

# EMISSIONS MODELING FOR SMRAQ: A SEASONAL AND REGIONAL EXAMPLE USING SMOKE

MARC R. HOUYOUX, CARLIE J. COATS, JR., ALISON EYTH, AND SUNNY LO  
MCNC Environmental Programs  
P.O. Box 12889  
Research Triangle Park, NC 27709-2889

## ABSTRACT

The Seasonal Model for Regional Air Quality (SMRAQ) project is an effort to model ozone formation and transport from mid-May to mid-September in 1995 for the Eastern two-thirds of the continental United States. The modeling grid is 48 columns by 50 rows at a 54 kilometer resolution. The models used for the project are the Multiscale Air Quality Simulation Platform (MAQSIP) for ozone chemistry and transport, MM5 for meteorology, and the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system for emissions. The emissions inventory was derived from the Ozone Transport Assessment Group (OTAG) inventory for 1995.

This paper describes the experiences and insight gained from emissions processing for the SMRAQ project. The huge emissions inventory required by the regional and seasonal features of SMRAQ unfavorably impacts emissions processing time, disk storage, and computer memory requirements. Additionally, the regional domain forces consideration of time zones, gridded meteorological data (such as temperature and wind), and other spatial data (such as land use and climate). These issues are described here in relation to the approach used for SMRAQ emissions processing.

The improved computational efficiency provided by the SMOKE modeling system, which uses sparse matrix algebraic techniques to reduce both computational and storage requirements, permits efficient emissions processing for seasonal modeling of regional ozone formation and transport. Comparisons between SMOKE and EMS-95 are described that indicate SMOKE requires 32 times less CPU resources than EMS-95 for a five-day case based on the OTAG 1990 inventory. For the 120 day SMRAQ episode, SMOKE is estimated to be able to process all emissions on an IBM 590 workstation in 26.7 CPU hours.

## INTRODUCTION

Urban and regional air quality modeling is widely used in decision making for environmental policy. The scale of early modeling efforts was focused on urban regions, perhaps 100 by 100 kilometers and on episodes of several days duration. Such efforts suggested that such modeling could not adequately account for regional meteorological phenomena that could potentially have large impact on ozone pollutant levels. Consequently, regional modeling efforts have been undertaken in which the modeling domain covers several states and the episodes have longer duration. Most recently, the Ozone Transport Assessment Group (OTAG) has focused on air quality modeling that includes 37 states, the District of Columbia, and parts of Canada. Episodes of one to two weeks have been run for four base years (1988, 1991, 1993, and 1995), for 1990, and for several strategies in the year 2007.

Although efforts such as OTAG can help answer many questions, one open question is the ability to model ozone formation and transport over the course of the entire ozone season. Such is the goal of the Seasonal Model for Regional Air Quality (SMRAQ) project sponsored by the Southeastern States Air Resource Managers (SESARM). The motivation for such an effort has been driven by the Southern Oxidant Study (SOS), a coalition of science communities and air quality managers concerned with air quality issues in a ten-state region of the Southeast [1].

For the SMRAQ project, ozone formation and transport is being modeled from mid-May to mid-September in 1995 for the Eastern two-thirds of the continental U.S. The modeling grid (Figure 1) is 48 columns by 50 rows at a 54 kilometer resolution using a Lambert projection, and it includes the same states and covers approximately the same region as the OTAG modeling domain. Other modeling grids may be used during the project in which sub-regions of the domain have finer meshes that would be

“nested” in the coarse grid shown in Figure 1. The chemistry mechanism used in the air-quality model is Carbon Bond-IV.

The seasonal feature of the SMRAQ project raises many interesting issues that have affected the selection of modeling tools. For example, experience from other regional modeling projects indicates that correctly executing the necessary models for a week long episode can take several weeks or even months of human and computer time. The SMRAQ effort seeks to complete a much longer modeling episode, but to do so in relatively the same amount of project time as these shorter efforts. The answer to such concerns lies in choosing fast and flexible computational tools that can be extensively automated, while maintaining or surpassing the level of accuracy for current air-quality modeling approaches.

