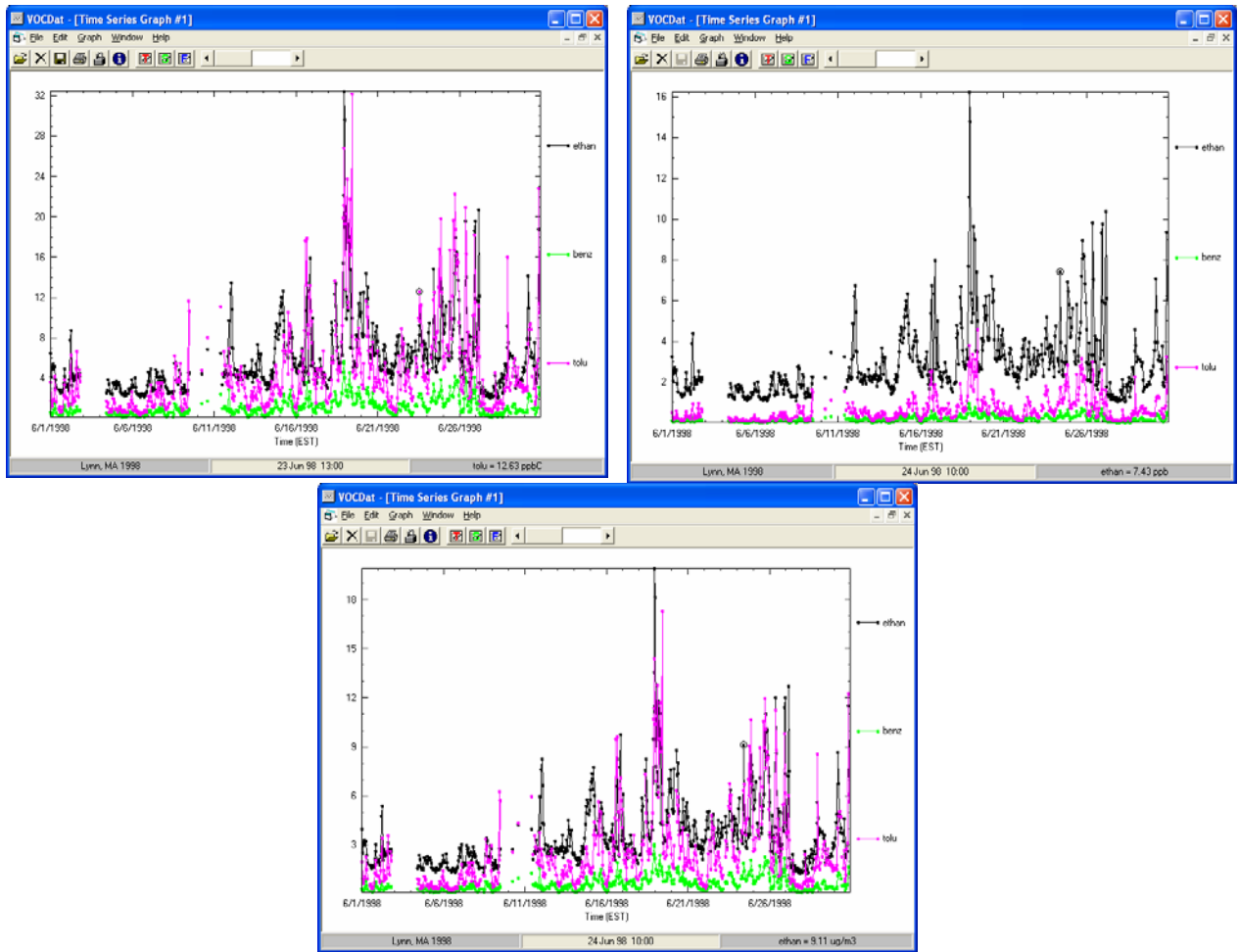


Figure 5-4. Example screen shots showing the same data in ppbC, ppb, and $\mu\text{g}/\text{m}^3$.

6. GENERAL DATA VALIDATION APPROACH AND EXAMPLES

The use of VOCDat has been demonstrated and many users have been trained through VOCDat training workshops (e.g., Hafner, 2003). Through these workshops, many examples of invalid data and validation procedures were demonstrated. This section provides general guidance for data validation as summarized in the workshops.

6.1 DATA VALIDATION OBJECTIVES

The objectives for data validation should include the following:

- Produce a database with values that are validated and of a known quality.
- Evaluate the internal, spatial, temporal, and physical consistency of the data.
- Intercompare data to identify errors, biases, or outliers.

Data that have been submitted to AQS or AIRS have been shown to still contain calibration data, data influenced by operating problems, species misidentifications, and contamination problems. Therefore, it is important to ensure that data submitted do not contain this type of invalid data. Likewise, it is important to inspect data from other sources that have been retrieved from AQS for these problems.

6.2 APPROACH

An overall approach to data validation is provided in **Table 6-1**. It is very important to proceed from the “big picture” to a closer view (e.g., proceeding from a month of data, to a week, and then to a day); this strategy is helpful in forming an overall understanding of the data.

Specifically, in Steps 2 through 7, the VOCDat user should perform the following:

1. Select an appropriate species file to match the data file.
2. Import or open the data file.
3. Create a time series plot of the first few species in the list to ensure that the data have been imported (or opened) correctly.
4. Check the species in the time series window or “show species list” for missing species (e.g., compare the species list with the PAMS list supplied by EPA for PAMS data or with the expected list of species supplied by the laboratory that supplied the data).
5. Modify the screening criteria to best fit the data and then run the auto screening checks. Print out the results and note samples that should be further investigated.
6. Plot and inspect the time series of every species and species group. This is necessary to identify outliers, calibration spikes, abrupt changes in concentrations, possible misidentification of peaks, and extended periods of unusually high or low concentrations. It is useful to plot species together that are primarily emitted by the same source (e.g., benzene and acetylene are both present in automobile exhaust). Or, plotting species

together that are emitted by different sources on a plot can also be useful (e.g., contrast isoprene with isopentane).

Table 6-1. List of steps for data validation.

Step	Description	Examples
1	Obtain and review as much supplemental information regarding data quality as possible	Sampling information (e.g., schedule, instrument type, sampling media), laboratory analysis procedures and MDLs, collocated measurements (e.g., meteorology, non-toxic species, data from other networks)
2	Import the data into software suitable for preparation of graphics and summary statistics	EPA-sponsored data visualization tools (e.g., VOCDat); statistical packages; spreadsheets
3	Apply visual/graphic methods to investigate the data	Time series, scatter, and fingerprint plots
4	Prepare and inspect summary statistics for species by site and year	Minimum, maximum, median, mean, standard deviation, and coefficient of variation.
5	Apply general screening criteria	Lower limits (i.e., background concentrations), expected ratios, species A > species B
6	Investigate internal consistency checks	Sum of identified species < total species (PM or VOC), PM _{2.5} < PM ₁₀ , cation/anion balances, correlations among species emitted by the same source
7	Investigate flagged data (i.e., data remain suspect or are deemed valid or invalid)	Review meta data including site maps, nearby emissions sources, meteorology; compare to historical data record at the site
8	Continue to investigate flagged data	Compare data to other sites and data sets
9	Assess data quality	Tabulate flagged samples/species; determine percentage of invalid or suspect samples
Convert data to AQS format. May be performed after Level 0, I, II, or III validation steps.		

7. Inspect all the fingerprints quickly and further inspect fingerprints of samples that have been flagged (i.e., identified as suspect or invalid) in time series or scatter plot analyses. Paging through the fingerprints one-by-one quickly allows the analyst to observe diurnal changes in species or species groups. The analyst should inspect hours surrounding suspect and invalid data to see if there was any “carryover” effect.

6.3 EXAMPLES OF TYPICAL PROBLEMS IDENTIFIED IN PAMS VOC DATA

This section discusses examples of data problems found during numerous investigations of PAMS and air toxics data collected throughout the country.

6.3.1 Species Misidentification

Species misidentification (i.e., the misassignment of concentrations) is often found in auto-GC data. Depending upon the concentration of the species involved and the frequency of the misidentification, data analyses could be affected. **Figure 6-1** shows a time series of hourly toluene, 2-methylpentane, and 3-methylheptane concentrations. The data show several periods for which the toluene chromatograph peak was mistakenly identified by the auto-GC system as either 2-methylheptane or 3-methylheptane. The topmost line in the time series plot is likely toluene. These samples exhibited low concentrations of toluene relative to other species. Since toluene is one of the most abundant hydrocarbons in ambient air, its low concentration was notable. Plotting species that elute near each other in the chromatograph facilitates this analysis.

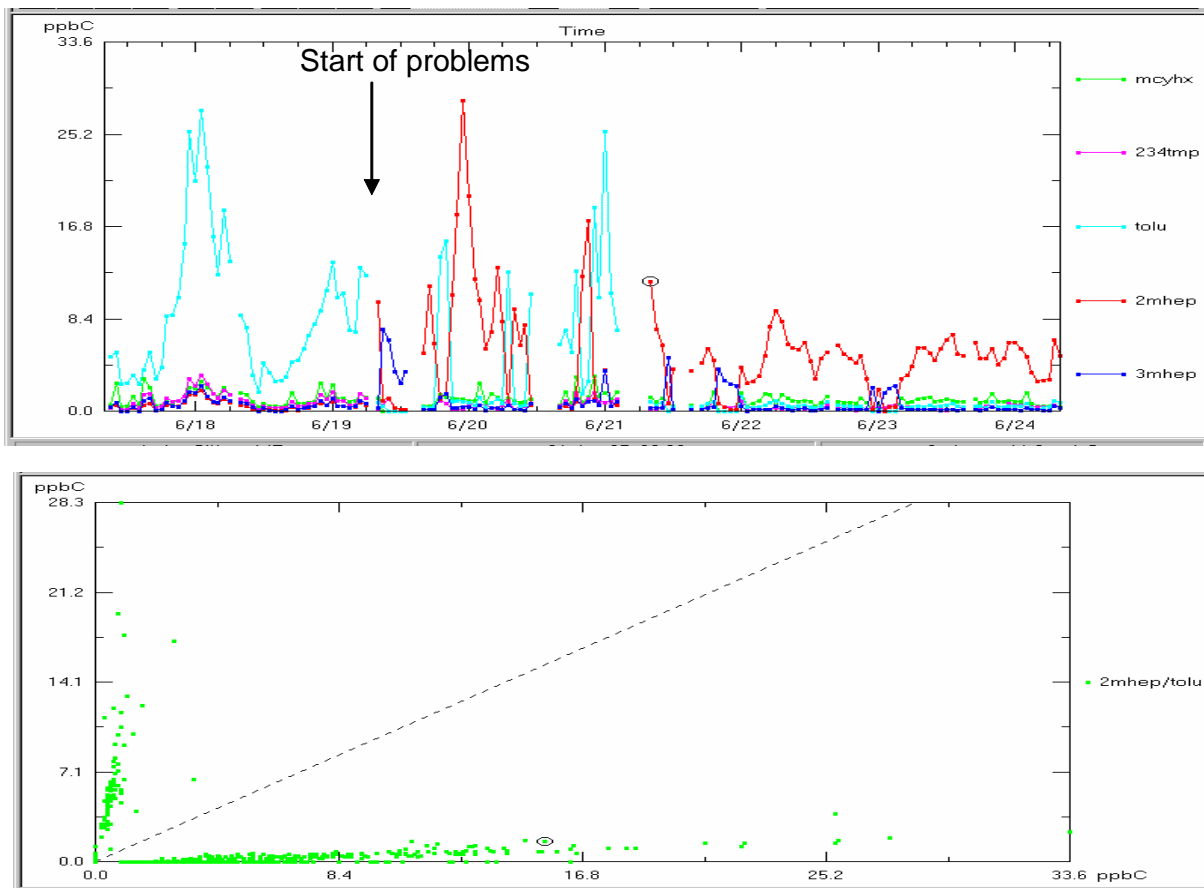


Figure 6-1. Example of misidentification of toluene, a typically abundant species, in hourly data collected at a PAMS site during June 1995 using time series and scatter plots.

