U.S. EPA's S2S-Tool v2 (SPECIATE-to-SMOKE Tool): User's Guide

Primary Author: Karl Seltzer (U.S. EPA) Date of this version: 09/06/2023

Contributors: Madeleine Strum, Alison Eyth, Heather Simon, Havala Pye, Ben Murphy (U.S. EPA)

Table of Contents

List of Acronyms and Abbreviations

AQM	Air Quality Model
CAMx	Comprehensive Air Quality Model with Extensions
CAP	Criteria Air Pollutant
CB6	Carbon Bond version 6
CH_4	Methane
CMAQ	Community Multiscale Air Quality model
CRACMM	Community Regional Atmospheric Chemistry Multiphase Mechanism
CROC	Condensible Reactive Organic Carbon
GROC	Gaseous Reactive Organic Carbon
HAP	Hazardous Air Pollutants
IVOC	Intermediate Volatility Organic Compounds
MO	Metal-Bound Oxygen
NBAFM	Naphthalene-Benzene-Acetaldehyde-Formaldehyde-Methanol
NEI	National Emissions Inventory
NMOG	Non-Methane Organic Gas
O_3	Ozone
OM	Organic Matter
$PM_{2.5}$	Particulate Matter with diameter smaller than 2.5 µm
PNCOM	Primary Non-Carbon Organic Matter
POA	Primary Organic Aerosol
POC	Primary Organic Carbon
RACM2	Regional Atmospheric Chemistry Mechanism version 2
ROC	Reactive Organic Carbon
S2S-Tool	SPECIATE-to-SMOKE Tool
SAPRC07	Statewide Air Pollution Research Center 2007 chemical mechanism
SCC	Source Classification Code
SMOKE	Sparse Matrix Operator Kerner Emissions
SOA	Secondary Organic Aerosol
SVOC	Semi-Volatile Organic Compounds
SV-POA	Semi-Volatile Primary Organic Aerosol
TOG	Total Organic Gas
TOM	Total Organic Matter

VOC	Volatile Organic Compound

1 Background

Air pollutant inventories are a key input variable for photochemical air quality models (AQMs) such as the Community Multiscale Air Quality model (CMAQ). At the U.S. EPA, inventories are compiled in-house, with the National Emissions Inventory[1] (NEI) being the most prominent and widely used within the U.S. air quality modeling community. The NEI is typically generated at the annual, county-level (exceptions include point source emissions), and all emissions are assigned to a Source Classification Code[2] (SCC). Each SCC represents a unique, category-specific process or function that emits air pollutants. Emissions must be further processed for input into AQMs, and one such process is speciating volatile organic compound (VOC) emissions and fine particulate matter ($PM_{2.5}$).

VOC emissions include "any compound of carbon, excluding carbon monoxide, carbon dioxide, carbonic acid, metallic carbides or carbonates, and ammonium carbonate, which participates in atmospheric photochemical reactions" (40 CFR 51.100(s)). A limited number of airborne organic compounds have been determined to have negligible photochemical reactivity and are excluded from the definition of VOC[3], but hundreds-of-thousands of individual species (Goldstein and Galbally, 2007) that span more than 10-orders of magnitude in volatility (Robinson et al., 2007) remain atmospherically relevant. Representing each individual species would be computationally impractical and the species information must be condensed for regulatory modeling purposes. Condensed chemical mechanisms, such as Carbon Bond 6 (CB6, Yarwood et al., 2010), SAPRC07 (Carter, 2010), the Regional Atmospheric Chemistry Mechanism version 2 (RACM2, Goliff et al., 2013), and the Community Regional Atmospheric Chemistry Multiphase Mechanism (CRACMM, Pye et al., 2023), are developed to represent this complex chemistry in AQMs. In these mechanisms, organic chemistry is reduced to select explicit and lumped species or functional groups that represent both directly emitted compounds and secondary compounds that form from reactions in the atmosphere, as well as corresponding chemical reactions.

The source-level speciation of VOC emissions varies considerably, and expertise is required to match organic profiles to VOC sources. SPECIATE[4] is the U.S. EPA's repository of organic gas and particulate matter (PM) speciation profiles of air pollution sources and v5.3 currently houses more than 3000 individual compounds and 2500 gas-phase profiles. Gas-phase organic profiles within SPECIATE are typically generated as Total Organic Gas (TOG) profiles, which include all exempt organics, in addition to regulatory defined VOCs. SCCs are mapped to SPECIATE profiles using a speciation cross-reference file (GSREF), which serves as input for the Sparse Matrix Operator Kerner Emissions (SMOKE[5]) Modeling System. Two other important input files for SMOKE include the "speciation profile file," or GSPRO, and the "pollutant-to-pollutant conversion factors file," or GSCNV. The GSPRO contains the factors that are used to disaggregate inventory pollutant emissions (e.g., VOC) into emissions of chemical mechanism-specific species required by an AQM. The GSCNV contains conversion factors that are used to translate inventory pollutant emissions (e.g., VOC) into other, related pollutants (e.g., TOG).

[1] https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei
[2] https://sor-scc-api.epa.gov/sccwebservices/sccsearch/docs/SCC-IntroToSCCs_2021.pdf
[3] https://www.epa.gov/ground-level-ozone-pollution/complete-list-voc-exemption-rules
[4] https://www.epa.gov/air-emissions-modeling/speciate
[5] https://www.cmascenter.org/smoke/

To facilitate the generation of the GSPRO, "mechanism mapping" must be performed. This process maps each of the individual compounds within SPECIATE to the appropriate chemical mechanism-specific specie(s), and this mapping is housed in the mechanism_4Import (hereafter mech4import) file for gases and the mech_PM file for particulate matter. Both files are input into the Speciation Tool[6], which generates GSPRO and GSCNV files.

The SPECIATE-to-SMOKE Tool (S2S-Tool) performs both the mapping process and the functions of the prior Speciation Tool, which generated GSCNV and GSPRO files. In this User's Guide, the description, input files, output files, and functionality of the S2S-Tool and mechanism-specific mapping scripts are documented. The S2S-Tool was developed and will be maintained by the U.S. EPA. The S2S-Tool both translates the most common functions of the existing Speciation Tool into a *python*-based collection of modules and adds new functionality to enable state-of-science speciation representation (e.g., semi-volatile primary organic aerosol source-based assignments). In addition, the description, methods applied, mechanisms considered, and assumptions for the in-house mechanism mapping scripts are included. This mapping process was previously performed manually, and these new scripts both streamline the mapping process and make it more transparent.

Below, the overall process of translating emission methods to photochemical modeling input is illustrated.

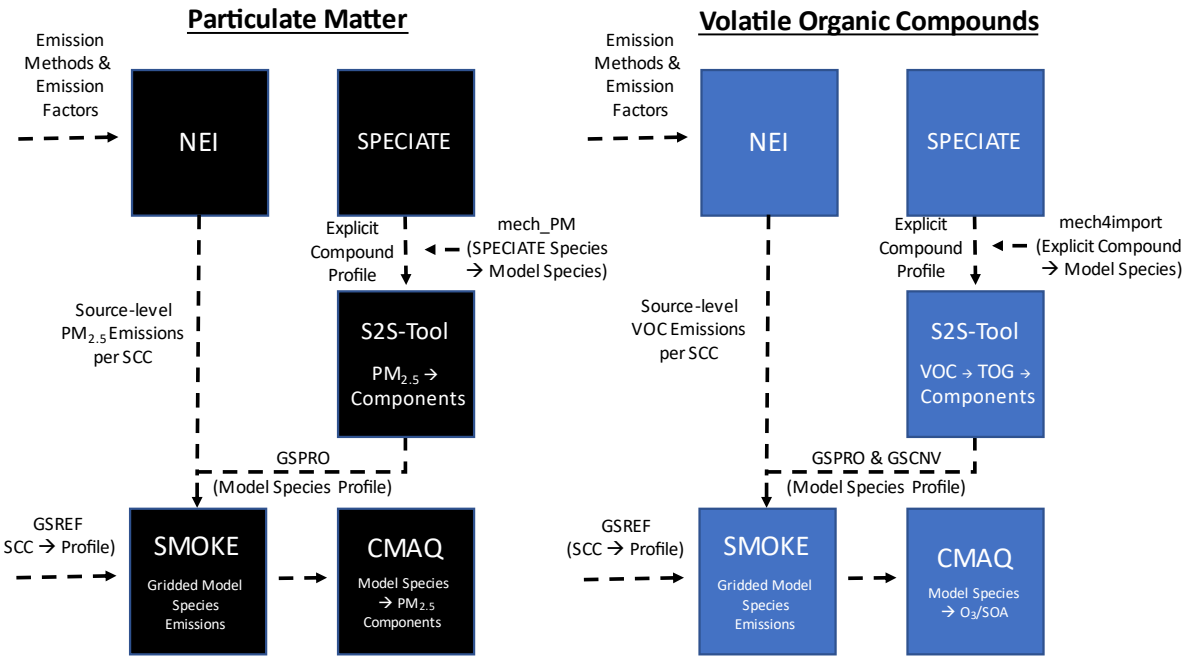


Figure 1: Workflow for generating gridded, model ready emissions for CMAQ at the U.S. EPA.

As shown, the S2S-Tool, plays a key role in generating gridded, model ready emissions and is housed between SPECIATE and SMOKE. The mechanism-specific mapping scripts used to generate the mech4import files currently serve as input for the S2S-Tool. In the future, these scripts will be incorporated into the Tool's infrastructure.

[6] https://github.com/CMASCenter/Speciation-Tool

1.1 Modeling Platform Speciation

The NEI houses annual emissions information for Criteria Air Pollutants (CAPs), precursors to CAPs (e.g., NH_3, VOC), and Hazardous Air Pollutants (HAPs). Hereafter, both CAPs and CAP precursors will collectively be referred to as CAPs. Often, pollutants that are classified as a HAP are also a CAP (e.g., benzene and formaldehyde are both HAPs and VOCs; lead and arsenic are both HAPs and $PM_{2.5}$ components). In a few instances, there are pollutants that are a HAP but not a CAP (e.g., methyl chloroform is a HAP but exempt from the definition of VOC). As such, HAPs in a modeling platform, which is defined as a dataset of gridded, model ready emissions for an AQM, can either be retrieved directly from the NEI or generated through speciation of VOC or $PM_{2.5}$.

Currently, there are two types of modeling platforms generated by the U.S. EPA. In both, CAP and HAP emissions are retrieved directly from the base inventory. However, the two platform types vary in the sector-specific speciation approaches and under what circumstances HAP emissions from the inventory are directly incorporated into the modeling platform. In a "CAP-only" modeling platform, select HAP emissions are retrieved from the base inventory for the mobile (i.e., both onroad and nonroad), nonpoint, and point fire sectors, while HAP emissions are generated through speciation of CAP emissions for the point sector. In a "HAP-CAP" modeling platform, select HAP emissions are retrieved from the base inventory for all sectors. The differences between the two speciation methods are predominately driven by the granularity of the inventory data (e.g., HAP emissions from point sources are often at the facility-level, whereas CAP emissions are at the process-level).

When HAP emissions are generated through CAP emission speciation, the method is defined as "criteria" speciation. When HAP emissions are pulled directly from the base inventory, the non-HAP mass of the CAP emissions can either use an "integrate" or a "no-integrate" speciation profile to characterize the non-HAP mass. In an "integrate" profile, the explicit HAPs within the chemical mechanism for the modeling platform are removed from the SPECIATE profile and the remaining mass is re-normalized to 100%, whereas a "no-integrate" profile does not feature the re-normalization step. Thus the "integrate" approach results in retaining total VOC or $PM_{2.5}$ mass in the inventory while the "no-integrate" approach can result in the sum of VOC or $PM_{2.5}$ species mass being larger or smaller than the VOC or $PM_{2.5}$ emissions in the underlying NEI. Below, a conceptual overview of the various speciation methods is illustrated.

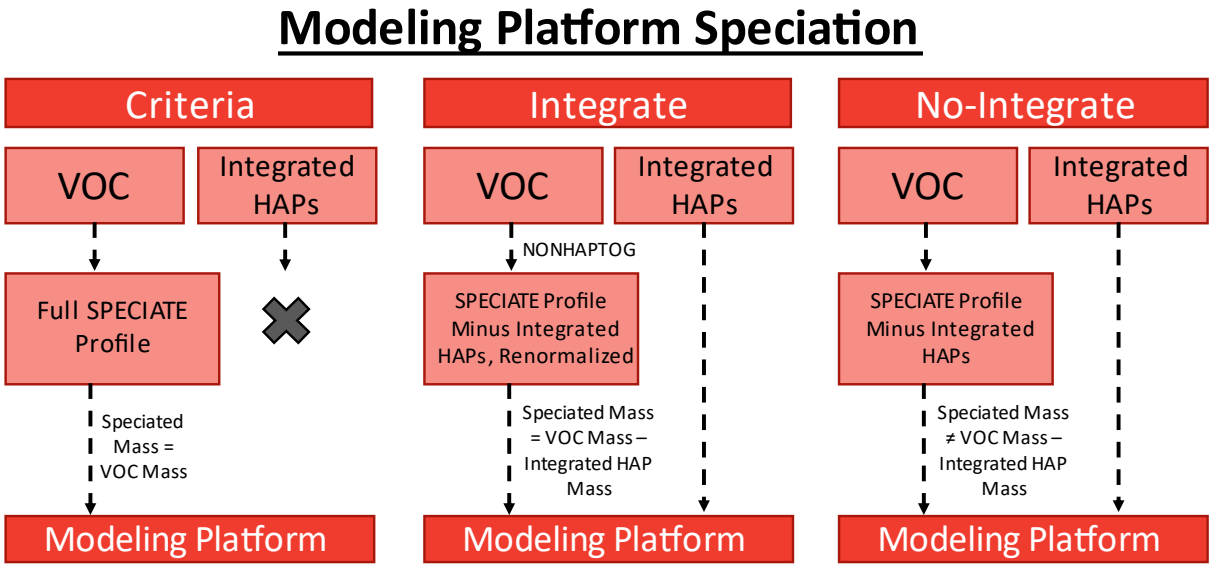


Figure 2: Workflow for speciating gaseous organic emissions at the U.S. EPA.

"No-integrate" profiles are almost exclusively used for point sector emissions in HAP-CAP modeling platforms. In the base inventory, some point source HAP emissions are retrieved from sources for which HAP and CAP emissions are not reported consistently (e.g., HAPs may be reported at the facility-level). Since all CAP emissions are reported at the process-level, it is sometimes not possible to reconcile the two dataset types (i.e., facility-level and process-level). In these circumstances, the inventoried HAP emissions are preferentially used over speciated, process-level CAP emissions. To avoid a potential double counting of HAP emissions, the HAPs retrieved from the base inventory are dropped from the SPECIATE profile and the profile is not renormalized since facility-level vs. process-level emission proportions are not know. Therefore, to incorporate facility-level HAP emissions, mass of CAP emissions at the process-level is not conserved.

1.2 Updates for v2

Several small changes were made in S2S-Tool v2, including new input files that accommodate SPECIATE v5.3, other miscellaneous updates to input files, incorporation of new methods for reporting the molecular weight of mechanism-species species in the GSPRO output files, and refinements to select mechanism mappers.

1.2.1 Input Files

To facilitate the updated *effective molecular weight* methods (see Section 1.2.2 and Section 2.5.4.1), the mechanism_mw file (see Section 2.1.8) was added. With this method update, the carbons file was no longer needed and has been deleted from the input directory. In addition, "ALK" was appended to the POA_mapping file (see Section 2.1.10) for all semi-volatile POA species for PM_CR1 files, which is consistent with the mechanism. Finally, all SPECIATE files (i.e., export_profiles, export_species, and export_species_properties) were updated to reflect SPECIATE v5.3 additions.

1.2.2 Modules

Two main updates were made to the modules within S2S-Tool for v2. First, the *gen_gspro_voc* function (see Section 2.5.4.1) was updated to report the mechanism-defined molecular weight of each model species. Previously, the molecular weight reported in each GSPRO was calculated using the representative

compound of each species in SPECIATE. However, issues can arise for highly lumped model species. By incorporating the mechanism-defined molecular weight into the calculations, these issues are reduced. Second, the *gen_gspro_pm* function (see Section 2.5.4.2) was updated to remove functionality of the PM-AE8 aerosol module. This aerosol module within CMAQ was temporary and no support is provided for its use, with all future updates focused on the development of CRACMM.

1.2.3 Mechanism Mappers

Several small updates were made to the mechanism mappers, with nearly all occurring for the Carbon Bond implementation of these scripts. First, all mechanism mappers were updated to account for a minor error in the counting of C-C double bonds in SMILES strings. Second, a mapper for CB7_CF2 was developed (see Section 3.3.3), which is identical to the CB6R4_CF2 mechanism mapper, except for the addition of alpha-pinene (APIN) as an explicit model species. Third, as recommended by the mechanism developer, a rule was added to all Carbon Bond mechanism mappers that limits 1 functional group assignment for each species, with the preferential selection of the functional group following this order: TOL > XYL > IOLE > OLE > ALDX > KET. Fourth, as recommended by the mechanism developer, a rule was added to all Carbon Bond mechanism mappers that eliminates occurances of molar assignments where OLE / PAR > 1. In those circumstances, OLE and PAR are recalculated as follows: OLE = (PAR + 2 OLE)/3 and then PAR = OLE. Finally, as recommended by the mechanism developer, a rule was added to all Carbon Bond mechanism mappers that eliminates occurances of molar assignments where KET / PAR > 0.333. In those circumstances, KET and PAR are recalculated as follows: KET = (PAR + KET)/4 and then PAR = 3*KET.

2 Contents of the S2S-Tool

The S2S-Tool is distributed through the U.S. EPA's GitHub and is entirely written in *python* using common libraries (e.g., pandas and NumPy). In Table 2, the folder structure and contents of the S2S-Tool's GitHub repository is listed, including a description of each file.

Folder	File	Description
.	S2S.main.py	Execution script and where all runtime input is assigned.
.	README.md	A README with an abbreviated description of the Tool.
.	contributing.md	A text file with a brief description of the Tool and contact information.
.	license.md	Software License (MIT License).
./documentation	202308XX_S2S-Tool_UsersGuide.pdf	The official S2S-Tool User's Guide.
./documentation	S2S-Tool_simplified.xlsx	Sample calculations for deriving a GSCNV and GSPRO.
./mechanism_mappers	CB6R3_AE7_mapper.py	Mechanism mapper for CB6R3_AE7 and CB6R5_AE7.
./mechanism_mappers	CB6R4_CF2_mapper.py	Mechanism mapper for CB6R4_CF2.
./mechanism_mappers	CB7_CF2_mapper.py	Mechanism mapper for CB7_CF2.
./mechanism_mappers	CB6R3_AE7_TRACER_mapper.py	Mechanism mapper for CB6R3_AE7_TRACER.
./mechanism_mappers	CRACMMv1.0_mapper.py	Mechanism mapper for CRACMMv1.0.
./mechanism_mappers	SAPRC07RC_AE7_mapper.py	Mechanism mapper for SAPRC07TC_AE7.
./mechanism_mappers	SAPRC07_CF2_mapper.py	Mechanism mapper for SAPRC07_CF2.
./modules	check_inputs.py	Python script that performs several QA checks on input files and runtime input.
./modules	gscnv.py	Python script that generates GSCNV files.
./modules	gspro.py	Python script that generates GSPRO files.
./input	camx_fcrs.profile.csv	Lists profiles for which FCRS is used in lieu of FPRM in CAMx.
./input	export_profiles.csv	Contents from the PROFILES table in SPECIATE.
./input	export_species.csv	Contents from the SPECIES table in SPECIATE.
./input	export_species_properties.csv	Contents from the SPECIES_PROPERTIES table in SPECIATE.
./input	gscnv_append.csv	Lists profiles that are appended to the end of GSCNV files when PROFILE_TYPE is set to CRITERIA to accommodate alternative emission inventories.
./input	mech_pm_ae5_ae6_cr1.txt	Aerosol module specific mapping of $PM_{2.5}$ components in SPECIATE to model species.
./input	mechanism_forImport_SPECIATEv5.3.csv	Chemical mechanism specific mapping of VOC components in SPECIATE to model species.
./input	mechanism_mw.csv	Molecular weight assignments for model species in all supported mechanisms.
./input	oxygen_metal_Ratios.csv	Oxygen-to-metal ratio of common oxides; used to derive metal-bound oxygen.

