Addendum

SPECIATE Version 5.1
Database Development Documentation

June 2020 **EPA/600/R-20/189**

EPA SPECIATE Workgroup
US Environmental Protection Agency

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

Executive Summary

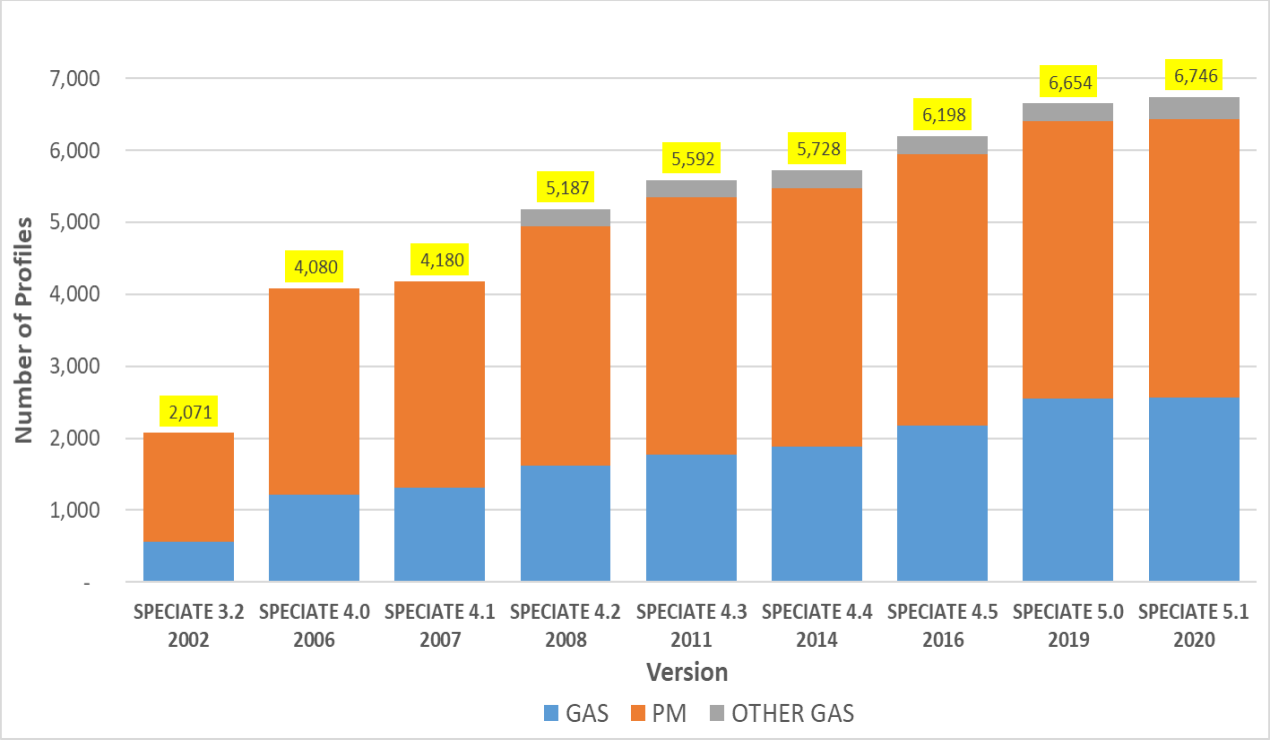
EPA is releasing an updated version of the SPECIATE database, SPECIATE 5.1, about a year after the release of SPECIATE 5.0. In lieu of full documentation, this document provides highlights of the revisions to SPECIATE 5.0. Full documentation of the SPECIATE program can be found in the SPECIATE documentation 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) adding PM species in the EPA's National Emissions Inventory (NEI); (3) developing black carbon assessments and particulate carbonaceous inventories; (4) estimating air toxic pollutant emissions from PM and organic gas primary emissions; (5) providing input to chemical mass balance (CMB) receptor models; and, (6) verifying profiles derived from ambient 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 (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.

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 5.1, 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 SPECIATE 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, quality assure the data and profiles, and coordinate improvements to the database structure and metadata fields. As newer SPECIATE versions are developed, improvements are made to the process as well as the data.

Processes that have been improved in SPECIATE include the method by which the SWG evaluates profiles for inclusion, a systematic approach to identify profile needs, updates to reference information, and improved documentation.

The EPA generated SPECIATE 5.1 by appending 16 organic gas profiles, 18 PM profiles and 58 other (mercury) profiles to the SPECIATE 5.0 database. In total, the SPECIATE 5.1 database includes 6,746 profiles. The organic gas profiles added include region-specific oil and gas profiles and new composites from existing SPECIATE profiles. The PM profiles include new fire profiles developed by EPA's biomass burning testing program and a sugar cane combustion profile from the literature. The mercury profiles include data for new source categories such as geothermal electricity generation and improved documentation for existing categories.

The EPA made structural improvements to the SPECIATE 5.0 database to improve the approach for storing multiple references. The EPA expanded the metadata fields to include additional information on profile quality and for continued support of the volatility basis set approach for air quality models. An additional vapor pressure field was added as well as a species identifier from EPA's Distributed Structure-Searchable Toxicity (DSSTox) Database.

The SPECIATE BROWSER has been updated to accommodate the new structure and data. A new simplified profile table has been added that provides the key fields for describing the profile, species, and providing the profile information most critical to users (weigh percent information). In addition, a "how to" guide has been posted to assist users with data searches.

Adding profiles to SPECIATE requires many layers of review including processes to prioritize and evaluate the data. For SPECIATE 5.1 we followed the quality assessment scoring system (resulting in a quality score or QSCORE) developed during SPECIATE 5.0 to evaluate profiles using criteria that cuts across many aspects of profile development and potential use such as measurement techniques, completeness, and source category needs. We added a descriptor field that maps the quality score to a quality category (e.g., excellent, good, fair).

In summary, the maintenance of SPECIATE requires continuous assessment. EPA SWG members have established a process to identify and prioritize need, find sound research sources which address that need, critically review those sources, and finally add that data to SPECIATE. This rigorous attention to maintaining quality and relevance has established SPECIATE as a uniquely positioned source of information for air quality analysts, modelers, researchers, specialists, as well as interested public officials and individuals.

This addendum discusses the profiles added and database changes to SPECIATE 5.1. Chapter 1 provides these changes; and Chapter 2 provides a description of the database tables and field names.

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. Marc Menetrez, the EPA Task Order Contract Officer Representative (TOCOR) for this project; the Alternate TOCOR is Dr. Madeleine Strum. The SWG is coordinated by Dr. Menetrez and staffed by air quality professionals from the EPA's Office of Research and Development (ORD) and the Office of Air and Radiation (OAR). As of May 2020, the workgroup members include:

SPECIATE WORKGROUP MEMBERSHIP, May 2020			
NAME	EPA OFFICE	EPA DIVISION	EXPERTISE/SPECIALIZATION
Souad Benromdhane	OAR/OAQPS	HEID	Health Benefits of Air Quality Management
Julia Black	OAR/OAQPS	AQAD	Data analysis and visualization
Casey Bray	OAR/OAQPS	SPPD	Emission factors and Speciation
Art Diem	OAR/OAQPS	AQAD	Air toxics and Speciation
Justine Geidosch	OAR/OTAQ	ASD	Mobile Source Emissions
Ingrid George	ORD/CEMM	AMCD	Emission Source Testing and Black Carbon
Michael Hays	ORD/CEMM	AMCD	Emission Source Testing
Brooke Hemming	ORD/CPHEA	HEEAD	Atmospheric Chemistry, Wildfire Emissions Chemistry, Climate Change/Air Quality
Amara Holder	ORD/CEMM	AMCD	Emission Source Testing and Black Carbon
Marc Menetrez, co-lead	ORD/CEMM	AESMD	Emission Source Speciation
Ben Murphy	ORD/CEMM	AESMD	Secondary Organic Aerosol Modeling
Libby Nessley	ORD/CEMM	EPD	QA Manager
George Pouliot	ORD/CEMM	AESMD	Emissions Modeling (Inventories and Platforms)
Havala Pye	ORD/CEMM	AESMD	VOC Chemistry Modeling
Venkatesh Rao	OAR/OAQPS	AQAD	Biomass Burning, Black Carbon Inventory and Agricultural sources; SPECIATE literature searches
Heather Simon	OAR/OAQPS	AQAD	Air Quality Modeling
Madeleine Strum, co-lead	OAR/OAQPS	AQAD	National Emissions Inventory and Emissions Modeling
Tiffany Yelverton	ORD/CEMM	AMCD	Air Pollution Control, Combustion, and Black Carbon

CONTENTS

LIST OF TABLES

LIST OF FIGURES

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
ASD	Assessment and Standards Division, EPA/ORD
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
CMB	chemical mass balance
CMV	commercial marine vessels
DRI	Desert Research Institute
EC	elemental carbon
EF	emission factor
EPA	Environmental Protection Agency
EPD	Ecosystem Processes Division, EPA/ORD
EPI	estimation program interface
ESP-CS	Electrostatic precipitator, cold-side (ESP is installed downstream of the air preheater)
ESP-HS	Electrostatic precipitator, hot-side
ES&T	Environmental Science and Technology
FF Baghouse	fabric filter baghouse
FGD	flue gas desulfurization
FID	flame ionization detector
GC-MS	gas chromatography-mass spectroscopy
GHG	greenhouse gas
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
kg	kilogram
km	kilometer
LDDV	light-duty diesel vehicle
LVOC	low volatility organic compound
LVP	low vapor pressure
mg	milligram
MO	metal-bound oxygen
MTBE	methyl t-butyl ether
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

PADEP	Pennsylvania Department of Environmental Protection
QA	quality assurance
QSCORE	profile quality score
ROG	reactive organic gas
PAHs	polycyclic aromatic hydrocarbons
PAMS	photochemical assessment monitoring station
PC boiler	pulverized coal boiler
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 compounds
POA	primary organic aerosols
RFG	reformulated gasoline
RTP	Research Triangle Park
SAROAD	Storage and Retrieval of Aerometric Data
SCR	Selective catalytic reduction
SDA	dry lime/spray dryer adsorber followed by a baghouse.
SNCR	selective noncatalytic reduction
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 work group
THC	total hydrocarbon
TOCOR	Task Order Contract Officer Representative
TOG	total organic gases
UTDEQ	Utah Department of Environmental Quality
VBS	volatility basis set
VOC	volatile organic compounds
WRAP	Western Regional Air Partnership
XRF	x-ray diffraction

CHAPTER I. SPECIATE 5.1 – Changes from SPECIATE 5.0

The purpose of this addendum is to document changes made to the database since SPECIATE 5.0 and describe the tables, queries, and data fields in the database.

SPECIATE 5.1 reflects three main changes from SPECIATE 5.0: 1) the addition of profiles, 2) changes to the database structure and field name changes, additions, and meta data additions and 3) updates to the browser. There were also minor changes to the QSCORE that are documented here.

A. Profile updates

Profiles were added for each of the 3 broad profile types: 1) Gas, 2) particulate matter (PM) and 3) OTHER.

The majority of the 16 gas profiles added to SPECIATE 5.1 are for the oil and gas sector. Three separate sources, the Western Regional Air Partnership (WRAP), the Pennsylvania Department of Environmental Protection (PADEP) and the Utah Department of Environmental Quality (UTDEQ) supported the development of region-specific composition of produced gas or flash gas. All of them are for total organic gases (TOG) and rely on the sum of species.

The WRAP efforts resulted in five composite profiles, four covering the Williston Basin in Montana and North Dakota, and one covering the Central Montana uplift. The four from the Williston basin are: produced gas composition from oil wells in the Williston Basin in Montana, produced gas composition from oil wells in the Williston Basin in North Dakota, flash gas composition from tanks at oil wells in the Williston Basin in Montana and flash gas composition from tanks at oil wells in the Williston Basin in North Dakota. The one from Central Montana uplift is produced gas composition from gas wells. These profiles are documented in a memorandum from John Grant and Amnon Bar-Ilon, of Ramboll (Ramboll, 2019).

The PADEP provided composition data on natural gas from gas wells collected for compliance with unconventional wells under state regulations. The vast array of data included data from different methods and different levels of documentation and speciation. Some of these data provided full gas speciation and we attempted to identify and use data we believed to be most useful and complete for developing county-specific composites. Due to time and resource constraints, we limited the data to three counties. In future SPECIATE versions, we may include additional counties. The three county composite gas profiles are for Butler, Greene and Washington counties.

The UTDEQ provided raw gas and flash gas TOG speciation profiles for the Uinta Basin resulting from Uinta Basin Composition study (Tran et. al., 2019). These data were provided to the SWG using the guidelines for data developers. Data from individual samples were provided, as well as six composites which were generated from the individual samples with a new statistical technique, isometric log-ratio transformation. The six composite profiles were added to SPECIATE 5.1. Four describe the composition of: i) flashed gas from oil wells, ii) flashed gas from gas wells, iii) raw gas from oil wells, and iv) raw gas from gas wells in the Uinta Basin. The two remaining profiles describe flashed gas from oil and gas wells including carbonyls, which were included in the measurements for a subset of 10 wells (5 gas wells and 5 oil wells).

Two gas profiles that are not in the oil and gas sector were added to SPECIATE 5.1. Profile 95331NEIHP blends existing profile 95331 with volatile organic compound (VOC)-to- hazardous air pollutant (HAP) emission factor (EF) ratios used for estimating HAP emissions for commercial marine vessels (CMV) for the 2017 NEI. Profile DIESEVP is a diesel headspace composite based on the mean of existing profiles 4702 and 4703, which are headspace profiles based on different brands of diesel.

Table 1. Gas Profiles Added to SPECIATE 5.1

PROFILE_CODE	PROFILE_NAME
WIL04	Oil and Gas - Produced Gas Composition from Oil Wells - Williston Basin Montana
WIL03	Oil and Gas - Produced Gas Composition from Oil Wells - Williston Basin North Dakota
WIL02	Oil and Gas - Flash Gas Composition from Tanks at Oil Wells - Williston Basin Montana
WIL01	Oil and Gas - Flash Gas Composition from Tanks at Oil Wells - Williston Basin North Dakota
CMU01	Oil and Gas - Produced Gas Composition from Gas Wells - Central Montana Uplift - Montana
PAGAS01	Oil and Gas-Produced Gas Composition from Gas Wells-Greene Co, PA
PAGAS02	Oil and Gas-Produced Gas Composition from Gas Wells-Butler Co, PA
PAGAS03	Oil and Gas-Produced Gas Composition from Gas Wells-Washington Co, PA
UTUBOGA	Flash Gas from Oil Tanks - Composite Uinta basin
UTUBOGB	Flash Gas from Condensate Tanks - Composite Uinta basin
UTUBOGC	Raw Gas from Oil Wells - Composite Uinta basin
UTUBOGD	Raw Gas from Gas Wells - Composite Uinta basin
UTUBOGE	Flash Gas from Oil Tanks - including Carbonyls - Composite Uinta basin
UTUBOGF	Flash Gas from Condensate Tanks - including Carbonyls - Composite Uinta basin
95331NEIHP	Marine Vessel - 95331 blend with CMV HAP
DIESEVP	Diesel Headspace Vapor Composite

Eighteen PM profiles were added to SPECIATE 5.1; one is correction to a profile that had been originally added to SPECIATE 4.2. Seventeen are new profiles that were added from two different studies. One of the studies produced a sugar cane burning $PM_{2.5}$ profile; this is the same study that was used to create a TOG profile for sugar cane burning in SPECIATE 5.0, therefore this $PM_{2.5}$ profile is a sibling to the SPECIATE 5.0 TOG profile. The second study was conducted by EPA's Office of Research and Development (ORD). Field and laboratory $PM_{2.5}$ composition data were collected for wildland fires using fuels from different regions of the United States. 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; reference information for these profiles will be added in the next version of SPECIATE. The SPECIATE work group (SWG) thought it would be useful to add these profiles to SPECIATE 5.1 as soon as the data were available to fill the need addressed by the assessment of profile needs addressed in Bray, et. al.

