SPECIATE - EPA's Database of Speciated Emission Profiles

Paper #229

AUTHOR

Lee L. Beck, Senior Project Engineer

U.S. Environmental Protection Agency (EPA), MD-E343-02, Air Pollution Prevention and Control Division, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711

ABSTRACT

SPECIATE is EPA's repository of Total Organic Compound (TOC) and Particulate Matter (PM) speciated profiles for a wide variety of sources. The profiles in this system are provided for air quality dispersion modeling and as a library for source-receptor and source apportionment type models. This recent initiative to update SPECIATE is needed because speciated emissions profiles continue to be developed and the data in the existing EPA database (SPECIATE 3.2) is becoming outdated.

SPECIATE v3.2 was posted to EPA's CHIEF web site in November 2002 with 2068 emission profiles. No new profiles had been added to SPECIATE until recently. SPECIATE v4.0 is now available. It contains more than 4300 source profiles. SPECIATE 4.0 currently exists as a Microsoft Access relational database but may be committed to stand-alone software with the development of a "front end." There also has been work to develop composite profiles and to map profiles to source classification codes (SCC). Other work related to EPA's SPECIATE database has been updated or new work has begun. This paper presents details about EPA's SPECIATE database and ancillary work being done through a collaboration involving EPA's Office of Research and Development (ORD) and the Office of Air Quality Planning and Standards (OAQPS) at Research Triangle Park, NC.

INTRODUCTION

Emissions from sources of air pollution have been sampled for several decades with the objective of determining "source strength," which is a term given to express the amount of pollution getting into the atmosphere. Source strength was usually measured at the source of emissions (e.g., industrial stacks or automobile exhaust pipes). Environmental scientists also study the composition of pollution in ambient air. The ambient concentrations are of interest because this is what we breathe. Ambient measurements also can provide clues as to where the emissions originated (the sources).

Initially, only the total mass of emissions was of interest. As the science of health and other environmental concerns progressed, the focus progressed from total mass to the physical and

chemical characteristics of the pollutants. Physical size of the particles in the air that we breathe is important. Emissions of particles (PM) that are small have a greater adverse impact on human health because they are inhaled and deeply trapped in the lungs.[1] Chemical composition of these particles is also of concern because it has been determined that some chemicals are more toxic than others.

Determining emission characteristics (e.g., different size categories of particles) is not a new idea. The focus on particulate matter (PM) changed from total mass to PM with aerodynamic diameters of less than 10 microns (PM10) several decades ago and more recently to PM2.5 (< 2.5 microns). The determination of chemical composition of emission streams has been evolving also. Certain metals and other toxic substances are routinely analyzed and reported because of their adverse effects on human health.

Finally, it has been determined that much of the atmospheric PM2.5 is a result of condensation into the particle phase after the emissions leave the source. Since ambient pollutant concentrations of are of primary concern, specialized equipment has been developed in an attempt to mimic what happens after the hot gases experience the cooling effects of the ambient air. One such device is called a dilution tunnel or "dilution sampler" system (DSS). The U.S. Environmental Protection Agency (EPA) has provided some enhancements such as electronic controls and a gas conditioning system to a DSS that was initially developed by the California Institute of Technology. [2] A major objective of the DSS is to separate out condensed PM from ash or crustal (primarily non-aerosol) PM. The gases are also collected by the DSS and reported as VOC (volatile organic compounds) or TOG (total organic gases). The term VOC is used to report the fraction of gaseous material collected that may contribute to the formation of tropospheric ozone and photochemical smog. TOC (total organic compound) and TOG are other ways of reporting emissions of organics to the atmosphere without regard to their potential for reacting to form ozone or smog.

As the dilution sampling of emissions and characterization continues to evolve there is a need to systematically categorize and report these large amounts of data. Today's widespread use of personal computers and database software has presented an ideal environment to develop such a system. EPA's database of speciated emission profiles is called SPECIATE (not an acronym but typically presented as all capital letters). EPA developed its first version of the SPECIATE database in 1999 and released it with a user-friendly front-end in 2002. [3] EPA's SPECIATE database has now been updated to address the large amount of speciated emission profiles generated since the release of SPECIATE 3.2. The new database, called SPECIATE 4.0 is a Microsoft Access database which can be made amenable to searches and queries, and the tabulated data is easily exported to Microsoft Excel for those not familiar with the Access software.

