Addendum

SPECIATE Version 5.4
Database Development Documentation

September 2025

EPA SPECIATE Workgroup
US Environmental Protection Agency

Abt Global
EPA Contract No. 68HERD21A002

EXECUTIVE SUMMARY

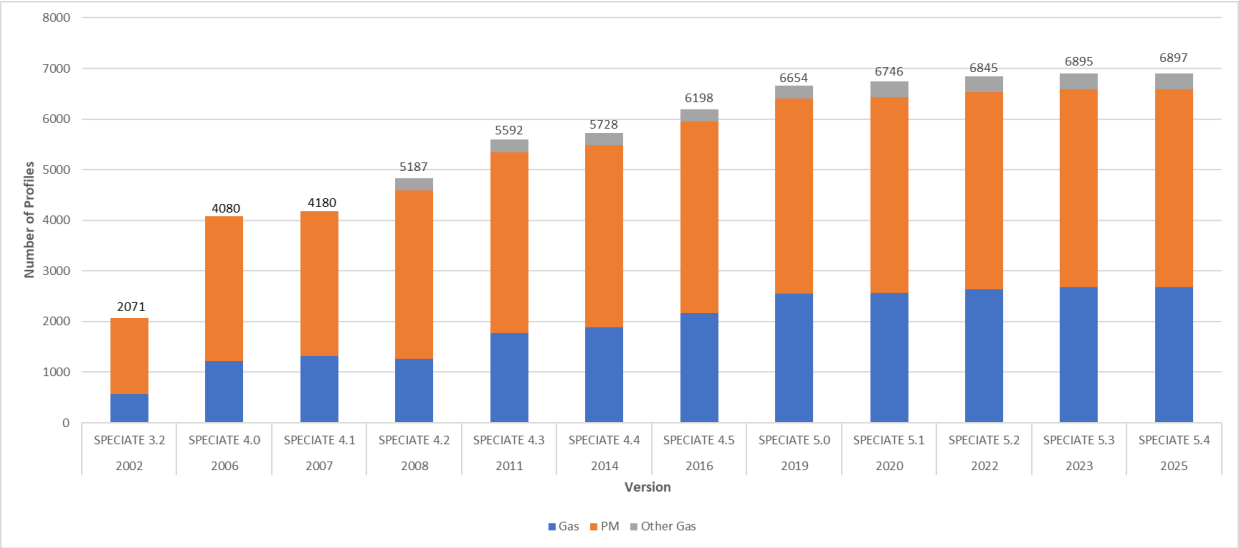
EPA is releasing an updated version of the SPECIATE database, SPECIATE v5.4. In lieu of full documentation, this addendum provides highlights of the revisions to the SPECIATE v5.3 database. Full documentation of the SPECIATE v5.0 database can be found in the SPECIATE documentation section of EPA's air emissions modeling website.

SPECIATE is the U.S. Environmental Protection Agency's (EPA) repository of speciation profiles for air pollution sources and provides the species makeup or composition of organic gas, particulate matter (PM) and other pollutants emitted from these sources. Uses of these source profiles include: (1) creating speciated emissions inventories for regional haze, PM, and photochemical air quality modeling; (2) developing black carbon assessments and particulate carbonaceous inventories; (3) estimating air toxic pollutant emissions from PM and organic gas primary emissions; (4) providing input to chemical mass balance (CMB) receptor models; and, (5) verifying profiles derived from ambient air measurements by multivariate receptor models (e.g., factor analysis and Positive Matrix Factorization).

EPA routinely uses SPECIATE data for development of air quality modeling platforms and for the National Emissions Inventory (NEI). For the NEI, SPECIATE data are used to estimate black carbon (which is assumed to be equivalent to elemental carbon) emissions as well as organic carbon, sulfate, and nitrate species of fine PM and to estimate hazardous air pollutants (HAPs) for some source categories. The SPECIATE database is also used as an input to HAP augmentation processes in EPA databases.

SPECIATE v3.2, released in 2002, was the first electronic version, a Microsoft Access® database. Periodically, EPA releases an updated version of SPECIATE that adds data to previous versions of the Microsoft Access® database. EPA also provides the data in a browser tool to allow users to browse and download profile information without the need to use Microsoft Access®. EPA is now releasing SPECIATE v5.4, both in Microsoft Access® format and in the SPECIATE browser. Figure ES-1 below shows the number of profiles in various releases of SPECIATE.

Figure ES-1. Number of profiles by SPECIATE version.



The development and update of the SPECIATE database is accomplished by a multi-office EPA SPECIATE Workgroup (SWG) comprised of staff from the Office of Research and Development (ORD), Office of Air and Radiation (OAR), and the contractor Abt Global.

The SWG members search for published data in reports and publications, select the data, evaluate, and quality assure the data and profiles. They also coordinate improvements to the database structure and metadata fields to support downstream needs such as mechanism mapping and new aerosol mechanisms. As newer SPECIATE versions are developed, improvements are made to the process as well as the data.

EPA generated SPECIATE v5.4 by appending one organic gas profile and one PM profile to the SPECIATE v5.3 database. In total, the SPECIATE v5.4 database includes 6,897 profiles. The new organic gas profile (95876) added to SPECIATE v5.4 is a composite profile represents swine farm and animal waste emissions. The new PM profile (95875) added to SPECIATE v5.4 is a composite tire wear PM profile. Both new profiles are based on peer-reviewed journal articles that SWG selected and decided they are critical for SPECIATE users.

Prior to SPECIATE v5.4, PM profiles added for air quality modeling are in PM-AE6, PM-AE8, or PM-CR1 model ready formats (chemical mechanism-specific profiles). Since EPA has developed the capability for the S2S-Tool ("SPECIATE to SMOKE", formerly the Speciation Tool) to create model ready format profiles from PM profiles, the SWG has decided to retain the original form of PM profile (e.g., 95875) added to SPECIATE v5.4 and any future PM profiles. This will allow SPECIATE developers to rely on the S2S-Tool to perform the calculations rather than having to do them manually and put the resulting model ready profile into the SPECIATE v5.4 database.

The SPECIATE v5.4 database also includes 23 new species in the SPECIES_PROPERTIES table. The field "PROFILE_INCLUDED_IN_2020_PLATFORM" in the PROFILES table is modified to "PROFILE_INCLUDED_IN_MODELING_PLATFORM" to indicate that a profile is included in the GSREF file that is part of the 2020 and/or 2022 modeling platform files. This modification allows accommodation for indicating future modeling platforms when they become available.

Chapter I of this addendum discusses the profiles added since SPECIATE v5.3, Chapter II provides a description of the database tables and field names, Chapter III lists potential considerations for the next version, and Chapter IV discusses quality assurance and data limitations.

SPECIATE WORKGROUP MEMBERS

EPA's SPECIATE program is made possible by the following organizations that fund and/or provide employee resources:

- EPA Center for Environmental Measurement and Modeling (CEMM)
- EPA Office of Air Quality Planning and Standards (OAQPS)
- EPA Office of Transportation and Air Quality (OTAQ)

The primary contact for the project is Dr. George Pouliot, the EPA Task Order Contracting Officer's Representative (TOCOR).

