Development of a Chemical Speciation Database and Software for Processing VOC Emissions for Air Quality Models

William P. L. Carter
College of Engineering Center for Environmental Research and Technology
University of California, Riverside, California 92521
carter@cert.ucr.edu

ABSTRACT

The chemical categories used in volatile organic compound speciation profiles in current emissions databases are inconsistent, include many categories that are not actual compounds, and assignments to these categories to different chemical mechanisms are inconsistent and difficult to update as new mechanisms or profiles are developed. To address this, we developed a chemical speciation database that includes assignments of actual chemical compounds to speciation categories used in EPA, California, and Texas organic emissions profiles. These chemical compounds were then assigned to model species in the Carbon Bond 4, RADM 2, and SAPRC-99 chemical mechanisms. Procedures and computer programs are being written that that these assignments for processing emissions for these mechanisms, and to allow for the capability of readily adding new model species or using SAPRC-99 model species whose mechanisms depend on the compounds they are representing. The programs are being written for inclusion into the SMOKE emissions system, but can also be run independently. The speciation assignments and profiles are available as Excel files that can be downloaded from the project web site at http://www.cert.ucr.edu/~carter/emitdb, and the programs and example calculations will be available there soon.

INTRODUCTION

Many hundreds of types of volatile organic compounds (VOCs) are emitted into the atmosphere, each having different atmospheric reaction rates, degradation mechanisms, and effects on formation of ozone and other atmospheric impacts such as secondary PM. Because of this, it is important that the chemical compositions of VOC emissions sources be appropriately represented in emissions databases, and that appropriate methods be utilized to represent these chemical differences in models. However, current models, emissions databases, and emissions processing systems have a number of deficiencies in this regard. These can be summarized as follows.

- The speciation profiles used to represent many types of VOC sources are out-of-date or of uncertain quality. No centralized database and procedures exist where new speciation information can be compiled, reviewed, or made available to modelers.
- Most models use highly condensed mechanisms that lump many different types of VOCS into a limited number of model species. Though generally adequate for predicting ozone

in current day scenarios, they are not adequate for assessing effects of VOC control strategies that result in changes in relative compositions of emitted VOCs.

- Current speciation databases have inconsistent chemical categorizations with many poorly defined and redundant classifications. This makes it difficult to process emissions for different mechanisms in a consistent manner, and has impeded implementing new or more chemically detailed mechanisms in models. It also makes it too easy to add low-quality, poorly defined speciation profiles to emissions databases.
- Current emissions processing databases use direct assignments of model species to emissions profiles. This also makes it difficult to implement new or chemically detailed mechanisms in models, or to represent individual chemicals separately in existing models, as is desired for toxics or VOC reactivity modeling.

Developing improved profiles and databases and procedures where they can be made available to modelers is obviously essential and will require a major and ongoing effort. However, the chemical categorization methods, model species assignments, and procedures for implementing these assignments into models are equally essential, since without them the models cannot take advantage of whatever speciation information is available.

For example, the SAPRC-99 chemical mechanism (Carter, 2000a) was designed to permit airshed models to incorporate the detailed chemical speciation information into models using a manageable number of model species that can vary depending on the needs of the model application. It gives the modeler the choice of which compounds are represented explicitly and which are lumped, and allows for reactions of lumped species in the model to be adjusted based on the specific set of compounds they represent. This makes it well-suited for VOC reactivity assessment or for air toxics modeling. However, deficiencies in the current emissions speciation assignment data and methods and emissions processing software have resulted in models not taking advantage of this "adjustable parameter" capability of the SAPRC-99 mechanism.

One important source of inconsistency in assignments of model species to current speciation profiles results from the direct assignments of model species to mixture categories. In order to do this, the chemical mechanism developer or implementer must make assumptions about what the mixtures actually contain. Since mechanisms are implemented at different times by different people, this can result in inconsistent assignments among mechanisms. A better approach would be to make model species assignments to individual compounds only, and derive the assignments for mixture categories based on assumed compositions for the mixtures that are the same regardless of which mechanism is processed. Not only is it more straightforward to implement new or more detailed mechanisms, it is also easier to update and improve the representation of the mixture categories as a separate effort when new information becomes available.

In order to address this problem, we developed a new chemical speciation categorization database with chemically consistent assignments for the SAPRC-99, Carbon Bond, and RADM-2 chemical mechanisms, and software to use these assignments to derive model species assignments for speciation profiles, using different chemical mechanisms and lumping schemes, including the adjustable parameter version of SAPRC-99. This is described in this paper.

DATABASE DEVELOPMENT AND ASSIGNMENTS

Compilation of Speciation Assignments

A list of all the chemical categories in the emissions speciation databases used by the EPA for processing emissions for Models-3 (Gipson, 2001), the EPA's Speciate 3.2 database, (Speciate, 2001) the Texas Natural Resource Conservation Commission (TNRCC) (Yarwood, 2002), and the California Air Resources Board (CARB) (Allen, 2001) was compiled. Although the EPA, California, and Texas databases all used 6-digit SAROAD numbers to define the categories, there were a sufficient number of inconsistencies between them that each had to be treated as a separate categorization in this project. Duplicate categories were merged, a master list containing unique categories was compiled, and these unique speciation categories were assigned to the categories used in the various emissions databases. The resulting database contained 1973 master speciation categories to represent the total of 3639 emissions categories from the 4 databases. Some new categories were added to the database to define the compositions of the mixtures and to represent the SAPRC-99 detailed model species that were not already in the emissions databases.

