Speciation Data Validation Analysis Tool (SDVAT)

Version 2.0

User's Guide

Prepared for:

Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Prepared by:

Ed Rickman
Linda Andrews
Jennifer Lloyd
Emaly Simone

Environmental Health and Safety Division
RTI International
Research Triangle Park, NC 27709

EPA Contract 68-D-02-065, Work Assignment 4-09
RTI Project 91U-8637-004-009-003

Copyright 2002, 2006 (pending EPA approval)

September 2006

[This page intentionally left blank.]

Speciation Data Validation Analysis Tool (SDVAT)

Version 2.0

User's Guide

Prepared for:

Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Prepared by:

Ed Rickman
Linda Andrews
Jennifer Lloyd
Emaly Simone

Environmental Health and Safety Division
RTI International
Research Triangle Park, NC 27709

EPA Contract 68-D-02-065, Work Assignment 4-09
RTI Project 91U-8637-004-009-003

Copyright 2002, 2006 (pending EPA approval)

September 2006

[This page intentionally left blank.]

Table of Contents

List of Figures

Acronym List

AIRS	Aerometric Information Retrieval System
AQS	Air Quality System
CAA	Clean Air Act
COC	chain of custody
EPA	U.S. Environmental Protection Agency
FRM	Federal Reference Method
IC	ion chromatography
MB	megabytes
MDL	minimum detection limit
MS	Microsoft
PM	particulate matter
POC	Parameter Occurrence Code
RAM	Random Access Memory
SDVAT	Speciation Data Validation Analysis Tool
SIP	State Implementation Plans
STN	Speciation Trends Network
XRF	X-ray fluorescence

[This page intentionally left blank.]

1.0 Project Background

The Clean Air Act (CAA) requires the U.S. Environmental Protection Agency (EPA) to revise or update air quality standards based on a review of the latest scientific information on known and potential human health effects associated with particulate matter (PM) levels found in ambient air. After reviewing the air quality criteria for PM and epidemiological data, the EPA found an association between ambient concentrations of PM and a range of serious health effects.

The PM$_{2.5}$ Chemical Speciation Trends Network (STN) was established in 1999 to characterize major aerosol mass components in urban areas of the United States and to provide an accountability mechanism to assess the effectiveness of emission mitigation programs. The study examines the fraction of PM with aerodynamic diameters less than or equal to 2.5 micrometers (PM$_{2.5}$). Since its inception, the program has grown from 12 air quality trends monitoring sites to approximately 54 sites. These sites use multiple speciation samplers and consistent sampling and analysis protocols to examine air quality trends over time. Additional state, local, and tribal sites provide additional data points to enhance the STN. These non-trends sites supply information for developing effective State Implementation Plans (SIPs). The sampling frequency and location of these sites are flexible to address regional and local issues.

1.1 The Speciation Data Validation Analysis Tool (SDVAT)

The contract laboratory provides speciation data to monitoring agency data contacts monthly. This data has already passed the contract laboratory's Level 0 and Level 1 validations (covering field and laboratory data quality) and is ready for higher-level validations by the monitoring agencies. These higher-level validations include the following:

- Data completeness
- Time series analyses
- Mass concentration reconstruction
- Pie chart species distributions
- Spatial and temporal variability analyses.

The Speciation Data Validation Analysis Tool (SDVAT) is designed to allow monitoring agency specialists to review and validate data from the monthly monitoring reports. The SDVAT imports data from the monthly Microsoft (MS) Excel spreadsheets, displays the data for modification and approval, and presents data analyses to the user in MS Excel charts and spreadsheets.

[This page intentionally left blank.]

2.0   System Overview

The SDVAT is an MS Access 2002 (XP) database program that uses MS Excel 2002 (XP) for graphing and analysis. Although the program may be installed on a local area network, the program was designed for a single-user environment. To avoid problems with data selection and record locking, multiple users should each have their own database copy.

2.1   System Requirements

The SDVAT program should be run on a personal computer with the following:

- Windows 2000 or Windows XP operating system.

- 256 megabytes (MB) of random access memory (RAM). Users who plan to review a large number of sites look at long time periods or run other applications while using the SDVAT program may want to add more memory.

- Pentium (500 MHz or higher) or similar processor.

- MS Excel 2002 (XP) (preferably MS Office XP). Users without MS Access 2002 (XP) may install the version of the application that includes the MS Access Run-Time package. This program has not been tested with other versions of MS Excel or MS Access.

- Sufficient free hard-drive space to install the program, data, and the run-time version of MS Access (if used).

This program may run on machines with less memory or slower processors, but it could cause poor performance. Not having adequate RAM can result in significantly slower operation while lower processor speed will generally cause a lesser effect.

2.2   Installation Procedures

The first (Section 2.2.1) is for use on computers that already have MS Access 2002 and MS Excel 2002 installed. The second (Section 2.2.2) is for use on computers that have MS Excel 2002, but do not have MS Access 2002 installed.

2.2.1   Systems with Both Access 2002 and Excel 2002

Users who already have MS Access 2002 installed simply need to copy the supplied MS Access 2002 database file (SDVAT_V2-0.mde from the "WithAccess" folder) to a suitable folder on their hard drive or network. Open the .mde file to start using the program. Remember to make a separate copy for each user.

If you are copying the SDVAT.mde file from a CD-ROM, be certain to remove the read-only property from the file (all files on CD have this property set). Otherwise the file will be read-only and no data can be added.

To do this in Windows Explorer, select the SDVAT_V2-0.mde file as copied to your hard drive. Then right mouse click and select "Properties." Under the "General" tab there is a "Read-only" check box. Make certain that this box is cleared (clear it if it is necessary).

2.2.2   Systems with MS Excel 2002, but Without MS Access 2002, Including Systems
with Earlier Versions of MS Access

Run the setup.exe file found in the "Without Access" folder on the CD and follow the on-screen prompts. The added programs may be removed (if desired) from the Add/Remove icon in the Windows Control Panel.

Note: Installation of the run-time version of MS Access 2002 on systems with earlier versions of MS Access (e.g., MS Access 95 or MS Access 97) may cause problems with file type associations in File Manager. File Manager will typically try to open an MS Access database in the version of MS Access that was used last. To prevent problems when using multiple MS Access versions, open the correct version of MS Access from the Programs menu, then open the database file from the File Open menu in MS Access.

2.2.3   Systems Without MS Excel 2002 or MS Access 2002

You will need to purchase and install a copy of MS Excel 2002 (or preferably MS Office XP Professional, which includes MS Excel 2002 and MS Access 2002), then install the SDVAT as described above.

3.0　Import Overview

The data from the contract laboratory are available to EPA and monitoring agency data contacts in a variety of formats. The SDVAT is designed to work with the main MS Excel spreadsheet for its input (i.e., the files without "summary" in the file name). These spreadsheets (i.e., the "B????Spreadsheet.xls" files on the Web site) were chosen because they contain lab and field parameters, unlike the summary spreadsheets, which contain only chemical speciation data. Because the speciation samplers and analysis laboratories produce so much information for each event, these spreadsheets may be difficult to interpret. The SDVAT is designed to manipulate this data and present it in a more understandable way without special analysis tools.

3.1　Loading Data from Spreadsheets

The first step is to download your sampling site's data files from the contract laboratory's external Web site. It's important to note that access to these files is limited to a designated site data contact. Other users who may need access to this data should contact the site data contact for a copy of the appropriate files. Once you have accessed the Web site and downloaded your site's files onto your hard drive, you are ready to load the information into the SDVAT system.

Go to the SDVAT main menu (Figure 3-1) and use the left mouse button to click on the box labeled "Import Data."
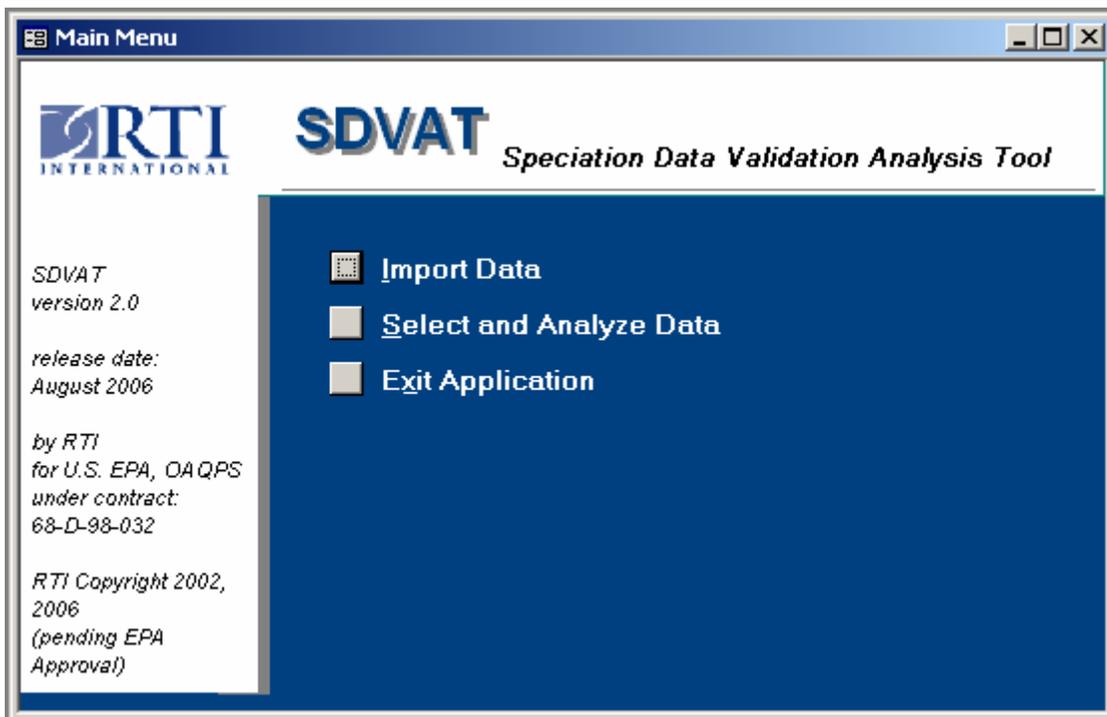

Figure 3-1. Main menu (Import Data).

