Users Manual for the
Emissions Speciation Database Files and Programs

William P. L. Carter
College of Engineering Center for Environmental Research and Technology
University of California, Riverside, CA 90521
carter@cert.ucr.edu

January 21, 2015

Summary

The structure and use of a database for making consistent assignments of model species for various chemical mechanisms when processing emissions input for various chemical mechanisms is described. This database can be used to assign model species for various chemical mechanisms to chemical categories used in five different anthropogenic total organic gas (TOG) profile databases or output by two different biogenic volatile organic compound emissions models. This involved creating a unified organic compound and mixture classification system, assigning compounds to mixtures, assigning model species to compounds for the various mechanisms, and developing procedures for processing these assignments and also making assignments for unknown, unassigned, or nonvolatile mass. The assignments are incorporated in two Excel spreadsheets that have macros to output the assignments in files for use with speciation processing systems. Fortran programs developed to use the assignments to make GSPRO files for SMOKE, and files output by the system can be imported directly into the Speciation Tool, are described. The use of the system for assigning model species to biogenic model outputs is also described. The database and associated programs and files are available at http://www.cert.ucr.edu/~carter/emitdb.

Contents

List of Tables

3

1. Introduction

Organic compounds differ significantly in their effects on air quality when they are emitted into the atmosphere, and air quality models need to take these differences into account when predicting the effects of organic compound emissions on formation of secondary pollutants such as ozone or particulate matter. The portion of the models that represents how reactions of primary pollutants such as organic compounds affect air quality is referred to as the chemical mechanism. Different mechanisms use different types and numbers of model species to represent the many hundreds of types of compounds that can be emitted, with the number of model species used to represent such emissions ranging from 10 for the Carbon Bond IV mechanism (Gery et al, 1988) to approximately 200 for the Master Chemical Mechanism (MCM) (MCM, 2015; Jenkin et al., 1997, 2003; Bloss et al, 2005). Organic compound emissions for various modeling episodes are generally given in terms of total mass of total organic gas (TOG) or reactive volatile organic compound (VOC) emissions[1] for the various locations and times in the modeling domain, with the speciation information contained in *speciation profiles* that give the mass fractions of various types of organic compounds emitted from the different sources. An important part of the air quality modeling process is determining which model species are used when processing VOC emissions data using the various profiles.

This is a large task because current speciation profile databases have on the order of 1000 or more chemical categories that need to be assigned model species. Different modeling groups use different databases containing organic speciation profile composition information as well as different chemical mechanisms. These speciation profile databases include many categories that are poorly characterized mixtures rather than single compounds, and also include categories for which the chemical compositions cannot be determined. Chemical mechanism developers or those implementing new mechanisms in air quality models can make different assumptions on how to treat these categories, yielding inconsistent treatment of mixtures with different mechanisms. An additional problem is that speciation profile databases have different chemical categorization systems, meaning that model species assignments made for different emissions databases may not necessarily be the same.

Model inputs for biogenic VOC emissions are generally prepared separately from outputs from models such as BEIS (Pierce et al, 1990; Schwede et al, 2005; SMOKE, 2014) and MEGAN (Guenther et al, 2012, MEGAN, 2014). In the case of BEIS v3 the model species assignments are made in files available from the EPA (EPA, 2008), while in the case of MEGAN v 2.1 the assignments for several mechanisms are built into the code. However, these assignments and procedures for adding new mechanisms are not documented.

To address these problems, we developed a mechanism assignment database and programs to make consistent assignments of model species for various mechanisms for TOG or VOC profiles in various emissions speciation database. These were originally described by Adelman et al (2005), and have undergone a number of revisions and enhancements since then, including adding support for new atmospheric chemical mechanisms, adding support for additional profile databases, revising the organization of the database to allow for use with larger mechanisms and to improve handling of unknown, unassigned, nonvolatile, and unreactive mass, and simplifying the programs used to incorporate

---

[1] The term "VOC" as used in a regulatory context in the United States refers to any volatile organic compound except for organic compounds that listed as exempt for regulation on the basis of low reactivity (U.S., 2014). Examples non-VOC organic compounds include methane, ethane, acetone, low-reactive fluorinated compounds, and certain compounds such as siloxanes that have been shown to be ozone inhibitors.

the mechanism speciation assignments when processing VOC emissions for atmospheric modeling. This database is available online (Carter, 2015a), and its development, assignments, uncertainties, and related issues are discussed by Carter (2015b). This users manual document describes the contents and structure of the database in more detail and describes the procedures and programs used to utilize the assignments in the database when processing VOC emissions for modeling using systems such as SMOKE (Smoke, 2013) or the Speciation Tool (Jimenez et al, 2013).

2. Description of the Database

This speciation assignment database is implemented in two Excel files, with the file SpecDB.xls containing the data needed to assign model species to the various chemical categories, and a separate file MechDB.xls containing the information about the model species used in the various chemical mechanisms[1]. SpecDB.xls also includes macros to output the assignments in files that can be used by other programs for processing emissions for modeling, such as SMOKE (SMOKE, 2013) or the Speciation Tool (Jiminez et al, 2013), as discussed later. The sheets contained in SpecDB.XLS are listed in Table 1, and the major macros used for data processing are listed in Table 2. The contents of MechDB.xls, which normally is not updated unless new mechanisms are added, are discussed in Section 5.1, below.

The sheet "Master List" in SpecDB.xls lists all the chemical categories used in the assignment database and gives information about the categories as indicated in Table 3. Each category has a unique ID code, referred to as the SpecDB "Unique ID" in the subsequent discussion, codes indicating the type of category, information on how it is assigned, and optional documentation information. Table 4 lists the various types of categories used in this database, the category codes used for each, and indicates how model species assignments for these categories are made.

The approach used for assigning model species was to make such assignments only to categories referring to single compounds, and not make such assignments directly to mixtures. For mixtures, the approach was to derive or estimate a composition for each mixture category in terms of individual compounds, and then use the model species for the compounds to derive those for mixtures. This requires treating the various categories in different ways, depending on whether they refer to compounds, known mixtures, or categories that cannot be speciated for various reasons. These are described further belowassignments for these categories are made. These are described further below

2.1. Category Name Mappings

The emissions profile databases or emissions models whose chemical speciation categories can be processed using the assignments in SpecDB.xls are listed in Table 5. Since these profile databases or models use category identification codes that are different from the unique ID codes used in on the "Master List" table and elsewhere in SpecDB.xls, the sheet "SAROAD Assignments" in SpecDB.xls gives the SpecDB unique ID code that corresponds to each category code for each profile database or emissions model listed in Table 5. The information on this sheet is summarized in Table 6. This is used by macros in SpecDB.xls to output files giving the model species assignments corresponding to the category codes in the database. Note that in some cases a single unique ID can be assigned to more than one profile database category, if it is judged that the profile category refers to the same compound or mixture.

---

[1] The earlier version of the database described by Adelman et al (2005) had all this information in a single file named EmitDB.xls,.

Table 1. List of worksheets and assignment tables in the SpecDB.xls file

Sheet name	Description
	General Information
Documentation	Lists the various worksheets in the file and describes columns of data where applicable. Also describes data contained in the MechDB.xls file and gives instructions on how to add new chemical mechanisms. Users should refer to this sheet for details on the worksheets not contained in this document. Also describes the font color conventions used throughout the workbooks.
Update History	Summarizes updates and modifications made to the database since it was first made available online.
Parameters and Commands	Contains the parameters used by the various macros in the worksheet and has control buttons for running the macros. The macros are listed in Table 2. Includes locations of output files and profiles used for representing unknown mixtures or unassigned compounds (see discussion), which can be modified by the user. Contains a list of supported mechanisms (with links to MechDB.xls) and allows the user to select which mechanisms are included in output files for the Speciation Tool. Contains a list of the supported speciation profile databases and the codes used to identify them. Contains additional information used by the macros or of potential interest to users such as numbers of categories, etc.
	Assignment Tables
Master List	Lists all of the chemical categories used in the database, giving the unique ID code, the description of the category, a code indicating the type of type of category as shown on Table 4, and information about mixtures for mixture categories. Mixture information includes the type of mixture and and notes and references for compositions. This also contains other information about the category such as whether it is used in any of the profile databases and a summary of assignment results made when the spreadsheet macros are run to process the assignments.
Compounds	Lists all categories in the database that are assigned to known single chemical compounds (type code 0). The information given for the compounds is listed in Table 7. For each compound this includes the unique ID, compound description, molecular weight, atoms numbers and atomic formulas, and codes indicating whether the compounds is volatile and assigned model species. Model species assignments are given for those compounds assigned model species, though such assignments were not made for all compounds. Structural information is also given for a number of compounds as indicated in Table 8.
Notes	Gives footnotes or codes used for some of the data columns in the "Master List" or "Compounds" sheet.
Atoms	This contains the atomic weights of the elements used for computing the molecular weights. The atomic weights are from IUPAC (2003).
Mix Profile Assignments	This gives the compositions of mixtures in terms of single chemical compound categories or categories whose compositions are defined and listed previously in the "Master List" sheet. One column gives the mixture name, another gives the compound or simpler mixture category, and a third gives the mass fraction. Additional columns are used for descriptions and to test that 100% of the mass is assigned.

Table 1 (continued)

Sheet name	Description
SAROAD assignments	This gives information needed to assign the chemical category codes used in the supported profile speciation databases to the unique ID codes used in this mechanism assignment database. The columns of data indicate the speciation profile database code, the category code or SAROAD number used in the profile database, the description of the category in the profile database, and the unique ID code used in this mechanism assignment database, which must match a code in the "Master Assignments" sheet described above. It also contains columns used for lookup purposes and summaries of results of processing by the speciation assignment macros.
Master Assignments	This sheet is updated and used by the macros for processing the speciation assignments and does not include any manually input assignment data. It contains columns giving the category ID, a compound contained in the category (or "UNKNOWN" or "NONVOL" if the composition is unknown), the weight fraction of the compound, and molecular weight and mechanism assignments for the compound.
SpTool Data	This contains data needed to prepare output files needed to implement the mechanism assignments for the Speciation Tool (Jimenez et al, 2013). In particular, this contains data needed to create the file SpeciesTable.csv, which must replace the "species_properties" data file when running the Speciation Tool with these assignments. This is necessary because the Speciation Tool must use the correct molecular weights for the categories when processing the model species assignments, and the molecular weights depend on the compositions assigned to the mixture in this mechanism speciation database. See Section 3.2.3.

2.2. Single Compound Categories

As indicated on Table 4, a majority of the categories in the merged dataset refer to single compounds, most but not all of which are assigned model species for the various chemical mechanisms. These categories are listed in the "Compounds" sheet of SpecDB.xls, which is where the chemical mechanism assignments are made. The information and assignments given for the various compounds is listed in Table 7, where the information is provided in all cases except as indicated otherwise in the table.

To aid in making and checking model species assignments, structural information is also given for many of the compounds, using the structural codes listed in Table 8. The table also indicates how many compounds were given a code of at least 1 for each code. Note that multifunctional or cyclic compounds may be assigned more than one code. These structural assignments were not made for all molecules, either because it had structures not listed on Table 8 or because it was not important in terms of mass contributions in the profiles and could not readily be determined from the atom numbers or previous model species assignments.

The assignments of model species for the compounds are also given in the "Compounds" sheet. Table 9 lists the chemical mechanisms for which the assignments were made and also indicates the formats of the assignments, which varied depending on the maximum numbers of model species used for each compound and whether reactivity weighting is used in any cases. The assignment columns are blank for compounds with no model species assignments for the mechanisms.

Table 2.   List of spreadsheet macros used in SpecDB to process mechanism speciation assignments and output files discussed in this document.

Macro	Description
	<u>Speciation Processing</u> (must be run whenever categories added or model species or mixture assignments are changed)
MergeAllCmpds	Determine mass fractions of individual compound categories listed in the "Compounds" sheet for all compound and mixture categories in the "Master List" sheet and put the results in the "Master Assignments" sheet. The mixture compositions are taken from the "Mix Profile Assignments" sheet.
CheckLMS	Checks all model species assignments in the compounds sheet in SpecDB.xls against the lists of model species for the mechanisms in MechDB.xls. Checks that compounds assigned for one lumped mechanism are assigned for all.
	<u>Produce files used as input to the ProfPre program</u>
OutputMasterCmpdAssts	Outputs file SpecAsn.dat giving the assignments of chemical compound categories to all categories in the database.
OutputMasterSARAssts	Outputs assignments of chemical categories used in the various profile speciation databases to categories used in this database. Output goes to the file *type*SARID.DAT, where *type* is the profile speciation database category listed in Table 2 of the main manuscript.
	<u>Produce files used as input to the ProfAsn program</u>
OutputSelectedMech	Outputs assignments files of model species for a selected mechanism. Files output are *MechID*ASN.dat and *MechID*MS.dat, where *MechID* is the code used for the mechanism (see Table 9.), the ASN.DAT file contains the model species assignments, and the MS.dat file lists the model species.
OutputUnknEMCfiles	Outputs chemical composition files for the mixtures used to make assignments for unknown mixtures or unassigned compounds, if the options to use these for this purpose are selected.
OutputAllMechs	Runs OutputSelectedMech for all mechanisms
	<u>Produces assignment files used by the Speciation Tool</u>
OutputForSpTool	Produces assignment files used by the Speciation Tool as described in Section 3.2.3. The names of these files can be specified in the "SpTool Data" sheet.
	<u>Produces model species assignment files selected categories</u>
OutputGSPRO	Produces GSPRO files that can be used by SMOKE for processing a selected category, with the option to give output in mass carbon rather than mass units. Useful for producing a GSPRO file to process BEIS 3 model output, which are in mass carbon units. The file name, the category ID (SARID), the mechanism ID and the output units are given in the "Parameters and Commands" sheet.

Table 3.   Information provided about all the chemical categories in the "Master List" sheet.

Column [a]	Description
Unique ID	The code identifying the category in the SpecDB database, which must be unique and not contain any blanks. For compounds with well-defined CAS numbers this can be a "C" + the CAS number, e.g., C74-82-8 as used for methane. Other categories are generally given designations based on the profile database ID and the category code it uses, as can be seen in Unique ID's given to many of the mixture categories. This code is then used in the "U.ID" column of the "SAROAD Assignments" sheet where the profile database categories are assigned Master List categories, as well as other places in the spreadsheet where this category is referenced.
Description	Description of the compound or mixture category. Not used for processing.
Type	Code number indicating the type of category, as given in Table 4. 0 is used for individual compounds, and it must also have an entry in the "Compounds" sheet. Code numbers 1 or 2 for mixtures whose compositions are specified on the "Mix Profile Assignments" sheet. Codes 3-9 are used for various types of unspeciated categories, and need not have entries in other sheets.
	<u>Compound or mixture assignments</u>
Type	The general type of assignment. "CMPD" means single compound, "SMX" or "CMP" means a mixture whose composition is specified, and "NVOL" means a nonvolatile unspeciated mixtures. This column is blank for unknown or unassigned mixtures presumed to be volatile.
Repres-entation	Unique ID of compound or mixture category that is used to represent the composition of this category. This can be one of the following: (1) The unique ID of the category itself if it is a single compound or if a mixture whose composition is given in the "Mix Profile Assignments" sheet; (2) the unique ID of a category listed above it in the "Master List" sheet if the assignments made for another category is used; (3) "UNKNOWN" if it is an unspeciated volatile mixture; or (4) "NONVOL" if it is an unspeciated nonvolatile mixture.
Code	Code number givng a subjective assessment of the quality of the mixture assignment. These include: 0 = compound representing itself; 1 = assignment considered appropriate; 2 = arbitrary assumption; 3 = guess that this composition is appropriate; and 4 = appropriatenss highly uncertain.
note1, note2, n3, n4, n5	Optional comments about the category or the composition assignment. "note1" and "note2" give comment text, and "n3" through "n5" give footnote codes where the text is defined in the "Notes" sheet. Not used for processing.
Used For	Indicates where this category is referred to elsewhere on this database (>1=yes, (blank)=no). Following codes used: S40 = used to make assignments for categories in SPECIATE 4 profile databases; SpT = used by the Speciation Tool; CA = used in the current California profile database; TX = used in the general Texas profile database; TXS = used in the Texas stationary source profile database; SAR = used in any category assignment in the "SAROAD assignments" sheet; Any = used anywhere. Codes for S40, CA, TX, or TXS may be greater than one if this SpecDB category is assigned to more than one profile category. Optional and not used for processing.

