Updates and Improvements to the SPECIATE Data and Program

William Hodan
Pacific Environmental Services, 5001 S. Miami Blvd., Suite 300, RTP, NC 27709
wmhodan@mactec.com

ABSTRACT

The Emission Factor and Inventory Group (EFIG) of EPA is currently involved in an effort to re-mission and restructure the SPECIATE program. In response to comments and suggestions from the user community and a general need to add current speciated emission profiles to the SPECIATE database, EFIG has formed an expert panel consisting of members of the SPECIATE user community and source profile researchers. The expert panel has generated a number of ideas and suggestions that have been compiled in a software plan for review by EFIG and EPA as a whole. The expert panel has given insight on topics including profile additions and updates, old profile deletion, expansion of existing references, profile normalization methodology, software platform changes, new profile identification, emission source prioritization and other issues that will be discussed in detail in this paper.

Concurrent with development of the software plan, a literature search was conducted to locate new emission profiles published in recent research. Approximately 100 applicable technical articles containing Total Organic Compound (TOC) and Particulate Matter (PM) emission profiles were located. Profiles from these technical articles will be incorporated into the next release of SPECIATE following the new format recommendations provided by the expert panel.

This paper will discuss the issues identified by the SPECIATE user community and the expert panel, and will describe the resolution of those issues. Also included will be a presentation of the improvements to be included in the next release of SPECIATE and the time frame for its development.

INTRODUCTION

In response to comments and suggestions from the user community and a general need to add current speciated emission profiles to the SPECIATE database, the Emission Factors and Inventory Group (EFIG) of EPA tasked PES to assemble an expert panel of emission profile developers and emission modelers to assist in compiling a software development plan for SPECIATE. PES interviewed local members of the panel to identify issues and suggestions for improvement of the current platform. Additionally, local EPA and the North Carolina Department of Environment and Natural Resources (NCDENR) Division of Air Quality (DAQ) personnel have been interviewed in an effort to develop an understanding of all the parcels of knowledge that must come together to satisfy the SPECIATE user community so that it remains the primary source of emission profile data. PES completed the first phase of formation of the expert panel by organizing and conducting a meeting of the panel which was held at the American

Association for Aerosol Research (AAAR) conference in Charlotte, NC on October 10, 2002. By conducting these interviews and meetings, PES has gained knowledge of the data and data types required by SPECIATE users, and the types of data available.

BODY

One point fundamental to understanding the types of data and format necessary to be included in SPECIATE is the difference in the two types of users: (1) Air Quality Modelers and (2) Source Apportionment/Receptor Modelers. Both types of SPECIATE users are discussed below.

Air Quality Modelers

Air Quality Modelers use SPECIATE primarily for ozone modeling, and therefore are particularly interested in the SPECIATE TOC module. Users of CMAQ, CAMx, REMSAD, and other air quality models require profiles that sum to 100%. For profiles that do not sum to 100%, the Air Quality Modeler must compensate for the missing percentage either by adjusting the existing percentages to "normalize" the profiles to 100%, or adding a percentage of "Unknown" to the profile.

Air Quality Modelers require Source Classification Code (SCC) assignments to profiles to enable them to speciate their emission inventory and subsequently run an emission model. For efficiency, the Air Quality Modelers prefer to obtain the SCC to profile assignment file as an individual file consisting of one-to-one assignment of profiles to SCCs.

Due to the current function and operation of air quality models, chemicals are grouped together based on one of several methods that are dependent on the type of emission model and chemical mechanism being used. This chemical grouping is indirectly dependent on the outdated system of Storage and Retrieval of Aerometric Data (SAROAD) codes. Although the air quality models do not require SAROAD codes specifically, SAROAD codes have historically been used as the mapping tool from individual chemical species to the "Carbon Bond" and other chemistry mechanisms. Although there is currently no centralized maintenance of the SAROAD codes, several State and other agencies have assigned SAROAD codes independently. This assignment process has resulted in differences in the SAROAD code system used by various agencies, including chemicals with multiple code assignments. The lack of maintenance of the SAROAD codes, and their resulting absence in SPECIATE, has caused difficulty and inconsistent mapping of the profile species to the corresponding chemical mechanism assignments.