The Import Data form (Figure 3-2) will appear.
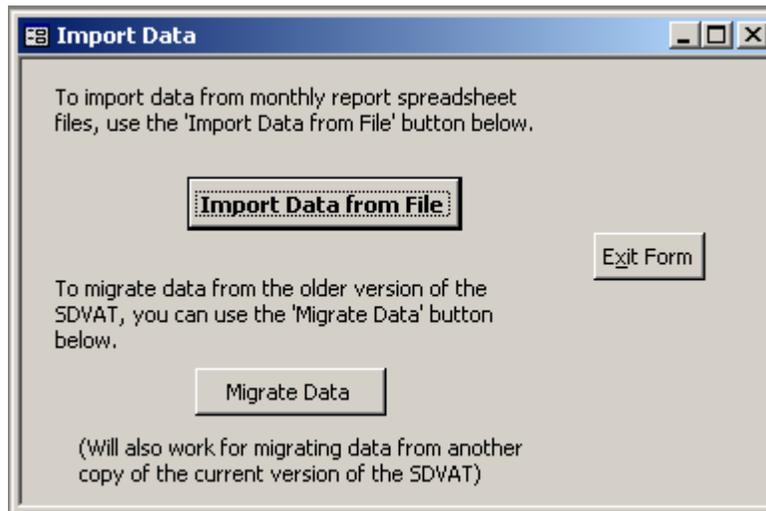
Figure 3-2. Import Data form.

This form has the following three buttons:

- "Import Data from File," which begins the import process.
- "Migrate Data," which imports data from the old SDVAT version.
- "Exit Form," which returns to the main menu.

After selecting the "Import Data from File" button, an Explorer window will open on your screen. Note that the default file location is the same folder where the SDVAT_V2-0.mde file is saved. Navigate your hard drive until you have found the folder that contains your speciation data form the Web site (Figure 3-3).
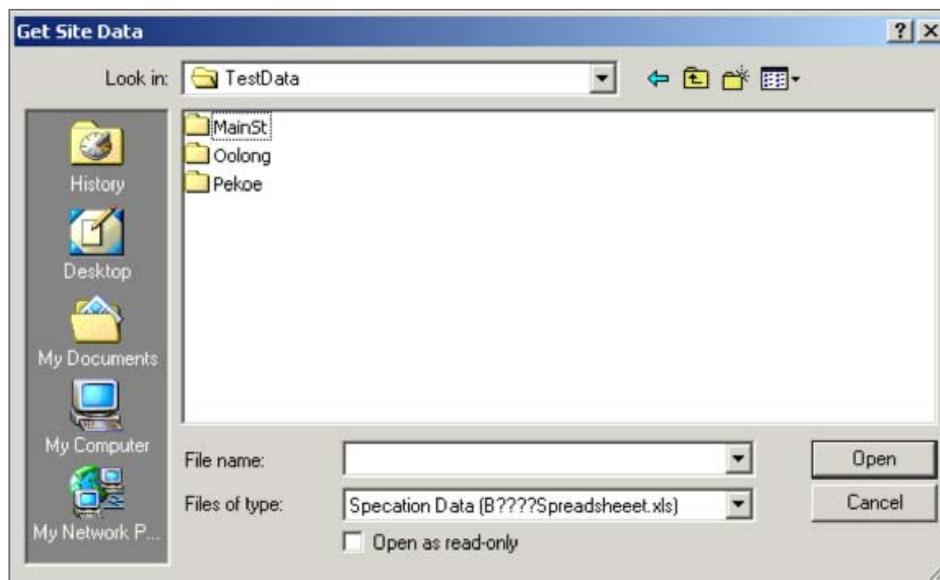
Figure 3-3. Explorer window (file location).

Using the left mouse button, double click on one of the folders containing the speciation data. The folder will open and reveal MS Excel spreadsheets that can be for imported (Figure 3-4). Please note that the default file type is the "B????spreadsheet.xls," which will find the original file names as produced by the contract laboratory. You may import MS Excel spreadsheets with other names (e.g., files that you may have modified and saved using the SDVAT) by selecting the "All Excel spreadsheets (*.xls)" from the file type dropdown box (Figure 3-4).
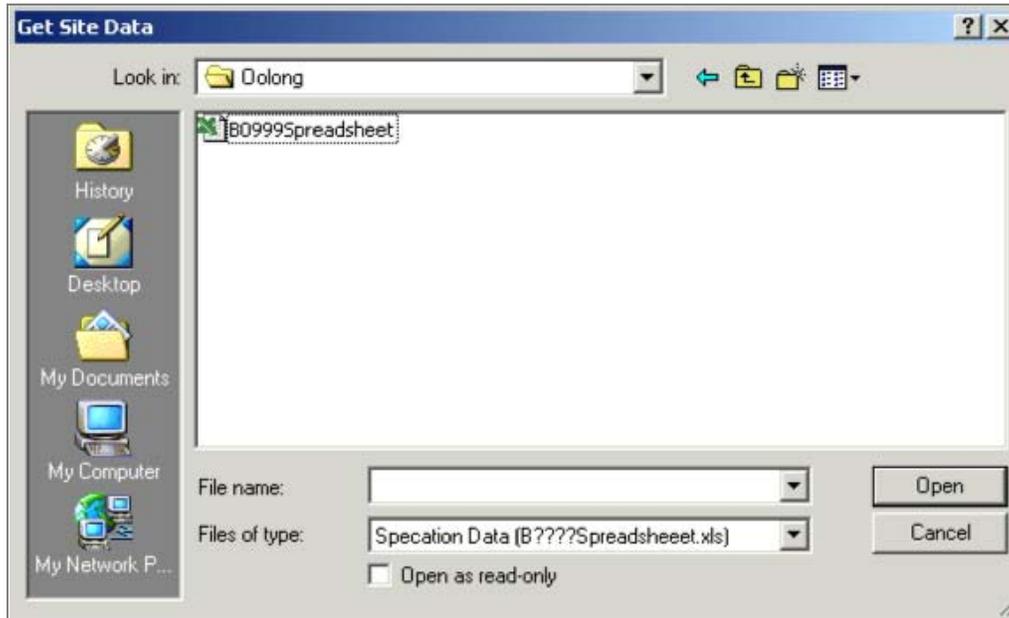
Figure 3-4. Explorer window (file type).

Use your mouse to select the spreadsheet that you want to import, and click the "Open" button. The Explorer window will disappear. Watch the status bar (Figure 3-5) in the bottom left of your computer screen. When "Form View" appears on the status bar, the import process has been completed.

Figure 3-5. Status bar.

You may repeat these steps to import additional spreadsheets into the database. To return to the main menu, use your mouse to select the "Exit Form" button from the Import Data form.

3.2   Migrating Data from a Previous Version

The "Migrate Data" button on the same Import Data form (Figure 3-2) will import data from a previous SDVAT program version into the current SDVAT program. Once you have clicked on the "Migrate Data" button, an Explorer window will appear. Please note that the default file type is the "Access Databases (*.mde)," which will find the SDVAT program file. You may also select the "Access Databases (*.mdb)" from the file type dropdown box to import data from MS Access databases with the "mdb" file extension (the import will only be successful if the database table is in the same format as the SDVAT).

If you do not have the previous version saved in the same location as the current SDVAT, navigate until you have found the folder that contains the older SDVAT version. Use your mouse to double click on the older SDVAT file from which you wish to migrate data. The Explorer window will disappear, the cursor will turn into an hourglass, and the status bar will show the current status of the migration.

During the data migration process, the program checks to see if there are any duplicate records in both versions of the SDVAT. If there are any duplicate records, you will be prompted about whether you want to replace the current records with those from the older version, keep the records in the current version and only migrate those from the older version that are not already in the current version, or cancel the migration (Figure 3-6).
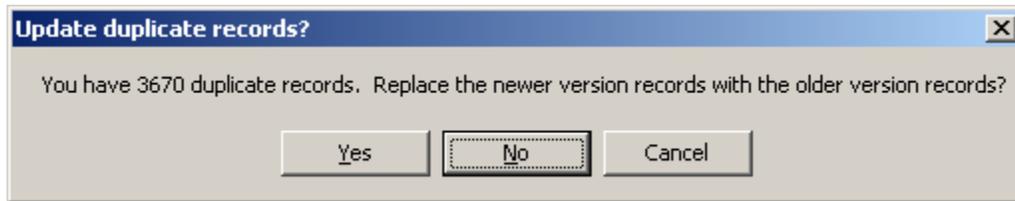
Figure 3-6. Duplicate records message.

When the migration is complete, a message box will appear (Figure 3-7). To exit the migration process, use the left mouse button to click "OK."

Figure 3-7. Message box.

To return to the main menu, use the left mouse button to select the "Exit Form" button from the Import Data form.

4.0 Filtering Data for Analysis

Speciation sampling networks generate large amounts of data; therefore, it is often useful to select subsets of this data for analysis and processing. This selection process is referred to as "filtering" the dataset. SDVAT contains an easy-to-use and powerful, form-based filter tool that allows you to filter data based on a variety of criteria (e.g., particular site, analysis, analyte, sampling period, sample type, and validity type).

