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

Speciation Tool User's Guide Version 5.0

RAMBOLL Bright ideas. Sustainable change.

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.

Ramboll
7250 Redwood Boulevard
Suite 105
Novato, CA 94945
USA

T +1 415 899 0700
https://ramboll.com

CONTENTS

APPENDICES

Table of Figures

Table of Tables

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

1.0 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:

$$NO2 \rightarrow NO + O$$
$$O + O2 \rightarrow O3$$
$$O3 + NO \rightarrow NO2$$

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, $PM_{2.5}$) to model species (e.g., formaldehyde as part of VOC, elemental carbon as part of $PM_{2.5}$)
- 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.

1.1 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)

2.0 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:

$$E_i = IP_k \times \sum_j SP_{k,j} \times MM_{j,i} \qquad (1)$$

$where$:
E_i $is\ the\ emission\ rate\ of$ $model\ species\ i$
$IP_k is\ the\ emission\ rate\ of$ $inventory\ pollutant\ k$
$SP_{k,j} is\ the\ speciation\ profile\ of$ $chemical\ consituents\ j\ for$ $inventory\ pollutant\ k$
$MM_{j,i} is\ the\ mechanism\ mapping\ from$ $chemical\ j\ to\ model\ species\ i$

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 ($PM_{2.5}$). 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.

2.1 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 ($c^* < 10^{2.5}$ µg m^{-3}) or IVOC ($10^{2.5} < c^* < 10^{6.5}$ 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

Mechanism Mapping Name	Purpose	Comments
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.

2.2 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 default data 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 $PM_{2.5}$ 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.

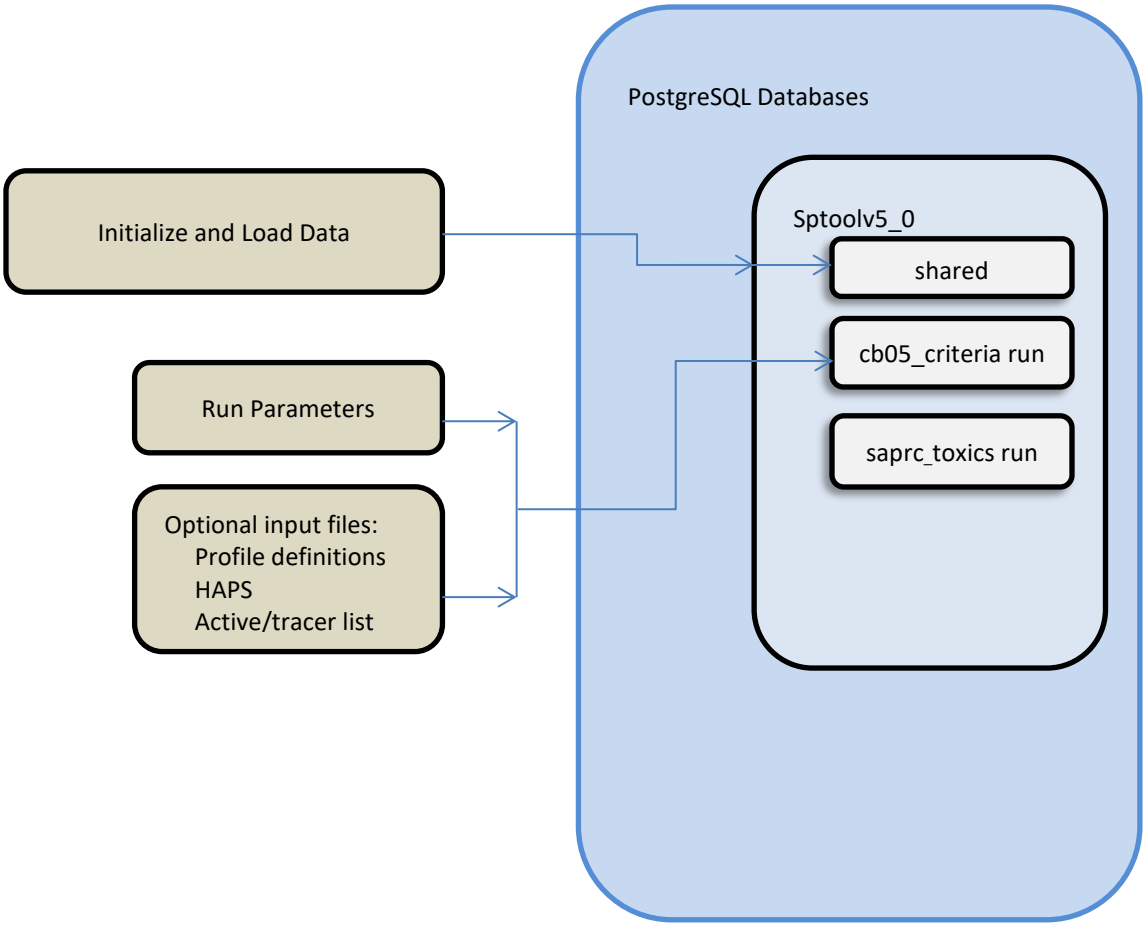


Exhibit 1. Speciation Tool Initialization and Run schemas.

3.0 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
    a. Run the initialization script Change to the run directory; cd /run
    b. 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.

3.1 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.
    a. Change to the Speciation Tool directory.
    b. Copy the Speciation Tool package to this directory.
    c. Extract and uncompress the zipped file.
3. Verify required software packages are available. See Appendix A for software packages.
    a. Run *./spool_reqd_checks.sh*
4. Set environment variables.
    a. Edit *Assigns.sptool*; set SPTOOL_HOME.
    b. source Assigns.sptool
5. Initialize the Speciation Tool database.
    a. Execute *./init_sptooldb_v5.0.csh*
6. Execute the provided test case.
    a. Change to the run directory; cd /run
    b. Run *./test_all.job*
7. Review and compare results.
    a. Change to output directory; cd ../output
    b. Compare results to provided test case files.

3.2 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.

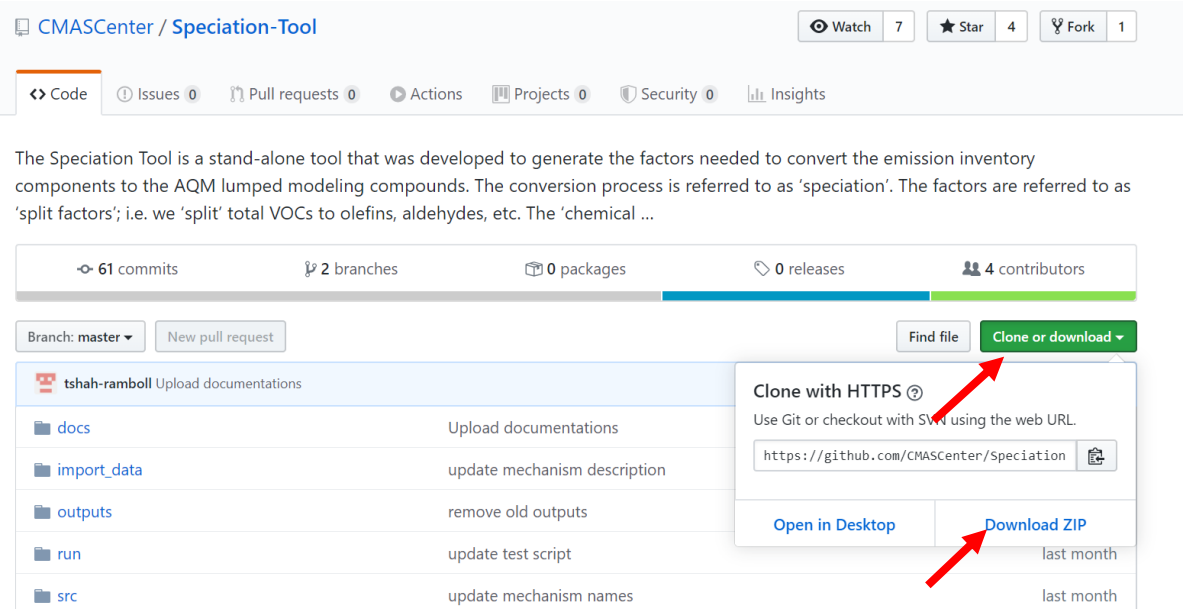


Exhibit 2. Downloading Speciation Tool from GitHub.

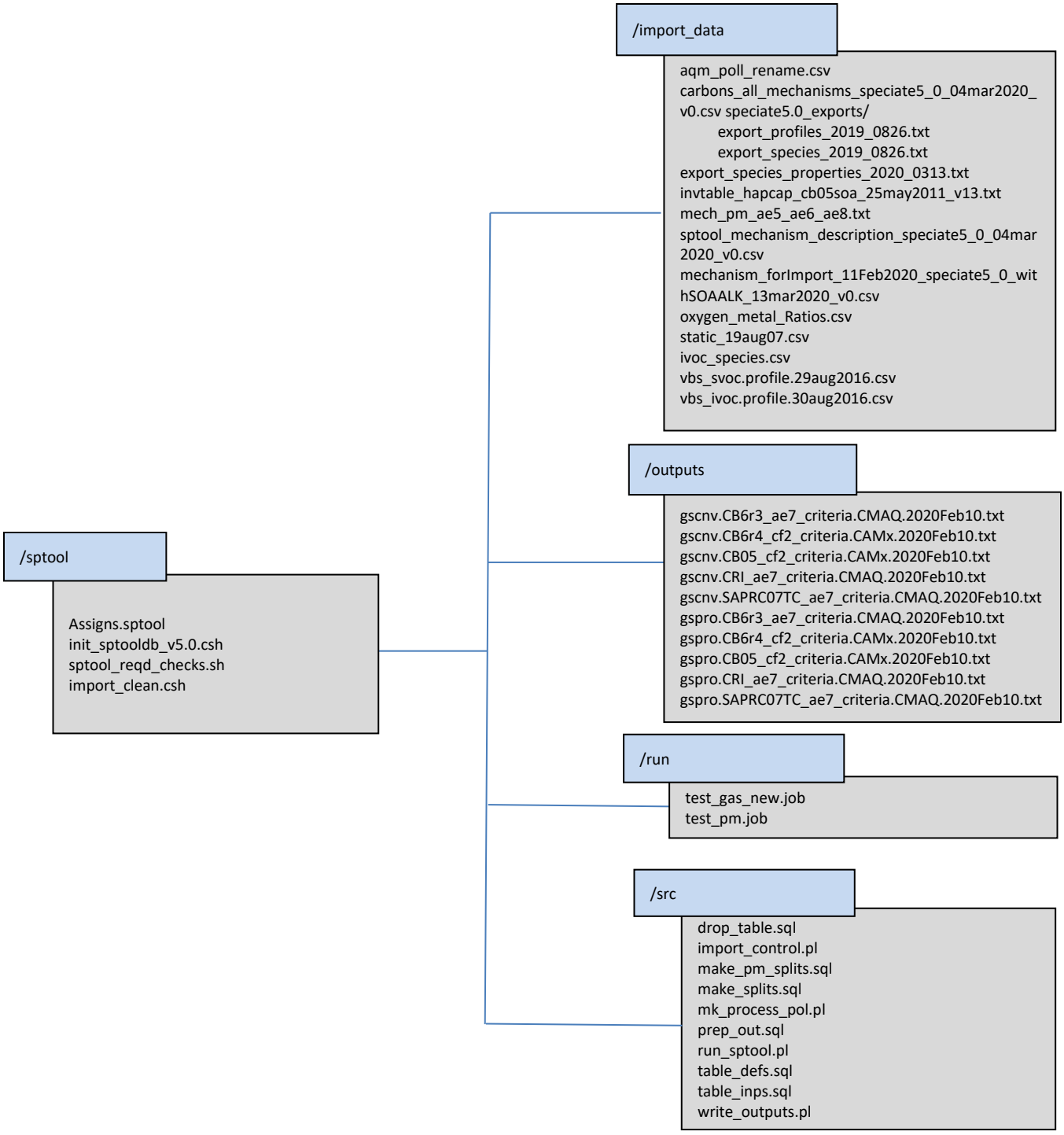


Exhibit 3. Speciation Tool Directory and File Names.

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

Folder	Files	Description
/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 $PM_{2.5}$ 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

3.3 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.
```

3.4 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.

3.4.1 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
```

3.4.2 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.

3.4.3 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.

3.4.4 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	$PM_{2.5}$ 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.

4.0 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.

4.1 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.

4.2 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'

        nspname          | nspowner |                nspacl
-------------------------+----------+---------------------------------------
pg_toast                 |        1 |
shared                   |      103 | {yshi=UC/yshi,=UC/yshi}
pg_temp_1                |        1 |
pg_catalog               |        1 | {postgres=UC/postgres,=U/postgres}
public                   |        1 | {postgres=UC/postgres,=UC/postgres}
information_schema       |        1 | {postgres=UC/postgres,=U/postgres}
camx_cb6r4_cf2_criteria  |      103 | {yshi=UC/yshi,=UC/yshi}
cmaq_cb6r3_ae7_criteria  |      103 | {yshi=UC/yshi,=UC/yshi}
```

4.3 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

Keyword	Description
Run Parameters	
	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.

4.3.1 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 $PM_{2.5}$ 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 $PM_{2.5}$ 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 $PM_{2.5}$ 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' $PM_{2.5}$ 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.

4.3.2 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.

4.3.3 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.

4.3.3.1 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

Metadata Keyword	Example
#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.

4.4 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.

4.4.1 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
vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv
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
vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv
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
vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv
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
```

4.4.2 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 $PM_{2.5}$ 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 $PM_{2.5}$ 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.

5.0 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 $PM_{2.5}$ speciation profiles.

5.1 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] 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] provides further details on the significance of these compounds.

5.1.1 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

[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

[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.

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.

5.1.1.1 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.