Table 2. PM Profiles Added to SPECIATE 5.1

PROFILE_CODE	PROFILE_NAME	TYPE
4746a	Diesel Exhaust - Bus at -10 oC, 4-stroke, oxidation catalyst - corrected	PM
SUGP02	Sugar Cane Pre-Harvest Burning Mexico	PM-AE6
95793	Forest Fire-Flaming-Oregon AE6	PM-AE6
95794	Forest Fire-Smoldering-Oregon AE6	PM-AE6
95795	Grass Fire-Flaming-Oregon AE6	PM-AE6
95796	Grass Fire-Smoldering-Oregon AE6	PM-AE6
95797	Grass Fire-Field-Oregon AE6	PM-AE6
95798	Forest Fire-Flaming-North Carolina AE6	PM-AE6
95799	Forest Fire-Smoldering-North Carolina AE6	PM-AE6

PROFILE_CODE	PROFILE_NAME	TYPE
95800	Grass Fire-Flaming-Kansas AE6	PM-AE6
95801	Grass Fire-Smoldering-Kansas AE6	PM-AE6
95802	Grass Fire-Field-Flaming-Kansas AE6	PM-AE6
95803	Grass Fire-Field-Smoldering-Kansas AE6	PM-AE6
95804	Forest Fire-Flaming-Montana AE6	PM-AE6
95805	Forest Fire-Smoldering-Montana AE6	PM-AE6
95806	Peat Fire-Smoldering-Minnesota AE6	PM-AE6
95807	Forest Fire Understory-Flaming-Minnesota AE6	PM-AE6
95808	Forest Fire Understory-Smoldering-Minnesota AE6	PM-AE6

In addition to gas and PM profiles, profiles of type "OTHER" were added to SPECIATE 5.1. Fifty-eight mercury speciation profiles were added based on a technical memo documenting the development of Mercury Speciation Factors for EPA's Air Emissions Modeling Programs. While some of these profiles are already in SPECIATE, they are not easily matched to the air emissions modeling platform profiles. Their inclusion in SPECIATE with profile codes that are consistent with those used in the modeling platform improves the accessibility and transparency of these data. The set of mercury profiles also includes data for some categories that were not previously available in the modeling platform or in SPECIATE.

Table 3. Mercury Profiles added to SPECIATE 5.1

Profile Code	Profile Description
EGUBIN00	Bituminous Coal, Coal Gasification
EGUBIN01	Bituminous Coal, PC Boiler with ESP-CS
EGUBIN02	Bituminous Coal and Pet. Coke, PC Boiler with ESP-CS
EGUBIN03	Bituminous Coal, PC Boiler with SNCR and ESP-CS
EGUBIN04	Bituminous Coal, PC Boiler with ESP-HS
EGUBIN05	Bituminous Coal, PC Boiler with PM Scrubber
EGUBIN06	Bituminous Coal, PC Boiler with Dry Sorbent Injection and ESP-CS
EGUBIN07	Bituminous Coal, PC Boiler with FF Baghouse
EGUBIN08	Bituminous Coal, PC Boiler with SDA/FF Baghouse
EGUBIN09	Bituminous Coal, PC Boiler with SCR and SDA/FF Baghouse
EGUBIN10	Bituminous Coal, PC Boiler with ESP-CS and Wet FGD
EGUBIN11	Bituminous Coal, PC Boiler with ESP-HS and Wet FGD
EGUBIN12	Bituminous Coal, PC Boiler with FF Baghouse and Wet FGD
EGUBIN13	Subbituminous Coal, PC Boiler with ESP-CS
EGUBIN14	Subbituminous Coal, PC Boiler with ESP-HS
EGUBIN15	Subbituminous Coal, PC Boiler with FF Baghouse
EGUBIN16	Subbituminous Coal, PC Boiler with PM Scrubber
EGUBIN17	Subbituminous Coal, PC Boiler with SDA/ESP
EGUBIN18	Subbituminous Coal, PC Boiler with SDA/FF Baghouse
EGUBIN19	Subbituminous Coal, PC Boiler with ESP-CS and Wet FGD
EGUBIN20	Subbituminous Coal, PC Boiler with ESP-HS and Wet FGD
EGUBIN21	Lignite Coal, PC Boiler with ESP-CS
EGUBIN22	Subbituminous Coal, Cyclone Boiler with PM Scrubber
EGUBIN23	Subbituminous Coal/Pet. Coke, Cyclone Boiler with ESP-HS
EGUBIN24	Lignite Coal, Cyclone Boiler with ESP-CS
EGUBIN25	Bituminous Coal/Pet. Coke, Fluidized Bed Combustor with SNCR and FF Baghouse
EGUBIN27	Bituminous Waste, Fluidized Bed Combustor with FF Baghouse
EGUBIN28	Lignite Coal, Fluidized Bed Combustor with ESP-CS

Profile Code	Profile Description
EGUBIN29	Lignite Coal, Fluidized Bed Combustor with FF Baghouse
EGUBIN30	Anthracite Waste, Fluidized Bed Combustor with FF Baghouse
EGUBIN31	Bituminous Coal, Stoker Boiler with SDA/FF Baghouse
EGUBIN33	Lignite Coal, PC Boiler with ESP-CS and FF Baghouse
EGUBIN34	Lignite Coal, PC Boiler with SDA/FF Baghouse
EGUBIN35	Lignite Coal, PC Boiler with PM Scrubber
EGUBIN36	Lignite Coal, PC Boiler with ESP-CS and Wet FGD
EGUBIN37	Bituminous Coal, Cyclone Boiler with Mechanical Collector
EGUBIN38	Bituminous Coal/Pet. Coke, Cyclone with ESP-CS and Wet FGD
EGUBIN39	Lignite Coal, Cyclone Boiler with SDA/FF Baghouse
EGUBIN40	Subbituminous Coal, Fluidized Bed Combustor with SNCR and FF Baghouse
EGUBIN41	Subbituminous Coal/Bituminous Coal, PC Boiler with ESP-CS
EGUBIN42	Subbituminous Coal/Bituminous Coal, PC Boiler with ESP-HS
EGUBIN43	Bituminous Coal/Pet. Coke, PC Boiler with FF Baghouse
EGUBIN44	Bituminous Coal/Subbituminous Coal, PC Boiler with FF Baghouse
HGCEM	Portland Cement Kiln Exhaust
HGCLI	Cement Clinker Cooler
HGCMB	Fuel Combustion
HGCRE	Cremation (humans and animals)
HGELE	Elemental (dental alloy, reagents, fluorescent lamp breakage, Portland cement raw material handling operations, artisanal scale gold mining)
HGGEO	Geothermal power plant (non-binary)
HGGLD	Large-scale gold production
HGHCL	Chlor-Alkali Plants
HGINC	Waste Incineration
HGIND	Industrial (average of non-comb. profiles)
HGMD	Mobile Diesel
HGMET	Metal Production (iron and steel production, non-ferrous metal production)
HGMG	Mobile Gasoline
HGMWI	Medical Waste Incineration
HGPETCOKE	Petroleum Coke Combustion

B. Database updates

In SPECIATE5.1 structural changes were made to profile references and their metadata. In addition, new fields were included in the PROFILES table, SPECIES table, and SPECIES_PROPERTIES tables. SPECIATE tables, queries and macros are described in Chapter II. SPECIATE field names/changes are described in Table 8 of Chapter II. The remainder of this section provides an overview to the key changes to the reference and species properties information.

