# SPECIATE: Guidelines for Data Developers

EPA Contract No. 68HERD21A002

EPA/600/X-22/182

Submitted to:

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

**The views expressed in this document are those of the authors and do not necessarily represent the views or policies of the U.S. EPA. This document has gone through review process within the Agency and is cleared for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.**

# CONTENTS



# LIST OF TABLES



# LIST OF FIGURES

## Acronyms and Abbreviations

| | |
|---|---|
| CAS | Chemical Abstracts Service |
| CMAQ | Community Multi-scale Air Quality Modeling System |
| DRI | Desert Research Institute |
| EC | elemental carbon |
| EPA | Environmental Protection Agency |
| EPI | estimation program interface |
| FID | flame ionization detector |
| GC-FID | gas chromatography-flame ionization detector |
| GC-MS | gas chromatography-mass spectroscopy |
| HAPs | hazardous air pollutants |
| ID | identification |
| kg | kilogram |
| LVP | low vapor pressure |
| mg | milligram |
| MO | metal-bound oxygen |
| MW | molecular weight |
| NMHC | non-methane hydrocarbons |
| NMOG | non-methane organic gas |
| OC | organic carbon |
| OM | organic matter |
| OPERA | OPEn structure-activity/property Relationship App |
| QA | quality assurance |
| QSCORE | profile quality score |
| ROG | reactive organic gas |
| PAHs | polycyclic aromatic hydrocarbons |
| PAMS | photochemical assessment monitoring station |
| PM | particulate matter |
| $PM_{10}$ | particulate matter with an aerodynamic diameter $\leq$ 10 micrometers |
| $PM_{2.5}$ | particulate matter with an aerodynamic diameter $\leq$ 2.5 micrometers |
| PNCOM | particulate non-carbon organic matter |
| POC | primary organic compounds |
| POA | primary organic aerosols |
| SAROAD | Storage and Retrieval of Aerometric Data |
| SMOKE | Sparse Matrix Operator Kernel Emissions (EPA emissions modeling tool) |
| SRS | Substance Registry System |
| SVOC | semi-volatile organic compounds |
| SWG | SPECIATE work group |
| THC | total hydrocarbon |
| TOG | total organic gases |
| VBS | volatility basis set |
| VOC | volatile organic compounds |
| XRF | x-ray diffraction |

# 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 to create 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.

The purpose of this document is to inform the research community about the content and quality expectations 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 or can use this document as a guide to publish work on PM or VOC speciation for use in the SPECIATE database by EPA.

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** | • Review these guidelines to ensure test plan will produce results suitable for the SPECIATE database |
| **Step 2** | • Conduct source testing and analysis for chemical compositions |
| **Step 3** | • 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** | • Voluntary use of the 5 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; (d) PROFILE_REFERENCE_CROSSWALK; and (d) REFERENCES |
| **Step 5** | • 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** | • 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 before EPA can use the data |

## 2. Speciation Profile Definition, Data Collection, and Completeness

Speciation profiles are chemical compositions of organic gas, PM, and other pollutants (e.g., mercury) emitted from sources of these pollutants. 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. The most desired profiles to add to SPECIATE for air quality modeling applications are profiles for total organic gases (TOG), particulate matter less than or equal to 2.5 micrometers in diameter (PM2.5), and mercury.

For organic gas profiles, weight percents reflect the composition of the organic gases portion of the emissions from the source measured. Species are normalized by the "master pollutant" which represents the TOG 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, total hydrocarbons (THC), 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 in the Code of Federal Regulations. Because TOG is the most inclusive, it is the most desirable master pollutant for organic gas profiles.

Table 1 provides the relationships among TOG[1], VOC[2], NMOG, THC[3], and 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.

**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, THC or NMHC.

The data for an organic gas profile should fully characterize the source emissions and should not consist of just a few species. Ideally the study should strive to measure 100 percent of the mass of organics emitted. 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) 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-15a (toxic VOC, EPA 2019)], 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 $PM_{10}$ profiles, though if both are created, the different compositions of the two size fractions is of interest. A reference for PM species that should considered are the elements reported by the IMPROVE and $PM_{2.5}$ Speciation Trends networks . 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)]. Also of interest for chemical transport modeling [e.g., the Community Multiscale Air Quality Modeling System (CMAQ)] are the CMAQ aerosol mechanism species (for aerosol module versions 6 and higher), 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. Currently for both version 6 and version 7 of the aerosol mechanism (AE6 and AE7), the PM species needed by the model are identical.

