

# Updates to the Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System and Integration with Models-3

**Marc R. Houyoux and Jeffrey M. Vukovich**  
Environmental Programs  
MCNC–North Carolina Supercomputing Center  
3021 Cornwallis Road  
Research Triangle Park, NC 27709-2889

## ABSTRACT

The Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system has been integrated with the Environmental Protection Agency's Models-3 air quality modeling system. This paper provides an overview of the emission processing capabilities of the revised SMOKE system (version 1.0), and describes how they can benefit the air quality modeling community. Like the SMOKE prototype, SMOKE version 1.0 processes area-source, biogenic, mobile-source, and point-source emissions data to create gridded, hourly emission files for air quality models. It also permits faster emission processing (in minutes rather than days or longer) and requires less disk space than comparable emission processors.

SMOKE version 1.0 is specifically designed for support of emission processing with user-selected chemical mechanisms and for assessing reactivity-based controls, such as reformulating an industrial process to use a volatile organic compound that has less ozone-forming potential. A number of additional benefits arise from these enhancements and from integration with Models-3: (1) the ability to process any pollutant (criteria, hazardous, etc.) for area and point sources, and to some degree for mobile sources; (2) the ability to output emissions for any model species for which the user can define speciation profiles; (3) multiple-country processing; (4) the ability to input more inventory formats; (4) direct communication with Models-3 to retrieve grid descriptions, episode information, and chemical mechanism details; (5) support of the Models-3 plume-in-grid preprocessor; and (6) a Models-3 user interface for SMOKE input file preparation.

## INTRODUCTION

Researchers at the North Carolina Supercomputing Center (NCSC) developed the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system prototype through a cooperative agreement with the U.S. Environmental Protection Agency (EPA) that explored improving the efficiency of emission data processing. SMOKE uses matrix-vector multiplication to significantly improve emission processing time, produce emissions results comparable to those from other processing systems, and do so without the need for third-party software such as SAS<sup>®</sup>.<sup>1,2</sup> SMOKE performs the core functions of emission data processing: the spatial allocation, temporal allocation, chemical speciation, and control of area-source, mobile-source, and point-source emissions, and the generation of biogenic emission estimates. SMOKE contains a driver for the MOBILE5a and MOBILE5b models,<sup>3,4</sup> and it also uses a reorganized version of the Urban Airshed Model–Biogenic Emission Inventory System, version 2 (UAM-BEIS2).<sup>5</sup> The SMOKE prototype was developed using the FORTRAN language.

We have developed SMOKE version 1.0 to further improve emissions processing and to make it useful for the Models-3 air quality modeling system.<sup>6</sup> EPA supported a project for this effort that had two primary purposes: First, we were to integrate the existing functionality of the SMOKE prototype into the Models-3 framework, and to add the few extra capabilities needed for processing emissions for EPA's Community Multiscale Air Quality (CMAQ) model.<sup>7</sup> Second, we were to add reactivity-based control capabilities to SMOKE, including the ability to process any pollutant and use any speciation profile.

The resulting integrated system permits SMOKE version 1.0 to be used from the Models-3 Framework, and allows SMOKE version 1.0 to read files prepared by a new Models-3 component, the "SMOKE Tool." The integrated system supports emission data processing for user-selected chemical mechanisms and reactivity assessments, and SMOKE can access and use the data objects (such as grid definition and chemical mechanisms) available from the Models-3 Science Manager database..

Figure 1 shows the relationship between SMOKE version 1.0 and the Models-3 Framework. The Model-3 Framework components that interact with SMOKE are the SMOKE Tool, the Science Manager, the Data Manager, and the Study Planner. The SMOKE Tool provides a user interface for preparing input files for SMOKE, such as the inventory files; the temporal, speciation, and gridding cross-reference files; the temporal and speciation profiles; the gridding surrogates; and the control input files. The Science Manager provides a user interface for defining grids and layer structures, defining the dates and times of the modeling episode, and defining chemical mechanisms. The Data Manager keeps track of SMOKE input and output files, and uses an object-oriented database with metadata that can include detailed descriptions of the data files. Finally, the Study Planner provides a user interface for selecting grids, layer structures, episodes, and chemical mechanisms that were defined using the Science Manager. The Study Planner also provides a user interface for assigning SMOKE files in the Data Manager to SMOKE programs and running the SMOKE programs to process the emissions data. One final note is that although SMOKE version 1.0 can interact closely with the Models-3 modeling system, it can also be run independently from Models-3 and does not require SAS.

