|  |
| --- |
| Prepared for:  Alison Eyth  Madeleine Strum  Benjamin Murphy  U.S. EPA  Under Contract to UNC  Prepared by:  Tejas Shah  Yuge Shi  Ross Beardsley  Greg Yarwood  Ramboll US Corporation  7250 Redwood Blvd., Suite 105  Novato, California 94945  June 2020  Month Year (or full date)  169000XXXX |

|  |
| --- |
| Intended for    Document type    Date |
| Speciation Tool User’s Guide Version 5.0  [Subject] |

|  |
| --- |
| Speciation Tool User’s Guide Version 5.0 |
|  |
|  |

|  |
| --- |
| Speciation Tool User’s Guide Version 5.0 |

|  |
| --- |
|  |

**ACKNOWLEDGEMENTS**

The authors of this report gratefully acknowledge Marc Houyoux, Madeleine Strum, Benjamin Murphy and Alison Eyth of the U.S. Environmental Protection Agency Office of Air Quality Planning and Standards for their contributions to this report.

|  |
| --- |
| Contents |

[List of Acronyms and Abbreviations 3](#_Toc44416458)

[1.0 Background 4](#_Toc44416459)

[1.1 Get the Latest Speciation Tool Version 5](#_Toc44416460)

[2.0 Technical Description 6](#_Toc44416461)

[2.1 Mechanisms Supported by Speciation Tool 6](#_Toc44416462)

[2.2 Coding Strategy 8](#_Toc44416463)

[3.0 Install and initialize the speciation tool 10](#_Toc44416464)

[3.1 Quick Start 10](#_Toc44416465)

[3.2 Install Speciation Tool 11](#_Toc44416466)

[3.3 Check for Required Software 14](#_Toc44416467)

[3.4 Initialize the Speciation Tool Database 15](#_Toc44416468)

[3.4.1 Update and Source the *Assigns.sptool* File 15](#_Toc44416469)

[3.4.2 Create the Speciation Tool Database 16](#_Toc44416470)

[3.4.3 Load the Speciation Tool *shared* Schema Data 17](#_Toc44416471)

[3.4.4 The Import Program 18](#_Toc44416472)

[4.0 Running the Speciation Tool 20](#_Toc44416473)

[4.1 The Run Script 20](#_Toc44416474)

[4.2 Run Name 20](#_Toc44416475)

[4.3 Run Control File 21](#_Toc44416476)

[4.3.1 Run Parameters 22](#_Toc44416477)

[4.3.2 Input Files 24](#_Toc44416478)

[4.3.3 Output Files 26](#_Toc44416479)

[4.4 Run the Speciation Tool 27](#_Toc44416480)

[4.4.1 Run-time Messages 27](#_Toc44416481)

[4.4.2 What happens when the script is running? 29](#_Toc44416482)

[5.0 Applications and Methodology 31](#_Toc44416483)

[5.1 Gas profile processing 31](#_Toc44416484)

[5.1.1 Run Type Options 31](#_Toc44416485)

[5.1.2 Example Use Cases for Gas Profiles 33](#_Toc44416486)

[5.1.3 Chemical Mechanisms 34](#_Toc44416487)

[5.2 PM2.5 Profile Processing 35](#_Toc44416488)

[5.2.1 Special Handling for CAMx Particulate Modeling 35](#_Toc44416489)

[5.2.2 Volatility Basis Set (VBS) 36](#_Toc44416490)

[6.0 FAQ 39](#_Toc44416491)

[6.1 How do I add a handful of speciation profile? 39](#_Toc44416492)

[6.2 Can I have more than one copy of the Speciation Tool database? 39](#_Toc44416493)

[6.3 How do I define a new chemical mechanism? 39](#_Toc44416494)

[6.3.1 Example of how to introduce a new mechanism, “SAPRC07TC\_ae7 with explicit benzene” 40](#_Toc44416495)

[6.4 What if I have a different INVTABLE? 40](#_Toc44416496)

[6.5 How do I add a new species to the database? 40](#_Toc44416497)

APPENDICES

Appendix A: PostgreSQL and Perl Installation Procedures

Appendix B: SHARED Schema File Formats

Appendix C: RUN Schema File Formats

Appendix D: PM2.5 Chemical Mechanism Definitions

Appendix E: Create AE6 Ready Profiles

Appendix F: Mechanism Mappings for CMAQ (AE7 and AE8) and CAMx (cf2)

Table of Figures

[Exhibit 1. Speciation Tool Initialization and Run schemas. 9](#_Toc44416498)

[Exhibit 2. Downloading Speciation Tool from GitHub. 11](#_Toc44416499)

[Exhibit 3. Speciation Tool Directory and File Names. 12](#_Toc44416500)

[Exhibit 4. Example of run-time messages. 28](#_Toc44416501)

Table of Tables

[Table 1. Mechanism mappings provided with Speciation Tool v5.0 7](#_Toc44416502)

[Table 2. Speciation Tool required data files. 13](#_Toc44416503)

[Table 3. Keyword list showing Speciation Tool shared data types. 19](#_Toc44416504)

[Table 4. Speciation Tool control file keywords. 21](#_Toc44416505)

[Table 5. Metadata keywords and example. 26](#_Toc44416506)

[Table 6. Program and module calls from the Speciation Tool program, run\_sptool.pl. 30](#_Toc44416507)

[Table 7. Examples of SPTOOL runs for different options for using HAPs 34](#_Toc44416508)

[Table 8. PM2.5 components from CMAQ to CAMx. 36](#_Toc44416509)

[Table 9. Example SVOC model species assignments for three profile codes 38](#_Toc44416510)

[Table 10. Example IVOC assignments for three profile codes 38](#_Toc44416511)

# List of Acronyms and Abbreviations

APIN Alpha-pinene

CAMx Comprehensive Air Quality Model with Extensions

CAP Criteria Air Pollutant

CB Carbon Bond

CEDS Community Emissions Data System

CH4 Methane

CMAQ Community Multiscale Air Quality

CMAS Community Modeling and Analysis System

CO Carbon Monoxide

CSV Comma-separated Values

CTM Chemical Transport Model

ETHA Ethane

EPA Environmental Protection Agency

HAP Hazardous Air Pollutants

NEI National Emissions Inventory

NMVOC Non-Methane VOC (VOC – methane)

NO Nitric Oxide

NO2 Nitrogen Dioxide

NOx Nitrogen Oxides

O3 Ozone

PEC Particulate Elemental Carbon

PM Particulate Matter

PM10 Particulate Matter with diameter smaller than 10 µm

PM2.5 Particulate Matter with diameter smaller than 2.5 µm

POC Particulate Organic Carbon

SAPRC Statewide Air Pollution Research Center

SMOKE Sparse Matrix Operator Kernel for Emissions

TOG Total Organic Gasses

TRI Toxic Release Inventory

VBS Volatility Basis Set

VOC Volatile Organic Compounds

# Background

Chemical Transport Models (CTMs) account for chemistry occurring in the atmosphere by using a chemical mechanism with multiple chemical reactions and species. For example, the following gas-phase chemical reactions involving nitric oxide (NO), nitrogen dioxide (NO2) and ozone (O3) are included in the chemical mechanisms used by CTMs:

NO2 → NO + O

O + O2 → O3

O3 + NO → NO2

Gas-phase chemical mechanisms that are widely used in CTMs include the Carbon Bond mechanisms (CB05, CB6) and Statewide Air Pollution Research Center mechanisms (SAPRC07, SAPRC11). Aerosol-phase chemistry schemes tend to be specific to individual CTMs. The Community Multiscale Air Quality Model (CMAQ) has aerosol schemes named AE6 and AE7 whereas the Comprehensive Air quality Model with extensions (CAMx) has aerosol schemes named CF2 and CMU.

Emission rates are an essential CTM input, and chemical species provided in the inputs must exactly match the model species of the CTM, although some CTMs like CMAQ allow for chemical mapping online. Emission input files are generated from data provided by emission inventories. However, emission inventories are prepared in terms of regulated pollutants such as carbon monoxide (CO), nitrogen oxides (NOx), volatile organic compounds (VOC), and particulate matter (PM). Some inventory pollutants exactly correspond to a single model species (e.g., CO) but most inventory pollutants correspond to several model species, e.g., inventory pollutant NOx corresponds to model species NO and NO2. Another potential complication is that chemical mechanisms tend to have different model species, especially for VOC, and so a CTM requires emission inputs with different model species for a simulation using the CB6 vs. SAPRC07 chemical mechanism.

The purpose of the Speciation Tool is to translate from emission inventory pollutants to CTM emission input species by:

* Creating “split factors” that allocate inventory pollutants (e.g., VOC, PM2.5) to model species (e.g., formaldehyde as part of VOC, elemental carbon as part of PM2.5)

Naming model species correctly (e.g., formaldehyde as HCHO or FORM) to be recognized by the CTM

The split factors output by the Speciation Tool are input data needed by emission processing software such as the Sparse-Matrix Operating Kernel for Emissions (SMOKE).

Generally, CTM emission inputs are created from emission inventories of criteria air pollutant (CAPS) such as the EPA’s National Emissions Inventory (NEI). However, the NEI also contains toxic air pollutants, also known as hazardous air pollutants (HAPS). It can be advantageous to combine information from CAPS and HAPS into a unified modeling emission inventory of toxic and other species. Taking formaldehyde as an example, processing the NEI VOC emission estimates for modelling will produce formaldehyde emissions estimates that could then be replaced by explicit estimates of formaldehyde emissions from the NEI. Implementing this strategy, named integration, requires coordinating the generation of split factors (by the Speciation Tool) with the emission processing (by SMOKE). The Speciation Tool supports the integration CAPS and HAPS emission estimates as an option.

## Get the Latest Speciation Tool Version

The Speciation Tool can be downloaded from a public GitHub repository <https://github.com/CMASCenter/Speciation-Tool>. The public repository stores source code, documentation, and tool inputs which enables version control and quick updates to the tool.

In the master branch the following folders/files are provided, and their descriptions are in parentheses:

* docs (folder contains documentation)
* import\_data (folder contains Speciation Tool input files)
* outputs (folder contains Speciation Tool outputs)
* run (folder contains run scripts)
* src (folder contains source code)
* Assigns.sptool (data file of environment variables described in 3.4.1)
* README.md (GitHub file – not a Speciation Tool file)
* \_config.yml (GitHub file – not a Speciation Tool file)
* import\_clean.csh (script described in Table 2)
* init\_sptooldb\_v5.0.csh (script described in Table 2)
* sptool\_reqd\_checks.sh (script described in Table 2)

# Technical Description

The Speciation Tool creates the chemical speciation input files used by the SMOKE emissions processor. Specifically, it creates the speciation profile file (GSPRO) and pollutant-to-pollutant conversion file (GSCNV). The GSPRO file defines conversion factors from inventory pollutant to model species for specific speciation profiles. The GSCNV file defines the conversion factor between VOC and total organic gasses (TOG) for each TOG speciation profile (TOG includes methane, ethane, acetone and other compounds that are excluded from VOC by EPA regulation). The Speciation Tool does not create the SMOKE speciation cross-reference file (GSREF); you must ensure that a consistent profile code labeling scheme is used in your GSPRO, GSCNV and GSREF files.

The fundamental equation for producing model species emissions from inventory pollutants is:

The Speciation Tool performs the summation shown in Equation 1. The summation contains a speciation profile (SP, e.g., the TOG composition of natural gas, or the PM composition of wood smoke) and the mechanism mapping (MM) must exist between every chemical constituent (*k*) contained in the speciation profile and all model species (*i*) contained in the chemical mechanism. The summation in Equation 1 is performed separately for each speciation profile to create a library of model species profiles that can be used to convert from emissions of inventory pollutants to emissions of model species.

The Speciation Tool is provided with speciation profile data from EPA’s SPECIATE database for both TOG and particulate matter less than 2.5 microns (PM2.5). SPECIATE uniquely identifies each chemical constituent of its profiles by number (a sequentially allocated integer, starting from 1) which in SPECIATE is called the SPECIES\_ID. Although SPECIATE also associates a chemical abstracts (CAS) code with many chemical constituents, CAS code does not constitute a unique identifier of SPECIATE profile constituents because some constituents can be, in addition to specific chemical compounds, broad groups that do not have a CAS code. The Speciation Tool identifies chemical constituents (of speciation profiles) by SPECIES\_ID and identifies model species (of chemical mechanisms) by name.

## Mechanisms Supported by Speciation Tool

The Speciation Tool repository includes mechanism mappings for gas- and aerosol-phase chemical mechanisms that are implemented in the CMAQ and CAMx CTMs (Table 1). Each mapping is named to identify both the gas-phase and aerosol mechanisms that it supports. Mappings for gas-phase mechanisms include the Carbon Bond mechanism (versions CB05, CB6), the Statewide Air Pollution Research Centre mechanism (versions SAPRC07 and SAPRC07T), the Common Representative Intermediates (CRI) and the Regional Atmospheric Chemistry Mechanism version 2 (RACM2) mechanisms. Model implementations of gas-phase mechanisms tend to be similar except for some differences in model species names and customizations to resolve hazardous air pollutants (HAPS). In contrast, model implementations for aerosols are highly specific to individual models. Aerosol mechanisms included in the Speciation Tool v5.0 are the AE6/AE7 and AE8 schemes of CMAQ v5.3 and the CAMx coarse-fine (CF2) scheme for CAMx versions 6.5 and newer. The AE7 mechanism is built on the AE6 and is identical in terms of model species and mechanism definition but requires that alpha-pinene (APIN) be separate from all other monoterpenes (TERP) and not included in TERP to avoid double counting.

Table 1. Mechanism mappings provided with Speciation Tool v5.0

| **Mechanism Mapping Name** | **Purpose** | **Comments** |
| --- | --- | --- |
| CB6r4\_cf2 | Support for CAMx v6.5 | Volatile organic species mapped for CAMx CB6r4 (also applicable for CB6r3 and CB6r2)  Semi-volatile organic species mapped to model species NVOL (c\* < 102.5 μg m-3) or IVOC (102.5 < c\* < 106.5 μg m-3)  Inorganic species mapped for CAMx CF aerosol scheme, version 2 (cf2) |
| CB6r3\_ae7 | Support for CMAQv5.3 | Volatile organic species mapped for CMAQ CB6r3 (with added model species NAPH and XYLMN) plus:  - added explicit alpha-pinene (APIN) model species  - acetic acid mapped to AACD  - formic acid mapped to FACD  Semi-volatile organic species mapped to IVOC and NVOL (as for CB6r4\_cf2)  SOAALK is added and comes from a separate tracer mechanism (see Table 4- “CB6R3\_AE7\_TRACER”)  Inorganic species mapped for CMAQ ae7 aerosol scheme |
| CB6r3\_ae8 | Support for AE8 in future CMAQ release | Volatile organic species mapped as CB6r3\_ae7 except model species XYLMN renamed to XYL  Semi-volatile organic species mapped to Volatility Basis Set (VBS) model species (SVOCN1…IVOCP6, IVOCP5ARO, etc.)  Inorganic species mapped for CMAQ ae8 scheme |
| SAPRC07\_cf2 | Support for CAMx v6.5 | Volatile organic species mapped for CAMx SAPRC07  Semi-volatile organic species mapped to IVOC and NVOL (as for CB6r4\_cf2)  Inorganic species mapped for CAMx cf2 scheme |
| SAPRC07TC\_ae7 | Support for CMAQv5.3 | Volatile organic species mapped for CMAQ SAPRC07TC with added explicit naphthalene (NAPH)  Semi-volatile organic species mapped to IVOC and NVOL  Inorganic species mapped for CMAQ ae7 scheme |
| SAPRC07TC\_ae8 | Prepare for ae8 in a future CMAQ release | Volatile organic species mapped as SAPRC07TC\_ae7  Semi-volatile organic species mapped to VBS model species (SVOCN1…IVOCP6, IVOCP5ARO, etc.)  Inorganic species mapped for CMAQ ae8 scheme |
| CRI\_ae7 | Support for CMAQv5.3 as-is | Volatile organic species mapped for CMAQ CRI with added explicit naphthalene (NAPH)  Semi-volatile organic species mapped to IVOC and NVOL  Inorganic species mapped for CMAQ ae7 scheme |
| CRI\_ae8 | Support for AE8 in future CMAQ release | Volatile organic species mapped as CRI\_ae7  Semi-volatile organic species mapped to VBS model species (SVOCN1…IVOCP6, IVOCP5ARO, etc.)  Inorganic species mapped for CMAQ ae8 scheme |
| RACM2\_ae7 | Support for CMAQv5.3 | Volatile organic species mapped for CMAQ RACM2 with added explicit naphthalene (NAPTHALENE) and methane (ECH4)  Semi-volatile organic species mapped to IVOC and NVOL  Inorganic species mapped for CMAQ ae7 scheme |
| RACM2\_ae8 | Support for AE8 in future CMAQ release | Volatile organic species mapped as RACM2\_ae7  Semi-volatile organic species mapped to VBS model species (SVOCN1…IVOCP6, IVOCP5ARO, etc.)  Inorganic species mapped for CMAQ ae8 scheme |
| CB05\_cf2 | Support for CAMx v6.5 | Volatile organic species mapped for CAMx CB05  Semi-volatile organic species mapped to IVOC and NVOL (as for CB6r4\_cf2)  Inorganic species mapped for CAMx cf2 scheme |

The Speciation Tool also produces factors to convert between mass of VOC and TOG; namely the VOC-to-TOG conversion factor (or TOG-to-VOC ratio as it is called in SPECIATE). These factors are needed because emission inventories often report VOC but speciation profiles are referenced to TOG. Since the organic gas speciation profiles are defined as compound fractions of TOG, it is straightforward for the Speciation Tool to compute the VOC-to-TOG conversion factor from a list of TOG compounds that are excluded from VOC, which is obtained from SPECIATE.

