



SPECIATE: Guidelines for Data Developers



June 2019

EPA Contract No. EP-BPA-17H-0012

EPA/600/R-19/085

Submitted to:

Dr. Marc Menetrez (E343-02)

Office of Research and Development

U.S. Environmental Protection Agency Research Triangle Park, NC 27711

Submitted by:

Abt Associates Inc.

Drs. Ying Hsu, Frank Divita, and Jonathan Dorn

6130 Executive Boulevard, Rockville, MD 20852 4907

CONTENTS

1.	Introduction	1
2.	Speciation Profile Definition, Data Collection, and Completeness	2
3.	Quality	4
4.	Format for Compiling Data	6
5.	Data Normalization	5
6.	Profile Quality Criteria Evaluation	7
	Quality Criteria Factors	7
	References	11
APPENDIX A.	Descriptive Data Dictionary (How to populate these fields for your data can be found in the template)	A-1

LIST OF TABLES

Table 1.	Relationships between TOG, VOC, NMOG, THC, and NMHC	2
Table 2.	Description of the Data Tables in the SPECIATE Data Template	6

LIST OF FIGURES

Figure 1.	Overview for Adding Speciation Data in EPA SPECIATE Database	1
-----------	--	---

1. Introduction

SPECIATE is the U.S. Environmental Protection Agency's (EPA) repository of speciation profiles of many types of air pollution sources. The profiles provide the species makeup or composition of organic gas (such as volatile organic compounds, or VOC), particulate matter (PM) and other pollutants emitted from these sources. Speciation profiles are used by EPA, other governmental and non-governmental agencies including international agencies, the regulated community and academia. Speciation profiles are used in creating speciated emissions inventories, including those needed for photochemical air quality modeling done in support of air quality management activities such as management of surface-level ozone, regional haze, and PM. Detailed documentation of SPECIATE is provided at EPA's [SPECIATE web page](#) (last accessed May 2019).

The purpose of this document is to inform the research community about the content and quality considerations of data so that the EPA can consider community-developed data for inclusion in SPECIATE. Researchers can provide these data voluntarily to EPA for consideration to be added to SPECIATE.

Figure 1 provides a quick-step guide for voluntary practices provided in these guidelines for speciation data that could be incorporated into the SPECIATE database.

Figure 1. Overview for Using Data Development Guidelines

Step 1	<ul style="list-style-type: none">• Review these guidelines to ensure test plan will produce results suitable to be used for SPECIATE database
Step 2	<ul style="list-style-type: none">• Conduct source testing and analytical analysis for chemical compositions
Step 3	<ul style="list-style-type: none">• As an option for proper formatting and metadata fields, you may voluntarily download the template workbook from EPA at the SPECIATE home page
Step 4	<ul style="list-style-type: none">• Voluntary use of the 4 tabs of the template workbook will organize speciation data and help you to ensure that the information needed by SPECIATE will be available; these tabs are (a) Raw Data; (b) PROFILES; (c) SPECIES; and (d) MasterReferenceList
Step 5	<ul style="list-style-type: none">• All questions about these guidelines, the workbook template, and notification that data are available for EPA to use can be sent to the SPECIATE Workgroup Email
Step 6	<ul style="list-style-type: none">• The EPA SPECIATE Workgroup (SWG) will respond to your inquiries and notifications as expeditiously as possible, and may have follow-up questions that would need to be answered for the EPA to use the data

2. Speciation Profile Definition, Data Collection, and Completeness

Speciation profiles are chemical compositions of organic gas, PM, and other pollutants emitted from these sources. In the SPECIATE database, profiles are presented as the weight percent of chemical species measured in a source-specific emission stream. The database also has optional fields that allow actual emission factors (in addition to fractional amounts of a “master pollutant”) to be included in SPECIATE.

For organic gas profiles, weight percents reflect the composition of the organic gases portion of the source measured. Species are normalized by the “master pollutant” which represents the total organic gases measured. A profile’s “master pollutant” can be any one of the following, depending on the available species and analytical methods: Total Organic Gases (TOG), Non-Methane Organic Gases (NMOG), VOC, or Non-Methane Hydrocarbons (NMHC). TOG are compounds of carbon, excluding carbon monoxide, carbon dioxide, carbonic acid, metallic carbides or carbonates, and ammonium carbonate. VOC profiles contain similar compounds as TOG profiles, except that VOC profiles exclude compounds that have negligible photochemical reactivity (i.e., exempt VOC compounds). The EPA definition of VOC and a list of exempt organic gases are available in Title 40, Chapter I, Subchapter C, Part 51, Subpart F, [Section 51.100](#) (*last accessed May 2019*) in the Code of Federal Regulations.

Table 1 provides the relationships among TOG¹, VOC², NMOG, THC³, and NMHC:

Table 1. Relationships among TOG, VOC, NMOG, THC, and NMHC.