./input	POA_mapping.csv	Mapping of all OM/OC/NCOM SPECIATE species to aerosol module specific species.
./input	POA_VolatilityBins.csv	Semi-volatile primary organic aerosol volatility assignments for all Category Level 1/2 combinations in SPECIATE.
./input	tbl_tox_MOVES_HAPS.csv	Gaseous integrated species (typically toxics) list from the MOVES mobile model.
./input	tbl_tox_NBAFM.csv	Default gaseous integrated species (e.g., typically NBAFM for CB6) list.
./input	tbl_tox_RESID_PM.csv	$PM_{2.5}$ integrated species list from the MOVES mobile model when speciating RESID_PM), which includes all $PM_{2.5}$ other than EC, SO4, and total organic matter (TOM).
./input	tbl_tox_TOM.csv	$PM_{2.5}$ integrated species list from the MOVES mobile model when speciating TOM.
./output	gscnv.*	GSCNV output from S2S-Tool executions.
./output	gspro.*	GSPRO output from S2S-Tool executions.

Table 1: Folder and file structure of the S2S-Tool's GitHub repository.

Below, the contents and structure of all output and input files, the description of all variables assigned at runtime, configurations currently available in the S2S-Tool, and a description of all modules/functions within the S2S-Tool are described.

2.1 Input Files

The contents and structure of all input files is described below. Note that the headers of each file are case sensitive, and the S2S-Tool will exit an execution if the header information is incorrect.

2.1.1 camx_fcrs.profile File

When compiling the GSPRO for $PM_{2.5}$ profiles, residual mass that is not assigned to an aerosol module species (e.g., PEC, PSO4, ROCP0) is allocated to PMOTHR when AQM is set to CMAQ (see Section 2.4) and FRPM (other primary emissions in the fine mode) when AQM is set to CAMX. However, if the profile is listed in the CAMx FCRS file, the residual $PM_{2.5}$ mass is allocated to FCRS (crustal emissions in the fine mode) rather than FPRM. Currently, all profiles listed in the CAMx FCRS file are composites of dust profiles and these assignments were made by the CAMx developers. This file is composed of a single column, with PROFILE_CODE as the header, that lists profiles where FCRS should be used in place of FPRM.

Column	Header	Description
1	PROFILE_CODE	SPECIATE PROFILE_CODE; profiles for which residual $PM_{2.5}$ is allocated to FCRS.

Table 2: File format for CAMx FCRS file.

2.1.2 export_profiles File

The export_profiles file contains profile properties that inform the output of the S2S-Tool. This information is stored in SPECIATE as components of the PROFILES table and is retrieved using the export_profiles query. Following the execution of the SPECIATE query, additional (i.e., manual) formatting may be necessary before running the S2S-Tool, including conversion of the file extension to ".csv". All SPECIATE profiles should be included within the export_profiles file, and the S2S-Tool filters the contents, which depend on the user-specified OUTPUT selected at runtime (see Section 2.4). This dataset is comma

delimited and composed of six columns: PROFILE_CODE, PROFILE_TYPE, ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO, TOGVOC, CATEGORY_LEVEL_1_Generation_Mechanism, and CATEGORY_LEVEL_2_Sector_Equipment.

Column	Header	Description
1	PROFILE_CODE	SPECIATE PROFILE_CODE.
2	PROFILE_TYPE	SPECIATE PROFILE_TYPE.
3	ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO	SPECIATE OM/OC ratio. This value is only used when converting PM-CR1 profiles to PM-AE6
4	TOGVOC	TOG-to-VOC ratio of profile.
5	CATEGORY_LEVEL_1_Generation_Mechanism	SPECIATE metadata.
6	CATEGORY_LEVEL_2_Sector_Equipment	SPECIATE metadata.

Table 3: File format for export_profiles file.

2.1.3 export_species File

The export_species file provides the weight percent of each SPECIATE species for each profile. This data is retrieved from the SPECIES table within SPECIATE, and it is generated using the export_species query. Following the execution of the query, additional formatting may be necessary including conversion of the file extension to ".csv". In addition, all values in the weight percent and PROFILE_CODE columns should "Convert to Number" (this function is available through the error checking notifications). All SPECIATE profiles should be included within the export_species file and the S2S-Tool filters the contents, as necessary. This dataset is comma delimited and composed of three columns: PROFILE_CODE, SPECIES_ID, and WEIGHT_PERCENT.

Column	Header	Description
1	PROFILE_CODE	SPECIATE PROFILE_CODE.
2	SPECIES_ID	SPECIATE SPECIES_ID.
3	WEIGHT_PERCENT	SPECIATE WEIGHT_PERCENT.

Table 4: File format for export_species file.

2.1.4 export_species_properties File

The export_species_properties file contains SPECIATE species properties (e.g., molecular weight, vapor pressure), and it is generated from the SPECIES_PROPERTIES table using the export_species_properties query. It should be noted that some of these variables (e.g., vapor pressure) are not used by the Tool. Following the execution of the query, additional formatting may be necessary including conversion of the file extension to ".csv". All entries within the SPECIES_PROPERTIES table should be included within the export_species_properties file. This dataset is comma delimited and composed of five columns: SPECIES_ID, HAPS, NonVOCTOG, SPEC_MW, and VP_Pascal_OPERA.

Column	Header	Description
1	SPECIES_ID	SPECIATE SPECIES_ID.
2	HAPS	TRUE/FALSE indicator if species is a HAP.
3	NonVOCTOG	TRUE/FALSE indicator if species is an exempt VOC.
4	SPEC_MW	Molecular weight of SPECIATE species.
5	VP_Pascal_OPERA	Vapor pressure (Pascal) of SPECIATE species.

Table 5: File format for export_species_properties file.

2.1.5 gscnv_append File

The gscnv_append file contains a list of SPECIATE profiles that are appended to the end of GSCNV files when PROFILE_TYPE is set to CRITERIA. These profiles feature different input and output pollutants than standard GSCNV files and are included to accommodate alternative emission inventories (e.g., the Canadian mobile profiles contain EVP_VOC and EXH_VOC). This dataset is comma delimited and composed of three columns: INPUT.POLL, OUTPUT.POLL, and PROFILE.

Column	Header	Description
1	INPUT.POLL	GSCNV input pollutant.
2	OUTPUT.POLL	GSCNV output pollutant.
3	PROFILE	SPECIATE PROFILE_CODE.

Table 6: File format for gscnv_append file.

2.1.6 mech_pm File

The mech_pm file contains the mapping of SPECIATE species to aerosol module specific model species. For example, SPECIES_ID 797 in SPECIATE is elemental carbon and is mapped to PEC (particulate elemental carbon; note: all mechanism-specific PM species are listed in Appendix A) in the PM-AE6 aerosol module. All mapping is 1-to-1, mass-based, and there is often overlap in the mapping between aerosol modules (e.g., SPECIES_ID = 797 is mapped to PEC in all modules). This dataset is comma delimited and composed of four columns: AQM, Mechanism, SPECIES_ID, and Species.

Column	Header	Description
1	AQM	Modeling indicator (options: CMAQ and CAMX).
2	Mechanism	Chemical mechanism (see “MECH_BASIS” Section 2.4).
3	SPECIES_ID	SPECIATE SPECIES_ID.
4	Species	Mechanism- specific, aerosol-phase species.

Table 7: File format for mech_pm file.

2.1.7 mechanism_forImport File

The mechanism_forImport (or mech4import) file contains the mapping of SPECIATE species to chemical mechanism specific gas-phase species. For example, SPECIES_ID 598 in SPECIATE is n-decane and is mapped to 10 PAR (paraffins) in CB6. All mapping is molar-based, one species may be mapped to a single model species or several, and there is often little to no overlap in the mapping between chemical mechanisms. The contents of this file can be generated using the mechanism-specific mapping scripts (see Section 3), is comma delimited, and composed of four columns: Mechanism, SPECIES_ID, Species, and Moles.

Column	Header	Description
1	Mechanism	Name of chemical mechanism (see Section 2.4).
2	SPECIES_ID	SPECIATE SPECIES_ID.
3	Species	Mechanism-specific, gas-phase species.
4	Moles	Number of mechanism-specific moles per SPECIATE species.

Table 8: File format for mechanism_forImport file.

2.1.8 mechanism_mw File

The mechanism_mw file contains the molecular weight of gas-phase species for all supported mechanisms. Previously, the molecular weight of each mechanism-specific species was calculated within the Tool, but these calculations could generate unrealistic values for highly lumped species. Now, the molecular weight of each species, as specified by each chemical mechanism, are fixed. This file is comma delimited, and composed of three columns: Mechanism, Species, and SPEC_WM.

Column	Header	Description
1	Mechanism	Name of chemical mechanism (see Section 2.4).
2	Species	Mechanism-specific, gas-phase species.
3	SPEC_MW	Molecular weight for each mechanism-specific, gas-phase species.

Table 9: File format for mechanism_mw file.

2.1.9 oxygen-metal_Ratios File

The Oxygen-to-Metal file contains the assumed oxygen-to-metal mass-based ratio for common metal species in $PM_{2.5}$ speciation profiles. This file is only used when generating the GSPRO for executions when the OUTPUT is set to PM (see Section 2.4) and the PROFILE_TYPE of the profile being processed is "PM" (i.e., not processed into a PM-AE6 or PM-CR1 profile). The information stored in this file helps perform the "PM-Protocol" (see Appendix B) and follows the methods first developed in Reff et al. (2009). This dataset is comma delimited and composed of three columns with the header indicating the contents of each column: SPECIES_ID, Species, and oxy/metal_ratio.

Column	Header	Description
1	SPECIES_ID	SPECIATE SPECIES_ID.
2	Species	Elemental symbol for SPECIATE species.
3	oxy/metal_ratio	Assumed oxygen-to-metal ratio of SPECIATE species.

Table 10: File format for mechanism_forImport file.

2.1.10 POA_mapping File

The POA Mapping file facilitates the conversion of different profile types (i.e., PM-AE6, PM-CR1) within the S2S-Tool. This file is only used when generating the GSPRO for executions when the OUTPUT is set to PM (see Section 2.4). The file includes all particulate organic matter SPECIATE species, an indicator if the species is organic carbon (OC), non-carbon organic matter (NCOM), or organic matter (OM), an integer for the volatility bin of the species (e.g., "1" if the saturation concentration (C*) of the species meets the following specifications: $10^{0.5} < \log_{10}(C^*) < 10^{1.5}$; "-999" if the species is nonvolatile; etc), and the module-specific mapping for each species. This dataset is comma delimited and composed of five columns with the header indicating the contents of each column: SPECIES_ID, OM/OC/NCOM, SV-POA, PM-AE6, and PM-CR1.

Column	Header	Description
1	SPECIES_ID	SPECIATE SPECIES_ID.
2	OM/OC/NCOM	OM, OC, NCOM indicator of SPECIATE species.
3	SV-POA	SV-POA volatility bin. If SPECIATE species is nonvolatile, value should be "-999".
4	PM-AE6	Mapping of SPECIATE species to the PM-AE6 aerosol-phase species.
5	PM-CR1	Mapping of SPECIATE species to the PM-CR1 aerosol-phase species.

Table 11: File format for POA Mapping file.

2.1.11 POA_VolatilityBins File

The POA Volatility Bins file facilitates the conversion of different profile types (i.e., PM-AE6, PM-CR1) within the S2S-Tool. This file is only used when generating the GSPRO for executions when the OUTPUT is set to PM (see Section 2.4). The file assigns a volatility distribution profile to each unique $PM_{2.5}$ SPECIATE entry of CATEGORY_LEVEL_1_Generation_Mechanism and CATEGORY_LEVEL_2_Sector_Equipment to a semi-volatile primary organic aerosol (SV-POA) volatility profile. For example, all Combustion; Mobile profiles are currently assigned to the "Gasoline" SV-POA profile based on data from Lu et al. (2020). This dataset is comma delimited and composed of six columns with the header indicating the contents of each column: CATEGORY_LEVEL_1_Generation_Mechanism, CATEGORY_LEVEL_2_Sector_Equipment -2, -1, 0, 1, 2, and Reference. Note: the values in columns "-2, -1, 0, 1, 2" are all fractional values whose sum should equal to one (i.e., the volatility distributions are normalized to mass of compounds with $\log_{10}C^* \leq 10^{2.5}$ µg m^{-3}).

Column	Header	Description
1	CATEGORY_LEVEL_1_Generation_Mechanism	SPECIATE metadata.
2	CATEGORY_LEVEL_2_Sector_Equipment	SPECIATE metadata.
3	-2	Weight percent for 10^{-2} µg/m^3 volatility bin.
4	-1	Weight percent for 10^{-1} µg/m^3 volatility bin.
5	0	Weight percent for 10^{0} µg/m^3 volatility bin.
6	1	Weight percent for 10^{1} µg/m^3 volatility bin.
7	2	Weight percent for 10^{2} µg/m^3 volatility bin.
8	Reference	Reference for volatility profile.

Table 12: File format for POA Mapping file.

2.1.12 tbl_tox File

The tbl_tox file is used when RUN_TYPE is INTEGRATE or NOINTEGRATE (see Section 2.4), which occurs when emitted species (e.g., HAPs) are pulled directly from the base inventory (see Section 1.1) in the modeling platform. In the tbl_tox file, a list of SPECIATE species informs the S2S-Tool which species to not include when performing calculations to generate the GSCNV and GSPRO for a given execution. For example, if one were to generate a set of GSCNV and GSPRO VOC files and incorporate naphthalene, benzene, acetaldehyde, formaldehyde, and methanol (NBAFM) from the base inventory (i.e., NAPHTH, BENZENE, ACETALD, FORMALD, METHANOL in the U.S. NEI) for a HAP-CAP modeling platform, these pollutants should not be included when speciating non-NBAFM VOC mass.

When speciating emissions from the MOVES model, additional emitted gas-phase species are imported directly to the modeling platform and should not be included when speciating VOC mass. As such, the tbl_tox file in these scenarios must expand to include the additional "integrated" species (e.g., 1,3-butadiene, ethylbenzene, ethanol). In addition, the MOVES model provides explicit emissions for particulate elemental carbon and sulfate, which are components of $PM_{2.5}$. To speciate the remaining portion of $PM_{2.5}$, MOVES outputs emissions of particulate total organic matter (TOM) and all residual mass (RESID_PM). To accommodate speciation of TOM and RESID_PM, two additional tbl_tox files are included in the S2S-Tool repository that enable the generation of GSPRO's for TOM and RESID_PM.

This dataset is comma delimited and composed of three columns with the header indicating the contents of each column: AQM, SPECIES_ID, and Inv.Species. Note that multiple Inv.Species entries may have

equivalent SPECIES_ID entries due to process-level emissions information (e.g., BENZENE, EVP__BENZENE, and EXH__BENZENE are all inventory species whose SPECIES_ID = 302).

Column	Header	Description
1	AQM	Modeling indicator (options: CMAQ and CAMX).
2	SPECIES_ID	SPECIATE SPECIES_ID of inventory species to be incorporated.
3	Inv.Species	Inventory species to be incorporated (SMOKE INVTABLE must match).

Table 13: File format for tbl_tox file.

2.2 Output Files

The contents and structure of all output files are described below.

2.2.1 GSCNV File

The GSCNV file is one of the primary outputs from the S2S-Tool and is an input to the SMOKE modeling system. Also known as the pollutant-to-pollutant conversion factors file, a GSCNV is required when there is a mismatch between the pollutant in the inventory and the pollutant for which the speciation profiles have been developed. Currently, this is only relevant for profiles within SPECIATE for which the PROFILE_TYPE is GAS. The most common example of pollutant-to-pollutant conversion within SMOKE is the conversion of VOC, which is a regulatorily defined collection of gas-phase organic emissions, to TOG, which is the standard gas-phase speciation profile type within SPECIATE.

A GSCNV file can begin with comment lines, which the S2S-Tool generates to provide metadata on the S2S-Tool runtime specifications that were used when generating the GSCNV file. Following the comment lines, the first required header line for SMOKE is "#BY PROFILE" (see Fig. 3). Following the "#BY PROFILE" line, a row is generated for each SPECIATE profile for which the PROFILE_TYPE = GAS. These rows each consist of four columns: the input pollutant, the output pollutant, the SPECIATE profile code, and the conversion factor (units of ton per ton).

```
#S2S_AQM              CMAQ
#S2S_CAMX_FCRS        Not Applicable
#S2S_MW               ./input/mechanism_mw.csv
#S2S_MECH_BASIS       CB6R3_AE7
#S2S_RUN_TYPE         CRITERIA
#S2S_RUN_DATE         2023-08-09
#S2S_TBL_TOX          Not Applicable
#BY PROFILE
VOC                   TOG                   2402                  1.03316458
VOC                   TOG                   3150                  1.08471635
```

Figure 3: Example GSCNV file.

In Fig. 3, the first entry indicates that the input pollutant is VOC, the output pollutant is TOG, and the SPECIATE profile 2402 features a conversion factor of 1.03316458 between the input and output pollutant. In other words:

$$TOG_{2402} = VOC_{2402} \times 1.03316458$$

As noted in the comments, the GSCNV illustrated in Fig. 3 was generated for "criteria" speciation methods. An exact replica of this file (except #S2S_RUN_TYPE = NOINTEGRATE; see Section 2.4) would be generated for "no-integrate" speciation methods. In contrast, an "integrate" speciation methods GSCNV would use NONHAPVOC and NONHAPTOG for the input and output pollutants, respectively. Contrary to the name,

NONHAPTOG does not represent all TOG minus HAPs. Rather, NONHAPTOG represents all TOG minus the species directly integrated into the modeling platform from the base inventory. For example, for a profile consisting of 30% decane, 15% acetone, 20% benzene, and 35% propanal, EPA's current "HAP-CAP" modeling platform would integrate benzene from the base inventory, but not propanal (even though it is a HAP) because propanal is not chosen to be integrated. Noting that acetone is an exempt VOC, the conversion factor for this profile following "integrate" speciation methods would be 1.230769 ($\frac{0.30+0.15+0.35}{0.3+0.35}$), computed using the sum of non-integrated (i.e., NONHAP) TOG weight percents in the numerator and the sum of non-integrated, VOC weight percents in the denominator.

For mobile sources generated using the MOVES model, the number of species integrated from the emissions model increases and includes species beyond HAPs (e.g., ethanol). In these circumstances, the tbl_tox input file (see Section 2.1.12) must be updated to reflect all species incorporated from a given MOVES run.

2.2.2 GSPRO File

The GSPRO file is the second primary output from the S2S-Tool and is also an input to the SMOKE modeling system. Also known as the speciation profile file, a GSPRO contains the factors that are used to separate aggregated pollutant emissions (e.g., VOC) into emissions of model species required by an AQM (e.g., ETOH, ACET for ethanol and acetone in CB6). Unlike the GSCNV, which only applies to gaseous profiles, a GSPRO is relevant for multiple profile types within SPECIATE: GAS, PM, PM-AE6, and PM-CR1.

A GSPRO file begins with comment lines and contains metadata on the S2S-Tool runtime specifications that were used to generate the file. Following the comment lines, a row is generated for each non-zero model species within a SPECIATE profile. These rows each consist of six columns: the SPECIATE profile code, the speciated pollutant, the model species, a mass-based split factor (mass-based fraction of the model species in the profile), the effective molecular weight of the model species ("divisor"), and the mass fraction. In all cases, the split factor and mass fraction are equivalent.

```
#S2S_AQM            CMAQ
#S2S_CAMX_FCRS      ./input/camx_fcrs.profile.csv
#S2S_MW             ./input/mechanism_mw.csv
#S2S_MECH_BASIS     CB6R3_AE7
#S2S_RUN_TYPE       CRITERIA
#S2S_RUN_DATE       2023-08-09
#S2S_TBL_TOX        Not Applicable
2402                TOG                 ACET       3.210000E-02  5.810000E+01  3.210000E-02
2402                TOG                 ALDX       1.160839E-03  5.810000E+01  1.160839E-03
2402                TOG                 ETOH       2.280000E-02  4.610000E+01  2.280000E-02
2402                TOG                 IVOC       1.170000E-02  2.405000E+02  1.170000E-02
2402                TOG                 KET        6.038634E-02  2.900000E+01  6.038634E-02
2402                TOG                 PAR        4.552528E-01  1.400000E+01  4.552528E-01
2402                TOG                 TOL        1.296000E-01  9.210000E+01  1.296000E-01
2402                TOG                 XYLMN      2.870000E-01  1.062000E+02  2.870000E-01
2402                TOG                 NMOG       1.000000E+00  1.000000E+00  1.000000E+00
```

Figure 4: Example GSPRO file.

In Fig. 4, the CB6R3_AE7 speciation profile for SPECIATE profile 2402 is illustrated. While the "raw" SPECIATE profile is composed of 28 separate entries, the GSPRO indicates that these 28 species are represented by 8 model species for the CB6R3_AE7 chemical mechanism when using "criteria" speciation methods (RUN_TYPE = CRITERIA; see Section 2.4). When the SMOKE modeling system is run, the SPECIATE profile is matched to an emissions source, usually based on the SCC assignment within the GSREF

speciation cross-reference file. For example, if a source emits 100 tons per year of VOC and is assigned to SPECIATE profile 2402 in the GSREF, SMOKE will combine information from the GSCNV and GSPRO to generate 3.32 tons per year of acetone (ACET) from this source $(100 \text{ tons VOC / year} \times 1.03316458 \times 3.21\%)$. In addition to the standard model species (see Appendix A), NMOG (Non-Methane Organic Gas) is appended to the end. Since profile 2402 does not contain methane, NMOG equals 100%.