## Coding Strategy

The Speciation Tool is a PostgreSQL database application with a Perl script interface. The primary operating system to run the Speciation Tool is UNIX. It is designed for concurrent multiple user access where each run has a unique user specified name. The Speciation Tool assigns the run name to a database schema for storing temporary tables and results. As a database convention, a schema is essentially a container that stores tables and functions in a selected database. Think of a schema as a folder in the Speciation Tool database. The *shared* schema in the Speciation Tool database, created during initialization, holds the imported data of the mechanism definitions, profile descriptions, profile definitions, and species properties. It also retains the functions which compute the speciation factors. Before running the Speciation Tool, PostgresSQL, Perl, and Perl libraries must be installed. Speciation Tool scripts must be run to confirm the correct installation of these ancillary tools and to initialize the database. A successful initialization will create the database, load the tool functions, create the *shared* schema with defined tables, and import defaultdata files to the *shared* schema. The instructions for these tasks are provided in Chapter 3.

Chapter 4, Running the Speciation Tool, provides the details for running the model; run parameters and options are described. A Speciation Tool run creates a named schema in the Speciation Tool database. Refer to the Exhibit 1 below; initialization creates the *sptoolv5\_0* PostgreSQL database with schema *shared* and each run will create an additional named schema based on the user specs. In this example two runs have been made; cb05\_criteria and saprc\_toxics. Chapter 5, Applications and Methodology, provides additional details of the Speciation Tool calculations. Chapter 6, FAQ, addresses both basic and advanced questions on using the Speciation Tool model.

The Speciation Tool can create AE6 profiles from raw non-AE6-ready PM2.5 profiles in SPECIATE, assign species to intermediate-volatility organic compounds (IVOC) for Volatility Basis Set (VBS) schemes using vapor pressure information, and process VBS profiles in SPECIATE. Appendix E describes the steps for creating AE6-ready profiles in the Speciation Tool.

PostgreSQL Databases

Sptoolv5\_0

shared

cb05\_criteria run

saprc\_toxics run

Run Parameters

Optional input files:

Profile definitions

HAPS

Active/tracer list

Initialize and Load Data

Exhibit 1. Speciation Tool Initialization and Run schemas.

# Install and initialize the speciation tool

The Speciation Tool is written in PostgreSQL using the Perl Database Interface (DBI) and CSV parser (Text::CSV). These software products are open source and available on the internet. You must install PostgreSQL and Perl prior to running the Speciation Tool. Appendix A has instructions for installing these packages and options required. During the initialization step, the input data are cleaned to strip out any special or new line characters which can make PostgreSQL to choke. This is carried out with the script import\_clean.csh which runs automatically during the initialization step.

The Speciation Tool must be initialized prior to making any runs. Initialization steps include:

* Get the Speciation Tool package
* Extract files
* Set the home directory
  1. Run the initialization script Change to the run directory; cd /run
  2. Run *./test\_all.job*

Run the initialization script

Following is a detailed description of each of these steps. In addition, a quick step-by-step approach is provided for those already familiar with software setup and database applications.

## Quick Start

For a quick start follow these steps. The details of each of these steps are provided in the following sections.

1. Create a Speciation Tool directory
2. Download the repository and extract the Speciation Tool installation files.
   1. Change to the Speciation Tool directory.
   2. Copy the Speciation Tool package to this directory.
   3. Extract and uncompress the zipped file.
3. Verify required software packages are available. See Appendix A for software packages.
   1. Run *./spool\_reqd\_checks.sh*
4. Set environment variables.
   1. Edit *Assigns.sptool*; set SPTOOL\_HOME.
   2. source Assigns.sptool
5. Initialize the Speciation Tool database.
   1. Execute *./init\_sptooldb\_v5.0.csh*
6. Execute the provided test case.
   1. Change to the run directory; cd /run
   2. Run *./test\_all.job*
7. Review and compare results.
   1. Change to output directory; cd ../output
   2. Compare results to provided test case files.

## Install Speciation Tool

To install the Speciation Tool, create a Speciation Tool home directory from which to work. In the example below it is */sptool* but you can specify any valid directory name. Download the Speciation Tool package to the directory from GitHub (<https://github.com/CMASCenter/Speciation-Tool>) in a ZIP file format as shown in Exhibit 2 and extract the Speciation Tool files.

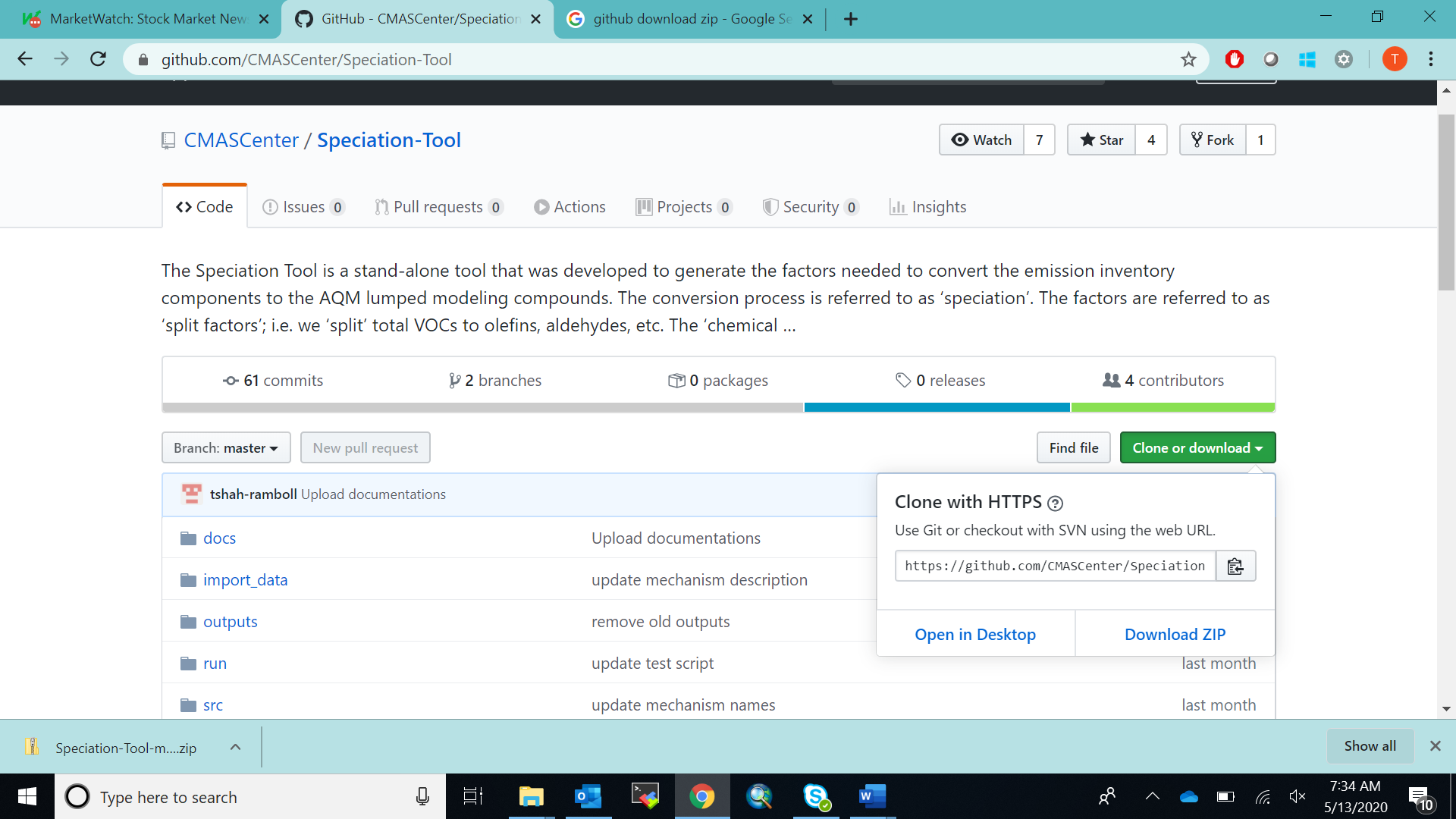


Exhibit 2. Downloading Speciation Tool from GitHub.

aqm\_poll\_rename.csv carbons\_all\_mechanisms\_speciate5\_0\_04mar2020\_v0.csv speciate5.0\_exports/

export\_profiles\_2019\_0826.txt

export\_species\_2019\_0826.txt

export\_species\_properties\_2020\_0313.txt

invtable\_hapcap\_cb05soa\_25may2011\_v13.txt

mech\_pm\_ae5\_ae6\_ae8.txt

sptool\_mechanism\_description\_speciate5\_0\_04mar2020\_v0.csv

mechanism\_forImport\_11Feb2020\_speciate5\_0\_withSOAALK\_13mar2020\_v0.csv

oxygen\_metal\_Ratios.csv

static\_19aug07.csv

ivoc\_species.csv

vbs\_svoc.profile.29aug2016.csv

vbs\_ivoc.profile.30aug2016.csv

gscnv.CB6r3\_ae7\_criteria.CMAQ.2020Feb10.txt

gscnv.CB6r4\_cf2\_criteria.CAMx.2020Feb10.txt

gscnv.CB05\_cf2\_criteria.CAMx.2020Feb10.txt

gscnv.CRI\_ae7\_criteria.CMAQ.2020Feb10.txt

gscnv.SAPRC07TC\_ae7\_criteria.CMAQ.2020Feb10.txt

gspro.CB6r3\_ae7\_criteria.CMAQ.2020Feb10.txt

gspro.CB6r4\_cf2\_criteria.CAMx.2020Feb10.txt

gspro.CB05\_cf2\_criteria.CAMx.2020Feb10.txt

gspro.CRI\_ae7\_criteria.CMAQ.2020Feb10.txt

gspro.SAPRC07TC\_ae7\_criteria.CMAQ.2020Feb10.txt

Assigns.sptool

init\_sptooldb\_v5.0.csh

sptool\_reqd\_checks.sh

import\_clean.csh

/import\_data

test\_gas\_new.job

test\_pm.job

drop\_table.sql

import\_control.pl

make\_pm\_splits.sql

make\_splits.sql

mk\_process\_pol.pl

prep\_out.sql

run\_sptool.pl

table\_defs.sql

table\_inps.sql

write\_outputs.pl

/sptool

/outputs

/run

/src

Exhibit 3. Speciation Tool Directory and File Names.

The data files available in the GitHub are listed in Table 2 along with their brief descriptions. The files and their purposes are described in the subsequent sections.

Table 2. Speciation Tool required data files.

| **Folder** | **Files** | **Description** |
| --- | --- | --- |
| /sptool | Assigns.sptool | Sets environment variables that are required for initialization |
|  | init\_sptooldb\_v5.0.csh | Initialization script that checks for required environment variables, creates the database, sets permissions, creates shared schema and table definitions, and imports data |
|  | sptool\_reqd\_checks.sh | Script to check the required software are installed properly |
|  | import\_clean.csh | Script to strip out any special or new line characters in the input data |
| /import\_data | speciate5.0\_exports | Folder containing SPECIATE inputs |
|  | aqm\_poll\_rename.csv | Cross reference file to rename model species names |
|  | camx\_fcrs.profile.lst | Provides list of profiles where FPRM is renamed to FCRS |
|  | carbons\_all\_mechanisms\_speciate5\_0\_04mar2020\_v0.csv | Provides number of carbons for model species |
|  | invtable\_hapcap\_cb05soa\_25may2011\_v13.txt | SMOKE Inventory Table |
|  | ivoc\_species.csv | Provides molecular weights for IVOC compounds |
|  | mech\_pm\_ae5\_ae6\_ae8.txt | File containing PM2.5 mechanism definition |
|  | sptool\_mechanism\_description\_speciate5\_0\_04mar2020\_v0.csv | File with description and data source for each chemical mechanism |
|  | mechanism\_forImport\_11Feb2020\_speciate5\_0\_withSOAALK\_13mar2020\_v0.csv | File containing gas chemical mechanism definition |
|  | oxygen\_metal\_Ratios.csv | Oxygen-to-Metal Ratio for metal species. Used to calculate metal bound oxygen for creating AE6-ready profiles |
|  | static\_19aug07.csv | Static profiles, profile weights do not change |
|  | vbs\_ivoc.profile.30aug2016.csv | IVOC factors by profile for VBS modeling |
|  | vbs\_svoc.profile.26sep2016.csv | SVOC factors by profile for VBS modeling |
| /output | gscnv.CB6r3\_ae7\_criteria.CMAQ.2020Feb10.txt | Sample SMOKE GSCNV output file. Provides VOC-to-TOG conversion factor |
|  | gspro.CB6r3\_ae7\_criteria.CMAQ.2020Feb10.txt | Sample SMOKE GSPRO speciation profile file |
| /run | test\_gas\_new.job | Creates run control file and executes Speciation Tool run script for gas mechanisms |
|  | test\_pm.job | Creates run control file and executes Speciation Tool run script for PM2.5 mechanisms |
| /src | drop\_table.sql | Executes PostgreSQL commands to drop database tables |
|  | import\_control.pl | Script to import run control |
|  | import\_rawdata.pl | Script to import shared data files |
|  | make\_pm\_splits.sql | Routine to generate and fill the temporary tables used to determine the PM2.5 split factors. |
|  | make\_splits.sql | Routine to generate and fill the temporary tables used to determine the split factors. |
|  | mk\_process\_pol.pl | Script to create Motor Vehicle (MV) process modes |
|  | prep\_out.sql | Executes PostgreSQL commands to populate tables for output |
|  | run\_sptool.pl | Main control script to run Speciation Tool |
|  | table\_defs.sql | Executes PostgreSQL commands to create shared database tables |
|  | table\_inps.sql | Executes PostgreSQL commands to create tables for scenario inputs |
|  | write\_outputs.pl | Script to write GSPRO and GSCNV output files |

## Check for Required Software

Before you begin the initialization, first verify that PostgreSQL and Perl are installed and accessible. Whoever is installing the Speciation Tool needs PostgreSQL permission to create a database. Execute the script *sptool\_reqd\_checks.sh* to check that the required software can be found and to verify user PostgreSQL authorizations.

> ./sptool\_reqd\_checks.sh

Following is an example of a successful check for the required software:

===== Speciation Tool Requirements Check =====

Checking the status of software requirements...

Status of required software:

[x] = Installed

[ ] = Not installed

[?] = Unable to determine, see notes

------PERL------

[x] Perl

[x] -DBI

[x] -DBD-Pg

[x] -Text-CSV

---POSTGRESQL---

[x] PostgreSQL

[x] -PL/pgSQL

Refer to the Speciation Tool User Guide Appendix A for installation procedures of the required software.

## Initialize the Speciation Tool Database

Prior to making any Speciation Tool runs the Speciation Tool must be initialized. First, set the required environment variables and then start the initialization.

### Update and Source the *Assigns.sptool* File

The *Assigns.sptool* file provided in the top level Speciation Tool home directory sets environment variables that are required for initialization. These include the Speciation Tool home directory, database name, and input file names of the default speciation data. The only environment variable that must be updated is SPTOOL\_HOME. This is installation dependent and must be set based on your system. Change the path name to correspond to the directory where the Speciation Tool top directory resides on your system. For example, if you install the Speciation Tool under the directory /disk4/models/emis, you would set the SPTOOL\_HOME environment variable like this:

setenv SPTOOL\_HOME /disk4/models/emis/sptool # Speciation Tool top level directory

The other environment variable that you might want to change is the Speciation Tool database name. The file provided in the distribution package has this variable set to sptoolv5\_0:

setenv SPTOOL\_DB sptoolv5\_0 # Speciation Tool Database name

You may change “sptoolv5\_0” to any valid PostgreSQL database name; just be sure it doesn’t currently exist. The command line ‘psql –l’ (dash lower case L) will display a list of existing PostgreSQL databases. Restrict the database name characters to a – z, 0 – 9, and the underscore “\_” with no imbedded blank characters.

When the *Assigns.sptool* file has been updated then ‘source’ the file to set the required environment variables.

> source Assigns.sptool

### Create the Speciation Tool Database

The initialization script, *init\_sptooldb\_v5.0.csh*, is a C-shell script that checks for required environment variables, executes a set of PostgreSQL commands to create the Speciation Tool database, sets database permissions, creates the *shared* schema and table definitions for the tool data, imports Speciation Tool SQL functions and data to the *shared* schema. The functions imported by the script perform the computations needed for generating the speciation profiles and pollutant-to-pollutant factors that are output by the program.

>./init\_sptooldb\_v5.0.csh

Messages are printed to the screen to reflect installation progress. The first part of a successful initialization will look similar to:

>./init\_sptooldb\_v5.0.csh

SPTOOL\_SRC\_HOME = /disk4/models/emis/sptool/src

New database: SPTOOL\_DB = sptoolv5\_0

SPTOOL\_USER = yourusername

POSTGRES\_BIN = /usr/local/pgsql/bin

CREATE DATABASE

Database sptoolv5\_0 created

CREATE SCHEMA

Shared schema created

GRANT

Create permissions granted on sptoolv5\_0

GRANT

All permissions granted on shared schema

Defining custom functions and initializing tables. ...working

Speciation Tool functions and tables successfully defined in sptoolv5\_0.

If you forget to set the environment variables you will get error messages as displayed below. Review the previous section on the *Assigns.sptool*, source the Assisgns file, and then rerun the initialization script.

SCRIPT ERROR: Required environment variable SPTOOL\_SRC\_HOME not set

in script init\_sptooldb\_v5.0.csh

SCRIPT ERROR: Required environment variable SPTOOL\_DB not set

in script init\_sptooldb\_v5.0.csh

ABORT: init\_sptooldb\_v5.0.csh script aborted with errors.

If you attempt to initialize the Speciation Tool with a database name that already exists, you will get the following message:

createdb: database creation failed: ERROR: database "sptoolv5\_0" already exists

ERROR: failed to create a new database sptoolv4\_0. This usually means the database already exists.

To replace the existing database type 'dropdb sptoolv5\_0' from the command line.

Delete the existing database if you want it replaced. Otherwise change the SPTOOL\_DB name in the Assigns.sptool file and provide a different database name. The command to delete the existing PostgreSQL database is:

> dropdb $SPTOOL\_DB

Once you either drop the existing database or change the assigned name you can start the initialization again.