Table 3 (continued)

Column [a]	Description
	Assignment Summary: Summary data produced by the MergeAllCmpds macro. Should not be filled in manually.
Mwt	Molecular weight of the compound or effective molecular weight (1/number of moles in 1 gram mixture) of known mixtures. Blank for unspeciated mixtures.
NVOL	Mass fraction of nonvolatile compounds or unspeciated mixtures (separate columns for each).
NOASN	Mass fraction of unassigned compounds.
Exempt	Mass fraction of exempt (non-VOC) compounds.
Unknown	Mass fraction of unspeciated mixtures (volatile and nonvolatile).

[a] Columns used for internal or quality control purposes are not listed. See the "Documentation" sheet in SpecDB.xls for a complete list of all columns used in the "Compounds" sheet.

2.3. Mixture and Other Categories

A significant number of categories, representing non-negligible fractions of the mass in profiles and emissions, do not refer to single compounds, but are either mixtures of varying degrees of specificity or could not be assigned in terms of individual compounds. The various of types of mixture and other categories, and the methods used to assign model species for them, are summarized in Table 4.

For the purpose of making model species assignments, the mixture and other categories can be classified as known mixtures, unknown volatile mixtures, and unknown nonvolatile mixtures. The "Representation" column in the "Master List" sheet (see Table 3) indicates whether this is an unspeciated mixture and if so what type or gives the ID of a mixture category for which composition assignments are made, which either the ID of the mixture itself or the ID of a previously defined category assumed to have the same composition.

The compositions of the known mixtures in terms of compounds or previously defined mixtures are given in the "Mix Profile Assignments" sheet, as indicated on Table 10. Alternatively, the compositions can be specified as being the same for those for other mixture categories or compounds, using the "Representation" column on the "Master List" sheet (see Table 3), but only if they are listed before them on the "Master List" sheet. For that reason, simple mixtures of isomers, whose compositions are all given in terms of compounds, are listed first, followed by more mixtures of similar compounds that may be useful in specifications compositions of more complex mixtures, and finally by complex mixture categories such as "mineral spirits". The database includes two special complex mixture categories "REPUNK" and "REPNOA", which are optionally can be used to derive model species for unknown mixtures or unassigned compounds as discussed in Section 3.2.

When the "MergeAllCmpds" macro (see Table 2) is run, the macro lists the compositions of all categories in terms of individual compounds (or "UNKNOWN" or "NONVOL" if applicable) in the "Master Assignments" sheet, along with the mass fractions, molecular weights, and compound index numbers used by the macros that output model species assignments. All the contents of this sheet areproduced by this macro; there are no directly input assignment data there. This macro also outputs summary information about the categories on the "Master List" sheet, as indicated on Table 3.

Table 4.   List of types of chemical categories in the database and how they are assigned model species.

Code	Description
0	Single compounds or mixtures of isomers of higher molecular weight (usually $C_{\geq 12}$) compounds that are treated as single compounds for the purpose of this database. Most of these are compounds assumed to be volatile and assigned model species, but some are assumed to be nonvolatile and some are unassigned.
1	Simple mixtures of no more than 8 isomers. Model species are assigned based on those in the compounds in the mixture. The compositions are given on the "Mix Profile Assignments" sheet.
2	Complex mixtures whose compositions are given in terms of compounds or simple mixtures. Model species are assigned based on those in the compounds in the mixture. The compositions are either given on the "Mix Profile Assignments" sheet or are derived from compositions of another compound or mixture categories.
3	Complex mixtures of unknown composition that may be volatile. These are either assigned the special model species "UNKN" on a mass basis, or the moles of model species are assigned based on the composition of the "REPUNK" mixture.
4	Unspeciated mixtures that are judged to be nonvolatile. These are assigned the special "NVOL" model species on a mass basis.
5	Categories that are not appropriate for VOC profiles but are presumed to be volatile. Primarily Carbon Bond model species. Represented in same way as other unknown volatile mixture categories (e.g., code=3: UNKN or REPUNK).
6	Categories that are poorly named, ambiguous, or refer to compounds whose structures could not be determined, but which are presumed to be volatile. Represented in same way as other unknown volatile mixture categories (e.g., code=3: UNKN or REPUNK).
7	Categories that are not appropriate for VOC profiles because they refer to elements, salts, or PM. These are presumed to be nonvolatile and are represented by "NVOL" on a mass basis.
8	Categories that refer to compounds whose structures could not be determined, but are expected to be nonvolatile and are represented by "NVOL" on a mass basis.
9	Categories used in emissions profile speciation databases that have not yet been classified or assigned for this database because their contributions to the profile databases are negligible. These can be classified and assigned later if they become non-negligible in future profiles. Represented in same way as unknown volatile mixture categories (e.g., code=3: UNKN or REPUNK).

[a] Reflects numbers of categories as of November, 2014. These will change as the database is updated.

Table 5.   List of VOC emissions profile databases and biogenic VOC categories currently supported by the chemical mechanism assignment database.

ID	Description and Source
S40	SPECIATE version 4.1 through 4.4 obtained from the SPECIATE web site (EPA, 2014a). Uses 1809 chemical categories using numbers ranging from 1 to 2786. Some categories used in earlier versions are not used in version 4.4 but are retained for use with older profiles. Gas profiles in SPECIATE 4.4 uses 1692 of these categories.
EPA	Earlier speciation database maintained by the U.S. EPA that uses 5-digit SAROAD numbers to identify chemical categories. Obtained from the EPA in 2001 (Gipson, 2001). This has been superceded by SPECIATE, but is retained for compatibility with older EPA profiles.
CA	Speciation database maintained for the State of California by the California Air Resources Board (CARB) as of December, 2013 (Yang, 2013). Uses SAROAD numbers originally based on those used in the earlier EPA speciation database, but with categories added and some inconsistencies with the EPA codes. Contains 1206 categories.
TX	Speciation database maintained for the State of Texas by the Texas Commission on Environmental Quality (TCEQ) as of August, 2013 (Thomas, 2013). Uses SAROAD numbers originally based on those used in the earlier EPA speciation database, but with categories added and some inconsistencies with the EPA codes. Contains 935 categories.
TXC	Stationary source speciation categories maintained for the State of Texas by the Texas Commission on Environmental Quality (TCEQ) as of August, 2013 (Smith, 2013). Uses 541 5-digit "Contam. Codes". Additional work is needed to complete assignments for these categories.
BE3	Chemical categories output by version 3 of the BEIS biogenic emissions model (Pierce et al, 1990; Schwede et al, 2005; SMOKE, 2014). These consist of 34 categories, 30 being individual VOC compounds and 4 being VOC mixtures.
ME2	Chemical categories output in the first stage of speciation processing by the MEGAN version 2.1 biogenic emissions model (Guenther et al, 2012, MEGAN, 2014). These consist of 18 VOC categories, 13 being individual compounds, and 5 being mixtures. MEGAN also includes a program that speciates these into 150 compounds and outputs model species for selected mechanisms, but this cannot be updated or used for other mechanisms because its assignments are built into the code. However, the source code for this program (MEGAN, 2014) was used to inform the assignments in this work.
SDB	Uses the same chemical categories as used in the model species assignment database developed in this work. Not actually used in current emissions speciation profile databases, but this allows for processing profiles or mixtures that use these categories.

Table 6.   Information about speciation categories used in supported profile databases contained in the "SAROAD Assignments" sheet.

Column	Description
Type	The two-three character code for the speciation profile database or emissions model that uses this category designation, such as "S40" for SPECIATE 4.x, "CA" for California SAROAD codes, "BE3" for BEIS-3, etc. Must match a code in the "Codes for Category Names" table in the "Parameters and Commands" sheet.
SAROAD	The designation for the category in the profile database or emissions model that is specified in the "Type" column. Examples include "717" for toluene in SPECIATE 4.x databases, "45202" for toluene in the California profile database, and "ETHO" for ethanol output by BEIS 3.
TypeID	A combination of the Type and SAROAD codes (*Type*:*SAROAD*) used for lookups.
U.ID	The SpecDB master category code that is assigned to this profile category. For example, the code "C108-88-3" is used for all categories referring to toluene (CAS no 108-88-3).
Comment	Optional comment about the category. Only used in a few cases.
	The data in the following columns are derived by macros and should not be filled in manually.
Master DB name	The name of the SpecDB category, taken from the "Master List" sheet.
Check	Should be blank unless the category entry in the "U.ID" column is not found on the "Master List" sheet.
SDB Seq	Column number for the category in the Master List. Used for processing.
Assignment Summary	Summary information about the assignment similar to that given on the "Master List" sheet (see Table 3).
	The remaining columns contain optional information that is not needed for processing. They can be filled in by copying the calculation cells on the row below the table to rows added to the table, and deleting contents of cells with #N/A.
Key	A combination of the type and Unique ID that can be used for lookups
Type	The category type code (0=compound, 1-2 = known mixture, etc.) from the "Master List" sheet.
Sptool	"1" if this is a Speciate 4.x category that is output in files to be read by the Speciation Tool.
S44, etc.	"1" if this has the database type indicated, where "S44" is SPECIATE 4.4, and "DMS" is SAPRC detailed model species category.,

Table 7.   Information provided about chemical compounds in the "Compounds" sheet.

Column [a]	Description
Unique ID	Category ID. Must match a type "0" category in the "Master List" sheet. In most cases this is based on the CAS number for the compound if it is known.
Description	Name(s) of the compound.
Mwt	Molecular weight calculated from the atom numbers and the atomic weights on the "Atoms" sheet.
Asn OK?	Indicates if compound is assigned model species for the lumped mechanisms. Note that nonvolatile compounds are considered to be assigned model species. 1="yes".
Asn MS?	Indicates if the compound is volatile and assigned model species for the lumped mechanisms. 1="yes".
Is VOL?	Is this compound judged to be sufficiently volatile to be appropriate to assign to a gas-phase VOC model species? 1="yes".
Unsp. Cmpd?	Indicates if the category is used for unspecified isomers that are expected to be assigned the same model species in the mechanisms used. This may not be indicated in all cases for racemic mixtures or mixtures of stereoisomers of cyclic compounds. "1"="yes".
EPA Exmt?	Indicates if the compound is VOC-exempt according to the U.S. definition of VOC (US, 2014). 1="yes".
CAS	The CAS number for the compound. This should be the same as the CAS number used to form the ID. Not given if the CAS number is unknown or not well defined.
Atom Numbers	Gives number of atoms for compound. Separate columns used for C, H, O, N, S, F, Cl, I, and Br. The "X1" and "X2" columns are used to indicate other types and numbers of atoms if present, and "OA" is used to give the total atomic weight of such elements.
Formula	Atomic formula, giving atom types and numbers (e.g., "C2H6" for ethane).
Atoms	List of atom symbols for types of atoms in the molecule. (e.g., "CH" for ethane)
Chemical Groups	Types of chemicals or number of types of groups in the molecule. These are listed on Table 8. This is not given for all compounds, as indicated on Table 8.
Type	Code for the type of compound for tabulation purposes (Carter, 2015b).
Comment	Comments relevant to assignments for this compound given in a few special cases. Not given for most compounds.
Structure	If this compound is represented by a SAPRC detailed model species whose mechanism is generated using the SAPRC mechanism generation system (Carter, 2000, 2010a), this gives the structure used by the mechanism generation system for this purpose. Structures are also given in this format for a few other compounds, for potential future use by this system. Not given for all compounds.
Assignment Data	Model species for the various chemical mechanisms. The chemical mechanisms, mechanism codes, and formats used for the assignments are given on Table 9. Not given for all compounds.

[a] Columns used for internal or quality control purposes are not listed. See the "Documentation" sheet in SpecDB.xls for a complete list of all columns used in the "Compounds" sheet.

Table 8.   List of chemical group codes assigned to compounds in the database to assist in making model species assignments for new mechanisms.

Code	Description	No. [a]
Unkn	Not yet classified	213
n-alk	Normal alkanes	35
alk	Other Alkanes	421
cyc	Number of non-aromatic rings	387
C:C	Number of C=C double bonds	362
bz	Number of benzene rings (including in PAH's)	334
nap	Number of additional aromatic rings in PAH's (-CH=CH-CH=CH-). (Note that naphthalene has 1 bz and 1 nap.)	73
C:::C	Number of C-C triple bonds	17
alco	Number of C-OH groups (not carbonyl or first OH on phenol)	155
ald	Number of aldehyde (CHO) groups (except the first aldehyde group on an aromatic ring)	53
ket	Number of ketone (C=O) groups except on esters and acids	75
a-cb	Alpha-dicrbonyl compounds	7
ester	Number of ester groups	194
ethr	Number of C-O-C groups (not carbonyl)	144
acid	Number of carboxylic acid groups	57
phen	Number of phenol groups (aromatic ring + first OH)	51
bald	Has aldehyde group on aromatic rings (aromatic ring + first HCO)	7
amin	Number of amine (-NH2, >NH, -N<) groups	31
haro	Number of heterocycle containing aromatic rings (e.g., furan rings) (rings, double bonds, and atoms counted elsewhere)	12
C-Cl	Number of C-Cl bonds	60
C-F	Number of C-F or Si-F bonds	43
C-Br	Number of C-Br bonds	11
SiO	Number of -O-Si- groups, where Si bonded to 2 C's.	9

[a] Number of compounds with at least one of the structures. Note that multifunctional compounds can have more than one code and numbers greater than one.

Table 9.   Chemical mechanisms currently supported by the database, and codes, numbers of model species, and formates of the assignment data in the "Compounds" sheet, as of November, 2014.

Code [a]	Species [b]	Format [c]	Mechanism, References and Notes
			Lumped Mechanisms [d]
S99L	34	2	SAPRC-99 lumped mechanism (Carter, 2000) as implemented in CMAQ [e]
S07L	36	2	Standard version of the SAPRC-07 lumped mechanism (Carter, 2010a,b)
S07T	47	2	Toxics version of SAPRC-07 as implemented in CMAQ (Hutzell et al, 2012)
CS07	20	3	Condensed SAPRC-07 (Carter, 2010c)
S11L	38	2	SAPRC-11 with standard SAPRC-07 lumping (Carter and Heo, 2013; Heo et al, 2014)
S11D	331	2	Semi-detailed version of SAPRC-11, SAPRC-11D (Heo et al, 2014)
CB4	13	4	Carbon Bond version 4 (Gery et al, 1988)
CB05	18	4	Carbon Bond version 5 (Yarwood et al, 2005).
RADM	23	2	RADM2 mechanism (Stockewll et al, 1990)
RACM	42	3	RACM version 2 (Goliff et al, 2013)
			Detailed or Explicit Mechanisms [f]
DMS	1104	1	SAPRC detailed model species (DMS) for which SAPRC-07 reactivity values have been derived (Carter, 2010a,b).
MCM	203	1	Master Chemical Mechanism as used in airshed model calculations provided by Ying (2013).

[a] Code used within the database to identify the mechanism

[b] Number of model species, including "NROG" and "NVOL" but not "NOASN" or "UNKN".

[c] Format used for model species assignments in the "Compounds" sheet of SpecDB.xls, as follows:

   1 Assignments are in a single column, with one mole model species assigned per mole of compound, except for "NVOL", which are on a mass basis.