When generating a GSPRO for "integrate" or "no-integrate" speciation methods, the speciated pollutant for a given OUTPUT, the collection of model species used in a SPECIATE profile, and the mass-based split factors are different from the above example and will be discussed further in Section 2.5.4. Furthermore, additional comment lines are included in the header of "integrate" GSPRO files. These comment lines list the emission inventory species that were removed from the speciation profiles, as provided in the tbl_tox file (see Section 2.1.12), and therefore not included in the definition of NONHAPVOC.

2.3 Supported Mechanisms

The S2S-Tool is currently capable of generating datasets compatible with several VOC and $PM_{2.5}$ representations in chemical mechanisms. The mechanism mapping names for the VOC-relevant speciation (i.e., the MECH_BASIS runtime input in the S2S-Tool; see Section 2.4) include CB6R3_AE7, CB6R5_AE7, CB7_AE7, CB6R4_CF2, CB7_CF2, CB6R3_AE7_TRACER, CRACMMv1.0, SAPRC07TC_AE7, and SAPRC07_CF2, and the mechanism mapping names for the $PM_{2.5}$-relevant speciation include PM-AE6, and PM-CR1.

A broader description of these mappings and their chemical treatments are included in the table below.

OUTPUT	MECH_BASIS	AQM	Comments
VOC	CB6R3_AE7	CMAQv5.3+	• All IVOCs ($10^{2.5} < \log_{10}(C^*) < 10^{6.5}$) mapped to IVOC. • All compounds with $\log_{10}(C^*) < 10^{2.5}$ mapped to NVOL.
VOC	CB6R5_AE7	CMAQv5.4+	• Same emitted/model species as CB6R3_AE7; updated model chemistry.
VOC	CB7_AE7	n/a	• Same emitted species as CB6R3_AE7; updated model chemistry.
VOC	CB6R4_CF2	CAMx	• All IVOCs ($10^{2.5} \le \log_{10}(C^*) < 10^{6.5}$) mapped to IVOC. • All compounds with $\log_{10}(C^*) < 10^{2.5}$ mapped to NVOL.
VOC	CB7_CF2	CAMx	• Same emitted species as CB6R4_CF2, except APIN added; updated model chemistry.
VOC	CB6R3_AE7_TRACER	CMAQv5.3+	• Assigns SPECIES_ID = 279 to ALD2_PRIMARY, 465 to FORM_PRIMARY, cyclic alkanes with nC > 6 and non-cyclic alkanes with nC > 8 to SOAALK, and all other compounds to NONBAF.
VOC	SAPRC07TC_AE7	CMAQv5.3+	• All IVOCs ($10^{2.5} < \log_{10}(C^*) < 10^{6.5}$) mapped to IVOC. • All compounds with $\log_{10}(C^*) < 10^{2.5}$ mapped to NVOL.
VOC	SAPRC07_CF2	CAMx	• All IVOCs ($10^{2.5} < \log_{10}(C^*) < 10^{6.5}$) mapped to IVOC. • All compounds with $\log_{10}(C^*) < 10^{2.5}$ mapped to NVOL.
VOC	CRACMMv1.0	CMAQv5.4+	• ROCIOXY: Oxygenated and $\log_{10}(C^*) \le 10^{6.5}$ • ROCP6ALK: $10^{5.5} \le \log_{10}(C^*) < 10^{6.5}$ • ROCP5ALK: $10^{4.5} \le \log_{10}(C^*) < 10^{5.5}$

			• ROCP4ALK: $10^{3.5} \leq \log_{10}(C^*) < 10^{4.5}$ • ROCP3ALK: $10^{2.5} \leq \log_{10}(C^*) < 10^{3.5}$ • ROCP2ALK: $10^{1.5} \leq \log_{10}(C^*) < 10^{2.5}$ • ROCP1ALK: $10^{0.5} \leq \log_{10}(C^*) < 10^{1.5}$ • ROCP0ALK: $10^{-0.5} \leq \log_{10}(C^*) < 10^{0.5}$ • ROCN1ALK: $10^{-1.5} \leq \log_{10}(C^*) < 10^{-0.5}$ • ROCN2ALK: $\log_{10}(C^*) < 10^{-1.5}$ • ROCP6ARO: aromatic and $10^{5.5} < \log_{10}(C^*) < 10^{6.5}$ • ROCP5ARO: aromatic and $10^{4.5} < \log_{10}(C^*) < 10^{5.5}$
PM	PM-AE6	CMAQv5.3+	• Generated using the "PM-Protocol" (see Appendix B). Organic matter is split between two species: POC and PNCOM.
PM	PM-CR1	CMAQv5.4+	• Generated using the "PM-Protocol" (see Appendix B). Organic matter is split among five species: ROCP2, ROCP1, ROCP0, ROCN1, ROCN2.

Table 14: VOC and $PM_{2.5}$ representations and their description currently supported by the S2S-Tool.

2.4 Runtime Inputs

All runtime inputs are user specified and populated in the *S2S.main.py* file. Currently, there are twelve variables listed as runtime inputs: MECH_BASIS, OUTPUT, RUN_TYPE, AQM, TOLERANCE, TOX_IN, FCRS_FILE, CAR_FILE, M4I_FILE, PMM_FILE, and TOX_FILE (Table 15).

Runtime Input	Options (Description)
MECH_BASIS	**CB6R3_AE7** (OUTPUT = VOC option) **CB6R5_AE7** (OUTPUT = VOC option) **CB7_AE7** (OUTPUT = VOC option) **CB6R4_CF2** (OUTPUT = VOC option) **CB7_CF2** (OUTPUT = VOC option) **CB6R3_AE7_TRACER** (OUTPUT = VOC option) **SAPRC07TC_AE7** (OUTPUT = VOC option) **SAPRC07_CF2** (OUTPUT = VOC option) **CRACMMv1.0** (OUTPUT = VOC option) **PM-AE6** (OUTPUT = PM option) **PM-CR1** (OUTPUT = PM option)
OUTPUT	**VOC** (Selected when the GSCNV/GSPRO output is for processing VOC emissions) **PM** (Selected when the GSPRO output is for processing $PM_{2.5}$ emissions)
RUN_TYPE	**CRITERIA** (Selected for "criteria" speciation of VOC/$PM_{2.5}$) **INTEGRATE** (Selected for "integrate" speciation of VOC/$PM_{2.5}$) **NOINTEGRATE** (Selected for "no-integrate" speciation of VOC)
AQM	**CMAQ** (Selected when the GSCNV/GSPRO output is for a CMAQ modeling platform) **CAMX** (Selected when the GSCNV/GSPRO output is for a CAMx modeling platform)
TOLERANCE	A numerical (e.g., 0.05 for 5%) value. Represents the acceptable deviation from 100% for the sum of SPECIATE species for gas-profiles. If a profile falls outside this range, it is not processed by the S2S-Tool and noted in the run log.
TOX_IN	**NONHAPTOG** (Represents the input pollutant for a RUN_TYPE = INTEGRATE execution)* **TOM** (Represents the input pollutant for a RUN_TYPE = INTEGRATE execution)*

	RESID_PM (Represents the input pollutant for a RUN_TYPE = INTEGRATE execution)*
FCRS_FILE	File name and location of the CAMx FCRS file.
MW_FILE	File name and location of the mechanism_mw file.
M4I_FILE	File name and location of the mechanism_forImport file.
PMM_FILE	File name and location of the mech_pm file.
TOX_FILE	File name and location of the tbl_tox file.

Table 15: Runtime input variables, options, and their description. Note that the options within the S2S-Tool are case-sensitive.
* Currently, NONHAPTOG is exclusively used when OUTPUT = VOC and RUN_TYPE = INTEGRATE. TOM and RESID_PM are exclusively used when OUTPUT = PM and RUN_TYPE = INTEGRATE, and exclusively used for speciating lumped components from MOVES.

2.4.1 Available Configurations

While the S2S-Tool is flexible and VOC or $PM_{2.5}$ representations not included in Table 16 can be processed (see Section 2.6), the Tool "out-of-the-box" is compatible with a set number of runtime configurations. These configurations fit the needs for developing "CAP-only" and "HAP-CAP" modeling platforms at the U.S. EPA, which includes profiles that fit "criteria," "integrate," and "no-integrate" speciation (see Section 1.1). In addition, as speciation of mobile source emissions now occurs outside of the MOVES modeling system, select functionalities are included within the S2S-Tool to facilitate speciation of MOVES' VOC, NONHAPTOG, TOM, and RESID_PM emissions.

OUTPUT	MECH_BASIS	RUN_TYPE	AQM	INPUT POLLUTANT*	TOX_FILE
VOC	CB6R3_AE7	CRITERIA	CMAQ	TOG	n/a
VOC	CB6R5_AE7	CRITERIA	CMAQ	TOG	n/a
VOC	CB7_AE7	CRITERIA	CMAQ	TOG	n/a
VOC	CRACMMv1.0	CRITERIA	CMAQ	TOG	n/a
VOC	SAPRC07TC_AE7	CRITERIA	CMAQ	TOG	n/a
VOC	CB6R3_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB6R5_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB7_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CRACMMv1.0	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	SAPRC07TC_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB6R3_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB6R5_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB7_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CRACMMv1.0	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	SAPRC07TC_AE7	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB6R3_AE7	NOINTEGRATE	CMAQ	TOG	tbl_tox_NBAFM.csv
VOC	CB6R5_AE7	NOINTEGRATE	CMAQ	TOG	tbl_tox_NBAFM.csv
VOC	CB7_AE7	NOINTEGRATE	CMAQ	TOG	tbl_tox_NBAFM.csv
VOC	CRACMMv1.0	NOINTEGRATE	CMAQ	TOG	tbl_tox_NBAFM.csv
VOC	SAPRC07TC_AE7	NOINTEGRATE	CMAQ	TOG	tbl_tox_NBAFM.csv
VOC	CB6R3_AE7_TRACER	CRITERIA	CMAQ	TOG	n/a
VOC	CB6R3_AE7_TRACER	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB6R3_AE7_TRACER	INTEGRATE	CMAQ	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB6R3_AE7	CRITERIA	CAMX	TOG	n/a
VOC	CB6R5_AE7	CRITERIA	CAMX	TOG	n/a
VOC	CB7_AE7	CRITERIA	CAMX	TOG	n/a
VOC	CRACMMv1.0	CRITERIA	CAMX	TOG	n/a
VOC	SAPRC07TC_AE7	CRITERIA	CAMX	TOG	n/a

VOC	CB6R3_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB6R5_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB7_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CRACMMv1.0	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	SAPRC07TC_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB6R3_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB6R5_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB7_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CRACMMv1.0	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	SAPRC07TC_AE7	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB6R3_AE7	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
VOC	CB6R5_AE7	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
VOC	CB7_AE7	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
VOC	CRACMMv1.0	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
VOC	SAPRC07TC_AE7	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
VOC	CB6R4_CF2	CRITERIA	CAMX	TOG	n/a
VOC	CB6R4_CF2	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB6R4_CF2	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB6R4_CF2	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
VOC	CB7_CF2	CRITERIA	CAMX	TOG	n/a
VOC	CB7_CF2	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	CB7_CF2	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	CB7_CF2	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
VOC	SAPRC07_CF2	CRITERIA	CAMX	TOG	n/a
VOC	SAPRC07_CF2	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_NBAFM.csv
VOC	SAPRC07_CF2	INTEGRATE	CAMX	NONHAPTOG	tbl_tox_MOVES_HAPS.csv
VOC	SAPRC07_CF2	NOINTEGRATE	CAMX	TOG	tbl_tox_NBAFM.csv
PM	PM-AE6	CRITERIA	CMAQ	PM2_5	n/a
PM	PM-CR1	CRITERIA	CMAQ	PM2_5	n/a
PM	PM-AE6	INTEGRATE**	CMAQ	TOM	tbl_tox_RESID_PM.csv
PM	PM-AE6	INTEGRATE**	CMAQ	RESID_PM	tbl_tox_TOM.csv
PM	PM-CR1	INTEGRATE**	CMAQ	TOM	tbl_tox_TOM.csv
PM	PM-CR1	INTEGRATE**	CMAQ	RESID_PM	tbl_tox_RESID_PM.csv
PM	PM-AE6	CRITERIA	CAMX	PM2_5	n/a
PM	PM-AE6	INTEGRATE**	CAMX	TOM	tbl_tox_TOM.csv
PM	PM-AE6	INTEGRATE**	CAMX	RESID_PM	tbl_tox_RESID_PM.csv

Table 16: Available "out-of-the-box" configurations for the S2S-Tool.

* INPUT POLLUTANT = the runtime TOX_IN specified by the user when RUN_TYPE = INTEGRATE.

** RUN_TYPE = INTEGRATE for OUTPUT = PM is only required for MOVES output. Speciation of $PM_{2.5}$ now occurs outside of the MOVES modeling system. EC and SO_4 are explicitly provided by MOVES, while Total Organic Matter (TOM = OC + NCOM) and all residual $PM_{2.5}$ mass (RESID_PM) must be speciated within SMOKE-MOVES. Too account for the variable $PM_{2.5}$ components contained within each aggregate pollutant (i.e., TOM and RESID_PM), pollutants must be integrated into base SPECIATE profiles.

2.5 Module Descriptions & Functions

The current S2S-Tool includes four *python* files and several input datasets (see Section 2.1). The *S2S.main.py* file is the main program, where the runtime inputs are set, and calls each of the other *python* modules and their functions. The *check_inputs.py*, *gscnv.py*, and *gspro.py* modules are located within the "./modules" directory and house several functions that perform quality assurance checks, generate the GSCNV files, and generate the GSPRO files, respectively. Below, the overall workflow of the S2S-Tool is illustrated, followed by a description of each program and the calculations performed in each.

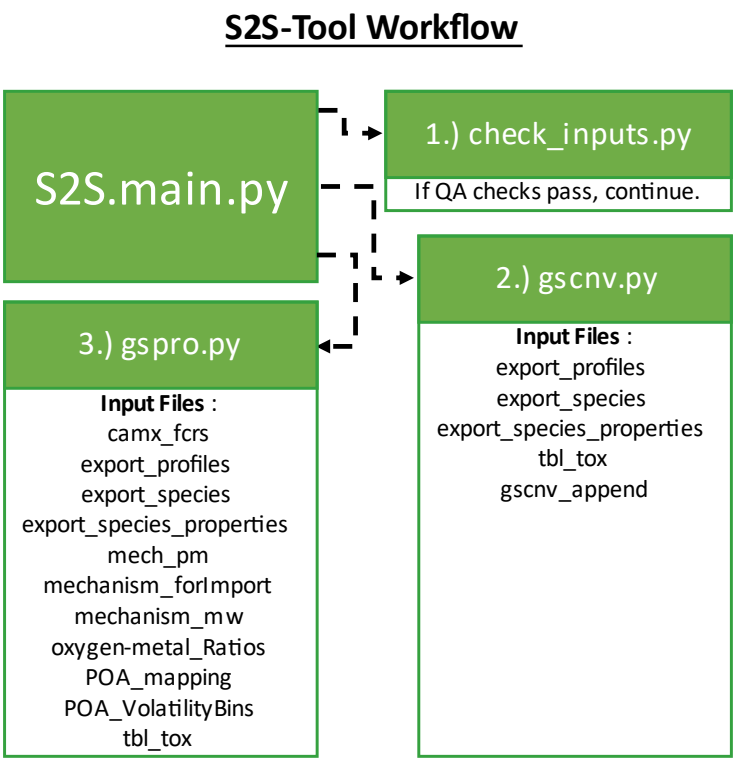


Figure 5: Current workflow in the S2S-Tool.

2.5.1 S2S.main.py

The *S2S.main.py* file is the main S2S-Tool program. Here, runtime inputs are provided following the section that begins with "### User Input" (see Table 15 for variables and options). Following the runtime inputs section, this file initiates several variables, reads the appropriate input files, filters input files to meet the needs of the selected runtime inputs, imports the S2S-Tool modules, runs quality assurance checks, generates a GSCNV file (if OUTPUT = VOC), and finally generates a GSPRO file.

2.5.2 check_inputs.py

The *check_inputs.py* module contains the first collection of functions called by *S2S.main.py*. These functions provide a check on the runtime inputs and the input files, and if an error is found, stops the execution with a printed statement indicating where and what issue is found. For example, one of the functions checks to ensure that the selected MECH_BASIS is compatible with the selected OUTPUT. A separate check ensures that all profiles included in export_profiles files are in export_species file. While not fully comprehensive, this series of checks should help ensure there are no fatal issues with the runtime inputs or the input files.

2.5.3 gscnv.py

Following the quality assurance checks, the S2S-Tool generates a GSCNV file if the OUTPUT is set to VOC. Within the *gscnv.py* module is two functions; the first generates the GSCNV file and the second adds a header to the file. In the *gen_gscnv* function, methods depend on the RUN_TYPE assignment. If the RUN_TYPE is set to CRITERIA or NOINTEGRATE, the input and output pollutants are always VOC and TOG, respectively (see Section 2.2). In contrast, if the RUN_TYPE is set to INTEGRATE, the input and output pollutants are always set to NONHAPVOC and NONHAPTOG, respectively (see Section 2.2).

In both scenarios, the function checks if the total weight percent of each profile is within the set TOLERANCE and passes over the profile if it is not within the set range. Following this check, if the RUN_TYPE is set to CRITERIA or NOINTEGRATE, a ratio of the mass of all organic gases (TOG) to the mass of all regulatorily defined VOCs[7] within a profile is derived. If the profile is entirely derived of an exempt VOC (e.g., tetrachloroethylene), the conversion factor is set to zero. If the RUN_TYPE is set to INTEGRATE, the profile is first filtered to remove all species listed in the tbl_tox file. Subsequently, a ratio of the mass of all organic gases (TOG) to the mass of all regulatorily defined VOCs within a profile is derived. Again, if the profile is entirely derived of an exempt VOC, the conversion factor is set to zero. Finally, if the RUN_TYPE is set to CRITERIA, the profiles listed in the gscnv_append file are appended to the end of the dataframe.

The second function in *gscnv.py* generates the header for the final GSCNV file, which includes several of the runtime inputs selected for the given Tool run (see Figure 3). Finally, the output filename includes a date stamp to facilitate tracking of file versions.

2.5.4 gspro.py

Following the *gscnv.py* module, the S2S-Tool executes the *gspro.py* module, which generates a GSPRO file. Within the *gspro.py* module are three functions; the first generates a GSPRO file if OUTPUT is set to VOC (*gen_gspro_voc*), the second generates a GSPRO file if OUTPUT is set to PM (*gen_gspro_pm*), and the third adds a header to the file.

2.5.4.1 gen_gspro_voc

The *gen_gspro_voc* function iterates over each relevant profile in the export_profiles file and appends processed data to a final dataframe. First, the function checks if the total weight percent of each profile is within the set TOLERANCE. If the profile is not within the set range, it is passed over and a notification is printed to notify the user that the profile was not processed. Next, if RUN_TYPE is set to CRITERIA, the profile is renormalized to ensure the summation of all weight percents equals one and the input pollutant (i.e., the speciated pollutant; see Section 2.2.2) is set to TOG. If RUN_TYPE is set to NOINTEGRATE, the profile is first renormalized to ensure the summation of all weight percents equals one, then filtered to remove all species listed in the tbl_tox file, and the input pollutant is set to TOG. Finally, if RUN_TYPE is set to INTEGRATE, the profile is renormalized to ensure the summation of all weight percents equals one, filtered to remove all species listed in the tbl_tox file, renormalized again to ensure the summation of all remaining weight percents equals one, and the input pollutant is set to NONHAPTOG.

The mole-based split factor (*MolBSF*) in moles of mechanism-specific model species *j* per mass of input pollutant (e.g., TOG) is then calculated using:

$$MolBSF_j = \sum_i Wght\%_i \div MW_i \times Moles_{i,j} \quad (1)$$

Where *Wght%* is the weight percent of compound *i* in the target profile (mass compound per mass TOG) from SPECIATE, MW_i is the *effective molecular weight* of compound *i*, and $Moles_{i,j}$ is the number of moles of model species *j* in compound *i* (i.e., the value provided in the mech4import file; see Section 2.1.7). New to S2S-Tool v2 is the introduction of the *effective molecular weight* of each compound. Previously, the MW_i was provided by the representative compound of each species in SPECIATE, but this can lead to issues for highly lumped model species. In photochemical modeling, the compounds that are mapped to a

[7] https://www.epa.gov/ground-level-ozone-pollution/complete-list-voc-exemption-rules

chemical mechanism model species adopt the properties of that model species, regardless of the structure or molecular weight of the explicit species in SPECIATE. This is particularly important for IVOCs, which, in general, are highly lumped and often feature significant oxidation. Analysis has shown that most compounds mapped to IVOC model species adopt a MW that is lower than the explicit species MWs. For these compounds, the use of the explicit species MW, as in the prior methods, overestimates the molar emissions (as defined by the chemical mechanism), which leads to an overestimate of the mass concentrations. By using the model species MWs, the number of moles (and hence mass) is reduced, mitigating this overestimation. This issue will likely decrease in time as future mechanisms expand the number of lumped species and better represent the complexity of emissions from sources.