### Load the Speciation Tool *shared* Schema Data

The final step in initializing the Speciation Tool is to load the data that is provided in the distribution package. This information includes species properties, chemical mechanism definitions, speciation profile definitions from SPECIATE 5.0 (i.e., the profile codes with species weight percents), and some additional general information that is shared by all Speciation Tool runs. The data are imported to the Speciation Tool database during initialization and stored in tables in the Speciation Tool PostgreSQL database *shared* schema.

Previously, importing relevant data from the SPECIATE database had been problematic due to special characters and new line character issues related to how the data are populated into Microsoft Access® database. To work around this issue, a macro in Access® was developed to get rid of special characters before importing data into the PostgreSQL; however, this was an iterative process as the universe of special characters changes whenever a new version of SPECIATE is released. In the Speciation Tool v5, a process is developed to translate the data tables from the SPECIATE Access® database into Speciation Tool inputs to eliminate the need for any manual data manipulation such as stripping out special characters and/or other manual steps previously needed. The process consists of export queries in the SPECIATE database to extract relevant data in the correct format and the shell script *import\_clean.csh* to strip out any special or new line characters which is executed during the initialization.

A database contains one or more named schemas, which in turn contain tables and other objects. The same table name can be used in different schemas without conflict; for example, in Exhibit 1 both cb05\_criteria and saprc\_toxics schemas contain the same table names. Unlike PostgreSQL databases, schemas are not rigidly separated; a user can access tables in any of the schemas within the same database. The Speciation Tool *shared* schema is accessed by all Speciation Tool runs; the tables are shared.

The input data files reside in the /import\_data subdirectory of the Speciation Tool home directory. The file names are assigned to environment variables in the *Assigns.sptool* file and do not need to be changed.

Messages are displayed during the initialization process (i.e., when you run “./init\_sptooldb\_v5.0.csh”) to indicate progress while importing the shared data. Below is a partial sample of a successful load. The time it takes to import the data will vary depending on your system, a few minutes is typical.

Importing mechanisms /disk4/models/emis/sptool/src/../import\_data/mechanism\_mar2013\_forimport.txt

Using transactions - import of data will abort on any error.

Reading data from /disk4/models/emis/sptool/src/../import\_data/mechanism\_mar2013\_forimport.txt...

1000 lines processed

2000 lines processed

…

18000 lines processed

...finished, imported 18981 lines.

…

Importing species properties /disk4/models/emis/sptool/src/../import\_data/export\_species\_properties\_2013\_0311.csv

Using transactions - import of data will abort on any error.

Reading data from /disk43/sptool/wa1\_05.2012/src/../import\_data/export\_species\_properties\_2013\_0311.csv...

1000 lines processed

2000 lines processed

...finished, imported 2274 lines.

Speciation Tool shared data successfully imported

Completed: Wed Sep 20 11:50:30 PDT 2016

The Speciation Tool uses the *shared* schema functions and tables in every run. Appendix B provides the table definitions that are stored in the *shared* schema.

If an error occurs while importing the data, the program will terminate with a message indicating which file and record caused the error. An example of an error is if an input file is not in the right format. A database rollback will be initiated and none of the data for the file with the error will be imported. The import stops at the first error.

### The Import Program

The initialization script executes the Perl program *import\_rawdata.pl* for each file imported to the *shared* schema. Each database table in the *shared* schema is associated with a keyword listed in Table 2 which identifies the table to populate. To import additional data to the *shared* schema tables (e.g., if you have additional profiles from those that are in SPECIATE 5.0) you can execute the Perl program from the command line. This will essentially append data to an existing table.

> perl $SPTOOL\_SRC\_HOME/import\_rawdata.pl $SPTOOL\_DB <table\_type> <input\_file>

The parameters include the database name, a keyword from Table 3, and the path/file name of the inputs records. The input data must conform to the table definitions in Appendix B. Any error will abort the import with a database rollback; no records will be appended to the table.

Table 3. Keyword list showing Speciation Tool shared data types.

|  |  |
| --- | --- |
| **Keyword** | **Description** |
| Mechanism | mechanism definition |
| mechanismPM | PM2.5 mechanism definition |
| mechanism\_description | description and data source for each chemical mechanism |
| invtable | SMOKE Inventory Table |
| gas\_profiles | gas profile description and historical information |
| gas\_profile\_weights | gas weight profiles |
| pm\_profiles | PM profile description and historical information |
| pm\_profile\_weights | PM weight profiles |
| rename\_species | rename model species names |
| species | species data: ids, names, MW, etc |
| carbons | number of carbons for model species |
| static | static profiles, profile weights do not change |
| camx\_fcrs | list of profiles where FPRM is renamed to FCRS |
| vbs\_svoc\_factors | SVOC factors by profile for VBS modeling |
| vbs\_ivoc\_factors | IVOC factors by profile for VBS modeling |
| ivoc\_species | molecular weights for IVOC compounds |

After a successful Speciation Tool initialization, the PosgreSQL *sptoolv5\_0* database will be created with the *shared* schema that holds the tool functions and default data.

# Running the Speciation Tool

This chapter discusses the mechanics of running the Speciation Tool, which include how to run the tool and the parameters and options available.

The command to run the Speciation Tool is:

perl $SPTOOL\_SRC\_HOME/run\_sptool.pl <database> <run\_name> <run\_control\_file>

The command line parameters include the database name, user specified run name, and a control file. The *run\_sptool.pl* program creates the run schema, reads and imports the specified control file, imports run-specific data files, executes the PostgreSQL functions to compute the split factors and conversion factors, and writes the output files.

Prior to running the Speciation Tool:

* Source the Assigns file (see 3.4.1)
* Create a Run Control file (see 4.3)
* Develop the input files that are required for the run (see 4.3.2)

The steps to run the Speciation Tool are explained in detail below.

## The Run Script

The Perl program *run\_sptool.pl* creates the output speciation profile files GSPRO and GSCNV. The program requires the source code directory and database name. Source the *Assigns.sptool* file to set the required environment variables SPTOOL\_HOME and SPTOOL\_DB. This file is discussed in the previous chapter on initializing the Speciation Tool.

If you receive an error “file not found” when running the script, it is probably because you have not set the Speciation Tool home directory or that the path is set to an invalid path name.

## Run Name

The run name is an identifier to uniquely label the Speciation Tool run. It is used to create a schema of that name in the Speciation Tool database. To create a new schema, specify a run name that has not previously been used. If you specify an existing run name then the existing schema is dropped and replaced by the new run.

The run name must meet the following conditions:

* Begin with an alphabetic character
* Must **not** include spaces in the name
* Must **not** be the *‘shared’*  schema
* Should be different from all existing run identifiers in the Speciation Tool, unless an existing run is meant to be replaced
* Must **not** be ‘public’ or begin with ‘pg\_’ which are reserved for system schemas

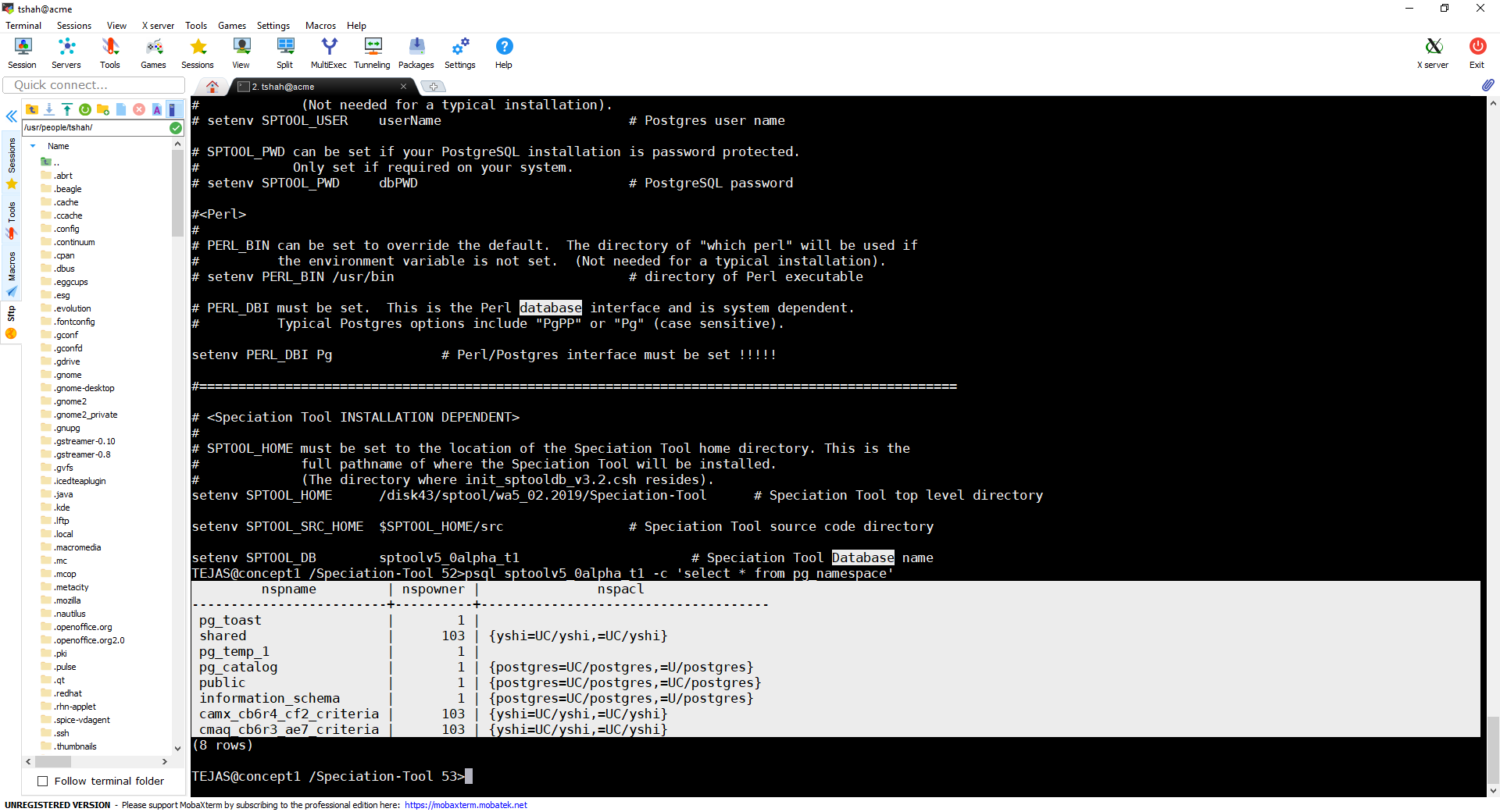
Must be less than 64 characters in length

The command to list existing schema (runs) in the database is:

psql <database\_name> -c 'select \* from pg\_namespace'

where <database\_name> is the Speciation Tool database name. The first column of the resulting list, labeled nspname, will include run names as well as system schema names. The last column, labeled nspacl, lists the owner of each schema.

>psql sptoolv5\_0 -c 'select \* from pg\_namespace'



## Run Control File

The run control file indicates the run parameters and the run-specific data files required for a Speciation Tool run. The format of the run control file is

<keyword>, <option>

Only lines that begin with a keyword are recognized by the database. All other lines are skipped allowing imbedded comments within the control file. Table 4 summarizes the control file keywords and options. Details for each keyword are provided in following sections.

Table 4. Speciation Tool control file keywords.

| **Keyword** | **Description** |
| --- | --- |
| **Run Parameters** | |
| MECH\_BASIS | Mechanism identifier/name.  Matches identifier in the shared schema mechanism table. Options provided with Speciation Tool include:  CB05\_CF2  CB6R3\_AE7  CB6R3\_AE7\_TRACER  CB6R3\_AE8  CB6R4\_CF2  CRI\_AE7  CRI\_AE8  RACM2\_AE7  RACM2\_AE8  SAPRC07\_CF2  SAPRC07TC\_AE7  SAPRC07TC\_AE8  AE6  AE8 |
| OUTPUT | Output options (default is VOC):  VOC  PM  STATIC |
| RUN\_TYPE | Specifies how split factors are developed. Options include:  CRITERIA  INTEGRATE  NOINTEGRATE  HAPLIST  VBS |
| AQM | Air quality model. Determines some of the pollutant names for the output files. Options:  CAMX  CMAQ |
| TOLERANCE | The acceptable deviation for which the sum of weight percentages for a profile definition is within 100%. The default is 5%. Applies only to gas profiles. |
| **Optional Input Files** | |
| TOX\_FILE | Toxic species path/file name. |
| PRIMARY\_FILE | Additional toxic entries path/file name. |
| PROC\_FILE | Mobile process modes path/file name. |
| PRO\_FILE | User-defined input profiles path/file name. |
| **Output Files** | |
| SPLITS\_OUT | Output speciation profiles (GSPRO) path/file name. |
| CNV\_OUT | Output conversion factors (GSCNV) path/file name. |

### Run Parameters

|  |  |  |
| --- | --- | --- |
| **MECH\_BASIS** | Example: | MECH\_BASIS, CB05\_cf2 |

This parameter identifies the mechanism mapping (see Table 1). The specified mechanism must match a mechanism name in one of the mechanism tables in the *shared* schema which are imported when the Speciation Tool is initialized. The data provided with the Speciation Tool include VOC mechanism definitions listed in Table 1 and PM2.5 mechanism definitions for AE6/AE7 (which has MECH\_BASIS = AE6) and AE8. The names of mechanism are revised to more transparently indicate their purpose by including both the chemical mechanism name and the aerosol option in the mapping title (see Appendix F). For example, the CB6 mappings for CMAQ AE7 and AE8 are titled CB6R3\_AE7 and CB6R3\_AE8, respectively, and the CB6 mapping for the CAMx two-mode coarse/fine aerosol option (CF2) is titled ‘CB6R4\_CF2’. CB6 mappings are named for the most recent CB6 revision present in the target model (R3 in CMAQ and R4 in CAMx). The \_TRACER mechanism (CB6R3\_AE7\_TRACER) is created for CMAQ to produce three additional species that are added to the base mechanism: ALD2\_PRIMARY, FORM\_PRIMARY, and SOAALK (see section 5.1).

|  |  |  |
| --- | --- | --- |
| **OUTPUT** | Example: | OUTPUT, VOC |

The OUTPUT keyword identifies which output profiles will be generated. The output options are “VOC” gas species, “PM” particulates, and “STATIC”. Only one option can be specified per run. STATIC represents the list of splits that do not change with profile weight definitions, such as the pollutants CO and NH3. It simply formats the imported static profiles to correspond to the output GSPRO file format.

The OUTPUT keyword is optional. If it is omitted from the run control file the default is to generate VOC factors.

|  |  |  |
| --- | --- | --- |
| **RUN\_TYPE** | Example: | RUN\_TYPE, CRITERIA |

The RUN\_TYPE options are: CRITERIA, INTEGRATE, NOINTEGRATE, HAPLIST, and VBS.

CRITERIA: All model species, including toxic species, are computed from criteria emissions.

INTEGRATE: Specific HAPs species are to be integrated from the HAP emissions in the emission inventory. Part of the criteria VOC mass is replaced with the specified HAP VOC mass. This involves subtracting specified HAP VOC emissions from the criteria VOC emissions to avoid double counting of VOC. Profiles are generated by removing the specified HAP species from the profile, and renormalizing. The resultant profile is NONHAPTOG.

NOINTEGRATE: The HAPs species from the inventory are not integrated. The criteria VOC mass retains the HAP VOC mass, however the speciation profile is adjusted to remove the specified HAPs. A user may choose the NOINTEGRATE approach when the HAPs in the inventory do not have the necessary details to provide a one-to-one match (e.g., they may be reported at a different unit or process than the VOC). The **active** HAPs (specified as explicit in the INVTABLE) are removed from the VOC profiles but the profiles are not renormalized, mass is not preserved, and profiles are generated as TOG.

HAPLIST: Generates records for the specified HAPs that define the SMOKE calculation of NONHAPTOG. This is needed whenever RUN\_TYPE is INTEGRATE or NOINTEGRATE.

VBS: The Volatility Basis Set (VBS) option generates semi-volatile organic compounds (SVOC) when generating PM2.5 profiles or intermediate-volatility organic compounds (IVOC) when generating VOC profiles. E.g., IVOC VBS species for CMAQ are: IVOC\_G, IVOC\_D, IVOC\_F and the SVOC VBS species for CMAQ are: P\_PVBi, P\_FVBi, P\_CVBi. This option generates a complete VOC and PM profiles depending on the choice of mechanism along with IVOC and SVOC compounds. Refer to Section 5.2.2 for details.

The different run types support different simulation options in SMOKE. For emissions processing of only VOCs (i.e., no HAP use) the CRITERIA option is used. The INTEGRATE or NOINTEGRATE options are used if the user would like to use the HAPs with the VOC in the SMOKE modeling. The INTEGRATE case will generate NONHAPVOC profiles in which the toxics mass is removed from the criteria VOC mass. The NOINTEGRATE case is selected when sources of the VOC and HAPs cannot be definitively matched. The NOINTEGRATE case assumes that the criteria VOC mass includes toxics mass, and therefore the toxics that are explicit in the chemical mechanism need to be removed from the TOG profiles so as not to double count these toxics in the inventory. Refer to the SMOKE User’s Manual (http://www.smoke-model.org) for more detailed information of how SMOKE handles integrating criteria VOC and toxics emissions.

Note: The Speciation Tool *shared* tables includes a default INVTABLE file. These data should be reviewed and potentially replaced with the INVTABLE that you are using for SMOKE modeling.

The only run types currently supported for PM2.5 output are CRITERIA and VBS.

|  |  |  |
| --- | --- | --- |
| **AQM** | Example: | AQM, CAMX |

The Air Quality Model (AQM) is used to determine the appropriate model species names in the output files. The Speciation Tool initialization data files support the AQM options CMAQ and CAMX. CMAQ stands for the Community Multiscale Air Quality Model, and the CAMX option stands for the Comprehensive Air Quality Model with extensions.

