Addendum

SPECIATE Version 5.2
Database Development Documentation

August 2022 **EPA/600/X-22/181/**

EPA SPECIATE Workgroup
US Environmental Protection Agency

Abt Associates
EPA Contract No. 68HERD21A002

Executive Summary

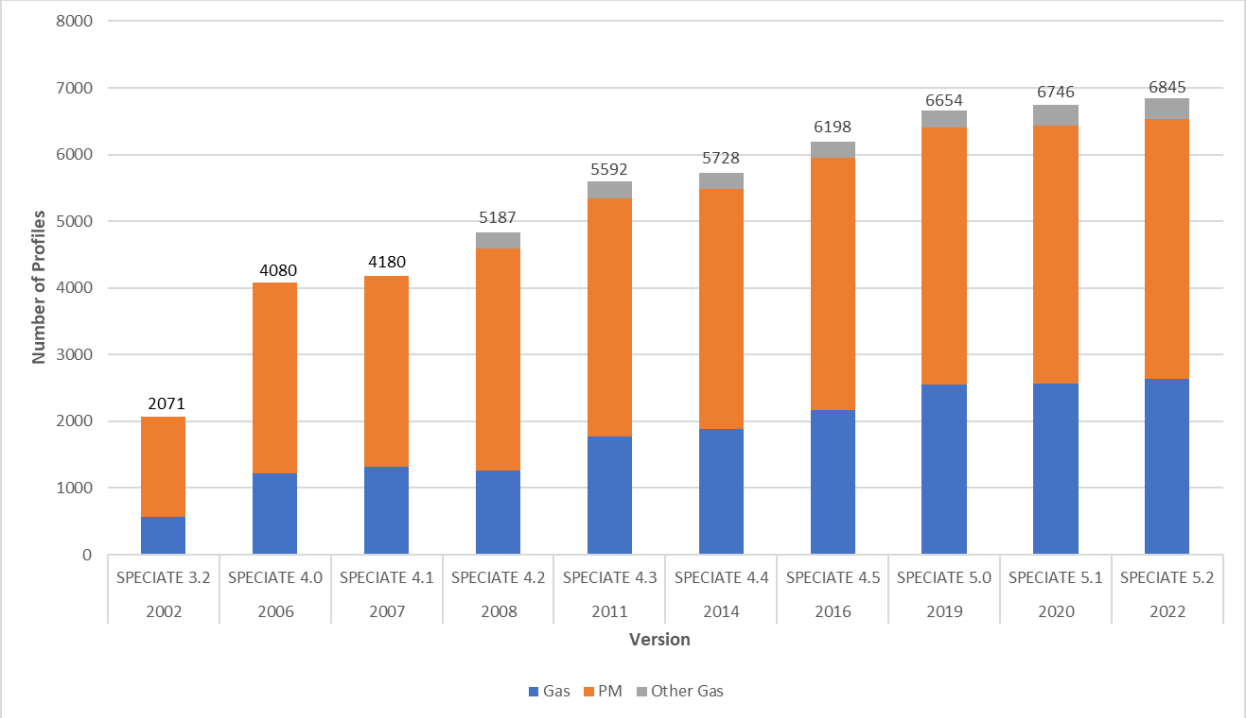
EPA is releasing an updated version of the SPECIATE database, SPECIATE 5.2. In lieu of full documentation, this document provides highlights of the revisions to the SPECIATE 5.1 database. Full documentation of the SPECIATE 5.0 database can be found in the SPECIATE documentation section of EPA's air emissions modeling website.

SPECIATE is the U.S. Environmental Protection Agency's (EPA) repository of speciation profiles of air pollution sources that provide the species makeup or composition of organic gas, particulate matter (PM) and other pollutants emitted from these sources. Some of the many uses of these source profiles include: (1) creating speciated emissions inventories for regional haze, PM, greenhouse gas (GHG), 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 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 3.2, 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 5.2, both in Microsoft Access® format and in the SPECIATE browser.

The figure below shows the number of profiles in various releases of SPECIATE.



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) and Office of Air and Radiation (OAR).

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 5.2 by appending 80 organic gas profiles and 19 PM profiles to the SPECIATE 5.1 database. In total, the SPECIATE 5.2 database includes 6,845 profiles. The new organic gas profiles added to SPECIATE 5.2 span a variety of sectors, including region-specific oil and gas, volatile chemical products, asphalt paving, dairy silage, mobile sources, biomass burning, and pesticides. The new PM profiles in SPECIATE 5.2 include a grass fire profile developed by EPA's biomass burning testing program and 17 mobile source profiles. Additionally, there was one organic gas profile (95112) and two PM profiles (95219 and 95220) with errors that have been corrected. The corrected profiles (95112a, 95219a, and 95220a), as well as the prior profiles, are included in SPECIATE 5.2.

The SPECIATE 5.2 database also includes 160 new species in the SPECIES_PROPERTIES table. Of these, 84 are new, volatility-defined, lumped species. New asphalt paving profiles (organic gas) and mobile source profiles (both organic gas and PM) use these new species.

In addition to updating many entries in the SPECIES_PROPERTIES table, EPA also added 7 new columns. These new fields specify representative compound structures for each species in SPECIATE and physiochemical properties predicted by the OPEn structure–activity/property Relationship App (OPERA) (Mansouri et al.,2018). These new representative compound fields were implemented to provide greater transparency and enable more automated mapping of individual species to chemical mechanisms, such as CRACMM. For a description on each of these additions, please see Table 6.

Change to the SPECIES table include:

(1) Revised the data in the INCLUDE_IN_SUM field for PM profiles with both ionic and atomic species in the same profile to be consistent with the EPA Speciation Tool used to develop speciated modeling emission inventories. 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. Since the Speciation Tool processes the profiles for modeling, EPA decided not to re-visit PM profiles that are in SPECIATE because the changes are likely small.