   2 Assignments are in two columns, the first giving the model species and the second giving the moles of model species per mole of compound.

   3 Assignments are given in four columns, with the first and third giving the model species and the 2nd and 4th giving the moles of model species per mole of compound. The 2nd and 3rd columns are blank if there is only one model species assigned for the compound.

   4 One column is used for each model species in the mechanism, where each column gives the moles assigned for the model species. For "NROG" and "NVOL" the columns give the mass assignments.

[d] Model species assignments are made for all the lumped mechanisms or for none.

[e] Model species for benzene and sesquiterpenes have been added.

[f] Some compounds where assignments are for the lumped mechanisms may not assignments of SAPRC detailed model species or MCM explicit species.

Table 10.   Information about mixture compositions given in the "Mix Profile Assignments" sheet.

Column	Description
Mix	The Unique ID for the mixture whose composition is being assigned
Component	The Unique ID for a compound or a previously defined mixture giving its composition. "Previously defined" means that the mixture used to derive the composition is above it in the "Master List" sheet. The order of mixture definitions in the "Mix Profile Assignment" sheet is not relevant to this.
Wt %.	The weight fraction of the mixture, given as fraction. Must sum up to exactly 1 (not 100).

The other fields are derived by calculation or macros and do not need to be entered manually.

Column	Description
IncC, IndM	Row numbers on "Compounds" or "Master List" sheet that are used by the macros for processing. These are automatically updated whenever these macros are run.
Description	The description of the component, taken from the "Master List" sheet. Not used for processing. Derived by calculation. [a]
Count	The number of components defined thus far for this mixture. Not used for processing.
Sum	The mass fraction of compounds defined thus far for this mixture. If the totals are greater than 1, the "minimum" and "maximum" numbers in the top columns would be something other than 100%. Not used for processing. Derived by calculation. [a]
Mixture Description	The description of the mixture from the "Master List" sheet. Only given for the first component of the mixture. Not used for processing. Derived by calculation. [a]
Seq	Used for sorting. Not used for processing.

[a] Formulas to calculate the data for these cells are given in the "Calc:" row below the table. They can be copied to the new cells to obtain the appropriate data, then converted to values by copying then using "paste special" and selecting "values".

3. Speciation Processing Procedures

The methods used to process the speciation assignments to produce the data used by emissions processing programs to assign model species to speciation profiles are discussed in this section. First we discuss the general methods applicable to all emissions speciation profile processing systems, and then we discuss those that are specific for processing emissions speciation using SMOKE (2013) and the Speciation Tool (Jimenez et al, 2013). These methods can be used with other emissions speciation profile processing systems by adapting or modifying the files produced for one of these two systems.

3.1. Treatment of Mass Assigned to Moles of Model Species

Most but not all of the mass in the speciation profiles are assigned to chemical categories that refer to compounds or mixtures of compounds that are assigned to moles of model species in the various mechanisms. For such categories the objective is to determine the moles of model species assigned per unit mass of emissions in these categories, and to output files containing this information that can be used by the speciation profile processing programs to determine moles of model species per unit mass of emissions for each profile. This involves the following steps:

1. Determine which category in the SpecDB database corresponds to each category used by the emissions speciation profile database being processed.

2. Determine mass fractions of individual chemical compound categories for all of the compound and mixture categories in the database and put the results in the "Master Assignments" sheet of SpecDB.

3. Use the molecular weights and molar model species assignments for the compounds on the "Compounds" sheet to determine the moles of model species per unit mass of each compound

4. Use the mass fractions of the compounds assigned to each SpecDB category and the moles of model species per mole of compound derived in the above step to determine moles of model species per unit mass of category.

5. Output the results in various files needed by the separate emissions speciation profile processing programs. Different files and programs are used when preparing for use with SMOKE or the Speciation Tool.

The second step in this process is done using the the "MergeAllCmpds" macro in SpecDB (see Table 2), which should be run whenever categories are added or assignments are changed. The other steps are done either by other macros in the spreadsheet or by separate Fortran programs as described later in this document.

Some emissions processing systems, such as the Speciation Tool (Jimenez et al, 2013) require a molecular weight in order to make model species assignments for the category. For this purpose, the effective molecular weight of a known mixture is defined as the mass of the mixture divided by the number of moles in that mass, which is given by 1 / sum (mass fraction of each component / molecular weight of the component). Note that the effective molecular weights of these mixtures are defined based on the molecular weights of all the components, not just those in the reactive and volatile fraction.

3.2. Treatment of Unknown, Nonvolatile and Unreactive Mass

Not all categories in the database are completely assigned to reactive model species. Some of the mass is treated as nonvolatile or as unreactive and therefore are not assigned to reactive model species, and some of the mass cannot be assigned to any compounds and must be treated as unknowns. Special model species NVOL, NROG, and UNKN, are used to represent this mass, with each being output in terms of mass fractions rather than molar assignments. This is because moles either cannot be assigned to these categories, or are poorly defined because compounds of unknown or widely varying molecular weights are being lumped together. These are discussed further below.

Non-volatile mass (NVOL): The special species "NVOL" is used to represent emitted mass that is to be treated as nonvolatile. This can include mass identified as individual compounds that are judged to be nonvolatile, as well as mass whose chemical composition is unknown but is characterized as nonvolatile. This is output as mass fractions rather than moles of model species because moles cannot be determined for the unspeciated nonvolatile categories and also because mass is a more appropriate unit for use when modeling condensed species such as organic particulate matter. For most model applications the nonvolatile species in the VOC profiles are probably not significant compared to primary PM assigned to PM profiles, so output of the NVOL mass can probably be ignored. But this may not be the case in all modeling scenarios.

Unreactive mass (NROG): The special species "NROG" was added to all mechanisms to represent identified chemical compounds that are judged to be unreactive and that do not have model species otherwise assigned to them in the mechanism. At present the database has no unspeciated mixtures that are judged to be unreactive, but should any be added in the future this species could be used for these as well. The Carbon Bond mechanisms use a model species ("UNR") to represent both unreactive carbon in reactive compounds and also carbon in reactive compounds, so NROG is not assigned in any compounds in those mechanisms. However, it is retained on the model species lists for those mechanisms for consistency.

The unreactive compounds represented by NROG are output as mass fractions of NROG rather than moles of reactive model species, similarly to the treatment of nonvolatile compounds represented by NVOL. This is because unreactive compounds have widely varying molecular weights, so moles of such model species are not particularly meaningful unless different model species are used for compounds of different molecular weight ranges.

Note that if modelers believe that it is useful to track unreactive moles, moles carbon or individual unreactive compounds they can use mechanisms that have separate model species for this purpose (as is the case for the Carbon Bond mechanisms). The system allows for the possibility of some compounds being represented as NROG mass in some mechanism and being assigned to molar model species in other mechanisms. In the latter case the unreactive model species are treated the same as reactive model species as discussed in the previous section. Otherwise, the models can either ignore this input of NROG mass or use the input and track the transport of the unreactive mass, depending on the desired application.

Unknown Mass (UNKN): The database and associated macros and programs and macros provide two options for treatment of unknown volatile mass. One is simply to output mass emissions of the special species "UNKN" to account for the unknown mass. This has to be output as mass fractions because moles cannot be determined when molecular weights are unknown. The models can either ignore this unknown mass, track its transport, or represent it by reactive model species as determined by the modeler. One approach in this regard is to normalize emissions assigned to model species, which is in effect assuming that unknown emissions have the same compositions as known emissions from all

sources. This is not a particularly good option because known emissions contain large fractions of low molecular weight compounds such as ethene that are not representative of the (generally higher molecular weight) compounds likely to be in unspeciated mixtures. Note that unknown mass is not negligible in current profiles, so ignoring it is not recommended because it introduces a consistent negative bias in terms of reactivity of emissions being modeled.

In order to give modelers a better option for treating unknown mass rather than ignoring it entirely or representing it by emissions from all sources, the speciation database system allows for the option of using a standard "unknown mixture" profile to determine model species to represent this unknown mass. Although the most appropriate composition for this mixture is highly uncertain, at least it allows the reactivity of unknown mixtures to be represented without introducing bias, and with consistent assumptions regarding the composition regardless of the mechanism used. If this option is chosen, then mass emissions for "UNKN" are not output, but instead the system outputs moles of model species assigned to this mixture. The results are the same as if all of the unspeciated mixture profiles other than those judged to be nonvolatile were assigned to the compounds in this "unknown mixture" profile. However, unknown mixture assignments are not part of mixture database itself but are made separately, either by the Fortran programs using data files output by SpecDB in the case of processing emissions for SMOKE, or by the macros that produce the speciation assignment files that are used by the Speciation Tool. This is discussed for the specific cases below.

To aid in the use of this second option, Adelman et al (2005) derived an "unknown mixture" profile to derive model species for unknown mixtures, and that is used for the current system without modification. This mixture, called "REPUNK", was derived based on estimates of the types of compounds likely to be in unknown mixtures, which are generally of higher molecular weight than those in better characterized emissions. Although the most appropriate composition for this mixture is highly uncertain, at least it allows the reactivity of unknown mixtures to be represented without introducing bias, and with consistent assumptions regarding the composition regardless of the mechanism used. If this option is chosen, then mass emissions for "UNKN" are not output, but instead the system outputs moles of model species assigned to this mixture. As with known mixtures, the moles of a model species per unit mass of the mixture are derived from the mass fractions of the compounds in the mixture times the moles of the model species per unit mass of the compound.

The system allows modelers the option of using other mixtures to derive model species to represent unknown mixtures in their model inputs. To do this, add the desired mixture on the "Master List" sheet and specify its composition to the "Mix Profile Assignments" sheet and as discussed in Section 5.3. The composition must consist only of volatile compounds that have model species assignments, though unreactive compounds can be included if appropriate. Then change the mixture name given at the location on the "Parameters and Commands" sheet designated as "Mixture for UNKN" from "REPUNK" to the name of the desired mixture. The "UpdateAll" macro should be re-run after these changes are made.

The system also allows modelers the option of having the emissions processing procedures output mass emissions of the special model species "UNKN" to represent unspeciated volatile mixtures rather than model species assigned to a mixture. For processing for SMOKE of other emissions systems using the ProfAsn program, this is done by not input records specifying compositions for "UNKN" as discussed in Section 4.1.1. For processing using the Speciation Tool, the "Mixture for UNKN" given on the "Parameters and Commands" sheet needs to be blank when the "OutputForSPtool" macro is run.

3.2.1. Treatment of Unassigned Compounds

Model species assignments were not made for all compounds. These unassigned compounds are treated differently than unknown mixtures because the types of compounds involved are expected to be quite different. The database and associated macros and programs and macros provide two options for treatment of compounds not assigned to model species. One is simply to output molar emissions of the special species "NOASN" to account for the moles of these unassigned compounds. The models can either ignore these unassigned moles, track their transport, or represent them by reactive model species as determined by the modeler. Note that unassigned mass is not significant in current profiles, though this may not always be the case when future profiles are developed.

In order to give modelers a better option for treating unassigned compounds rather than ignoring them entirely, the system allows for the option of using a standard "unassigned mixture" profile to determine a distribution of model species to represent the moles of these unassigned compounds. Unlike the treatment of unknown mass, the model species are assigned on a mole per mole basis rather than as moles per unit mass, with the same number of moles of model species being used per mole of unassigned compound regardless of the molecular weight of the unassigned compound. The moles of model species per mole of compound is derived from the moles of model species per total moles of compounds in the "unassigned mixture." This is because higher molecular weight compounds tend to be less reactive than lower molecular weight compounds, and making assignments in the same way as done with unknown mixtures would increase the moles of model species and therefore reactivity as the molecular weight of the unassigned molecule increases. Although uncertain, this treatment on a molar basis removes the biases introduced by using this approach.

To aid in the use of this option, a standard "unassigned mixture" profile, called REPNOA, is included as mixture category in the SpecDB database. This is different than the mixture used for unknown mixtures because the distributions of compounds represented are expected to be different. For this purpose, Adelman et al (2005) derived a mixture consisting of a molecule-weighted average of compounds represented by SAPRC-99 mechanism in each SAPRC-99 chemical class. This mixture is used in the current system except that compounds with fewer than 3 carbons and nonvolatile compounds are removed. Although the appropriateness of this mixture is highly uncertain, at least it is better than introducing biases by ignoring the reactivity of unassigned compounds entirely, or treating them inconsistently when using different mechanisms. Modelers can chose different mixtures for this purpose if desired.

If use of a mixture than REPNOA is desired for representing unassigned compounds, then the specification of "Mixture for NOASN" in the "Parameters and Commands" sheet needs to be changed to specify a different mixture, and that mixture should be specified for NOASN when ProfAsn is run (see Section 4.1.1). That mixture needs to be included on the "Master List" sheet and its composition to the "Mix Profile Assignments" sheet and as discussed in Section 5.3. The mixture can include only volatile compounds that are assigned model species or designated as unreactive. If it is desired to output moles of "NOASN" rather than model species for this mixture, the NOASN input can be omitted when ProfAsn is run the "Mixture for NOASN" name on the "Parameters and Commands" sheet should be blank when the "Output for SPtool" macro is run.

3.2.2. Processing Procedures and Files for SMOKE

The SMOKE emissions processing system uses GSPRO files to assign moles of model species to mass emissions of various profiles, so the objective of processing for SMOKE consists of producing GSPRO files for various speciation profile databases for various mechanisms. These files cannot be output by SpecDB.xls because the compositions of the individual profiles are not part of the SpecDB

database. Instead, SpecDB.xls outputs assignment files that Fortran programs included in this distribution can use to produce GSPRO files from input files containing compositions of the profiles and the profile database's speciation categories.

The SpecDB assignment files used for this purpose are output by various macros within SpecDB.xls (see Table 2) which are run using controls in the "Parameters and Commands" sheet. These files, and the macros producing them, are as follows:

- *ProfDBcode*SARID.DAT[1] contains the mappings of the chemical categories used in the supported profile databases to the category IDs used in SpecDB. The first part of the file name, *ProfDBcode* is the code for the profile database, as indicated in the first column in Table 5. For example, the mappings for the SPECIATE 4.x categories to the SpecDB categories are in the file S40SARID.DAT. There is one such file for each profile database listed in Table 5. These are output by the "OutputMasterSARAssts" macro.

- SpecAsn.DAT contains the compositions of all the speciation categories in SpecDB in terms of individual compounds in the database. These are needed in order to obtain model species assignments for the categories, since model species assignments are only made to categories designated as individual compounds. This file is independent of the profile database or chemical mechanism. It is output by the "OutputMasterCmpdAssts" macro.

- *MechID*MS.DAT contains the list of model species for each supported chemical mechanism. The first part of the file name, *MechID* is the code used for the mechanism, as indicated by the first column in Table 9, and there is one such file for each mechanism listed on the table. For example, CB05MS.DAT lists the model species in the CB05 mechanism. The information in the file includes the model species name, the number of carbons, and the molecular weight. These files are output by the "OutputAllMechs" macro.

- *MechID*asn.DAT contains the model species assignments for the mechanism identified by *MechID* for each individual compound category in the database, and also other information about the compounds. As with the MS.DAT files, *MechID* is the code for the mechanism as indicated in Table 9 and there is one such file for each supported mechanism. For example, CB05ASN.DAT gives the model species assignments for the CB05 mechanism. In addition to the model species assignment, these files contain the molecular weight, carbon number, and description of the compound, and also an indication of whether the compound is on the EPA's exempt list. The carbon numbers and descriptions are not used for speciation processing, but the programs need to know if the compound is on the VOC exempt list in order to derive VOC-to-TOG factors (see discussion below). The formats of these files depend on the type of mechanism, as discussed in Section 4.2. These files are also output by the "OutputAllMechs" macro.