Source Apportionment/Receptor Modelers

Source Apportionment or Receptor Modelers use emission profiles in an effort to determine the source of emissions collected at ambient monitoring sites. This is done by analysis of certain "fingerprint" species and speciated emission profiles in models such as

EPA's Chemical Mass Balance (CMB). The knowledge that only certain types of sources emit particular chemicals ("fingerprints") can assist in apportioning or tracing ambient samples to the sources that emitted them. These modelers primarily use the SPECIATE PM module, but are also interested in TOC profiles for analyzing emissions where the difference between gas and particle chemistry become intermingled.

To a Receptor Modeler, one of the most important aspects of an emission profile is the associated documentation. It is important to know what method was used to determine the elemental and organic carbon fractions, what species were tested for, the mass fraction of the sample that was ultimately speciated, exactly what type of source was tested, the specific type of emission testing procedure conducted, information on normalization and uncertainty, and any other specific information available. Receptor Modeling requires this specific knowledge because the modeler may wish to use several similar emission profiles with subtle source type differences in a battery of model runs before the model converges to an acceptable degree.

Receptor models require profiles that express the speciated chemical abundances in terms of the mass fraction of the chemical in the total emissions stream. The emission stream may be parsed into multiple size fractions. However, if the emission stream is parsed into size fractions, each specie in the profile should be reported as a mass fraction of each respective size fraction, and the total mass of each size fraction collected should also be reported. If any type of normalization is factored into the profile before it is included in a published reference, it is important that this be expressed in the profile literature for review by the Receptor Modeler so that the information can be included in SPECIATE and used by the Receptor Modeler. Information concerning whether the profile is based on a measurement of total mass emission during the source profile measurement or a form of reconstructed mass is important information associated with the notion of profile normalization.

Profile uncertainties associated with each specie in an emission profile are essential to the Receptor Modeler. Additionally, the method used to calculate the uncertainty is necessary. There are two types of uncertainties that contribute to the overall uncertainty of emission profiles: (1) Statistical uncertainty resulting from multiple source tests, and (2) Instrument calibration or test measurement uncertainty, which is difficulty to quantify. There is no known standard operating procedure associated with emission profile uncertainty calculation, so it is imperative that information about the method used to calculate the uncertainty of each profile is made available to the Receptor Modeler.

Addressing Data and Program Requirements

PES believes supplying both Air Quality Modelers and Source Apportionment/Receptor Modelers with the data required can be accomplished by providing profile data in its most basic format, together with tools for processing the data into the format necessary for the application. These tools consist of the inclusion of all available information with the profile so the user can determine its relevance to the

modeling application, and addition of some processing capability to the SPECIATE program to convert profile data into the desired format.

The current SPECIATE (v. 3.2) available on EPA's CHIEF website supplies profile information to users via separate TOC and PM modules. As described earlier in this paper, Air Quality Modelers primarily use the TOC module and Source Apportionment/Receptor Modelers primarily use the PM module. The differences in the data format and content requirements of these groups and discussion amongst interested parties at the AAAR conference expert panel meeting points to the necessity for separation of the current SPECIATE into two separate tools. The efforts to serve Air Quality Modelers and Source Apportionment/Receptor Modelers using one tool (SPECIATE), has resulted in difficulties for both groups. It was suggested at the AAAR conference expert panel meeting that SPECIATE should be re-missioned to support the needs of the Source Apportionment/Receptor Modeling group, and that a completely different product be designed to support the needs of Air Quality Modelers. This suggestion was met with support at the expert panel meeting. With this in mind, the remainder of this paper will address recommendations that PES presented to representatives from both modeling groups, and expert panel reactions.

The following recommendations for SPECIATE are made in consideration of the needs of both the Air Quality and Receptor Modelers:

(1) Update SPECIATE with new PM and TOC profiles and delete out-dated profiles.
(2) Expand existing profile reference information.
(3) Do not normalize profiles for inclusion in SPECIATE.
(4) Add profile-to-SCC assignment list to SPECIATE.
(5) Develop an automated process to create model inputs for each model chemical mechanism.
(6) Maintain the list of SAROAD codes, and include SAROAD assignments in SPECIATE.
(7) Convert SPECIATE to a web-based system.
(8) Revise the SPECIATE PM module to be more flexible regarding size ranges.
(9) Add a mechanism for users to add their profiles into SPECIATE.
(10) Create a routine for profile comparison.
(11) Expand search capability to any field.
(12) Make profile technical papers available for review.

