EPA-CMB8.2 Users Manual

EPA-452/R-04-011
December 2004

EPA-CMB8.2 Users Manual

By:

C. Thomas Coulter
Air Quality Modeling Group
Emissions, Monitoring & Analysis Division
Office of Air Quality Planning & Standards
Research Triangle Park, NC 27711

US. Environmental Protection Agency
Office of Air Quality Planning & Standards
Emissions, Monitoring & Analysis Division
Air Quality Modeling Group

ACKNOWLEDGMENTS

CMB8 was originally developed as part of Contracts 5D1607NAEX and 5D2724NAEX between EPA's National Exposure Research Laboratory (NERL) and the Desert Research Institute (DRI) of the University and Community College System of Nevada. The manual was originally drafted as part of Contract 5D1808NAEX with EPA's Office of Air Quality Planning & Standards (OAQPS). The Project Officers were C. Thomas Coulter[1] and Charles W. Lewis[2]. Substantial contributions to initial development of the CMB8 model were made by DRI staff members John G. Watson, Norman F. Robinson, Judith C. Chow, Eric M. Fujita, and Douglas H. Lowenthal. Teri L. Conner[2] and Robert D. Willis[3] also made important contributions to CMB8's initial development.

Tom Coulter and Donna Kenski (then at EPA Region 5) spent considerable time troubleshooting CMB8 in May 1999. An effort was initiated by EPA in 1999 to repair and redesign CMB8. In September 1999, Pacific Environmental Services (PES) began work to repair and enhance CMB8 under contract 9D-1844-NTSA, an effort that led to the redesigned CMB8.2. CMB8.2 was developed principally by Vickie Kriegsman and Robert Wagoner PES (now MACTEC), Research Triangle Park, NC. Additional issues and operational problems were discovered and documented in early 2000. Subsequently, Ideal Software, Inc. (Apex, NC) was retained under contracts 2D-6228-NTSA (July 2002) and 4D-6389-NTSA (July 2004). Ideal Software engineers successfully compiled the model's C++/Fortran-based DLL (the first since DRI to do so), and resolved many behavioral problems with the model, now called EPA-CMB8.2. Ideal's John Scalco created many of the runtime features EPA-CMB8.2 now exhibits through its (Delphi) enhanced User Interface. Tom Coulter reorganized and refactored C++ and Fortran code for the DLL, reworked error handling, and removed obsolete variables and subroutines. With careful assistance from Ed Anderson of Science Applications International Corp., Tom updated the LINPACK library to LAPACK v3.0 and modified the Fortran code to accommodate it. Tom also served as Contracting Officer Representative for these contracts, assisted with the Delphi

[1] Office of Air Quality Planning & Standards, Research Triangle Park, NC 27711

[2] National Exposure Research Laboratory, Office of Research & Development, Research Triangle Park, NC 27711

[3] ManTech Environmental Technology, Inc., Research Triangle Park, NC 27709

reprogramming, and rewrote the manual. Tom's diligent work resulted in a level of documentation and disclosure unknown in the history of the CMB model.

EPA-CMB8.2 has its foundation in the previous versions that span two decades. OAQPS' involvement in supporting CMB development, dating back to the mid-80s, was initiated under the guidance of Tom Pace. The present author is indebted to the many contributors to the earlier work.

During September - December 2004, EPA-CMB8.2 and its documentation (this Users Manual and its companion *Protocol for Applying and Validating the CMB Model for* $PM_{2.5}$ *and VOC*) were subjected to scientific peer review under EPA Contract 4D-6097-NTSX. EPA is grateful for the Peer Review panel - Jamie Schauer, Donna Kenski, Robert Willis - and its diligent work and helpful suggestions for both the model and its documentation. Further suggestions from users are welcome, and may be directed to Tom Coulter at
Coulter.Tom@epa.gov.

DISCLAIMER

This manual was reviewed by EPA for publication. The information presented here does not necessarily express the views or policies of EPA. Any mention of trade names or commercial hardware and software in this document does not constitute endorsement of these products. No explicit or implied warranties are given for the software and data sets described in this document.

Abstract

The Chemical Mass Balance (CMB) air quality model is one of several receptor models that have been applied to air resources management. EPA-CMB8.2 incorporates the upgrade features that CMB8 has over CMB7, but also corrects errors/problems identified with CMB8 and adds enhancements for a more robust and user-friendly system. EPA-CMB8.2 is a 32-bit (Windows® 9x and higher) version of CMB modeling software that substantially facilitates the estimation of source contributions to speciated PM_{10} (particles with aerodynamic diameters nominally less than 10µm), $PM_{2.5}$ (particles with aerodynamic diameters nominally less than 2.5µm), and Volatile Organic Compounds (VOC) data sets. EPA-CMB8.2 features: (1) full use of Windows® (32-bit) for file access/management, (2) a *tabbed page interface* that eases the necessary progression for doing a CMB calculation, (3) multiple, indexed arrays for selecting fitting sources and species, (4) versatile display capability for ambient data and source profiles, (5) mouse-overs and on-line help screens, (6) increased attention to volatile organic compounds (VOC) applications, (7) correction of some flaws in the previous version (CMB7), (8) flexible options for input/output data formats, (9) addition of a more accurate least squares computational algorithm, (10) upgraded linear algebra library, (11) a new treatment of source collinearity, and (12) choice of criteria for determining *best fit*.

This manual introduces EPA-CMB8.2 and its development history. It describes hardware and software requirements and shows how to install EPA-CMB8.2 on a personal computer. It explains EPA-CMB8.2 menu options and input and output file formats. The manual provides a step-by step tutorial of EPA-CMB8.2 operations using an example data set provided with the model. Performance measures are briefly described, though their use in practical applications is deferred to a separate application and validation protocol. A comprehensive list of references is included for those desiring more information about CMB, its utility and applications.

Table of Contents

List of Figures

1. INTRODUCTION

The Chemical Mass Balance (CMB) air quality model is one of several receptor models that have been applied to air resources management. Receptor models use the chemical and physical characteristics of gases and particles measured at source and receptor to both identify the presence of and to quantify source contributions to receptor concentrations. Receptor models are generally contrasted with dispersion models that use pollutant emissions rate estimates, meteorological transport, and chemical transformation mechanisms to estimate the contribution of each source to receptor concentrations. The two types of models are complementary, with each type having strengths that compensate for the weaknesses of the other.

This manual describes how to operate EPA-CMB8.2 modeling software to calculate source contributions to ambient PM_{10} (particles with aerodynamic diameters nominally less than 10µm), $PM_{2.5}$ (particles with aerodynamic diameters nominally less than 2.5µm), and volatile organic compounds (VOC).

A separate applications and validation protocol (EPA, 2004) describes how to apply EPA-CMB8.2 to specific situations and how to evaluate its outputs. Several review articles, books, and conference proceedings provide additional information about the CMB and other receptor models (Chow *et al.*, 1993; Gordon, 1980, 1988; Hopke and Dattner, 1982 ; Hopke, 1985, 1991; Pace, 1986, 1991; Stevens and Pace, 1984; Watson, 1979, 1984; Watson *et al.*, 1989, 1990, 1991).

1.1 EPA-CMB8.2 Features

EPA-CMB8.2 replaces CMB7 (EPA, 1990; Watson *et al.*, 1990) as a more convenient method of estimating contributions from different sources to ambient chemical concentrations (Coulter and Scalco, 2005). EPA-CMB8.2 returns the same results as CMB7, but it operates in a Windows®-based environment and accepts inputs and creates outputs in a wider variety of formats than CMB7. The major EPA-CMB8.2 enhancements are:

Windows®-based, event-driven operations: EPA-CMB8.2 makes full use of Windows® (32-bit) features, including a *tabbed interface* that facilitates the necessary progression for doing a CMB calculation. Commands may be executed with hot-keys or toolbar buttons, and features are described via *mouse-overs* and context sensitive on-line help screens. EPA-CMB8.2 also offers flexible options for input/output data formats. Input formats are compatible with output files from EPA's source profile library SPECIATE (*www.epa.gov/ttn/chief/*).

Multiple arrays for fitting species and fitting sources: Up to ten indexed arrays of fitting source profiles and fitting species may be specified in input data selection files. Different arrays can be selected during EPA-CMB8.2 operation. Upon session exit, an option is provided to conveniently save (update) or rename selection files to reflect arrays that are added, modified or deleted during the session.

Britt-Luecke algorithm: A general solution to the least squares estimation problem that includes uncertainty in all the variables (i.e., the source compositions as well as the ambient concentrations) is available. While an approximation to the Britt-Luecke iteration scheme (Britt and Luecke, 1973) was used in CMB7, exercise of Britt-Luecke algorithm option in EPA-CMB8.2 allows solution without approximation.

Improved collinearity diagnostics: The uncertainty/similarity clusters have been replaced with a singular value decomposition eligible space treatment that allows the user to define an acceptable error and an acceptable collinearity among weighted source profiles.

Better handling of VOC applications: EPA-CMB8.2 gives the user control in adjusting collinearity parameters which in CMB7 are "hard-wired" and not necessarily optimum for every application. Values for these parameters were chosen in CMB7 to be compatible with characteristics of particulate mass measurements, but they may not be as well suited to CMB solutions involving VOC.

Search for best fit: Using a user-selected weighted optimization of performance measures, EPA-CMB8.2 can systematically check up to 10 possible *paired* combinations of fitting species and sources arrays as it searches for a maximum of an empirical composite measure. The best fit arrays are then indicated in their respective windows.

User-selected preferences: In EPA-CMB8.2, the user may set options for maximum iterations for convergence, eligible space tolerances, positions of decimal points in output, receptor concentration units, special calculation alternatives, and performance measure weights for use in Best Fit mode.

Negative source contributions: EPA-CMB8.2 calculations can be set to eliminate negative contributions.

Improved memory management: EPA-CMB8.2 memory is limited only by the available RAM on the host computer, not by pre-set memory limitations.

Upgraded linear algebra library: The linear algebra library that EPA-CMB8.2 uses to perform its effective variance, least-squares regressions has been updated with LAPACK v3.0.

Versatile graphic display capability: For ambient samples and source profiles, bar charts for species concentrations can be displayed within EPA-CMB8.2, which is useful for visual inspection. These can be cut from their windows and pasted into other Windows® documents.

Context-sensitive on-line help: Context-sensitive on-line help is accessible directly from the User Interface.

Flexible input and output formats: comma-separated value (CSV), xBASE (DBF), and worksheet (WKS) formats are supported as input files, in addition to the formatted, blank-delimited ASCII text files (TXT) supported by CMB7. Output files formats are ASCII and CSV (which ports nicely to Microsoft Excel®).

File handling: EPA-CMB8.2 differs from CMB7 in several ways with regard to the files used by each. EPA-CMB8.2 does not support CMB6 style ambient data and source profile data files. Control File (formerly filenames file), source profile, ambient data, and sample selection file formats differ slightly from CMB7. CMB7 source profile and ambient concentration data files can be read directly by EPA-CMB8.2, however, so backward compatibility is assured. Different Control Files can be loaded during the same session, obviating the need to terminate the application. In EPA-CMB8.2 graphical output is not provided as HPGL text files. Instead output can be printed through Windows® or copied to the clipboard for insertion into documents. Text output can also be directed to the printer, the clipboard, or a report file. A feature of EPA-CMB8.2 is that the computational machinery files (*.exe, *.dll, etc.) need not reside in the same folder as either the input or output files. This facilitates file management.

The naming structure for (*optional*) selection files has been changed to one that is more logical and intuitive - a convention that meshes with the one used for naming the *required* input data files:

	Previously	EPA-CMB8.2
Source **PR**ofile selection file:	SO*.sel	**PR***.sel
SPecies selection file:	PO*.sel	**SP***.sel
Ambient **D**ata (sample) selection file:	DS*.sel	**AD***.sel

1.2 Chemical Mass Balance Overview

The CMB receptor model (Friedlander, 1973; Cooper and Watson, 1980; Gordon, 1980, 1988; Watson, 1984; Watson *et al.*, 1984; 1990; 1991; Hidy and Venkataraman, 1996) consists of a solution to linear equations that express each receptor chemical concentration as a linear sum of products of source profile abundances and source contributions. For each run of CMB, the model fits speciated data from a specified group of sources to corresponding data from a particular receptor (sample). The source profile abundances (i.e., the mass fraction of a chemical or other property in the emissions from each source type) and the receptor concentrations, with appropriate uncertainty estimates, serve as input data to CMB. The output consists of the amount contributed by each source type represented by a profile to the total mass, as well as to each chemical species. CMB calculates values for the contributions from each source and the uncertainties of those values. CMB is applicable to multi-species data sets, the most common of which are chemically-characterized PM_{10}, $PM_{2.5}$, and Volatile Organic Compounds (VOC). The theory of CMB is described in Appendix A.

The CMB modeling procedure requires: 1) identification of the contributing source types; 2) selection of chemical species or other properties to be included in the calculation; 3) knowledge of the fraction of each of the chemical species which is contained in each source type (source profiles); 4) estimation of the uncertainty in both ambient concentrations and source profiles; and 5) solution of the chemical mass balance equations. The CMB approach is implicit in all factor analysis and multiple linear regression models that intend to quantitatively estimate source contributions (Watson, 1984). These models attempt to derive source profiles from the covariation in space and/or time of many different samples of atmospheric constituents that originate in different sources. These profiles are then used in a CMB solution to quantify source contributions to each ambient sample.

Several solution methods have been proposed for the CMB equations: 1) single unique species to represent each source (tracer solution) (Miller *et al.*, 1972); 2) linear programming solution (Hougland, 1983); 3) ordinary weighted least squares, weighting only by uncertainty of ambient measurements (Friedlander, 1973; Gartrell and Friedlander, 1975); 4) ridge regression weighted least squares (Williamson and DuBose, 1983); 5) partial least squares (Larson and Vong, 1989; Vong *et al.*, 1988); 6) neural networks (Song and Hopke, 1996); and 7) effective variance weighted least squares (Watson *et al.*, 1984).

The effective variance weighted solution is generally applied because it: 1) theoretically yields the most likely solutions to the CMB equations, providing model assumptions are met; 2) uses all available chemical measurements, not just so-called “tracer” species; 3) analytically estimates the uncertainty of the source contributions based on uncertainty of both the ambient concentrations and source profiles; and 4) gives greater influence to chemical species with lower uncertainty in both the source and receptor measurements than to species with higher uncertainty. The effective variance is a simplification of a more exact, but less practical, generalized least squares solution proposed by Britt and Luecke (1973).

CMB model assumptions are: 1) compositions of source emissions are constant over the period of ambient and source sampling; 2) chemical species do not react with each other (i.e., they add linearly); 3) all sources with a potential for contributing to the receptor have been identified and have had their emissions characterized; 4) the number of sources or source categories is less than or equal to the number of species; 5) the source profiles are linearly independent of each other; and 6) measurement uncertainties are random, uncorrelated, and normally distributed.

The degree to which these assumptions are met in applications depends to a large extent on the particle and gas properties measured at source and receptor. CMB performance is examined generically by applying analytical and randomized testing methods, and specifically for each application by following an applications and validation protocol. The six assumptions are fairly restrictive and they will never be totally complied with in actual practice. Fortunately, CMB can tolerate reasonable deviations from these assumptions, though these deviations increase the stated uncertainties of the source contribution estimates (Cheng and Hopke, 1989; Currie *et al.*, 1984; deCesar *et al.*, 1985, 1986; Dzubay *et al.*, 1984; Gordon *et al.*, 1981; Henry, 1982, 1992; Javitz and Watson, 1986; Javitz *et al.*, 1988a, 1988b; Kim and Henry,1989; Lowenthal *et al.*, 1987, 1988a, 1988b, 1988c, 1992, 1994; Watson, 1979).

The formalized protocol for CMB application and validation (EPA, 2004 is applicable to the apportionment of gaseous organic compounds and particles (Watson *et al.*, 1994a; Fujita *et al.*, 1994). This seven-step protocol: 1) determines model applicability; 2) selects a variety of profiles to represent identified contributors; 3) evaluates model outputs and performance measures; 4) identifies and evaluates deviations from model assumptions; 5) identifies and corrects model input deficiencies; 6) verifies consistency and stability of source contribution estimates; and 7) evaluates CMB results with respect to other data analysis and source assessment methods.

CMB is intended to complement rather than replace other data analysis and modeling methods. CMB helps explain observations that have been made; it does not predict ambient impacts from sources as do dispersion models. When source contributions are proportional to emissions, as they often are for PM and VOC, then a source-specific proportional rollback (Barth, 1970; Cass and McCrae, 1981; Chang and Weinstock, 1975; deNevers and Morris, 1975) is used to estimate the effects of emissions reductions. Similarly, when a secondary compound apportioned by CMB is known to be limited by a certain precursor, a proportional rollback is used on the controlling precursor. The most widespread use of CMB over the past decade has been to justify emissions reduction measures in PM_{10} non-attainment areas. More recently, CMB has been coupled with extinction efficiency receptor models (Lowenthal *et al.*, 1994; Watson and Chow, 1994) to estimate source contributions to light extinction and with aerosol equilibrium models (Watson *et al.*, 1994b) to estimate the effects of ammonia and oxides of nitrogen emissions reductions on secondary nitrates.

CMB does not explicitly treat profiles that change between source and receptor (assumption #2 above).[1] Most applications use source profiles measured at the source, with at most dilution to ambient temperatures and <1 minute of aging prior to collection to allow for condensation and rapid transformation. Profiles have been "aged" prior to submission to CMB using aerosol and gas chemistry models to simulate changes between source and receptor (Friedlander, 1981; Lin and Milford, 1994; Venkatraman and Friedlander, 1994). These models are often overly

[1] For a discussion of special approaches for treating secondary formation with EPA-CMB8.2, refer to EPA, 2004

simplified, and require additional assumptions regarding chemical mechanisms, relative transformation and deposition rates, mixing volumes, and transport times.

CMB requires species with different abundances in different source types. The consistency of a species abundance is more important than the uniqueness for source quantification. The uniqueness is useful to identify which sources to include in a CMB analysis. Combining particle and gas properties for source emissions, normalized to NMHC (non-methane hydrocarbon) or $PM_{2.5}$ mass emissions, could assist the apportionment of both VOC and $PM_{2.5}$.

New analytical methods, however, such as isotopic abundances, specific organic compounds, and single particle morphology may be used in CMB when they have been applied to source and receptor samples to more precisely differentiate among contributions from different sub-types. CMB performs tests on ambient data and source profiles that tell how well source-type contributions can be resolved from each other for different combinations of source profiles and chemical measurements.

CMB quantifies contributions from chemically distinct source-types rather than contributions from individual emitters. Sources with similar chemical and physical properties cannot be distinguished from each other by CMB. CMB model calculates source contribution estimates for each individual ambient sample. The combination of source profiles that best explains the ambient measurements may differ from one sample to the next owing to differences in emission rates (e.g., some days may have wood-stove burning bans in effect and others will not), wind directions (e.g., a downwind point source would not be expected to be contributing at an upwind sampling site), and changes in emissions compositions (e.g., different gasoline characteristics and engine performance in winter and summer may result in different profiles).

1.3 CMB Software History

The CMB receptor model was first applied by Winchester and Nifong (1971), Hidy and Friedlander (1972), and Kneip *et al.* (1973). The original applications used unique chemical species associated with each source-type, the so-called "tracer" solution. Friedlander (1973) introduced the ordinary weighted least-squares solution to the CMB equations, and this had the advantages of relaxing the constraint of a unique species in each source-type and of providing estimates of uncertainties associated with the source contributions. The ordinary weighted least squares solution was limited in that only the uncertainties of the receptor concentrations were considered; the uncertainties of the source profiles, which are typically much higher than the uncertainties of the receptor concentrations, were neglected.

The first interactive, user-oriented software for CMB was programmed in 1978 at the Oregon Graduate Center in Fortran IV on a PRIME 300 minicomputer (Watson, 1979). The PRIME 300 was limited to 3 megabytes of storage and 64 kilobytes of random access memory. CMB Versions 1 through 6 updated this original version and were subject to many of the limitations dictated by the original computing system. CMB7 was completely rewritten in a combination of the C and Fortran languages for DOS-based PCs with floating-point coprocessors, hard disk systems with tens of megabytes storage, and available memory of 640 kilobytes. CMB8 was developed but not officially released by EPA. CMB8 created a user interface for CMB7 calculations using the Borland Delphi object oriented language.

EPA-CMB8.2 incorporates the upgrade features that CMB8 has over CMB7, but also corrects errors/problems identified with CMB8 and adds enhancements for a more robust and user-friendly system. The source code, executable, and test cases are available from EPA's website (www.epa.gov/scram001).

1.4 Organization of the Users Manual

Section 1 introduces EPA-CMB8.2 and the scope of this manual. Section 2 describes hardware requirements and related files, and describes how to install EPA-CMB8.2 on a personal computer. Section 3 describes EPA-CMB8.2 key model features while Section 4 documents input and output file formats. Section 5 provides a step-by step tutorial of EPA-CMB8.2 operations using a test case from example data sets provided with the model. Performance measures are briefly described in Section 6, though their use in practical applications is deferred to the application/validation protocol (EPA, 2004). Section 7 includes a bibliography of CMB-related literature, including references cited throughout this manual.

2. SOFTWARE INSTALLATION

This section describes the hardware requirements, computer programs, and installation procedures for EPA-CMB8.2.

2.1 Hardware and Operating System

The minimum requirements for running EPA-CMB8.2 software are:

- IBM® PC compatible desktop, portable, or laptop computer with 386 processor and 16MB RAM
- Hard disk drive with 4 megabytes of storage
- Windows® 9x or higher operating system

The *recommended* hardware configuration is:

- IBM® compatible Intel Pentium® microcomputer with 64MB of RAM and 100MB storage.
- Super VGA video graphics adapter and monitor.
- Graphics capable printer.
- Windows® XP or NT 4.0 operating system.

2.2 EPA-CMB8.2 Software and Related Files

The EPA-CMB8.2 software, as well as this manual, can be retrieved from the EPA's Support Center For Regulatory Air Models (SCRAM) website:

www.epa.gov/scram001

The following files are available and can be downloaded as needed:

- ***EPA-CMB82.zip*****:** A ZIPped file that contains the EPA-CMB8.2 executable, its companion DLL file, and a help file (Section 2.3). This installation is compatible for all applications using Windows® 9x or higher.
- ***EPA-CMB82 test.zip*****:** A ZIPped file that contains all files needed for the test case described in Section 5 of this manual. Included are $PM_{2.5}$ data (ambient and source profile) from several sites in California's San Joaquin Valley Air Quality Study (SJVAQS; Chow *et al.*, 1990; 1992)

- ***EPA-CMB82 Manual.pdf*:** An Adobe Acrobat® version of this users manual. A color printer supporting PostScript fonts is recommended. Use this manual to learn EPA-CMB8.2 features and operating methods.

- ***CMB Protocol.pdf*:** An Adobe Acrobat® version of the *Protocol for Applying and Validating the CMB Model for $PM_{2.5}$ and VOC* (EPA, 2004). A color printer supporting PostScript fonts is recommended. This protocol is an important companion document that provides useful guidance on interpreting CMB's diagnostic statistics and on assessing the integrity of its apportionments.

- ***Source82.zip*:** A compressed file of EPA-CMB8.2 source code. This file preserves the source code for further updates and allows it to be inspected for scientific verification. Most users do not need this file.

  EPA-CMB8.2 software is written in the Fortran, C++, and Delphi (Pascal) computer languages. The Fortran and C++ code (Appendix B) are compiled into a main DLL which is called at run time by the Delphi client (executable). This Delphi client handles the user interface, and is produced using the Delphi 7 compiler from Borland Software Corporation (Appendix D).

Note: The examples given in this manual are specific to the Windows® 95 (and higher) installation and use of the 32-bit software. The compressed files for test data sets and any documentation can be obtained by unZIPping the files listed above into a suitable folder. It is recommended that you also create a folder **\test case** and extract EPA-CMB82 test.zip. This test case is used in Section 5 of this manual.

2.3 Installing EPA-CMB8.2 Software

Extract the compressed EPA-CMB82.zip into a suitable folder (e.g., \CMB8.2\). When successfully installed, the following files will appear:

EPACMB82.exe Executable file.

EPACMB82.hlp[2] The context-sensitive help file.

CMB82.dll Dynamically Linked Library file; called at runtime by the executable.

EPACMB82.ini[3] Initializes file access information for use in the next session.

Using the right mouse button, a Shortcut can created for the executable (EPACMB82.exe) and moved to your desktop.

In addition, several "scratch" files will be created at run time and are stored temporarily in this working directory. When the user advances off the **Select Input Files** screen, the following *proient* (direct access) files are created in the executable directory:

AMBdirect.dat (binary file read by Fortran code; location of ambient data directory)

PROdirect.dat (binary file read by Fortran code; location of source profile data directory)

Once a **Run** is made, the following files are also created in this directory:

SumDirect.dat (binary file read by Fortran code; created only when a fit calculation is made)

$TEMPOUT.txt (ASCII buffer that stores results used by the Delphi User Interface)

These 4 scratch files are destroyed when the model is terminated normally.

[2] This help file, accessed by the Delphi client at run time, is generated by a help compiler from the help project file EPACMB82.hpj, which contains EPACMB82.rtf, etc. (provided with the source code).

[3] This initialization file will appear once EPA-CMB8.2 is executed.

3. EPA-CMB8.2 OPERATION

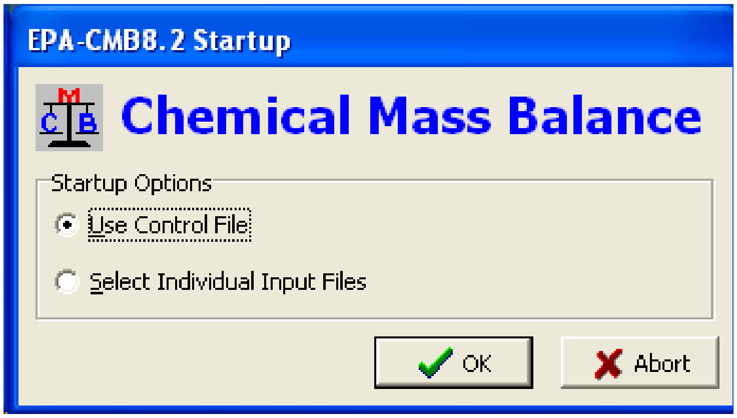
This section describes the EPA-CMB8.2 model commands. Although written using an object oriented programming language, the previous (i.e., CMB8) interface used an old-style, menu-driven approach. This design presented several buttons that each launched a separate form, and some buttons were redundant on different forms. Unfortunately, this approach did not take advantage of the underlying object oriented programming language of Delphi to create an *event-driven* system.

EPA-CMB8.2 uses an approach that features a logically organized menu, a system tool bar containing buttons for frequently used functions such as opening files and printing reports, a tabbed page interface, and a status bar at the top of each screen. This type of presentation is clear and eliminates the confusion associated with multiple buttons for the same functions. The tabbed page interface consolidates the numerous forms dictated by the EPA-CMB8.2 source code. A tabbed interface also provides users with visual clues regarding the logical progression of steps necessary to run the model. These improvements provide a Windows® look-and-feel to the CMB software that is familiar to most users.

Another benefit to this design is that the source code is much easier to maintain. Instead of maintaining separate programming modules (called *units* in Delphi) for each form, the source code is contained in a single unit. A list of run-time error messages has been compiled (Appendix E).

3.1 Input files

Figure 3.1 shows the screen that first appears when EPA-CMB8.2 is launched.



Startup Options

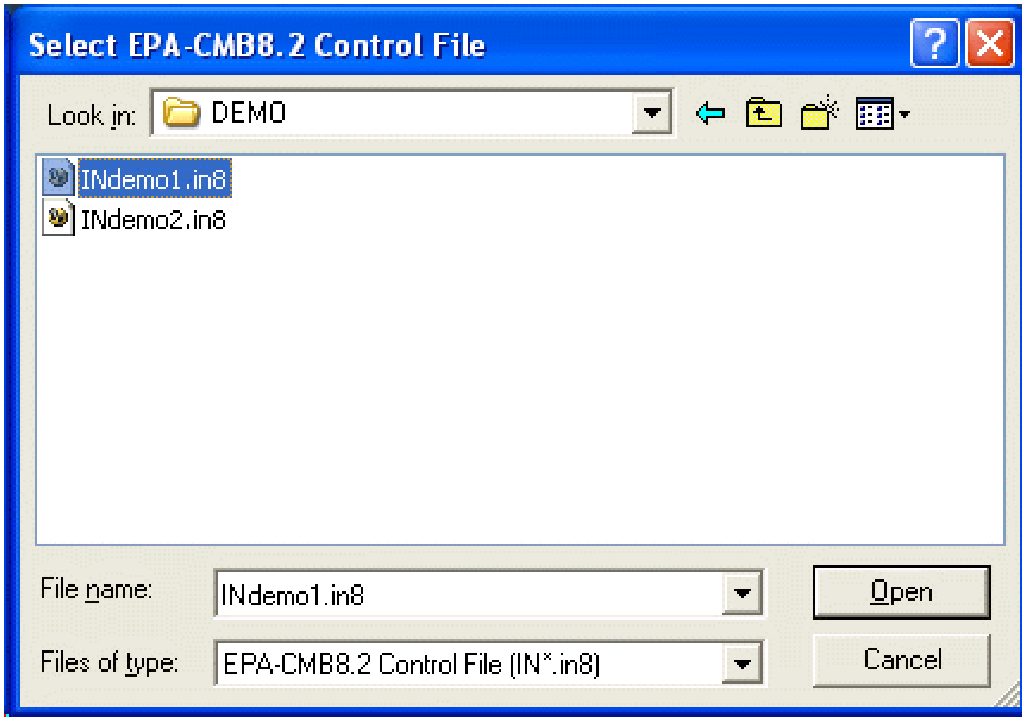
● Use Control File

○ Select Individual Input Files

Figure 3.1 Launching EPA-CMB8.2

Most users will have prepared a *Control File* (Section 4.2.1) for a particular CMB application. Control Files are commonly used in air quality models to specify input files that will be invoked during runtime. Input files listed in this file are described in Section 4. Selection of this mode brings up a Windows® browse box for selecting a Control File. If a Control File has not been prepared and the user simply wants to run CMB with particular (freely associated) input files, the other mode should be selected. Clicking on *Cancel* returns you to the (previous) Startup dialog.

If *Use Control File* was selected, a browse dialog appears as shown in Figure 3.2. This dialog allows the user to find a particular Control File. Once a Control File is chosen, the Select Input Files window appears, as depicted in Figure 3.3.



Look in: DEMO

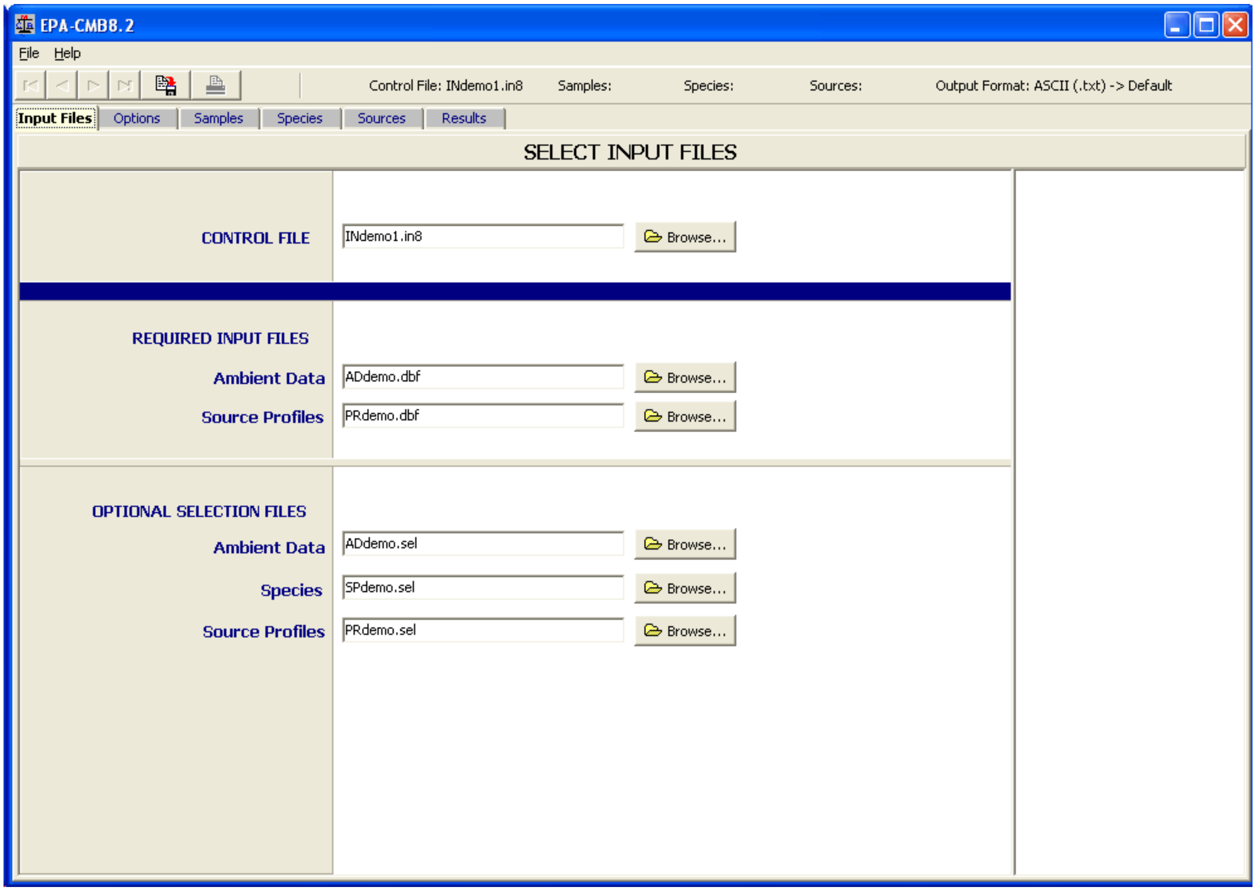
File name: INdemo1.in8

Files of type: EPA-CMB8.2 Control File (IN*.in8)

Figure 3.2 Browse Dialog for Selecting a Control File

In the Select Input Files window, the name of the Control File[4] is prominently displayed across the top, and the various input files it directs appear below. The Control File in use during any session also appears on all screens in the status bar at the top. Note that even though a Control File has been selected at this point, another one can easily be chosen via its browse dialog. As mentioned earlier, different Control Files can be loaded repeatedly during the same session, obviating the need to terminate the application. For any particular Control File, any of the input files may also be changed or removed via respective browse functions. EPA-CMB8.2 also gives the user the option from this screen to create a new Control File (with the new input file(s)) by using the File/Save function in the upper left-hand corner. Note, however, that once the initial Control File has been modified, its name will no longer appear on the top of the screen and must be reselected (via browse), assuming that a new file was created (saved).

[4] See Section 4.2.1 for more details on the structure and function of the Control File.



Control File: INdemo1.in8 Samples: Species: Sources: Output Format: ASCII (.txt) -> Default

SELECT INPUT FILES

CONTROL FILE: INdemo1.in8

REQUIRED INPUT FILES

Ambient Data: ADdemo.dbf

Source Profiles: PRdemo.dbf

OPTIONAL SELECTION FILES

Ambient Data: ADdemo.sel

Species: SPdemo.sel

Source Profiles: PRdemo.sel

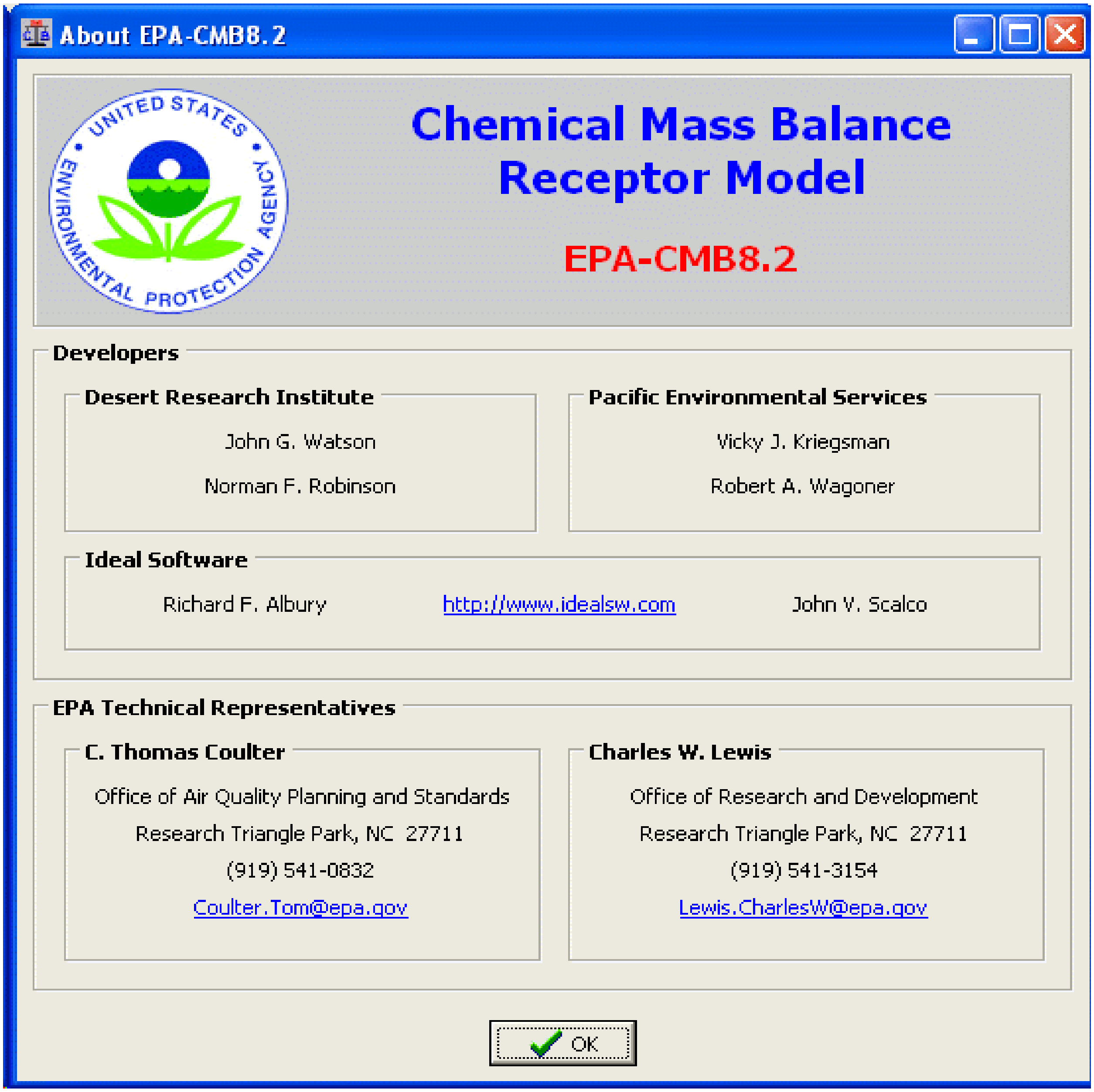
Figure 3.3 Select Input Files Screen

If at startup (Fig. 3.1) the user opts not to use a Control File, a *null* Select Input Files screen appears as illustrated in Figure 3.3, except that all the input file names are absent (including the Control File name). Files are selected via their respective browse dialogs and can be located in any directory. Note that EPA-CMB8.2 will accept *.csv, *.dbf, and *.wks data formats for ambient sample (AD*) and source profile (PR*) files, which are the only input files *required* by the model. The selection files are optional.

Even if a particular Control File has been loaded, different input files can be selected via their respective browse dialogs. If this is done, the user can create a new Control File by clicking the File icon at the top and following the prompts. This same functionality applies to the optional selection files (Sections 3.3 - 3.5).

Note that the 'Help' button on the top toolbar presents the option 'Contents'. Clicking this loads and presents the on-line help for EPA-CMB8.2. You can navigate through this system to find help on a variety of operational topics. Note that pressing **F1** on any screen presents help for **that** screen (on-line help is in this sense context-sensitive).