DEVELOPMENT OF SPECIATE 4.0

Several parallel activities took place concurrently as we began development of the SPECIATE database: formation of a workgroup for project management, identification of Access fields and functions, acquisition of published and unpublished data (literature search) and development of a Quality Assurance Project Plan (QAPP).

SPECIATE Work Group

There is wide spread interest in SPECIATE because of its usefulness in various research and regulatory groups. To accommodate these diverse interests a workgroup was formed. The workgroup comprises 40-50 EPA personnel primarily from the Office of Air Quality Planning and Standards (OAQPS) and the Office of Research and Development (ORD). Workgroup personnel are predominantly scientists, engineers, and other environmental professionals who are involved with research or the development of emission inventories. Personnel in the private sector (e.g., contractors) are present at the meetings on an as-needed and appropriate basis to discuss their areas of contribution to the effort.

The research component contains modelers who are looking to SPECIATE for inputs to models that can evaluate the impact of certain species of compounds on ozone formation (e.g., reactivity classes of VOC). Other modelers use source-receptor modeling to attempt to identify the source of emissions measured in the ambient air (e.g., by using chemical "markers"). Other workgroup members have responsibilities in areas such as emission inventory development or policy. Still others are experts that have the ability to discuss technical issues such as sampling and analytical methodology (and attendant limitations).

The workgroup meets on a recurring basis, usually monthly, and on an as-needed basis to discuss issues and make recommendations. Individuals who are unable to attend are given the opportunity to review work products and comment through frequent email exchanges in areas of interest and expertise. Individuals not in the commuting area (i.e., Research Triangle Park, NC) are patched in by telephone.

Identification of Access Fields

One of the first functions in the planning of SPECIATE 4.0 was to develop a database design. The design accommodated the desired modifications and additions to the SPECIATE functionality. The modifications were based on the suggestions from the October 2002 meeting of the SPECIATE Expert Panel held at the American Association for Aerosol Research conference in Charlotte, NC [4], as well as additional recommendations provided by the SPECIATE Workgroup. The major elements in the design appear in 8 tables:

- The PM_PROFILE table includes, but is not limited to, profile number, name, notes

on the profile, and descriptive information about the profile such as sum of species, test method, and normalization basis. Also included in this table are the data quality ratings discussed later in this paper.

- PM_SPECIE table includes the specie identification number, the profile number associated with the specie, the percentage of the specie in the profile, the uncertainty associated with the percentage value, the method used to determine uncertainty, and a description of the analysis method used to determine the species percentage in the profile.

- The REFERENCE table includes information that characterizes the reference documents associated with the profiles, including whether or not a particular reference is the primary reference or not (thus allowing multiple and unlimited references for any profile) and a description of the documentation. The reference document itself is in some instances stored in the REFERENCE table.

- The GAS_PROFILE table includes, but is not limited to, profile number, name, notes on the profile, and descriptive information about the profile such as, sum of species, test method, and normalization basis. Also included in this table are the data quality ratings discussed later in this paper.

- The GAS_SPECIE table includes the specie identification number, the profile number associated with the specie the percentage of the specie in the profile, the uncertainty associated with the percentage value, the method used to determine uncertainty, and a description of the analysis method used to determine the species percentage in the profile.

- The KEYWORD table includes descriptive keywords of profiles. This information can be used in keyword-based searches for profiles.

- The SPECIE_PROPERTIES table includes the identifying numbers associated with the compounds that are species in the database, as well as other characteristic information such as molecular weight.

- The MNEMONIC table includes abbreviated profile names used, for instance, in chemical mass balance (CMB) receptor models.