CONTENTS

LIST OF FIGURES AND TABLES

ACRONYMS AND ABBREVIATIONS

AESMD	Atmospheric and Environmental Systems Modeling Division, EPA/ORD
AMCD	Air Methods and Characterization Division, EPA/ORD
AQAD	Air Quality Assessment Division, EPA/OAR
CARB	California Air Resources Board
CAS	Chemical Abstracts Service
CEMM	Center for Environmental Measurement and Modeling, EPA/ORD
CMAQ	Community Multi-scale Air Quality Modeling System
CPHEA	Center for Public Health and Environmental Assessment, EPA/ORD
CROC	Condensed reactive organic carbon defined as saturation concentration < $10^{2.5}$ µg m^{-3}
CMB	chemical mass balance
CRACMM	Community Regional Atmospheric Chemistry Multiphase Mechanism
DRI	Desert Research Institute
EC	elemental carbon
EPA	Environmental Protection Agency
EPI	estimation program interface
GROC	Gaseous reactive organic carbon defined as saturation concentration > $10^{2.5}$ µg m^{-3}
HAPs	hazardous air pollutants
HDDV	heavy-duty diesel vehicle
HEEAD	Health and Environmental Effects Assessment Division, EPA/ORD
HEID	Health and Environmental Impacts Division, EPA/OAR
ID	identification
LDDV	light-duty diesel vehicle
MBO	metal-bound oxygen
MW	molecular weight
NEI	National Emissions Inventory
NMHC	non-methane hydrocarbons
NMOG	non-methane organic gas
OAQPS	Office of Air Quality Planning and Standards, EPA/OAR
OAR	Office of Air and Radiation, EPA
OC	organic carbon
OM	organic matter
OPERA	OPEn structure-activity/property Relationship App
ORD	Office of Research and Development, EPA
OTAQ	Office of Transportation and Air Quality, EPA
PFAS	per- and polyfluoroalkyl substances
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 ≤ 10 micrometers
$PM_{2.5}$	particulate matter with an aerodynamic diameter ≤ 2.5 micrometers
PNCOM	particulate non-carbon organic matter
POC	primary organic carbon
POM	particulate organic matter
SAROAD	Storage and Retrieval of Aerometric Data
SMOKE	Sparse Matrix Operator Kernel Emissions (EPA emissions modeling tool)
SOA	secondary organic aerosol

SRS	Substance Registry System
SVOC	semi-volatile organic compounds
SWG	SPECIATE Workgroup
TOCOR	Task Order Contracting Officer's Representative
TOG	total organic gases
VBS	volatility basis set
VCP	volatile chemical products
VOC	volatile organic compounds
XRF	X-ray fluorescence

CHAPTER I. SPECIATE v5.4 – Changes from SPECIATE v5.3

SPECIATE is EPA's repository of total organic gas and particulate matter (PM) speciation profiles of air pollution sources. Uses of these source profiles include: (1) creating speciated emissions inventories for regional haze, PM, and photochemical air quality modeling; (2) estimating hazardous and toxic air pollutant emissions from PM and primary emissions of organic gases; (3) providing input to chemical mass balance receptor models; and (4) verifying profiles derived from ambient measurements by multivariate receptor models. Current and previous publicly released versions of SPECIATE are available for download from the EPA's SPECIATE website. EPA also provides the current version of the database as a web-based application ("The SPECIATE Browser") designed using the Qlik platform. This web-based data browser allows users to view and filter profile data, including the weight percents of species or any metadata field, and export selected records into Excel workbooks. In addition, users of the browser can view profile weight percents for individual profiles using various charts for visualization and make comparisons across profiles.

The purpose of this addendum is to document changes made to the database since SPECIATE v5.3 and describe the tables, queries, and data fields in the database.

SPECIATE v5.4 includes three main changes from SPECIATE v5.3: A) the addition of profiles, B) new species used in the newly added profiles, and C) the modified field in the PROFILES table "PROFILE_INCLUDED_IN_MODELING_PLATFORM" to indicate that a profile is included in the GSREF file that is part of the 2020 and/or 2022 modeling platform files.

A. Profile additions

In the SPECIATE v5.4 database, there is one PM profile, and one organic gas profile added to the SPECIATE v5.4 database. Both new profiles are based on peer-reviewed journal articles that SWG selected and decided they are critical for SPECIATE users. The new PM profile (95875) is a composite tire wear PM profile. As the tailpipe emissions from the mobile sources are reducing, due to recent regulations and increasing population of hybrid and full electric vehicles, tire wear PM emissions have become a significant source of air pollution from motor vehicles. Therefore, SWG recommended adding the composite tire wear PM profile. The newly added organic gas profile (95876) is a composite profile represents swine farm and animal waste emissions. With the inclusion of new data in the compositing process, Profile 95876 added 85 new species that largely improves the detailed speciation of the previous Swine Farm and Animal Waste Profile 95241. In total, the SPECIATE v5.4 database includes 6,897 profiles.

Table 1. New Profiles Added to SPECIATE v5.4

PROFILE CODE	PROFILE_NAME	PROFILE TYPE
95875	Composite Tire Wear	PM
95876	Composite Swine Farm and Animal Waste	NMOG

CHAPTER II. SPECIATE v5.4 Database – Tables, Queries and Fields

This chapter describes the organization of the SPECIATE v5.4 database and presents the tables, queries, macros, and field names.

A. Key Tables

The key tables, listed in Table 2 below, provide the data in SPECIATE as well metadata about the profiles and links to supporting information used to develop the profiles. The fields in these tables are described in Section F of this Chapter.

Table 2. Key Tables in the SPECIATE v5.4 Database

Table	Description
PROFILES	Provides the metadata for the profiles, other than the references. One row per profile code.
PROFILE_REFERENCE _CROSSWALK	Provides each profile with one or more reference code(s) reflecting the references used for the profile. Allows more than one row per profile. Also allows the same reference code to be used for difference profiles (one reference can cover multiple profiles)
REFERENCES	Provides metadata and a cross walk for reference codes (paper citation, report, etc.). All references are included in this table. One row per reference.
SPECIES	Provides the species weight percent for each profile, as well as other information, where available, such as the emission rate, emission factors, etc. One row per profile/species combination.
SPECIES_PROPERTIES	Key table: identifiers (SPECIES_ID) for each of the pollutants in the database and metadata for each species.
SPECIES_SYNONYMS	Supporting data: contains SPECIES_NAMES and synonyms (where available) for each SPECIES_ID. Used by macro “mcrConcatenateSpecies” to produce the fields SPECIES_NAME that has a pipe delimited list of species synonyms for each Species ID.

B. Concatenated Tables

The tables listed in Table 3 provide SPECIATE reference information and SPECIES names in a format that supports the database browser and VIEW queries.

Table 3. Concatenated Tables in the SPECIATE v5.4 Database

Table	Description	Notes
tblProfileAndConcatRefs	Provides a profile reference crosswalk that concatenates multiple references for the same profile (when a profile has more than one reference) and produces/populates up to 3 link fields per profile. One row per profile. Includes several fields from the PROFILE table that do not need concatenation to make the table more useful as a standalone table	Created by macro mcrContenatedReferences. -The current approach used by the macro limits the number of references per profile to 3.
tblSpeciesAndConcatSynonyms	SPECIES_ID – to- multiple SPECIES_NAMES in a format that is one row per species. The SPECIES_NAMES has all names in the SPECIES_NAME field plus any others in the SPECIES_SYNONYMS table. It was created by concatenating all the synonyms for each SPECIES_ID. This table also includes the SPECIES_NAME field from the SPECIES_PROPERTIES table.	- Created by macro mcrConcatenateSpecies Concatenating all the species synonyms makes it easier to search in the browser and VIEW queries by species name.