The tools used for SMRAQ were selected with these facts in mind. The models used for the project are the Multiscale Air Quality Simulation Platform (MAQSIP) for ozone chemistry and transport [2], MM5 for meteorology [3], and the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system for emissions [4,5,6,7]. The SMOKE system as applied to the SMRAQ project is the focus of this paper. First, we describe SMOKE, followed by a summary of the emissions inventory used for SMRAQ. Next, we address emissions processing issues arising from the regional and seasonal nature of this work. Finally, we compare emissions processing of SMOKE to that of the Emissions Modeling System -95 (EMS-95)[8], and describe the emissions experiences in SMRAQ from a computational resources perspective.

## EMISSIONS PROCESSING

### Background

Emissions are a major and essential input to air-quality models. These emissions either take the form of anthropogenic (man-made) emissions or biogenic emissions. Anthropogenic emissions processing can be described as the conversion of inventory emissions to the grid and temporal resolution expected by the air-quality model. The sources are defined based on the type of activity that creates the pollutant. Biogenic emissions processing can be defined as converting land use information, which quantifies the amounts and types of plant life in a region, to emissions of model species. The anthropogenic sources are divided into three primary types: point, area, and mobile sources. Because these are the sources that can be regulated, they get the most attention during emissions processing.

The emissions processor must have each source uniquely defined for processing. *Point sources* usually represent specific stationary waste stacks coming from specific industrial activities such as chemical manufacturing or utilities. Consequently, point sources are typically defined for emissions processing in terms of their county, process, plant, and stack. *Area sources* usually represent all of the emissions of a particular type in a specific region, for example, an estimate of the total dry-cleaning emissions for a county. They are defined by the county and the process identifiers. The term *mobile sources* is often used to categorize the emissions from motor vehicles, but the notion of a source is difficult to define in this case. One reason for this difficulty is that the raw data used in mobile emissions processing is not emissions, but vehicle-miles-traveled (VMT), which is stored for either an entire county, or a section of road known as a “link.” Mobile sources are therefore considered to be either *link sources* or *non-link sources* (alternatively: *on-net* and *off-net*), the primary difference being how the final emissions get allocated on the modeling grid. During emissions processing, the mobile-source VMT is converted to what is essentially inventory pollutant emissions using *emission factors*, which typically are created by running the MOBILE5 model.

The standard emission models used today are the Emissions Preprocessor System 2.0 [EPS 2.0] [9], the Flexible Regional Emissions Data System (FREDS) [10], and EMS-95. The goal of each of these systems is to convert the source-level emissions to hourly emissions of model-species in each grid cell. Such conversions consist of multiplying several factors to the various sources in steps called *temporalization*, *speciation*, and *gridding*. An additional *apply controls* step can exist if emissions values are being reduced using control factors that represent actual ways in which emissions could be reduced. These steps are the components of emissions processing, and they are typically performed separately for point, area, and mobile sources.

The performance of these models is inefficient because of the way in which the conversion factors are applied. At each stage, a given processor tries to find the properties of each source (e.g., state/county code, process code) in a *cross-reference table*, which lists the source properties with an associated *profile number*. When a match is made, the profile number is used as an index into a *profile table*, which provides the conversion factors to be used for the source.

For the standard emissions systems, this “lookup” process is inefficient for several reasons. First, the sequential approach typically used does not adapt well to modern workstation or supercomputing architectures. Moreover, the lookups are often based on matching character strings (e.g., the process code is stored, and therefore matched, as character string ‘0005’ instead of the number 5), which is a process approximately 100 times more computationally expensive than integer matching. Also, algorithms are used that perform an enormous amount of redundant processing. For examples we have seen, the factors for gridding a given area source are typically recalculated up to 360 times for each day of each control strategy modeled.

Many other disadvantages of these systems exist as well. One that can be a nuisance for a large modeling project is the use of machine-specific binary output-file formats, making it difficult to use more than a single computing platform. Large modeling efforts are likely to require several different computing platforms, each suited for specific tasks. If the model-ready emissions inputs can only be used by certain computers, modelers cannot take full advantage of these heterogeneous computing environments.

### **SMOKE Modeling System**

The SMOKE system available at the time of this writing is a working prototype, which has been developed by MCNC’s Environmental Programs group with funding from a cooperative agreement with the Environmental Protection Agency (EPA). Results from SMOKE have been validated by comparisons with both EPS 2.0 and EMS-95. The EPS 2.0 comparisons were performed on the domain used for North Carolina’s State Implementation Plan modeling, and the EMS-95 comparisons have been made on the OTAG 1990 and 1995 basecases. Further modifications are expected based on the SMRAQ modeling, which will help advance SMOKE’s capabilities. Even given its current prototype status, various features make it an excellent choice for urban, regional, and seasonal emissions processing.

