The Use of SMOKE to Process Multi-pollutant Inventories - Integration of Hazardous Air Pollutant And Volatile Organic Compound Emissions

Madeleine Strum and Laurel Driver
Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency, Research Triangle Park, NC 27711
strum.madeleine@epa.gov

Gerald Gipson
National Exposure and Risk Laboratory, Office of Research and Development
U.S. Environmental Protection Agency, Research Triangle Park, NC 27711

William G. Benjey[*]
Atmospheric Sciences Modeling Division
Air Resources Laboratory
National Oceanic and Atmospheric Administration
Research Triangle Park, NC 27711

Rich Cook
Office of Transporation and Air Quality
U.S. Environmental Protection Agency
Ann Arbor, MI 48103

Marc Houyoux and Catherine Seppanen
Carolina Environmental Program, University of North Carolina at Chapel Hill,
c/o Environmental Modeling Center, 3021 Cornwallis Road, RTP, NC 27709

Greg Stella
Alpine Geophysics, LLC
Arvada, CO 80007

ABSTRACT

The Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system is being updated to process multi-pollutant (criteria and toxics) emissions. SMOKE will be able to import both criteria and toxic pollutant inventories, to combine them, and to prepare the model-ready emission inputs for a variety of air quality models selected by the user. These functions will allow SMOKE to support air quality modeling of criteria and toxic pollutants with the Community Multiscale Air Quality (CMAQ) model and the Regulatory Modeling System for Aerosols and Deposition (REMSAD). The recent release of SMOKE version 1.5 Beta (hereafter referred to as SMOKE 1.5b) has come a long way toward meeting these goals.

This paper describes the new functions of the SMOKE 1.5b release pertaining to the combination of criteria and toxics inventories for mobile sources only (there are additional capabilities but they will not be described here). Numerous toxics that are on EPA's list of hazardous air pollutants (HAPs) are also volatile organic compounds (VOC) or particulate matter (PM). The focus of this paper is the overlap in HAPs that are also VOC. This paper will discuss the design of the options provided in

---

[*] On assignment to the National Exposure Research Laboratory, U.S. Environmental Protection Agency

SMOKE that will allow the use of mobile source emissions from the National Emission Inventory (NEI) for HAPs and the NEI for criteria pollutants to develop the model species needed for ozone and PM modeling, as well as for toxics modeling. These options provide for optimal use of the toxics inventory data in situations in which inventory estimation methods are consistent for both VOC and HAPs; for example, developing on-road emissions data using the MOBILE6.2 emissions model. For inventories that have questionable consistency between the NEI HAP and VOC, the user may choose not to combine the HAP and VOC estimates; SMOKE will allow for this option as well.

The paper also contains preliminary results from SMOKE test cases that show SMOKE performance and the impact on model species emission estimates of integrating the on-road HAP and VOC emissions generated using MOBILE6.2 in SMOKE.

INTRODUCTION

The Environmental Protection Agency (EPA)'s Office of Air Quality Planning and Standards (EPA/OAQPS) generally uses separate inventories and air quality models to estimate ozone and PM versus toxic air pollutant ambient air concentrations. Great strides have been made to integrate the criteria and toxics inventories for the 1999 inventory year, in particular with regard to mobile source emissions. Additional efforts to integrate the inventories are expected for the 2002 inventory year. Nonetheless, at the present time, differences in these inventories, and differences in the model ready emissions required for the various air quality models, have necessitated the development of separate emission modeling systems.

For national and regional scale ozone and PM modeling applications, EPA/OAQPS is currently using the NEI for Criteria Pollutants (http://www.epa.gov/ttn/chief/net/index.html), formerly called the National Emission Trends inventory, with Eulerian grid models such as the Community Multiscale Air Quality (CMAQ) model[1] and Regulatory Modeling System for Aerosols and Deposition (REMSAD)[2]. EPA uses the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system[3] to prepare the model-ready inputs for these and other grid models. The general inventory development and emissions modeling approach have been described earlier.[4]

For toxics modeling applications, EPA/OAQPS is currently using the NEI for HAPs (http://www.epa.gov/ttn/chief/net/index.html), formerly called National Toxics Inventory, with Gaussian-based models such as the Assessment System for Population Exposure Nationwide (ASPEN)[5] for national-scale and the Industrial Source Complex Short Term model (ISCST3)[6] for urban-scale applications. The EPA has been using a separate emission processing tool, the Emission Modeling System for Hazardous Air Pollutants (EMS-HAP)[7] to process the HAPs from the NEI for HAPs and a diesel PM inventory derived from the PM in the NEI for criteria pollutants to prepare the model-ready inputs for these models.

Two limitations of the EMS-HAP/ASPEN modeling system (that also apply to the EMS-HAP/ISCST3 modeling system) that were emphasized during the Science Advisory Board review of the National-Scale Air Toxics Assessment for 1996 are the air quality model's inability to model transport of any pollutant beyond the 50 kilometers from its original emission point and its inability to simulate non-linear chemistry effects.[8]

For this reason, and for EPA to the advance tools to allow for the combined chemistry one-atmosphere modeling approach, EPA has embarked on a project to modify SMOKE to add the capability to process multi-pollutant inventories for grid models. The project also includes the incorporation of the MOBILE6 model for highway emissions into SMOKE, which includes the final version of MOBILE6.0 and draft versions of MOBILE6.1 and MOBILE6.2. The incorporation of MOBILE6 was previously reported, but was not included, in a public website release until the SMOKE 1.5 Alpha (February 2003) and Beta (March 2003) releases.[9]

On March 14, 2003, SMOKE 1.5b was released on the Community Modeling and Analysis System (CMAS) Clearinghouse website (http://www.cmascenter.org/modelclear.html). This update contains several new functions and enhancements. This paper addresses the new functions of SMOKE

1.5b associated with the processing of multi-pollutant inventories for on-road and nonroad mobile source emissions, for the CMAQ model. In conjunction with the SMOKE changes, enhancements are being made to CMAQ which allow several gaseous toxics to be modeled.[10] Future work on SMOKE will result in updates in order to address stationary source toxics and criteria emissions and the REMSAD air quality model.