(2) Populated the INCLUDE_IN_SUM field in the Species table for all profiles. For example, EC1, EC2, and EC3 are part of EC and their weight % are not added to the INCLUDE_IN_SUM to avoid double counting. All species in TOG/VOC profiles are marked "Yes."

Change to the PROFILES table include:

(1) Updated hyperlinks in the DOC_LINK and Q_LINK fields from "ftp://newftp.epa.gov" to "https://gaftp.epa.gov". This change allows current web browsers to access the target files.

Change to the SPECIES_PROPERTIES table include:

(1) The HAPS field in the SPECIES_PROPERTIES table was corrected for 59 species. Of the 59 species changed, 56 of them are polycylic aromatic hydrocarbons. Polycyclic aromatic hydrocarbons qualify as POM (polycyclic organic matter), which regulatorily includes organic compounds with more than one benzene ring, and which have a boiling point greater than or equal to 100 ºC. Since naphthalene, the simplest POM, has a boiling point of 200 ºC, naphthalene and all other PAHs are listed as HAPs.

Finally, new profile types (PROFILE_TYPE = PM-AE8 and PM-CR1) have been added to the database. These new profiles expand upon the PM-AE6 species by further disaggregating the OC and PNCOM portions based on volatility. The PM-AE8 type facilitates the use of the AE8 module and the PM-CR1 type facilitates the use of the CRACMM (Pye et al. in prep) chemical mechanism. It should be noted that the PM-AE8 PROFILE_TYPE replaces the prior PM-VBS PROFILE_TYPE.

Chapter I of this addendum discusses the profiles added since SPECIATE 5.1 and the database changes, such as tables and field names, Chapter II provides a description of the database tables and field names, and Chapter III lists potential considerations for the next version.

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 Contract Officer Representative (TOCOR) The workgroup members include:

SPECIATE WORKGROUP MEMBERSHIP			
NAME	EPA OFFICE	EPA DIVISION	EXPERTISE/SPECIALIZATION
Amara Holder	ORD/CEMM	AMCD	Emissions Source Testing and Black Carbon
Art Diem	OAR/OAQPS	AQAD	Co-Lead, HAPS, and Air Toxics
Ben Murphy	ORD/CEMM	AESMD	Secondary Organic Aerosol Modeling
Brooke Hemming	ORD/CPHEA	HEEAD	Climate Change and Black Carbon
Casey Myers	OAR/OAQPS	SPPD	Emission Source Speciation, Stationary Source Emission Factors
Claudia Toro	OAR/OTAQ	FEL	Mobile Source Emissions
Emma D'Ambro	ORD/CEMM	AESMD	PFAS Modeling, CRACMM Development
George Pouliot	ORD/CEMM	AESMD	Emissions Modeling (Inventories and Platforms)
Havala Pye	ORD/CEMM	AESMD	Secondary Organic Aerosol Modeling, Solvents, CRACMM Development
Heather Simon	OAR/OAQPS	AQAD	Air Quality Modeling
Ingrid George	ORD/CEMM	AMCD	Emissions Source Testing and Black Carbon
Ivan Piletic	ORD/CEMM	AESMD	CRACMM Development
Jenia McBrian	ORD/CEMM	AESMD	QA Manager
Julia Black	OAR/OAQPS	AQAD	SPECIATE Browser
Justine Geidosch	OAR/OTAQ	FEL	Mobile Source Emissions
Karl Seltzer	OAR/OAQPS	AQAD	Co-Lead, Secondary Organic Aerosol Modeling, Solvents, CRACMM Development
Michael Hays	ORD/CEMM	AMCD	Emission Source Testing
Rich Cook	OAQ/OTAQ	FEL	Mobile Source Emissions
Sara Farrell	ORD/CEMM	AESMD	CRACMM Development
Souad Benromdhane	OAR/OAQPS	HEID	Health Benefits of Air Quality Management
Venkatesh Rao	OAR/OAQPS	AQAD	Biomass Burning and Black Carbon Inventory

CONTENTS

LIST OF 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
GHG	greenhouse gas
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 Contract Officer Representative
TOG	total organic gases
VBS	volatility basis set
VCP	volatile chemical products
VOC	volatile organic compounds
XRF	X-ray fluorescence

CHAPTER I. SPECIATE 5.2 – Changes from SPECIATE 5.1

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 for regional haze, PM, greenhouse gases, 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"). This web-based data browser, which is hosted on epa.gov, is designed using the Qlik platform, allows users to view and filter profile data, including the weight percents of species, from 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 comparisons across profiles.

The purpose of this addendum is to document changes made to the database since SPECIATE 5.1 and describe the tables, queries, and data fields in the database.

SPECIATE 5.2 reflects two main changes from SPECIATE 5.1: A) the addition of profiles and B) other structural and meta data updates to the data base.

A. Profile additions

In the SPECIATE 5.2 database, there are 80 new organic gas profiles. Many are new Volatile Chemical Product (VCP) profiles that were generated using the VCPy framework developed by Seltzer et al., (2021). These emissions are increasingly important in highly populated, urban settings and include everyday items, such as personal care products, general cleaners, architectural coatings, pesticides, adhesives, and printing inks.

Other newly added organic gas profiles span a variety of sectors, including region-specific oil and gas, dairy silage, asphalt paving, mobile sources, biomass burning, and pesticides. The oil and gas profiles are composites based on existing SPECIATE profiles; the dairy silage profile is based on a mass transfer model (Hafner et al., 2012); the asphalt paving profiles are based on comprehensive emissions characterization experiments (Khare et al., 2020), including intermediate/semi-volatile organic compounds (I/SVOCs); the mobile source profiles are based on a collaborative effort between EPA 's Office or Research and Development (ORD) and Office of Transportation and Air Quality (OTAQ); the biomass burning profiles are retrieved from the work of Koss et al., 2018, and the pesticide profile is developed using data from a California Air Resources Board pesticide usage report.