Note that clicking on the Help button also presents the option “About”. This button brings down a banner page as shown in Figure 3.5. It is good practice to verify this against the postings on EPA’s modeling web site (Section 2.2) to assure that the most recent revision is being used. EPA-CMB8.2 is being continually improved as users respond with recommendations or difficulties. There is also information on the model developers, as well as EPA project officers.



Chemical Mass Balance
Receptor Model

EPA-CMB8.2

Developers

Desert Research Institute

John G. Watson

Norman F. Robinson

Pacific Environmental Services

Vicky J. Kriegsman

Robert A. Wagoner

Ideal Software

Richard F. Albury   http://www.idealsw.com   John V. Scalco

EPA Technical Representatives

C. Thomas Coulter

Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711
(919) 541-0832
Coulter.Tom@epa.gov

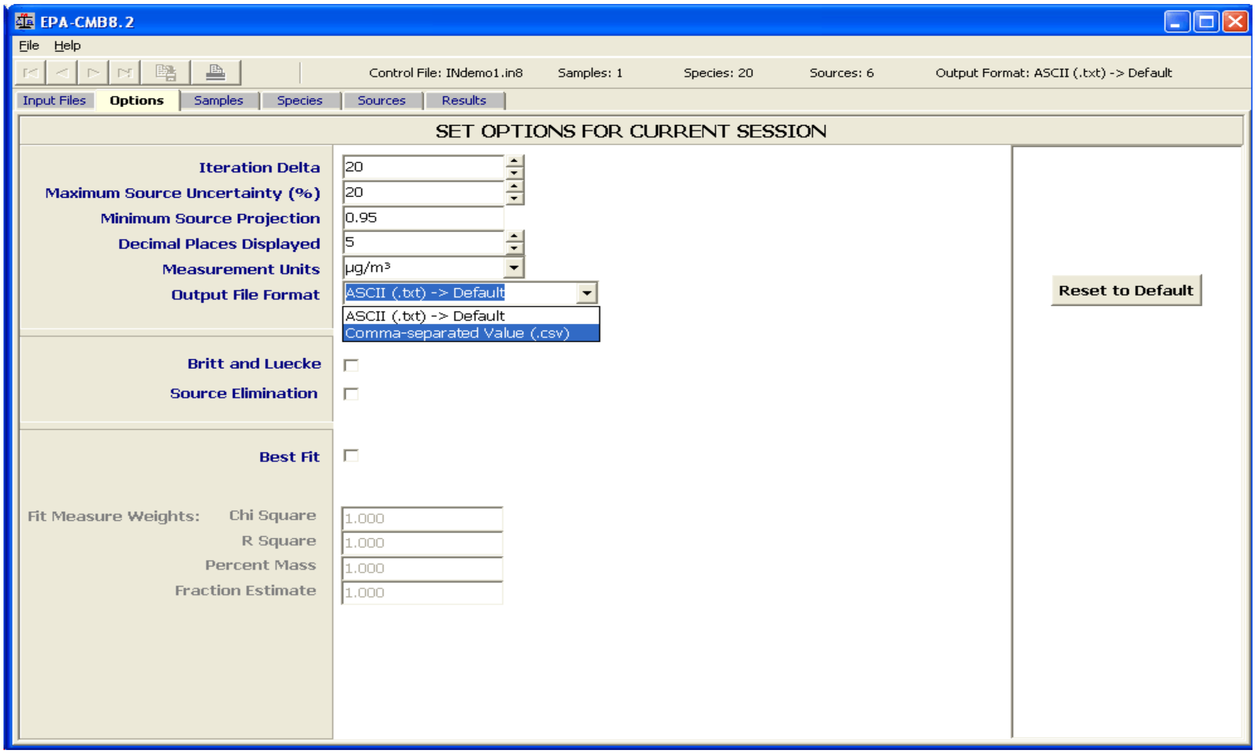
Charles W. Lewis

Office of Research and Development
Research Triangle Park, NC 27711
(919) 541-3154
Lewis.CharlesW@epa.gov

Figure 3.4 Banner Page for EPA0CMB8.2

3.2 Options

Several options are available in EPA-CMB8.2 that are selected from the Options tab (Fig. 3.4), where various values and selections may be changed from their default values. Note that the Control File name is clearly displayed on the status bar (top).



Control File: INdemo1.in8 Samples: 1 Species: 20 Sources: 6 Output Format: ASCII (.txt) -> Default

SET OPTIONS FOR CURRENT SESSION

Iteration Delta: 20
Maximum Source Uncertainty (%): 20
Minimum Source Projection: 0.95
Decimal Places Displayed: 5
Measurement Units: µg/m³
Output File Format: ASCII (.txt) -> Default
ASCII (.txt) -> Default
Comma-separated Value (.csv)

Britt and Luecke □
Source Elimination □

Best Fit □

Fit Measure Weights: Chi Square 1.000
R Square 1.000
Percent Mass 1.000
Fraction Estimate 1.000

Figure 3.5 Options for Current Session

Iteration Delta. This parameter sets the maximum number of iterations EPA-CMB8.2 will attempt to arrive at a solution. If no convergence can be achieved, there is probably excessive collinearity for this sample and combination of fitting sources. Its value is adjusted via the *spinners*. (Must be >0; no theoretical upper limit; default = 20)

Maximum Source Uncertainty / Minium Source Projection. These parameters allow the eligible space collinearity evaluation method of Henry (1992) to be implemented with each CMB calculation (Section 6.2). The eligible space method uses: 1) maximum source uncertainty; and 2) minimum source projection on the eligible space. The maximum source uncertainty is a *threshold* expressed as a percentage of the total measured mass and is adjustable via the *spinners* (default = 20% ; acceptable range 0 - 100). The minimum source projection is set to a default value of 0.95 (acceptable range 0.0 - 1.0), but can be changed in the display field. See Section 6.1.2 for more discussion of eligible space.

Decimal Places Displayed. This parameter sets the number of decimal places displayed in the output window and output files. This depends on the units used in the input data files. For example, data reported in ng/m^3 require fewer decimal places than values expressed in µg/m^3. Reducing this value from 5 to 4 accommodates most PM$_{2.5}$ mass and chemical concentrations expressed in µg/m^3. A value of 1 or 2 is best for concentrations expressed in ng/m^3 or for VOC expressed in ppbC or µg/m^3. This setting affects the display columns for source contributions estimates, measured species concentrations (ambient samples and source profiles), calculated contributions by species, as well as for inverse singular values. This parameter may be adjusted by using the *spinners*. The default value is 5 and the maximum value is 6.

Units. The units used for reporting results may be changed via a pull-down menu. Other typical units are available, or one may be created (the number of characters is limited to 5 or less).

Output File Format. The file format for spreadsheet-type output is selected in the pull-down box. As discussed in Section 4.5, the default is ASCII (txt); comma-separated value (CSV) is also available (which ports nicely to Microsoft Excel®). This selection is echoed on the status bar at the top of the screen.

Britt and Luecke. Checking this box applies the Britt and Luecke (1973) linear least squares solution that is explained by Watson *et al.* (1984) when applied to CMB calculations. This option allows the source profiles used in the fit calculation to vary, and enables a general solution to the least squares estimation that includes uncertainty in all the variables (i.e., the source compositions as well as the ambient concentrations). The default (option disabled) is the same approximation to the Britt-Luecke algorithm used in CMB7. Note that while the exact Britt-Luecke algorithm must generate a fit whose χ^2 value is equal to or better (i.e., smaller) than that from the approximation algorithm, there is no guarantee that the solution with the better χ^2 will be superior in terms of its physical meaning. Invocation of this option affects the fit obtained and, as in the case of EPA-CMB8.2's new treatment of collinearity, user experience is necessary to judge the utility in exercising the new Britt-Luecke algorithm. The Britt-Luecke algorithm, as implemented in EPA-CMB8.2, has not undergone comprehensive testing, and is not recommended for inexperienced users. Its inclusion as an option is mainly intended to provide the opportunity for interested advanced users to perform research investigations needed to establish its future viability. Note also that species concentrations that will appear in the Main Report reflect this algorithm's modification to the source profile matrix. The individual species concentrations for each source which appear in the (spreadsheet-type) output file are calculated using the UNmodified source profile matrix (and therefore will be different). This was done to maintain continuity with CMB6.

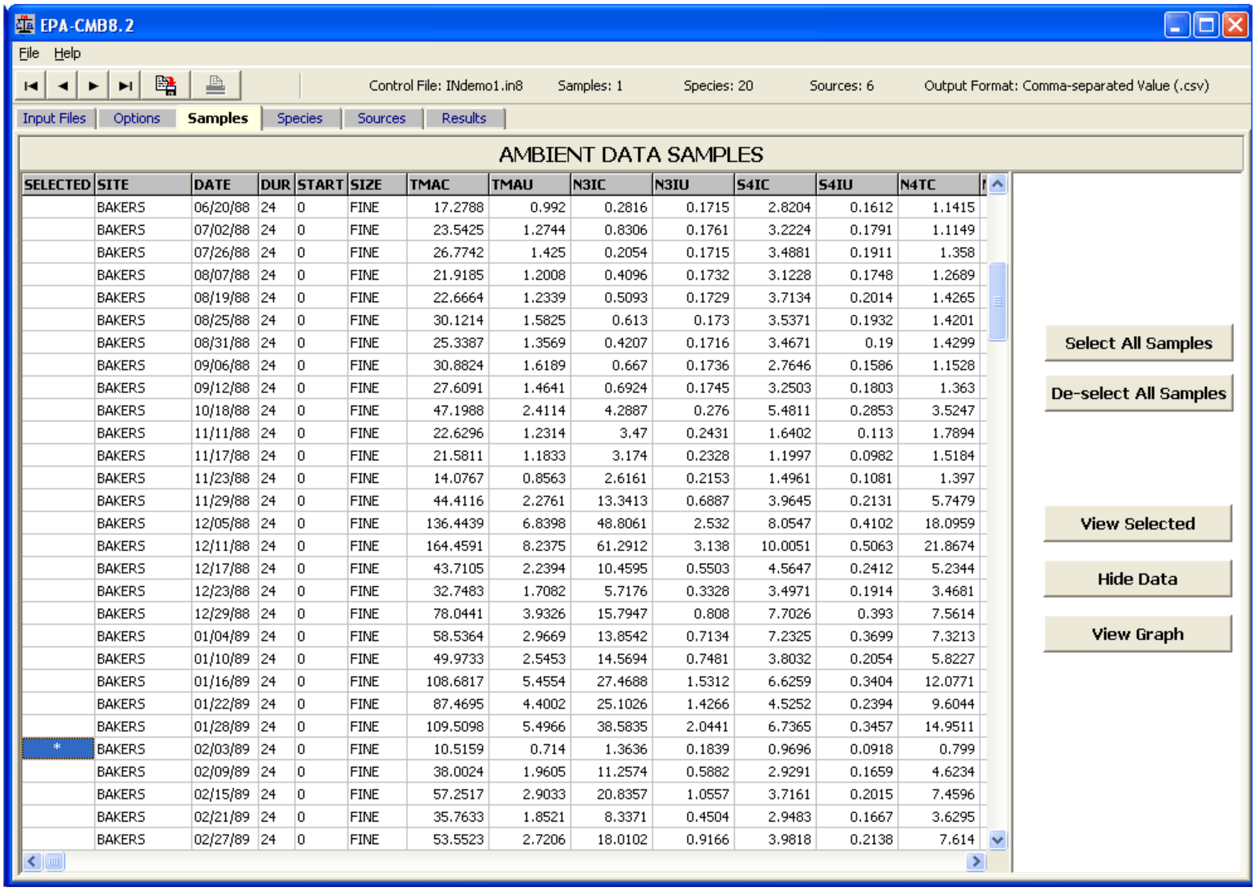
Source Elimination. Checking this box eliminates *negative* source contributions from the calculation, one at a time. After each fit attempt, if any sources have negative contributions, the source with the largest negative contribution is eliminated and another fit is attempted. This process is repeated until EPA-CMB8.2 finds no sources with negative contributions. Invocation of this option affects the fit obtained by effectively removing collinear sources (Section 6.1.2).

Best Fit. Checking this box causes the program to cycle through the corresponding *pairs* (same array index) of fitting species and source profile arrays specified in the source and species selection input windows until the best composite Fit Measure has been achieved. When Best Fit is invoked, EPA-CMB8.2 ignores any arrays of species and sources that may have been selected. The first fitting species array is paired with the first fitting sources array, and so on. EPA-CMB8.2 only attempts a search for a best fit among available corresponding pairs. Any arrays without a corresponding array to make a pair are ignored. The fit with the largest Fit Measure is then displayed and becomes the current fit. After a Best Fit has been made, the fitting species and fitting sources arrays will be tagged (highlighted) in their respective windows. The Fit Measure algorithm is described in Section 6.3.

Fit Measure Weight. These are the weights (coefficients) applied to each of the performance measures chi square, r-square, percent mass, and fraction of eligible sources (number in eligible space divided by number of fitting sources). Adjustment of these weights is not enabled in EPA-CMB8.2 unless Best Fit is invoked. Positive values between 0 and 1 may be entered by typing into the appropriate display fields. Defaults are 1.0 for each performance measure weight. See Section 6.1 for more discussion.

3.3 Select Ambient Data Samples

The screen for selecting a subset of samples on which source apportionment will be performed is shown in Figure 3.6. If an (optional) AD*.sel file is being directed from the Control File, one or more samples may appear selected initially (see Section 4.2.4). Otherwise, individual samples are "tagged" by clicking in the respective fields under "SELECTED". Clicking again deselects the sample. Use of Select /Clear All Samples may also be used to help establish the desired list of samples. Note that a counter displays on the status bar (top) the number of samples selected at any given time. As indicated, the collection date, duration, start hour, and size fraction (particles) are displayed for each sample. Show/Hide Data toggle between modes in which speciated data (alternating concentration and uncertainty) for the samples are either shown or masked. Toggling between View Selected/View All determines whether data will be displayed for selected (tagged) samples only, or for all samples in the list. Clicking View Graph will provide a bar chart for any sample (selected or not) for which any field is filled with blue (note that because of the physical constraints of the graph, some distortion will occur if the number of species exceeds ~25). This graph is useful to verify that input data files have been properly read. Note that use of the "VCR" control buttons on the top toolbar can help navigate down a long list of samples.



Control File: INdemo1.in8 Samples: 1 Species: 20 Sources: 6 Output Format: Comma-separated Value (.csv)

AMBIENT DATA SAMPLES

SELECTED	SITE	DATE	DUR	START	SIZE	TMAC	TMAU	N3IC	N3IU	S4IC	S4IU	N4TC
	BAKERS	06/20/88	24	0	FINE	17.2788	0.992	0.2816	0.1715	2.8204	0.1612	1.1415
	BAKERS	07/02/88	24	0	FINE	23.5425	1.2744	0.8306	0.1761	3.2224	0.1791	1.1149
	BAKERS	07/26/88	24	0	FINE	26.7742	1.425	0.2054	0.1715	3.4881	0.1911	1.358
	BAKERS	08/07/88	24	0	FINE	21.9185	1.2008	0.4096	0.1732	3.1228	0.1748	1.2689
	BAKERS	08/19/88	24	0	FINE	22.6664	1.2339	0.5093	0.1729	3.7134	0.2014	1.4265
	BAKERS	08/25/88	24	0	FINE	30.1214	1.5825	0.613	0.173	3.5371	0.1932	1.4201
	BAKERS	08/31/88	24	0	FINE	25.3387	1.3569	0.4207	0.1716	3.4671	0.19	1.4299
	BAKERS	09/06/88	24	0	FINE	30.8824	1.6189	0.667	0.1736	2.7646	0.1586	1.1528
	BAKERS	09/12/88	24	0	FINE	27.6091	1.4641	0.6924	0.1745	3.2503	0.1803	1.363
	BAKERS	10/18/88	24	0	FINE	47.1988	2.4114	4.2887	0.276	5.4811	0.2853	3.5247
	BAKERS	11/11/88	24	0	FINE	22.6296	1.2314	3.47	0.2431	1.6402	0.113	1.7894
	BAKERS	11/17/88	24	0	FINE	21.5811	1.1833	3.174	0.2328	1.1997	0.0982	1.5184
	BAKERS	11/23/88	24	0	FINE	14.0767	0.8563	2.6161	0.2153	1.4961	0.1081	1.397
	BAKERS	11/29/88	24	0	FINE	44.4116	2.2761	13.3413	0.6887	3.9645	0.2131	5.7479
	BAKERS	12/05/88	24	0	FINE	136.4439	6.8398	48.8061	2.532	8.0547	0.4102	18.0959
	BAKERS	12/11/88	24	0	FINE	164.4591	8.2375	61.2912	3.138	10.0051	0.5063	21.8674
	BAKERS	12/17/88	24	0	FINE	43.7105	2.2394	10.4595	0.5503	4.5647	0.2412	5.2344
	BAKERS	12/23/88	24	0	FINE	32.7483	1.7082	5.7176	0.3328	3.4971	0.1914	3.4681
	BAKERS	12/29/88	24	0	FINE	78.0441	3.9326	15.7947	0.808	7.7026	0.393	7.5614
	BAKERS	01/04/89	24	0	FINE	58.5364	2.9669	13.8542	0.7134	7.2325	0.3699	7.3213
	BAKERS	01/10/89	24	0	FINE	49.9733	2.5453	14.5694	0.7481	3.8032	0.2054	5.8227
	BAKERS	01/16/89	24	0	FINE	108.6817	5.4554	27.4688	1.5312	6.6259	0.3404	12.0771
	BAKERS	01/22/89	24	0	FINE	87.4695	4.4002	25.1026	1.4266	4.5252	0.2394	9.6044
	BAKERS	01/28/89	24	0	FINE	109.5098	5.4966	38.5835	2.0441	6.7365	0.3457	14.9511
*	BAKERS	02/03/89	24	0	FINE	10.5159	0.714	1.3636	0.1839	0.9696	0.0918	0.799
	BAKERS	02/09/89	24	0	FINE	38.0024	1.9605	11.2574	0.5882	2.9291	0.1659	4.6234
	BAKERS	02/15/89	24	0	FINE	57.2517	2.9033	20.8357	1.0557	3.7161	0.2015	7.4596
	BAKERS	02/21/89	24	0	FINE	35.7633	1.8521	8.3371	0.4504	2.9483	0.1667	3.6295
	BAKERS	02/27/89	24	0	FINE	53.5523	2.7206	18.0102	0.9166	3.9818	0.2138	7.614

Figure 3.6 Ambient Data Selection

There will be times when you data will include (*dichotomous*) measurements for Fine and Coarse size fractions. When this is the case and you want to apportion Fine and Coarse samples in a batch run, you must take care that ***compatible sample pairs*** (matched by site ID, date, sampling duration, and start hour) are selected. These pairs may be Fine/Coarse ... or Coarse/Fine ...

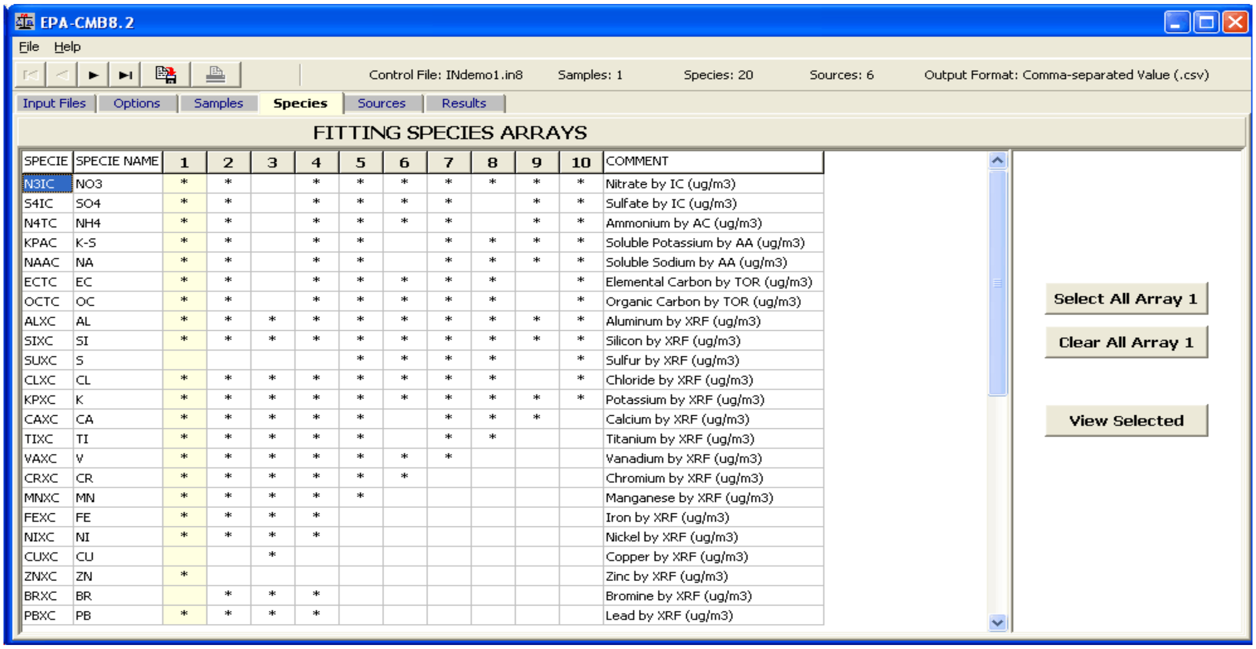
Upon exit, EPA-CMB8.2 will detect changes that may have occurred in arrays initialized by the optional input files during the session:

1) If an initial array of selected samples (as directed by AD*.sel) has been modified, the user will be prompted and asked if AD*.sel should be updated (overwritten). The file may also be conveniently renamed.

2) If no selection file was in use but an array of tagged samples has been created during the session, the user will be prompted and asked whether a samples selection file should be saved. If so, an appropriate name (e.g., AD*.sel) should be entered; the extension .sel will be appended automatically.

3.4 Select Fitting Species

Fitting species are used in the calculation of source contribution estimates. Species not included in this calculation are termed *floating* species (Section 4.3.2). The comparison of calculated and measured values for floating species is part of the model validation process. Fitting species should be selected that are major or unique components of the source types influencing the receptor concentrations. The screen for selecting fitting species is shown in Figure 3.7. This screen is initiated by data read from the (optional) species selection file



Control File: INdemo1.in8 Samples: 1 Species: 20 Sources: 6 Output Format: Comma-separated Value (.csv)

FITTING SPECIES ARRAYS

SPECIE	SPECIE NAME	1	2	3	4	5	6	7	8	9	10	COMMENT
N3IC	NO3	*	*		*	*	*	*	*	*	*	Nitrate by IC (ug/m3)
S4IC	SO4	*	*		*	*	*	*		*	*	Sulfate by IC (ug/m3)
N4TC	NH4	*	*		*	*	*	*		*	*	Ammonium by AC (ug/m3)
KPAC	K-S	*	*		*	*		*	*	*	*	Soluble Potassium by AA (ug/m3)
NAAC	NA	*	*		*	*		*	*	*	*	Soluble Sodium by AA (ug/m3)
ECTC	EC	*	*		*	*	*	*	*		*	Elemental Carbon by TOR (ug/m3)
OCTC	OC	*	*		*	*	*	*	*		*	Organic Carbon by TOR (ug/m3)
ALXC	AL	*	*	*	*	*	*	*	*	*	*	Aluminum by XRF (ug/m3)
SIXC	SI	*	*	*	*	*	*	*	*	*	*	Silicon by XRF (ug/m3)
SUXC	S					*	*	*	*		*	Sulfur by XRF (ug/m3)
CLXC	CL	*	*	*	*	*	*	*	*		*	Chloride by XRF (ug/m3)
KPXC	K	*	*	*	*	*	*	*	*	*	*	Potassium by XRF (ug/m3)
CAXC	CA	*	*	*	*	*		*	*	*		Calcium by XRF (ug/m3)
TIXC	TI	*	*	*	*	*		*	*			Titanium by XRF (ug/m3)
VAXC	V	*	*	*	*	*	*	*				Vanadium by XRF (ug/m3)
CRXC	CR	*	*	*	*	*	*					Chromium by XRF (ug/m3)
MNXC	MN	*	*	*	*	*						Manganese by XRF (ug/m3)
FEXC	FE	*	*	*	*							Iron by XRF (ug/m3)
NIXC	NI	*	*	*	*							Nickel by XRF (ug/m3)
CUXC	CU			*								Copper by XRF (ug/m3)
ZNXC	ZN	*										Zinc by XRF (ug/m3)
BRXC	BR		*	*	*							Bromine by XRF (ug/m3)
PBXC	PB	*	*	*	*							Lead by XRF (ug/m3)

Figure 3.7 Fitting Species Arrays

(SP*.sel; see Section 4.2.4). Species contained in the selection file are listed down the left-handside and a field of up to 10 arrays is provided. At startup, the first array (in a series) will always be initially selected as a default. Other arrays may be selected by clicking on the array index (1 - 10). Within a given array that is first activated by clicking its index number, species may be added or removed by clicking in the appropriate field. Select/Clear All Array X may also help in configuring a selection array. Note that use of the "VCR" control buttons can help navigate down a long list of species, and that a counter displays on the status bar (top) the number of species tagged for any selected array. For a selected array, toggling between View Selected/View All determines which species will be displayed. This can be handy for a long list of species that would be impossible to display on the screen. If comments are provided in the selection file, they will be displayed on the right-hand side.

Multiple arrays for fitting species are useful when CMB calculations are performed on samples from several locations or during different times of the year that have different contributors. They are also used by the Best Fit option to cycle through different source combinations until the weighted Fit Measure is optimized (Section 3.2).

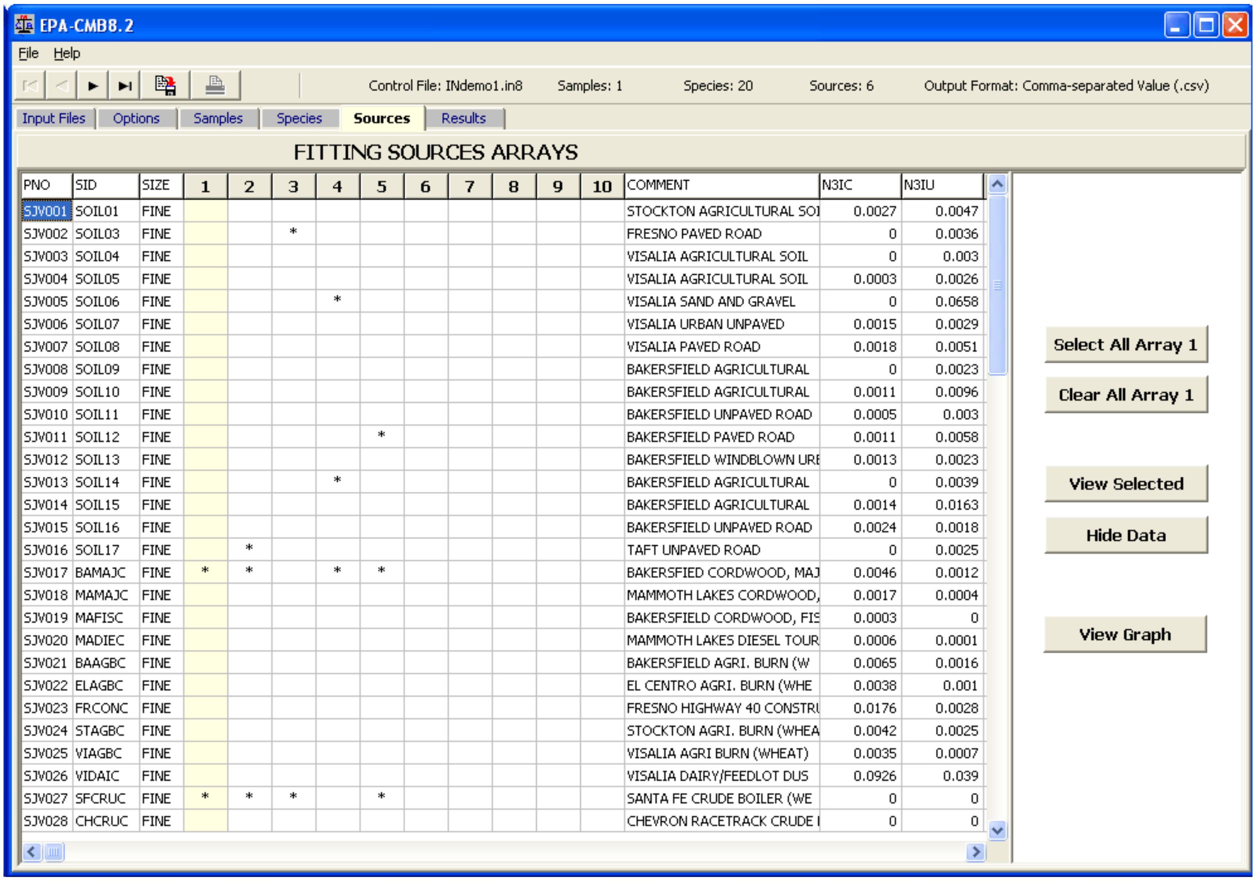
Upon exit, EPA-CMB8.2 will detect changes that may have occurred in arrays initialized by the optional input files during the session:

1) If an initial array of selected species (as directed by SP*.sel) has been modified, the user will be prompted and asked if SP*.sel should be updated (overwritten). The file may also be conveniently renamed.

2) If no selection file was in use but an array of tagged species has been created during the session, the user will be prompted and asked whether a samples selection file should be saved. If so, an appropriate name (e.g., SP*.sel) should be entered; the extension .sel will be appended automatically.

3.5 Select Fitting Sources

Fitting source profiles are included in the CMB calculation. The user should select profiles that represent the emissions most likely to influence receptor concentrations. Several profiles may be available that represent the same source type, but only one of these is usually used as a fitting source. Profiles of similar chemical composition are often found to be collinear when two or more are selected as fitting sources. The screen for selecting fitting sources is shown in Figure 3.8. This screen is initiated with data read from the (optional) species selection file (PR*.sel; see Section 4.2.4). Source profiles contained in the selection file are listed down the left-hand side



Control File: INdemo1.in8 Samples: 1 Species: 20 Sources: 6 Output Format: Comma-separated Value (.csv)

PNO	SID	SIZE	1	2	3	4	5	6	7	8	9	10	COMMENT	N3IC	N3IU
SJV001	SOIL01	FINE											STOCKTON AGRICULTURAL SOI	0.0027	0.0047
SJV002	SOIL03	FINE			*								FRESNO PAVED ROAD	0	0.0036
SJV003	SOIL04	FINE											VISALIA AGRICULTURAL SOIL	0	0.003
SJV004	SOIL05	FINE											VISALIA AGRICULTURAL SOIL	0.0003	0.0026
SJV005	SOIL06	FINE				*							VISALIA SAND AND GRAVEL	0	0.0658
SJV006	SOIL07	FINE											VISALIA URBAN UNPAVED	0.0015	0.0029
SJV007	SOIL08	FINE											VISALIA PAVED ROAD	0.0018	0.0051
SJV008	SOIL09	FINE											BAKERSFIELD AGRICULTURAL	0	0.0023
SJV009	SOIL10	FINE											BAKERSFIELD AGRICULTURAL	0.0011	0.0096
SJV010	SOIL11	FINE											BAKERSFIELD UNPAVED ROAD	0.0005	0.003
SJV011	SOIL12	FINE					*						BAKERSFIELD PAVED ROAD	0.0011	0.0058
SJV012	SOIL13	FINE											BAKERSFIELD WINDBLOWN URE	0.0013	0.0023
SJV013	SOIL14	FINE				*							BAKERSFIELD AGRICULTURAL	0	0.0039
SJV014	SOIL15	FINE											BAKERSFIELD AGRICULTURAL	0.0014	0.0163
SJV015	SOIL16	FINE											BAKERSFIELD UNPAVED ROAD	0.0024	0.0018
SJV016	SOIL17	FINE		*									TAFT UNPAVED ROAD	0	0.0025
SJV017	BAMAJC	FINE	*	*		*	*						BAKERSFIED CORDWOOD, MAJ	0.0046	0.0012
SJV018	MAMAJC	FINE											MAMMOTH LAKES CORDWOOD,	0.0017	0.0004
SJV019	MAFISC	FINE											BAKERSFIELD CORDWOOD, FIS	0.0003	0
SJV020	MADIEC	FINE											MAMMOTH LAKES DIESEL TOUR	0.0006	0.0001
SJV021	BAAGBC	FINE											BAKERSFIELD AGRI. BURN (W	0.0065	0.0016
SJV022	ELAGBC	FINE											EL CENTRO AGRI. BURN (WHE	0.0038	0.001
SJV023	FRCONC	FINE											FRESNO HIGHWAY 40 CONSTRU	0.0176	0.0028
SJV024	STAGBC	FINE											STOCKTON AGRI. BURN (WHEA	0.0042	0.0025
SJV025	VIAGBC	FINE											VISALIA AGRI BURN (WHEAT)	0.0035	0.0007
SJV026	VIDAIC	FINE											VISALIA DAIRY/FEEDLOT DUS	0.0926	0.039
SJV027	SFCRUC	FINE	*	*	*		*						SANTA FE CRUDE BOILER (WE	0	0
SJV028	CHCRUC	FINE											CHEVRON RACETRACK CRUDE I	0	0

Figure 3.8 Fitting Sources Arrays

and a field of up to 10 arrays is provided. As with fitting species, arrays are selected by clicking on the array index (1 - 10). Within a given array, source profiles may be added or removed by clicking in the appropriate field. Select/Clear All Array X may also help in configuring a selection array. Note that use of the "VCR" control buttons can help navigate down a long list of sources, and that a counter displays on the status bar (top) the number of source profiles tagged for any selected array. For a selected array, toggling between View Selected/View All determines which source profiles will be displayed. This can be handy for a long list of sources that would be impossible to display on the screen. If comments are provided in the selection file, they will be displayed on the right-hand side.

As for ambient samples, clicking View Graph will provide a bar chart for any source profile (selected or not) for which any field is filled with blue (note that because of the physical constraints of the graph, some distortion will occur if the number of species exceeds ~25). This graph is useful for visual inspection; it helps to verify that input data files have been properly read and to identify abundant components in each profile. View Grid returns to the array screen.

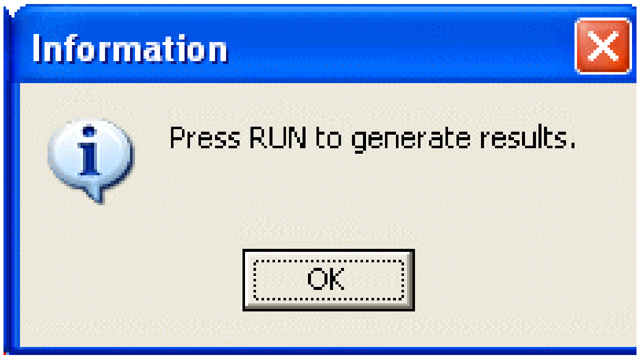
Upon exit, EPA-CMB8.2 will detect changes that may have occurred in arrays initialized by the optional input files during the session:

1) If an initial array of selected source profiles (as directed by PR*.sel) has been modified, the user will be prompted and asked if PR*.sel should be updated (overwritten). The file may also be conveniently renamed.

2) If no selection file was in use but an array of tagged source profiles has been created during the session, the user will be prompted and asked whether a samples selection file should be saved. If so, an appropriate name (e.g., PR*.sel) should be entered; the extension .sel will be appended automatically.

3.6 Calculation Results

Once options have been set, one or more samples selected, as well as a suitable array of fitting species and source profiles, EPA-CMB8.2 is ready to do a calculation. The Results screen shown in Figure 3.9 is where fitting results and statistics are reported for examination. When the screen is first viewed, the user is prompted to click on Run in order to initiate a calculation.



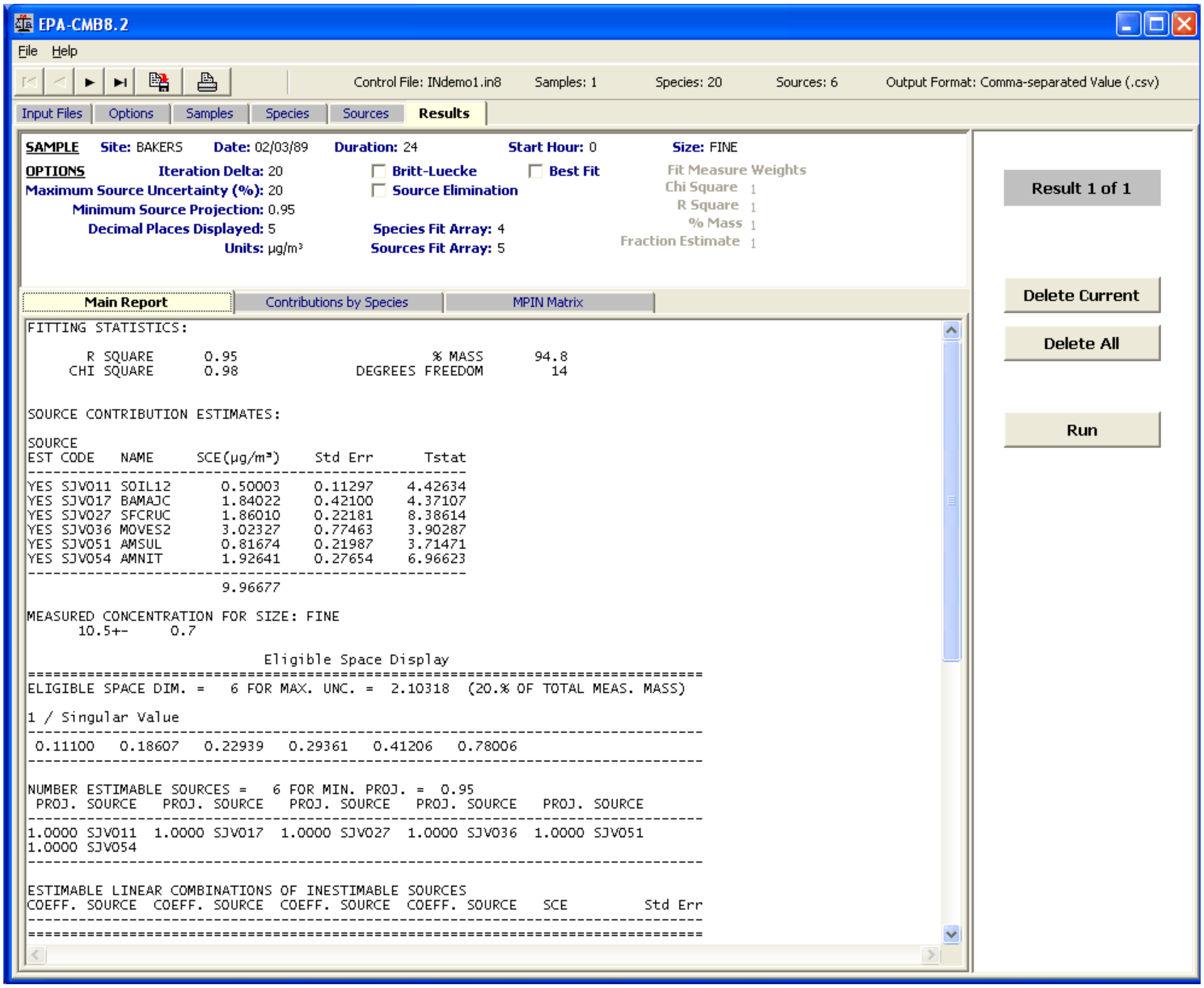
Press RUN to generate results.

Figure 3.9 Calculation Results Tab

When Run is invoked, EPA-CMB8.2 performs the least-squares estimation of source contribution estimates and performance measures on the selected sample data using the designated fitting species and source profiles. Note also that if more fitting sources than fitting species have been selected, a warning appears and the user is forced to reconfigure.