In the scatter plot shown in Figure 6-1, toluene was plotted versus 2-methylheptane. The plot shows an odd bifurcation; data points would typically be present in the region of the plot between the two extreme edges. Another example of toluene misidentification is shown in **Figure 6-2**.

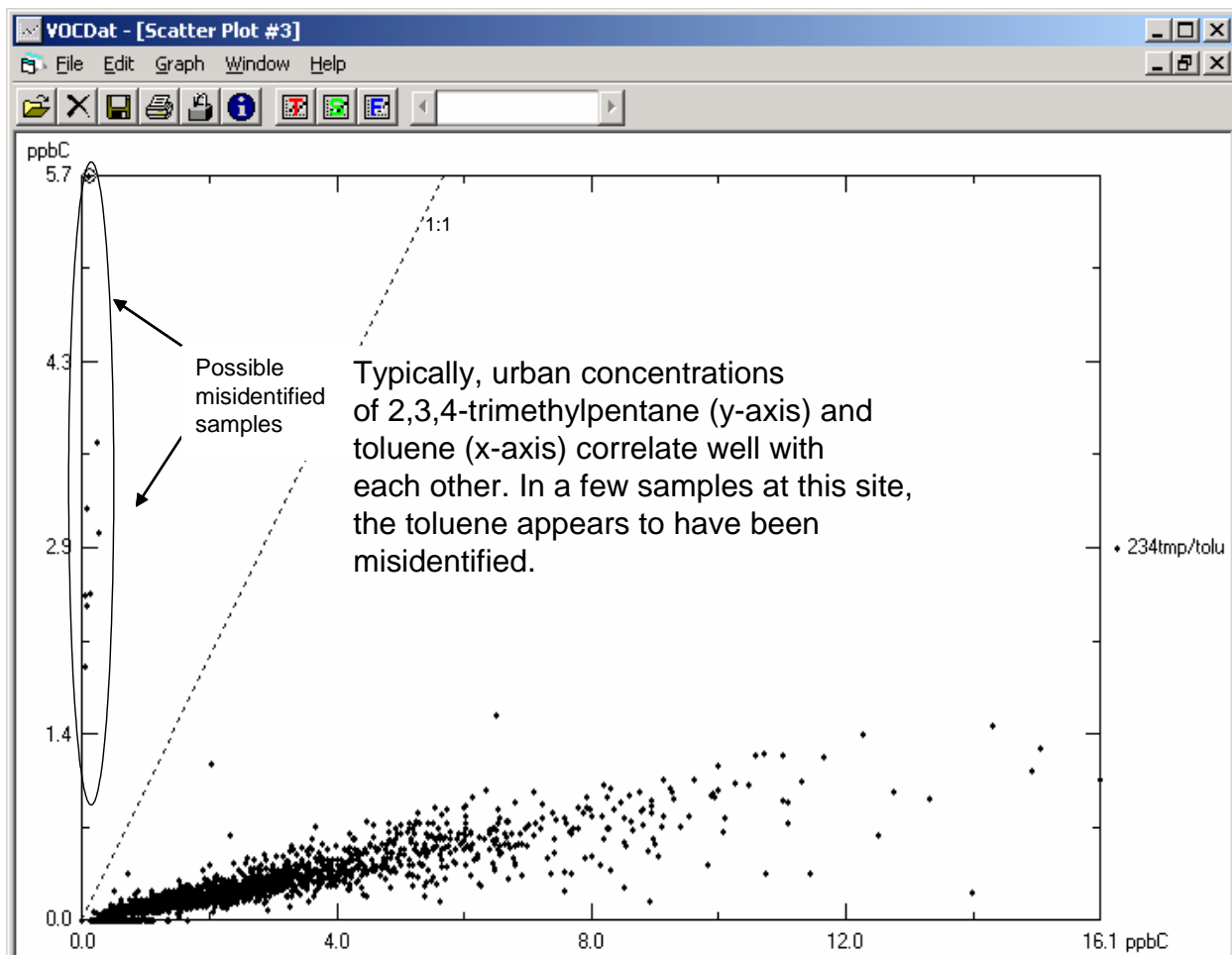


Figure 6-2. Example of potentially misidentified toluene concentrations.

Misidentification problems with 2-methylpentane and 3-methylpentane peaks, both relatively abundant hydrocarbons in ambient air, are also commonly noted in PAMS data (Figure 6-3).

In all these data sets, the “raw” data would need to be reevaluated by the reporting agency and corrected.

6.3.2 Calibration Data

Periodically during the operation of auto-GCs, calibration of the instrument is required. Typically, this involves injecting a calibration gas into the instrument during a selected hour instead of injecting an ambient sample. Calibration data should be flagged by the operating system and not reported in the ambient database; however, the calibration data have not always been removed from the database before submittal to AQS. Failure to remove these data before data analysis could result in erroneous conclusions. For example, the summary statistics for hours when calibrations are performed may be significantly different with and without the

calibration data. **Figure 6-4** shows fingerprint plots of typical and calibration gas data at a site. The calibration data should not have been reported in the ambient database and should be removed.

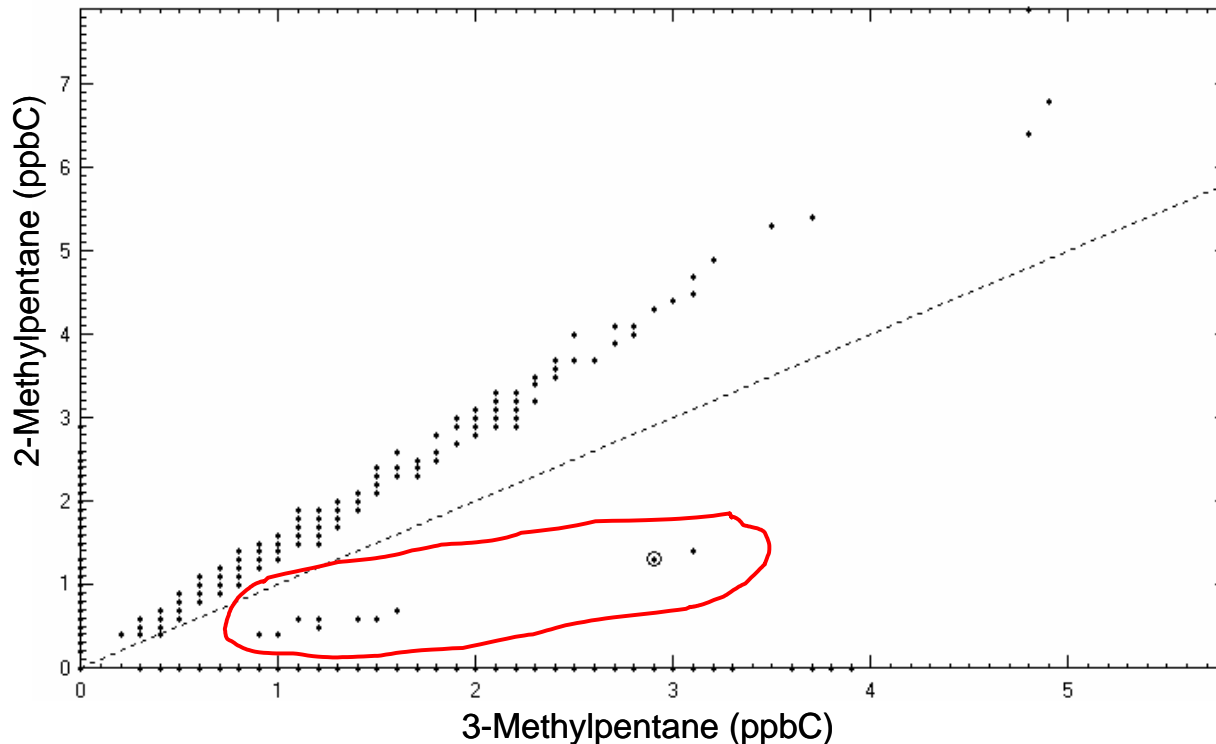


Figure 6-3. Example of possible misidentification of 2-methylpentane and 3-methylpentane concentrations in a PAMS data set.

Sometimes data collected following a calibration are affected by “carryover”. Carryover results when one (or more) compounds remain in the chromatographic system for more than one cycle. Relatively high concentrations of the heavier species after a calibration and then a gradual decrease over the next few hours are typical of carryover. **Figure 6-5** illustrates the effect of calibration gas carryover on n-decane and n-undecane; 1,2,3-trimethylbenzene, unaffected by carryover, is shown in the figure for contrast. The calibration standard readily available in the early PAMS program contained the target species at concentrations of about 30 ppbC. Since this level is typically an order of magnitude higher than mean ambient values at many sites, carryover is a valid concern. The affected species should be flagged as suspect; corrections for the carryover should be made by the reporting agency before submittal to AQS.