Species	Definition
TOG	= VOC + exempt compounds (e.g., methane, ethane, various chlorinated fluorocarbons, acetone, perchloroethylene, volatile methyl siloxanes, and other compounds listed in the regulatory definition of VOC provided below).
TOG	= NMOG + methane
THC	= NMHC + methane [contain only hydrocarbons (i.e., not oxygenated compounds like aldehydes) due to gas chromatography-flame ionization detector (GC-FID) measurement technique]
NMOG	= NMHC + oxygenated compounds

A metadata field (the MASTER_POLLUTANT field) in the SPECIATE database indicates whether a profile is based on TOG, NMOG, VOC, or NMHC.

-
- ¹ TOG means “compounds of carbon, excluding carbon monoxide, carbon dioxide, carbonic acid, metallic carbides or carbonates, and ammonium carbonate.” TOG includes all organic gas compounds emitted to the atmosphere, including the low reactivity, or “exempt VOC” compounds (e.g., methane, ethane, various chlorinated fluorocarbons, acetone, perchloroethylene, volatile methyl siloxanes, etc.). TOG also includes low volatility or “low vapor pressure” (LVP) organic compounds (e.g., some petroleum distillate mixtures). TOG includes all organic compounds that can become airborne (through evaporation, sublimation, as aerosols, etc.), excluding carbon monoxide, carbon dioxide, carbonic acid, metallic carbides or carbonates, and ammonium carbonate.
 - ² VOC means any compounds of carbon that participate in atmospheric photochemical reactions, excluding methane, ethane, acetone, carbon monoxide, carbon dioxide, carbonic acid, metallic carbides or carbonates, and ammonium carbonate. VOC, additionally, exclude numerous exempt compounds that can be found in the Electronic Code of Federal Regulations under Title 40, Chapter I, Subchapter C, Part 51, Subpart F, §51.100. The list of exempt compounds is updated when new compounds are added through rulemaking.
 - ³ THC means organic compounds, as measured by gas chromatography-flame ionization detector (GC-FID). Notably, an FID measures carbon and hydrogen.

The data for a SPECIATE profile should fully characterize the source emissions and should not consist of just a few species. If there are major components missing from a profile, it will mischaracterize the composition of a source. Ideally, profiles should be based on TOG as the “master pollutant” and include methane and all organic functional groups (e.g., alkanes, alkenes, aromatics, carbonyls, etc.) associated with the sources. For example, TOG profiles from combustion sources should include alkanes, alkenes, aromatics, carbonyls, and semi-volatile organic compounds (SVOC), if possible. As another example, methanol, is a major component of emissions from pulp and paper industry sources and should not be missing from profiles for key sources in that industry.

A starting point for determining which compounds to measure is to find a similar source in the SPECIATE database. Ambient data monitoring networks are another source of information. The target list of compounds measured by the Photochemical Assessment Monitoring Stations (PAMS, EPA 1998) is a good reference for organic gas species that may be present. The [Ambient Monitoring Technology Information Center \(AMTIC\)](#) (*last accessed May 2019*) website posts the current PAMS [target compound list](#). However, it is important to note that, depending on the source, additional species may also be present (or some PAMS species may not be present). Additional species can be found in standard EPA test methods [e.g., TO-11A (carbonyls, EPA 1999a), TO-13A (SVOC/polyaromatic hydrocarbons (PAHs), EPA 1999b), TO-15 (toxic VOC, EPA 1999c)], posted on the [air toxics monitoring methods page](#) of the [Ambient Monitoring Technology Information Center \(AMTIC\)](#) website. A single instrument or measurement protocol cannot measure all TOG species that are needed for a complete speciation profile. Thus, to develop a speciation profile that could be useful for SPECIATE, it is likely that multiple instruments are needed to fully characterize organic gasses emitted from sources.

PM profiles should also be as complete as possible. For SPECIATE, they need to include the size fraction of the PM being speciated (SPECIATE uses the LOWER_SIZE and UPPER_SIZE metadata fields to store the size fraction). For air quality modeling purposes, PM_{2.5} profiles are generally more widely used than PM₁₀ profiles, though if both are created, the different compositions of the two size fractions is of interest. A reference for PM species that should be looked for are the elements reported by the [IMPROVE](#) and [PM_{2.5} Speciation Trends](#) networks (*last accessed May 2019*). PM species of interest are water-soluble ions (sulfates and nitrates at a minimum, plus ammonium, potassium, sodium, chloride, fluoride, phosphate, calcium, and magnesium), SVOC, and carbon fractions [Organic Carbon (OC), and Elemental Carbon (EC) (used interchangeably with black carbon)], preferably with further breakdowns of OC and EC. Also of interest for chemical transport modeling (e.g., the Community Multiscale Air Quality Modeling System (CMAQ)) are the CMAQ5.2 AE6 aerosol mechanism species, which include several discrete ions and atoms as provided in the SPECIATE 5.0 documentation (Table G-1) posted on EPA’s [SPECIATE web page](#).