The effective molecular weight (*MWi*) for each explicit species within a profile is calculated as follows:

$$MW_i = \sum_j \left(MW_j \times Moles_{i,j} \right) \quad (2)$$

Where MW_j is the molecular weight for each model species *j* and $Moles_{i,j}$ is the number of moles of model species *j* in compound *i* (i.e., the value provided in the mech4import file; see Section 2.1.7).

Next, the mass-based split factor (*MassBSF*) for each mechanism-specific model species *j* is then calculated using:

$$MassBSF_j = MolBSF_j \times MW_j \quad (3)$$

Where *MWj* is the molecular weight of model species *j* (Section 2.1.8). The units for *MassBSF* are grams of model species *j* per mass of emitted TOG. A simplified example using these methods can be found in ./documentation/S2S-Tool_simplified.xlsx.

Finally, the NMOG fraction for each profile is appended for emissions mass tracking in SMOKE. NMOG is calculated as the total organic gas mass fraction, minus methane and minus all species listed in the tbl_tox file if RUN_TYPE is set to INTEGRATE or NOINTEGRATE.

2.5.4.2 gen_gspro_pm

The *gen_gspro_pm* function also iterates over each relevant profile in the export_profiles file and appends processed data to a final dataframe. First, if the PROFILE_TYPE of the profile set to be processed is PM, the profile is processed into a "PM-ready" profile (see Appendix B). Next, if RUN_TYPE is set to CRITERIA, the input pollutant (i.e., the speciated pollutant; see Section 2.2.2) is set to PM2_5. If RUN_TYPE is set to INTEGRATE, the profile is filtered to remove all species listed in the tbl_tox file, renormalized to ensure the summation of all remaining weight percents equals one, and the input pollutant is set to the TOX_IN runtime variable. As a reminder, RUN_TYPE = INTEGRATE for OUTPUT = PM is only relevant when processing MOVES output (see Table 16) and RUN_TYPE = NOINTEGRATE is not an available option when OUTPUT = PM.

Next, organic matter is processed, as necessary, to accommodate output of profiles compatible with the AE6 (PM-AE6) aerosol module and CRACMMv1.0 (PM-CR1). Except for organic matter related species, all profile types (i.e., PM-AE6 and PM-CR1) are interchangeable (e.g., regardless of profile type, PEC, primary elemental carbon, should be an equivalent fraction for a given profile processed). For PM-AE6 profiles, all organic matter is routed to either POC or PNCOM (i.e., OM = POC + PNCOM). In PM-CR1 profiles, all organic matter (i.e., carbon and non-carbon organic matter are summed for each volatility bin) is routed to five separate model species (ROCP2ALK, ROCP1ALK, ROCP0ALK, ROCN1ALK, ROCN2ALK) for organic

matter. Assignments depend on volatility and whether carbon is contained in the modeled species (e.g., organic matter with a saturation concentration (C*) that is < $10^{1.5}$ µg m^{-3} and > $10^{0.5}$ µg m^{-3} would be routed to ROCP1ALK (with the middle two characters reflecting the sign of the exponent of the saturation concentration; P: positive; N: negative).

A profile within SPECIATE and processed by the S2S-Tool can be output to one of two profile types (i.e., PM-AE6 or PM-CR1) using the ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO, CATEGORY_LEVEL_1_Generation_Mechanism and CATEGORY_LEVEL_2_Sector_Equipment metadata in the export_species_properties file and SV-POA profile in the POA Volatility Bins file. For example, to translate from a PM-CR1 profile type, which does not differentiate carbon and non-carbon organic matter, to a PM-AE6 profile type, the ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO is used to split the organic matter. Separately, to translate from a PM-AE6 profile type, which does not resolve organic matter by volatility, to a PM-CR1 profile, the appropriate SV-POA profile from the POA Volatility Bins file is extracted using the CATEGORY_LEVEL_1_Generation_Mechanism and CATEGORY_LEVEL_2_Sector_Equipment metadata for the target profile.

Following treatment of organic matter, the *gen_gspro_pm* function accounts for the various ways chlorine, calcium, magnesium, potassium, and sodium are reported in SPECIATE profiles (atomic vs. ionic). Then, residual PM mass is assigned to PMOTHR if the AQM runtime variable is assigned CMAQ and FPRM if it is assigned CAMX. However, if the target profile is listed in the CAMx FCRS file (see Section 2.1.1), the residual PM mass is assigned to FCRS if the AQM runtime variable is assigned CAMX. Finally, if the AQM runtime variable is assigned CAMX, POA is added by summing POC and PNCOM (when AQM is CAMX and OUTPUT is PM, MECH_BASIS must be PM-AE6), and PNCOM is then dropped from the profile.

2.6 Running the S2S-Tool

To run the S2S-Tool, the contents of the S2S-Tool repository should be copied to either a local or remote machine. Next, following the available configurations listed in Table 16, the runtime input fields in the *S2S.main.py* script should be updated to meet the need of the user. Finally, the *S2S.main.py* script should be executed. All modules and functions are performed in memory (i.e., no databases need to be initialized nor shared schema uploaded) and should only take a few minutes to complete on a standard machine. Output from the run will be available in the "output" directory and feature a date stamp to facilitate version control.

3 Mechanism-Specific Mapping Scripts

3.1 Recent Gas-Phase Mapping Updates

In recent years, an updated mech4import file has been generated by external contractors with each release of SPECIATE. These updates incorporate new species that have been added to SPECIATE and revisions to supported gas- (e.g., CB05, CB6, CB7, SAPRC07, SAPRC07TC, CRI) and aerosol-phase (e.g., AE7, CF2) mechanisms. More recently, methods to leverage emitted species metadata and automatically map to mechanism species have been developed internally at EPA for use with CRACMM (Pye et al., 2023). The current public release of the mech4import file can be retrieved from the Speciation Tool GitHub repository, and CRACMM files are available on the CRACMM GitHub repository[8].

In the Speciation Tool v4.5 release, a volatility dimension (Odum et al., 1996) was incorporated into the mapping process. The updated mapping routed compounds that met the semi-volatile organic compound (SVOC; $3x10^{-1}$ µg m^{-3} < C* < $3x10^{2}$ µg m^{-3}) and intermediate-volatility organic compound (IVOC; $3x10^{2}$ µg m^{-3} < C* < $3x10^{6}$ µg m^{-3}) definitions to NVOL and IVOC, respectively. Here, C* is the effective saturation concentration of the pure semivolatile compound (Pankow, 1994; Odum et al., 1996; Kroll and Seinfeld, 2008) and is a measure of volatility often used in the atmospheric chemistry community. In addition, UNK was added to account for inorganic species and other species that could not be mapped. These species (i.e., NVOL, IVOC, UNK) were not previously treated within the gas-phase chemical mechanisms.

In preparation for the SPECIATE v5.0 release, the Speciation Tool mapping process was again updated to incorporate the expected species associated with semivolatile POA. The update disaggregated the NVOL component into four volatility bins (SVOCN1, SVOCP0, SVOCP1, SVOCP2) and the IVOC into six volatility bins that include information on compound structure (IVOCP3, IVOCP4, IVOCP5, IVOCP6, IVOCP5ARO, IVOCP6ARO). In addition, several compounds were assigned to newly explicit model species for select mechanisms: naphthalene, alpha-pinene, acetic acid, formic acid, 1,3-butadiene, acrolein, 1,2,4-trimethyl benzene, ethanol, o-, m-, p-xylene, propane, sesquiterpenes, and toluene.

3.2 General Methods

In SPECIATE v5.2, a *Representative Compounds* field was introduced for all entries in the database (Pye et al., 2023). The primary purpose of a *Representative Compound* is to robustly identify each species, thus facilitating the prediction of physiochemical properties (e.g., molecular weight, vapor pressure, k_{OH} rate coefficient, Henry's Law constant; using OPERA, Mansouri et al., 2018) and allow for automated parsing of information necessary for mechanism mapping. In most circumstances, the *Representative Compound* and SPECIES_NAME are equivalent. Other times, the SPECIES_NAME is non-specific, but does convey certain chemistry relevant details (e.g., "Isomers of octane" are assigned to "Octane"). Finally, some entries are broad and minimal, if any, information regarding the compound structure is available (e.g., "Aggregated VOCs" and "Aggregated exempt compounds"). In these circumstances, *Decane* is selected as the default *Representative Compound*, since C10 compounds are the approximate median entry in SPECIATE (both in terms of molecular weight and structure), unless other identifying information can be collected from the SPECIES_NAME (e.g., "Aggregated VOCs" were assigned to *Decane*, but "Aggregated exempt compounds" were assigned to *Acetone* since it is one of the most abundantly emitted exempt VOCs).

[8] https://github.com/USEPA/CRACMM

The newly developed mechanism mapper(s) utilize the predicted physiochemical properties of each SPECIATE entry to make chemical mechanism specific assignments. These properties include the molecular weight, vapor pressure, k_{OH} rate constant, and SMILES string of the *Representative Compound*. The molecular weight and vapor pressure are used to calculate the effective saturation concentration (C*).

Each mechanism mapper features three main components:

1. The *Representative Compound* SMILES string of each SPECIATE entry is analyzed (hereafter referred to as the "chemical assessment"). This includes, but is not limited to:
    a. Counting the number of carbon, oxygen, and nitrogen atoms.
    b. Counting functional groups, such as number of ketones, aldehydes, benzene rings, acid, and alcohol groups.
    c. Determining if the compound features a ring or is a monoterpene.
2. All explicit chemical mechanism species are considered for assignment.
3. All lumped chemical mechanism species are considered for assignment.

In the following sections, more detail on each of the supported mechanisms that currently feature mechanism mappers are discussed.

3.3 Chemical Mechanisms

3.3.1 CB6R3_AE7 and CB6R5_AE7 and CB7_AE7

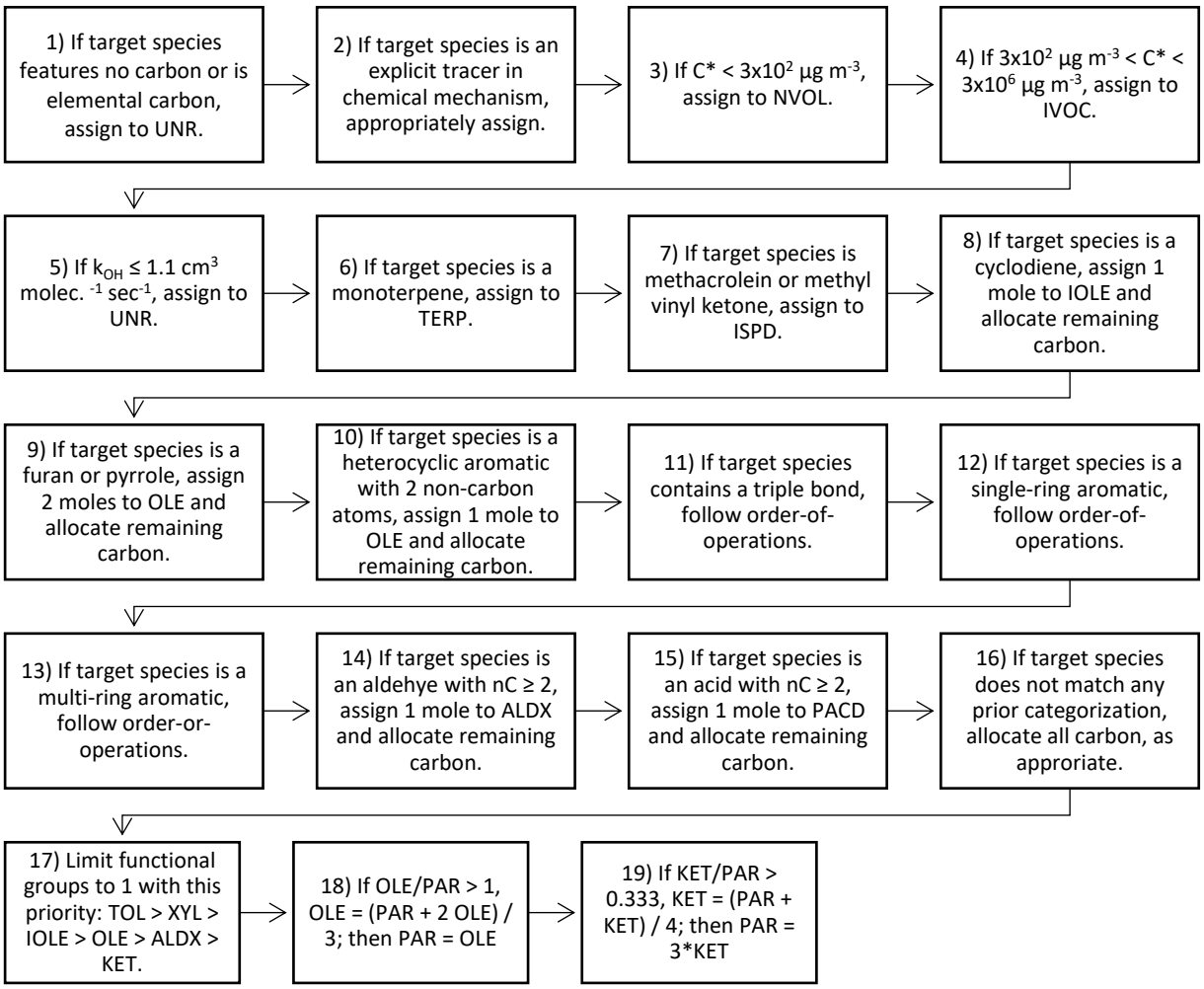
When compared to CB6R3_AE7, the CB6R5_AE7 and CB7_AE7 gas-phase chemical mechanism features updates to reaction rates and an added chemical reaction (Yarwood et al., 2020). No new chemical species were added. As such, equivalent gas-phase mechanism mapping scripts can be used for CB6R3_AE7 and CB6R5_AE7. Mapping for the AE7 aerosol-phase module within CMAQ introduced the NVOL and IVOC tracers. As noted above, NVOL and IVOC were defined to include semi-volatile organic compounds (SVOC; $3x10^{-1}$ µg m^{-3} < C* < $3x10^{2}$ µg m^{-3}) and intermediate-volatility organic compounds (IVOC; $3x10^{2}$ µg m^{-3} < C* < $3x10^{6}$ µg m^{-3}), respectively. A description and various properties for all emitted gaseous organic model species from the CB6R3_AE7, CB6R5_AE7, and CB7_AE7 chemical mechanisms are shown below.

Model Species	Description	nC	Tracer Type	SMILES
AACD	Acetic Acid	2	Explicit	CC(O)=O
ACET	Acetone	3	Explicit	CC(C)=O
ALD2	Acetaldehyde	2	Explicit	CC=O
APIN	alpha-pinene	10	Explicit	CC1=CCC2CC1C2(C)C
BENZ	Benzene	6	Explicit	C1=CC=CC=C1
CH4	Methane	1	Explicit	C
ETH	Ethene	2	Explicit	C=C
ETHA	Ethane	2	Explicit	CC
ETHY	Ethyne	2	Explicit	C#C
ETOH	Ethanol	2	Explicit	CCO
FACD	Formic acid	1	Explicit	OC=O
FORM	Formaldehyde	1	Explicit	C=O
GLY	Glyoxal	2	Explicit	O=CC=O
GLYD	Glycolaldehyde	2	Explicit	OCC=O

Model Species	Description	nC	Tracer Type	SMILES
ISOP	Isoprene	5	Explicit	CC(=C)C=C
MEOH	Methanol	1	Explicit	CO
MGLY	Methylglyoxal	3	Explicit	CC(=O)C=O
NAPH	Naphthalene	10	Explicit	C1=CC2=CC=CC=C2C=C1
PRPA	Propane	3	Explicit	CCC
ALDX	Propionaldehyde and higher aldehydes	2	Lumped	-
CAT1	Methyl-catechols	7	Lumped	-
CRES	Cresols	7	Lumped	-
CRON	Nitro-cresols	7	Lumped	-
IOLE	Internal olefin carbon bond (R-C=C-R)	4	Lumped	-
ISPD	Isoprene products	4	Lumped	-
IVOC	Intermediate-Volatile Organic Compounds	12	Lumped	-
KET	Ketone carbon bond (C=O)	1	Lumped	-
NVOL	Nonvolatile	18	Lumped	-
OLE	Terminal olefin carbon bond (R-C=C)	2	Lumped	-
PACD	Peroxyacetic and higher peroxycarboxylic acids	2	Lumped	-
PAR	Paraffin carbon bond (C-C)	1	Lumped	-
TERP	Monoterpenes	10	Lumped	-
TOL	Toluene and other monoalkyl aromatics	7	Lumped	-
UNR	Unreactive	1	Lumped	-
XYLMN	Xylene and other polyalkyl aromatics	8	Lumped	-

Table 17: Model species, description, number of carbons, tracer type, and SMILES string (where applicable) for all emitted organics in the CB6R3_AE7, CB6R5_AE7, CB7_AE7 chemical mechanisms.

Below, the order-of-operations for assignment of a target species are outlined and subsequently discussed in more detail.



1) If a target species features no carbon or is elemental carbon, it is assigned to UNR.
2) If a target species is an explicit tracer in the chemical mechanism, it is appropriately assigned. This is determined using the SMILES string and applies to AACD, ACET, ALD2, BENZ, CH4, ETH, ETHA, ETHY, ETOH, FACD, FORM, GLY, GLYD, ISOP, MEOH, MGLY, NAPH, PRPA, and APIN.
3) If the saturation concentration (C*) of the target species is < $3x10^2$ µg m^{-3}, it is assigned to NVOL. Note: like prior mechanism mapping methods, double mapping for compounds assigned to NVOL is not performed.
4) If the C* of the target species is $3x10^2$ µg m^{-3} < C* < $3x10^6$ µg m^{-3}, it is assigned to IVOC. Note: like prior mechanism mapping methods, double mapping for compounds assigned to IVOC is not performed.
5) If the k_{OH} of the target species is ≤ 1.1 x 10^{-12} cm^3 molec. $^{-1}$ sec^{-1}, it is assigned to UNR.
6) If the target species is a monoterpene (nC = 10, nH = 18, nO = 1 or nC = 10 and nH = 16), and not alpha-pinene, it is assigned to TERP.
7) If the target species is methacrolein or methyl vinyl ketone (nC = 4, nH = 6, nO = 1, number of ketone groups = 1), it is assigned to ISPD.
8) If the target species is a cyclodiene (number of C=C = 2 and contains a ring), assign 1 mole to IOLE and count remaining carbons (nC – 4). If nC – 4 > 0, assign remaining carbon to OLE, IOLE, KET, or PAR, depending on structure.