The key to SMOKE’s performance is the sorting of the emission sources into a predetermined, fixed order. This organizing step lets SMOKE avoid the redundant lookups that plague other systems. To add further efficiency, SMOKE uses integer-based tables and efficient binary searches rather than character-string based tables and linear-search lookups.

In SMOKE, each sorted list of sources (one each for point, area, and mobile) is treated as a vector of emissions values. The location of each source in either the point sources, area sources, or mobile sources vector is its source number. Other inventory parameters belonging to the sources, such as state/county codes, process codes, and plant codes, are stored in source-attribute vectors. During the emissions processing, the various sets of factors are organized into matrices used to operate on the emissions vectors. For example, matrices are created to apply the speciation factors and gridding factors. The matrices that result have only a few non-zero entries per row, and the storage and multiplication algorithms take advantage of this fact. SMOKE gets a substantial further improvement by multiplying these matrices together only once (or at least only once per control strategy) rather than using them separately and repeatedly.

The basic components of the SMOKE modeling system are the area, biogenic, mobile, and point source submodels and the merge processor, as indicated in Figure 2. Each component submodel tends to follow the organization indicated in Figure 3:

- **Data structuring:** Import and sort the inventory into some “standard” order, making annualized inventory totals and source attributes into column vectors { emissions, county, process, ... } (this needs to be done once per inventory);
- **Temporal model:** Compute hourly column vectors { emissions<sub>t</sub> } of the inventory pollutants for each episode (once per episode; may be factor based or may be a “real model”);
- **Gridding model:** Compute gridding matrix  $G$  (once per grid);
- **Speciation model:** Compute speciation matrix  $S$  (once per chemical mechanism); and
- **Control model:** Compute control matrix  $C$  for this control strategy (once per control strategy).

Referring back to Figure 2:

- **Merge processor:** Compute the product matrix  $T_B = G S$  for the base case or  $T_C = G S C$  for a control strategy. Then compute and accumulate all four categories of model-ready emissions {  $\sum T_C$  emissions<sub>t</sub> } for each hour of each episode of each control strategy.

The exceptions to this pattern of submodel structure are the biogenic and mobile submodels, and (to a lesser extent) the plume rise treatment in the point source submodel. Additional detail on all of the

submodels has been provided in other documentation [4,5,6,7]. Here it will suffice to say that the processing approach described above realizes large performance improvements, and such performance details for the SMRAQ modeling effort will be described later in this document.

An additional feature of SMOKE is that it uses the Models-3/EDSS I/O API [11,12] file format. This format permits architecture-independent binary files. These files are portable to all of the major UNIX computing platforms, without needing a conversion step. The binary storage makes the file size smaller than the same data stored in ASCII or SAS data formats. The I/O API format also allows SMOKE to take advantage of a unified set of analysis and visualization tools designed specifically for this format.

## **EMISSIONS INVENTORY**

As part of the OTAG project, an emissions inventory was developed for 1995 by E.H. Pechan and Associates (EHP). Originally, SMRAQ participants planned on directly using the 1995 OTAG inventory, however, problems in this inventory have led to some deviations from that original course. The SMRAQ work is now using an improved version of this inventory, created by repairing flaws in the original inventory found as part of the OTAG modeling effort. In addition, some sources are also further modified from this improved OTAG inventory to be more suitable for seasonal modeling, and plans exist to investigate the importance of other possible changes. This section summarizes the 1995 OTAG inventory, and then describes modifications for the SMRAQ project.

### **OTAG Inventory**

The states represented by the 1995 OTAG inventory have a western boundary consisting of Texas, Oklahoma, Kansas, Nebraska, North Dakota, and South Dakota. The inventory is for these states, all states east of these, and some portions of Canada. When the inventory was built by EHP, first it was initialized with the 1990 national inventory, and then the emissions values were grown to estimated values for a 1995 average summer weekday.

Point source emissions were grown using Bureau of Economic Analysis (BEA) economic earnings data for all point sources except those sources that could be matched to records in the Department of Energy's (DOE) fuel-use database. For these sources, which represented approximately 55 percent of the utility sources, a ratio of *estimated* 1995 fuel-use data and 1990 fuel-use data was used to grow the emissions values. The 1995 fuel-use data was estimated from data for previous years because the 1995 fuel-use had not yet been released by DOE. In addition, the fuel-use adjustments converted these utility emissions to an average-July-weekday estimate, instead of an average-summer-weekday. The approach for these sources is intended to more accurately estimate the utilities' emissions, because the utilities contribute a significant portion of the NO<sub>x</sub> emissions in the point sources inventory.