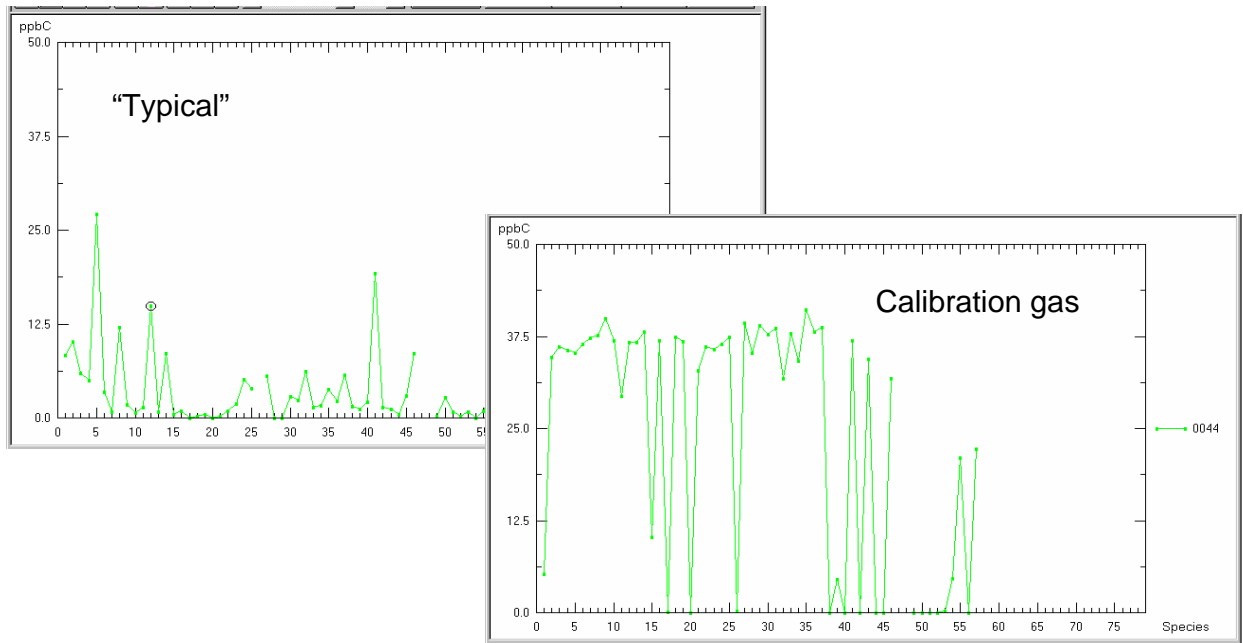


Figure 6-4. Example of typical and calibration fingerprints observed at a PAMS site.

6.3.3 Contamination

Sample contamination may occur and may not be identified immediately. For example, a new shelter installed at a PAMS site in the Northeast during 1995 was off-gassing n-pentane and i-pentane (or species identified as such) and the concentrations of these species were up to several orders of magnitude higher than typical ambient levels. **Figure 6-6** shows a scatter plot of i-pentane and n-butane at that site at start up. The n-butane concentrations were the ‘expected’ concentrations for the rural site location. After thorough detective work by the state operators and the regional EPA, the source of the high concentrations of the pentanes was found to be the blowing agent used in the manufacture of the shelter. The NMHC, paraffin, and pentane concentrations should be considered invalid. To salvage data, an analyst could use the linear relationship between n-butane and i-pentane observed in the lower portion of Figure 6-6 to estimate the ambient i-pentane concentrations. This could be repeated for n-pentane, the paraffin, and NMHC totals.

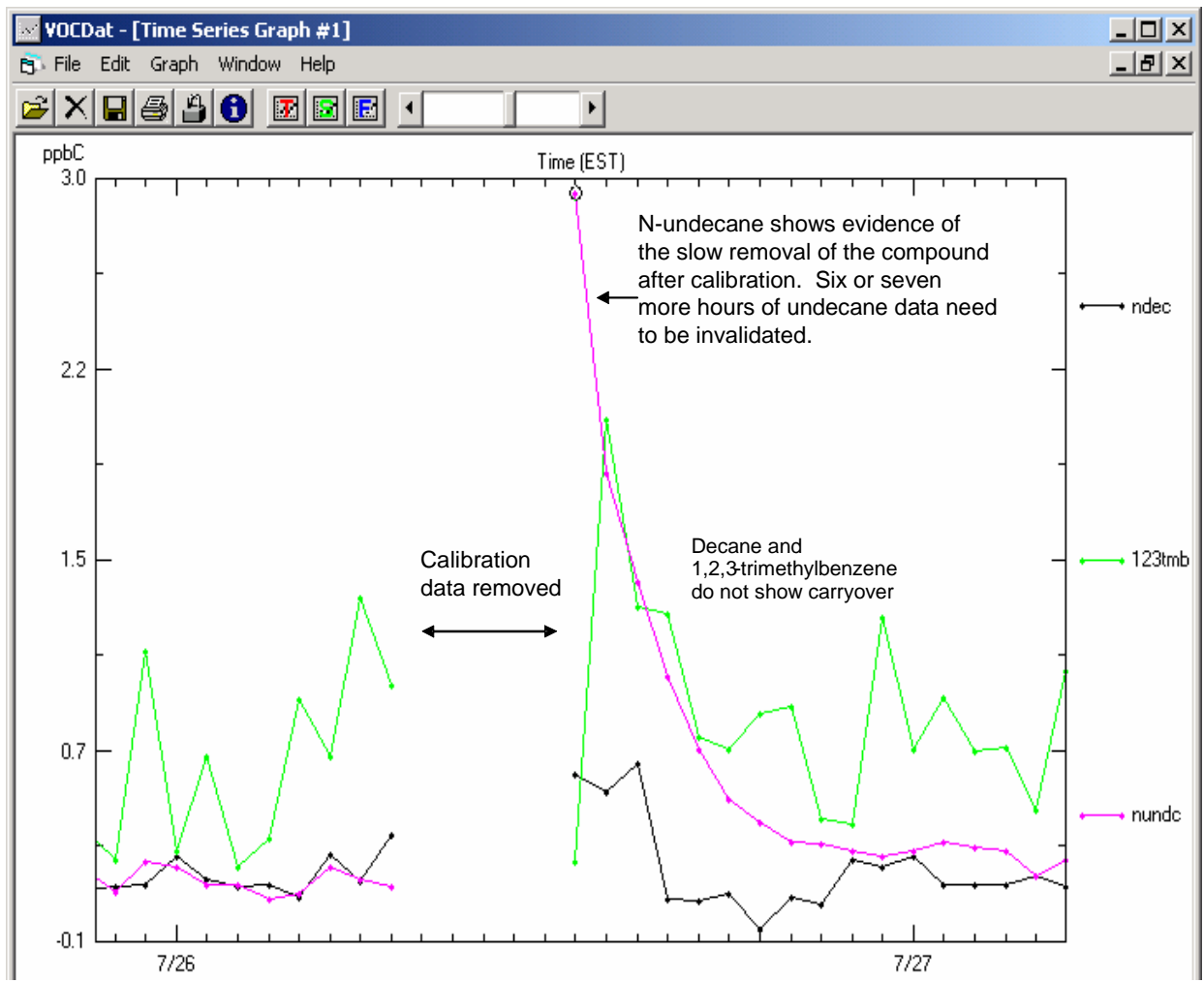


Figure 6-5. Example of calibration gas carryover and its effect on n-decane and n-undecane concentrations for several hours after the calibration.

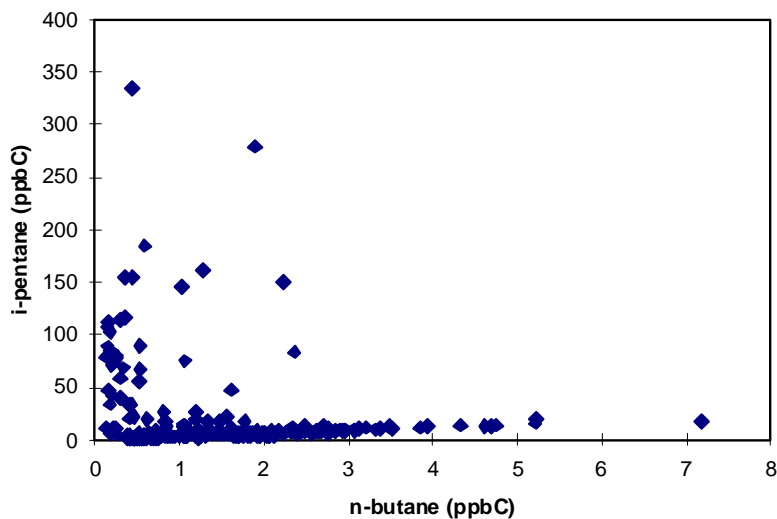


Figure 6-6. Concentrations of i-pentane and n-butane measured at a PAMS site during June 1995 showing high concentrations of i-pentane relative to n-butane. The source of the high concentrations of i-pentane (and n-pentane) was found to be the blowing agent used in the manufacture of the shelter.

Figure 6-7 shows relatively high concentrations of several hydrocarbons that exhibit a gradual decline over the first few days of measurements. In this case, it appears that contamination of either the GC column or the sampling system had occurred. The data collected during June 1 through 4 probably are not representative of ambient air and should be invalidated.

6.3.4 Miscellaneous Data Problems

Figure 6-8 shows another example of “odd” data reported to AIRS. The 1,2,4-trimethylbenzene concentrations never drop below about 4 ppbC unlike other aromatic species (e.g., toluene). This may indicate a higher detection limit for this species than for the other PAMS target species. For these data, only concentrations above the apparent detection limit should be considered in analyses. Failure to note this difference in detection limit can affect how the analyst views the relative abundance of the target species and its importance to sample composition and ozone formation potential.