For PM profiles, test results from dilution sampling trains are recommended for use in SPECIATE, since these results come closest to representing the composition of emissions in the ambient air. The ideal normalization basis for a PM profile is the gravimetric mass collected on a PM Teflon™ filter. This is because that approach is consistent with the PM emission factor measurements. If the gravimetric mass is not available, then the sum of fully speciated compounds [including derived mass such as a non-carbon organic mass (NCOM)] can be used as the normalization basis to calculate a PM profile.

In addition to the weight percent of species in profiles, available information on the analytical uncertainty for individual test profiles should be quantified and described separately. An ideal source testing campaign should quantify sampling and analytical uncertainties. Sampling uncertainties can be calculated

by sampling multiple replicates from the same source under the same condition. Analytical uncertainties can be quantified by measuring the same sample numerous times and calculating the standard deviation.

When replicate tests are performed and then compiled to construct a profile, the weight percents of the profile are computed as follows: (1) for six or more tests, the median weight percents of each species should be used; (2) for five or fewer, the mean of each species can be used to represent the weight percent of that species. In both cases, the species weight percents should then be renormalized such that the profile's weight percents sum to 100 percent. An estimate of the variability of each species (e.g. standard deviation) should be provided in the metadata (UNCERTAINTY_PERCENT field). The method used to estimate the central tendency (mean, median, or other method) is documented in the metadata (PROFILE_NOTES field). The method for computing the variability is also documented as metadata (UNCERTAINTY_METHOD field).

3. Quality

Researchers should understand that EPA strives to use data of “good” to “excellent” quality for SPECIATE. The SWG is a team of scientists and engineers that reviews data quality prior to the data being accepted for inclusion in SPECIATE. The team uses a quality score (QSCORE) based on a set of criteria to determine a perceived overall quality of a speciation dataset. Because there are so many complex and variable aspects to collecting speciation data, the QSCORE approach provides leeway for capturing that complexity because a simplistic black-and-white set of criteria would not be useful. The SWG assigns a QSCORE to data being considered for inclusion in the database, and it is recorded in SPECIATE if the data are accepted for inclusion. More information on the quality rating protocol is available in Section 6 of this document.

To capture the data quality in the QSCORE, supporting information about the measurements is critically important. Ideally the supporting information is a peer-reviewed research paper or a report that fully describes the source, sampling methods and conditions, analytical methods, quality assurance methods, uncertainties and assumptions, in addition to providing complete and relevant data. The SPECIATE database provides a sufficient structure to thoroughly document profiles and their underlying analyses. Thus, better quality data will include the supporting information that will allow the EPA to populate the various data fields in SPECIATE as thoroughly as possible. The fields described in Appendix A provide the research community with a list of such supporting information to consider when performing measurement research and documenting the results. Researchers are encouraged to contact the SWG (email SPECIATE_WG@epa.gov) to ask questions or solicit advice.

Key considerations for improving the overall quality of the measurements and resulting data are as follows:

Choose Appropriate Measurement Methods – Reviewers experienced in analytical methods and application of speciation profiles will need to determine if characteristic compounds are present and properly measured. Sampling and analytical procedures need to be specific to the source and documented as thoroughly as possible. For example, the EPA Method TO-14 is not an appropriate method for dairy farm emission speciation. Since this method was developed to test industrial sources, fatty acids and other important organic species were not included in the target species list.

Select Methods with Appropriate Measurement Precision – Low precision is expected for certain species; the data quality ratings should reflect this issue. EPA standard test methods [e.g., TO-11A (carbonyls, EPA 1999a), TO-13A (SVOC/PAHs, EPA 1999b), TO-15 (toxic VOC, EPA 1999c)] are

recommended for accurate chemical analyses. Note that olefinic aldehydes such as acrolein and crotonaldehyde degrade partially and form unknown species. This is due to a loss of carbonyl-dinitrophenylhydrazine (DNPH) derivative from the reaction of atmospheric ozone on DNPH-coated silica gel cartridges while sampling ambient air. This bias can be eliminated when sampling for carbonyls with the application of an ozone scrubber system (potassium iodide (KI)-coated denuder) preceding the DNPH cartridge (TO-11A, EPA 1999a).

Overall Confidence in the Measurements – Results obtained from the test program should be consistent with expectations for that source, and if not, the differences should be sufficiently accounted for. For example, in an U.S. Air Force sponsored study (AFIERA/RSEQ, 1998) measuring aircraft exhaust compositions, a brief discussion in the measurement section showed that the contractor measured essentially the same concentrations of target compounds in the background air as in the samples collected from aircraft exhaust. As a result, toxic species were reported at relatively low emission rates in this study. In cases where there are significant unexplainable results, the data should not be included in the SPECIATE database.