Area and mobile source data were grown in a more straightforward way. All area source emissions were grown using the BEA economic earnings data. The mobile source VMT were grown from 1990 values to 1995 values based on the ratio of 1995 Highway Performance Monitoring System (HPMS) data to 1990 HPMS data for those states that supplied the needed data. For other states, 1994 HPMS data was used, with adjustments made to reflect changes from 1994 to 1995. Additional detail on the OTAG inventories can be obtained from [13].

### **SMRAQ Inventory**

Some features of the OTAG inventory focus on an average-July weekday or on an average-summer weekday, but because the SMRAQ episode spans May through September, such an inventory may need modification to suit this larger time span. For example, the inventory could be modified such that month-specific, average-day emissions estimates are used for each month. The ability to make such modifications is limited, however, by the availability of data suitable for making such adjustments. In addition, equally strong are the arguments that (1) an average-month-weekday approach is more representative of reality, given accurate adjustment factors, or (2) such an approach will create undesirably abrupt input changes for the air-quality model, suggesting that average-season-weekday emissions are more useful. Given these considerations, an initial approach of the SMRAQ work has been adopted, which is described in the following paragraphs.

**Point sources.** The July-specific portions of the OTAG inventory are the least appropriate for use in a seasonal episode; therefore, emphasis has been given to applying the fuel-use information from EHP and DOE to remove the July feature of the inventory. The current approach is to retain the average summer weekday sources and to adjust the July-specific sources to each month of the seasonal episode. This

approach maintains the original intent of using the 1995 OTAG inventory, but in a way consistent with seasonal modeling.

No hour-specific emissions are being used in the base SMRAQ modeling. For those sources which were hour-specific for OTAG, the inventory has been updated to ensure that the daily emissions in the inventory files are correct. Future SMRAQ emissions inventory approaches will consider incorporating the DOE's Continuous Emissions Monitoring (CEM) data and investigating the accuracy of the temporal profiles when compared to such data.

**Area sources.** No modifications are planned for the average-summer-weekday inventory. Two reasons have contributed to this approach. First, it seems wise to limit the complexity of the inventory for the initial SMRAQ modeling. Given the exploratory nature of the project, future work may seek to investigate the impact of modifying the area source inventory on a monthly basis. Second, little reliable data are available for supporting a month-specific approach for area sources.

**Mobile sources.** The mobile source raw input files used here are taken directly from the 1995 OTAG basecase. These inputs are the VMT per county, the VMT mix for each county and road class, the speeds for each county and road class, and the MOBILE5 inputs used for creating the emission factors. The VMT data are based on an average summer day.

Consideration has been given to removing the average-summer-day adjustment, and then applying a month-specific adjustment. We have obtained the appropriate factors for modifying the average-summer-day VMT to an average-annual-day basis. We have also obtained factors from several SMRAQ states that can be used to create month-specific VMT estimates; however, most states are not represented by these data. Consequently, our current approach will be to use the average summer day VMT values for the actual modeling, and examine the impact of month-specific VMT using a sensitivity study. The MOBILE5 emission factors are calculated using a SMOKE processor that drives the MOBILE5 model. Version 5a.1 of MOBILE5 is used here, and this is the same version as used for the OTAG mobile processing. As in OTAG, both diurnal and non-diurnal emissions are calculated and then aggregated for each hour of the episode.

Mobile source processing requires consideration of temperature. Because of the large spatial extent of the modeling domain and long duration of the seasonal episode, gridded and hourly temperature inputs will be used by SMOKE. The temperatures will be taken from the MM5 simulations (run primarily for the air-quality model), and they will be interpolated to values representing the temperature at 1.5 meters above the surface. Diurnal emissions will be based on the minimum and maximum temperature for each day in each county, as part of the SMOKE processing.

**Other inventory features.** The cross-references and profiles used for gridding, temporalization, and speciation have been created based on those used for OTAG. The gridding surrogates, of course, are entirely different because of the different modeling domain, projection, and grid scale. Instead of the ~12 kilometer lat-lon projection of surrogates used for the OTAG domain, the SMRAQ domain requires 54 kilometer Lambert projection of surrogates.