CAS identification numbers were determined for all the speciation categories where these are well defined and uniquely identify a specific chemical or mixture, and were used to identify these categories using a "C"+CAS number designation (e.g., "C74-84-0" for ethane). The Chemfinder.com and ChemExper.com web sites were used as the primary means for this purpose, with the SciFinder Scholar program, which directly accesses the CAS database, being used for compounds not found on the other databases until that resource became unavailable as a free resource for academics in early 2002. Chemicals that do not have CAS numbers that we could determine or mixtures for which CAS numbers are not assigned or poorly defined were given separate identification codes based on SAROAD numbers used in EPA or other database (e.g., "S1-43155" for "isomers of heptadecane", based on the EPA SAROAD number, or "S2-97000" for "C4-naphthalenes", based on the CARB chemical codes.)

Since 585 the categorizations referred to mixtures, with varying degrees of uncertainty as to their actual compositions, each of the categories were classified as either individual compounds, mixtures of isomers ("simple mixtures"), mixtures of compounds that are not necessarily isomers ("complex mixtures"), and poorly defined categories that could not be associated with any particular VOC or VOC mixture. Each of these types of categories is treated separately in the database as described below. The current database has 1388 individual compounds, 367 simple mixtures, 186 complex mixtures, and 32 poorly defined classes.

For each speciation category identified as an individual compound, the chemical formula (numbers of each type of atom) was determined and used as the basis for computing the molecular weight, and other relevant information about the compound was compiled. This included its CAS number, its VOC status (VOC exempt or reactive according to EPA's regulations), and our judgment as to whether the compound is sufficiently volatile to undergo gas-phase reactions. The latter judgments were largely subjective, and if there was considered to be any uncertainty or even partial volatility the compound was not flagged as nonvolatile. For

example, nonadecane (n-C_{19}) was classified as volatile, while eicosane (n-C_{20}) was classified as nonvolatile. This classification is relevant to emissions processing for models because it is inappropriate to represent non-volatile compounds using volatile model species. This is discussed further below.

Since model species assignments should be made only to compounds, it is necessary to assign compounds to all mixture categories for model species for these categories to be derived. For each speciation category identified as a mixture of isomers, a distribution of up to 8 individual compounds was chosen to represent the mixture for modeling purposes. The mixture was chosen based on either existing assignments in the SAPRC-99 mechanism or using a representative set of compounds of this type in the database. Composition assignments were made for as many of the complex mixture speciation categories as possible. In most cases, a separate set of "profiles" were used to represent the compositions of the mixtures, though in a few cases representation by a single compound or simple mixture was considered to be appropriate. The representative profiles were taken from existing inventories or estimated in this work, and their derivations are documented in the database. Flags were used to indicate which assignments of compounds to simple mixtures are considered to be arbitrary (which was the case for the majority) or particularly uncertain, and comments were used to indicate how the complex mixture compositions were derived, and the assumptions made in this process.

Some of the 32 poorly defined emissions categories were identified as polymers or other substances that had no volatility, so these were identified as such so that their mass would not be counted when making assignments for emissions. Others included Carbon Bond model species that do not correspond to any particular compound or mixture, or had ambiguous, unknown, or incorrect compound designations. These are treated as unknown mixtures for emissions processing purposes.

Chemical Mechanism Assignments

Model species assignments were made for the SAPRC-99 (Carter, 2000a,b), Carbon Bond 4 (CB4) (Gery et al, 1988, 1989), and RADM-2 (Stockwell et al, 1990, Middleton et al, 1990) mechanisms to as many of the 1388 individual compounds as feasible. In the case of SAPRC-99, assignments were made for three levels of detail, to permit use of both the detailed and condensed version of the mechanism and to aid in the implementation of other condensed “lumped molecule” mechanisms. These were made as follows:

SAPRC-99 Detailed. The SAPRC-99 mechanism has over 600 “detailed model species” for which separate mechanism assignments have been made (Carter, 2000a). Some of these represent single compounds explicitly, while others groups of compounds considered to have very similar mechanisms (or for which we have insufficient information to represent mechanism differences), such as “branched C9 alkanes” or “C10 dialkyl benzenes”. Assignments of these model species were made to as many of the compounds as practical based on previous assignments to emissions categories, or added for this work. Not all compounds could be clearly associated with detailed model species, and such assignments were made to only 78% of these compounds. However, most of the compounds for which assignments could not be made are only

minor constituents of most profiles. The assigned compounds represent approximately 92% of the mass in the Speciate 3.2 profiles (computed by weighing each profile equally), and 94% of the in a profile representing emissions totals in an EPA Models-3 database (EPA, 1998).

SAPRC-99 Emissions Groups. In order to expedite and make more consistent mechanism assignments for "lumped molecule" condensed mechanisms such as SAPRC-99 and RADM-2, compounds were also assigned to "emissions groups", which in turn can be assigned to model species, using weighting factors if desired. This is similar to the approach used when making emissions assignments for the RADM-2 mechanism that was developed by Middleton et al (1990), except that an expanded number of groups was employed. These are listed on Table 1. Each SAPRC-99 detailed model species was also assigned an emissions group, and these assignments were used to derive the emissions group assignments to the compounds for which detailed model species were assigned. However, there are many compounds for which no appropriate detailed model species could be assigned that could be assigned to an emissions group, and attempts were made to make emissions group assignments to as many of such compounds as possible. This resulted in a total of 94% of the compounds being assigned to an emissions group, and this accounts for all but 0.002% of the mass in the Speciate 3.2 profiles with each profile weighed equally, or less than 0.001% of the mass in a profile representing emissions totals into a modeling domain (EPA, 1998).