- *UnknMixID*.EMC and *UnasnMixID*.EMC contain the category IDs and weight fractions of the individual compounds in the mixtures used to make model species assignments for unknown mixtures and unassigned compounds, if the options to use these mixtures is chosen. *UnknMixID* and *UnasnMixID* are the names of these mixtures, which are "REPUNK" and "REPNOA" by default. The "Parameters and Commands" sheet give the names of the categories used for these mixtures, or blanks if they are not used, as indicated by comments in the sheet. These files also contain flags indicating whether the compounds are volatile or VOC exempt[2] and the names or

---

[1] Portions of file names or options that are variable are indicated using *a smaller font and italics* throughout this document.
[2] Note that the presence of exempt VOCs in mixtures used to derived model species for unknown mixtures or unassigned compounds does not affect how the system calculates exempt fractions in profiles. It is assumed that unknown mixtures have no exempt compounds, and the exempt status of unassigned

descriptions of the compounds. These files are output by the "OutputUnknEMCfiles" macro, which does nothing if the names of both of these mixtures are blank.

All of the macros used to output the files listed above are executed when the "UpdateAll" macro is run on the "Parameters and Commands" sheet, which first runs the macros to compile and update the assignments that need to be run when the database is modified. This assures that the output files incorporate the current assignments, and that they are internally consistent. For example, if the category lists in the "Master List" or "Compounds" sheet are modified or sorted, the data in the output files would not be valid unless the "MergeAllCmpds" macro is run first.

Adelman et al (2005) describe several Fortran programs to use output files from an earlier version of this SpecDB database to process emissions profiles for SMOKE. These programs have subsequently been modified and simplified, and the capability to produce modified SAPRC mechanisms with variable lumping is no longer supported[1]. Currently, the necessary processing to prepare input files for SMOKE from files output by SpecDB.xls is done by running a single program, ProfAsn, though ProfAsn requires use of a separate program, ProfPre, as part of its processing.

More detailed information about running the ProfAsn and ProfPre and the files they produce is given in Section 4.1. Briefly, ProfAsn prepares the GSPRO files used by SMOKE for assigning model species for a given mechanism to profiles from a speciation profile database. The name of file(s) containing the profile data, the designation of the chemical mechanism and the profile database category formats are specified when ProfAsn is run. If mixtures are to be used to assign model species to unknown mass or unassigned compounds, the names and compositions of these mixtures are also included in ProfAsn input. In addition to outputting a GSPRO file with model species assignments for all the profiles input (which can come from multiple files from different profile databases), it also outputs a CSV file giving various summary information about the profiles. This includes the mass inputs for the profiles (which should be 100% if the profiles are properly normalized), the mass fractions assigned to reactive, inert, and nonvolatile model species, mass fractions of exempt compounds (see discussion of VOC-to-TOG ratios, below) and the mass fractions that are unknown mixtures or unassigned compounds and how they are represented.

ProfAsn calls a separate program, ProfPre, to determine the mass distribution of individual compounds in the profiles from the distributions of categories in the profile database. This involves first determining which SpecDB category corresponds to each profile category used, and then determining which compounds are assigned to the mixture categories. Users do not need to run ProfPre separately unless they want to retain the files containing the profile compositions in terms of individual chemical compounds in the SpecDB database. These are needed by ProfAsn but not for subsequent emissions processing by emissions processing systems. But this information may be of interest in some cases.

3.2.3. Processing Procedures for the Speciation Tool

The Speciation Tool (Jimenez et al, 2013) is an application written in PostgreSQL to prepare speciation profile data for model input. The current version of the Speciation Tool is 3.1, and the software

---

compounds is based on the assignment for the compound itself. Therefore, exempt fractions in the mixtures used for unknowns or unassigned compounds are ignored.

[1] This capability was implemented by a program called MechPro for preparing CMAQ mechanism definition files for versions of SAPRC-99 with differing lumping approaches (Adelman et al, 2005). It was only used in a few special studies and the files necessary to use it for updated versions of SAPRC have not been prepared. This capability can be restored in the future if there is sufficient interest and support for preparing the necessary files for updated versions of SAPRC.

and the data files it uses can be downloaded from the CMAS web site (CMAS, 2014), with the files in the current distribution dated July 23, 2013. SpecDB.xls contains a macro called OutputForSPtool to output the assignment data into four CSV files that can be imported into the Speciation Tool in order to implement the assignments for selected mechanisms in SpecDB. The controls to run this macro and the names and locations of the output files are contained in the "SpTool data" sheet in SpecDB.xls. The mechanisms to be included in the output for the Speciation Tool, and the names of mixtures to be used to derive model species for unknown mass or unassigned moles, or whether such mixtures are used at all, can be controlled by changing options in the "Parameters and Commands" sheet. Comments on these sheets indicate where these options are given.

The four files that are output for the Speciation Tool, and the tables in the Speciation Tool database that they update, are discussed below.

The "Species Table" file contains updated information about the chemical categories used by the Speciation Tool. The data are contained in the "SpTool" data sheet, with the data in most fields taken from the "export_species_properties_2013_0311.csv".file in the Speciation Tool distribution (CMAS, 2014). The default name for this file is "SpeciesTable.csv", output in a subfolder called "SPTool". This is used to update the "species" table in the Speciation Tool documentation, and the data fields in this table are listed in Table B-12 of Jimenez et al (2013). The only data that are updated concern the "molecular_weight", "volatile_mw" "nonvoctog" fields, as discussed below. Note that some of the fields described in Table B-12 of Jimenez et al (2013) did not contain any data in this file in the Speciation Tool distribution, so apparently not all of these data fields are actually used by this system.

The two molecular weight fields are changed to the appropriate molecular weight for the category based on the assignments and options in SpecDB.xls. For compounds this is the same as the molecular weight for the compound, for mixtures it is the effective molecular weight of the mixture[1], and for unknown mixtures that represented by "NVOL" or "UNKN" it is set to 1, since assignments of NVOL and UNKN are output on a mass basis. If a mixture is used to derive model species for unknown volatile mixture categories then the effective molecular weight of this mixture is used for the molecular weight of these categories. Categories that do not have assignments in SpecDB[2] are assigned molecular weight values of -99. The "nonvoctog" field is set to 1 if the category refers to an exempt compound or to a mixture that has more than 50% exempt compounds, and zero otherwise[3]. However, as discussed in the following section, it may not be appropriate to use these to derive VOC-to-TOG ratios used for processing VOC inventory data for modeling.

The "Mechanism List" file contains the summary information about the mechanism that is used as a replacement to the "mechanism description" table described in Table B-6 of Jimenez et al (2013). The default name for this file is "MechList.csv", also output in the "SPTool" subfolder. This only needs to be updated if a new mechanism is added or its description changes, but is automatically output whenever the other files are created. This replaces the data in the "mechanism_description.14may2013.txt" file in the Speciation Tool distribution (CMAS, 2014).

---

[1] The effective molecular weight of a mixture is the reciprocal of the total number of moles per unit mass of compounds in the mixture.

[2] None of these unassigned categories are used in any of the gas profiles in SPECIATE 4.3, but their data as contained in the "export_species_properties_2013_0311.csv".file are copied to the output file.

[3] The Speciation Tool database structure does not permit for non-integer values for the "nonvoctog" field, which means that fractions of exempt may not be properly defined for mixtures. However, the highest mass fraction of exempts in mixture categories that are not 100% exempt is less than 5%. Therefore, categories with less than 5% exempts on a mass basis are treated as having no exempts in the Speciation Tool output.

The "Model Species List" file contains the list of model species in the supported mechanism that is a replacement to the "carbons" table as described in Table B-1 of Jimenez et al (2013). This table is so named because it also contains the carbon numbers for these model species, though they are not well defined for mechanisms other than Carbon Bond, are not needed for model species assignments for any mechanisms, and its use to estimate mass of emissions from model species assignments is not recommended. The default name for this file is "MSClist.csv", output in the "SPTool" subfolder. The model species lists and their carbon numbers are taken from the sheet for the mechanisms in the MechDB.xls file. Note that they include the special model species "NROG" and "NVOL" for all mechanisms. Note that NROG is not needed for the Carbon Bond mechanisms but is included on their model species lists for compatibility with the speciation processing macros. If no mixtures are used for unknown mixtures or unassigned species it also includes model species "UNKN" and/or "NOASN", again with carbon numbers of zero. All of these special model species are given carbon numbers of zero because the carbon numbers are either unknown or so variable as to be meaningless. This replaces the data in the file "carbons_14may2013.csv" in the distribution.

The "Model Assignment File" contains the model species assignments for the SPECIATE 4.3 chemical categories, to be used as a replacement to the "mechanism" table described in Table B-5 of Jimenez et al (2013). The default name for this file is "MechAsn.csv", output in the "SPTool" subfolder. The assignments are given in a mole per mole basis, so it is important that the molecular weights in the "Species Table" file be consistent with these or incorrect assignments will be made for mixtures. Note that amounts of the special model species "NROG", "NVOL" and "UNKN" are output on a mass basis, so if these are assigned to compound categories with molecular weights, the amounts output are set to the molecular weight of the compound so the molecular weights will cancel out and the output will be in mass units. Note also that unknown mixture categories represented by "NVOL" or "UNKN" are assigned molecular weights of 1, so the moles per mole output in the assignment file are 1 in these cases. This is used as a replacement for the file "mechanism_14may2013_forimport.csv" in the distribution. For most mechanisms these data were produced using an earlier version of this mechanism assignment speciation database, though the molecular weights in the distributed "species" table was not be appropriate for some of the chemical categories.

In order to incorporate the assignments in SpecDB.xls when running the Speciation Tool, the user must import all four files discussed above into the Speciation Tool database as described in the Speciation Tool documentation (Jimenez et al, 2013). This will need to be done each time a mechanism is added or the model species or mixture assignments are modified. Note that it is important to export the updated molecular weights in the "Species Table" file as well as the files with the model species assignments, or incorrect assignments may be made for profiles containing SPECIATE categories that refer to known or unknown mixtures. Although the molecular weights for compounds will not change, those for known mixtures will depend on the compositions of the mixtures, and those for unknown mixtures may change if a mixture is used to derive model species for unknown mixtures, and if the composition of that mixture changes.

3.2.4. Processing Procedures for BEIS 3 Biogenic Model Outputs

The BEIS 3 biogenic emissions model outputs mass carbon emissions of biogenic compounds using the "BE3" chemical categories (Table 5). These are processed using SMOKE using GSPRO files that give assignments of model species to these categories. These are different than the GSPRO files produced by the ProfAsn for anthropogenic emissions as discussed previously because the TOG emissions in anthropogenic profiles are quantified as quantified by mass, while the biogenic VOC emissions output by BEIS are quantified by mass carbon. Because of this, SpecDB.xls has separate procedures to prepare GSPRO files for use with BIES 3 output.

As indicated in Table 2, there is a macro called "OutputGSPRO" that can be used to output categories of a selected type with the option of processing data in mass carbon units. The parameters needed by the macro, which consist of the output file name, the Category (SAROAD) ID, the mechanism(s) to use, and the units are given near the top of the "Parameters and Commands" sheet by the control to run the macro. SpecDB.xls as distributed has parameters appropriate for producing a GSPRO file for use with BEIS 3, with a file name of "BEIS3.GSP", a category ID (SARID) of "BE3" (see Table 5), a mechanism ID of "all" (meaning produce assignments for all mechanisms with a "1" in the "use" column of the mechanism listing table in the "Parameters and Commands" sheet), and units of "C", meaning process data in mass carbon units. The user can change the file name if desired and also give the name or code identifying a single mechanism (matching the "MechID" or "Code" column on the mechanism listing table in this sheet) if a assignments for single mechanism is to be output, but the "SARID" must be kept as "BE3" and the "Units" must be kept at "C" output assignments for BEIS 3. The format of the GSPRO file produced by this macro is discussed in Section 4.2.3.

3.2.5. Processing Procedures for MEGAN 2 Biogenic Model Outputs

At present, the SpecDB system does not have separate procedures for processing MEGAN 2 model outputs because currently the MEGAN system uses built-in model species assignments that cannot be updated using the SpecDB system. However, the files output for processing anthropogenic profiles by SMOKE, discussed in Section 3.2.2, can be used to assign model species to the 18 biogenic VOC categories as discussed by Carter (2015b). Since MEGAN outputs biogenic emissions in mass units, GSPRO files prepared for anthropogenic emissions can be used for this purpose, except that the code of "ME2" should be used to identify the profile database category, and the MEGAN category names given in the "SAROAD Assignments" sheet should be used rather than anthropogenic profile ID codes. It is recommended that future versions of MEGAN take model species assignments from files, as is the case for the Speciation Tool. This would allow for direct integration of model species assignments output by future versions of SpecDB.xls.

The "OutputGSPRO" macro (Table 2), discussed above, can also be used to produce model species assignment files for processing MEGAN output. The parameters needed by the macro, which consist of the output file name, the Category (SAROAD) ID, the mechanism(s) to use, and the units are given near the top of the "Parameters and Commands" sheet by the control to run the macro. Although SpecDB.xls as distributed has parameters appropriate for producing a GSPRO for BEIS 3, these can be modified to produce GSPRO files for processing MEGAN output. For this purpose the "SARID" column should be changed to "ME2" (see Table 5), and the "Units" column must be either blank or contain "Mass". The "mech" parameter can either be "all" (or blank) to process all mechanisms with a "1" in the "use" column in the table of mechanisms, or the code for a single mechanism, which must match one of the mechanisms in the "MechID" or "Code" column. The format of the GSPRO file produced by this macro is discussed in Section 4.2.3.

3.2.6. Other Emissions Processing Systems

The current system does not include programs and files for processing these assignments directly for use with emissions speciation and processing systems other than SMOKE or the Speciation Tool. However, the files output for SMOKE or the Speciation Tool contain all the mechanism-dependent information that these systems should need for speciation processing. The files produced for the most similar system can be formatted as needed for this purpose. The ability to output prepare assignments can be added in the future if required.

4. Program and Files Details

4.1. Program Descriptions

The SpecDB system as distributed includes two Fortran programs, ProfPro and SpecPro, for processing outputs of SpecDB.xls and MechDB.xls for use in emissions speciation processing. These programs and their use are described in this section. Examples of using these programs and testing their implementation are discussed in Section 4.4.

These two programs supercede the five programs described by Adelman et al (2005) for speciation processing and optionally revising mechanisms based on assignments in the speciation database. These programs are still available for download at the speciation database web site (Carter, 2015a), and the files output by the current version of SpecDB.xls are still compatible with use of these programs. However, only ProfPro and SpecPro as described below are actually needed to assign model species to emissions profiles; the other three are only needed if modelers want to run models with modified versions of the SAPRC-99 mechanism with different lumping approaches and using different mixtures to derive mechanisms of lumped species. In practice few modelers have been interested in using this capability, and the current programs and files do not support SAPRC-07 or later versions of SAPRC. Therefore, until there is renewed interest in this mechanism modification capability, the programs not needed to assign model species to profiles will no longer be updated or supported.

4.1.1. ProfAsn Program

In the present system, a program called ProfAsn is the only program needed to be run directly by the user in order to process profile data using the assignments output by SpecDB.xls for use with SMOKE. The five types of assignment files output by SpecDB.xls and required for processing profiles using ProfAsn are described in the "Processing Procedures and Files for SMOKE" section, above. In most cases, ProfAsn is run by using the command

ProfAsn *inputfile MechID Outputfile*

Where *inputfile* is the full path name of an input file containing or referencing the profile data, *MechID* is the code for the chemical mechanism as given in the first column in Table 9, and *Outputfile* is the path and name, but not the extension, used for the output files. The files output include *Outputfile*.GSP containing the data in the GSPRO file and *Outputfile*.CSV containing summary information for the profiles that are processed.