5.1.1.2 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
```

5.1.1.3 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.

5.1.1.4 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.

5.1.1.5 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 10^3 to 10^6 µg/m^3. 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] 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.

5.1.2 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.

[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).

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 HAPs[1]	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 NOINTEGRATE[1]	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).

5.1.3 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.

5.2 $PM_{2.5}$ Profile Processing

The Speciation Tool provides speciation profile entries (GSPRO) for $PM_{2.5}$. 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 $PM_{2.5}$ 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 $PM_{2.5}$ 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.

5.2.1 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. $PM_{2.5}$ 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.

5.2.2 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:

1. 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.
2. 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.
3. 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	CAMx CMAQ Model Species	Saturation Concentration Bin (µg/m3) 0 bin_0	1 bin_1	10 bin_2	100 bin_3	1000 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 CAMx	Model Species CMAQ	IVOC/NMOG
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

6.0 FAQ

6.1 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.

6.2 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.

6.3 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 $PM_{2.5}$ 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.

6.3.1 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".

6.4 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).

6.5 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
- Linux
  - RedHat/CentOS/Fedora/Scientific families Linux
  - Debian GNU/Linux
  - Ubuntu 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

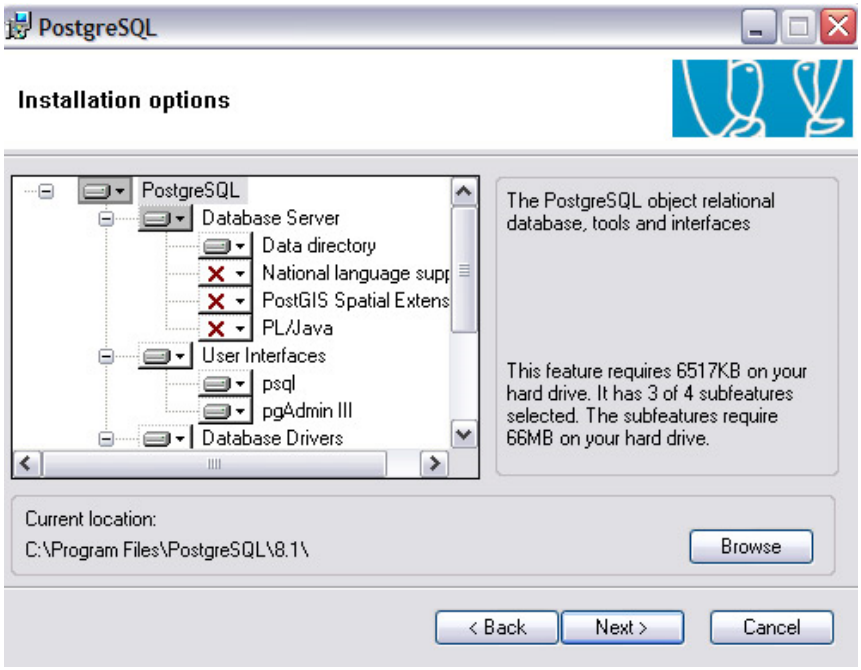
   convert drive_letter: /fs:ntfs.

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.

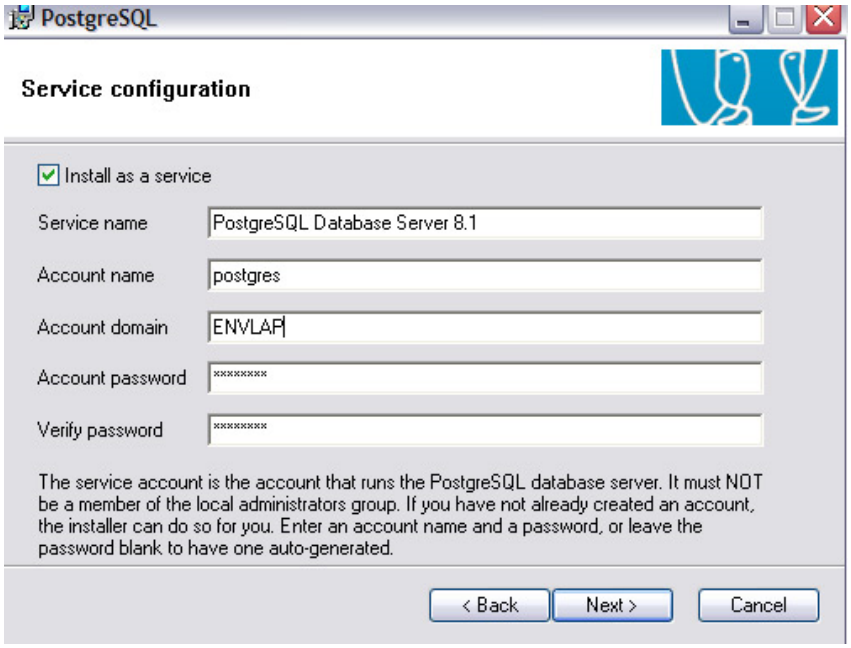
Install PostgreSQL using the Installation Wizard

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).

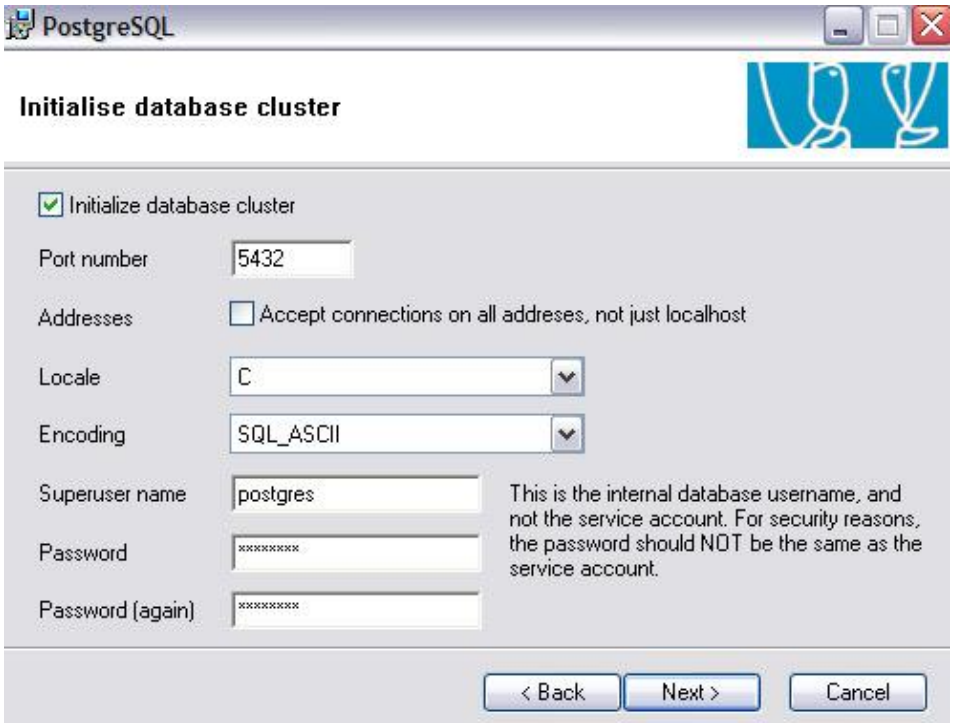
1. Unzip the PostgreSQL software that you downloaded.
2. Execute (double-click on) postgresql-8.1.msi. This will start the installation wizard.
3. Select the language you want to use for the installer and click Start.
4. Click Next to continue the installation following the Welcome message.
5. Click 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.



7. 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?"

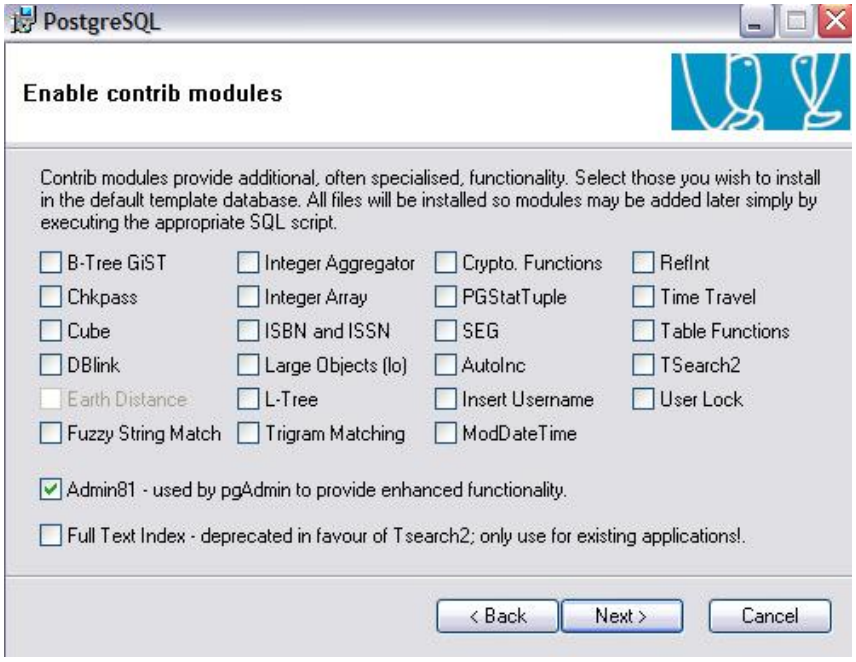


8. In the Initialize database cluster window provide the superuser name postgres and a password. Click Next to continue.



9. 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.

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



11. 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

4. Open pg_hba.conf file (text format) in the database's directory.
5. 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
```

6. 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/m^3)
o_loading	Numeric (10,2)	Organic loading (ug/m^3)
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 shared tbl_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 shared tbl_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	$PM_{2.5}$ 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 $PM_{2.5}$	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].

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 $PM_{2.5}$ 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.

[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)

Step 2 – Calculate sulfate if only sulfur is present in the profile
Many of the raw profiles contained a value for either SO_4 or S, but not both. If a profile has SO_4 and not S, the S does not need to be computed. However, if the profile has S but not SO_4, then SO_4 must be computed stoichiometrically from the S weight percentage as follows:

$$SO_4 = \left(\frac{96}{32}\right) * S \qquad (1)$$

The preference for using SO_4 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 (SO_4) and ammonium (NH_4) concentrations or percentages.

$$PH2O = 0.24 * (PSO4 + PNH4) \qquad (2)$$

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 SO_4 in the profile. This approach assumes that SO_4 binds to the metals preferentially over the oxygen.

Unadjusted MO is computed as

$$MO_{unadjusted} = \sum_{El}^{N} Ox_{El} \times E_{El} \qquad (3)$$

where Ox_{El} is the oxygen-to-metal ratio for metal El as shown Table E-2, and E_{El} 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 metal[1]	Oxide Form 1	Oxide Form 2	Oxide Form 3	oxygen/ metal ratio
Na	22.99	Na_2O			0.348
Mg	24.31	MgO			0.658
Al	26.98	Al_2O_3			0.889
Si	28.09	SiO_2			1.139
P	30.97	P_2O_3	P_2O_5		1.033
K	39.10	K_20			0.205
Ca	40.08	CaO			0.399
Ti	47.87	TiO2			0.669
V	50.94	V_2O_5			0.785
Cr	52.00	Cr_2O_3	CrO_3		0.692
Mn	54.94	MnO	MnO_2	Mn_2O_7	0.631
Fe	55.85	FeO	Fe_2O_3		0.358
Co	58.93	CoO	Co_2O_3		0.339
Ni	58.69	NiO			0.273
Cu	63.55	CuO			0.252
Zn	65.39	ZnO			0.245
Ga	69.72	Ga_2O_3			0.344
As	74.92	As_20_3	As_2O_5		0.427
Se	78.96	SeO	SeO_2	SeO_3	0.405
Rb	85.47	Rb_2O			0.094
Sr	87.62	SrO			0.183
Zr	91.22	ZrO_2			0.351
Mo	95.94	MoO_2	MoO_3		0.417
Pd	106.42	PdO	PdO_2		0.226
Ag	107.87	Ag_2O			0.074
Cd	112.41	CdO			0.142
In	114.82	In_2O_3			0.209
Sn	118.71	SnO	SnO_2		0.202
Sb	121.76	Sb_2O_3	Sb_2O_5		0.263
Ba	137.33	BaO			0.117
La	138.91	La_2O_3			0.173
Ce	140.12	Ce_2O_3	CeO_2		0.200
Hg	200.59	Hg2O	HgO		0.060
Pb	207.20	PbO	PbO_2		0.116

To adjust MO based on preferential combining of SO_4 over oxygen, compute the available SO_4 for binding with metals, which is the SO_4 remaining after fully neutralizing the NH_4 in the profile.

$$Neutralized\ SO_4^{2-} = \frac{0.5 * 96}{18} \times\ E_{NH_4^+} \qquad (4)$$

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

The non-neutralized SO_4 is the remainder from the SO_4 in the profile.

$$Non\_Neutralized\_SO_4^{2-} = E_{SO_4^{2-}} - Neutralized\ SO_4^{2-} \qquad (5)$$

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

$$MO_{adjusted} = MO_{unadjusted} \qquad (6)$$

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

$$MO_{adjusted} = MO_{unadjusted} - Non\_Neutralized\ SO_4^{2-} \times \frac{16}{96} \qquad (7)$$

If $MO_{adjusted}$ is negative due to an excess of $Non\_Neutralized\ SO_4^{2-}$, set $MO_{adjusted}$ 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):

$$PNCOM = 0.25 * POC \qquad (8)$$

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

$$PNCOM = 0.7 * POC \qquad (9)$$

For wood-fired boilers and all other sources

$$PNCOM = 0.4 * POC \qquad (10)$$

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, SO_4^{2-}	Gap filling procedure: If the profile has sulfur (species ID = 700) but no sulfate, then compute sulfate stoichiometrically (SO_4^{2-}=96/32*S)
PNO3	613	Nitrate, NO_3^-	
PNH4	784	Ammonium, NH_4^+	
PNCOM	2669	non-carbon organic matter	Computed from OC based on the (OM to OC ratio) which is a function of the source characteristics and is based on the Reff et al. (2009) default assignments: Mobile exhaust (combustion): 1.25 Wood combustion sources except wood fired boilers: 1.7 All other sources including wood fired boilers: 1.4 If a particular study uses a different ratio than the default (e.g., the Kansas City study profiles use 1.2 instead of 1.25), then that ratio would be used in place of the default.
PFE	488	Iron	
PAL	292	Aluminum	
PSI	694	Silicon	
PTI	715	Titanium	

Species Name	Species ID	Species Description, Chemical Formula	Comments/Updated Mappings From
PCA	2303	Calcium ion Ca^{2+}	This is a change from the Speciation Tool version 4.0 which used the atom (329). Gap filling procedure: If Species ID 2303 isn't present and Calcium (Species ID 329) is present, then use Species ID 329. If neither Species ID 2303 nor Species ID 329 are present but calcium oxide (CaO Species ID 2847) is present, then Ca^{2+} = 40/56* CaO.
PMG	2772	Magnesium ion Mg^{2+}	This is a change from the Speciation Tool version 4.0 which used the Magnesium atom (Species ID 525). Gap filling procedure: If Species ID 2772 isn't present and Magnesium atom (Species ID 525) is present, then use Species ID 525. If neither Species ID 2772 nor Species ID 525 are present but Magnesium Oxide, MgO (2852) is present, then Mg^{2+}= 24/40* MgO
PK	2302	Potassium ion K^{+}	This is a change from the Speciation Tool version 4.0 which used the Potassium atom (Species ID 669). Gap filling procedure: If Species ID 2302 isn't present and Potassium atom (Species ID 669) is present, then use Potassium atom (Species ID 669).
PMN	526	Manganese	

Species Name	Species ID	Species Description, Chemical Formula	Comments/Updated Mappings From
PNA	785	Sodium ion Na^+	This is a change from the Speciation Tool version 4.0 which used the Sodium atom (Species ID 696). Gap filling procedure: If Sodium ion (Species ID 785) isn't present and Sodium atom (Species ID 696) is present, then use Sodium atom (Species ID 696).
PCL	337	Chloride ion	This is a change from the Speciation Tool version 4.0 which used the Chloride atom (Species ID 795). Gap filling procedure: If Chloride ion (Species ID 337) isn't present and Chlorine atom (Species ID 795) is present, then use Chlorine atom (Species ID 795).
PH2O	2668	Water	Computed for non-combustion and non-high temperature sources
PMO	2671	$PM_{2.5}$ not in other AE6 species	Optional for PM-AE6 profile in SPECIATE but computed in the Speciation Tool. (can compute or leave out). Computed from 100-sum of other species.
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	

Species Name	Species ID	Species Description, Chemical Formula	Comments/Updated Mappings From
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)

MEMO

To **Alison Eyth, Madeleine Strum and Benjamin Murphy, EPA**

From **Ross Beardsley, Tejas Shah and Greg Yarwood**

Subject **Speciation Tool v5.0 Mechanism Mappings for CMAQ (AE7 and AE8) and CAMx (cf2)**

Attachments:

A. Memo: Mapping SPECIATE 5.0 compounds for photochemical models

B. Memo: Classification of SPECIATE 4.5 compounds as a function of estimated vapor pressure

C. Memo: Species Mappings for CB6 and CB05 for use with SPECIATE 4.5

Summary

Ramboll created new mechanism mappings[1] for the Speciation Tool v5.0 to accommodate recent changes in the gas speciation profiles of the EPA's SPECIATE 5.0 repository and specifically support the revised aerosol modules (AE7 and AE8) of the Community Multiscale Air Quality Model v5.3 (CMAQ). CMAQ mappings were created for the Carbon Bond 6 (CB6), SAPRC07TC, and the Common Representative Intermediates (CRI)[2] gas phase mechanisms. Minor revisions to mappings for the Comprehensive Air Quality Model with Extensions (CAMx)[3] mechanisms CB05, CB6 and SAPRC07 provide more consistent treatment of low volatility organic compounds (LVOC) across models. Detailed descriptions of the methods used to develop the mechanism mappings for each chemical mechanism were provided previously[4] and are included here as attachments for completeness.

The names of mechanism mappings are revised to more transparently indicate 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).

A summary of the mechanism mappings in Speciation Tool v5.0 is provided in **Table 1.** The mechanism implementation details are available in respective model User's Guides. For example, CMAQ chemical mechanisms are described on CMAQ GitHub[5] and CAMx mechanism details are available in CAMx User's Guide[3]. **Table 2** lists chemical species names with their description common to all CB6 mechanisms. Similarly, **Table 3** and **Table 4** presents list of chemical species for SAPRC07 and CRI mechanisms respectively.

March 24, 2020

Ramboll
7250 Redwood Boulevard
Suite 105
Novato, CA 94945
USA

T +1 415 899 0700
F +1 415 899 0707
www.ramboll.com

[1] "Mechanism mapping" refers to the translating the detailed chemical constituents contained in the gas speciation profiles of air pollution sources in EPA's SPECIATE 5.0 repository to the "model species" used by condensed chemical mechanisms of photochemical air quality models. Mechanism mapping is implemented in software called the "Speciation Tool."

[2] http://mcm.leeds.ac.uk/CRI/

[3] http://www.camx.com/

[4] Ramboll technical memo to EPA titled "Mapping SPECIATE 5.0 compounds for photochemical models" dated October 11, 2018.