Models can use different names for the same mechanism species. For example, ‘other’ PM2.5 emissions are mapped to the model species name “PMOTHR” in CMAQ and “FPRM” in CAMx. Tables tbl\_static and tbl\_rename\_species in the shared schema include fields that identify model species names. If another model uses different model species name, then those cases need to be included in these tables. The shared schema table tbl\_rename\_species allows you to export model-specific species names different from the ones specified in the mechanism table. For example, ethene in the SAPRC mechanisms is stored as “ETHE” in the mechanism table and is mapped to the model species name “ETHENE” for CMAQ. An optional input called SPECIES\_RENAME needs to be set to specify a species rename mapping file or you can execute the Perl program from the command line described in section 3.4.4 to import the file and load it into tbl\_rename\_species.

|  |  |  |
| --- | --- | --- |
| **TOLERANCE** | Example: | TOLERANCE, 3 |

Each profile is defined as the sum of its components and is quantified as the percent contribution of each component to total organic gases. Ideally the sum of the percent components for each profile should equal 100 percent, which means that all mass is accounted for and assigned to individual species. However, due to round-off or inaccurate profile definitions the sum of the profile components do not always add to 100%. The optional TOLERANCE keyword defines an acceptable deviation from 100%, with the default hard-coded in the Speciation Tool at 5%. Any profile definition where the sum of the percent is outside the acceptable tolerance will **not** be output in the run. In other words, all profiles that sum to less than 95% or greater than 105% are excluded from the output and a warning is printed (see section 4.4.1). This keyword only applies to VOC profiles.

If the sum of the weight percentages for a particular profile is not 100% and nothing is done to correct the profile definition, then the resulting speciation profile will drop or add mass if used in emissions modeling. The Speciation Tool renormalizes all gas profiles whose sum is within the tolerance thereby preserving VOC mass. Profiles outside of the tolerance are dropped at run time. These profiles are not included in the model outputs but they do continue to reside in the *shared* schema tables.

### Input Files

File formats for all of the run specific input files are provided in Appendix C.

|  |  |  |
| --- | --- | --- |
| **PRO\_FILE** | Example: | PRO\_FILE, ctl/prfwts\_02coalstudy.dat |

The PRO\_FILE keyword is used to specify a file of gas profiles. Specifying this option will override using the gas profiles defined in the *shared* schema. The profiles provided in this input file are the profiles for which the model will generate the splits and conversion factors. PRO\_FILE is an optional keyword in the control file. The default is to use the profile definitions in the *shared* schema. At this time, only one set of profiles is used per run; either the *shared* data or the data provided using the PRO\_FILE keyword. If speciation profiles are required from both sources, then separate runs are required.

|  |  |  |
| --- | --- | --- |
| **TOX\_FILE** | Example: | TOX\_FILE, ctl/toxics\_coalstudy.dat |

The toxics file is required for run types INTEGRATE, NOINTEGRATE, and HAPLIST. The file lists the toxic compounds that are to be used from the inventory. The Speciation Tool uses the information in the Inventory Table, imported to the *shared* schema during initialization, to determine whether the toxics are to be treated as active or tracer compounds. An active compound is defined with ‘Y’ in the INVTABLE *explicit* field. Active species are included in the chemical mechanism with chemical feedback, while tracer species are included in the mechanism with no chemical feedback. Tracer toxics species mass is double counted.

For the INTEGRATE case (i.e., RUN\_TYPE, INTEGRATE) the active and tracer toxic species are removed from the VOC profiles, whereas for the NOINTEGRATE case, only active toxic species are removed from the VOC profiles. The toxic species are removed based on SPECIES\_ID provided in the TOX\_FILE.

For the case where HAPLIST is the specified RUN\_TYPE, the active HAPs produce a single one-to-one mapping from HAPs inventory pollutant to model compound. The tracer HAPs generate the same mapping as the active, but in addition, the HAPs inventory pollutant is mapped to the original VOC profile compounds to support the INTEGRATE case. The HAPLIST case also generates a ‘nointegrate’ HAP by appending the suffix “\_NOI” to the model compound (for example, BENZENE\_NOI).

The toxics file is only applicable to VOC output.

|  |  |  |
| --- | --- | --- |
| **PRIMARY\_FILE** | Example: | PRIMARY\_FILE, ctl/primary\_toxics\_coalstudy.dat |

You can add or override toxic entries by providing a primary toxic list. This file is only used when the run type is HAPLIST. If it is specified for any other run type the data are ignored.

The overwrite flag carried in this table determines whether a toxic species should be added (“N”) or replaced (“Y”) in the GSPRO file. This capability allows the Speciation Tool to support the one-to-many toxics species mapping and primary toxic profiles. The file format and example are as follows:

<Inv poll> <AQM name> <splitfac> <overwrite flag>

FORMALD, FORM\_PRIMARY, 1.0, N

ACETALD, ALD2\_PRIMARY, 1.0, N

XYLS, MXYL, 0.52, Y

XYLS, OXYL, 0.16, Y

XYLS, PXYL, 0.16, Y

The overwrite flag in the last column determines whether a toxic species should be added (“N”) or replaced (“Y”) in the GSPRO file. XYLS (species id = 507) in the input toxic table has been assigned to DONT\_USE (any place holder would suffice). With the overwrite flag turned on, the Speciation Tool will remove the “XYLS DONT\_USE” entry and add the three XYLS entries shown above. The FORM\_PRIMARY and ALD2\_PRIMARY will be added to the GSPRO.

|  |  |  |
| --- | --- | --- |
| **PROC\_FILE** | Example: | PROC\_FILE, process\_mode.dat |

The PROC\_FILE keyword specifies an optional input file which provides mobile source emission modes for profiles that represent mobile processing. This feature is provided to support SMOKE requirements of mobile source emission modes. An example record in this file is “4674, EXH” where profile code 4674 applies to exhaust emissions. When the process file is provided then additional records are generated in the GSCNV output file. For example, profile 4674 would include a VOC to TOG record as well as an EXH\_VOC to EXH\_TOG record; with the same conversion factors.

### Output Files

|  |  |  |
| --- | --- | --- |
| **SPLITS\_OUT** | Example: | SPLITS\_OUT, outputs/gspro\_02coalstudy.dat |

The SPLITS\_OUT keyword is used to optionally specify a path and file name for the model splits results written in SMOKE GSPRO format. Paths can be relative paths from the Speciation Tool main directory (as shown) or absolute paths. This is an optional keyword. If omitted the output file name is derived from the run control parameter specifications and the run date and is written to the relative path /outputs. An example of a default output filename is outputs/gspro\_CB05\_NOINTEGRATE\_CAMX\_26Jun2011.

|  |  |  |
| --- | --- | --- |
| **CNV \_OUT** | Example: | CNV\_OUT, outputs/gscnv\_02coalstudy.dat |

The CNV\_OUT keyword is used to optionally specify a path and file name for the model TOG/VOC conversion factors written in SMOKE GSCNV format. Paths can be relative paths from the Speciation Tool main directory (as shown) or absolute paths. This is an optional keyword. If omitted the output file name is derived from the run control parameter specifications and the run date and is written to the relative path /outputs. An example of a default output filename is outputs/gscnv\_CB05\_NOINTEGRATE\_CAMX\_26Jun2011.

|  |  |  |
| --- | --- | --- |
| **IVOC\_BIN\_OUT** | Example: | IVOC\_BIN\_OUT, TRUE |

The IVOC\_BIN\_OUT keyword is used to specify whether to produce an optional intermediate chemical mechanism file for VBS scheme that contains mapping of IVOC compounds to VBS model species. The intermediate provides IVOC binning for each species based on vapor pressure information in SPECIATE. The file includes speciate id, species name, and assigned VBS model species for each IVOC compound. This file is a first step in developing a complete chemical mechanism assignment for VBS scheme in AQM and subsequently requires offline review to develop a complete mapping file for each chemical mechanism. For example, naphthalene is explicit in CMAQ, and it will not be mapped based on its vapor pressure when using CMAQ and therefore needs to be changed in this intermediate file.

#### Header Records: Metadata

The Speciation Tool output files, GSPRO and GSCNV, contain header records which are referred to as ‘metadata’. Each of the output files has the same common metadata keywords. The purpose of the metadata records is to provide a summary of the data and parameters that were used to generate the outputs. Table 5 provides the list of metadata keywords for an example run. The below example is for a VOC GSPRO with no HAP use. Note that an INTEGRATE GSPRO would include a list of toxics that are integrated in the form of #NHAP headers.

Table 5. Metadata keywords and example.

| **Metadata Keyword** | **Example** |
| --- | --- |
| #SPTOOL\_AQM | CMAQ |
| #SPTOOL\_CARBONS | carbons\_all\_mechanisms\_speciate5\_0\_04mar2020\_v0.csv |
| #SPTOOL\_GAS\_PROFILES | Qry\_gas\_specie\_export.txt |
| #SPTOOL\_INVTABLE | invtable\_hapcap\_NATA\_2016\_19mar2020\_nf\_v11.txt |
| #SPTOOL\_MECH | CB6R3\_AE7 |
| #SPTOOL\_PM\_PROFILES | Not Applicable |
| #SPTOOL\_PROCESS | process\_poll.csv |
| #SPTOOL\_STATIC | Not Applicable |
| #SPTOOL\_RUN\_TYPE | CRITERIA |
| #SPTOOL\_CAMX\_FCRS | Not Applicable |
| #SPTOOL\_VBS\_SVOC\_FACTORS | Not Applicable |
| #SPTOOL\_VBS\_IVOC\_FACTORS | Not Applicable |

If the keyword is not applicable for a particular run then the corresponding value is set to "Not Applicable". For example, if the output type is specified as PM then the metadata record #SPTOOL\_GAS\_PROFILES is set to “Not Applicable”. These headers can be used as a guide for merging different GSPRO tables. Before combining outputs for toxics TOG profiles with NONHAPTOG profiles one should confirm that the same chemical mechanism was specified for each run and that the same input files for the VOC calculations were used for each run.

## Run the Speciation Tool

To run the Speciation Tool, enter the following command:

perl $SPTOOL\_SRC\_HOME/run\_sptool.pl $SPTOOL\_DB <run\_name> <run\_control\_file>

Where:

$SPTOOL\_DB is the Speciation Tool database name

<run\_name> is the user assigned run name

<run\_control\_file> is the name and location of the control file which defines the run parameters and input and output file names.

### Run-time Messages

The Speciation Tool writes a number of messages to standard output that report the program progress. The messages identify key steps as they are completed either from the Perl scripts or the PostgreSQL functions. The difference between the message types is easily detected as all of the PostgreSQL messages begin with the word “NOTICE”. Exhibit 4 provides an example of a typical list of run-time messages. This example is for the run identified as “cb05\_integrate”. If a run name is specified for a case that has been previously run, then additional messages are written at the start of the program that indicate existing tables are being dropped. For example, “NOTICE: drop cascades to table cb05\_integrate.tmp\_metadataset”.

|  |
| --- |
| Created cb05\_integrate schema  Granted permissions on cb05\_integrate schema  Set path successfully to cb05\_integrate schema  Input tables created in cb05\_integrate schema  Completed importing User Profile Weights file user\_prof\_wts\_2540.txt  Imported run control file run\_ctl\_cb05\_integrate.txt in cb05\_integrate schema  NOTICE: Type of Output is VOC  NOTICE: AQM is CAMX  NOTICE: Type of run is INTEGRATE  NOTICE: Mechanism basis is CB05  NOTICE: Profile Tolerance is 5  NOTICE: ...establishing profile weights  NOTICE: ...renormalizing profile weights  NOTICE: ...establishing mechanism  NOTICE: ...calculating moles per gram emissions  NOTICE: ...calculating mole percent  NOTICE: ...calculating moles per mole emissions  NOTICE: ...summing on AQM pollutant  NOTICE: ...calculating mole weight percent  NOTICE: ...calculating average molecular weight by specie  NOTICE: ...calculating average molecular weight by AQM  Completed splits calculations  Completed output preparations  Completed writing the output files for run cb05\_integrate |

Exhibit 4. Example of run-time messages.

There are a number of different warning messages that are displayed during processing. The most frequent message occurs when a species exists in the *shared* schema species table that has not been defined in the specified chemical mechanism. Below is a small example of these messages which provide the Species ID number and name. There are many species that are currently undefined in the chemical mechanisms. This is only a concern if a profile references one of them.

NOTICE: WARNING review vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv

NOTICE: WARNING: CB6 mechanism undefined for the following species.

NOTICE: The following will be set to UNK if referenced.

NOTICE: WARNING: SPECIES ID 2673 2,6-dimethylheptane, propylcyclopentane

NOTICE: WARNING: SPECIES ID 2674 Trans-3-hexene; 3-methylcyclopentene

NOTICE: WARNING: SPECIES ID 2675 2-methyl-2-hexene; cis-3-heptene

NOTICE: WARNING: SPECIES ID 2676 3-methyl-trans-3-hexene; Trans-2-heptene

Profiles are dropped during processing when the weight percent sum is outside the specified tolerance. The dropped profiles are reported to standard output. Below is a small example of these warnings.

NOTICE: ...establishing profile weights

NOTICE: WARNING review vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv

NOTICE: WARNING The following profiles were dropped.

NOTICE: WARNING Total weight percent is outside the tolerance of 5 percent.

NOTICE: WARNING: PROFILE ID 4742 82.87586700 percent Diesel Exhaust - Bus at -10 oC, 2-stroke

NOTICE: WARNING: PROFILE ID 4743 83.00138200 percent Diesel Exhaust - Bus at 20 oC, 2-stroke

NOTICE: WARNING: PROFILE ID 4744 84.50961700 percent Diesel Exhaust - Bus at -10 oC, 2-stroke, oxidatio

Many of these listed profiles have a profile definition with a percent sum of 100%. The problem is not with the profile definition but the incomplete species data. Any species with an undefined molecular weight is dropped from the profile since the molecular weight is required to compute the splits factors. This dropped percent due to missing molecular weights sometimes takes the total profile percent sum outside the accepted tolerance. This type of problem can be resolved by updating the species properties molecular weights.

The complete list of undefined species in a chemical mechanism can be quite long. An additional list of undefined species that are referenced by a profile, whose weight percent will be assigned to UNK is also provided. Thus, profiles with undefined species are NOT dropped, but rather the undefined species is assigned to UNK. Because neither CMAQ nor CAMx use UNK the result is that the percent of the mass assigned to UNK is not used (and effectively the UNK portion of the VOC or PM using the profile is dropped). Below is a small subset of the complete list of species that are not defined in the chemical mechanism.

NOTICE:

NOTICE: WARNING review vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv

NOTICE: WARNING: CB6 mechanism undefined for the following species.

NOTICE: The following species are referenced by profiles and will be set to UNK.

NOTICE: WARNING: SPECIES ID 1014 2-pentylfuran

NOTICE: WARNING: SPECIES ID 2715 Cyclopropane

NOTICE: WARNING: SPECIES ID 2716 1-Chloro-3-methylbenzene

### What happens when the script is running?

The *run\_sptool.pl* program creates the run schema, reads and imports the run control file, imports the run-specific data files, executes the PostgreSQL functions that compute split factors and conversion factors, and writes the output files. After the script has been run, the run-specific schema retains the run parameters, input data, intermediate calculations, and the final results. Appendix C lists the run-based schema table formats for the input data and a list of the database tables created during a run.

The Speciation Tool is driven by the Perl program *run\_sptool.pl*. Table 6 provides a summary of the program steps.

Table 6. Program and module calls from the Speciation Tool program, run\_sptool.pl.

| **Command line** | **Language** | **Description** |
| --- | --- | --- |
| SELECT inputs\_createtables | SQL | Executes PostgreSQL commands in table\_inps.sql:  Creates tables in the run\_name schema. |
| import\_control.pl | Perl | Imports the run control data and optional user specified input data. |
| SELECT MakeSplits | SQL | If VOC processing:  Executes PostgreSQL commands in make\_splits.sql:  Creates tables in the run\_name schema.  Performs data validity checks.  Performs numerous calculations to generate the gas split factors. |
| SELECT MakePMSplits | SQL | If PM2.5 processing:  Executes PostgreSQL commands in make\_pm\_splits.sql:  Creates tables in the run\_name schema.  Performs data validity checks.  Performs calculations to generate the PM2.5 split factors |
| SELECT PrepOut | SQL | Executes PostgreSQL commands in prep\_out.sql:  Creates tables in the run\_name schema.  Generates the output splits and conversion factors data. |
| write\_outputs.pl | Perl | Extracts from the run\_name schema the split factors and conversion factors data and writes the output GSPRO and GSCNV files. |

In addition to the above commands numerous data and processing checks are performed in the run script.

# Applications and Methodology

The Speciation Tool generates emission modeling speciation profiles formatted for SMOKE. The data files provided with the Tool include gas profiles, PM profiles, and a comprehensive species list extracted from EPA’s SPECIATE 5.0 database. In addition, chemical mechanisms used in Air Quality Models CMAQ and CAMx are also provided. These data are the building blocks to generating speciation profiles.

The Speciation Tool can produce either gas profiles or PM2.5 speciation profiles.

## Gas profile processing

The Speciation Tool is designed to support the availability of both CAPS and HAPS, to either integrate the HAPs in the chemical mechanism or not, and to include both active and tracer species in modeling. It is designed to generate the GSPRO and GSCNV speciation input files for the SMOKE model. Refer to the SMOKE User’s Manual for detailed information of how SMOKE handles integrating the criteria VOC and toxics inventories.

The Speciation Tool *shared* schema includes the table *tbl\_invtable* which carries the imported default INVTABLE data. These data should be reviewed to verify that they correspond to the SMOKE INVTABLE that will be used in your SMOKE modeling.

The speciation Tool includes VOC mechanism definitions listed in Table 1. The mechanisms are named to identify their purpose by including both the chemical mechanism name and the aerosol option in the mapping title. For example, the CB6 mappings for CMAQ AE7 and AE8 are titled CB6R3\_AE7 and CB6R3\_AE8, respectively, and the CB6 mapping for the CAMx two-mode coarse/fine aerosol option (CF2) is titled ‘CB6R4\_CF2’. CB6 mappings are named for the most recent CB6 revision present in the target model (R3 in CMAQ and R4 in CAMx). Refer to the Appendix F for detailed information on the development of mechanism mappings for CMAQ and CAMx.