MAJOR NEW FUNCTIONS

In this paper, we will discuss the design for the new functions of SMOKE 1.5b that will process on-road and nonroad toxics and criteria inventories for the CMAQ model. These functions allow SMOKE to: 1) import criteria and toxics mobile inventories, and, 2) combine the two inventories such that toxics emissions can be used to replace VOC emissions to model ozone PM with the carbon bond 4 (CB4) mechanism or to model toxics pollutants using a toxics CB4 mechanism with CMAQ.

Importing of Criteria and Toxics Inventories

While the SMOKE model can process any pollutant, it has not been able to handle the multitude of HAPs in the NEI. The NEI for HAPs contains approximately 400 distinct pollutants that represent the 188 HAPs from Section 112(b) of the Clean Air Act (see http://www.epa.gov/ttn/atw/188polls.html for the list). In addition to memory and efficiency issues involved in processing such a large number of pollutants, there is an issue of how to design a system to use both sets of emissions to support the "one atmosphere" concept whereby all pollutants (toxics and criteria) can be modeled with the appropriate chemistry. The latter issue will be addressed in the next subsection of the paper.

Significant changes have been made to SMOKE's inventory import (Smkinven) routines. SMOKE 1.5b reads two separate inventories: a criteria inventory which can be input into SMOKE in the same formats SMOKE currently supports (i.e., inventory data analyzer [IDA]) and a toxics inventory which utilizes a new format called the SMOKE toxics format. The latter format requires that each pollutant be in a separate record (line) of the file. Table 1 shows the format.

Table 1. Nonroad mobile toxic inventory format

Column	Column name	Type	Description
A	State FIPS code	Integer	Two digit FIPS code for state (leading zeros not needed)
B	County FIPS code	Integer	Three digit FIPS code for county (leading zeros not needed)
C	SCC	10-character string	Ten character source classification category (SCC) code
D	CAS number	10-character string	Inventory Pollutant Code (usually a Chemical Abstract Systems [CAS] number)
F	Annual emissions value	Real	Emissions [tons/year]
G	Average-day emissions value	Real	Emissions [tons/average day] (optional)
H	Control Efficiency	Real	Control Efficiency (give value of 0-100; use –9 if missing -will result in SMOKE default of 0)
I	Rule Effectiveness	Real	Rule Effectiveness (give value of 0-100; use –9 if missing -will result in SMOKE default of 100)
J	Rule Penetration	Real	Rule Penetration (give value of 0-100; use –9 if missing - will result in SMOKE default of 100)
K	Comment	String of any length	Comment (optional, not read by SMOKE) Need to start with an "!" sign.

The 1999 NEI for HAPs contains 14.1 million inventory records for the entire nation. Such a large number of inventory records required a different memory handling approach during data import, which involves reading in and processing the inventory by groups (currently set to 1 million records at a time) and using temporary files.

New functions performed in the import stage are the (1) selection of pollutants, (2) aggregation/disaggregation of pollutants, and (3) the assignment of information needed for combining the toxics and criteria pollutants (discussed in the next section). The user controls how these functions are performed through the "Inventory Table" ancillary input file (SMOKE file name INVTABLE). This ancillary input file replaces the "SIPOLS" and "ACTVNAMS" ancillary input files that were used in previous versions of SMOKE to specify valid pollutants and activities in the input inventory file. The Inventory Table performs this in addition to the new functions. For purposes of aggregation/disaggregation, the Inventory Table file is analogous to EMS-HAP's "Haptable" file.

The selection function allows the user to subset the input inventory to only those particular pollutants to be sent through the SMOKE system for modeling. Aggregation of several pollutants to one pollutant is accomplished by having several inventory pollutant code (CAS) numbers assigned to the same inventory pollutant name. For example, the various compounds of lead reported in the inventory can be aggregated into "lead." Disaggregation is accomplished by including split factors in the Inventory Table, and assigning the same inventory pollutant code (CAS) number to multiple pollutant names. For example, mercury from on-road sources can be disaggregated (or speciated) into the CMAQ model species of divalent, divalent particulate, and elemental mercury. The aggregation/dissaggregation function may also be done at the source category code (SCC)-level using SMOKE's speciation processing step. We chose to have this function at the import stage in addition to the speciation stage, because it allows for inventory-wide speciation, and permits creating unique pollutants that will simplify pollutant-specific reports of the different downstream steps. Upfront aggregation also reduces the potentially large number of the toxics pollutants fed into the other SMOKE programs. Finally, the assignment of information needed for combining the toxics and criteria pollutants allows user control over how SMOKE will combine toxics and criteria inventories without duplicating mass, which is discussed in the next section.

The development of wrapper subroutines for the Input/Output Applications Programming Interface (I/O API), which is used in data storage and access, is another significant development included with SMOKE 1.5b. We developed these tools to permit the reading and handling of the multitude of pollutants in the inventory. These wrappers will support more than 120 variables by using the standard I/O API library and managing multiple files (called a file set). The wrappers may also be useful for CMAQ.

Combination of Toxics and Criteria Inventories

The major design issue addressed in SMOKE 1.5b is the combination of toxics and criteria inventories without duplicating the mass of VOC (i.e., benzene, formaldehyde, acetaldehyde). While the SMOKE code is general and can handle many chemical mechanisms, the input files that describe the mechanism must be specific for each mechanism. To date, we have created mechanism-specific files for two mechanisms that combine toxics and criteria inventories: (1) a CB4 mechanism and (2) a modified CB4 mechanism, "toxics CB4." The toxics CB4 allows concentrations for particular explicit VOC toxics to be generated. The toxics CB4 mechanism is similar to the original CB4 except that additional model species are added to allow for explicit treatment of particular toxic pollutants within the chemical mechanism. These explicitly treated toxics are formaldehyde, acetaldehyde, 1,3 butadiene and acrolein. They were chosen based on the importance of these to the ozone chemistry. They were also chosen based on the effort it would take to revise CB4 to solve for the other, non-explicit pollutants, such as benzene, outside the mechanism using pseudo-first order reaction rates. These other model species are referred to as "outside the mechanism," whereas the explicit toxics are "inside the mechanism." [10] The distinction of whether the particular toxic is "inside" or "outside" the chemical mechanism is important because it affects the way the VOC and toxic pollutants are speciated in SMOKE. Table 2 shows a list of the mechanism species currently in the toxics CB4 mechanism. Note that in toxics CB4, FORM from CB4 is split into FORM (explicit formaldehyde) and FORMSURR; ALD2 from CB4 is split into ALD2 and ALD2SURR. The toxics mechanism species were selected such that all of the gaseous HAPs (or toxic VOCs) on the list of the urban 33 HAPs from the integrated air toxics urban strategy (http://www.epa.gov/ttn/atw/urban/urbanpg.html) could be estimated. New outside-the-mechanism species may be added as additional rate constants are determined.