[5] https://github.com/USEPA/CMAQ/tree/master/CCTM/src/MECHS

Table 1. Summary of mechanism mappings in Speciation Tool v5.0

Mechanism Mapping Name	Purpose	Summary Description
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 ($c^* < 10^{2.5}$ µg m^{-3}) or IVOC ($c^* < 10^{6.5}$ 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) 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 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
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

Table 2. Species names, descriptions and carbon numbers common to all CB6 Mechanisms

Species Name	Description	Carbon #
ACET	Acetone	3
ALD2	Acetaldehyde	2
ALDX	Propionaldehyde and higher aldehydes	2
BENZ	Benzene	6
CH4	Methane	1

Species Name	Description	Carbon #
ETH	Ethene	2
ETHA	Ethane	2
ETHY	Ethyne	2
ETOH	Ethanol	2
FORM	Formaldehyde	1
IOLE	Internal olefin carbon bond (R-C=C-R)	4
ISOP	Isoprene	5
KET	Ketone carbon bond (C=O)	1
MEOH	Methanol	1
OLE	Terminal olefin carbon bond (R-C=C)	2
PAR	Paraffin carbon bond (C-C)	1
PRPA	Propane	3
TERP	Monoterpenes	10
TOL	Toluene and other monoalkyl aromatics	7
XYL	Xylene and other polyalkyl aromatics	8

Table 3. Species names, descriptions and carbon numbers common to all SAPRC07 Mechanisms

Species Name	Description	Carbon #
ACET	Acetone	3
ACYE	Acetylene (ethyne)	2
ALK1	Alkanes and other non-aromatic compounds that react only with OH and have kOH (OH radical rate constant) between 2 and 5 x 10^2 ppm^{-1} min^{-1} (primarily ethane)	2
ALK2	Alkanes and other non-aromatic compounds that react only with OH and have kOH between 5 x 10^2 and 2.5 x 10^3 ppm^{-1} min^{-1} (primarily propane)	3
ALK3	Alkanes and other non-aromatic compounds that react only with OH and have kOH between 2.5 x 10^3 and 5 x 10^3 ppm^{-1} min^{-1}	4
ALK4	Alkanes and other non-aromatic compounds that react only with OH and have kOH between 5 x 10^3 and 1 x 10^4 ppm^{-1} min^{-1}	5
ALK5	Alkanes and other non-aromatic compounds that react only with OH and have kOH greater than 1 x 10^4 ppm-1 min-1	8
ARO1	Aromatics with kOH < 2x10^4 ppm^{-1} min^{-1}	7
ARO2	Aromatics with kOH > 2x10^4 ppm^{-1} min^{-1}	9
BACL	Biacetyl	4
BALD	Aromatic aldehydes	7
BENZ	Benzene	6
CCHO	Acetaldehyde	2
CH4	Methane	1
CRES	Phenols and Cresols	7
ETHE	Ethene	2
FACD	Formic Acid	1

Species Name	Description	Carbon #
GLY	Glyoxal	2
HCHO	Formaldehyde	1
IPRD	Lumped isoprene product species	5
ISOP	Isoprene	5
MACR	Methacrolein	4
MEK	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-13} but slower than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	4
MEOH	Methanol	1
MGLY	Methyl Glyoxal	3
MVK	Methyl Vinyl Ketone	4
OLE1	Alkenes (other than ethene) with kOH < 7x10^4 ppm^{-1} min^{-1}	5
OLE2	Alkenes with kOH > 7x10^4 ppm^{-1} min^{-1}	5
PACD	Peroxyacetic and higher peroxycarboxylic acids	3
PRD2	Ketones and other non-aldehyde oxygenated products that react with OH radicals faster than 5 x 10^{-12} cm^3 $molec^{-2}$ sec^{-1}	6
RCHO	Propionaldehyde and larger aldehydes	3
TERP	Terpenes	10
RNO3	Lumped Organic Nitrates	6

Table 4. Species names, descriptions and carbon numbers common to all CRI Mechanisms

Species Name	Description	Carbon #
CH4	Methane	1
C2H6	Ethane	2
C3H8	Propane	3
NC4H10	Butane and larger alkanes	4
C2H4	Ethene	2
C3H6	Propene and larger terminal alkenes	3
TBUT2ENE	Trans-2-butene and other internal alkenes	4
APINENE	Alpha-pinene and similar terpenes	10
BPINENE	Beta-pinene and similar terpenes	10
C5H8	Isoprene	5
C2H2	Ethyne	2
BENZENE	Benzene	6
TOLUENE	Toluene and other monoalkyl aromatics	7
OXYL	Xylene and other polyalkyl aromatics	8
CH3OH	Methanol	1
C2H5OH	Ethanol	2
HCHO	Formaldehyde	1

Species Name	Description	Carbon #
CH3CHO	Acetaldehyde	2
C2H5CHO	Propionaldehyde and larger aldehydes	3
CH3COCH3	Acetone	3
MEK	Methyl Ethyl Ketone	5
UCARB10	Lumped oxygenated compounds	4
HCOOH	Formic acid	1
CH3CO2H	Acetic acid and larger carboxylic acids	2

Mapping of Semi-Volatile and Intermediate Volatility Organic Compounds

Some gas profiles of the SPECIATE 5.0 repository allocate mass to semi-volatile (SVOC) and intermediate volatility organic compounds (IVOC) that tend to form secondary organic aerosol (SOA) which tends to reduce their influence on oxidant chemistry. For this reason, species classified as SVOC or IVOC are mapped to model species that do not participate in gas-phase chemistry but are accounted for by SOA chemistry schemes. The classification of SPECIATE compounds by volatility is based on estimated saturation concentration (C*) as described previously.[6] As shown in **Table 5,** species with log (C*) less than or equal to 6.5 will not participate in oxidant chemistry.

SPECIATE SVOC and IVOC are mapped using two distinct methods that account for differences in the SOA schemes of CMAQ and CAMx. For CMAQ AE7 and CAMx CF2, SVOC and IVOC are mapped to model species NVOL and IVOC, respectively. For CMAQ AE8, the SPECIATE SVOC and IVOC are mapped to a larger number of CMAQ model species designed to work with the CMAQ VBS SOA scheme.[7] **Table 5** lists the CMAQ or CAMx model species used for SPECIATE SVOC and IVOC according to their binned volatility.

Table 5. Mapping of SPECIATE SVOC and IVOC to CMAQ and CAMx model species according to volatility bin

SVOC and IVOC Volatility Bin Upper Threshold in log(C*)	CMAQ AE8 VBS Species[1]	CMAQ AE7 and CAMx CFv2.2 Model Species
-0.5	SVOCN1	NVOL[2]
0.5	SVOCP0	
1.5	SVOCP1	
2.5	SVOCP2	
3.5	IVOCP3	IVOC
4.5	IVOCP4	
5.5	IVOCP5, IVOCP5ARO	
6.5	IVOCP6, IVOCP6ARO	

1 IVOCP5ARO and IVOCP6ARO specifically represent intermediate volatility aromatic compounds with C* of 10^5 and 10^6 µg/m^3, respectively
2 The upper threshold in log(C*) for NVOL was previously 3.5 in CAMx mappings, but was revised to 2.5 for consistency with the CMAQ VBS bins.

[6] Ramboll technical memo to EPA titled "Mapping SPECIATE 5.0 compounds for photochemical models" dated October 11, 2018.
[7] Robinson, A.L., Donahue, N.M., Shrivastava, M.K., Weitkamp, E.A., Sage, A.M., Grieshop, A.P., Lane, T.E., Pierce, J.R. and Pandis, S.N., 2007. Rethinking organic aerosols: Semivolatile emissions and photochemical aging. Science, 315(5816), pp.1259-1262.

Some gas profiles in SPECIATE 5.0 contain CMAQ AE8 VBS species from **Table 5** directly in the profile creating a conceptual linkage of SPECIATE profiles to a specific aerosol scheme. AE8 VBS model species present in SPECIATE are treated by the Speciation Tool as one-to-one mappings for CB6r3_AE8 and mappings to either IVOC or NVOL for CB6r3_AE7 and CB6r4_CF2.

The mechanism mapping named CB6r4_CF2 can be used safely for other CB6 mechanisms in CAMx, namely CB6r3 and CB6r2h.

New Explicit CMAQ Model Species for CB6r3, CRI, and SAPRC07TC

New explicit model species are added to the mechanism mappings for CMAQ CB6r3, CRI, and SAPRC07TC chemical mechanisms with AE7 and AE8 aerosol options. The mappings intended for other models (e.g., CAMx) do not have these CMAQ-specific changes.

Naphthalene (SPECIATE ID: 611) is now mapped to the explicit species NAPH (naphthalene) in all CMAQ mappings to support CMAQv5.3. Naphthalene was previously mapped to XYL in the CMAQ CB6r3 mappings, and thus XYL is renamed as XYLMN (xylene minus naphthalene) in CB6r3_AE7 to match a naming convention of CMAQv5.3. Future CMAQ versions will simplify the name of XYLMN to XYL and accordingly CB6r3_AE8 uses the name XYL. In the CB6r4_CF2 mapping naphthalene is mapped to IVOC (rather than XYL) based on its estimated low volatility.

Alpha-pinene (SPECIATE ID: 1083) generally is mapped to the CB6 model species TERP (terpenes) but is now mapped to the explicit species APIN in the CMAQ CB6r3 mechanism mappings for AE7 and AE8.[8] We did not adopt a "terpene minus alpha-pinene" naming convention and so the TERP species remains in the CB6r3 mechanism mappings. Alpha-pinene is already a model species in CRI ('APINENE') and SAPRC07TC (see **Creation of the CMAQ SAPRC07TC mapping**).

Explicit (one-to-one) mappings were added for acetic acid (to AACD) and formic acid (to FACD) in the CMAQ CB6r3 mappings. Previously, acetic acid was mapped to PAR + UNR and formic acid to UNR. The recommended mappings of acetic and formic acid for CAMx were not changed because these species are predominantly secondary in the atmosphere and the CAMx developer assessed that changing the mappings would incur cost (larger emission files) for little benefit (little change in model concentration).

Creation of the CMAQ SAPRC07TC mapping

Previous versions of the Speciation Tool contained mechanism mappings for SAPRC07 but not SAPRC07TC, which includes many additional explicit model species to represent air toxics. The previous SAPRC07 mappings are revised to include the following explicit model species of SAPRC07TC: 1,3-butadiene (BDE13), acrolein (ACRO), a-pinene (APIN), 1,2,4-trimethyl benzene (B124), ethanol (ETOH), o-xylene (OXYL), m-xylene (MXYL), p-xylene (PXYL), propane (PRPE), sesquiterpenes (SESQ), and toluene (TOLU). Creating new mappings was straightforward for species that tend to be resolved in SPECIATE profiles, e.g., propane, 1,3-butadiene, toluene. Where SPECIATE profiles identified a compound mixture that clearly corresponds to a few explicit SAPRC07TC model species we made assumptions, e.g., that a previous mapping to XYL can be replaced by equal fractions of OXYL, MXYL and PXYL. The updated mappings are shown in **Table 6**. Where SPECIATE profiles identified complex mixtures that might contain an indeterminate amount of an explicit SAPRC07TC model species (e.g., "undefined petroleum distillates") we made no assumption and the previous mapping remains unchanged.

[8] See documentation here: https://github.com/USEPA/CMAQ/blob/master/DOCS/Release_Notes/biogenic_apinene.md

Table 6. Updated SAPRC07TC mappings of compound mixtures containing explicit model species

Species ID	Description	SAPRC07TC Mapping
326	C9 aromatics	0.2857*ARO1 + 0.6429*ARO2 + 0.0714*B124
507	Isomers of xylene	0.25*ARO1 + 0.25*MXYL + 0.25*OXYL + 0.25*PXYL
580	Misc. trimethylbenzenes	0.9*ARO2 + 0.1*B124
755	Trimethylbenzenes (mixed)	0.9*ARO2 + 0.1*B124
1962	C-3 Compounds	0.5*ALK2 + 0.5*PRPE

ATTACHMENT A

Mapping SPECIATE 5.0 compounds for photochemical models

RAMBOLL

ENVIRONMENT
& HEALTH

MEMO

To **Alison Eyth and Madeleine Strum, OAQPS, EPA**

From **Ross Beardsley, Tejas Shah and Greg Yarwood**

Subject **Mapping SPECIATE 5.0 compounds for photochemical models**

October 11, 2018

Ramboll
7250 Redwood Boulevard
Suite 105
Novato, CA 94945
USA

T +1 415 899 0700
F +1 415 899 0707
www.ramboll.com

Summary

"Mechanism mapping" refers to translating the detailed chemical constituents identified in "speciation profiles" to the "model species" used by condensed chemical mechanisms of photochemical air quality models. Mechanism mapping is implemented in software called the "Speciation Tool." Ramboll reviewed version 5.0 of the Environmental Protection Agency's (EPA's) SPECIATE database of speciation profiles and created mechanism mappings for newly added and previously unmapped chemical species. We estimated the vapor pressure (p_i^o) of these compounds and those identified the low volatility organic compounds (LVOC) were not mapped to gas phase mechanisms. After removing LVOC, mechanism mappings for the remaining new or previously unmapped species were created for the Carbon Bond (CB), SAPRC07, and the Common Representative Intermediates (CRI) [1] gas phase mechanisms using methods developed previously.[2]

CLASSIFYING COMPOUNDS AS NVOL, IVOC, OR VOC

Some of the newly added or unmapped organic gas profiles within SPECIATE 5.0 repository allocate mass to LVOC that react in the atmosphere and rapidly move to the particulate phase limiting their influence on gas phase chemistry. For this reason, previous mechanism mappings for the Speciation Tool assigned some compounds to the non-volatile model species NVOL, removing them from the gas phase mechanism. However, the methodology for assigning compounds to NVOL was not well defined because the p_i^o of these compounds was not available within SPECIATE. A scheme was created previously to refine the identification of LVOCs within the SPECIATE 4.5 database and classify these compounds as NVOL or intermediate-volatility (IVOC) based on the definition of IVOC proposed by Robinson et al.[3] for the Volatility Basis Set SOA model. This task describes the

[1] http://mcm.leeds.ac.uk/CRI/

[2] Ramboll Environ technical memos to EPA titled "Species Mappings for CB6 and CB05 for use with SPECIATE 4.5" and "Classification of SPECIATE 4.5 compounds as a function of estimated vapor pressure" dated September 27, 2016 and September 29, 2017, respectively.

[3] Robinson, A.L, N. Donahue, M. Shrivastava, E.A. Weitkamp, A.M. Sage, A.P. Grieshop, T.E. Lane, J.R. Pierce, and S.N. Pandis. "Rethinking Organic Aerosols: Semivolatile Emissions and Photochemical Aging." Science 315, Issue 5816 (March 2, 2007): 1259-1262. doi:10.1126/science:1133061

application of this scheme to the newly added or previously unmapped species in SPECIATE 5.0.

Estimating Volatility

The MPBPWIN model (v1.43) within the EPA's Estimation Programs Interface (EPI) Suite (https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface) was utilized to estimate the p_i^o of compounds in SPECIATE 5.0 based on CAS# or simplified molecular-input line-entry system (SMILES) string provided in the SPECIATE species properties table. For compounds without a CAS# or SMILES string in SPECIATE, a SMILES string or estimate of p_i^o was obtained from an online repository (such as PubChem or ChemSpider) if readily available. In some cases the p_i^o of a surrogate compound was used, e.g. 1-undecene (ID: 1082) used for C11 Olefins (ID: 1930). Inorganic species, which were primarily mapped to NVOL previously, and other species lacking sufficient information to classify were assigned no mapping to ensure that they are not incorrectly treated as non-volatile organic species in the SOA modeling schemes of AQMs. Organic particulate matter species were assigned to NVOL (e.g. Organic Carbon II). The species assigned no mapping and the organic particulate matter species mapped to NVOL are listed in Table A1 and Table A2 of Appendix A, respectively.

In summary, the methodology used for estimating p_i^o of compounds in the SPECIATE database was as follows:

1. Use CAS# or SMILES string from the SPECIATE species property table.
2. If a CAS# or SMILES string is not available in SPECIATE, acquire SMILES string or p_i^o estimate from online repository such as PubChem (https://pubchem.ncbi.nlm.nih.gov/) or ChemSpider (http://www.chemspider.com/).
3. If a SMILES string cannot be readily obtained, assign the p_i^o of a surrogate.
4. If the properties of the species are unknown and a SMILES string or surrogate cannot be reasonably assigned, do not assign a mapping.

After estimating p_i^o, the saturation concentration of each species (C_i^*) was calculated at 298 K following Equation 1[4]:

$$C_i^* = \frac{10^6 MW_i p_i^o}{760RT} \quad (1)$$

where, MW_i is the molecular weight of each compound from the SPECIATE 5.0 database[5], R is the ideal gas constant (m^3 atm mol^{-1} K^{-1}), and T is the temperature (K).

Then, the species were classified as NVOL, IVOC or VOC using the volatility bins shown in Table 1, which are based on the volatility range propsed for IVOC in Robinson et al[3].