Each of these recommendations is discussed in detail below.

SPECIATE Data Update

A total of 178 journal articles were collected by PES and reviewed to determine their relevance to SPECIATE. Most of the articles collected were published between 1990 and 2002 in peer-reviewed journals. PES prepared a reference list of these 178

articles. The reference list is accompanied by a one-page review/summary of each article that explains the type of profile data contained in the reference and any specific information of interest. Since the focus of the data collection task for the previous work assignment was to locate and obtain PM profile data, most of the data collected were for PM sources. The one-page review/summaries consist of the following:

- reference citation;
- emission category (PM including size fractions, TOC);
- description of profile data available in the journal article;
- for PM profiles, whether or not the article distinguishes between elemental and organic carbon;
- description of uncertainty information (if available);
- and any additional information that appears relevant, or is specific to the testing procedure conducted.

A review of the 178 journal articles yielded about 100 that PES believes are suitable for addition to SPECIATE. Many articles were not found to be suitable for SPECIATE because they only reported results of ambient monitoring. Only source-specific data is considered relevant to the SPECIATE data model.

Expand Existing Profiles

The expert panel discussed the importance of retaining existing SPECIATE data, and what steps could be taken to make the existing data more useful. The issue at hand concerning the older profiles in SPECIATE is whether the data are relevant to emission sources today. Much of the data in the older SPECIATE profiles are the result of testing and/or studies that were conducted in the 1980's, and in some cases the 1970's. If it is determined that the older profiles are relevant to emission sources today, then existing profile reference and notes information should be expanded to satisfy all requirements of the new SPECIATE data model. The panel agreed that there is no reason to delete old profiles from the existing data since this data may retain some value, but it was suggested that the year of the profile sample should be added to each profile record so the user can readily determine its value on a case-by-case basis.

Normalization

In order to supply the information necessary for both types of SPECIATE users, the speciated profiles must be in the most elemental format available, (i.e. no normalization of profiles or grouping of chemicals under chemical types such as "C-10 COMPOUNDS".) In cases where the profile literature supplies the profile in a normalized format, research, phone calls, or other contacts should be made to try to find the original test data to include in SPECIATE. Although this "basic data" approach would cause problems for Air Quality Modelers, presentation of the basic data and all the associated documentation in SPECIATE is the most logical format for Source Apportionment/Receptor Modelers. Presentation of normalized profiles in addition to

raw profile data in a separate system would be one possible solution for satisfying the data needs of Air Quality Modelers.

Profile-to-SCC Assignments

Although the individual profiles available in SPECIATE should be in their most elemental format, Air Quality Modelers have a need for a simple profile-to-SCC assignment file. Traditionally this file has been maintained separately from SPECIATE. Maintenance and update of this file should be part of the separate tool that has been suggested to support Air Quality Modelers. The profile-to-SCC assignment file should contain a one-to-one assignment of profiles to SCCs based on the EPA's recommended assignment. Once the file is saved, it could be opened and edited as necessary by the user to make changes or substitutions to the file where necessary.

Model Inputs for Individual Chemical Mechanisms

The use of various chemical mechanisms in Air Quality Modeling requires that emission profiles such as the components of TOC profiles in SPECIATE be categorized by their chemical characteristics prior to model input. The mechanism used depends on the type of emission model to be used. This categorization step is not currently addressed by SPECIATE. PES has discovered through meetings with EPA, NCDENR, and the SPECIATE expert panel that this necessary intermediate step between emission profiles and model inputs has never been completely standardized by EPA or any other agency. PES believes that a tool developed specifically for Air Quality Modeling purposes would provide the medium necessary for housing and distribution of the data and procedures necessary to bridge the gap between emission profiles and emission models.