A description of the elements in the tables is presented in Tables 1-8.[5]

Table 1. PM_PROFILE TABLE

Field	Description
P_NUMBER	PM Profile Number (Primary Key)
NAME	PM Profile Name
QUALITY	Quality rating (A-E) of the profile (related to the products of the J, V and D ratings, see "Development of Data Quality Indicators" for explanation)
CONTROLS	Emission Controls Description
P_DATE	Date profile added
NOTES	Notes
TOTAL	Sum of species percentages for a given profile, excluding organic species in PM and inorganic gases in PM profiles.
MASTER_POL	Indicates the pollutant to be used in calculation. Allowed value: 'PM' In the future, other values may be allowed (e.g., PM_PRI, PM_FIL, PM_CON)
T_METHOD	Description of sampling method
NORM_BASIS	Description of how profile was normalized
ORIG_COMPO	Specifies whether the profile is original or composite. Allowed values: 'C','O'
STANDARD	Indicates whether the profile is provided by EPA SPECIATE (standard) or user-added. The database is constructed to allow users to add profiles.
ORGANICS	Indicates whether or not the profile provides speciated organics in PM
INCL_GAS	Indicates whether or not the profile includes inorganic gas species (e.g., sulfur dioxide, hydrogen sulfide, oxides of nitrogen, etc.)
TEST_YEAR	Indicates year testing was conducted
J_RATING	Objective expert judgment rating based on general merit (see "Development of Data Quality Indicators" for explanation)
V_RATING	Vintage based on TEST_YEAR field (see "Development of Data Quality Indicators" for explanation)
D_RATING	Data quality rating based on number of observations, robustness (see "Development of Data Quality Indicators" for explanation)
REGION	Geographic region of testing
LOWER_SIZE	Identifies lower end of aerodynamic diameter particle size, micrometers
UPPER_SIZE	Identifies upper end of aerodynamic diameter particle size, micrometers
SIBLING	GAS Profile number; samples taken from the same source and study, if available.
OLD_PROFILES	Was the profile taken from SPECIATE 3.2?
SIMPLIFIED	Is the profile a PM Simplified Profile?

Table 2. PM_SPECIE TABLE

Field	Description
ID	Unique Identifier (Primary Key)
SPECIE_ID	Specie Identifier (The same as ID in SPECIE_PROPERTIES)
P_NUMBER	PM Profile number (Link to PM_Profile Table)
WEIGHT_PER	Weight percent of pollutant (%)
UNCERTAINT	Uncertainty percent of pollutant
UNC_METHOD	Description of method used to calculate uncertainty
ANLYMETHOD	Description of analytical method (e.g., X-ray fluorescence spectroscopy, ion chromatography, etc.)

Table 3. REFERENCE TABLE

Field	Description
ID	Unique Identifier
P_TYPE	Indicates PM or GAS. Allowed values: P (PM), G (Gas)
P_NUMBER	Profile number (Link to PM_PROFILE and GAS_PROFILE tables)
DATA_ORIGN	Source of data (e.g., EPA APPCD, Schauer, ARB, DRI, literature, etc.)
PRIMARY	Designates a reference as primary. When a profile is based on multiple references, this field allows one reference to be tagged as the primary reference.
DESCRIPTIO	Stores the descriptive information about the profile.
DOCUMENT	Complete reference citation.

Table 4. GAS_PROFILE TABLE

Field	Description
P_NUMBER	GAS Profile Number (Primary Key)
NAME	GAS Profile Name
QUALITY	Quality rating (A-E) of the profile (related to the products of the J, V and D ratings, see "Development of Data Quality Indicators" for explanation)
CONTROLS	Emission Controls description
P_DATE	Date profile added
NOTES	Notes
TOTAL	Sum of organic species percentages for a given profile
MASTER_POL	Indicates the pollutant to be used in calculation. Allowed values: 'VOC', 'TOG'. When methane was not measured in a study, ethane, acetone and other non-VOCs are removed from the profile and it is defined as a VOC profile.
T_METHOD	Description of sampling method
NORM_BASIS	Description of how profile was normalized
ORIG_COMPO	Specifies whether the profile is original or composite. Allowed values: 'C','O'
STANDARD	Indicates whether the profile is provided by EPA SPECIATE (standard) or user-added. The database is constructed to allow users to add profiles.
TEST_YEAR	Indicates year testing was conducted
J_RATING	Objective expert judgment rating based on general merit (see "Development of Data Quality Indicators" for explanation)
V_RATING	Vintage based on TEST_YEAR field (see "Development of Data Quality Indicators" for
D_RATING	Data quality rating based on number of observations, robustness (see "Development of Data Quality Indicators" for explanation)
REGION	Geographic region of testing