In addition to the benefits of using SMOKE from the Models-3 Framework, many other improvements have been made to SMOKE that will advance air quality modeling efforts. These are the following:

- Processing for any inventory pollutant
- Processing for user-defined split factors and model species
- Multicountry processing for up to 10 countries
- Support of multiple inventory formats
- Plume-in-grid processing
- Reactivity controls
- Dynamic memory allocation

These improvements and their expected benefits are described in more detail in the sections following the background. Last, we will describe future plans for further enhancements.

## **BACKGROUND**

Before we describe the improvements to SMOKE, we provide a description of the SMOKE prototype's capabilities. The prototype processes the standard four emissions source categories (area, biogenic, mobile, and point sources), and it processes only nitrogen oxides (NO<sub>x</sub>), volatile organic compounds (VOC) or reactive organic gases (ROG), and carbon monoxide (CO). For all source categories, the purpose of SMOKE processing is to convert the summary resolution of the inventory emissions to the gridded, hourly, model species values needed by an air quality model. Because of the inventory pollutants it processes, the SMOKE prototype has been used only for ozone air quality modeling.

SMOKE uses profile tables and cross-reference tables to convert the emissions resolution. The profile tables contain the factors for converting emissions from one resolution to another. For example, for converting from daily emissions to hourly emissions, SMOKE uses 24 factors (one for each hour). Each profile entry is assigned a profile number. The cross-reference tables assign the profiles to each source. They contain source characteristics and the profile numbers to use when SMOKE matches the source characteristics in this file to a source. The cross-reference tables are applied to the sources in a stepwise

manner, such that the most specific entry is always applied. For example, if a cross-reference entry were available for a specific plant, stack, and source category code (SCC), SMOKE would apply it instead of the cross-reference entry that matched that source only by SCC.

SMOKE arranges the inventory emissions data into vectors sorted by the source characteristics. In applying the various profiles to the emissions data, the sorted order is utilized to improve the processing efficiency. The factors from the profiles and surrogates are stored in matrices: a speciation matrix, a gridding matrix, and a control matrix. The temporal assignments convert the emissions vectors to hourly emissions vectors. For point sources, SMOKE also computes a layer fraction matrix for allocating elevated sources to the model layers. To create the hourly, gridded, model-species files for the air quality model, a merge processor combines the hourly emissions vectors with the various matrices using an efficient vector-matrix multiplication algorithm.

Each source category requires a slightly different processing approach. For area sources, emissions data are available by county and SCC as an average-daily value or annual value. Weekly (day-of-week) temporal profiles convert from an average day to an average day of the week; for example, from an estimated annual-average, daily emissions value to an average Monday, Tuesday, Wednesday, etc. Diurnal temporal profiles convert from daily emissions to hourly emissions. SMOKE assigns the temporal profiles to each source using cross-reference tables. For speciation, SMOKE divides the inventory  $\text{NO}_x$  into 90% NO and 10%  $\text{NO}_2$  uniformly across the inventory based on the molecular weight of  $\text{NO}_2$  (i.e.,  $\text{NO} = (0.9/46)\text{NO}_x$ ), and it converts ROG emissions to model species using speciation profiles that contain pollutant-to-species factors. SMOKE assigns the sources a speciation profile through another cross-reference table. It also allocates the county-total data to each grid cell using a spatial surrogate such as population, housing, water, or agriculture. The surrogate that is used depends on the SCC, and SMOKE makes this assignment through another cross-reference table. Finally, SMOKE combines the intermediate files created by each of these steps to create an hourly, gridded, model-species area-source emissions file.

For biogenic sources, the basic inventory data are gridded land use data. These data indicate the types and amounts of vegetation in each grid cell. SMOKE associates these land use data with the BEIS2 emission factors, which convert the land use value to a “normalized” emissions value. SMOKE then adjusts this emissions value by meteorological conditions (temperature and solar radiation). SMOKE can also use county-total land use data, and allocate these data to grid cells using gridding surrogates.

For mobile sources, the basic inventory data are (1) daily vehicle miles traveled (VMT) by county, road type, and optionally road segment (link), and (2) inputs to the MOBILE5 emission factor generation program. SMOKE processes mobile sources using a system similar to that for area sources, based on temporal profiles and cross-references, speciation profiles and cross-references, and gridding surrogates and cross-references. Mobile-source processing is different from area-source processing, however, in that SMOKE computes the emissions by multiplying the VMT with the emission factors from the MOBILE5 model. SMOKE computes the emissions in this way for each hour and by county and/or state, road type, vehicle type, and emission process (e.g., exhaust, evaporative). Also, the VMT values are computed for each vehicle type from the county and road type using factors called VMT mix data. SMOKE drives the MOBILE5 emission factor program using hourly, gridded temperature data from a meteorology model, and the emission factors are applied for each source based on the temperatures at that source.