SAROAD Code Assignments

As explained earlier, SAROAD codes are necessary for the assignment of emission profiles to the appropriate mechanisms for the purposes of Air Quality Modeling. There are constant changes in the interpretation of chemical mechanisms as research expands the understanding of complex chemical mechanisms present in atmospheric chemistry. The SAROAD code system is a mechanism for grouping chemicals with similar atmospheric reactivity characteristics. As the knowledge of atmospheric chemistry expands, it will become increasingly possible to model specific chemicals, and therefore, grouping of like chemicals will become less necessary. This trend, however, will be very gradual, and SAROAD code based chemical mapping is not going to become obsolete in the foreseeable future. In recent years, SAROAD numbers have been indiscriminately defined by many researchers with no one organization ensuring their uniqueness and accuracy. Thus, today there are many redundant and unresolved SAROAD numbers. Dr. William Carter at UC-Riverside has reviewed the SAROAD numbers and reportedly resolved duplicate numbers[1]. One possibility for resolution of the SAROAD numbers would be for EPA to conduct an independent quality assurance review of Dr. Carter's SAROAD database. After this formal review of the SAROAD codes, the list of codes could be published on the EPA's website, and be

incorporated in the tool to be developed for Air Quality Modeling. This list of SAROAD codes should be centrally maintained because of its critical role in ozone modeling.

Chemical Abstract Service (CAS) numbers are currently assigned to specific chemical species in SPECIATE, but grouped chemical species and mixtures which are common to emission profiles often cannot be characterized by CAS numbers. The system of SAROAD codes (when it was maintained) historically recognized and systematically assigned codes to these groups. The CAS numbering system should be maintained in the SPECIATE system as it is developed for Source Apportionment/Receptor Modelers so that individual chemical species with more than one name can be easily identified as the same chemical. Retaining the CAS numbers in SPECIATE does not present any apparent maintenance problems, and would be beneficial for chemical cataloging.

Web-based SPECIATE

SPECIATE is now distributed to the user community via a download and install on individual PC's. Since it is visioned as a central repository of speciated profile data, it has been suggested that a web-based system may better serve the needs of the user community. This would allow for central control of the data, but also allow users to extract the most current data from the database to develop input files for various modeling scenarios. Some of the benefits that could be seen by changing to a web-based platform include the following:

- *An Intuitive Interface*
  Typically, a web system is a more intuitive interface for users to understand. Users have become familiar with basic Internet controls and how they are used. A web system also provides unlimited scrollable screen real estate, which allows information to be spread out in a more understandable way. Links can be provided to allow additional information to be available as needed, but not always displayed on the screen. This ease of use generally translates into less training.

- *Updates and Revisions*
  A web-system allows for simple distribution of regular updates and revisions to data or the interface. Changes can be completed easily in one centralized system, and be seen and used immediately by users.

- *Automatic System Emails*
  Users can be given an on-line option to 'subscribe' and be notified of system changes. System email can be utilized in a powerful and useful way to facilitate communication with users.

- *Easy Access*
  A web-based system would provide a platform for users in geographically diverse areas to easily access the system.

- *User Maintenance Requirements Minimized*
  PC applications require each user to have sufficient system resources. A web system would be designed as a 'thin client' application requiring minimal local resources. The only software required by the user is a web browser.

- *Automatic Error Trapping*
  System error details would be automatically logged and emailed to the system administrator and programmer when they occur. This allows the programmer to respond to problems as soon as possible, and removes the responsibility for the system administrator to receive error information and notify the programmer.

The idea of converting SPECIATE to a web-based system was suggested to the expert panel at the AAAR conference. There was no clear support by the expert panel for or against this suggestion, although the expert panel showed support for some of the attributes that would be inherent in a web-based system. The expert panel was particularly interested in the idea of more frequent updates and revisions to the SPECIATE data.

PM Module Database Structure

The SPECIATE PM module needs to be less rigid so that it can accommodate emission profile data in the different formats supplied in scientific literature. As discussed earlier, the current SPECIATE model only has housed data for 4 specific size fractions of PM particles. The size range of interest to the receptor modeler depends on the source of the emissions. Additionally, as further research is conducted into the health effects of particles of different size ranges, the trend is showing that the smaller particles such as those less than 2.5μm may cause significant health problems. Today, much of the emission research being conducted includes particle size ranges of 1μm, or less. Concentrations of various pollutants in PM profiles vary widely depending on the size range of the PM being measured. A pollutant that makes up only 5% of the 0 to 2.5μm size range may actually represent 80% of the 0 to 1μm size range. Thus, it is important to make SPECIATE more flexible regarding the size fraction of PM maintained in the database.