C. Other Tables

There are several other tables as listed in Table 4 in the SPECIATE database that are used for reference. These include legacy tables that have been in previous versions of SPECIATE; they are kept with SPECIATE v5.4 for completeness.

Table 4. Other Tables in the SPECIATE v5.4 Database

Table	Description	Notes
List of SVOC Splitting Factors	Supporting table (legacy). This is an old table that is more of a toolbox from the Schauer profiles (Schauer et al, 1998; Schauer et al 1999) that was useful to determine phase. It is documented in the SPECIATE 5.0 Final Report, Appendix D.	The species names in this table may not exactly match those in SPECIES_PROPERTIES because TMS was removed from all the names.
MNEMONIC	A lookup table (legacy) that relates Desert Research Institute (DRI) profiles in SPECIATE to a DRI profile code and Chemical Mass Balance (CMB) model identifier.	
Oxide Forms	Supporting table: Provides oxide forms and oxygen to metal ratios used to compute metal-bound oxygen (MBO) needed for PM profile mass reconstruction.	The MBO computation is not done in the SPECIATE database, but rather in the MS Excel workbooks that contain underlying calculations of how weight percents are computed.
REVISION_LIST	Supporting table: Provides all the revisions made to the database.	Primarily for EPA SPECIATE developers to track/share changes to the database.
tblLastUpdated	Derived table: Provides the date that the derived tables tblProfileAndConcatRefs and tblSpeciesAndConcatSynonyms were last created. Produced by the same macros that create these tables.	

D. Queries

Queries in the database serve different functions including: 1) providing easy access to readable data summaries that contain descriptions and metadata; 2) providing input to the S2S-Tool, software that translates SPECIATE profiles into profiles that are chemical mechanism-specific and generates SMOKE input files; and 3) allowing users to perform quality assurance (QA). The queries summarized in Table 5 are primarily those that provide readable data summaries and inputs to the S2S-Tool.

Most VIEW queries provide simplified and complex views of the profile/species/weight percent information with different levels of metadata to allow a user to view the data with appropriate metadata fields (e.g., species names and profile names).

Table 5. Queries in the SPECIATE v5.4 Database

Query Name	Purpose
View GAS Profiles	View nearly all fields in SPECIATE for all gas profile types (GAS, GAS-VBS) in a flattened view; one row per profile/species
View PM Profiles	View nearly all fields in SPECIATE for all PM profile types (PM, PM-AE6, PM-AE8, PM-CR1, PM-Simplified) in a flattened view one row per profile/species
View Other Profiles	View nearly all fields in SPECIATE for profile type “OTHER” in a flattened view one row per profile/species

View Gas Profiles-Simplified	Same as View GAS Profiles but fewer fields; enables you to focus on the weight percent information without all the metadata
View PM Profiles-Simplified	Same as View PM Profiles but fewer fields; enables you to focus on the weight percent information without all the metadata
View Other Profiles-Simplified	Same as View Other Profiles but fewer fields; enables you to focus on the weight percent information without all the metadata
MasterReferenceListQuery	Provides more detailed reference information associated with each profile. Up to 3 rows per profile (if the profile has 3 references)
IdentifySpeciesDuplicates	Creates an extract of all duplicate rows. This is possible because of the naming convention "-duplicate" in the SPECIES_NAMES.
export_profiles	Query developed to extract select fields from the PROFILES table for use in the S2S-Tool.
export_species	Query developed to extract select fields from the SPECIES table for use in the S2S-Tool.
export_species_properties	Query developed to extract select fields from the SPECIES_PROPERTIES table for use in the S2S-Tool.
ProfilesMissingWeights	This informational/QA query lists all profiles that have missing WEIGHT_PERCENT values. The only profiles/species that should result from this query come from profiles with Profile_Type = "OTHER"
qryProfilesAfterDate	This informational query lists all profiles after a date entered by the user. Enter date as Month/Day/Year (e.g., 1/1/2020)
SPECIES_SYNONYM_SELECTOR	Used to support the PROFILE FORM for entering profile data into the database. This query allows species entry by the name in the SPECIES_PROPERTIES table or a synonym that is in the SPECIES_SYNONYMS table.

E. Macros

There are two macros in the SPECIATE database. These create tables that contain concatenated references and species names described in Table 5. The Macro "mcrContenatedReferences" creates the table called "tblProfileAndConcatRefs" and the macro "mcrConcatenateSpecies" creates the table "tblSpeciesAndConcatSynonyms."

F. Data Dictionary – Field names and Descriptions

Table 6 provides a list of fields in the SPECIATE database tables and Table 7 provides a list of fields in the supporting derived tables that are used in the VIEW queries and browser.

Table 6. Descriptive Data Dictionary

Field Name	Data Type	Description
PROFILES Table		
PROFILE_CODE	Text	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.
PROFILE_NAME	Text	Profile Name
PROFILE_TYPE	Text	Indicates type of profile: PM-AE6, PM-AE8, PM-CR1, PM-Simplified, PM, GAS, GAS-VBS and OTHER.
MASTER_POLLUTANT	Text	Indicates the pollutant being speciated
PROFILE_INCLUDED_IN_MODELING_PLATFORM	Text	Indicates that the profile is included in the GSREF file that is part of the 2020 and/or 2022 platform files
QSCORE	Number	Profile quality score out of 30 points total for measurement study. 22-30 = excellent. 16-21 = good. 8-15 = fair. 7 or less = poor. See Appendix A.

Field Name	Data Type	Description
QSCORE_DESC	Text	Description of the numeric QSCORE rating.
QUALITY	Text	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	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.
TEST_METHOD	Long Text	Description of sampling/test method for overall profile.
NORMALIZATION_BASIS	Text	Description of how profile was normalized.
ORIGINAL_COMPOSITE	Text	Specifies whether the profile is original, a composite of SPECIATE profiles, or study composite. Allowed values: 'O','C','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. At this time all data are 'YES'
INCLUDES_INORGANIC GAS	Yes/No	Indicates the presence or absence of inorganic gas species in the profile (e.g., sulfur dioxide, hydrogen sulfide, oxides of nitrogen, etc.)
TEST_YEAR	Text	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 for an explanation)
VINTAGE_RATING	Number	Vintage based on TEST_YEAR field (see Chapter II.D of the SPECIATE 5.0 document for an explanation)
DATA_QUANTITY_RATING	Number	Data sample size rating based on number of observations, robustness (see Chapter II.D of the SPECIATE 5.0 documentation for an explanation)
REGION	Text	Geographic region of relevance
SAMPLES	Text	Number of samples (separate experiments or measurements) used to make the profile.
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	GAS or PM Profile number taken from the same study, if exists
VERSION	Text	SPECIATE database version in which the profile was added
TOG_to_VOC RATIO	Number	Ratio of TOG mass to VOC mass, computed by either (1) or (2) (1) sum(all species%) / (sum(all species%) -sum(nonVOC)%) (2) sum(all species%) / sum (VOC species%)
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_LOADING_ug_per_m3	Number	Organic loading during sampling in units of micrograms/m^3
CATEGORY_LEVEL_1_Generation _Mechanism	Text	The mechanism by which emissions are generated by the emissions source. (See Appendix F of the SPECIATE 5.0 document for details)