For point sources, the basic inventory data are average-day or annual emissions for each county, facility (plant), device, process, stack, and SCC. SMOKE also uses temporal and speciation profiles and cross-references for point sources. Gridding surrogates are not required because, by definition, point sources have coordinates available for allocating their emissions to the appropriate grid cell. In addition, whereas SMOKE treats the other source category emissions as occurring at the surface, point sources have a vertical component. Many of the sources have tall stacks and the gases often exit at high velocity and temperature, causing the plumes to extend into the model layers above the first layer. The plume rise of these sources must be computed, and a method using a Briggs algorithm has been devised to make such computations.<sup>8</sup> Primarily, temperature, boundary layer height, pressure, and wind (direction and

speed) affect the plume rise. In many cases, day- or hour-specific emissions are available for point sources, and SMOKE can import these and override temporal allocation from the temporal profiles.

SMOKE intermediate and output files are in a format known as Network Common Data Format (NetCDF), which is used by the CMAQ. In programs expected to work in Models-3, the NetCDF files are created with a library called the Input/Output Applications Programming Interface (I/O API). The benefits of I/O API NetCDF files are (1) they are portable to any computer platform; (2) they are binary files and are therefore smaller than other portable formats such as text files; (3) they are self-describing files because they contain a header with information about the date, time, grid, and other important file characteristics; and (4) the data are direct-access, which means that each variable can be accessed without the need to scroll through other variables.

## SMOKE VERSION 1.0 PROGRAMS

Figure 2 shows the SMOKE version 1.0 programs and how they relate to one another. One major difference between SMOKE version 1.0 and the SMOKE prototype is that the programs are now shared by multiple source categories, resulting in many fewer programs. This enhancement was possible in part because of the dynamic memory allocation described later. The reduced number of programs permits more code reuse than in the prototype; this was beneficial during code development and will benefit code maintenance in the future. Note that although the programs can be used for multiple source categories, they must be run for one source category at a time. Below we summarize the SMOKE programs and their major purposes.

- **Smkinven:** Imports the inventory emission and VMT inventory data created by the SMOKE Tool or by the user.
- **Rawbio:** Imports county land use and gridding surrogates to create normalized gridded biogenic emissions.
- **Grdbio:** Imports gridded land use to create normalized gridded biogenic emissions.
- **Grdmat:** Creates the gridding matrix. For all source categories, this program uses the inventory files generated by Smkinven. For point sources, it uses source coordinates; for area sources, it uses gridding surrogates and cross-reference files; for mobile sources, it uses gridding surrogates and cross-reference files for nonlink sources, and link coordinates for link sources. For mobile sources, it creates an ungridding matrix for deriving source-specific temperatures from the gridded temperatures.
- **Premobl:** Preprocesses the temperature data for mobile-source emission data processing when MOBILE5 is being used to generate emission factors for use with VMT data.
- **Emisfac:** Drives the emission factor model(s). This version drives only the MOBILE5a or MOBILE5b models.
- **Elevpoint:** Reads in a list of major point sources (MPSs) and plume-in-grid (PinG) sources and creates a STACK\_MEPSE file for the Plume Dynamics Model (PDM), which is a preprocessor of PinG sources for the CMAQ. Elevpoint also creates a text file that other SMOKE programs use to identify MPSs and PinG sources.
- **Temporal:** Creates hourly emissions files for area, mobile, and point sources. Applies temporal profiles and cross-reference data to the inventory files from Smkinven. It also reads in SMOKE-formatted day- and hour-specific emissions (although there is currently no way to generate these in SMOKE version 1.0). For mobile sources, it applies the emission factors from Emisfac to the VMT data based on an emission factor cross-reference file, the gridded temperatures, the ungridding matrix from Grdmat, and the preprocessed temperatures from Premobl.
- **Spcmat:** Creates the speciation matrix. Applies speciation profiles and cross-reference files for area, mobile, or point sources, using inventory files from Smkinven. It outputs only species that

are associated with pollutants appearing in the inventory. It creates both a mass-based and mole-based speciation matrix for each run.