SAPRC-99 Condensed. Carter (2000a) developed a condensed version of SAPRC-99 suitable for implementation in Models-3/CMAQ and other current airshed models. This mechanism uses 30 model species to represent the various classes of VOCs, using a lumping scheme similar to, but is slightly more detailed than, that used in RADM-2 (Stockwell et al, 1990). The assignments for this mechanism were made by assigning model species to the SAPRC-99 emissions groups as indicated on Table 1. Since the SAPRC-99 condensed assignments were based directly on those for the emission groups, the same fraction of the number and mass of compounds were assigned as indicated above for the emissions groups.

RADM-2. The RADM-2 mechanism was developed by Stockwell et al (1990) for use in acid deposition and regional models, and has been implemented in the CMAQ modeling system. Middleton et al (1990) derived emissions assignments to this mechanism by assigning each speciation category (which as discussed above included mixtures as well as compounds) to one of 32 emissions groups, which were then assigned to RADM-2 model species. In a number of cases reactivity weighting factors were used to represent estimated differences in kinetic reactivity between the emissions groups and the model species. For this work, RADM-2 emissions groups that do not represent mixtures were assigned to most of the SAPRC-99 emissions groups, and the RADM-2 mechanism assignments were based on those for the corresponding RADM-2 emissions group. This is indicated on Table 1. However, there were some SAPRC-99 groups that did not correspond to RADM groups, and assignments were made in this work to improve the representation of theses species in this mechanism. In addition, the assignment to the "styrenes" group was modified based on SAPRC-99 calculations indicating the reactivity of styrene is better represented by a "toluene" model species alone than a mixture of a toluene and olefin species. These assignments are indicated on Table 1, where footnotes indicate the basis for the assignments that were made.

Table 1. SAPRC-99 emissions groups and assignments of SAPRC-99 lumped and RADM-2 model species to each.

Emissions Groups		Lumped Mech. Asst's	
ID	Description	SAPRC-99	RADM-2 [a]
CH4	Methane	CH4	CH4
ETHA	Ethane	ALK1	ETH
PROP	Propane	ALK2	0.519 HC3
ALK2	Alkanes (kOH = 0.05-0.25 x 10^4 ppm^{-1} min^{-1})	ALK2	0.440 HC3 [b]
ALK3	Alkanes (kOH =0.25-0.50 x 10^4 ppm^{-1} min^{-1})	ALK3	0.964 HC3
ALK4	Alkanes (kOH =0.50-1.00 x 10^4 ppm^{-1} min^{-1})	ALK4	0.956 HC5
ALK5	Alkanes (kOH =1.00-2.00 x 10^4 ppm^{-1} min^{-1})	ALK5	0.945 HC8
ALK6	Alkanes (kOH>2.00 x 10^4 ppm^{-1} min^{-1})	ALK5	1.141 HC8
ETHE	Ethene	ETHE	OL2
OLE1	Alkenes (Primary) (also primary allenes)	OLE1	OLT
OLE2	Alkenes (Internal) (also internal allenes)	OLE2	OLI
OLEM	Alkenes with multiple double bonds (excluding allenes)	OLE2	OLI
13BDE	1,3-Butadiene	OLE2	OLI
ISOP	Isoprene	ISOP	ISO [c]
TERP	Terpenes with one double bond	TERP	OLI
TERP2	Terpenes with two or more double bonds	TERP	OLI
BENZ	Benzene	0.295 ARO1	0.293 TOL
HBEN	Halo and nitrobenzenes	0.295 ARO1	0.293 TOL
ARO1	Aromatics (kOH <2 x 10^4 ppm^{-1} min^{-1})	ARO1	TOL
ARO2	Aromatics (kOH >2 x 10^4 ppm^{-1} min^{-1})	ARO2	XYL
NAPT	Naphthalenes	ARO2	XYL
PHEN	Phenols	PHEN	CSL [d]
CRES	Cresols	CRES	CSL [c]
STYR	Styrenes	OLE2	TOL [e]
HCHO	Formaldehyde	HCHO	HCHO
CCHO	Acetaldehyde	CCHO	ALD
RCHO	Higher Aldehydes	RCHO	ALD
AALD	Aromatic Aldehydes	BALD	NR [f]
ACET	Acetone	ACET	0.253 KET
KET1	Ketones (kOH <0.73 x 10^4 ppm^{-1} min^{-1})	MEK	KET
KET2	Ketones (kOH >0.73 x 10^4 ppm^{-1} min^{-1})	PRD2	0.956 HC5
HCOOH	Formic Acid	ALK2	ORA1
CCOOH	Acetic Acid	ALK2	ORA2 [c]
RCOOH	Higher organic acids	ALK2	ORA2 [d]
ACTYL	Acetylene	ALK2	0.343 HC3
ALYN1	Terminal Alkynes	OLE1	OLT
ALYN2	Internal Alkynes	OLE2	OLI
PERC	Perchloroethylene	ALK1	0.078 HC3 [g]
TCE	Trichloroethylene	ALK3	1.198 HC3 [h]
HALKE	Other Haloalkenes	ALK4	HC5 [j]
OTH1	Others (kOH =0.02-0.05 x 10^4 ppm^{-1} min^{-1})	ALK1	0.109 HC3 [b]
OTH2	Others (kOH =0.05-0.25 x 10^4 ppm^{-1} min^{-1})	ALK2	0.404 HC3