Field Name	Data Type	Description
CATEGORY_LEVEL_2_Sector_Equipment	Text	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 SPECIATE 5.0 document for details)
CATEGORY_LEVEL_3_ Fuel_Product	Text	This category provides the highest level of detail for the profile categorization. (See Appendix F of the SPECIATE 5.0 document for details)
MASTER_POLLUTANT_EMISSION_RATE	Number	PM or GAS emission rate (emission factor), if available
MASTER_POLL_EMISSION_RATE_UNIT	Text	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, 1.7 for biomass combustion (other than wood fired boilers), 1.4 for wood fired boilers and all others, with some exceptions.
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 PM-AE6 profile would not exceed 100%
CREATED BY	Text	Person who added the profile
CREATED Date	Date/Time	Date the profile was added
MODIFIED BY	Text	Person who modified the profile
MODIFIED DATE	Date/Time	Date the profile was added
REVIEWED BY	Text	Person who reviewed the profile
REVIEWED DATE	Date/Time	Date the profile was reviewed
Data_Origin	Text	Origin of data. This is the same as DATA_ORIGIN which is in the KEYWORD_REFERENCE table in SPECIATE 5.0
Keywords	Text	List of ideas and topics that define the content of the profile. This is the same as KEYWORD which is in the KEYWORD_REFERENCE table in SPECIATE 5.0
DOC_LINK	Text	A link to an excel workbook showing how the profile was developed or a zip folder that contains documentation not readily available on the internet in addition to the workbook
Q_LINK	Text	A link to the QSCORE rating documentation for a profile/reference
SPECIES Table		
PROFILE_CODE	Text	Unique Identifier links to PROFILES table.
SPECIES_ID	Number	Species Identifier (Same as in SPECIES_PROPERTIES table)
WEIGHT_PERCENT	Number	Weight percent of pollutant (%)
UNCERTAINTY_PERCENT	Number	Uncertainty percent of pollutant (%)
UNCERTAINTY_METHOD	Long Text	Description of method used to calculate uncertainty
ANALYTICAL_METHOD	Text	Description of analytical method (e.g., X-ray fluorescence spectroscopy, ion chromatography)
INCLUDE_IN_SUM	Text	Indicates (Yes or No) 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) to be consistent with the S2S-Tool. The ionic form of Ca, Mg, K, Na, and Cl take precedence over the atomic form and the atomic form is used only if the ionic form is missing.

Field Name	Data Type	Description
PHASE	Text	Indicates 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	Species emission rate units (e.g., mg/mile)
PROFILE_REFERENCE_CROSS WALK table		
PROFILE_CODE	Text	Unique Identifier links to PROFILES table.
REF_Code	Text	Code representing a unique SPECIATE Reference. That reference could be a paper from the literature, report, memorandum, personal communication or other for which there is a reference in the REFERENCE table
REFERENCES table		
REF_Code	Text	Code representing a unique SPECIATE Reference. That reference could be a paper from the literature, report, memorandum, personal communication or other for which there is a reference in the REFERENCE table
REFERENCE	Long Text	Complete reference citation including a Digital Object Identifier (DOI), where available
REF_DESCRIPTION	Long Text	Stores the descriptive information about the reference.
LINK	Hyperlink	Hyperlink to the reference (or abstract if it is under copywrite)
SPECIES_PROPERTIES Table		
SPECIES_ID	Number	Unique Identifier for a speciated compound or mixture (Species)
CAS	Text	Chemical Abstracts Service (CAS) number associated with the species (with hyphens) (blank if no CAS)
CAS no hyphen	Text	Same as the CAS, without the hyphen
ALT_CAS	Text	This is used when there are multiple CAS or a CAS was changed (retired) and is no longer used. There may be more than one ALT_CAS, and if so, they are separated by a semicolon
DSSTox_ID		Unique Identifier for a chemical in EPA's Distributed Structure-Searchable Toxicity (DSSTox) Database
SAROAD	Text	Storage and Retrieval of Aerometric Data (SAROAD) code
SRS ID	Text	EPA Substance Registry Service (SRS) Chemical Identifier
NonVOCTOG	Yes/No	Is this species regarded as a volatile organic compound (VOC)? The VOC definition is from 40 CFR. §51.100
PAMS	Yes/No	Is PAMS pollutant? (Yes or No)
HAPS	Yes/No	Is Hazardous Air Pollutant (HAP)? (Yes or No) HAPs are defined 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.
SPECIES_NAME	Text	Species Name
REPRESENTATIVE_COMPOUND_NAME	Text	Representative compound structures are assigned for all compounds in the SPECIATE database (Pye et. al., 2023). Representative structures are listed by name and linked to a DTXSID identifier. In other cases, a representative or major compound in a mixture (e.g., vinegar or crude oil) is listed.
REPRESENTATIVE_COMPOUND_DSSTox_ID	Text	The DSSTox Substance Identifier for the representative compound available in the EPA Chemicals Dashboard (https://comptox.epa.gov/dashboard/).

Field Name	Data Type	Description
SYMBOL	Text	Standard chemical abbreviation
SPEC_MW	Number	Species molecular weight
NOTE	Long Text	Notes about the SPECIES_ID or its properties
Molecular Formula	Text	Molecular formula
OXYGEN_to_CARBON_RATIO	Number	Ratio of oxygen atoms to carbon atoms
Smiles Notation	Text	Smiles notation of the REPRESENTATIVE_COMPOUND_NAME.
VP_Pascal_EPI	Number	Vapor Pressure in units of Pascals from the EPISUITE model (recommended by SWG member Ben Murphy, EPA/ORD/NERL)
VP_Pascal_UM	Number	Vapor Pressure in units of Pascals from UManSysProp tool (uses the EVAPORATION algorithm, slightly updated) http://umansysprop.seaes.manchester.ac.uk/tool/vapour_pressure
VP_Pascal_OPERA	Number	Vapor pressure, in Pascal, as predicted by the OPERA model (DOI: 10.1186/s13321-018-0263-1)
ATMOSPHERIC_HYDROXYLATION_RATE_(AOH)_CM3/MOLECULE-SEC_OPERA	Number	Atmospheric hydroxylation rate, in cm^3 $molec^{-1}$ sec^{-1}, as predicted by the OPERA model.
HENRYS_LAW_ATM-M3/MOLE_OPERA	Number	Henry's Law, in atm m^3 $mole^{-1}$, as predicted by the OPERA model.
OCTANOL_AIR_PARTITION_COEFF_LOGKOA_OPERA	Number	Octanol-Air partitioning coefficient, which is unitless, as predicted by the OPERA model.
OCTANOL_WATER_PARTITION_LOGP_OPERA	Number	Octanol-water partitioning coefficient, which is unitless, as predicted by the OPERA model.
Duplicate_ID	Text	Indicates whether this compound is the same as one covered by a different SPECIES_ID and the lowest number of the SPECIES_ID belonging to this duplicate pair.
REPCMPSCORE	Number	A score indicating degree of confidence in the assignment of the representative compound to the SPECIATE species (high=most confident). See additional below.*
SPECIES_SYNONYMS table		
ID1		Unique value for this table
SPECIES_ID	Number	Species Identifier (Same as in SPECIES_PROPERTIES table)
Descriptor	Text	Species name or synonym
Convention	Text	Originally set up to provide origin of the synonym, but the value is "Preexisting" for all data

*REPCMPSCORE: A score indicating the level of confidence in the representative compound assignment:

- 4 - Species is already fully specified in database and representative compound is an exact match to SPECIATE. The representative compound is the exact species.
- 3 - Species is matched in the dashboard, but there are inconsistencies in the representative compound properties and some SPECIATE fields. SPECIATE fields updated to match the representative compound.
- 2 - Species does not have a DTSXID or automated match by name in the dashboard. The species is manually matched to a representative species based on some information. For example, C12 alkanes are matched to dodecane, terpenes are mapped to d-limonene.
- 1 - Species does not have a DTSXID or automated match by name and the species is nebulously defined, possibly a mixture. The species is manually mapped with a low degree of certainty. For

example, "aggregated VOCs" is mapped to decane. These mappings may be reexamined in the future if they are significant contributors by mass.