Most new PM profile added to the SPECIATE 5.2 database are composite mobile source profiles based on a collaborative effort between EPA 's Office or Research and Development (ORD) and Office of Transportation and Air Quality (OTAQ). These composite profiles are based on existing relevant profile(s), volatility resolved SVOCs from more recent studies, and an assigned, lumped volatility distribution for remaining OC+NCOM mass. The new mobile profiles include "CROC" in the PROFILE_CODE. There is one new PM-AE6 biomass burning profile that is based on a study conducted by EPA's Office of Research and Development (ORD). For this profile, field and laboratory $PM_{2.5}$ composition data were collected for grass fires using fuels from the state of Kansas. The focus of this research was to develop $PM_{2.5}$ composition data by combustion phase. These data were added to SPECIATE prior to manuscript completion and reference information for these profiles will be added in the next version of SPECIATE.

The Southern Ute profiles (SUIROGCT and SUIROGWT) applied to Archuleta and La Plata counties in southwestern Colorado and were developed from data provided in Tables 19 and 20 of the report by Oakley Hayes, Matt Wampler, Danny Powers titled "Final Report for 2017 Southern Ute Indian Tribe

Comprehensive Emissions Inventory for Criteria Pollutants, Hazardous Air Pollutants, and Greenhouse Gases."

A composite coal bed methane produced water profile, CBMPWWY, was developed by compositing a subset of the SPECIATE 5.0 pond profiles associated with coal bed methane wells. The SPECIATE 5.0 pond profiles were developed based on the publication: "Lyman, Seth N, Marc L Mansfield, Huy NQ Tran, Jordan D Evans, Colleen Jones, Trevor O'Neil, Ric Bowers, Ann Smith, and Cara Keslar. 2018. 'Emissions of Organic Compounds from Produced Water Ponds I: Characteristics and Speciation', Science of the Total Environment, 619: 896-905

The DJTFLR95 profile, DJ Condensate Flare Profile with DRE 95%, filled a need for the flared condensate and produced water tanks for Colorado's oil and gas operations. This profile was developed using the same approach as was used for the FLR99 SPECIATE 4.5 profiles, but instead of using profile 8949 for the uncombusted gas, it uses the Denver-Julesburg Basin Condensate composite (95398) and it quantifies the combustion by-products based on a 95% DRE. The approach for combining profile 95398 with combustion by-products based on the TCEQ's flare study (Allen, David T, and Vincent M Torres, University of Texas, Austin. 2011. 'TCEQ 2010 Flare Study Final Report', Texas Commission on Environmental Quality, https://www.tceq.texas.gov/assets/public/implementation/air/rules/Flare/2010flarestudy/2010- flare-study-final-report.pdf) and is the same as used in the workbook for the FLR** SPECIATE4.5 profiles. The approach uses the analysis developed by Ramboll (Ramboll and EPA, 2017).

Table 1. New Profiles Added to SPECIATE 5.2

PROFILE_CODE	PROFILE_TYPE	PROFILE_NAME
100CROC	PM-CR1	Onroad Gasoline Cold Start
100GROC	GAS	Onroad Gasoline Pre-Tier 2 Cold Start
101CROC	PM-CR1	Onroad Gasoline Running
101GROC	GAS	Onroad Gasoline Tier 2 Cold Start
102CROC	PM-CR1	Onroad Diesel Non-DPF Idle
102GROC	GAS	Onroad Gasoline Pre-Tier 2 Running
103CROC	PM-CR1	Onroad Diesel Non-DPF Start/Run
103GROC	GAS	Onroad Gasoline Tier 2 Running
104CROC	PM-CR1	Onroad Diesel DPF
104GROC	GAS	Onroad Diesel Non-DPF
PROFILE_CODE		**PROFILE_NAME**
105CROC	PM-CR1	Onroad CNG Non-OCR
105GROC	GAS	Onroad Diesel DPF
106CROC	PM-CR1	Onroad CNG OCR
106GROC	GAS	Onroad Diesel 2010+
107CROC	PM-CR1	Nonroad Gasoline
107GROC	GAS	Onroad CNG Bus
108CROC	PM-CR1	Nonroad Diesel Non-DPF
108GROC	GAS	Onroad E85 Cold Start
109CROC	PM-CR1	Nonroad Diesel DPF
109GROC	GAS	Onroad E85 Running
110CROC	PM-CR1	Nonroad Diesel Airport
110GROC	GAS	Nonroad E0 2 stroke
111CROC	PM-CR1	Nonroad Gas Turbine Aircraft
111GROC	GAS	Nonroad E10 2 stroke
112CROC	PM-CR1	Nonroad Diesel Rail
112GROC	GAS	Nonroad E0 4 stroke