For PM profiles, test results from dilution sampling trains are recommended for use in SPECIATE, since these results more closely represent 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 particulate non-carbon organic matter (PNCOM)] can be used as the normalization basis to calculate a PM profile.

For mercury profiles, elemental mercury, divalent gaseous mercury and particulate mercury should be included where present. A method that has been used to measure these species is the Ontario Hydro method (ASTM Standard Method 6784-16).

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 multiple tests are performed and then compiled to construct a representative composite profile, the weight percents of the species may be computed using the arithmetic mean, geometric mean, or median of the weight percents. Generally, the median is useful where there are a large number of samples (e.g., six or more) and where there are outliers that can skew the mean. The arithmetic mean is generally used for a small number of samples but can be chosen for a larger set, if there are no outliers. The geometric mean is also a method that can be used and is particularly useful computing the central tendency for a set of largely varying values (e.g., order of magnitude) and where zero or negative values may be ignored. The method used to estimate the central tendency is important metadata and can be provided in the PROFILE_NOTES field. In addition, an estimate of the variability of each species (e.g. standard deviation) would be documented in the UNCERTAINTY_PERCENT field along with the method (UNCERTAINTY_METHOD field).

## 3. Quality

Researchers should understand that the EPA strives to use data of "good" to "excellent" quality for SPECIATE. The SPECIATE Work Group (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. Using EPA-approved and updated measurement methods would also be a bonus. For example, 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 not included in the target species list.

**Select Methods with Appropriate Measurement Precision** – Low precision is expected for certain species; the QSCORE should reflect this issue. EPA standard test methods [e.g., TO-11A (carbonyls, EPA 1999a), TO-13A (SVOC/PAHs, EPA 1999b), TO-15a (toxic VOC, EPA 2019)] 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 a 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 a 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 TOG profile for normalization.

Characterization of the source is also important, including the sampling location. For oil and gas, for example, the sample and sampling location should be appropriate for characterizing the intended source and documented. Ideally, the data developer would additionally describe the processes represented by the tested source based on source classification codes .

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 percentages 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 percent 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), $PM_{10}$, 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 PNCOM, 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 Addendum for SPECIATE 5.1 which can be accessed from the SPECIATE website. To facilitate review and use of data, researchers should compile their data in this format.

If researchers voluntarily use this format 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 5.2 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. Two resources for chemical abstract numbers and synonyms are the Substance Registry Service and EPA chemical dashboard. |
| 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. If available, species emission factors can be provided in this table. If your data includes SPECIES that are not currently in the SPECIES_PROPERTIES table, then use the CAS number for this table and EPA will add the SPECIES and assign a SPECIES_ID. |
| PROFILE_ REFERENCE_ CROSSWALK | This table contains the profile codes and reference codes. For consistent naming convention purpose, the convention for the REF_Code field is lastname concatenated with date but can also be an organization name (e.g., EPA2020); developers can leave it to EPA for determining unique profile codes and reference codes. |
| REFERENCES | This table includes reference codes, reference, study description, and hyperlink for the publications. There may be more than one reference document for each profile (but no more than 3); each reference document goes into a separate row of the table. |

## 6. Profile Quality Criteria Evaluation

The quality criteria factors, referred to as 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. 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:

22-30 = excellent
16-21 = good
8-15 = fair
7 or less = poor

Each numerical ranking (QSCORE) is added to the SPECIATE database along with the description of the value (QSCORE_DESC).

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 payed 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?<br><br>For organic gas, does the profile include a total amount of gaseous organic compounds (TOG), TOG should include<br>(1) methane;<br>(2) alkanes, alkenes and aromatic VOC;<br>(3) alcohols;<br>(4) aldehydes.<br>$PM_{2.5}$ should include critical pollutants such as<br>(1) EC and OC;<br>(2) sulfate/nitrate/$NH4^+$ ions;<br>(3) metals/inorganics.<br>Higher scores are given if PAHs and SVOCs are also available.<br><br><u>Is there complete speciation data of Hg?</u><br>Hg should include:<br>(1) Elemental mercury ($Hg^0$)<br>(2) Reactive Gas mercury (a.k.a. ionic)<br>(3) Particulate form<br>Scoring guidance for Hg profiles: One species=2, Two species=6, all three species=10 | 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 researchers 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?<br><br>For organic gas, does the profile include a total amount of gaseous organic compounds (TOG), TOG should include<br>(1) methane;<br>(2) alkanes, alkenes and aromatic VOC;<br>(3) alcohols;<br>(4) aldehydes.<br>$PM_{2.5}$ should include critical pollutants such as<br>(1) EC and OC;<br>(2) sulfate/nitrate/$NH4^+$ ions;<br>(3) metals/inorganics.<br>Higher scores are given if PAHs and SVOCs are also available.<br><br>Is there complete speciation data of Hg?<br>Hg should include:<br>(1) Elemental mercury ($Hg^0$)<br>(2) Reactive Gas mercury (a.k.a. ionic)<br>(3) Particulate form<br>Scoring guidance for Hg profiles: One species=2, Two species=6, all three species=10 | 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 | 1-3 |