Table 2. Toxics CB4 mechanism species

NAME	Inside or Outside Toxic Mechanism Species	Comments
PAR		Same as in CB4
TOL		Same as in CB4
XYL		Same as in CB4
OLE		Same as in CB4
FORM	inside	Contains only formaldehyde
FORMSURR		Non formaldehyde part of CB4 FORM
ALD2	inside	Contains only acetaldehyde and those compounds that explicitly form acetaldehyde rapidly such as t-2-butene, c-2-butene,t-2-pentene.
ALD2SURR		Non-acetaldehyde part of CB4 ALD2. Includes higher-level aldehydes
BUTADIENE	inside	1,3 butadiene
ACROLEIN	inside	acrolein
BENZENE	outside	benzene
ACRYLONITRILE	outside	acrylonitrile
CARBON_TET	outside	carbon tetrachloride
CHLOROFORM	outside	chloroform
ETHYLENE_DIBROMI	outside	ethylene dibromide
PROPYLENE_DICHL	outside	propylene dichloride
ETHYLENE_DICHLOR	outside	ethylene dichloride
ETHYLENE_OXIDE	outside	Ethylene oxide
MECL	outside	methylene chloride
PERC	outside	perchloroethylene
TRICHLOROETHYLEN	outside	trichloroethylene
VINYL_CHLORIDE	outside	vinyl chloride
DICHLOROPROPENE1	outside	1, 3-dichloropropene
TETCHLETHANE1122	outside	1, 1, 2, 2-tetrachloroethane

There are two approaches that SMOKE uses to combine toxics and criteria inventories to eliminate duplicate VOC mass. We denote the approaches as the "integrate" and "no-integrate" cases. For the "integrate" case, for each source, the VOC provides the non-HAP part of the mass, the HAP inventory provides the HAP-part of the VOC mass, and the total VOC mass is conserved (neither increased nor decreased). The HAPs and non-HAP part of the VOC are mapped to the appropriate model species in the chemical mechanism (i.e., either CB4 or toxics CB4 mechanism). This allows more accurate model species estimates to be made because the HAPs may be generated with a more detailed methodology than is done using the traditional speciation of VOC. This is particularly true for some of the toxics automatically output by MOBILE6.2 such as benzene[†]. For the "no-integrate" case, the VOC HAP will not be used except in the toxics CB4 mechanism, where it will be used to provide the

---

[†] MOBILE6.2 estimates some benzene, 1,3 butadiene, acrolein, formaldehyde, acetaldehyde and methyl tert-butyl ether (MTBE) explicitly, while others are estimated based on user-provided information on emission factors or toxic fractions of total organic gases (TOG,VOC, or PM.)

mass of the toxics model species only. In this situation, the criteria VOC will be speciated to provide only those model species that are not explicit toxics. This case is needed when it is impractical or invalid to integrate the inventories, as is discussed later.

The Integrate Case

A schematic of the "integrate" case is depicted in Figure 1. For both mechanisms, SMOKE computes emissions for a new pollutant group: NONHAPVOC, which contains that portion of the VOC that are not toxics VOC. SMOKE computes NONHAPVOC for each source by subtracting the sum of the toxics VOC from the VOC emissions on a source-by-source basis. SMOKE can also create NONHAPTOG (which can occur if the user provides TOG, total organic gas, emissions instead of VOC emissions or SMOKE runs the MOBILE6.2 model to create TOG emissions), but since we intend to use VOC in the input inventory and compute VOC in MOBILE6 along with the various toxics, we will refer to only NONHAPVOC in the remainder of the paper. The user must identify in the Inventory Table the particular toxics VOC to subtract, and these toxics must be consistent across an inventory (i.e., the same set of toxics VOC must be used for all nonroad sources for a particular SMOKE run).