4.1 Getting Started

Open the filter tool by selecting "Select and Analyze Data" from the main menu (Figure 4-1).
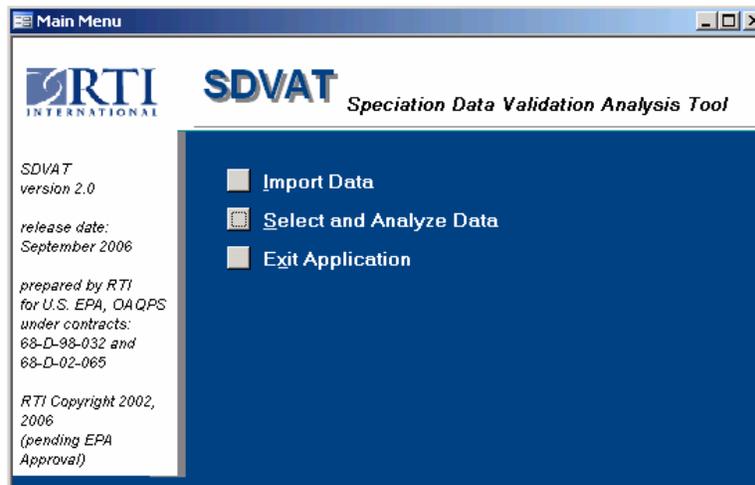
**Figure 4-1. Main menu (Select and Analyze Data).**

After selecting this option, the Choose Criteria to Filter the Data form (Figure 4-2) will appear. This form describes the steps involved in selecting and analyzing the data. To continue on to the first step, select the "Continue" button. To return to the main menu, click on the "Exit" button.
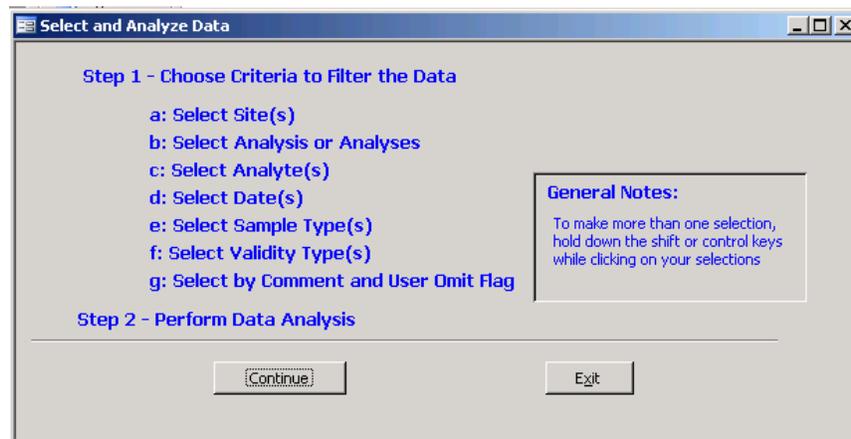
Figure 4-2. Choose Criteria to Filter the Data form.

In addition to providing an overview of the process, this form also offers some helpful notes on selecting items from lists on the filter forms.

- To select multiple items that are next to each other in the list, use your mouse to select the first item. Next, press and hold the "Shift" key with one hand. While holding down the "Shift" key, use your mouse to click on the last item in the series. This will select the first item, the last item, and all of the items listed in between.

- To select multiple items that are not next to each other in the list, use your mouse to select the first item. Then, press down on the "Control" (Ctrl) key and use your mouse to click on each of the other items you want to select.

4.2   Selecting Sites

The next form (Select Site(s) form, Figure 4-3) contains a list of the air quality monitoring sites in the SDVAT database. The sites are listed alphabetically by site name. The Air Quality System (AQS) code (previously known as the Aerometric Information Retrieval System [AIRS] code) and the AQS Parameter Occurrence Code (POC) are displayed to the right of the site name.

Figure 4-3. Select Site(s) form.

To select a site, left mouse click on the site name. Sites that are not visible on the form may be viewed by using the scroll bar to the right of the POC. You may also select a site with the mouse and use the arrow keys to move this selection up or down on the list. Use your mouse and keyboard as described in Section 4.1 to select multiple sites.

The following are five buttons on this form:

- "Select All Sites," which selects all of the sites in the list for analysis.

- "Clear All Sites," which clears all selected sites on this form and allows you to restart.

- "Start Over," which clears the criteria on all of the selection forms and restarts the selection process. On this form, the button serves the same purpose as the "Clear All Sites" button; however, this button will return you to this form in later forms.

- "Exit Form," which allows you to quit the process at any time.

- "Next –>>," which signals that your choices are complete and displays the next form.

4.3   Selecting Analyses

After selecting the sites from the Select Site(s) form and pressing the "Next –>>" button, the Select Analysis or Analyses form (Figure 4-4) appears.
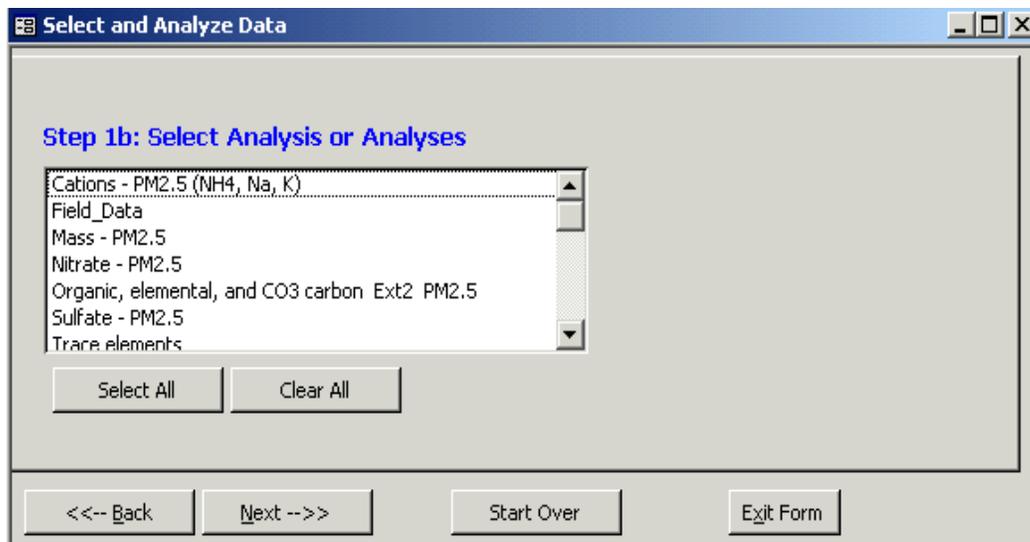
Figure 4-4. Select Analysis or Analyses form.

The analyses performed on the PM$_{2.5}$ samples are alphabetically listed in this form. To permit review of field sampler data (e.g., flow rate, sample volume) submitted to the contract laboratory, a "Field_Data" analysis has been added to the list of analyses as shown in Figure 4-4. Use your mouse to select the desired analysis. Additional choices may be viewed using the scroll bar. Multiple analyses may be selected as described in Section 4.1.

The following are six buttons on this form:

- "Select All," which selects all of the analyses in the list.
- "Clear All," which clears all of the selections on this form.
- "Exit Form," which quits the filtering process.
- "Start Over," which clears all of the entries and returns you to the Select Site(s) form (Figure 4-3).
- "Next –>>," which indicates that your choices are complete and moves you to the Select Analyte(s) form (Figure 4-5).
- "<<– Back," which returns you to the previous form (Select Site(s) form, Figure 4-3).

4.4   Selecting Analytes

The Select Analyte(s) form (Figure 4-5) will be displayed after the analyses have been selected.

Figure 4-5. Select Analyte(s) form.

This form shows an alphabetical list of analytes. The analytes are analysis-specific and vary depending on which analyses were selected on the previous form. For example, selecting the "Nitrate – PM2.5" analysis will provide the analyte "Nitrate," whereas selecting "Trace elements" will yield a choice of 48 different analytes.

Use your left mouse button to select the desired analyte. More choices may be viewed using the scroll bar. Multiple analytes may be selected as described in Section 4.1. The other six buttons are described above.

4.5   Selecting Dates

Once the analytes have been selected, the Select Date(s) form (Figure 4-6) will be displayed.

Figure 4-6. Select Date(s) form.

On this form, sampling periods may be selected by "Earliest and/or Latest Start Date"or "Earliest and/or Latest Scheduled Date."

Every sampling event has a scheduled date. The scheduled date is the date when the sampling should occur. The start date is the date when sampling occurs. These dates are often the same, but occasionally the samples are collected on a different date. Some samples, such as blanks, may not have a start date.

Blanks are sampling events that are designed to measure the background PM$_{2.5}$ levels that the sampling media may encounter in the laboratory, in the field, and in transit (Section 4.6).

The dates in parentheses represent the earliest (or latest) date available for the selected sites, analyses, and analytes. Pressing the button to the left of each input box will copy this date into the box. You may enter alternate dates using the keyboard. All dates should be entered as month/day/year (e.g., 5/11/00). You may specify criteria for one or more of the date fields to select the sampling period. If all of the boxes are left blank, then all records (from the first sampling event to the most recent) will be selected.

The five buttons on this form include the following:
- "->,"which copies the date in parenthesis into the input box.
- "Exit Form, "which stops the filtering process.
- "Start Over," which clears all of the entries and returns you to the Select Site(s) form (Figure 4-3).
- "Next –>>," which indicates that your choices are complete and moves you to the Select Sample Type(s) form (Figure 4-7).
- "<<– Back," which returns you to the Select Analyte(s) form (Figure 4-5).

4.6   Selecting Sample Types

The next form to be displayed is the Select Sample Type(s) form (Figure 4-7).
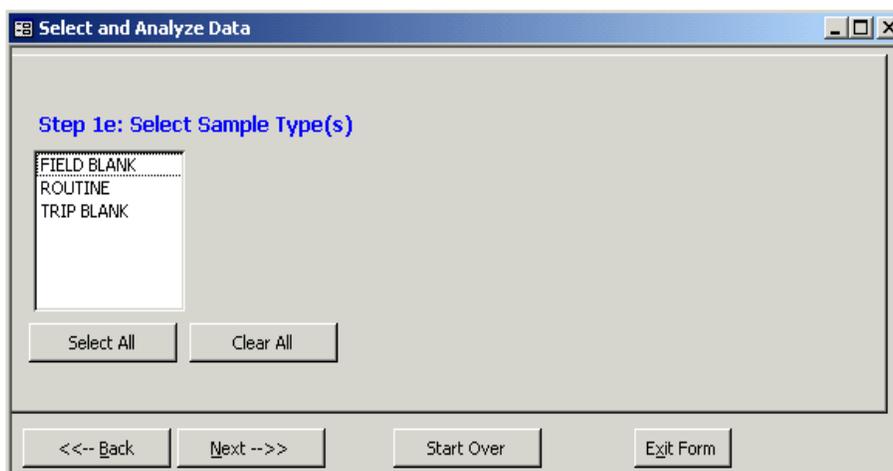
Figure 4-7. Select Sample Type(s) form.