9) If the target species is a furan or a pyrrole, assign 2 moles to OLE and count remaining carbons (nC – 4). If nC – 4 > 0, assign remaining carbon to OLE, IOLE, KET, or PAR, depending on structure.
10) If the target species is a heterocyclic aromatic compound with 2 non-carbon atoms, assign 1 mole to OLE and count remaining carbons (nC – 2). If nC – 2 > 0, assign remaining carbon to OLE, IOLE, KET, or PAR, depending on structure.
11) If the target species contains a triple bond, do the following order-of-operations:
    a. If the target species contains no other reactive functional groups beyond a triple bond, assign 1 mole to OLE and count remaining carbons (nC – 2). If nC – 2 > 0, assign remaining carbon to PAR.
    b. Else, assign all triple bonds to PAR and count remaining carbons (nC - number of C-C triple bonds). If nC - number of C-C triple bonds > 0, assign remaining carbon to OLE, IOLE, KET, or PAR, depending on structure.
12) If the target species is a single-ring aromatic (number of benzene rings = 1), do the following order-of-operations:
    a. If the target species is a methyl-catechol (nC ≥ 7, number of alcohol groups ≥ 2), it is assigned to CAT1.
    b. If the target species is a nitro-cresol (nC ≥ 7, nN ≥ 1, number of alcohol groups ≥ 1), it is assigned to CRON.
    c. If the target species is an isomer of cresol ((nC = 7, number of alcohol groups = 1), it is assigned to CRES.
    d. If the target species is halogenated and features 4 or more halogens, assigned 6 moles to UNR.
    e. If the target species is halogenated and features more than 1 halogen, assign 1 mole to PAR and 5 moles to UNR.
    f. If the target species is toluene or another monoalkyl aromatic (nC ≥ 7, number of benzene rings = 1, number of branches on benzene ring = 1), assign 1 mole to TOL and count remaining carbons (nC – 7). If nC – 7 > 0, assign remaining carbon to PAR.
    g. If the target species is an isomer of xylene or another polyalkyl aromatic (nC ≥ 8, number of benzene rings = 1, number of branches on benzene ring > 1), assign 1 mole to XYLMN and count remaining carbons (nC – 8). If nC – 8 > 0, assign remaining carbon to PAR.
    h. For all remaining target species, assign 1 mole to BENZ and count remaining carbons (nC – 6). If nC – 6 > 0, assign remaining carbon to PAR.
13) If the target species is a multi-ring aromatic (number of benzene rings > 1 and nO = 0), it is likely assigned to NVOL, IVOC, or NAPH. If not, assign carbon to OLE, IOLE, KET, or PAR, depending on structure.
14) If the target species is an aldehyde with nC ≥ 2, assign 1 mole to ALDX and count remaining carbons (nC – 2). If nC – 2 > 0, assign remaining carbon to OLE, IOLE, KET, or PAR, depending on structure.
15) If the target species is an acid with nC ≥ 2, assign 1 mole to PACD and count remaining carbons (nC – 2). If nC – 2 > 0, assign remaining carbon to PAR.
16) For all remaining target species, assign all carbon to OLE, IOLE, KET, or PAR, depending on structure.
17) Following assignments, and per recommendations by mechanism developer, limit functional groups to 1 with this priority: TOL > XYL > IOLE > OLE > ALDX > KET. Assign all carbon mass not assigned to a functional group to PAR.
18) Following assignments, and per recommendations by mechanism developer, if ratio of OLE/PAR > 1, recalculate: OLE = (PAR + 2 OLE) / 3; then assign PAR = OLE.

19) Following assignments, and per recommendations by mechanism developer, if ratio of KET/PAR > 0.333, recalculate: KET = (PAR + KET) / 4; then PAR = 3*KET.

3.3.2 CB6R4_CF2

The CB6R4_CF2 chemical mechanism is exclusively used in the CAMx photochemical model and features significant overlap with the CB6R3_AE7 chemical mechanism. Specifically, several explicit species and one lumped species is excluded from the CB6R4_CF2 chemical mechanism when compared to CB6R3_AE7. Both utilize the NVOL and IVOC tracers, which are defined to include semi-volatile organic compounds (SVOC; $3x10^{-1}$ µg m^{-3} < C* < $3x10^{2}$ µg m^{-3}) and intermediate-volatility organic compounds (IVOC; $3x10^{2}$ µg m^{-3} < C* < $3x10^{6}$ µg m^{-3}), respectively. A description and various properties for all emitted organics from the CB6R4_CF2 chemical mechanism is shown below.

Model Species	Description	nC	Tracer Type	SMILES
ACET	Acetone	3	Explicit	CC(C)=O
ALD2	Acetaldehyde	2	Explicit	CC=O
BENZ	Benzene	6	Explicit	C1=CC=CC=C1
CH4	Methane	1	Explicit	C
ETH	Ethene	2	Explicit	C=C
ETHA	Ethane	2	Explicit	CC
ETHY	Ethyne	2	Explicit	C#C
ETOH	Ethanol	2	Explicit	CCO
FORM	Formaldehyde	1	Explicit	C=O
ISOP	Isoprene	5	Explicit	CC(=C)C=C
MEOH	Methanol	1	Explicit	CO
PRPA	Propane	3	Explicit	CCC
ALDX	Propionaldehyde and higher aldehydes	2	Lumped	-
CAT1	Methyl-catechols	6	Lumped	-
CRES	Cresols	6	Lumped	-
CRON	Nitro-cresols	6	Lumped	-
IOLE	Internal olefin carbon bond (R-C=C-R)	4	Lumped	-
IVOC	Intermediate-Volatile Organic Compounds	12	Lumped	-
KET	Ketone carbon bond (C=O)	1	Lumped	-
NVOL	Nonvolatile	18	Lumped	-
OLE	Terminal olefin carbon bond (R-C=C)	2	Lumped	-
PACD	Peroxyacetic and higher peroxycarboxylic acids	2	Lumped	-
PAR	Paraffin carbon bond (C-C)	1	Lumped	-
TERP	Monoterpenes	10	Lumped	-
TOL	Toluene and other monoalkyl aromatics	7	Lumped	-
UNR	Unreactive	1	Lumped	-
XYL	Xylene and other polyalkyl aromatics	8	Lumped	-

Table 18: Model species, description, number of carbons, tracer type, and SMILES string (where applicable) for all emitted organics in the CB6R4_CF2 chemical mechanism.

Since the only changes between CB6R4_CF2 and CB6R3_AE7 relate to the mapping of several explicit species and one lumped species, Step 2 from Section 3.3.1 must be updated and Step 7 must be deleted. For CB6R4_CF2, Step 2 should read as:

2) If a target species is an explicit tracer in the chemical mechanism, it is appropriately assigned. This is determined using the SMILES string and applies to ACET, ALD2, BENZ, CH4, ETH, ETHA, ETHY, ETOH, FORM, ISOP, MEOH, and PRPA.

3.3.3 CB7_CF2

The CB7_CF2 chemical mechanism is exclusively used in the CAMx photochemical model and features significant overlap with the CB6R4_CF2 chemical mechanism. Specifically, alpha-pinene (APIN) was added to CB7_CF2. Both utilize the NVOL and IVOC tracers, which are defined to include semi-volatile organic compounds (SVOC; $3x10^{-1}$ µg m^{-3} < C* < $3x10^{2}$ µg m^{-3}) and intermediate-volatility organic compounds (IVOC; $3x10^{2}$ µg m^{-3} < C* < $3x10^{6}$ µg m^{-3}), respectively. A description and various properties for all emitted organics from the CB7_CF2 chemical mechanism is shown below.

Model Species	Description	nC	Tracer Type	SMILES
ACET	Acetone	3	Explicit	CC(C)=O
ALD2	Acetaldehyde	2	Explicit	CC=O
APIN	alpha-pinene	10	Explicit	CC1=CCC2CC1C2(C)C
BENZ	Benzene	6	Explicit	C1=CC=CC=C1
CH4	Methane	1	Explicit	C
ETH	Ethene	2	Explicit	C=C
ETHA	Ethane	2	Explicit	CC
ETHY	Ethyne	2	Explicit	C#C
ETOH	Ethanol	2	Explicit	CCO
FORM	Formaldehyde	1	Explicit	C=O
ISOP	Isoprene	5	Explicit	CC(=C)C=C
MEOH	Methanol	1	Explicit	CO
PRPA	Propane	3	Explicit	CCC
ALDX	Propionaldehyde and higher aldehydes	2	Lumped	-
CAT1	Methyl-catechols	6	Lumped	-
CRES	Cresols	6	Lumped	-
CRON	Nitro-cresols	6	Lumped	-
IOLE	Internal olefin carbon bond (R-C=C-R)	4	Lumped	-
IVOC	Intermediate-Volatile Organic Compounds	12	Lumped	-
KET	Ketone carbon bond (C=O)	1	Lumped	-
NVOL	Nonvolatile	18	Lumped	-
OLE	Terminal olefin carbon bond (R-C=C)	2	Lumped	-
PACD	Peroxyacetic and higher peroxycarboxylic acids	2	Lumped	-
PAR	Paraffin carbon bond (C-C)	1	Lumped	-
TERP	Monoterpenes	10	Lumped	-
TOL	Toluene and other monoalkyl aromatics	7	Lumped	-
UNR	Unreactive	1	Lumped	-
XYL	Xylene and other polyalkyl aromatics	8	Lumped	-

Table 19: Model species, description, number of carbons, tracer type, and SMILES string (where applicable) for all emitted organics in the CB6R4_CF2 chemical mechanism.

Since the only changes between CB6R4_CF2 and CB6R3_AE7 relate to the mapping of several explicit species and one lumped species, Step 2 from Section 3.3.1 must be updated and Step 7 must be deleted. For CB6R4_CF2, Step 2 should read as:

3) If a target species is an explicit tracer in the chemical mechanism, it is appropriately assigned. This is determined using the SMILES string and applies to ACET, ALD2, APIN, BENZ, CH4, ETH, ETHA, ETHY, ETOH, FORM, ISOP, MEOH, and PRPA.

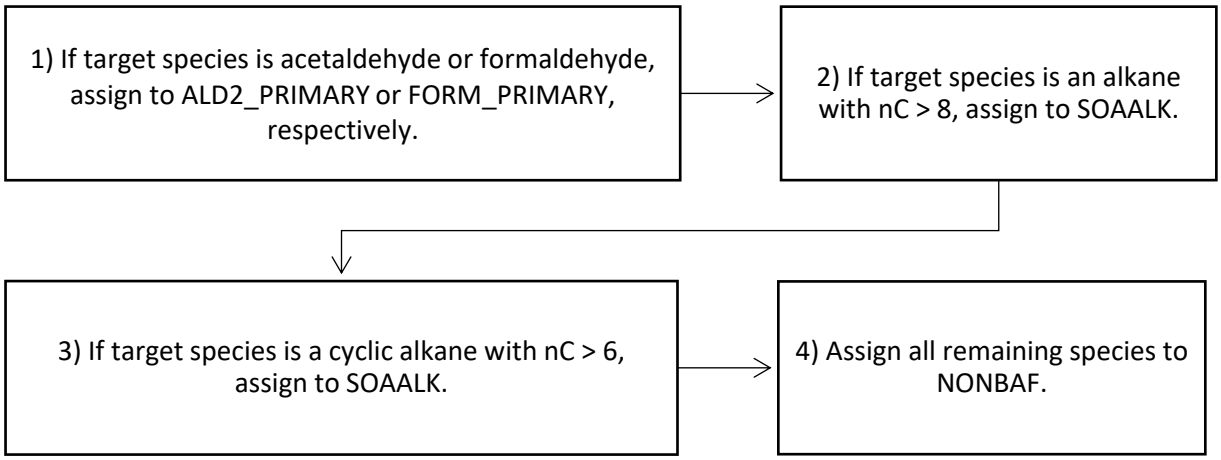
3.3.4 CB6R3_AE7_TRACER

The CB6R3_AE7_TRACER chemical mechanism serves two purposes. First, it assigns SOAALK using the methods described below (Pye and Pouliot, 2012). Second, it creates acetaldehyde and formaldehyde tracers, which are used in HAP-CAP modeling (see Section 1.1). A description of the tracers contained in the CB6R3_AE7_TRACER chemical mechanism is shown below.

Model Species	Description	Tracer Type	SMILES
ALD2_PRIMARY	Acetaldehyde	Explicit	CC=O
FORM_PRIMARY	Formaldehyde	Explicit	C=O
NONBAF	Not SOAALK, Acetaldehyde, Formaldehyde	Lumped	-
SOAALK	Long-chain, SOA precursor alkanes	Lumped	-

Table 20: Model species, description, tracer type, and SMILES string (where applicable) for all species organics in the CB6R3_AE7_TRACER chemical mechanisms. Note: NONBAF is dropped from emissions generation process.

The order-of-operations for assignment of a target species using the CB6R3_AE7_TRACER is outlined below.



In the above diagram, the definition of SOAALK follows Pye and Pouliot (2012), which defines SOAALK as C6 and larger cyclic and C8 and larger linear/branched alkanes.

3.3.5 CRACMMv1.0

The Community Regional Atmospheric Chemistry Multiphase Mechanism (CRACMM) v1.0 (Pye et al., 2023, https://doi.org/USEPA/CRACMM) is a chemical mechanism that builds off the history of the Regional Atmospheric Chemistry Mechanism (RACM; Stockwell et al., 1997) and RACM version 2 (Goliff et al., 2013). Unlike the Carbon Bond series of mechanisms, CRACMM retains the carbon backbone of emitted species, and unlike all prior chemical mechanisms, CRACMM fully integrates the gas- and aerosol-phase portions of atmospheric chemistry. This integration results in all condensible or soluble precursors to SOA being formed directly as gas-phase products or react heterogeneously and form SOA. Therefore, formation of SOA in CRACMM removes mass from the gas phase. Part of the development of CRACMM

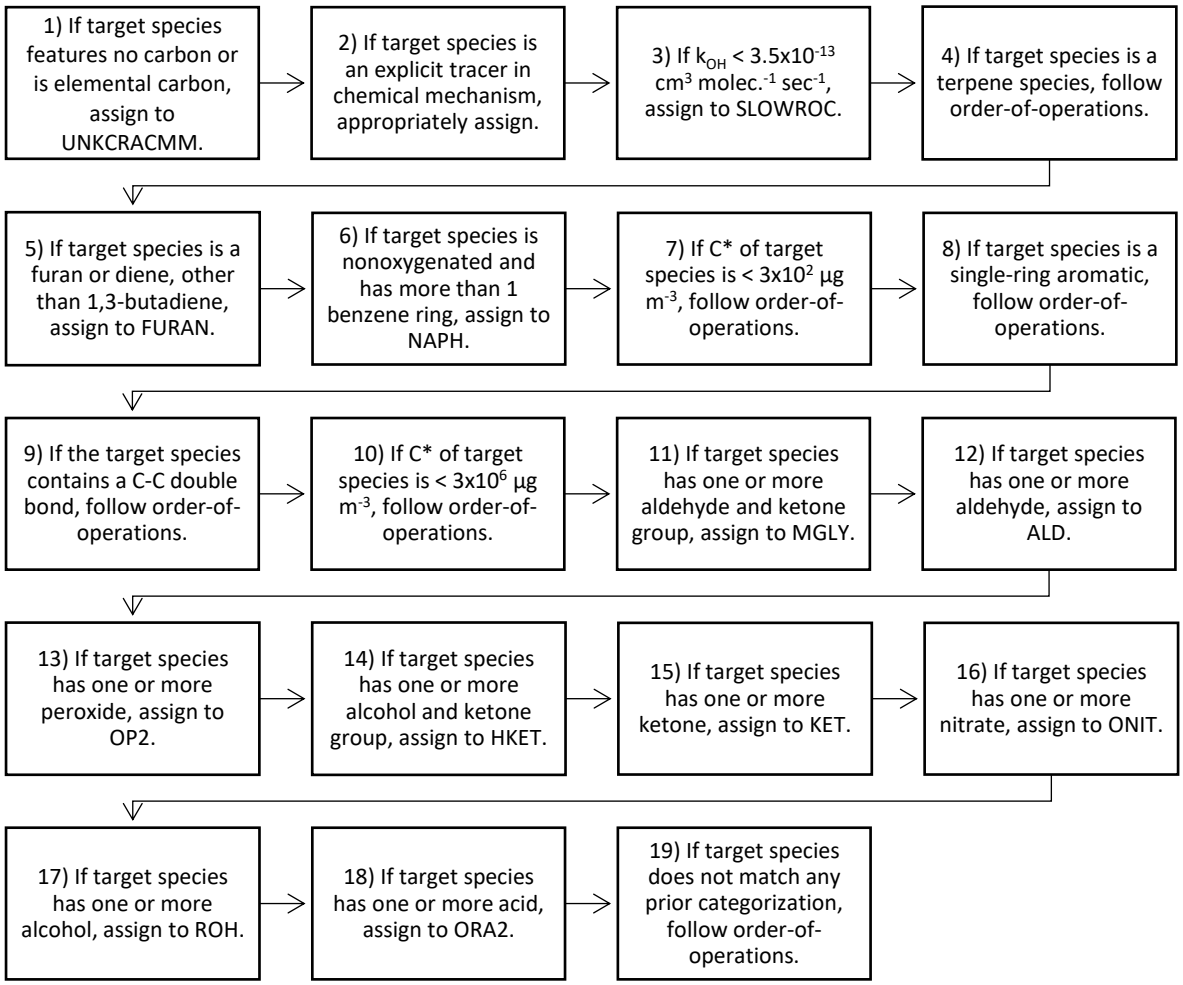
included coding of an emission mapper. A description and various properties for all emitted organics from the CRACMMv1.0 chemical mechanism is shown below.

Model Species	Description	nC	Tracer Type	SMILES
ACD	Acetaldehyde	2	Explicit	CC=O
ACE	Acetylene	2	Explicit	C#C
ACRO	Acrolein	3	Explicit	C=CC=O
ACT	Acetone	3	Explicit	CC(=O)C
ALD	Higher aldehydes	3	Lumped	-
API	Alpha-pinene monoterpenes	10	Lumped	-
BALD	Benzaldehyde and aromatic aldehydes	7	Lumped	-
BDE13	1,3-butadiene	4	Explicit	C=CC=C
BEN	Benzene	6	Explicit	C1=CC=CC=C1
CSL	Cresols	9	Lumped	-
DCB1	Unsaturated dicarbonyls	5	Lumped	-
ECH4	Methane	1	Explicit	C
EOH	Ethanol	2	Explicit	CCO
ETE	Ethene	2	Explicit	C=C
ETEG	Ethylene Glycol	2	Explicit	C(CO)O
ETH	Ethane	2	Explicit	CC
FURAN	Furans and other dienes	5	Lumped	-
GLY	Glyoxal	2	Explicit	C(=O)C=O
HC10	“alkanes”; 6.8E-12 < k_{OH}	10	Lumped	
HC3	“alkanes”; k_{OH} < 3.4E-12	3	Lumped	
HC5	“alkanes”; 3.4E-12 ≤ k_{OH} ≤ 6.8E-12	5	Lumped	
HCHO	Formaldehyde	1	Explicit	C=O
HKET	Hydroxy ketones	3	Lumped	-
ISO	Isoprene	5	Explicit	CC(=C)C=C
KET	Other ketones	5	Lumped	-
LIM	Limonene monoterpenes	10	Lumped	-
MACR	Methacrolein and Crotonaldehyde	4	Lumped	-
MCT	Methylcatechol	7	Lumped	-
MEK	Methyl ethyl ketone	4	Explicit	CCC(=O)C
MGLY	Methylglyoxal and similar species	3	Lumped	-
MOH	Methanol	1	Explicit	CO
MVK	Methyl vinyl ketone	4	Explicit	CC(=O)C=C
NAPH	PAH with 2 aromatic rings	10	Lumped	-
OLI	Internal alkenes	5	Lumped	-
OLT	Terminal alkenes	3	Lumped	-
ONIT	Organic nitrates	4	Lumped	-
ORA1	Formic Acid	1	Explicit	C(=O)O
ORA2	Acetic acid and higher acids	2	Lumped	-
PHEN	Phenols	6	Lumped	-
PROG	Propylene Glycol	3	Explicit	CC(CO)O
ROCIOXY	ROC; log(C*) < 6.5 and O/C > 0.1	12	Lumped	-
ROCN1ALK	ROC; -1.5 < log(C*) < -0.5	29	Lumped	-

Model Species	Description	nC	Tracer Type	SMILES
ROCN2ALK	ROC; log(C*) < -1.5	30	Lumped	-
ROCP0ALK	ROC; -0.5 < log(C*) < 0.5	28	Lumped	-
ROCP1ALK	ROC; 0.5 < log(C*) < 1.5	27	Lumped	-
ROCP2ALK	ROC; 1.5 < log(C*) < 2.5	24	Lumped	-
ROCP3ALK	ROC; 2.5 < log(C*) < 3.5	21	Lumped	-
ROCP4ALK	ROC; 3.5 < log(C*) < 4.5	18	Lumped	-
ROCP5ALK	ROC; 4.5 < log(C*) < 5.5	14	Lumped	-
ROCP5ARO	Aromatic ROC; 4.5 < log(C*) < 5.5	14	Lumped	-
ROCP6ALK	ROC; 5.5 < log(C*) < 6.5	13	Lumped	-
ROCP6ARO	Aromatic ROC; 5.5 < log(C*) < 6.5	13	Lumped	-
ROH	C3 and higher alcohols	3	Lumped	-
SESQ	Sesquiterpenes	15	Lumped	-
SLOWROC	Low reactivity gas	1	Lumped	-
TOL	Toluene	7	Explicit	CC1=CC=CC=C1
UALD	Unsaturated aldehydes	5	Lumped	-
UNKCRACMM	Unknown	1	Lumped	-
UNKKOH	Unknown k_{OH}	1	Lumped	-
UNKSMILES	Unknown SMILES	1	Lumped	-
XYE	o-, p-xylenes and less reactive aromatics	8	Lumped	-
XYM	m-xylene and more reactive aromatics	8	Lumped	-

Table 21: Model species, description, number of carbons, tracer type, and SMILES string (where applicable) for all emitted organics in the CRACMMv1.0 chemical mechanism.