Table 1 (continued)

Emissions Groups		Lumped Mech. Asst's	
ID	Description	SAPRC-99	RADM-2 [a]
OTH3	Others (kOH =0.25-0.5 x 10^4 ppm^{-1} min^{-1})	ALK3	1.215 HC3
OTH4	Others (kOH =0.5-1.0 x 10^4 ppm^{-1} min^{-1})	ALK4	1.075 HC5
OTH5	Others (kOH =1-2 x 10^4 ppm^{-1} min^{-1})	ALK5	1.011 HC8
OTH6	Others (kOH >2 x 10^4 ppm^{-1} min^{-1})	ALK5	1.141 HC8 [k]
MEOH	Methanol	MEOH	0.402 HC3 [h]
ETOH	Ethanol	ALK3	1.198 HC3 [h]
INHIB	Inhibitors	NROG	NR [f]
GLY	Glyoxal	GLY	GLY [c]
MGLY	Methyl Glyoxal	MGLY	MGLY [c]
BACL	Biacetyl	BACL	MGLY [c]
MACR	Methacrolein	MACR	MACR [d]
UALD	Other Unsaturated Aldehydes	IPRD	MACR [d]
UKET	Unsaturated Ketones	MVK	MVK [d]
PHOT	Other photoreactive	BACL	MGLY [m]
PAH	Polycyclic aromatic hydrocarbons	ARO2	XYL
BOAT	Pthallates and benzoates	ARO1	HC8 [n]
MARO	Miscellaneous other heteroatom-containing aromatics	ARO1	TOL [n]
MRCT	Miscellaneous heteroatom-containing non-aromatic reactive compounds	ALK5	HC5 [n]
AMIN	Amines and amides	ALK5	HC8 [n]
INERT	Unreactive	NROG	NR
NVOL	Non-volatile	Not represented	

[a] Unless noted otherwise, this is based on model species assignments given by Middleton et al (1990).

[b] The weighting factor is derived using the methods of Middleton et al (1990) based on an estimated OH rate constant in the middle of the range for this group.

[c] This compound or group of compounds is represented in the RADM2 mechanism, but was not on the list of assignments given by Middleton et al (1990).

[d] This is the most similar model species to this group of compounds. Not on list of assignments given by Middleton et al (1990).

[e] Middleton et al (1990) represents styrenes as a mixture of toluene (TOL) and 1-alkenes (OLT). However, calculations using the SAPRC-99 mechanism that was evaluated using experimental data of Carter et al (1999) give reactivities somewhat more consistent to that of toluene than 1-alkenes or toluene + 1-alkenes.

[f] Benzaldehyde and other inhibitors are treated as unreactive in the RADM2 mechanism.

[g] The generic haloalkene RADM2 assignment was based on this compound.

[h] The weighting factor computed from the OH rate constants for the compound and model species using the method of Middleton et al (1990). Because of the relative importance of this compound, it is considered to be preferable to the generic assignment used by Middleton et al (1990).

[j] The average of the OH rate constants for 1,2-dichloroethylene and vinyl chloride is close to the OH rate constant for the HC4 model species. The RADM2 assignment was based on perchloroethylene, which is much less reactive.

[k] This group is not used by Middleton et al (1990). The weighting factor for ALK6 is used.

[m] Based on the representation of this group in lumped SAPRC-99

[n] This is a very rough and uncertain estimate of which model species is most appropriate to this class.

Since the RADM-2 assignments for the compounds were derived directly from those for the SAPRC-99 emissions groups, the fractions of numbers of compounds and mass of profiles represented by the RADM-2 assignments are the same as those for the SAPRC-99 lumped groups indicated above.

Carbon Bond 4 (CB4). Although somewhat out of date, the CB4 mechanism developed by Gery et al (1988, 1889) continues to be widely used because of its computational efficiency and the fact that many emissions databases are already processed for this mechanism. Unlike the mechanisms discussed above, this uses a "lumped structure" approach where different model species are used to represent different parts of the molecule as if they reacted independently. For this reason, the emissions groups used for lumped SAPRC-99 and RADM-2 assignments could not be used. Assignments for this mechanism were based primarily on those used for the Texas emissions database (Yarwood, 2002), with most other CB4 assignments from databases received previously from the EPA and other sources. However, a number of new CB4 assignments had to be made for this project, and in those cases we attempted to be consistent with the methodologies used in the existing databases. Because existing emissions speciation databases have reasonably comprehensive assignments for CB4, we were able to make CB4 assignments for 94% of the compounds in the database. This accounts for all but around 0.003% of the volatile mass in Speciate 3.2 profiles with each profile weighed equally and in a profile representing emissions totals into a modeling domain (EPA, 1998).

Unfortunately, the most appropriate CB4 assignment is not always obvious, and different groups of compounds that were added at different times or represented in different databases sometimes employ different methods. An example is high molecular weight compounds such as alkanes that a detailed mechanism would predict have decreasing molar reactivity with the size of the molecule (Carter, 2000a), but are represented by increasing numbers of reactive PAR species in the CB4 mechanism. In some cases the carbons beyond a certain point are not represented as reactive, and in others they are. Although the complete set of CB4 assignments should probably be re-assessed and made consistent, doing so was beyond the scope of this project. Instead, we attempted, as much as possible, to utilize assignments in existing databases, and use our judgments only when no assignments are available. Footnotes in the database indicate the basis or source of all the CB4 assignments.