- **Cntlmat:** Creates one or more control matrices (multiplicative, additive, projection, or reactivity) for area, mobile, or point sources, using inventory files from Smkinven. It uses control packets to determine which controls to apply and the values of factors for those controls.
- **Laypoint:** Creates a layer-fractions matrix for point sources based on the stack characteristics, gridded meteorology data, and adjustments needed for MPSs and PinG sources based on the Elevpoint outputs.
- **Tmpbio:** Applies meteorology adjustments to the normalized emissions from Rawbio or Grdbio to create hourly, gridded, speciated biogenic emissions. The speciation is done based on a speciation profile provided as another input.
- **Smkmerge:** Combines inventory (annual) or hourly emissions data with gridding matrices, speciation matrices, and control matrices for any combination of anthropogenic source categories or biogenic emissions. Can write 2-D or 3-D model-ready emissions file depending on whether or not layer fractions from Laypoint are used. Can output plume-in-grid outputs if PinG identification from Elevpoint is used as input. Can also output state and/or county totals.
- **Grwinven:** Combines the projection and/or control matrices with an inventory file, and writes out another inventory file in SMOKE format and/or in the Inventory Data Analyzer (IDA) format that can be read by the SMOKE Tool.

## SMOKE ENHANCEMENTS FOR VERSION 1.0

### Processing for Any Inventory Pollutant

To support the flexibility inherent in Models-3 for modeling ozone, particulates, or toxics, SMOKE was enhanced to support user-selected inventory pollutants. The SMOKE programs do not have any pollutant names or characteristics coded directly into the programs; rather, this information is provided by the input files when the user runs the programs. There are two exceptions to this rule. The first exception is for mobile sources, which rely on emission factors from the MOBILE5 program. MOBILE5 can compute emission factors only for CO, NO<sub>x</sub>, and one of VOC, total hydrocarbons (THC), nonmethane hydrocarbons (NMHC), total organic gases (TOG), or nonmethane organic gases (NMOG). The second exception is for biogenic emissions, for which the BEIS2 model (and its SMOKE version) can compute emissions based on nitric oxide, terpenes, and “other VOCs” (i.e., other than terpenes).

The enhancements affect the SMOKE programs responsible for inventory import, temporal processing, speciation matrix generation, control matrix generation, and merge processing. The resulting system can import and process emissions of any inventory pollutant for area, mobile, and point sources. For mobile sources and for pollutants other than CO, NO<sub>x</sub>, and VOC, the emission values must be computed outside of SMOKE for each county, road type, and (optionally) link. SMOKE processes these emissions much like it processes area sources. In Smkinven (the data import program), the pollutant names are obtained from the inventory files, and these are written to a SMOKE inventory file for each source category. The other programs get the pollutant names from the SMOKE inventory file.

There are two limitations on the user in specifying pollutants for SMOKE. First, the pollutant names must be 16 characters or fewer. Second, because I/O API files are limited to 120 variables, there is a limit to the number of pollutants that SMOKE can handle in a single run. This limit varies by source category because different source categories require storage of different numbers of variables other than the emissions values. For area sources, SMOKE can handle up to 19 pollutants, for mobile sources, 54, and for point sources, 15.

These improvements to SMOKE provide a consistent emission data processing approach for criteria, particulate, and toxic pollutants. Since any pollutants can be processed by the system (given the necessary data), SMOKE can be used in future work that uses pollutants for which inventories are not yet available.

## Processing for User-defined Split Factors and Model Species

SMOKE version 1.0 can also support user-defined split factors and any model species. The programs are designed to get all of the information about model species from the input files, including species names, which pollutant(s) they are associated with, how to match the speciation profiles to the sources in the inventory, and the speciation factors.

These enhancements affect the programs responsible for speciation matrix generation and merge processing. The Spcmat program gets the names and other characteristics of the species from the speciation profiles file. Since SMOKE does not have explicit instructions as to which pollutants are speciated and which are not, the speciation profiles file must also designate which pollutants do not get split into multiple species. The same limitations imposed by the I/O API on pollutants also apply for the model species: species names are limited to 16 characters, and the output files are limited to 120 model species. The 120-species limit applies for each source category and for totals across all source categories in the final model-ready input files.

These improvements will allow SMOKE users to prepare emissions for different chemical mechanisms, such as the carbon bond IV (CB-IV)<sup>9</sup> or Regional Acid Deposition Model (RADM)<sup>10</sup> mechanisms. In addition, users can develop experimental chemical mechanisms and process emissions for them. Although this benefit is not necessarily useful for routine emissions processing for regulatory purposes, it is useful to air quality model researchers who evaluate different chemical mechanisms in air quality models. Furthermore, these changes are necessary for supporting reactivity controls, as described below.