Table 1. Saturation concentration (C_i^*, µg m^{-3}) bins used in volatility classification

NVOL	$C_i^* < 3 \times 10^3$
IVOC	$3 \times 10^3 \leq C_i^* < 3 \times 10^6$
VOC	$3 \times 10^6 \leq C_i^*$

[4] Odum, J.R., T. Hoffmann, F. Bowman, D. Collins, R.C. Flagan, and J.H. Seinfeld. "Gas/Particle Partitioning and Secondary Organic Aerosol Yields." Environmental Science & Technology 30, no. 8 (July 1, 1996): 2580–85. doi:10.1021/es950943+.

[5] In cases in which the molecular weight was missing from the database or incorrect, a corrected or assigned molecular weight was used in these calculations.

Discussion of Volatility Estimates

The estimated C_i^* of SPECIATE 5.0 species are shown in Figure 1 as a function of molecular weight along with the defined volatility bins for VOC, IVOC, and NVOL. Using estimated vapor pressure to classify SPECIATE compounds allows for systematic identification, classification and removal of complex, low-volatility compounds.

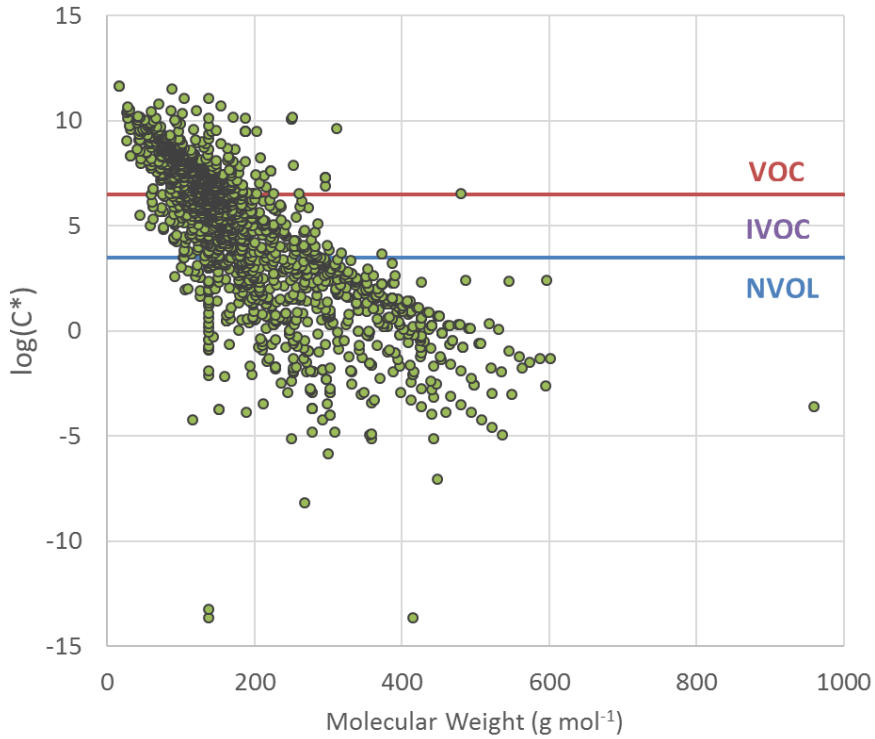


Figure 1. Estimated saturation concentrations (C_i^*) of species in SPECIATE 5.0 at 298 K as a function of molecular weight.

The number of species from SPECIATE 5.0 assigned to each bin is shown in Table 2, and approximately 50% are assigned to IVOC or NVOL based on their estimated p_i^o.

Table 2. Number of SPECIATE 4.5 compounds assigned to each volatility bin or not mapped.

Volatility Bin	Number of Compounds
VOC	1265
IVOC	619
NVOL	786
Not mapped*	124
Total	**2794**

* Inorganic species and other unknown species with insufficient information to characterize

MAPPING LVOC FOR ALL GAS-PHASE MECHANISMS

All gas-phase mechanism mappings for Speciation Tool version 5.0 use a new systematic methodology for mapping low volatility compounds. Compounds with very low vapor pressure are mapped to model species NVOL and intermediate volatility compounds are mapped to IVOC. By reviewing which compounds are mapped to IVOC and NVOL, representative molecular weights (MW) and numbers of carbon atoms (C#) were assigned to these species (Table 3). Consequently, the Speciation Tool version 5.0 produces speciation profiles that compute IVOC and NVOL in moles, which is consistent with all other model species in the gas-phase mechanisms. For consistency, air quality models that make use of IVOC and/or NVOL emissions from Speciation Tool version 5.0 should make use of the representative properties given in Table 3.

Table 3. Representative molecular weights and numbers of carbon atoms for model species IVOC and NVOL

Model Species	Representative C#	Representative MW (g/mol)
IVOC	12	160
NVOC	18	260

MAPPING MIXTURES FOR ALL MECHANISMS

Several species in SPECIATE 5.0 are not a single species but represent a mixture of compounds, e.g. 1-methylcyclohexene; 4-methylheptane (ID:2677). For these compounds the average molecular weight of the compounds was used in estimating vapor pressure. Similarly, the mapping assignments of each compound in the mixture were averaged to get the mapping of the mixture.

MAPPING INORGANIC AND ORGANIC PARTICULATE MATTER SPECIES FOR ALL MECHANISMS

Inorganic species and species with unknown properties, which previously were primarily mapped to NVOL, were assigned no mapping to ensure that they are not incorrectly treated as non-volatile organic species by air quality models. These species are now implicitly mapped to UNK by the Speciation Tool along with all other species that are not mapped. Organic particulate matter species were mapped to NVOL. The species assigned no mapping and the organic particulate matter species mapped to NVOL are listed in Table A1 and Table A2 of Appendix A, respectively.

GAS-PHASE MECHANISM MAPPINGS

After the classification of the LVOCs contained in SPECIATE 5.0 database in Task 1, new gas phase mechanism mappings were created for the remaining VOCs based on the guidelines created for SPECIATE 4.5.[2]

Carbon Bond (CB) Mechanisms

Organic gases are mapped to the CB mechanism either as explicitly represented individual compounds (e.g. ALD2 for acetaldehyde), or as a combination of model species that represent common structural groups (e.g. ALDX for other aldehydes, PAR for alkyl groups). Table 4 lists the explicit and structural model species in the CB05 and CB6 mechanisms, each of which represents a defined number of carbon atoms. CB6 contains four more explicit model species than CB05 and an additional structural group to represent ketones.

Table 4. Model Species in the CB05 and CB6 chemical mechanisms

Model Species Name	Description	Number of Carbons	Included in CB05 (structural mapping)	Included in CB6
Explicit model species				
ACET	Acetone (propanone)	3	**No** (3 PAR)	Yes
ALD2	Acetaldehyde (ethanal)	2	Yes	Yes
BENZ	Benzene	6	**No** (1 PAR, 5 UNR)	Yes
CH4	Methane	1	Yes	Yes
ETH	Ethene (ethylene)	2	Yes	Yes
ETHA	Ethane	2	Yes	Yes
ETHY	Ethyne (acetylene)	2	**No** (1 PAR, 1 UNR)	Yes
ETOH	Ethanol	2	Yes	Yes
FORM	Formaldehyde (methanal)	1	Yes	Yes
ISOP	Isoprene (2-methyl-1,3-butadiene)	5	Yes	Yes
MEOH	Methanol	1	Yes	Yes
PRPA	Propane	3	**No** (1.5 PAR, 1.5 UNR)	Yes
Common Structural groups				
ALDX	Higher aldehyde group (-C-CHO)	2	Yes	Yes
IOLE	Internal olefin group (R_1R_2>C=C<R_3R_4)	4	Yes	Yes
KET	Ketone group (R_1R_2 >C=O)	1	**No** (1 PAR)	Yes
OLE	Terminal olefin group (R_1R_2>C=C)	2	Yes	Yes
PAR	Paraffinic group (R_1-C<R_2R_3)	1	Yes	Yes
TERP	Monoterpenes	10	Yes	Yes
TOL	Toluene and other monoalkyl aromatics	7	Yes	Yes
UNR	Unreactive carbon groups (e.g., halogenated carbons)	1	Yes	Yes
XYL	Xylene and other polyalkyl aromatics	8	Yes	Yes

CB6 and CB05 mappings of new VOCs (excluding species mapped to IVOC and NVOL) were created using the general guidelines for mapping provided in Table 5, and mapping guidelines for difficult to map compound classes and structural groups shown in Table 6.

Table 5. General guidelines for mapping using CB6 structural model species.

CB6 Species Name	Number of Carbons	Represents
ALDX	2	Aldehyde group. ALDX represents 2 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. propionaldehyde is ALDX + PAR
IOLE	4	Internal olefin group. IOLE represents 4 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. 2-pentene isomers are IOLE + PAR. *Exceptions:* • IOLE with 2 carbon branches on both sides of the double bond are downgraded to OLE
KET	1	Ketone group. KET represents 1 carbon and additional carbons are represented as alkyl groups (mostly PAR), e.g. butanone is 3 PAR + KET
OLE	2	Terminal olefin group. OLE represents 2 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. propene is OLE + PAR. Alkyne group, e.g. butyne isomers are OLE + 2 PAR.
PAR	1	Alkanes and alkyl groups. PAR represents 1 carbon, e.g. butane is 4 PAR. See UNR for exceptions.
TERP	10	All monoterpenes are represented as 1 TERP.
TOL	7	Toluene and other monoalkyl aromatics. TOL represents 7 carbons and any additional carbons are represented as alkyl groups (mostly PAR), e.g. ethylbenzene is TOL + PAR. Cresols are represented as TOL and PAR. Styrenes are represented using TOL, OLE and PAR.
UNR	1	Unreactive carbons are 1 UNR such as quaternary alkyl groups (e.g., neo-pentane is 4 PAR + UNR), carboxylic acid groups (e.g., acetic acid is PAR + UNR), ester groups (e.g., methyl acetate is 2 PAR + UNR), halogenated carbons (e.g., trichloroethane isomers are 2 UNR), carbons of nitrile groups (-C≡N)
XYL	8	Xylene isomers and other polyalkyl aromatics. XYL represents 8 carbons and any additional carbons are represented as alkyl groups (mostly PAR), e.g. trimethylbenzene isomers are XYL + PAR
IVOC	12	Intermediate volatility organic compounds with saturation concentrations (C_i^*) in the range: $3 \times 10^3 \mu g\ m^{-3} \leq C_i^* < 3 \times 10^6 \mu g\ m^{-3}$
NVOL	18	Low volatility organic compounds with saturation concentrations (C_i^*) in the range: $C_i^*\ \mu g\ m^{-3} < 3 \times 10^3$

Table 6. Mapping guidelines for some difficult to map compound classes and structural groups.

Compound Class/Structural group	CB model species representation
Chlorobenzenes and other halogenated benzenes	Guideline: • 3 or less halogens – 1 PAR, 5 UNR • 4 or more halogens – 6 UNR Examples: • 1,3,5-Chlorobenzene – 1 PAR, 5 UNR • Tetrachlorobenzenes – 6 UNR
Cyclodienes	Guideline: • 1 IOLE with additional carbons represented as alkyl groups (generally PAR) Examples: • Methylcyclopentadiene – 1 IOLE, 2 PAR • Methylcyclohexadiene – 1 IOLE, 3 PAR
Furans/Pyrroles	Guideline: • 2 OLE with additional carbons represented as alkyl groups (generally PAR) Examples: • 2-Butylfuran – 2 OLE, 4 PAR • 2-Pentylfuran – 2 OLE, 5 PAR • Pyrrole – 2 OLE • 1-Methylpyrrole – 2 OLE, 1 PAR
Heterocyclic aromatic compounds containing 2 non-carbon atoms	Guideline: • 1 OLE with remaining carbons represented as alkyl groups (generally PAR) Examples: • Ethylpyrazine – 1 OLE, 4 PAR • 1-methylpyrazole – 1 OLE, 2 PAR • 4,5-Dimethyloxazole – 1 OLE, 3 PAR
Triple bond(s)	Guideline: • Triple bonds are treated as PAR unless they are the only reactive functional group. If a compound contains more than one triple bond and no other reactive functional groups, then one of the triple bonds is treated as OLE with additional carbons treated as alkyl groups. Examples: • 1-Penten-3-yne – 1 OLE, 3 PAR • 1,5-Hexadien-3-yne – 2 OLE, 2 PAR • 1,6-Heptadiyne – 1 OLE, 5 PAR
Alkane isomers and cycloalkanes	Guideline: • UNR is calculated as the inverse of the carbon number (1/number of carbon) with the remaining treated as PAR. This includes alkane branches on rings and other compounds that do not have an explicitly defined structure • Cycloalkanes are assumed to be cyclohexane with additional carbons treated as described above unless explicitly defined Examples: • Branched C7 Alkanes – 6.86 PAR, 0.14 UNR • C10 Cycloalkanes – 9.75 PAR, 0.25 UNR

After applying the volatility classification summarized in Table 2, 50% of the existing CB mappings for SPECIATE 5.0 were assigned to IVOC or NVOL. Of the remaining 50%, 124 inorganic species were assigned no mapping (and consequently mapped to UNK by the Speciation Tool). The updated CB mappings were incorporated in the Speciation Tool.

SAPRC07 Mechanism

Mappings for the SAPRC07 mechanism were created for species added in SPECIATE 5.0 with the LVOC assigned to IVOC or NVOL as described previously for CB mechanisms. VOCs are represented in SAPRC by lumped model species classified by structure (e.g., alkane, alkene, etc.) and OH radical rate constant

(kOH). Existing species were used as examples to create mappings for the newly added and previously unmapped SPECIATE 5.0 compounds by analogy. For compounds where no analogous mappings existed, the AopWin v1.92 model in EPI Suite was used to estimate kOH and mappings were assigned based on structure and estimated kOH. The updated mappings were incorporated in the Speciation Tool.

CRI Mechanism

The CRI mechanism is a condensed mechanism developed from the Master Chemical Mechanism version 3.1[6]. There are several versions of the CRI mechanism with differing levels of condensation and the mappings described here are specific to version CRIv2r5. SPECIATE 5.0 VOCs were mapped by translating the updated SAPRC07 mappings (and CB06 mappings in some cases). All alkanes with 4 or more carbons were mapped to NC4H10 with the mapping factors equal to the ratio of the MW of the VOC to the MW of butane, which is consistent with the VOC mass weighting approach used in the development of the CRI mechanism[7,8]. Organic nitrates (RNO3 in SAPRC07) were treated as alkanes, and terpene mappings in SAPRC (TERP) were split evenly between APINENE and BPINENE except the compounds α-pinene and β-pinene. The CRI mechanism mappings for most terpenes could be improved.

The mapping guidelines are provided in Table 7. The updated mappings were incorporated in the Speciation Tool.

[6] http://mcm.leeds.ac.uk/CRI/

[7] Personal communication with CRI Developer Mike Jenkin

[8] Watson, L.A., D.E. Shallcross, S.R. Utembe, and M.E. Jenkin, A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 2: Gas phase mechanism reduction, Atmos. Environ., 42 7196-7204, 2008.

Table 7. Translation from SAPRC07 mappings to CRI mappings.

CRIv2r5 Species	SAPRC07 Translation
CH4	CH4
C2H6	ALK1
C3H8	ALK2
NC4H10	ALK3 + ALK4 + ALK5 (mapping weighted by the ratio of the MW of the compounds to the MW of butane)
C2H4	ETHE
C3H6	OLE1
TBUT2ENE	OLE2
APINENE	TERP/2
BPINENE	TERP/2
C5H8	ISOP
C2H2	ACYE
BENZENE	BENZ
TOLUENE	(ARO1 – BENZ*) + CRES + BALD
OXYL	ARO2
CH3OH	MEOH
C2H5OH	ETOH*
HCHO	HCHO
CH3CHO	CCHO
C2H5CHO	RCHO + GLY
CH3COCH3	ACET
MEK	MEK + PRD2
UCARB10	MACR + MVK + IPRD + BACL + MGLY
HCOOH	FACD
CH3CO2H	AACD
NROG	NROG
NVOL	NVOL

* CB06 model species

Appendix A

Inorganic species and species with unknown properties in SPECIATE 5.0 that were not assigned a mechanism mapping

Table A1. Inorganic species and species with unknown properties in SPECIATE 5.0 that were not assigned a mechanism mapping.