Using the definitions of sources provided in the background above, the approximate number of sources for the point source, area source, and mobile source inventories are 123 thousand, 237 thousand, and 21 thousand, respectively. These values are relevant to SMOKE emissions processing because they determine the length of the vectors used to store and process the emissions data, which in turn affects the processing performance and memory requirements.

## **REGIONAL AND SEASONAL MODELING ISSUES**

Most of the conventional experience in emissions modeling has been for single- or several-day episodes for urban-scale regions. There is some experience with studies, such as OTAG, that are regional and have episodes lasting up to a couple of weeks. The focus on these types of scenarios has led to a number of common assumptions which are no longer viable when emissions processing is performed at regional scales spatially or at seasonal scales temporally. Emissions processing systems such as SMOKE, must therefore contend with the following issues in order to fully address the potential needs of regional and seasonal modeling efforts.

**Time zones.** The temporal profiles are meant to be applied to sources in terms of local time. For example, if a motor-vehicles temporal profile indicates that morning rush hour should begin at 8 AM, then it must begin at 8 AM in all time zones in the modeling domain. Since the hours in the emissions output

file can be defined only by a single time zone, the emissions processing must take into account the time zone of each source when applying the temporal profiles.

SMOKE addresses these needs by using a time zone cross-reference file, which associates a particular state or county with its time zone. Most states are entirely within one time zone, and therefore require only a single record in this cross-reference. Some states are divided by two time zones, but most of the counties in the state are entirely within a single time zone, requiring each state and county code (for that state) to have a record in the cross-reference file. In a few rare cases, a county is divided by a time zone, in which case the user must choose for which time zone the county will be assigned during input file preparation.

**Vertical resolution.** Earlier modeling efforts have typically used coarse vertical resolutions, but the accuracy of this approach for regional applications has been questioned. As a result, the SMRAQ effort plans to use between 22 and 30 vertical layers during the modeling. Doing so affects the emissions processing, because point sources are normally treated as having a pollutant plume, which is allocated to one or more of the vertical layers of the domain. MAQSIP requires a three-dimensional, gridded emissions input file. This means that the plume rise calculations must be done as an emissions processing step prior to input into the air-quality model. Such an approach has two advantages: (1) it enables quality assurances of the plume rise calculations before the air-quality modeling, and (2) it permits the plume rise calculation to be done once per inventory and meteorology data set, instead of being calculated for every run of the air-quality model.

For the SMRAQ work, SMOKE computes plume rise for each source, using a modified Briggs algorithm. SMOKE has the ability to treat all sources as potentially elevated, and because of the fine vertical resolution of the modeling grid, such an approach has been taken here. SMOKE generates an hourly, layered file containing the percentage of each source's emissions assigned to each layer, which is used to create the three-dimensional file needed by MAQSIP. This method actually allocates most of the sources entirely to the lowest modeling layer, but more sources extend above the lowest layer than would have if elevated sources were set using the typical plume-rise-cutoff approach.

**Meteorology.** Various meteorological conditions serve as input to some emissions processing stages. For point-source emissions, plume rise is affected by several parameters including temperature, wind speed and direction, and boundary layer height; mobile-source emissions are affected by temperature; and biogenic emissions are affected by temperature and solar radiation. These meteorological features vary widely over regional domains and seasonal episodes, which has led to emissions modeling using the same meteorology data as are used for the air-quality modeling. When this hourly, gridded data are used, emissions processing becomes much more complex than the urban-scale approach of a single temperature per hour.

For SMRAQ, SMOKE uses gridded, hourly meteorological data for elevated point sources, mobile sources, and biogenic sources. In order to treat all sources as (potentially) elevated, a processor in SMOKE uses the meteorological data to calculate the fraction of the plume in each layer. This processor is required once for each emissions inventory and meteorological file. For all practical purposes, it can be considered a meteorology processor.

In the mobile-source processing, the gridded, hourly temperatures are first interpolated from the original meteorological data to temperatures at 1.5 meters. Then, the temperatures are aggregated from several grid cells to a temperature for each source. In other words, these "source temperatures" are based on the temperatures of all grid cells intersecting the source county or link, using the amount of each grid cell intersecting the source as a weighting factor. In practice, the mobile gridding step creates an "ungridding" matrix, which is later applied to the gridded temperature data in order to calculate hourly temperatures for each source. More details on the mobile source treatment of temperatures can be found in [6]. For SMRAQ, biogenics processing uses the 1.5 meter, gridded, hourly temperature data for each grid cell directly.