- 0 - Not mapped. This value is only temporarily used in development versions of SPECIATE.
- Note: For highly uncertain VOCs, decane is the recommended default surrogate. Decane (C10) falls within the standard VOC range and represents a compound that is likely to be collected during VOC measurement. Unidentified gas-phase compounds (from the UCM) tend to fall in the IVOC range. Decane falls at the less volatile end of the VOC range and its ozone formation is a conservative (low) estimate. Since decane is more volatile than IVOCs its SOA formation is also conservative (low). Species between C8 and C12 would likely be acceptable surrogates as well. Species receiving this treatment should have a REPCMPSCORE of 1.

Table 7. Fields Used to Support the VIEW Queries and Browser

Field Name	Data Type	Description
tblProfileAndConcatRefs Table		
PROFILE_CODE	Text	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.
PROFILE_NAME	Text	Profile Name
VERSION	Text	SPECIATE database version in which profile was added
PROFILE_TYPE	Text	Indicates type of profile: PM-AE6, PM-AE8, PM-CR1, PM-Simplified, PM, GAS, and OTHER
Data_Origin	Text	Origin of data. This is the same as DATA_ORIGIN in the KEYWORD_REFERENCE table in SPECIATE 5.0.
REF_Codes	Text	Indicates the pollutant to be used in the calculation.
REFERENCES	Long Text	Concatenation of each REFERENCE for the profile. Each reference is separated by a carriage return. Can have up to 3 references per profile.
REF_DESCRIPTIONS	Long Text	Concatenation of each REF_DESCRIPTION for each reference for the profile. Each REF_DESCRIPTION is separated by a carriage return. Can have up to 3 references per profile.
Keywords	Text	List of ideas and topics that define the content of the profile. This is the same as KEYWORD in the KEYWORD_REFERENCE table in SPECIATE 5.0
LINK1	Hyperlink	Hyperlink to the documentation for the first reference
LINK2	Hyperlink	Hyperlink to the documentation for the second reference
LINK3	Hyperlink	Hyperlink to the documentation for the third reference
tblSpeciesAndConcatSynonyms Table		
SPECIES_ID	Number	Species Identifier (Same as in SPECIES_PROPERTIES table)
SPECIES_NAME	Text	Species Name
SPECIES_NAMES	Long Text	Concatenation of all synonyms in the Synonyms table for the same SPECIES_ID

CHAPTER III. Future Considerations for the Next Version

The current plan is to release a new version of SPECIATE within two years. In the next release, planned enhancements to the SPECIATE program and database include:

- Adding more high-priority profiles that can lead to better performance in air quality modeling and potentially improve the NEI for HAPs;
- Adding relevant species properties to the database such as atmospheric photochemical reactivity to form ozone and estimates of secondary organic aerosol yields;
- Conducting literature searches for PM, organic gas, and mercury speciation;
- Reaching out to the research community for high quality and high priority data to include in SPECIATE; and
- Maintaining/updating the standard operating procedures for developing and enhancing SPECIATE, as needed.

CHAPTER IV. Quality Assurance and Data Limitations

- This work was conducted under the Agency's Quality Assurance (QA) program for environmental information, with an approved Quality Assurance Project Plan for *SPECIATE Version 5.2 and later* (J-AESMD-ESAB-C-0-60, approved 12/2021). Independent QA audits were not deemed necessary.
- This addendum has been reviewed by the ORD/CEMM/AESMD Quality Assurance Manager and it has been determined to be consistent with EPA Category B quality assurance requirements.
- All existing data sets considered for inclusion in the database were reviewed by the SPECIATE team following the *SOP for Getting Data into SPECIATE* (J-AESMD-0016020-OP-1-0, approved 5/2019).
- Any limitations to the data contained in the database are clearly stated within the database itself. Each profile is assigned a set of ratings described in Chapter II Section D of the SPECIATE 5.0 Final Report (EPA/600/R-19/098) The limitations are captured with the following fields: Vintage Rating, Data Quantity Rating, Quality Rating (QSCORE), and Judgement Rating.

REFERENCES

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

Pye, H.O.T., Place, B.K., Murphy, B.N., Seltzer, K.M., D'Ambro, E.L., Allen, C., Piletic, I., Farrell, S., Schwantes, R.H., Coggon, M.M., Saunders, E., Xu, L., Sarwar, G., Hutzell, B., Foley, K.M., Pouliot, G., Stockwell, W.R. Linking gas, particulate, and toxic endpoints to air emissions in the Community Regional Atmospheric Chemistry Multiphase Mechanism (CRACMM), Atmos. Chem. Phys., 23, 5043–5099, https://doi.org/10.5194/acp-23-5043-2023, 2023.

Murphy, B. N., Sonntag, D., Seltzer, K. M., Pye, H. O. T., Allen, C., Murray, E., Toro, C., Gentner, D. R., Huang, C., Jathar, S. H., Li, L., May, A. A., and Robinson, A. L.: Reactive Organic Carbon Air Emissions from Mobile Sources in the United States, Atmos. Chem. Phys., 23, 13469–13483, 2023 https://doi.org/10.5194/acp-23-13469-2023.

APPENDIX A. Profile Quality Criteria Evaluation

The Quality Score (QSCORE) provides 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.

The evaluation framework serves as a guide for reviewers to assign quality value points to the areas of a study deemed most important for use in SPECIATE. The framework is comprehensive, but also easy to understand and apply. Points are assigned based on answers to a series of questions and the sum of the points is the evaluation score. An ideal evaluation score would have 30 (Data from Measurements) or 29 (Data from other Methods) desired criteria (points). Points are additive, influencing, but not necessarily distinguishing the study. Highly ranked publications or reports receive priority for inclusion in the SPECIATE database. The QSCORE total points are valued as follows:

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

Each ranking is added to the workbook developed for the profile. The QSCORE documents are stored on SharePoint and the ftp site. The Q_Link field in the PROFILES table provides the link to the document on the ftp site.

To provide the hyperlink, the folder will change with the version. The base is: 'ftp://gaftp.epa.gov/air/emismod/SPECIATE_supportingdata/' then the version and file name. The folder for the current SPECIATE database versions is 5_4, so the link would be:
ftp://gaftp.epa.gov/air/emismod/SPECIATE_supportingdata/v5_4 /*name of Q score file
Composite profiles developed from existing profiles in SPECIATE may not always have a QSCORE. If there is a paper (i.e., Reff, et. al., 2009) that documents the development of the composites, then that paper is QSCORED using the "DATA FROM OTHER METHODS (Blended) which produces a maximum score of 29). If there is not a paper and all profiles being composited have the same QSCORE (e.g., CARB composites) then the composite can take the value of the individual profiles (e.g., CARB consumer products based on the 2010 survey update that were added to SPECIATE 5.0). If there is no paper and the profiles being composited have different QSCORES (SUG03 profile built from SUG01 and SUG02) then the QSCORE is left NULL.