113CROC	PM-CR1	Nonroad Diesel Marine
113GROC	GAS	Nonroad E10 4 stroke
114CROC	PM-CR1	Nonroad Residual Oil Marine
114GROC	GAS	Nonroad Diesel Pre-Tier 1
115GROC	GAS	Nonroad Diesel Tier 1
116GROC	GAS	Nonroad Diesel Tier 2-3
117GROC	GAS	Nonroad LPG
118GROC	GAS	Nonroad Gasoline Airport
119GROC	GAS	Nonroad Diesel Airport
120GROC	GAS	Nonroad Gas Turbine Aircraft
121GROC	GAS	Nonroad Diesel Marine
122GROC	GAS	Nonroad Residual Oil Marine
95112a	GAS	Unburned Gasoline Exhaust - Tunnel Study - Revised
95120a	GAS	Volatility Resolved Liquid Diesel - California composite
95219a	GAS	CNG transit bus exhaust from a lean-burn engine -no aftertreatment
95220a	GAS	CNG transit bus exhaust from a lean-burn engine – oxidation catalyst
95809	PM-AE6	Grass Fire-Field-Kansas AE6
95810	GAS	Volatile Chemical Products - Adhesives and Sealants - 2000 Representation
95811	GAS	Volatile Chemical Products - Adhesives and Sealants - 2015 Representation
95812	GAS	Volatile Chemical Products - Aerosol Coatings - 2010 Representation
95813	GAS	Volatile Chemical Products - Allied Paint Products - 2000 Representation
95814	GAS	Volatile Chemical Products - Allied Paint Products - 2015 Representation
95815	GAS	Volatile Chemical Products - Architectural Coatings - 2005 Representation
95816	GAS	Volatile Chemical Products - Architectural Coatings - 2014 Representation
PROFILE_CODE		**PROFILE_NAME**
95817	GAS	Volatile Chemical Products - Cleaning Products - 2000 Representation
95818	GAS	Volatile Chemical Products - Cleaning Products - 2015 Representation
95819	GAS	Volatile Chemical Products - Fuels and Lighter - 2000 Representation
95820	GAS	Volatile Chemical Products - Fuels and Lighter - 2015 Representation
95821	GAS	Volatile Chemical Products - Miscellaneous Products - 2000 Representation
95822	GAS	Volatile Chemical Products - Miscellaneous Products - 2015 Representation
95823	GAS	Volatile Chemical Products - Personal Care Products - 2000 Representation
95824	GAS	Volatile Chemical Products - Personal Care Products - 2015 Representation

95825	GAS	Volatile Chemical Products - All FIFRA Related Products - 2000 Representation
95826	GAS	Volatile Chemical Products - All FIFRA Related Products - 2015 Representation
95828	GAS	Pesticides - Agricultural
95829	GAS	Asphalt Paving: 60 degC
95830	GAS	Asphalt Paving: 140 degC
95831	GAS	Asphalt Shingles: 75 degC Lamp Off
95832	GAS	Asphalt Shingles: 75 degC Lamp On
95833	GAS	Asphalt Sealant: 75 degC Lamp Off
95834	GAS	Asphalt Sealant: 75 degC Lamp On
95835	GAS	Roofing Asphalt: 75 degC Lamp Off
95836	GAS	Roofing Asphalt: 75 degC Lamp On
95837	GAS	Printing Evaporation Loss - General - CARB Profile 517
95838	GAS	Dairies-Silage-Hafner-mass transfer model
95839	GAS	Personal Care Products - VCPy.v2.0
95840	GAS	Automotive Aftermarket Products - VCPy.v2.0
95841	GAS	Household Products - VCPy.v2.0
95842	GAS	Degreasing - VCPy.v2.0
95843	GAS	Coatings and Related Products - VCPy.v2.0
95844	GAS	Ponderosa Pine - Laboratory Wildfire
95845	GAS	Lodgepole pine - Laboratory Wildfire
95846	GAS	Douglas Fir - Laboratory Wildfire
95847	GAS	Subalpine Fir - Laboratory Wildfire
95848	GAS	Engelmann Spruce - Laboratory Wildfire
95849	GAS	Loblolly Pine - Laboratory Wildfire
95850	GAS	Jeffrey Pine - Laboratory Wildfire
95851	GAS	Juniper - Laboratory Wildfire
95852	GAS	Sage - Laboratory Wildfire
95853	GAS	Manzanita, unpolluted site - Laboratory Wildfire
95854	GAS	Manzanita, polluted site - Laboratory Wildfire
95855	GAS	Chamise, unpolluted site - Laboratory Wildfire
95856	GAS	Chamise, polluted site - Laboratory Wildfire
95857	GAS	Ceanothus - Laboratory Wildfire
PROFILE_CODE		**PROFILE_NAME**
95858	GAS	Bear Grass - Laboratory Wildfire
95859	GAS	Rice Straw - Laboratory Wildfire
95860	GAS	Peat - Laboratory Wildfire
95861	GAS	Laboratory Wildfire Composite - Various non-grass fuel types
CBMPWWY	GAS	Coal Bed Methane Produced Water Profile - WY ponds
DJTFLR95	GAS	DJ Condensate Flare Profile with DRE 95%
SUIROGCT	GAS	Flash Gas from Condensate Tanks - Composite Southern Ute Indian Reservation
SUIROGWT	GAS	Flash Gas from Produced Water Tanks - Composite Southern Ute Indian Reservation

B. Database updates

In SPECIATE5.2, new fields were added to the SPECIES_PROPERTIES table and are described further below. SPECIATE tables, queries, and macros, some of which were updated, are described in Chapter II.

The SPECIATE field name changes are described in Table 8 of Chapter II. The remainder of this section provides an overview of the key changes to the reference and species properties information.

Change to the SPECIES table include:

(1) Revised the data in the INCLUDE_IN_SUM field for PM profiles with both ionic and atomic species in the same profile to be consistent with the EPA Speciation Tool used to develop speciated modeling emission inventories. 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. Since the Speciation Tool processes the profiles for modeling, EPA decided not to re-visit PM profiles that are in SPECIATE because the changes are likely small.

(2) Populated the INCLUDE_IN_SUM field in the Species table for all profiles. For example, EC1, EC2, and EC3 are part of EC and their weight % are not added to the INCLUDE_IN_SUM to avoid double counting. All species in TOG/VOC profiles are marked "Yes."

Change to the PROFILES table include:

(1) Updated hyperlinks in the DOC_LINK and Q_LINK fields from “ftp://newftp.epa.gov" to “https://gaftp.epa.gov”. This change allows current web browsers to access the target files.