**Seasonality and geography.** Seasonal parameters can be applied that adjust emission values or VMT to be applicable for a specific season of the year. These factors attempt to account for variations in emissions arising from meteorological differences and behavioral differences generally defined by the four seasons of the year. For example, utility emissions may have seasonal factors which indicate emissions should be greater during the summer and winter months, when cooling and heating systems draw more electricity. Another example is VMT increasing in coastal communities because of the greater traffic caused by vacationers. A third example is biogenic emissions using seasonal emission factors for estimating emissions.

Assuming that this is a valid approach, a problem arises in attempting to define in which season a specific geographic region falls for a specific date. All would agree that the month of May is a spring month. But would all agree that May in Florida is a spring month? Perhaps it is a summer month? Clearly, the definition of season depends on geography. Furthermore, the summer adjustment factors for one region may not be valid for another. These issues are currently not addressed explicitly by any existing emissions processing systems, including SMOKE. If seasonal factors are to be rigorously applied for seasonal emissions processing, future emissions modeling efforts must consider these issues.

**Process and file management.** Traditionally, air-quality simulations, and therefore the emissions processing steps, have been processed on a day-by-day basis. This means that all components of the emissions processing system must be run repeatedly for each different day of the modeling episode. For longer episodes, especially for seasonal emissions processing, this results in an extremely large number of individual executions of the many separate programs that make up emissions processing systems. For example, using SMOKE with this approach (and SMOKE already has many fewer steps than other systems), there are 11 one-time processors, and 11 other processors per day. For the 120 day episode, this would result in 1331 processing steps. In addition to keeping track of these processing stages, the many intermediate files would have to be maintained as well.

Ideally, an entire seasonal episode could be modeled in just a few periods - perhaps by months or weeks. However, most operating systems limit file sizes to two gigabytes, and this limit can be reached for intermediate and output files containing emissions for more than several days. Another factor limiting the period length is the desire to quality assure the initial days of the simulation before expending resources on the latter days. Given these considerations, emissions processing (and meteorology and air-quality simulation) for the SMRAQ work is done using five-day periods. This approach will reduce the number of individual emissions processing steps to 275, which is significantly better than 1331. Not only are the number of steps reduced, but fewer time-dependent intermediate and final files will be created.

In addition to the five-day time period, emissions processing for SMRAQ is taking advantage of the Environmental Decision Support System (EDSS) Study Planner [12,14] for process- and file-management. The use of this system for SMRAQ is described in another paper from this conference [15].

## **EMISSIONS PROCESSING RESOURCES**

This section describes the resources which are currently required by SMOKE for SMRAQ emissions processing. In order to give some context for this description, we first compare the requirements and performance between EMS-95 and SMOKE for the OTAG 1990 basecase. The version of EMS-95 used for comparison has already been optimized several times during the OTAG modeling. After this analysis, the resources required for SMRAQ emissions are presented and discussed. All performance numbers were compiled with an IBM 590, which for both systems was fully dedicated to the emissions processing. All file output and SAS temporary space used local disks.

Table 1 summarizes the comparison of resources between EMS-95 and SMOKE. Data on CPU time, wall clock time, and file sizes have been compiled in three stages: processors run once per episode, once per day, and for a hypothetical five-day case. The five-day case is used to give a better sense of the total processing requirements for a given scenario, and it more easily permits comparison with numbers compiled for the five-day SMRAQ processing statistics. The numbers used to compile the totals, however, are taken directly from processing of the OTAG 1990 inventory. In the actual OTAG 1990 basecase, EMS-95 was run for three days for point and area sources, and for every day of each scenario for mobile sources. The SMOKE system was run using a cutoff for elevated point sources versus low-level point sources, in the same manner as EMS-95, instead of using the method described above for which all point sources are potentially elevated. While the emissions outputs from the two systems are essentially the same, SMOKE's performance is clearly superior in all respects.