The \_TRACER mechanism is created for CMAQ to produce three additional species that are added to the base mechanism: ALD2\_PRIMARY, FORM\_PRIMARY, and SOAALK. The \_PRIMARYs are used by multi-pollutant and multi-pollutant lite version of CMAQ.[[1]](#footnote-2) SOAALK is needed by CMAQ so that the SOA module in CMAQ can predict the formation of secondary organic aerosols independently of the ozone formation chemistry. Long-chain (C6 to C19) alkanes and small (2–4 ring) polycyclic aromatic hydrocarbons (PAHs), many of which are intermediate volatility organic compounds (IVOCs), can produce secondary organic aerosol (SOA) in relatively high amounts. Pye and Pouliot (2012)[[2]](#footnote-3) provides further details on the significance of these compounds.

### Run Type Options

The different Speciation Tool run options for gas profile processing are: CRITERIA, INTEGRATE, NOINTEGRATE, HAPLIST, and VBS. CRITERIA is specified if no additional HAPS inventory is included in the modeling. The INTEGRATE and NOINTEGRATE options determine how a separate HAPS inventory will be handled in the SMOKE model by either subtracting the HAPS portion from the criteria VOC or not. The HAPLIST option generates the HAPS records used in SMOKE for those HAPs used to compute NONHAPVOC. VBS is a CRITERIA run where a portion of the non-methane mass is assigned to IVOC compounds.

For any one sector, there could be a need for GSPRO records created from multiple run types. The following subsections discuss the different run types. Section 5.1.2 provides an example of the run types needed to generate the gas GSPRO for different use cases for CB6r3\_ae7.

A workbook *Sample\_VOC\_profile\_calcs.xlsx* is provided in an electronic supplement showing processing steps in the Speciation Tool and sample calculation for the run type of CRITERIA.

#### CRITERIA

A run type of CRITERIA means that all VOC model species, including toxic VOC species, are computed from criteria VOC emissions. This option is specified when there are no toxics emissions included in the inventory used for modeling or when the user does not wish to use the toxics emissions, but instead create all model species, including those that match explicit toxics inventory species, from the speciation of VOC.

#### INTEGRATE

A run type of INTEGRATE means that the specified HAPs species are integrated from using the emissions of these HAPs in the emission inventory. Part of the criteria VOC mass will be replaced with the HAP VOC mass from the HAP emission inventory. SMOKE computes the NONHAPVOC mass from the criteria VOC mass by subtracting the HAP VOC mass for each source.

The Speciation Tool creates the HAPs list using the Inventory Table data which is imported during initialization. The HAPs list is composed of species/compounds in the Inventory Table where the field Keep = Y **AND** the field VOCTOG = (V **OR** T). The Speciation Tool removes the HAPs from the TOG profiles and renormalizes to generate NONHAPTOG profiles. The GSCNV output file will contain NONHAPVOC to NONHAPTOG factors.

The GSPRO header records list all of the toxic species that are subtracted from TOG to estimate NONHAPTOG. If the Inventory Table includes process mode information for on-road mobile modeling then some pollutant names will also be appended to the process mode. For example,

#NHAP NONHAPTOG BENZENE

#NHAP NONHAPTOG EVP\_\_BENZENE

#### NOINTEGRATE

In the NOINTEGRATE case both the VOC and toxics are included in emissions processing. However, due to an inability to line up HAPs with the VOC by source, the HAPs and VOC are both used. The criteria VOC includes the HAP VOC mass, but the profiles generated for this run type removes the HAP mass. In generating the speciation profiles, the specified HAPs are removed from the VOC profiles but the profiles are not renormalized, mass is not preserved, and profiles are generated as VOC.

#### HAPLIST

This option generates GSPRO records for the HAPs that define the NONHAPVOC for both the INTEGRATE and NOINTEGRATE cases. All HAPS are written to profile number ‘0000’ which is often used as the default profile ID in SMOKE processing.

#### VBS

The VBS (Volatility Basis Set) option computes semi-volatile OC (SVOC) species fractions (split factors) from primary OA species (POA= POC plus PNCOM), as well as intermediate volatility organic compound (IVOC) species from total non-methane organic compounds (TNMOC) to support the VBS scheme in CTMs. The output includes IVOC/SVOC species along with the full suite of model species depending on the choice of mechanism.

For gas processing, the VBS option produces intermediate-volatility organic compounds (IVOC) and adjusted non-methane organic gas (NMOG) components. IVOC represents VOC with volatility ranging from 103 to 106 μg/m3. The Speciation Tool default input data includes a table of profile codes and the fraction of NMOG that is assigned to IVOC for each profile, because the IVOC fraction (i.e., IVOC/TNMOC) can vary by source category.[[3]](#footnote-4) When IVOC is added the VOC model species is reduced so that total NMOG mass is conserved. Currently (2020) about 20% of gas profiles output IVOC compounds.

### Example Use Cases for Gas Profiles

For gas profiles, multiple runs of the Speciation Tool may be required to create the GSPRO files you need for running SMOKE, particularly if you want to use different options for HAPs for different inventory sectors. Table 7 provides examples of the different runs needed for different use cases when using CB6R3AE7 as your mechanism. This mechanism contains SOAALK which is not produced during the same run as the other gas model species. Therefore, even when not using HAPs you need multiple runs of the tool for this mechanism.

Table 7. Examples of SPTOOL runs for different options for using HAPs

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case** | **Run type** | **MECH\_BASIS** | **Notes** |
| Do not use HAPs | CRITERIA | CB6R3AE7 | To create TOG profiles |
|  |  | CB6R3AE7\_TRACER | To create SOAALK.  Can also use to create ALD2\_PRIMARY and FORM\_PRIMARY from VOC emissions, if needed. |
| INTEGRATE HAPs1 | HAPLIST | CB6R3AE7 | To map inventory HAPs to model species |
|  |  | CB6R3AE7\_TRACER | To map inventory formaldehyde and acetaldehyde to FORM\_PRIMARY and ALD2\_PRIMARY, respectively |
|  | INTEGRATE | CB6R3AE7 | To create NONHAPTOG profiles |
|  |  | CB6R3AE7\_TRACER | To create SOAALK |
| Use HAPs with NOINTEGRATE1 | HAPLIST | CB6R3AE7 | To map HAPs in the inventory to model species |
|  |  | CB6R3AE7\_TRACER | To map inventory formaldehyde and acetaldehyde to FORM\_PRIMARY and ALD2\_PRIMARY, respectively |
|  | NOINTEGRATE | CB6R3AE7 | To create TOG profiles that drop the HAPs |
|  |  | CB6R3AE7\_TRACER | To create SOAALK |

1 For INTEGRATE or NOINTEGRATE runs, the user selects which HAPs to use.

When using HAPs, it is likely to have multiple use cases within the same sector, for example, you may integrate some sources and use the no HAPs for the other sources. This would require the run types listed in the “Do not use HAPs” and “INTEGRATE” use cases. You may also choose to integrate different HAPs for different sectors, which would also require separate INTEGRATE runs of the Speciation Tool, each using a different INVTABLE (the input that identifies the HAPs you choose to integrate).

### Chemical Mechanisms

During initialization the Speciation Tool imports default data (profiles, species list, chemical mechanism, etc.) into the *shared* schema. However, the Speciation Tool is not tailored to a specific chemical mechanism such as CB05. If the chemical mechanism you require is not available in the Speciation Tool database you can import a new mechanism definition, assuming you can supply the new chemical mechanism assignments. Refer to the Chapter FAQ for these details.

The Speciation Tool includes mechanism mapping of the IVOC compounds to VBS model species for chemical mechanisms supporting VBS, so it processes the SPECIATE profile of type ‘GAS-VBS’ for those mechanisms (example, CB6R3\_AE8). The Speciation Tool can also produce an optional intermediate file that contains mapping of IVOC compounds to VBS model species. The intermediate file provides IVOC binning for each species based on vapor pressure information in SPECIATE. This file is a first step in developing a complete chemical mechanism assignment for the VBS scheme in the AQM and subsequently requires offline review to develop a complete mapping file for each chemical mechanism.

When using SAPRC or CRI mechanisms keep in mind that the carbon content of VOC or TOG emissions is not conserved by the split factors, that the output mass fractions in the GSPRO file are sensitive to the number of carbons assumed for each species, and that carbon numbers are not well-defined for SAPRC and CRI lumped species (i.e., model species that represent several compounds with potentially different numbers of carbons). This uncertainty does not influence the output modeling emissions in molar units. Output emissions reported in mass units (such as speciated SMOKE reports) do inherit this uncertainty and, therefore, should be considered as qualitative rather than quantitative information.

## PM2.5 Profile Processing

The Speciation Tool provides speciation profile entries (GSPRO) for PM2.5. The CRITERIA and VBS run options are supported for PM processing. The VBS option is needed to compute SVOC species fractions from primary OA species (POA= POC plus PNCOM) to support VBS schemes in the CTM. The chemical mechanism definitions imported during initialization include the CMAQ aerosol mechanisms AE6/AE7 (AE6 and AE7 are the same for PM profile species) and AE8. Appendix D provides these mechanism definitions. The Speciation Tool processes profiles with a lower bound size of 0 and an upper bound size of 2.5, all profiles outside these conditions are ignored.

The PM2.5 chemical mechanism definitions include a list of species and the AQM pollutant name to be assigned any unspecified PM mass (e.g., PMOTHR). In the AE6/AE7 mechanism, the SPECIATE profile of type ‘PM’ and ‘PM-AE6’ are processed and, in the AE8 mechanism, profiles of type ‘PM’, ‘PM-AE6’, and ‘PM-VBS’ are processed. For each profile that matches the selection criteria, only the species that are specifically defined in the chemical mechanism definition will be included in the outputs. Each PM2.5 chemical mechanism definition is required to have one record specifying the AQM pollutant to compute. The unspecified mass component of each profile is defined as 1.0 minus the sum of the mass fractions of the mechanism compounds and is assigned to the ‘compute’ AQM pollutant. In the AE6/AE7 and AE8 chemical mechanisms the unspecified mass is assigned to PMOTHR.

The profiles of type ‘PM’ go through a series of steps to create AE6 profiles which are described in Appendix E and a sample calculation shown in the workbook *Sample\_make\_AE6\_profile\_calcs.xlsx* provided in an electronic supplement.

When using the AE6/AE7 and AE8 mechanisms, if both ion and atom form of an element are present in the same profile, the Speciation Tool preferably uses the ion form to develop the AE6 profile. This is a change from the previous version of Speciation Tool (v 4.0 and earlier) which used the atom form. If an ion form is not present and an atom form is present, then the atom form is used. If neither ion nor atom forms are present but an oxide form of the element is present, then corresponding AE6 species is computed stoichiometrically from the oxide form.

### Special Handling for CAMx Particulate Modeling

As stated above the AE6 and AE8 aerosol mechanisms were developed specifically for CMAQ. In order to support CAMx PM speciation scheme (scheme CAMx CF), the Speciation Tool calculates the PM model species required by CAMx from the AE6 definition. Table 8 shows the relationship between CAMx CF scheme model species and the AE6 compounds.

Table 8. PM2.5 components from CMAQ to CAMx.

| **CMAQ AE6 compounds** | **CAMx CF** |
| --- | --- |
| PAL | sum FPRM |
| PCA |
| PFE |
| PK |
| PMG |
| PMN |
| PMOTHR |
| PSI |
| PTI |
| PNCOM | sum POA |
| POC |
| PEC | PEC |
| PNA | NA |
| PH2O | PH2O |
| PNH4 | PNH4 |
| PNO3 | PNO3 |
| PSO4 | PSO4 |
| PCL | PCL |
| POC | POC\* |

\* Note that POC is accounted for under the CAMx POA compound. Although CAMx ignores POC, it is still included as a separate compound to assist with model performance evaluation.

### Volatility Basis Set (VBS)

In order for downstream CTMs to simulate the semivolatile partitioning of primary organic emissions, the conventional organic aerosol species (POA for CAMx or POC and PNCOM for CMAQ) which are assumed to be nonvolatile must be distributed to SVOC species (example, species 3245-3254 in Table D1 for CMAQ). To do so, “VBS Profile” must be assumed for each source. There are several places in the emission input generation process where this distribution can be applied:

Source profiles in the SPECIATE database can include particle SVOC species. Profiles of this sort are available starting in SPECIATE 5.0 for gasoline and diesel vehicles as well as jet engines. These profiles are thus labelled “PM-VBS”. However, the majority of profiles available in SPECIATE do not have SVOCs resolved.

VBS profiles may be applied during the Speciation Tool execution (the topic of this section). Two pieces of information are needed to accomplish this task: (1) specific volatility distribution to be applied for a certain source types as discussed below, and (2) a table identifying the profiles that fit those defined source types. For example, to distribute biomass-burning sources, Speciation Tool needs to know which profiles to modify and the SVOC factors to modify them with.

Some CTMs (e.g., CMAQ) allow for online mapping and scaling of emissions input species to model species. The benefit to applying the VBS distribution at this stage is that emissions do not have to be reprocessed when a distribution is modified. However, it does mean that emission inputs must be provided for many sources independently so that the correct VBS profile can be applied to each.

With the VBS run-type option (#2 above), the Speciation Tool computes SVOC species fractions (split factors) from primary OA species (POA= POC plus PNCOM), as well as intermediate volatility organic compound (IVOC) species from total non-methane organic compounds (TNMOC).

The Speciation Tool includes three distinct sets of SVOC model species to represent emissions from cooking meat (charbroiling), biomass-burning and other anthropogenic sources. Each SVOC model species set has five model species with different saturation pressure ranging from zero (non-volatile) to 1000 μg/m3 (most volatile). The Speciation Tool uses separate input file to support generating SVOC model species, specifically assignment of each profile code to one of the three emission categories (cooking induced, biomass-burning, other anthropogenic), and SVOC species fractions for each saturation concentration bin. Table 9 provides examples of the input data required for SVOC generation. The Speciation Tool Github includes this data in the file “vbs\_svoc.profile.26sep2016.csv” which is imported to table “tbl\_vbs\_svoc\_factors”.

IVOC represents VOC with volatility ranging from 103 to 106 μg/m3. Three IVOC model species are used to represent the source categories gasoline engine exhaust, diesel engine exhaust, and biomass-burning. The IVOC fraction (i.e., IVOC/TNMOC) may vary by source category. The Speciation Tool uses a separate input data to support generating IVOC model species, specifically the source category type of each profile code (which is reflected in the model species name) and the IVOC fraction of TNMOC. Table 10 provides several examples of the input data required for IVOC generation. The Speciation Tool Github includes this information in the file “vbs\_ivoc.profile.30aug2016.csv” which is imported to table “tbl\_vbs\_ivoc\_nmogfactors”.

The file identifying these profiles was created prior to the SPECIATE 5.0 release, so this input file may need to be modified to make sure all profiles are listed if using this feature of the Speciation Tool. In future versions of the Speciation Tool, the SPECIATE Categorization fields can be used instead of a table of profiles, which would alleviate the need to update the input file with every new version of SPECIATE.

Also, the Speciation Tool must have SVOC and IVOC species fractions for each saturation concentration bin. SVOC species will replace the traditional POA for CAMx or POC and PNCOM for CMAQ. The VBS run-type option currently requires that all the five bin fractions must sum to one, though this mass conservation does not necessarily have to be required since the inventory may have missed SVOCs due to dilute conditions in the emissions sample. Most likely as specific VBS profiles are developed, they will in fact not necessarily conserve mass. However, the current approach used by the Speciation Tool requires mass conservation and will compute the sum during processing and will abort if a record is found where the sum of the fractional components is not equal to 1.0.

Table 9. Example SVOC model species assignments for three profile codes

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Profile Code | Profile Description | Emission Category |  | Saturation Concentration Bin (μg/m3) | | | | |
| CAMx  CMAQ | 0 | 1 | 10 | 100 | 1000 |
| Model Species | bin\_0 | bin\_1 | bin\_2 | bin\_3 | bin\_4 |
| 91106 | Diesel engines | Anthropogenic | PAP*i*  P\_PVB*i* | 0.03 | 0.25 | 0.37 | 0.24 | 0.11 |
| 91102 | Wildfires | Biomass Burning | PFP*i*  P\_FVB*i* | 0.2 | 0.1 | 0.1 | 0.2 | 0.4 |
| 91116 | Charbroiling | Cooking-induced | PCP*i*  P\_CVB*i* | 0.35 | 0.35 | 0.1 | 0.1 | 0.1 |

Table 10. Example IVOC assignments for three profile codes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Profile Code | Profile Description | Emission Category | Model Species | | IVOC/NMOG |
| CAMx | CMAQ |
| 5345 | Gasoline Exhaust - E10 splash blend gasoline | Gasoline Engines | IVOG | IVOC\_G | 0.25 |
| 8774 | Diesel Exhaust Emissions from Pre-2007 Model Year Heavy-Duty Diesel Trucks | Diesel Engines | IVOD | IVOC\_D | 0.20 |
| 8743 | Composite Profile - Forest Fires | Biomass Burning | IVOB | IVOC\_F | 0.20 |
| 5650 | Residential Wood Combustion | Wood Burning | IVOB | IVOC\_F | 0.07 |

# FAQ

## How do I add a handful of speciation profile?

The easiest method of processing one or more new profiles is to use the run control keyword pro\_file. The profile definition is imported to a run schema table. If this table is populated, then only profiles in the run schema are processed.

## Can I have more than one copy of the Speciation Tool database?

Yes. If you want to rerun the Speciation Tool initialization and keep an already existing instance of the Speciation Tool database then all you need to do is make a copy of the *Assigns.sptool* file and change the database name variable SPTOOL\_DB. ‘Source’ the new Assigns file and run the initialization program. A new database is created and the specified data files are imported.