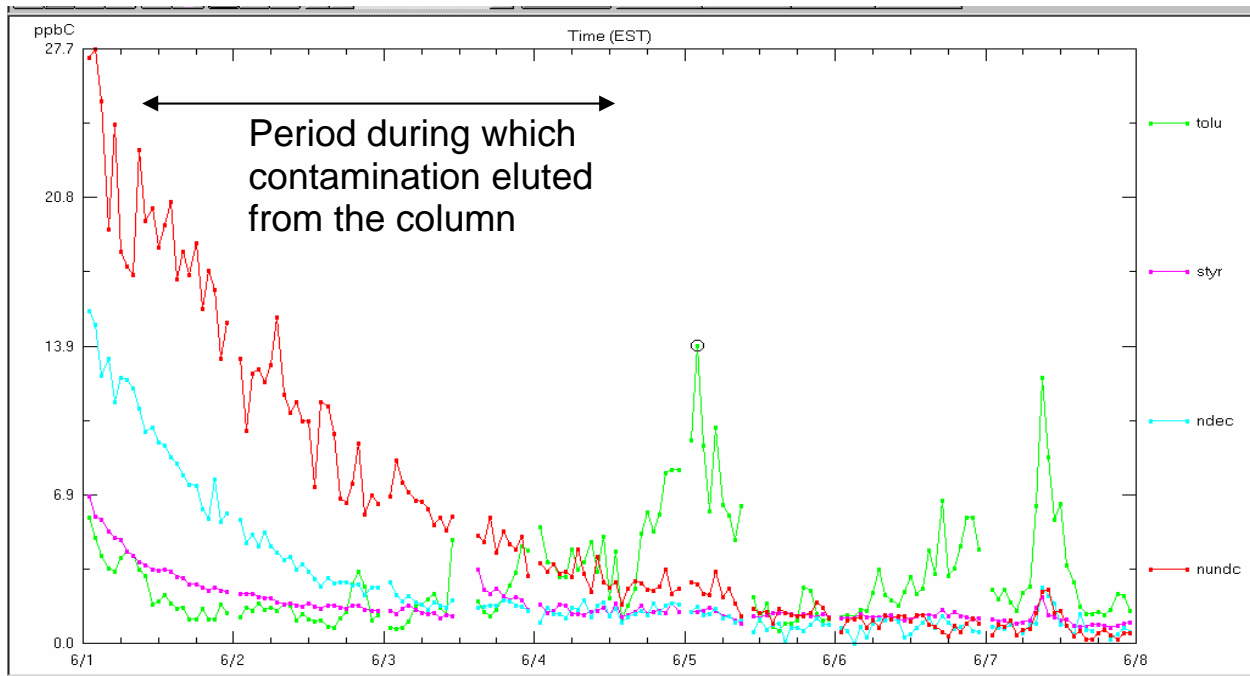


Figure 6-7. Example of possible contamination of the column in the GC at the beginning of the month involving several species at a PAMS site during June 1996. Hourly concentration data are shown for toluene (tolu), styrene (styr), n-decane (ndec), and n-undecane (nundc).

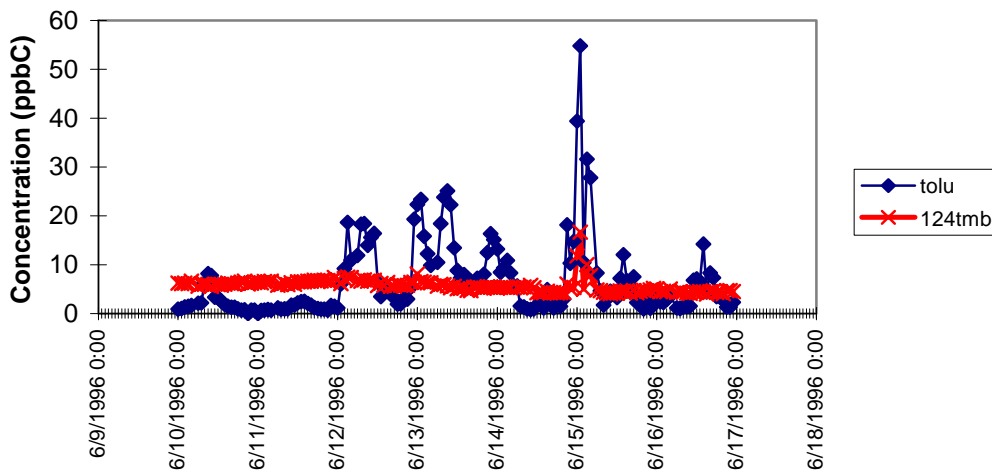


Figure 6-8. Example of high detection limit observed for 1,2,4-trimethylbenzene at a PAMS site during 1995; toluene concentrations are shown for contrast. The relative importance of this species should be evaluated with full consideration of its high detection limit compared to other species. Hourly concentration data are shown.

Figure 6-9 shows an example of an abrupt change in a few species and their relationships among one another. The p- and o-ethyltoluene concentrations were typically high together when m-ethyltoluene concentrations were reported as 0 ppbC (possible misidentification?) during July. In August, this occurrence was not noted. These data were reinvestigated by the reporting agency.

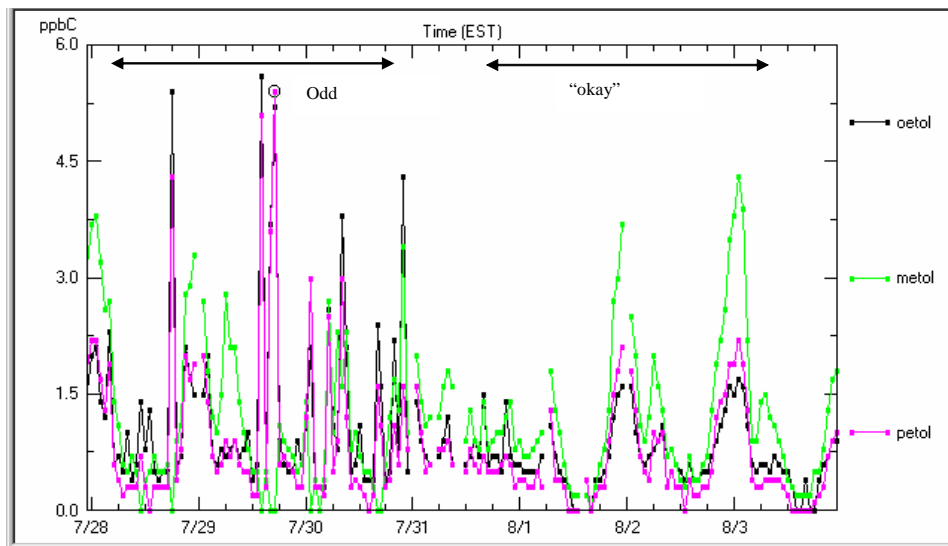


Figure 6-9. Example of an analytical system change between two months that affected the relationship among three isomers.

Figure 6-10 shows three years worth of air toxics data (24-hr samples) in which a few problems are noted. In the example, the TNMOC was not reported after June 2000 and no data were reported at all in 2001. In 2002, one sample had unusually high concentrations of unidentified mass in January with much lower values for the rest of the year. For a sample later in the year, the sum of identified species exceeded the reported total. The next step would be to see if the TNMOC data are available for 2000, if data are available for 2001, and if the two samples in 2002 with problems may be a reporting error rather than something wrong with the data.

An example comparison of concentrations of long-lived species measured at an urban site and background concentrations measured at remote northern hemisphere sites is shown in **Figure 6-11**. Significant spikes and dips in concentrations are circled in red. Most of the time, concentrations at this monitor were equal to or greater than background concentrations, as expected. Concentrations well below the background level were flagged as suspect. Concentrations well above the background may be due to local sources. **Table 6-2** lists background concentrations for some common gas-phase air toxics (McCarthy et al., 2004).

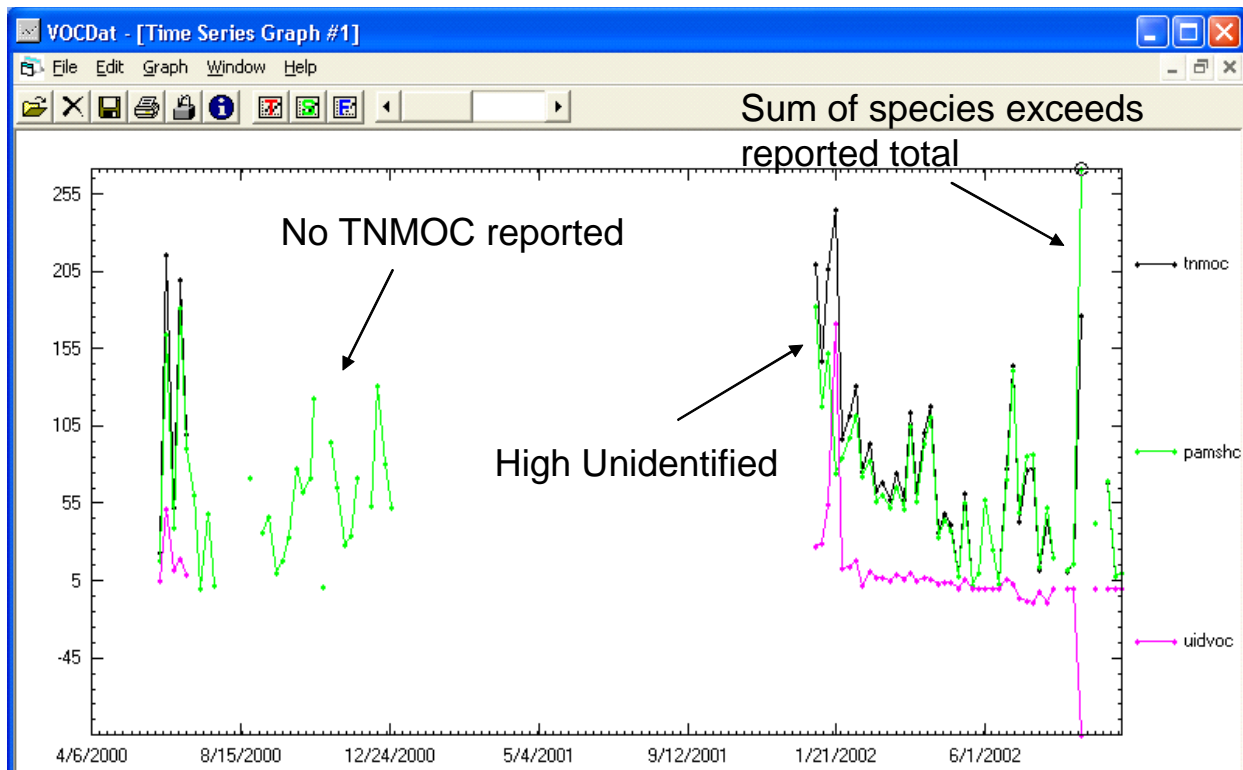


Figure 6-10. Three years of 24-hr air toxics/PAMS concentrations (ppbC) with some problems noted.