Speciation Assignment Data Files

The current speciation and mechanism assignment data are incorporated in an Excel file that can be downloaded from our project web site (Carter, 2004). The file includes a "documentation" sheet that describes in detail the format of all the other sheets, and also a number of Excel macros that can be used to process the data or output ASCII files with the mechanism assignments, as discussed below. Separate sheets are used for compounds, mixtures, mixture assignment profiles, assignments of master database categories to SAROAD numbers and Speciate 3.2 symbols, emissions group assignments to SAPRC-99 detailed model species, SAPRC-99 lumped and RADM-2 model species assignments to emissions groups, etc.

Since the speciation processing programs, discussed in the following section, cannot take the data directly from the Excel file, macros were written to output the assignment data in various ASCII files that the programs can read. These include the following: (1) Files giving assignments of SAROAD numbers or other chemical categories used in current emission databases to master chemical categories used in this database. There is one such file for each source of profiles used in the present database, as discussed in the following section. (2) A file giving the distribution of compounds for all the master chemical categories for which compositions have been assigned. (3) A file giving the molecular weights and descriptions for all the chemical compounds in the database. (4) Files giving the assignments of chemical compounds to model species for each mechanisms for which assignments were made. As indicated above, these currently include detailed and lumped SAPRC-99, RADM, and Carbon Bond 4. (5) Files giving the carbon numbers and molecular weights for the model species in the lumped mechanisms.

Note that these ASCII files are not intended to be the major repository of the data, nor are they updated directly. New assignments are made by updating the Excel file and re-running the macro to create these files.

Speciation Profile Database

As part of this project we also compiled the speciation profiles from various databases for use with this system. These included the following:

Speciate 3.2. Speciate is EPA's repository of Total Organic Compound and Particulate Matter speciated profiles for a variety of sources for use in source apportionment studies (EPA, 2002), but for this project only the organic profiles were used. The latest version available, 3.2 (dated November, 2002), was used for this project. The data were made available to us in a database file provided by Ron Ryan of the EPA. This database has a total of 567 profiles, numbered “0000” through “8539”. To uniquely identify these in our database (because other databases use similar numeric designations), these profiles were identified as “S32-0000” through “S32-8539”. In addition to compositions in terms of Speciate chemical categories, the database had descriptions and reference and other documentation information for most of the categories (EPA, 2002). This documentation was also incorporated into our database.

EPA. Profiles used for emission processing for Models-3 by the EPA were obtained from Gipson (2001). These used a categorization based on SAROAD numbers, which were assigned to the categories in our database based on the descriptions of the categories as discussed above. A total of 392 profiles, numbered “0000” through “9047” and identified as “EPA-0000”, etc., were obtained from this database. There was no additional documentation information available with these profiles other than the description line. The profiles through 8400 were essentially the same as the Speciate 3.1 profiles with the same number, but the 43 other profiles were different. Those profiles that are considered to be the same as Speciate 3.2 profiles were flagged as such.

California. The California Air Resources Board (CARB) has a largely independent effort for maintaining and updating speciation profiles, and also uses a SAROAD categorization

system that is incompatible with that used by the EPA. The California speciation profile data are available on the CARB web site (CARB, 2003), and these data, dated 3/19/2003, were downloaded and used for this program. A total of 373 profiles, numbered "0012" through "1949" (designated "CA-0012", etc.) were in this dataset, of which only 28 were the same as Speciate profiles. All these profiles had a description and also a reference citation or note, which were incorporated into the database.

Texas. Yarwood (2002) provided speciation profile data used in the Texas emissions databases maintained by the Texas Natural Resource Conservation Commission (TNRCC). These consisted of 377 profiles, numbered "0000" through "9047" and also "D201", "H2K2S", "H2KGS", "H2KGV", and "H2KWT". These were designated with the prefix "TX-" in our database. However, the numbered profiles turned out to be essentially the same as the EPA profiles with the same number, and thus the only unique profiles were the "D201" and the "H2K" profiles. Unfortunately, no description or other documentation information was made available with these profiles. We are aware that the TNRCC has developed some updated profile for modeling in Texas (Cantu, personal communication, 2003), but these have not yet been incorporated into our database.

Categorization Assignment Profiles. A total of 30 composition profiles were developed in this work for assigning compounds to complex mixture SAROAD or Speciate categories. These are included in the profile database for reference purposes, along with associated descriptions and notes indicating how the profiles were derived.

Other Profiles. The other profiles incorporated in this database include two that were used to designate base case reactive organic gas (ROG) emissions in previous work (Carter, 1994, 2000b), a special profile derived as a proposed means to represent unknown VOC mass, and another proposed as a means to represent compounds with no model species assignments.

These profile data are incorporated in a separate Excel file that can also be downloaded from our project web site (Carter, 2004). In order that their original categorization information be retained, the data are stored using the native chemical categorization systems of the database from which they were taken. A macro in the Excel file outputs the composition data and available documentation information into separate ASCII files that can serve as inputs to the speciation processing programs, discussed below, to derive the assignments of compounds or model species to these profiles.