Consideration of Source Category-specific Issues – For certain source categories such as the pulp and paper industry, oxygenated compounds contribute significantly to organic gas emissions. The generic total hydrocarbon (THC) method using an FID calibrated with hydrocarbon standards (e.g. hexane) does not properly characterize the total TOG or VOC emissions. For processes whose emissions are dominated by methanol, this compound (and other oxygenated species) should be sampled and quantified separately using GC calibrated with a methanol standard (see Someshwar, 2003). Due to poor detector performance, the emission rates measured for THC were observed to be less than those measured specifically for methanol using an appropriate standard. Consequently, for this case, the THC is not suitable to serve as the normalization basis for this organic gas profile. The solution is to collect fully speciated data using appropriate methods and to consolidate all organic gases into a total organic gas profile for normalization.

Speciation profiles developed from the following methods are less desirable for inclusion in SPECIATE:

1. Samples from combustion sources not collected by dilution sampling;
2. Low total speciated percentage (less than 80% for both organic gases and PM);
3. PM profiles normalized by the “sum of species” mass, which assumes profiles of this type are fully speciated;
4. Any noticeable outliers or other unreasonable test results; and
5. Unpublished data from an author/institution unfamiliar to the SWG.

The research community can further review the QSCORE criteria questions (Section 6) that EPA uses to assign QSCORE ratings to profiles. This information can be used to improve study design and implementation to obtain higher quality results that better meet downstream user needs such as the SPECIATE database.

4. Data Normalization

Because the base measurement unit in the database is weight percent, data processing for SPECIATE requires normalization, which is the process for calculating the species fractions from the total mass (e.g., VOC or PM_{2.5}) that is sampled. The method used for profile normalization should be clearly documented, and the rationale for selecting the normalization basis should be stated. The normalization basis should be documented in the metadata (NORMALIZATION_BASIS) field and the rationale could be provided on the raw data tab of the Microsoft Excel® template workbook discussed in Section 5 of this document.

Normalization of organic gas data should be on a mass basis (i.e. mass species/mass TOG; emission rate species/emission rate TOG). Volume carbon basis is not a recommended normalization approach because assumptions are needed regarding the composition of unresolved species. Mole fractions should be converted to mass fractions. Whenever possible, researchers should use a normalization basis of total gas chromatography (GC)-elutable organic gases.

Normalization of PM data should be size-specific. Ideally, the profile will be normalized on total PM (with a specified upper size limit), PM₁₀, or PM_{2.5}. However, normalization based on other size fractions can also be accommodated in SPECIATE. Profiles normalized on total gravimetric mass are preferred. If sum of species is used, the major chemical components (sulfate, nitrate, ammonium, EC, OC with estimated NCOM, soil elements with estimated or measured oxides) should be present. Consult Reff, et. al. for additional details on the estimated chemical components.

5. Format for Compiling Data

The SPECIATE database is a Microsoft Access® relational database. The current SPECIATE data structure is documented in the final report for SPECIATE 5.0 which can be accessed from the [SPECIATE website](#). To facilitate review and use of data, researchers can compile data in a form that can be easily added to the SPECIATE database. This section provides a format that the EPA can use to import data into the SPECIATE database and metadata fields that must be present for the data to be usable by the EPA.

If researchers voluntarily use this formation to compile their speciation data, the information should be filled in as completely as possible, including references, test methods, analytical methods, Chemical Abstracts Service (CAS) numbers, data quality ratings, normalization basis, etc. To facilitate proper formatting, the EPA has provided an annotated Microsoft Excel® template workbook on the [SPECIATE website](#), which provides the fields needed for the current version (SPECIATE 5.0) of the database. The data should be compiled by populating the tabs of the template Microsoft Excel® workbook described in Table 2.

Table 2. Description of the Data Tables in the SPECIATE Data Template

Tab Name	Description
RAWDATA	This table contains the data from your study from which you compute weight percents to use for the SPECIATE tables. If the data are in a publication, the RAWDATA tab would identify the table numbers from the publication that are associated with data in the template. If some species are inferred, provide the method and/or assumptions used for the inferred values. Include formulas in this tab to document the steps you did to manipulate the data to convert to weight percents. The rationale for the normalization basis should be provided in this tab. The format is not specified; however, providing CAS numbers (where available) and SPECIES_ID are useful for identifying each species in the profile. You can use the supporting SPECIATE_PROPERTIES table that is available in the template to determine the SPECIES_ID for each species. Another resource for chemical abstract numbers and synonyms is the Substance Registry Service .
PROFILES	This table includes metadata about the profile. There are several fields in this table that allow researchers to provide documentation of the emissions source being measured, sampling conditions/methods and other notes that help others better understand the profile. Where appropriate, this documentation would include fuel type, operating parameters, emissions controls, and type of facility. Other metadata includes the normalization basis, geographic region (particularly important if the source characteristics are region-specific) and date of test. Non-detects or incomplete analyses should be documented in the PROFILE_NOTES so that the reader fully understands the analytical results. The specific fields in this table are described in the template.