Table A1. Evaluation Questions and Assigned Points for the QSCORE Evaluation of "Measured Source Profiles". The Maximum Score is 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? (how easy to translate the data from the paper to SPECIATE-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/$NH4^+$ ions; (3) metals/inorganics. Higher scores are given if PAHs and SVOCs are also available. Is there complete speciation data of Hg? Hg should include: (1) Elemental mercury (Hg^0) (2) Reactive Gas mercury (a.k.a. ionic) (3) Particulate form 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	1-3

	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. 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)

Separate QSCORE evaluation criteria were developed for data collected using other methods. Other methods include any paper where the researchers did not directly measure what they report in the paper. Examples of other methods include: Urbanski 2014 (compiling others' work) and profile for flares (FLR99) that estimated the composition from a test of propylene (Allen et al 2011).

Table A2. Evaluation Questions and Assigned Points for a "Data from Other Methods - Blended" QSCORE Evaluation. The Maximum Score is 29.

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/$NH4^+$ ions; (3) metals/inorganics. Higher scores are given if PAHs and SVOCs are also available. Hg should include:	1-10

	(1) Elemental mercury (Hg^0) (2) Reactive Gas mercury (a.k.a. ionic) (3) Particulate form Scoring guidance for Hg profiles: One species=2, Two species=6, all three species=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

APPENDIX B. The PM Protocol

Background and Purpose

SPECIATE is EPA's repository of total organic gas and particulate matter (PM) speciation profiles of air pollution sources. Some of the many uses of these source profiles include: (1) creating speciated emissions inventories; (2) estimating hazardous and toxic air pollutant emissions from PM and primary emissions of organic gases; (3) providing input to chemical mass balance receptor models; and (4) verifying profiles derived from ambient measurements by multivariate receptor models (e.g., factor analysis and positive matrix factorization). Here, methods for processing raw PM2.5 profile data into a format suitable for import into SPECIATE and use in the Community Multiscale Air Quality (CMAQ) photochemical air quality modeling are described.

PROFILE_TYPE options for PM2.5 profiles within SPECIATE include PM-AE6, PM-AE8, PM-CR1, PM-Simplified, and PM. PM-AE6 profiles are post-processed PM2.5 profiles developed using the "AE6-Protocol" (Reff et al., 2009) and were first introduced in SPECIATE v4.3. These profiles include species that directly map to tracers used by the AERO6 (AE6) module of CMAQ. Except for the organic matter components, PM-AE6, PM-CR1 are operationally the same. Mathematically, the relationship among the organic matter components for these three profile types is as follows:

$$POC + PNCOM\ (PM - AE6\ species)$$

$$= ROCN2 + ROCN1 + ROCP0 + ROCP1 + ROCP2\ (PM - CR1\ species)$$

In PM-CR1, both POC and PNCOM are split among several tracers that vary by volatility. Each volatility bin spans an order of magnitude, and the value is represented by a letter and number. The letter "N" indicates negative, and "P" indicates positive. The number following "N" or "P" indicates the log base-10 value. For example, N2 = 10-2 and P1 = 101. In PM-CR1 profiles, the organic carbon and non-carbon organic mass are summed together (and thus represented as organic matter; OM). PM-Simplified profiles are limited to five components: nitrate (613), OC (626), sulfate (699), EC (797), and PM other (1884). These components are aggregated from a profile that has been post-processed into a PM-AE6 profile. PM profiles include other, non-mechanism-specific components, such as explicit semi-volatiles and polycyclic aromatic hydrocarbons. Here, the PM Protocol describes the methods used to generate profiles suitable for import into SPECIATE, as well as the post-processing steps necessary to generate PM-AE6 and PM-CR1 profiles.

"PM-Ready" Species

The S2S-Tool, and formerly the Speciation Tool, translates SPECIATE profiles into profiles that are chemical mechanism-specific and generates several files that are input into SMOKE. For each PM2.5 PROFILE_TYPE, the S2S-Tool maps SPECIATE species to chemical-mechanism specific species and allocates all remaining mass to an "Other" Category. This functionality allows SPECIATE developers to include more species than are contained within aerosol modules.

Below, all "PM-ready" species in SPECIATE are listed. These species are the only entries from a profile that are used when generating a GSPRO when MECH_BASIS is PM-AE6 or PM-CR1. In other words, if additional species are included in a PM profile (e.g., polycyclic aromatic hydrocarbons), those species and their weight percent are not included in subsequent calculations. It should be noted that this list includes additional species that are included in the calculation of "Other" PM2.5 mass and are not explicit in chemical mechanisms or modules used by the U.S. EPA. Please note that for some species (calcium, magnesium, potassium, sodium, and chloride), there is an order-of-operations for species selection. In some studies, the atomic form of a metal is measured, but not the ionic form. The atomic form results from the use of x-ray fluorescence (XRF) as the measurement technique and the ionic form results from the use of ion chromatography.

It should be noted that for a profile to be used in air quality modeling using the AE6 mechanism, the profile must have either PH_2O or PNCOM. This is a requirement of the S2S-Tool which prepares the PM-AE6 speciation profiles in SPECIATE for SMOKE.