Normally *inputfile* would contain the following records, where the first two are optional and the third is required and can be included multiple times to merge profile data from various sources:

UNKN *UnknMixFile*
NOASN *NoAsnFile*
INPUT=*ProfDataFile DataFormatCode SDBfile*

Note that the spaces are required where indicated and that there is no space between "INPUT=" and *ProfDataFile*.

The UNKN and NOASN input records must be included before any "INPUT=" records if it is desired that the profiles be processed using mixtures to assign model species to unknown mixtures or unassigned compounds. If these are not used, then mass of unknown categories are output as mass of the special "UNKN" model species, and moles of unassigned compounds are output as moles of the

"NOASN" species. If these are used, then moles of model species assigned to these mixtures are output. In this case, *UnknMixFile* and *NoAsnFile* give the name of .EMC files giving mass fractions of the individual compounds of these respective mixtures. By default these files are named REPUNK.EMC and REPNOA.EMC and are produced using the "OutputUnknEMCfiles" macro in SpecDB.xls. If the UNKN record is omitted or given after the INPUT= record then the mass fractions of unknown mixtures in the profiles are output as mass fractions of the special UNKN model species. Likewise, if the NOASN record is omitted or given after the INPUT= record then the moles of unassigned compounds are output as moles of the special model species NOASN. If these records are included before the INPUT= record, then moles of model species assigned to the mixtures are output are output instead of mass of UNKN or moles of NOASN.

The INPUT=*ProfDataFile DataFormatCode SDBfile* record is used to tell ProfAsn where to get the profile data to process and how to process it. Here, *ProfDataFile* is the full path name of the profile data file extracted from the profile database, *DataFormatCode* is the code indicating the categorization system used, and *SDBfile* is the name of the processed profile data file that ProfAsn needs to make the assignments. *DataFormatCode* is required and must match one of the codes given in the first column of Table 5, and can be "S40" for SPECIATE 4.x profiles, "CA" for profiles in the California database, etc. ProfAsn then uses these as input to make a system call to run the separate ProfPre program to produce an intermediate .SDB file containing the profile data as mass fractions of individual compounds in SpecDB, which ProfAsn then reads in order to make the model species assignments for the profiles. *SDBfile* is optional and if given is the name of the intermediate .SDB file, without the .SDB extension, which is added. If *SDBfile* is not specified, the name TMP.SDB is used for the intermediate file, which is over-written when each INPUT= line is processed. At least one INPUT= record must be included or no profiles would be processed, and multiple such records can be used to merge profiles from different sources (including from databases using different categorizations) to include in the output files.

Note that the profile names "UNKN" and "NOASN" are reserved for use for assigning model species to unknown mass or unassigned moles, and must not be used for any other purpose. This means that compositions of such mixtures can be given in files named in "INPUT=" records rather than UNKN or NOASN records, as long as they are given those names. But the recommended procedure is to use the UNKN and NOASN input to keep those separate from the profile database.

ProfAsn can also be used to take input directly from processed mixture or profile data files that give compositions of mass fractions of individual compounds in the SpecDB database. As discussed in Section 4.2, these can be either .EMC files containing data for single profiles or mixtures, or .SDB files containing data for multiple profiles. EMC files for the mixtures used for unknown mixtures or unassigned compounds are output directly by SpecDB.xls using the "OutputUnknEMCfiles" macro, or can be produced by the ProfPro program for any profile as documented by Adelman et al (2005). SDB files are output by the ProfPre program using profile data as discussed below.

To input data for single profiles contained in .EMC files, the *inputfile* used by ProfAsn can contain records consisting of the profile ID code and the .EMC file name, separated by a space. The UNKN and NOASN input discussed above are actually examples of this. This is not expected to be normally used because it is generally more convenient to use input files containing data from multiple profiles rather than using separate input files for each profiles. However, this may be convenient if special or replacement profiles are used.

If the profile data have already been processed by ProfPre and are in a .SDB file that it outputs, and if the .SDB file contains profiles named UNKN or NOASN if desired for processing unknown mixtures or unassigned compounds, then input can be taken directly from the .SDB file by using it instead of a separate *inputfile* specified in the command line. The program looks at the extension of the input file

and if it is .SDB then it assumes that this is the case, and no separate input file is used. Note that this means that input files that are not to be used in this way should not be given the extension .SDB.

4.1.2.   ProfPre Program

The ProfAsn program calls a separate ProfPre program in order to convert profiles given as mass fractions of chemical categories in the various profile databases into mass fractions of individual compounds in the SpecDB database. It is not normally run directly by the user unless it is desired to examine or preserve the files giving the profiles in terms mass fractions of the chemical compounds. It is run using the command

ProfPre *ProfDataFile DataFormatCode SDBfile*

where *ProfDataFile* is the name of the input file with the profile data, *DataFormatCode* is the code for the profile database categorization system as discussed above in conjunction with the input for ProfAsn. *SDBfile* is optional and contains the name given to the output file, with the extension .SDB added to it. If *SDBfile* is not specified (as would be the case if it is not given on the INPUT= record when ProfPre is called by ProfAsn) then the output goes a temporary file named TMP.SDB. Note that the temporary file TMP.SDB is retained if ProfPre is called by ProfAsn, but it is over-written the next time ProfAsn is run.

4.2. File Descriptions

4.2.1.   Profile Data Input Files

Regardless of which *DataFormatCode* is used, the profile data files are expected to consist of multiple records with three fields each, separated by one or more spaces. The first field is the profile ID number or code, which is treated as a character string with a maximum length of 16 characters and must be left justified. (If it is necessary to use longer profile names, then the length of the PROFID variable in ProfAsn.for and ProfPre.for, and source lines that depend on this length, would need to be modified.) The second field is the chemical category code, which must be a valid code in the profile database indicated by *DataFormatCode*. This is also treated as a character string even though some databases use numbers to indicate the category. The third data filed is the mass fraction of the category in the profile. All records containing data for a particular profile must be together in the file, since the ProfPre program used the change in the profile ID code as an indication that all input for a profile is completed and processing of the profile should begin. Therefore, depending on the profile database, it may be necessary to sort profile input data by profile ID before using them to create the *ProfDataFile*s.

4.2.2.   GSPRO Output Files for Anthropogenic TOG Profiles

The format and contents of the GSPRO files output using ProfAsn is somewhat different than output by the previous program (SpecPro) documented by Adelman et al (2005), and is subject to change depending on feedback received from modelers who will use these files. These are comma separated data files containing five data fields, as follows:

1. Profile ID as specified in the input *ProfDataFile*;

2. The code "TOG" which is assumed for all profiles;

3. The model species name for the mechanism, which might include NVOL or NROG for non-volatile or unreactive mass and UNKN or NOASN for unknown mass or unassigned moles if there are no profiles previously input named UNKN or NOASN;

29

4. The moles of model species per gram of profile emissions for the regular model species in the mechanisms, or the mass fractions for NVOL, NROG, UNKN, or NOASN if applicable.

5. The "denominator" quantity for the mass to mole conversion, which is always 1.0 in the GSPRO files output by ProfAsn. See discussion below.

6. The mass fraction of emitted compounds that are represented by this model species. This is the same as the data in the 4th field for NVOL, NROG, or UNKN, but different for model species assigned on a molar basis.

The data in the 4th and 5th fields are somewhat different than those in GSPRO files output by the previous version of the program, SpecPro (Adelman et al, 2005), which correspond more closely to GSPRO files used in most existing modeling database. However, this does not affect the assignments of moles of model species per unit mass of emissions of the profile because that is determined by the ratio of the number in the field 4 to the number in field 5. The concept behind the original format is that the data in the 4th field is the moles of model species per mole of profile while the data in the 5th field is the molecular weight of the compounds in the profile. However, "moles" and "molecular weight" is not particularly well defined for profiles that contain mixtures, and cannot be defined if the profiles contain unspeciated mixtures. Using field 4 to simply represent moles of model species per gram of profiles bypasses these problems and eliminates the need to define the molecular weight for mixture profiles. If the "molecular weight" field 5 always contains 1, the result of the speciation process is the same, and the GSPRO files do not have to contain quantities that are poorly defined.

The mass fraction quantities in column 6 should sum up to one if the mass in the input profiles are properly normalized. For "lumped molecule" mechanisms that use a single model species to represent a compound, this is just the total mass of compounds represented by this model species. However, this does not work for if the mechanisms use "reactivity weighting" for some compounds[1] or if they use more than one model species to represent a compound, as is the case for the Carbon Bond mechanisms. To assure that the mass fraction quantities add up to one, the mass of the compound is distributed among the model species representing it using the formula

$$\text{Mass fraction (i)} = \text{Moles(i)} \cdot \text{Mwt(i)} / \Sigma_j \text{Moles(j)} \cdot \text{Mwt(j)}$$

where Mass fraction (i) is the mass fraction of model species i assigned to the compound, Moles(i) is the moles of the model species used to represent a mole of the compound, Mwt(i) is the molecular weight assigned to the model species by the chemical mechanism developer, and the sum in the denominator is over all the model species used to represent the compound. The Mwt(i) values for the model species are taken from the assignment file *MechID*MS.DAT, discussed above, that gives the list of model species. Note that the molecular weights of the model species cancel out when processing for mechanisms that use only one model species to represent a compound. Note also that in general the mass represented by a model species will depend on the compound being represented, so it is not a property of the model species itself[2].

---

[1] Reactivity weighting in lumped molecule mechanisms consists of use of variable numbers of moles of model species to represent a compound, depending on its reactivity. This is used for a few species in the Lumped SAPRC mechanisms and for many species in the RADM2 and RACM2 mechanisms.

[2] This is because the a given model species may represent different compounds with a range of molecular weights, and reactivity weighting may be used for some compounds and mechanisms. The variability is smaller for the Carbon Bond mechanisms since the model species represent all the carbons in the compounds. However the molecular weight per carbon varies from compound to compound, especially for compounds that are not hydrocarbons.

The previous program, SpecPro (Adelman et al, 2005), used output data in a fixed column format using column widths that were exactly the same as found in GSPRO files obtained at the time by the EPA. The current program outputs these files in CSV format, though they are not given a CSV extension by default.

Although the total of all mass fractions in emissions speciation databases should sum up to 1 or 100% in all profiles, this isn't always the case in current databases. ProfPre does not modify these masses so if the mass in the profiles is not normalized in the input file they won't be normalized in the output .SDB file. However, ProfAsn always normalizes the mass fractions for the profiles before producing the GSPRO files. If the total mass fractions in any profiles are deliberately less than 1 to represent unreactive or nonvolatile mass, then the profile should be normalized by adding unreactive or nonvolatile categories, which exist in all known emissions speciation databases. We can think of no scientifically justifiable reason for deliberately having a total mass fraction greater than 1 in any profile, so we will not enable this practice by permitting the system to process such input without normalizing. The .CSV output file produced by ProfAsn gives the total masses of the profiles, discussed below, and can be used to identify the problem cases.

Another possible cause of profile masses not summing up to 1 in the ProfAsn output is that the profile data in the *ProfDataFile* were not sorted properly, and data for some of the profiles may not be all together. This was found to occur in files directly output by both the SPECIATE and California databases. The ProfPre program uses the change in profile IDs as an indicator that the input for a profile is completed, so if the data for a profile are in two separate places in the input file it will output two profiles with the same IDs, both with low total weights. Therefore, it is important that the profile data files be properly sorted before using them as input to this system.

Note that if UNKN or UNASN mixtures are used to process unknown mass or unassigned moles, the GSPRO file also contains model species assigned to these profiles as if they are separate profiles named UNKN or NOASN. These would normally appear at the beginning of the GSPRO file since they need to be input and processed before any profiles containing unknown mixtures or unassigned compounds. These can be removed from the output files by changing the value of the "GSPUNK" logical variable in ProfAsn.for as indicated by comments in this file.

4.2.3.   GSPRO Output Files Produced by the "OutputGSPRO" macro

The GSPRO file output by the OutputGSPRO macro in SpecDB.xls are consistent with the GSPRO file provided by the EPA for use with BEIS 3 (EPA, 2014b) if the "C" units option is selected, and can be used to give assignments for other types of VOC categories if desired. They have similar data but a somewhat different format than GSPRO files output by ProfAsn as discussed aboveCurrently, the BEIS 3 GSPRO files output by SpecDB.xls are comma separated data files starting with two comment lines identifying the file and the date it was created, followed by records with six fields of data, as follows:

1. The chemical mechanism ID as given in the first column of the mechanism listing table in the "Parameters and Commands" sheet. The file produced by OutputGSPRO has records for only one chemical mechanism, but all these records still contain the mechanism ID.

2. The category code as given in the "SAROAD assignments" sheet. For example, for assignments for BEIS 3 output this would be BEIS 3 category names and for MEGAN output this would be MEGAN category names.

3. The model species name.

4. The moles of model species per mole of emissions VOCs in the category. For mixtures, this is calculated as the sum of moles of model species for each compound per unit mass of mixture, divided by the number of moles in the same mass of mixture.

5. If the option to give assignments to process mass carbon, as appropriate for BEIS 3, is selected, this contains the grams of carbon per unit mole of emissions of VOCs in the BEIS 3 category. For compounds this is the number of carbons times the atomic weight of carbon. For mixtures this is the atomic weight of carbon x the number of carbons per unit mass of mixture divided by the total number of moles in the same mass. If the option to give assignments for mass, this is the molecular weight of the compound or the ratio of the total mass to total number of moles for mixtures.

6. The mass of emissions represented by this model species per mass carbon emitted. This is calculated in the same way as the mass fraction field for anthropogenic GSPRO files discussed in the previous section except that if the option to give assignments to process mass carbon is selected, this is multiplied by a correction factor to give mass from mass carbon.

SMOKE calculates the moles of model species per gram carbon emissions as the ratio of field 5 to field 5, as is the case for the anthropogenic GSPRO files discussed in the previous section. If the option to process mass carbon is selected, the use of carbons per mole in field 5 and the correction factor in field 6 corrects for the fact that BEIS gives emissions as mass carbon rather than true mass.

4.2.4. ProfAsn Listing CSV files

When ProfAsn is run it outputs CSV file with the same name as the GSPRO file (except for the extension) containing summary information about the profiles that were processed. This includes the input mass (which should be 1 for properly normalized profiles), the mass fractions assigned to normal model species, unreactive and nonvolatile mass fractions, mass fractions of unassigned compounds, mass fractions of unknown mixtures, and mass fractions of exempt (non-VOC) compounds. It also gives the number of individual compounds in the processed profiles and the number of model species (normal and special) output in the GSPRO file for them. If mixtures are used to assign model species to unknown mass or unassigned moles the mass fractions of unknown mixtures or unassigned compounds are output as negative numbers, and the assigned and unreactive mass fractions refer to those in the mixtures. The file also indicates which profiles were processed using mixtures for unknown mass or unassigned compounds and indicates the input files used to derive their compositions.

4.2.5. Processed Single Mixture or Profile Composition Files

Files in EMC format. EMC files give compositions of a single mixture or profile in terms of chemical compounds in the SpecDB database. These files for the mixtures used for unknown mixtures or unassigned compounds are output directly by SpecDB.xls using the "OutputUnknEMCfiles" macro, or they can be produced by the ProfPro program for any profile as documented by Adelman et al (2005). These files can be used as input to the ProfAsn program as discussed above. These files contain records with comments containing available documentation information, a record with a "." in the first column indicating start of composition input, then records containing the code used in SpecDB for the chemical compound category, the mass fraction of the compound in the mixture or profile, flags indicating if the compound is volatile or a VOC (not exempt) (1 for yes, 0 for no in both cases), and a description of the compound. Records starting with a "!" in the first column are ignored. These files as output by SpecDB.xls and ProfPro have the data according to the Fortran format code (A16,1X,G12.6,I3,I4,2X,A).