This form displays a list of up to four sample types.
- A "FIELD BLANK" is a sampling event. During this event, filters are sent to the site, placed in a functioning air sampler for a few minutes, and returned to the laboratory for analysis. A "FIELD BLANK" measures the background analyte levels on the filters from normal shipping, loading, unloading, and analysis procedures.
- A "ROUTINE" sample is collected during a standard sampling event. During this event, filters are sent to the site, placed in a functioning air sampler for 24 hours, and returned to the laboratory for analysis.

- A "TRIP BLANK" is when the filters are sent to a site and returned to the laboratory. These filters are never placed in the air sampler. This blank is designed to measure the background levels on the filters from shipping and analysis procedures.

- An "UNSAMPLED BLANK" is a set of filters that was originally scheduled as a "ROUTINE" sampling event, but was not actually sampled. Because it is unclear if these filters were handled as a "FIELD BLANK" or a "TRIP BLANK," they receive a separate designation.

Use your mouse to click on the desired sample type(s). Multiple sample types may be selected as described in Section 4.1.

There following are the six buttons that appear on this form:

- "Select All," which selects all of the sample types in the list.

- "Clear All," which clears all of the selections on this form.

- "Exit Form," which stops the filtering process.

- "Start Over," which clears all of the entries and returns you to the Select Site(s) form (Figure 4-3).

- "Next –>>," which indicates that your choices are complete and moves you to the Select Validity Type(s) form (Figure 4-8).

- "<<– Back," which returns you to the Select Date(s) form (Figure 4-6).

4.7 Selecting Validity Types

After selecting the appropriate sample types, the Select Validity Type(s) form (Figure 4-8) will be displayed.

Figure 4-8. Select Validity Type(s) form.

This form allows you to further narrow the results within your selected sampling period. You may choose to view the following:

- "Valid Data Only," which selects only valid data. These data were collected using whole filters, functioning air samplers, and proper laboratory procedures.

- "Invalid Data Only," which selects only invalid data. The analysis laboratory has marked some of the sampling data invalid due to equipment failures, damage during shipment, or laboratory error. Some of the filters were analyzed before they were flagged as being invalid. These data points will be displayed in the final data tables. Other samples were flagged before analysis. These data will appear as null values in the final data tables.

- "All Data," which selects the valid and invalid data for the sampling period.

Please note that the contract laboratory determines the validity that's discussed in this section. As the data reviewer, you will have a chance to omit certain sampling events, analyses, and analytes from the final data. Please see Section 6.1 for details.

This form has the following four buttons:

- "Exit Form," which stops the filtering process.

- "Start Over," which clears all of the entries and returns you to the Select Site(s) form (Figure 4-3).

- "Next –>>," which indicates that your choices are complete and moves you to the Comments form (Figure 4-9).

- "<<– Back," which returns you to the Select Sample Type(s) form (Figure 4-7).

4.8   Selecting User Comments

The next form is the Comments form (Figure 4-9).

Figure 4-9. Comments form.

Later in this program (Section 6.1), you will be given the opportunity to view, edit, invalidate, and comment on the data from your sites. This form allows you to filter data based on your comments and other edits previously added.

This form contains the following four buttons:

- "Exit Form," which stops the filtering process.

- "Start Over," which clears all of the entries and returns you to the Select Site(s) form (Figure 4-3).

- "Next –>>," which indicates that your choices are complete and moves you to the Perform Data Analysis form (Figure 5-1).

- "<<– Back," which returns you to the Select Validity Type(s) form (Figure 4-8).

4.8.1   Comments Bar

The Comments bar (Figure 4-10) allows you to filter data based on comments you have previously entered about the sampling events at your sites.

Figure 4-10. Comments bar.

There are many ways to use the Comments bar to filter the data from your sites, including the following:

- **Straight Text Entry**. You may filter the data to search for a specific word in the comments section. For example, if you want to examine all data collected near forest fires, you might enter "fire" in the Comments box. This would filter the data for all records containing the word "fire."

- **Wildcards (*)**. You may use this feature to search for all records containing a specific grouping of letters in the comments section. For example, entering "fire*" would filter all records containing words that start with "fire," such as "fire," "fires," and "fireman." Wildcards may be used before, after, or in the middle of a group of letters.

- **Single-Letter Wildcards (?).** This feature will filter the data for comments containing a specific grouping of letters and one unknown letter. For example, entering "fire?" would reveal all records containing "fires," and "fired," but not those records containing "fireman" or "fire."

- **Excluding Specific Records and Records Without Comments**. To exclude the records containing a specific word, enter the phrase "not like" before the word in the Comments bar. For example, to exclude all of the records that contain the word "fire," enter the phrase "not like fire" on the Comments bar. Please note that this technique will exclude all records without comments. Wildcards may also be used to exclude records containing a specific grouping of letters.

- **Excluding All Comments**. You may exclude all records with comments, by entering "is null" in the Comments bar. This will select all records that do not have comments.

- **Excluding Specific Records**. This technique allows you to exclude the records with specific comments and include the records without comments. Enter the phrase "not like," followed by the comment to be excluded, and enter "or is null" after the word. For example, to exclude records containing the word "fireman," type "not like fireman or is null" on the Comments bar. This will select all of the records (with and without comments) that do not contain the word "fireman." Wildcards may be used.

- **Including Only Records With Comments**. To filter only those records with comments, you may enter a "*" into the comments bar or type in the phrase "is not null."

- **All Records**. To select all records from your site, regardless of the existence of comments, leave the Comments bar blank.

4.8.2   Omissions

The Omission box (Figure 4-11) is on the left side of the Comments form.
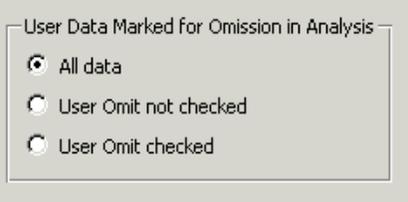
Figure 4-11. Omission box.

In the Add/Edit Data section (Section 6.1), you can omit data from certain sampling events. This box allows you to filter the data based on the status of the omission box. You may include the following:

- All data
- User Omit not checked (filters for data you have not omitted)
- User Omit checked (filters for data that you have omitted).

[This page intentionally left blank.]

5.0 Data Analysis Overview

The main purpose of the SDVAT is to facilitate state and local agencies with Level 2 and Level 3 data reviews. Such reviews involve evaluating the following quality indicators:

- Data completeness
- Time–series analyses
- Mass concentration reconstruction
- Species distribution
- Spatial and temporal variations
- Integration of local meteorological data.

The data filtering screens (Section 4) allow filtering of data for the data analysis tools, which are described in this section, and various data utilities, which are described in Section 6. These tools allow you to select date, numeric, and non-numeric data; however, only numeric data and numeric data paired with dates (for time–series analyses) can be used with the data analysis tools described in this section.

For the purpose of data analysis, it is important to note that sampling data are frequently missing their actual start dates. It is possible that sites failed to supply this information or it was not appropriate for the sample type, such as for field and trip blanks. To prevent problems when combining data that have a start date with data that are missing a start date, all the data analysis tools use an event date selected as the following:

- If the start date is available (routine samples only), it is used as the date.
- Otherwise, the scheduled date is used as the date.

Sections 5.1 through 5.6 describe how the SDVAT can be used to perform analyses for each of the indicators listed above. These analyses may be accessed though the Perform Data Analysis form (Figure 5-1). This form is accessed from the last step in filtering data (Section 4.8).

Figure 5-1. Perform Data Analysis form.

The buttons on the left of this form allow access to the analysis tools and will be discussed in this section. The buttons on the right allow access to various Data Utilities that are discussed in Section 6. The following three navigation buttons are at the bottom of this form:

- <<– Back," which return to the Comments form (Figure 4-9) for changing selection criteria.

- "Start Over," which clears all of the filtering criteria and returns to the Select Site(s) form (Figure 4-3).

- "Exit Form," which quits the filtering and analysis processes.

5.1   Data Completeness Analysis

Data completeness is the ratio of valid data to total data. This is used as an indicator of data capture and is separately calculated for each analyte/parameter. It is expressed as "percent valid," and is calculated as the following:

$$Percent\,valid = \frac{Number\ of\ valid\ results}{Number\ of\ total\ results} \times 100\%$$

Where

"Valid" data has its [OVERALL_INVALID] field equal to FALSE or NULL

Please note that this definition of data completeness is slightly different from the Federal Reference Method (FRM) completeness criterion, which is based on the ratio of valid samples to scheduled samples. Because we normally provide null data values for events that were scheduled, but not completed, these two definitions are the same if data for all scheduled events were imported.

To start this analysis, use your mouse to click on the button labeled "Data Completeness" on the Perform Data Analysis form (Figure 5-1). For routine samples, $PM_{2.5}$ concentration ($\mu g/m^3$) is used in the calculation. For non-routine samples (e.g., field or laboratory blanks), gravimetric mass (μg) is used.

5.1.1   Data Selection for Data Completeness

Selection criteria for data completeness analysis are selected using the data filtering forms (described in Section 4). Note that not all criteria are applicable for this analysis; Table 5-1 indicates which criteria may be used.

Table 5-1. Criteria for the Data Completeness Analysis

Selection Criteria	Applicable to Analysis
Site(s)	✓
Analysis/Analyses	✓
Analyte(s)	✓
Date(s)	✓
Sample Type	✓
Validity Type	Valid and total samples counted regardless of validity type specified
With Comment (Filter by user-specified comment)	✓
Analysis Omit (Filter by user-specified flag)	✓

5.1.2 Output for Data Completeness

Spreadsheets and bar charts are generated for the data completeness analysis. A single workbook is created that contains separate worksheet pages and charts for each location (Location Name/AIRS Code/POC), analysis, channel, and sample type.

The resulting spreadsheet(s) will contain the following outputs:

- Location Name
- AIRS Code
- POC
- Sample Type
- Channel Name
- Analysis
- Analyte
- Total Count
- Valid Count
- Percent Valid.