Table B1. PM Components to Include in a PM-AE6 Profile

Species Name	SPECIES_ID	Description
POC	626	Organic Carbon
PEC	797	Elemental Carbon
PSO4	699	Sulfate
PNO3	613	Nitrate
PNH4	784	Ammonium
PNCOM	2669	Non-Carbon Organic Matter
PFE	488	Iron
PAL	292	Aluminum
PSI	694	Silicon
PTI	715	Titanium
PCA	2303*	Calcium
PMG	2772*	Magnesium
PK	2302*	Potassium
PMN	526	Manganese
PNA	785*	Sodium
PCL	337*	Chloride
PH2O	2668	Particulate Water
PMOTHR	2671	Other PM
POCP2	3245	POCP2, C* = $1e^2$ ug m^{-3}
POCP1	3246	POCP1, C* = $1e^1$ ug m^{-3}
POCP0	3247	POCP0, C* = $1e^0$ ug m^{-3}
POCN1	3248	POCN1, C* = $1e^{-1}$ ug m^{-3}
POCN2	3249	POCN2, C* = $1e^{-2}$ ug m^{-3}
PNCOMP2	3250	PNCOMP2, C* = $1e^2$ ug m^{-3}
PNCOMP1	3251	PNCOMP1, C* = $1e^1$ ug m^{-3}
PNCOMP0	3252	PNCOMP0, C* = $1e^0$ ug m^{-3}
PNCOMN1	3253	PNCOMN1, C* = $1e^{-1}$ ug m^{-3}
PNCOMN2	3254	PNCOMN2, C* = $1e^{-2}$ ug m^{-3}
ROCN1ARO	3331	ROCN1ARO, Single-Ring Aromatics, C* = $1e^{-1}$ ug m^{-3}
ROCP0ARO	3332	ROCP0ARO, Single-Ring Aromatics, C* = $1e^0$ ug m^{-3}
ROCP1ARO	3333	ROCP1ARO, Single-Ring Aromatics, C* = $1e^1$ ug m^{-3}
ROCP2ARO	3334	ROCP2ARO, Single-Ring Aromatics, C* = $1e^2$ ug m^{-3}
ROCN2PAH	3341	ROCN2PAH, PAH, C* = $1e^{-2}$ ug m^{-3}
ROCN1PAH	3342	ROCN1PAH, PAH, C* = $1e^{-1}$ ug m^{-3}
ROCP0PAH	3343	ROCP0PAH, PAH, C* = $1e^0$ ug m^{-3}
ROCP1PAH	3344	ROCP1PAH, PAH, C* = $1e^1$ ug m^{-3}
ROCP2PAH	3345	ROCP2PAH, PAH, C* = $1e^2$ ug m^{-3}
ROCN2SULF	3350	ROCN2SULF, Sulfur-containing Hydrocarbons, C* = $1e^{-2}$ ug m^{-3}
ROCN1SULF	3351	ROCN1SULF, Sulfur-containing Hydrocarbons, C* = $1e^{-1}$ ug m^{-3}
ROCP0SULF	3352	ROCP0SULF, Sulfur-containing Hydrocarbons, C* = $1e^0$ ug m^{-3}
ROCP1SULF	3353	ROCP1SULF, Sulfur-containing Hydrocarbons, C* = $1e^1$ ug m^{-3}
ROCP2SULF	3354	ROCP2SULF, Sulfur-containing Hydrocarbons, C* = $1e^2$ ug m^{-3}
ROCN2OXY	3361	ROCN2OXY, Oxygenated Organics, C* = $1e^{-2}$ µg m^{-3}
ROCN1OXY	3362	ROCN1OXY, Oxygenated Organics, C* = $1e^{-1}$ µg m^{-3}
ROCP0OXY	3363	ROCP0OXY, Oxygenated Organics, C* = $1e^0$ µg m^{-3}
ROCP1OXY	3364	ROCP1OXY, Oxygenated Organics, C* = $1e^1$ µg m^{-3}

Species Name	SPECIES_ID	Description
ROCP2OXY	3365	ROCP2OXY, Oxygenated Organics, C* = $1e^{2}$ µg m^{-3}
ROCN2CYC	3372	ROCN2CYC, Cyclic Hydrocarbons, C* = $1e^{-2}$ µg m^{-3}
ROCN1CYC	3373	ROCN1CYC, Cyclic Hydrocarbons, C* = $1e^{-1}$ µg m^{-3}
ROCP0CYC	3374	ROCP0CYC, Cyclic Hydrocarbons, C* = $1e^{0}$ µg m^{-3}
ROCP1CYC	3375	ROCP1CYC, Cyclic Hydrocarbons, C* = $1e^{1}$ µg m^{-3}
ROCP2CYC	3376	ROCP2CYC, Cyclic Hydrocarbons, C* = $1e^{2}$ µg m^{-3}
ROCN2BRN	3383	ROCN2BRN, Branched Hydrocarbons, C* = $1e^{-2}$ µg m^{-3}
ROCN1BRN	3384	ROCN1BRN, Branched Hydrocarbons, C* = $1e^{-1}$ µg m^{-3}
ROCP0BRN	3385	ROCP0BRN, Branched Hydrocarbons, C* = $1e^{0}$ µg m^{-3}
ROCP1BRN	3386	ROCP1BRN, Branched Hydrocarbons, C* = $1e^{1}$ µg m^{-3}
ROCP2BRN	3387	ROCP2BRN, Branched Hydrocarbons, C* = $1e^{2}$ µg m^{-3}
ROCN2ALK	3394	ROCN2ALK, Linear Hydrocarbons, C* = $1e^{-2}$ µg m^{-3}
ROCN1ALK	3395	ROCN1ALK, Linear Hydrocarbons, C* = $1e^{-1}$ µg m^{-3}
ROCP0ALK	3396	ROCP0ALK, Linear Hydrocarbons, C* = $1e^{0}$ µg m^{-3}
ROCP1ALK	3397	ROCP1ALK, Linear Hydrocarbons, C* = $1e^{1}$ µg m^{-3}
ROCP2ALK	3398	ROCP2ALK, Linear Hydrocarbons, C* = $1e^{2}$ µg m^{-3}
--	666	Phosphorus
--	767	Vanadium
--	347	Chromium
--	379	Cobalt
--	612	Nickel
--	380	Copper
--	778	Zinc
--	468	Gallium
--	298	Arsenic
--	693	Selenium
--	689	Rubidium
--	397	Strontium
--	779	Zirconium
--	586	Molybdenum
--	649	Palladium
--	695	Silver
--	328	Cadmium
--	487	Indium
--	714	Tin
--	296	Antimony
--	300	Barium
--	519	Lanthanum
--	1861	Cerium
--	528	Mercury
--	520	Lead

"PM-Ready" species. These are the exclusive species that can be included in profiles whose PROFILE_TYPE is PM-AE6 and PM-CR1.

*** For calcium, if a profile does not have 2303 but does have 329, 329 should be used. For magnesium, if a profile does not have 2772 but does have 525, 525 should be used. For potassium, if a profile does not have 2302 but does have 669, 669 should be used. For sodium, if a profile does not have 785 but does have 696, 696 should be used. For chloride, if a profile does not have 337 but does have 795, 795 should be used.**

It should be noted that for a profile to be used in air quality modeling using the AE6 mechanism, the profile must have either PH2O or PNCOM.

Instructions for creating "PM-Ready," PM-AE6, and PM-CR1 profiles for SPECIATE

Step 1 – Read the reference (i.e., paper or report) and supplemental information carefully to get the mass fraction information and determine if some species should not be included. Note the measurement methods and whether the source is controlled.

Step 2: Map species in the reference to SPECIATE species using the appropriate SPECIES_IDs.

Step 3: Determine if OC needs to be adjusted due to "artifacts." Artifacts may include volatile components that condense in the sampler. These should not be counted as PM because they are in the gas phase and are not emitted from the source as condensed PM.

A non-zero back up filter measurement does provide evidence for positive artifacts and can be quantitatively used to adjust measurements by subtracting the backup from the primary filter. It is possible that some of the mass on the back-up could be mass desorbed from the primary filter.

If a quantitative estimate of "true" OC or an adjustment to compute it is provided in the paper, then that should be used to adjust OC. If neither are available, "engineering judgement" may be used to estimate "true" OC as the difference between the primary and secondary filter measurements.

If there is no adjustment provided or too uncertain, and there appear to be artifacts, then OC can be adjusted later if the mass exceeds 100% after adding in the other relevant species (e.g., PH2O) that may not be contained in the paper.

Step 4: Add particulate water, PH2O (SPECIES_ID = 2668), per methods outlined in Reff et. al., 2009.

Source Type	Particulate Water (PH_2O) calculation
Combustion and other high temperature sources.	0%
All other sources	24% of the sum of PSO4 and PNH4.

Particulate water calculations.