In SPECIATE 5.1 the reference structure was redesigned to more easily accommodate situations where a single reference could be used by many different profiles, and where a single profile can have several references (though we limit the number of references used by a single profile to three). In previous versions of SPECIATE, references were manually repeated across the different profiles they applied to (sometimes with typographical errors).

The new structure was implemented by adding two key new tables. One table maps a profile to one or more reference codes, while the second table maps each reference code to the reference description, reference citation and reference link. The new reference code field (Ref_Code) was developed to map each profile to one or more reference without having to type the full reference (since so many profiles share the same reference, using the full reference for the crosswalk is prone to typographical errors).

Given that the new structure provides multiple references per profile, i.e., more than one row per profile, we developed a macro to concatenate the reference information to display as one row per profile. This way, a SPECIATE user could see all references for a given profile along with all of the other meta data for that profile. The resultant concatenated table emulates the "KEYWORD_REFERENCE" table in previous versions of SPECIATE, which had been prepared manually without the supporting reference structure described above. The concatenation macro also takes the links for each of the references (up to a maximum of three) and puts them into separate fields labeled LINK1, LINK2, LINK3. The separation of these links allows them to be used as hyperlinks in both the SPECIATE database and the browser. Any time a new profile is added, the macro must be run to update the concatenated reference information which is used for the view profile queries in the database.

This reference structure change required us to create the new Ref_Code field. We generally used the last name of first author and date, where available. We also needed to create a unique reference description for references where a different description was used depending on the profile. Since the new structure requires reference description to be a property of the reference, we had to change many reference descriptions to conform to the new structure. We discovered that in many situations, the reference description was not a reference description but rather additional profile notes or a repeat of the profile notes, so this change involved improving the reference description for profiles added in previous SPECIATE versions.

In addition, we checked and updated many reference links used by profiles added in previous SPECIATE versions.

The species properties information changes included numerous new fields, including an alternative Chemical Abstract Service (CAS) field (ALT_CAS) that includes previously used CAS numbers for the species. We also added a link to EPA's CompTox Chemistry Dashboard database (Williams et al., 2017) through the addition of the DSSTox Substance ID, which is provided in the new DSSTox_ID field. Another vapor pressure field (VP_Pascal_OPERA) based on the OPEn structure-activity/property Relationship App (OPERA) approach (Mansouri et. al., 2018) was also added to the database using the batch search feature of the EPA's CompTox Chemistry Dashboard. We also updated the vapor pressures using the Estimation Program Interface (EPI) method to be consistent with the data developed for the Speciation Tool by Ramboll (Ramboll, 2018), and we revised some of the UMANSYSPROP model values (e.g., removed some values for pure metals or pure carbon that did not make sense).

Similar to the table of concatenated profile references, we created a derived table providing concatenated species names, separated by the double pipe symbol (||). This allows use of one row per SPECIES which is needed for the View queries and browser. This was done by expanding the species synonym table to include all species names and numerous synonyms. We then used a concatenation macro to create a derived table (tblSpeciesAndConcatSynonyms) in which all synonyms for a specific SPECIES_ID are concatenated into a SPECIES_NAMES field. As more synonyms are added, the macro could be run to update the SPECIES_NAMES field in the derived table.

We also worked with the Substance Registry System (SRS) to identify species duplicates. We added a field indicating whether a SPECIES_ID contains a duplicate (DUPLICATE_ID). We populated it for each duplicate pair or triplicate with the value "dup" concatenated with the SPECIES_ID that is the lowest value. In addition, for the higher SPECIES_ID for which there is a lower SPECIES ID duplicate, we added the "-duplicate" to the SPECIES_NAME field. We developed a query to produce a list of duplicates ("Identify Species Duplicates")- this list contains only the higher SPECIES_ID value(s) of the duplicate or triplicate duplicate group.

Our work with the SRS team also highlighted the need for corrections to a handful of species with 'TMS' in the name. Upon further research, it was determined that these species, which were from older versions

of the database, were trimethylsilyl derivatives of those compounds because derivatization was used to measure polar compounds using gas chromatography-mass spectrometry (GC-MS). Upon discussion, it was determined that the 'TMS' would be dropped from those species names and those species with duplicates would be labeled as such.

Another metadata improvement we made was to populate the categorization fields for the remaining PM coarse PM profiles in the SPECIATE database. The categorization fields were first added in SPECIATE 5.0 (see Chapter II E of the SPECIATE 5.0 documentation), and we in that version we did not populate them for some PM profiles (i.e., where the upper size limit is more than 2.5 microns).

C. QSCORE updates

The 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. We changed the ranks for the QSCORE point values and added a description field (QSCORE_DESC) that provides a description of the range.

22-30 = excellent (previously 20-30)
16−21 = good (previously 12-19)
8-15 = fair (previously 5-11)
7 or less = poor (previously <4)

When determining the QSCORE for the mercury profiles added to SPECIATE 5.1, we recognized that some of the questions/point values were more geared towards organic gas and PM profiles, particularly the ten points for completeness, considering mercury contains only three species (elemental gas, divalent gas, and particulate). As a result, we added specific guidelines for mercury for that question.

D. Browser updates

The SPECIATE data browser was updated to provide a user guide and additional tables of information. A new table listing the profiles and metadata (but not species information) was added. The browser provides two data tables showing the profiles with species information, 1) a table with all the fields (which is the same as the SPECIATE 5.0 browser) and 2) a new table that has fewer fields. Fields with links (e.g., reference links) are now active links in the browser and will take you to the link location when you click on it.