3.6.1 Main Report

When Run in invoked, EPA-CMB8.2 attempts fits of all samples selected. If more than one sample is selected, the model runs in a *batch mode*.[5] As each sample is apportioned, results are successively written to the output window as part of the Main Report (Figure 3.10). For any



Control File: INdemo1.in8   Samples: 1   Species: 20   Sources: 6   Output Format: Comma-separated Value (.csv)

SAMPLE Site: BAKERS Date: 02/03/89 Duration: 24 Start Hour: 0 Size: FINE

OPTIONS Iteration Delta: 20 □ Britt-Luecke □ Best Fit
Maximum Source Uncertainty (%): 20 □ Source Elimination
Minimum Source Projection: 0.95
Decimal Places Displayed: 5 Species Fit Array: 4
Units: µg/m³ Sources Fit Array: 5

Fit Measure Weights
Chi Square 1
R Square 1
% Mass 1
Fraction Estimate 1

```
FITTING STATISTICS:

       R SQUARE      0.95                     % MASS       94.8
     CHI SQUARE      0.98               DEGREES FREEDOM      14


SOURCE CONTRIBUTION ESTIMATES:

SOURCE
EST CODE   NAME     SCE(µg/m³)    Std Err      Tstat
----------------------------------------------------
YES SJV011 SOIL12      0.50003    0.11297    4.42634
YES SJV017 BAMAJC      1.84022    0.42100    4.37107
YES SJV027 SFCRUC      1.86010    0.22181    8.38614
YES SJV036 MOVES2      3.02327    0.77463    3.90287
YES SJV051 AMSUL       0.81674    0.21987    3.71471
YES SJV054 AMNIT       1.92641    0.27654    6.96623
----------------------------------------------------
                       9.96677

MEASURED CONCENTRATION FOR SIZE: FINE
      10.5+-     0.7

                            Eligible Space Display
================================================================================
ELIGIBLE SPACE DIM. =   6 FOR MAX. UNC. =  2.10318  (20.% OF TOTAL MEAS. MASS)

1 / Singular Value
--------------------------------------------------------------------------------
 0.11100   0.18607   0.22939   0.29361   0.41206   0.78006
--------------------------------------------------------------------------------

NUMBER ESTIMABLE SOURCES =   6 FOR MIN. PROJ. =  0.95
 PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE
--------------------------------------------------------------------------------
1.0000 SJV011  1.0000 SJV017  1.0000 SJV027  1.0000 SJV036  1.0000 SJV051
1.0000 SJV054
--------------------------------------------------------------------------------

ESTIMABLE LINEAR COMBINATIONS OF INESTIMABLE SOURCES
COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE   SCE         Std Err
--------------------------------------------------------------------------------
================================================================================
```

Figure 3.10 Calculation Results - Main Report

(*current*) result displayed in the output window, the header at the top reflects basic identifying information pertaining to the sample that has been apportioned and an echo of the options settings. The species and source profile fitting array indices are also indicated. If more than one sample was apportioned, the number will be reflected in the box in the upper right-hand corner. Viewing any sample result in a series is quite easy with the use of the VCR buttons on the toolbar. Using appropriate buttons, the current or all results may be deleted from the output window. The report may be printed for the current or all sample results via the *print* button on

[5] Note that in batch mode, certain warning /error messages that require user intervention are suppressed.

the toolbar; it is also possible to print to Adobe Acrobat® (PDF). A file may be created via the Print To File option. The header for this file is embellished with an echo of all input files used in the calculation - another feature of EPA-CMB 8.2.

In analyzing data that may have been collected in sampling networks that employ dichotomous samplers, it is common to input speciated data for two complementary size fractions, *fine* and *coarse* (Section 4.2.2). If EPA-CMB8.2 detects that it has such a sample pair for a given period, it will assume them to be complementary. When a fit is performed, an additional report is created in which results are summed for the *fine* and *coarse* fractions to give the Total (frequently = PM_{10}). In this case, the fitting array indices are the same as for the *fine* and *coarse* component samples, as is the value for Degrees of Freedom. For % Mass explained, the value is determined as follows:

$$\%\ Mass = \frac{\left(mass_{SF1} + mass_{SF2}\right)_{calc.}}{\left(mass_{SF1} + mass_{SF2}\right)_{meas.}}$$

where SF = size fraction

For a given sampling site and period, the samples tagged for analysis must be presented as alternating dichotomous pairs, i.e., Fine/Coarse; Fine/Coarse; Fine/Coarse; ... or Coarse/Fine; Coarse, Fine; Coarse/Fine ... The ‘Total’ report is always appended to that for the 2nd in the dichotomous pair.

In the Main report are several information blocks which present performance measures (Section 6.1.1). First are fitting statistics: r^2, χ^2, percent mass explained, and degrees of freedom. Next is a block that presents the most basic results from EPA-CMB8.2: source contribution estimates. For each source selected is presented a source contribution estimate (SCE) in user-chosen units (Section 3.2), standard errors, and values for T-stat. The series of SCEs are summed to provide a convenient check on the % mass explained value (EPA-CMB8.2 feature):

$$\%\ Mass = \left[\frac{\sum SCEs}{total\ measured\ conc.}\right] X\ 100$$

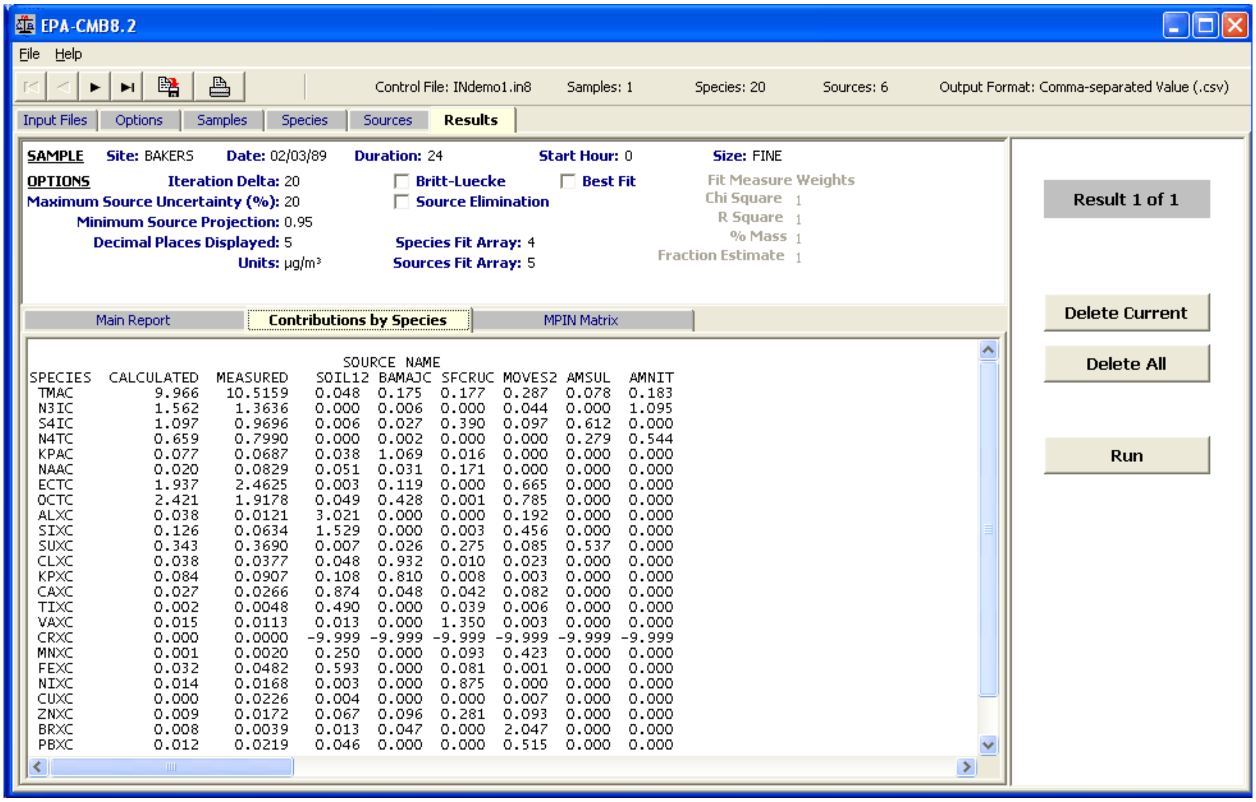
In making this check, be aware that the full uncertainty of the measured concentration value is not displayed in the Main Report. The field ‘EST’ under ‘SOURCE’ indicates (YES or NO) whether a source’s contribution was estimable in EPA-CMB8.2's attempt at a fit using the settings in Options. The next block is the *Eligible Space Collinearity Display*: an echo of the measured concentration and error for the sample, eligible space dimension for the chosen maximum uncertainty (Section 3.2), inverse singular values, the number of estimable sources for the chosen minimum source projection (Section 3.2), and estimable linear combinations of inestimable sources. The concepts of *estimable sources* and *estimable space* are discussed in Section 6.1.2. Finally, there is a block detailing species concentrations. Shown for each species (fitting species are tagged with asterisks) is its measured and calculated mass and uncertainty, the ratio of calculated/measured mass (± uncertainty factor), and the ratio of the signed residual (calculated - measured mass) to the uncertainty of that residual. See Section 6.1.3 for more details.

Of note is the way EPA-CMB8.2 handles missing values in source and receptor files (designated by **–99.** in place of the value). When a fitting species value is missing from either an ambient sample or source profile, that species is automatically removed from the calculation and the species selection flag (ordinarily an asterisk) is set to "M" in the report output file. See Section 4.2.2 & 4.2.3; Appendix F.

For any of the information blocks in the main report, the presence of a series of asterisks for a numerical value field represents an overflow condition. Reducing the number of decimal places displayed (Section 3.2) should correct the problem.

3.6.2 Contributions by Species

Beyond the traditional apportionment results generated by EPA-CMB8.2, it is also of interest to analyze the way in which pollutant mass is distributed among sources by species. This distribution is presented in the report Contributions by Species (Figure 3.11). This report



Control File: INdemo1.in8 Samples: 1 Species: 20 Sources: 6 Output Format: Comma-separated Value (.csv)

SAMPLE **Site:** BAKERS **Date:** 02/03/89 **Duration:** 24 **Start Hour:** 0 **Size:** FINE

OPTIONS **Iteration Delta:** 20 □ **Britt-Luecke** □ **Best Fit**

Maximum Source Uncertainty (%): 20 □ **Source Elimination**

Minimum Source Projection: 0.95

Decimal Places Displayed: 5 **Species Fit Array:** 4

Units: µg/m³ **Sources Fit Array:** 5

Fit Measure Weights
Chi Square 1
R Square 1
% Mass 1
Fraction Estimate 1

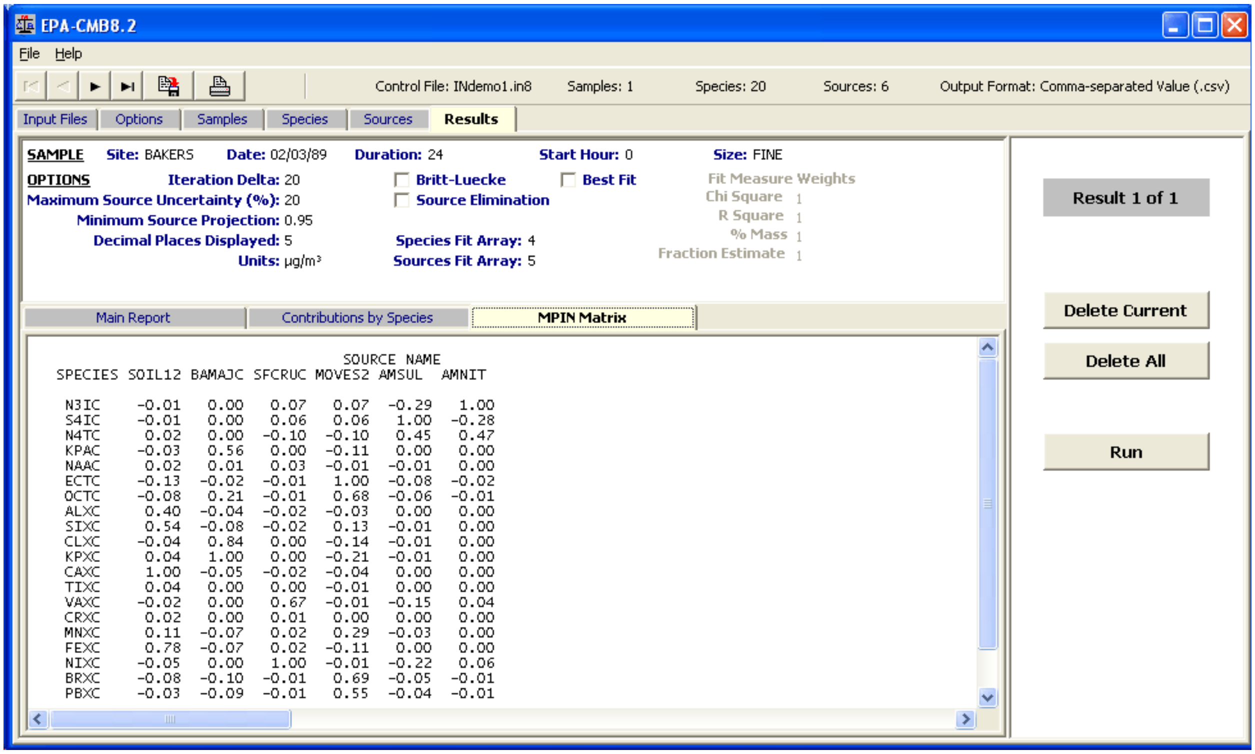
```
                                    SOURCE  NAME
SPECIES  CALCULATED  MEASURED   SOIL12 BAMAJC SFCRUC MOVES2 AMSUL  AMNIT
 TMAC         9.966   10.5159   0.048  0.175  0.177  0.287  0.078  0.183
 N3IC         1.562    1.3636   0.000  0.006  0.000  0.044  0.000  1.095
 S4IC         1.097    0.9696   0.006  0.027  0.390  0.097  0.612  0.000
 N4TC         0.659    0.7990   0.000  0.002  0.000  0.000  0.279  0.544
 KPAC         0.077    0.0687   0.038  1.069  0.016  0.000  0.000  0.000
 NAAC         0.020    0.0829   0.051  0.031  0.171  0.000  0.000  0.000
 ECTC         1.937    2.4625   0.003  0.119  0.000  0.665  0.000  0.000
 OCTC         2.421    1.9178   0.049  0.428  0.001  0.785  0.000  0.000
 ALXC         0.038    0.0121   3.021  0.000  0.000  0.192  0.000  0.000
 SIXC         0.126    0.0634   1.529  0.000  0.003  0.456  0.000  0.000
 SUXC         0.343    0.3690   0.007  0.026  0.275  0.085  0.537  0.000
 CLXC         0.038    0.0377   0.048  0.932  0.010  0.023  0.000  0.000
 KPXC         0.084    0.0907   0.108  0.810  0.008  0.003  0.000  0.000
 CAXC         0.027    0.0266   0.874  0.048  0.042  0.082  0.000  0.000
 TIXC         0.002    0.0048   0.490  0.000  0.039  0.006  0.000  0.000
 VAXC         0.015    0.0113   0.013  0.000  1.350  0.003  0.000  0.000
 CRXC         0.000    0.0000  -9.999 -9.999 -9.999 -9.999 -9.999 -9.999
 MNXC         0.001    0.0020   0.250  0.000  0.093  0.423  0.000  0.000
 FEXC         0.032    0.0482   0.593  0.000  0.081  0.001  0.000  0.000
 NIXC         0.014    0.0168   0.003  0.000  0.875  0.000  0.000  0.000
 CUXC         0.000    0.0226   0.004  0.000  0.000  0.007  0.000  0.000
 ZNXC         0.009    0.0172   0.067  0.096  0.281  0.093  0.000  0.000
 BRXC         0.008    0.0039   0.013  0.047  0.000  2.047  0.000  0.000
 PBXC         0.012    0.0219   0.046  0.000  0.000  0.515  0.000  0.000
```

Figure 3.11 Calculation Results - Contributions by Species

provides one more dimension to the Species Concentrations block in the Main Report. These results are useful when source contributions to species other than total mass are of interest. The report also indicates which sources are the major and minor contributors to each species. Since values in this report are ratios of calculated species concentrations to the measured total species concentration, multiplying the values by their respective *measured* value and summing will confirm the values listed in the sum of calculated species contributions column (left-hand side). For convenience, both the calculated and measured columns are from the Main Report are reproduced here - another feature of EPA-CMB8.2. As with the Main Report, a print-out of Contributions by Species may be obtained for the current sample result via the *print* button on the toolbar, and a file may be created via the Print To File option. For more information see Section 6.2.

3.6.3 Modified Pseudo-Inverse Normalized (MPIN) Matrix

Another report that may be of interest is the Modified Pseudo-Inverse Normalized (MPIN) matrix (Figure 3.12). The MPIN matrix identifies which fitting species have the largest influence on the source contribution estimates from each profile (Section 6.2). Examining these weights suggests sensitivity tests to determine the extent to which source contributions vary with changes in profile abundances or the selection of fitting species. As with the Main Report and Contributions by Species, a print-out of the MPIN matrix may be obtained for the current sample result via the *print* button on the toolbar, and a file may be created via the Print To File option.



Control File: INdemo1.in8 Samples: 1 Species: 20 Sources: 6 Output Format: Comma-separated Value (.csv)

SAMPLE Site: BAKERS Date: 02/03/89 Duration: 24 Start Hour: 0 Size: FINE

OPTIONS Iteration Delta: 20
Maximum Source Uncertainty (%): 20
Minimum Source Projection: 0.95
Decimal Places Displayed: 5
Units: µg/m³

□ Britt-Luecke □ Best Fit
□ Source Elimination

Species Fit Array: 4
Sources Fit Array: 5

Fit Measure Weights
Chi Square 1
R Square 1
% Mass 1
Fraction Estimate 1

```
                           SOURCE  NAME
SPECIES  SOIL12  BAMAJC  SFCRUC  MOVES2  AMSUL   AMNIT

  N3IC     -0.01    0.00    0.07    0.07   -0.29    1.00
  S4IC     -0.01    0.00    0.06    0.06    1.00   -0.28
  N4TC      0.02    0.00   -0.10   -0.10    0.45    0.47
  KPAC     -0.03    0.56    0.00   -0.11    0.00    0.00
  NAAC      0.02    0.01    0.03   -0.01   -0.01    0.00
  ECTC     -0.13   -0.02   -0.01    1.00   -0.08   -0.02
  OCTC     -0.08    0.21   -0.01    0.68   -0.06   -0.01
  ALXC      0.40   -0.04   -0.02   -0.03    0.00    0.00
  SIXC      0.54   -0.08   -0.02    0.13   -0.01    0.00
  CLXC     -0.04    0.84    0.00   -0.14   -0.01    0.00
  KPXC      0.04    1.00    0.00   -0.21   -0.01    0.00
  CAXC      1.00   -0.05   -0.02   -0.04    0.00    0.00
  TIXC      0.04    0.00    0.00   -0.01    0.00    0.00
  VAXC     -0.02    0.00    0.67   -0.01   -0.15    0.04
  CRXC      0.02    0.00    0.01    0.00    0.00    0.00
  MNXC      0.11   -0.07    0.02    0.29   -0.03    0.00
  FEXC      0.78   -0.07    0.02   -0.11    0.00    0.00
  NIXC     -0.05    0.00    1.00   -0.01   -0.22    0.06
  BRXC     -0.08   -0.10   -0.01    0.69   -0.05   -0.01
  PBXC     -0.03   -0.09   -0.01    0.55   -0.04   -0.01
```

Figure 3.12 Calculation Results - MPIN Matrix

4. INPUT AND OUTPUT FILES

This section describes the structure of EPA-CMB8.2 input and output files and methods of generating these files. Each type of input file structure is illustrated with one of the test data sets packaged with EPA-CMB8.2.

4.1 File Naming Conventions

EPA-CMB8.2 input and output files can have any file name with a three-character extension that indicates the file type. A suggested naming convention is **PP*.ext**, where:

- **PP:** Type of file. Common definitions are:

  IN Control file identifying specific input data files.

  AD **A**mbient **D**ata. Selection file initiates sample selection from the ambient data file for apportionment during an CMB session; data file contains the measured ambient concentrations and their uncertainty values.

  PR Source **PR**ofile. Selection file identifies initial fitting profiles and source profile descriptions; data file contains mass-fraction chemical abundances and their uncertainties.

  SP **SP**ecies selection file identifies initial fitting species for the CMB session.

- ***** Study identifier. This code allows separate studies to be distinguished from one another. EPA-CMB8.2 allows Windows® flexibility for this name string (i.e., it is not character-limited).

- **ext** Extension that also identifies file type or format. The following file extensions are recognized by EPA-CMB8.2:

  in8 Input control (ASCII text) file. EPA-CMB8.2 lists files with this extension in the Control File browse window at startup.

  sel Fitting profiles, fitting species, and sample selection (ASCII text) files. EPA-CMB8.2 recognizes files with this extension as containing initial selections that can be entered external to the program. This extension applies only to the PR, SP, and AD file types.

  csv Ambient data or source profile comma-separated value ASCII text file. Each field is separated by a comma. Comma-delimited ASCII data base output files are written with this extension.

  dbf Data base file generated by dBase or FoxPro compatible data management software. Most commonly used spreadsheets offer this as an output option. dBase or FoxPro output files are written with this extension.

txt Ambient data or source profile data blank-delimited ASCII text file. Formatted, blank-delimited ASCII data base output files may be created with these extensions.

wks Lotus 1-2-3 version 1 spreadsheet format. Most commonly used spreadsheets offer this as an output option. This is the most useful output format for the data base output file when source contribution estimates will be analyzed using a spreadsheet.

Note that if neither input file (ambient data or source profile data) is supplied in ASCII format (*.txt), EPA-CMB8.2 converts any of the **csv**, **dbf**, and **wks** input data files to the blank-delimited (**txt**) files which are actually used by the program. These **txt** files are created "on the fly" as soon as the user moves off of the Input Files Screen (Figure 3.3), and will appear in the same subdirectory that stores any of the **csv**, **dbf**, and/or **wks** files. Such **txt** files created from **dbf** files are nicely formatted and easy to read. If any **txt** files are supplied in the subdirectory but not directed for use as input files in the Control File, they will be overwritten (replaced) by new versions created by EPA-CMB8.2. If, however, any **txt** files are supplied and directed for use as input data by the Control File, they will be retained (not modified).

4.2 Input File Relationship

Six data files are normally used for input to EPA-CMB8.2, the first of which is a *control* file that directs EPA-CMB8.2 to five specific files. Three of the files are optional *selection* files, which provide substantial user convenience by establishing commonly used arrays and sample subsets that would otherwise need to be initialized each time the model is run. The remaining two - the ambient and source profile data files - are *required* by EPA-CMB8.2. Figure 4.1 presents the relationship of the files whose descriptions appear in the following subsections.

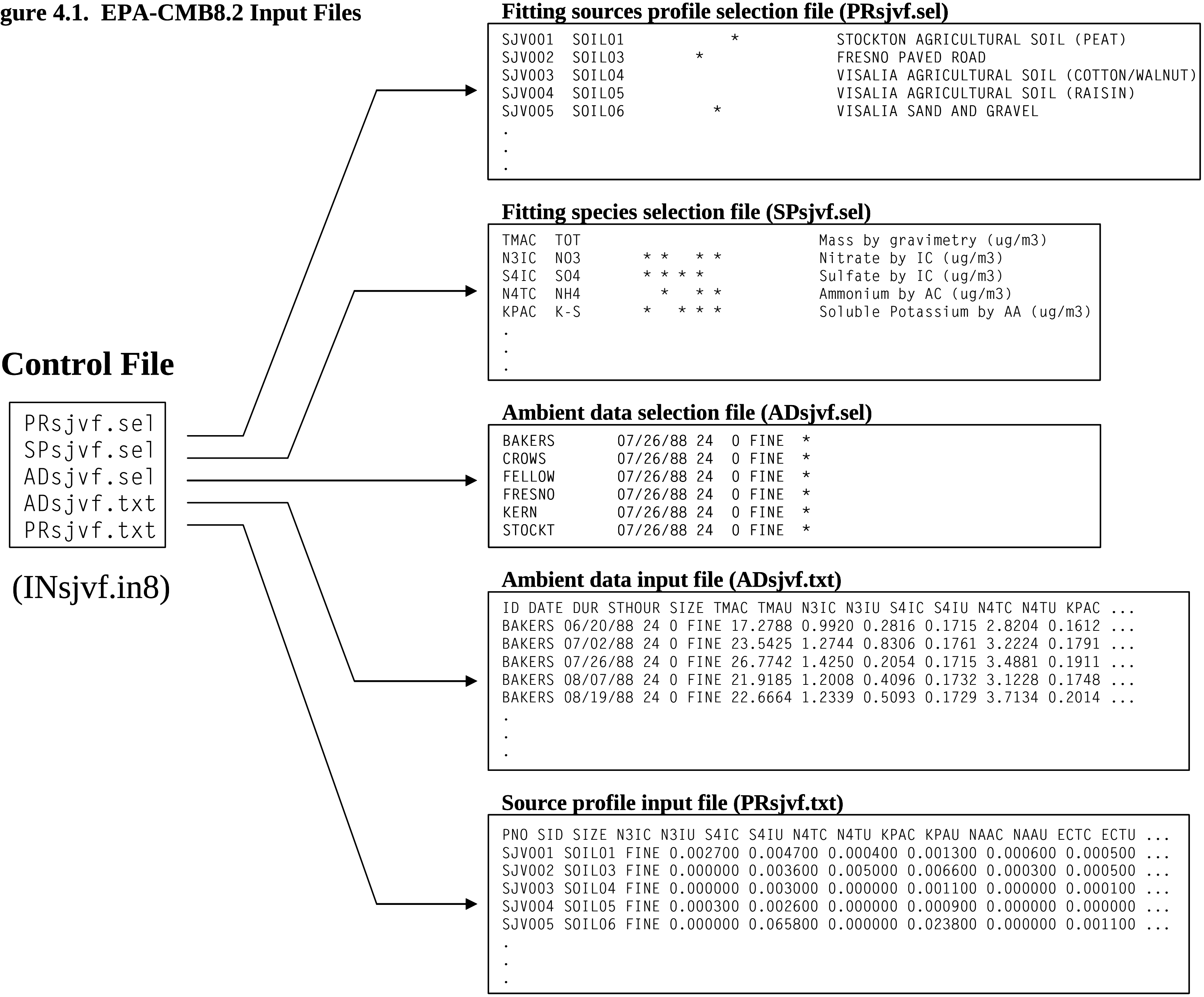
4.2.1 Control File: IN*.in8

This fixed format file contains a list of the names of EPA-CMB8.2 input data files, all of which must reside in the same directory that stores the Control File itself. This filename (e.g., *INsjvf.i*n8, exemplified in Figure 4.1) consists of five lines as shown below. These lines, in succession, contain the names of the files which are described in the following subsections. If a selection file is absent, the corresponding line in the Control File should be labeled with one or more characters, e.g., a series of asterisks ('******') - or any name that doesn't reside in the Control File directory. Here's an example:

```
PRsjvf.sel

SPsjvf.sel

*******

ADsjvf.csv

PRsjvf.csv
```

File name entries should be left justified and in the sequence shown. In EPA-CMB8.2, the only restriction on file names is that they are acceptable to the operating system. This means that extended file names may be used. The utility of the Control File is to save the effort of keying in the input filenames individually. If a Control File is not used at startup, EPA-CMB8.2 will accept the names of individual data input files *on the fly*, provided they are compatible with each other.

Figure 4.1. EPA-CMB8.2 Input Files



Fitting sources profile selection file (PRsjvf.sel)

```
SJV001  SOIL01            *           STOCKTON AGRICULTURAL SOIL (PEAT)
SJV002  SOIL03        *               FRESNO PAVED ROAD
SJV003  SOIL04                        VISALIA AGRICULTURAL SOIL (COTTON/WALNUT)
SJV004  SOIL05                        VISALIA AGRICULTURAL SOIL (RAISIN)
SJV005  SOIL06          *             VISALIA SAND AND GRAVEL
.
.
.
```

Fitting species selection file (SPsjvf.sel)

```
TMAC  TOT                           Mass by gravimetry (ug/m3)
N3IC  NO3       * *   * *           Nitrate by IC (ug/m3)
S4IC  SO4       * * * *             Sulfate by IC (ug/m3)
N4TC  NH4         *   * *           Ammonium by AC (ug/m3)
KPAC  K-S       *   * * *           Soluble Potassium by AA (ug/m3)
.
.
.
```

Ambient data selection file (ADsjvf.sel)

```
BAKERS       07/26/88 24  0 FINE  *
CROWS        07/26/88 24  0 FINE  *
FELLOW       07/26/88 24  0 FINE  *
FRESNO       07/26/88 24  0 FINE  *
KERN         07/26/88 24  0 FINE  *
STOCKT       07/26/88 24  0 FINE  *
```

Ambient data input file (ADsjvf.txt)

```
ID DATE DUR STHOUR SIZE TMAC TMAU N3IC N3IU S4IC S4IU N4TC N4TU KPAC ...
BAKERS 06/20/88 24 0 FINE 17.2788 0.9920 0.2816 0.1715 2.8204 0.1612 ...
BAKERS 07/02/88 24 0 FINE 23.5425 1.2744 0.8306 0.1761 3.2224 0.1791 ...
BAKERS 07/26/88 24 0 FINE 26.7742 1.4250 0.2054 0.1715 3.4881 0.1911 ...
BAKERS 08/07/88 24 0 FINE 21.9185 1.2008 0.4096 0.1732 3.1228 0.1748 ...
BAKERS 08/19/88 24 0 FINE 22.6664 1.2339 0.5093 0.1729 3.7134 0.2014 ...
.
.
.
```

Source profile input file (PRsjvf.txt)

```
PNO SID SIZE N3IC N3IU S4IC S4IU N4TC N4TU KPAC KPAU NAAC NAAU ECTC ECTU ...
SJV001 SOIL01 FINE 0.002700 0.004700 0.000400 0.001300 0.000600 0.000500 ...
SJV002 SOIL03 FINE 0.000000 0.003600 0.005000 0.006600 0.000300 0.000500 ...
SJV003 SOIL04 FINE 0.000000 0.003000 0.000000 0.001100 0.000000 0.000100 ...
SJV004 SOIL05 FINE 0.000300 0.002600 0.000000 0.000900 0.000000 0.000000 ...
SJV005 SOIL06 FINE 0.000000 0.065800 0.000000 0.023800 0.000000 0.001100 ...
.
.
.
```

4.2.2 Ambient Data Input File (AD*.csv, AD*.dbf, AD*.txt, AD*.wks)

Ambient data files may be formatted as comma-separated values in ASCII text (*.csv), xBASE (*.dbf), blank-delimited ASCII text (*.txt), or Lotus Worksheet (*.wks). The **csv** and **dbf** formats are preferred, as they are easier to prepare in spreadsheet (e.g., Microsoft Excel®, Corel QuatroPro®, Lotus 123) and data base (e.g., Microsoft Access®, dBASE) software than the other formats. The **wks** format creates large files and requires substantial translation time for EPA-CMB8.2 input and output, so it is the least desirable of these alternatives. The TXT format is most consistent with CMB7, so older CMB7 data files can be used for EPA-CMB8.2 input without modification. NB: if using TXT format, make sure the file's not tab-delimited! Recall from Section 4.1 that under some circumstances, EPA-CMB8.2 will create an ambient data input file in **txt** (ASCII) format "on the fly", and that it is actually this format that the model uses for calculations. The appropriate file extension must be associated with each format, as EPA-CMB8.2 recognizes the file type by this extension.

Examples of all supported file types are provided with the *EPA-CMB8.2test.ZIP* test data. Following is an example of the *ADsjvf.csv* file:

```
ID,DATE,DUR,STHOUR,SIZE,TMAC,TMAU,N3IC,N3IU,.....,PBXC,PBXU
BAKERS,06/20/88,24,0,FINE,17.2788,0.9920,0.2816,0.1715,.....,0.0236,0.0052
```

Of the first 5 field names in the header, note that as currently configured, EPA-CMB8.2 limits the first 2 field names to 4 characters; the last 3 fields must be named identically as indicated above. The "total" pair (e.g., TMAC & TMAU) preceding the species list differentiates the AD*.* header from that for PR*.*; there is no practical limitation for the pair names. All subsequent species names in the header are restricted to 6 characters.

The delimited forms of this file do not require fixed format spacing, only that a comma (or a blank character for TXT files) separate each field from prior and subsequent fields. The 1st line contains the field identifiers, followed by (beginning in field 6) the species codes, which occur in ***pairs*** (*concentration*, followed by *uncertainty*). Note that the name (*code*) for the concentration component of each pair must correspond identically (including *case*) with its counterpart in both the species selection files (SP*.sel) described in Section 4.2.4. and the source profile data input file (PR*.*) described in Section 4.2.3. Note that it is on this line where EPA-CMB8.2 gets the *labels* that appear in the species selection window (not from the species selection file, SP*.sel). Regardless of how the AD*.sel file is organized in terms of the sequence of ambient samples listed, in the sample selection screen they will appear according to the sequence dictated by the ambient data file (e.g., AD*.txt). Note also this line is the source of all species labels appearing in the Main Report. Beginning with the 2nd line are the actual data, one line per ambient sample, starting with the Site ID in Field 1. Note also that it is this ambient data input file that controls the sequence of ambient samples displayed in the samples selection window (not the samples selection file, AD*.sel). The records for each sample are formatted as follows:

Field 1: Site ID (up to 12 characters)
Field 2: Sampling date (up to 8 characters)
Field 3: Sample duration (up to 2 characters)
Field 4: Sample start hour (up to 2 characters)
Field 5: Particle size fraction (up to **6** characters)
Field 6: Total Mass concentration (any number of characters in integer, floating point, or exponential format)
Field 7: Uncertainty of total mass concentration (same format as Field 6)
Field 8+2n: Concentrations of chemical species (same format as Field 6), where n = 0, 1, 2, ...
Field 9+2n: Uncertainty of species concentrations (same format as Field 6), where n = 0, 1, 2, ...

EPA-CMB8.2 always assumes that Field 6 is the *total* mass concentration, and it does not use this as a fitting species. Uncertainty values in fields 7 and 9+2n are in the same units as the measured concentration values. For EPA-CMB8.2, the total number of ambient data records can reach into the thousands, limited only by computer memory. This makes it especially useful for examining multi-species hourly data obtained from automated gas chromatographs and time-of-flight mass spectrometers. For particles, up to four different size fraction identifiers may be used for the same sampling site and period, and the user can select mnemonics that suit individual purposes.[6] The size fraction names FINE and COARSE are reserved for the $PM_{2.5}$ and coarse particle (PM_{10} to $PM_{2.5}$) size fractions that are commonly measured in PM_{10} source assessment studies. As mentioned in Section 3.6.1, when these size fraction identifiers (i.e., ‘FINE’ & ‘COARSE’) are used, an additional report is produced that sums the *fine* and *coarse* source contribution estimate(s) (and uncertainties) to provide the estimate(s) for ‘Total’ (frequently = PM_{10}). Any other designator can be placed in the size column for non-segregated samples, such as “PM25” or “VOC”. Where semi-volatile materials are being apportioned, the particle (PART) and gas (GAS) phases are good designations.

Positive uncertainty values (> 0) must be assigned to all (non-missing) chemical concentrations used as fitting species. EPA-CMB8.2 will return an error message when it detects (in the input data file for ambient samples) a value for uncertainty that is less than or equal to zero for any species concentration value not flagged as missing. A species for which the concentration value is missing (i.e., invalid) cannot be used as a fitting species for that sample. Missing values for either concentrations or associated uncertainty must be designated by placing **-99.** in the appropriate field in the data file. If a species is to be flagged as “missing”, the value **-99.** must appear in the concentration field; substituting **-99.** for its uncertainty is optional because, as mentioned in Section 3.6.1, the species will automatically be removed from the calculation (though they will appear in the Main Report). For any species, when the fields for missing values species are substituted in this way, **-99.0000** will appear in applicable fields under “MEASURED” in the Main Report so long as decimal places displayed (Section 3.2) is set to a value < 5. When the species concentration is flagged with **-99.**, the value for CALCULATED/MEASURED in the Main Report will always be 0.00. The value for this diagnostic as well as for RESIDUAL/UNCERTAINTY will in this case be meaningless. See Appendix F for a discussion of EPA-CMB8.2's interpretation and treatment of input data conditions.

[6] If more that 4 size ranges are supplied for a given sampling site and period, the following error message from the DLL will appear (Appendix E):
Number of fitting sources = 0 > Number of fitting species = 0, followed by:
The number of fitting sources must be positive and <= number of fitting species.

4.2.3 Source Profile Input File (PR*.csv, PR*.dbf, PR*.txt, PR*.wks)

Source profile data files may be formatted as comma-separated values in ASCII text (**csv**), xBase (**dbf**), blank-delimited ASCII text (**txt**), or Lotus Worksheets (**wks**). The **csv** and **dbf** formats are the most portable and easily prepared. NB: if using TXT format, make sure the file's not tab-delimited! Recall from Section 4.1 that under some circumstances, EPA-CMB8.2 will create a source profile input file in **txt** (ASCII) format "on the fly", and that it is actually this format that the model uses for calculations. The appropriate file extension must be associated with each format, as EPA-CMB8.2 recognizes the file type by this extension. Examples of all supported file types are provided with EPA-CMB8.2's test case (*EPA-CMB82test.zip*). Following is an example of the *PRsjvf.csv* file:

```
PNO,SID,SIZE,N3IC,N3IU,.....,PBXC,PBXU
SJV001,SOIL01,FINE,0.002700,0.004700,.......,0.000000,0.000000
```

As mentioned for the AD*.* file *(Section 4.2.2), all species names in the header are restricted to 6 characters.

The delimited forms of this file do not require fixed format spacing, only that a comma (or a blank character for TXT files) separate each field from prior and subsequent fields. The 1st line contains the field identifiers followed by the species codes (fields > 3), which can be up to 6 alphameric characters in length. As with the ambient data input file, these codes appear in abundance / uncertainty *pairs*, and that the name (*code*) for the concentration component of each pair must correspond identically (including *case*) with its counterpart in both the species selection files (SP*.sel) described in Section 4.2.4. and the ambient data input file (PR*.*) described in Section 4.2.3. Beginning with the 2nd line are the actual data, one line per source profile. The first two fields are the *source code* and *mnemonic*, respectively. The source code must correspond identically with that used in the source profile selection file (PR*.sel) described in Section 4.2.4 (and note that these two fields are what appear as the left two columns in the sources selection window). The limitations on each field are:

Field 1:	Profile number or source code (up to 6 characters)
Field 2:	Source mnemonic (up to 8 characters)
Field 3:	Particle size fraction (up to **6** characters)
Field 4+2n:	Fraction of species in primary mass of source emissions (floating point or exponential format), where n = 0, 1, 2, ...
Field 5+2n:	Uncertainty of fraction of species in primary mass of source emissions (same format as Field 4), where n = 0, 1, 2, ...

If the name (*code*) used in field 1 does not match identically (including *case*) its counterpart in the source profile selection file, the source will not appear on the Sources selection screen when EPA-CMB8.2 is run, and thus not be available for use in a calculation.

Source profile abundances are expressed in fractions of total mass, not in percent. In the example input file snippet above for source profile SJV001, the NO_3^- abundance is 0.27% of the total mass, and its uncertainty is 0.47%. Unlike the ambient data file, the source profile input file does not contain a mass concentration field because all species abundances have been divided by this mass. The abundances and their uncertainties are typically *unitless*, but in certain applications can be represented as concentrations. The user will have to account for this latter case in interpreting the SCEs that are computed. The total number of records included depends on the number of species, number of sources, and size of the computer memory.

From one to four different size fraction identifiers may be used, but these must be the same as those used in the ambient data and sample selection files. Missing values for chemical species in source profile files can be replaced by a best estimate with a large uncertainty if they are to be used as fitting species. Missing values must be flagged with **–99.** if the species is not intended for use as fitting species. Species with mass fractions so flagged will automatically be removed from the calculation (though they will appear in the Main Report). While uncertainty values for species in source profiles are allowed to be ≤ 0.0, some effort should be made to supply values > 0.0. Default values of 0.0 for the fraction and 0.0001 to 0.01 for the uncertainty are often chosen for species that are expected to be present in small abundances. This indicates that the species is present in source emissions at a concentration less than 0.01% to 1%. A smaller value may be appropriate for certain source types and species. See Appendix F for a discussion of EPA-CMB8.2's interpretation and treatment of input data conditions.