## Reactivity Controls

The premise behind reactivity controls is that ozone formation can be affected by changing the types of VOCs released from emission sources. Different chemical compounds have different components, represented in air quality models as chemical species, and these components have different tendencies to contribute to ozone formation. This tendency is the *reactivity* of the chemical component, which contributes to the reactivity of the chemical compound. The concept of a reactivity control is that by changing the chemical compounds used in emission-producing activities in order to reduce the reactivity, ozone formation can be inhibited and air quality can be improved.

To date, there is no evidence as to whether reactivity controls work in practice. The best way to determine whether reactivity controls might be effective is through modeling. Therefore, part of the motivation for enhancing SMOKE was to provide an emission data processing tool that is capable of processing emissions to model the effects of reactivity controls. *Reactivity assessment* is a term used to refer to the evaluation through modeling of whether a particular compound or source is a good candidate for chemical compound replacement.

Several issues are important when addressing the emissions processing requirements for reactivity assessments. First, reactivity assessment involves the replacement of one compound in the inventory by another compound, and this replacement can impact emission projections, the total magnitude of the inventory pollutants, and the associated SCCs. The degree to which the replacement compound is used across the industry (market penetration) may vary in time and space, which affects the future-year emissions. Also, the replacement compound may be needed in much greater or much smaller amounts, and this affects the total inventory emissions. Finally, if a different process is required to use the different compound, the SCC for that source may change.

Second, the scale of the reactivity assessment is important; it could be local, statewide, or national. The local case could involve investigating reactivity for one source. The statewide case could be implementing a change in compound based on reactivity considerations for a State Implementation Plan, and this would affect sources across the state. The national case could involve EPA's investigating the formulation of nationally distributed consumer products.

Third, exemptions from controls for certain sources must be permitted as part of an emissions control strategy. These exemptions can occur when a reactivity assessment determines that certain compounds and/or processes do not significantly affect pollution formation.

To address these issues, an emissions processing system must be able to target changes in a VOC for specific classes of VOC emissions, and to account for the spatial and temporal considerations implied by market penetration issues. Furthermore, when replacement options are being investigated, the correct replacement operations need to be facilitated by the emissions processing system. These operations include selecting sources, changing underlying pollutant emissions, changing SCCs, correctly projecting future-year emissions based in part on market penetration issues, and appropriately speciating emissions for the new compound. Because new speciation may need to use a different speciation mechanism than was used for the uncontrolled emissions, flexible speciation is needed to support reactivity controls.

During the design of SMOKE version 1.0, we developed the following formulation for applying reactivity controls to inventory VOC emissions in SMOKE. The vector of total model species emissions,  $\bar{E}_{ty}$ , in year  $y$  from a source is given by

$$\bar{E}_{ty} = \bar{E}_{ry} + \bar{E}_y \quad (1)$$

where

$\bar{E}_{ry}$  = emissions contribution from reactivity-controlled emissions for the source in year  $y$  [tons]

$\bar{E}_y$  = emissions contribution from non-reactivity-controlled emissions for the source in year  $y$  [tons]

These components are given by:

$$\bar{E}_{ry} = V_r P_{ry} k_{ry} \bar{s}_r \quad (2)$$

$$\bar{E}_y = V P_y (1 - k_{ry}) \bar{s} \quad (3)$$

and

$$k_{ry} = \text{MIN}(1, n_y m_r) \quad (4)$$

where the subscript  $r$  indicates a value prescribed by a reactivity-based control, and

$V$  = base-year VOC emissions for the source [tons]

$P_y$  = projection factor from base year to year  $y$  [tons in year  $y$  / tons in base year]

$k_{ry}$  = reactivity-control market penetration in year  $y$  [fraction]

$\bar{s}$  = array of speciation factors needed to convert VOC emissions to model species emissions

$n_y$  = number of years between base year and year  $y$

$m_r$  = market penetration rate (0.0 to 1.0) [fraction/year]

In some cases, a reactivity-defined base-year VOC emissions value may not be needed, in which case  $V_r$  would default to  $V$ . Similarly, the reactivity-defined projection factor,  $P_{ry}$ , may default to the base-year projection factor,  $P$ . If  $y$  is equal to the base year, then Equation (2) goes to zero,  $P_y$  goes to 1, and  $\bar{E}_{ty}$  equals base-year speciated emissions,  $V \bar{s}$ .