Below, the order-of-operations for assignment of a target species are outlined and subsequently discussed in more detail (adapted from the work of Pye et al., 2023).



1) If a target species features no carbon or is elemental carbon, it is assigned to UNKCRACMM.
2) If a target species is an explicit tracer in the chemical mechanism, it is appropriately assigned. This is determined using the SMILES string and applies to ACD, ACE, ACT, BEN, ECH4, EOH, ETE, ETEG, ETH, HCHO, ISO, MOH, ORA1, OP1, MEK, MVK, TOL, BCE13, ACRO, GLY, and PROG.
3) If the k_{OH} of the target species is < 3.5x10^{-13} cm^3 molec.$^{-1}$ sec^{-1}, it is assigned to SLOWROC.
4) If the target species is a monoterpene, do the following order-of-operations:
    a. If the target species has nC = 15, nH = 24, and the number of C=C ≥ 1, it is assigned to SESQ.
    b. If the target species is a monoterpene and the number of C=C = 1, it is assigned to API.
    c. If the target species is a monoterpene and the number of C=C > 1, it is assigned to LIM.
5) If the target is a furan other than 1,3-butadiene, it is assigned to FURAN.
6) If the target has more than one benzene ring and nO = 0, it is assigned to NAPH.
7) If the C* of the target species is < 3x10^2 µg m^{-3}, it is assigned to a semi-volatile organic compound model species:
    a. If C* < 3x10^{-2} µg m^{-3}, it is assigned to ROCN2ALK.
    b. If 3x10^{-2} µg m^{-3} < C* < 3x10^{-1} µg m^{-3}, it is assigned to ROCN1ALK.
    c. If 3x10^{-1} µg m^{-3} < C* < 3x10^0 µg m^{-3}, it is assigned to ROCP0ALK.
    d. If 3x10^0 µg m^{-3} < C* < 3x10^1 µg m^{-3}, it is assigned to ROCP1ALK.
    e. If 3x10^1 µg m^{-3} < C* < 3x10^2 µg m^{-3}, it is assigned to ROCP2ALK.

8) If the target species is a single-ring aromatic, do the following order-of-operations:
    a. If the target species contains an aldehyde, it is assigned to BALD.
    b. If the target species contains 2 or more alcohol and nC ≥ 7, it is assigned to MCT.
    c. If the target species contains 1 alcohol and nC ≥ 7, it is assigned to CSL.
    d. If the target species contains 1 or more alcohol and nC = 6, it is assigned to PHEN.
    e. If C* < $3x10^5$ µg m^{-3}, it is assigned to ROCP5ARO.
    f. If $3x10^5$ µg m^{-3} < C* < $3x10^6$ µg m^{-3}, it is assigned to ROCP6ARO.
    g. If the k_{OH} of the target species is ≤ 1.02 * $1.43x10^{-11}$ cm^3 $molec.^{-1}$ sec^{-1}, it is assigned to XYE.
    h. Else, it is assigned to XYM
9) If the number of C=C ≥ 1 for the target species, do the following order-of-operations:
    a. If C* < $3x10^5$ µg m^{-3}, it is assigned to ROCP5ARO.
    b. If $3x10^5$ µg m^{-3} < C* < $3x10^6$ µg m^{-3}, it is assigned to ROCP6ARO.
    c. If the number of C=C ≥ 2 for the target species, it is assigned to FURAN.
    d. If the number of C=C = 1 and there are more than 1 carbonyl groups, it is assigned to DCB1.
    e. If the number of C=C = 1, nC =4, and there is one aldehyde group, it is assigned to MACR.
    f. If the number of C=C = 1 and there is one or more aldehyde group, it is assigned to UALD.
    g. If the number of C=C = 1 and it is a terminal olefin carbon bond, it is assigned to OLT.
    h. If the number of C=C = 1, it is assigned to OLI.
10) If the C* of the target species is C* < $3x10^6$ µg m^{-3}, do the following order-of-operations:
    a. If C* < $3x10^6$ µg m^{-3} and nO/nC ≥ 0.1, or if nSi > 0, it is assigned to ROCIOXY.
    b. If C* < $3x10^3$ µg m^{-3}, it is assigned to ROCP3ALK.
    c. If $3x10^3$ µg m^{-3} < C* < $3x10^4$ µg m^{-3}, it is assigned to ROCP4ALK.
    d. If $3x10^4$ µg m^{-3} < C* < $3x10^5$ µg m^{-3}, it is assigned to ROCP5ALK.
    e. If $3x10^5$ µg m^{-3} < C* < $3x10^6$ µg m^{-3}, it is assigned to ROCP6ALK.
11) If the number of ketone groups ≥ 1 and number of aldehyde groups ≥ 1 in the target species, it is assigned to MGLY.
12) If the number of aldehyde groups ≥ 1 in the target species, it is assigned to ALD.
13) If the number of peroxide groups ≥ 1 in the target species, it is assigned to OP2.
14) If the number of alcohol groups ≥ 1 and number of ketone groups ≥ 1 in the target species, it is assigned to HKET.
15) If the number of ketone groups ≥ 1 in the target species, it is assigned to KET.
16) If the number of nitrate groups ≥ 1 in the target species, it is assigned to ONIT.
17) If the number of alcohol groups ≥ 1 in the target species, it is assigned to ROH.
18) If the number of acid groups ≥ 1 in the target species, it is assigned to ORA2.
19) For all remaining target species, do the following order-of-operations:
    a. If the k_{OH} of the target species is < $3.4x10^{-12}$ cm^3 $molec.^{-1}$ sec^{-1}, it is assigned to HC3.
    b. If the k_{OH} of the target species is $3.4x10^{-12}$ ≤ k_{OH} ≤ $6.8x10^{-12}$ cm^3 $molec.^{-1}$ sec^{-1}, it is assigned to HC5.
    c. If the k_{OH} of the target species is > $6.8x10^{-12}$ cm^3 $molec.^{-1}$ sec^{-1}, it is assigned to HC10.

3.3.6 SAPRC07TC_AE7

Prior to Speciation Tool v5.0, there was no mechanism mapping for SAPRC07TC. When compared to SAPRC07, SAPRC07TC added explicit model species, largely to better represent air toxics. These additional model species include 1,3-butadiene, acrolein, a-pinene, 1,2,4-trimethy benzene, ethanol, o-, m-, p-xylene, propane, sesquiterpenes, and toluene. As with other chemical mechanism that feature the AE7

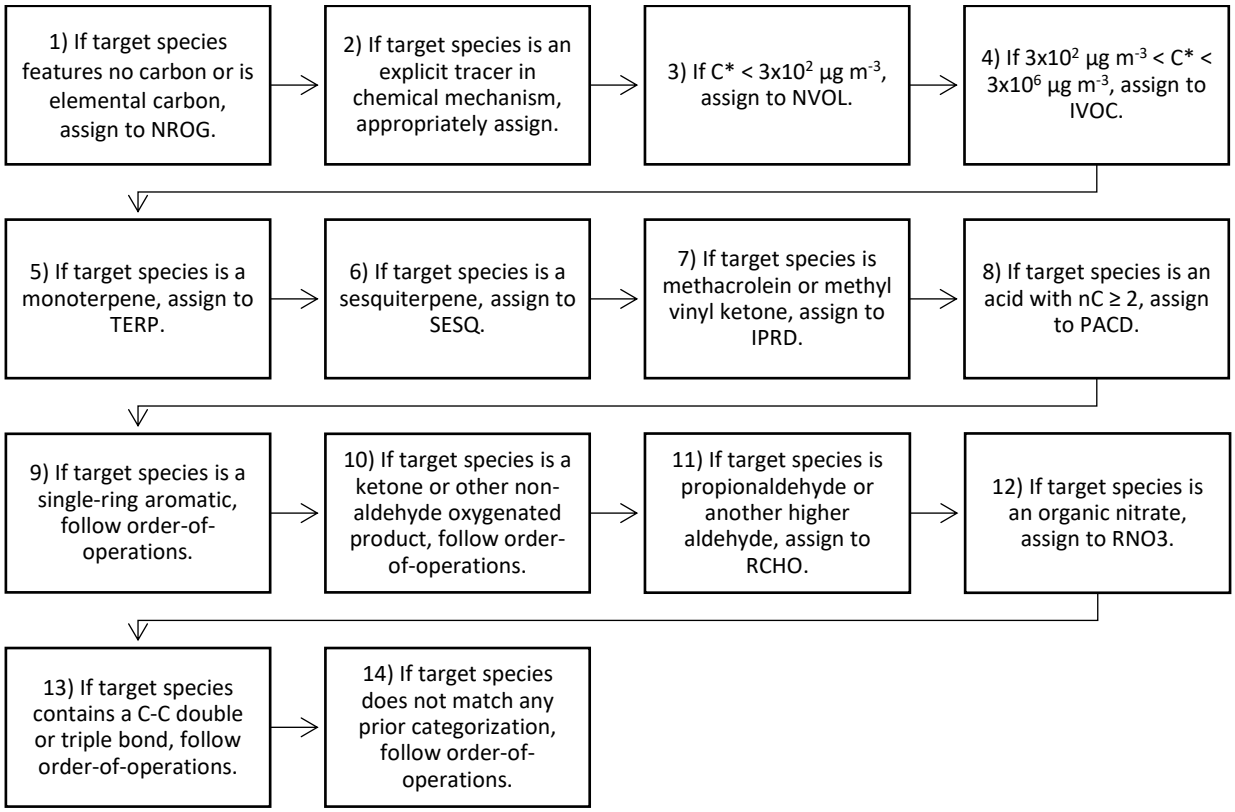
aerosol-phase module within CMAQ, the NVOL and IVOC tracers were defined to include semi-volatile organic compounds and intermediate-volatility organic compounds, respectively. A description and various properties for all emitted organics from the SAPRC07TC_AE7 chemical mechanism are shown below.

Model Species	Description	nC	Tracer Type	SMILES
AACD	Acetic Acid	2	Explicit	CC(O)=O
ACET	Acetone	3	Explicit	CC(C)=O
ACRO	Acrolein	3	Explicit	C=CC=O
ACYE	Acetylene	2	Explicit	C#C
APIN	alpha-Pinene	10	Explicit	CC1=CCC2CC1C2(C)C
B124	1,2,4-trimethyl benzene	9	Explicit	CC1=CC(C)=C(C)C=C1
BACL	Biacetyl	4	Explicit	CC(=O)C(C)=O
BDE13	1,3-butadiene	4	Explicit	C=CC=C
BENZ	Benzene	6	Explicit	C1=CC=CC=C1
CCHO	Acetaldehyde	2	Explicit	CC=O
CH4	Methane	1	Explicit	C
ETHE	Ethene	2	Explicit	C=C
ETOH	Ethanol	2	Explicit	CCO
FACD	Formic Acid	1	Explicit	OC=O
GLY	Glyoxal	2	Explicit	O=CC=O
HCHO	Formaldehyde	1	Explicit	C=O
ISOP	Isoprene	5	Explicit	CC(=C)C=C
MACR	Methacrolein	4	Explicit	CC(=C)C=O
MEOH	Methanol	1	Explicit	CO
MGLY	Methyl Glyoxal	3	Explicit	CC(=O)C=O
MVK	Methyl Vinyl Ketone	4	Explicit	CC(=O)C=C
MXYL	m-Xylene	8	Explicit	CC1=CC(C)=CC=C1
NAPH	Naphthalene	10	Explicit	C1=CC2=CC=CC=C2C=C1
OXYL	o-Xylene	8	Explicit	CC1=C(C)C=CC=C1
PRPE	Propane	3	Explicit	CCC
PXYL	p-Xylene	8	Explicit	CC1=CC=C(C)C=C1
TOLU	Toluene	7	Explicit	CC1=CC=CC=C1
ALK1	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2 and 5 x 10^2 ppm^{-1} min^{-1} (primarily ethane)	2	Lumped	-
ALK2	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^2 and 2.5 x 10^3 ppm^{-1} min^{-1} (primarily propane)	3	Lumped	-
ALK3	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2.5 x 10^3 and 5 x 10^3 ppm^{-1} min^{-1}	4	Lumped	-
ALK4	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^3 and 1 x 10^4 ppm^{-1} min^{-1}	5	Lumped	-

Model Species	Description	nC	Tracer Type	SMILES
ALK5	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} greater than 1 x 10^4 ppm^{-1} min^{-1}	8	Lumped	-
ARO1	Aromatics with k_{OH} < 2x10^4 ppm^{-1} min^{-1}	7	Lumped	-
ARO2	Aromatics with k_{OH} > 2x10^4 ppm^{-1} min^{-1}	9	Lumped	-
BALD	Aromatic aldehydes	7	Lumped	-
CRES	Phenols and Cresols	7	Lumped	-
IPRD	Lumped Isoprene product species	5	Lumped	-
IVOC	Intermediate-Volatile Organic Compounds	12	Lumped	-
MEK	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-13} but slower than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	4	Lumped	-
NROG	Nonreactive Organic Gas	2	Lumped	-
NVOL	Nonvolatile	18	Lumped	-
OLE1	Alkenes (other than ethene) with k_{OH} < 7 x 10^4 ppm^{-1} min^{-1}	5	Lumped	-
OLE2	Alkenes with k_{OH} > 7 x 10^4 ppm^{-1} min^{-1}	5	Lumped	-
PACD	Peroxyacetic and higher peroxycarboxylic acids	3	Lumped	-
PRD2	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	6	Lumped	-
RCHO	Propionaldehyde and larger aldehydes	3	Lumped	-
RNO3	Lumped Organic Nitrates	6	Lumped	-
SESQ	Sesquiterpenes	15	Lumped	-
TERP	Terpenes	10	Lumped	-

Table 22: Model species, description, number of carbons, tracer type, and SMILES string (where applicable) for all emitted organics in the SAPRC07TC_AE7 chemical mechanism.

Below, the order-of-operations for assignment of a target species are outlined and subsequently discussed in more detail.



1) If a target species features no carbon or is elemental carbon, it is assigned to NROG.
2) If a target species is an explicit tracer in the chemical mechanism, it is appropriately assigned. This is determined using the SMILES string and applies to AACD, ACET, ACRO, ACYE, B124, BALC, BDE13, BENZ, CCHO, CH4, ETHE, ETOH, FACD, HCHO, GLY, ISOP, MACR, MEOH, MGLY, MVK, NAPH, PRPE, APIN, TOLU, MXYL, OXYL, and PXYL.
3) If the C* of the target species is < $3x10^2$ µg m^{-3}, it is assigned to NVOL. Note: Like prior mechanism mapping methods, double mapping for compounds assigned to NVOL is not performed.
4) If the C* of the target species is $3x10^2$ µg m^{-3} < C* < $3x10^6$ µg m^{-3}, it is assigned to IVOC. Note: Like prior mechanism mapping methods, double mapping for compounds assigned to IVOC is not performed.
5) If the target species is a monoterpene (nC = 10, nH = 18, nO = 1 or nC = 10 and nH = 16), and not alpha-pinene, it is assigned to TERP.
6) If the target species is a sesquiterpene (nC = 15, nH = 24, number of C=C = 2), it is assigned to SESQ.
7) If the target species is methacrolein or methyl vinyl ketone (nC = 4, nH = 6, nO = 1, number of ketone groups = 1), it is assigned to ISPD.
8) If the target species is an acid with nC ≥ 2, it is assigned to PACD.
9) If the target species is a single-ring aromatic (number of benzene rings = 1), do the following order-of-operations:
    a. If the target species is an isomer of cresol or phenol (6 ≤ nC ≤ 7, number of alcohol groups = 1), it is assigned to CRES.
    b. If the target species contains an aldehyde, it is assigned to BALD.
    c. If the k_{OH} of the target species is ≤ 2.4E4 ppm^{-1} min^{-1}, it is assigned to ARO1.
    d. Else (i.e., k_{OH} of the target species is > 2.4E4 ppm^{-1} min^{-1}), it is assigned to ARO2.

10) If the target species is a ketone or other non-aldehyde oxygenated product (number of ketone groups = 1, number of aldehyde groups = 0), do the following order-of-operations:
    e. If the k_{OH} of the target species is ≥ 5.0E-12 cm^3 $molec.^{-1}$ sec^{-1}, it is assigned to PRD2.
    f. If the k_{OH} of the target species is ≤ 5.0E-12 cm^3 $molec.^{-1}$ sec^{-1} and ≥ 5.0E-13 cm^3 $molec.^{-1}$ sec^{-1}, it is assigned to MEK.
    g. Else (i.e., k_{OH} of the target species is < 5.0E-13 cm^3 $molec.^{-1}$ sec^{-1}), it is assigned to NROG.
11) If the target species is propionaldehyde or another higher aldehyde (nC ≥ 2 and number of aldehyde groups ≥ 1), it is assigned to RCHO.
12) If the target species is an organic nitrate, it is assigned to RNO3.
13) If the target species contains a double or triple carbon-carbon bond, do the following order-of-operations:
    h. If the k_{OH} of the target species is ≤ 7.0E4 ppm^{-1} min^{-1}, it is assigned to OLE1.
    i. Else (i.e., k_{OH} of the target species is > 7.0E4 ppm^{-1} min^{-1}), it is assigned to OLE2.
14) For all remaining target species, do the following order-of-operations:
    j. If the k_{OH} of the target species is ≤ 2.0E2 ppm^{-1} min^{-1}, it is assigned to NROG.
    k. If the k_{OH} of the target species is ≤ 5.0E2 ppm^{-1} min^{-1} and > 2.0E2 ppm^{-1} min^{-1}, it is assigned to ALK1.
    l. If the k_{OH} of the target species is ≤ 2.5E3 ppm^{-1} min^{-1} and > 5.0E2 ppm^{-1} min^{-1}, it is assigned to ALK2.
    m. If the k_{OH} of the target species is ≤ 5.0E3 ppm^{-1} min^{-1} and > 2.5E3 ppm^{-1} min^{-1}, it is assigned to ALK3.
    n. If the k_{OH} of the target species is ≤ 1.0E4 ppm^{-1} min^{-1} and > 5.0E3 ppm^{-1} min^{-1}, it is assigned to ALK4.
    o. If the k_{OH} of the target species is > 1.0E4 ppm^{-1} min^{-1}, it is assigned to ALK5.
    p. Else, all remaining species are assigned to NROG.

3.3.7 SAPRC07_CF2

The SAPRC07_CF2 chemical mechanism is exclusively used in the CAMx photochemical model and features significant overlap with the SAPRC07TC_AE7 chemical mechanism. Specifically, several explicit species and one lumped species is excluded from the SAPRC07_CF2 chemical mechanism when compared to SAPRC07TC _AE7. Both utilize the NVOL and IVOC tracers, which are defined to include semi-volatile organic compounds (SVOC; $3x10^{-1}$ µg m^{-3} < C* < $3x10^{2}$ µg m^{-3}) and intermediate-volatility organic compounds (IVOC; $3x10^{2}$ µg m^{-3} < C* < $3x10^{6}$ µg m^{-3}), respectively. A description and various properties for all emitted organics from the SAPRC07_CF2 chemical mechanism is shown below.

Model Species	Description	nC	Tracer Type	SMILES
AACD	Acetic Acid	2	Explicit	CC(O)=O
ACET	Acetone	3	Explicit	CC(C)=O
ACYE	Acetylene	2	Explicit	C#C
BACL	Biacetyl	4	Explicit	CC(=O)C(C)=O
BENZ	Benzene	6	Explicit	C1=CC=CC=C1
CCHO	Acetaldehyde	2	Explicit	CC=O
CH4	Methane	1	Explicit	C
ETHE	Ethene	2	Explicit	C=C
FACD	Formic Acid	1	Explicit	OC=O

Model Species	Description	nC	Tracer Type	SMILES
GLY	Glyoxal	2	Explicit	O=CC=O
HCHO	Formaldehyde	1	Explicit	C=O
ISOP	Isoprene	5	Explicit	CC(=C)C=C
MACR	Methacrolein	4	Explicit	CC(=C)C=O
MEOH	Methanol	1	Explicit	CO
MGLY	Methyl Glyoxal	3	Explicit	CC(=O)C=O
MVK	Methyl Vinyl Ketone	4	Explicit	CC(=O)C=C
ALK1	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2 and 5 x 10^2 ppm^{-1} min^{-1} (primarily ethane)	2	Lumped	-
ALK2	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^2 and 2.5 x 10^3 ppm^{-1} min^{-1} (primarily propane)	3	Lumped	-
ALK3	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2.5 x 10^3 and 5 x 10^3 ppm^{-1} min^{-1}	4	Lumped	-
ALK4	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^3 and 1 x 10^4 ppm^{-1} min^{-1}	5	Lumped	-
ALK5	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} greater than 1 x 10^4 ppm^{-1} min^{-1}	8	Lumped	-
ARO1	Aromatics with k_{OH} < 2x10^4 ppm^{-1} min^{-1}	7	Lumped	-
ARO2	Aromatics with k_{OH} > 2x10^4 ppm^{-1} min^{-1}	9	Lumped	-
BALD	Aromatic aldehydes	7	Lumped	-
CRES	Phenols and Cresols	7	Lumped	-
IPRD	Lumped Isoprene product species	5	Lumped	-
IVOC	Intermediate-Volatile Organic Compounds	12	Lumped	-
MEK	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-13} but slower than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	4	Lumped	-
NROG	Nonreactive Organic Gas	2	Lumped	-
NVOL	Nonvolatile	18	Lumped	-
OLE1	Alkenes (other than ethene) with k_{OH} < 7 x 10^4 ppm^{-1} min^{-1}	5	Lumped	-
OLE2	Alkenes with k_{OH} > 7 x 10^4 ppm^{-1} min^{-1}	5	Lumped	-
PACD	Peroxyacetic and higher peroxycarboxylic acids	3	Lumped	-
PRD2	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	6	Lumped	-
RCHO	Propionaldehyde and larger aldehydes	3	Lumped	-
RNO3	Lumped Organic Nitrates	6	Lumped	-
TERP	Terpenes	10	Lumped	-

Table 23: Model species, description, number of carbons, tracer type, and SMILES string (where applicable) for all emitted organics in the SAPRC07TC_AE7 chemical mechanism.