This is the approach to take if you want to modify any of the shared data. In particular, you may have different versions of the INVTABLE that you want to support. Each version can be represented in a different database.

## How do I define a new chemical mechanism?

To introduce a new chemical mechanism to the Speciation Tool first define the assignments to the mechanism table with a unique mechanism name (different from existing mechanisms) and import the data to the shared schema mechanism table (tbl\_mechanism for VOC processing, tbl\_pm\_mechanism for PM2.5 processing).

Next, the *shared* schema carbon table (tbl\_carbons) may need to be updated to include any new model species names for VOC processing. The last step is to update the mechanism description table (tbl\_mechanism\_description) also in the *shared* schema.

For examples of the tables used to define new chemical mechanisms look under the import\_data directory in the Speciation Tool distribution package for the following files:

tbl\_mechanism: mechanism\_forImport\_11Feb2020\_speciate5\_0\_withSOAALK\_13mar2020\_v0.csv

tbl\_carbons: carbons\_all\_mechanisms\_speciate5\_0\_04mar2020\_v0.csv

tbl\_mechanism\_description: mechanism\_description.30Mar2020.txt

Refer to Appendix B for the file formats.

The new mechanism data can be imported with the *import\_rawdata.pl* Perl program. This needs to be run 3 times for the 3 different table\_types (mechanism, carbon, mechanism description)

perl $SPTOOL\_SRC\_HOME/import\_rawdata.pl $SPTOOL\_DB table\_type input\_file

table\_type - is the keyword from Table 1 (in this example it will be either mechanism, carbons, or mechanism\_description)

input\_file - contains the specified new records

Alternatively, rather than appending records to three separate shared schema tables (tbl\_mechanism, tbl\_carbons, tbl\_mechanism\_description) a new version of the Speciation Tool can be generated (refer to the question above). The new mechanism definition data could be imported as part of the initialization process. These data could be appended to existing files or not – depending on your processing preferences.

### Example of how to introduce a new mechanism, “SAPRC07TC\_ae7 with explicit benzene”

Update the mechanism table:

Extract the SAPRC07TC\_ae7entries from the import file:

grep ^ SAPRC07TC\_ae7 mechanism\_forImport\_11Feb2020\_speciate5\_0\_withSOAALK\_13mar2020\_v0.csv > mechanism\_ SAPRC07TC\_ae7\_wexpl\_benz.txt

Edit the new file:

Change SAPRC07TC\_ae7 to SAPRC07TC\_ae7\_BENZ (on all records)

Delete existing benzene (specie\_id = 302) assignments of “ARO1” and “NROG”

Add explicit benzene entry “SAPRC07TC\_ae7\_BENZ, 302, BENZ, 1.”

Import the new mechanism file to the shared mechanism table:

perl import\_rawdata.pl database\_name mechanism mechanism\_SAPRC07TC\_ae7\_wexpl\_benz.txt

Update the carbon table:

Create a file with the record “SAPRC07TC\_ae7\_BENZ,BENZ,6”

Import the file to the shared carbon table

perl import\_rawdata.pl database\_name carbons <new\_carbon\_file.txt>

Update the mechanism description table:

Create a file with the record “SAPRC07TC\_ae7\_BENZ, SAPRC07TC\_ae7 with explicit benzene, Y, <description>,,”

Import the file to the shared mechanism description table

perl import\_rawdata.pl database\_name mechanism\_description <new\_file.txt>

The mechanism description table carries the field “NONSOA Flag” which defines whether the mechanism treats SOA explicitly. The CB and SAPRC mechanisms all have NONSOA Flag set to “Y”. Only the SOA\_CAMX45 mechanism has the flag set to “N”.

## What if I have a different INVTABLE?

The SMOKE INVTABLE data is imported as part of the Speciation Tool initialization. Currently there is not an optional input of this data for a single run. If you only need your version of the INVTABLE then replace the INVTABLE filename that is referenced in the Assigns.sptool before you initialize the database.

If you need multiple versions for your processing applications, then you will need multiple versions of the Speciation Tool database; each importing a different INVTABLE during initialization. (Refer to 6.2 above).

## How do I add a new species to the database?

To add a single record to the data base from the command line use an INSERT statement. The *shared* schema *tbl\_species* carries almost twenty fields (exported from the SPECIATE database). The Speciation Tool references only a few of these fields. The required fields are SPECIES\_ID, NAME, SPEC\_MW, and NonVOCTOG. Here is an example to add a single record to the table with only the required fields:

psql –d $SPTOOL\_DB –c “INSERT INTO shared.tbl\_species (species\_id,name,spec\_mw,nonvoctog)

VALUES (‘EG1’, ‘Example Name’, 70.273,FALSE)”

The SPECIES\_ID must be unique in the table; you will get an error if you attempt to add an already existing SPECIES\_ID. The NonVOCTOG field is type Boolean and should be set true if a TOG species is not regarded as a VOC.

**Appendix A**

**PostgreSQL and Perl**

**Installation Procedures**

**Linux Installation**

This section provides the details for installing PostgreSQL and the required Perl modules (DBI, DBD-Pg, and Text-CVS) via the Linux command line. Before continuing it is recommended that you discuss these software requirements with your system administrator as you may need to be logged in as root.

**Status of Requirements**

Begin by checking what is or is not installed on your system. Execute the script *sptool\_reqd\_checks.sh* which is provided in the Speciation Tool package. The script checks for the required software and user PostgreSQL authorizations.

> ./sptool\_reqd\_checks.sh

Following is an example of a successful check for the required software:

===== Speciation Tool Requirements Check =====

Checking the status of software requirements...

Status of required software:

[x] = Installed

[ ] = Not installed

[?] = Unable to determine, see notes

------PERL------

[x] Perl

[x] -DBI

[x] -DBD-Pg

[x] -Text-CSV

---POSTGRESQL---

[x] PostgreSQL

[x] -PL/pgSQL

Refer to the Speciation Tool User Guide Appendix A for installation procedures of the required software.

Missing software/modules are indicated by a [ ] whereas modules/software that is installed and ready to be used are marked with [x]. Note that if PostgreSQL is erroneously shown as not installed, restarting the PostgreSQL service can fix many of the reasons that might cause this to occur.

**Perl**

For the purposes of this guide, it is assumed that Perl has already been installed on your system. Almost every modern Linux distribution comes with Perl installed. To determine if Perl is installed, enter the following at the Linux command line:

# perl –v

If Perl is installed, information about the version will be displayed. Additional interfaces and modules for Perl need to be installed for the Speciation Tool to read text files and communicate with the database. Instructions for the installation of these modules follow the PostgreSQL installation section. PostgreSQL must be installed prior to installing the Perl database driver and interface.

**PostgreSQL Installation**

The Yum software package manager is an easy to use utility that installs, updates and removes software packages on RPM-based systems. This utility is found on RedHat, CentOS, Fedora, and other RPM-based Linux distributions. You can use yum to install PostgreSQL and the required Perl modules.

Note: For systems with different installation procedures/requirements, PostgreSQL pre-compiled binaries can be found at <http://www.postgresql.org/>. Click on the Downloads tab to review the available products. From <http://www.postgresql.org/download/>, choose the binary package corresponding to your operating system. As of writing, binary packages are available for the following operating systems:

* BSD
  + FreeBSD
  + [OpenBSD](http://www.postgresql.org/download/openbsd/)
* Linux
  + [RedHat/CentOS/Fedora/Scientific](http://www.postgresql.org/download/linux/redhat/) families Linux
  + [Debian](http://www.postgresql.org/download/linux/debian/) GNU/Linux
  + [Ubuntu](http://www.postgresql.org/download/linux/ubuntu/) Linux
  + [SuSE](http://www.postgresql.org/download/linux/suse/) and OpenSuSE
  + Other Linux
* Mac OS X
* Solaris

Windows

**Install PostgreSQL using the Installation Wizard**

To install PostgreSQL using Yum, you will probably need to be logged in as root. To download and install PostgreSQL, execute the following command:

# yum install postgresql

If prompted, enter ‘y’ to confirm the installation. The package will download and install; additional packages may be automatically selected for install in order to resolve dependencies.

**Install Perl database interface modules**

Three Perl modules are required to read in text files and communicate with the database used by the Speciation Tool. They are:

Perl DBI (perl-DBI) The Perl Database Interface (DBI) allows communication between the Speciation Tool scripts and PostgreSQL database.

Perl DBD-Pg (or perl-DBD-Pg) This is a PostgreSQL-specific database driver for the DBI module to allow Perl to communicate with the PostgreSQL database using the DBI.

or Perl DBD:PgPP

Perl Text-CSV (perl-Text-CSV) This module allows for Perl scripts to parse and create CSV (Comma Separated Values) files.

Use Yum to download and install (if needed) all three modules. Enter the command:

# yum install perl-DBI perl-DBD-Pg perl-Text-CSV

Enter ‘y’ when prompted to confirm the installation of these packages.

Note: If errors about packages not being found are displayed, check the spelling of the package names in your command; they are case-sensitive.

**Start the PostgreSQL Service and Prepare it for the Speciation Tool**

To start the PostgreSQL database service, execute the following command:

# /etc/init.d/postgresql start

Change to the user who will be running the Speciation Tool:

# su <user name here>

Start the terminal-based front-end to PostgreSQL (psql) with the default postgres user:

> psql –U postgres

Create PostgreSQL user(s) with database create option for each of the users that will be running the Speciation Tool. The user name should correspond to the Linux user account name. The command is:

=# CREATE USER <user name here> WITH CREATEDB;

Exit psql by entering “\q”:

=# \q

At this point you should be ready to follow the Speciation Tool set up procedures (refer to Chapter 3). Run the script *sptool\_reqd\_checks.sh* to check that all of the packages are available to the Speciation Tool.

**Windows Installation**

This section provides the details for installing PostgreSQL and the Perl Database Interface (DBI, DBD:PgPP) in a Windows environment. Perl must be installed prior to PostgreSQL in order for the PostgreSQL Installation Wizard to offer the procedural language PL/perl.

Steps for Windows Installation:

1. Install Perl
2. Install PostgreSQL
3. Install additional Perl modules

**Perl Installation**

1. If necessary, download ActivePerl software from http://www.activestate.com/
2. Unzip ActivePerl and click Installer.bat.
3. A DOS window will open, read and accept the license agreement.
4. Enter the top-level directory for Perl installation.
5. Accept or decline Perl features by replying yes/no.
6. Proceed to start the installation.
7. Restart your computer before you proceed with the PostgreSQL installation.

**PostgreSQL Installation**

PostgreSQL can be found at <http://www.postgresql.org/>. Click on the Downloads tab to review the products available. From <http://www.postgresql.org/download/> click FTP Browser followed by selecting the win32 subdirectory. From this subdirectory download postgresql-8.1.4-1.zip.

**Prior to installing PostgreSQL**

Check the file system of the hard drive on which PostgreSQL will be installed. The NTFS file system has a comprehensive access control system and offers the reparse point functionality to implement tablespaces used by PostgreSQL. The PostgreSQL installer package will not initialize a database cluster on anything but an NTFS partition. FAT and FAT32 file systems do not offer such reliability. It is necessary that a FAT or FAT32 disk be converted to NTFS prior to installation of PostgreSQL.

To determine the file system of the hard drive, through Windows Explorer, highlight the drive, right click and select Properties. The General tab of the Local Disk Properties window displays the current file system of the selected drive.

To convert FAT/FAT32 to NTFS:

1. open a DOS window
2. from the DOS command prompt enter the command

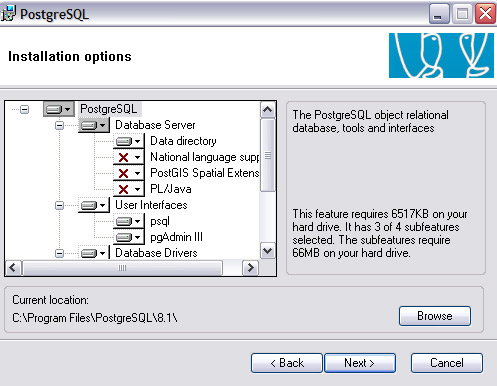
convert drive\_letter: /fs:ntfs.

For example, typing convert D: /fs:ntfs would format the D: drive to the ntfs file system format. It is a good practice to backup all important data prior to the conversion.

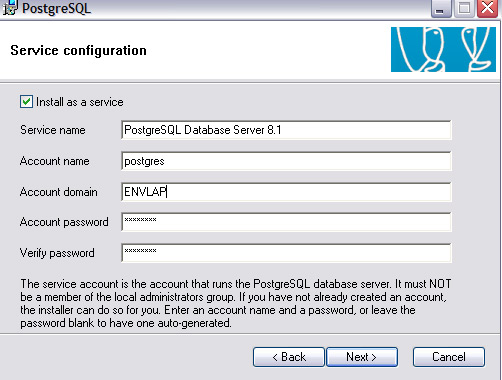
**Install PostgreSQL using the Installation Wizard**

A complete description of the installation process can be found at <https://www.2ndquadrant.com/en/resources/postgresql-installer-2ndquadrant/> (accessed on 06/27/2020).

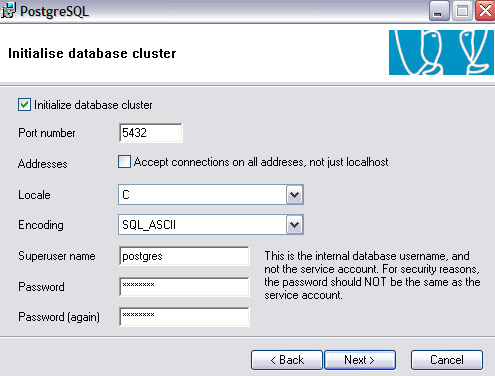
1. Unzip the PostgreSQL software that you downloaded.
2. Execute (double-click on) postgresql-8.1.msi. This will start the installation wizard.
3. Select the language you want to use for the installer and click Start.
4. Click Next to continue the installation following the Welcome message.
5. Click Next to continue following the Installation notes.
6. In the Installation options window, verify that “Database Server” > “Data directory”, “User Interfaces”, and “Database drivers” are selected. You may change the PostgreSQL installation directory by clicking on Browse and specifying the directory of your choice.



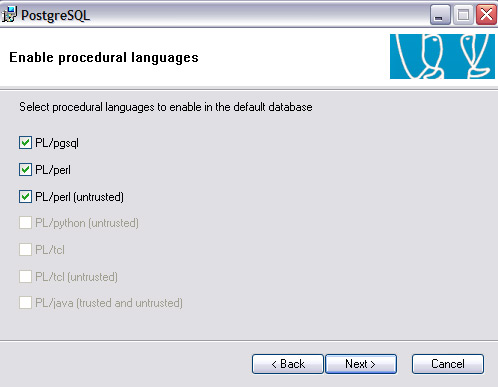
1. In the Service configuration window install PostgreSQL as a service by checking the box “Install as a service”. Provide and verify a password for the account name “postgres”. This is only for an administrator to run the PostgreSQL database server. Click Next to continue. Click Yes if you get the error message “Would you like the account to be created for you?”



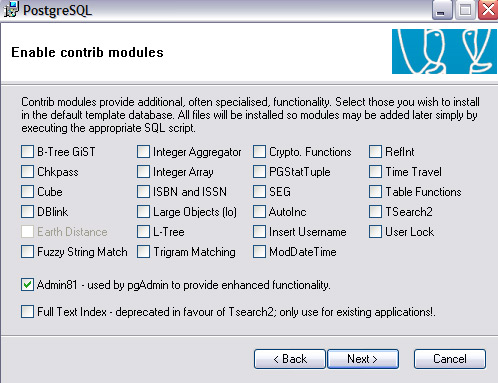
1. In the Initialize database cluster window provide the superuser name postgres and a password. Click Next to continue.



1. In the Enable procedural languages window select PL/pgsql, PL/perl and PL/perl (untrusted) to be enabled in the default database. Click Next to continue.



1. Click Next to use the default setting of Admin81.



1. Click Next to start the actual installation.

Please note that these steps are applicable to PostgreSQL version 8.1. Consult the PostgreSQL website if installing a later version.

**PostgreSQL Client Authentication**

PostgreSQL offers a number of different client authentication methods. The method used to authenticate a particular client connection can be selected on the basis of (client) host address, database, and user. The Speciation-PC Tool is set to trust users by running on the local host without a password. To allow this setting in PostgreSQL, users will have to modify a client authentication setting. The client authentication is controlled by a configuration file, which traditionally is named pg\_hba.conf and is stored in the database's data directory; i.e. C:\Program Files\PostgreSQL\8.1\data.

To change the client authentication

1. Open pg\_hba.conf file (text format) in the database’s directory.
2. Modify the configuration file by adding a “trust” entry in “METHOD” field and save.

#TYPE DATABASE USER CIDR-ADDRESS METHOD

host all all 127.0.0.1/32 trust

1. If this step is done after PostgreSQL service is running, stop the service, reload configuration and restart the service (Click Start Menu>Programs>PostgreSQL 8.1> Stop service, Reload configuration, Start service). Otherwise, only click reload configuration.

**Install Perl database interface Modules**

The following procedures, utilizing the PPM, require a connection to the internet. Alternatively, these packages can be downloaded manually and loaded offline.

1. Open a DOS window and start Perl’s Programmer’s Package Manager by entering “ppm” on the command line.
2. At the ppm command prompt, install DBI by entering “install DBI”.
3. At the ppm command prompt, install DBD:PgPP by typing “install DBD-PgPP”.
4. At the ppm command prompt, install Text-CSV by typing “install Text-CSV”.

At this point, if all directions have been followed without error, you should be ready to follow the Speciation Tool set up procedures.

**Appendix B**

**SHARED Schema File Formats**

A field name in blue indicates an indexed field.