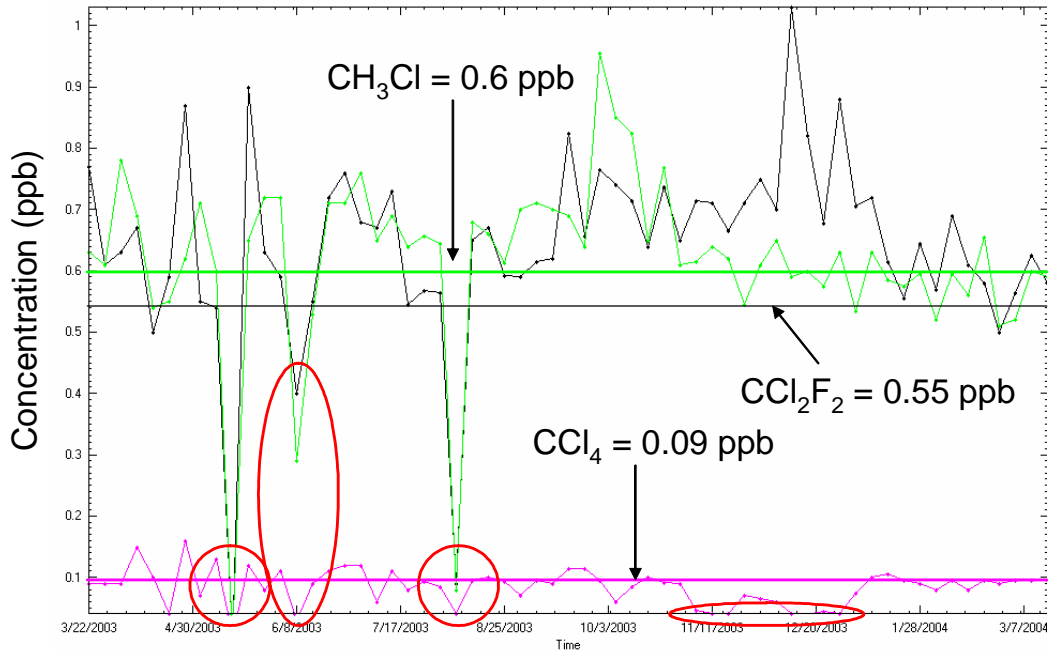


Figure 6-11. Concentrations (ppb) of carbon tetrachloride (CCl₄, pink), dichlorodifluoromethane (CCl₂F₂, black), and methyl chloride (CH₃Cl, green) at an urban site in 2003 and 2004. Known northern hemisphere background concentrations of each species were taken from internet sources (http://cdiac.esd.ornl.gov/pns/current_ghg.html and <http://www.cmdl.noaa.gov/publications/annrpt24/khalil.htm>) and plotted as a line on top of the ambient measurements. Dips of concentrations well below background concentrations are circled in red.

Table 6-2. Remote northern hemisphere background concentrations of selected air toxics taken from McCarthy et al. (2004) and other sources.

HAP	Remote Average (ppt)	Remote Minimum (ppt)	Remote Maximum (ppt)	Source
Acetaldehyde	90	60	120	Singh et al. (2001)
Benzene	44.44	13.3	84.4	S.A. Montzka, NOAA-CMDL (unpublished data)
Carbon tetrachloride	99.02	90	110	S.A. Montzka, NOAA-CMDL (unpublished data)
Chloroform	9.43	6.2	15.2	S.A. Montzka, NOAA-CMDL (unpublished data)
Formaldehyde	150	100	200	ATSDR (1992), Elinder (1985), and Zhou et al. (1996)
Methylene chloride	24.65	20.1	34.4	S.A. Montzka, NOAA-CMDL (unpublished data)
Tetrachloroethylene	3.21	1.5	10.2	S.A. Montzka, NOAA-CMDL (unpublished data)
Trichlorofluoromethane	255			http://cdiac.esd.ornl.gov/pns/current_ghg.html
Dichlorodifluoromethane	541			http://cdiac.esd.ornl.gov/pns/current_ghg.html
Trichlorotrifluoroethane	80			http://cdiac.esd.ornl.gov/pns/current_ghg.html
1,1,1-trichloroethane	33			http://cdiac.esd.ornl.gov/pns/current_ghg.html
Methyl chloride	600	500	700	Khalil and Rasmussen (1993; 1999)

7. ADDITIONAL FEATURES

In support of the NATTS, two new features have been added to VOCDat software: (1) the ability to import files from freeware software designed to process and screen aethalometer data and (2) the ability to prepare precision reports for AQS. These features are discussed in this section.

7.1 PROCESSING AND SCREENING AETHALOMETER DATA

The Aethalometer Data Masher is a freeware software program (**Figure 7-1**) available to facilitate the processing, validation, and preparation of AQS-ready Aethalometer data. This program was jointly developed by Jay Turner at the University of Washington at St. Louis and George Allen of NESCAUM. The Data Masher is stand alone (i.e., it does not require other software in order to operate) but has limited documentation and technical support.

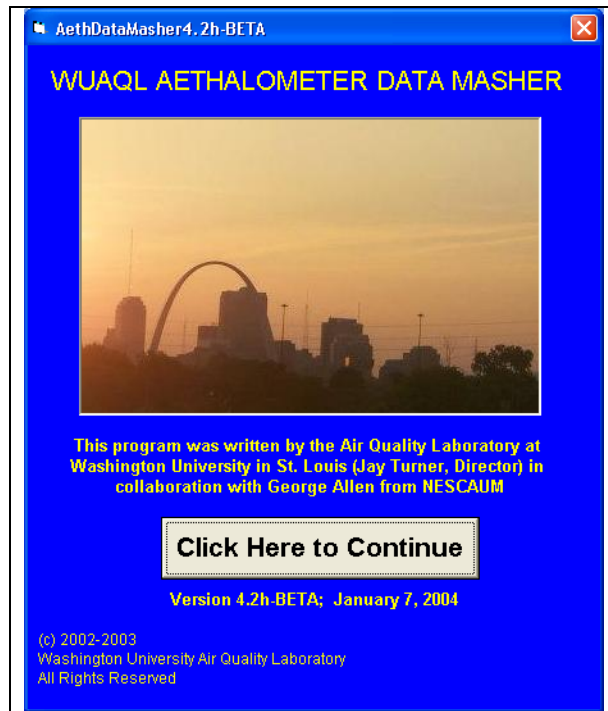


Figure 7-1. Entry screen for the Aethalometer Data Masher.

7.1.1 Data Masher Features

The Aethalometer Data Masher is designed to read up to 100 days of raw 5-minute Aethalometer data files from diskette or data logger files and to write out validated 5-minute and 1-hour black carbon (BC) data as comma separated variable (CSV) format files that are readable by other software packages such as VOCDat. The Data Masher input description is as follows:

- Multiple 5-minute BC data files can be imported, up to 3.3 months at a time (i.e., 112 days between first and last data record).
- Files do not need to be in any temporal order.
- Files can be read directly from the Aethalometer data disk or serial logger file.
- Data must be 5 minute intervals, un-compressed format, and from a 1 or 2-channel Aethalometer, not the 7-channel version.

The Data Masher also performs several data screening steps:

- Voids hours with less than 75% of possible data intervals.
- Fills in missing hourly data records with null values in order to make a complete data record.
- Screens for (and flags) various problems at the 5-minute data level such as large negative BC values adjacent to the start or end of hour, very large negative BC values at any time, and excessively low sensor channel lamp readings from tape advance failure or dirty optics.

All of the actions taken during data screening are documented in a log file as shown in **Figure 7-2**. An important data processing note is that when the Aethalometer data disk is changed, a new file with the same file name is created. Therefore, care must be taken to create unique file names when copying multiple Aethalometer data disks to a single directory for “mashing” or some file names could be overwritten resulting in missing data. A recommended solution to this naming dilemma is to rename the most recent BC*.CSV file on the target directory after each data disk is copied. Care must be taken to ensure that the new names meet the BC*.CSV format needed by the Data Masher. New names can be as simple as adding a letter to the end of the filename: for example, rename BC111302.CSV to BC111302a.CSV.

7.1.2 Validating the Data

To facilitate visual inspection of the BC data, VOCDat has been modified to import files from the Data Masher. The data output by the Data Masher are provided in a CSV file with a header record containing the AIRS parameter code for BC (e.g., **Figure 7-3**).

To import these files into VOCDat, select the appropriate species file (i.e., one that contains the parameter code for the selected BC data) and then select “File – Import Data – ASCII”. A pop-up window will appear as shown in **Figure 7-4**. For the Data Masher files, the user selects comma delimited format, number of header records = 1, head record containing species identification codes (IDs) = 1, species IDs are AIRS parameter codes, the date and time fields are separate, with the date field in column 1 and the time field in column 2. Once the user presses “OK”, the data are processed into a form that VOCDat can then plot. The user is prompted to save the file.

```

WUAQL AETHALOMETER DATA MASHER LOG FILE
Software Version 4.2c; October 7, 2003
Compilation Date/Time = 10/7/2003 8:45:43 AM
Process aethalometer raw data...
Raw Data Files(s)... 1 total files
C:\-ch masher test input data.csv
Output File #1= C:\PF05_1-ch masher test input data.csv
Data Columns: Date & Time, Date, Time, 5-Min BC, Validation Code
Output File #2= C:\PF60_1-ch masher test input data.csv
Data Columns: Date & Time, Date, Time, 1-Hour BC, Validation Code
ALL CONCENTRATIONS IN INPUT FILES ARE NANOGRAMS PER CUBIC METER
ALL CONCENTRATIONS IN OUTPUT FILES ARE MICROGRAMS PER CUBIC METER
Start Time = 07/09/2003 08:45:00
Stop Time = 09/09/2003 10:45:00
Correctly formatted data records = 14188
Theoretical data records in time interval = 17881
5-minute average data processing completed;
start 1-hour average processing...
WARNING - BC concentration less than -0.1 nanograms per cubic meter
at 08/04/2003 12:00:00 - void this record and one previous and one
subsequent 5-minute record
Aethalometer data processing completed

```

Figure 7-2. Example Aethalometer Data Masher log file.

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

Figure 7-3. Example hourly averaged PM_{2.5} standard temperature and pressure (STP) BC data (AIRS parameter code 84313).

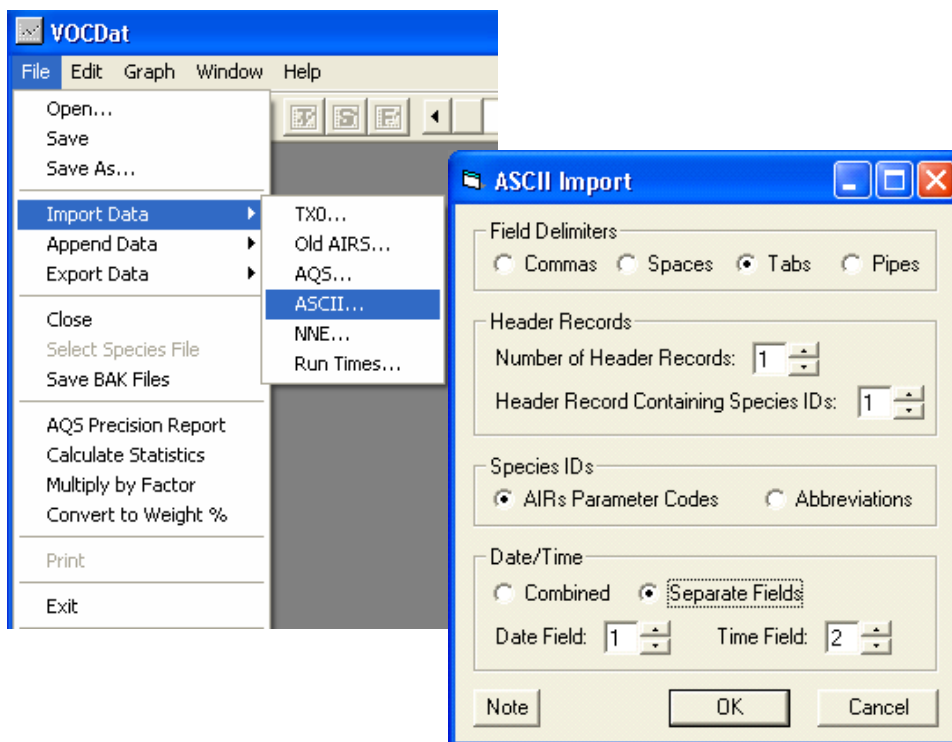


Figure 7-4. VOCDat ASCII file import pop-up windows.