SPECIES	This table includes the SPECIES_ID, the profile code associated with the species, the percentage of the species 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 percentages in the profile.
MasterReferenceList -to enter	<p>This table includes keywords for the profile and information that characterizes the reference document(s) associated with the profile, including whether a particular reference is the primary reference. There may be more than one reference document for each profile; each reference document goes into a separate row of the table.</p> <p>The SPECIATE MasterReferenceList table is included in the template (called "SPECIATE5.0 MRL-donotchange"). If your data comes from a reference already used in SPECIATE, then use the exact text from the "NEW REFERENCE" field from the SPECIATE MasterReferenceList when filling out the information in the MasterReferenceList for your profile.</p>

6. Profile Quality Criteria Evaluation

Quality Criteria Factors

The Quality Criteria Factors (QSCORE) provide an evaluation framework to easily recognize and assign value points to indicators of a strong, well-planned and executed study, which is presented in a complete and logical manner. The presentation of air emission profile data can be in the form of a peer-reviewed publication or report. This information is provided so that the research community can better understand the features of a higher quality speciation study.

The QSCORE framework guides EPA data reviewers to assign quality value points to the aspects of the study deemed most important for use in SPECIATE. The framework is meant to be comprehensive, but should also be easy to understand and apply, not rigid or overly detailed. The QSCORE evaluation is based on a series of questions with points assigned to each question. An ideal QSCORE would have 30 (Data from Measurements) or 29 (Data from other Methods) desired criteria (points). The points are additive, influencing, but not necessarily distinguishing the study. The QSCORE total points are valued as follows:

20-30 = excellent

12-19 = good

5-11 = fair

<4 = poor

Each numerical ranking (QSCORE) is added to the SPECIATE Database.

DATA FROM MEASUREMENTS - (Ideal score of 30)

No.	Question	Total Points
1	Are data from a peer-reviewed publication?	1
2	Is the source U.S. based or does it relate to a National Emissions Inventory (NEI) source?	1
3	Is the author well known or affiliated with a well-known research organization in conducting speciated source measurements?	1

4	Is the emission source current, are up-to-date technologies employed (collection, measurement, analysis)?	1
5	Is subject source identified as “priority” source (see, for example, the study: Bray, et. al. ¹)	1
6	Were data collected under an established quality system or sufficiently addressed /are QA/QC activities associated with the data collection/measurements included in the publication or supplementary information?	1
7	Sampling Design	
7a	Is the sampling design discussed logically (logic behind the experiments)?	1
7b	Are the data limitations clear (i.e., can the reviewer easily figure them out or are they explicitly stated)?	1
7c	Are assumptions clearly stated? (e.g., fireplace is representative of typical fireplace found throughout the country)	1
7d	Are samples capturing the natural variability of the sources?	1
8	Measurement Methodologies	
8a	Is measurement instrumentation presented or referenced?	1
8b	Are the data limitations clear?	1
8c	Were measurements taken using standard methods [EPA, National Institute of Standards and Technology (NIST)], and applicable/up-to-date technologies, methods, and instrumentation?	1
8d	Are replicate measurements done (duplicate or triplicate)? (Measurement methods using duplicate or triplicate collection implies that the study paid attention to data accuracy, representation and reproducibility. This attention should be viewed as an advantage.)	1
9	Data reduction procedures (statistics)	
9a	Are standard deviations (SDs) presented in the paper? (SDs are needed in the profile or we would contact the PI to get it.)	1
9b	Are SDs acceptable for the type of source and pollutants measured?	1
9c	Are the data ready for listing? (i.e., data are already in emission factor form, not in need of conversion or clarification; units consistently used throughout the publication; appropriate number of significant figures reported?)	1
9d	Is there complete speciation data of PM or organic gas provided? For organic gas, does the profile include a total amount of gaseous organic compounds (TOG), TOG should include (1) methane; (2) alkanes, alkenes and aromatic VOC; (3) alcohols; (4) aldehydes. PM _{2.5} should include critical pollutants such as (1) EC and OC; (2) sulfate/nitrate/NH ₄ ⁺ ions; (3) metals/inorganics. Higher scores are given if PAHs and SVOCs are also available.	1-10
10	The overall evaluation should ask; is the paper transparent with regards to describing sampling, test methods and data manipulation? Did the clarity and purpose of this paper leave a positive impression? (This element is meant to be based on the EPA reviewer’s impression of the paper, not a hard-fast scale, and may vary from one reviewer to another.)	1-3

1. Bray, et. al. 2019. Bray, C.D., Strum, M., Simon, H., Riddick, L., Kosusko, M., Menetrez, M., Hays, M.D., Rao, V., 2019. An Assessment of Important SPECIATE Profiles in the EPA Emissions Modeling Platform and Current Data Gaps. Atmospheric Environment 207, 93-104. DOI: 10.1016/j.atmosenv.2019.03.013

DATA FROM OTHER METHODS (Blended) (Ideal score of 29)

OTHER METHODS: Any paper where the researches did not directly measure what they report in the paper. Examples of other methods: Urbanski 2014 (putting together others' work), profile for flares (FLR99) that estimated the composition from a test of propylene.