(2) Two new PM profile have been added to the database: PM-AE8 and PM-CR1. Only PM-CR1 profiles were added to this version of SPECIATE, however previous PM-VBS profiles were renamed to PROFILE_TYPE = PM-AE8. Both profile types are derived using the AE6-protocol (See APPENDIX B) but differ in the treatment of organic matter. For PM-AE8, the POC and PNCOM from PM-AE6 profiles is distributed across a series of volatility bins. For PM-CR1, organic matter is reported as the PM species (i.e., it does not differentiate aerosol carbon and aerosol non-carbon) and distributes this mass across a series of volatility bins. The volatility bins for which this organic mass is distributed includes species with a saturation concentration (C*) up to 320 µg m-3 (operationally defined as Condensable Reactive Organic Carbon; CROC). These profiles are consistent with the species used in the AE8 module of CMAQ and the Community Regional Atmospheric Chemistry Multiphase Mechanism (CRACMM) chemical mechanism (Pye et al., in prep.), respectively. It should be noted that new profiles should be added as PM-AE6, PM-AE8, or PM-CR1 once, and that all three versions of the same profile do not need to be added to SPECIATE. The Speciation Tool will handle the translation between different profile types. Preferentially, profiles will be added to SPECIATE as PM-CR1, then PM-AE8, and finally PM-AE6. This order of preference highlights the desire to retain the volatility information from the underlying measurement study, if available. If a study does not report volatility information, default volatility profiles will be applied in the Speciation Tool if the chemical mechanism of the photochemical model requires such information

Change to the SPECIES_PROPERTIES table include:

(1) The HAPS field in the SPECIES_PROPERTIES table was corrected for 59 species. Of the 59 species changed, 56 of them are polycylic aromatic hydrocarbons. Polycyclic aromatic hydrocarbons qualify as POM (polycyclic organic matter), which regulatorily includes organic compounds with more than one benzene ring, and which have a boiling point greater than or equal to 100 ºC. Since naphthalene, the simplest POM, has a boiling point of 200 ºC, naphthalene and all other PAHs are listed as HAPs.

(2) Several fields were added: REPRESENTATIVE_COMPOUND_NAME, REPRESENTATIVE_COMPOUND_DSSTox_ID, ATMOSPHERIC_HYDROXYLATION_RATE_(AOH)_CM3/MOLECULE-SEC_OPERA, HENRYS_LAW_ATM-M3/MOLE_OPERA, OCTANOL_AIR_PARTITION_COEFF_LOGKOA_OPERA,

OCTANOL_WATER_PARTITION_LOGP_OPERA, REPCMPSCORE. For more information on these entries, please see later sections of this document.

CHAPTER II. SPECIATE 5.2 Database – Tables, Queries and Fields

This chapter describes the organization of the SPECIATE 5.2 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 meta data 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 5.2 Database

Table	Description	Notes
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 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 meta data 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 meta data for each species.	Added 84 new volatility and functionality defined, lumped species and 7 new columns - REPRESENTATIVE_COMPOUND_NAME, REPRESENTATIVE_COMPOUND_DSSTox_ID, ATMOSPHERIC_HYDROXYLATION_RATE_(AOH)_CM3/MOLECULE-SEC_OPERA, HENRYS_LAW_ATM-M3/MOLE_OPERA, OCTANOL_AIR_PARTITION_COEFF_LOGKOA_OPERA, OCTANOL_WATER_PARTITION_LOGP_OPERA, REPCMPSCORE
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 delimitered list of species synonyms for each Species ID.	Several synonyms were added to this table. All species names in the SPECIES_PROPERTIES table were added plus additional synonyms that were identified from EPA's SRS

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 5.2 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. -Due to the current approach used by the macro, you cannot have more than 3 references per profile
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 By concatenating all the species synonyms, it will be easier to search the in the browser or 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 SPECIATE5.2 for completeness.

Table 4. Other Tables in the SPECIATE 5.2 Database

Table	Description	Notes
List of SVOC Splitting Factors	Supporting table (legacy). This is an old table that is really 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 mass reconstruction for PM profiles	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 included in the database serve different functions as follows: 1) provide easy access to readable data summaries that contain descriptions and meta data, 2) provide input to the Speciation Tool, and 3) allow users to perform quality assurance (QA). The queries summarized in Table 5 are primarily those that provide readable data summaries and inputs to the Speciation 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 5.2 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 many fewer fields; enables you to focus on the weight percent information without all the meta data
View PM Profiles-Simplified	Same as View PM Profiles but many fewer fields; enables you to focus on the weight percent information without all the meta data
View Other Profiles-Simplified	Same as View Other Profiles but many fewer fields; enables you to focus on the weight percent information without all the meta data
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 Speciation Tool.
export_species	Query developed to extract select fields from the SPECIES table for use in the Speciation Tool.
export_species_properties	Query developed to extract select fields from the SPECIES_PROPERTIES table for use in the Speciation 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 you to choose to enter a species by the name in 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 the 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
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.
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, 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 this 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)

Field Name	Data Type	Description
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 that a profile was added to
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)
CATEGORY_LEVEL_2_Sector_Eq uipment	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 AE6 profile would not exceed 100%
CREATED BY	Text	Person who added this profile
CREATED Date	Date/Time	Date the profile was added
MODIFIED BY	Text	Person who modified this profile
MODIFIED DATE	Date/Time	Date the profile was added
REVIEWED BY	Text	Person who reviewed this profile
REVIEWED DATE	Date/Time	Date the profile was reviewed
Data_Origin	Text	Origin of data. This is the same as DATA_ORIGIN which was in the KEYWORD_REFERENCE table in SPECIATE 5.0

Field Name	Data Type	Description
Keywords	Text	List of ideas and topics that define what your content is about. This is the same as KEYWORD which was 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 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) so not all species should be included to sum mass. To be consistent with Speciation 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.
PHASE	Text	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	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

Field Name	Data Type	Description
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 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 were assigned for all compounds in the SPECIATE database (Pye et. al., in prep.). 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)
REPRESENTATIVE_COMPOUND_DSSTox_ID	Text	The DSSTox Substance Identifier for the representative compound available in the EPA Chemicals Dashboard (https://comptox.epa.gov/dashboard/).
SYMBOL	Text	Standard chemical abbreviation
SPEC_MW	Number	Species molecular weight
NOTE	Long Text	Note (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.