Figure 1. Integrate Case in SMOKE

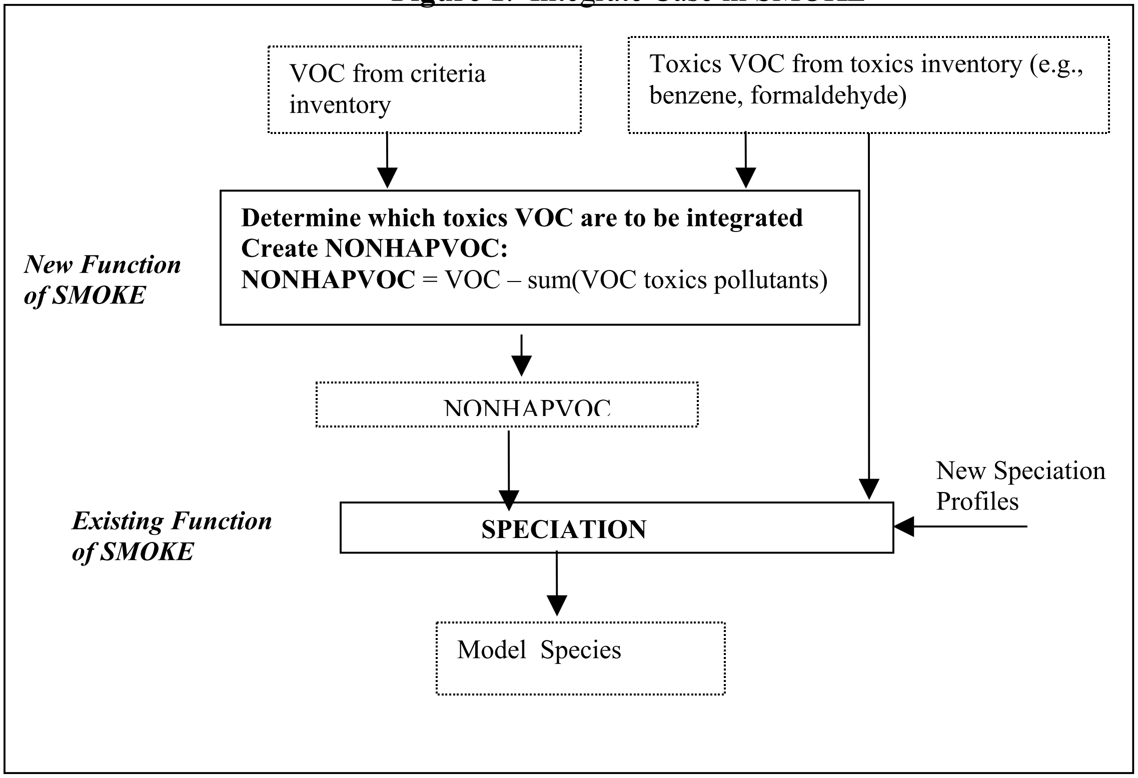


Table 3 lists the HAP that we subtracted. We chose these because these are the HAPs that EPA generated emission estimates for in the the on-road and nonroad 1999 NEI for HAPs.[11, 12] Other toxics VOC that may be included in the on-road or nonroad 1999 NEI for HAPs (e.g., on the basis of a State submittal) would not be in Table 3. The Inventory Table (INVTABLE) files and General Speciation Profile (GSPRO) files provided with SMOKE 1.5b were designed based on the subtraction of these particular HAPs from the VOC from the criteria inventory

Table 3. Toxics VOC Subtracted from VOC for the Integrate Case

ON-ROAD: Toxics VOC to subtract from criteria VOC		**NONROAD:** Toxics VOC to subtract from criteria VOC	
CAS No.	Pollutant Name	CAS No	Pollutant Name
71432	benzene	71432	benzene

50000	formaldehyde	50000	formaldehyde
1634044	MTBE (methyl tertiary butyl ether)	1634044	MTBE (methyl tertiary butyl ether)
106990	1,3 butadiene	106990	1,3 butadiene
107028	acrolein	107028	acrolein
75070	acetaldehyde	75070	acetaldehyde
100414	ethylbenzene	100414	ethylbenzene
110543	hexane	110543	hexane
100425	styrene	100425	styrene
108883	toluene	108883	toluene
1330207	xylenes.	1330207	xylenes.
540841	2,2,4 trimethylpentane (iso octane)	540841	2,2,4 trimethylpentane (iso octane)
123386	propionaldehyde	123386	propionaldehyde
191242	Benzo[g,h,I]perylene	191242	benzo[g,h,I]perylene
83329	Acenaphthene	83329	acenaphthene
208968	Acenaphthylene	208968	acenaphthylene
56553	Benz-a-anthracene	56553	benz-a-anthracene
205992	Benzo[b] fluoranthene	205992	benzo[b]fluoranthene
207089	Benzo[k]fluoranthene	207089	benzo[k]fluoranthene
218019	Chrysene	218019	chrysene
53703	Dibenz[a,h]anthracene	53703	dibenz[a,h]anthracene
193395	Indeno[1,2,3-cd]pyrene	193395	indeno[1,2,3-cd]pyrene
120127	Anthracene	120127	anthracene
86737	Fluorene	86737	fluorene
91203	Naphthalene	91203	naphthalene
85018	Phenanthrene	85018	phenanthrene
		108952	phenol
		78933	MEK (methyl ethyl ketone)
		108907	chlorobenzene
		67561	methanol

Different lists are needed for on-road versus nonroad source categories because additional HAPs are computed in the NEI for HAPs for some nonroad source categories. Although we could have added the additional nonroad HAPs to the on-road list (there are no emissions to subtract for these), we decided to keep the lists separate because some of the HAPs that are not computed by EPA for on-road sources (e.g., MEK) may well be emitted, and we do not want to adjust the speciation profiles to exclude these. Another issue with the above table is the presence of 13 of the 16 PAHs estimated from mobile sources. Note that the PAHs are often emitted in both particulate and gaseous phases. We have made a simplifying assumption to treat all PAHs as toxics VOC except for pyrene, benzo(a)pyrene and fluoranthene.

The next step of the integrate case is chemical speciation. SMOKE speciates NONHAPVOC or NONHAPTOG and maps the toxics VOC from the toxics inventory into the appropriate model species. Different speciation profiles are needed for the CB4 and toxics CB4 cases because they have different model species. Using this approach, we needed to develop new speciation profiles for the new pollutants, NONHAPVOC and the individual toxics that result from the integration. We developed the GSPRO ancillary input files to SMOKE that contain speciation profiles to map the pollutants to model species for the CB4 and toxics CB4 mechanisms for the integrate case. These are included with SMOKE 1.5b. To develop the NONHAPVOC profiles, the VOC profiles from the SPECIATE 3.1 (http://www.epa.gov/ttnchie1/software/speciate/index.html) database were obtained, and the HAP VOCs that are listed in Table 3 were zeroed out in the appropriate (those pertaining to on-road and nonroad sources) profiles. The SPECIATE-based profiles were then renormalized and the individual pollutant species were aggregated into the model species to generate the GSPRO files. Note that SMOKE's integration approach is very general in that the user can define his/her own NONHAPVOC and use alternative chemical mechanisms. However, it is critical that the speciation profiles for NONHAPVOC be developed so that they are consistent with the list of HAPs that are subtracted from criteria VOC. For example, if benzene is a HAP that is subtracted from VOC to compute NONHAPVOC, then the NONHAPVOC speciation profiles should not include the mass of benzene.

Table 4 contains sample records from the GSPRO file included with SMOKE 1.5b for the CB4 mechanism. This file contains the factors used to separate aggregated (e.g. VOC or NONHAPVOC) or individual (e.g., benzene, toluene) inventory pollutant emissions totals into emissions of model species for the air quality model. SMOKE uses the "split factor" to split the inventory pollutant into its model species then divides by the "divisor" (which is sometimes a molecular weight) to compute the moles of model species per gram of pollutant. Note that in the eighth and ninth records in Table 4, benzene, which comes from the toxics inventory, is mapped to PAR (parrafins group) and NR (non-reactive). With the integrate case, VOC is not speciated for any sources that are integrated, only NONHAPVOC and the toxics VOC from the toxics inventory are used to compute model-species emissions.