Table 5. GAS_SPECIE TABLE

Field	Description
ID	Unique Identifier (Primary Key)
SPECIE_ID	Species Identifier (Must be the same as ID in SPECIE_PROPERTIES)
P_NUMBER	GAS Profile Number (Link to GAS_PROFILE table)
PERCENT	Weight percent of pollutant (%)
UNCERTAINT	Uncertainty percent of pollutant
UNC_METHOD	Description of method used to calculate uncertainty
ANLYMETHOD	Description of Analytical method (e.g., GC/FID, GC/MS, HPLC/UV, etc.)

Table 6. KEYWORD TABLE

Field	Description
ID	Unique Identifier (Primary Key)
P_TYPE	Indicates PM or GAS. Allowed values: P, G
P_NUMBER	Profile Number (Link to PM_PROFILE and GAS_PROFILE Tables)
KEYWORD	Keyword describing profile

Table 7. SPECIE_PROPERTIES TABLE

Field	Description
ID	Unique Identifier (Primary Key) (Link to PM_SPECIES and GAS_SPECIES tables)
CAS	Chemical Abstract Service number assigned to pollutant (with hyphens) (blank if no CAS)
EPA_ID	EPA Chemical Identifier; to be provided by EPA Substance Registry System for species without CAS numbers.
SAROAD	Storage and Retrieval of Aerometric Data (SAROAD) code.
PAMS	Is PAMS pollutant? (Yes or No)
HAPS	Is Hazardous Air Pollutant? (Yes or No)
NAME	Pollutant name
SYMBOL	Standard chemical abbreviation (provided by Eric Fujita, DRI)
SPEC_MW	Species molecular weight
NONVOCTOG	Is the species a non-volatile organic gas?
EPAITN	EPA Internal Tracking Number
SPECIATETEMP	SPECIATE Temporary ID

Table 8. MNEMONIC TABLE

Field	Description
ID	Unique Identifier (Primary Key)
P_TYPE	Indicates PM or GAS. Allowed values: P (PM), G (Gas)
P_NUMBER	Profile number (Link to PM_PROFILE and GAS_PROFILE tables)
DRI_PNUMBR	DRI profile number (Original DRI profile numbers)
MNEMONIC	Alphanumeric Code unique to each profile. Used in CMB input files.

The Microsoft Access database is used at this time only as a repository for the data. However, the reason for selection of the Access platform as the repository is so that relationships can be assigned for database queries. After the database is fully functional and adequate peer review is given to the data, consideration will be given to developing a graphical user interface so that users can easily parse out the data that they need and export it to their application or to spreadsheets.

Database Population Methodology

The literature search identified hundreds of potentially significant articles that contained some amount of speciated emissions data. A spreadsheet was developed as a tool for management of the findings. Items in the spreadsheet included source (e.g., publication), authors, abstract or completed work, data format (databases, spreadsheets, word processing files), source type (open burning, coal combustion, diesel automobiles, etc.), PM or gas, number of profiles, internal

reviewer (several reviewers were needed), recommendation (include or not), as well as a complete reference to allow retrieval of the source. The spreadsheet was continually refined and a column was included to indicate priority and estimated level of effort needed to incorporate the data into SPECIATE. There also was a notes field to allow brief comments. The prioritization for data inclusion was reviewed by the Workgroup and discussed at the meetings. Ultimately, a list was developed based on data quality, perceived need for the data by the user community, ease of incorporation (e.g., some data were already formatted in a way that easily allowed electronic incorporation), and available time and resources.

Some sources of extensive data allowed for easy incorporation of profiles. The most significant of these was the extensive work on speciation already performed by the Desert Research Institute (DRI). Other sources of large amounts of data were the Texas Commission on Environmental Quality, [6] the California Air Resources Board, [7] universities specializing in data speciation (e.g., the work of Jamie Schauer at the University of Wisconsin) and of course the research being performed at EPA.