Once you have clicked on the "Data Completeness" button, an Explorer window will appear. Use the Explorer window to find the folder in which you want to save the workbook. Name the file, and use your mouse to click on the "Save" button. When the function is complete, the status bar in the bottom left of your screen will read "Form View." Use your Explorer and MS Excel programs to access the workbook. Figure 5-2 shows an example of the bar chart generated by the selected data.
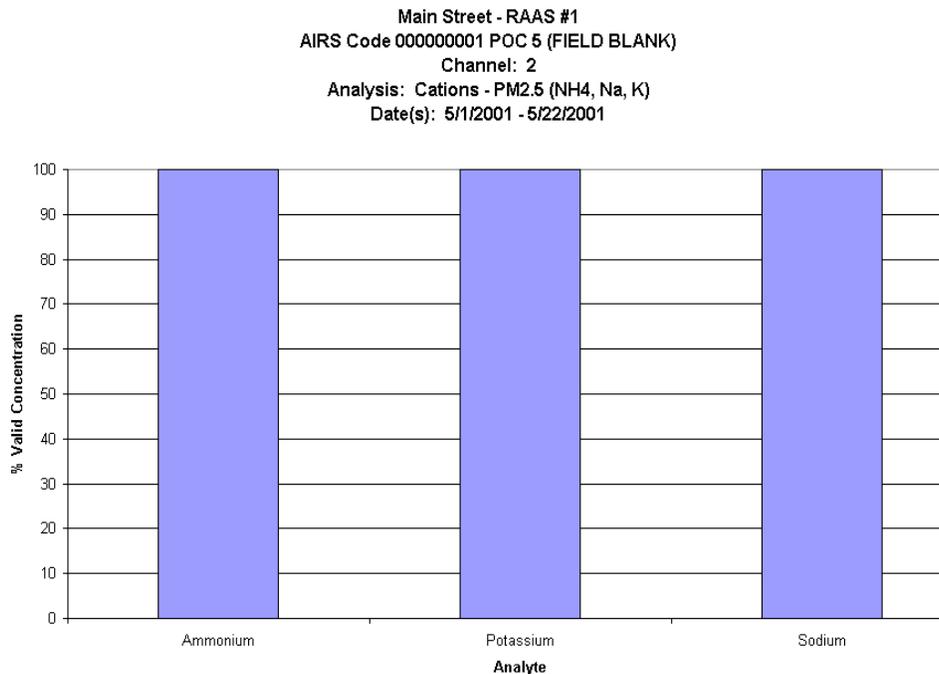

Figure 5-2. Bar chart generated by the data completeness analysis.

Please note that although automated scaling of the x-axis is employed by the software, it does not allow font sizes below 8 points. Users may want to manually set the x-axis font size in the resulting MS Excel spreadsheet(s) for the "Trace Elements" analysis to permit viewing of all labels. We recommend a 7-point font size.

5.2   Time–Series Analysis

Time–series analyses are achieved by plotting the mass, concentration, or other measured value(s) of an analyte versus time. The SDVAT allows the user to choose one of the following formats:

- XY (Scatter) Plot (Figure 5-4). Data points are plotted on the x- and y-axes, showing relationships among the numeric values and possible clustering of data. Data points are connected by lines.

- Stacked Column Plot (Figure 5-5). Data is plotted in columns, with each new series stacked on top of the previous. This shows the relationship of individual items to the whole.

To begin the time–series analysis, use your mouse to click on the "Time Series" button on the Perform Data Analysis form (Figure 5-1).

No additional calculations are performed with the data used for the time–series charts. The $PM_{2.5}$ concentration (for routine samples) or mass (for other samples) is plotted as it is reported in the database. The output worksheet contains a separate column for each analyte in the analysis and a separate row for each sampling event.

5.2.1   Data Selection for Time Series

Selection criteria for time–series analyses are selected using the data filtering forms (described in Section 4). Note that not all criteria are applicable for this analysis; Table 5-2 indicates which criteria may be used.

Table 5-2. Criteria for the Time–Series Analysis

Selection Criteria	Applicable to Analysis
Site(s)	✓
Analysis/Analyses	✓
Analyte(s)	✓
Date(s)	✓
Sample Type	✓
Validity Type	✓
With Comment (Filter by user-specified comment)	✓
Analysis Omit (Filter by user-specified flag)	✓

5.2.2   Output for Time Series

Once the user has clicked on the "Time Series" button, the Time Series Select Plot Type form (Figure 5-3) will appear.
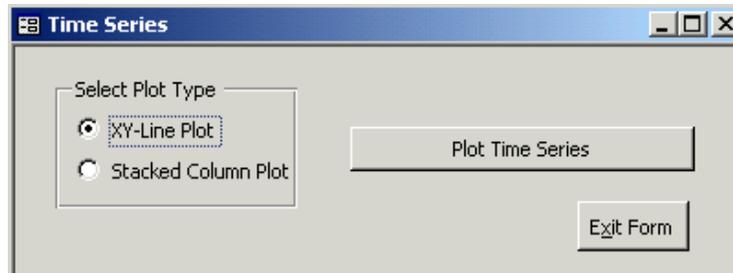
**Figure 5-3. Time Series Select Plot Type form.**

The Time Series Select Plot Type form gives the user a choice of displaying output as an xy-line plot or as a stacked column plot. Use your mouse to click in the circle next to the desired plot type. After selecting the appropriate plot type, use your mouse to click on the "Plot Time Series" button. An Explorer window will appear. Select the folder in which you wish to save the workbook, name the file, and click the "Save" button. The SDVAT will generate the spreadsheet(s) and chart(s). When "Form View" appears on the status bar in the lower left of the form, processing is complete. The workbook may be accessed using Explorer and MS Excel programs. The "Exit Form" button on the Time Series Select Plot Type form will stop this function and return you to the Perform Data Analysis form.

A single workbook is created containing separate worksheet pages and charts for each location (Location Name/AIRS Code/POC), analysis, channel, sample type, and Unit (Differs depending on sample type).

The resulting spreadsheet(s) will contain the following output:

- Location Name
- AIRS Code
- POC
- Sample Type
- Channel Name
- Analysis
- Analyte (Separate column/series created for each analyte)
- Overall Invalid
- Date
- Value Type (Concentration or mass)
- Units.

Figure 5-4 shows an example of an xy-line plot.

Main Street - RAAS #1
AIRS Code 000000001 POC 5 (ROUTINE)
Channel: 2
Analysis: Cations - PM2.5 (NH4, Na, K)
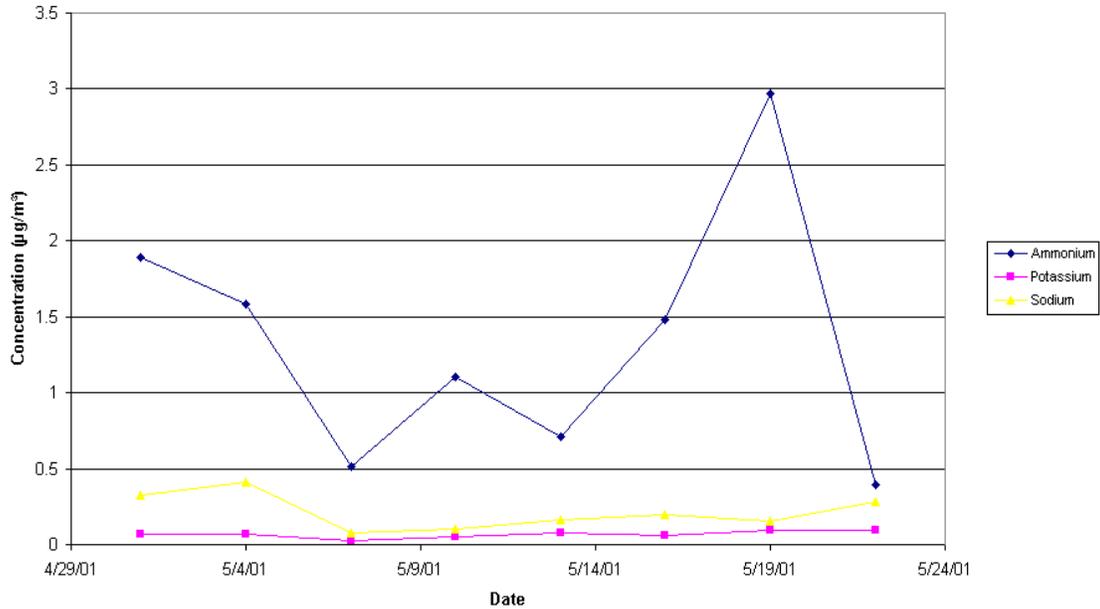
Figure 5-4. XY-line plot generated by the time–series analysis.

Figure 5-5 displays a stacked column plot.

Main Street - RAAS #1
AIRS Code 000000001 POC 5 (ROUTINE)
Channel: 2
Analysis: Cations - PM2.5 (NH4, Na, K)
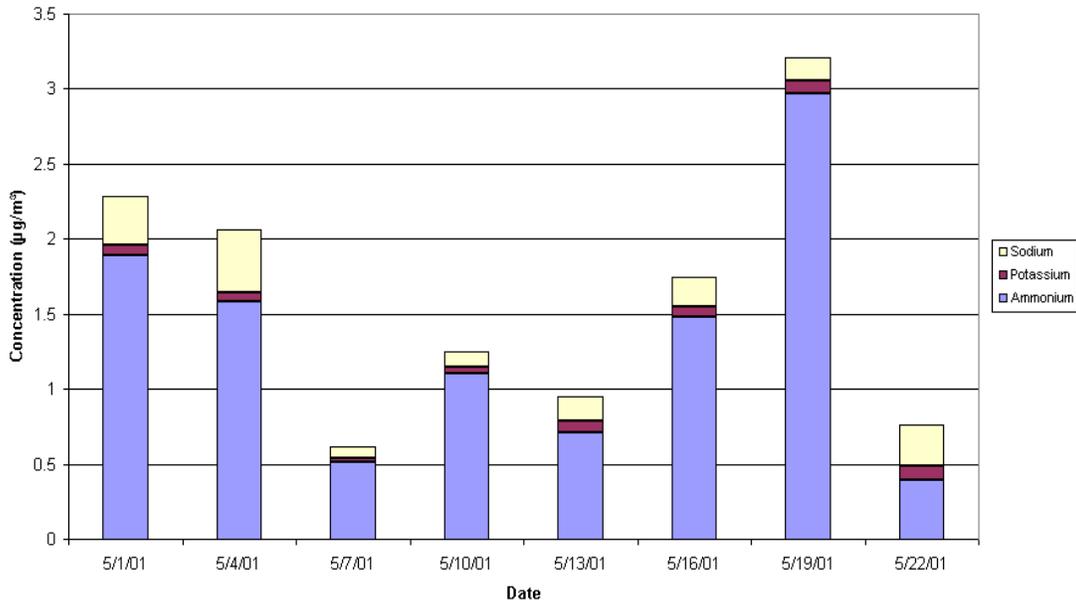
Figure 5-5. Stacked column plot generated by the time–series analysis.