Table 4. Sample GSPRO file records* For the CB4 Integrate Case

Profile Number	Pollutant Name	Model Species	Split Factor	Divisor
1097A	NONHAPVOC	ETH	0.011007047	1
1097A	NONHAPVOC	FORM	0.001110391	1
1097A	NONHAPVOC	NR	0.004786372	1
1097A	NONHAPVOC	OLE	0.003541792	1
1097A	NONHAPVOC	PAR	0.022544043	1
1097A	NONHAPVOC	TOL	7.25675E-05	1
1097A	NONHAPVOC	XYL	7.75354E-05	1
0000	BENZENE	PAR	1	78.11
0000	BENZENE	NR	5	78.11
0000	FORM	FORM	1	30.03
0000	BUTADIENE	OLE	2	54.09
0000	ACROLEIN	OLE	0.5	56.06
0000	ACROLEIN	ALD2	1	56.06

*The Mass fraction column is not reported; it is not used by SMOKE for the example explained in this paper

Table 5 contains sample records from the GSPRO file included with SMOKE 1.5b for the toxics CB4 mechanism.

Table 5. Sample GSPRO file records* For the Toxics CB4 Integrate Case

Number	Pollutant Name	Species	Split Factor	Divisor
0000	BENZENE	BENZENE	1	78.11
0000	BENZENE	PAR	1	78.11
0000	BENZENE	NR	5	78.11
0000	FORM	FORM	1	30.03
0001A	NONHAPVOC	PAR	0.057377025	1
1097A	NONHAPVOC	ETH	0.011007047	1
1097A	NONHAPVOC	FORMSUR	0.001110391	1
1097A	NONHAPVOC	NR	0.004786372	1
1097A	NONHAPVOC	OLE	0.003541792	1
1097A	NONHAPVOC	PAR	0.022544043	1
1097A	NONHAPVOC	TOL	7.25675E-05	1
1097A	NONHAPVOC	XYL	7.75354E-05	1

*The Mass fraction column is not reported; it is not used by SMOKE for the example explained in this paper

Table 5's example records show a number of the additional model species present in the toxics CB4 mechanism. Also note that benzene from the inventory is mapped to both the non-toxics (PAR) and toxics (BENZENE) mechanism species. While this appears to be double counting benzene mass (indeed SMOKE is taking benzene mass twice), the proper accounting of mass is done within the toxics CB4 mechanism in the air quality model. This is because the toxics CB4 approximates the chemical loss outside of the chemical mechanism. The benzene that gets mapped to PAR does not affect the overall mass of benzene other than its role in the chemical loss mechanism that will affect the reaction rate. Formaldehyde, however, is not double counted in SMOKE, because it is not mapped to two model species – a feature of all pollutants that are treated explicitly inside the chemical mechanism. Differences between the outside the chemical mechanism and inside the chemical mechanism toxics have played a role in the design for the no-integrate case, which is discussed in the next section.

As mentioned earlier, there are times when it is impractical or inappropriate to integrate the criteria and toxics inventories. Integration should only be done for sources in which VOC and HAP estimates were made on a consistent basis (e.g., same activity information) and when the source has all of the gaseous HAP estimates. This will always be true when SMOKE is driving the MOBILE6 model because MOBILE6 is internally consistent. It is also true for EPA-generated nonroad emission sources other than military aircraft emissions, which is not a category that EPA computes emissions for in the NEI for HAPs. When using the NEI for criteria and NEI for HAPs, it is not appropriate to integrate when a State agency replaces EPA-generated data for criteria and not HAPs or vice versa. SMOKE provides the flexibility to allow the user to exempt particular sources from the integrate approach through an ancillary NHAPEXCLUDE file supplied by the user. Those sources listed in the NHAPEXCLUDE file are then treated under the "No-Integrate" case as discussed below.

No-Integrate Case

The key premise for the no-integrate case is that the VOC inventory should be used to provide the chemical mechanism model species except in cases where the toxic pollutant itself is of interest for air quality modeling. Figure 2 depicts a schematic of the non-integrate case for the CB4 and toxics CB4 mechanisms. In both situations, SMOKE does not compute NONHAPVOC for those inventory sources that have been selected for no integration. For the CB4 mechanism, SMOKE will map the VOC to the standard CB4 species and drop the HAP VOC. The Inventory Table allows the user to tell SMOKE which pollutants are considered HAP VOC. Note that other toxics that are not HAP VOC (e.g., mercury) can still be processed through SMOKE; they will not be dropped. When using the CB4 mechanism, SMOKE can use the same CB4 speciation profiles as are currently used for ozone and PM modeling for speciation of the VOC for any source getting the no-integrate approach.

The no-integrate case for the toxics CB4 mechanism is more complicated because we have chosen to use the toxics inventory only for the toxics model species (e.g., BENZENE and FORM). In the no-integrate case, the VOC inventory is used for the model species that are not explicit toxics (e.g., FORMSURR and PAR). SMOKE will therefore drop only the toxic emissions for those VOC toxics that are not themselves model species. Thus, neither formaldehyde (a model species that is inside-the-mechanism) nor benzene (a model species that is outside-the-mechanism) will be dropped. In the speciation step, the toxics inventory pollutants that are toxics model species will be speciated (mapped) to themselves. The VOC will be speciated to the other (non-toxics) model species. To avoid duplication of mass, the VOC speciation profiles must be adjusted to remove those toxics species that are inside the chemical mechanism: formaldehyde, acetaldehyde, acrolein and butadiene. Otherwise, their mass would be double counted since the toxics inventory is providing their mass. It is not necessary to remove the outside-the-mechanism toxics, such as benzene, because the VOC is being used to generate the benzene mass that influences the chemistry (that is, the mass from benzene that goes to PAR as seen in Table 4.)