No.	Question	Total Points
1	Are data from a peer-reviewed publication?	1
2	Is the source U.S. based or does it relate to a National Emissions Inventory (NEI) source?	1
3	Is the author well known or affiliated with a well-known research organization in conducting speciated source measurements or analyses?	1
4	Is the emission source current, are up-to-date technologies employed (collection, measurement, analysis)?	1
5	Is subject source identified as "priority" source (see, for example, the study: Bray, et. al. ¹)	1
6	Composite Data Development	
6a	Are data based on an established, acceptable methodology?	2
6b	If any of the values or data are based on assumptions or calculations are they clearly documented?	2
6c	Was post-processing used for the data? If so, is it novel, reasonable or widely accepted?	2
7	Is there complete speciation data of PM or organic gas provided? For organic gas, does the profile include a total amount of gaseous organic compounds (TOG), TOG should include (1) methane; (2) alkanes, alkenes and aromatic VOC; (3) alcohols; (4) aldehydes. PM _{2.5} should include critical pollutants such as (1) EC and OC; (2) sulfate/nitrate/NH ₄ ⁺ ions; (3) metals/inorganics. Higher scores are given if PAHs and SVOCs are also available.	1-10
8	Are assumptions clearly stated? (i.e., fireplace is representative of typical fireplace found throughout the country)	2
9	Data reduction procedures (statistics)	
9a	Are standard deviations (SDs) presented in the paper? (SDs are needed in the profile or we would contact the PI to get it.)	1
9b	Are SDs acceptable for the type of source and pollutants measured?	1

9c	Are the data ready for listing? (i.e., data are already in emission factor form, not in need of conversion or clarification; units consistently used throughout the publication; appropriate number of significant figures reported?)	1
10	The overall evaluation should ask; is the paper transparent with regards to describing sampling, test methods and data manipulation? Did the clarity and purpose of this paper leave a positive impression? (This element is meant to be based on the EPA reviewer's impression of the paper, not a hard-fast scale, and may vary from one reviewer to another.)	1-3

1. Bray, et. al. 2019. Bray, C.D., Strum, M., Simon, H., Riddick, L., Kosusko, M., Menetrez, M., Hays, M.D., Rao, V., 2019. An Assessment of Important SPECIATE Profiles in the EPA Emissions Modeling Platform and Current Data Gaps. Atmospheric Environment 207, 93-104. DOI: 10.1016/j.atmosenv.2019.03.013

References

- AFIERA/RSEQ, 1998. *Aircraft Engine and Auxiliary Power Unit Emissions Testing for the US Air Force*, Environmental Quality Management Inc, and Roy F. Weston Inc., December 1998.
- Bray, et. al. 2019. Bray, C.D., Strum, M., Simon, H., Riddick, L., Kosusko, M., Menetrez, M., Hays, M.D., Rao, V., 2019. An Assessment of Important SPECIATE Profiles in the EPA Emissions Modeling Platform and Current Data Gaps. *Atmospheric Environment* 207, 93-104. DOI: 10.1016/j.atmosenv.2019.03.013
- EPA, 2002. *Draft Guidelines for the Development of Total Organic Compound and Particulate Matter Chemical Profiles*, developed by Emission Factors and Inventory Group, U.S. EPA, September 25, 2002.
- EPA (PAMS), 1998. *Technical Assistance Document for Sampling and Analysis of Ozone Precursors*, EPA/600-R-98/161, September 1998.
- EPA (TO-11A), 1999a. *Determination of Formaldehyde in Ambient Air Using Adsorbent Cartridge Followed by High Performance Liquid Chromatography (HPLC)*, EPA/625/R-96/010b, January 1999.
- EPA (TO-13A), 1999b. *Determination of Polycyclic Aromatic Hydrocarbons (PAHs) in Ambient Air Using Gas Chromatography/Mass Spectrometry (GC/MS)*, EPA/625/R-96/010b, January 1999.
- EPA (TO-15), 1999c. *Determination of Volatile Organic Compounds (VOCs) in Air Collected in Specially-Prepared Canisters and Analyzed by Gas Chromatography/Mass Spectrometry (GC/MS)*, EPA/625/R-96/010b, January 1999.
- Reff et al., 2009: Reff, A., Bhawe, P.V., Simon, H., Pace, T.G., Pouliot, G.A., Mobley, J.D., and Houyoux, M., Emissions Inventory of PM_{2.5} Trace Elements across the United States, *Environmental Science and Technology*, 43: 5790–5796, 2009.
- Someshwar, 2003. Arun Someshwar, *Compilation of ‘Air Toxic’ and Total Hydrocarbon Emissions Data for Sources at Kraft, Sulfite and Non-Chemical Pulp Mills – an Update*, Technical Bulletin No. 858, National Council for Air and Stream Improvement, February, 2003.
- Urbanski, 2014, Urbanski, S., Wildland Fire Emissions, *Carbon, and Climate: Emission Factors*, Forest Ecology and Management, 317, 51–60, 2014.
- Watson and Chow, 2002. Watson, J. and J. Chow, *Considerations in Identifying and Compiling PM and VOC Source Profiles for the SPECIATE Database*, Desert Research Institute, August, 2002.