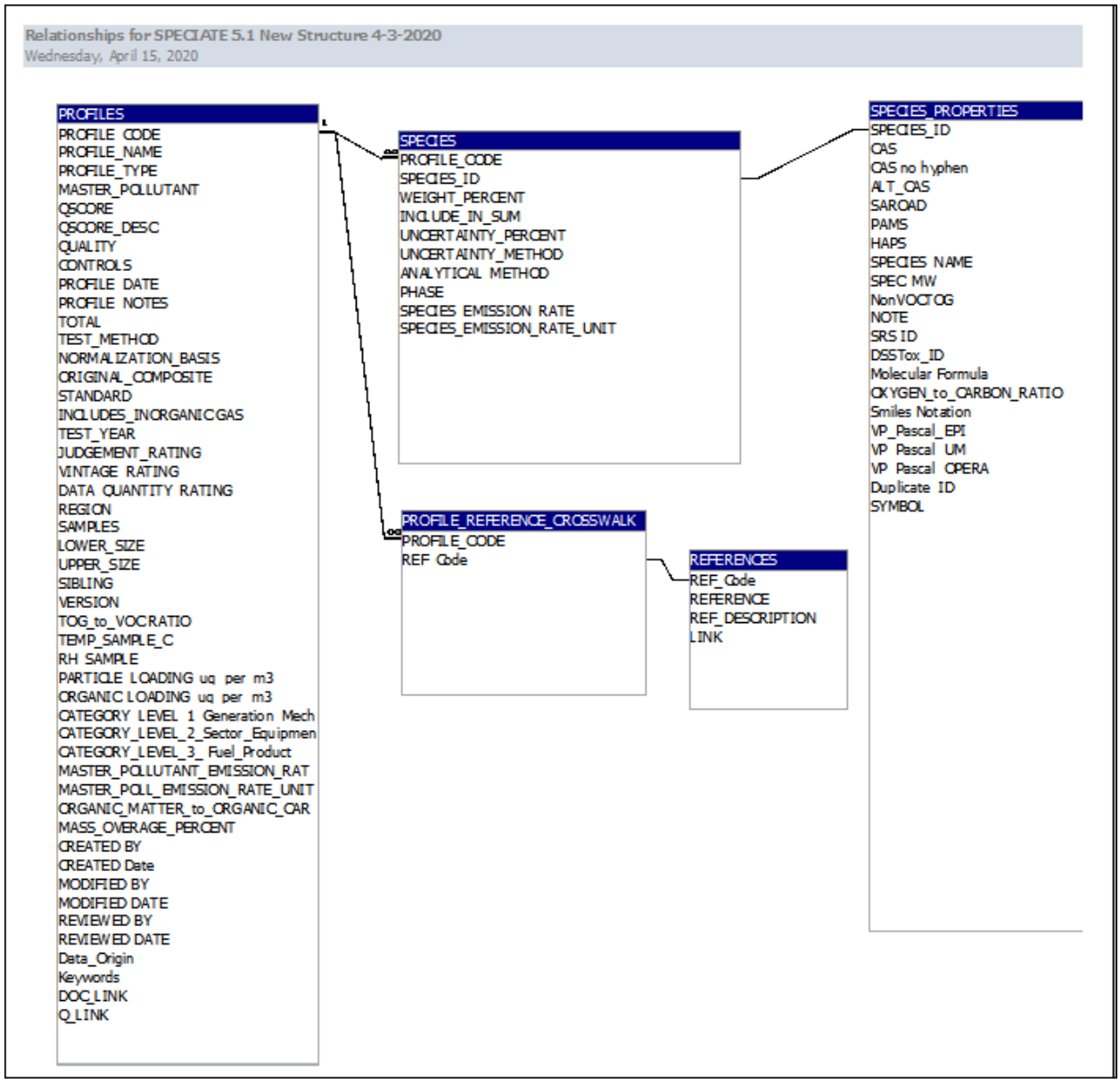
Chapter II. SPECIATE 5.1 Database – Tables, Queries and Fields

This chapter describes the organization of the SPECIATE 5.1 database and describes the tables, queries, macros and field names.

The SPECIATE 5.1 database is a data repository housed in a Microsoft Access® relational database file that contains the new profiles from the current SPECIATE 5.0 process and all previous versions. In order to use the SPECIATE 5.0 database, Microsoft Access 2002® or a newer version must be installed. The current SPECIATE database and other relevant documentation can be downloaded from EPA's Air Emissions Modeling website: https://www.epa.gov/air-emissions-modeling (*last accessed April 2020*). The direct link to the SPECIATE page is: https://www.epa.gov/air-emissions-modeling/speciate *(last accessed April 2020).* To facilitate inspection of the data by persons without detailed database manipulation skills, queries are available that link the key tables together to allow the user to view the fields in these tables when the queries are run. The View_GAS_Profiles query has a filter to display the organic gas profiles [TOG, reactive organic gas (ROG), VOC, non-methane organic gas (NMOG), non-methane hydrocarbons (NMHC) and GAS-volatility basis set (VBS)]. The View_PM_Profiles query allows the user to view all of the PM profile types (PM, PM-AE6, PM-VBS) and the View Other Profiles query allows the user to view mercury, oxides of nitrogen (NO_X) and other profiles. For SPECIATE 5.1 we added simplified View profiles that provide a minimum number of fields for quick viewing of the species names and weight percents without carrying all the profile and species meta data. We also added export queries that export data needed for input into the Speciation Tool.

The data may also be obtained through the SPECIATE 5.0 data browser on EPA's SPECIATE webpage, *(last accessed June 2020)*, where all fields in the SPECIATE database are provided. This new web-based data browser, 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 Microsoft Excel®. In addition, users of the browser can create custom tables that provide only the fields of interest and can view profile weight percents for individual profiles in a stacked bar chart format for visualization and comparisons across profiles. Users can also download data easily from the browser as a Microsoft Excel ® spreadsheet.

Figure 1. SPECIATE 5.1 Data Diagram



A key difference from SPECIATE 5.0 is that the table called "KEYWORD_REFERENCE" is no longer used, and the fields previously in that table are now in the PROFILES table or are built from fields in the REFERENCES table.

A. Key Tables

The key tables, listed in Table 4 below, provide the data in SPECIATE as well as links to supporting information used to develop the profiles. The fields in these tables are described in Section F of this Chapter.

Table 4. Key Tables in the SPECIATE 5.1 Database

Table	Description	Notes
PROFILES	Provides the metadata for the profiles, other than the references. One row per profile code.	New fields: QSCORE_DESC, DOC_LINK, Q_LINK Fields moved from the KEYWORD_REFERENCE in previous SPECIATE versions: Data_Origin, keywords (formerly called "KEYWORD")
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)	This is a new table
REFERENCES	Provides meta data for the references code. All references are included in this table. One row per reference.	This is a new table
SPECIES	Provides information on the species used in each profile such as the weight percent, emission rate, emission factors, etc. One row per profile/species combination.	New fields: INCLUDE_IN_SUM
SPECIES_PROPERTIES	Key table: identifiers (SPECIES_ID) for each of the pollutants in the database and meta data for each species.	New fields: CAS no hyphen, ALT_CAS, DSSTox_ID, VP_Pascal_OPERA, Duplicate_ID
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

These tables shown in Table 5 provide SPECIATE reference information and SPECIES names in a format that supports the database browser and VIEW queries.

Table 5. Concatenated Tables in the SPECIATE 5.1 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.	- Created by macro mcrConcatenateSpecies By concatenating all the species synonyms, it will be easier to search the in the

	This table also includes the SPECIES_NAME field from the SPECIES_PROPERTIES table.	browser or VIEW queries by species name.

C. Other Tables

There are several other tables (See Table 6) in the SPECIATE database used for reference. These include legacy tables that have been in numerous previous versions of SPECIATE; they are kept with SPECIATE5.1 for completeness.

Table 6. Other Tables in the SPECIATE 5.1 Database

Table	Description	Notes
List of Non-VOC with Species Names	Supporting table: List of Non-VOC species. These are Species that don't meet the regulatory definition of VOC, per the Code of Federal Regulations Title 40, Chapter I Subchapter C,Part 51, Subpart F, §51.100	Revised from SPECIATE 5.0 by adding the name (previously just had CAS) and the SPECIES_ID. Includes some Species that are not in SPECIATE (SPECIES_ID is null for these)
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 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 the 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 many different functions: 1) easy access to readable data summaries that contain descriptions and meta data, 2) provide input to Speciation Tool, and 3) quality assurance (QA). The queries summarized in Table 7 are primarily those that provide readable data summaries and provide 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).

The queries are summarized in Table 7.