Please note that to allow for a wide range of measured values without omitting extreme (including negative) values from the chart, an automatic scaling of the y-axis is used. Users may want to manually set the y-axis ranges in the resulting MS Excel spreadsheets to the same scale to permit better comparisons between sites.

5.3   Mass Concentration Reconstruction Analysis

Mass concentration reconstruction compares the total $PM_{2.5}$ gravimetric results to the total results from carbon, ion chromatography (IC), and X-ray fluorescence (XRF) analyses (excluding sodium [S], ammonium [Na], and potassium [K], which are also included in the IC results). Please note that concentrations are used rather than masses because samplers have different flow rates (and thus different collection rates) in channels used for different analyses. Mass concentration reconstruction is only performed for routine samples (i.e., blanks are never used).

Reconstructed mass concentration (μg/m^3) is calculated as follows:

$$\text{Re} \, constructed \; Mass \, Concentration = \frac{\sum Anions(IC) + \sum Cations(IC) + Total \, Carbon}{+ \sum Trace \, Elements(XRF), excluding \, S, Na, K}$$

Where

Anions	=	Nitrate and sulfate
Cations	=	Potassium, ammonium, and sodium
Trace Elements	=	Aluminum, antimony, arsenic, barium, bromine, cadmium, calcium, cerium, cesium, chlorine, chromium, cobalt, copper, europium, gallium, gold, hafnium, indium, iridium, iron, lanthanum, lead, magnesium, manganese, mercury, molybdenum, nickel, niobium, phosphorus, rubidium, samarium, scandium, selenium, silicon, silver, strontium, tantalum, terbium, tin, titanium, vanadium, wolfram, yttrium, zinc, and zirconium

If any one component of the calculation is missing, then the analysis routine is aborted. The user is notified with a message box.

To begin the analysis, click on the "Mass Concentration Reconstruction" button on the Perform Data Analysis form.

5.3.1   Data Selection for Mass Concentration Reconstruction

Selection criteria for mass reconstruction analysis are selected using the data filtering forms (described in Section 4). Note that not all criteria are applicable for this analysis; Table 5-3 indicates which criteria may be used.

Table 5-3. Criteria for the Mass Concentration Reconstruction Analysis

Selection Criteria	Applicable to Analysis
Site(s)	✓
Analysis/Analyses	Not applicable
Analyte(s)	Not applicable

Selection Criteria	Applicable to Analysis
Date(s)	✓
Sample Type	Routine samples only
Validity Type	✓
With Comment (Filter by user-specified comment)	✓
Analysis Omit (Filter by user-specified flag)	✓

5.3.2   Output for Mass Concentration Reconstruction

Once you have clicked on the "Mass Concentration Reconstruction" button, an Explorer window will appear. Select the folder in which you would like to save the workbook, name the file, and click the "Save" button. The SDVAT will generate the spreadsheet(s) and graph(s). When the process is complete, the words "Form View" will appear on the status bar in the bottom left of the screen.

The output for this analysis consists of an xy (scatter) plot (Figure 5-6) of gravimetric mass concentration (x-axis) versus reconstructed mass concentration (y-axis). The graph includes a linear regression trendline and displays the slope-intercept formula for the trendline, as well as the regression coefficient squared (R^2) value. The slope-intercept formula is represented by $y = mx + b$, where m is the slope, and b is the y-intercept.
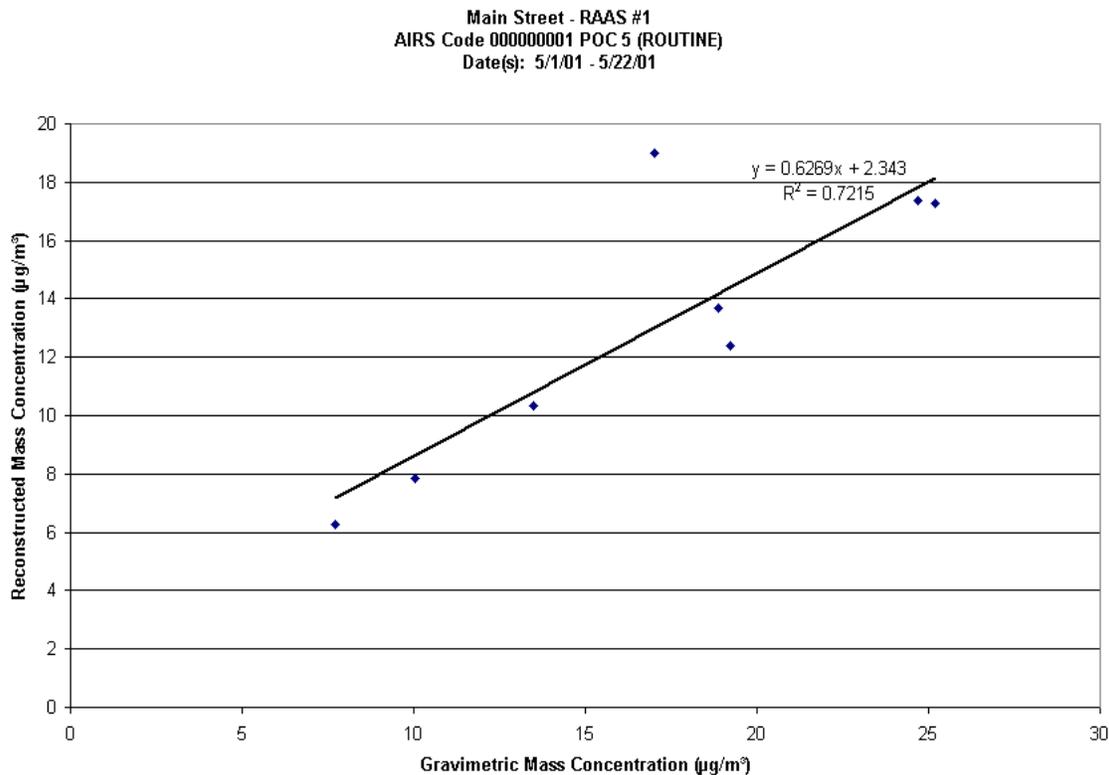
Figure 5-6. XY (scatter) plot generated by the mass concentration reconstruction analysis.

A single MS Excel workbook is created containing separate worksheet pages and charts for each location (Location Name/AIRS Code/POC) selected by the user. The accompanying spreadsheet(s) contain the following output:

- Location Name
- AIRS Code
- POC
- Date
- Gravimetric Mass Concentration (μg/m^3)
- Reconstructed Mass Concentration (μg/m^3)
- Difference (Gravimetric mass concentration minus reconstructed mass concentration, μg/m^3)
- Units
- Count of Analytes (Number of analytes used to calculate the reconstructed mass concentration).

5.4   Species Distributions Analysis

The species distributions analysis provides a graphic representation of the composition of PM$_{2.5}$ sampling results. Similar to the Mass Concentration Reconstruction function, this analysis is performed only on concentrations from routine samples.

The distribution results are comprised of the following "major elements:"

- Nitrate (Total)
- Sulfate
- Ammonium
- Organic carbon
- Elemental carbon
- Crustal component (Calculated value)
- Other (Calculated value).

Crustal Component (fine soil) is calculated as described by the Standard Aerosol-Type Equations on the Interagency Monitoring of Protected Visual Environments' (IMPROVE's) Web site, http://vista.cira.colostate.edu/improve/Tools/AerTypeEqs.htm). If any of the components are missing in the calculation, the value for Crustal Component is not calculated (i.e., the value is set to NULL). The brackets in the following equation indicate the mass concentration (μg/m^3) of the aerosol element:

$$Crustal\ Component = 2.2[Al] + 2.49[Si] + 1.63[Ca] + 2.42[Fe] + 1.94[Ti]$$

"Other" represents the amount of each sample not included by the major elements or crustal component. It is calculated by subtracting the sum of the mass concentration (μg/m^3) for each of the components from the gravimetric mass concentration (μg/m^3). If the calculated value for "Other" is negative, then a value of zero is reported. If any of the components are missing in the calculation, the value for "Other" is not calculated (i.e., set to NULL).

$$Other = Gravimetric\ Mass - \left(NO_3^- + SO_4^{2-} + NH_4^+ + OC + EC + Crustal\ Component \right)$$

5.4.1 Data Selection for Species Distribution

Selection criteria for species distribution analysis are selected using the data filtering forms (described in Section 4). Note that not all criteria are applicable for this analysis; Table 5-4 indicates which criteria may be used.

Table 5-4. Criteria for the Species Distributions Analysis

Selection Criteria	Applicable to Analysis
Site(s)	✓
Analysis/Analyses	Not applicable
Analyte(s)	Not applicable
Date(s)	✓
Sample Type	Routine samples only
Validity Type	✓
With Comment (Filter by user-specified comment)	✓
Analysis Omit (Filter by user-specified flag)	✓

5.4.2 Output for Species Distribution

Once the "Species Distributions (Major Elements)" button has been selected, the Select Plot Type form (Figure 5-7) will appear.
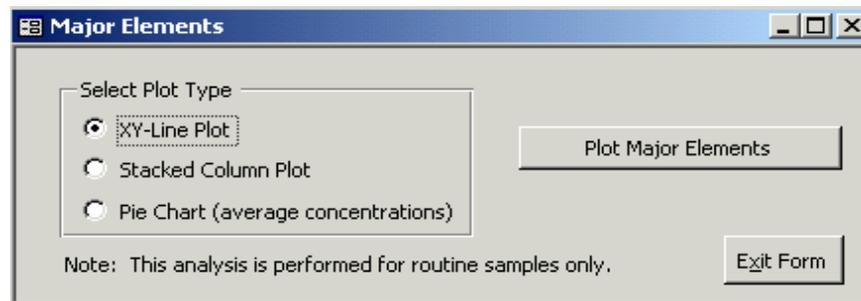
Figure 5-7. Select Plot Type form.