Data Import

Two methods should be considered for the input of outside data to SPECIATE. The simplest would be for individual users to import data into their own copy of SPECIATE for personal use. (This option would not be possible with a web-based version of SPECIATE). This would require no QA since the user would be responsible for its accuracy. The expert panel addressed the possibility of adding this capability to SPECIATE, but there was little support for this suggestion. In response to the suggestion of this capability, some members of the expert panel felt that it may impede the development of the maintained SPECIATE program, since profile developers may be less likely to submit useful data to the EPA, and just maintain their own copy of the program.

The second method is for data to be loaded into a common SPECIATE database that would be available to the public, either through the Internet or multiple downloads and/or updates. In this case, any new data would have to be quality assured prior to inclusion in the database. The simplest test would be if the data were published, it would qualify for inclusion. The most rigorous would be to have it reviewed by the advisory panel prior to inclusion.

Profile Comparison

SPECIATE users have identified a need for a tool that could be used to compare similar profiles in SPECIATE. As more profiles are added to SPECIATE, it will become increasingly important to provide SPECIATE users with a tool for profile comparison. It may be necessary to compare two or more PM or TOC profiles against each other, or to compare two or more PM size fractions from the same or different testing protocols against each other. A tool could be developed and added to SPECIATE for the analysis of the similarities and differences of profiles. A visual tool capable of producing graphical and tabular comparisons would be the most useful.

Search Capability

Currently, SPECIATE allows users to run queries to select profiles based on profile name, number, component chemical species, or keywords. The limited size of the current database can be efficiently searched using this capability. PES expects that a significant number of new profiles will be added to SPECIATE resulting from the current and continued research effort. The additional profiles will increase the size of the database enough that it may become difficult to locate all profiles relevant to a particular source. Alternatively, the user may wish to locate only profiles that used a certain test method for a particular source, and find it difficult to pare down the results using the current query tool. The expert panel discussed search capability options and added that the year of the actual test resulting in each profile would be a useful search capability.

Availability of Research Papers

PES investigated the possibility of adding PDF versions of the profile journal articles to SPECIATE to make access to all relevant information easy for the SPECIATE user. Due to copyright restrictions, costs are prohibitive to distribute any form of the journal articles with SPECIATE.

SUMMARY

The issues and recommendations presented in this paper are considered to be a work in progress. In many cases, the interviews that PES conducted with EPA and NCDENR personnel and discussions at the expert panel meeting yielded conflicting or inconclusive opinions. At the same time, these conflicts can be resolved by further expert panel review and discussion. After the expert panel meeting on October 10, 2002, PES distributed the minutes from the meeting, which also included a list of outstanding issues

that have not yet been addressed or fully resolved by the expert panel. A list of some of the key outstanding issues is included below.

- Additional information is needed to make a decision regarding the mechanism(s) for users to add profiles to SPECIATE.
- Specific criteria and methodology should be agreed upon to review and/or screen profiles for applicability.
- Details regarding possible data normalization schemes are necessary.
- Information and documentation on profile uncertainty is important to emission modelers. Should profiles with no associated uncertainty be removed altogether, or flagged somehow for filtering within SPECIATE? Additionally, none of the profiles in the current SPECIATE TOC module have uncertainties associated with them.
- What should be done in the case of existing PM profiles, where the values for at least one of the size ranges has merely been copied from another size range that was actually measured?

There is not a current work assignment underway to implement the changes to SPECIATE as discussed in this paper; however, PES is currently working on an EPA project to develop improved emission inventory split factors for PM2.5 Air Quality Modeling. This effort involves evaluation of PM2.5 profile data from a recent literature search. The results of the current effort will serve a dual purpose, as they will improve current split factors, and much of the resulting data will be directly applicable to SPECIATE.

REFERENCES

1) Carter, W. 2002. University of California Riverside, *email communication with Greg Pagett of PES*.

KEYWORDS

SPECIATE
Source Apportionment
Receptor Modeling
Air Quality Modeling
Emission Profile
PM
TOC

ACKNOWLEDGEMENTS

PES would like to recognize the contributions of the SPECIATE expert panel to the SPECIATE improvement effort. The members of the panel are as follows: Judy Chow (DRI), Tom Coulter (AQMD & ORD), Mark Janssen (LADCO), Donna Kenski (LADCO), Chuck Lewis (ORD), Joellen Lewtas (EPA Region 10), Tom Pace (EFIG), Ron Ryan (EFIG), James Schauer (Univ. of WI), Paul Shapiro (ORD), Dean Smith (ORD), and John Watson (DRI).