In certain cases in which an uncertainty value >0 is applied to a species whose abundance = 0, the uncertainty represents the lower quantifiable limit of the measurement. A zero value for abundance means that the true value is something between zero and the detection limit. Zeros are important in some source profiles when they occur in other source profiles. This makes that species a marker for the source in which it occurs. The uncertainty for the zero adds to the effective variance weighting. If this uncertainty is high and the source contribution estimate is high, the influence of that species is reduced by the weighting.

4.2.4 Source (PR*.sel), Species (SP*.sel), and Sample Selection (AD*.sel) Input Files

The *optional* source, species and sample selection files provide initial selection arrays that do not have to be entered from the program each time a EPA-CMB8.2 session is begun. These files limit the profiles, species, and ambient data records to those listed in the selection file arrays, even though a larger number may be included in the ambient and source profile data files. This means that the data files need not be edited when only subsets of variables are desired for a specific EPA-CMB8.2 modeling session. Variable definitions can also be documented in these files (comment fields). The selection files also dictate the sequence in which species and sources appear in the output. The total species list is that common to the species selection file, the ambient data file, and the source profiles file. The total sources list is that common to the sources selection file and the source profiles file. The total ambient samples list is that common to the ambient sample selection file and the ambient data file. As mentioned in Sections 3.3, 3.4 & 3.5, if any of the selection (*.sel) files are modified during a session, EPA-CMB-8.2 will detect the change and, upon program exit, allow the opportunity to update (overwrite) these files. The file may also be conveniently renamed.

Following is an example of the *source profile* selection file *PRsjvf.sel*. In this selection file, as well as the two that follow, the first two lines are shown only for field location (they are not part of the file itself):

```
0        1         2         3         4
1234567890123456789012345678901234567890
SJV001  SOIL01            *           STOCKTON AGRICULTURAL SOIL (PEAT)
SJV002  SOIL03        *               FRESNO PAVED ROAD
SJV003  SOIL04                        VISALIA AG SOIL (COTTON/WALNUT)
SJV004  SOIL05                        VISALIA AGRICULTURAL SOIL (RAISIN)
SJV005  SOIL06          *             VISALIA SAND AND GRAVEL
SJV006  SOIL07                        VISALIA URBAN UNPAVED
SJV007  SOIL08                        VISALIA PAVED ROAD
SJV008  SOIL09                        BAKERSFIELD AGRICULTURAL SOIL, ALKALINE
SJV009  SOIL10                        BAKERSFIELD AG SOIL, SANDY LOAM
SJV010  SOIL11                        BAKERSFIELD UNPAVED ROAD (OILDALE)
SJV011  SOIL12    *                   BAKERSFIELD PAVED ROAD
SJV012  SOIL13                        BAKERSFIELD WINDBLOWN URBAN UNPAVED
SJV013  SOIL14          *             BAKERSFIELD AG SOIL, WASCO SANDY LOAM
SJV014  SOIL15                        BAKERSFIELD AG SOIL, CAJON SANDY LOAM
SJV015  SOIL16                        BAKERSFIELD UNPAVED ROAD (RESIDENTIAL)
SJV016  SOIL17      *                 TAFT UNPAVED ROAD
SJV017  BAMAJC    * * * * *           BAKERSFIED CORDWOOD, MAJESTIC FIREPLACE
SJV018  MAMAJC                        MAMMOTH LAKES WOOD, MAJESTIC FIREPLACE
SJV019  MAFISC                        BAKERSFIELD WOOD, FISHER MAMA BEAR STOVE
SJV020  MADIEC                        MAMMOTH LAKES DIESEL TOUR BUSES (IDLING)
SJV021  BAAGBC                        BAKERSFIELD AG BURN (WHEAT AND BARLEY)
SJV022  ELAGBC                        EL CENTRO AGRI. BURN (WHEAT)
SJV023  FRCONC                        FRESNO HIGHWAY 40 CONSTRUCTION
SJV024  STAGBC                        STOCKTON AGRI. BURN (WHEAT)
SJV025  VIAGBC                        VISALIA AGRI BURN (WHEAT)
SJV026  VIDAIC                        VISALIA DAIRY/FEEDLOT DUST
SJV027  SFCRUC    * * * * *           SANTA FE CRUDE BOILER
```

```
SJV028  CHCRUC                        CHEVRON RACETRACK CRUDE BOILER
SJV029  MOTIBC                        MODESTO TIRE POWER PLANT
SJV030  SCRRFC                        STANISLAUS RESOURCE RECOVERY FACILITY
SJV031  CDCEMT                        NBS CEMENT DUST
SJV032  CDRKCR                        ROCK CRUSHING 1987 SCAB
SJV033  CDSAPL                        SANDBLASTING AND PLASTERING
SJV034  MARINE                        MARINE
SJV035  MOVES1                        MOVES-SS(NEA-E,WOB,T42,TVMT)
SJV036  MOVES2    * * * * *           MOVES-SS(NEA-E,WOB,WOT,TVMT)
SJV038  MOVES3                        MOVES-SCAB(ARB-E,WOB,WOT,CM)
SJV039  MOVES4                        MOVES-SCAB(NEA-E,WOB,WOT,CM)
SJV040  MOVES5                        MOVES-SCAB(NEA-E,WB1,T42,CM)
SJV041  MPGYPU                        GYPSUM DUST, (TOTAL FROM CASO4)
SJV051  AMSUL     * * * * *           AMMONIUM SULFATE
SJV052  AMBSUL                        AMMONIUM BISULFATE
SJV053  H2SO4                         SULFURIC ACID
SJV054  AMNIT     * * * * *           AMMONIUM NITRATE
SJV055  HNO3                          NITRIC ACID
SJV056  NANO3       * * * *           SODIUM NITRATE
SJV057  MVDEN1                        50% DIESEL, 20% LEADED, 30% UNLEADED
SJV058  MVDEN2                        75% DIESEL, 15% LEADED, 10% UNLEADED
SJV059  MVDEN3                        85% DIESEL, 10% LEADED,  5% UNLEADED
SJV060  OC                            PURE ORGANIC CARBON
SJV061  LIME                          LIMESTONE
SJV062  SOIL28                        CROWS LANDING AGRI.
SJV063  SOIL29                        CROWS LANDING PAVED ROAD
SJV064  SOIL30                        KERN UNPAVED ROAD
SJV065  SOIL31                        KERN AGRI.
```

As with the source profile input file, the first fields are the source profile code and mnemonic, respectively. A source code with up to six characters is located in Columns 1 to 6. Columns 9 to 16 are available for an eight-character profile name (mnemonic). The names used in both fields must correspond identically with those used in the source profile input file (Section 4.2.3). Note that these two fields are what appear as the left two columns in the sources selection window. Asterisks in Column 19, 21, 23, 25, 27, 29, 31, 33, 35 and 37 designate initial arrays (independent sets) of fitting sources when EPA-CMB8.2 is executed. The maximum number of sources is essentially unlimited. Comments can be added to this file beginning in column 39 to document the source profiles.

For each source listed in the selection file, some information must appear in either (1) the mnemonic field, (2) the selection array field (at least one column must be tagged with ‘*****’), or (3) the comment field in order for the source to appear on the selection screen in EPA-CMB8.2. If the name used in field 1 does not match identically (including *case*) its counterpart in the source profile input file, or if any of the aforementioned conditions is not met, the source will not appear on the Sources selection screen when EPA-CMB8.2 is run, and thus not be available for use in a calculation.

Following is an example of the *species* selection file *SPsjvf.sel*:

```
         1         2         3         4
1234567890123456789012345678901234567890
  TMAC  TOT                             Mass by gravimetry (ug/m3)
  N3IC  NO3       * *   *               Nitrate by IC (ug/m3)
  S4IC  SO4       * *   *               Sulfate by IC (ug/m3)
  N4TC  NH4       * *   *               Ammonium by AC (ug/m3)
  KPAC  K-S       * *   *               Soluble Potassium by AA (ug/m3)
  NAAC  NA        * *   *               Soluble Sodium by AA (ug/m3)
  ECTC  EC        * *   *               Elemental Carbon by TOR (ug/m3)
  OCTC  OC        * *   *               Organic Carbon by TOR (ug/m3)
  ALXC  AL        * * * *               Aluminum by XRF (ug/m3)
  SIXC  SI        * * * *               Silicon by XRF (ug/m3)
  SUXC  S                               Sulfur by XRF (ug/m3)
  CLXC  CL        * * * *               Chloride by XRF (ug/m3)
  KPXC  K         * * * *               Potassium by XRF (ug/m3)
  CAXC  CA        * * * *               Calcium by XRF (ug/m3)
  TIXC  TI        * * * *               Titanium by XRF (ug/m3)
  VAXC  V         * * * *               Vanadium by XRF (ug/m3)
  CRXC  CR        * * * *               Chromium by XRF (ug/m3)
  MNXC  MN        * * * *               Manganese by XRF (ug/m3)
  FEXC  FE        * * * *               Iron by XRF (ug/m3)
  NIXC  NI        * * * *               Nickel by XRF (ug/m3)
  CUXC  CU            *                 Copper by XRF (ug/m3)
  ZNXC  ZN            *                 Zinc by XRF (ug/m3)
  BRXC  BR        *   * *               Bromine by XRF (ug/m3)
  PBXC  PB        * * * *               Lead by XRF (ug/m3)
```

A species *code* with up to six characters is located in Columns 1 to 6, and must correspond identically with that used in both the ambient data input file (Section 4.2.2) and source profile input data file (Section 4.2.3). Columns 9 to 16 are available for an eight-character species name, which is optional (this is the only place they can be installed). Note that these two fields are what appear as the left two columns in the species selection window. Asterisks in Column 19, 21, 23, 25, 27, 29, 31, 33, 35 and 37 designate initial arrays (independent sets) of fitting species when EPA-CMB8.2 is executed. The maximum number of species is essentially unlimited. Comments can be added to this file beginning in column 39 to document the meaning and units of the chemical components; this is the only place they can be installed.

As mentioned above for each source listed in the source selection file, for each species listed in the species selection file, some information must appear in either (1) the name field, (2) the selection array field (at least one column must be tagged with '*'), or (3) the comment field in order for the species to appear on the selection screen in EPA-CMB8.2. This constraint notwithstanding, the name field is truly optional. Note that if the species code in field 1 doesn't match exactly (including *case*) its counterpart in ***both*** the ambient data input file and the source profile input file, or is missing, the species will **not** appear in the selection screen when the model is run regardless of any other condition. If the species does not appear on the selection screen when EPA-CMB8.2 is run, it will not be available for use in a calculation.

For the ambient data records (sample) selection file, columns 1 through 12 are for the Site ID, columns 14 through 21 are for the date, columns 23 and 24 for the sample duration, columns 26 and 27 for the sample start hour, and columns 29-33 for the particle size fraction, if appropriate. Intermediate columns should be blank. An asterisk in column 36 initializes (selects) a record for processing.

Following is an example of the ambient data records selection file *ADsjvf.sel*:

```
          1         2         3         4
12345678980123456789012345678901234567890
BAKERS        07/26/88 24  0 FINE    *
CROWS         07/26/88 24  0 FINE    *
FELLOW        07/26/88 24  0 FINE    *
FRESNO        07/26/88 24  0 FINE    *
KERN          07/26/88 24  0 FINE    *
STOCKT        07/26/88 24  0 FINE    *
```

Note that the tagging asterisks are placed in column 36 (not 35, as in CMB8.0). This change is to maintain spacing for the new 6-character field for size fraction (i.e., COARSE). Note also that regardless of the sequence of ambient samples listed, they will appear in the sample selection window according to the sequence dictated by the ambient data input file (e.g., AD*.txt); see Section 4.2.2.

4.3 Output Files

Both general report (a listing) and data base output files are produced by EPA-CMB8.2.

4.3.1 Report Output File

The report output file is generated from the Main Report screen (Section 3.6.1) and presents the source contribution estimates, standard errors, model performance measures, and measured and calculated chemical concentrations for each sample (Section 3.6). The report written to the output file is identical to that which appears in the Output window during an interactive modeling session. It is in ASCII text format and can be imported into word processing programs to document the source contributions calculated for each sample. All information needed to independently repeat the source apportionment is contained in this report, including an echo of the *.in8 input file that was used in its generation. Examples of the report are shown in Section 5 & 6.

4.3.2 Data Base Output File

The data base (spreadsheet-type) output file records the contribution of each source-type to a particular species in a single data record, one record per species. Sample identifiers and model performance measures are also included in each record. As EPA-CMB8.2 is currently configured, this file may be generated in blank-delimited (*.txt) or comma-separated value (*.csv) formats.[7] The file structure is:

[7] Binary-type formats (e.g., DBF & WKS) are disabled in EPA-CMB8.2.

Field 1:	Species Code
Field 2:	Species Name
Field 3:	Fitting flag; a '*' indicates a fitting species, while a '_' indicates a floating species
Field 4:	Sampling site identifier
Field 5:	Sampling date
Field 6:	Sample start hour
Field 7:	Sample duration
Field 8:	Particle size fraction
Field 9:	Measured species concentration
Field 10:	Uncertainty of measured species concentration
Field 11:	Calculated species concentration
Field 12	Uncertainty of calculated species concentration
Field 13:	R-square value
Field 14:	Chi-square value
Field 15:	Percent of measured mass
Field 16+2n:	Source contribution estimate, n = 0, 1, 2, ...
Field 17+2n:	Standard error of source contribution estimate, n = 0, 1, 2, ...

Fields 1, 2, and 4 through 8 record the sample information. Fields 3 and 11 through 15 provide information about the EPA-CMB8.2 calculation.[8] The remaining fields correspond to each source profile in the PR*.* data file and contain the source contribution estimates and standard errors for these sources. A value of -99. is recorded when a profile was not used in the calculation.

The first record in this output file contains the field identifiers. All subsequent records contain data. Fields 16+2n and 17+2n are labeled with source codes and source names, respectively.

4.4 Creating Data Input Files

When using a Control File, note that if several data input files (e.g., AD*.* & PR*.*) exist for which the only difference is *format*, EPA-CMB8.2 assumes that the files will be named identically except for the extension. For blank-delimited and comma-separated value input files, there are three common methods of creating EPA-CMB8.2 input files: 1) manually entering the data in the correct format using a text editor or word processing program; 2) editing existing input files with a text editor or word processing program; or 3) transferring files from computerized data bases.

A text editor or word processor in text mode can be used to type entire input files. It is best to bring the example files into the editor, then insert the new values in the same locations as the existing values by using the editor in *TYPEOVER* mode. Spaces between fields should be entered with the space bar; tabs should not be set. Each line should be terminated with the

[8] Fields 11 & 12 (calculated species concentration/uncertainty) are a new feature in EPA-CMB8.2 that make the output data file more robust.

ENTER key rather than using the wraparound feature present in many editors. No blank lines at the end of the file should be present. Completed files should be saved with an appropriate filename. A wide variety of very good ASCII text editors are available for creating and manipulating input files.

When data files have been prepared for other applications (e.g., source profiles may be common to several different data sets), these files may be cut and pasted to produce the needed input data files. Owing to differences in individual editing programs, the user is should consult the manual for the editing program to be used for directions on opening a copy of the existing file, deleting and adding material, saving the changes, and renaming the file. When using word processors (e.g., MSWord or WordPerfect), the files must be saved as DOS text with line breaks. Otherwise, extra information is included in the files that EPA-CMB8.2 cannot read.

Input files are most easily produced with spreadsheet or data base software. Many source profile and ambient data sets are available in data base management formats. Selections of data, field names, and data structure can be easily made by the data base software. These can be saved using the Save As or Export selections from the File menu. The CSV, DBF, and WKS formats can be selected from the “Save as type” option box that usually appears in the “Save As” window.

4.5 Reading Output Files

Main Report (ASCII) text files can be read directly into a word processing application where the detailed output for each sample can usually be displayed on a single page with columns aligned using a *non-proportional* font, e.g., a New Courier or Letter Gothic 8-point to 10-point. Such a fixed pitch font in which every character occupies the same space is necessary for columns to be correctly aligned. As mentioned in Section 4.3.2, data base (spreadsheet-type) output files from EPA-CMB8.2 are formatted either as ASCII (default) or comma-separated value (CSV), and this selection must be made on the Options screen before a calculation is made. The output file will be stored in the directory in which the Control File resides (which can, of course, be changed in the browse dialog). These output files can be opened directly by data base or spreadsheet applications, e.g., Microsoft’s Excel®, and from there exported into any of several different formats, e.g., **xls**. The contents of the EPA-CMB8.2 output screen can also be selected and copied to the clipboard for pasting into other Windows® applications. Graphs made with EPA-CMB8.2 (Section 3.3 & 3.5) can be printed directly from the screen or copied to the Windows® clipboard via copy command, then pasted into a text box or frame in a word processing application.

5. USING EPA-CMB8.2: TEST CASES

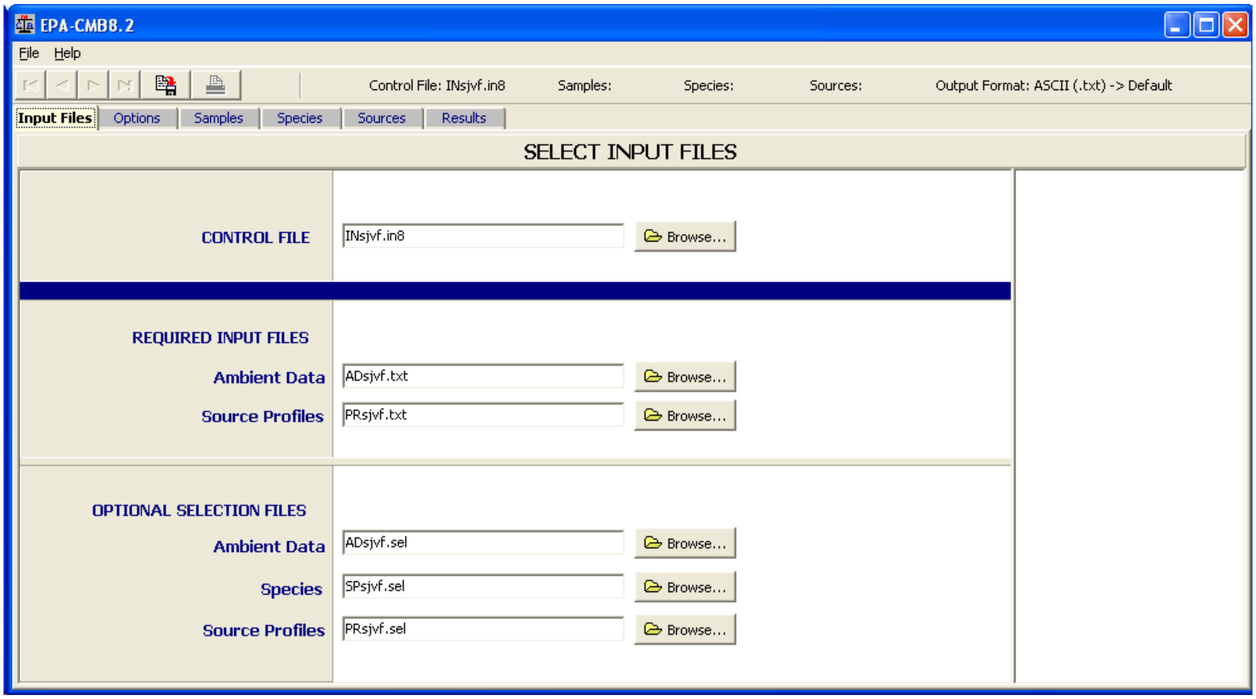
This section illustrates EPA-CMB8.2 commands and operations using the San Joaquin Valley, CA, $PM_{2.5}$ data set. Other test data sets (Section 2.2) are available on EPA's modeling website (www.epa.gov/scram001) as examples of additional data base file formats and for independent practice in EPA-CMB8.2 application and validation. These examples are most effective when accompanied by actual application of EPA-CMB8.2 on the user's computer.

5.1 Starting EPA-CMB8.2

Start EPA-CMB8.2 by double-clicking on the EPA-CMB8.2 icon on your desktop (or by selecting it from the Start menu). The first screen that will appear is depicted in Figure 3.1. As discussed in Section 3.1, at this point EPA-CMB8.2 requires the user to choose between using a stored Control File or to choose input files (which are otherwise directed by the Control File) individually.

5.1.1 Control File Operation

Select the Control File mode and use the browse feature to access the test case folder. A screen like the one shown in Figure 3.2 will then appear. Select INsjvf.in8 and open it. A screen as shown in Figure 5.1 will appear. Select the options screen and use the defaults as shown in Figure 3.4. Choose the Select Samples screen and click View Selected to present only the tagged samples. Note for each sample are the following attributes: selection status, site,



Control File: INsjvf.in8 Samples: Species: Sources: Output Format: ASCII (.txt) -> Default

SELECT INPUT FILES

CONTROL FILE: INsjvf.in8

REQUIRED INPUT FILES

Ambient Data: ADsjvf.txt

Source Profiles: PRsjvf.txt

OPTIONAL SELECTION FILES

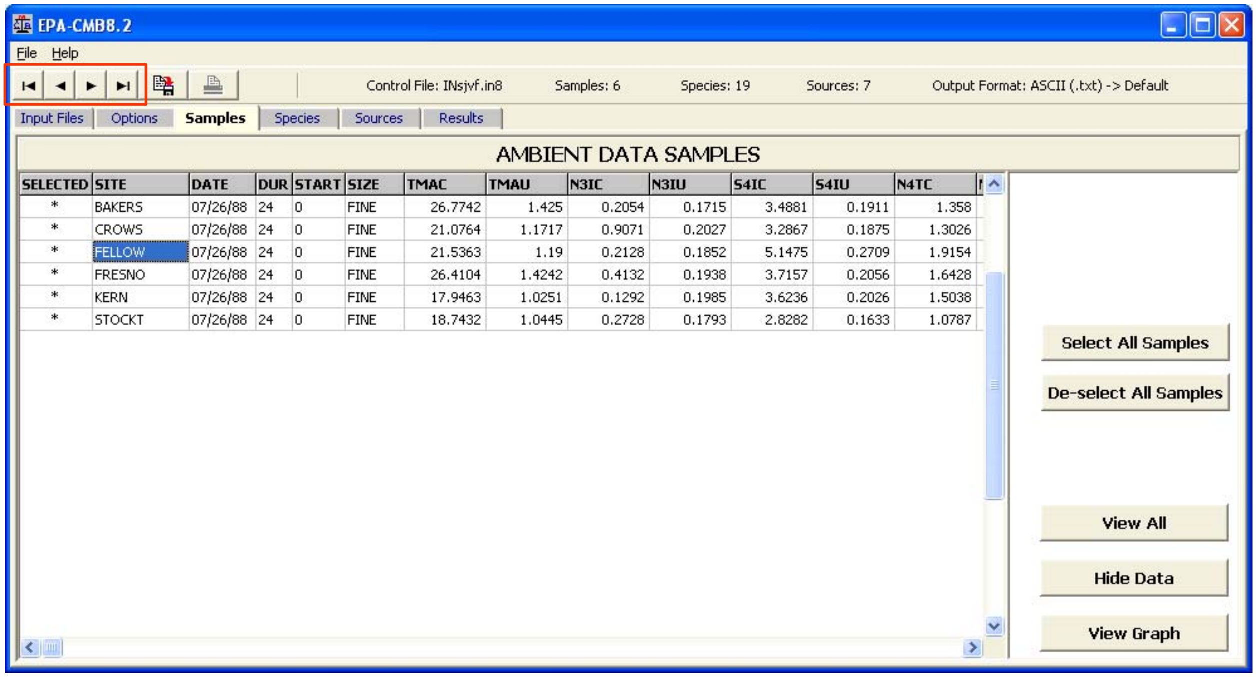
Ambient Data: ADsjvf.sel

Species: SPsjvf.sel

Source Profiles: PRsjvf.sel

Figure 5.1 Input Files for Test Case - Control File Mode

sample date, sampling duration (hours), start hour, and size fraction (Figure 5.2). This is followed by a series of fields for measured concentration and uncertainty for total mass and for constituent species in the sample. Measured values for species may be viewed by moving to the right through the field using the horizontal scroll, and these values may also be hidden (Hide Data).



Control File: INsjvf.in8 Samples: 6 Species: 19 Sources: 7 Output Format: ASCII (.txt) -> Default

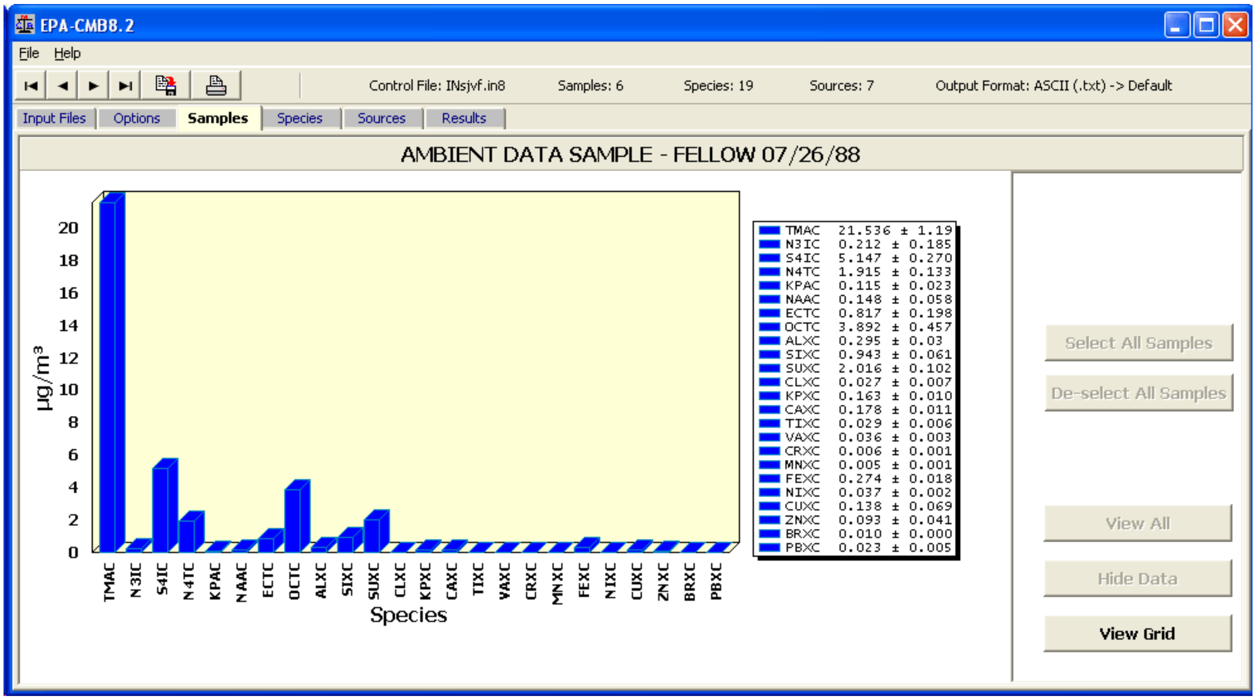
AMBIENT DATA SAMPLES

SELECTED	SITE	DATE	DUR	START	SIZE	TMAC	TMAU	N3IC	N3IU	S4IC	S4IU	N4TC
*	BAKERS	07/26/88	24	0	FINE	26.7742	1.425	0.2054	0.1715	3.4881	0.1911	1.358
*	CROWS	07/26/88	24	0	FINE	21.0764	1.1717	0.9071	0.2027	3.2867	0.1875	1.3026
*	FELLOW	07/26/88	24	0	FINE	21.5363	1.19	0.2128	0.1852	5.1475	0.2709	1.9154
*	FRESNO	07/26/88	24	0	FINE	26.4104	1.4242	0.4132	0.1938	3.7157	0.2056	1.6428
*	KERN	07/26/88	24	0	FINE	17.9463	1.0251	0.1292	0.1985	3.6236	0.2026	1.5038
*	STOCKT	07/26/88	24	0	FINE	18.7432	1.0445	0.2728	0.1793	2.8282	0.1633	1.0787

Figure 5.2 Ambient Samples for Test Case

Note that the "VCR" control buttons (red rectangle on toolbar) can be used to facilitate navigation through the selection field, for purposes of selecting or de-selecting individual samples.

Highlight any cell for the ambient sample for Fellows, CA (FELLOWS 07/26/88), as shown in Figure 5.2, and see a graphical representation of the sample (View Graph), as shown in Figure 5.3. This graph may be printed from the print icon on that screen. Note that you may step through the list of samples by using the "VCR" control buttons to view their graphical representations. You may then return to the Sample screen (View Grid).



Control File: INsjvf.in8 Samples: 6 Species: 19 Sources: 7 Output Format: ASCII (.txt) -> Default

AMBIENT DATA SAMPLE - FELLOW 07/26/88

▬	TMAC	21.536 ± 1.19
▬	N3IC	0.212 ± 0.185
▬	S4IC	5.147 ± 0.270
▬	N4TC	1.915 ± 0.133
▬	KPAC	0.115 ± 0.023
▬	NAAC	0.148 ± 0.058
▬	ECTC	0.817 ± 0.198
▬	OCTC	3.892 ± 0.457
▬	ALXC	0.295 ± 0.03
▬	SIXC	0.943 ± 0.061
▬	SUXC	2.016 ± 0.102
▬	CLXC	0.027 ± 0.007
▬	KPXC	0.163 ± 0.010
▬	CAXC	0.178 ± 0.011
▬	TIXC	0.029 ± 0.006
▬	VAXC	0.036 ± 0.003
▬	CRXC	0.006 ± 0.001
▬	MNXC	0.005 ± 0.001
▬	FEXC	0.274 ± 0.018
▬	NIXC	0.037 ± 0.002
▬	CUXC	0.138 ± 0.069
▬	ZNXC	0.093 ± 0.041
▬	BRXC	0.010 ± 0.000
▬	PBXC	0.023 ± 0.005

Figure 5.3 Speciation Graph of Ambient Sample (Test Case)

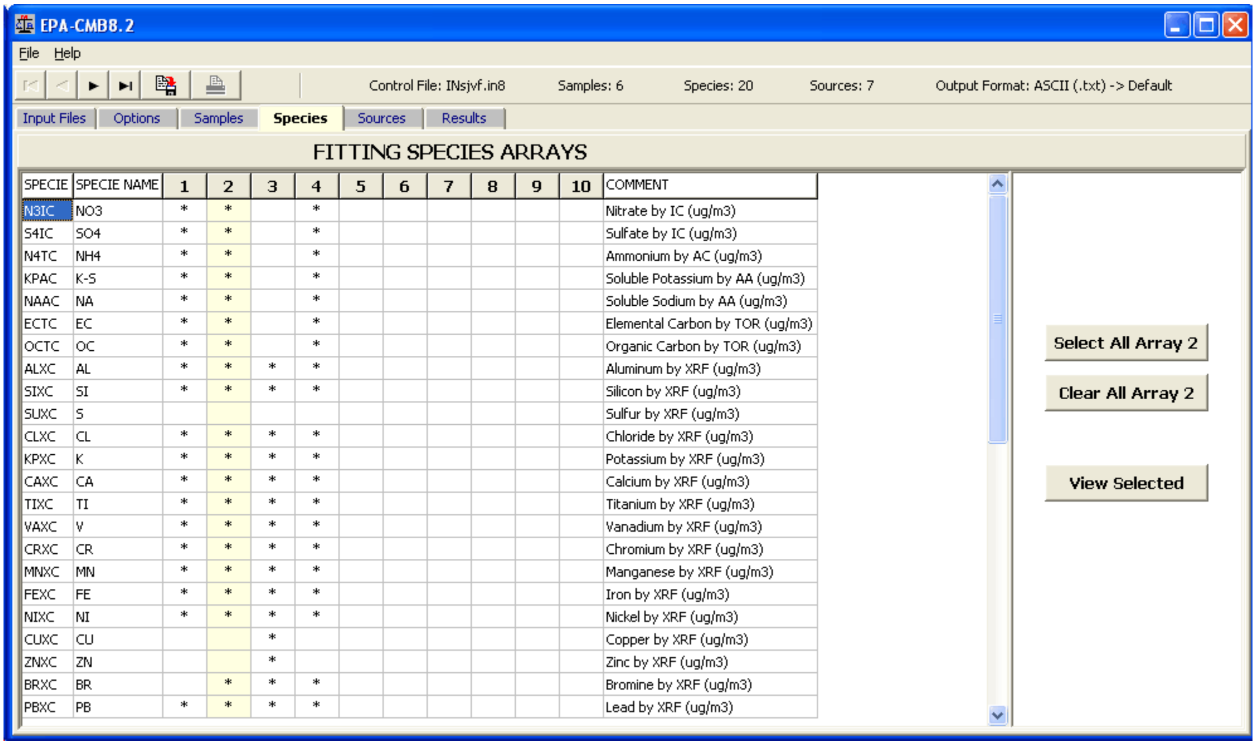
Finally, note also that the counters on the status bar (top) reflect the number of ambient samples, fitting species, or source profiles that are selected on their respective screens.[9]

Go to the Species screen to select an array of fitting species for the calculation (Figure 5.4). As mentioned in Section 3.4, Array 1 will be always be initiated (highlighted) by default at startup. Different arrays are selected by clicking on the index (integer) that labels each array. Click on 2 to select the 20 fitting species "tagged" in Array 2 (yellow shading on screen view).

The number of species tagged in a selected array is indicated by the counter on the status bar at the top of the screen. If only tagged species are of interest, choose View Selected to show tagged species for the selected array (which toggles back via View All). Comments will appear if present in the selection file (SP*.sel).

Note that the "VCR" control buttons operate in the same way as in the ambient samples screen to facilitate navigation through a series of species in a selected array.

[9] Note that at startup, initial counter values will appear as directed from any selection files (PR*.sel, SP*.sel, AD*.sel) that have been loaded.



Control File: INsjvf.in8 Samples: 6 Species: 20 Sources: 7 Output Format: ASCII (.txt) -> Default

Input Files | Options | Samples | Species | Sources | Results

FITTING SPECIES ARRAYS

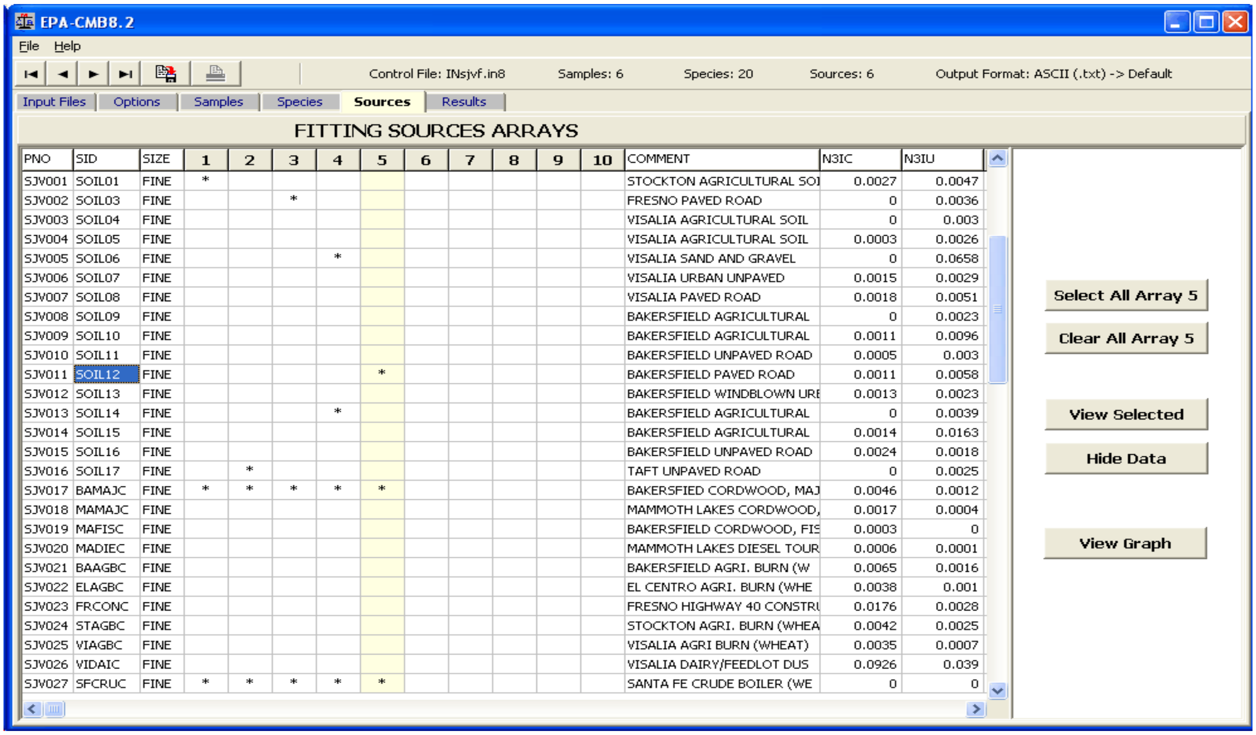
SPECIE	SPECIE NAME	1	2	3	4	5	6	7	8	9	10	COMMENT
N3IC	NO3	*	*		*							Nitrate by IC (ug/m3)
S4IC	SO4	*	*		*							Sulfate by IC (ug/m3)
N4TC	NH4	*	*		*							Ammonium by AC (ug/m3)
KPAC	K-S	*	*		*							Soluble Potassium by AA (ug/m3)
NAAC	NA	*	*		*							Soluble Sodium by AA (ug/m3)
ECTC	EC	*	*		*							Elemental Carbon by TOR (ug/m3)
OCTC	OC	*	*		*							Organic Carbon by TOR (ug/m3)
ALXC	AL	*	*	*	*							Aluminum by XRF (ug/m3)
SIXC	SI	*	*	*	*							Silicon by XRF (ug/m3)
SUXC	S											Sulfur by XRF (ug/m3)
CLXC	CL	*	*	*	*							Chloride by XRF (ug/m3)
KPXC	K	*	*	*	*							Potassium by XRF (ug/m3)
CAXC	CA	*	*	*	*							Calcium by XRF (ug/m3)
TIXC	TI	*	*	*	*							Titanium by XRF (ug/m3)
VAXC	V	*	*	*	*							Vanadium by XRF (ug/m3)
CRXC	CR	*	*	*	*							Chromium by XRF (ug/m3)
MNXC	MN	*	*	*	*							Manganese by XRF (ug/m3)
FEXC	FE	*	*	*	*							Iron by XRF (ug/m3)
NIXC	NI	*	*	*	*							Nickel by XRF (ug/m3)
CUXC	CU			*								Copper by XRF (ug/m3)
ZNXC	ZN			*								Zinc by XRF (ug/m3)
BRXC	BR		*	*	*							Bromine by XRF (ug/m3)
PBXC	PB	*	*	*	*							Lead by XRF (ug/m3)

Figure 5.4 Fitting Species Selection for Test Case

Go to the Sources screen to select an array of fitting sources. As mentioned in Section 3.4, Array 1 will be always be initiated (highlighted in) by default at startup. For fitting sources, use the 6 shown in Array 5 on the screen (Figure 5.5) by clicking on its index (label) at the top. Note that the “VCR” buttons operate in the same way as in the ambient samples screen to facilitate navigation through a series of species in a given array.

Highlight any cell for any source and see a graphical representation of the profile (View Graph), as shown in Figure 5.6 for Soil 12. This graph may be printed from the print icon on that screen. As with the graphs of ambient samples, note that you may step through the list of sources in a selected array by using the “VCR” control buttons to view their graphical representations. You may then return to the Source screen (View Grid).

The number of sources “tagged” in a selected array is indicated by the counter on the status bar at the top of the screen. If only tagged sources are of interest, choose View Selected to show tagged sources for the selected array (which toggles back via View All). Comments will appear if present in the selection file (PR*.sel). The comments field is followed by a series of fields for abundance and uncertainty for all species in the source. The View Selected button shows only the tagged sources(s) for the array selected, and values for species may be viewed by exploring the field to the right by using the horizontal scroll. These values may also be hidden (Hide Data).