Files in SDB format. SDB files give compositions of multiple mixtures or profiles in terms of chemical compounds in the SpecDB database. These files are produced using the ProfPre program and are

used as input to the ProfAsn program as discussed above. The first record of these files contains the code "MULTI-PROF" followed by a space and the name of the input profile data file and profile category code used as input to ProfPre. Input for each profile starts with a record with "..", followed by a record with the profile ID code, then followed by records with the profile compositions, where each record contains the compound ID in the SpecDB database and the mass fraction of the compound in the mixture. Additional fields can be included with the profile composition data as in .EMC files, and .SDB files can be created by putting together multiple .EMC files with the comments stripped and the ".." codes and the profile ID codes added at the start. However, the only data that are used when these files are input to the ProfAsn program are the profile and compound IDs and the mass fractions.

4.2.6.   Profile Database Category Name Assignment Files

Files named *ProfDBcode*SARID.DAT contain the mappings of the chemical categories used in the supported profile database indicated by *ProfDBcode* to the category IDs used in SpecDB. The first set of records in these files are comments that are ignored by the programs, terminated by a record with a "." in the first column. The subsequent records consist of the chemical category name used by the profile database (as a character string, left justified), followed by one or more spaces, then the SpecDB category name assigned to it. It is needed as input to ProfPre for all profile database categories except for SDB.

4.2.7.   SpecAsn.DAT Assignment File

This file contains compositions of all the speciation categories in SpecDB in terms of individual compounds in the database. The first set of records in these files are comments that are ignored by the programs, terminated by a record with a "." in the first column. The subsequent records consist of the SpecDB master category name, the compound category (or "UNKNOWN" or "NONVOL" if it is a volatile or nonvolatile unspeciated mixture) and the mass fraction of the compound in the category. Single compound and unknown mixture categories have only a single record with a mass fraction of 1, while known mixture categories have records for each compound in the mixture, where the mass fractions should sum up to 1 for the entire mixture. This is needed by ProfPre for processing profiles in any database.

4.2.8.   Model Species Name Files

The files named *MechID*MS.DAT contain the list of model species for the mechanism identified as *MechID*. The first set of records in these files are comments that are ignored by the programs, terminated by a record with a "." in the first column. The subsequent records contain the list of model species, one for each record, with 3 fields of data with the Fortran format code (A16,I4,F7.2). The first field is the model species name, the second is the number of carbons, and the third is the molecular weight of the model species. These files are used by ProfAsn, though it does not use the carbon numbers in its processing.

4.2.9.   Model Species Assignment Files

The files named *MechID*ASN.DAT contain the assignments of model species for each compound for the mechanism identified as *MechID*, as well as other information about the compounds. The first set of records in these files are parameters or comments that are ignored by the programs, terminated by a record with a "." in the first column. This input must include a record containing either "FORMAT=1", "FORMAT=2", "FORMAT=2 *nCol*" (where *nCol* is the number of sets of columns of assignment data, which is not given if there is only one set of columns) or "FORMAT=3" to indicate how the assignment data are formatted. If the file contains a "FORMAT=3" record, the input must also contain a record with "NUMSPEC=*n*, where *n* is the number of model species. Other input in these records before the "." are

treated as comments and are ignored. The subsequent records contain the assignment data, whose format depend on the type of mechanism as indicated by the "type" filed in the list of mechanisms in the "Parameters and Commands" sheet of SpecDB.

"FORMAT=1" is used mechanisms that have only one mole of model species assigned for each compound with model species assignments, with no reactivity weighting. The data fields in the assignment records are the SpecDB ID code for the compound, given with the Fortran Format Code (A16,F8.3,I4,,I3,2X, A16,1X,A). The data filelds are the SpecDB ID code for the compound, the molecular weight, the number of carbons, a flag (1 or 0) indicating if the compound is a VOC, the model species name and a description of the compound.

"FORMAT=2" is used for mechanisms that usually have only one or a relatively small number of model species assigned per compound, and may have reactivity weighting for some compounds. Most of these mechanisms have only one model species assigned, but may have reactivity weighting so the moles of model species per mole of compound need to be specified. In some cases, e.g., CSAPRC07, there can be more than one model species assigned for some compounds, so more than one set of columns of assignment data are needed. The number of sets of columns of assignment data are specified on the *nCol* option on the FORMAT=2 record (e.g., "FORMAT=2 2" for mechanisms with two sets of columns). The assignment data are given with the Fortran format code (A16,F8.3,I4,,I3,2X,*nCol*(A16,F7.3,1X),A), where *nCol* is the number of sets of columns of assignment data. The data fields are the SpecDB ID code for the compound, the molecular weight, the number of carbons, a flag (1 or 0) indicating if the compound is a VOC, the model species name and the moles of model species per mole of compound (repeated for each column of assignment data), and a description of the compound.

"FORMAT=3" is used for mechanisms, such as the Carbon Bond mechanisms, that have relatively small numbers of model species and that use more than one model species for many compounds. The Fortran format code for the assignment data are (A16,F8.3,I3,1X,A1,2X,I2,1X, *n*F7.3,1X,A), where *n* is the number of model species input (including NROG and NVOL but not counting UNKN or NOASN if applicable). This is specified in the "NUMSPEC=*n*" parameter record as indicated above. The data fields consist of the compound ID code, the molecular weight, the number of carbons, an assignment code, a flag indicating whether the compound is a VOC, the moles of each model species per mole of compound, and the description of the compound. The assignment code can be either "A" for assigned, "U" for unassigned (no model species assigned) or "N" for nonvolatile.

Note that mechanisms whose assignment data are output in formats 1 and 2 can have model species with names as long as 16 characters. However, mechanisms whose assignment data are output in format 3, which consist only of Carbon Bond mechanisms, can have model species names no longer than 8 characters. This is because of the formatting of the fixed column output files that give these assignments. This is not a problem with the current versions of the Carbon Bond mechanisms.

These files are used by ProfAsn to make the model species assignments for the mechanism. Note that the numbers of carbons and the compound description fields are ignored by ProfAsn, but the other data are all used.

4.2.10. Speciation Tool Import Data Files

The four files output by SpecDB.xls used to import data into the Speciation Tool are discussed in Section 3.2.3. These are all comma separated value files. The data contained in these files are also discussed in the documentation for the Speciation Tool (Jimenez et al, 2013) for the corresponding data table in the Speciation Tool database.

4.3. Files in the Current Distribution

The current distribution of speciation database assignment files and programs, and input and output files needed to process representative profiles with selected mechanisms are contained in the file SpecDB.zip, available at http://www.cert.ucr.edu/~carter/emitdb. The files for this distribution are listed in Table 11. All the files except the source files are contained in a single directory and only the files needed to process emissions from SPECIATE 4, Texas, or California databases for the CB05, and the lumped mechanisms listed in Table 9 are included in this distribution. The SpecDB.xls and MechDB.xls spreadsheets are included in SpecDB.zip and can also be directly downloaded from this website.

4.4. Example of Use of the Programs

The distribution contains a file called S44-PROF.DAT that contains profile speciation assignments extracted directly from the GAS_SPECIES sheet of the SPECIATE 4.4 Access database file (EPA, 2014a), but ordered as required by ProfPre, and sorted so the data for each profile are together. To make speciation assignments for any mechanism, an input file called S44-PROF.IN is included containing the following records:

```
UNKN REPUNK.EMC
NOASN REPNOA.EMC
INPUT=S44-PROF.DAT S40
```

To produce a GSPRO file for SMOKE for these profiles and the Carbon Bond 05 mechanism, use the command

```
ProfAsn S44-PROF.IN CB05 S44-CB05
```

where the output file name "S44-CB05" identifies both the profiles and the mechanism. This will produce an output file S44-CB05.GSP with the GSPRO data and S44-CB05.CSV with the profile summaries. A file called TMP.SDB is also produced giving the profile data in terms of mass fractions of individual compounds in the SpecDB database. This is used internally by ProfAsn but is not needed for subsequent emissions processing, and it is over-written the next time ProfAsn is run or when another INPUT= record is processed. If the user wishes to save this file using a name identifying the profiles, it could give it an appropriate name at the end of the INPUT= record, eg., use "INPUT=S44-PROF S40 S44-PROF".

If the profiles are to be processed without using the REPUNK and REPNOA mixtures to assign model species for unknown mixtures or unassigned compounds, the UNKN and NOASN records can be removed from the .IN file, leaving only the "INPUT=S44-PROF.DAT S40" record, and giving the same command to run ProfAsn as indicated above. In this case, the same output files are produced if the commands

```
ProfPre S44-PROF.DAT S40 S44-PROF
ProfAsn S44-PROF.SDB CB05 S44-CB05
```

are given. In this latter case, a file S44-PROF.SDB containing the individual compounds in the profiles is also produced. In these cases, the S44-PROF.GSP file does not contain model species assigned to unknown mixtures or unassigned compounds, and instead contains mass fractions of UNKN or moles of NOASN for such profiles.

The distribution also contains a file called CA-PROF.DAT containing the profiles in the current California profile database (Yang, 2013) and an input file called CA-PROF.IN for processing these profiles for any mechanism. The commands to process these profiles, or to process SPECIATE or

Table 11.   Files contained in the current distribution of the updated speciation database programs and files.

File	Description
	Files in the root directory
	Documentation
SpecDB.pdf	The current version of this document, in PDF format.
	Database Files
SpecDB.xls	Main speciation database assignment file.
MechDB.xls	Mechanism description file containing lists of model species and related information for the supported mechanisms
	Mechanism-independent assignment files output by SpecDB.xls
SpecAsn.DAT	Assignments of speciation database categories to single compounds
S40SARID.DAT CASARID.DAT TXSARID.DAT	Assignments of chemical categories in SPECIATE 4, California, and profile databases to SpecDB chemical categories.
BE3SARID.DAT ME2SARID.DAT	Assignments of chemical categories output by the BEIS 3 and MEGAN 2.1 models to SpecDB chemical categories. Note, however, that ProfPre and ProfAsn should not be used to produce GSPRO files from with BEIS 3 outputs that have data in mass carbon units. BEIS3.GSP should be used instead.
	Assignment files for mechanisms output by SpecDB.xls for selected mechanisms
*MS.DAT *asn.DAT	Model species and assignments for the various mechanisms. Such files are given for CB4; CB05; SAPRC-99, SAPRC-07 and SAPRC-11 with standard lumping (S99L, S07L and S11L); toxics lumping (S07L) semi-detailed SAPRC-11 (S11D); RADM2 (RADM); and RACM2 (RACM) mechanisms.
BEIS3.GSP	Model species assignments for BEIS 3 categories for the various mechanisms, based on BEIS 3 outputs being in units of mass carbons. The distributed file includes assignments for the same mechanisms as listed above for *MS.DAT and .ASN.
	Files giving compositions for representing unknown or unassigned reactive mass in model simulations in terms of mass emissions of compounds
REPUNK.EMC	Composition of mixture recommended as appropriate for unknown mass
REPNOA.EMC	Composition of mixture recommended as appropriate for unassigned mass
	Executable and batch files
ProfAsn.exe ProfPre.exe	The executable for the ProfAsn and ProfPre programs as produced by BLDALL.bat in the SOURCE subfolder
EXAMPLE.BAT	Batch file for running examples for the programs. All input and output files for these examples are included in the distribution.
	Input files used to run examples
S44-PROF.DAT	Compositions of all SPECIATE 4.3 gas profiles in terms of SPECIATE 4 chemical categories. Used as input to ProfPre (called by ProfAsn)

Table 11. (continued)

File	Description
CA-PROF.DAT	Composition of California gas profiles as of December, 2013 (Yang, 2013). Used as input to ProfPre (called by ProfAsn)
S44-PROF.IN CA-PROF.IN	SpecAsn input file to process SPECIATE 4.3 or California profiles using REPUNK and REPNOA to represent unknown mass or unassigned moles.

<u>Intermediate and output files created by running EXAMPLE.BAT</u>

File	Description
S44-CB05.GSP CA-CB05.GSP	GSPRO files for the SPECIATE or California profiles for the CB05 mechanism that can be used by SMOKE or similar programs for emissions processing.
S44-S07T.GSP CA-S07T.GSP	As above, but for the SAPRC-07T mechanism.
-.CSV	Profile summary files output by ProfAsn for the SPECIATE or California profiles and the CB05 or SAPRC-11 mechanism (e.g., S44-CB05.CSV, etc).
TMP.SDB	File produced by ProfPre containing assignments of SDB compounds to the last set of profiles that was processed by ProfAsn.

<u>Files in the SpTool Directory (to import into the Speciation Tool database)</u>

File	Description
SpeciesTable.csv	Species table file to import into the "species" table of the database
MechList.csv	Mechanism list file to import into the "mechanism description" table of the database
MSClist.csv	Model species list file to import into the "carbons" table of the database
MechAsn.csv	Model assignment file to import into the "mechanisms" table of the database

<u>Filies in the SOURCE subdirectory</u>

File	Description
BLDALL.BAT	Batch file to compile and link programs using gfortran for Windows
ProfAsn.for	Main source file for ProfAsn
ProfPre.FOR	Main source file for ProfPre
STSUBS.FOR	String-handling subroutines used by both programs
stsubs.o	Object file for STSUBS.FOR produced by BLDALL.BAT

California profiles for other mechanisms, are analogous to those discussed above. The distribution also contains a file called EXAMPLE.BAT to run these examples for SPECIATE and California profiles for the CB05 and SAPRC-11 (standard lumping) mechanisms.

4.5. Installation and Implementation Testing

The distributed files and programs in this system can be installed and tested on a Windows-based computer system using the following procedure, which must be done in the order indicated.

Installation:

1. Download the file SpecDB.zip from http://www.cert.ucr.edu/~carter/emitdb.

2. Extract the files to a selected location on your computer, being sure to use an extraction procedure that maintains the directory structure (i.e., does not put all files in one folder). The files should be as indicated in Table 11.

3. IMPORTANT: Open SpecDB.xls. Change the contents of the cell named "OutPath" in the "Parameters and Commands" sheet (currently at cell B3, but in any case near the top of the sheet) to the location on your computer to where you copied the files. The contents of this cell are in red font to indicate that it can be updated by the user. Do not use a trailing "\". This does not have to be the location of SpecDB.xls and MechDB.xls (though it will be if the files were extracted as indicated above) but does need to be the location of the input and executable files listed on Table 11. Save SpecDB.xls once you made this change.

Testing:

4. Copy the files *MS.DAT, *asn.DAT and all files with extensions .GSP, EMC, and .CSV to a separate location so they will not be over-written when the tests and examples are run. These can then be used to compare with the files produced by running the tests and examples. Alternatively, extract all the files to a separate location for this purpose.

5. Open SpecDB.xls. Run the "UpdateAll" macro using the control at the "Parameters and Commands" sheet, near the top of the table labeled "Macros Used in Sheets" (currently at cell B49 but subject to change if mechanisms or category codes are added). This will call all the macros to create the assignments for the supported mechanisms and categorization systems and output them in the *MS.DAT and *ASN.DAT files in output location you specified in Step 3, above. The macro will output two messages, the first indicating the update is completed and the second indicating that the files are output. Note that the "selected mechanisms" are those with a red "1" in the "Use" column of the "Supported Mechanisms" table on the "Parameters and Commands" sheet. Compare these files with the copies you created in Step 4. Do not change the data in the "Use" column in the mechanism list table when running this test so the output files created when running this test will be comparable with those in the distribution.

6. Run the "Output for SPtool" macro using the control near the top of the "Parameters and Commands" sheet, currently at cells F7-H7. It will give a message stating that several CSV files (described in Section 3.2.3) were created in the SPtool folder at the output location you specified in Step 3, above. Compare these with the with the copies you created in Step 4.