Once the data are imported, the user can create time series plots of the BC data (an example is provided in **Figure 7-5**). The user can now flag single data points or groups of data points and change the QC code associated with the flagged data. AIRS null codes can be attached to invalidated data. To enhance the validation process, VOCDat allows the user to plot the BC data with other air quality and meteorological parameters collected during the same time period such as ozone, NO_x, and wind direction.

After validation, the user can export the BC data file into AQS R2 format for import into the AQS system. The export procedure within VOCDat works similarly to the import procedure.

7.1.3 Obtaining the Aethalometer Data Masher

The Aethalometer Data Masher can be obtained by contacting Jay Turner via e-mail: JRTurner@seas.wustl.edu. For the Data Masher, registration is recommended so that the developers know who to contact in order to inform users of updates and bug fixes. Please read the documentation fully before asking for support for these “freeware” applications, since support is currently unfunded. But if a problem encountered cannot be solved using the documentation, or a problem is suspected with the program, then please contact George Allen, at NESCAUM (gallen2@nescaum.org) for the Aethalometer Data Masher for assistance.

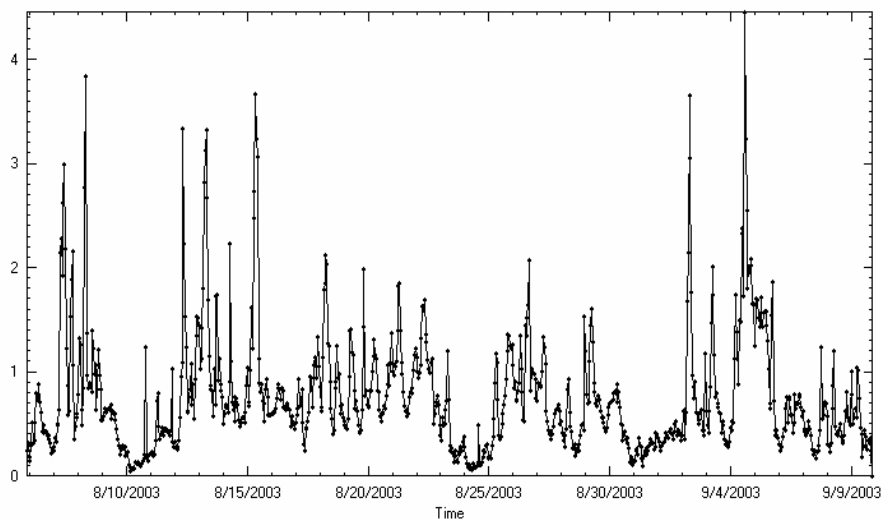


Figure 7-5. Example time series plot of 1-hr average BC data (units are $\mu\text{g}/\text{m}^3$).

7.2 PRECISION REPORT PREPARATION

Another new VOCDat feature is the ability to read precision input data from a tab- or comma-delimited file and to write an AQS Precision Report (*.RP). At this time, the input format must conform to a specific format and some user input is required to the input file.

Figure 7-6 shows the required format for the precision data. Based on an example file provided to us some time ago, the user needs to set up the data as shown in the figure with an AIRS site code, AIRS parameter code, sample duration code, unit code, method code, and parameter occurrence code (POC).

To generate a precision report using VOCDat:

- In a spreadsheet program (such as Microsoft Excel or Lotus 123), use a file called “AQS Precision Data Template.xls” as a template for the data². Organize the duplicate and replicate data in the format shown. Data for multiple AIRS site codes can be entered into this spreadsheet.
- Save the newly created file as a .txt (tab delimited) or .CSV (comma separated values) file; e.g., “Example AQS Precision Data Input.txt”.
- Open VOCDat and select “File: AQS Precision Report”; e.g., choose “Example AQS Precision Data Input.txt”.
- The AQS Precision Report will be written in AQS-ready format to the directory from which the input file resided with the same name as the input file and a RP suffix; e.g., “Example AQS Precision Data Input.RP”.

² The template is supplied with the latest VOCDat download, and after installation it will be found in the c:/program files/VOCDat/demo files directory.

AIRS Code	Parameter Code	Sample Duration Code	Unit Code	Precision Sample ID: POC ID: Method Code:	1	2	3
					5	1	2
250092006	43206	7	82	7/6/1994 0:00	2	1.97	1.99
250092006	43206	7	82	7/7/1994 0:00	2.49	2.46	2.48
250092006	43206	7	82	7/8/1994 0:00	0.59		0.59
250092006	43206	7	82	7/9/1994 0:00	1.62	1.65	1.63
250092006	43206	7	82	7/10/1994 0:00	1.75	1.74	
250092006	43206	7	82	7/11/1994 0:00	1.81	1.79	1.75
250092006	43206	7	82	7/12/1994 0:00	2.08	2.1	2.06
250092006	43206	7	82	7/13/1994 0:00	1.96	1.95	1.95
250092006	43206	7	82	7/14/1994 0:00	1.4	1.36	1.32
250092006	43206	7	82	7/15/1994 0:00	1.03		1.05
250092006	43206	7	82	7/16/1994 0:00	0.63	0.65	0.68
250092006	43206	7	82	7/17/1994 0:00	0.84	0.8	
250092006	43206	7	82	7/18/1994 0:00	0.59	0.54	0.55
250092006	43206	7	82	7/19/1994 0:00	0.66	0.68	0.73
250092006	43206	7	82	7/20/1994 0:00	0.97		0.94
250092006	43206	7	82	7/21/1994 0:00	1.5	1.48	1.46
250092006	43206	7	82	7/22/1994 0:00	2.49	2.49	2.52

Figure 7-6. Example of the data format required for preparation of an AQS Precision Report.

This is a preliminary effort to provide a means with which to create a precision report; more review and feedback is needed to make this process easier and more effective for the users. Also needed is verification of the conventions for POC numbering and more examples of how duplicate and replicate data are delivered to the state and local agencies. Additional assistance to the users could be developed in future versions of the program to provide selection boxes for duration code, unit code, and method code selection. Comments on formats and user needs are welcomed.

8. REFERENCES

- Agency for Toxic Substances and Disease Registry (1992) Toxicological profile for Beryllium. Draft prepared for U.S. Public Health Service.
- Atkinson R. (1989) Kinetics and mechanisms of the gas-phase reactions of the hydroxyl radical with organic compounds, Monograph 1. *J. Phys. Chem. Ref. Data*.
- Atkinson R. (1994) Gas-phase tropospheric chemistry of organic compounds, Monograph 2. *J. Phys. Chem. Ref. Data*.
- Carter W.P.L. (1994) Development of ozone reactivity scales for volatile organic compounds. *J. Air & Waste Manag. Assoc.* **44**, 881-899.
- Carter W.P.L. (2000) Documentation of the SAPRC-99 chemical mechanism for VOC reactivity assessment: volume 1 of 2, documentation text. Final report prepared for the California Air Resources Board, Sacramento, CA by Air Pollution Research Center and College of Engineering, Center for Environmental Research and Technology, University of California, Riverside, CA, May.
- Carter W.P.L. (2001) The SAPRC-99 chemical mechanism and updated reactivity scales. Final report prepared for the California Air Resources Board on contracts 92-329 and 95-308, available at <<http://pah.cert.ucr.edu/~carter/>>.
- Elinder C.-G. (1985) Cadmium: Uses, Occurrence, and Intake. *Cadmium and Health: A Toxicological and Epidemiological Appraisal*, CRC Press, Inc., Boca Raton, Florida.
- Hafner H.R. (2003) VOCDat toxics training: Chicago, IL, Research Triangle Park, NC, Albany, NY, and San Francisco, CA. Training Manual prepared for U.S. Environmental Protection Agency, Research Triangle Park, NC by Sonoma Technology, Inc., Petaluma, CA, STI-902600-2364-TM, May and June.
- Hafner H.R. and McCarthy M.C. (2004) Phase III air toxics data analysis workbook. Workbook prepared for Lake Michigan Air Directors Consortium, Des Plaines, IL, by Sonoma Technology, Inc., Petaluma, CA, STI-903553-2592-WB, August.
- Khalil M.A.K. and Rasmussen R.A. (1993) Arctic haze: patterns and relationships to regional signatures of trace gases. *Global Biogeochemical Cycles* **7** (1), 27-36.
- Khalil M.A.K. and Rasmussen R.A. (1999) Atmospheric methyl chloride. *Atmos. Environ.* **33** (8), 1305-1321.
- Main H.H. and Roberts P.T. (2000) PAMS data analysis workbook: illustrating the use of PAMS data to support ozone control programs. Prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-900243-1987-FWB, September.

- McCarthy M.C., Hafner H.R., and Montzka S.A. (2004) Background concentrations of 18 core air toxics in the northern hemisphere. *J. Air & Waste Manage. Assoc.* (in progress) (STI-2589).
- Singh H.B., Chen Y., Staudt A., Jacob D.J., Blake D.R., Heikes B.G., and Snow J.W. (2001) Evidence from the Pacific troposphere for large global sources of oxygenated organic compounds. *Nature* **410**, 1078-1081.
- 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.
- U.S. Environmental Protection Agency (1994) Photochemical assessment monitoring stations implementation manual. Report prepared by Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-454/B-93-051, March.
- Zhou X., Lee Y.-N., Newman L., Chen X., and Mopper K. (1996) Tropospheric formaldehyde concentration at the Mauna Loa Observatory during the Mauna Loa Observatory Photochemistry Experiment 2. *J. Geophys. Res.* **101** (D9), 14711-14720.

APPENDIX A

IMPORT FILE FORMATS

This appendix contains examples of the file formats currently supported by VOCDat for importing: (a) TX0 files (prepared by Pete Brunelli, Connecticut, Department of Environmental Protection), (b) EPA AIRS AMP 370, (c) NARSTO-Northeast data bulletin board CSV files, (d) AQS R-2, and (e) ASCII.