Since the only changes between SAPRC07_CF2 and SAPRC07_AE7 relate to the mapping of several explicit species and one lumped species, Step 2 from Section 3.3.6 must be updated and Step 6 must be deleted. For SAPRC07_CF2, Step 2 should read as:

2) If a target species is an explicit tracer in the chemical mechanism, it is appropriately assigned. This is determined using the SMILES string and applies to AACD, ACET, ACYE, BALC, BENZ, CCHO, CH4, ETHE, FACD, HCHO, GLY, ISOP, MACR, MEOH, MGLY, and MVK.

4 Common Applications

4.1 Exporting and utilizing new data from SPECIATE

If a new version of SPECIATE is released, or if new profiles within the database require an updated GSCNV/GSPRO, several input files may require updates. These include the export_profiles, export_species, and export_species_properties files. To generate these files, the export_profiles, export_species, and export_species_properties queries within SPECIATE should be run, the contents should be post-processed to meet the formatting requirements specified in Sections 2.1.2-2.1.4, and the files should be saved as .csv files.

4.2 Adding capabilities for a new gas-phase mechanism

If new emitted gas-phase species are added to an existing chemical mechanism or if a fully new mechanism is made available, updates to input files are necessary. This includes the mechanism_mw file (see Section 2.1.8) and the mech4import file (see Section 2.1.7). For example, if n-decane were made an explicit species (e.g., NDEC) in CB6R5_AE7, a row must be added to the mechanism_mw file for NDEC with a specified molecular weight. Next, the mech4import file would be updated to ensure SPECIES_ID 598 in SPECIATE (n-decane) is mapped to 1 NDEC for CB6R5_AE7. If an entirely new gas-phase mechanism were to be introduced, the entire list of unique species in SPECIATE must be mapped to model species.

4.3 Adding capabilities for a new aerosol-phase mechanism

If new emitted aerosol-phase species are added to an existing chemical mechanism (e.g., a series of oxygenated primary organic aerosol species is added) or if a fully new mechanism is made available. If either of these scenarios arise and updates are needed, the mech_pm file (see Section 2.1.5) and the POA mapping file (see Section 2.1.10) must be updated. In the mech_pm file, the new mechanism must be defined and the SPECIATE species must be mapped to a new mechanism-specific species. For example, say CRACMM were to add oxygenated POA to the base mechanism in a future version (e.g., PM-CR2). Additional lines would be required in the mech_pm file, such as “CMAQ,PM-CR2,3363,ROCP0OXY”. Next, a new column titled “PM-CR2” must be added to the POA mapping file and the row corresponding to 3363 (i.e., the SPECIATE_ID) should read: “3363,OM,0,POC,ROCP0ALK,ROCP0OXY” (see Section 2.1.10).

4.4 Add capabilities for new VOC/PM integrated species

If the user desires additional species to be integrated in the platform speciation process (see Section 1.1), additional species may need to be added to the tbl_tox file. For example, if acrolein were made an integrated species when running CMAQ, in all scenarios (i.e., not just integrated when processing emissions from MOVES), SPECIES_ID = 283 and Species = ACROLEIN would be added to the generic (i.e., tbl_tox_NBAFM.csv) tbl_tox file.

4.5 Updating POA volatility bin assignments

If new data were to become available regarding the volatility distribution of POA for a particular source, updates to the POA Volatility Bins file (see Section 2.1.11) may be warranted. For example, say a new default POA volatility profile were available for commercial cooking and a user wanted to incorporate this profile into the S2S-Tool workflow. The user would simply need to update the relevant rows (e.g., “Combustion, Cooking”) in the POA Volatility Bins file and re-run the Tool.

5 References

Carter, W. P. L.: Development of the SAPRC-07 chemical mechanism, Atmospheric Environment, 44, 5324-5335, doi: 10.1016/j.atmosenv.2010.01.026, 2010.

Goliff, W. S., Stockwell, W. R., Lawson, C. V.: The regional atmospheric chemistry mechanism, version 2, Atmospheric Environment, 68, 174-185, doi: 10.1016/j.atmosenv.2012.11.038, 2013.

Kroll, J. H., Seinfeld, J. H.: Chemistry of secondary organic aerosol: Formation and evolution of low-volatility organics in the atmosphere, Atmospheric Environment, 42, 3593-3624, 10.1016/j.atmosenv.2008.01.003, 2008.

Lu, Q., Murphy, B. N., Qin, M., Adams, P. J., Zhao, Y., Pye, H. O. T., Efstathiou, C., Allen, C., Robinson, A. L.: Simulation of organic aerosol formation during the CalNex study: updated mobile emissions and secondary organic aerosol parameterization for intermediate-volatility organic compounds, Atmos. Chem. Phys., 20, 4313–4332, https://doi.org/10.5194/acp-20-4313-2020, 2020.

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

Odum, J. R., Hoffmann, T., Bowman, F., Collins, D., Flagan, R. C., Seinfeld, J. H.: Gas/Particle Partitioning and Secondary Organic Aerosol Yields, Environ. Sci. Technol., 30, 2580-2585, 1996.

Pankow, J. F.: An absorption model of the gas/particle partitioning involved in the formation of secondary organic aerosol, Atmos. Environ., 28, 2, 189-193, 1994.

Pye, H. O. T., Pouliot, G. A.: Modeling the Role of Alkanes, Polycyclic Aromatic Hydrocarbons, and Their Oligomers in Secondary Organic Aerosol Formation, Environ. Sci. & Tech., 46, 6041−6047, doi: 10.1021/es300409w, 2012.

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

Reff, A., Bhave, P. V., Simon, H., Pace, T. G., Pouliot, G. A., Mobley, J. D., Houyoux, M.: Emissions Inventory of $PM_{2.5}$ Trace Elements across the United States, Environ. Sci. Technol., 43, 5790-5796, doi: 10.1021/es802930x, 2009.

Stockwell, W. R., Kirchner, F., Kuhn, M.: A new mechanism for regional atmospheric chemistry modeling, Journal of Geophysical Research, 102, D22, 25847-25879, 1997.

Yarwood, G., Whitten, G. Z., Jung, J., Heo, G., Allen, D. T.: Development, Evaluation and Testing of Version 6 of the Carbon Bond Chemical Mechanism (CB6), ENVIRON International Corporation, 2010.

Yarwood, G., Shi, Y., Beardsley, R.: Impact of CB6r5 Mechanism Changes on Air Pollutant Modeling in Texas Final Report, Prepared for Texas Commission on Environmental Quality, 2020.

Appendix A: Mechanism-specific species

Mechanism	Model Species	Description	Tracer Type
CB6R3_AE7, CB6R5_AE7, CB7_AE7	AACD	Acetic Acid	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ACET	Acetone	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ALD2	Acetaldehyde	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	APIN	alpha-pinene	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	BENZ	Benzene	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	CH4	Methane	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ETH	Ethene	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ETHA	Ethane	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ETHY	Ethyne	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ETOH	Ethanol	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	FACD	Formic acid	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	FORM	Formaldehyde	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	GLY	Glyoxal	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	GLYD	Glycolaldehyde	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ISOP	Isoprene	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	MEOH	Methanol	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	MGLY	Methylglyoxal	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	NAPH	Naphthalene	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	PRPA	Propane	Explicit
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ALDX	Propionaldehyde and higher aldehydes	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	CAT1	Methyl-catechols	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	CRES	Cresols	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	CRON	Nitro-cresols	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	IOLE	Internal olefin carbon bond (R-C=C-R)	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	ISPD	Isoprene products	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	IVOC	Intermediate-Volatile Organic Compounds	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	KET	Ketone carbon bond (C=O)	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	NVOL	Nonvolatile	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	OLE	Terminal olefin carbon bond (R-C=C)	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	PACD	Peroxyacetic and higher peroxycarboxylic acids	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	PAR	Paraffin carbon bond (C-C)	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	TERP	Monoterpenes	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	TOL	Toluene and other monoalkyl aromatics	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	UNR	Unreactive	Lumped
CB6R3_AE7, CB6R5_AE7, CB7_AE7	XYLMN	Xylene and other polyalkyl aromatics	Lumped
CB6R4_CF2	ACET	Acetone	Explicit
CB6R4_CF2	ALD2	Acetaldehyde	Explicit
CB6R4_CF2	BENZ	Benzene	Explicit
CB6R4_CF2	CH4	Methane	Explicit
CB6R4_CF2	ETH	Ethene	Explicit
CB6R4_CF2	ETHA	Ethane	Explicit

Mechanism	Model Species	Description	Tracer Type
CB6R4_CF2	ETHY	Ethyne	Explicit
CB6R4_CF2	ETOH	Ethanol	Explicit
CB6R4_CF2	FORM	Formaldehyde	Explicit
CB6R4_CF2	ISOP	Isoprene	Explicit
CB6R4_CF2	MEOH	Methanol	Explicit
CB6R4_CF2	PRPA	Propane	Explicit
CB6R4_CF2	ALDX	Propionaldehyde and higher aldehydes	Lumped
CB6R4_CF2	CAT1	Methyl-catechols	Lumped
CB6R4_CF2	CRES	Cresols	Lumped
CB6R4_CF2	CRON	Nitro-cresols	Lumped
CB6R4_CF2	IOLE	Internal olefin carbon bond (R-C=C-R)	Lumped
CB6R4_CF2	IVOC	Intermediate-Volatile Organic Compounds	Lumped
CB6R4_CF2	KET	Ketone carbon bond (C=O)	Lumped
CB6R4_CF2	NVOL	Nonvolatile	Lumped
CB6R4_CF2	OLE	Terminal olefin carbon bond (R-C=C)	Lumped
CB6R4_CF2	PACD	Peroxyacetic and higher peroxycarboxylic acids	Lumped
CB6R4_CF2	PAR	Paraffin carbon bond (C-C)	Lumped
CB6R4_CF2	TERP	Monoterpenes	Lumped
CB6R4_CF2	TOL	Toluene and other monoalkyl aromatics	Lumped
CB6R4_CF2	UNR	Unreactive	Lumped
CB6R4_CF2	XYL	Xylene and other polyalkyl aromatics	Lumped
CB7_CF2	ACET	Acetone	Explicit
CB7_CF2	ALD2	Acetaldehyde	Explicit
CB7_CF2	APIN	alpha-pinene	Explicit
CB7_CF2	BENZ	Benzene	Explicit
CB7_CF2	CH4	Methane	Explicit
CB7_CF2	ETH	Ethene	Explicit
CB7_CF2	ETHA	Ethane	Explicit
CB7_CF2	ETHY	Ethyne	Explicit
CB7_CF2	ETOH	Ethanol	Explicit
CB7_CF2	FORM	Formaldehyde	Explicit
CB7_CF2	ISOP	Isoprene	Explicit
CB7_CF2	MEOH	Methanol	Explicit
CB7_CF2	PRPA	Propane	Explicit
CB7_CF2	ALDX	Propionaldehyde and higher aldehydes	Lumped
CB7_CF2	CAT1	Methyl-catechols	Lumped
CB7_CF2	CRES	Cresols	Lumped
CB7_CF2	CRON	Nitro-cresols	Lumped
CB7_CF2	IOLE	Internal olefin carbon bond (R-C=C-R)	Lumped
CB7_CF2	IVOC	Intermediate-Volatile Organic Compounds	Lumped
CB7_CF2	KET	Ketone carbon bond (C=O)	Lumped
CB7_CF2	NVOL	Nonvolatile	Lumped
CB7_CF2	OLE	Terminal olefin carbon bond (R-C=C)	Lumped

Mechanism	Model Species	Description	Tracer Type
CB7_CF2	PACD	Peroxyacetic and higher peroxycarboxylic acids	Lumped
CB7_CF2	PAR	Paraffin carbon bond (C-C)	Lumped
CB7_CF2	TERP	Monoterpenes	Lumped
CB7_CF2	TOL	Toluene and other monoalkyl aromatics	Lumped
CB7_CF2	UNR	Unreactive	Lumped
CB7_CF2	XYL	Xylene and other polyalkyl aromatics	Lumped
CB6R3_AE7_TRACER	ALD2_PRIMARY	Acetaldehyde	Explicit
CB6R3_AE7_TRACER	FORM_PRIMARY	Formaldehyde	Explicit
CB6R3_AE7_TRACER	NONBAF	Not SOAALK, Acetaldehyde, Formaldehyde	Lumped
CB6R3_AE7_TRACER	SOAALK	Long-chain, SOA precursor alkanes	Lumped
CRACMMv1.0	ACD	Acetaldehyde	Explicit
CRACMMv1.0	ACE	Acetylene	Explicit
CRACMMv1.0	ACRO	Acrolein	Explicit
CRACMMv1.0	ACT	Acetone	Explicit
CRACMMv1.0	ALD	Higher aldehydes	Lumped
CRACMMv1.0	API	Alpha-pinene monoterpenes	Lumped
CRACMMv1.0	BALD	Benzaldehyde and aromatic aldehydes	Lumped
CRACMMv1.0	BDE13	1,3-butadiene	Explicit
CRACMMv1.0	BEN	Benzene	Explicit
CRACMMv1.0	CSL	Cresols	Lumped
CRACMMv1.0	DCB1	Unsaturated dicarbonyls	Lumped
CRACMMv1.0	ECH4	Methane	Explicit
CRACMMv1.0	EOH	Ethanol	Explicit
CRACMMv1.0	ETE	Ethene	Explicit
CRACMMv1.0	ETEG	Ethylene Glycol	Explicit
CRACMMv1.0	ETH	Ethane	Explicit
CRACMMv1.0	FURAN	Furans and other dienes	Lumped
CRACMMv1.0	GLY	Glyoxal	Explicit
CRACMMv1.0	HC10	“alkanes”; 6.8E-12 < k_{OH}	Lumped
CRACMMv1.0	HC3	“alkanes”; k_{OH} < 3.4E-12	Lumped
CRACMMv1.0	HC5	“alkanes”; 3.4E-12 ≤ k_{OH} ≤ 6.8E-12	Lumped
CRACMMv1.0	HCHO	Formaldehyde	Explicit
CRACMMv1.0	HKET	Hydroxy ketones	Lumped
CRACMMv1.0	ISO	Isoprene	Explicit
CRACMMv1.0	KET	Other ketones	Lumped
CRACMMv1.0	LIM	Limonene monoterpenes	Lumped
CRACMMv1.0	MACR	Methacrolein and Crotonaldehyde	Lumped
CRACMMv1.0	MCT	Methylcatechol	Lumped
CRACMMv1.0	MEK	Methyl ethyl ketone	Explicit
CRACMMv1.0	MGLY	Methylglyoxal and similar species	Lumped
CRACMMv1.0	MOH	Methanol	Explicit
CRACMMv1.0	MVK	Methyl vinyl ketone	Explicit
CRACMMv1.0	NAPH	PAH with 2 aromatic rings	Lumped
CRACMMv1.0	OLI	Internal alkenes	Lumped

Mechanism	Model Species	Description	Tracer Type
CRACMMv1.0	OLT	Terminal alkenes	Lumped
CRACMMv1.0	ONIT	Organic nitrates	Lumped
CRACMMv1.0	ORA1	Formic Acid	Explicit
CRACMMv1.0	ORA2	Acetic acid and higher acids	Lumped
CRACMMv1.0	PHEN	Phenols	Lumped
CRACMMv1.0	PROG	Propylene Glycol	Explicit
CRACMMv1.0	ROCIOXY	ROC; log(C*) < 6.5 and O/C > 0.1	Lumped
CRACMMv1.0	ROCN1ALK	ROC; -1.5 < log(C*) < -0.5	Lumped
CRACMMv1.0	ROCN2ALK	ROC; log(C*) < -1.5	Lumped
CRACMMv1.0	ROCP0ALK	ROC; -0.5 < log(C*) < 0.5	Lumped
CRACMMv1.0	ROCP1ALK	ROC; 0.5 < log(C*) < 1.5	Lumped
CRACMMv1.0	ROCP2ALK	ROC; 1.5 < log(C*) < 2.5	Lumped
CRACMMv1.0	ROCP3ALK	ROC; 2.5 < log(C*) < 3.5	Lumped
CRACMMv1.0	ROCP4ALK	ROC; 3.5 < log(C*) < 4.5	Lumped
CRACMMv1.0	ROCP5ALK	ROC; 4.5 < log(C*) < 5.5	Lumped
CRACMMv1.0	ROCP5ARO	Aromatic ROC; 4.5 < log(C*) < 5.5	Lumped
CRACMMv1.0	ROCP6ALK	ROC; 5.5 < log(C*) < 6.5	Lumped
CRACMMv1.0	ROCP6ARO	Aromatic ROC; 5.5 < log(C*) < 6.5	Lumped
CRACMMv1.0	ROH	C3 and higher alcohols	Lumped
CRACMMv1.0	SESQ	Sesquiterpenes	Lumped
CRACMMv1.0	SLOWROC	Low reactivity gas	Lumped
CRACMMv1.0	TOL	Toluene	Explicit
CRACMMv1.0	UALD	Unsaturated aldehydes	Lumped
CRACMMv1.0	UNKCRACMM	Unknown	Lumped
CRACMMv1.0	UNKKOH	Unknown k_{OH}	Lumped
CRACMMv1.0	UNKSMILES	Unknown SMILES	Lumped
CRACMMv1.0	XYE	o-, p-xylenes and less reactive aromatics	Lumped
CRACMMv1.0	XYM	m-xylene and more reactive aromatics	Lumped
SAPRC07TC_AE7	AACD	Acetic Acid	Explicit
SAPRC07TC_AE7	ACET	Acetone	Explicit
SAPRC07TC_AE7	ACRO	Acrolein	Explicit
SAPRC07TC_AE7	ACYE	Acetylene	Explicit
SAPRC07TC_AE7	APIN	alpha-Pinene	Explicit
SAPRC07TC_AE7	B124	1,2,4-trimethyl benzene	Explicit
SAPRC07TC_AE7	BACL	Biacetyl	Explicit
SAPRC07TC_AE7	BDE13	1,3-butadiene	Explicit
SAPRC07TC_AE7	BENZ	Benzene	Explicit
SAPRC07TC_AE7	CCHO	Acetaldehyde	Explicit
SAPRC07TC_AE7	CH4	Methane	Explicit
SAPRC07TC_AE7	ETHE	Ethene	Explicit
SAPRC07TC_AE7	ETOH	Ethanol	Explicit
SAPRC07TC_AE7	FACD	Formic Acid	Explicit
SAPRC07TC_AE7	GLY	Glyoxal	Explicit
SAPRC07TC_AE7	HCHO	Formaldehyde	Explicit
SAPRC07TC_AE7	ISOP	Isoprene	Explicit