Table 7. Queries in the SPECIATE 5.1 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-VBS, 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 much 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 much 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 much fewer fields; enables you to focus on the weight percent information without all the meta data
MasterReferenceListQuery	Allows you to get more detailed reference information associated with each profile. Up to 3 rows per profile (if the profile has 3 references)
IdentifySpeciesDuplicates	Allows you to extract all duplicate rows. This is possible because of the naming convention "-duplicate" in the SPECIES_NAMES.
export_profiles	Query developed by Ramboll to extract select fields from the PROFILES table for use in the Speciation Tool.
export_species	Query developed by Ramboll to extract select fields from the SPECIES table for use in the Speciation Tool.
export_species_properties	Query developed by Ramboll to extract select fields from the SPECIES_PROPERTIES table for use in the Speciation Tool.
IdentifySpeciesDuplicates	Query that makes a list of any duplicates or triplicates (results in a list of the highest SPECIES_ID of the duplicate or triplicate pair).
ProfilesMissingWeights	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	Informational query that lists all profiles after a date entered by the user. Enter date as Month/Day/Year (e.g., 1/1/2020)
SPECIES_SYNONYM_SELECTOR	Supports 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

In addition to adding several species to the PROFILES table, SPECIES table and SPECIES_PROPERTIES table, we also renamed and reused some of the fields. The NAME field was changed to SPECIES_NAME and we dropped the EPA_ID which was duplicative with the SRS ID.

Table 8 provides a list of fields in the SPECIATE database tables and Table 9 provides a list of fields in the Supporting derived tables that are used in the VIEW queries and browser.

Table 8. Descriptive Data Dictionary

Field Name	Data Type	Description	New ?
PROFILES Table			
PROFILE_CODE	Text	Profile Code - alphanumeric. Should be 10 characters or less 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-VBS, 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.	
QSCORE_DESC	Text	Description of the numeric QSCORE rating.	New
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 (see Chapter IV.G “Avoiding Double Counting Compounds” of this report for rationale).	
TEST_METHOD	Long Text	Description of sampling/test method for overall profile	
NORMALIZATION_BASIS	Text	Description of how profile was normalized (see Chapter IV.F for details)	
ORIGINAL_COMPOSITE	Text	Specifies whether the profile is original, composite of SPECIATE profiles or study composite. Allowed values: 'C','O','SC'. The option for study composite, SC, added in SPECIATE5.0, means composite was developed in the study.	
STANDARD	Yes/No	Indicates whether the profile is provided by EPA SPECIATE (standard) or user-added. The database is constructed to allow users to add profiles. 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 document 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 document 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	

Field Name	Data Type	Description	New ?
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_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 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	Revised
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	Revised
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	New
Q_LINK	Text	A link to the QSCORE rating documentation for a profile/reference	New
SPECIES Table			

Field Name	Data Type	Description	New ?
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, No or blank; blank is yes) 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.	New
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	New
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	
SAROAD	Text	Storage and Retrieval of Aerometric Data (SAROAD) code	
PAMS	Yes/No	Is PAMS pollutant? (Yes or No)	

Field Name	Data Type	Description	New ?
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	
SYMBOL	Text	Standard chemical abbreviation	
SPEC_MW	Number	Species molecular weight	
NonVOCTOG	Yes/No	Is this species regarded as a volatile organic compound (VOC)? The VOC definition is from 40 CFR. §51.100	
NOTE	Long Text	Note (notes) about the SPECIES_ID or its properties	
SRS ID	Text	EPA Substance Registry Service (SRS) Chemical Identifier	
DSSTox_ID		Unique Identifier for a chemical in the EPA's Distributed Structure-Searchable Toxicity (DSSTox) Database	
Molecular Formula	Text	Molecular formula	
OXYGEN_to_CARBON_RATIO	Number	Ratio of oxygen atoms to carbon atoms	
Smiles Notation	Text	Smiles notation	
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 from the OPERA model (DOI: 10.1186/s13321-018-0263-1)	
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.	
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	

1 Length – maximum number of characters allowed.

Table 9. Additional Fields Used in the VIEW Queries and Browser

Field Name	Data Type	Description	New?
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	

Field Name	Data Type	Description	New?
VERSION	Text	SPECIATE database version that a profile was added to	
PROFILE_TYPE	Text	Indicates type of profile: PM-AE6, PM-VBS, PM-Simplified, PM, GAS, GAS-VBS 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.	New
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.	New
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 the all synonyms in the Synonyms table for the same SPECIES_ID	

1 Length – maximum number of characters allowed.

CHAPTER III. Future Considerations for the Next Version

We expect to continue to improve and upgrade SPECIATE. We hope to release new versions of SPECIATE when several new items have been incorporated since the last release. In our next release, we plan on enhancing the SPECIATE program and updating the database. This includes:

- Completing the standard operating procedures for developing and enhancing SPECIATE
- Adding more high-priority profiles based on the current assessment and on future assessments that result from using speciation information in air quality modeling
- Improving VBS information and guidelines for several source categories
- Adding relevant species properties to the database such as atmospheric photochemical reactivity to form ozone
- 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
- Continuing to reach out to the research community for providing high quality and high priority data to SPECIATE

References

.

Bray et. al., 2019, 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

Mansouri et al., 2018: Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A. J., OPERA models for predicting physicochemical properties and environmental fate endpoints, *Journal of Cheminformatics* **2012** *10* (10), 1-19, DOI: 10.1186/s13321-018-0263-1.

Ramboll, 2018: Memorandum: "Mapping SPECIATE 5.0 compounds for photochemical models" and excel workbook "EPA_volatility_mappings_for_SPECIATE5_10Oct18.xlsx" To: Alison Eyth and Madeleine Strum, OAQPS, EPA From: Ross Beardsley, Tejas Shah and Greg Yarwood. October 11, 2018 and accompanying workbook. Available at: https://github.com/CMASCenter/Speciation-Tool/ (*last accessed April 2020*)

Ramboll, 2019: Memorandum: "Gas Composition Profile Results from the WRAP OGWG Survey" To: Western Regional Air Partnership Oil (WRAP) and Gas Working Group (OGWG), From: John Grant and Amnon Bar-Ilon. July 18, 2019. https://www.wrapair2.org/pdf/WRAP_OGWG_GasComp_18Jul2019.pdf

Reff et al., 2009: Reff, A., Bhave, P.V., Simon, H., Pace, T.G., Pouliot, G.A., Mobley, J.D., and Houyoux, M., Emissions Inventory of $PM_{2.5}$ Trace Elements across the United States, *Environmental Science and Technology*, 43: 5790–5796, 2009.

Schauer et al., 1998: Schauer, J.J., Kleeman, M.J., Cass, G.R., and Simoneit, B.R.T., Characterization and Control of Organic Compounds Emitted from Air Pollution Sources, California Air Resources Board Contract Number 93-329, 1998.

Schauer et al., 1999: Schauer, J.J., Kleeman, M.J., Cass, G.R, and Simoneit, B.R.T., Measurement of Emissions from Air Pollution Sources. 2. C1 through C30 Organic Compounds from Medium Duty Diesel Trucks, *Environmental Science and Technology*, 33: 1578-1587, 1999.

Tran et al. 2019: Tran, T., Lyman, S., Tran, H., Composition of Volatile Organic Compound Emissions from Oil and Gas Wells in the Uinta Basin. Final Report Submitted to Lexie Wilson, Utah Division of Air Quality, November 2019. https://documents.deq.utah.gov/air-quality/planning/technical-analysis/DAQ-2020-004826.pdf

Williams et al., 2017: Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A. D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I. Wambaugh, J., F., Judson, R. S., Ri, A. D., Mansouri, K., Bac, N.C, G.R, and Richard, A.M., The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, *Journal of Cheminformatics*, 9:61, 1-27, 2017.