SPECIATION PROCESSING PROGRAMS

The chemical mechanism assignments for most emissions processing systems such as SMOKE (MCNC, 2000) are generally incorporated in a single file that assigns model species directly to profiles. (This is called the GSPRO file in the SMOKE system.) At the present time the publicly available emissions software or modeling software does not include programs that will produce new profile-to-model species assignment files that would be needed if the profiles, speciation assignments, or chemical mechanisms are changed. Without such software, the advantages of the updated speciation assignments discussed above are not available to the

modeler, and changes to the chemical mechanism or incorporating updated profile data used by models is much more difficult.

In order to address this problem, we are developing a series of FORTRAN programs that can be used to implement these speciation assignments into the SMOKE (MCNC, 2000) emissions processing system for use with the CMAQ model. In addition to utilizing the speciation assignments and profile database discussed to produce profile-to-model species assignments that can be incorporated directly into SMOKE, they also include programs that permit the use of the adjustable parameter and adjustable lumping version of the detailed SAPRC-99 mechanism (Carter, 2000a,b 20004). The programs that currently have been written are referred to as the Profile Preprocessor (ProfPro), the Speciation Preprocessor (SpecPro), the Emissions Summary Processor (EmitSum), and the Mechanism Processor (MechPro). The relationships between these programs and relevant programs and files in the SMOKE and the CMAQ systems are shown in Figure 1. An overview of these programs is given below.

The Profile Preprocessor (ProfPro) program converts profile composition data with the native chemical classes used in existing profile databases to compositions in terms of specific chemical compounds. Mixture categories are resolved into compounds using the mixture composition assignments in the speciation database. The compounds are identified in the output files using the master speciation database designations, which in most cases are based on the compounds' CAS numbers. This program must be run to process profile data for input to the other programs, discussed below, but need be run only once for each profile. Its input and output files are mechanism-independent.

The Speciation Processer (SpecPro) program is used to produce the chemical mechanism assignments for a selected set of profiles and a selected chemical mechanism. The mechanism assignments can be output in several formats, but for use with SMOKE the principal function is the GSPRO file that SMOKE uses to make model species assignments. The program can also output a file giving the VOC to TOG conversion factors based on the fraction of compounds in the profile that are on the EPA exempt list, using the information in our compound database. The profile composition data are input in terms of distributions of individual compounds, as produced by ProfPro. The program provides several options for representing unknown or unassigned mass, including assuming it is unreactive, assuming it has the same composition of the known or assigned portions of the profile, or using separate profiles to represent unassigned mass or unknown compounds. (Two separate profiles are provided to demonstrate this capability.) The assignments of model species to the compounds are taken from the defaults for the given with the speciation database, but the user has the option of over-riding these assignments if it is desired to have selected compounds or groups of compounds represented by different lumped or explicit model species. This capability is useful for processing emissions for toxics modeling and reactivity assessment, where the compounds of particular interest are generally represented separately in the model.

The Total Emissions Profile Summation (EmitSum) program is used to produce composition file giving distribution of chemical compounds emitted into the model domain, given the total emissions amounts for all the profiles in the domain. The input file giving total emissions of profiles can produced from the SMOKE report files, though the program to perform