**Table B1. tbl\_carbons.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| mechanism | Character 20 | Mechanism name |
| aqm\_poll | Character 20 | Model species identifier/name |
| num\_carbons | Numeric (5,2) | Number of carbons for model species |

**Table B2. tbl\_profile\_weights.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| profile\_id | Character 20 | Profile identifier |
| specie\_id | Character 20 | Unique species identifier |
| percent | Numeric (10,6) | Percentage of pollutant in profile |
| uncertainty | Numeric (10,6) | Uncertainty percent of pollutant |
| unc\_method | Character 100 | Description of method used to determine uncertainty |
| analytic\_method | Character 500 | Description of analytical method |

**Table B3. tbl\_profiles.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| profile\_id | Character 20 | Unique profile identifier |
| profile\_name | Character 200 | Profile name |
| profile\_type | Character 20 | Profile type |
| total | Numeric (10,5) | Sum of profile percentages |
| master\_poll | Character 20 | Basis of profile |
| norm\_basis | Character 100 | Normalization description |
| composite | Character 1 | Indicates if profile is original or composite; O or C |
| standard | Boolean | Standard or user added profile |
| incl\_gas | Boolean | Includes inorganic gas |
| test\_year | Character 50 | Year testing was completed |
| j\_rating | Numeric (10,2) | Judgment rating based on general merit |
| v\_rating | Numeric (10,2) | Vintage Rating |
| d\_rating | Numeric (10,2) | Data quality rating based on number of observations |
| region | Character 100 | Geographic region of relevance |
| samples | Character 100 | Samples |
| lower\_size | Numeric (10,4) | Lower size |
| upper\_size | Numeric (10,4) | Upper size |
| sibling | Character 20 | Profile\_ID of PM profile from same study |
| version | Character 5 | Version |
| voc\_to\_tog | Numeric (12,7) | VOC to TOG conversion factor |
| t\_sample | Numeric (10,2) | Temperature of sample (C) |
| rh\_sample | Numeric (10,2) | Relative humidity of sample |
| p\_loading | Numeric (10,2) | Particle loading (ug/m3) |
| o\_loading | Numeric (10,2) | Organic loading (ug/m3) |
| gen\_mechanism | Character 100 | Generation mechanism |
| sec\_equipment | Character 100 | Sector equipment |
| fuel\_product | Character 100 | Fuel product |
| ms\_poll\_rate | Numeric (10,2) | Master pollutant emission rate |
| ms\_poll\_unit | Character 10 | Master pollutant emission rate unit |
| om\_to\_oc | Numeric (10,2) | Organic matter to organic carbon ratio |
| mass\_overage\_pct | Numeric (15,12) | Mass overage percent |

**Table B4. tbl\_invtable.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| eminv\_poll | Character 12 | Emission inventory pollutant name |
| mode | Character 3 | Process mode |
| poll\_code | Character 16 | Pollutant code |
| specie\_id | Character 20 | Species identifier |
| reactivity | Character 20 | Reactivity group |
| keep | Character 20 | Keep flag; Y or N |
| factor | Character 20 | Adjustment factor |
| voc | Character 20 | VOC or TOG component flag |
| model | Character 20 | Model species flag |
| explicit | Character 20 | Explicit in mechanism flag |
| activity | Character 20 | Data type flag; Y indicates activity (not emissions) |
| nti | Character 20 | Identifies HAPs on the Clean Air List |
| unit | Character 20 | Units |
| description | Character 50 | Inventory data description |
| cas\_description | Character 50 | CAS pollutant description |

**Table B5. tbl\_mechanism.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| mechanism | Character 20 | Mechanism name |
| specie\_id | Character 20 | Unique species identifier |
| aqm\_poll | Character 20 | Model species identifier/name |
| moles\_per\_mole | Numeric (20,12) | The moles per mole |

**Table B6. tbl\_mechanism\_description.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| mechanism | Character 20 | Mechanism name |
| description | Character 256 | Description |
| nonsoaflag | Character 1 | Flag: “N” SOA mechanism  “Y” nonSOA mechanism |
| origin | Character 300 | Originating organization |
| reference | Character 100 | References |
| comment | Character 500 | Comment |

**Table B7. tbl\_metadata.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| keyword | Character 20 | Metadata keyword |
| datival | Character 256 | Corresponding value (file) |
| version | Character 20 | Data version |

**Table B8. tbl\_pm\_mechanism.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| mechanism | Character 20 | Mechanism name |
| specie\_id | Character 20 | Unique species identifier |
| aqm\_poll | Character 20 | Model species identifier/name |
| qualify | Boolean | Indicates which species determines if a profile qualifies for the specified mechanism |
| compute | Boolean | If true, compute aqm\_poll (usually PMFINE or PMOTHR) |

**Table B11. tbl\_rename\_species**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| aq\_model | Character 10 | AQM Model |
| mechanism | Character 20 | Mechanism name |
| eminv\_poll | Character 20 | Compound name |
| aqm\_poll | Character 20 | Replacement model species identifier/name |

**Table B12. tbl\_species.**

| **Field Name** | **Field Type** | **Description** |
| --- | --- | --- |
| specie\_id | Character 20 | Unique species identifier |
| specie\_name | Character 100 | Species name |
| CAS | Character 50 | CAS identifier |
| epaid | Character 50 | EPA identifier |
| saroad | Character 10 | Old SAROAD code |
| pams | Boolean | Yes or No to PAMS |
| haps | Boolean | Yes or No to HAPs |
| symbol | Character 10 | Symbolic name |
| molecular\_weight | Numeric (20,12) | Molecular weight of specie |
| non\_voctog | Boolean | Yes or No to non-volatile organic gas |
| non\_vol\_wt | Character 20 | Non-volatile weight |
| unknown\_wt | Character 20 | Unknown weight |
| unassign\_wt | Character 20 | Unassigned weight |
| exempt\_wt | Character 20 | Exempt weight |
| volatile\_mw | Numeric (20,12) | Volatile molecular weight |
| num\_carbons | Numeric (20,12) | Number of carbon bonds |
| epa\_itn | Character 20 | EPA internal tracking number |
| comment | Character 50 | Comments |
| vp\_epi | Numeric (20,12) | Vapor pressure from EPI |
| vp\_um | Numeric (20,12) | Vapor pressure from UM |

**Table B13. tbl\_static.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| profile\_id | Character 20 | Profile identifier |
| eminv\_poll | Character 20 | Emission inventory pollutant code |
| aqm\_poll | Character 20 | Model species identifier/name |
| split\_factor | Numeric (20,10) | Mole based split factor (numerator) |
| divisor | Numeric (20,10) | Denominator of the mole based factor |
| mass\_fraction | Numeric (20,10) | Mass fraction |
| aq\_model | Character 10 | Air Quality Model |

**Table B14. tbl\_camx\_fcrs.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| profile\_id | Character 20 | Profile identifier |

**Table B15. tbl\_vbs\_ivoc\_nmogfactors.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| profile\_id | Character 20 | Profile identifier |
| cmaq\_ivocname | Character 20 | CMAQ IVOC compound name |
| camx\_ivocname | Character 20 | CAMx IVOC compound name |
| nmogfraction | Numeric (10,6) | Fraction of NMOG assigned to IVOC compounds |

**Table B16. tbl\_vbs\_ivoc\_species.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| aqm | Character 20 | Air Quality Model name |
| specie\_id | Character 20 | IVOC compound name (assigned in Table B15) |
| molwt | Numeric (10,6) | Molecular weight of IVOC compound |

**Table B17. tbl\_vbs\_svoc\_factors.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| profile\_id | Character 20 | Profile identifier |
| cmaq\_svocname | Character 20 | CMAQ SVOC compound name |
| camx\_svocname | Character 20 | CAMx SVOC compound name |
| bin0 | Numeric (10,6) | Fraction of POA assigned to SVOC\_bin0 |
| bin1 | Numeric (10,6) | Fraction of POA assigned to SVOC\_bin1 |
| bin2 | Numeric (10,6) | Fraction of POA assigned to SVOC\_bin2 |
| bin3 | Numeric (10,6) | Fraction of POA assigned to SVOC\_bin3 |
| bin4 | Numeric (10,6) | Fraction of POA assigned to SVOC\_bin4 |

**Appendix C**

**RUN Schema File Formats**

**Table C1. tbl\_gas\_process.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| profile\_id | Character 20 | Profile identifier |
| process | Character 20 | Process mode |

**Table C2. tbl\_primary.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| aqminv\_poll | Character 20 | Inventory pollutant name |
| aqm\_add | Character 20 | Compound name to add |
| split\_factor | Numeric (12,8) | Split factor for added compound |
| writeflag | Character 1 | Flag: “N” – compound should be added  “Y” – compound should be replaced |

**Table C3. tbl\_run\_control.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| keyword | Character 20 | Key word identifying function or file name |
| datival | Character | Corresponding value for the key word |

**Table C4. tbl\_toxics.**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Field Type** | **Description** |
| aqm\_model | Character 20 | AQM name |
| specie\_id | Character 20 | Unique species identifier |
| aqm\_poll | Character20 | Model species identifier/name |
| num\_carbons\* | Numeric (6,3) | Number of carbons for model species |
| active\* | Character 1 | Options: A – active, T - tracer |

\* Not a user input; assigned by the system.

**Table C5. tbl\_user\_profile\_wts.**

|  |  |  |
| --- | --- | --- |
| Field Name | Field Type | Description |
| profile\_id | Character 20 | Profile identifier |
| specie\_id | Character 20 | Unique species identifier |
| percent | Numeric (10,6) | Percentage of pollutant in profile |

**Table C6. Table names in the run schema created by the Speciation Tool for internal calculations.**

|  |  |
| --- | --- |
| **Table Name** | **Description** |
| tmp\_actox | Checks for profiles with 100% active toxics. |
| tmp\_aqm\_carbons | Initialized with sharedtbl\_carbons. Records are added for user specified toxic species. |
| tmp\_calcs\_byaqm | Stores intermediate calculations by mechanism, profileId, and AQM pollutant. Includes moles per gram, moles model species per mole emissions, and average molecular weight. |
| tmp\_calcs\_byspc | Stores intermediate calculations by mechanism, profileId, specieId, and AQM pollutant. Includes moles per gram, moles model species per mole emissions, mole weight percent of model species, grams per mole, and average molecular weight. |
| tmp\_calcs\_haps | Stores intermediate calculations by mechanism, profileId, specieId, emission inventory pollutant, and AQM pollutant for the user specified HAPs species. Includes moles per mole, moles per gram, and average molecular weight. |
| tmp\_calcs\_haps\_null | Stores the calculated average molecular weights for those records in tmp\_calcs\_haps with undefined average molecular weights. |
| tmp\_camxpm\_splits | Stores PM split factors for CAMx AQM. |
| tmp\_error | Stores run error messages. |
| tmp\_gscnv | Stores the output converserion factor data. |
| tmp\_gspro | Stores the output splits factor data. |
| tmp\_haps | Stores the list of HAPs extracted from the INVTABLE. |
| tmp\_header | Stores the NONHAPTOG header entries. |
| tmp\_invtable | Used for checking output pollutant name widths specified in INVTABLE. |
| tmp\_mechanism | Initialized with the user selected mechanism from sharedtbl\_mechanism or tbl\_pm\_mechanism. Records are removed or inserted depending on run parameters. |
| tmp\_metadataset | Stores the metadata written to the header of the output files. |
| tmp\_prfwts | Stores the working profile weights |
| tmp\_pm\_mechanism | Mechanism definition for user specified PM mechanism run. |
| tmp\_pm\_splits | PM2.5 split factors. |
| tmp\_prfwts | Initialized with tmp\_raw\_profiles. Data records are removed and renormalized depending upon run parameters. |
| tmp\_profile\_list | Stores the unique list of PM profiles. |
| tmp\_qa\_carbons | Stores AQM compounds where no carbon data has been specified. |
| tmp\_qa\_mechanism | Stores species with no corresponding mechanism definition. |
| tmp\_raw\_profiles | Initialized with either shared.tbl\_gas\_profile\_weights, tbl\_pm\_profile\_weights, or tbl\_user\_profile\_wts if user provided. |
| tmp\_species | Stores the verified species list. |
| tmp\_species\_carbons | Stores the calculated number of carbons for each species. |
| tmp\_spcinp | Stores user specified species that have invalid molecular weights. |
| tmp\_sumnmog | Stores the sums of NMOG components |
| tmp\_sums | Stores the profile weights percent sums during renormalization. |
| tmp\_vbs\_svoc\_factors | Stores the SVOC factors for VBS processing |
| tmp2\_qa\_mechanism | Stores species for selected profiles with no mechanism definition. |

**Appendix D**

**PM2.5 Chemical Mechanism Definitions**

Table D1 represents the AE6 and AE8 chemical mechanism definitions. Table D2, prepared by EPA, provided the information used to define Table D1.

**Table D1. Chemical Mechanisms AE6 (or AE7) and AE8 as input to the Speciation Tool.**

| **Chemical Mechanism** | **SpecieID** | **AQM Pollutant** | **Qualify** | **Compute** |
| --- | --- | --- | --- | --- |
| **AE6** | | | | |
| AE6 | 626 | POC | F | F |
| AE6 | 797 | PEC | F | F |
| AE6 | 699 | PSO4 | F | F |
| AE6 | 613 | PNO3 | F | F |
| AE6 | 784 | PNH4 | F | F |
| AE6 | 2669 | PNCOM | T | F |
| AE6 | 488 | PFE | F | F |
| AE6 | 292 | PAL | F | F |
| AE6 | 694 | PSI | F | F |
| AE6 | 715 | PTI | F | F |
| AE6 | 2303 | PCA | F | F |
| AE6 | 2772 | PMG | F | F |
| AE6 | 2302 | PK | F | F |
| AE6 | 526 | PMN | F | F |
| AE6 | 785 | PNA | F | F |
| AE6 | 337 | PCL | F | F |
| AE6 | 2668 | PH2O | T | F |
| AE6 |  | PMOTHR | F | T |
| **AE8** | | | | |
| AE8 | 626 | PMOCN2 | F | F |
| AE8 | 797 | PEC | F | F |
| AE8 | 699 | PSO4 | F | F |
| AE8 | 613 | PNO3 | F | F |
| AE8 | 784 | PNH4 | F | F |
| AE8 | 2669 | PMNCOMN2 | T | F |
| AE8 | 488 | PFE | F | F |
| AE8 | 292 | PAL | F | F |
| AE8 | 694 | PSI | F | F |
| AE8 | 715 | PTI | F | F |
| AE8 | 2303 | PCA | F | F |
| AE8 | 2772 | PMG | F | F |
| AE8 | 2302 | PK | F | F |
| AE8 | 526 | PMN | F | F |
| AE8 | 785 | PNA | F | F |
| AE8 | 337 | PCL | F | F |
| AE8 | 2668 | PH2O | T | F |
| AE8 |  | PMOTHR | F | T |
| AE8 | 3245 | POCP2 | F | F |
| AE8 | 3246 | POCP1 | F | F |
| AE8 | 3247 | POCP0 | F | F |
| AE8 | 3248 | POCN1 | F | F |
| AE8 | 3249 | POCN2 | F | F |
| AE8 | 3250 | PNCOMP2 | F | F |
| AE8 | 3251 | PNCOMP1 | F | F |
| AE8 | 3252 | PNCOMP0 | F | F |
| AE8 | 3253 | PNCOMN1 | F | F |
| AE8 | 3254 | PNCOMN2 | F | F |

**Table D2. Chemical Mechanisms AE6**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SPECIATE**  **speciesID** | **Species Description** | **Species Name** | **Calculation** | **Notes** |
|  | CMAQ AE6: |  |  |  |
| 626 | organic carbon | POC | explicit | measured |
| 797 | elemental carbon | PEC | explicit | measured |
| 699 | sulfate | PSO4 | explicit | measured |
| 613 | nitrate | PNO3 | explicit | measured |
| 784 | ammonium | PNH4 | explicit | measured |
| 2669 | non-carbon organic matter | PNCOM | explicit | PNCOM = POC\*(OM/OC Ratio - 1)  where OM/OC ratio is 1.25 for motor vehicle exhaust, 1.7 for wood combustion, 1.4 for other sources. |
| 488 | iron | PFE | explicit | measured |
| 292 | aluminum | PAL | explicit | measured |
| 694 | silica | PSI | explicit | measured |
| 715 | titanium | PTI | explicit | measured |
| 2303 | calcium | PCA | explicit | measured |
| 2772 | magnesium | PMG | explicit | measured |
| 2302 | potassium | PK | explicit | measured |
| 526 | manganese | PMN | explicit | measured |
| 785 | sodium | PNA | explicit | measured |
| 337 | chloride | PCL | explicit | measured |
| 2668 | water | PH2O | explicit | 0.24\*(PNH4+PSO4) for non-combustion  sources, 0 for combustion sources or use measured value of hydrated water, if available |
|  | unspeciated PM2.5 | PMOTHR | 1 - (sum of 17 species) | n/a |

\* A note on atom/ion duplicate pair in the same profile: There are duplicate pairs of ions and atoms analysed with different methods in the same AE6 profile. If both ion and atom form are present in the same profile, the Speciation Tool preferably uses ion to populate corresponding AE6 species. This is a change from the previous version of Speciation Tool (v 4.0 and earlier) which used the atom. If an ion is not present and atom is present, then atom is used. If neither ion nor atom are present but oxide form of the element is present, then corresponding AE6 species is computed stoichiometrically from the oxide.

**Appendix E**

**Create AE6 Ready Profiles**

This section describes the procedure for creating AE6-ready profiles in the Speciation Tool by performing additional calculations on the measured or composited PM profiles data in SPECIATE. The creation of these profiles largely follows the approach in Reff et. al (2009)[[4]](#footnote-5).