Species ID	Name/Description
292	Aluminum
294	Ammonia
296	Antimony
298	Arsenic
300	Barium
307	Bromine
328	Cadmium
329	Calcium
337	Chloride ion
339	Chlorine
347	Chromium
379	Cobalt
380	Copper
436	Total carbon
468	Gallium
477	Gold
487	Indium
488	Iron
519	Lanthanum
520	Lead
525	Magnesium
526	Manganese
528	Mercury
586	Molybdenum
612	Nickel
613	Nitrate
649	Palladium
665	Phosphate
666	Phosphorus
669	Potassium
689	Rubidium
693	Selenium
694	Silicon
695	Silver
696	Sodium
697	Strontium
699	Sulfate
700	Sulfur

Species ID	Name/Description
712	Thallium
714	Tin
715	Titanium
765	Uranium
767	Vanadium
777	Yttrium
778	Zinc
779	Zirconium
784	Ammonium
785	Sodium ion
788	Carbonate
794	Elemental carbon I
795	Chlorine atom
796	Elemental carbon III
797	Elemental Carbon
810	Bromine Atom
830	Sulfur dioxide
831	Hydrogen Sulfide
843	Sum of PM species
1027	Steroid-w , also noted as sterow
1166	Carbon dioxide
1190	Elemental carbon II
1423	Steroid-m , also noted as sterom
1680	Nitrite
1839	Beryllium
1847	Tellurium
1849	Dysprosium
1850	Lithium
1851	Neodymium
1852	Niobium
1853	Praseodymium
1854	Samarium
1855	Scandium
1856	Tantalum
1857	Terbium
1858	Thorium
1859	Tungsten
1860	Boron
1861	Cerium

Species ID	Name/Description
1862	Cesium
1863	Europium
1864	Gadolinium
1865	Germanium
1866	Hafnium
1867	Holmium
1868	Ytterbium
1869	Bismuth
1870	Iodine atom
1871	Fluoride ion
1872	hydrogen phosphate
1884	PM other
1919	Butyl
2032	Copper naphthenate
2302	Potassium ion
2303	Calcium ion
2586	Particulate Mercury
2587	Oxidized Mercury
2588	Elemental Mercury
2604	Reactive gas-phase Mercury
2605	Nitrogen Monoxide (Nitric Oxide)
2606	Nitrogen Dioxide
2607	Nitrous acid
2663	Protactinium
2664	Erbium
2665	Platinum
2668	Particulate Water
2670	Metal-bound Oxygen
2671	Other Unspeciated PM2.5
2672	Chromium(VI)
2686	Iridium
2687	Hydrated sulfate
2772	Magnesium ion
2839	Oxalate anion
2840	Acetate anion
2841	Formate anion (formic acid, ion)
2847	Calcium Oxide
2848	Aluminum Oxide (or Alumina)
2849	Iron Oxide

Species ID	Name/Description
2850	Potassium Carbonate
2851	Sodium Carbonate
2852	Magnesium Oxide
2853	Sulfur Trioxide
2854	Silica
2855	Titanium Oxide
2856	Lutetium
3041	Ruthenium
3042	Rhodium
3152	Corrosion Inhibitor

Table A2. Organic particulate matter species in SPECIATE 5.0 that were mapped to NVOL.

Species ID	Name/Description
626	Organic carbon
789	Organic carbon II
790	Organic carbon III
791	Organic carbon IV
792	Pyrolyzed organic carbon
1027	Steroid-w , also noted as sterow
1183	Organic carbon I
1423	Steroid-m , also noted as sterom
2298	Volatile Carbon
2669	Particulate Non-Carbon Organic Matter

ATTACHMENT B

Classification of SPECIATE 4.5 compounds as a function of estimated vapor pressure

September 29, 2017

MEMORANDUM

To: Alison Eyth and Madeleine Strum, OAQPS, EPA
From: Ross Beardsley, Tejas Shah and Greg Yarwood, Ramboll Environ
Subject: Classification of SPECIATE 4.5 compounds as a function of estimated vapor pressure

---

Summary

Ramboll Environ (RE) implemented a scheme to identify and classify low volatility organic compounds (LVOC) contained within the gas speciation profiles of the Environmental Protection Agency's (EPA's) SPECIATE version 4.5 repository using estimated vapor pressure (p_i^o). These LVOC are not well represented by the model species of condensed gas phase mechanisms (e.g. CB6) and have little impact on atmospheric oxidants because their degradation products tend to be incorporate into aerosols. The purpose of this work is to remove these compounds from the gas phase mechanism and add LVOC emissions that can be utilized by secondary organic aerosol (SOA) schemes, such as the volatility basis set (VBS).

After classifying the LVOC species in the SPECIATE 4.5 repository, the Carbon Bond (CB) and SAPRC07 mechanism mappings of the remaining VOCs were reviewed and updated. Species added in SPECIATE 4.5 were mapped for CB6 and CB05 mechanisms previously[1], but existing mappings were not verified. Additionally, SAPRC07 mechanism mappings were created for the new species in SPECIATE 4.5 and some existing mappings were updated for consistency. Finally, VOC mappings were created for the model species of the Common Representative Intermediates (CRI) mechanism.[2] This memorandum is organized according to the specific project tasks listed in the Scope of Work (SOW).

Task 1: Implement a new scheme to classify compounds as NVOL, IVOC, or VOC

Some of the organic gas profiles within SPECIATE 4.5 repository allocate mass to LVOC that rapidly move to the particulate phase limiting their influence on gas phase chemistry. For this reason, current mechanism mappings for the Speciation Tool assign some compounds to the non-volatile model species NVOL, removing them from the gas phase mechanism. However, the methodology for assigning compounds to NVOL is not well defined because the p_i^o of these compounds are not available within SPECIATE. Additionally, refined classification of these compounds as NVOL or intermediate-volatility (IVOC) could improve support for SOA modeling schemes in air quality models (AQMs) such as CMAQ (http://www.cmascenter.org/cmaq/) and CAMx (http://www.camx.com).

[1] Ramboll Environ technical memo to EPA "Species Mappings for CB6 and CB05 for use with SPECIATE 4.5" dated September 27, 2016
[2] http://mcm.leeds.ac.uk/CRI/

Ramboll Environ, 773 San Marin Drive, Suite 2115, Novato, CA 94998
V +1 415.899.0700 F +1 415.899.0707

This task describes the implementation of a volatility based framework for classifying SPECIATE compounds by estimated p_i^o using the volatility groupings of the VBS SOA scheme in CAMx.

Methods

The MPBPWIN model (v1.43) within the EPA's Estimation Programs Interface (EPI) Suite (https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface) was utilized to estimate the p_i^o of compounds in SPECIATE 4.5 based on CAS# or simplified molecular-input line-entry system (SMILES) string provided in the SPECIATE species properties table. For compounds without a CAS# or SMILES string in SPECIATE, a SMILES string or EPI estimate of p_i^o was obtained from an online repository (such as PubChem or ChemSpider) if readily available. In some cases the p_i^o of a surrogate compound was used, e.g. 1-undecene (ID: 1082) used for C11 Olefins (ID: 1930). Some species in the database lack sufficient information to assign a SMILES string or surrogate (e.g. Species ID# 2285, Unknown 1), and these species were assumed to be volatile organic compounds (VOC). Inorganic species, which were primarily mapped to NVOL previously, were assigned no mapping to ensure that they are not incorrectly treated as non-volatile organic species in the SOA modeling schemes of AQMs.

In summary, the methodology used for estimating p_i^o of compounds in the SPECIATE database was as follows:

1. Use CAS# or SMILES string from the SPECIATE species property table.
2. If a CAS# or SMILES string is not available in SPECIATE, acquire SMILES string or EPI p_i^o estimate from online repository such as PubChem (https://pubchem.ncbi.nlm.nih.gov/) or ChemSpider (http://www.chemspider.com/).
3. If a SMILES string cannot be readily obtained, assign the p_i^o of a surrogate.
4. If the properties of the species are unknown and a SMILES string or surrogate cannot be reasonably assigned, classify the species as VOC. [3]

After estimating p_i^o, the saturation concentration of each species (C_i^*) was calculated at 298 K following Equation 1[4]:

$$C_i^* = \frac{10^6 MW_i p_i^o}{760RT} \qquad (1)$$

where, MW_i is the molecular weight of each compound from the SPECIATE 4.5 database, R is the ideal gas constant (m^3 atm mol^{-1} K^{-1}), and T is the temperature (K).

[3] The vapor pressures of 'Tetrabutyl orthotitanate' (ID: 2249) could not be estimated by the MPBPVP model of EPI Suite, but is low volatility and thus was assigned to NVOL instead of VOC.

[4] Odum, J.R., T. Hoffmann, F. Bowman, D. Collins, R.C. Flagan, and J.H. Seinfeld. "Gas/Particle Partitioning and Secondary Organic Aerosol Yields." Environmental Science & Technology 30, no. 8 (July 1, 1996): 2580–85. doi:10.1021/es950943+.

Then, the species were classified as NVOL, IVOC or VOC using the volatility bins shown in Table 1, which are based on the volatility range for IVOC from the VBS SOA scheme in CAMx.

Table 1. Saturation concentration (C_i^*, µg m^{-3}) bins used in volatility classification

NVOL	$C_i^* < 3 \times 10^3$
IVOC	$3 \times 10^3 \leq C_i^* < 3 \times 10^6$
VOC	$3 \times 10^6 \leq C_i^*$

Finally, after classifying compounds using the p_i^o estimated in EPI Suite, The Estimation of Vapour Pressure of Organics, Accounting for Temperature Intramolecular, and Non-addivity Effects (EVAPORATION) method of Compernolle et al.[5] was used to estimate the subcooled liquid vapor pressure ($p_{L,i}^o$) of a subset of compounds that were then compared to the EPI p_i^o estimates. The EVAPORATION method was more recently developed and was designed to specifically handle complex, polyfunctional molecules that are relevant to SOA formation. However, the scope of the model is somewhat limited by the functional groups it can handle (e.g. does not treat aromatics), and so this method could not be employed for all compounds. For this reason, the model was only applied to a subset of compounds that are within the scope of the model and whose EPI estimated C_i^* is within approximately one order of mangitude of the boundary of the NVOL/IVOC and IVOC/VOC bins.

Results and Discussion

The estimated C_i^* are shown in Figure 1A and 1B as a function of molecular weight for compounds previously treated as NVOL and as VOC, respectively, along with the newly defined volatility bins for VOC, IVOC, and NVOL. As can be seen, the lack of a defined methodology in previous mappings resulted in compounds with p_i^o corresponding to the VOC volatility bin being mapped to NVOL, and vice versa. Using estimated vapor pressure to classify SPECIATE compounds allows for systematic identification, classification and removal of complex, low-volatility compounds.

[5] Compernolle, S., K. Ceulemans, and J.-F. Müller, EVAPORATION: a new vapor pressure estimation method for organic molecules including non-additivity and intramolecular interactions, Atmos. Chem. Phys., 19431-9450, 2011. Accessed at: http://tropo.aeronomie.be/models/evaporation_run.htm

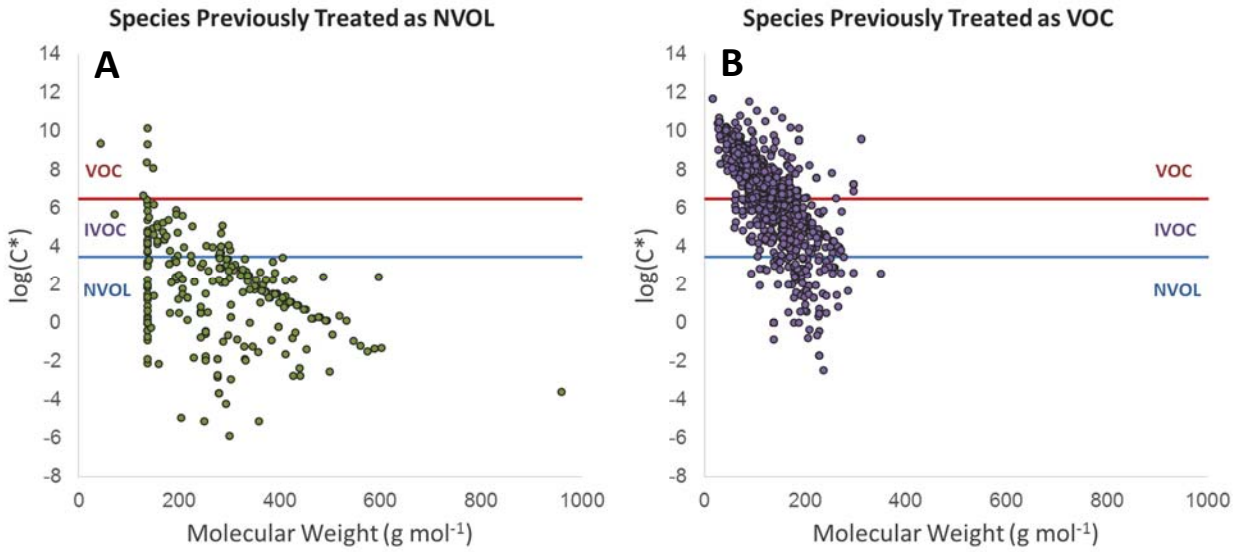


Figure 1. Estimated saturation concentrations (C_i^*) at 298 K as a function of molecular weight shown separately for species previously treated as NVOL (A) and as VOC (B).

The number of species from SPECIATE 4.5 assigned to each bin is shown in Table 2, and approximately 40% were assigned to IVOC or NVOL based on their estimated p_i^o. There are still a number of species that were not mapped for SPECIATE 4.4. The unmapped compounds are listed in Table A1 of Appendix A, and many of these compounds are inorganic or low volatility organic compounds (likely IVOC or NVOL).

Table 2. Number of SPECIATE 4.5 compounds assigned to each volatility bin or not mapped.

Volatility Bin	Number of Compounds
VOC	1180
IVOC	459
NVOL	302
Not mapped*	30
Total	**1971**

* Inorganic species

Saturation concentrations estimated by EPI Suite and the EVAPORATION method are compared in Figure 2 for a subset of compounds. In general, the estimated p_i^o from EPI suite are higher than the corresponding EVAPORATION method estimates, especially for low volatility compounds, as is reflected in the slope of the intercept-free linear regression line. This results in a shift to higher volatility bins for some the classification of some compounds. More specifically, the use of saturation concentrations from EVAPORATION would result in 15% of the VOC compounds in the subset examined to be assigned to IVOC, and 12% of the IVOC species in the subset to be assigned to NVOL. However, there is relatively good overall agreement between the estimates of the two models for these purposes (R^2 = 0.81), and the broad applicability of the EPI Suite and the simplicity of estimating the p_i^o of a large number of compounds (using batch mode) makes it an effective tool for implementing the volatility classification. However, future application of a more recent and accurate method may be valuable.

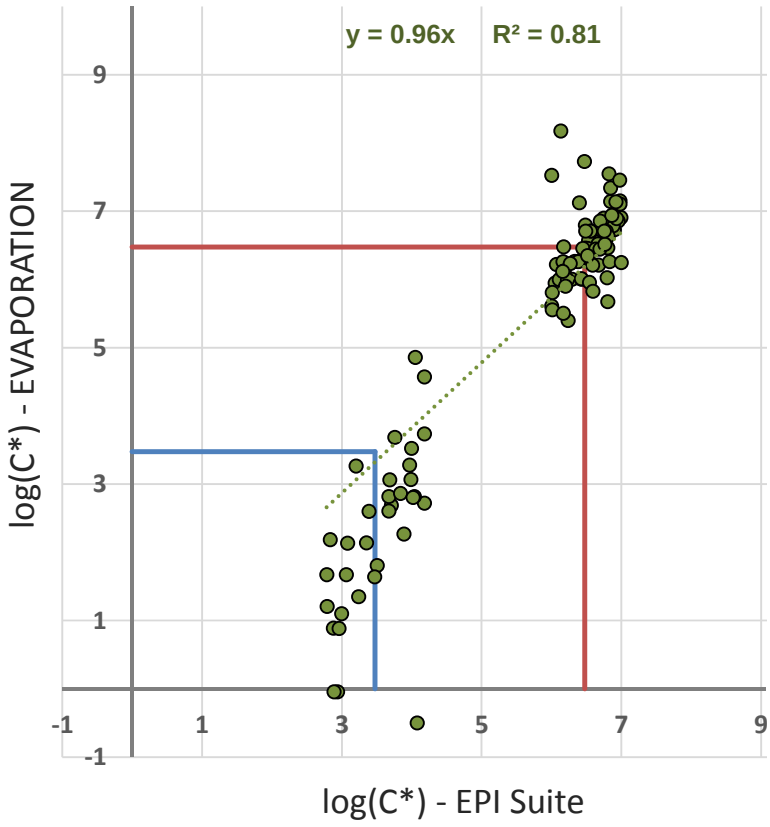


Figure 2. Comparison of the estimated Saturation Concentrations (C_i^*) at 298 K from the EPI Suite (x-axis) and the EVAPORATION method (y-axis).