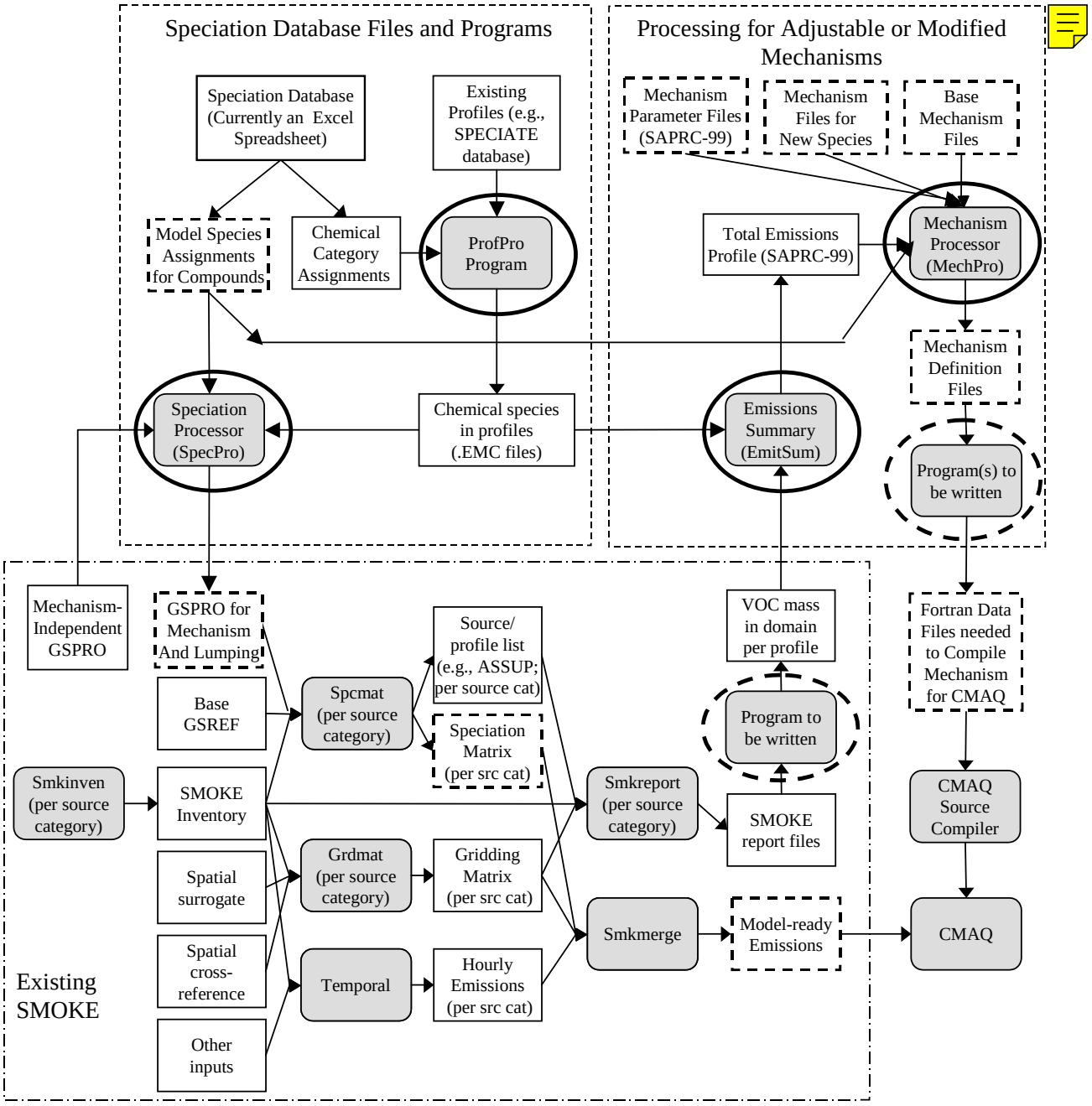

Figure 1. Diagram showing relationships between the speciation database programs and files and the SMOKE emissions processing systems. Mechanism-dependent files are indicated by dashed borders. The programs discussed here are indicated with bold ovals, with dashed ovals showing programs that need to be written to complete the system.

this specific function for SMOKE still needs to be written. The emissions summary profile output by EmitSum is needed by the mechanism processor program when processing mechanisms, such as SAPRC-99, where parameters for certain model species depend on the distribution of compounds emitted into the modeling domain. This program is not needed when processing for "fixed parameter" mechanisms such as CB4 or RADM or the fixed parameter version of SAPRC-99. However, a total emissions summation profile may be useful for other purposes, such as assessing relative contributions of various compounds to emissions, deriving “base case” mixtures for reactivity or other analyses, etc.

The Mechanism Processor (MechPro) program is used to prepare mechanism files (giving reactions and rate constants or kinetic parameters) when processing for adjustable SAPRC-99 or other mechanisms where the parameters for some model species depend on the distribution of compounds they represent. It can also be used to prepare mechanisms where new model species have been added to represent selected compounds. In the case of SAPRC-99 the program can derive the reactions of the new or adjustable parameter model species from mechanistic parameter files. For other mechanisms, the reactions of the new species must be given in input files. The mechanisms for the new or adjustable species are merged with the base mechanism (whose reactions are input in a separate file), and used to create complete mechanism file specifying the reactions and kinetic parameters in the format used by the CMAQ mechanism compilation system, or optionally in the format used by the SAPRC programs (Carter, 2001). Although the output files contain all the information needed to integrate the mechanism into a model, a separate program or series of programs, not yet completed, is needed to convert the mechanism information into the specific files needed to compile a mechanism into CMAQ.

Further information and preliminary documentation for these programs is available at the project web site (Carter, 2004). These programs have been tested and run on a PC platform using but have not yet been ported to Unix systems where the SMOKE software is implemented, nor have all their features been completely debugged. Current versions of the program, with test calculations processing SAPRC-99 and CB4 mechanisms with adjustable parameter or additional model species, are available on request (see also Carter, 2004). Work on completing the implementation of these programs into the SMOKE system, in collaboration with researchers at the Carolina Environmental Programs at the University of North Carolina, is underway.

DISCUSSION AND CONCLUSIONS

Although we believe this project represents an important step forward in rationalizing the chemical compound and model species assignments for speciation profiles in emissions databases, a number of issues need to be resolved and additional work is needed before this objective is fully realized. These are summarized below.

- The integration of the software and data files developed in this project into the SMOKE emissions processing needs to be completed, debugged, documented, and made publicly available. It should also be made available for use with other emissions processing systems besides SMOKE.

- The assignments incorporated in the current database need to undergo independent peer review. The assignment of compounds to the various mixture categories used in current profiles database was in many cases based on highly arbitrary assumptions or guesses that need independent review and improvement. In addition, although making frequent cross-checks was an important part of this effort, almost certainly some mistakes were made when going through the many thousands of categories and profiles.
- The judgments as to which compounds are represented as non-volatile also need to be reviewed, preferably based on knowledge or estimates of their actual vapor pressures. The current estimates are subjective and often arbitrary judgments.
- If the CB4 mechanism is to continue to be used, its assignments need to be reviewed and made to reflect consistent assignment criteria. It may be appropriate as part of this effort to revise and update these criteria to better reflect current mechanistic knowledge.
- Thought needs to be given on how best to represent unknown mass in profiles. The common practice of ignoring it or assuming it is unreactive is probably unacceptable. The best approach is probably to use profiles to represent what our best guess is what may be in this unknown mass. Included in our profile database is proposed "unknown mass" profile that was derived from n_C>6, Mwt>120 compounds in Speciate all-profile average profile. This can serve as a starting point in this regard, but needs to improved.
- Thought needs to be given on how best to represent unassigned compounds in profiles, since ignoring them or assuming they are unreactive is also unacceptable. Although the database includes a profile provided this purpose, based on the molecule-weighed average of compounds represented by SAPRC-99 mechanism in each SAPRC-99 chemical class, almost certainly this could be improved. Ultimately assigning all the compounds to model species would be the goal, but presently this is not feasible.
- Improving the categorization and mechanism assignment and implementation process as discussed above does not reduce the need to continue research to improve the quality of existing profiles. Non-volatile compounds and poorly defined categories need to be removed, and new profiles that include such categories should not be accepted.