Field Name	Data Type	Description
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 was already fully specified in database and representative compound is an exact match to SPECIATE. The representative compound is the exact species.
- 3 - Species was matched in the dashboard, but there were 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 was manually matched to a representative species based on some information. For example, C12 alkanes were matched to dodecane, terpenes were 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 was manually mapped with a low degree of certainty. For example, "aggregated VOCs" was mapped to decane. These mappings may be reexamined in the future if they are significant contributors by mass.
- 0 - Not mapped. This value should only be used temporarily 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 that a profile was added to

Field Name	Data Type	Description
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 which was in the KEYWORD_REFERENCE table in SPECIATE 5.0.
REF_Codes	Text	Indicates the pollutant to be used in 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 what your content is about. This is the same as KEYWORD which was 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

Our current plan is to release a new version of SPECIATE each year. In our next release, we plan on enhancing the SPECIATE program and updating the database by:

- Maintaining/updating the standard operating procedures for developing and enhancing SPECIATE, as needed
- Adding more high-priority profiles that can lead to better performance in air quality modeling and potentially improve the NEI for HAPs
- Supplementing existing PM-AE6 profiles that are frequently used in modeling platforms with source-specific PM-CR1 profiles that incorporate volatility information
- Adding relevant species properties to the database such as atmospheric photochemical reactivity to form ozone and estimates of secondary organic aerosol yields.
- Improving the literature searches for PM and VOC speciation, and adding search terms for mercury speciation
- Continuing to improve the browser tool based on comments that we receive from its use
- Continue to reach out to the research community for providing high quality and high priority data to SPECIATE

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* (J-AESMD-ESAB-C-0-60, approved 12/2021). Independent QA audits were not deemed necessary.
- This report 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

Hafner, S.D., Montes, F., Rotz, C.A., A mass transfer model for VOC emission from silage, Atmospheric Environment, 54 (2012) 134-140. doi: 10.1016/j.atmosenv.2012.03.005

Khare, P., Machesky, J., Soto, R., He, M., Presto, A.A., Gentner, D.R. "Asphalt-related emissions are a major missing nontraditional source of secondary organic aerosol precursors." Science Advances, 6: eabb9785, 2020.

Koss, A.R., Sekimoto, K., Gilman, J.B., Selimovic, V., Coggon, M.M., Zarzana, K.J., Yuan, B., Lerner, B.M., Brown, S.S., Jimenez, J.L., Krechmer, J., Roberts, J.M., Warneke, C., Yokelson, R.J., and de Gouw, J. "Non-methane organic gas emissions from biomass burning: identification, quantification, and emission factors from PTR-ToF during the FIREX 2016 laboratory experiment." Atmos. Chem. Phys., 18, 3299–3319, 2018, doi: 10.5194/acp-18-3299-2018.

Mansouri, K., Grulke, C.M., Judson, R.S. *et al.* OPERA models for predicting physicochemical properties and environmental fate endpoints. *J Cheminform,* **10:** 10, 2018. Doi: 10.1186/s13321-018-0263-1.

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) version 1.0, in preparation.

Seltzer, K.M., Pennington, E., Rao, V., Murphy, B.N., Strum, M., Isaacs, K.K., and Pye, H. O. T., Reactive organic carbon emissions from volatile chemical products, Atmos. Chem. Phys., 21, 5079–5100, 2021, doi: 10.5194/acp-21-5079-2021.

APPENDIX A. Profile Quality Criteria Evaluation

The Quality Score (QSCORE) provide an evaluation framework to easily recognize and assign value points to indicators of a strong, well planned and executed study, which is presented in a complete and logical manner. The presentation of air emission profile data can be in the form of a peer-reviewed publication, or report.

The evaluation framework is meant to guide the reviewer to assign quality value points to the areas 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 and overly detailed. A point to each question adds-up to an evaluation score. An ideal point score would have 30 (Data from Measurements) or 29 (Data from other Methods) desired criteria (points). Each point or points is additive, influencing, but not necessarily distinguishing the study. The publication or report should be ranked as high as possible for inclusion into 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://newftp.epa.gov/air/emismod/SPECIATE_supportingdata/' then the version and file name. The folder for the current SPECIATE database versions is 5_1, so the link would be:
ftp://newftp.epa.gov/air/emismod/SPECIATE_supportingdata/v5_1 /*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 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)

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. Protocol for Developing AE6-ready $PM_{2.5}$ Speciation Profiles for Inclusion in SPECIATE

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 for regional haze, PM, greenhouse gases, 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 (e.g., factor analysis and positive matrix factorization). Here, methods for processing raw $PM_{2.5}$ profile data into a format suitable for use in the Community Multiscale Air Quality (CMAQ) photochemical air quality modeling are described.

PROFILE_TYPE options for $PM_{2.5}$ profiles within SPECIATE include PM-AE6, PM-AE8, PM-CR1, PM-Simplified, and PM. PM-AE6 profiles are PM profiles developed using the "AE6 protocol" (Reff et al., 2009) were first introduced in SPECIATE v4.3. These profiles include species that directly map to species used by the AERO6 (AE6) module of CMAQ.