CREATING TX0 FILES
Prepared by Pete Brunelli, Connecticut DEP

Turbochrome *.TX0 files can be used as source data for VOCDat, and the resulting data files can be viewed in graphical (line, scatter, and fingerprint) form. Preparing these files in Turbochrome involves the use of a custom report format. The report format looks like this:

```
"=====
"Software Version:","4.0<3H19>"
"Sample Name:","a", "Time:","7/7/95","01:31 PM"
"Sample Number:","4", "Study:","Stafford PAMS"
"Operator:","pcb"

"Instrument:","970_-_0","Channel:","A","A/D mV Range:","1000
"AutoSampler:",""
"Rack/Vial:","0,0

"Interface Serial #:","9198571137","Data Acquisition Time:","5/30/95","01:46 PM"
"Delay Time:","0.00", "min."
"End Time:","48.00", "min."
"Sampling Rate:","3.1250", "pts/sec"

"Raw Data File:","C:\95VOC\F0695\F65A004.RAW"
"Result File:","C:\95VOC\F0695\RST\F65A004.RST"
"Inst Method:","C:\95VOC\F0695\F0695PL1","from", "C:\95VOC\F0695\RST\F65A004.RST"
"Proc Method:","C:\95VOC\F0695\F0695PL1.mth"
"Calib Method:","C:\95VOC\F0695\F0695PL2.mth"
"Sequence File:","C:\95VOC\F0695\F0695Z.SEQ"

"Sample Volume:","1.0000", "cc", "Area Reject:","0.000000
"Sample Amount:","1.0000","Dilution Factor",1.0000

"=====

"DEFAULT REPORT"
"Component","Raw","Time",
"Name","[ppbC]","[min]",
"-----"
"          ", 53.3893, 7.94
"43202 ethane", 4.7798, 8.71
"43203 ethene", 0.8353, 9.63
"43204 propane", 3.3920, 12.52
"43205 propene", 0.3395, 19.89
"43214 i-butane", 0.5947, 23.06
"43212 butane", 1.5257, 24.28
```

```

"43206 acetylen    ", 1.0790, 25.32
"43216 t2butene   ", 0.0000, 29.62
"43280 1butene    ", 0.0000, 30.15
"                ", 1.0748, 30.80
"43217 c2butene   ", 0.0000, 31.59
"43242 Cpentane   ", 0.0000, 32.83
"43221 i-pentan   ", 2.2020, 33.12
"43220 pentane    ", 1.2539, 33.95
"43228 2M2btene   ", 0.0000, 35.21
"43283 Cpentene   ", 0.0000, 36.55
"43226 t2pente   ", 0.5161, 36.80
"43282 3M1btene   ", 0.5804, 37.42
"43224 1pente     ", 0.3603, 37.76
"43227 c2pente    ", 0.3134, 38.35
"43244 22DMbtan   ", 0.0000, 39.89
"43284 23DMbtan   ", 0.2519, 40.47
"43285 2Mpente    ", 1.1536, 40.58
"43230 3Mpente    ", 0.5068, 40.70
"                ", 0.2481, 41.37
"43243 isoprene   ", 1.1994, 41.50
"43234 4M1ptene   ", 0.0000, 43.28
"43246 2M1ptene   ", 1.2673, 43.89
"                ", 0.4135, 45.14
"-----"
"", 77.2767,""

"=====
"CT02 EST 09 013 1001 Stafford CT Shenipsit State Forest"
"=====

```

The report has a component name, raw amount, and retention time field. The footer is laid out with a unique site ID, the time zone of the chromatography system, the AIRS site ID, and a verbose description. It is very simple and should not be altered. **ONLY CHANGE THE FOOTER!!** Devise your own site ID in the same format and modify the remainder of the line as necessary.

If you are not used to generating these files, here are a few hints.

- They are generated during data processing or reprocessing.
- If you do not want hard copies, set up a printer on NUL: or EPT: and make it the default printer. Also, you can specify the device in the sequence, but setting the null printer as the default is safer.
- Make sure that the ASCII formatted file option is enabled in the report options list.
- Look for the tx0 or tx1 files in the same directory as the result files.

**EXAMPLE AIRS AMP 370 RAW DATA CONVERSION FILE
(SAMPLING TIMES LESS THAN 24 HOURS) (*.AIR)**

```

125009200642602110070829606150030019 0012 0012 0010 0009 0010 0009 0008      I
125009200642602110070829606150830005 0004 0003 0003 0002 0003 0003 0004      I
125009200642602110070829606151630004 0006 0011 0008 0016 0016 0010 0011      I
125009200642602110070829606160030009 0009 0014 0012 0009 0006 0004 0006      I
125009200642602110070829606160830003 0006 0003 0003 0003 0003 0003 0003      I
125009200642602110070829606161630004 0008 0006 0006 0011 0016 0013 0016      I
125009200642602110070829606170030023 0026 0011 0007 0009 0016 0027 0031      I
125009200642602110070829606170830024 0021 0014 0010 0025 0030 0017 0014      I
125009200642602110070829606171630016 0016 0019 0020 0020      0035 0049      I
125009200642602110070829606180030040 0051 0025 0008 0006 0007 0007 0006      I
125009200642602110070829606180830006 0008 0006 0006 0012 0011 0006 0004      I
125009200642602110070829606181630005 0007 0007 0009 0013 0017 0014 0023      I
125009200642602110070829606190030018 0016 0018 0012 0020 0020 0019 0012      I
125009200642602110070829606190830006 0007 0008 0009 0010 0014 0014 0016      I
125009200642602110070829606191630021 0022 0030 0039 0035 0041 0026 0025      I
125009200642602110070829606200030026 0020 0018 0013 0012 0013 0012 0013      I
125009200642602110070829606200830013 0011 0010 0008 0007 0008 0006 0004      I
↑↑      ↑      ↑↑      ↑      ↑      ↑      ↑↑↑      ↑↑↑      ↑
a b      c      d e f g      h i j      k

```

Where:

- a transaction type (1)
- b state (2), county (3), site (4) code
- c parameter code (5)
- d POC (1), interval code (1)
- e units code (3)
- f method code (3)
- g date (YYMMDD)
- h start hour (2)
- i decimal point indicator (1)
- j sample data occurs 8 times (4), validity flag (1)
- k action code

Source: AIRS users guide volume AQ1, page 4-232

EXAMPLE EXCERPT SHOWING NARSTO-NORTHEAST BULLETIN BOARD DATA FORMAT (*.CSV)

```
"HYDROCARBON PAMS VOC DATA FOR THE STATE OF RHODE ISLAND DURING THE MONTH OF AUGUST,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,  
"DATE OF LAST REVISION OF DATA 06/11/96, AVERAGING PERIOD: 180(min),,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,  
Val,SITE,DATE,EST,DURATION,SAMPLER ID,CANISTER,RECORD,acety,,ethyl,,ethan,,prpyl,,propa,,isbta,,ndec,,nundc,,tnmoc,,form,,acet,,aceta,  
Lvl,,mm/dd/yy,hh/mm,(min),,,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,QC,ppbC,  
0,440071010,8/1/95,2:00,180,-999,-999,0,2.3,0,2.4,0,5.3,0,0.6,0,4.8,0,1.9,0,0.4,0,3,0,0,0,0,0,5.8,0,0,0,2.2,0,0,0,0,0,0,0,0.1,0  
0,440071010,8/1/95,5:00,180,-999,-999,0,6,0,7.2,0,9.4,0,2.2,0,10.6,0,6.6,0,2.4,0,12.9,0,0.6,0,0.5,0,23.6,0,0.7,0,9.5,0,4.4,0,0.9,0  
0,440071010,8/1/95,8:00,180,-999,-999,0,4.4,0,4.8,0,7.7,0,1.5,0,7.6,0,4,0,1.4,0,8.1,0,0,0,0,0,14,0,0.3,0,5.7,0,6.9,0,0.2,0,0,0,0.7  
0,440071010,8/1/95,11:00,180,-999,-999,0,1.4,0,1.3,0,5,0,0.6,0,4,0,1.2,0,0.7,0,2.5,0,0,0,0,0,3.7,0,0,0,1.6,0,11.8,0,0,0,0,0,0,0  
0,440071010,8/1/95,14:00,180,-999,-999,0,1.6,0,1.4,0,6,0,0.7,0,4.5,0,1.8,0,1,0,4.1,0,0,0,0,0,6.4,0,0,0,2.5,0,10.4,0,0.1,0,0,0,0  
0,440071010,8/1/95,17:00,180,-999,-999,0,2.8,0,2.2,0,5.9,0,0.9,0,6,0,2.1,0,1.3,0,4.8,0,0,0,0,0,7.7,0,0.2,0,3.2,0,5.9,0,0.3,0,0.1,0  
0,440071010,8/1/95,20:00,180,-999,-999,0,2,0,1.9,0,5.3,0,0.9,0,5.8,0,4.2,0,1.7,0,10.5,0,0.7,0,0.7,0,19.8,0,0.7,0,8.5,0,0.7,0,0.9,0  
0,440071010,8/1/95,23:00,180,-999,-999,0,1.4,0,1.2,0,5.2,0,0.5,0,4.6,0,1.2,0,0.6,0,2.6,0,0,0,0,0,4.2,0,0,0,1.8,0,0,0,0,0,0,0  
0,440071010,8/2/95,2:00,180,-999,-999,0,2.2,0,2.2,0,9.6,0,0.7,0,7,0,3.2,0,0.9,0,7.3,0,0,0,0.2,0,13.9,0,0.4,0,5.4,0,0,0,0.4,0,0.2,0
```

Note: VOCDat users should use the electronic version of the example as a template for preparing their data for import into VOCDat. The example file (sample NNE file.csv) is provided with other demonstration files.

EXAMPLE ASCII FILE FORMAT

site	date		canister	cal100file	filelocation	sample.time	totalsPE1	METHANE	ethan ETHANE	ethyl ETHYLENE
BT	5/17/99	1:00 AM	AQ-211	PE1152.CAL	PE1.04R	PM	50.91	37.81	3.9	1.12
BT	5/21/99	2:00 AM	AQ-187	PE1152.CAL	PE1.54R	PM	68.45	43	3.98	3.23
BT	5/25/99	3:00 AM	AQ-162	PE1152.CAL	PE1.53R	PM	67.33	41.61	3.51	0.43
BT	6/2/99	4:00 AM	AQ-046	PE1152.CAL	PE1.12R	PM	52.61	37.54	2.33	0
BT	6/8/99	5:00 AM	AQ-033	PE1152.CAL	PE1.14R	PM	61.15	32.96	0	0.92
BT	6/16/99	6:00 AM	AQ-134	PE1175.CAL	PE1.78R	PM	63.6	38.02	3.58	1.24
BT	6/18/99	7:00 AM	AQ-168	PE1175.CAL	PE1.76R	PM	53.09	37.51	3.09	0
BT	6/22/99	8:00 AM	AQ-002	PE1175.CAL	PE1.11R	PM	48.46	34.63	2.11	0.74
BT	6/24/99	9:00 AM	AQ-045	PE1175.CAL	PE1.12R	PM	45.97	31.15	2.42	0.71

The requirements on ASCII file formats are as follows:

- Data must be arranged with each row representing a separate sample.
- Only data from one site at a time may be imported (here, BT).
- Data should be sorted by date/time and should not contain any duplicate records.
- Acceptable delimiters are commas, spaces, tabs, or pipes (e.g., Excel exports comma separated variable *.CSV files and tab-delimited *.TXT files).
- Any number of header records may be present. One of the header records must contain species identifications expressed as AIRS parameter codes or species abbreviations (in the example above, there are two header rows with species names in row 1).
- The species identifications may be AIRS parameter codes or abbreviations (names in row 1 of this example match the species.txt file). Whichever one is used, the parameter codes or abbreviations in the species.txt and/or the species2.txt files must match exactly or the data will not be imported.