7. Near the top of the "Parameters and Commands" sheet there should be a section with cells labeled "GSPRO File Name", "SARID", "Mech" and "Units", which should contain, in red font, "BEIS3.GSP", "BE3", "all", and "C", respectivey (currently at rows 9 and 10). If they do not, change them accordingly. Then run the "Output GSPRO" macro whose control is located to the right of these cells. It should output a message stating that the file "BEIS3.GSP" was created in

the location you specified in Step 3, above. Compare with this file with the copy you created in Step 4. Be sure that the "Use" column in the mechanisms table, discussed above, were not changed or the contents of this file will not be comparable with the version in the distribution.

8. Open a DOS command window and navigate to the folder where you extracted the files, which should be the same as that you specified in Step 3, above. Then run the batch file "EXMPLE.BAT" that should be at this location, along with the executables, along with ProfAsn.exe and ProfPre.exe. Note that the example runs ProfAsn four times, two each for Speciate 4.4 and California profiles for both the SAPRC07T and CB05 mechanisms, producing .CSV and .GSP files named S44-CB05, S44-S07T, CA-CB05, and CA-S07T. Note also that in each case ProfAsn.exe calls ProfPre.exe as part of its processing of the data. Compare these with the copies you created in Step 4.

An analogous procedure can be used for installing the system on Unix systems, but this has not been tested. Installations on Unix systems would require minor changes to the source files, as described in the following section. We welcome input from users who have installed and tested these programs on a Unix system, and advice on how to update this users guide accordingly.

4.6. Programming Considerations

The distributed files, which are available at http://www.cert.ucr.edu/~carter/emitdb, include the source code for ProfAsn and ProfPre, which are contained in the files ProfAsn.FOR, ProfPre.for, and STSUBS.FOR, where the latter contains subroutines used by both programs. These are compiled using the GNU "gfortran" compiler for Windows systems, and a batch file for compiling these programs and linking them so they will run on systems that do not have the gcc libraries installed are included in the file BLDALL.BAT in the same folder with the source files. The executables are also included in the distribution, and should be able to run on any Windows system, though they have only been tested using Windows XP and 7. These source files should compile and run on Lunux or Unix systems, though the call to SYSTEM in ProfAsn.for would need modified to run ProfPre in its proper location. We should be notified if the modelers have other programs with running these programs on their systems, so we can modify the programs and documentation as needed.

5. Adding Mechanisms and Profile Databases

The current database was developed primarily to support use of mechanisms and emissions profile databases primarily used for tropospheric regional modeling in the United States. However, if suitably updated with assignments for appropriate mechanisms or speciation profile databases, it could well be used for global, international, or other modeling applications. Even if it is restricted to regional air quality modeling in the U.S., it would need to be updated as new mechanisms are developed and speciation profiles are updated or added with new compounds or chemical categories. In this section, we discuss the steps that need to be taken when adding new mechanisms or speciation profile categorization systems.

In the discussion below it is assumed that the user is familiar with Excel spreadsheets and the concept of Excel worksheets and how to define range names. In this context, the word "table" is used to refer to the portion of an Excel worksheet that contains information of the same type in tabulation form in rows and columns.

5.1. Adding or Updating Mechanisms

In the context of this discussion, a "mechanism" refers only to which model species are assigned to emitted organic compounds, not to how they react or the products that they form. If a mechanism update only includes changes to rate constants and reactions of the model species, but not the model species assigned to emissions, then no change is needed to the speciation assignments in this database so no update is needed. However, if the updated mechanism changes model species assignments for any compound, as may happen if a compound that was previously lumped is now represented by its own model species, then it must be considered to be a different mechanism from the perspective of this database, even if no other changes are made other than for one compound. In this case, the recommended procedure is to add the updated mechanism as if it is an entirely different mechanism, so that emissions for use with models with the previous mechanism can still be processed. However, the assignments made for the previous version of the mechanism can greatly simplify the process of making assignments for the new version.

The steps involved in adding a new or updated mechanism with different lumping assignments are given in the following sections.

5.1.1.   Additions to the SpecDB.XLS "Parameters and Commands" Sheet

Near the top of the "Parameters and Commands" sheet there is a table called "Supported Mechanisms". Add a row where desired in this table, generally following the row for the previous version of this mechanism if applicable or before the row for "Atoms" if there is no previous mechanism of this type. It does not matter where it is added provided that there are now blanks separating the rows of supported mechanisms. Copy the calculations in the "ok" and the "Description" columns (in blue font) to the corresponding locations in the new row. At this point it does not matter what is shown in the new row. Add the data in the "MechID", "Type" and "Code" columns as follows:

"MechID": This is a short, single-word name that identifies the mechanism as it is (or will be) generally known in the modeling community. The only restriction is that it is nonblank and have no leading or trailing blanks, is different from those used for the other mechanisms (including previous versions that are retained) and that it is legal for use as a sheet name in an Excel spreadsheet. The latter requirement means that it cannot contain some characters (e.g. "\" and "*") and cannot be too long.

"Type": This is a numerical code indicating how model species are assigned to compounds in the mechanism, which depends primarily on how many model species may be required to represent a compound, and whether any compound uses more than one mole of a model species. It also indicates whether this mechanism should be assumed to have assignments for all compounds that are assigned in other mechanisms. The type codes are as described on Table 12.

"Code": This is a 3-5 character alphanumeric code that also uniquely identifies the mechanism and that is legal to use as the first part of an Excel range name. These are used to define ranges in the "Compound" sheet where the assignments for the particular mechanism are given.

5.1.2.   Additions to MechDB.xls

The MechDB.xls file contains summary information about the supported mechanisms, both in a summary table given in the "Summary" sheet, and also in a separate worksheet for the mechanism containing information about the mechanism and the list of model species that can be assigned to organic emissions. To add a mechanism, a row for the mechanism must be added to the summary table in the

Table 12.   Mechanism type codes describing how mechanism assignment data are given in SpecDB.xls and whether assignments must be made for all compounds with assignments in other lumped mechanisms.

Type	Description
-1	Mechanism has one mole of model species for all assigned compounds. There may be no model species assignments for some compounds that have assignments for the lumped mechanisms. This is generally used for mechanisms, such as MCM, that do not have lumped model species for compounds that are not explicitly represented. The assignments for this mechanism are given in a single column in the "Compounds" sheet.
≥0	Mechanism is treated as a "lumped" mechanism that has assignments for all compounds that have assignments for the other lumped mechanisms. Model species assignments must be made for all compounds in the "Compounds" sheet except for those flagged as being unassigned for all mechanisms.
0	Compounds are assigned exactly one mole of a model species. This can be used if the mechanism has no "reactivity weighting" or compounds represented using more than one model species. The assignments for this mechanism are given in a single column in the "Compounds" sheet.
1	Compounds are assigned a single model species, but not necessarily on a mole-for-mole basis. This is generally used for mechanisms that employ "reactivity weighting" for some compounds, where assignments to something other than one mole of model species is used so one model species can represent compounds of varying reactivity. The assignments for this mechanism are given in two columns in the "Compounds" sheet, the first one giving the model species name and the second given the moles model species assigned to the compound.
2-9	Compounds can be assigned more than one model species, but no more than the code number. This is used for mechanisms has can have more than one model species representing a compound, but not more than 10. The assignments are given in 2 x Type numbers of columns, the first giving the model species name and second other giving the moles of model species assigned. The first set of columns must be nonblank for assigned compounds; the remainder are nonblank only if needed.
≥10	Compounds in general are assigned more than one model species, and the assignments are given with one column in the "Compounds" sheet for each model species, giving the moles of that model species that is used for the compound. Blank or zero in the column means that that model species is not used for the compound. This is used primarily for Carbon Bond-type mechanisms and also the "atoms" pseudo-mechanism. The type number gives the number of columns and model species used. Note that columns are needed only for those model species used to represent emitted organics. It is recommended that this number be calculated using the "COUNTA()" spreadsheet function using as input the column headers for the mechanism in the "Compounds" sheet, since this will assure that the number will remain correct if columns are added or deleted. However, this cannot be done until after the columns and column headers are added to this "Compounds" sheet as described below.

"Summary" sheet and a new sheet needs to be created and filled out for the mechanism, as discussed below.

Mechanism worksheet. The best way to create a worksheet in MechDB.xls for a new mechanism is to take a sheet for an already implemented mechanism that is the most like the one being added and make a copy of this sheet by dragging the sheet tab to another part of the spreadsheet while holding down the "control" key. This will assure that the new sheet has the range names that must be in the mechanism sheet for the mechanism to be processed. Generally something other than a "SAPRC" mechanism should be used unless this is a SAPRC mechanism. This will make a new sheet with the necessary range names that the system requires. Ignore any warning messages about too many characters in cells, if applicable. The new sheet can also be created manually, but in this case the range names need to be defined as indicated on Table 13. Rename the new or copied sheet to be the same as the mechanism name given in the "MechID" column in the "Parameters and Commands" sheet of SpecDB.xls as discussed above. Fill out the general mechanism information as indicated in the top part of Table 13 as appropriate. Then fill out a model species listing table, where each row describes model species that may be assigned to emitted organics, and the columns contain at a minimum the information given on the bottom part of Table 13. The special model species NVOL and NROG be included, even if the mechanism does not need to use the NROG model species. Additional columns can be included in this table to give further information about the model species, such as descriptions of the model species, names of the species used internally in the model code, non-integer carbon numbers, and other information that may vary from mechanism to mechanism. Examples can be seen by examining the tables for the various mechanisms that are already implemented. If the sheet were created from that for a similar mechanism, it may already have the columns that may be appropriate for the new one.

Mechanism Summary Table. An entry for the mechanism must also be added to the mechanism listing table in the "Summary" sheet of MechDB.xls. Add a row for the new mechanism, preferably in the same location as the row was added in the "Parameters and Commands" sheet of SpecDB.xls as described in the previous mechanism. The "MechID" and "Type" column should be the same as the mechanism summary table in the "Parameters and Commands" sheet of SpecDB.xls, though the data in the "Type" column is not actually used. The "ok" column should contain a "1" if this mechanism is considered ready for use when processing emissions data and blank otherwise. It is blank for "Atoms" because that is not actually a mechanism, and for "SAPRC-DMS" and "MCM" because they do not have assignments for all compounds for which assignments are made for lumped mechanisms. Copy the calculated data (in blue) from a row with another mechanism to fill out the row for the new mechanism.

To see if this mechanism summary table in MechDB.xls was set up correctly, check the mechanism listing table in the "Parameters and Commands" sheet of SpecDB.xls. It should have the appropriate information in the "ok" and "Description" columns. If it does not, check to be sure that the MechID name on this table matches that in the table in MechDB.xls.

5.1.3. Additions to the "Compounds" sheet of SpecDB.xls

Find the columns in the "Compounds" sheet in SpecDB.xls that has the model species assignments for the mechanisms, and add the number of columns needed for this new mechanism. The numbers of columns needed are based on the "Type" code given to the mechanism as shown on Table 12. It is recommended that this be done by copying the assignment columns for a previously-implemented mechanism that is most similar to this one, such as a previous version if this is an update.

Select the cell in the column heading immediately above the row with the first compound in the first column with the mechanism assignments, and give it the range name MechAsn*Code*, where *Code* is the mechanism assignment code given for the mechanism in the "Code" column of the mechanism

Table 13.   Information about a mechanism and its model species that are included in the mechanism's worksheet in MechDB.xls.

Label	Description	Range Name [a]
	General Mechanism Information	
Mechanism ID	Sheet name and also MechID as given in the mechanism table in the "Parameters and Commands" sheet in SpecDB.xls	ModSpeName
Assignment Code	Code entered in the "Code" column in the table of mechanisms in the "Parameters and Commands" sheet of SpecDB.xls and used to identify the assignments in the "Compounds" sheet as discussed above	MechAsnCode
Description	Short description of the mechanism that is output in the mechanism summary table produced for the Speciation Tool as discussed in Section 3.2.3, and also shows up in the mechanism tables in the "Parameters and Commands" sheet of SpecDB.xls and the summary table of MechDB.xls.	MechDesc
Version	The version of the mechanism, which can be given for information purposes but is not used by the current system. Optional.	MechVer
Assignment Desc.	Information about the assignments such as the date of the assignment. This is optional and not used by the current system	MechAsnDesc
Output Parm ID	Logical value set to TRUE if the mechanism has model species with adjustable parameters and FALSE if not. This is not used by the current system and is only currently applicable for some SAPRC mechanisms. Optional.	MechAsnOutput ParmID
No. Species	Number of model species in the model species listing table, including the required "NVOL" and "NROG" species. It is recommended that this be derived using the "COUNTA()" spreadsheet function, using the model species name listing as input. This will assure that the number is correct even if species are added or removed from the list. It is particularly important in mechanisms given a "Type" code of 10 or more in the mechanism table in the "Parameters and Commands" sheet of SpecDB.xls.	nMechModSpe
	Model Species Table [b]	
Name or Emit Name	Name of the model species that is to be output to the emissions system. It should include all model species that are assigned to compounds in the "Compounds" sheet, but does not need to include model species used for other purposes.	ModSpeName
MWt	The molecular weight in grams per mole assigned to the model species. A molecular weight of 1 must be used for NROG and NVOL, since they are output on a mass basis.	ModSpeMWt
nC or Int nC	Carbon number, which must be given as an integer. If the mechanism documentation gives a non-integer carbon number an optional separate column can be given with the documented carbon number, but this would not be used for processing.	ModSpeNC

43

Label	Description	Range Name [a]
Mol/Gm UNKN	The numbers of moles model species per gram of the mixture used to represent unknown mixtures. The data should initially be blank in this column because they are computed by macros.	ModSpeUnkn
Mol/Gm UNKN	The numbers of moles of model species used to represent a mole of unassigned compounds computed using the composition of the mixture used to represent unassigned compounds. The data should initially be blank in this column because they are computed by macros.	ModSpeNoasn

[a] These do not need to be given if the mechanism worksheet is created from that for another mechanism in a way that preserves range names. For the model species table the range name should be given to the column header cell, in the row immediately above the information about the first model species.

[b] Additional columns can be included to give optional or other information about the model species, though this is not used for processing. Those listed on this table give the minimum data required.

summary table in the "Parameters and Commands" sheet. For example, the header row for the first column with the SAPRC99 assignments was given the range name "MechAsnS99L".

If this is a mechanism with Type code of 10 or greater, label all the column headers with the model species names in the order they appear in the model species listing in the mechanism's sheet in MechDB.xls, being sure to include NROG and NVOL, even if NROG is not used. Then go to the entry in the table for this mechanism in the "Parameters and Commands" sheet and enter a "COUNTA()" spreadsheet function in the "Type" column, and have it count the number of model species names in the column headers for this mechanism. This will assure that the correct number of model species will be processed, or model species will be missed or inappropriate data may be output.

If this is a lumped mechanism, i.e., a mechanism given any "Type" code other than -1 (see Table 12), model species assignments must be given for all compounds that have assignments for the other lumped mechanisms. If any of these are missing the macros processing the assignments will fail with an error message giving the mechanism and the compound ID for the first instance found. This can be a very time consuming process because assignments need to be made for (currently) over 1800 compounds, but is greatly expedited if assignments for a similar mechanism are used as the starting point. If there is no sufficiently similar mechanism, the following steps may be helpful.

- Compounds that have no assignments for other lumped mechanisms, i.e., those with a "1" in the "none" column under "assigned" near the first set of columns in the table, can be ignored.

- Compounds that are indicated as non-volatile, i.e., assigned the model species "NVOL" in other mechanism and have the appropriate flags in the "is VOL" and "is NVOL" columns near the first sets of columns in the table, should be assigned the NVOL model species unless the mechanism has other appropriate model species for these non-volatile compounds.