US EPA, 2019, SPECIATE 5.0, EPA/600/R-19/098, https://www.epa.gov/air-emissions-modeling/speciate-4.

US EPA, 2020, Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA

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

DATA FROM MEASUREMENTS - (Ideal score of 30)

No.	Question	Total Points
1	Are data from a peer-reviewed publication?	1
2	Is the source U.S. based or does it relate to a National Emissions Inventory (NEI) source?	1
3	Is the author well known or affiliated with a well-known research organization in conducting speciated source measurements?	1
4	Is the emission source current, are up-to-date technologies employed (collection, measurement, analysis)?	1
5	Is subject source identified as “priority” source (see, for example, the study: Bray, et. al.[1])	1
6	Were data collected under an established quality system or sufficiently addressed /are QA/QC activities associated with the data collection/measurements included in the publication or supplementary information?	1
7	Sampling Design	
7a	Is the sampling design discussed logically (logic behind the experiments)?	1
7b	Are the data limitations clear (i.e., can the reviewer easily figure them out or are they explicitly stated)?	1
7c	Are assumptions clearly stated? (e.g., fireplace is representative of typical fireplace found throughout the country	1
7d	Are samples capturing the natural variability of the sources?	1
8	Measurement Methodologies	
8a	Is measurement instrumentation presented or referenced?	1
8b	Are the data limitations clear?	1

8c	Were measurements taken using standard methods [EPA, National Institute of Standards and Technology (NIST)], and applicable/up-to-date technologies, methods, and instrumentation?	1
8d	Are replicate measurements done (duplicate or triplicate)? (Measurement methods using duplicate or triplicate collection implies that the study payed attention to data accuracy, representation and reproducibility. This attention should be viewed as an advantage.)	1
9	Data reduction procedures (statistics)	
9a	Are standard deviations (SDs) presented in the paper? (SDs are needed in the profile or we would contact the PI to get it.)	1
9b	Are SDs acceptable for the type of source and pollutants measured?	1
9c	Are the data ready for listing? (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) (Ideal score of 29)

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

No.	Question	Total Points
1	Are data from a peer-reviewed publication?	1
2	Is the source U.S. based or does it relate to a National Emissions Inventory (NEI) source?	1
3	Is the author well known or affiliated with a well-known research organization in conducting speciated source measurements or analyses?	1
4	Is the emission source current, are up-to-date technologies employed (collection, measurement, analysis)?	1
5	Is subject source identified as "priority" source (see, for example, the study: Bray, et. al.[1])	1
6	Composite Data Development	
6a	Are data based on an established, acceptable methodology?	2
6b	If any of the values or data are based on assumptions or calculations are they clearly documented?	2
6c	Was post-processing used for the data? If so, is it novel, reasonable or widely accepted?	2
7	Is there complete speciation data of PM or organic gas provided? For organic gas, does the profile include a total amount of gaseous organic compounds (TOG), TOG should include (1) methane; (2) alkanes, alkenes and aromatic VOC; (3) alcohols; (4) aldehydes. $PM_{2.5}$ should include critical pollutants such as (1) EC and OC; (2) sulfate/nitrate/$NH4^+$ ions; (3) metals/inorganics. Higher scores are given if PAHs and SVOCs are also available. 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
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 the EPA's repository of volatile organic gas and particulate matter (PM) speciation profiles of air pollution sources. Among the many uses of speciation data, these emission source profiles are used to create speciated emissions inventories for photochemical air quality modeling. In particular, volatile organic compounds (VOC) and $PM_{2.5}$ from emission inventories are speciated into the model species required by the chemical and aerosol mechanisms. This document concerns $PM_{2.5}$ profiles used for air quality modeling.

SPECIATE houses different types of $PM_{2.5}$ profiles: PM, PM-SIMPLIFIED, PM-AE6 and PM-VBS. Profiles of type PM contain all the species provided by a measurement study or a combination (composite) of measurement studies. The other PM profile types are profiles derived from a measurement study or composite in order to provide the species required for an air quality modeling aerosol mechanism. The PM-SIMPLIFIED profiles are for the AE5 aerosol mechanism and are computed by keeping only AE5 species (elemental carbon, organic carbon, sulfate and nitrate) and creating a PM Other species calculated as the 100 – sum of AE5 species' weight percent. The PM-AE6 and PM-VBS aerosol mechanisms have additional species not typically measured and that need to be computed. The PM-AE6 profiles include non-carbon organic mass (PNCOM) and/or particulate water (PH2O). The PM-AE6 profiles were first put into the SPECIATE database in SPECIATE 4.3 (series that begins with profile code "91") by Reff, et.al. (2009)[1]. They were developed by compositing pre-existing SPECIATE profiles, computing PNCOM and PH2O and ensuring mass conservation. Documentation on the steps taken are provided in the Supplemental Information of Reff, et. al. (2009). Since that initial work, there have been additional $PM_{2.5}$ speciation data published in the literature for which SPECIATE developers used Reff's approach so that they could be used in the AE6 aerosol mechanism.

The Speciation Tool is software that creates a complete set of emissions modeling ready profiles for input into SMOKE from the profiles in SPECIATE. For AE6 profiles, the Speciation Tool takes the subset of SPECIATE profiles in which the additional species for AE6 were computed (i.e., "AE6-ready" profiles), maps the species IDs in the SPECIATE profile to the AE6 species names and creates "PM Other" (the remainder of the mass not assigned to AE6 species) as 100 – sum of AE6 species. The ability for the Speciation Tool to create AE6 profiles for the modeling allows SPECIATE developers to include AE6-ready profiles that contain all of the measured species plus the additional calculated species so that only one version of the profile needs to go into the SPECIATE database. EPA is also developing the capability for the Speciation Tool to create AE6 profiles from a profile type of PM. This will allow SPECIATE developers to rely on the Speciation Tool to perform the calculations rather than having to do them manually and put the resulting AE6 profile into the SPECIATE database.

The purpose of this protocol is to document the procedure for creating AE6-ready profiles by either of two ways, 1) performing additional calculations on the measured or composited data and including the AE6-ready profile in SPECIATE, or 2) through running the Speciation Tool on PM profiles in SPECIATE. In either way, the creation of these would largely follow the approach in Reff et. al. This will ensure more consistency and transparency in AE6 profiles that are put into SPECIATE or computed by the Speciation Tool. Changes to the Reff et al. approach will be identified.

[1] "Emissions Inventory of $PM_{2.5}$ Trace Elements across the United States"; Adam Reff, Prakash V. Bhave, Heather Simon, Thompson G. Pace, George A. Pouliot, J. David Mobley, and Marc Houyoux; *Environmental Science & Technology* **2009** *43* (15), 5790-5796; DOI: 10.1021/es802930x (Supplemental Information)