| | | |
|---|---|---|
| | impression of the paper, not a hard-fast scale, and may vary from one reviewer to another.) | |

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-15a), 2019, *Determination of Volatile Organic Compounds (VOCs) in Air Collected in Specially-Prepared Canisters and Analyzed by Gas Chromatography-Mass Spectrometry (GC/MS)*, , September, 2019 (https://nepis.epa.gov/Exe/ZyPDF.cgi/P100YDPO.PDF?Dockey=P100YDPO.PDF). *(last accessed June 2020)*

Reff et al., 2009: Reff, A., Bhave, 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. Should be 10 characters or less 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 being speciated. Options for organic gases are described in Section 2, above. PM profiles use "PM" | |
| QSCORE | Number | | Profile quality score out of 30 points total. 22-30 = excellent. 16-21 = good. 8-15 = fair. 7 or less = poor. | Yes |
| QSCORE_DESC | Text | 255 | Description of the numeric QSCORE rating. | 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 document for an explanation | |
| CONTROLS | Text | 150 | Emission Controls Description | |
| PROFILE_DATE | Date/ Time | | Date profile added (MM/DD/YYYY) | |
| PROFILE_NOTES | Long Text | | 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 | | 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 | Long Text | | 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 | | 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. | Yes |
| INCLUDES_INORGANIC GAS | Yes/No | | Indicates the presence or absence of inorganic gas species in this profile (e.g., sulfur dioxide, hydrogen sulfide, oxides of nitrogen, etc.) | |

[4] Length – maximum number of characters allowed.

| Field Name | Data Type | Length[4] | Description | Will EPA provide |
|---|---|---|---|---|
| TEST_YEAR | Text | 50 | Indicates year testing was completed | |
| JUDGEMENT_RATING | Number | | Subjective expert judgement rating based on general merit (see Chapter II.D of the SPECIATE 5.0 Documentation) | |
| VINTAGE_RATING | Number | | Vintage based on TEST_YEAR field (see Chapter II.D of the SPECIATE 5.0 Documentation) | |
| DATA_QUANTITY_RATING | Number | | 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 | 5 | Number of samples (separate experiments or measurements) taken | |
| LOWER_SIZE | Number | | Identifies lower end of aerodynamic diameter particle size, micrometers | |
| UPPER_SIZE | Number | | Identifies upper end of aerodynamic diameter particle size, micrometers | |
| SIBLING | Text | 10 | GAS or PM Profile number taken from the same study, if exists | |
| VERSION | Text | 5 | SPECIATE database version that a profile was added to | Yes |
| TOG_to_VOC RATIO | Number | | Ratio of TOG mass to VOC mass, computed by either (1) or (2) below:<br>(1) sum(all species%) / (sum(all species%) -sum(nonVOC)%)<br>(2) sum(all species%) / sum (VOC species%) | Yes |
| TEMP_SAMPLE_C | Number | | Temperature while samples were taken, in degrees Celsius | |
| RH_SAMPLE | Number | | Relative humidity while samples were taken. | |
| PARTICLE_LOADING_ug_per_m3 | Number | | PM loading during sampling in units of micrograms/$m^3$ | |
| ORGANIC CARBON_LOADING_ug_per_m3 | Number | | 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 | | PM or GAS emission rate (also known as emission factor), if available | |
| MASTER_POLLUTANT_EMISSION_RATE_UNIT | Text | 50 | PM or GAS emission rate units (e.g., mg/mile), if available | |
| ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO | Number | | 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. | |