The Species Distributions Select Plot Type form, allows the user to choose the following output formats:

- XY-Line Plot (Figure 5-8). Data points are plotted on the x- and y-axes, showing relationships among the numeric values and possible clustering of data. Data points are connected by lines.

- Stacked Column Plot (Figure 5-9). Data are plotted in columns, with each new series stacked on top of the previous. This shows the relationship of individual items to the whole.

- Pie Chart (Figure 5-10). Averaged data are presented as a pie chart, along with percentages.

The xy-scatter charts and stacked column charts will display all available values, regardless of missing values; however, the pie chart displays averages calculated only from complete records. In other words, if any component is missing, then the entire record is omitted from the pie chart analysis.

To begin the Species Distributions Analysis, use the mouse to select the plot type, and click on the "Plot Major Elements" button. An Explorer window will appear. Select the folder in which you would like to

save the spreadsheet, name the file, and click the "Save" button. The SDVAT will generate the spreadsheet(s) and graph(s). When the process is complete, the words "Form View" will appear on the status bar in the bottom left of the screen.

To quit the analysis process and return to the Perform Data Analysis form, click on the "Exit Form" button.

A single workbook is created containing separate worksheet pages and charts for each location (Location Name/AIRS Code/POC) selected by the user. The resulting spreadsheet(s) will contain the following columns for each row:

- Location Name
- AIRS Code
- POC
- Sample Type
- Value Type
- Units
- Date
- Analyte (Separate column/series created for each of the major elements).
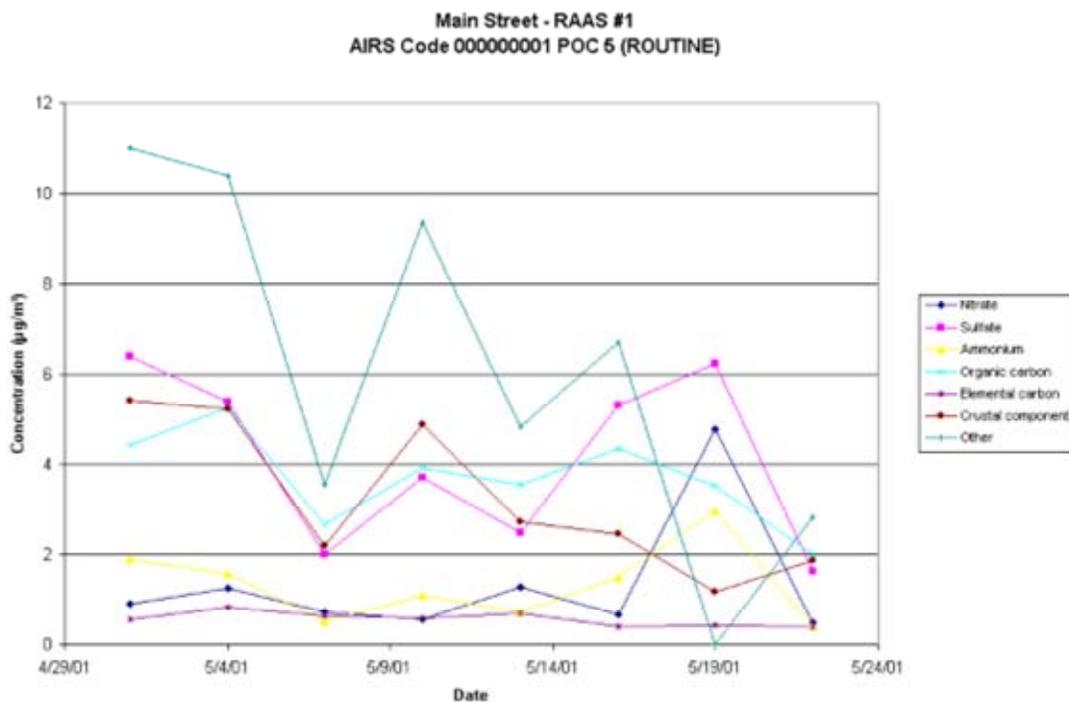
Figure 5-8. XY-line plot generated by the species distributions analysis.
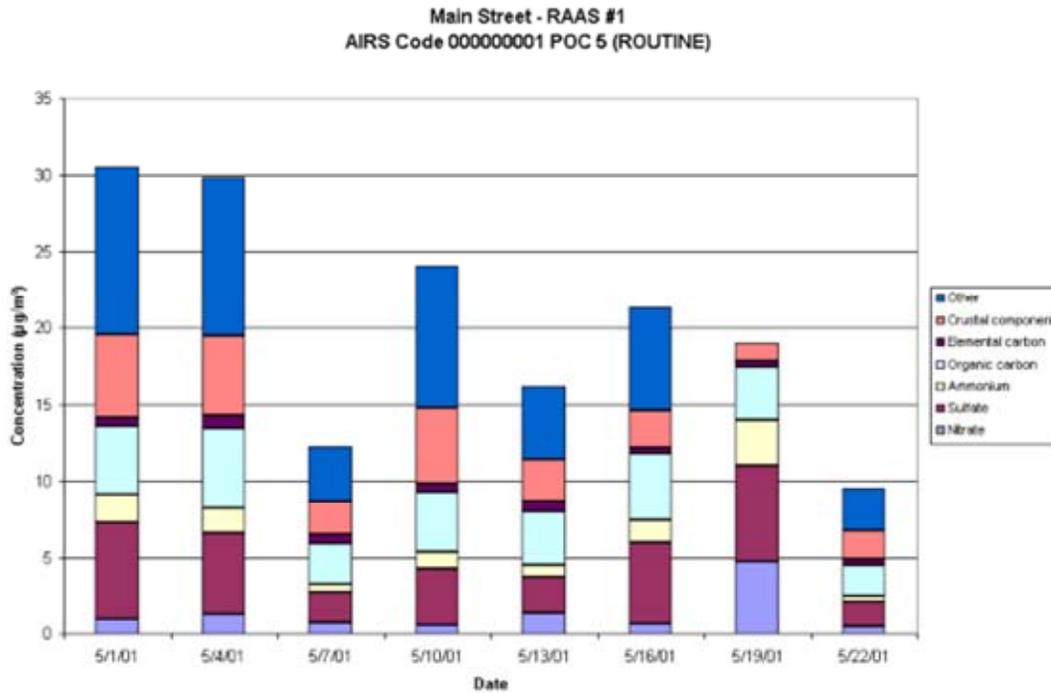
Figure 5-9. Stacked column plot generated by the species distributions analysis.
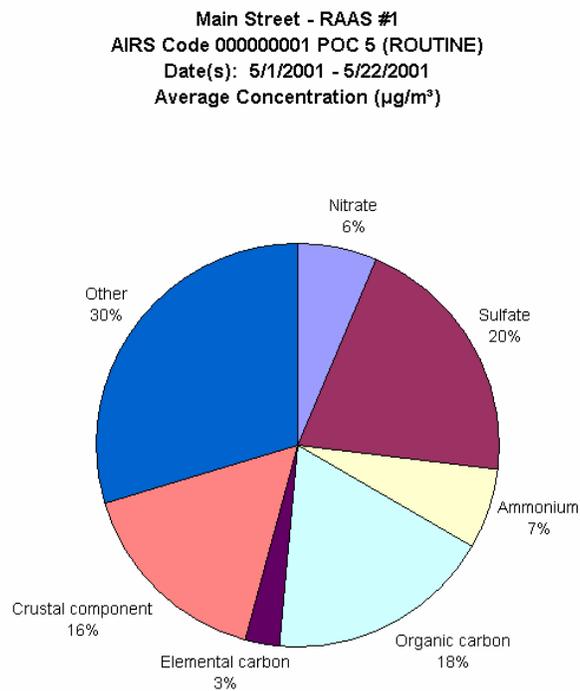
Figure 5-10. Pie chart generated by the species distributions analysis.

5.5 Spatial and Temporal Variations

Temporal variability analyses are discussed under time–series analyses (Section 5.2). The SDVAT does not currently support true spatial analysis because this would require using mapping software currently not available to this system; however, the SDVAT allows users to export filtered datasets so that they can perform spatial analyses using their own software. See Section 6.3 for instructions on data export routines.

5.6 Integration of Local Meteorological Data

The SDVAT provides several ways to support integration of local meteorological data, including the following:

- Users can import meteorological data from MS Excel spreadsheets, using the same input file structure as the speciation report data (Section 3).

- Users can enter the data by hand using a database entry form (Section 6.1.2).

[This page intentionally left blank.]

6.0 Data Utilities Overview

In addition to the ability to select and analyze data, the SDVAT also contains a number of utilities to help users manage data. These include the ability to edit, delete, export, and add comments to the filtered data. These tasks are accessible through the Perform Data Analysis form (Figure 6-1).

Figure 6-1. Perform Data Analysis form (Data Utilities).

The four buttons on the left of this form perform analyses on the filtered data, as discussed in Section 5. The four buttons on the right of the form allow the user to add/edit, delete, export, and modify comments to the filtered data. These buttons are discussed in Sections 6.1 through 6.4. The following three navigation buttons are on the bottom of the form:

- "<<– Back," which allows you to modify the filtering criteria by returning to the Comments form (Figure 4-9).
- "Start Over," which clears all filtering criteria and returns you to the Select Site(s) form (Figure 4-3).
- "Exit Form," which quits the filtering and analysis processes.