Recommendation

The MPBPVP model of the EPI Suite was found to be an effective tool for implementing the volatility classification scheme described in this memo due to the simplicity of the approach and ability to handle a wide range of compounds. It is recommended that the EPI p_i^o classification is used for future additions to the SPECIATE database unless a more accurate, easily applied method capable of handling a wide range of compounds is made available. Now that a systematic method has been implemented for mapping low volatility compounds, mappings could be created for a number of compounds that were not mapped for SPECIATEv4.4 (listed in Table A1).

Task 2: Review all Carbon Bond VOC mappings for consistency with current mapping guidelines

After the classification of the LVOCs in SPECIATE 4.5 in Task 1, a systematic review of the existing CB6 and CB05 mechanism mappings was performed for the remaining VOCs based on the guidelines created for SPECIATE 4.5[1]. In addition, updated guidelines for mapping IVOC, NVOL, and inorganic species were established and applied.

Methods

CB Model Species

Organic gases are mapped to the CB mechanism either as explicitly represented individual compounds (e.g. ALD2 for acetaldehyde), or as a combination of model species that represent common structural groups (e.g. ALDX for other aldehydes, PAR for alkyl groups). Table 3 lists all of the explicit and structural model species in the CB05 and CB6 mechanisms, each of which represents a defined number of carbon atoms. CB6 contains four more explicit model species than CB05 and an additional structural group to represent ketones.

Table 3. Model Species in the CB05 and CB6 chemical mechanisms

Model Species Name	Description	Number of Carbons	Included in CB05 (structural mapping)	Included in CB6
Explicit model species				
ACET	Acetone (propanone)	3	**No** (3 PAR)	Yes
ALD2	Acetaldehyde (ethanal)	2	Yes	Yes
BENZ	Benzene	6	**No** (1 PAR, 5 UNR)	Yes
CH4	Methane	1	Yes	Yes
ETH	Ethene (ethylene)	2	Yes	Yes
ETHA	Ethane	2	Yes	Yes
ETHY	Ethyne (acetylene)	2	**No** (1 PAR, 1 UNR)	Yes
ETOH	Ethanol	2	Yes	Yes
FORM	Formaldehyde (methanal)	1	Yes	Yes
ISOP	Isoprene (2-methyl-1,3-butadiene)	5	Yes	Yes
MEOH	Methanol	1	Yes	Yes
PRPA	Propane	3	**No** (1.5 PAR, 1.5 UNR)	Yes
Common Structural groups				
ALDX	Higher aldehyde group (-C-CHO)	2	Yes	Yes
IOLE	Internal olefin group (R_1R_2>C=C<R_3R_4)	4	Yes	Yes
KET	Ketone group (R_1R_2 >C=O)	1	**No** (1 PAR)	Yes
OLE	Terminal olefin group (R_1R_2>C=C)	2	Yes	Yes
PAR	Paraffinic group (R_1-C<R_2R_3)	1	Yes	Yes
TERP	Monoterpenes	10	Yes	Yes
TOL	Toluene and other monoalkyl aromatics	7	Yes	Yes

Model Species Name	Description	Number of Carbons	Included in CB05 (structural mapping)	Included in CB6
UNR	Unreactive carbon groups (e.g., halogenated carbons)	1	Yes	Yes
XYL	Xylene and other polyalkyl aromatics	8	Yes	Yes

Updated mapping guidelines for non-explicit species

CB and all other gas-phase mechanism mappings for Speciation Tool version 4.5 use a new systematic methodology for mapping low volatility compounds (see Task 1). Compounds with very low vapor pressure are mapped to model species NVOL and intermediate volatility compounds are mapped to IVOC. By reviewing which compounds are mapped to IVOC and NVOL, representative molecular weights (MW) and numbers of carbon atoms (C#) were assigned to these species (Table 4). Consequently, the Speciation Tool version 4.5 produces speciation profiles that compute IVOC and NVOL in moles, which is consistent with all other model species in the gas-phase mechanisms. This is a change from previous versions of the Speciation Tool that generate profiles which compute NVOL emissions in grams. Air quality models that make use of IVOC and/or NVOL emissions from Speciation Tool version 4.5 can make use of the representative properties given in Table 4.

Table 4. Representative molecular weights and numbers of carbon atoms for model species IVOC and NVOL

Model Species	Representative C#	Representative MW (g/mol)
IVOC	12	160
NVOC	18	260

In addition, the mappings of inorganic species, which were primarily mapped to NVOL previously, were removed in order to ensure that they are not incorrectly treated as non-volatile organic species in the SOA modeling schemes of AQMs. These species are now implicitly mapped to UNK by the Speciation Tool along with all other species that are not mapped.

After the volatility classification of Task 1 was implemented, the CB6 and CB05 mappings of VOCs (excluding species mapped to IVOC and NVOL) were systematically reviewed using the general guidelines for mapping provided in Table 5, and mapping guidelines for difficult to map compound classes and structural groups shown in Table 6.

Table 5. General guidelines for mapping using CB6 structural model species.

CB6 Species Name	Number of Carbons	Represents
ALDX	2	Aldehyde group. ALDX represents 2 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. propionaldehyde is ALDX + PAR
IOLE	4	Internal olefin group. IOLE represents 4 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. 2-pentene isomers are IOLE + PAR. *Exceptions:* • IOLE with 2 carbon branches on both sides of the double bond are downgraded to OLE
KET	1	Ketone group. KET represents 1 carbon and additional carbons are represented as alkyl groups (mostly PAR), e.g. butanone is 3 PAR + KET
OLE	2	Terminal olefin group. OLE represents 2 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. propene is OLE + PAR. Alkyne group, e.g. butyne isomers are OLE + 2 PAR.
PAR	1	Alkanes and alkyl groups. PAR represents 1 carbon, e.g. butane is 4 PAR. See UNR for exceptions.
TERP	10	All monoterpenes are represented as 1 TERP.
TOL	7	Toluene and other monoalkyl aromatics. TOL represents 7 carbons and any additional carbons are represented as alkyl groups (mostly PAR), e.g. ethylbenzene is TOL + PAR. Cresols are represented as TOL and PAR. Styrenes are represented using TOL, OLE and PAR.
UNR	1	Unreactive carbons are 1 UNR such as quaternary alkyl groups (e.g., neo-pentane is 4 PAR + UNR), carboxylic acid groups (e.g., acetic acid is PAR + UNR), ester groups (e.g., methyl acetate is 2 PAR + UNR), halogenated carbons (e.g., trichloroethane isomers are 2 UNR), carbons of nitrile groups (-C≡N)
XYL	8	Xylene isomers and other polyalkyl aromatics. XYL represents 8 carbons and any additional carbons are represented as alkyl groups (mostly PAR), e.g. trimethylbenzene isomers are XYL + PAR
IVOC	12	Intermediate volatility organic compounds with saturation concentrations (C_i^*) in the range: $3 \times 10^3 \mu g\ m^{-3} \leq C_i^* < 3 \times 10^6 \mu g\ m^{-3}$
NVOL	18	Low volatility organic compounds with saturation concentrations (C_i^*) in the range: $C_i^*\ \mu g\ m^{-3} < 3 \times 10^3$

Table 6. Mapping guidelines for some difficult to map compound classes and structural groups

Compound Class/Structural group	CB model species representation
Chlorobenzenes and other halogenated benzenes	Guideline: • 3 or less halogens – 1 PAR, 5 UNR • 4 or more halogens – 6 UNR Examples: • 1,3,5-Chlorobenzene – 1 PAR, 5 UNR • Tetrachlorobenzenes – 6 UNR
Cyclodienes	Guideline: • 1 IOLE with additional carbons represented as alkyl groups (generally PAR) Examples:

	• Methylcyclopentadiene – 1 IOLE, 2 PAR • Methylcyclohexadiene – 1 IOLE, 3 PAR
Furans/Pyrroles	Guideline: • 2 OLE with additional carbons represented as alkyl groups (generally PAR) Examples: • 2-Butylfuran – 2 OLE, 4 PAR • 2-Pentylfuran – 2 OLE, 5 PAR • Pyrrole – 2 OLE • 1-Methylpyrrole – 2 OLE, 1 PAR
Heterocyclic aromatic compounds containing 2 non-carbon atoms	Guideline: • 1 OLE with remaining carbons represented as alkyl groups (generally PAR) Examples: • Ethylpyrazine – 1 OLE, 4 PAR • 1-methylpyrazole – 1 OLE, 2 PAR • 4,5-Dimethyloxazole – 1 OLE, 3 PAR
Triple bond(s)	Guideline: • Triple bonds are treated as PAR unless they are the only reactive functional group. If a compound contains more than one triple bond and no other reactive functional groups, then one of the triple bonds is treated as OLE with additional carbons treated as alkyl groups. Examples: • 1-Penten-3-yne – 1 OLE, 3 PAR • 1,5-Hexadien-3-yne – 2 OLE, 2 PAR • 1,6-Heptadiyne – 1 OLE, 5 PAR

Results

Following the volatility classification in Task 1, 40% of the existing CB mappings for SPECIATE 4.5 were assigned to IVOC or NVOL (Table 2). Of the remaining 60%, 30 inorganic species were assigned no mapping (and consequently mapped to UNK by the Speciation Tool) and 16 mappings were updated for consistency with the current guidelines. Molecular weights were added for 15 compounds. The updated mappings were incorporated in the Speciation Tool.

Task 3: Update SAPRC07 mappings

Mappings for the SAPRC07 mechanism were created for species added in SPECIATE 4.5 and low volatility species were assigned to IVOC or NVOL. VOCs are represented in SAPRC by lumped model species classified by structure and OH radical rate constant (kOH). Existing species were used as examples to create mappings for the added SPECIATE 4.5 compounds by analogy. For compounds where no analogous mappings existed, the AopWin v1.92 model in EPI Suite was used to estimate kOH. Excluding the species assigned to NVOL or IVOC, 140 new mappings were created and 15 were updated for consistency. The updated mappings were incorporated in the Speciation Tool.

Task 4: Create mappings for the CR1v2r5 mechanism

Methods and Results

The CRI mechanism is a condensed mechanism developed from the Master Chemical Mechanism version 3.1[6]. SPECIATE 4.5 VOCs were mapped by translating the updated SAPRC07 mappings (and CB06 mappings in some cases). All alkanes with 4 or more carbons were mapped to NC4H10 with the mapping factors equal to the ratio of the MW of the VOC to the MW of butane, which is consistent with the VOC mass weighting approach used in the development of the CRI mechanism[7,8]. Organic nitrates (RNO3 in SAPRC07) were treated as alkanes, and terpene mappings in SAPRC (TERP) were split evenly between APINENE and BPINENE except the compounds α-pinene and β-pinene. The mapping guidelines are provided in Table 7. The updated mappings were incorporated in the Speciation Tool.

Table 7. Translation from SAPRC07 mappings to CRI mappings

CRIv2r5 Species	SAPRC07 Translation
CH4	CH4
C2H6	ALK1
C3H8	ALK2
NC4H10	ALK3 + ALK4 + ALK5 (mapping weighted by the ratio of the MW of the compounds to the MW of butane)
C2H4	ETHE
C3H6	OLE1
TBUT2ENE	OLE2
APINENE	TERP/2
BPINENE	TERP/2
C5H8	ISOP
C2H2	ACYE
BENZENE	BENZ
TOLUENE	(ARO1 – BENZ*) + CRES + BALD
OXYL	ARO2
CH3OH	MEOH
C2H5OH	ETOH*
HCHO	HCHO
CH3CHO	CCHO
C2H5CHO	RCHO + GLY
CH3COCH3	ACET
MEK	MEK + PRD2
UCARB10	MACR + MVK + IPRD + BACL + MGLY
HCOOH	FACD

[6] http://mcm.leeds.ac.uk/CRI/

[7] Personal communication with CRI Developer Mike Jenkin

[8] Watson, L.A., D.E. Shallcross, S.R. Utembe, and M.E. Jenkin, A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 2: Gas phase mechanism reduction, Atmos. Environ., 42 7196-7204, 2008.

CRIv2r5 Species	SAPRC07 Translation
CH3CO2H	AACD
NROG	NROG
NVOL	NVOL

* CB06 model species

Recommendation

- Refine the current approach of mapping most terpenes evenly between APINENE and BPINENE.

Appendix A

Unmapped Species from SPECIATEv4.4 that Remain Unmapped for v4.5

Table A1. Unmapped species from SPECIATEv4.4 that remain unmapped for v4.5

Species ID, Name	Species ID, Name	Species ID, Name
2838: 3-Methylcholanthrene	2495: 2,4-Dimethylbenzoic acid	1016: G-decanolactone
2839: Oxalate anion	2496: 2,5-Dimethylbenzoic acid	1017: C27-20R5a(H),14(H),17(H)-cholestane , also noted as c27rch
2840: Acetate anion	2497: Myristoleic acid	1019: C27-20R5a(H),14a(H),17a(H)-cholestane , also noted as c27rac
2841: Formate anion (formic acid, ion)	2498: Trans-2-decenoic acid	1021: 1-Hexadecene
2847: Calcium Oxide	2499: o-Toluic acid	1022: C28-20R5a(H),14(H),17(H)-ergostane
2848: Aluminum Oxide (or Alumina)	2500: Maleic acid	1024: C28-20R5a(H),14a(H),17a(H)-ergostane
2849: Iron Oxide	2501: p-Toluic acid	1025: C29-20S5a(H),14a(H),17a(H)-stigmastane
2850: Potassium Carbonate	2502: m-Toluic acid	1026: C29-20R5a(H),14(H),17(H)-stigmastane
2851: Sodium Carbonate	2503: Salicylic acid	1027: Steroid-w , also noted as sterow
2852: Magnesium Oxide	2504: Ergosterol	1044: 4-formyl-guaiacol-TMS , also noted as f4gucl
2853: Sulfur Trioxide	2505: Tetradecanedioic acid	1056: Me-succinic acid-TMS
2854: Silica	2506: 1-Nitrobenzoepyrene	1066: Sitostane
2855: Titanium Oxide	2507: 20S 24RS-Ethylcholestane	1120: G-nonanoic lactone
2856: Lutetium	2508: 20R 24R-Ethylcholestane	1123: Trans-2-heptenal
2857: Oleic Acid (or cis-9-Octadecenoic Acid)	2509: 2-nitroanthracene	1127: 7-Hexadecene
2858: trans-Oleic acid (trans-9-Octadecenoic acid)	2510: 3-Nitrobenzoepyrene	1166: Carbon dioxide
2859: 2-Octadecenoic acid	2511: C19-tricyclic terpane	1174: Benzofluoranthenes
2860: 2-pentacosenoic acid	2512: C20-tricyclic terpane	1183: Organic carbon I
2861: Eicosanedioic acid	2513: C21-tricyclic terpane	1190: Elemental carbon II
2862: n-Octadecanol	2514: C22-tricyclic terpane	1194: Undecanoic-g-lactone
2863: 1-Nonadecanol	2515: C21--sterane	1254: A-Methylbiphenyl
2864: n-Eicosanol	2516: C21--sterane	1257: A-MePyMeFl
2865: Henicosanal	2517: C22--sterane	1267: B-Methylbiphenyl
2866: Docosanal	2518: C23--dimethyl-a-butylpodocarpane	1269: B-methylphenanthrene
2867: Tricosanal	2519: C24--dimethyl-a-methylbutylpodocarpane	1270: B-Methylpyrene
2868: Tetracosanal	2520: C25-tricyclic terpane	1275: Benzo(b+j+k)fluoranthene
2869: Pentacosanal	2521: C26-tricyclic terpane	1280: B-MePyMeFl
2870: 14-Methylhexadecanoic acid	2522: C26-tricyclic triterpane-22R	1281: Benzonaphthothiophene
2871: 16-Methyloctadecanoic acid	2523: C26-tricyclic triterpane-22S	1289: C-Methylbiphenyl
2872: n-Docasanoic acid	2524: C27-20S--cholestane	1290: C-methylfluorene
2873: Heptacosanedioic acid	2525: C27--Trisnorhopane	1293: 1-MeFl+C-MePyFl
2874: Methyl hexadecanoate	2526: C27--Trisnorneohopane	1296: C-MePyMeFl
2875: Methyl 14-methylhexadecanoate	2527: C27-tetracyclicterpane-22R	1312: D-MePyMeFl
2876: Methyl nonadecanoate	2528: C27-tetracyclicterpane-22S	1313: D-Trimethylnaphthalene
2877: Methyl heneicosanoate	2529: C28-20R--methylcholestane	1315: E-Methylpyrene