SPECIATE houses different types of PM2.5 profiles: PM, PM-SIMPLIFIED, PM-AE6 and PM-VBS. Profiles of type PM contain all the species provided by a measurement study or a combination (composite) of measurement studies. The other PM profile types are profiles derived from a measurement study or composite in order to provide the species required for an air quality modeling aerosol mechanism. The PM-SIMPLIFIED profiles are for the AE5 aerosol mechanism and are computed by keeping only AE5 species (elemental carbon, organic carbon, sulfate and nitrate) and creating a PM Other species (PMFINE) calculated as the 100 – sum of AE5 species’ weight percent. The PM-AE6 and PM-VBS aerosol mechanisms have additional species not typically measured and that need to be computed. The PM-AE6 profiles include non-carbon organic mass (PNCOM) and/or water (PH2O). The PM-AE6 profiles were first put into the SPECIATE database in SPECIATE 4.3 (series 91XXX) by Reff, et.al. (2009). They were developed by compositing pre-existing SPECIATE profiles, computing PNCOM and PH2O and ensuring mass conservation. Documentation on the steps taken are provided in the supplemental information of Reff, et. al. (2009). Since that initial work, there have been additional PM2.5 speciation data published in the literature for which SPECIATE developers used Reff’s approach so that they could be used in the AE6 aerosol mechanism.

For AE6 profiles, the Speciation Tool takes the subset of SPECIATE profiles in which the additional species for AE6 were computed (i.e., “AE6-ready” profiles), maps the species IDs in the SPECIATE profile to the AE6 species names and creates “PM Other” (the remainder of the mass not assigned to AE6 species) as 100 – sum of AE6 species. Previous version of the Speciation Tool could generate SMOKE-ready profiles only from PM2.5 profiles that are ‘AE6-ready’. As indicated above, AE6-ready means that the PM2.5 profile must contain either PH2O or PNCOM, both of which are computed not measured. Therefore, when a new PM profile is added to SPECIATE, it was necessary to include both the original ‘raw’ PM profile and AE6-ready profile which must be manually calculated using methods described in Reff et al.

To automate this manual process, the Speciation Tool was enhanced to add capability to create AE6-ready profiles from a profile type of PM. The ability for the Speciation Tool to create AE6 profiles for the modeling allows SPECIATE developers to rely on the Speciation Tool to perform the calculations rather than having to do them manually and put the resulting AE6 profile into the SPECIATE database. The section below describes the procedure for creating AE6-ready profiles.

A list of steps programmed in the Speciation Tool to generate AE6-ready profiles are described below. These steps are based on the Reff et al. paper and include the steps of computing PH2O and PNCOM and adjusting other AE6 model species.

One issue that will need to be corrected in a subsequent version of the Speciation Tool is how the weight percentages are summed when reconstructing profile mass. The Speciation Tool currently sums the same form of the metal (i.e., ion) as are used in the AE6, but it should be using the atomic form. This should not have a significant impact, and it has no impact for profiles that only have ion or atom forms (but not both). It is likely also not to have a significant impact for profiles that have both ion and atom as the the difference in weight percents is not expected to be very large.

**Step 1 - Generate a list of PM2.5 profiles in the SPECIATE database that are not AE6-ready by scanning the PROFILE\_TYPE field.**

The PROFILES table of SPECIATE 5.0 database includes the PROFILE\_TYPE field to identify different PM profile types which indicates whether a profile is AE6-ready or not. The Speciation Tool scans this field to generate a list of PM profiles that are not AE6-ready.

**Step 2 – Calculate sulfate if only sulfur is present in the profile**

Many of the raw profiles contained a value for either SO4 or S, but not both. If a profile has SO4 and not S, the S does not need to be computed. However, if the profile has S but not SO4, then SO4 must be computed stoichiometrically from the S weight percentage as follows:

|  |  |  |
| --- | --- | --- |
|  |  | *(1)* |

The preference for using SO4 rather than S is explained by Reff, et al., section S3.7.5 as being due to higher accuracy of ion chromatography than XRF.

**Step 3 – Add particulate water, PH20. Note that this is SPECIES ID 2668 in SPECIATE***.*

The approach here is from the supplemental information from Reff, et al., section S3.7.1. Particulate water (PH2O) is zero for combustion and other high temperature sources because the water is likely to be emitted in the vapor phase. Table E-1 provides a list of profiles with zero PH2O emissions. The combustion and other higher temperature sources are identified based on the SPECIATE profile categorization fields.

SPECIATE contains three profile categorization fields in the PROFILES table to provide readily searchable metadata about the emission source covered by the profile. The fields describe the emission source in terms of emission generation mechanism (level 1), sector and/or equipment (level 2) and fuel and/or product (level 3). The level 1 field “CATEGORY\_LEVEL\_1\_Generation\_Mechanism” identifies the mechanism by which emissions are generated by the emissions source. There are 9 options for this field: Ash, Atomization, Background-air, Chemical Reaction, Combustion, Dust, Microbial, Miscellaneous, and Volatilization. The Speciation Tool scans the category level 1 field to identify combustion profiles and set them to zero PH2O emissions.

**Table E-1. List of PM2.5 profiles where PH2O need to be set to zero**

| **Type of Source** | **Particulate Water (PH2O) calculation** |
| --- | --- |
| CATEGORY\_LEVEL\_1\_Generation\_Mechanism = “Combustion” identifies combustion and other high temperature sources, where water is likely to be emitted in the vapor phase | 0 |
| All other sources | 24% of the sum of sulfate (PSO4) and ammonium (PNH4) concentrations or percentages |

For all other profiles, PH2O is calculated as 24% of the sum of sulfate (SO4) and ammonium (NH4)concentrations or percentages.

|  |  |  |
| --- | --- | --- |
|  |  | *(2)* |

**Step 3 – Add Metal Bound Oxygen, MO. Note that this is SPECIES ID 2670 in SPECIATE.**

While MO is not an AE6 species, it needs to be computed and included in the profile (unless it is 0) to enable a check for total mass fraction <= 100%. The approach to compute MO follows Section S.3.7.2 in from Reff, et al., which is to stoichiometrically combine oxygen with the metals, and then adjust the MO downward based on the amount of available in the profile. This approach assumes that SO4 binds to the metals preferentially over the oxygen.

Unadjusted MO is computed as

|  |  |  |
| --- | --- | --- |
|  |  | *(3)* |

where OxEl is the oxygen-to-metal ratio for metal El as shown Table E-2, and EEl is the emission of metal El, **except for Na, Ca, Mg and K**. For these 4 metals, the EEl should reflect the difference between the atom form of the metal and the ion form. If, for Na, Ca, Mg, and K, the profile has only one form (atom or ion but not both) then the EEl should be set to 0. Also, if the difference is negative, it should be set to 0. Note that for metal oxides with multiple forms an average oxygen to metal ratio across all forms is used.

**Table E-2. Assumed oxide forms of each metal and resulting mean oxygen-to-metal ratio used in Eq**Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.**uation 1**

| **Species** | **MW of metal1** | **Oxide Form 1** | **Oxide Form 2** | **Oxide Form 3** | **oxygen/metal ratio** |
| --- | --- | --- | --- | --- | --- |
| Na | 22.99 | Na2O |  |  | 0.348 |
| Mg | 24.31 | MgO |  |  | 0.658 |
| Al | 26.98 | Al2O3 |  |  | 0.889 |
| Si | 28.09 | SiO2 |  |  | 1.139 |
| P | 30.97 | P2O3 | P2O5 |  | 1.033 |
| K | 39.10 | K20 |  |  | 0.205 |
| Ca | 40.08 | CaO |  |  | 0.399 |
| Ti | 47.87 | TiO2 |  |  | 0.669 |
| V | 50.94 | V2O5 |  |  | 0.785 |
| Cr | 52.00 | Cr2O3 | CrO3 |  | 0.692 |
| Mn | 54.94 | MnO | MnO2 | Mn2O7 | 0.631 |
| Fe | 55.85 | FeO | Fe2O3 |  | 0.358 |
| Co | 58.93 | CoO | Co2O3 |  | 0.339 |
| Ni | 58.69 | NiO |  |  | 0.273 |
| Cu | 63.55 | CuO |  |  | 0.252 |
| Zn | 65.39 | ZnO |  |  | 0.245 |
| Ga | 69.72 | Ga2O3 |  |  | 0.344 |
| As | 74.92 | As203 | As2O5 |  | 0.427 |
| Se | 78.96 | SeO | SeO2 | SeO3 | 0.405 |
| Rb | 85.47 | Rb2O |  |  | 0.094 |
| Sr | 87.62 | SrO |  |  | 0.183 |
| Zr | 91.22 | ZrO2 |  |  | 0.351 |
| Mo | 95.94 | MoO2 | MoO3 |  | 0.417 |
| Pd | 106.42 | PdO | PdO2 |  | 0.226 |
| Ag | 107.87 | Ag2O |  |  | 0.074 |
| Cd | 112.41 | CdO |  |  | 0.142 |
| In | 114.82 | In2O3 |  |  | 0.209 |
| Sn | 118.71 | SnO | SnO2 |  | 0.202 |
| Sb | 121.76 | Sb2O3 | Sb2O5 |  | 0.263 |
| Ba | 137.33 | BaO |  |  | 0.117 |
| La | 138.91 | La2O3 |  |  | 0.173 |
| Ce | 140.12 | Ce2O3 | CeO2 |  | 0.200 |
| Hg | 200.59 | Hg2O | HgO |  | 0.060 |
| Pb | 207.20 | PbO | PbO2 |  | 0.116 |

To adjust MO based on preferential combining of SO4 over oxygen, compute the available SO4 for binding with metals, which is the SO4 remaining after fully neutralizing the NH4 in the profile.

|  |  |  |
| --- | --- | --- |
|  |  | *(4)* |

where is the mass of NH4in the profile.

The non-neutralized SO4 is the remainder from the SO4 in the profile.

|  |  |  |
| --- | --- | --- |
|  |  | *(5)* |

If

|  |  |  |
| --- | --- | --- |
|  |  | *(6)* |

If

|  |  |  |
| --- | --- | --- |
|  |  | *(7)* |

If is negative due to an excess of , set equal to zero because all the available sulfate would preferentially bond with metals, and there won’t be any metal bound oxygen (MO).

**Step 4 – Add Particulate Non-Carbon Organic Matter (PNCOM)**

Every profile that has particulate organic carbon (POC) must have PNCOM computed from POC. We use the OM/OC (OM = organic matter) ratios provided in the section S.3.7.3 of Reff, et al. These values are provided in the box below for three types of sources.

For on-road and non-road motor vehicle exhaust profiles (e.g., the HDDV Exhaust, Nonroad Gasoline Exhaust, On-road Gasoline Exhaust, and LDDV Exhaust source categories):

|  |  |  |
| --- | --- | --- |
|  |  | *(8)* |

For wood combustion sources other than wood-fired boilers (e.g., wildfires, agricultural burning, residential wood combustion, prescribed burning, slash burning):

|  |  |  |
| --- | --- | --- |
|  |  | *(9)* |

For wood-fired boilers and all other sources

|  |  |  |
| --- | --- | --- |
|  |  | *(10)* |

The PNCOM calculation requires assigning each PM profile to one of the above three source categories to apply correct OM/OC ratio. The Speciation Tool uses the SPECIATE profile categorization fields to assign each profile to the above three source categories. Equation 8 is applied to onroad/nonroad exhaust profiles which are identified using CATEGORY\_LEVEL\_1= “Combustion” and CATEGORY\_LEVEL\_2 contains “Mobile”. Equation 9 applies to wood combustion sources except boilers which are identified using CATEGORY\_LEVEL\_2 contains “biomass burning” and does not contain “boiler”. Equation 10 applies to all other profiles not covered by Equation 8 and 9.

**Step 5 – Check for Sum of PM2.5 Weight Fractions over 100%**

No adjustments need to be made if the weight fraction is less than 101%. If the mass is over 100% then:

* Assume there are POC artifacts in the profile. Adjust POC and PNCOM down by the same multiplier until the sum of weight fractions is 100%

If POC and PNCOM adjustment will not generate a profile with 100% weight percent sum then adjust all species down (i.e., normalize all weight percent) to get the sum to be 100%.

After the calculation of PH2O, MO and PNCOM, the Speciation Tool sums weight fraction of the reconstructed mass making sure that there is no double counted species. For example, particulate PAHs double count PNCOM. There is no indication in SPECIATE 5.0 of which species may be double counted with PNCOM. The Speciation Tool uses a list of species, shown in the Table E-3, to sum which is based on the AE6 species (including gap-filled species as indicated in the last column of the Table E-3) list plus other metal species that are not explicit in AE6 but may be present in the profile. The other metal species come from Table E-2. As noted earlier, the species chosen for the summation are the same as the AE6 but actually should not be in the case of species that have both atom and ion forms. For those, the atom would give a more complete weight percent, so if both atom and ion are present, the atom should be used. This will be changed in a subsequent update of the Speciation Tool.

**Table E-3. Species to include from sum of PM2.5 mass**

| **Species Name** | **Species ID** | **Species Description, Chemical Formula** | **Comments/Updated Mappings From** |
| --- | --- | --- | --- |
| POC | 626 | Organic carbon | This is obtained from the measurement study, but the weight percent may need to be adjusted downward when creating an AE6 profile if the sum of the species’ weight percents exceed 100. The adjustment assumes that the POC included measurement artifacts and is adjusted to achieve mass conservation. |
| PEC | 797 | Elemental carbon |  |
| PSO4 | 699 | Sulfate, SO42- | Gap filling procedure: If the profile has sulfur (species ID = 700) but no sulfate, then compute sulfate stoichiometrically (SO42-=96/32\*S) |
| PNO3 | 613 | Nitrate, NO3- |  |
| PNH4 | 784 | Ammonium, NH4+ |  |
| PNCOM | 2669 | non-carbon organic matter | Computed from OC based on the (OM to OC ratio) which is a function of the source characteristics and is based on the Reff et al. (2009) default assignments:  Mobile exhaust (combustion): 1.25  Wood combustion sources except wood fired boilers: 1.7  All other sources including wood fired boilers: 1.4  If a particular study uses a different ratio than the default (e.g., the Kansas City study profiles use 1.2 instead of 1.25), then that ratio would be used in place of the default. |
| PFE | 488 | Iron |  |
| PAL | 292 | Aluminum |  |
| PSI | 694 | Silicon |  |
| PTI | 715 | Titanium |  |
| PCA | 2303 | Calcium ion Ca2+ | This is a change from the Speciation Tool version 4.0 which used the atom (329). Gap filling procedure:  If Species ID 2303 isn’t present and Calcium (Species ID 329) is present, then use Species ID 329.  If neither Species ID 2303 nor Species ID 329 are present but calcium oxide (CaO Species ID 2847) is present, then Ca2+ = 40/56\* CaO. |
| PMG | 2772 | Magnesium ion Mg2+ | This is a change from the Speciation Tool version 4.0 which used the Magnesium atom (Species ID 525). Gap filling procedure:  If Species ID 2772 isn’t present and Magnesium atom (Species ID 525) is present, then use Species ID 525.  If neither Species ID 2772 nor Species ID 525 are present but Magnesium Oxide, MgO (2852) is present, then Mg2+= 24/40\* MgO |
| PK | 2302 | Potassium ion K+ | This is a change from the Speciation Tool version 4.0 which used the Potassium atom (Species ID 669). Gap filling procedure: If Species ID 2302 isn’t present and Potassium atom (Species ID 669) is present, then use Potassium atom (Species ID 669). |
| PMN | 526 | Manganese |  |
| PNA | 785 | Sodium ion Na+ | This is a change from the Speciation Tool version 4.0 which used the Sodium atom (Species ID 696). Gap filling procedure: If Sodium ion (Species ID 785) isn’t present and Sodium atom (Species ID 696) is present, then use Sodium atom (Species ID 696). |
| PCL | 337 | Chloride ion | This is a change from the Speciation Tool version 4.0 which used the Chloride atom (Species ID 795). Gap filling procedure: If Chloride ion (Species ID 337) isn’t present and Chlorine atom (Species ID 795) is present, then use Chlorine atom (Species ID 795). |
| PH2O | 2668 | Water | Computed for non-combustion and non-high temperature sources |
| PMO | 2671 | PM2.5 not in other AE6 species | Optional for PM-AE6 profile in SPECIATE but computed in the Speciation Tool. (can compute or leave out). Computed from 100-sum of other species. |
| P | 666 | phosphorous |  |
| V | 767 | vanadium |  |
| Cr | 347 | chromium |  |
| Co | 379 | cobalt |  |
| Ni | 612 | nickel |  |
| Cu | 380 | copper |  |
| Zn | 778 | zinc |  |
| Ga | 468 | Gallium |  |
| As | 298 | Arsenic |  |
| Se | 693 | selenium |  |
| Rb | 689 | Rubidium |  |
| Sr | 697 | Strontium |  |
| Zr | 779 | Zirconium |  |
| Mo | 586 | Molybdenum |  |
| Pd | 649 | Palladium |  |
| Ag | 695 | Silver |  |
| Cd | 328 | Cadmium |  |
| In | 487 | Indium |  |
| Sn | 714 | Tin |  |
| Sb | 296 | Antimony |  |
| Ba | 300 | Barium |  |
| La | 519 | Lanthanum |  |
| Ce | 1861 | Cerium |  |
| Hg | 528 | Mercury |  |
| Pb | 520 | Lead |  |

**Appendix F**

**Mechanism Mappings for CMAQ (AE7 and AE8) and CAMx (cf2)**

1. The tracer mechanism is basically:

   Species 279=ALD2\_PRIMARY

   Species 465= FORM\_PRIMARY

   Many other species are mapped to SOAALK

   All other species are mapped to NONBAF [↑](#footnote-ref-2)
2. Pye, H. O., & Pouliot, G. A. (2012). Modeling the role of alkanes, polycyclic aromatic hydrocarbons, and their oligomers in secondary organic aerosol formation. Environmental science & technology, 46(11), 6041-6047. [↑](#footnote-ref-3)
3. Jathar, S. H. et al. Unspeciated organic emissions from combustion sources and their influence on the secondary organic aerosol budget in the United States. Proc Natl Acad Sci. USA 111, 10473–10478 (2014). [↑](#footnote-ref-4)
4. “Emissions Inventory of PM2.5 Trace Elements across the United States”; Adam Reff, Prakash V. Bhave, Heather Simon, Thompson G. Pace, George A. Pouliot, J. David Mobley, and Marc Houyoux; Environmental Science & Technology 2009 43 (15), 5790-5796; DOI: 10.1021/es802930x (Supplemental Information) [↑](#footnote-ref-5)