6.1 Add/Edit Selected Data

The Add/Edit Selected Data function allows you to view, edit, and comment on each individual record in the filtered dataset. The Site Data List form (Figure 6-2) displays a list of these records and can be accessed through the "Add/Edit Selected Data" button on the Perform Data Analysis form.

Figure 6-2. Site Data List form.

Abbreviated records are presented horizontally on the Site Data List form. Additional abbreviated records may be accessed using the scroll bar on the right of the form. The complete records (View Site Data by Record form in Figure 6-3) may be viewed by using your mouse to click on the "View Data" button on the right side of each record. To return to the Perform Data Analysis form, click on the "Close Form" button.

Each abbreviated record contains the following:

- Location Name
- Sampling Chain of custody (COC) ID
- AIRS Code
- POC
- Sample Type
- Sample Volume
- Result Type
- Scheduled Date
- Start Date
- Sampling Configuration
- Channel Name
- SDVAT User Comment
- Omit for Analysis Box
- Analysis
- Analyte
- Analyte Mass
- Numeric Value

- Date Value
- Overall Invalid check box
- View Data button

Figure 6-3. View Site Data by Record form.

In addition to the information from the abbreviated record, the complete record in Figure 6-3 (View Site Data by Record form) shows the following:

- Delivery Batch
- Delivery Order
- Report Date
- Resubmittal check box
- Sampling Module ID
- Sampler Part
- Text Value
- All Event, Shipping, Flow, Module Disassembly, and Analysis Flags
- Validity Status
- Validation check boxes
- AIRS Codes
- Non-detect check box
- Uncertainty (mass and concentration)
- Detection Limit (minimum detection limit [MDL], mass and concentration)

To return to the Site Data List form, click on the "Close Form" button in the top right of the complete record.

6.1.1   Modify Records

The existing records may be modified at any time from the abbreviated or complete record forms. To change the data in one of the boxes or to add a comment in the SDVAT User Comment box, use your mouse to move the cursor into the box. Then use the keyboard to delete the old information and enter the changes. Please refer to Section 6.4 for instructions on modifying all of the comments for a subset of data. To check a box (e.g., the Non-detect box), use the mouse to click in the box. The check will disappear if you click in the box again.

6.1.2   Add Records

To add a new record, go to the Site Data List form (Figure 6-2) and use the mouse to click on the "Add New Record" button. This will bring you to the next blank record. Use your mouse and the keyboard to enter the new data. Use the "View Data" button to open the "View Site Data by Record" form (Figure 6-3), which shows additional fields for this record.

6.1.3   Delete Records

To delete a record, use the mouse to click on the bar (Figure 6-4) on the left of the Location Name on the Site Data List form. This will select, or highlight, the record. Press the "Delete" key. The computer will ask you if you want to permanently delete one or more records. Click the "Yes" button. You may use the keyboard and mouse as described in Section 4.1 to select multiple records for deletion. Please refer to Section 6.2 for additional techniques for deleting records.
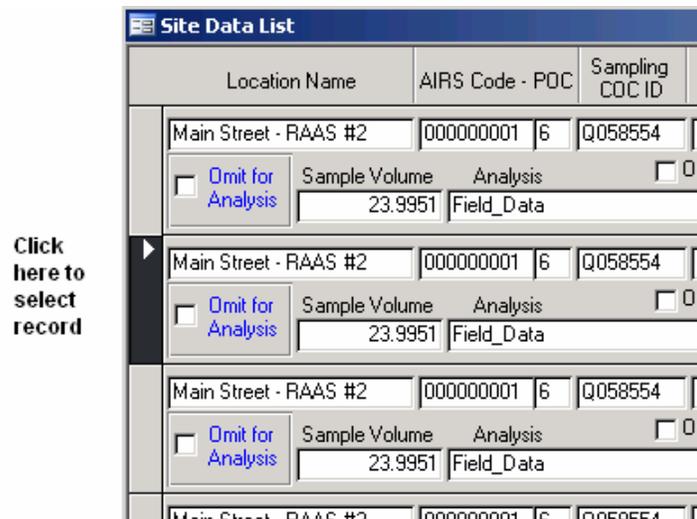

Figure 6-4. Select record.

6.2   Delete Selected Data

The "Delete Selected Data" function (Figure 6-1) allows the user to delete an entire subset of data from the database. Use the filtering process to isolate the unwanted data, then use your mouse to click on the "Delete Selected Data" button. A new form (Confirm Delete message in Figure 6-5) will appear and ask you if you want to permanently remove the selected data. Clicking on the "Yes" button will permanently

delete all of the selected data from the database. If you click on the "No" or "Cancel" button, the data will remain in the database and you will return to the Site Data List form.
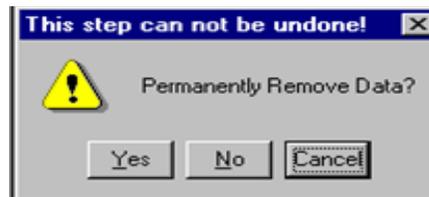
Figure 6-5. Confirm Delete message.

6.3   Export Selected Data

To generate a spreadsheet of data from the filtered, edited subset, use your mouse to click on the "Export Selected Data to SS" button on the Perform Data Analysis form. An Explorer window (Figure 6-6) will appear on your screen.
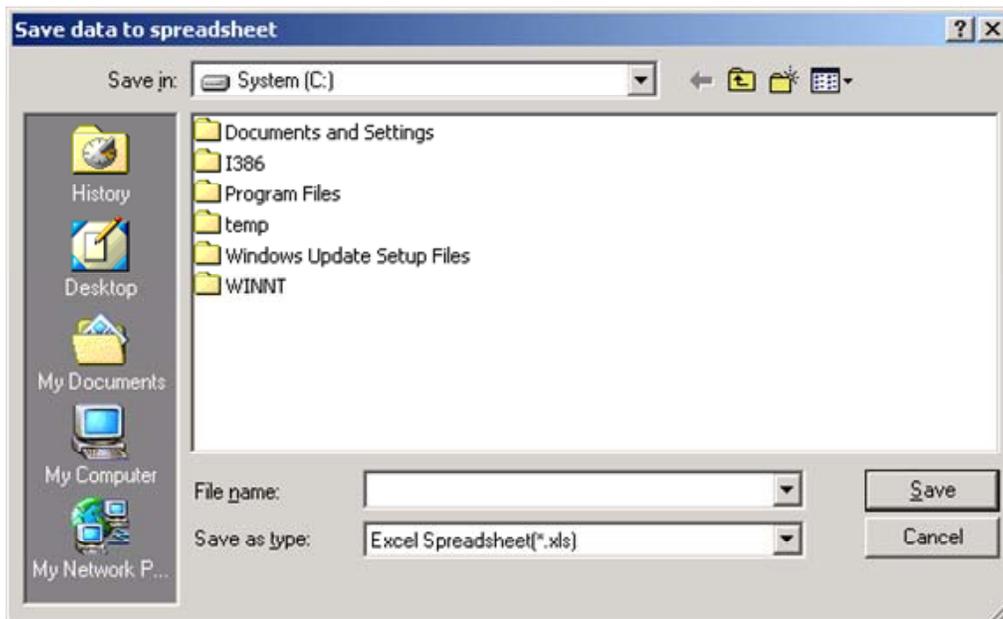
Figure 6-6. Explorer window (export spreadsheet to file).

Choose the folder in which you want to save your spreadsheet, name the file, and use your mouse to click on the "Save" button. When the status bar in the bottom left of your screen reads "Form View," the export process is complete and the spreadsheet is now saved onto your hard drive. You may use MS Excel programs to access the spreadsheet or re-import it into the SDVAT using the "Import Data" functions (Section 3).

6.4   Modify Comments

The "Modify Comments" function allows you to modify or replace all of the SDVAT User Comment fields in a subset of records. After using your mouse to click on the "Modify Comments" button, the Edit Comments form will appear (Figure 6-7).
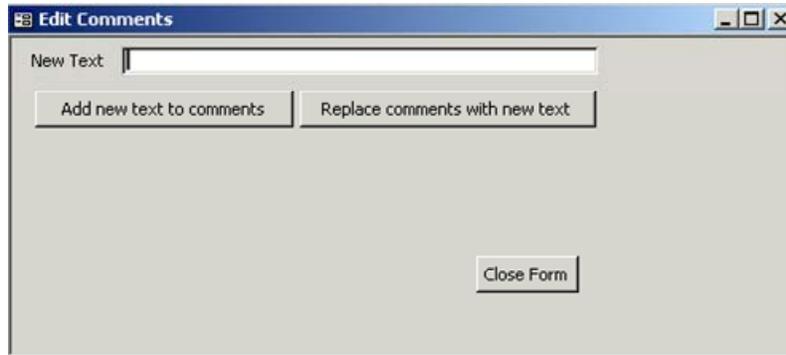
Figure 6-7. Edit Comments form.

This form offers you the following four choices:

- The "Close Form" button will cancel the request and return you to the Perform Data Analysis form.

- To add additional information to any comments already in the SDVAT User Comment field, type the new information into the "New Text" box and use the mouse to click on the "Add new text to comments" button. When the process is complete, the Addition Completed message (Figure 6-8) will appear. The number on the form will vary based on the number of records in your subset.
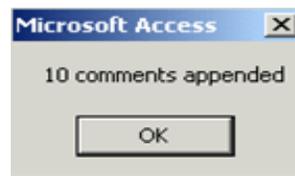
Figure 6-8. Addition Completed message.

- To clear the existing comments in the SDVAT User Comment field and replace these comments with a new one, enter the new comment in the "New Text" box and use your mouse to click on the "Replace comments with new text" button. When the process is complete, Figure 6-9 (Modification Completed message) will appear.

Figure 6-9. Modification Completed message.

- You may also use the "Replace Comments" function to clear all of the existing comments in your subset. Leave the "New Text" box empty, and click on the "Replace comments with new text" button. A Permanently Remove Data message (Figure 6-10) will appear.

Figure 6-10. Permanently Remove Data message.

- Use your mouse to click on the "Yes" button. The Modification Completed message (Figure 6-9) will appear when the process is complete.

[This page intentionally left blank.]