| Field Name | Data Type | Length[4] | Description | Will EPA provide |
|---|---|---|---|---|
| MASS_OVERAGE_PERCENT | Number | | 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 | Yes |
| CREATED DATE | Date/ Time | | Date the profile was added | Yes |
| MODIFIED BY | Text | 50 | Person who modified this profile | Yes |
| 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 |
| Data_Origin | Text | 50 | Source of data (e.g., EPA Air Pollution Prevention and Control Division (APPCD), CARB, DRI, NPRI, Literature) | |
| Keywords | Text | 255 | Keywords describing a profile. | |
| DOC_LINK | Hyperlink | | Link to the workbook and/or any documentation on the EPA SPECIATE ftp site | Yes |
| Q_LINK | Hyperlink | | Link to the QSCORE documentation on the EPA SPECIATE ftp site | Yes |
| **SPECIES Table** | | | | |
| PROFILE_CODE | Text | 10 | Unique Identifier links to PROFILES table. | Yes |
| SPECIES_ID | Number | | Species Identifier (The same as ID in SPECIES_PROPERTIES table) | |
| WEIGHT_PERCENT | Number | | Weight percent of pollutant (%) | |
| INCLUDE_IN_SUM | Text | 3 | This is needed to indicate whether the species should be used in calculating the sum of the weight percents (in many PM profiles there could be overlapping species such as PAHs and PNCOM/POC or calcium atom and calcium ion) so not all species should be included to sum mass. | Yes |
| UNCERTAINTY_PERCENT | Number | | Uncertainty percent of pollutant (%) | |
| UNCERTAINTY_METHOD | Text | | 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 | | Species emission rate (also known as emission factor) | |
| SPECIES_EMISSION_RATE_UNIT | Text | 50 | Species emission rate units (e.g., mg/mile) | |
| **PROFILE_REFERENCE_CROSSWALK Table** | | | | |
| PROFILE_CODE | Text | 10 | Unique Identifier links to PROFILES table. | Yes |
| REF_Code | Text | 255 | Unique Identifier links to REFERENCES table. | Yes |

| Field Name | Data Type | Length[4] | Description | Will EPA provide |
|---|---|---|---|---|
| **REFERENCES Table** | | | | |
| REF_CODE | Text | 255 | Unique reference code links to PROFILE_REFERENCE_CROSSWALK table. | Yes |
| REFERENCE | Long Text | | Complete reference citation including a digital object identifier (doi), where available | |
| REF_DESCRIPTION | Long Text | | Stores the descriptive information about the profile. | |
| LINK | Hyperlink | | Link to the citations such as reports and journal articles (ok to repeat doi from the REFERENCE field). | |
| **SPECIES_PROPERTIES Table** | | | | |
| SPECIES_ID | Number | | Unique Identifier (Link to SPECIES table) | Yes |
| CAS | Text | 255 | Chemical Abstracts Service (CAS) number assigned to pollutant (with hyphens) (blank if no CAS) | Yes |
| SAROAD | Text | 5 | Storage and Retrieval of Aerometric Data (SAROAD) code | Yes |
| PAMS | Yes/No | | Is PAMS pollutant? (Yes or No) | Yes |
| HAPS | Yes/No | | 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. | Yes |
| SPECIES_NAME | Text | 255 | Species name | Yes |
| SPEC_MW | Number | | Species molecular weight | Yes |
| NonVOCTOG | Yes/No | | Is this species regarded as a volatile organic compound (VOC)? The VOC definition is from 40 CFR §51.100 | Yes |
| NOTE | Text | | Note (notes) about the SPECIES_ID or its properties | Yes |
| SRS ID | Text | 255 | EPA Substance Registry Service Chemical Identifier | Yes |
| DSSTox_ID | Text | 255 | The DSSTox Substance Identifier, a unique identifier associated with a substance on the EPA Distributed Structure-Searchable Toxicity (DSSTox) Database | Yes |
| Molecular Formula | Text | 255 | Molecular formula | Yes |
| OXYGEN_to_CARBON_RATIO | Number | | Ratio of oxygen atoms to carbon atoms | Yes |
| Smiles Notation | Text | 255 | 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 |
| VP_Pascal_OPERA | Number | | Vapor Pressure in units of Pascals from OPERA model (obtained from EPA Chemical Dashboard) | Yes |
| Duplicate_ID | Text | 255 | Identify which Species ID in this table represent the same compound | Yes |
| SYMBOL | Text | 255 | Standard chemical abbreviation | Yes |