- Go through the list of explicitly represented compounds in the mechanism, find the compounds in the table (using either the CAS number or the molecular formula to aid the search) and assign it the corresponding model species.

- It may be useful to make a table mapping model species in the new mechanism to those in a more detailed mechanism (if available), so assignments for the new mechanism can be derived

44

from those for the previous mechanism using the VLOOKUP() spreadsheet function. Cases where there is no direct correspondence can be excluded to avoid inappropriate assignments, but hopefully this method may be useful for large numbers of compounds.

- The remaining compounds can be sorted by atom numbers, molecular weights, structural codes, estimated kOH values, or model species assigned for other mechanisms to aid making assignments by means of chemical judgment.

These assignments obviously should be made by someone familiar with the objectives of the intentions of the developer(s) of the mechanism being implemented. Close consultation with the mechanism developers may be required unless the new mechanism has only a few well-defined model species assignment differences than a previously implemented mechanism.

5.2. Adding or Updating Profile Databases

The only characteristic of speciation profile databases that is relevant to data in SpecDB.xls is the set of codes used to identify the chemical categories when defining the compositions of these profiles. No changes need to be made to SpecDB.xls or its output files when emissions speciation profiles are added or modified, as long as the same set of chemical category codes are used to specify the compositions. However, SpecDB.xls would need to be updated if new chemical categories are added, as may occur if a new compound is identified that was not previously used in the profile database, or if the codes used to identify the categories in the profiles are changed in any way.

The currently supported profile categorization systems are given in the table labeled "Codes for Category names in supported emissions profile databases" given in the "Parameters and Commands" sheet of SpecDB.xls, immediately under the table listing the mechanisms. Note that each two-or three character code refers to a system of identifying compounds of chemical categories in profile databases, each of which could be used in more than one database. These codes must be specified in the input to the ProfPre and ProfAsn programs when processing profile data so the programs will know which chemical categories are being referenced. Note that use of the S40 code (for Speciate 4.x categories) is implied when processing assignments for import into the Speciation Tool as described in Section 3.2.3.

A new categorization system does not need to be added to the system when codes are added or deleted from existing categorization systems, as long as the meanings of existing codes for the system are not changed. It is recommended that categories that are no longer used when profile databases are updated not be deleted so that older profiles can still be processed, as long as the code is not "recycled" for another purpose (which is a very bad idea).

The profile category assignments are implemented in a table in the "SAROAD Assignments" sheet of SpecDB.xls, whose contents are described in Table 6. If a new category and category code is added to a profile database, then a new row of data can be added to this table assigning a SpecDB unique ID code to the profile database code. However, if the compound or mixture category that needs to be added has no corresponding category in the SpecDB "Master List" on the database, then the new category will have to be added as described in the following section. The unique ID code for this new category can then be used where indicated on Table 6.

A new categorization system will need to be added if the profile database or emissions speciation program uses different category codes than those listed on Table 5 or subsequently added. It also is needed if at least one code is changed to refer to a different compound or mixture in an existing categorization system if it is desired to still be able to process profiles using the previous version of the system. In this case, a unique code of no more than 4 characters must be assigned to this new system and

added to the "Codes for Category Names" table in the "Parameters and Commands" sheet. There should be no blanks rows in the table since mappings for categories following a blank row are not output by the macros that create the category name mapping files. The mappings for all the categories for the new system should then be added when to the "SAROAD Assignments" sheet as indicated in Table 6.

Anytime the data in the "SAROAD Assignments" sheet is updated, the "OutputMasterSARAssts" macro needs to be run. The best way to do this is to run the "UpdateAll" macro, which will execute this as well as all other macros that need to be run whenever the database is updated and updated data files are to be output. "UpdateAll" must be used if any new categories have to be added to the "Master List", "Compounds" or "Mix Profile Assignment" sheets, as discussed below.

5.3. Adding New Speciation Categories

In most cases, the master category listing in SpecDB.xls will already include compounds and mixture categories that are already on the master list in the database, so these can be used in the mappings in the "SAROAD" assignment sheet for this database. It is important that duplicate categories not be added, especially if the category refers to an individual chemical compound. If the new category refers to a compound, first look for the CAS number to be sure it is not already listed, and if not found then sort the compounds list in the "Compounds" sheet by atomic formula and check each of the listed compounds that have same formula. If it is a mixture category, try to find an existing mixture categories in the "Master List" sheet that should be similar and appropriately assigned the same composition before adding new mixture categories. Otherwise, the database will become unnecessarily large and the possibility of inconsistencies in assignments for different databases will increase.

However, in some cases new compounds or mixtures may have to be added to SpecDB.xls in order process new categories in new or updated speciation databases or models. These need to be added to the "Master List" sheet in all cases, and in some cases to the "Compounds" or "Mix Profile Assignment" sheets as discussed below.

All categories need to have an entry in the master category table in the "Master List" sheet of SpecDB.xls. The information that must or can optionally be provided for each category is described in Table 3. The row containing the information for the new category should be added in an appropriate place in this table, without the table having any blank rows in the table. Generally the categories are ordered first by type (compound, mixture, etc.), with the compounds listed first by carbon number and then by molecular weight in ascending order. The macros processing the assignment data require that the compounds be given before mixtures (except for the special mixtures "UNKNOWN" and "NONVOL"), followed by mixtures whose compositions are specified only by compounds, then by mixtures whose compositions are specified only by categories that are above them on the list. The ordering for unspeciated mixture categories are not important, but generally they are given last. There must be no blank rows in the table before the greyed-out row at the end of the table.

5.3.1.   Adding Single Compound Categories

If the new category refers to an individual chemical compound or (for higher molecular weight compounds) a mixture of isomers that is always treated the same way, then it must also be included in the "Compounds" sheet where compound-specific information and model species assignments are given. The Unique ID assigned to compounds should be formed from "C" + its CAS number, unless the compound has no CAS number of no CAS number uniquely identifies it. This can be checked by Googling the compound and being sure that only one CAS number shows up for it in several listings, and then Googling the CAS number and being sure that only this compound shows up for it. If the CAS number is

unknown or poorly defined (as is the case for some compounds in the database), it should be given another unique designation, which should not look like a CAS number.

The information that must or optionally be given for each compound category is indicated in Table 7. The "Description" field must uniquely identify the compound and isomer and trade names or acronyms should be avoided. Although structural code information and a Smiles string is optional, it is a good idea to include it because it aids making model species assignments when new mechanisms are added, and it may also help making assignments for this new compound.

If the compound is judged to be non-volatile then the compound should be assigned the "NVOL" model species for all mechanisms where this is appropriate (currently all the supported mechanisms). It then needs to be assigned model species in all the lumped mechanisms (i.e., all mechanisms with Type codes >0 as indicated on Table 12), unless it is absent from or only a very minor constituent of all profiles. The easiest way to do this is to find the most similar compound already on the list and give this new compound the same model species assignments. This procedure will not work for explicit mechanisms such as MCM or SAPRC DMS. No assignments should be given for those mechanisms unless it is known that they have a model species for this compound or (in the case of SAPRC DMS) an isomeric mixture that includes this compound.

If the new category refers to a mixture of isomers then it is generally better to add the category as as an isomeric mixture category as described in the following section, then add entries for all or at least representative isomers as compounds, and then define the composition of this category in terms of these isomers. However, this is not necessary for optical isomers that are treated separately in any profile database or mechanism, and can be impractical for mixtures of isomers of high molecular weight compounds that can have many isomers that are usually not identified or treated differently in any mechanism. However, if the category designated as a compound refers to an isomeric mixture other than optical isomers, then the "Unsp. Cmpd" field in the "Compounds" table (see Table 8) should indicate that this is the case.

5.3.2. Adding Known Mixture Categories

Mixture categories whose compositions are known or specified are indicated as such in on the "Master List" sheet, as shown on Table 3. This is indicated by the category being given a type code of "1" for mixtures of isomers and "2" for other mixtures, an assignment type code of "CMPD", "SMX" or "CMP", and a nonblank entry in the assignment "Representation" field. The "Representation" field can refer to a single compound or another mixture category that appears before this one in the "Master List" table or it can contain the same ID as used for this new category. In the former case the composition of the mixture is the same as used for the other category, which can refer to a single compound category if it is judged to be appropriate to assign it model species for that compound for all mechanisms. In the latter case the composition of this new category must be given in the "Mix Profile Assignment" sheet as discussed below.

The information given about the mixture compositions consists of rows of data in the "Mix Profile Assignment" sheet as indicated in Table 10. The assignments can be given in any order, but all assignments for a given mixture must be given together, and there should be no blank rows in the table. The compositions can be given in terms of compounds or other known mixture categories, but all categories used must be above it in the "Master List" sheet or the macro compiling the assignments will fail. Note that the compositions must be given as mass fractions, though this is the same as mole fractions for isomeric mixtures. The macros do not check for 100% of the mass being assigned, but the workbook includes calculation cells that can be used to check for this, as indicated on Table 10. Do not enter mass fractions that do not sum up to exactly 100%.

It is recommended that the "Notes" columns contain information to document how the composition of the mixture was derived and the assignment "code" column contain the code indicating a subjective judgment of the quality of the assignment, as indicated on Table 3. Although not needed for processing, it could be helpful for those who want to know how the assignments were derived and is good database practice.

5.3.3. Adding Unknown Mixture Categories

Categories whose chemical compositions are unknown or not yet assigned are also indicated as such in on the "Master List" sheet, as shown on Table 3.If the mixture is judged to be volatile then "NVOL" should be entered in the assignment "Type" column, and "NONVOL" in the assignment "Representation" column. If it is judged to be volatile, the assignment "Type" column should be blank and the "Representation" column should be "Unknown".

6. References

Adelman, Z. J. Vukovich, and W. P. L. Carter (2005): "Integration of the SAPRC Chemical Mechanism in the SMOKE Emissions Processor for the CMAQ/Models-3 Airshed Model," Final report to American Chemistry Council Agreement No. 1846. Available at http://www.cert.ucr.edu/~carter /absts.htm#smokespr

Carter, W. P. L. (2000): "Documentation of the SAPRC-99 Chemical Mechanism for VOC Reactivity Assessment," Report to the California Air Resources Board, Contracts 92-329 and 95-308, May 8. Available at http://www.cert.ucr.edu/~carter/absts.htm#saprc99.

Carter, W. P. L. (2010a): "Development of the SAPRC-07 Chemical Mechanism and Updated Ozone Reactivity Scales," Final report to the California Air Resources Board Contract No. 03-318. January 27. Available at www.cert.ucr.edu/~carter/SAPRC.

Carter, W. P. L. (2010b): "Development of the SAPRC-07 Chemical Mechanism," Atmospheric Environment, 44, 5324-5335.

Carter, W. P. L. (2010c): "Development of a Condensed SAPRC-07 Chemical Mechanism, Atmospheric Environment, 44, 5336-5345.

Carter, W. P. L. (2015a): "Development of an Improved Chemical Speciation Database for Processing Emissions of Volatile Organic Compounds for Air Quality Models," web site at http://www.cert. ucr.edu/~carter/emitdb.

Carter, W. P. L. (2015b): "Development of a Database for Chemical Mechanism Assignments for Volatile Organic Emissions," Submitted to the *Journal of the Air & Waste Management Association*, Jamuary. Manuscript available upon request.

Carter, W. P. L. and G. Heo (2013): "Development of Revised SAPRC Aromatic Mechanisms," Atmos. Eviron., 77, 404-414.

CMAS (2014): Community Modeling and Analysis Center website at https://www.cmascenter.org/. Last accessed January, 2014. Page with Speciation Tool download was last modified on 7/24/13.

EPA (2014a): SPECIATE Version 4.4, Speciation database available at http://www.epa.gov/ttnchie1/ software/speciate/. Last modified March 20, 2014; database is dated February, 2014.

EPA (2014b): File gspro_cmaq_cb05_soa_2011ec_v6_11f_beis_nf.txt, giving assignments of model species to BEIS 3 categories for various mechanisms. Included in ftp://ftp.epa.gov/ EmisInventory/2011v6/v1platform/data_common_to_all_years/ge_dat_for_2011ec.zip. Accessed November 13, 2014.

Gery, M. W., G. Z. Whitten, and J. P. Killus (1988): "Development and Testing of the CBM-IV for Urban and Regional Modeling," EPA-600/ 3-88-012, January.

Goliff, W. S., Stockwell, W. R. and Lawson, C. V. (2013): "The regional atmospheric chemistry mechanism, version 2," Atmos. Environ. 68, 174-185.

Guenther, A. B., X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons, and X. Wang (2012): The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471–1492,

Heo, G., W. P. L. Carter, and Qi Ying (2014): "Environmental Chamber Experiments and CMAQ Modeling to Improve Mechanisms to Model Ozone Formation from HRVOCs," Final Report for Texas AQRP Project 12-006, February. Available at http://aqrp.ceer.utexas.edu/ viewprojectsFY12-13.cfm?Prop_Num=12-006

Hutzell, W. T., D.J. Luecken, K.W. Appel, and W.P.L. Carter (2012): "Interpreting predictions from the SAPRC07 mechanism based on regional and continental simulations," Atmos. Environ., 46, 417-429.

Jimenez, M., U Nopmongcol, G. Yarwood (2013): Speciation Tool Users Guide, Version 3.1, Prepared for U.S. EPA, July. Available at http://www.cmascenter.org/help/model_docs/speciation_tool /3.1/Sptool_UG_V3.1.pdf.

MCM (2015): The Master Chemical Mechanism, available at http://mcm.leeds.ac.uk/MCM/, accessed January, 2015.

MEGAN (2014): Model of Emissions of Gases and Aerosols from Nature (MEGAN), web page at http://acd.ucar.edu/~guenther/MEGAN/MEGAN.htm, undated.

Pierce, T, B. Lamb, and A. van Meter (1990): Development of a Biogenic Emissions Inventory System for Regional Scale Air Pollution Models," presented at the 83rd Air and Waste Management Association Annual Meeting, Pittsburgh, PA, June.

Schwede, D., G. Pouliot, and T. E. Pierce. (2005). Changes to the biogenic emissions inventory system version 3 (BEIS3). 4th Annual CMAS Models-3 Users' Conference, September 26: 28.

SMOKE (2013): Official SMOKE website hosted by the Center for Environmental Modeling for Policy Development at the University of North Carolina at Chapel Hill. Available at http://www. cmascenter.org/smoke/. Last modified December 4.

SMOKE (2014): Section 6.12. of SMOKE Core Programs: Normbeis3 Description. SMOKE web site at http://www.cmascenter.org/smoke/documentation/3.1/html/ch06s12.html. Last modified date unknown. Accessed February, 2014.

Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang (1990): "The Second Generation Regional Acid Deposition Model Chemical Mechanism for Regional Air Quality Modeling," J. Geophys. Res. 95, 16343-16376.

Thomas, R. (2013): Ron Thomas, Texas Commission on Environmental Quality, Austin, Texas, personal communication, August 22.

US (2014): U.S. Code of Federal Regulations: Title 40: Protection of Environment; Part 51 - Requirements for Preparation, Adoption, and Submittal of Implementation Plans; Subpart F— Procedural Requirements, 51.000, Definitions. Available at http://www.ecfr.gov/cgi-bin/text-idx?c=ecfr&rgn=div8&view=text&node=40:2.0.1.1.2.3.8.1&idno=40. Current as of January 16 when accessed.

Yang, Wenli (2013): California Emissions profiles. Wenli Yang, California Air Resources Board, personal communication, December, 2013.

Yarwood, G., S. Rao, M. Yocke, and G. Whitten (2005): "Updates to the carbon bond chemical mechanism: CB05. Final report to the U.S. EPA, RT-0400675. Available at http://www.camx.com/publ/pdfs/CB05_Final_Report_120805.pdf.

Ying (2013): MCM emissions assignments, Qi Ying, Texas A&M University, personal communication, April 1.