Sources for which it is assumed that there is no particulate water include: Agricultural Burning, Bituminous Combustion, Calcium Carbide Furnace, Charbroiling, Charcoal Manufacturing, Distillate Oil Combustion, Electric Arc Furnace, Ferromanganese Furnace, Glass Furnace, HDDV Exhaust, Heat Treating, Kraft Recovery Furnace, LDDV Exhaust, Lignite Combustion, Lime Kiln, Meat Frying, Natural Gas Combustion, Nonroad Gasoline Exhaust, Onroad Gasoline Exhaust, Open Hearth Furnace, Prescribed Burning, Process Gas Combustion, Pulp & Paper Mills, Residential Coal Combustion, Residential Natural Gas Combustion, Residential Wood Combustion, Residual Oil Combustion, Sintering Furnace, Slash Burning, Sludge Combustion, Solid Waste Combustion, Sub-Bituminous Combustion, Wildfires, and Wood Fired Boiler.

Step 5: For ammonium sulfate production or ammonium nitrate production, add ammonium per methods outlined in Reff et. al., 2009. These are computed stoichiometrically assuming $(NH_4)_2SO_4$ for ammonium sulfate production and NH_4NO_3 for ammonium nitrate production.

If ammonium is computed, document it in the NOTES field of the SPECIATE database.

Step 6: Ensure consistency between sulfate and sulfur. If a profile has sulfate and not sulfur, the sulfur does not need to be computed. However, if it has sulfur and not sulfate, it should be computed as follows:

$$SO_4 = \left(\frac{96}{32}\right) * S$$

If sulfate is computed, document it in the NOTES field of the SPECIATE database.

Step 7: Add metal bound oxygen, MO (SPECIES_ID = 2670), per methods outlined in Reff et. al., 2009.

The approach stoichiometrically combines oxygen with the measured metals and then adjusts the MO downward based on the amount of available sulfate in the profile. This approach assumes that the sulfates bind to the metals preferentially over the oxygen. A modest change from Reff et. al., 2009 is to only use the difference between the atomic and ionic masses for Na, Ca, Mg and K, since the ionic version would not be the portion bound to oxygen.

Unadjusted MO is computed as follows:

$$\boldsymbol{MO_{unadjusted} = \sum_{i}^{N} Ox_i \times E_i}$$

where Ox_i is the oxygen-to-metal ratio for metal *i* (see Table B2) and E_i is the weight percent of metal *i*, except for Na, Ca, Mg and K. For these metals, the E_i should reflect the difference between the atomic and ionic forms of the metal. If a profile has only one reported (i.e., only atomic *or* ionic) value for of Na, Ca, Mg, or K, then the E_i should be set to 0. In addition, if the difference is negative, it should be set to 0.

To adjust MO based on the assumed preferential binding to sulfate over oxygen, the mass of SO_4 in the profile following neutralization with NH_4 is calculated as follows:

$$\boldsymbol{Neutralized\ SO_4 = \frac{0.5 * 96}{18} \times\ E_{NH_4}}$$

where E_{NH_4} is the weight percent of NH_4 in the profile. The non-neutralized sulfate is then calculated as follows:

$$\boldsymbol{Non - Neutralized\ SO_4 =\ E_{SO_4} -\ Neutralized\ SO_4}$$

If $Non - Neutralized\ SO_4 < 0$,

$$\boldsymbol{MO_{adjusted} =\ MO_{unadjusted}}$$

If $Non - Neutralized\ SO_4 > 0$

$$\boldsymbol{MO_{adjusted} =\ MO_{unadjusted} -\ Non - Neutralized\ SO_4 \times \frac{16}{96}}$$

$$\boldsymbol{If\ MO_{adjusted} < 0,\ MO_{adjusted} = 0}$$

Table B2: Assumed Oxide Forms of Each Metal and Resulting Mean Oxygen-to-Metal Ratio

Species	MW	Oxide Form 1	Oxide Form 2	Oxide Form 3	Oxygen/Metal Ratio
Na	22.99	Na_2O			0.348
Mg	24.31	MgO			0.658
Al	26.98	Al_2O_3			0.889
Si	28.09	SiO_2			1.139
P	30.97	P_2O_3	P_2O_5		1.033
K	39.10	K_2O			0.205
Ca	40.08	CaO			0.399
Ti	47.87	TiO_2			0.669

Species	MW	Oxide Form 1	Oxide Form 2	Oxide Form 3	Oxygen/Metal Ratio
V	50.94	V_2O_5			0.785
Cr	52.00	Cr_2O_3	CrO_3		0.692
Mn	54.94	MnO	MnO_2	Mn_2O_7	0.631
Fe	55.85	FeO	Fe_2O_3		0.358
Co	58.93	CoO	Co_2O_3		0.339
Ni	58.69	NiO			0.273
Cu	63.55	CuO			0.252
Zn	65.39	ZnO			0.245
Ga	69.72	Ga_2O_3			0.344
As	74.92	As_2O_3	As_2O_5		0.427
Se	78.96	SeO	SeO_2	SeO_3	0.405
Rb	85.47	Rb_2O			0.094
Sr	87.62	SrO			0.183
Zr	91.22	ZrO_2			0.351
Mo	95.94	MoO_2	MoO_3		0.417
Pd	106.42	PdO	PdO_2		0.226
Ag	107.87	Ag_2O			0.074
Cd	112.41	CdO			0.142
In	114.82	In_2O_3			0.209
Sn	118.71	SnO	SnO_2		0.202
Sb	121.76	Sb_2O_3	Sb_2O_5		0.263
Ba	137.33	BaO			0.117
La	138.91	La_2O_3			0.173
Ce	140.12	Ce_2O_3	CeO_2		0.200
Hg	200.59	Hg_2O	HgO		0.060
Pb	207.20	PbO	PbO_2		0.116

Oxygen-to-metal ratio for "PM-Ready" metals.

Step 8: If the profile has POC, add particulate non-carbon organic matter (PNCOM). If available, the value reported in the reference paper should be used. Otherwise, "engineering judgement" may be used to estimate the OM/OC ratio or the user can apply the default values provided below. This value should propagate to the ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO field in SPECIATE. In addition, indicate in the NOTES field how PNCOM was computed.

Source Type	PNCOM Computation	OM-to-OC Ratio
Onroad and Nonroad motor vehicle exhaust profiles.	$PNCOM = 0.25 * POC$	1.25
Wood combustion sources other than wood-fired boilers (e.g., wildfires).	$PNCOM = 0.7 * POC$	1.7
All other sources, including wood-fired boilers.	$PNCOM = 0.4 * POC$	1.4

Source-specific, default OM-to-OC ratios.

Step 9: Sum all "PM-ready" species (see Table 24). If summation is less than 100%, assign all remaining mass to PMOTHR (SPECIES_ID = 2671). If summation is 100%, no adjustments are necessary. If the summation is greater than 100%, then the following steps should be completed.

1) Double check the paper to see if there are reported POC artifacts. If the paper does not quantitatively report this information, adjust POC and PNCOM down by the same multiplier until the sum of weight fractions is 100%. If POC and PNCOM are scaled to zero and the profiles is still greater than 100%, then adjust all species down to get the sum to be 100%.

2) If POC artifacts have already been corrected for, then adjust all species down (i.e., normalize all weight percents) to get the sum to be 100%.

Note that sulfur should not be included in the above calculations if sulfate is available.

Step 10: If the profile has POC and this mass can be resolved by volatility, the mass assigned to POC and PNCOM should be summed and reallocated to the appropriate volatility-resolved species. Preferentially, this mass should be allocated to SPECIES_ID species 3394 – 3398, which would make the PROFILE_TYPE = PM-CR1.