Figure 2. No-Integrate Cases for CB4 and Toxics CB4 Mechanisms

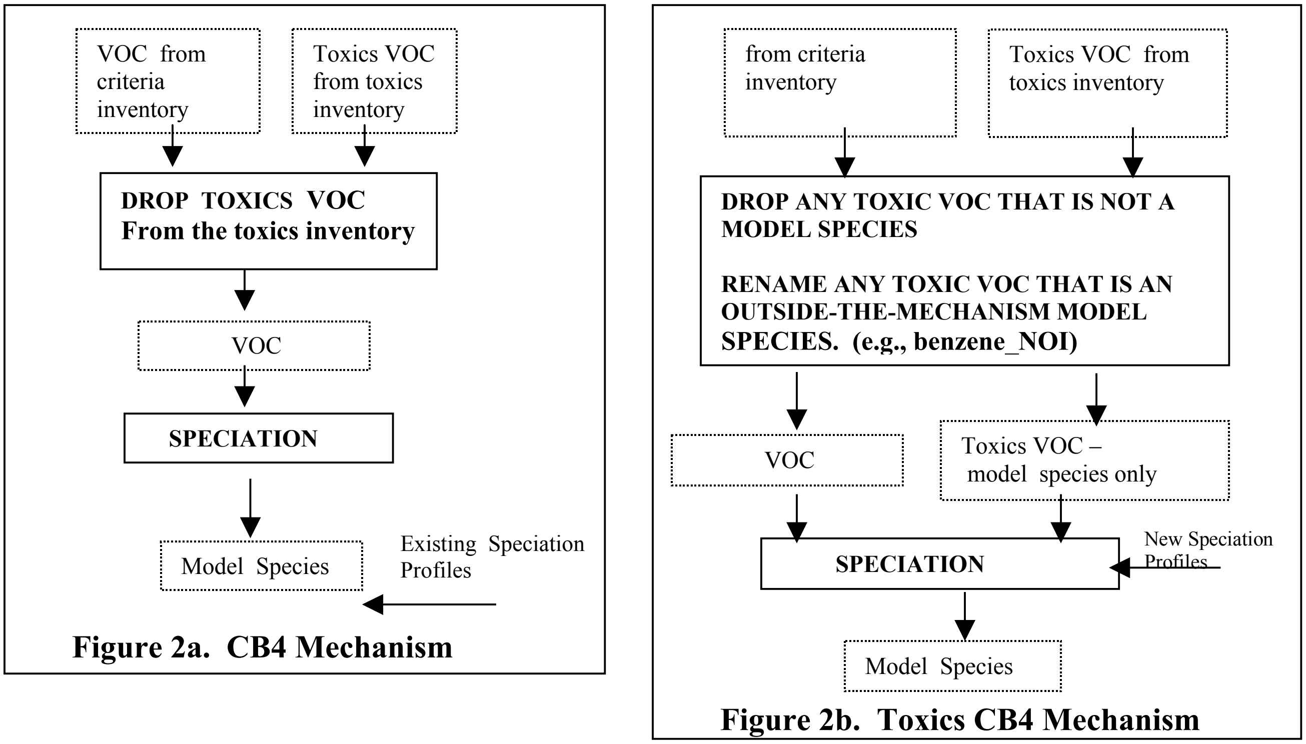


Figure 2a. CB4 Mechanism

Figure 2b. Toxics CB4 Mechanism

The benzene from the toxics inventory is used only for the model species "BENZENE." The VOC speciation profiles in this case will not sum to 1 (unless a particular profile contains none of the inside-the-chemistry species).

To provide the maximum flexibility, SMOKE allows the use of the integrate case for some sources and the no-integrate for other sources within the same run, since there may be some sources that are not amenable to the integration because of inconsistent inventory development approaches. For the toxics CB4 case, that means an outside-the-mechanism species such as benzene would be mapped differently for the integrate and no-integrate cases. For the integrate case, benzene is mapped to both benzene and PAR, but for the no-integrate case, benzene is only mapped to benzene. To be able to handle both cases within a single run, SMOKE changes the name of benzene to benzene_NOI (no-integrate) for the no-integrate case. The speciation profile must then contain an entry to map benzene_NOI to benzene. Note that SMOKE needs to know which of the toxics species are mechanism species, and which are outside-the-mechanism versus inside-the-mechanism species. SMOKE gets this information from user-provided codes in the Inventory Table.

PRELIMINARY TEST RESULTS

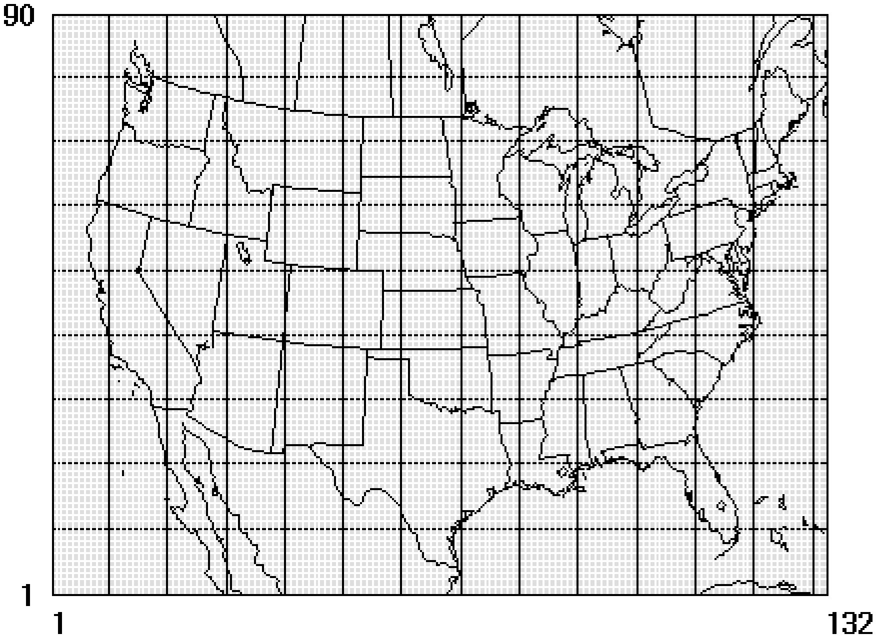
Testing of the code was performed using the following mobile source inventories:

- 1996 NEI for criteria pollutants
- 1999 (version 3 draft) NEI for HAPs
- 1999 vehicle miles traveled (VMT) with the 1999 MOBILE6.2 input files used to generate the 1999 on-road emissions.

Note that the 1999 NEI for criteria pollutants was not available in IDA format in time for the test runs. As a result of having the different-year toxics and criteria emission inventories, the testing could not be used to study the impact of the integration of the pre-computed emissions. It could be used to gain information on run time, memory and disk space requirements. As reported earlier, the incorporation of MOBILE6 into SMOKE caused a large increase in SMOKE run time, necessitating spatial and temporal temperature averaging to reduce the number of MOBILE6 runs.[9] The availability of 1999 VMT and 1999 MOBILE6.2 input files made it possible to examine at the impact of the integration for on-road sources when SMOKE runs the MOBILE6 model to generate on-road emissions.

Two national 36-kilometer grid, 7-day episodic test cases were used: one for the CB4 and another for the toxics-CB4 mechanisms, using the domain shown in Figure 3, below.