The ultimate success of this project requires the widespread adoption and use of this database (or one developed from it) when emissions profiles are developed and updated in the future. This will require development of standard procedures to update and add to the database as needed, and central maintenance of the database by an appropriate and recognized authority. Otherwise, the database will either not be used or will soon become outdated and something resembling the current disorganized and inconsistent system will evolve again.

ACKNOWLEDGEMENTS

This work was funded by in part by the University of Houston for a model development project directed by Dr. Anthony Haymet, and in part by a subcontract to the Carolina Environmental Programs at the University of North Carolina (UNC) for a project funded by the American Chemistry Council (ACC). The collaborators on the ACC/UNC project, who are

responsible for the implementation of the database and programs discussed here into SMOKE are Marc Houyoux, now at the EPA, and Jeff Vukovich of UNC. The author thanks Marc Houyoux for helpful discussions concerning the design and implementation of the programs, Ron Ryan of the EPA for providing the Speciate 3.2 data in a useable format, Deborah Luecken and Gerry Gipson of the EPA for providing the EPA's profile data, Greg Yarwood of Environ for providing the Texas profile data and helpful discussions and input into the design of the database, Paul Allen of the California ARB for helpful discussions regarding the CARB emissions database, and Chao-Jung Chien of our laboratories for ongoing assistance in implementing the mechanism processing output into CMAQ.

REFERENCES

CARB (2003): Organic Gas Speciation Profiles provided by the California Air Resources Board at http://arbis.arb.ca.gov/emisinv/speciate/speciate.htm. Downloaded file named ORGPROF_03_19_03.xls and dated 3/19/2003.

Carter, W. P. L., D. Luo, and I. L. Malkina (1999): 'Investigation of the Atmospheric Impacts and Ozone Formation Potential of Styrene," Report to the Styrene Information and Research Center. March 10.

Carter, W. P. L. (2000a): "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://cert.ucr.edu/~carter/absts.htm#saprc99 and http://www.cert.ucr.edu/~carter/ reactdat.htm.

Carter, W. P. L. (2000b): "Implementation of the SAPRC-99 Chemical Mechanism into the Models-3 Framework," Report to the United States Environmental Protection Agency, January 29. Available at http://www.cert.ucr.edu/~carter/absts.htm#s99mod3.

Carter, W. P. L. (2001): "SAPRC-99 Mechanism Files and Associated Programs and Examples," http://www.cert.ucr.edu/~carter/SAPRC99.htm.

Carter, W. P. L. (2004): "Current Project Information Page: Development of an Improved Chemical Speciation Database for Processing Emissions of Volatile Organic Compounds for Air Quality Models," http://www.cert.ucr.edu/~carter/emitdb.

EPA (1998): Emissions profile received by email from Deborah Luecken of the EPA on Dec 31, 1998. The email message indicated that they contain SAROAD emission totals for a 36km evaluation grid, are based on the 1995-NET EPA US inventory, and contain the emission summaries for the US part of the grid.

MCNC (2000): "Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System," Available at http://www.emc.mcnc.org/products/smoke/.

Middleton, P., W. R. Stockwell, and W. P. L. Carter (1990): “Aggregation and Analysis of Volatile Organic Compound Emissions for Regional Modeling,” Atmos. Environ., 24A, 1107-1133.

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.

Gipson (2001). PROFILE.VOC.DAT file, used for processing VOC emissions data for Models-3. Received from Gerald L. Gipson, EPA, Research Triangle Park, NC,. November, 2001

Gery, M. W., G. Z. Whitten, J. P. Killus, and M. C. Dodge (1989): “A Photochemical Mechanism for Urban and Regional Scale Computer Modeling,”, J. Geophys. Res, 94, 12,925.

MCNC (2000): "Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System," Available at http://www.emc.mcnc.org/products/smoke/.

Middleton, P., W. R. Stockwell, and W. P. L. Carter (1990): “Aggregation and Analysis of Volatile Organic Compound Emissions for Regional Modeling,” Atmos. Environ., 24A, 1107-1133.

EPA (2002). Speciate 3.2 database available at http://www.epa.gov/ttn/chief/software/ speciate as of early 2004. Dated November 3, 2002.

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.

Yarwood (2002). Texas Natural Resource Conservation Commission emissions species data file "compound.database.eps2x.17May02", provided by Greg Yarwood, Environ corporation, Novato, CA, Latest version provided June 18, 2002

KEY WORDS

- Emissions Speciation Database
- Emissions Speciation Profiles
- TOG Speciation
- VOC Speciation
- Mixture Compositions
- Chemical Mechanisms
- Emissions Processing for Models
- Mechanism implementation in Models
- SMOKE
- SAPRC-99 Mechanism
- RADM-2 Mechanism
- Carbon Bond 4 Mechanism
- SAROAD
- SPECIATE