The date and time of the samples must be specified. This information can be combined in a date/time field or provided separately (here, the date and time are separate with date in column 2 and time in column 3). VOCDat does not make any conversions of the time field. If your data are in daylight time, please change the data to standard time before importing to VOCDat. VOCDat assumes the data are in standard time for exporting to AQS format.

APPENDIX B

EXPORT FILE FORMATS FOR DATA EXPLORATION

This appendix contains examples of the file formats currently produced by VOCDat for exporting that are useful for further exploration of the data: (a) DAT files, (b) TXT files, and (c) graph export TXT files. Graph export *.CSV files are identical to the *.TXT files except that a comma is used as a delimiter instead of a tab.

EXCERPT OF VOCDAT EXPORT OPTION "DAT"

Year	Month	Day	Hour	DOW	WDWE	QC	QCS	acety	ethyl
1995	7	1	23	6	2	0	0	2	3
1995	7	2	5	7	2	0	0	2	2
1995	7	2	12	7	2	7	7	.	.
1995	7	2	16	7	2	0	9	.	1
1995	7	5	5	3	1	0	7	3	4
1995	7	5	12	3	1	0	0	2	.
1995	7	5	16	3	1	0	0	1	1
1995	7	10	23	1	1	0	0	.	2
1995	7	11	5	2	1	0	0	2	3
1995	7	11	12	2	1	0	0	1	1
1995	7	11	16	2	1	0	0	.	.
1995	7	14	5	5	1	0	0	3	6
1995	7	14	12	5	1	0	0	2	3

Data in the .DAT file are tab-delimited where:

DOW = day of week with 1=Monday and 7=Sunday

WDWE = weekday/weekend with Monday through Friday = 1, Saturday and Sunday = 2

QC = quality control flag for the record

QCS = "worst" quality control flag for an individual species in the sample, record QC not necessarily changed

Missing data are signified by a "."

EXCERPT OF VOCDAT EXPORT OPTION "TXT"

[Project Information]

[Site Information]
 Name = Lynn, MA 1998
 Code = 2006
 AIRS Code = 250092006
 Latitude (ddmmss) =
 Longitude (ddmmss) =
 Latitude (deg) = 0
 Longitude (deg) = 0
 UTM Northing (km) = 0
 UTM Easting (km) = 0
 UTM Grid Zone = 0
 Elevation (ft) = 0
 Elevation (m) = 0

[File Information]

Creator = VOCDat
 Version = 2.16
 Begin Date = 6/1/98 00:00:00
 End Date = 6/30/98 23:00:00
 Number of Days = 30
 Number of Records = 720
 Number of Data Fields = 75
 Time Zone = EST
 Sampling Interval (hr) = 1
 Sample Duration Code =
 Sampling Frequency Code =

[QC Codes]

0 = Valid
 1 = 9986 CALIBRATION
 7 = Suspect
 8 = Invalid
 9 = Missing

[Data]

Val	Start	Start	Sample	Record	Code:					
Lvl	Time	Hour	Time	QC	Worst	ethan	ethyl	propa	prpyl	
					QC	ppbC	ppbC	ppbC	ppbC	
0	6/1/98 0:00	6/1/98 0:00	0:40	0	0	6.52	0.56	2.68	0.43	43202
0	6/1/98 1:00	6/1/98 1:00	0:40	0	0	4.13	0.25	1.73	0	43203
0	6/1/98 2:00	6/1/98 2:00	0:40	0	0	5.1	0.18	2.35	0.12	43204
0	6/1/98 3:00	6/1/98 3:00	0:40	0	0	5.01	0.28	1.69	0.26	43205
0	6/1/98 4:00	6/1/98 4:00	0:40	0	0	5.47	0.68	2.2	0.34	

These files look very similar to the VOCDat (.VOC) files but no longer contain the individual species QC codes. Data are tab-delimited.

EXAMPLE GRAPH EXPORT FORMAT (*.TXT)

This example is from a time series of ethane (ethan) and ethene (ethyl) from a file called temp.voc.

temp.VOC	ethan	ppbC
23 Jun 98 23:00		6.94
24 Jun 98 00:00		6.01
24 Jun 98 01:00		4.57
24 Jun 98 02:00		5.54
24 Jun 98 03:00		8.55
24 Jun 98 04:00		9.56
24 Jun 98 06:00		6.32
24 Jun 98 07:00		6.10
24 Jun 98 08:00		6.00
24 Jun 98 09:00		5.04
24 Jun 98 10:00		14.86
24 Jun 98 11:00		7.96
24 Jun 98 12:00		6.09
24 Jun 98 13:00		4.72
24 Jun 98 14:00		5.24

temp.VOC	ethyl	ppbC
23 Jun 98 23:00		1.37
24 Jun 98 00:00		1.30
24 Jun 98 01:00		0.57
24 Jun 98 02:00		1.02
24 Jun 98 03:00		1.19
24 Jun 98 04:00		1.54
24 Jun 98 06:00		2.75
24 Jun 98 07:00		3.32
24 Jun 98 08:00		3.61
24 Jun 98 09:00		3.11
24 Jun 98 10:00		5.28
24 Jun 98 11:00		4.58
24 Jun 98 12:00		1.87
24 Jun 98 13:00		1.77
24 Jun 98 14:00		2.18
24 Jun 98 15:00		1.87
24 Jun 98 16:00		3.99
24 Jun 98 17:00		4.11

etc.

This example is from a scatter plot of benzene (benz) versus acetylene (acety) concentrations. Data were from a file called temp.VOC.

temp.VOC	temp.VOC
benz	acety
ppbC	ppbC
0.65	0.82
0.38	0.42
0.37	0.57
0.44	0.51
0.55	0.65
0.87	1.82
1.62	3.44
1.03	1.19
0.57	0.71
0.36	0.44
0.36	0.61
0.47	0.71
0.34	0.45
0.31	0.63
0.38	0.51
0.49	0.70
0.92	1.16
1.43	1.68
1.17	1.51
0.83	0.90
0.83	0.90
0.67	0.82
0.58	0.73

APPENDIX C

ANNOTATED CONFIGURATION FILE

All configurations, such as the screening criteria, graph colors, and the positions of graphs in the VOCDat window, can be saved for future sessions by choosing “Windows - Save Configuration”. You will then be prompted for a file name. A unique file name may be given to any number of configurations. Custom configurations can be restored in a future session (for different sets of graphs if desired) by choosing “Windows - Restore Configuration”. This appendix contains an annotated example; annotations are inserted in *bold italics*. The file can be edited in a text editing program and it is sometimes useful to change screening criteria or default file locations using this method.

[Program Information]

DataFilePath =
SpeciesFilename = C:\Program Files\VOCDat\Species\species.txt
Password =
RecentFilename1 = C:\Data\California\sanjose.VOC
RecentFiletitle1 = sanjose.VOC

*the location of commonly used data files can be specified here
this is the location of the default species file
if a password is desired, this is where to set it
VOCDat remembers the 4 most recently accessed files*

.
RecentFilename4 = C:\Data\Maine\Bangor.VOC
RecentFiletitle4 = Bangor.VOC
QC0Name = Valid
QC1Name = 9986 CALIBRATION
QC2Name = 9986 CALIBRATION
QC3Name = 9980 MACHINE MALFUNCTION (MALF)
QC4Name = 9988 POWER FAILURE (POWR)
QC5Name = 9993 MAINTENANCE/ROUTINE REPAIRS
QC6Name = Failed QC
QC7Name = Suspect
QC8Name = Invalid
QC9Name = Missing
QC0Show = True
QC1Show = False
QC2Show = False
QC3Show = False
QC4Show = False
QC5Show = False
QC6Show = False
QC7Show = False
QC8Show = False
QC9Show = False
QC0Color = 65280
QC1Color = 16711680

user defined QC codes 1-6

*instructions to VOCDat on whether to show data or not once
flagged with these QC codes*

*a series of instructions to VOCDat regarding color of data points
we do not recommend changing these*

MarkColor = 8421504
ShowToolbar =-1
ShowStatusbar =-1
SaveBAK = -1

[AutoQC1]
1 ,acety, 0.5
1 ,ethan, 0.5
1 ,propa, 0.5
1 ,nbuta, 0.5
1 ,ispna, 0.5
1 ,npnta, 0.5
0 ,nhexa, 0.5
1 ,benz, 0.5
1 ,tolu, 0.5
1 ,m/pxy, 0.5

[AutoQC2]
1 ,ethyl,ethan, 0 , 0
1 ,prpyl,propa, 0 , 0
1 ,npnta,ispna, 0 , 0
1 ,oxyl,m/pxy, 0 , 0
1 ,benz,tolu, 0 , 0
1 ,3mpna,2mpna, 0 , 0
0 ,, 0 , 0
1 ,tnmoc,uidvoc, 1000 , 200
1 ,ethan,benz, 2 , 2
1 ,uidvoc,tnmoc, 50 , 0

whether to show the tool and status bars or not

whether to create back up files when using the save file command, where 0 = no and -1 = yes

*abundant species concentrations screening criteria
0 = ignore, 1 = use; species abbreviation; concentration*

*concentration comparison screening criteria
0 = ignore, 1 = use; species abbreviation; concentration value 1, value 2*

[AutoQC3]

1 ,acety, 3
1 ,ethan, 3
1 ,propa, 3
1 ,nbuta, 3
1 ,ispna, 3
1 ,npnta, 3
1 ,nhexa, 3
1 ,benz, 3
1 ,tolu, 3
1 ,m/pxy, 3

concentration variability screening criteria
0 = ignore, 1 = use; species abbreviation; number of
standard deviations

[Window Configuration]

MainFormPos = 1335 , 540 , 13470 , 10095

position of the VOCDat window