Species ID, Name	Species ID, Name	Species ID, Name
2878: Methyl heptacosanoate	2530: C28-20S--methylcholestane	1317: 2-ethyl-1-methylnaphthalene
2879: Ethyl docosanoate	2531: C28-20S--methylcholestane	1321: F-Methylpyrene
2880: Ethyl tetracosanoate	2532: C28-20R-Ba-diasterane	1326: G-Methylpyrene
2881: 9,12-Octadecenoic acid, methyl ester	2533: C28-20S-Ba-diasterane	1328: H-Trimethylnaphthalene
2882: Methyl eicosenoate	2534: C29-20S-Ba-diasterane	1352: 1,2,8-trimethylnaphthalene
2883: Methyl 2-docosenoate	2535: C28-aaB-Bisnorhopane	1357: Nitro-benzo(a)anthracene
2884: Behenic acid, methyl ester (or Methyl behenate; Methyl docosanoate; n-Docosanoic acid methyl ester)	2536: C28-tetracyclicterpane-22R	1387: 18a(H),21(H)-22,29,30-Trisnorhopane
2885: Methylparaben	2537: C28-tetracyclicterpane-22S	1390: 17a(H),18a(H),21(H)-28,30-Bisnorhopane
2886: 2-Phenylpropanoic acid	2538: C29-aB-Norneohopane	1391: 17a(H),21(H)-30-Norhopane , also noted as ab30nh
2887: Matairesinol	2539: C30-Tricyclicterpane-22R	1392: 17(H),21a(H)-30-Norhopane , also noted as ba30nh
2888: Cinnamaldehyde	2540: C30-Tricyclicterpane-22S	1393: 18a(H),21(H)-30-Norneohopane
2889: Hydroxyquinol (or 1,2,4-Benzenetriol; 1,2,4-Trihydroxybenzene)	2541: C32-22R--bishomohopane	1396: 17(H),21a(H)-hopane , also noted as bahop
2890: Phloroglucinol	2542: C32-22S--bishomohopane	1397: 22S-17a(H),21(H)-30-Homohopane , also noted as sabhhp
2891: Pyrogallol (or 1,2,3-Trihydroxybenzene; Pyrogallic acid)	2543: C33-22R--trishomohopane	1398: 22R-17a(H),21(H)-30-Homohopane , also noted as rabhhp
2892: 1,4-Naphthalenedione	2544: C33-22S--trishomohopane	1402: 22S-17a(H),21(H)-30,31,32-Trisomohopane
2893: Methoxynaphthol (or 2-Methoxy-1-naphthol)	2545: C34-22R--tetrakishomohopane	1407: C27-20S-13a(H),17(H)-diasterane
2894: 1,43,6-Dianhydro-D-mannitol	2546: C34-22S--tetrakishomohopane	1408: C27-20R-13a(H),17(H)-diasterane
2674: Trans-3-hexene; 3-methylcyclopentene	2547: C35-22R--pentakishomohopane	1409: C28-20S-13(H),17a(H)-diasterane
2675: 2-methyl-2-hexene; cis-3-heptene	2548: C35-22S--pentakishomohopane	1410: C29-20R-13a(H),17(H)-diasterane
2676: 3-methyl-trans-3-hexene; Trans-2-heptene	2549: Trisnorhopane	1411: C27-20S5a(H),14a(H)-cholestane
2677: 1-methylcyclohexene; 4-methylheptane	2550: c29 20R--ethylcholestane	1413: C27-20S5a(H),14(H),17(H)-cholestane , also noted as c27sbc
2678: 1-pentene; 2-butyne	2551: c28 20R--methylcholestane	1415: Ster45+40(cholestane)u
2679: Cyclohexene; 3-methylhexane	2552: c29 20r--ethylcholestane	1416: C28-20S5a(H),14a(H),17a(H)-ergostane
2680: 2,4,4-trimethyl-1-pentene; 2,3-dimethyl-2-pentene	2553: 22R-17a(H),21(H)-30-homohopane	1418: Ergostane , also noted as ergos
2681: 4-methyl-cis-2-pentene; 2-methylpentane (isohexane)	2554: 22S-17a(H),21(H)-30-homohopane	1423: Steroid-m , also noted as sterom
2682: 2,3-dimethylheptane; 2-methyloctane	2555: 4-Nitropyrene	1437: 4-ethyl-syringol-TMS , also noted as e4syrg
2683: Trans-1,3-dimethylcyclohexane; Cis-1,4-dimethylcyclohexane	2556: 2-Nitrofluorene	1438: 4-ethyl-guaiacol-TMS , also noted as e4gucl

Species ID, Name	Species ID, Name	Species ID, Name
2685: 4-nonene	2557: C27-20R5a(H),14a(H),17a(H)-cholestane	1460: Propylgyaiacol
2686: Iridium	2558: 20R--cholestane	2345: 1,3-Dimethylnaphthalene
2687: Hydrated sulfate	2561: 1-methoxy-2-Propanone	2347: anteiso-Hentriacontane
2648: 1-methylfluoranthene; 3-methylfluoranthene	2565: 2,2-Dimethylpropanal	2348: anteiso-Heptacosane
2649: Picene	2569: 1,3-dimethyldibenzothiophene	797: Elemental Carbon
2650: Dibenzoa,hpyrene	2586: Particulate Mercury	2349: anteiso-Hexacosane
2651: Heptatriacontane	2587: Oxidized Mercury	2350: anteiso-Nonacosane
2652: Octatriacontane	2588: Elemental Mercury	2351: anteiso-Octacosane
2653: Nonatriacontane	2298: Volatile Carbon	2352: anteiso-Pentacosane
2654: 2,3-dimethoxybenzoic acid	2302: Potassium ion	2353: anteiso-Tetracosane
2655: 2,5-dimethoxybenzoic acid	2303: Calcium ion	2354: iso-Dotriacontane
2656: 3,5-dimethoxybenzoic acid	2340: abieta-6,8,11,13-tetraen-18-oic acid	2356: iso-Heptacosane
2657: 2,4-dimethoxybenzoic acid	2342: Iso-tritriacontane	2357: iso-Hexacosane
2658: Palustric acid	2343: Ethylphenols	2358: iso-Octacosane
2660: Benzoafluoranthene	2344: Myosmine	2359: iso-Pentacosane
2661: Malonic acid (or 1,3-Propanedioic acid)	2035: Cyclopenta-anthracenes	2360: iso-Tetracosane
2662: Syringic acid	2060: Dimethylcyclobutanone	2361: Linoleic acid, 9,12-Octadecadienoic acid (9Z,12Z)-
2663: Protactinium	2147: Methyl C12 ester	2362: Monopalmitin, Hexadecanoic acid, monoester with 1,2,3-propanetriol
2664: Erbium	2222: Pentenyne	2363: Monostearin, Glyceryl monostearate
2665: Platinum	2296: 6,7-Dimethoxy-coumarin	2364: Hexatriacontane
2668: Particulate Water	292: Aluminum	2365: Pinonic acid
2669: Particulate Non-Carbon Organic Matter	294: Ammonia	2366: Stigmasterol
2670: Metal-bound Oxygen	296: Antimony	2369: N-Methylolacrylamide
2671: Other Unspeciated PM2.5	298: Arsenic	2370: Pentachloroethane
2672: Chromium(VI)	300: Barium	2371: Phosgene
2374: 1-Nonadecene	307: Bromine	2714: Trans-2-octene; Trans-1,2-dimethylcyclohexane
2375: 1-Eicosene	328: Cadmium	2739: Bromochloromethane
2376: 1-Heneicosene	329: Calcium	2742: 3,5-Dimethoxyphenol
2377: 1-Docosene	337: Chloride ion	2743: Syringealdehyde
2378: 1-Tricosene	347: Chromium	2744: 1,2-Benzenedicarboxylic acid, 4-methyl-
2379: 1-Tetracosene	379: Cobalt	2745: 1,2,3-Benzenetricarboxylic Acid
2380: 1-Pentacosene	380: Copper	2746: 1,2,4-Benzenetricarboxylic Acid
2381: 1-Hexacosene	436: Total carbon	2747: 1,3,5-Benzenetricarboxylic Acid

Species ID, Name	Species ID, Name	Species ID, Name
2382: 1-Heptacosene	468: Gallium	2748: 1,2,3,4-Benzenetetracaroboxylic Acid
2383: 1-Octacosene	477: Gold	2749: Galactosan
2384: 14-Methyl-hexadecanoic acid	487: Indium	2750: 1,6-anhydro-beta-D-mannopyranose (Mannosan)
2385: n-Hentriacontanoic acid	488: Iron	2751: Glycerol monooleate (9-Octadecenoic acid (9Z)-, 2,3-dihydroxypropyl ester)
2386: n-Dotriacontanoic acid	519: Lanthanum	2752: 1-Monostearin (Octadecanoic acid, 2,3-dihydroxypropyl ester)
2387: n-Tritriacontanoic acid	520: Lead	2753: Cholestanol (Cholestan-3-ol, Dihydrocholesterol)
2388: n-Tetratriacontanoic acid	525: Magnesium	2765: 1-Nitro-2-methylnaphthalene
2389: n-Pentatriacontanoic acid	526: Manganese	2766: 22S-17a(H),21b(H)-29-Homohopane
2390: n-Hexatriacontanoic acid	528: Mercury	2767: 22R-17a(H),21b(H)-29-Homohopane
2391: 10-Undecenoic acid	586: Molybdenum	2768: 22S-17a(H),21b(H)-29,30-Bishomohopane
2392: Tetradecenoic acid	612: Nickel	2769: 22R-17a(H),21b(H)-29,30-Bishomohopane
2393: Hexadecenoic acid	613: Nitrate	2770: 22R-17a(H),21b(H)-29,30,31-Trishomohopane
2394: Nonadecenoic acid	626: Organic carbon	2772: Magnesium ion
2395: Eicosenoic acid	649: Palladium	2773: Heptachlor
2396: Heneicosenoic acid	665: Phosphate	2774: Aldrin
2397: Docosenoic acid	666: Phosphorus	2775: Heptachlor epoxide
2398: Tricosenoic acid	669: Potassium	2776: Endosulfan
2399: Tetracosenoic acid	689: Rubidium	2777: 2,2,5-Trichlorobiphenyl
2400: Hexacosenoic acid	693: Selenium	2778: 2,4,5-Trichlorobiphenyl
2401: Undecanedioic acid	694: Silicon	2779: 2,4,4-Trichlorobiphenyl
2402: Hexadecanedioic acid	695: Silver	2780: 2,2,5,5-Tetrachlorobiphenyl
2403: Docosanedioic acid	696: Sodium	2781: 2,2,3,5-Tetrachlorobiphenyl
2404: Tetracosanedioic acid	697: Strontium	2782: 2,2,4,5,5-Pentachlorobiphenyl
2405: Pentacosanedioic acid	699: Sulfate	2783: 2,2,3,4,5,6-Hexachlorobiphenyl
2406: Hexacosanedioic acid	700: Sulfur	2784: 2,2,4,4,5,5-Hexachlorobiphenyl
2407: Methyl heptadecanoate	712: Thallium	2785: 2,2,3,4,4,5,5-Heptachlorobiphenyl
2408: Methyl eicosanoate	714: Tin	2786: Decachlorobiphenyl
2409: Heneicosanoic acid	715: Titanium	2895: Coumarin
2410: Methyl docosanoate	765: Uranium	2896: Methoxyhydroxycoumarin
2411: Methyl tricosanoate	767: Vanadium	2897: 5-Acetoxymethyl-2-furaldehyde
2412: Methyl tetracosanoate	777: Yttrium	2898: 16,17-Bisnordehydroabietic acid
2413: Methyl pentacosanoate	778: Zinc	2899: Secodehydroabietic acid
2414: Methyl hexacosanoate	779: Zirconium	2900: 19-Norabieta-8,11,13-triene

Species ID, Name	Species ID, Name	Species ID, Name
2415: Methyl octacosanoate	784: Ammonium	2901: Methyl deisopropyldehydroabietate
2416: Methyl nonacosanoate	785: Sodium ion	2902: Pimarinal
2417: Methyl triacontanoate	788: Carbonate	2903: Methyl 8,15-pimaradien-18-oate
2418: Methyl hentriacontanoate	789: Organic carbon II	2904: Methyl iso-pimarate
2419: Methyl dotriacontanoate	790: Organic carbon III	2905: Methyl 16,17-bisnordehydroabietate
2420: Methyl tetratriacontanoate	791: Organic carbon IV	2906: Dehydroabietal
2421: Methyl cis-9-octadecenoate	792: Pyrolyzed organic carbon	2907: Juvabione
2422: Methyl trans-9-octadecenoate	794: Elemental carbon I	2908: Todomatuic acid
2423: Methyl 9,12-octadecadienoate	795: Chlorine atom	2909: Friedelin
2424: Methyl tetracosenoate	796: Elemental carbon III	2910: trans-Methoxy-iso-eugenol
2425: Methyl hexacosenoate	810: Bromine Atom	2911: 7-Oxo-abieta-8,11,13,15-tetraen-18-oic acid
2426: 4-Vinylguaiacol	830: Sulfur dioxide	2912: Manoyl oxide
2427: Vanillic acid	831: Hydrogen Sulfide	2913: 18-Norisopimara-4(19),7,15-triene
2428: Methyl homovanillate	843: Sum of PM species	2914: Neoabietic acid
2429: Methoxyeugenol	853: Benz(a)anthracene-7,12-dione	2915: Pinostrobin chalcone
2430: cis-Methoxy-iso-eugenol	861: 9,10-dihydrobenzo(a)pyrene-7(8H)-one	2916: Umbelliferone
2431: Methoxybenzenediols	862: Benzanthrone; 7H-benzdeanthracen-7-one	2917: Monomethyl inositol
2432: Hydroxyacetophenone	865: 2,3-Benzofluorene	2918: Conidendrin
2433: 3,4,5-Trimethoxybenzoic acid	866: 5&6-methylchrysene	2919: Methyl-2-deoxomatairesinol
2434: Benzenepropanoic acid	869: 1,4-chrysenequinone	2920: Benzenetriols
2435: Diguaiacyl ethanes	870: 4H-cyclopenta(def)phenanthrene	2921: Ethyl hexacosanate
2436: Syringyl guaiacyl ethane	872: Dibenz(ah+ac)anthracene	2922: 20-Methyldocosanoic acid
2437: Disyringyl methane	875: 1,7-dimethylnaphthalene	2923: Octacosanal
2438: Disyringyl ethane	894: 7-methylbenzo(a)pyrene	2924: Hexacosanal
2439: 2-Deoxomatairesinol	897: C-Methylpyrene	2925: Eicosanal
2440: Phenylnaphthalene	910: 1,3-dinitronaphthalene	2926: 1-Hydroxyheptadecane
2441: 1-Naphthalenol	911: 1,3-dinitropyrene	2927: 1-triacontene
2442: 2-naphthalenol	912: 1,5-dinitronaphthalene	2928: Allobetul-2-ene
2443: Methylnaphthol	913: 1,6-dinitropyrene	2929: Allobetulone
2444: 1,43,6-Dianhydro--D-glucopyranose	914: 1,8-dinitronaphthalene	2930: Allobetulin (or (18)-19,28-Epoxyoleanan-3-ol)
2445: Tetramethoxyisoflavone	915: 1,8-dinitropyrene	2931: Betulin
2446: Dibenzofuranols	916: 1-nitronaphthalene	2932: Manool
2447: Benzonaphthofurans	917: 1-nitropyrene	2933: Dehydrojuvabione
2448: Deisopropyldehydroabietic acid	918: 2,7-dinitrofluorene	2934: 1-Heneicosanol
2449: 16-Nordehydroabietic acid	919: 2-nitrobiphenyl	2935: Syringaresinol dimethyl ether
2450: Secodehydroabietic acids	920: 2-nitronaphthalene	2958: 2,3,7,8-Tetrachlorodibenzo-p-dioxin (or 2,3,7,8-TCDD)