APPENDIX A. Descriptive Data Dictionary (How to populate these fields for your data can be found in the template)

Field Name	Data Type	Length ⁴	Description	Will EPA provide
PROFILES Table				
PROFILE_CODE	Text	10	Profile Code - alphanumeric. Ideally less than 7 characters for mobile profiles and less than 10 characters for others due to emissions model (e.g., SMOKE) field length limitations	Yes
PROFILE_NAME	Text	255	Profile Name – use a unique name that describes the source.	
PROFILE_TYPE	Text	20	Indicates type of profile: PM-AE6, PM-VBS, PM-Simplified, PM, GAS, GAS-VBS and OTHER	
MASTER_POLLUTANT	Text	25	Indicates the pollutant to be used in calculation. Options for organic gases are described in Section 2, above. PM profiles use “PM”	
QSCORE	Number	2	Profile data quality score out of 30 points total. 20-30 = excellent. 12-19 = good. 5-11 = fair. Less than 5 = poor.	Yes
QUALITY	Text	3	Overall Quality Rating (A-E) based on Vintage Rating and Data Quantity Rating, see Chapter II.D of the SPECIATE 5.0 documentation for an explanation of how it is determined	
CONTROLS	Text	150	Emission Controls Description	
PROFILE_DATE	Date/Time	10	Date profile added (MM/DD/YYYY)	
PROFILE_NOTES	Memo		Notes about the source and how data were put together. Examples include method for compositing, descriptions about the overall procedures and/or study purpose.	
TOTAL	Number	6	Sum of species percentages for a given profile, excluding organic species, inorganic gases, and elemental sulfur in individual PM profiles (see Chapter IV.G of the SPECIATE 5.0 documentation- “Avoiding Double Counting Compounds” for rationale).	
TEST_METHOD	Memo		Description of sampling/test method for overall profile	
NORMALIZATION_BASIS	Text	100	Description of how profile was normalized (see Chapter IV.F of the SPECIATE 5.0 documentation report for details; see also Section 4 of this document)	
ORIGINAL_COMPOSITE	Text	2	Specifies whether the profile is original, composite of SPECIATE profiles or study composite. Allowed values: 'C','O','SC'. The option for study composite, SC, added in SPECIATE5.0, means composite was developed in the study.	
STANDARD	Yes/No	1	Indicates whether the profile is provided by EPA SPECIATE (standard) or user-added. The database is constructed to allow users to add profiles in the future.	

⁴ Length – maximum number of characters allowed.

Field Name	Data Type	Length ⁴	Description	Will EPA provide
INCLUDE_INORGANIC GAS	Yes/No	1	Indicates the presence or absence of inorganic gas species in this profile (e.g., sulfur dioxide, hydrogen sulfide, oxides of nitrogen, etc.)	
TEST_YEAR	Text	50	Indicates year testing was completed	
JUDGEMENT_RATING	Number	4	Subjective expert judgement rating based on general merit (see Chapter II.D of the SPECIATE 5.0 Documentation)	
VINTAGE_RATING	Number	4	Vintage based on TEST_YEAR field (see Chapter II.D of the SPECIATE 5.0 Documentation)	
DATA_QUANTITY_RATING	Number	4	Data sample size rating based on number of observations, robustness (see Chapter II.D of the SPECIATE 5.0 Documentation)	
REGION	Text	50	Geographic region of relevance	
SAMPLES	Text	255	Number of samples (separate experiments or measurements) taken	
LOWER_SIZE	Number	5	Identifies lower end of aerodynamic diameter particle size, micrometers	
UPPER_SIZE	Number	5	Identifies upper end of aerodynamic diameter particle size, micrometers	
SIBLING	Text	25	GAS or PM Profile number taken from the same study, if exists	
VERSION	Text	5	SPECIATE database version that a profile was added to	
TOG_to_VOC_RATIO	Number	6	Ratio of TOG mass to VOC mass, computed as: $100\% / (100\% - \text{sum}(\text{nonVOC})\%)$	Yes
TEMP_SAMPLE_C	Number	6	Temperature while samples were taken, in degrees Celsius	
RH_SAMPLE	Number	6	Relative humidity while samples were taken.	
PARTICLE_LOADING_ug_per_m3	Number	6	PM loading during sampling in units of micrograms/m ³	
ORGANIC_LOADING_ug_per_m3	Number	6	Organic loading during sampling in units of micrograms/m ³	
CATEGORY_LEVEL_1_Generation_Mechanism	Text	255	The mechanism by which emissions are generated by the emissions source. (See Appendix F of the SPECIATE5.0 documentation for details)	
CATEGORY_LEVEL_2_Sector_Equipment	Text	255	This category provides more detail on the emissions generation category by including the sector and/or equipment or process used to generate the emissions. (See Appendix F of the SPECIATE5.0 documentation for details)	
CATEGORY_LEVEL_3_Fuel_Product	Text	255	This category provides the highest level of detail for the profile categorization. (See Appendix F of the SPECIATE5.0 documentation for details)	
MASTER_POLLUTANT_EMISSION_RATE	Number	6	PM or GAS emission rate, if available	
MASTER_POLLUTANT_EMISSION_RATE_UNIT	Text	50	PM or GAS emission rate units, if available	