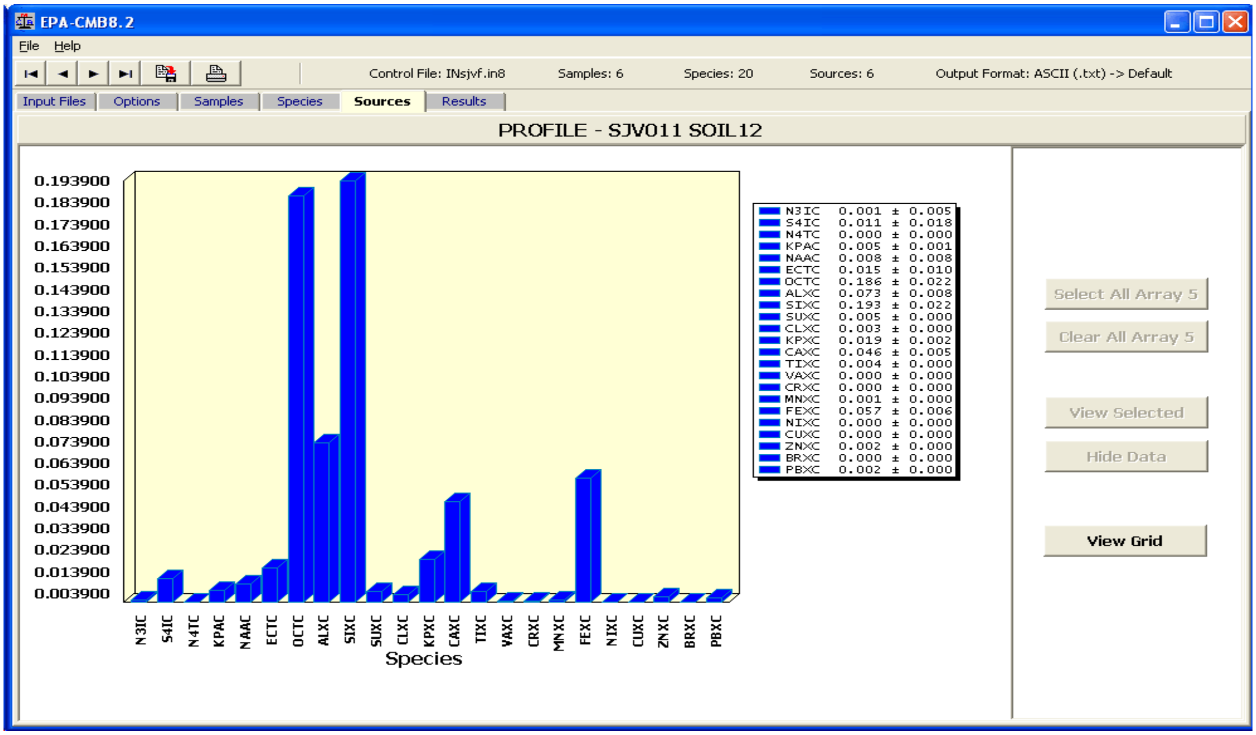


Control File: INsjvf.in8 Samples: 6 Species: 20 Sources: 6 Output Format: ASCII (.txt) -> Default

FITTING SOURCES ARRAYS

PNO	SID	SIZE	1	2	3	4	5	6	7	8	9	10	COMMENT	N3IC	N3IU
SJV001	SOIL01	FINE	*										STOCKTON AGRICULTURAL SOI	0.0027	0.0047
SJV002	SOIL03	FINE			*								FRESNO PAVED ROAD	0	0.0036
SJV003	SOIL04	FINE											VISALIA AGRICULTURAL SOIL	0	0.003
SJV004	SOIL05	FINE											VISALIA AGRICULTURAL SOIL	0.0003	0.0026
SJV005	SOIL06	FINE				*							VISALIA SAND AND GRAVEL	0	0.0658
SJV006	SOIL07	FINE											VISALIA URBAN UNPAVED	0.0015	0.0029
SJV007	SOIL08	FINE											VISALIA PAVED ROAD	0.0018	0.0051
SJV008	SOIL09	FINE											BAKERSFIELD AGRICULTURAL	0	0.0023
SJV009	SOIL10	FINE											BAKERSFIELD AGRICULTURAL	0.0011	0.0096
SJV010	SOIL11	FINE											BAKERSFIELD UNPAVED ROAD	0.0005	0.003
SJV011	SOIL12	FINE					*						BAKERSFIELD PAVED ROAD	0.0011	0.0058
SJV012	SOIL13	FINE											BAKERSFIELD WINDBLOWN URE	0.0013	0.0023
SJV013	SOIL14	FINE				*							BAKERSFIELD AGRICULTURAL	0	0.0039
SJV014	SOIL15	FINE											BAKERSFIELD AGRICULTURAL	0.0014	0.0163
SJV015	SOIL16	FINE											BAKERSFIELD UNPAVED ROAD	0.0024	0.0018
SJV016	SOIL17	FINE		*									TAFT UNPAVED ROAD	0	0.0025
SJV017	BAMAJC	FINE	*	*	*	*	*						BAKERSFIED CORDWOOD, MAJ	0.0046	0.0012
SJV018	MAMAJC	FINE											MAMMOTH LAKES CORDWOOD,	0.0017	0.0004
SJV019	MAFISC	FINE											BAKERSFIELD CORDWOOD, FIS	0.0003	0
SJV020	MADIEC	FINE											MAMMOTH LAKES DIESEL TOUR	0.0006	0.0001
SJV021	BAAGBC	FINE											BAKERSFIELD AGRI. BURN (W	0.0065	0.0016
SJV022	ELAGBC	FINE											EL CENTRO AGRI. BURN (WHE	0.0038	0.001
SJV023	FRCONC	FINE											FRESNO HIGHWAY 40 CONSTRU	0.0176	0.0028
SJV024	STAGBC	FINE											STOCKTON AGRI. BURN (WHEA	0.0042	0.0025
SJV025	VIAGBC	FINE											VISALIA AGRI BURN (WHEAT)	0.0035	0.0007
SJV026	VIDAIC	FINE											VISALIA DAIRY/FEEDLOT DUS	0.0926	0.039
SJV027	SFCRUC	FINE	*	*	*	*	*						SANTA FE CRUDE BOILER (WE	0	0

Figure 5.5 Fitting Source Profile Selection for Test Case



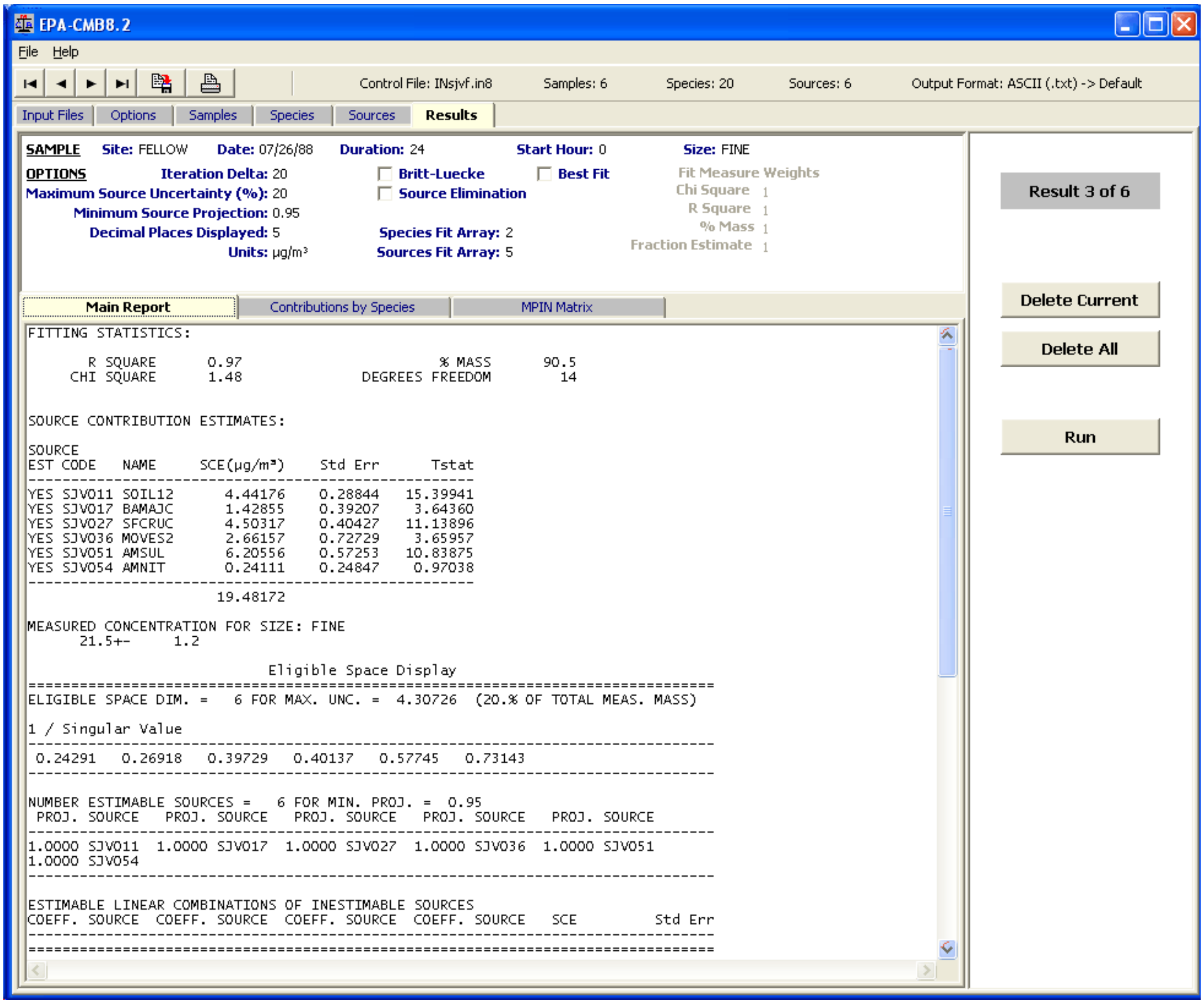
Control File: INsjvf.in8 Samples: 6 Species: 20 Sources: 6 Output Format: ASCII (.txt) -> Default

PROFILE - SJV011 SOIL12

```
N3IC  0.001 ± 0.005
S4IC  0.011 ± 0.018
N4TC  0.000 ± 0.000
KPAC  0.005 ± 0.001
NAAC  0.008 ± 0.008
ECTC  0.015 ± 0.010
OCTC  0.186 ± 0.022
ALXC  0.073 ± 0.008
SIXC  0.193 ± 0.022
SUXC  0.005 ± 0.000
CLXC  0.003 ± 0.000
KPXC  0.019 ± 0.002
CAXC  0.046 ± 0.005
TIXC  0.004 ± 0.000
VAXC  0.000 ± 0.000
CRXC  0.000 ± 0.000
MNXC  0.001 ± 0.000
FEXC  0.057 ± 0.006
NIXC  0.000 ± 0.000
CUXC  0.000 ± 0.000
ZNXC  0.002 ± 0.000
BRXC  0.000 ± 0.000
PBXC  0.002 ± 0.000
```

Figure 5.6 Speciation Graph of Selected Source Profile (Test Case)

Go to the Results screen to perform a calculation on the selected samples. (Note on the status bar that 6 samples have been selected, as well as 20 fitting species and 6 fitting sources.) As the information window reminds you, the Run button must be clicked in order to execute the calculation (Figure 3.9). In this example, the model was run in a *batch* mode (>1 sample selected; EPA-CMB8.2 calculations will always be performed for all samples selected). When the calculation is finished, the Main Report appears showing the result for the first sample (i.e., Bakers 07/26/88). Using the "VCR" buttons (upper left-hand corner), advance through the series of results for each sample until the results for Fellows, CA (FELLOW 07/26/88) is displayed, as shown in Figure 5.7. Note that all the key information for the calculation is displayed in the header: 6 samples, 20 fitting species and 6 sources.



Control File: INsjvf.in8   Samples: 6   Species: 20   Sources: 6   Output Format: ASCII (.txt) -> Default

SAMPLE Site: FELLOW Date: 07/26/88 Duration: 24 Start Hour: 0 Size: FINE

OPTIONS Iteration Delta: 20

Maximum Source Uncertainty (%): 20

Minimum Source Projection: 0.95

Decimal Places Displayed: 5

Units: µg/m³

□ Britt-Luecke

□ Source Elimination

□ Best Fit

Species Fit Array: 2

Sources Fit Array: 5

Fit Measure Weights

Chi Square 1

R Square 1

% Mass 1

Fraction Estimate 1

```
FITTING STATISTICS:

       R SQUARE      0.97                    % MASS      90.5
     CHI SQUARE      1.48            DEGREES FREEDOM       14


SOURCE CONTRIBUTION ESTIMATES:

SOURCE
EST CODE   NAME     SCE(µg/m³)    Std Err      Tstat
----------------------------------------------------
YES SJV011 SOIL12      4.44176    0.28844   15.39941
YES SJV017 BAMAJC      1.42855    0.39207    3.64360
YES SJV027 SFCRUC      4.50317    0.40427   11.13896
YES SJV036 MOVES2      2.66157    0.72729    3.65957
YES SJV051 AMSUL       6.20556    0.57253   10.83875
YES SJV054 AMNIT       0.24111    0.24847    0.97038
----------------------------------------------------
                      19.48172

MEASURED CONCENTRATION FOR SIZE: FINE
      21.5+-     1.2

                             Eligible Space Display
==============================================================================
ELIGIBLE SPACE DIM. =   6 FOR MAX. UNC. =  4.30726  (20.% OF TOTAL MEAS. MASS)

1 / Singular Value
------------------------------------------------------------------------------
 0.24291   0.26918   0.39729   0.40137   0.57745   0.73143
------------------------------------------------------------------------------

NUMBER ESTIMABLE SOURCES =   6 FOR MIN. PROJ. =  0.95
 PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE
------------------------------------------------------------------------------
1.0000 SJV011  1.0000 SJV017  1.0000 SJV027  1.0000 SJV036  1.0000 SJV051
1.0000 SJV054
------------------------------------------------------------------------------

ESTIMABLE LINEAR COMBINATIONS OF INESTIMABLE SOURCES
COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE   SCE         Std Err
------------------------------------------------------------------------------
==============================================================================
```

Figure 5.7 Main Report - Test Case

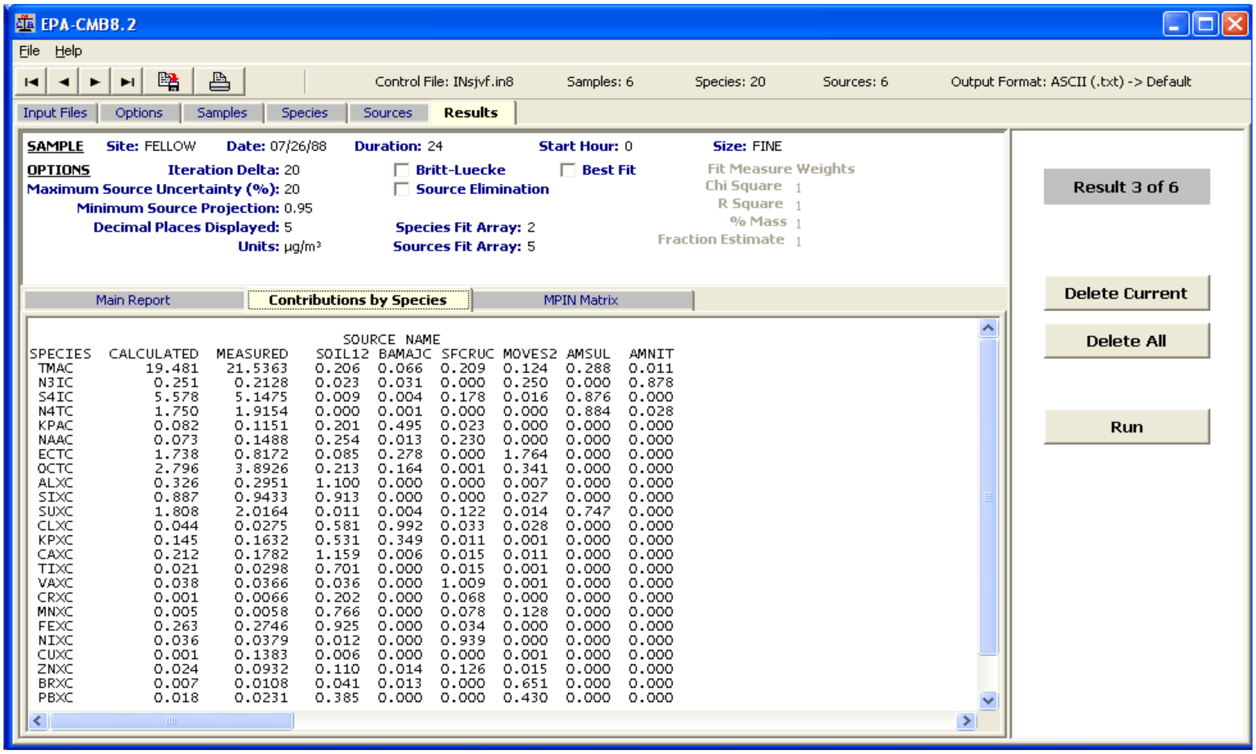
Note that settings for various options are echoed in the header, including the Britt &Luecke flag, indicating whether or not the Britt and Luecke (1973) solution was selected, as well as whether or not the fit was obtained using Source Elimination. As mentioned in Section 3.6.1, the field ‘EST’ under ‘SOURCE’ indicates (YES or NO) whether a source’s contribution was estimable in EPA-CMB8.2's attempt at a fit using the settings in Options. The concept of *estimable sources* is discussed in Section 6.1.2.

Note that for the Fellows sample, all 6 sources were *estimable* (“YES”), given their uncertainty. Acceptable uncertainty is specified in the Options screen (Section 3.2). The % mass explained can be readily verified by dividing the sum of the source contribution estimates (19.48 μgm^{-3}) by the measured mass (21.5 μgm^{-3}). Though Best Fit was not invoked, the default Fit Measure Weights are displayed.

Examine the species concentrations display (farther down in the Main Report). Particular attention should be paid to the ratios: calculated/measured and residual/uncertainty. See Section 6.1.3 for a detailed discussion of these and other performance measures.

Print this report in ASCII format by clicking on the Printer icon on the toolbar. An ASCII (txt) file may also be created by choosing the Print to File option, then OK, which brings up a dialog box for file disposition. As stated in Section 3.6.1, the header for this file is embellished with an echo of all input files used in the calculation.

Go to the Contributions by Species screen, which is shown in Figure 5.8. Note that the header information indicating the calculation parameters is carried through to this screen. Note



Control File: INsjvf.in8 Samples: 6 Species: 20 Sources: 6 Output Format: ASCII (.txt) -> Default

SAMPLE Site: FELLOW Date: 07/26/88 Duration: 24 Start Hour: 0 Size: FINE

OPTIONS Iteration Delta: 20
Maximum Source Uncertainty (%): 20
Minimum Source Projection: 0.95
Decimal Places Displayed: 5
Units: μg/m³

□ Britt-Luecke □ Best Fit
□ Source Elimination
Species Fit Array: 2
Sources Fit Array: 5

Fit Measure Weights
Chi Square 1
R Square 1
% Mass 1
Fraction Estimate 1

```
                                     SOURCE  NAME
SPECIES   CALCULATED   MEASURED    SOIL12 BAMAJC SFCRUC MOVES2 AMSUL   AMNIT
 TMAC         19.481    21.5363    0.206  0.066  0.209  0.124  0.288  0.011
 N3IC          0.251     0.2128    0.023  0.031  0.000  0.250  0.000  0.878
 S4IC          5.578     5.1475    0.009  0.004  0.178  0.016  0.876  0.000
 N4TC          1.750     1.9154    0.000  0.001  0.000  0.000  0.884  0.028
 KPAC          0.082     0.1151    0.201  0.495  0.023  0.000  0.000  0.000
 NAAC          0.073     0.1488    0.254  0.013  0.230  0.000  0.000  0.000
 ECTC          1.738     0.8172    0.085  0.278  0.000  1.764  0.000  0.000
 OCTC          2.796     3.8926    0.213  0.164  0.001  0.341  0.000  0.000
 ALXC          0.326     0.2951    1.100  0.000  0.000  0.007  0.000  0.000
 SIXC          0.887     0.9433    0.913  0.000  0.000  0.027  0.000  0.000
 SUXC          1.808     2.0164    0.011  0.004  0.122  0.014  0.747  0.000
 CLXC          0.044     0.0275    0.581  0.992  0.033  0.028  0.000  0.000
 KPXC          0.145     0.1632    0.531  0.349  0.011  0.001  0.000  0.000
 CAXC          0.212     0.1782    1.159  0.006  0.015  0.011  0.000  0.000
 TIXC          0.021     0.0298    0.701  0.000  0.015  0.001  0.000  0.000
 VAXC          0.038     0.0366    0.036  0.000  1.009  0.001  0.000  0.000
 CRXC          0.001     0.0066    0.202  0.000  0.068  0.000  0.000  0.000
 MNXC          0.005     0.0058    0.766  0.000  0.078  0.128  0.000  0.000
 FEXC          0.263     0.2746    0.925  0.000  0.034  0.000  0.000  0.000
 NIXC          0.036     0.0379    0.012  0.000  0.939  0.000  0.000  0.000
 CUXC          0.001     0.1383    0.006  0.000  0.000  0.001  0.000  0.000
 ZNXC          0.024     0.0932    0.110  0.014  0.126  0.015  0.000  0.000
 BRXC          0.007     0.0108    0.041  0.013  0.000  0.651  0.000  0.000
 PBXC          0.018     0.0231    0.385  0.000  0.000  0.430  0.000  0.000
```

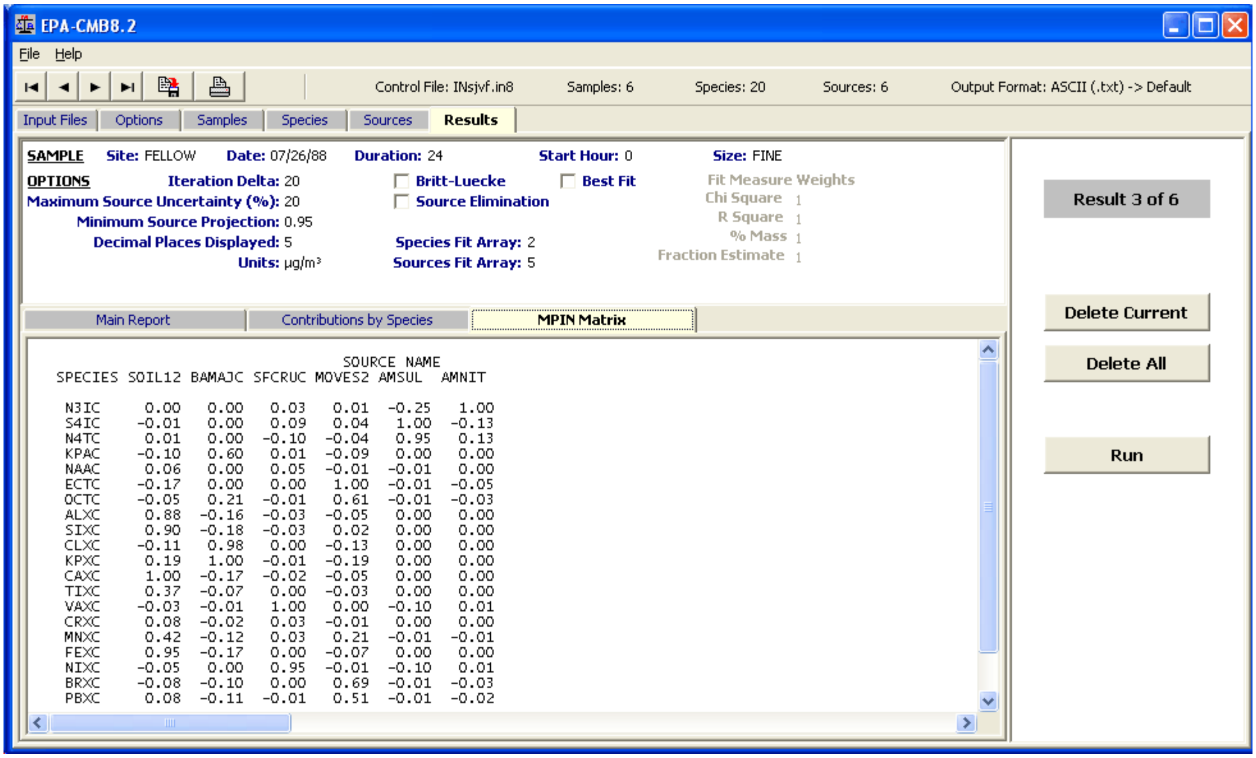
Figure 5.8 Contribution by Species - Test Case

also that the total measured mass for the sample also appears on the top line (TMAC). The source names that appear are the mnemonics provided in the source profile input file (Section 4.2.3).

For convenience, both the calculated and measured concentration for each species are repeated from the Main Report. All other values in the matrix are *ratios*. For any species, the calculated value may be confirmed (regenerated) by multiplying each ratio by the measured mass for **that** species, and then summing all the products.

This report may be printed in ASCII format by clicking on the Printer icon on the toolbar. Depending on the number of sources selected in the calculation, printing may be impractical (the right-hand side of the file may be truncated or there may be wrap-around problems). Alternatively, a formatted ASCII (txt) file may also be created by choosing the Print to File option, then OK, which brings up a dialog box for file disposition.

Go to the MPIN Matrix screen, which is shown in Figure 5.9. Note that the header information indicating the calculation parameters is also carried through to this screen. As with the Contributions by species screen, the source names that appear are the mnemonics provided in the source profile input file (Section 4.2.3). Print this report by clicking on the Printer icon on the toolbar. Depending on the number of sources selected in the calculation, printing may be impractical (the right-hand side of the file may be truncated or there may be wrap-around problems). Alternatively, a formatted ASCII file may also be created by choosing the Print to File option, then OK, which brings up a dialog box for file disposition.



Control File: INsjvf.in8 Samples: 6 Species: 20 Sources: 6 Output Format: ASCII (.txt) -> Default

SAMPLE Site: FELLOW Date: 07/26/88 Duration: 24 Start Hour: 0 Size: FINE

OPTIONS Iteration Delta: 20

Maximum Source Uncertainty (%): 20

Minimum Source Projection: 0.95

Decimal Places Displayed: 5

Units: µg/m³

□ Britt-Luecke □ Best Fit

□ Source Elimination

Species Fit Array: 2

Sources Fit Array: 5

Fit Measure Weights

Chi Square 1

R Square 1

% Mass 1

Fraction Estimate 1

```
                                        SOURCE  NAME
   SPECIES  SOIL12  BAMAJC  SFCRUC  MOVES2  AMSUL   AMNIT

    N3IC      0.00    0.00    0.03    0.01   -0.25    1.00
    S4IC     -0.01    0.00    0.09    0.04    1.00   -0.13
    N4TC      0.01    0.00   -0.10   -0.04    0.95    0.13
    KPAC     -0.10    0.60    0.01   -0.09    0.00    0.00
    NAAC      0.06    0.00    0.05   -0.01   -0.01    0.00
    ECTC     -0.17    0.00    0.00    1.00   -0.01   -0.05
    OCTC     -0.05    0.21   -0.01    0.61   -0.01   -0.03
    ALXC      0.88   -0.16   -0.03   -0.05    0.00    0.00
    SIXC      0.90   -0.18   -0.03    0.02    0.00    0.00
    CLXC     -0.11    0.98    0.00   -0.13    0.00    0.00
    KPXC      0.19    1.00   -0.01   -0.19    0.00    0.00
    CAXC      1.00   -0.17   -0.02   -0.05    0.00    0.00
    TIXC      0.37   -0.07    0.00   -0.03    0.00    0.00
    VAXC     -0.03   -0.01    1.00    0.00   -0.10    0.01
    CRXC      0.08   -0.02    0.03   -0.01    0.00    0.00
    MNXC      0.42   -0.12    0.03    0.21   -0.01   -0.01
    FEXC      0.95   -0.17    0.00   -0.07    0.00    0.00
    NIXC     -0.05    0.00    0.95   -0.01   -0.10    0.01
    BRXC     -0.08   -0.10    0.00    0.69   -0.01   -0.03
    PBXC      0.08   -0.11   -0.01    0.51   -0.01   -0.02
```

Figure 5.9 MPIN Matrix - Test Case

Once satisfied with the results, the output from EPA-CMB8.2 may be exported by clicking on the File icon, then using the Save Results dialog. As discussed in Section 4.5, the output format will be as selected in Options (default is ASCII).

The output file is a spreadsheet-type format and contains the contribution of each source to each measured species in each sample (Section 4.3.2). Fitting species are identified by an asterisk in the third column, and performance measures follow. Source contribution estimates and their standard errors are presented in subsequent columns, identified by mnemonics in the first row. Contributions from non-fitting profiles for a sample are identified as **–99.** Common spreadsheet data analysis tools can be used to interrogate files and select records for different chemical species, to group contributions from different profiles representing the same source type, to calculate average contributions, and to plot results

Close down EPA-CMB8.2 before moving on to the next exercise.

5.1.2 Individual Input File Operation

Restart EPA-CMB8.2, select the Individual Input Files mode, and a *null* screen like the one shown in Figure 5.10 will appear, except that the Control File name is absent and the file name boxes are blank. As described in Section 3.1, files are selected via their respective browse dialogs and *can be retrieved from any directory*. For ambient and source profile data files, select the ADsjvf.dbf and PRsjvf.dbf files, respectively. For the optional ambient data, species and source profiles selection files, select the same ones used in the previous exercise, i.e., ADsjvf.sel, SPsjvf.sel, and PRsjvf.sel, respectively.

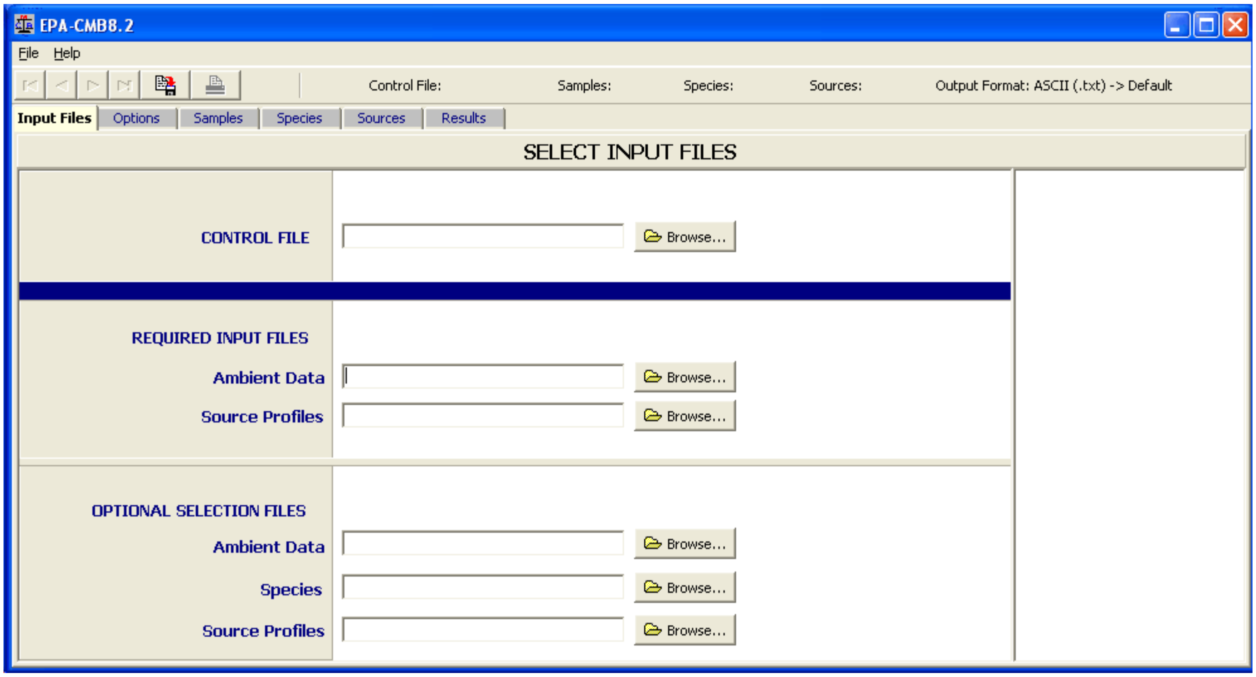
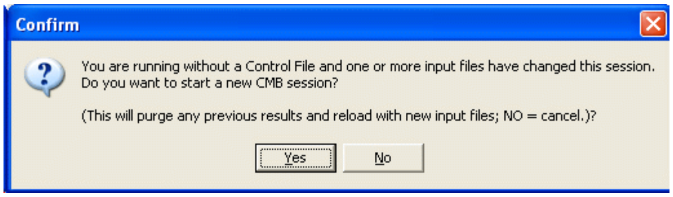


Figure 5.10 Input Files for Test Case - Individual Files Mode

From here, EPA-CMB8.2 will function exactly as it did in the previous exercise. After options are set, a sample(s) chosen, and arrays of fitting species and source profiles are selected, a calculation can be performed. Note that you can also go back and change any of the previous 5 files on the Input Files screen. If a selection file is changed, the following message will appear, indicating that EPA-CMB8.2 has detected a change in operating status. If the change is as you want, confirm the change (answer 'Yes') and continue. Note also that if no selection file(s) is(are) used, nothing will be initially selected and selection arrays will be blank.



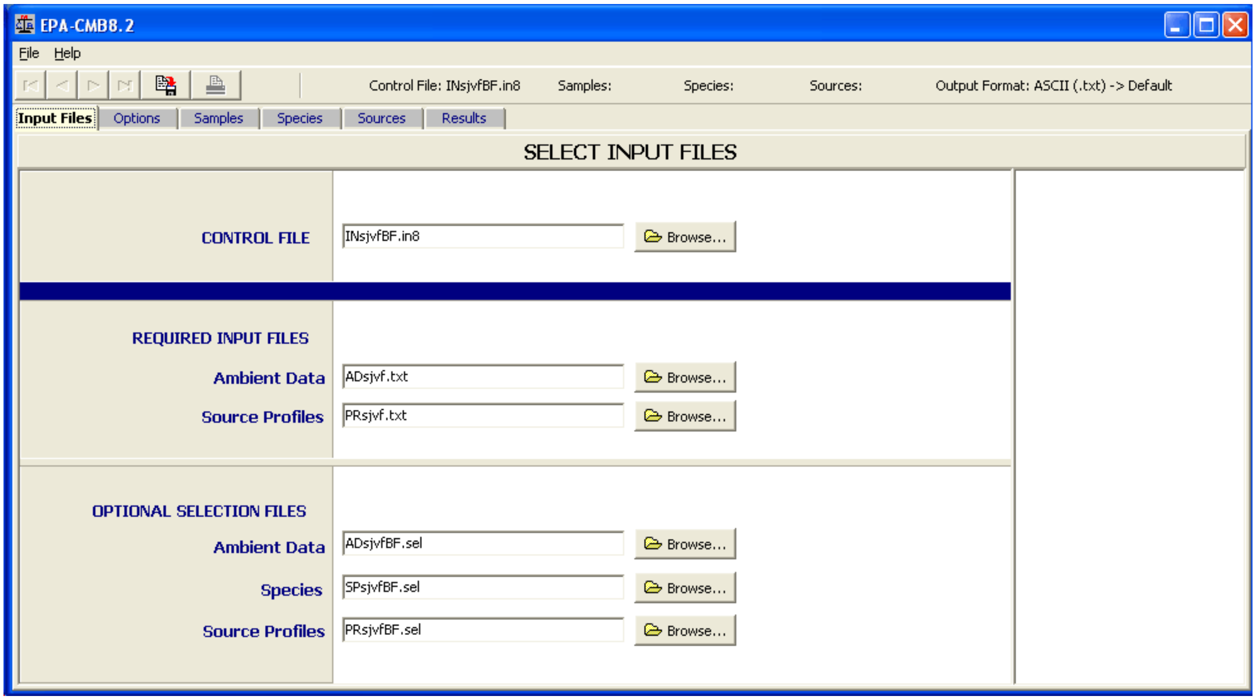
You are running without a Control File and one or more input files have changed this session. Do you want to start a new CMB session?

(This will purge any previous results and reload with new input files; NO = cancel.)?

Close down EPA- CMB8.2 before moving on to the next exercise.

5.2 Best Fit Option

Restart EPA-CMB8.2 and select the Control File mode. Select INsjvfBF.in8 and Open. A screen as shown in Figure 5.11 will appear. Go to the Options screen and select the Best Fit option. The default Fit Measure weights will be initialized at 1.000. Go to the Samples screen and note that 15 are initially selected. Click on View Selected to show only these 15 samples. Click on De-select All Samples and then click on Select All to reestablish the list of 206 ambient samples. Again click De-select All Samples and, using the "VCR" forward arrow or the vertical scroll arrow on the right -hand side of the screen, move down the list and select the **FRESNO 02/27/89** sample.



Control File: INsjvfBF.in8 Samples: Species: Sources: Output Format: ASCII (.txt) -> Default

CONTROL FILE: INsjvfBF.in8

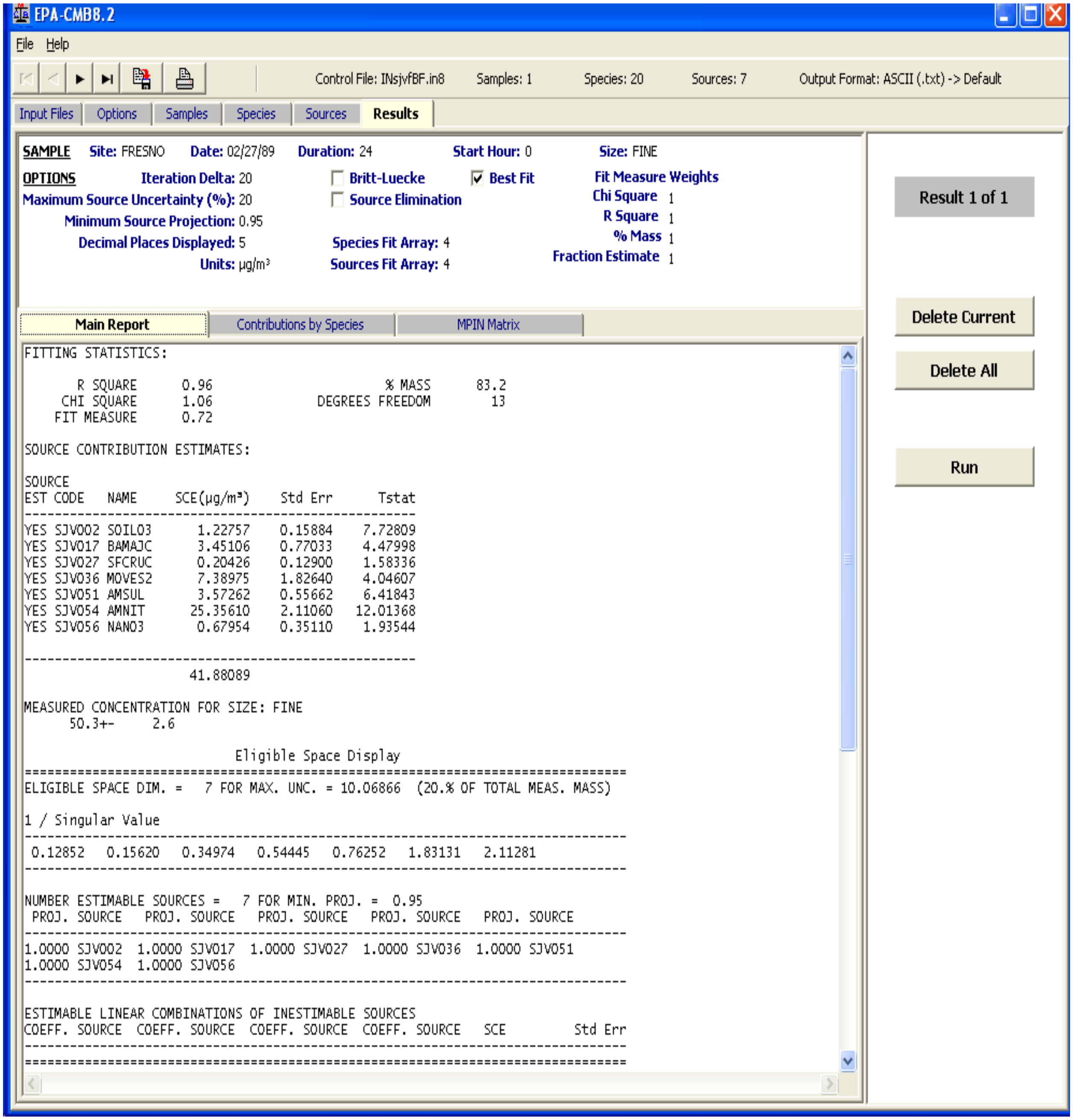
REQUIRED INPUT FILES
Ambient Data: ADsjvf.txt
Source Profiles: PRsjvf.txt

OPTIONAL SELECTION FILES
Ambient Data: ADsjvfBF.sel
Species: SPsjvfBF.sel
Source Profiles: PRsjvfBF.sel

Figure 5.11 Input Files for Testing the Best Fit Option

Go to the Results screen and perform a calculation using the Best Fit option selected. It does not matter which fitting sources and species arrays are selected because, as described in Section 3.2, EPA-CMB8.2 will first reset the array indices to 1 and then step through corresponding array pairs looking for a maximum of the Fit Measure, using the weights supplied on the Options screen (Section 6.3). If nonconvergence is encountered, the message "COLUMN NUMBER X of AFIT = 0" may appear (see Appendix E). In this case, "COLUMN NUMBER" refers to the source profile. Choose 'OK' and continue.