Species ID, Name	Species ID, Name	Species ID, Name
2451: Levopimaric acid	921: 3-nitrobiphenyl	2959: 1,2,3,7,8-Pentachlorodibenzo-p-dioxin (or 1,2,3,7,8-PeCDD)
2452: 18-Norabieta-8,11,13-triene	922: 3-nitrofluoranthene	2960: 1,2,3,6,7,8-Hexachlorodibenzo-p-dioxin (or 1,2,3,6,7,8-HxCDD)
2453: 19-norabieta-4,8,11,13-tetraene	923: 3-nitrophenanthrene	2961: 1,2,3,7,8,9-Hexachlorodibenzo-p-dioxin (or 1,2,3,7,8,9-HxCDD)
2454: 18-norabieta-4(19),8,11,13-tetraene	924: 4-nitrobiphenyl	2962: 1,2,3,4,6,7,8-Heptachlorodibenzo-p-dioxin (or 1,2,3,4,6,7,8-HpCDD)
2455: Dehydroabietane	925: 4-nitrophenanthrene	2963: 1,2,3,4,6,7,8,9-Octachlorodibenzo-p-dioxin (or OCDD)
2456: Methyl 6,8,11,13-abietatetraen-18-oate	926: 6-nitrobenzapyrene	2964: 2,3,7,8-Tetrachlorodibenzofuran (or 2,3,7,8-TCDF)
2457: Methyl 8,11,13,15-abietatetraen-18-oate	927: 6-nitrochrysene	2965: 1,2,3,7,8-Pentachlorodibenzofuran (or 1,2,3,7,8-PeCDF)
2458: Methyl dehydroabietate	928: 7-nitrobenz(a)anthracene	2966: 2,3,4,7,8-Pentachlorodibenzofuran (or 2,3,4,7,8-PeCDF)
2459: Methyl abietate	929: 9,10-dinitroanthracene	2967: 1,2,3,4,7,8-Hexachlorodibenzofuran
2460: Methyl-7-oxodehydroabietate	930: 9-nitroanthracene	2968: 1,2,3,6,7,8-Hexachlorodibenzofuran (or 1,2,3,6,7,8-HxCDF)
2461: Stigmasta-3,5-dien-7-one	931: 9-nitrophenanthrene	2969: 1,2,3,7,8,9-Hexachlorodibenzofuran
2462: Stigmasta-4,6-dien-3-one	932: G-Trimethylnaphthalene	2970: 2,3,4,6,7,8-Hexachlorodibenzofuran
2602: Heptadecane; Pristane	933: Abietic acid-TMS	2971: 1,2,3,4,6,7,8,9-Octachlorodibenzofuran (or OCDF)
2603: C29-20S5a(H),14(H),17(H)-stigmastane	936: Azelaic acid-TMS (Nonanedioic acid)	2972: N-Nitroso-N-diphenylamine
2604: Reactive gas-phase Mercury	938: 1,2,4-butanetriol-TMS	2973: N-Nitrosodimethylamine
2605: Nitrogen Monoxide (Nitric Oxide)	939: Cholesterol-TMS , also noted as chlsrl	2974: N-Nitrosomorpholine
2606: Nitrogen Dioxide	940: Cis-pinonic acid-TMS	2975: 22R-17a(H),21(H)-30,31,32-Trisomohopane
2607: Nitrous acid	942: Dehydroabietic acid-TMS	2976: Tetracontane
2617: Propylcyclopentane; Ethylcyclohexane	943: Docosanoic acid-TMS	2977: 1-Methyl-5-nitronaphthalene
2463: Stigmastan-3-ol	944: Eicosanoic acid-TMS	2978: 2,7-Dinitro-9-fluorenone
2464: Stigmastan-3-one	945: Elaidic acid-TMS	2979: 2-Nitrophenanthrene
2465: Beta-amyrone	946: Glutaric acid-TMS	2980: 2-Nitropyrene
2466: Alpha-amyrone	948: Henicosanoic acid	2981: 1-Methyl-4-nitronaphthalene
2467: Methyl indanones	949: Heptanedioic acid-TMS	2982: 2-Methyl-4-nitronaphthalene
2468: Delta-tocopherol; 8-methyltocol	953: Isophthalic acid-TMS	2983: 5-Nitroacenaphthene

Species ID, Name	Species ID, Name	Species ID, Name
2469: Beta-tocopherol; 5,8-dimethyltocol	955: Levoglucosan-TMS	2984: 1-Methyl-6-nitronaphthalene
2470: Gama-tocopherol; 7,8-dimethyltocol	960: Oleic acid-TMS	2985: Benzobfluorene
2472: Benzobkjfluoranthene	963: Phthalic acid-TMS (1,2-Benzenedicarboxylic Acid)	2986: Dibenzoa,epyrene
2473: 1-methylfluoranthene, C-methylpyrenefluoranthene	964: Picolinic acid-TMS	2987: Dibenz(a,h)acridine or Dibenzo(a,h)acridine
2475: Hexadecane & Norpristane	965: Sitosterol-TMS	2988: Dibenzo(a,i)pyrene or Dibenz(a,i)pyrene
2476: 8-abietic acid	971: Tetracosanoic acid-TMS	2989: Dibenz(a,j)acridine or Dibenzo(a,j)acridine
2477: 2,3 & 3,5-dimethylbenzoic acid	972: Tricosanoic acid-TMS	2990: Dibenz(a,j)anthracene or Dibenzo(a,j)anthracene; 3,4,5,6-Dibenzanthracene
2478: C28-20S5a(H),14(H),17(H)-ergostane	973: 2-nitrofluoranthene	2991: Dibenzo(a,l)pyrene or Dibenzo(def,p)chrysene; 1,2,9,10-Dibenzopyrene
2479: C27-20R5a(H),14a(H),17a(H)-cholestane & C29-20S13(H),17a(H)-diasterane	979: Butylated Hydroxytoluene	2992: Dibenzo(b,k)fluoranthene or Naphth(2,3-e)acephenanthrylene
2480: C29-20R5a(H),14a(H),17a(H)-stigmastane	986: Nitrobap	2993: 7H-Dibenzo(c,g)carbazole
2481: C27-20R5a(H),14(H)-cholestane	988: Nitropyrene	2994: 7,12-Dimethylbenz(a)anthracene
2482: C27-20R-13(H),17a(H)-diasterane	989: C27-tetracyclic terpane	2995: 4H-Cyclopentadefphenanthrene
2483: 17a(H),21(H)-22,29,30-Trisnorhopane	990: C28-tetracyclic terpane	2996: 5-Methylchrysene5-Methylchrysene
2484: 17a(H),18a(H),21(H)-25,28,30-Trisnorhopane	993: 17a(H),18a(H),21(H)-25,28,30-Trisnorhopane ,	2997: 9-phenylanthracene
2485: 2,3,5 & I-trimethylnaphthalene	994: 17a(H),21(H)-22,29,30-Trisnorhopane	3041: Ruthenium
2486: C-methylpyrene & methylfluoranthene	999: 17a(H),21(H)-Hopane	3042: Rhodium
2487: Methylpyrenesfluoranthenes	1000: 17(H),21(H)-Hopane	
2488: 2-methylglutaric acid	1004: 22S-17(H),21(H)-Hopane	
2489: 3-Methyladipic acid	1005: 22S-17a(H),21(H)-30,31-Bishomohopane	
2490: 2,6-dimethoxybenzoic acid	1006: 22R-17a(H),21(H)-30,31-Bishomohopane	
2491: Dodecanedioic acid	1008: 22R-17a(H),21(H)-30,31,32-Trishomohopane	
2492: 2,6-Dimethylbenzoic acid	1010: C27-20S-13(H),17a(H)-diasterane	
2493: 3-Methylglutaric acid	1011: C27-20R-13(H),17(H)-diasterane	
2494: 3,4-Dimethylbenzoic acid	1015: Caprolactone	

ATTACHMENT C

Species Mappings for CB6 and CB05 for use with SPECIATE 4.5

September 27, 2016

MEMORANDUM

To: Alison Eyth and Madeleine Strum, OAQPS, EPA
From: Ross Beardsley and Greg Yarwood, Ramboll Environ
Subject: Species Mappings for CB6 and CB05 for use with SPECIATE 4.5

Summary

Ramboll Environ (RE) reviewed version 4.5 of the SPECIATE database, and created CB05 and CB6 mechanism species mappings for newly added compounds. In addition, the mapping guidelines for Carbon Bond (CB) mechanisms were expanded to promote consistency in current and future work.

Background

The Environmental Protection Agency's SPECIATE repository contains gas and particulate matter speciation profiles of air pollution sources, which are used in the generation of emissions data for air quality models (AQM) such as CMAQ (http://www.cmascenter.org/cmaq/) and CAMx (http://www.camx.com). However, the condensed chemical mechanisms used within these photochemical models utilize fewer species than SPECIATE to represent gas phase chemistry, and thus the SPECIATE compounds must be assigned to the AQM model species of the condensed mechanisms. A chemical mapping is used to show the representation of organic chemical species by the model compounds of the condensed mechanisms.

This memorandum describes how chemical mappings were developed from SPECIATE 4.5 compounds to model species of the CB mechanism, specifically CB05 (http://www.camx.com/publ/pdfs/CB05_Final_Report_120805.pdf) and CB6 (http://aqrp.ceer.utexas.edu/projectinfoFY12_13/12-012/12-012%20Final%20Report.pdf).

Methods

CB Model Species

Organic gases are mapped to the CB mechanism either as explicitly represented individual compounds (e.g. ALD2 for acetaldehyde), or as a combination of model species that represent common structural groups (e.g. ALDX for other aldehydes, PAR for alkyl groups). Table 1 lists all of the explicit and structural model species in CB05 and CB6 mechanisms, each of which represents a defined number of carbon atoms allowing for carbon to be conserved in all cases. CB6 contains four more explicit model species than CB05 and an additional structural group to represent ketones. The CB05 representation of the five additional CB6 species is provided in the '*Included in CB05*' column of Table 1.

Ramboll Environ, 773 San Marin Drive, Suite 2115, Novato, CA 94998
V +1 415.899.0700 F +1 415.899.0707

In addition to the explicit and structural species, there are two model species that are used to represent organic gases that are not treated by the CB mechanism:

NVOL – Very low volatility SPECIATE compounds that reside predominantly in the particle phase and should be excluded from the gas phase mechanism. These compounds are mapped by setting NVOL equal to the molecular weight (e.g. decabromodiphenyl oxide is mapped as 959.2 NVOL), which allows for the total mass of all NVOL to be determined.

UNK – Compounds that are unable to be mapped to CB using the available model species. This approach should be avoided unless absolutely necessary, and will lead to a warning message in the speciation tool.

Table 1. Model species in the CB05 and CB6 chemical mechanisms.

Model Species Name	Description	Number of Carbons	Included in CB05 (structural mapping)	Included in CB6
Explicit model species				
ACET	Acetone (propanone)	3	**No** (3 PAR)	Yes
ALD2	Acetaldehyde (ethanal)	2	Yes	Yes
BENZ	Benzene	6	**No** (1 PAR, 5 UNR)	Yes
CH4	Methane	1	Yes	Yes
ETH	Ethene (ethylene)	2	Yes	Yes
ETHA	Ethane	2	Yes	Yes
ETHY	Ethyne (acetylene)	2	**No** (1 PAR, 1 UNR)	Yes
ETOH	Ethanol	2	Yes	Yes
FORM	Formaldehyde (methanal)	1	Yes	Yes
ISOP	Isoprene (2-methyl-1,3-butadiene)	5	Yes	Yes
MEOH	Methanol	1	Yes	Yes
PRPA	Propane	3	**No** (1.5 PAR, 1.5 UNR)	Yes
Common Structural groups				
ALDX	Higher aldehyde group (-C-CHO)	2	Yes	Yes
IOLE	Internal olefin group (R_1R_2>C=C<R_3R_4)	4	Yes	Yes
KET	Ketone group (R_1R_2 >C=O)	1	**No** (1 PAR)	Yes
OLE	Terminal olefin group (R_1R_2>C=C)	2	Yes	Yes
PAR	Paraffinic group (R_1-C<R_2R_3)	1	Yes	Yes
TERP	Monoterpenes	10	Yes	Yes
TOL	Toluene and other monoalkyl aromatics	7	Yes	Yes
UNR	Unreactive carbon groups (e.g., halogenated carbons)	1	Yes	Yes
XYL	Xylene and other polyalkyl aromatics	8	Yes	Yes
Not mapped to CB model species				
NVOL	Very low volatility compounds	*	Yes	Yes
UNK	Unknown	*	Yes	Yes

* Each NVOL represents 1 g mol^{-1} and low volatility compounds are assigned to NVOL based on molecular weight. UNK is unmapped and thus does not represent any carbon.

Mapping guidelines for non-explicit organic gases using CB model species

SPECIATE compounds that are not treated explicitly are mapped to CB model species that represent common structural groups. Table 2 lists the carbon number and general mapping guidelines for each of the structure model species.

Table 2. General Guidelines for mapping using CB6 structural model species.

CB6 Species Name	Number of Carbons	Represents
ALDX	2	Aldehyde group. ALDX represents 2 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. propionaldehyde is ALDX + PAR
IOLE	4	Internal olefin group. IOLE represents 4 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. 2-pentene isomers are IOLE + PAR. *Exceptions:* • IOLE with 2 carbon branches on both sides of the double bond are downgraded to OLE
KET	1	Ketone group. KET represents 1 carbon and additional carbons are represented as alkyl groups (mostly PAR), e.g. butanone is 3 PAR + KET
OLE	2	Terminal olefin group. OLE represents 2 carbons and additional carbons are represented as alkyl groups (mostly PAR), e.g. propene is OLE + PAR. Alkyne group, e.g. butyne isomers are OLE + 2 PAR.
PAR	1	Alkanes and alkyl groups. PAR represents 1 carbon, e.g. butane is 4 PAR. See UNR for exceptions.
TERP	10	All monoterpenes are represented as 1 TERP.
TOL	7	Toluene and other monoalkyl aromatics. TOL represents 7 carbons and any additional carbons are represented as alkyl groups (mostly PAR), e.g. ethylbenzene is TOL + PAR. Cresols are represented as TOL and PAR. Styrenes are represented using TOL, OLE and PAR.
UNR	1	Unreactive carbons are 1 UNR such as quaternary alkyl groups (e.g., neo-pentane is 4 PAR + UNR), carboxylic acid groups (e.g., acetic acid is PAR + UNR), ester groups (e.g., methyl acetate is 2 PAR + UNR), halogenated carbons (e.g., trichloroethane isomers are 2 UNR), carbons of nitrile groups (-C≡N).
XYL	8	Xylene isomers and other polyalkyl aromatics. XYL represents 8 carbons and any additional carbons are represented as alkyl groups (mostly PAR), e.g. trimethylbenzene isomers are XYL + PAR

Some compounds that are multifunctional and/or include hetero-atoms lack obvious CB mappings. We developed guidelines for some of these compound classes to promote consistent representation in this work and future revisions. Approaches for several compound classes are explained in Table 3. We developed guidelines as needed to address newly added species in SPECIATE 4.5 but did not systematically review existing mappings for “difficult to assign” compounds that could benefit from developing a guideline.

Table 3. Mapping guidelines for some difficult to map compound classes and structural groups

Compound Class/Structural group	CB model species representation
Chlorobenzenes and other halogenated benzenes	Guideline: • 3 or less halogens – 1 PAR, 5 UNR • 4 or more halogens – 6 UNR Examples: • 1,3,5-Chlorobenzene – 1 PAR, 5 UNR • Tetrachlorobenzenes – 6 UNR
Cyclodienes	Guideline: • 1 IOLE with additional carbons represented as alkyl groups (generally PAR) Examples: • Methylcyclopentadiene – 1 IOLE, 2 PAR • Methylcyclohexadiene – 1 IOLE, 3 PAR
Furans/Pyrroles	Guideline: • 2 OLE with additional carbons represented as alkyl groups (generally PAR) Examples: • 2-Butylfuran – 2 OLE, 4 PAR • 2-Pentylfuran – 2 OLE, 5 PAR • Pyrrole – 2 OLE • 1-Methylpyrrole – 2 OLE, 1 PAR
Heterocyclic aromatic compounds containing 2 non-carbon atoms	Guideline: • 1 OLE with remaining carbons represented as alkyl groups (generally PAR) Examples: • Ethylpyrazine – 1 OLE, 4 PAR • 1-methylpyrazole – 1 OLE, 2 PAR • 4,5-Dimethyloxazole – 1 OLE, 3 PAR
Triple bond(s)	Guideline: • Triple bonds are treated as PAR unless they are the only reactive functional group. If a compound contains more than one triple bond and no other reactive functional groups, then one of the triple bonds is treated as OLE with additional carbons treated as alkyl groups. Examples: • 1-Penten-3-yne – 1 OLE, 3 PAR • 1,5-Hexadien-3-yne – 2 OLE, 2 PAR • 1,6-Heptadiyne – 1 OLE, 5 PAR

These guidelines were used to map the new species from SPEICATE4.5, and also to revise some previously mapped compounds. Overall, a total of 175 new species from SPECIATEv4.5 were mapped and 7 previously mapped species were revised based on the new guidelines.

Recommendation

1. Complete a systematic review of the mapping of all species to ensure conformity with current mapping guidelines. The assignments of existing compounds that are similar to new species were reviewed and revised to promote consistency in mapping approaches, but the majority of existing species mappings were not reviewed as it was outside the scope of this work.
2. Develop a methodology for classifying and tracking larger organic compounds based on their volatility (semi, intermediate, or low volatility) to improve support for secondary organic aerosol (SOA) modeling using the volatility basis set (VBS) SOA model, which is available in both CMAQ and CAMx. A preliminary investigation of the possibility of doing so has been performed, and is discussed in a separate memorandum.