Figure 3: National domain



The runs were performed on Linux PCs with 2GB DDR RAM and an Athlon 1600+ 1.4GHz MP Processor. The computers run RedHat Linux 7.2 with XFS file system and have SCSI hard drives for output data and input data via the UNC network.

Figure 4 summarizes the performance of the nonroad runs. Note that for nonroad emission processing, only precomputed emissions were used, since SMOKE does not run the NONROAD model at this time. The temporal allocation and model-ready merge processing steps are per-day, averaged over seven days. The CPU usage times are provided to indicate which programs are more disk-intensive and whose times could be reduced by using faster disks and reducing competition from other processing running on the same computer or network. In our test cases, we were unable to get exclusive access to the computer network, and so any output was competing with approximately 60 other processors available on the Linux cluster that were writing to the same shared disk space. We anticipate that performance would be increased on other platforms that do not have such competition in writing to disk.

Based on these runs times, a national annual run could be very time intensive. The temporal allocation and model-ready merge steps would need to be repeated a minimum of 100 days (7 days per month and 16 holidays and day-after holidays). Using the numbers above, this would take 47 hours, 53 minutes for Temporal and 17 hours, 23 minutes for model-ready merging for approximately 3 days using 1 CPU. This time could be further reduced using multiple processors – only three processors would complete an annual/national run in 1 day. Another issue encountered during the testing was the size of the temporal output files, which were over 2 gigabytes per day. Such a large intermediate file will necessitate updating the SMOKE scripts to delete these files after each day has been run.

Figure 4. Summary of nonroad mobile processing performance

Nonroad Inventory Summary:

- -14.1 million inventory records
- 517,465 unique sources
- 43 pollutants, 30 kept as toxic VOC for integration with the VOC from the criteria inventory

Run Times:

Processing step	Wall clock time (MM:SS)	CPU usage
Inventory import	27:52	91%
Chemical speciation	3:27	58%
Spatial allocation	2:40	21%
Temporal allocation (per day)	28:44	14%
Model-ready merge (per day)	10:26	57%

For the on-road runs, all States with the exception of California and North Carolina utilized MOBILE6.2 in SMOKE. The on-road performance is summarized in Figure 5. The performance total when running for CB4 only and the toxics CB4 are provided separately. One difference in the runs was the CPU percent usage, which we could not control. The CB4 case needed to run only CO, NO_x, and VOC in MOBILE6.2, while the toxics CB4 case was run for CO, NO_x, VOC, all PM species (SO_2, NH_3, PM_{10}, $PM_{2.5}$) and all intrinsic and user-defined toxics species listed in Table 3 for on-road mobile sources, and this accounts for the timing differences for chemical speciation, emission factor generation, temporal allocation, and reporting.

As part of an effort to reduce processing times, the temperatures preprocessed by SMOKE’s Premobl program and used for MOBILE6.2 runs by SMOKE’s Emisfac program were temporally averaged over 1 week. This averaging provided 24-hour temporal profiles for that week, used for generating all emission factors. They were also spatially averaged over all counties within a state that share MOBILE6.2 emission factor input files.

Using the weekly temperature configuration, an annual run would need to be run for 365 days, though the emission factors would be run for 52 weeks. Based on the numbers above, CB4-only processing would require 21 hours 13 minutes for creating emission factors, 128 hours, 15 minutes for temporal allocation, and 81 hours 55 minutes for model-ready merge. The CB4-toxics processing took significantly longer, though part of that time was due to smaller CPU percent utilization for the processors. For an annual run based on the numbers above, a toxics-CB4 case would take 53 hours, 7 minutes for creating emission factors, 308 hours for temporal allocation, and an estimated 82 hours for model-ready merge. The model-ready processing step was not run for the CB4-toxics, but we estimate that the total processing time would be similar because the number of output model species was the same in both cases. In summary, annual/national CB4 processing would take approximately nine (9) and a half days using a single processor, and annual/national toxics-CB4 processing would take about twice as long (18 days) using a single processor. These runs times could be improved by using computational resources with multiple processors or with faster disk access and therefore higher CPU percent utilization.

Figure 5. Summary of on-road mobile processing performance

On-road inventory summary:

- 770,005 inventory records
- 220,362 unique sources
- 40 pollutants and activities (output from Smkinven)

Run Times for CB4 without toxics:

Processing step	Wall clock time (MM:SS)	CPU usage
Inventory import	1:48	90%
Chemical speciation	5:11	56%
Spatial allocation	0:23	57%
Emission factor setup	0:04	51%
Meteorology preprocessing	1:29	83%
Emission factor creation (per week)	24:29	93%
Temporal allocation (per day)	21:05	24%
Model-ready merge (per day)	13:28	53%
Reporting	4:33	26%

Run Times for toxicsCB4:

Processing step	Wall clock time (MM:SS)	CPU usage
Inventory import	1:48	90%
Chemical speciation	17:02	56%
Spatial allocation	0:20	60%
Emission factor setup	0:02	76%
Meteorology preprocessing	1:25	82%
Emission factor creation (per week)	1:52:26	83%
Temporal allocation (per day)	50:38	19%
Model-ready merge (per day)	unavailable	unavailable
Reporting	13:36	23

If the temperatures had not been temporally averaged for generating emission factors, then the run time would be the same for each day of processing instead of each week. Thus, without taking advantage of the temperature averaging, the annual run time would be seven times longer.

Another consideration in testing was to determine the impact of integrating the two inventories for the CB4 case. This was only possible for the on-road mobile runs in which we used the MOBILE6.2 model and therefore had a consistent basis on which to perform the integration of toxics and criteria VOC. A comparison was made of the CB4 model species estimates with and without the toxics integration. At this time, the results are preliminary.

Table 6 summarizes the percent change in moles of model species by vehicle type for each of the CB4 VOC model species (except the nonreactive "species" which is not a model species, but which results from the speciation profiles). The most dramatic differences show in the light duty diesel emissions, with a 138% increase in ALD2, 381% increase in FORM and 76% decrease in TOL.

Table 6: Changes in model-ready species emissions by vehicle type.