When EPA-CMB8.2 is finished, Arrays 4 for fitting species and fitting sources will have been selected as giving the best overall performance measure (value = 0.72), as shown in Figure 5.12. Returning to the species and source profile selection screens will reveal that indeed the 4th array is selected for each. Note that the Fit Measure Weights are echoed on the Results screen. These will also be echoed in printed reports.



Control File: INsjvfBF.in8   Samples: 1   Species: 20   Sources: 7   Output Format: ASCII (.txt) -> Default

SAMPLE Site: FRESNO Date: 02/27/89 Duration: 24 Start Hour: 0 Size: FINE

OPTIONS Iteration Delta: 20 □ Britt-Luecke ☑ Best Fit

Maximum Source Uncertainty (%): 20 □ Source Elimination

Minimum Source Projection: 0.95

Decimal Places Displayed: 5

Units: µg/m³

Species Fit Array: 4

Sources Fit Array: 4

Fit Measure Weights

Chi Square 1

R Square 1

% Mass 1

Fraction Estimate 1

```
FITTING STATISTICS:

      R SQUARE     0.96                   % MASS      83.2
    CHI SQUARE     1.06             DEGREES FREEDOM     13
   FIT MEASURE     0.72

SOURCE CONTRIBUTION ESTIMATES:

SOURCE
EST CODE    NAME     SCE(µg/m³)    Std Err      Tstat
----------------------------------------------------
YES SJV002 SOIL03      1.22757    0.15884    7.72809
YES SJV017 BAMAJC      3.45106    0.77033    4.47998
YES SJV027 SFCRUC      0.20426    0.12900    1.58336
YES SJV036 MOVES2      7.38975    1.82640    4.04607
YES SJV051 AMSUL       3.57262    0.55662    6.41843
YES SJV054 AMNIT      25.35610    2.11060   12.01368
YES SJV056 NANO3       0.67954    0.35110    1.93544

----------------------------------------------------
                      41.88089

MEASURED CONCENTRATION FOR SIZE: FINE
      50.3+-     2.6

                            Eligible Space Display
================================================================================
ELIGIBLE SPACE DIM. =   7 FOR MAX. UNC. = 10.06866  (20.% OF TOTAL MEAS. MASS)

1 / Singular Value
--------------------------------------------------------------------------------
 0.12852   0.15620   0.34974   0.54445   0.76252   1.83131   2.11281
--------------------------------------------------------------------------------

NUMBER ESTIMABLE SOURCES =   7 FOR MIN. PROJ. =  0.95
 PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE
--------------------------------------------------------------------------------
1.0000 SJV002  1.0000 SJV017  1.0000 SJV027  1.0000 SJV036  1.0000 SJV051
1.0000 SJV054  1.0000 SJV056
--------------------------------------------------------------------------------

ESTIMABLE LINEAR COMBINATIONS OF INESTIMABLE SOURCES
COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE   SCE         Std Err
--------------------------------------------------------------------------------
================================================================================
```

Figure 5.12 Results for Test Case Using Best Fit Option

6. EPA-CMB8.2 PERFORMANCE MEASURES

This section describes the different performance measures that are used to evaluate the validity of source contribution estimates. Greater detail on the use of the performance measures is presented in EPA (2004).

6.1 Main Report.

As discussed earlier (Section 3.6.1), the performance measures are presented in four separate groups in the Main Report produced by EPA-CMB-8.2: 1) fitting statistics; 2) source contribution estimates, (3) eligible space display; and (4) the species concentration display. Each of these displays is discussed below.

6.1.1 Source Contribution Estimates and Fitting Statistics

An example of a source contribution table display is shown below for the Fresno, CA sample (San Joaquin Valley data set):

```
                     Chemical Mass Balance Version EPA-CMB8.2
                              Report Date: 10/25/2004

SAMPLE:                        OPTIONS:                                INPUT FILES:

SITE:          FRESNO         BRITT & LUECKE:          No            INsjvfBF.in8
SAMPLE DATE:   02/27/89       SOURCE ELIMINATION:      No            PRsjvfBF.sel
DURATION:      24             BEST FIT:                Yes           SPsjvfBF.sel
START HOUR:    0              FIT MEASURE WEIGHTS:                   ADsjvfBF.sel
SIZE:          FINE               R Square:            1               ADsjvf.txt
                                  Chi Square:          1               PRsjvf.txt
Species Array: 4                  % Mass:              1
Sources Array: 4                  Fraction Estimate:   1

FITTING STATISTICS:

       R SQUARE      0.96                        % MASS      83.2

     CHI SQUARE      1.06             DEGREES FREEDOM          13

    FIT MEASURE      1.228

SOURCE CONTRIBUTION ESTIMATES:

SOURCE
EST CODE   NAME     SCE(μg/m³)    Std Err      Tstat
----------------------------------------------------
YES SJV002 SOIL03      1.22757    0.15884    7.72809
YES SJV017 BAMAJC      3.45106    0.77033    4.47998
YES SJV027 SFCRUC      0.20426    0.12900    1.58336
YES SJV036 MOVES2      7.38975    1.82640    4.04607
YES SJV051 AMSUL       3.57262    0.55662    6.41843
YES SJV054 AMNIT      25.35610    2.11060   12.01368
YES SJV056 NANO3       0.67954    0.35110    1.93544
----------------------------------------------------
                      41.88090

MEASURED CONCENTRATION FOR SIZE: FINE
      50.3+-     2.6
```

Source contribution estimates are the main output of EPA-CMB8.2. The sum of these concentrations approximates the total mass concentration. Negative source contribution estimates are not physically meaningful, but they can occur when a source profile is collinear with another profile or when the source contribution is close to zero. Collinearity is usually identified in the eligible sources display. When the absolute value of a positive or negative source contribution estimate is less than its standard error, the source contribution is undetectable. Two or three times the standard error may be taken as an upper limit of the source contribution in this case.

The standard errors reflect the uncertainty of the ambient data, the source profiles, and the amount of collinearity among different profiles. Standard errors should be reported with every source contribution estimate. *The standard error is a single standard deviation.* There is about a 66% probability that the true source contribution is within one standard error and about a 95% probability that the true contribution is within two standard errors of the source contribution estimate.

The T-statistic (Tstat) is the ratio of the source contribution estimate to the standard error. A Tstat value less than 2.0 indicates that the source contribution estimate is at or below a detection limit. Low Tstat values for several source contributions may be caused by collinearities among their profiles; this will be indicated in the Eligible Space Collinearity Display (Section 6.2).

The reduced chi-square, degrees of freedom[10], R-square, percent mass, and fit measure are performance measures for the least squares calculation. The chi-square is the weighted sum of squares of the differences between the calculated and measured fitting species concentrations. The weighting is inversely proportional to the squares of the uncertainty in the source profiles and ambient data for each species. Ideally, there would be no difference between calculated and measured species concentrations and chi-square would equal zero. A value less than 1 indicates a very good fit to the data, while values between 1 and 2 are acceptable. Chi-square values greater than 4 indicate that one or more species concentrations are not well explained by the source contribution estimates. The degrees of freedom equal the number of fitting species minus the number of fitting sources. The degrees of freedom is needed when statistical significance tests are applied to the chi-square value.

The R-square is the fraction of the variance in the measured concentrations that is explained by the variance in the calculated species concentrations. It is determined by a linear regression of measured versus model-calculated values for the fitting species. R-square ranges from 0 to 1.0. The closer the value is to 1.0, the better the source contribution estimates explain the measured concentrations. When R-square is less than 0.8, the source contribution estimates do not explain the observations very well with the fitting source profiles and/or species.

[10] The degrees of freedom (DF = no. fitting species - no. fitting sources) value is used to determine the statistical significance for a given confidence level (e.g., 99%). The test statistic is computed by normalizing (dividing) the χ^2 value by DF. With typical degrees of freedom of 13 to 20 using commonly measured ions, carbon and elements, a χ^2/DF value less than about 3 is not significant at the 99% confidence level. When many fitting species are used (as when apportioning organic compounds), the critical value for the test statistic (χ^2/DF) becomes smaller (e.g., <2).

Percent mass is the percent ratio of the sum of the model-calculated source contribution estimates to the measured mass concentration. This ratio should equal 100%, although values ranging from 80 to 120% are acceptable. If the measured mass is very low (< 5 to 10 µg/m^3), percent mass may be outside of this range because the uncertainty of the mass measurement is on the order of 1 to 2 µg/m^3.

6.1.2 Eligible Space Collinearity Display

Maximum source uncertainty (Section 3.2) defines the eligible space as that spanned by eigenvectors with inverse singular values less than or equal to the maximum source uncertainty. Sources lying within the eligible space may be estimated with an uncertainty less than the maximum source uncertainty. In practice, this strict criterion of inclusion is relaxed somewhat so that an estimable source is defined to be one whose projection into the eligible space is at least the minimum source projection. Inestimable sources have small projections within the eligible space. Certain linear combinations of inestimable sources may be estimable, and the program lists any of these that may exist. This may be understood as removing uncertainty by combining collinear sources. Different values for the maximum source uncertainty (ranging from 0 to 100%) and minimum source projections (ranging from 0 to 1.0) may be adjusted on the Options screen and will be retained for subsequent source contribution calculations during the session.

The eligible space display identifies the potential for collinearity and the potential reductions in standard errors in the source contribution estimates when source profiles are combined. An example appears below for the same Fresno, CA sample:

```
                   Eligible Space Collinearity Display

================================================================================
ELIGIBLE SPACE DIM. =   7 FOR MAX. UNC. = 10.06866  (20.% OF TOTAL MEAS. MASS)

1 / Singular Value
--------------------------------------------------------------------------------
 0.12852   0.15620   0.34974   0.54445   0.76252   1.83131   2.11281
--------------------------------------------------------------------------------

NUMBER ESTIMABLE SOURCES =   7 FOR MIN. PROJ. =  0.95
 PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE   PROJ. SOURCE
--------------------------------------------------------------------------------
1.0000 SJV002  1.0000 SJV017  1.0000 SJV027  1.0000 SJV036  1.0000 SJV051
1.0000 SJV054  1.0000 SJV056
--------------------------------------------------------------------------------

ESTIMABLE LINEAR COMBINATIONS OF INESTIMABLE SOURCES
COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE  COEFF. SOURCE   SCE         Std Err
--------------------------------------------------------------------------------
```

Henry's (1992) eligible space treatment uses the maximum source uncertainty, expressed as a percentage of the total measured mass, and the minimum source projection. These may be changed from their default values of 20% and 0.95, respectively, in the Options menu. As stated earlier, the maximum source uncertainty defines a space, called the *eligible space*, to be that spanned by those eigenvectors with inverse singular values less than or equal to the maximum source uncertainty. The first part of this display gives the eligible space dimension and the uncertainty used in its calculation. This is followed by a listing of the inverse singular values.

It was mentioned in Section 1.1 that EPA-CMB8.2 gives the user control in adjusting collinearity parameters which in CMB7 are "hard-wired" and not necessarily optimum for every application. Historically these parameters were chosen in CMB7 to be compatible with characteristics of particulate mass measurements, particularly those made by X-ray fluorescence. On an absolute basis with different measurement units, uncertainties associated with VOC may be up to an order of magnitude *lower* than those for particles. In an effort to resolve this discrepancy, the maximum source uncertainty (threshold set in Options; Section 3.2) in EPA-CMB8.2 is expressed as a *percentage* (“unit neutral”).

The next part of the display gives the number of estimable sources, the minimum source projection used in the calculation, and the projections of each profile vector into the *eligible space*. Inestimable sources are caused by excessive similarity (collinearity) among the source profiles or by high uncertainties in the individual source profiles. The standard errors associated with the source contribution estimates of one or more inestimable sources are usually very large, often too large to allow an adequate separation of these source contributions to be made. Inestimable sources will not appear if the two above-stated conditions (i.e., collinearity or large std. error of the SCE) do not occur. An absence of inestimable sources means that the source contributions can be resolved in the specific application. Since ambient data uncertainties, and relative levels of source contributions, vary from sample to sample, it is possible that a given set of profiles may appear in the ineligible (inestimable) space for one set of ambient data, but not for another set. For this reason, it is impossible to decide *a priori* that a set of profiles is collinear or not. The decision must be made for each set of data and each set of profiles combined with those data.

If collinearity is the cause of these excessive standard errors, then certain linear combinations of inestimable sources may be estimable, and the final part of the display lists these, if any exist. This may be understood as removing uncertainty by combining collinear sources. This linear combination may be more useful than the individual source contribution estimates if the standard error of the linear combination is substantially lower than the standard errors of each source contribution estimate. However, the treatment does not allow differentiation among the contribution estimates of the sources contained in the linear combination. Also, as stated above for the individual source data, the number of decimal places used in the presentation of the inverse singular values is that set in the Options menu. For more discussion of collinearity, see Section 4.1 of EPA, 2004.

6.1.3 Species Concentration Display

An example of the species concentration display is shown below for the same Fresno, CA sample:

```
================================================================================
SPECIES CONCENTRATIONS:
                                                                 CALCULATED    RESIDUAL
                                                                 ----------   -----------
SPECIES      FIT       MEASURED          CALCULATED               MEASURED    UNCERTAINTY
--------------------------------------------------------------------------------

TMAC   TMAU      50.34330+- 2.56520  41.88089+- 2.77273   0.83+- 0.07     -2.2
N3IC   N3IU   *  19.26080+- 0.97930  20.31037+- 1.97128   1.05+- 0.12      0.5
S4IC   S4IU   *   2.87790+- 0.16530   2.92384+- 0.36943   1.02+- 0.14      0.1
N4TC   N4TU   *   7.04960+- 0.36360   6.69662+- 0.58473   0.95+- 0.10     -0.5
KPAC   KPAU   *   0.14960+- 0.02350   0.14187+- 0.08541   0.95+- 0.59     -0.1
NAAC   NAAU   *   0.19820+- 0.05660   0.19290+- 0.07618   0.97+- 0.47     -0.1
ECTC   ECTU   *   4.55270+- 0.59790   4.57617+- 1.47563   1.01+- 0.35      0.0
OCTC   OCTU   *   5.99850+- 0.84490   5.42610+- 1.80669   0.90+- 0.33     -0.3
ALXC   ALXU   *   0.06410+- 0.02420   0.11887+- 0.01347   1.85+- 0.73      2.0
SIXC   SIXU   *   0.18690+- 0.03920   0.33528+- 0.10710   1.79+- 0.69      1.3
SUXC   SUXU       1.09520+- 0.05650   0.97978+- 0.12324   0.89+- 0.12     -0.9
CLXC   CLXU   *   0.06410+- 0.00800   0.07043+- 0.02214   1.10+- 0.37      0.3
KPXC   KPXU   *   0.16950+- 0.01070   0.16243+- 0.04288   0.96+- 0.26     -0.2
CAXC   CAXU   *   0.04500+- 0.00710   0.04837+- 0.00768   1.07+- 0.24      0.3
TIXC   TIXU   *   0.00060<  0.01930   0.00648<  0.00103  10.80< *****      0.3
VAXC   VAXU   *   0.00160<  0.00810   0.00212<  0.00041   1.32<  6.70      0.1
CRXC   CRXU   *   0.00200+- 0.00170   0.00039+- 0.00015   0.19+- 0.18     -0.9
MNXC   MNXU   *   0.00490+- 0.00090   0.00356+- 0.00178   0.73+- 0.39     -0.7
FEXC   FEXU   *   0.11250+- 0.01290   0.07588+- 0.00860   0.67+- 0.11     -2.4
NIXC   NIXU   *   0.00170+- 0.00100   0.00174+- 0.00024   1.02+- 0.62      0.0
CUXC   CUXU       0.02140<  0.06790   0.00062<  0.00022   0.03<  0.09     -0.3
ZNXC   ZNXU       0.02950<  0.04030   0.01050<  0.00250   0.36<  0.49     -0.5
BRXC   BRXU   *   0.01660+- 0.00100   0.01998+- 0.01123   1.20+- 0.68      0.3
PBXC   PBXU   *   0.03990+- 0.00560   0.03198+- 0.01530   0.80+- 0.40     -0.5
--------------------------------------------------------------------------------
```

This display shows how well the individual ambient concentrations are reproduced by the source contribution estimates. This display offers clues concerning which sources might be missing or which ones do not belong in the calculation. Fitting species are marked with an asterisk in the column labeled ' FIT '. Note that the symbol < flags measured values less than the uncertainty. The block of asterisks indicate numerical overflow, meaning that the number is too large to be displayed in the allocated field.

The column labeled CALCULATED / MEASURED displays, for each species, the ratio of the calculated (C) to measured (M) concentrations ± the standard error of the ratio for every chemical species with measured data:

$$std.\,err = \frac{C}{M}\sqrt{\left(\frac{err_C}{C}\right)^2 + \left(\frac{err_M}{M}\right)^2}$$

$$= \sqrt{\left(\frac{err_C}{M}\right)^2 + \left(\frac{C\ err_M}{M^2}\right)^2}\ ,$$

where:

err_C = error associated with the calculated concentration, and

err_M = error associated with the measured concentration.

The column labeled RESIDUAL / UNCERTAINTY displays, for each species, the ratio of the signed difference between the calculated and measured concentrations (residual = calculated - measured) divided by the uncertainty (standard error) of that residual:

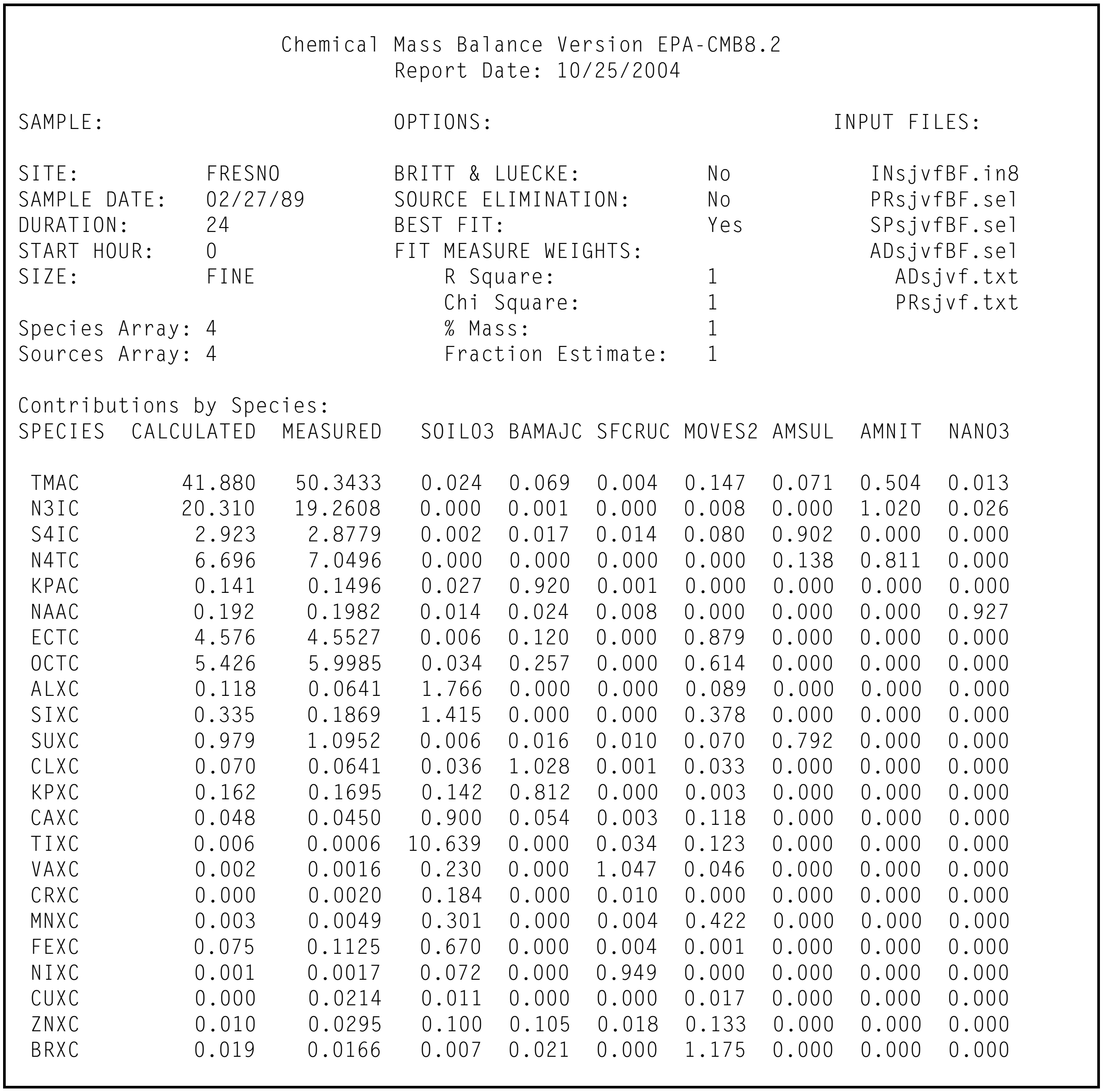
$$std.\,err = \sqrt{err_C^2 + err_M^2}$$

Note that the uncertainty is the square root of the sum of the squares of the uncertainty in the calculated and measured concentrations. The ratio specifies the number of uncertainty intervals by which the calculated and measured concentrations differ. When the absolute value of this ratio exceeds 2, the residual is significant. If it is positive, then one or more of the profiles is contributing too much to that species. If it is negative, then there is an insufficient contribution to that species and a source may be missing. The sum of the squared ratio for fitting species divided by the degrees of freedom yields the chi square. The highest ratio values for fitting species are the cause of high chi square values. Also, as above for the individual source data, the number of decimal places used in the presentation of the species data is that set in the Options menu.

As an example, if the calculated/measured ratio for TI was 10.8, this would indicate that TI is overestimated (over-explained) by EPA-CMB8.2 by an order of magnitude. While this condition could be within uncertainty limits, this behavior suggests that TI might be a candidate for removal as a fitting species, or that a source having a source contributing a large amount of TI might be a candidate for removal as a fitting source.

6.2 Additional Performance Measures

Another analysis available from the Results screen is the Contributions by Species report, which is a table showing the fraction of each species' calculated ambient concentration contributed by each source in the fit. The sources that are major contributors to each species can be determined by examining this display. An example of this display is shown below for the Fresno, CA sample, where it appears that the titanium concentration is substantially overestimated by the SOIL03 profile. While this condition could be within uncertainty limits, it might be advisable to substitute another profile with a lower titanium abundance, assuming the alternative profile is not unsuitable in other ways.



```
                     Chemical Mass Balance Version EPA-CMB8.2
                              Report Date: 10/25/2004

SAMPLE:                       OPTIONS:                             INPUT FILES:

SITE:          FRESNO         BRITT & LUECKE:          No             INsjvfBF.in8
SAMPLE DATE:   02/27/89       SOURCE ELIMINATION:      No             PRsjvfBF.sel
DURATION:      24             BEST FIT:                Yes            SPsjvfBF.sel
START HOUR:    0              FIT MEASURE WEIGHTS:                    ADsjvfBF.sel
SIZE:          FINE               R Square:            1                ADsjvf.txt
                                  Chi Square:          1                PRsjvf.txt
Species Array: 4                  % Mass:              1
Sources Array: 4                  Fraction Estimate:   1

Contributions by Species:
SPECIES  CALCULATED  MEASURED   SOIL03 BAMAJC SFCRUC MOVES2 AMSUL  AMNIT  NANO3

 TMAC        41.880   50.3433   0.024  0.069  0.004  0.147  0.071  0.504  0.013
 N3IC        20.310   19.2608   0.000  0.001  0.000  0.008  0.000  1.020  0.026
 S4IC         2.923    2.8779   0.002  0.017  0.014  0.080  0.902  0.000  0.000
 N4TC         6.696    7.0496   0.000  0.000  0.000  0.000  0.138  0.811  0.000
 KPAC         0.141    0.1496   0.027  0.920  0.001  0.000  0.000  0.000  0.000
 NAAC         0.192    0.1982   0.014  0.024  0.008  0.000  0.000  0.000  0.927
 ECTC         4.576    4.5527   0.006  0.120  0.000  0.879  0.000  0.000  0.000
 OCTC         5.426    5.9985   0.034  0.257  0.000  0.614  0.000  0.000  0.000
 ALXC         0.118    0.0641   1.766  0.000  0.000  0.089  0.000  0.000  0.000
 SIXC         0.335    0.1869   1.415  0.000  0.000  0.378  0.000  0.000  0.000
 SUXC         0.979    1.0952   0.006  0.016  0.010  0.070  0.792  0.000  0.000
 CLXC         0.070    0.0641   0.036  1.028  0.001  0.033  0.000  0.000  0.000
 KPXC         0.162    0.1695   0.142  0.812  0.000  0.003  0.000  0.000  0.000
 CAXC         0.048    0.0450   0.900  0.054  0.003  0.118  0.000  0.000  0.000
 TIXC         0.006    0.0006  10.639  0.000  0.034  0.123  0.000  0.000  0.000
 VAXC         0.002    0.0016   0.230  0.000  1.047  0.046  0.000  0.000  0.000
 CRXC         0.000    0.0020   0.184  0.000  0.010  0.000  0.000  0.000  0.000
 MNXC         0.003    0.0049   0.301  0.000  0.004  0.422  0.000  0.000  0.000
 FEXC         0.075    0.1125   0.670  0.000  0.004  0.001  0.000  0.000  0.000
 NIXC         0.001    0.0017   0.072  0.000  0.949  0.000  0.000  0.000  0.000
 CUXC         0.000    0.0214   0.011  0.000  0.000  0.017  0.000  0.000  0.000
 ZNXC         0.010    0.0295   0.100  0.105  0.018  0.133  0.000  0.000  0.000
 BRXC         0.019    0.0166   0.007  0.021  0.000  1.175  0.000  0.000  0.000
```

Yet another diagnostic available from the Results screen is the transpose of the normalized modified pseudo-inverse matrix (MPIN). This matrix indicates the degree of influence each species concentration has on the contribution and standard error of the corresponding source category. MPIN is normalized such that it takes on values from -1 to 1. Species with MPIN absolute values of 0.5 to 1.0 are considered influential species. Noninfluential species have MPIN absolute values of 0.3 or less. Species with absolute values between 0.3 and 0.5 are ambiguous but should generally be considered noninfluential. There are a number of useful references on identifying influential species (e.g., Belsley *et al.*, 1980; Kim and Henry, 1999).

An example display of this diagnostic for the Fresno, CA sample is shown below:

```
                      Chemical Mass Balance Version EPA-CMB8.2
                               Report Date: 10/25/2004

SAMPLE:                      OPTIONS:                            INPUT FILES:

SITE:          FRESNO        BRITT & LUECKE:          No            INsjvfBF.in8
SAMPLE DATE:   02/27/89      SOURCE ELIMINATION:      No            PRsjvfBF.sel
DURATION:      24            BEST FIT:                Yes           SPsjvfBF.sel
START HOUR:    0             FIT MEASURE WEIGHTS:                   ADsjvfBF.sel
SIZE:          FINE              R Square:            1               ADsjvf.txt
                                 Chi Square:          1               PRsjvf.txt
Species Array: 4                 % Mass:              1
Sources Array: 4                 Fraction Estimate:   1

MPIN Matrix:
SPECIES SOIL03 BAMAJC SFCRUC MOVES2 AMSUL  AMNIT  NANO3

N3IC    -0.01   0.00   0.01   0.04  -0.11   1.00   0.05
S4IC     0.00   0.00   0.00   0.01   1.00  -0.20   0.01
N4TC     0.01   0.00  -0.01  -0.04   0.12   0.90  -0.06
KPAC    -0.02   0.50   0.00  -0.08   0.00   0.00   0.00
NAAC     0.00   0.00   0.00   0.00   0.01  -0.12   1.00
ECTC    -0.17  -0.01   0.00   1.00  -0.08   0.01   0.00
OCTC    -0.11   0.16   0.00   0.71  -0.06   0.01   0.00
ALXC     0.83  -0.06  -0.02  -0.06   0.01   0.00   0.00
SIXC     0.43  -0.06  -0.01   0.17  -0.01   0.00   0.00
CLXC    -0.05   0.90   0.00  -0.10  -0.01   0.00  -0.01
KPXC     0.04   1.00   0.00  -0.17   0.00   0.00  -0.01
CAXC     0.76   0.01  -0.01   0.06  -0.01   0.00   0.00
TIXC     0.07   0.00   0.00  -0.01   0.00   0.00   0.00
VAXC     0.00   0.00   0.13   0.00  -0.01   0.00   0.00
CRXC     0.04   0.00   0.01  -0.01   0.00   0.00   0.00
MNXC     0.08  -0.06   0.00   0.39  -0.03   0.00   0.00
FEXC     1.00  -0.06   0.00  -0.17   0.01   0.00   0.00
NIXC    -0.02   0.00   1.00   0.00  -0.06   0.01  -0.01
BRXC    -0.11  -0.07   0.00   0.70  -0.06   0.00   0.00
PBXC    -0.05  -0.08   0.00   0.68  -0.06   0.00   0.00
```

6.3 Best Fit Measure

As discussed in Section 3.2, when more that one species or sources selection arrays are provided, EPA-CMB8.2 may be run in a Best Fit mode to have the model iterate among possible combinations of arrays and determine which results in an overall best fit. To do this, EPA-CMB8.2 uses a criterion called the Best Fit Measure that is calculated for each possible array pair combination. The Best Fit Measure is a linear combination of four factors related to fundamental performance measures that are combined as follows:

$$FM = \frac{W_1\left(\frac{1}{\chi^2}\right) + W_2 R^2 + W_3\left(\frac{\%\ mass}{100}\right) + W_4\left(FracEst\right)}{W_1 + W_2 + W_3 + W_4} \qquad \text{\% mass explained} \leq 100$$

$$FM = \frac{W_1\left(\frac{1}{\chi^2}\right) + W_2 R^2 + W_3\left(\frac{100}{\%\ mass}\right) + W_4\left(FracEst\right)}{W_1 + W_2 + W_3 + W_4} \qquad \text{\% mass explained} > 100$$

Where

FM = Fit Measure

χ^2, R^2, and % mass explained are the performance measures calculated by EPA-CMB8.2

FracEst is the ratio of the number of estimable fitting sources to the total number of fitting sources.

W_1 = χ^2 weight

W_2 = R^2 weight

W_3 = % mass weight

W_4 = fraction estimated weight

As mentioned in Section 3.2, these weights are set to 1.0 by default, and may be changed in the Options screen. And as mentioned in Section 5.1.3, for a given set of (non-negative) weight values, a larger value for FM corresponds to a better fit by EPA-CMB8.2.

7. REFERENCES

Anderson, E, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney and D. Sorensen, 1999. LAPACK Users' Guide, Third Edition. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; ISBN 0-89871-447-8 (paperback).

Barth, D., 1970. Federal motor vehicle emissions goals for CO, HC, and NO_x based on desired air quality levels. *JAPCA* **20:** 519.

Belsley, D.A., E. Kuh and R.E. Welsch, 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, Inc., New York.

Britt, H.I. and R.H. Luecke, 1973. The estimation of parameters in nonlinear, implicit models. *Technometrics* **15:** 233.

Cass, G.R. and G.J. McRae, 1981. Minimizing the cost of air pollution control. *Environ. Sci. Technol.* **15:** 748-57.

Chang, T.Y. and B. Weinstock, 1975. Generalized rollback modeling for urban air pollution control. *JAPCA* **25:** 1033-7.

Cheng, M.D. and P.K. Hopke, 1989. Identification of markers for chemical mass balance receptor model. *Atmos. Environ.* **23:** 1373-84.

Chow, J.C., J.G. Watson, D.H. Lowenthal, L.C. Pritchett and L.W. Richards, 1990. San Joaquin Valley Air Quality Study, Phase 2: PM_{10} modeling and analysis, Volume I: Receptor modeling source apportionment, Final Report. Report No. DRI 8929.1F, prepared by Desert Research Institute, Reno, NV.

Chow, J.C., J.G. Watson, D.H. Lowenthal, P.A. Solomon, K.L. Magliano, S.D. Ziman and L.W. Richards, 1992. PM_{10} source apportionment in California's San Joaquin Valley. *Atmos. Environ.* **26A:** 3335-54.

Chow, J.C., J.G. Watson, D.M. Ono and C.V. Mathai, 1993. PM_{10} standards and nontraditional particulate source controls: A summary of the A&WMA/EPA international specialty conference. *JAWMA* **43:** 74-84.

Cooper, J.A. and J.G. Watson, 1980. Receptor oriented methods of air particulate source apportionment. *JAPCA* **30:** 1116-25.

Coulter, C.T. and J.V. Scalco, 2005. Chemical Mass Balance Software: EPA-CMB8.2. Proceedings of A&WMA's 98th Conference & Exhibition; Minneapolis, MN, June 21-24, 2005.

Currie, L.A., R.W. Gerlach, C.W. Lewis, W.D. Balfour, J.A. Cooper, S.L. Dattner, R.T. deCesar, G.E. Gordon, S.L. Heisler, P.K. Hopke, J.J. Shah, G.D. Thurston and H.J. Williamson, 1984. Interlaboratory comparison of source apportionment procedures: Results for simulated data sets. *Atmos. Environ.* **18:** 1517.

deCesar, R.T., S.A. Edgerton, M.A.K. Khalil and R.A. Rasmussen, 1985. Sensitivity analysis of mass balance receptor modeling: methyl chloride as an indicator of wood smoke. *Chemosphere* **14:** 1495-501.

deCesar, R.T., S.A. Edgerton, M.A.K. Khalil and R.A. Rasmussen, 1986. A tool for designing receptor model studies to apportion source impacts with specified precision. In *Transactions, Receptor Methods for Source Apportionment: Real World Issues and Applications,* Pace, T.G., editor. Air Pollution Control Association, Pittsburgh, PA. pp. 56-67.

deNevers, N. and J.R. Morris, 1975. Rollback modeling: basic and modified. *JAPCA* **25:** 943.

Dzubay, T.G., R.K. Stevens, W.D. Balfour, H.J. Williamson, J.A. Cooper, J.E. Core, R.T. deCesar, E.R. Crutcher, S.L. Dattner, B.L. Davis, S.L. Heisler, J.J. Shah, P.K. Hopke and D.L. Johnson, 1984. Interlaboratory comparison of receptor model results for Houston aerosol. *Atmos. Environ.* **18:** 1555.

Environmental Protection Agency, 1990. Receptor Model Technical Series, Volume III (1989 Revision). CMB7 User's Manual. Report Number EPA-450/4-90-004. Office of Air Quality Planning and Standards, Research Triangle Park, NC.

Environmental Protection Agency, 2004. Protocol for Applying and Validating the CMB Model for PM2.5 and VOC. Report No. EPA-451/R-04-001. December 2004. Office of Air Quality Planning & Standards, Research Triangle Park, NC.

Friedlander, S.K., 1973. Chemical element balances and identification of air pollution sources. *Environ. Sci. Technol.* **7:** 235-40.

Friedlander, S.K., 1981. New developments in receptor modeling theory. In *Atmospheric Aerosol: Source/Air Quality Relationships,* Macias, E.S. and Hopke, P.K., editors. American Chemical Society, Washington, D.C. pp. 1-19.

Fujita, E.M., J.G. Watson, J.C. Chow and Z. Lu, 1994. Validation of the chemical mass balance receptor model applied to hydrocarbon source apportionment in the Southern California Air Quality Study. *Environ. Sci. Technol.* **28:** 1633-49.

Gartrell, G. and S.K. Friedlander, 1975. Relating particulate pollution to sources: The 1972 California Aerosol Characterization Study. *Atmos. Environ.* **9:** 279-99.

Gordon, G.E., 1980. Receptor models. *Environ. Sci. Technol.* **14:** 792-800.

Gordon, G.E., 1988. Receptor models. *Environ. Sci. Technol.* **22:** 1132-1142.

Gordon, G.E., W.H. Zoller, G.S. Kowalczyk and S.W. Rheingrover, 1981. Composition of source components needed for aerosol receptor models. In *Atmospheric Aerosol: Source/ Air Quality Relationships,* Macias, E.S. and Hopke, P.K., editors. American Chemical Society, Washington, D.C. pp. 51-74.

Henry, R.C., 1982. Stability analysis of receptor models that use least squares fitting. In *Receptor Models Applied to Contemporary Air Pollution Problems,* Hopke, P.K. and Dattner, S.L., editors. Air Pollution Control Association, Pittsburgh, PA. pp. 141-62.

Henry, R.C., 1992. Dealing with near collinearity in chemical mass balance receptor models. *Atmos. Environ.* **26A:** 933-8.

Hidy, G.M. and S.K. Friedlander, 1972. The nature of the Los Angeles aerosol. In *Second International Clean Air Congress,* Washington, D.C.

Hidy, G.M. and C. Venkataraman, 1996. The chemical mass balance method for estimating atmospheric particle sources in Southern California. *Chem. Eng. Comm.* **151:** 187-209.

Hopke, P.K., 1985. *Receptor Modeling in Environmental Chemistry.* John Wiley & Sons, New York, NY.

Hopke, P.K., 1991. *Receptor Modeling for Air Quality Management.* Elsevier Press, Amsterdam, The Netherlands.

Hopke, P.K. and S.L. Dattner, 1982. *Receptor Models Applied to Contemporary Pollution Problems.* Air & Waste Management Association, Pittsburgh, PA.

Hougland, E.S., 1983. Chemical element balance by linear programming. *73rd Annual Meeting of the Air Pollution Control Association,* Atlanta, GA.

Javitz, H.S. and J.G. Watson, 1986. Methods of receptor model evaluation and validation. In *Transactions, Methods for Source Apportionment: Real World Issues and Applications,* Pace, T.G, editor. Air Pollution Control Association, Pittsburgh, PA.

Javitz, H.S., J.G. Watson, J.P. Guertin and P.K. Mueller, 1988a. Results of a receptor modeling feasibility study. *JAPCA* **38:** 661.

Javitz, H.S., J.G. Watson and N.F. Robinson, 1988b. Performance of the chemical mass balance model with simulated local-scale aerosols. *Atmos. Environ.* **22:** 2309-22.

Kim, B.M. and R.C. Henry, 1989. Analysis of multicollinearity indicators and influential species for chemical mass balance receptor model. In *Transactions, Receptor Models in Air Resources Management,* Watson, J.G., editor. Air & Waste Management Association, Pittsburgh, PA. pp. 379-90.

Kim, B.M. and R.C. Henry, 1999. Diagnostics for Determining Influential Species in the Chemical Mass Balance Receptor Model. *JAWMA* **49:** 1449-1455.

Kneip, T.J., M.T. Kleinman and M. Eisenbud, 1973. Relative contribution of emission sources to the total airborne particulates in New York City. In *Third International Clean Air Congress,* Dusseldorf, FRG.

Larson, T.V. and R.J. Vong, 1989. Partial least squares regression methodology: Application to source receptor modeling. In *Transactions, Receptor Models in Air Resources Management,* Watson, J.G., editor. Air & Waste Management Association, Pittsburgh, PA. pp. 391-403.