Field Name	Data Type	Length ⁴	Description	Will EPA provide
ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO	Number	4	OM/OC ratio to calculate OM emissions. OM/OC ratio of 1.25 for motor vehicle exhaust, 1.4 for coal combustion, 1.70 for biomass combustion (other than wood fired boilers), 1.40 for wood fired boilers and all others, with some exceptions.	
MASS_OVERAGE_PERCENT	Number	6	Sum of species percentages that is over 100% calculated only for PM_AE6 profiles for which the mass of the measured OC and computed PNCOM was reduced so that the AE6 profile would not exceed 100%	
CREATED BY	Text	50	Person who added this profile	
CREATED DATE	Date/Time		Date the profile was added	
MODIFIED BY	Text	50	Person who modified this profile	
MODIFIED DATE	Date/Time		Date the profile was added	Yes
REVIEWED BY	Text	50	Person who reviewed this profile	Yes
REVIEWED DATE	Date/Time		Date the profile was reviewed	Yes
SPECIES Table				
PROFILE_CODE	Text	10	Unique Identifier links to PROFILES table.	Yes
SPECIES_ID	Number	5	Species Identifier (The same as ID in SPECIES_PROPERTIES table)	
WEIGHT_PERCENT	Number		Weight percent of pollutant (%)	
UNCERTAINTY_PERCENT	Number		Uncertainty percent of pollutant (%)	
UNCERTAINTY_METHOD	Memo		Description of method used to calculate uncertainty	
ANALYTICAL_METHOD	Text	100	Description of analytical method (e.g., X-ray fluorescence spectroscopy, ion chromatography)	
PHASE	Text	50	Indicate whether emissions were measured for PM, gaseous, or both phases.	
SPECIES_EMISSION_RATE	Number	6	Species emission rate	
SPECIES_EMISSION_RATE_UNIT	Text	50	Species emission rate units (e.g., mg/mile)	
KEYWORD_REFERENCE Table				
PROFILE_CODE	Text	10	Unique Identifier links to PROFILES table.	Yes
DATA_ORIGIN	Text	50	Source of data (e.g., EPA Air Pollution Prevention and Control Division (APPCD), Schauer, CARB, DRI, NPRI, Literature)	
REF_PRIMARY	Yes/No		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.	
REF_DESCRIPTION	Memo		Stores the descriptive information about the profile.	
REF_DOCUMENTS	Memo		Complete reference citation. Some profiles have multiple citations such as reports and journal articles.	
KEYWORD	Text		Keywords describing a profile.	
SPECIES_PROPERTIES Table				
SPECIES_ID	Number	9	Unique Identifier (Link to SPECIES table)	
CAS	Text	50	Chemical Abstracts Service (CAS) number assigned to pollutant (with hyphens) (blank if no CAS)	

Field Name	Data Type	Length ⁴	Description	Will EPA provide
EPA_ID	Text	50	EPA Chemical Identifier; provided by EPA Substance Registry Service (SRS) for species without CAS numbers	
SAROAD	Text	5	Storage and Retrieval of Aerometric Data (SAROAD) code	
PAMS	Yes/No	1	Is PAMS pollutant? (Yes or No)	
HAPS	Yes/No	1	Is Hazardous Air Pollutant (HAP)? (Yes or No) HAPs are defined in in the Clean Air Act, Section 112(b), changes to that list are in the Code of Federal Regulations (CFR), Title 40, Part 63. Current list is on EPA website.	
NAME	Text	255	Species name	
SYMBOL	Text	9	Standard chemical abbreviation	
SPEC_MW	Number	6	Species molecular weight	
NonVOCTOG	Yes/No	1	Is this species regarded as a volatile organic compound (VOC)? The VOC definition is from 40 CFR §51.100	
NOTE	Memo	250	Note (notes) about the SPECIES_ID or its properties	
SRS ID	Text	50	EPA SRS Chemical Identifier	Yes
Molecular Formula	Text	50	Molecular formula	Yes
OXYGEN_to_CARBON_RATIO	Number		Ratio of oxygen atoms to carbon atoms	
Smiles Notation	Text	10	Smiles notation	Yes
VP_Pascal_EPI	Number		Vapor Pressure in units of Pascals from the EPISUITE model	Yes
VP_Pascal_UM	Number		Vapor Pressure in units of Pascals from UManSysProp tool (uses the EVAPORATION algorithm, slightly updated)	Yes