AE6 Species

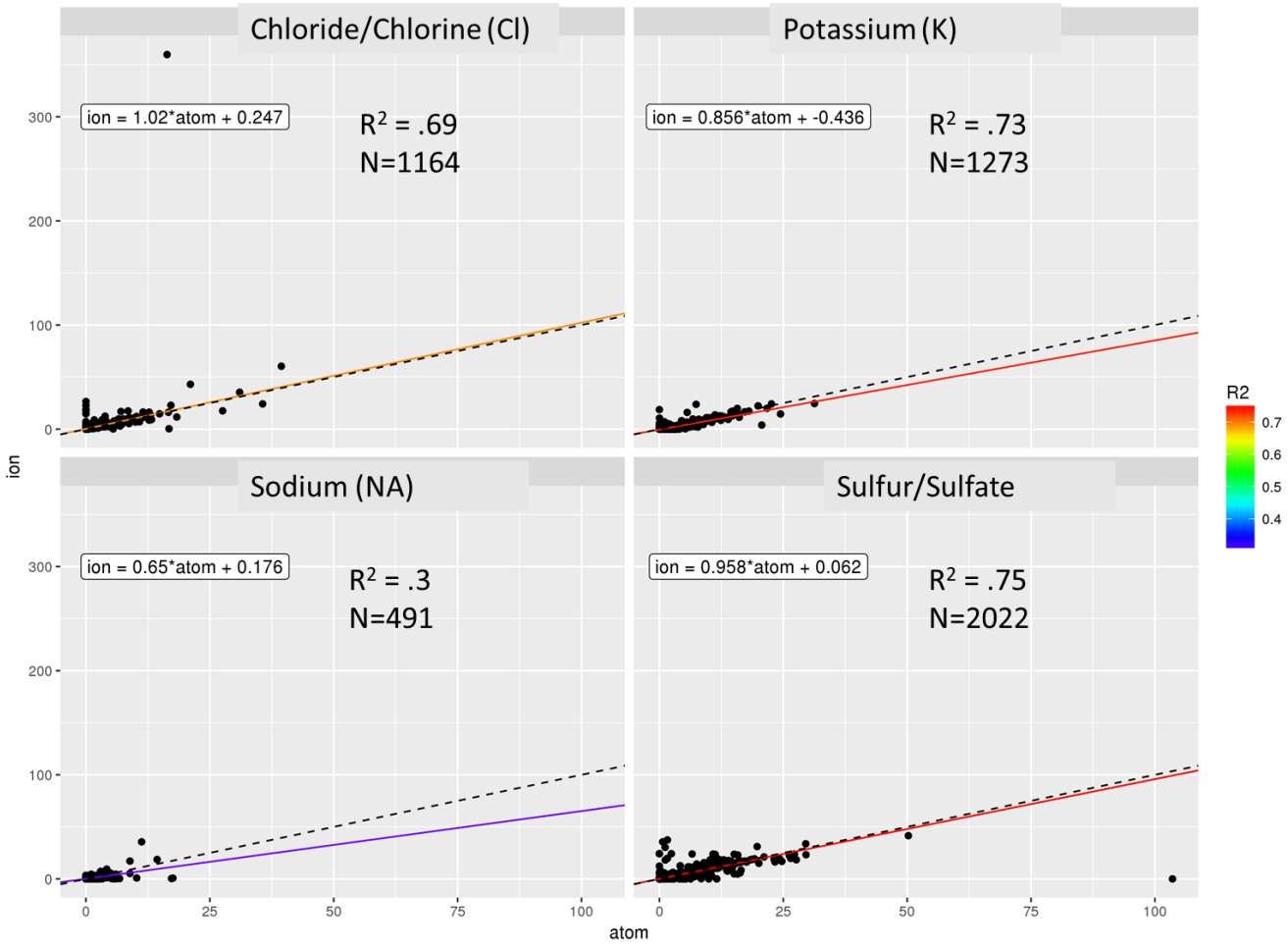
Table B-1 shows the AE6 species used in CMAQ and how the species in the SPECIATE database maps to each of them. This mapping uses the Speciation Tool to create the SMOKE-ready speciation profiles. The comments column indicates which species are typically not found in the literature and thus need to be computed by this protocol. These species are PNCOM and PH20. The comments also describe the gap filling procedures for when the exact species ID is not present in the profile, but a different form (i.e., atomic vs ionic) of the pollutant is available. Gap filling procedures are provided for sodium (Na), magnesium (Mg), chlorine (Cl), calcium (Ca), potassium (K) and sulfate (S). 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.

Table B-1: PM Model Species: AE6

Species Name	Species ID	Species Description, Chemical Formula	Comments/Updated Mappings From
POC	626	Organic carbon	This is obtained from the measurement study, but the weight percent may need to be adjusted downward when creating an AE6 profile if the sum of the species' weight percents exceed 100. The adjustment assumes that the POC included measurement artifacts and is adjusted to achieve mass conservation.
PEC	797	Elemental carbon	
PSO4	699	Sulfate, SO_4^{2-}	Gap filling procedure: If the profile has sulfur (species ID = 700) but no sulfate, then compute sulfate stoichiometrically (SO_4^{2-}=96/32*S)
PNO3	613	Nitrate, NO_3^-	
PNH4	784	Ammonium, NH_4^+	
PNCOM	2669	non-carbon organic matter	Computed from OC based on the (OM to OC ratio) which is a function of the source characteristics and is based on the Reff et. al. (2009) default assignments: Mobile exhaust (combustion): 1.25 Wood combustion sources except wood fired boilers: 1.7 All other sources including wood fired boilers: 1.4 If a particular study uses a different ratio than the default (e.g., the Kansas City study profiles use 1.2 instead of 1.25), then that ratio would be used in place of the default.
PFE	488	Iron	
PAL	292	Aluminum	
PSI	694	Silicon	
PTI	715	Titanium	

Species Name	Species ID	Species Description, Chemical Formula	Comments/Updated Mappings From
PCA	2303	Calcium ion Ca^{2+}	This is a change from the Speciation Tool version 4.0 which used the atom (329). Gap filling procedure: If Species ID 2303 isn't present and Calcium (Species ID 329) is present, then use Species ID 329. If neither Species ID 2303 nor Species ID 329 are present but calcium oxide (CaO Species ID 2847) is present, then Ca^{2+} = 40/56* CaO.
PMG	2772	Magnesium ion Mg^{2+}	This is a change from the Speciation Tool version 4.0 which used the Magnesium atom (Species ID 525). Gap filling procedure: If Species ID 2772 isn't present and Magnesium atom (Species ID 525) is present, then use Species ID 525. If neither Species ID 2772 nor Species ID 525 are present but Magnesium Oxide, MgO (2852) is present, then Mg^{2+}= 24/40* MgO
PK	2302	Potassium ion K^{+}	This is a change from the Speciation Tool version 4.0 which used the Potassium atom (Species ID 669). Gap filling procedure: If Species ID 2302 isn't present and Potassium atom (Species ID 669) is present, then use Potassium atom (Species ID 669).
PMN	526	Manganese	
PNA	785	Sodium ion Na^{+}	This is a change from the Speciation Tool version 4.0 which used the Sodium atom (Species ID 696). Gap filling procedure: If Sodium ion (Species ID 785) isn't present and Sodium atom (Species ID 696) is present, then use Sodium atom (Species ID 696).
PCL	337	Chloride ion	This is a change from the Speciation Tool version 4.0 which used the Chloride atom (Species ID 795). Gap filling procedure: If Chloride ion (Species ID 337) isn't present and Chlorine atom (Species ID 795) is present, then use Chlorine atom (Species ID 795).
PH2O	2668	Water	Computed for non-combustion and non-high temperature sources
PMO	2671	$PM_{2.5}$ not in other AE6 species	Optional for PM-AE6 profile in SPECIATE but computed in the Speciation Tool. (can compute or leave out). Computed from 100-sum of other species.

Figure B-1. Regression of Ion and Atom Forms for Profiles which Contain Data for Both



The comments column also indicates gap filling techniques to use if the measurements from the literature are in a different form (i.e., atomic instead of ionic) than the AE6 species. Also, POC and PNCOM species may need to be adjusted from the values in the paper. Adjustment of these may be needed to account for artifacts on the organic carbon (OC) measurement, or when the sum of weight percent across all unique species exceeds 100%. It should be noted that in order for a profile to be used in air quality modeling using the AE6 mechanism, it 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, PH20. 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 0 PH2O 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

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

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_{El} 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.

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

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.

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

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

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

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

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

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

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
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.