Vehicle type	ALD2 change	ETH change	FORM change	ISOP change	OLE change	PAR change	TOL change	XYL change
Light duty gas vehicles	9.2%	19.8%	42.2%	19.3%	13.5%	7.9%	12.7%	14.2%
Light duty gas trucks 1	10.4%	20.3%	54.8%	19.9%	17.2%	8.4%	12.7%	14.6%
Light duty gas trucks 2	14.0%	23.9%	119.1%	22.8%	17.4%	6.5%	4.1%	12.4%
Heavy duty gas vehicles	9.3%	17.7%	78.0%	16.4%	21.6%	6.0%	20.3%	18.8%
Light duty diesel vehicles	-20.7%	1.3%	-52.5%	0%	20.5%	4.3%	5.5%	25.8%
Light duty diesel trucks	137.9%	0%	380.5%	0%	49.0%	33.6%	-76.0%	-18.6%

Heavy duty diesel vehicles	7.1%	19.6%	28.7%	18.6%	13.3%	6.5%	13.3%	14.3%
Motorcycles	-20.6%	1.1%	-52.4%	0%	20.5%	4.3%	5.3%	29.4%

These differences create inventory-wide mass differences in the model species. The inventory-summed differences in M-moles/day are as follows:

ALD2 increased	2.2 M-moles	(12.0%)
ETH increased	4.3 M-moles	(20.7%)
FORM increased	4.4 M-moles	(56.5%)
ISOP increased	0.06M-moles	(21.3%)
OLE increased	2.7 M-moles	(15.0%)
PAR increased	51M-moles	(7.4%)
TOL increased	1.5 M-moles	(8.0%)
XYL increased	1.7 M-moles	(11.5%)
NR (non-reactive) *decreased*	94 M-moles	(-56.4%)

CONCLUSIONS

SMOKE has been modified to process multi-pollutant mobile source inventories for CMAQ. This modified version was released as SMOKE 1.5 Beta (SMOKE1.5b). The key design issues solved in this release were the addition of a sparse inventory format containing a large number of pollutants and the ability to combine it with the criteria inventory in the case of overlapping pollutants. There are other functions included in 1.5 Beta that were not discussed in the paper, and work continues to add the capability to process multi-pollutant inventories for additional source categories. Based on the testing, the use of toxics inventories for national and annual runs will be somewhat time intensive, and will require significant disk space requirements. Deleting intermediate files for each day after the model-ready emissions have been created can improve these disk space requirements.

Preliminary results were obtained on the impact of integration on the model species for mobile source inventories generated using MOBILE6 in SMOKE. The preliminary results showed an increase in overall reactive VOC emissions. This varied across model species and vehicle types. In general, the reactive model species increased emissions across the inventory from 7% (parrafins) to 57% (formaldehyde group). These differences vary by vehicle type, which will cause spatial variations by state, county, and road type.

In addition to continued testing on the results of integration, future work on SMOKE related to processing multi-pollutant inventories will be to release a version that will process the stationary and mobile toxics and criteria inventories. Also, we plan to explore parallel processing and other ways to reduce processing time, memory and disk space requirements.

DISCLAIMER

This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for presentation and publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

REFERENCES

1. *Science Algorithms of the EPA Models-3 Community MultiScale Air Quality (CMAQ) Modeling System*, U. S. Environmental Protection Agency, Research Triangle Park, NC, 1999; EPA-600/R-99/030.

2. *User's Guide to the Regional Modeling System for Aerosols and Deposition (REMSAD), Version 7,* ICF Consulting, July 2002.

3. Houyoux, M.; Vukovich, J.; Brandmeyer, J. *Sparse Matrix Operator Kernal Emissions Modeling System (SMOKE) User Manual*, Version 1.1.2 Draft, MCNC-North Carolina Supercomuting Center Environmental Programs, 2000. (Updates at http://www.cmascenter.org/modelclear.html#smoke)

4. Possiel, N.; Stella, G.; Ryan, R.; Pace, T.;Benjey, W.; Beidler, A.; Kinnee, E.; Houyoux, M.; Adelman, Z.. "Development of an Anthropogenic Emissions Inventory for Annual Nationwide Models-3/CMAQ Simulations of Ozone and Aerosols", Presented at the International Emission Inventory Conference, "One Atmosphere, One Inventory, Many Challenges." of the Environmental Protection Agency, Denver, CO, May 2001.

5. *User's Guide for the Assessment System for Population Exposure Nationwide (ASPEN, Version 1.1) Model*, U. S. Environmental Protection Agency, Research Triangle Park, NC, 2000; EPA-454/R-00-017

6. *User's Guide for the Industrial Source Complex (ISC3) Dispersion Models*, U. S. Environmental Protection Agency, Research Triangle Park, NC, 1995; EPA-454/B-95-003a.

7. *User's Guide for the Emissions Modeling System for Hazardous Air Pollutants (EMS-HAP, Version 2.0)*, U. S. Environmental Protection Agency, Research Triangle Park, NC, 2002; EPA-454/B-02-001.

8. *NATA – Evaluating the National-Scale Air Toxics Assessment 1996 Data – An SAB Advisory, An Advisory by the EPA Science Advisory Board (SAB)*, U. S. Environmental Protection Agency, Washington, DC, 2001; EPA-SAB-EC-ADV-02-001.

9. Seppanen, C.; Houyoux, M.; Strum, M.; Stella, G.; Possiel, N.; "Design and Application of SMOKE Integration with MOBILE6", Presented at the International Emission Inventory Conference, "Emission Inventories - Partnering for the Future," Atlanta, GA, April, 2002.

10. Gerald Gipson, Environmental Protection Agency, Office of Research and Development, Personal Communication, Spring 2002.

11. Documentation for the 1999 National Emissions Inventory Version 2.0 for Criteria Air Pollutants and Draft Version 3.0 for HAPs, On-Road Sources, U. S. Environmental Protection Agency, Research Triangle Park, NC, October 2002, (ftp://ftp.epa.gov/EmisInventory/draftnei99ver3/haps/documentation/onroad/or_doc99v2_oct02.pdf)

12. Documentation for Aircraft, Commercial Marine Vessel, Locomotive, and Other Nonroad Components of the National Emissions Inventory, Volume 1, Methodology, U. S. Environmental Protection Agency, Research Triangle Park, NC, November 2002, ftp://ftp.epa.gov/EmisInventory/finalnei99ver2/criteria/documentation/nonroad/nonroad_voli_nov2002.pdf

KEYWORD

SMOKE

emission modeling
emission processing
EMS-HAP
MOBILE6