Extent of Data Included in SPECIATE 4.0

SPECIATE 3.2 contained 1503 PM and 565 gas profiles. SPECIATE 4.0 added over 1200 new PM and more than 500 gas profiles. Additionally, 2475 simplified profiles have been added. The simplified profiles are for the same sources as the complex profiles but address a subset of the data that is the only fraction of concern to many modelers, thereby simplifying data acquisition.

Other Information Developed as Part of the SPECIATE Update

In addition to adding additional profiles to SPECIATE, there was a need to address several related issues:

Development of data quality indicators

When the Workgroup began planning for the SPECIATE update it became apparent that some methodology was needed to give at least a qualitative indication of the perception of data quality. This was because some of the data appeared in refereed journal articles that had undergone extensive peer review and other data was obtained from sources where the quality control was not obvious. There was a temptation to include only peer-reviewed data, but this would limit the amount of data considered and consequently the usefulness of the database. The decision was made to include all data and accompany the data with quality indicators and the reference so that the user would have information on which to base a judgment on the appropriateness for his purposes. The procedure adopted is as follows [8]:

- *V-rating (profile vintage)* is based on the vintage of the profile which reflects measurement technology and methodology. Inclusion of the V-rating is based on an assumption that more recent data are based on newer and presumably more accurate

measurement technology. For profiles before year 1980 – score = 1, 1980-1990 score = 2, 1991-1995 score = 3, 1996-2000 score = 4 and after Year 2000 score = 5. These data are housed in the V_RATING field in the PM and gas profile tables.

- *D-rating (number of samples)* is given a "1" (poor) to "4" (excellent) rating. The underlying assumption for the D-rating is that it allows an indication of precision of the methodology and the consistency or the source measured. Larger sample sizes also allow for more comprehensive statistical evaluation of the data. This category is rated based on the number of samples: # of samples > 10 score = 4; 5-9 samples score = 3; 3-4 and composite samples score = 2; 1-2 or unknown # of samples score = 1. Composite samples can be defined in at least two ways. One is where there is a mixing of individual samples prior to analysis. Another is when one number is given to represent a source where more than one sample was taken but individual results are not presented. These data are housed in the D_RATING field in the PM and gas profile tables.

- *Final Score* = (V-rating) x (D-rating). This rating is found in the PM and/or gas profile tables under the field names QUALITY, V_RATING, and D_RATING, respectively. As discussed above, SPECIATE 3.2 profiles will not have entries for V_RATING or D_RATING (or J_RATING shown below), however, they retain their legacy quality rating expressed numerically (5 = A, 1=E). This convention readily distinguishes any numeric value in the QUALITY field as a hold-over from SPECIATE 3.2. EPA SPECIATE 3.2 Documentation does not identify how the quality ratings were selected

Profile quality is rated from A (excellent) to E (poor) as shown in Table 9.

Table 9. Overall Profile Quality Ratings

Profile Quality	Final Score Ranges
A	17-20
B	13-16
C	9-12
D	5-8
E	<5

- *J-rating (expert judgment)* is given a "1" (poor) to "5" (excellent) rating. This value is based on the information underlying each profile including, but not limited to:
  - Profile composition;
  - Relative ratios of species within the profile;
  - Sum of the speciated mass fractions; and
  - Supporting documentation.

Development of composite profiles

When discussing SPECIATE within the modeling subgroup of the Workgroup, it became apparent that there is a need for a single number for use in model algorithms rather than to provide many different profiles which are present for some major categories. The development of the single number was termed "compositing" and the numbers were entered into the database as "composites." The categories chosen for compositing are:

- Agricultural Burning
- Agricultural Soil
- Construction Dust
- Diesel PM
- Forest Fires and Prescribed Burns
- Gasoline Exhaust PM
- Paved Road Dust
- Quarrying Dust
- Residential Coal
- Residential Wood Combustion
- Unpaved Road Dust
- Utility & Industrial Boilers

Each of the categories chosen for compositing was selected because there was a significant amount of data reported for the categories. In some cases the variability in the data reported was also large. For these reasons, the data sets were studied and a rationale was determined for arriving at one number that would represent each category. Written descriptions were provided for each rationale chosen. The composite number gives the modeler a choice to either use the composite number or to use number(s) of his choosing from the SPECIATE database.