Mechanism	Model Species	Description	Tracer Type
SAPRC07TC_AE7	MACR	Methacrolein	Explicit
SAPRC07TC_AE7	MEOH	Methanol	Explicit
SAPRC07TC_AE7	MGLY	Methyl Glyoxal	Explicit
SAPRC07TC_AE7	MVK	Methyl Vinyl Ketone	Explicit
SAPRC07TC_AE7	MXYL	m-Xylene	Explicit
SAPRC07TC_AE7	NAPH	Naphthalene	Explicit
SAPRC07TC_AE7	OXYL	o-Xylene	Explicit
SAPRC07TC_AE7	PRPE	Propane	Explicit
SAPRC07TC_AE7	PXYL	p-Xylene	Explicit
SAPRC07TC_AE7	TOLU	Toluene	Explicit
SAPRC07TC_AE7	ALK1	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2 and 5 x 10^2 ppm^{-1} min^{-1} (primarily ethane)	Lumped
SAPRC07TC_AE7	ALK2	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^2 and 2.5 x 10^3 ppm^{-1} min^{-1} (primarily propane)	Lumped
SAPRC07TC_AE7	ALK3	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2.5 x 10^3 and 5 x 10^3 ppm^{-1} min^{-1}	Lumped
SAPRC07TC_AE7	ALK4	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^3 and 1 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07TC_AE7	ALK5	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} greater than 1 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07TC_AE7	ARO1	Aromatics with k_{OH} < 2x10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07TC_AE7	ARO2	Aromatics with k_{OH} > 2x10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07TC_AE7	BALD	Aromatic aldehydes	Lumped
SAPRC07TC_AE7	CRES	Phenols and Cresols	Lumped
SAPRC07TC_AE7	IPRD	Lumped Isoprene product species	Lumped
SAPRC07TC_AE7	IVOC	Intermediate-Volatile Organic Compounds	Lumped
SAPRC07TC_AE7	MEK	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-13} but slower than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	Lumped
SAPRC07TC_AE7	NROG	Nonreactive Organic Gas	Lumped
SAPRC07TC_AE7	NVOL	Nonvolatile	Lumped
SAPRC07TC_AE7	OLE1	Alkenes (other than ethene) with k_{OH} < 7 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07TC_AE7	OLE2	Alkenes with k_{OH} > 7 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07TC_AE7	PACD	Peroxyacetic and higher peroxycarboxylic acids	Lumped

Mechanism	Model Species	Description	Tracer Type
SAPRC07TC_AE7	PRD2	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	Lumped
SAPRC07TC_AE7	RCHO	Propionaldehyde and larger aldehydes	Lumped
SAPRC07TC_AE7	RNO3	Lumped Organic Nitrates	Lumped
SAPRC07TC_AE7	SESQ	Sesquiterpenes	Lumped
SAPRC07TC_AE7	TERP	Terpenes	Lumped
SAPRC07_CF2	AACD	Acetic Acid	Explicit
SAPRC07_CF2	ACET	Acetone	Explicit
SAPRC07_CF2	ACYE	Acetylene	Explicit
SAPRC07_CF2	BACL	Biacetyl	Explicit
SAPRC07_CF2	BENZ	Benzene	Explicit
SAPRC07_CF2	CCHO	Acetaldehyde	Explicit
SAPRC07_CF2	CH4	Methane	Explicit
SAPRC07_CF2	ETHE	Ethene	Explicit
SAPRC07_CF2	FACD	Formic Acid	Explicit
SAPRC07_CF2	GLY	Glyoxal	Explicit
SAPRC07_CF2	HCHO	Formaldehyde	Explicit
SAPRC07_CF2	ISOP	Isoprene	Explicit
SAPRC07_CF2	MACR	Methacrolein	Explicit
SAPRC07_CF2	MEOH	Methanol	Explicit
SAPRC07_CF2	MGLY	Methyl Glyoxal	Explicit
SAPRC07_CF2	MVK	Methyl Vinyl Ketone	Explicit
SAPRC07_CF2	ALK1	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2 and 5 x 10^2 ppm^{-1} min^{-1} (primarily ethane)	Lumped
SAPRC07_CF2	ALK2	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^2 and 2.5 x 10^3 ppm^{-1} min^{-1} (primarily propane)	Lumped
SAPRC07_CF2	ALK3	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 2.5 x 10^3 and 5 x 10^3 ppm^{-1} min^{-1}	Lumped
SAPRC07_CF2	ALK4	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} between 5 x 10^3 and 1 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07_CF2	ALK5	Alkanes and other non-aromatic compounds that react only with OH and have a k_{OH} greater than 1 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07_CF2	ARO1	Aromatics with k_{OH} < 2x10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07_CF2	ARO2	Aromatics with k_{OH} > 2x10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07_CF2	BALD	Aromatic aldehydes	Lumped

Mechanism	Model Species	Description	Tracer Type
SAPRC07_CF2	CRES	Phenols and Cresols	Lumped
SAPRC07_CF2	IPRD	Lumped Isoprene product species	Lumped
SAPRC07_CF2	IVOC	Intermediate-Volatile Organic Compounds	Lumped
SAPRC07_CF2	MEK	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-13} but slower than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	Lumped
SAPRC07_CF2	NROG	Nonreactive Organic Gas	Lumped
SAPRC07_CF2	NVOL	Nonvolatile	Lumped
SAPRC07_CF2	OLE1	Alkenes (other than ethene) with k_{OH} < 7 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07_CF2	OLE2	Alkenes with k_{OH} > 7 x 10^4 ppm^{-1} min^{-1}	Lumped
SAPRC07_CF2	PACD	Peroxyacetic and higher peroxycarboxylic acids	Lumped
SAPRC07_CF2	PRD2	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	Lumped
SAPRC07_CF2	RCHO	Propionaldehyde and larger aldehydes	Lumped
SAPRC07_CF2	RNO3	Lumped Organic Nitrates	Lumped
SAPRC07_CF2	TERP	Terpenes	Lumped
PM-AE6 (CMAQ)	POC	Organic carbon	--
PM-AE6 (CMAQ)	PEC	Elemental Carbon	--
PM-AE6 (CMAQ)	PSO4	Sulfate	--
PM-AE6 (CMAQ)	PNO3	Nitrate	--
PM-AE6 (CMAQ)	PNH4	Ammonium	--
PM-AE6 (CMAQ)	PNCOM	Particulate Non-Carbon Organic Matter	--
PM-AE6 (CMAQ)	PFE	Iron	--
PM-AE6 (CMAQ)	PAL	Aluminum	--
PM-AE6 (CMAQ)	PSI	Silicon	--
PM-AE6 (CMAQ)	PTI	Titanium	--
PM-AE6 (CMAQ)	PCA	Calcium	--
PM-AE6 (CMAQ)	PMG	Magnesium	--
PM-AE6 (CMAQ)	PK	Potassium	--
PM-AE6 (CMAQ)	PMN	Manganese	--
PM-AE6 (CMAQ)	PNA	Sodium	--
PM-AE6 (CMAQ)	PCL	Chloride	--
PM-AE6 (CMAQ)	PH2O	Particulate Water	--
PM-AE6 (CMAQ)	PMOTHR	Other	--
PM-CR1 (CMAQ)	PEC	Elemental Carbon	--
PM-CR1 (CMAQ)	PSO4	Sulfate	--
PM-CR1 (CMAQ)	PNO3	Nitrate	--
PM-CR1 (CMAQ)	PNH4	Ammonium	--
PM-CR1 (CMAQ)	PFE	Iron	--
PM-CR1 (CMAQ)	PAL	Aluminum	--

Mechanism	Model Species	Description	Tracer Type
PM-CR1 (CMAQ)	PSI	Silicon	--
PM-CR1 (CMAQ)	PTI	Titanium	--
PM-CR1 (CMAQ)	PCA	Calcium	--
PM-CR1 (CMAQ)	PMG	Magnesium	--
PM-CR1 (CMAQ)	PK	Potassium	--
PM-CR1 (CMAQ)	PMN	Manganese	--
PM-CR1 (CMAQ)	PNA	Sodium	--
PM-CR1 (CMAQ)	PCL	Chloride	--
PM-CR1 (CMAQ)	PH2O	Particulate Water	--
PM-CR1 (CMAQ)	PMOTHR	Other	--
PM-CR1 (CMAQ)	ROCN2	ROCN2ALK, Linear Hydrocarbons, C* = 1e-2 ug m-3	--
PM-CR1 (CMAQ)	ROCN1	ROCN1ALK, Linear Hydrocarbons, C* = 1e-1 ug m-3	--
PM-CR1 (CMAQ)	ROCP0	ROCP0ALK, Linear Hydrocarbons, C* = 1e0 ug m-3	--
PM-CR1 (CMAQ)	ROCP1	ROCP1ALK, Linear Hydrocarbons, C* = 1e1 ug m-3	--
PM-CR1 (CMAQ)	ROCP2	ROCP2ALK, Linear Hydrocarbons, C* = 1e2 ug m-3	--
PM-AE6 (CAMx)	POC	Organic carbon	--
PM-AE6 (CAMx)	PEC	Elemental Carbon	--
PM-AE6 (CAMx)	PSO4	Sulfate	--
PM-AE6 (CAMx)	PNO3	Nitrate	--
PM-AE6 (CAMx)	PNH4	Ammonium	--
PM-AE6 (CAMx)	NA	Sodium	--
PM-AE6 (CAMx)	PCL	Chloride	--
PM-AE6 (CAMx)	PH2O	Particulate Water	--
PM-AE6 (CAMx)	PNCOM	Particulate Non-Carbon Organic Matter	--
PM-AE6 (CAMx)	FPRM	Other, Fine Mode	--
PM-AE6 (CAMx)	FCRS	Crustal Emissions, Fine Mode	--

Appendix B: The PM Protocol

Background and Purpose

SPECIATE is EPA's repository of total organic gas and particulate matter (PM) speciation profiles of air pollution sources. Some of the many uses of these source profiles include: (1) creating speciated emissions inventories; (2) estimating hazardous and toxic air pollutant emissions from PM and primary emissions of organic gases; (3) providing input to chemical mass balance receptor models; and (4) verifying profiles derived from ambient measurements by multivariate receptor models (e.g., factor analysis and positive matrix factorization). Here, methods for processing raw $PM_{2.5}$ profile data into a format suitable for import into SPECIATE and use in the Community Multiscale Air Quality (CMAQ) photochemical air quality modeling are described.

PROFILE_TYPE options for $PM_{2.5}$ profiles within SPECIATE include PM-AE6, PM-AE8, PM-CR1, PM-Simplified, and PM. PM-AE6 profiles are post-processed $PM_{2.5}$ profiles developed using the "AE6-Protocol" (Reff et al., 2009) and were first introduced in SPECIATE v4.3. These profiles include species that directly map to tracers used by the AERO6 (AE6) module of CMAQ. Except for the organic matter components, PM-AE6, PM-CR1 are operationally the same. Mathematically, the relationship among the organic matter components for these three profile types is as follows:

$$POC + PNCOM\ (PM - AE6\ species)$$
$$= ROCN2 + ROCN1 + ROCP0 + ROCP1 + ROCP2\ (PM - CR1\ species)$$

In PM-CR1, both POC and PNCOM are split among several tracers that vary by volatility. Each volatility bin spans an order of magnitude, and the value is represented by a letter and number. The letter "N" indicates negative, and "P" indicates positive. The number following "N" or "P" indicates the log base-10 value. For example, N2 = 10^{-2} and P1 = 10^{1}. In PM-CR1 profiles, the organic carbon and non-carbon organic mass are summed together (and thus represented as organic matter; OM). PM-Simplified profiles are limited to five components: nitrate (613), OC (626), sulfate (699), EC (797), and PM other (1884). These components are aggregated from a profile that has been post-processed into a PM-AE6 profile. PM profiles include other, non-mechanism-specific components, such as explicit semi-volatiles and polycyclic aromatic hydrocarbons. Here, the PM Protocol describes the methods used to generate profiles suitable for import into SPECIATE, as well as the post-processing steps necessary to generate PM-AE6 and PM-CR1 profiles.

"PM-Ready" Species

The S2S-Tool, and formerly the Speciation Tool, translates SPECIATE profiles into profiles that are chemical mechanism-specific and generates several files that are input into SMOKE. For each $PM_{2.5}$ PROFILE_TYPE, the S2S-Tool maps SPECIATE species to chemical-mechanism specific species and allocates all remaining mass to an "Other" Category. This functionality allows SPECIATE developers to include more species than are contained within aerosol modules.

Below, all "PM-ready" species in SPECIATE are listed. These species are the only entries from a profile that are used when generating a GSPRO when MECH_BASIS is PM-AE6 or PM-CR1. In other words, if additional species are included in a PM profile (e.g., polycyclic aromatic hydrocarbons), those species and their weight percent are not included in subsequent calculations. It should be noted that this list includes additional species that are included in the calculation of "Other" $PM_{2.5}$ mass and are not explicit in

chemical mechanisms or modules used by the U.S. EPA. Please note that for some species (calcium, magnesium, potassium, sodium, and chloride), there is an order-of-operations for species selection. In some studies, the atomic form of a metal is measured, but not the ionic form. The atomic form results from the use of x-ray fluorescence (XRF) as the measurement technique and the ionic form results from the use of ion chromatography.

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

Species Name	SPECIES_ID	Description
ROCP0SULF	3352	ROCP0SULF, Sulfur-containing Hydrocarbons, C* = $1e^{0}$ ug m^{-3}
ROCP1SULF	3353	ROCP1SULF, Sulfur-containing Hydrocarbons, C* = $1e^{1}$ ug m^{-3}
ROCP2SULF	3354	ROCP2SULF, Sulfur-containing Hydrocarbons, C* = $1e^{2}$ ug m^{-3}
ROCN2OXY	3361	ROCN2OXY, Oxygenated Organics, C* = $1e^{-2}$ µg m^{-3}
ROCN1OXY	3362	ROCN1OXY, Oxygenated Organics, C* = $1e^{-1}$ µg m^{-3}
ROCP0OXY	3363	ROCP0OXY, Oxygenated Organics, C* = $1e^{0}$ µg m^{-3}
ROCP1OXY	3364	ROCP1OXY, Oxygenated Organics, C* = $1e^{1}$ µg m^{-3}
ROCP2OXY	3365	ROCP2OXY, Oxygenated Organics, C* = $1e^{2}$ µg m^{-3}
ROCN2CYC	3372	ROCN2CYC, Cyclic Hydrocarbons, C* = $1e^{-2}$ µg m^{-3}
ROCN1CYC	3373	ROCN1CYC, Cyclic Hydrocarbons, C* = $1e^{-1}$ µg m^{-3}
ROCP0CYC	3374	ROCP0CYC, Cyclic Hydrocarbons, C* = $1e^{0}$ µg m^{-3}
ROCP1CYC	3375	ROCP1CYC, Cyclic Hydrocarbons, C* = $1e^{1}$ µg m^{-3}
ROCP2CYC	3376	ROCP2CYC, Cyclic Hydrocarbons, C* = $1e^{2}$ µg m^{-3}
ROCN2BRN	3383	ROCN2BRN, Branched Hydrocarbons, C* = $1e^{-2}$ µg m^{-3}
ROCN1BRN	3384	ROCN1BRN, Branched Hydrocarbons, C* = $1e^{-1}$ µg m^{-3}
ROCP0BRN	3385	ROCP0BRN, Branched Hydrocarbons, C* = $1e^{0}$ µg m^{-3}
ROCP1BRN	3386	ROCP1BRN, Branched Hydrocarbons, C* = $1e^{1}$ µg m^{-3}
ROCP2BRN	3387	ROCP2BRN, Branched Hydrocarbons, C* = $1e^{2}$ µg m^{-3}
ROCN2ALK	3394	ROCN2ALK, Linear Hydrocarbons, C* = $1e^{-2}$ µg m^{-3}
ROCN1ALK	3395	ROCN1ALK, Linear Hydrocarbons, C* = $1e^{-1}$ µg m^{-3}
ROCP0ALK	3396	ROCP0ALK, Linear Hydrocarbons, C* = $1e^{0}$ µg m^{-3}
ROCP1ALK	3397	ROCP1ALK, Linear Hydrocarbons, C* = $1e^{1}$ µg m^{-3}
ROCP2ALK	3398	ROCP2ALK, Linear Hydrocarbons, C* = $1e^{2}$ µg m^{-3}
--	666	Phosphorus
--	767	Vanadium
--	347	Chromium
--	379	Cobalt
--	612	Nickel
--	380	Copper
--	778	Zinc
--	468	Gallium
--	298	Arsenic
--	693	Selenium
--	689	Rubidium
--	397	Strontium
--	779	Zirconium
--	586	Molybdenum
--	649	Palladium
--	695	Silver
--	328	Cadmium
--	487	Indium
--	714	Tin
--	296	Antimony
--	300	Barium
--	519	Lanthanum

Species Name	SPECIES_ID	Description
--	1861	Cerium
--	528	Mercury
--	520	Lead

Table 24: "PM-Ready" species. These are the exclusive species that can be included in profiles whose PROFILE_TYPE is PM-AE6 and PM-CR1.
*** For calcium, if a profile does not have 2303 but does have 329, 329 should be used. For magnesium, if a profile does not have 2772 but does have 525, 525 should be used. For potassium, if a profile does not have 2302 but does have 669, 669 should be used. For sodium, if a profile does not have 785 but does have 696, 696 should be used. For chloride, if a profile does not have 337 but does have 795, 795 should be used.**

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

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

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

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

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

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

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

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

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

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

Table 25: Particulate water calculations.

Sources for which it is assumed that there is no particulate water include: Agricultural Burning, Bituminous Combustion, Calcium Carbide Furnace, Charbroiling, Charcoal Manufacturing, Distillate Oil Combustion, Electric Arc Furnace, Ferromanganese Furnace, Glass Furnace, HDDV Exhaust, Heat Treating, Kraft Recovery Furnace, LDDV Exhaust, Lignite Combustion, Lime Kiln, Meat Frying, Natural Gas Combustion,

Nonroad Gasoline Exhaust, Onroad Gasoline Exhaust, Open Hearth Furnace, Prescribed Burning, Process Gas Combustion, Pulp & Paper Mills, Residential Coal Combustion, Residential Natural Gas Combustion, Residential Wood Combustion, Residual Oil Combustion, Sintering Furnace, Slash Burning, Sludge Combustion, Solid Waste Combustion, Sub-Bituminous Combustion, Wildfires, and Wood Fired Boiler.

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

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

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

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

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

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

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

Unadjusted MO is computed as follows:

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

where Ox_i is the oxygen-to-metal ratio for metal *i* (see Table 26: **Oxygen-to-metal ratio for "PM-Ready" metals.**) and E_i is the weight percent of metal *i*, except for Na, Ca, Mg and K. For these metals, the E_i should reflect the difference between the atomic and ionic forms of the metal. If a profile has only one reported (i.e., only atomic *or* ionic) value for of Na, Ca, Mg, or K, then the E_i should be set to 0. In addition, if the difference is negative, it should be set to 0.

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

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

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

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

If $Non - Neutralized\ SO_4 < 0$,

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

If $Non - Neutralized\ SO_4 > 0$

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

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

Species	MW	Oxide Form 1	Oxide Form 2	Oxide Form 3	Oxygen/Metal Ratio
Na	22.99	Na_2O			0.348
Mg	24.31	MgO			0.658
Al	26.98	Al_2O_3			0.889
Si	28.09	SiO_2			1.139
P	30.97	P_2O_3	P_2O_5		1.033
K	39.10	K_2O			0.205
Ca	40.08	CaO			0.399
Ti	47.87	TiO_2			0.669
V	50.94	V_2O_5			0.785
Cr	52.00	Cr_2O_3	CrO_3		0.692
Mn	54.94	MnO	MnO_2	Mn_2O_7	0.631
Fe	55.85	FeO	Fe_2O_3		0.358
Co	58.93	CoO	Co_2O_3		0.339
Ni	58.69	NiO			0.273
Cu	63.55	CuO			0.252
Zn	65.39	ZnO			0.245
Ga	69.72	Ga_2O_3			0.344
As	74.92	As_2O_3	As_2O_5		0.427
Se	78.96	SeO	SeO_2	SeO_3	0.405
Rb	85.47	Rb_2O			0.094
Sr	87.62	SrO			0.183
Zr	91.22	ZrO_2			0.351
Mo	95.94	MoO_2	MoO_3		0.417
Pd	106.42	PdO	PdO_2		0.226
Ag	107.87	Ag_2O			0.074
Cd	112.41	CdO			0.142
In	114.82	In_2O_3			0.209
Sn	118.71	SnO	SnO_2		0.202
Sb	121.76	Sb_2O_3	Sb_2O_5		0.263
Ba	137.33	BaO			0.117
La	138.91	La_2O_3			0.173
Ce	140.12	Ce_2O_3	CeO_2		0.200
Hg	200.59	Hg_2O	HgO		0.060
Pb	207.20	PbO	PbO_2		0.116

Table 26: Oxygen-to-metal ratio for "PM-Ready" metals.

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

propagate to the ORGANIC_MATTER_to_ORGANIC_CARBON_RATIO field in SPECIATE. In addition, indicate in the NOTES field how PNCOM was computed.

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

Table 27: Source-specific, default OM-to-OC ratios.

Step 9: Sum all "PM-ready" species (see Table 24**: "PM-Ready" species. These are the exclusive species that can be included in profiles whose PROFILE_TYPE is PM-AE6 and PM-CR1.**). If summation is less than 100%, assign all remaining mass to PMOTHR (SPECIES_ID = 2671). If summation is 100%, no adjustments are necessary. If the summation is greater than 100%, then the following steps should be completed.

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

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

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