In other cases, the user might want to ignore the market penetration, and model the speciation changes and base-year VOC emissions from the reactivity controls in the base year. In this case, an option is available in the inventory projection program, Grwinven, to do this, which can be accomplished by setting  $n_y m_r$  to 1.

To support reactivity assessments, the SMOKE programs were modified to support any inventory pollutant and any speciation profile, as described above. A new type of controls, "reactivity controls,"

was added to the control processing for all source categories; these controls are designed to address all of the major issues identified above for reactivity assessments. Reactivity controls generate a new type of control matrix, the reactivity matrix. Finally, we enhanced the merge processing to permit the reactivity matrix to be merged with the emissions, speciation matrix, gridding matrix, and other control matrices, to produce model-ready emissions.

These enhancements to SMOKE provide a tool for exploring reactivity controls and modeling their effect on air quality. This will help researchers determine whether reactivity controls offer any benefits, which ultimately could provide another mechanism for controlling emissions to improve air quality.

### **Plume-in-Grid Processing**

SMOKE has been updated to support plume-in-grid processing for the CMAQ. PinG processing provides a more detailed approach for major elevated point sources than is offered by the plume rise approaches that consider plume rise only at the resolution of the grid cell.

The SMOKE prototype can treat all point sources as potentially elevated point sources, and it computes plume rise above the first layer only when the physics of the meteorology and stack parameters result in elevated plumes. SMOKE version 1.0 maintains this capability, but also permits the user to define MPSs and PinG sources. When MPSs are defined, SMOKE treats only those sources as potentially elevated. When PinG sources are defined, the Elevpoint program creates a stack parameters file that is used by the Models-3 PDM, and Smkmerge (the merge program) outputs an hourly PinG source emissions file used by the CMAQ. The PDM is a preprocessor for the CMAQ and performs PinG preprocessing using meteorology data and the stack parameters file.

In the SMOKE Version 1.0, the SMOKE user must define the MPSs and PinG sources outside of SMOKE. This is one of the functions of the SMOKE Tool. This approach, however, does not permit hour-specific emissions data to be used as one of the criteria for determining PinG sources. This limitation will be removed in future versions of SMOKE.

### **Multicountry Processing for up to 10 Countries**

SMOKE has been updated to more fully support multiple countries. Instead of using a 5-digit Federal Information Processing Standards (FIPS) code and assuming that the default country is the United States, SMOKE now uses a 6-digit country, state, and county code. The first digit is the country code, which can have a value of 0 to 9 (hence the 10-country limit). Digits 2 and 3 are for the state code, and digits 4-6 are for the county. The user can use the three sections of the 6-digit code for regions other than countries, states, and counties, such as provinces, nonattainment regions, or air basins. An input file supplies the definitions of these codes, and the default file is configured for the United States, Canada, and Mexico (countries 0, 1, and 2). However, this file can be configured to support nonstandard regions as well.

### **Support of Multiple Inventory Formats**

For area and point sources, the SMOKE prototype provided support for the Emissions Preprocessor System 2.0 (EPS2.0) and Emissions Modeling System-95 (EMS-95) inventory formats. For mobile sources, the SMOKE prototype had its own format, which was similar to the EMS-95 format. For area sources, these formats were fully supported for the needs of most users. For point sources, support of these formats was limited by the fact that the SMOKE prototype defined a point source as a unique combination of FIPS code, SCC, plant, and stack. In reality, EPS2.0 point sources could additionally be defined by "points" and "segments." The SMOKE prototype ignored these source characteristics and instead aggregated emissions across multiple EPS2.0 sources with the same FIPS code, SCC, plant, and stack. Furthermore, the SMOKE prototype supported only integer-valued plant and stack codes, although the EPS2.0 format permits character strings for these fields. In the EMS-95 format, sources are defined as a FIPS code, plant, stack, device, process, and SCC. The SMOKE prototype aggregates

emissions across device and process to fit the definition of a SMOKE source. It does permit character fields in the plant and stack fields, but it assigns an internal integer value for these fields.

In addition to supporting the EPS2.0 and EMS-95 formats, SMOKE version 1.0 also imports the IDA formats for area, mobile, and point sources. This format is output by the SMOKE Tool, and it is the same as the National Emissions Trends (NET) format with the addition of a header that defines the country and pollutant names. Because point sources are defined differently for each of these formats, SMOKE lets the file format determine what the definition of a point source should be inside the programs. Furthermore, for all file formats, the Smkinven program can combine multiple files of the same format.