Lin, C. and J.B. Milford, 1994. Decay-adjusted chemical mass balance receptor modeling for volatile organic compounds. *Atmos. Environ.* **28:** 3261-76.

Lowenthal, D.H., R.C. Hanumara, K.A. Rahn and L.A. Currie, 1987. Effects of systematic error, estimates and uncertainties in chemical mass balance apportionments: Quail Roost II revisited. *Atmos. Environ.* **21:** 501-10.

Lowenthal, D.H. and K.A. Rahn, 1988a. Reproducibility of regional apportionments of pollution aerosol in the Northeastern United States. *Atmos. Environ.* **22:** 1829-33.

Lowenthal, D.H. and K.A. Rahn, 1988b. Tests of regional elemental tracers of pollution aerosols. 2. Sensitivity of signatures and apportionments to variations in operating parameters. *Atmos. Environ.* **22:** 420-6.

Lowenthal, D.H., K.R. Wunschel and K.A. Rahn, 1988c. Tests of regional elemental tracers of pollution aerosols. 1. Distinctness of regional signatures, stability during transport, and empirical validation. *Environ. Sci. Technol.* **22:** 413-20.

Lowenthal, D.H., J.C. Chow, J.G. Watson, G.R. Neuroth, R.B. Robbins, B.P. Shafritz and R.J. Countess, 1992. The effects of collinearity on the ability to determine aerosol contributions from diesel- and gasoline-powered vehicles using the chemical mass balance model. *Atmos. Environ.* **26A:** 2341-51.

Lowenthal, D.H., B. Zielinska, J.C. Chow, J.G. Watson, M. Gautam, D.H. Ferguson, G.R. Neuroth and K.D. Stevens, 1994. Characterization of heavy-duty diesel vehicle emissions. *Atmos. Environ.* **28:** 731-44.

Miller, M.S., S.K. Friedlander and G.M. Hidy, 1972. A chemical element balance for the Pasadena aerosol. *J. Colloid Interface Sci.* **39:** 165-76.

Pace, T.G., 1986. *Transactions, Receptor Methods for Source Apportionment: Real World Issues and Applications.* Air Pollution Control Association, Pittsburgh, PA.

Pace, T.G., 1991. Receptor modeling in the context of ambient air quality standard for particulate matter. In *Receptor Modeling for Air Quality Management,* Hopke, P.K., editor. Elsevier, Amsterdam, The Netherlands. pp. 255-97.

Song, X.H. and P.K. Hopke, 1996. Solving the chemical mass balance problem using an artificial neural network. *Environ. Sci. Technol.* **30:** 531

Stevens, R.K. and T.G. Pace, 1984. Review of the mathematical and empirical receptor models workshop (Quail Roost II). *Atmos. Environ.* **18:** 1499-506.

Venkataraman, C. and Friedlander, S.K., 1994. Source resolution of fine particulate polycyclic aromatic hydrocarbons using a receptor model modified for reactivity. *JAWMA* **44:** 1103-8.

Vong, R.J., P. Geladi, S. Wold and K. Esbensen, 1988. Source contributions to ambient aerosol calculated by discriminant partial least squares regression (PLS). *J. Chemometrics* **2:** 281-96.

Watson, J.G., 1979. Chemical element balance receptor model methodology for assessing the sources of fine and total particulate matter in Portland, Oregon. Ph.D. Dissertation. Oregon Graduate Center, Beaverton, OR.

Watson, J.G., 1984. Overview of receptor model principles. *JAPCA* **34:** 619-23.

Watson, J.G. and J.C. Chow, 1994. Clear sky visibility as a challenge for society. *Annual Rev. Energy Environ.* **19:** 241-66.

Watson, J.G., Cooper, J.A. and J.J. Huntzicker, 1984. The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atmos. Environ.* **18:** 1347-55.

Watson, J.G., J.C. Chow and C.V. Mathai, 1989. Receptor models in air resources management: A summary of the APCA international specialty conference. *JAPCA* **39:** 419-26.

Watson, J.G., N.F. Robinson, J.C. Chow, R.C. Henry, B.M. Kim, T.G. Pace, E.L. Meyer and Q. Nguyen, 1990. The USEPA/DRI chemical mass balance receptor model, CMB 7.0. *Environ. Software* **5:** 38-49.

Watson, J.G., J.C. Chow, Z. Lu, E.M. Fujita, D.H. Lowenthal and D.R. Lawson, 1994a. Chemical mass balance source apportionment of PM_{10} during the Southern California Air Quality Study. *Aerosol Sci. Technol.* **21:** 1-36.

Watson, J.G., J.C. Chow, F.W. Lurmann and S. Misarra, 1994b. Ammonium nitrate, nitric acid, and ammonia equilibrium in wintertime Phoenix, Arizona. *JAWMA* **44:** 405-12.

Williamson, H.J. and D.A. Dubose, 1983. Receptor model technical series, Volume III: User's manual for chemical mass balance model. Report No. EPA-450/4-83-014, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Winchester, J.W. and G.D. Nifong, 1971. Water pollution in Lake Michigan by trace elements from aerosol fallout. *Water Air and Soil Pollution* **1:** 50-64.

APPENDIX A

THEORY OF THE CHEMICAL MASS BALANCE RECEPTOR MODEL

A.1 INTRODUCTION

Receptor models use the chemical and physical characteristics of gases and particles measured at source and receptor to both identify the presence of and to quantify source contributions to the receptor. The particle characteristics must be such that: 1) they are present in different proportions in different source emissions; 2) these proportions remain relatively constant for each source type; and 3) changes in these proportions between source and receptor are negligible or can be approximated.

Common types of receptor models include: (1) chemical mass balance (CMB); (2) principal component analysis (PCA, otherwise known as factor analysis); and (3) multiple linear regression (MLR). Descriptions of these models and some of their variations are given in Henry *et al.* (1984) and Hopke (1991). Chemical mass balance is the fundamental receptor model, with all other approaches (including PCA and MLR) based on the use of the mass balance concept.

The chemical mass balance consists of a least squares solution to a set of linear equations which expresses each receptor concentration of a chemical species as a linear sum of products of source profile species and source contributions. The source profile species abundances (i.e., the fractional amount of the species in the emissions from each source-type) and the receptor concentrations, with appropriate uncertainty estimates, serve as input data to the CMB model. The output consists of the amount contributed by each source-type to each chemical species. The model calculates values for the contributions from each source and the uncertainties of those values. Input data uncertainties are used both to weight the importance of input data values in the solution and to calculate the uncertainties of the source contributions.

A.2 CMB Mathematics

The source contribution (S_j) present at a receptor during a sampling period of length T due to a source j with constant emission rate E_j is

$$S_j = D_j \bullet E_j \qquad \text{(A-1)}$$

where:

$$D_j = \int_0^T d\left[\vec{u}(t), \sigma(t), \vec{x}_j\right] dt \qquad \text{(A-2)}$$

is a dispersion factor depending on wind velocity (u), atmospheric stability (σ), and the location of source j with respect to the receptor (x_j). All parameters in Equation A-2 vary with time, so the instantaneous dispersion factor, D_j, must be integrated over time period T (Watson, 1979).

Various forms for D_j have been proposed (Pasquill, 1974; Benarie, 1976; Seinfeld and Pandis, 1998), some including provisions for chemical reactions, removal, and specialized topography. None are completely adequate to describe the complicated, random nature of dispersion in the atmosphere. The advantage of receptor models is that an exact knowledge of D_j is unnecessary.

If a number of sources, J, exists and there is no interaction between their emissions to cause mass removal, the total mass measured at the receptor, C, will be a linear sum of the contributions from the individual sources.

$$C = \sum_{j=1}^{J} D_j \bullet E_j = \sum_{j=1}^{J} S_j \qquad \text{(A-3)}$$

Similarly, the concentration of elemental component i, C_i, will be

$$C_i = \sum_{j=1}^{J} F_{ij} \bullet S_j \qquad i = 1, 2, \ldots I \qquad \text{(A-4)}$$

where: F_{ij} = the fraction of source contribution S_j composed of element i. The number of chemical species (I) must be greater than or equal to the number of sources (J) for a unique solution to these equations.

Solutions to the CMB equations consist of: (1) a tracer solution; (2) a linear programming solution; (3) an ordinary weighted least squares solution with or without an intercept; (4) a ridge regression weighted least squares solution with or without an intercept; and (5) an effective variance least squares solution with or without an intercept. An estimate of the uncertainty associated with the source contributions is an integral part of several of these solution methods.

Weighted linear least squares solutions are preferable to the tracer and linear programming solutions because: (1) theoretically they yield the most likely solution to the CMB equations, providing model assumptions are met; (2) they can make use of all available chemical measurements, not just the so-called tracer species; (3) they are capable of analytically estimating the uncertainty of the source contributions; and (4) there is, in practice, no such thing as a "tracer." The effective variance solution developed and tested by Watson *et al.* (1984): (1) provides realistic estimates of the uncertainties of the source contributions (owing to its incorporation of both source profile and receptor data uncertainties); and (2) gives greater influence to chemical species with smaller values for uncertainty in both the source and receptor measurements than to species with higher values for uncertainty. The effective variance solution is derived by minimizing the weighted sums of the squares of the differences between the measured and calculated values of C_i and F_{ij} (Britt and Luecke, 1973; Watson *et al.*, 1984). The solution algorithm is an iterative procedure which calculates a new set of S_j based on the S_j estimated from the previous iteration. It is carried out by the following steps expressed in matrix notation. A superscript ***k*** is used to designate the value of a variable at the k^{th} iteration.

1. Set initial estimate of the source contributions equal to zero.

$$S_j^{k=0} = 0 \qquad j = 1, 2, \ldots J \tag{A-5}$$

2. Calculate the diagonal components of the effective variance matrix, **V**$\mathbf{V}_\mathbf{e}$. All off-diagonal components of this matrix are equal to zero.

$$\boldsymbol{V}_{e_{ii}}^{k} = \sigma_{C_i}^2 + \sum \left(S_j^k\right)^2 \bullet \sigma_{F_{ij}}^2 \tag{A-6}$$

3. Calculate the k+1 value of S_j.

$$S^{k+1} = \left(\boldsymbol{F}^T \left(\boldsymbol{V}_e^k\right)^{-1} \boldsymbol{F}\right)^{-1} \boldsymbol{F}^T \left(\boldsymbol{V}_e^k\right)^{-1} C \tag{A-7}$$

4. Test the (k+1)th iteration of the S_j against the kth iteration. If any one differs by more than 1%, then perform the next iteration. If all differ by less than 1%, then terminate the algorithm.

$$\begin{aligned} &if \ \left|\left(S_j^{k+1} - S_j^k\right) \middle/ S_j^{k+1}\right| > 0.01, \text{ go to step 2} \\ &if \ \left|\left(S_j^{k+1} - S_j^k\right) \middle/ S_j^{k+1}\right| \leq 0.01, \text{ go to step 5} \end{aligned} \tag{A-8}$$

5. Assign the (k+1)th iteration to S_j and σ_{S_j}. All other calculations are performed with these final values.

$$\sigma_{S_j}^2 = \left[\boldsymbol{F}^T \left(\boldsymbol{V}_e^{k+1}\right)^{-1} \boldsymbol{F}_{jj}\right]^{-1} \qquad j = 1, 2, \ldots J \tag{A-9}$$

where: C = $(C_1 \ldots C_I)^T$, a column vector with C_i as the ith component

S = $(S_1 \ldots S_J)^T$, a column vector with S_j as the jth component

F = an I x J matrix of F_{ij}, the source composition matrix

σ_{C_i} = one standard deviation uncertainty of the C_i measurement

$\sigma_{\boldsymbol{F}_{ij}}$ = one standard deviation uncertainty of the F_{ij} measurement

$\boldsymbol{V}_e$ = diagonal matrix of effective variances

The effective variance solution algorithm is very general, and it reduces to most of the solutions cited above with the following modifications:

- When $\sigma_{\boldsymbol{F}_{ij}}$ is set equal to zero, the solution reduces to the ordinary weighted least squares solution.
- When $\sigma_{\boldsymbol{F}_{ij}}$ is set equal to the same constant value, the solution reduces to the unweighted least squares solution.
- When a column is added to the ***F*** matrix with all values equal to 1, an intercept term is computed for the variable corresponding to this column.
- When the number of source profiles equals the number of species (I = J), and if the selected species are present only in a single, exclusive source profile, the solution reduces to the tracer solution.
- When the expression $\left(\boldsymbol{F}^T(\boldsymbol{V}_e^k)^{-1}\boldsymbol{F}\right)$ is rewritten as $\left(\boldsymbol{F}^T(\boldsymbol{V}_e^k)^{-1}\boldsymbol{F}\text{-}\varphi\,\boldsymbol{I}\right)$, with **φ** equal to some non-zero number, known as the *smoothing parameter*, and ***I*** equal to the identity matrix, the solution becomes the ridge regression solution (Williamson and DuBose, 1983 and Henry *et al.*, 1984).

Formulas for the performance measures are:

Reduced chi square = $$\chi^2 = \frac{1}{I - J}\sum_{i=1}^{I}\left[\left(\mathbf{C}_i - \sum_{j=1}^{J}\boldsymbol{F}_{\mathbf{ij}}\mathbf{S}_{\mathbf{j}}\right)^{\mathbf{2}} \Big/ V_{e_{ii}}\right] \qquad \text{(A-10)}$$

Percent Mass = $100\left(\sum_{j=1}^{J} S_j\right) \Big/ C_t$, where C_t is the total measured mass (A-11)

R square = $1 - \left[(I - J)\chi^2\right] \Big/ \left[\sum_{i=1}^{I} C_i^2 \Big/ V_{e_{ii}}\right]$ (A-12)

Modified Pseudo-Inverse Matrix = $\left(\boldsymbol{F}^T\left(\boldsymbol{V_e}\right)^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\left(\boldsymbol{V}_e\right)^{-1/2}$ (A-13)

The Singular Value Decomposition of the weighted **F** matrix is given by (Henry, 1992)

$$\boldsymbol{V}_e^{1/2}\boldsymbol{F} = \boldsymbol{UDV}^T \tag{A-14}$$

where ***U*** and ***V*** are I X I and J X J orthogonal matrices, respectively, and where **D** is a diagonal matrix with J nonzero and positive elements called the singular values of the decomposition. The columns of ***V*** are called the eigenvectors of the composition and their components are associated with the source types.

References

Benarie, M.M., 1976. *Urban Air Pollution Modeling Without Computers.* U.S. Environmental Protection Agency, Research Triangle Park, NC.

Britt, H.I. and R.H. Luecke, 1973. The estimation of parameters in nonlinear, implicit models. *Technometrics* **15,** 233-233.

Henry, R.C., C.W. Lewis, P.K. Hopke, and H.J. Williamson, 1984. Review of receptor model fundamentals. *Atmos. Environ.*, **18**, 1507-1515.

Henry, R.C., 1992. Dealing with near collinearity in chemical mass balance receptor models. *Atmos. Environ.* **26A,** 933-938.

Hopke, P.K. (Ed.), 1991. *Receptor Modeling for Air Quality Management.* Elsevier Science Publishing Company, Inc. New York, NY. 329pp.

Pasquill, F., 1974. *Atmospheric Diffusion.* Ellis Horwood, Chichester, England.

Seinfeld, J.H. and S.N. Pandis, 1998. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change.* John Wiley & Sons, New York, NY.

Watson, J.G., 1979. *Chemical Element Balance Receptor Model Methodology for Assessing the Sources of Fine and Total Particulate Matter*. Ph.D. Dissertation, Oregon Graduate Center, Beaverton, OR. University Microfilms International, Ann Arbor, MI.

Watson, J.G., Cooper, J.A., and J.J. Huntzicker, 1984. The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atmos. Environ.* **18,** 1347-1355.

Williamson, H.J. and D.A. Dubose, 1983. Receptor model technical series, Volume III: User's manual for chemical mass balance model. Report No. EPA-450/4-83-014**,** U.S. Environmental Protection Agency, Research Triangle Park, NC.

APPENDIX B

Source Code for EPA-CMB8.2: Fortran & C++

The following is an inventory and brief description of the source code used to build the main DLL (Dynamic Link Library) for EPA-CMB8.2:

Anatomy of the DLL Source Code

Files:	**Function**[1]
CMB82.def	Definition file that defines ***CMB82C*** EXPORT to CMB82.dll
CMB82a.for	Contains fRun and all routines called directly by fRun
f. fRun	Defines cases and assigns integer values consistent with those defined in CMB82.c; selects 6 cases:
s. FitAlloc	
s. CMBSVD	Performs the bulk of the numerical calculations, including the singular value decomposition (SVD). Key fit statistics are computed, along with the modified pseudo inverse (mPIN) matrix.
s. ChecksFit	Checks the fit for sources to eliminate.
s. FracEst	Determines the number of estimable sources.
s. PData	Finishes computations and writes them to the Results screen.
s. SScont	Computes species contributions to each sources SCE (source contribution estimate).
s. PIN	Prepares mPIN matrix for output.
CMB82b.for	Supplemental; contains all routines not directly called by fRun.
s. AmbData	
s. GetAmbDirec	
f. Initialize	Declares variables, dimensions arrays & defines Fortran unit numbers for direct-access (binary) scratch files AmbDir.dat; ProDir.dat, SumDir.dat.
s. LoadATOT	
s. MATMUL	
s. MATSIN	
s. OpenAmbDir	Opens ambient data file and specifies record lengths via RECL stmt.
s. OpenProDir	Opens source profile data file and specifies record length via RECL stmt.
s. OpenSumDir	Opens summary data file and specifies record lengths via RECL stmt.
f. ProRec	
f. SetAmbRec	
s. SETiWORK	
s. SETWORK	
s. SfitHead	Creates header for the source fit display.
s. SfitOut	Creates DB (spreadsheet-type) output file.
s. SUBATOL	
s. WinMessage	References **MessageBoxC_(message)** in CMB82.c.
s. Write2ErrBuf	
s. Write2OutBuf	
s. XLCunc	Computes source uncertainties.
CMB82c.c	Visual C++ file; critical interface for Delphi; the main program file for EPA-CMB8.2
CMB82.inc	Include file for variables used globally in EPA-CMB8.2
WKShead.h	Header file referenced by CMB82.c for (Lotus) WKS output

[1] Historically, the CMB8.2 code has not been well documented. For this reason, even after inspection, it's not always clear what certain routines are doing. As experience is gained, the functionality notes should become more complete.

Organization of the source code (Fortran & C).

The 4 Fortran files associated with CMB8.0 (CMB85_A.for, CMB85_B.for, CMB85_C.for & CMB85F.for) have been reorganized and merged into 2 files: the main **CMB82a.for** & the supplemental **CMB82b.for**. The Fortran portion of the source code consists entirely of 28 modules (24 subroutines and 4 functions). In the Fortran code, there is no “main program unit” (though function ***fRun*** in CMB82a.for initially controls calls to all Fortran routines); the C-code file **CMB82.c** (not shown) serves as the main program. In an effort to purge legacy code, 28 routines that were unnecessary in CMB8.0 were removed, as were many variables associated with the remaining routines. All routines called by fRun are in CMB82a.for; all others are in CMB82b.for. The Fortran routines are listed here, along with other routines they may call subsequently:

CMB82a.for (8 routines, in calling sequence)

f. fRun

calls:

s. FitAlloc	⇨	SUBATOL, AmbData, SETiWORK, SETWORK
s. CMBSVD[2]	⇨	WinMessage, ProRec, MATMUL, MATSIN, Write2ErrBuf, XLCunc, OpenSumDir
s. ChecksFit	⇨	SETiWORK, SETWORK
s. FracEst		
s. PData[3]	⇨	OpenSumDir, Write2OutBuf, AmbData, GetAmbDirec
s. SScont	⇨	Write2OutBuf, AmbData, ProRec
s. PIN	⇨	AmbData, Write2OutBuf

CMB82b.for (20 routines, alphabetically)

s. AmbData	⇨	OpenAmbDir
s. GetAmbDirec	⇨	SUBATOL
f. Initialize[4]	⇨	LoadATOT
s. LoadATOT	⇨	OpenProDir
s. MATMUL		
s. MATSIN[5]		
s. OpenAmbDir		
s. OpenProDir		
s. OpenSumDir		
f. ProRec		
f. SetAmbRec[3]	⇨	OpenAmbDir, SUBATOL
s. SETiWORK	⇨	ProRec
s. SETWORK		
s. SFitHead[3]		
s. SFitOut[3]	⇨	OpenSumDir, ProRec
s. SUBATOL		
s. WinMessage	⇨	MessageBoxC_(message)[6]
s. Write2ErrBuf		
s. Write2OutBuf	⇨	WinMessage
s. XLCunc		

[2] Calls LAPACK’s SGESVD

[3] Calls LAPACK’s SGESVD (only if numinestsource > 0)

[4] Called by CMB82.c

[5] Calls LAPACK’s SPOTRF & SPOTRI

[6] This string is in CMB82.c

Figure B-1 shows conceptually the general direction and flow across the main program units. The Delphi system that controls the GUI interacts with CMB82.c (via CMB82.dll). CMB82.c is critical for the Delphi interface, and serves as the main program unit that controls actions within the computation machinery in the Fortran code. The occurrence of specific **case**s in CMB82.c triggers fRun to make specific calls in CMB82a.for. While the intricacies of the entire logic flow aren't shown, you should get a *macro* sense of how program flow is directed. For orientation, the appearance of certain screen views in EPA-CMB8.2 are indicated.

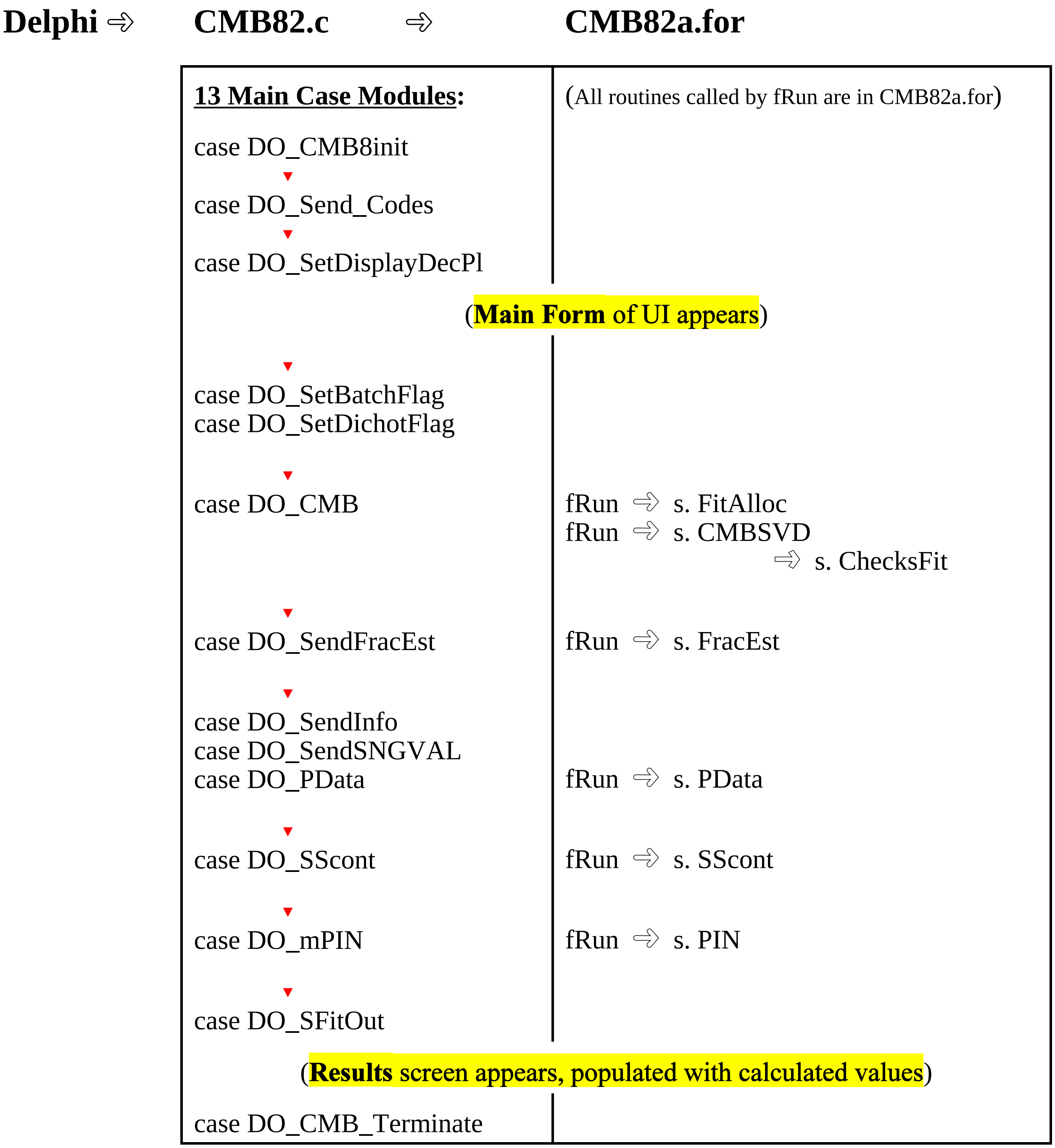


Figure B-1. Conceptual flow diagram for program sequence in EPA-CMB8.2.

Building the DLL for EPA-CMB8.2

The dll is compiled from code written in Fortran and C++ languages. This dll is resident in the installation subdirectory for EPA-CMB8.2 and is used by the Delphi client (executable) at run time (Appendix D). The 32 bit **dll**, CMB82.dll, is created with Compaq® Visual Fortran and Microsoft Visual C/C++ using the following instructions:

Assumption:

- Visual Fortran 6.6B[7] is installed (you may have to download the 6.5 ⇨ 6.6 and 6.6 ⇨ 6.6B or 6.6C updates from Compaq's website: http://compaq.com/fortran/visual/updates.html)
- Visual C/C++ v.6 (http://msdn.microsoft.com/visualc/) is installed.
- Intel® Fortran Compiler for Windows®
  (Available @ http://www.intel.com/software/products/compilers/fwin/index.htm) Intel® 7.1 (15 September 2004) is advisable; its installation executable is W_FC_PC_7.1.027.exe.[8]

Directions:

- Launch Visual Fortran Developer Studio (or Visual C++; they're quite similar for this purpose)
- Create a new Fortran Dynamic Link Library (File/New), calling the project *CMB82* and locating it somewhere like C:\CMB\Develop\CMB82. This will create a folder with this name.[9]
- In the one and only step in the App Wizard, select the option which reads "A DLL application with exported symbols" (not "An empty DLL application"), click Finish, and then click OK in the dialog that appears (which echoes the pathway to the project folder).
- Using Windows Explorer®, copy the CMB82 source code to the folder you just created (e.g., C:\CMB\Develop\CMB82).

The 8 source code files are:

CMB82.def
CMB82a.for
CMB82b.for
CMB82.c
CMB82.inc
WKShead.h
LAPACK.lib[10]
BLAS.lib[10]

[7] It may also be assumed that Visual Studio service pack 4 or 5 is installed. Note that Microsoft Visual C++ v6.0 has been upgraded to Visual C++.NET. However, please also note that Visual C++.NET reportedly does not integrate with Compaq® Visual Fortran.

[8] Once this compiler is installed, you will enable it within Visual Fortran by going to Tools | Customize and click Add-ins and Macro Files' check the Intel® Fortran Compiler Build Tool, and Close the dialog.
Note that, while later versions of the Intel® compiler are available, these require .NET versions of MS Visual Studio and MS Visual C++.

[9] These files will also be added:

CMB82.dsp	(project file)
CMB82.dsw	(workspace file)
CMB82.f90	(template for Fortran subroutine)
CMB82.opt	(project options file; appears after Workspace is closed)
CMB82.ncb	(browse information file built whenever a project is loaded; the tree view in the ClassView tab uses this file to populate itself)
CMB82.plg	(project log file; added after Build - F7)

[10] See Appendix C for information on LAPACK and for instructions on building the LAPACK *static* library.

1) switch back to Developer Studio
2) make sure the Workspace pane is visible, selecting Workspace from the View menu
3) click on FileView tab (not the ClassView tab)
4) expand the tree view until the one and only Fortran file (CMB82.f90) is visible under the folder Source Files
   - select this file in the tree and delete it
   - delete the Resource Files folder (it's not needed)
5) right-click on the ***project*** (the node which reads "CMB82 files") and select Add Files to Project
   - change the Files of type filter to All Files (***.***)
   - select CMB82.f90 and delete it
   - select all the source files listed above and click OK[11]
6) right-click on the ***project*** and select Settings.
7) select **All Configurations** in the **Settings For:** combo box
   - click on the **Fortran tab**
     - select Compatibility in the Category: combo box
       - check I/O Format and List Directed I/O Spacing in the PowerStation® 4.0 Compatibility Options listbox (Libraries should already be checked)
     - select Libraries in the Category: combo box
       - select Multi-threaded in the Use run-time library: combo box
8) select **Win32 Debug** in the **Settings For:** combo box
       - select Debug Multi-threaded in the Use run-time library: combo box
     - click on the **C++ tab**
       - select General from the Category: combo box
         - select Program Database in the Debug info: combo box
   - click OK to exit Project Settings[12]

9) click OK to exit Project Settings
10) press F7 to build[13]

The compiling and linking functions will proceed and details will be echoed to the output window at the bottom of the screen. The volume of the product DLL is minimized since CVF will only link the parts of the LAPACK library needed by EPA-CMB8.2; the debug version of the DLL should be of order 900KB; the release version is of order 700KB. Since the default configuration is Win32 Debug, the file *CMB82.dll* will be stored in the Debug folder within the project folder of similar name. To confirm/set the DLL destination,
click the Link tab under Project|Settings, select General in the Category, and adjust the path & name in the Output file name field.

[11] Take care to select only the 8 source files listed.

[12] When the Workspace is closed, *.opt will be added to the CMB82 project folder.

[13] To hold/use these settings for future builds (in Visual Fortran or Visual C++):
- select Options from the Tool menu
- click on the right scroll arrow in the upper right of the dialog until the Workspace tab is visible
- click on the Workspace tab
- check both of the “Reload ...” checkboxes
- File
  - Open Workspace
    - click on the project folder (e.g., CMB82)
- F7 (The output window at the bottom will echo linking details for any source files that were change since the last build.)

Completion of Build (F7) will add *.plg to the project folder.

APPENDIX C

The Linear Algebra Library for EPA-CMB8.2: LAPACK

In order to solve the effective variance, least-squares regression that CMB employs for an apportionment, notably the Singular Value Decomposition (SVD), the Fortran code accesses a library of linear algebra routines. CMB8.0 relies on the linear algebra library LINPACK.[14] LINPACK was designed for supercomputers in use in the 1970s and early 1980s and is a collection of Fortran subroutines that analyze and solve linear equations and linear least-squares problems. The package solves linear systems whose matrices are general, banded, symmetric indefinite, symmetric positive definite, triangular, and tridiagonal square. In addition, the package computes the QR and SVDs of rectangular matrices and applies them to least-squares problems. LINPACK uses column-oriented algorithms to increase efficiency by preserving locality of reference.

This library was up updated to LAPACK. LAPACK was developed by a consortium of university and government laboratories, and is the industry-standard subprogram package offering an extensive set of linear system and eigenproblem solvers. Reference:

Anderson, E, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney and D. Sorensen, 1999. ***LAPACK Users' Guide, Third Edition***. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; ISBN 0-89871-447-8 (paperback) http://www.siam.org/

LAPACK is written in Fortran77 and provides routines for solving systems of simultaneous linear equations, least-squares solutions of linear systems of equations, eigenvalue problems, and singular value problems. The associated matrix factorizations (LU, Cholesky, QR, SVD, Schur, generalized Schur) are also provided, as are related computations such as reordering of the Schur factorizations and estimating condition numbers. In all areas, similar functionality is provided for real and complex matrices, in both single and double precision.

The original goal of the LAPACK project was to make the widely used LINPACK libraries run efficiently on shared-memory vector and parallel processors. On these machines, LINPACK is inefficient because its memory access patterns disregard the multi-layered memory hierarchies of the machines, thereby spending too much time moving data instead of doing useful floating-point operations. LAPACK addresses this problem by reorganizing the algorithms to use block matrix operations, such as matrix multiplication, in the innermost loops. These block operations can be optimized for each architecture to account for the memory hierarchy (shared memory), and so provide a transportable way to achieve high efficiency on diverse modern machines. We use the term "transportable" instead of "portable" because, for fastest possible performance, LAPACK requires that highly optimized block matrix operations be already implemented on each machine. LAPACK routines are written so that as much as possible of the computation is performed by calls to the Basic Linear Algebra Subprograms (BLAS)[15]. While LINPACK is based on the vector operation kernels of the Level 1 BLAS, LAPACK was designed at the outset to exploit the Level 3 BLAS - a set of specifications for Fortran subprograms that do various types of matrix multiplication and the solution of triangular systems with multiple right-hand sides. Because of the coarse granularity of the Level 3 BLAS operations, their use promotes high efficiency on many high-performance computers, particularly if specially coded implementations are provided

[14] Developed by Jack Dongarra, Jim Bunch, Cleve Moler and Pete Stewart; 1 Feb 1984. Available from NETLIB: http://www.netlib.org/linpack/

[15] The BLAS library includes the industry-standard Basic Linear Algebra Subprograms for Level 1 (vector-vector), Level 2 (matrix-vector), and Level 3 (matrix-matrix) applications.

by the manufacturer. LAPACK has been thoroughly tested on many types of computers, and a list of known problems, bugs and compiler errors for LAPACK is maintained on NETLIB: http://www.netlib.org/lapack/release_notes.html

The LAPACK SVD (s. SGESVD) is better numerically than its counterpart in LINPACK (s. SSVDC). For this reason, as well as for certain increases in efficiency and “modularity” that lends itself to accommodating LAPACK updates, EPA-CMB8.2 has been modified to use the LAPACK library. As an additional consequence of this upgrade, LINPK.for (which contained a series of LINPACK routines used by CMB8.0) has been removed as a source file; it is replaced by a pair of companion *static* libraries: LAPACK.lib and BLAS.lib (Basic Linear Algebra Subprograms). These libraries can be reconstructed for building EPA-CMB8.2's main DLL (Appendix B) at any time that a new LAPACK becomes available. Building instructions are provided in this appendix.

The LAPACK routines are freely available and are copyrighted. The entire suite of (single-precision) routines for LAPACK v3.0 (latest version), including its requisite Level 1, 2 & 3 BLAS, was updated 31 May 2000 and is listed and described here (download ***lapack.pc.df.zip***): http://www.netlib.org/lapack/single/index.html

At run time, EPA-CMB8.2 directly calls these 3 Fortran routines in LAPACK (see Appendix B):

SGESVD[16] (called by s. CMBSVD & s. PData in CMB82a.for)
SPOTRF[17] (called by s. MATSIN in CMB82b.for)
SPOTRI[18] “ ” ” “ ”

These 3 routines subsequently call (or at least reference) the following 22 Fortran routines (listed alphabetically) - some in LAPACK, and others in BLAS (as indicated):

LAPACK		BLAS
ILAENV	SLASET	LSAME
SBDSQR	SLAUUM	SGEMM
SGEBRD	SORGBR	SSYRK
SGELQF	SORGLQ	STRSM
SGEQRF	SORGQR	XERBLA
SLACPY	SORMBR	
SLAMCH	SPOTF2	
SLANGE	STRTRI	
SLASCL		

[16] Computes the singular value decomposition (SVD) of a real M by N matrix A, optionally computing the left and/or right singular vectors. The SVD is written $A = U * \Sigma * V^T$, where Σ is an M by N matrix which is zero except for min(m,n) diagonal elements, U is an M by N orthogonal / unitary matrix, and V is an N by N orthogonal / unitary matrix. The diagonal elements of Σ are the singular values of A; they are real and non-negative, and are returned in descending order. The first min(m,n) columns of U and V are the left and right singular vectors of A.

[17] Computes the Cholesky factorization of a real, positive definite matrix A. The factorization has the form $A = U^H * U$, where U is an upper triangular matrix.

[18] Computes the inverse of the matrix A computed by SPOTRF.

Building the *static* LAPACK/BLAS Libraries for EPA-CMB8.2

To build the *static* LAPACK and companion BLAS libraries for EPA-CMB8.2 (see Appendix B), first obtain the 1291 LAPACK routines (12.2MB), and the 141 BLAS routines (966KB) from the source listed above.

Assumption:

- Visual Fortran 6.6B (or later)[19] is installed (you may have to download the 6.5 ⇨ 6.6 and 6.6 ⇨ 6.6B or 6.6C updates from Compaq's website: http://compaq.com/fortran/visual/updates.html)

- Intel® Fortran Compiler for Windows® (see footnote 8 in Appendix B)

Directions:

- Launch Visual Fortran Developer Studio (or Visual C++; they're quite similar for this purpose)
- Create a new Fortran Static Library (File/New), calling the Workspace (and 1st project) *LAPACK* and locating it somewhere like C:\CMB\Develop\CMB82. This will create a folder with this name.[20]

Build the LAPACK library

1) Using Windows Explorer®, copy the LAPACK source code (***.f** files) to the folder you just created (e.g., C:\CMB\Develop\LAPACK).
2) switch back to Developer Studio
3) make sure the Workspace pane is visible, selecting Workspace from the View menu
4) click on FileView tab (not the ClassView tab)
5) expand the tree view delete the Header Files folder (it's not needed)
6) right-click on the ***project*** (the node which reads "LAPACK files") and select Add Files to Project
   - change the Files of type: filter to Fortran Files
   - select all the source files listed in the Look in: window and click OK[21]

[19] It may also be assumed that Visual Studio service pack 4 or 5 is installed. Note that Microsoft Visual C++ v6.0 has been upgraded to Visual C++.NET. However, please also note that Visual C++.NET reportedly does not integrate with Compaq Visual Fortran. You will likely have to use the Visual C++.NET environment separately to compile the C code, and then use the newer linker to link the application.

[20] Eventually, all these files will also be added by CVF:

LAPACK.dsp	(project file)
LAPACK.dsw	(workspace file)
LAPACK.ncb	(browse information file built whenever a project is loaded; the tree view in the ClassView tab uses this file to populate itself)
LAPACK.plg	(project log file)

[21] This will load the 1291 source files into the LAPACK project.

Build the companion BLAS library

- right-click on the ***workspace*** (the node which reads Workspace 'LAPACK':) and select Add New Project to Workspace
  - select Fortran Static Library, enter "BLAS" for Project name:, and click OK twice.
- expand the tree view delete the Header Files folder (it's not needed)
- Using Windows Explorer®, copy the BLAS source code (***.f** files) to the folder you just created (e.g., C:\CMB\Develop\CMB82\BLAS).
- switch back to Developer Studio
  - right click on the BLAS project and select Add Files to Project. Select the \BLAS folder and change the Files of type: filter to Fortran Files
    - select all the source files listed in the Look in: window and click OK[22]

7) **For both static library projects - LAPACK & BLAS - the Library Settings must generally be set to match those for the main project ('CMB82' - see Appendix B). For both LAPACK & BLAS:**

- right-click on the ***project*** and select Settings.
- select **All Configurations** in the **Settings For:** combo box
  - click on the **Fortran tab**
    - select Libraries in the Category: combo box
      - select Multi-threaded in the Use run-time library: combo box
- click OK to exit Project Settings[23]

8) Build the libraries for both - LAPACK & BLAS - projects.

- click Project\Set Active Project ☞ LAPACK; press F7 to build[24]
- click Project\Set Active Project ☞ BLAS; press F7 to build

As the compiling and linking functions will proceed for each build command, details will be echoed to the output window at the bottom of the screen. Since the default configuration is Win32 Debug, the file(s) *name.lib* will be stored in the Debug folder within the respective project folder of similar name (this can be changed by clicking on the Library tab, then entering an alternate path/file name in the Output file name: field).

[22] This will load the 141 source files into the BLAS project.

[23] When the Workspace is closed, *.opt will be added to the LAPACK project folder.

[24] Note that compilation of the 1291 LAPACK routines will take some time. Occasional *informational* messages of the following type may be disregarded:

```
Info: This directive is not supported in this platform.
CDEC$         NOVECTOR
---------------^
```

Completion of Build (F7) will add *.plg to the project folder.