Development of PM simplified profiles

PM-simplified profiles are those based on full PM profiles collapsed to five species: Elemental Carbon, Organic Carbon, Sulfate, Nitrate, and PMFine (sum of all other species). PM-simplified profiles are used by specific models using the simplified, five-species approach. Simplified PM profiles were prepared for the new SPECIATE profiles, and were also developed for the SPECIATE 3.2 profiles, subject to certain restrictions. For the new composite profiles, simplified profiles were prepared after the composite profiles were introduced into the database. This additional field in SPECIATE 4.0 database will help modelers using the five-species approach to use the data without additional analysis and recalculation. SPECIATE 4.0 reports 2,475 simplified PM profiles.

Mapping of profiles to SCCs

Another item typically needed by modelers is a way to link the profile to the source. This is typically done by using a source category classification or "SCC" index.

Resources were not available to map every profile to an appropriate SCC. However, we were able to link speciation profiles covering the SCCs that account for 80% of the 2002 National Emission Inventory (NEI) emissions. To do this, we prioritized the SCC by VOC and PM2.5 emissions and identified 146 and 20 SCCs, respectively.

CONCLUSION

EPA has provided the first update of its SPECIATE database in more than 6 years. The number of speciated profiles has more than doubled. In addition to the increase in the number of profiles added, there also is a new database structure which enhances the number of criteria by which each profile can be characterized. Data quality indicators have been added to assist the user in data selection, and simplified PM profiles have been assembled to help modelers who are interested in only a subset of the data. It is believed that the more recent data, presumably gathered using more recent measurement technology and reported using data quality indicators, is an improvement over the older data presented in SPECIATE 3.2.

ACKNOWLEDGEMENTS

Much of the work done in the development and population of the SPECIATE 4.0 database was performed by personnel of E.H. Pechan and Associates. The principal investigator for Pechan was Mr. David Holoman. This work was built on the original database of SPECIATE 3.2, which was developed under the direction of Mr. Ronald Ryan, Emissions, Monitoring, and Analysis Division, Office of Air Quality Planning and Standards, EPA.

REFERENCES

1. Monitoring for Fine Particulate Matter. Eiseman, Elisa. Rand Corporation. Santa Monica, California, Report No. MR-974-OSTP. http://www.rand.org/pubs/monograph_reports/MR974/MR974.Ch1.html (accessed February 2006)
2. Hildemann, L.M., Cass, G.R., and Markowski, G.R. (1989) A Dilution Stack sampler for Collection of Organic Aerosol Emissions: Design, Characterization, and Field Tests, *Aerosol Sci. Technol*., **10**, 193-204.
3. United States Environmental Protection Agency. http://www.epa.gov/ttn/chief/software/speciate/ (accessed January 2006).
4. Meeting Highlights. Speciate Expert Panel Kickoff Meeting. American Association for Aerosol Research. October 10, 2002
5. SPECIATE 4.0 Speciation Profile Development Documentation Draft Final Report, Holoman, D.W., Hsu, Y., and Roe, S., E.H. Pechan & Associates, Durham, NC. 2005
6. Cantu, G. Speciation of Point Source VOC Emissions of Ambient Air Quality Modeling. Texas Commission on Environmental Quality, July 3003

7. Background Material: Speciation Profiles and Size Fractions. See http://www.arb.ca.gov/ei/speciate/speciate.htm (accessed February 2006)
8. SPECIATE 4.0 Quality Management Plan (QMP) / Quality Assurance Project Plan (QAPP), Holoman, D.W., E.H. Pechan & Associates, Durham, NC. 2006

KEY WORDS

SPECIATE, emissions, profiles, speciated, aerosol, particulate matter, PM, PM 2.5, VOC, TOG, TOC, modeling, EPA, database