The addition of the IDA format to the formats that SMOKE can read will allow more people to get their data working with the SMOKE system more easily. The SMOKE Tool also can import the various formats listed above, and convert them to the IDA format. However, if users do not want to use the SMOKE Tool or cannot because they do not have SAS, SMOKE can import the different formats directly.

### **Dynamic Memory Allocation**

The SMOKE prototype requires the user to recompile the SMOKE programs for each inventory and/or modeling grid. SMOKE version 1.0 has been written in FORTRAN 90, which permits dynamic allocation of computer memory at program run time. This means that users need to compile the programs only once for their computer, or can even use program executables compiled on another computer of the same platform. It also means that the SMOKE programs require less memory than comparable programs from the SMOKE prototype.

Dynamic memory allocation is particularly useful when SMOKE is being used for a new study, and sources are constantly being added to or removed from the inventory. It is also useful when the user is processing emissions on several grids, because the same programs can be used for all grids. Dynamic allocation is what is done for most software, so by adding this to SMOKE we have simply brought it up to the standards of modern computer software. Nevertheless, it is a significant improvement that current users of the SMOKE prototype will especially appreciate.

### **FUTURE PLANS**

A number of important parts of emission data processing will not be available in SMOKE version 1.0. First, day- and hour-specific data import will not be available for point sources. This means that data such as the Continuous Emissions Monitoring (CEM) data cannot be imported to be used by the Temporal program. Second, the SMOKE quality assurance programs have not been completely updated to work with the new SMOKE system. To some extent, the existing programs can be compiled and run for CO, NO<sub>x</sub>, and VOC, but this does not provide many needed quality assurance steps. Third, the identification of MPSs and PinG sources can be done only outside of SMOKE, and not within SMOKE where SMOKE could use (when they are added) the day- and hour-specific emissions as part of the criteria for selecting these sources. Finally, output for the UAM formats are not supported in this version of SMOKE.

Removing the above limitations is the short-term goal for SMOKE development. In addition, we plan to support even more advanced chemical mechanisms, such as the morphecule mechanism being developed by Harvey Jeffries at UNC-Chapel Hill. We also hope to remove or lessen the limitations on the number of inventory pollutants and model species. Further, we plan to port SMOKE to other platforms, including Linux, Windows NT, and all of the UNIX platforms; currently, SMOKE version 1.0 can run on Sun and SGI UNIX workstations. Finally, we would eventually like to permit SMOKE to drive other emission factor models such as EPA's PART5.<sup>11</sup>