APPENDIX D

The User Interface for EPA-CMB8.2: Delphi

The following lists and describes the source code, on which the Delphi[1] client that serves as the User Interface (UI) for EPA-CMB8.2, is based:

Anatomy of the Delphi Source Code[2]

EPACMB82.res The project resource file. This is a binary file which normally contains the version info resource (if required) and the application's main icon. Changes to the version info and the application icon in Delphi's Project Options dialog will result in changes to this file.

EPACMB82.dpr The Delphi project file. This is actually source code for the application: the project file moniker is a misnomer.

EPACMB82.cfg The project configuration file. Stores the project configuration settings, primarily compiler and linker settings.

EPACMB82.dof The Delphi options file. Stores the project option settings, such as additional compiler and linker settings, directories, conditional directives, and version information.

StartupFrm.pas / StartupFrm.frm The source code and form file for displaying the startup screen, the first form shown when the application is run.

AboutFrm.pas / AboutFrm.dfm The source code and form file for displaying the banner 'About EPA-CMB8.2'.

BLWarningFrm.pas / BLWarningFrm.frm The source code and form file for displaying the Britt-Luecke warning message.

WaitFrm.pas / WaitFrm.dfm The source code and form file for displaying the temporary 'Please wait' message, shown when loading input data.

MainFrm.pas / MainFrm.dfm The source code and form file for displaying the main CMB model screen (Select Input Files; Options; Sample; Species; Sources; Results). About half of the application's logic can be found in the source code for the main form.

[1] Borland Software Corporation

[2] The bulk of the application's code is in **MainFrm.pas** and **CMBInternals.pas**. Delphi file types are described on p. D - 3.

PrintOptionsFrm.pas / PrintOptionsFrm.dfm The source code and form file for displaying the Print window enabled on the Results screen when the results are in the buffer. Most of the logic for printing reports or saving them to a text file can be found in the source code for the Print Options form.

SaveResultsFrm.pas / SaveResultsFrm.dfm The source code and form file for displaying the Save Results screen. This form allows saving of the current record or all included records displayed in the Main Report buffer. Most of the logic for saving result data to a text file can be found in the source code for the Save Results form.

CMBInternals.pas Controls all Delphi setup and interaction with CMB82.c in calls to the C++/Fortran DLL. Also contains a number of math routines, e.g. Best Fit, and the code which refines the output report generated by s. PData in CMB82a.for.

General notes for Delphi file types:

***.pas** Delphi source code file, called a *unit* (source code module), which is written in Pascal. Each application may contain several unit source files, which typically contain most of the code for the application.

***.dfm** Delphi form file. A project-related file, *.dfm contains the description of the properties of the *form* and the components it owns. This description may be present in text *form* (a format very suitable for version control) or in a compressed binary format. Each *form* file represents a single *form*, which usually corresponds to a window or dialog box in an application. The Delphi IDE (Integrated Development Environment) allows you to view and edit all *form* files as text, and to save *form* files as either text or binary. Each application has at least one *form* and, while not all *unit* files (e.g., a math library of functions and type information) have a corresponding *form* file, each *form* has a corresponding *unit* file (by default, having the same name as the *form* file). In the Delphi IDE, if a *unit* defines a form (i.e., uses a visual component such as TForm), a *form* file is automatically created. N.B.: If a unit file (e.g., OptionsDlg.pas) defines a form and therefore has a corresponding form file (e.g., OptionsDlg.dfm), **both** must be resident in the working directory in order to open either in the IDE or Delphi will complain.

***.dpr** Delphi project file. Another source file, *.dpr is explicitly created by association with a project in the Delphi IDE. The project file (which corresponds to the ”main” program file in traditional Pascal) organizes the unit files into an application. The Delphi IDE automatically creates and maintains a unique project file for each application. The project file may be created by selecting a new project from the File menu in Delphi, e.g., File | New Application. When the gallery dialog appears, double click on the icon entitled 'Application'. Delphi will then create a new project file for you. It will create a file called unit1.pas (and unit1.dfm) as a default because Delphi assumes that you want to have a main form in your application.

***.dof** Delphi options file. Contains the current settings for project options, such as compiler and linker settings, directories, conditional directives, command-line parameters and version information. These are set using the Project Options dialog box (Project | Options), and Delphi saves them in text form for easy maintenance, version control, and sharing. Each project has an associated options file with the same name as the project (*.dpr) file.

***.res** Delphi resource file. A project-related file, it contains the version information resource (if required) and the application's main icon. This file may also contain other resources used within the application but these are preserved as is. Do not delete this file if your application contains any references to it.

***.cfg** Project configuration file. This file stores project configuration settings and has the same name as the project file.

Construction of the Delphi executable for EPA-CMB8.2

The Delphi client (executable) calls *CMB82.dll* (Appendix B) at runtime. This client controls the user interface for EPA-CMB8.2 and is created with Delphi 7, a product of Borland Software Corporation (http://www.borland.com/delphi/), using the following instructions:.

Assumptions:

- Delphi 6 or 7 is installed.

Ordinarily, the compiler should be optimized for a build:
Project | Options | Compiler | Code generation: check *Optimization*

Directions:

- Ensure that all of the following 19 Delphi source files are stored in a single working directory (folder):

EPACMB8.2.res
EPACMB82.dpr
EPACMB82.cfg
EPACMB8.2.dof

StartUpFrm.pas	StartUpFrm.dfm
AboutFrm.pas	AboutFrm.dfm
BLWarningFrm.pas	BLWarningFrm.dfm
WaitFrm.pas	WaitFrm.dfm
MainFrm.pas	MainFrm.dfm
PrintOptionsFrm.pas	PrintOptions.Frm.dfm
SaveResultsFrm.pas	SaveResultsFrm.dfm

CMB82Internals.pas[3]

- Launch Delphi.
- Select **Open Project** from the **File** menu.
- Browse to the folder that contains the source files.
- Select the file **EPACMB82.dpr** and click **OK**.
- Select **Build EPACMB82** from the **Project** menu.
- The built executable should now appear in the working directory (folder), along with the following compiled *unit* files (*.dcu):

[3] N.B.: The original call in CMB8.0 to CMB8wn32.dll from:
UNIT1.PAS(157): arg11, arg12 : Pointer) : LongInt; stdcall; External 'cmb8wn32.dll' name '_CMB80C@52'

was changed in EPA-CMB8.2 to a call to CMB82.dll from:
CMBInternals.pas(260): arg12: Pointer): Integer; stdcall; external 'CMB82.dll' name 'CMB82c'

WaitFrm.dcu
StartUpFrm.dcu
SaveResultsFrm.dcu
PrintOptions.Frm.dcu
MainFrm.dcu
BLWarningFrm.dcu
AboutFrm.dcu
CMBInternals.dcu

These intermediate *.dcu files , which are binary images for each *unit* file, are necessary to create the final executable file (EPACMB82.exe). However, after the **.exe** has been created, you may delete them. The next time you compile the application, the Delphi IDE will regenerate them automatically.

APPENDIX E

Warning and Error Messages from EPA-CMB8.2

There are many different warning and error messages that EPA-CMB8.2 may produce to indicate that an input file is not named properly or does not follow the proper data format, that a model run is not set up correctly, or that a run did not produce satisfactory results. Most of the error messages that involve user intervention include suggestions for appropriate corrective actions. See Appendix F for a more detailed discussion of EPA-CMB8.2's behavior that may result in certain error conditions. A **[]** indicates a numeric value or alphameric string.

CMBInternals.pas[1]

Error reading source selection filename. Please refer to the manual for file naming convention.
Error reading species selection filename. Please refer to the manual for file naming convention.
Error reading ambient data selection filename. Please refer to the manual for file naming convention.

Error reading ambient sample data filename. Please refer to the manual for file naming convention.
The ambient sample data file does not exist.

Error reading source profiles data filename. Please refer to the manual for file naming convention.
The source profiles data file does not exist.

**** NO CONVERGENCE AFTER [] ITERATIONS ****

AKT*VEFFIN*AK MATRIX NEEDS IMPROVEMENT. CHANGE FITTING SOURCES OR FITTING SPECIES.[2]

The number of fitting sources [] must be positive and <= the number of fitting species [].

Memory allocation error. Try shutting down other applications or increasing RAM.
File Open Error. Check that the file exists or try rebooting.
Species Set Error. Please see documentation regarding input file construction.
Sources Set Error. Please see documentation regarding input file construction.
The number of fitting sources must be positive and <= number of fitting species.

[1] These messages originate from the Delphi client.

[2] If a model run does not converge, the results of that run will not be reliable. This does not have any negative impact on other successful runs within the same EPA-CMB8.2 session, so the other results may still be used.

Positive uncertainties are required for all fitting species and sources.
This error should not occur. Contact Tom Coulter - Coulter.Tom@epa.gov
The fitting source profiles matrix (**Afit**) contained a column of all zeros. Check the source profiles data file or change the fitting sources.
The choices for fitting species and sources have produced a near-singular matrix. There is most likely collinearity between two or more of the fitting sources.
The fitting algorithm failed to converge. There is most likely collinearity between two or more of the fitting sources.
File Read Error. Please see documentation regarding input file construction.
Unknown error.
File Close Error. If your disk is full, try freeing up some space, or try rebooting.
File Write Error. If your disk is full, try freeing up some space, or try rebooting.
Ambient Data Error. Please see documentation regarding input file construction.
Source Profiles Error. Please see documentation regarding input file construction.
Error reading species selection file. Please see documentation regarding input file construction.
Error reading source selection file. Please see documentation regarding input file construction.
Error reading ambient data selection file. Please see documentation regarding input file construction.
Error reading ambient data file. Please see documentation regarding input file construction.
Error reading source profiles file. Please see documentation regarding input file construction.
Memory free error. Try shutting down other applications or increasing RAM.

MainFrm.pas[1]
The ambient sample data file is empty!
The ambient sample data file is open!
Got an empty line of data from the ambient sample data file!
The number of data values for this sample is wrong!
Didn"t get any data from the ambient sample data file!
The source profiles data file is empty!
The source profiles data file is open!
The source profiles selection file [] does not exist. Please edit the Control File and try again.
The species selection file [] does not exist. Please edit the Control File and try again.
The ambient data selection file [] does not exist. Please edit the Control File and try again.
The ambient sample data file [] does not exist. Please edit the Control File and try again.
The source profiles data file [] does not exist. Please edit the Control File and try again.
Exactly *ONE* sample must be selected for the Best Fit option.
The number of fitting species selected must be greater than or equal to the number of fitting sources selected.

The number of size ranges in the selected samples exceeds 4 for the same site and sampling period.

CMB82.c[3]

Can't open CMB8.2 ambient data file: []

(=> The file doesn’t exist, is open and locked (located?) elsewhere, or is read-only.)

Incorrect number of tokens in header line of file []. Should be an odd number > 5

(=> The header line in the ambient sample data file is invalid.).

Total number of species = []

(Total #spp <= 0)

Can't open CMB8.2 source profiles data file: []

(=> The file doesn’t exist, is open and locked (located?) elsewhere, or is read-only.)

Can't create CMB8.2 direct access temporary file

(=> The file doesn’t exist, is open and locked (located?) elsewhere, is read-only, or the user’s machine is out of free disk space)

Memory allocation error

(=> too many tokens)

Number of tokens in line [] of file [] not equal to number of tokens in header line[4]

Total number of profiles data records = []

(Total #src profiles <=0)

Memory allocation error for [] from line []

Pointer [] not defined in ptinfo, line = []

Unknown pointer [] from line []

Memory overrun for pointer [] from line []; size = []

Unknown action [] from line []

Not able to open output file: []

[3] These messages originate from the C / Fortran DLL.

[4] The file listed may be either an ambient data (AD*.*) file or source profile (PR*.*) file; see Appendix F.

CMB82a.for[2]
(all in s.CMBSVD)

NUMBER OF FITTING SOURCES = [] > NUMBER OF FITTING SPECIES = []

FITTING SPECIES [] HAS ZERO AMBIENT DATA UNCERTAINTY.
Replace uncertainty with a positive detection limit or remove the species from the fitting species list..

No record for source profile code [] and size []

Column (source) Number [] of Afit = 0[5]

AKT*VeffIn*AK matrix needs improvement. Change fitting sources or fitting species.[6]

NO CONVERGENCE AFTER [] ITERATIONS.

[5] This error condition is caused by a column of the fitting matrix containing all zeros. The column corresponds to one of the fitting sources. A zero column for a fitting source means that all of the fitting species' fractions are zero for that source. The fitting algorithm can't handle a zero column in the fitting matrix (it's the matrix equivalent of division by zero).

[6] The AKT*VeffIn*AK matrix is used internally by the fitting algorithm. If one of its columns consists of elements close to zero, the algorithm has problems because of computer round-off error, in which case MATSIN returns info = 1. This condition is usually caused by collinearity among fitting sources. Changing fitting sources and/or fitting species is intended to remove the collinearity.

APPENDIX F

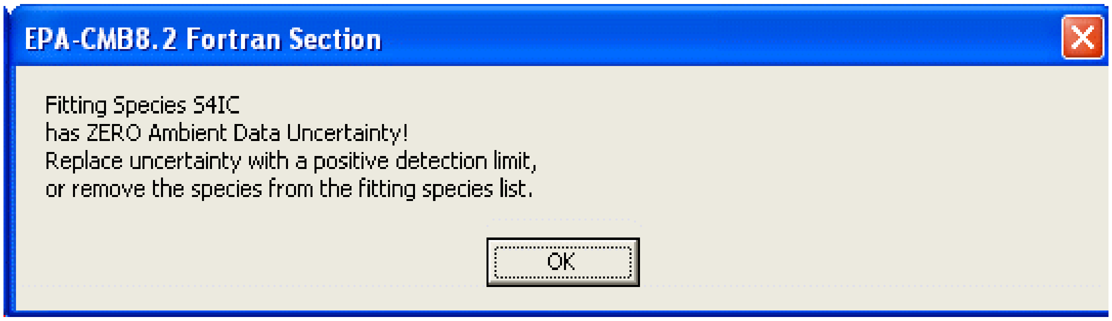
Data Input Issues for EPA-CMB8.2

In Section 3.6.1 we mentioned that EPA-CMB8.2 detects “missing values” for species concentrations by their flags (**-99.**) in the input files, and that when these flags are encountered, the species is “automatically removed from the calculation” and the selection flag in the Main Report is changed from an asterisk (*****) to ‘**M**’. Some discussion of EPA-CMB8.2 interpretation and treatment of data values encountered in either ambient sample or source profiles is in order. The first thing to note is that EPA-CMB8.2's behavior when it encounters certain kinds of data values in these two input sources is not parallel, and this is so for reasons that will be explained below.

Ambient (Sample) Data.

We stated in Section 4.2.2 that a species for which the concentration is missing (i.e., invalid) cannot be used as a fitting species for that sample and that concentration value must be substituted by **-99.** in the input data file. We also stated that the species will automatically be removed from the calculation. But what does that mean? By “removed from the calculation”, we mean that they are removed as fitting species but concentrations are calculated in the total list of source contributions so long as entries appear in the source profiles. Recall in Section 3.2 and 4.3.2 we mentioned *floating* species. These are species that participate in the CMB calculation (apportionment) but are not used as fitting species. When a species’ mass is missing in a particular ambient sample, CMB will attempt to apportion mass for that species as long as it is designated as a fitting species (Species Selection screen, Section 3.4) and EPA-CMB8.2 “finds” it in the source profiles. In this case, the species is treated as a floating species, and mass is apportioned. While in the Main Report there are appropriate indicators that the species **is** missing, a calculated mass appears in the Main Report to tell us what we should expect in the ambient sample (which allows us to estimate a value for the missing data), but more importantly it tells us if the apportionment is reasonable.

If, however, a legitimate (≥ **0**) value for the species concentration is entered, then the associated value for uncertainty MUST be greater than zero or else the following error message will be returned :


EPA-CMB8.2 Fortran Section

Fitting Species S4IC
has ZERO Ambient Data Uncertainty!
Replace uncertainty with a positive detection limit,
or remove the species from the fitting species list.

and execution will be halted during the first iteration attempt. To implement the effective variance least squares solution, the ambient uncertainty can never be zero because the first step in the iteration assumes that all source contributions are zero. That is, in the first iteration, there is no contribution to the effective variance from any of the source uncertainties (this is equivalent to the ordinary weighted least squares method described by Friedlander, 1973). Thus, EPA-CMB8.2 starts out by dividing by zero if there is no ambient data uncertainty for a fitting species and the program terminates.

Source Profile Data.

For source profile input data files, we stated in Section 4.2.3 that missing *mass fraction* values must be substituted by **-99.** and that the associated species will automatically be removed from the calculation. By this we mean that they are removed as fitting species and **no concentrations are calculated**. If a mass fraction is substituted by **-99.** in the source profile data file, when EPA-CMB8.2 is run the value appearing for that species in the Main Report under "CALCULATED" will be a large negative value (if the value for number of decimal places displayed is set low enough; Section 3.2). The values listed for the diagnostics CALCULATED/MEASURED and RESIDUAL/UNCERTAINTY will, of course, be meaningless. The value for the species in question listed in the Contribution by Species for the affected source will also be meaningless.

As we said in Section 4.2.3, while uncertainty values for species in source profiles are allowed to be ≤ 0.0, some effort should be made to supply values > 0.0. However, for appropriate reasons, the same kinds of behavior (error messages, etc.) do not occur with equivalent input data conditions as described above for the ambient data files. For example, if for a particular species in a source profile record the mass fraction is valid (i.e., ≥ 0.0), EPA-CMB8.2 allows an associated uncertainty value to be ≤ 0.0. No error message will appear and EPA-CMB8.2 will execute normally.

Recall from Appendix A the effective variance weighting to which all of the ambient concentrations are normalized. In Equation A-6, the effective variance is the concentration uncertainty squared plus the sum of source contribution squared times the corresponding elemental mass fraction uncertainty squared:

$$V_{e_{ii}}^{k} = \sigma_{C_i}^{2} + \Sigma(S_j^{k})^2 \bullet \sigma_{F_{ij}}^{2}$$

Consider a situation for a woodsmoke profile in which the uncertainty for K (potassium) was input as ≤ 0.0 for a mass fraction > 0.0.

If the uncertainty value for K < 0.0 (e.g., -99.), this uncertainty in the woodsmoke profile may overwhelm all of the other profile uncertainties (which are less than one because profiles are normalized to unity). This would effectively remove K as a fitting species, because of division by the effective variance resulting in a very low weight for this species. This will also propagate as a very high uncertainty on the source contribution estimate and on the calculated value for K.

If the uncertainty for K = 0.0, there are other components of the effective variance from the K uncertainty in the ambient sample as well as from the K uncertainties in the other profiles, so there is not division by zero just because K uncertainty in the woodsmoke profile is zero.

You may also obtain an *unweighted* least squares solution by setting all of the source profile uncertainties to zero and all of the ambient concentration uncertainties to the same value (e.g., 1.0). This gives the solution that you would get by multiple linear regression in a spreadsheet with the intercept forced through zero. The solution in this case is dominated by the species with the highest concentrations. Species such as Se (selenium) and V (vanadium) might get no representation with an unweighted least squares solution.

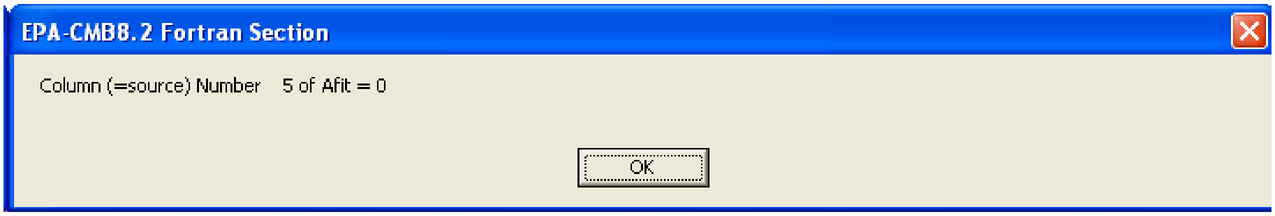
For source profiles, in general, you wouldn't enter a value of **-99.** for a species' uncertainty unless you also entered -99. for its abundance. A value of zero for a species' uncertainty might be used for testing of least-squares approaches or for convenience, as illustrated in the example above. In EPA-CMB8.2 the decision was explicitly made NOT to force ("hardwire") source profile uncertainties > 0.0 for the following reasons:

1. As noted above, zero uncertainties in the source profiles default to the ordinary weighted least squares (OWLS) method. This is a solution that some may still use. (Given what is known about the dominance of source profile variability on the error, it has been shown that the OWLS solution is biased compared with an effective variance solution.)

2. When data are unavailable for some components that are not likely to be in a source profile, you should insert zeros for the value and the uncertainty. An example is in the NFRAQS data set (Section 2.2), where we have lots of organic compounds for wood burning and cooking, and diesel and gasoline vehicles, but none for the soil or the road salt or the secondary ammonium nitrate and ammonium sulfate. You could put in a zero entry and a very low uncertainty, e.g., $1x10^{-7}$, to indicate a detection limit. But this is essentially no different from zero when it is squared and added into the effective variance with the uncertainties from those profiles that have real values. It is also effectively zero when compared to the ambient data uncertainty.

Source Profile Matrix: Afit

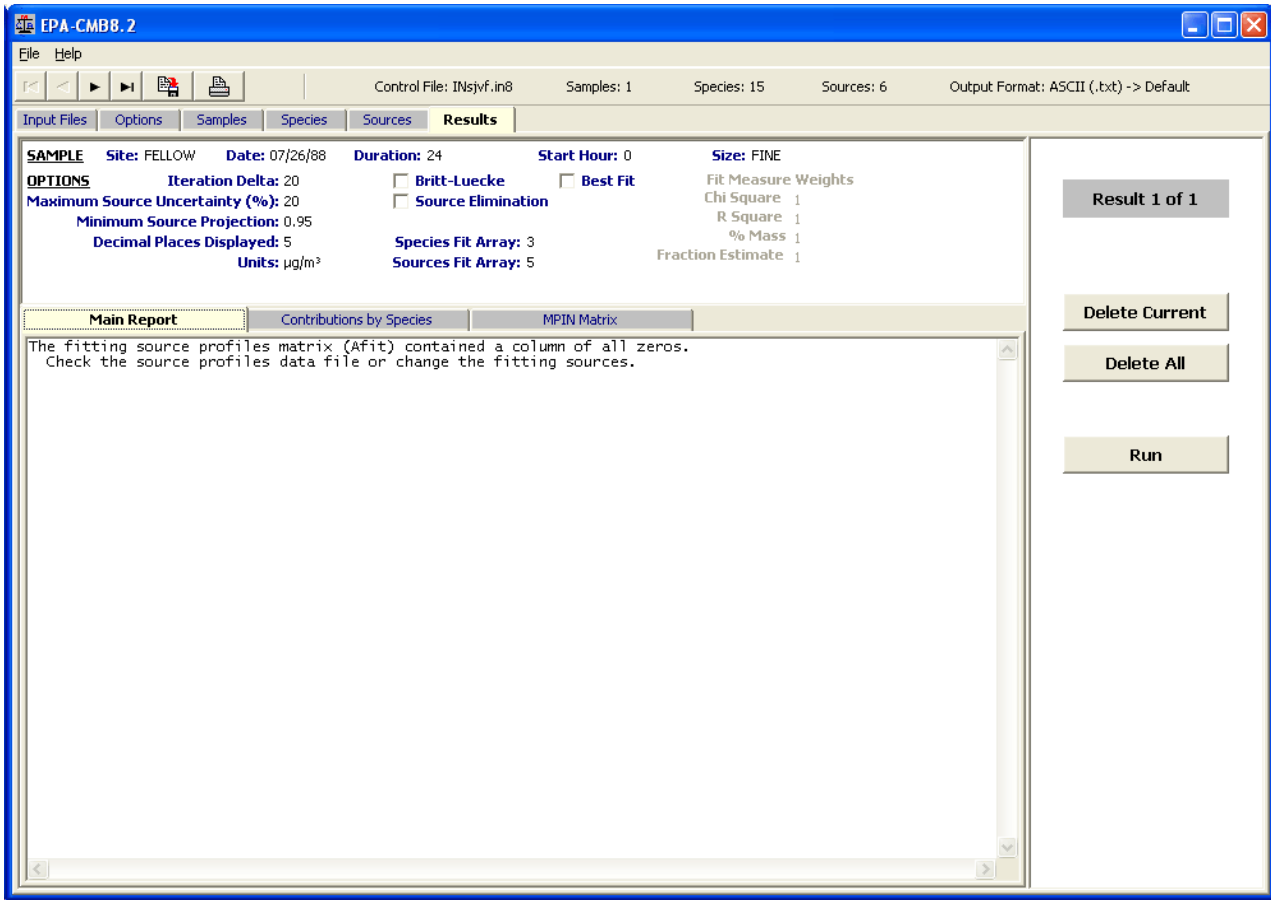
Sometimes a special situation may arise that affects the choice of fitting species (Section 3.4). This situation is likely to occur when *single constituent sources* are used as fitting sources. Single constituent sources are described elsewhere (e.g., EPA, 2004; Appendix G) but essentially involve the designation of secondary compounds as "sources" to allow EPA-CMB8.2 to account for secondarily formed materials observed in ambient samples. For such modeling scenarios, secondary compounds are involved that require selection of the constituent cations or anions. If, for example, you have sodium nitrate as a fitting source, but haven't selected (included) sodium or nitrate as fitting species, you will get a zero column in the source profiles (AFIT) matrix. When this happens, the following error message can be returned when a calculation is attempted:



Column (=source) Number 5 of Afit = 0

When 'OK' is clicked, the screen shown below appears. When EPA-CMB8.2 attempted to apportion the fitting sources to the ambient sample, it expected certain species for one or more single constituent sources (secondary compounds) that were excluded in the fitting species array.

This message will commonly appear when EPA-CMB8.2 is run in the *Best Fit* mode (Section 3.2 & 5.2). As EPA-CMB8.2 advances through various corresponding pairs of fitting species and source profile arrays, it is apt to encounter inappropriate combinations. The model will advance through all possible pairs, but will produce invalid results for the pairs that are ill-suited.



Control File: INsjvf.in8 Samples: 1 Species: 15 Sources: 6 Output Format: ASCII (.txt) -> Default

SAMPLE Site: FELLOW Date: 07/26/88 Duration: 24 Start Hour: 0 Size: FINE

OPTIONS Iteration Delta: 20 □ Britt-Luecke □ Best Fit

Maximum Source Uncertainty (%): 20 □ Source Elimination

Minimum Source Projection: 0.95

Decimal Places Displayed: 5 Species Fit Array: 3

Units: µg/m³ Sources Fit Array: 5

Fit Measure Weights

Chi Square 1

R Square 1

% Mass 1

Fraction Estimate 1

```
The fitting source profiles matrix (Afit) contained a column of all zeros.
  Check the source profiles data file or change the fitting sources.
```

Logic suggests that the 2nd line of the message should read: "... or change the fitting species." Specifically for this example, the error message might have said (or been interpreted to say): "Select a fitting species that corresponds to a positive entry for that species in the NaNO3 profile." This error message will also be triggered for situations when there are many organic species and a soil profile when we eliminate all of the geological elements. We usually have zero entries for the organics in the soil profile because we don't measure them. It also happens when a biogenic VOC source profile is included but there's no isoprene, and for many of the single constituent solvent and coating profiles (e.g., toluene, acetone, etc.).

Using EPA-CMB8.2 with Factors besides Chemical Abundance: e.g., Wind Direction.

As discussed above, an entry of **-99.** for species abundance (mass fraction) in a source profile automatically removes that species from the source contribution calculation. It does not remove it from the concentration calculation if there is also an entry for that species in the ambient data file. This allowance of discrepancy, i.e., a valid (>0.0) value for a constituent in the ambient sample data file and a value of **-99.** for its abundance in the source profile file(s), can be exploited for cases when we want to include other information in the ambient data file that is not appropriate for the source profile. For example, wind direction (WD) can be included to indicate that an upwind source profile should be used or not used. The **-99.** for this variable (factor) in the source profile file does not allow it to be used as a fitting species, but the direction is still easily available for consultation when picking sources. This is more efficient (and convenient) than using a separate data base or look-up table for WD. It also allows the source contributions to be stratified by wind direction (or other variable such as light scattering, temperature, or wind speed) by including it in the ambient sample data (receptor) file.

The structure exists to display such external information corresponding to each sample that is not, and cannot be, used as a fitting species. For example, one of the ways we recommend to choose a profile for a sample is to determine if a source is upwind or downwind. For the sample period(s) in question, the period-averaged WD can be imported directly into the ambient sample data file as a "constituent" of each sample. The data file can thus be configured with a WD variable into which you put the resultant wind direction, and set up a corresponding variable in the source profile with a **-99.** value for "abundance". This information can be easily viewed on-screen along with the species' concentration information (Section 3.3). If, for example, you see that the sample data indicate that it is downwind (± 5°) from a coal-fired power plant, and you also see a lot of selenium in this sample which doesn't appear in previous samples with prevailing winds from opposite or very different directions, you may decide to include the power plant profile, or retain a source combination you've already set up with that profile in it.

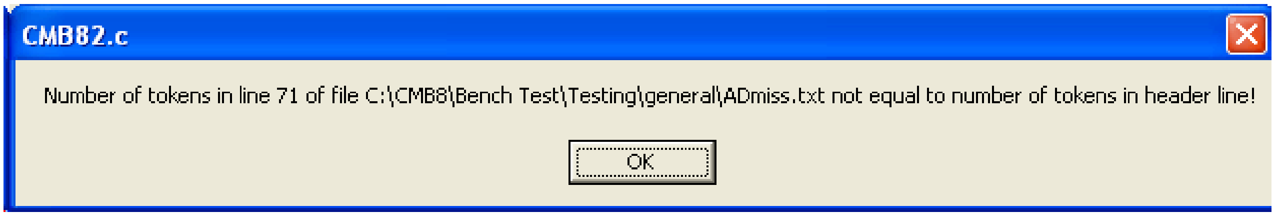
Highest wind speed during a sampling period is also good consideration for questions such as "Do I want the road dust or the windblown desert dust profile in this sample?" It's also useful when you have other data, like CO and NO_x, VOC, etc. at the receptor that might help you to choose (and justify the choice of) a certain profile or source type, but which doesn't have an entry in any of your source profiles. By including these variables as sample "constituents", it puts all the information in front of you rather than forcing you to analyze from several different (external) sources of information.

ABSENT DATA

One condition in which EPA-CMB8.2's behavior is parallel with respect to ambient sample or source profile data is that of "absent data". This condition is represented not by a flag for "missing" (invalid) data but rather the complete omission of the value from the input data record.

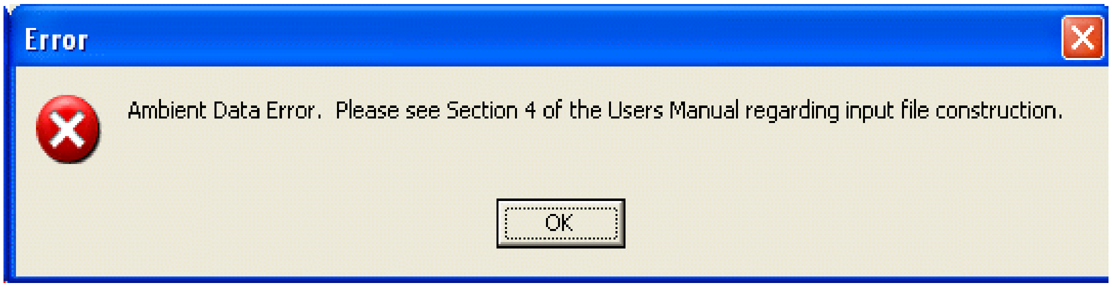
Ambient Sample Input File (AD*.*):

If a value for a species concentration, for example, is absent (omitted) from the ambient sample input data file, an error message such as the following will be returned:


CMB82.c

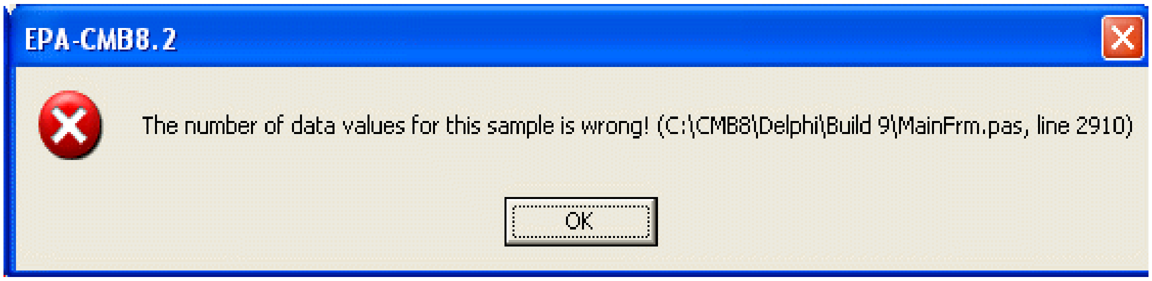
Number of tokens in line 71 of file C:\CMB8\Bench Test\Testing\general\ADmiss.txt not equal to number of tokens in header line!

as soon as EPA-CMB8.2 control is moved off of the Input Files screen. As indicated, on line 71 of the ambient data input data file there is a missing (= absent) value. The following message appears immediately after 'OK' is clicked on the message above:



Ambient Data Error. Please see Section 4 of the Users Manual regarding input file construction.

The following message then appears immediately after ‘OK’ is clicked on the message above:

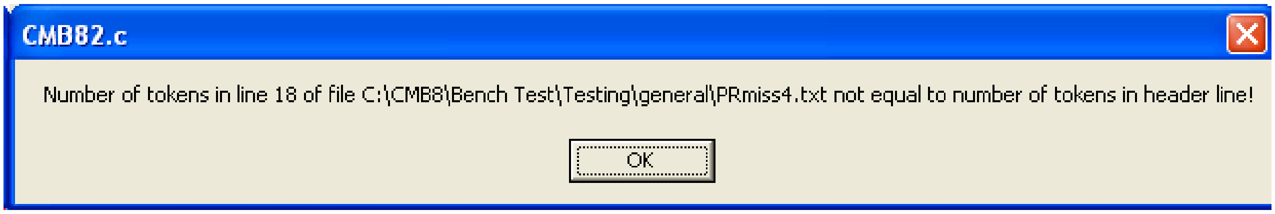


The number of data values for this sample is wrong! (C:\CMB8\Delphi\Build 9\MainFrm.pas, line 2910)

EPA-CMB8.2 must then be closed and an appropriate edition made to the input file after careful examination.

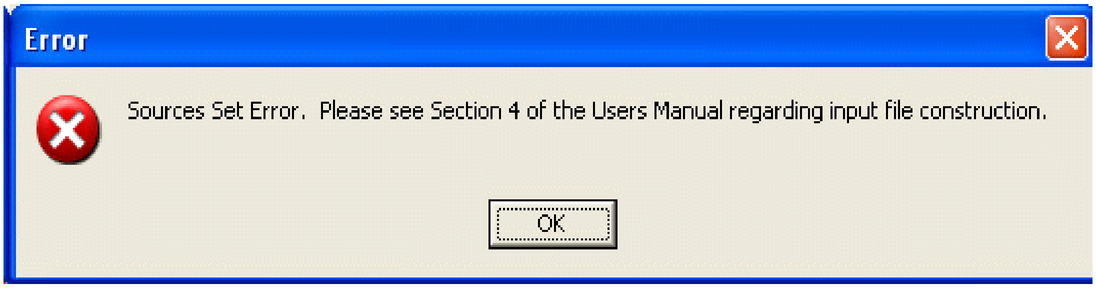
Source Profile Input File (PR*.*):

If a value for a species concentration, for example, is absent (omitted) from the source profile input data file, an error message such as the following will be returned:



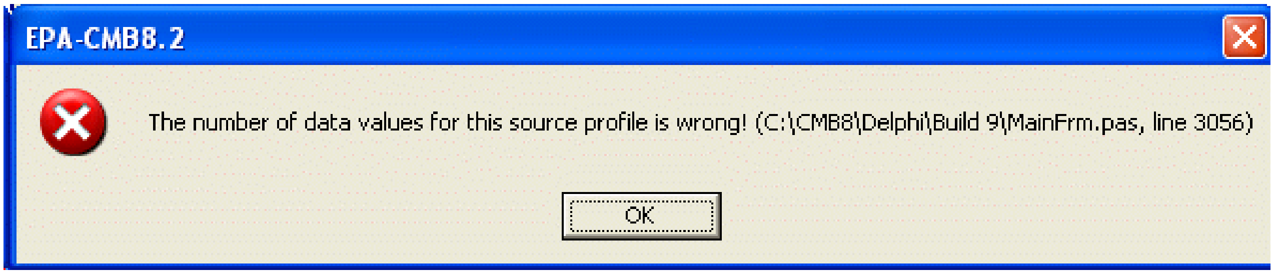
Number of tokens in line 18 of file C:\CMB8\Bench Test\Testing\general\PRmiss4.txt not equal to number of tokens in header line!

as soon as EPA-CMB8.2 control is moved off of the Input Files screen. As indicated, on line 18 of the source profile input data file there is a missing (= absent) value. The following message appears immediately after ‘OK’ is clicked on the message above:



Sources Set Error. Please see Section 4 of the Users Manual regarding input file construction.

The following message then appears immediately after ‘OK’ is clicked on the message above:



The number of data values for this source profile is wrong! (C:\CMB8\Delphi\Build 9\MainFrm.pas, line 3056)

As for the example illustrated above for the ambient data file, EPA-CMB8.2 must be closed and an appropriate edit be done on the input file.

APPENDIX G

Notes on EPA-CMB8.2 Diagnostics

EPA-CMB8.2's Main Report: EST

In the top portion of the Main Report is the diagnostic attribute “EST”, which may be either “YES” or “NO”. The manual states in several places (e.g., Sections 3.6.1 & 5.1.1) that “The field ‘EST’ under ‘SOURCE’ indicates (YES or NO) whether a source’s contribution was estimable in EPA-CMB8.2's attempt at a fit, using the settings in Options.” This is also stated in Table 4.2-1 of the *Protocol for Applying and Validating the CMB Model for PM-2.5 and VOC.* (EPA, 2004). In particular, it is first order indication of whether or not the *profile/source contribution estimate combination*[1] meets or does not meet the criteria that have been set up in Options. For any particular CMB fit, it is possible to set values for Maximum Source Uncertainty and Minimum Source Projection in Options to indicate all YESs or all NOs (see below).

We currently don't know how those settings translate into useful information, as indicated by the YESs and NOs. Unless you're researching how CMB responds to different combinations of source profiles and uncertainties, this parameter may not be very useful. The most important evaluation parameter is the *uncertainty of the source contribution estimate*, which is a function of the input uncertainties and the collinearity (variance inflation). You'll notice that when you have a negative source contribution, it usually has a very high uncertainty and there is another compensating source contribution (usually a very high positive value) that also has a high uncertainty. This is a good indication that these are collinear. This is why the model gives the uncertainties and why they should be considered in decision-making, as described in the *Protocol for Applying and Validating the CMB Model for PM-2.5 and VOC.* How to use this diagnostic in a practical setting has not been sufficiently studied. For now, it is really there for research purposes and not for everyday application.

In *sensitivity testing* for the two Options parameters mentioned above, the following has been seen:

Maximum Source Uncertainty.

The default is value is 20%. As this value is reduced, the occurrence of NOs increases. Once this value is set to zero, EST = NO for all sources. Conversely, as this value is increased, the occurrence of YESs will increase.

Minimum Source Projection.

The default value is 0.95%. As this value is substantially reduced, the occurrence of YESs increases. As it approaches zero, EST = YES for all sources. If set to 1.00, EST = NO for all sources.

The array (pattern) of YES/NO under EST is fairly sensitive to settings of Maximum Source Uncertainty; it is fairly insensitive to settings for Minimum Source Projection. However, adjustment of these two parameters has an independent (and opposite) effect on the EST variable.

[1] This “combination” is important because a profile may be well separated for a high contribution, but it may fall within the uncertainties of the other profiles for a low contribution.

United States
Environmental Protection
Agency

Office of Air Quality Planning and Standards
Air Quality Strategies and Standards Division
Research Triangle Park, NC

Publication No. EPA-452/R-04-011
December 2004