Both PM-AE8 and PM-CR1 profiles utilize the "AE6-protocol," but append additional steps that enable POC and PNCOM to be resolved along a volatility distribution. Mathematically, the relationship among the three profile types is as follows:

$$\begin{aligned} &POC + PNCOM\ (AE6\ species) \\ &= POCN2 + POCN1 + POCP0 + POCP1 + POCP2 + PNCOMN2 + PNCOMN1 + PNCOMP0 \\ &\quad + PNCOMP1 + PNCOMP2\ (AE8\ species) \\ &= ROCN2ALK + ROCN1ALK + ROCP0ALK + ROCP1ALK + ROCP2ALK\ (CRACMMv1\ species) \end{aligned}$$

Except for the organic matter components, PM-AE6, PM-AE8, and PM-CR1 are operationally the same. In PM-AE8 and PM-CR1, both POC and PNCOM are split among several species 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 = 10^{1}. In PM-AE8, the organic carbon and non-carbon organic mass are split, like PM-AE6, whereas the two are summed together in PM-CR1 (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). Finally, PM profiles include other, non-mechanism-specific components, such as explicit semi-volatiles and polycyclic aromatic hydrocarbons.

The Speciation Tool is software that translates SPECIATE profiles into profiles that are chemical mechanism-specific and generates several files that are input into SMOKE. For PM-AE6 profiles, the Speciation Tool maps the species IDs in the SPECIATE profile to only AE6 species and allocates all remaining mass to "PM Other." This functionality allows SPECIATE developers to include more species in "AE6-ready" profiles than are contained within the AERO6 module.

AE6 Species

Below, Table B-1 maps each SPECIES_ID in SPECIATE to compounds that are explicit in AERO6, as well as additional species that are included in the calculation of $PM_{2.5}$ mass (i.e., other "AE6-ready" species). This mapping is used in the Speciation Tool to create the SMOKE-ready speciation profiles. 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. In some studies, the atomic form 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. Gap filling is needed because the study may have measured only the atomic form of the metal, but the model uses the ionic form. Rather than putting in a 0 for the ionic form, the weight percent of the atomic form is used. Some profiles have both atomic and ionic forms and when doing a regression, we found that other than Na which has a poor regression coefficient, the weight percents of the ion/atomic forms closely follow each other (see Figure B-1). Note that Mg and Ca did not have sufficient data points for a meaningful regression and are not shown. Finally, the comments in Table B-1 indicate if there were changes made to the mapping from version 4.2 of the Speciation Tool. The changes to the mapping from version 4.2 of the Speciation Tool are made because in CMAQ, the AE6 uses the ionic form of several metals whereas version 4.2 assigned the atomic form.

Species Name	SPECIES_ID	Description	AE6 Species?
POC	626	Organic Carbon	YES
PEC	797	Elemental Carbon	YES
PSO4	699	Sulfate	YES
PNO3	613	Nitrate	YES
PNH4	784	Ammonium	YES
PNCOM	2669	Non-Carbon Organic Matter	YES
PFE	488	Iron	YES
PAL	292	Aluminum	YES
PSI	694	Silicon	YES
PTI	715	Titanium	YES
PCA	2303*	Calcium	YES
PMG	2772@	Magnesium	YES
PK	2302#	Potassium	YES
PMN	526	Manganese	YES
PNA	785%	Sodium	YES
PCL	337&	Chloride	YES
PH2O	2668	Particulate Water	YES
PMOTHR	2671	Other PM	YES
--	666	Phosphorus	NO
--	767	Vanadium	NO
--	347	Chromium	NO
--	379	Cobalt	NO
--	612	Nickel	NO
--	380	Copper	NO
--	778	Zinc	NO
--	468	Gallium	NO
--	298	Arsenic	NO
--	693	Selenium	NO
--	689	Rubidium	NO
--	397	Strontium	NO
--	779	Zirconium	NO
--	586	Molybdenum	NO
--	649	Palladium	NO
--	695	Silver	NO

Species Name	SPECIES_ID	Description	AE6 Species?
--	328	Cadmium	NO
--	487	Indium	NO
--	714	Tin	NO
--	296	Antimony	NO
--	300	Barium	NO
--	519	Lanthanum	NO
--	1861	Cerium	NO
--	528	Mercury	NO
--	520	Lead	NO

Table B-1. PM components to include in a PM-AE6 profile.
*** If a profile does not have 2303 but does have 329, 329 should be used.**
@ If a profile does not have 2772 but does have 525, 525 should be used.
 # If a profile does not have 2302 but does have 669, 669 should be used.
% If a profile does not have 785 but does have 696, 696 should be used.
& If a profile does not have 337 but does have 795, 795 should be used.
Note that all species included in a PM-AE6 profile but not an AE6 species is routed to PMO (2671) in the Speciation Tool.

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. This is a requirement of the Speciation Tool which prepares the PM-AE6 speciation profiles in SPECIATE for SMOKE.

Instructions for Creating AE6 Profiles for Inclusion in 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 due to comments in the paper. Note the measurement methods (can be different for different species), whether the source is controlled, and if so using what measures.

Step 2: Map species in the reference to SPECIATE species and assign Species IDs

Step 3: Determine if OC needs to be adjusted due to "artifacts."

Artifacts are volatiles 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.

We believe that a non-zero back up filter measurement does provide evidence for positive artifacts and *may* be able to be quantitatively used to adjust by subtracting the backup from the primary filter. However, if the two filters provide similar values, and the difference results in very small OC with high uncertainty, then that difference value should not be quantitatively used to estimate "true" OC because of the high uncertainty. 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 use this to adjust OC. If neither are available from the paper, a judgement should be made on a case-by-case basis on whether or not to estimate "true" OC as the difference between the primary and secondary filter measurements. The guidance here is that if the primary filter and back up filter measurements are close, then it is not appropriate to use the difference (a very small number) as the "true" OC.

If there is no adjustment provided or is too uncertain (masses of primary and secondary are similar), and there appear to be artifacts, then OC can be adjusted later if the mass exceeds 100% after adding in the other AE6 species that are not contained in the paper.

Step 4: ADD particulate water, PH2O. Note that this is SPECIES ID 2668 in SPECIATE.