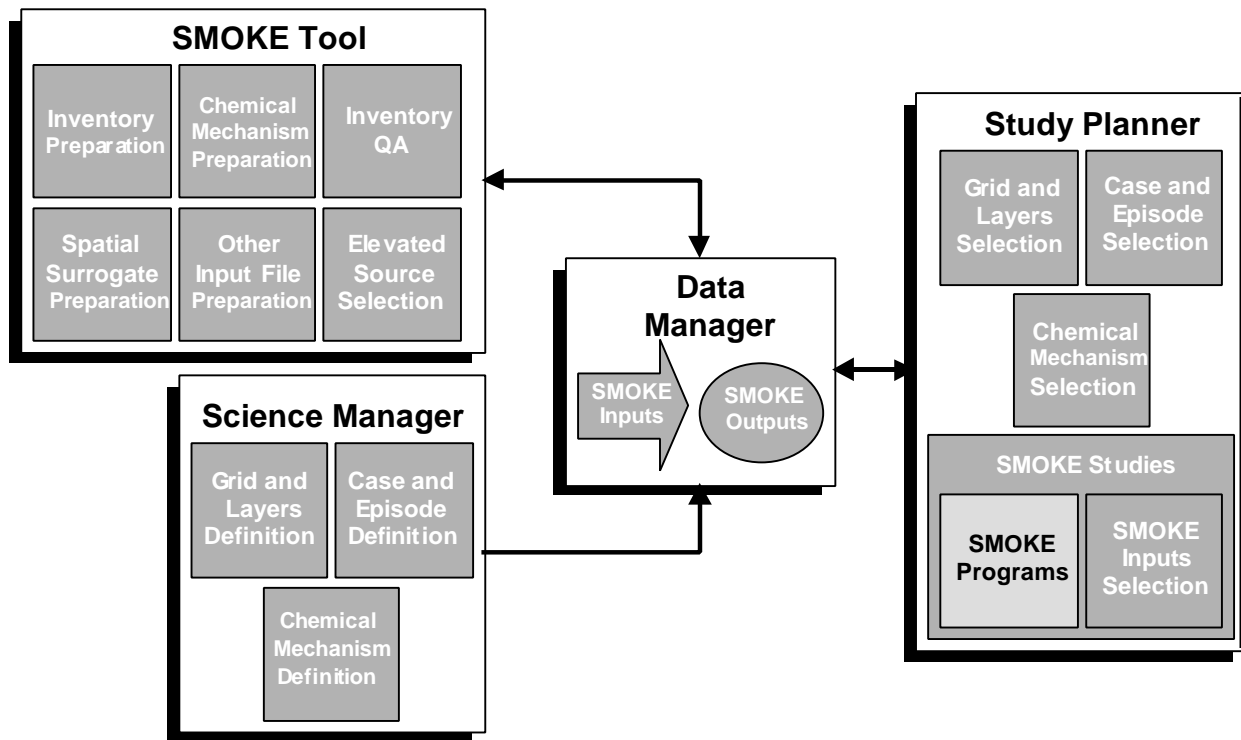
## DISCLAIMER

The U.S. Environmental Protection Agency through the Modeling Systems Analysis Branch funded and managed the research described here under Requisition/Reference number QT-RT-98-001208. This paper has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## REFERENCES

1. Coats, C. J., Jr.; Houyoux, M. R. "Fast Emissions Modeling with the Sparse Matrix Operator Kernel Emissions Modeling System," Presented at *The Emissions Inventory: Key to Planning, Permits, Compliance, and Reporting*, Air & Waste Management Association, New Orleans, LA, September 1996.
2. Houyoux, M. R.; Coats, C. J., Jr.; Eyth, A.; Lo, S. "Emissions Modeling for SMRAQ: A Seasonal and Regional Example Using SMOKE," Presented at *Computing in Environmental Resource Management*, Air & Waste Management Association, Research Triangle Park, NC, December 1996.
3. U.S. Environmental Protection Agency. *User's Guide to MOBILE5 (Mobile Source Emission Factor Model)*; Office of Air and Radiation/Office of Mobile Sources, U.S. EPA, Ann Arbor, MI, May 1994; Rep. EPA-AA-AQAB-94-01, 174 pp.
4. U.S. Environmental Protection Agency. EPA/OMS MOBILE5 Vehicle Emission Modeling Software: MOBILE5b; URL: <<http://www.epa.gov/oms/m5.htm#m5b>>, 1996.
5. Pierce, T.; Geron, C.; Bender, L.; Dennis, R.; Tonnesen, G.; Guenther, A. "Influence of increased isoprene emissions on regional ozone modeling," *J. Geophys. Res.* **1998**, *103-D19*, 25611-25629.
6. U.S. Environmental Protection Agency. *EPA Third-Generation Air Quality Modeling System, Models-3, Volume 9b: User Manual*; U.S. EPA, Research Triangle Park, NC, and U.S. EPA Systems Development Center, Arlington, VA, 1998; Rep. EPA 600/R-98/069(b), 420 pp.
7. U.S. EPA, *EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b: User Manual*, U.S. Environmental Protection Agency, Research Triangle Park, NC, and U.S. EPA Systems Development Center, Arlington, VA, 1998, Rep. EPA 600/R-98/069(b), 420 pp.
8. Briggs, G. A. "Plume rise and buoyancy effects," In *Atmospheric Science and Power Production*; Randerson, D., Ed.; Weather Service Nuclear Support Office, National Oceanic and Atmospheric Administration; Published by the Office of Scientific and Technical Information, U.S. Department of Energy, Oak Ridge, TN, 1984; DOE/TIC-27601 (DE84005177).
9. Gery, M. W.; Whitten, G. Z.; Killus, J. P.; Dodge, M. C. "A photochemical kinetics mechanism for urban and regional scale photochemical modeling," *J. Geophys. Res.*, **1989**, *94*, 12925-12956.
10. Stockwell, W. R.; Middleton, P.; Chang, J. S.; Tang, X. "The second generation regional acid deposition model chemical mechanism for regional air quality modeling," *J. Geophys. Res.* **1990**, *95*, 16343-16367.
11. U.S. Environmental Protection Agency. *Draft User's Guide to PART5: A Program for Calculating Particle Emissions for Motor Vehicles*; Office of Air and Radiation/Office of Mobile Sources, U.S. EPA, Ann Arbor, MI, February 1995; Rep. EPA-AA-AQAB-94-02, 47 pp.

**Figure 1.** The relationship between SMOKE version 1.0 and the Models-3 Framework.



**Figure 2.** SMOKE version 1.0 programs and their interaction.