The approach here is from the supplemental information from Reff, et. al, section S3.7.1

Type of Source	Particulate Water (PH2O) calculation
Combustion and other high temperature sources, where water is likely to be emitted in the vapor phase	0
All other sources	24% of the sum of sulfate (PSO4) and ammonium (PNH4) concentrations or percentages

Sources for which we assume zero PH_2O emissions are:

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 Reff et. al. Section 3.7.4. These are imputed 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: Make sure there is consistency in sulfate and sulfur. If a profile has sulfate and not sulfur, the sulfur does not need to be computed, but if it has sulfur but not sulfate it should be computed as follows:

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

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

Step 7: Add Metal Bound Oxygen, MO. Note that this is SPECIES ID 2670 in SPECIATE.

While MO is not an AE6 species, it needs to be computed and included in the profile (unless it is 0) to enable a check for total mass fraction <= 100%.

The approach to compute MO follows Section S.3.7.2 in from Reff, et. al., which is to stoichiometrically combine oxygen with the metals, and then adjust 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 change from the Reff, et. al., approach is to use only 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

$$\boldsymbol{MO_{unadjusted} = \sum_{El}^{N} Ox_{El} \times E_{El}} \quad (1)$$

where Ox_{El} is the oxygen-to-metal ratio for metal El (Table B-2), and E_{El} is the emission of metal El, **except for Na, Ca, Mg and K**. For these 4 metals, the E_{El} should reflect the difference between the atom form of

the metal and the ion form. If, for Na, Ca, Mg, and K, the profile has only one form (atom or ion but not both) then the E_{EI} should be set to 0. Also, if the difference is negative, it should be set to 0.

Note that for metals in which there are multiple forms of the MO compound, an average of the oxygen to metal ratios across all forms is used.

To adjust MO based on preferential combining of sulfate over oxygen, compute the available sulfate for binding with metals, which is the sulfate remaining after fully neutralizing the NH_4^+ in the profile.

$$\boldsymbol{Neutralized\ SO_4^{2-}} = \frac{\mathbf{0.5 * 96}}{\mathbf{18}} \times \boldsymbol{E_{NH_4^+}} \quad (2)$$

Where $E_{NH_4^+}$ is the mass of NH_4^+ in the profile.

The non-neutralized sulfate is the remainder from the sulfate in the profile.

$$\boldsymbol{Non\_Neutralized\_SO_4^{2-}} = \boldsymbol{E_{SO_4^{2-}}} - \boldsymbol{Neutralized\ SO_4^{2-}} \quad (3)$$

If $Non\_Neutralized\ SO_4^{2-} < 0$,

$$\boldsymbol{MO_{adjusted}} = \boldsymbol{MO_{unadjusted}} \quad (4)$$

If $Non\_Neutralized\ SO_4^{2-} > 0$

$$\boldsymbol{MO_{adjusted}} = \boldsymbol{MO_{unadjusted}} - \boldsymbol{Non\_Neutralized\ SO_4^{2-}} \times \frac{\mathbf{16}}{\mathbf{96}} \quad (5)$$

$$\boldsymbol{If\ MO_{adjusted} < 0,\ MO_{adjusted} = 0} \quad (6)$$

If the difference is >0 between atom and ion for NA, Ca, Mg, and K, use that for the MO calculation. Otherwise set the MO for these metals to 0.

Table B-2: Assumed Oxide Forms of Each Metal and Resulting Mean Oxygen-to-Metal Ratio Used in Equation 1

Species	MW of metal[1]	Oxide Form 1	Oxide Form 2	Oxide Form 3	Oxygen/Metal Ratio
Na (Use difference between atom and ion)	22.99	Na_2O			0.348
Mg (Use difference between atom and ion)	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 (Use difference between atom and ion)	39.10	K_2O			0.205
Ca (Use difference between atom and ion)	40.08	CaO			0.399
Ti	47.87	TiO2			0.669
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

Species	MW of metal[1]	Oxide Form 1	Oxide Form 2	Oxide Form 3	Oxygen/Metal Ratio
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	Hg2O	HgO		0.060
Pb	207.20	PbO	PbO_2		0.116

Step 8: Add particulate non-carbon organic matter (PNCOM)

Every profile that has POC must have PNCOM computed from POC. If the paper (also check the supplemental information) provides a factor to compute this, use the value provided in the paper. Otherwise, use the default values provided in section S.3.7.3 of Reff, et, al. These values are provided in the box below. Populate the ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO field in SPECIATE as 1 plus the fraction used (e.g., the default values are provided in the table below). Also, indicate in the NOTES field of the SPECIATE database how PNCOM was computed.

Type of Source	Computation of PNCOM	ORGANIC_MATTER_ to_ORGANIC_CARBON _RATIO
Onroad and Nonroad motor vehicle exhaust profiles (e.g., the HDDV Exhaust, Nonroad Gasoline Exhaust, Onroad Gasoline Exhaust, and LDDV Exhaust source categories):	$PNCOM = 0.25 * POC$	1.25
Wood combustion sources other than wood-fired boilers (e.g., wildfires, agricultural burning, residential wood combustion, prescribed burning, slash burning	$PNCOM = 0.7 * POC$	1.7
Wood-fired boilers and ALL OTHER SOURCES	$PNCOM = 0.4 * POC$	1.4

Step 9: Check for sum of $PM_{2.5}$ weight fractions over 100%

No adjustments need to be made if the weight fraction is less than 101%.

In this check, Sulfur should be excluded because it is double counted with sulfate. If the mass is still over 100% then:

1) Double check the paper to see if there are POC artifacts. If so and there is no quantitative information in the paper, **then adjust POC and PNCOM down by the same multiplier until the sum of weight fractions is 100%.**

2) If POC artifacts have already been corrected for, there is not likely to be POC artifacts or POC is already very low and adjusting it would not reduce the total to 100%, then adjust all species down (i.e., normalize all weight percents) to get the sum to be 100%. If any of these adjustments are made, it should be documented in the NOTES.