The CPU requirements for the two systems show significant differences. In all cases, SMOKE outperforms EMS-95 by a wide margin. For the hypothetical five-day case, point source processing is 15 times faster, area sources 18 times, and mobile sources 60 times. Overall, SMOKE requires 32 times less CPU resources than EMS-95. One major portion of the mobile source one-time processing total, creating the emissions factors, has been excluded because these EMS-95 performance numbers were not available for the IBM. Both EMS-95 and SMOKE drive the MOBILE5 program to get these factors, but assumptions made to do so result in very different performance. Experience using a DEC Alpha indicates that EMS-95 takes 24 to 48 hours to generate its emission factors for 1990, whereas SMOKE takes about 40 minutes.

The wall clock times show an even greater advantage for SMOKE. Again for the five-day case, point sources are processed 27 times faster, area sources 38 times, and mobile sources 68 times. The reason is that the SAS-based EMS-95 system is slowed by constant use of temporary SAS files. The CPU utilization of the most time consuming EMS-95 processors were typically in the range of 30 to 50 percent, suggesting that the processors spend a great deal of time waiting for I/O. In contrast, the CPU utilization of the SMOKE processors is typically about 75 to 99 percent.

The file size requirements are also better for SMOKE. For the five-day example, SMOKE requires 63.7 percent less disk space than EMS-95 for point source files, 79.8 percent for area sources, and 64.7 percent for mobile sources. Overall, SMOKE requires 70.1 less disk space for files. In addition, SMOKE does not require the 2-5 GB of disk space used by EMS-95 as temporary SAS work-space.

Table 2 summarizes the resource requirements for the SMRAQ domain using the 1990 OTAG inventory. These values differ from the OTAG totals for several reasons. In general, some file sizes and processing times are smaller because there are many fewer grid cells in the SMRAQ domain than in the OTAG domain. The emission factor generation numbers have also been added to the mobile sources once-per-scenario values. The point source intermediate and output file sizes are larger than for OTAG because of the approach taken for elevated sources. The layer fractions file contains plume fractions for all sources, for each layer, and for every hour, and such a large amount of data results in a 142 MB file per day. The process time for point sources is also affected. Generating the layer fractions file takes 7 minutes of the total 11 minutes per day. In addition, the point source output file sizes include the two-dimensional and three-dimensional gridded emissions output files. The input file totals do not include the meteorological data file sizes.

This case has been run for testing and QA purposes only, and it is not an actual scenario being used for the SMRAQ project. Additionally, the five-day case remains hypothetical, because as for OTAG, all area source emissions and non-elevated point sources emissions must be processed for only three days (weekday, Saturday, and Sunday).

Using the 1990 performance data, we have developed estimates of the computing resources required for modeling emissions for the entire 120 day 1995 SMRAQ episode. Our experience indicates that SMOKE performance for the 1990 and 1995 inventory is very comparable. The following estimates assume that area sources use the weekday/Saturday/Sunday approach of OTAG, and therefore require only three days of individual area-source processing. Point sources use this approach as well, but for three days of each of the 5 months; however, the elevated plume rise fractions must be calculated for all 120 days. The mobile, biogenic, and merge steps require processing for all 120 days.

Our current estimate of required CPU resources using SMOKE is 26.7 hours, which includes all processing for all four source categories; meteorology-based, fractionated, hourly plume rise for all sources; and merging to the final 22 layer emissions input file. The associated wall-clock time is 29.1 hours. The disk space requirements to store all input, intermediate, and output files are 40 GB, but of course, this data will be compressed and archived through the course of the project, and it will not actually be necessary to have all output data available at all times.

Using EMS-95 under the same assumptions, we estimate emissions processing for this episode would take 530 hours of CPU time and 739 hours of wall-clock time. These values do not include the biogenics processing, the generation of the emission factors, the merging, or the fractionated plume rise. The file storage is estimated to require 54.5 GB of disk space, again assuming no data compression.

The primary disadvantage incurred by the modeling approaches in SMOKE is that the processors have large computer memory requirements when run for the entire 37 state domain. The maximum required memory for the executables used in the OTAG 1990 evaluations for the processors is 93 MB for point, 49 MB for area, and 304 MB for mobile. These requirements were reported by the UNIX command "size." In practice, computers with 128 MB of RAM are sufficient to run the processors, because they can make use of disk space (swap space) to account for the additional memory needed. Computers with less memory can be used if the inventory is processed in two or three sections.

## CONCLUSIONS

Seasonal emissions processing for regional ozone formation and transport requires consideration of issues not necessarily a part of modeling for other types of episodes. These include not only the technical issues such as time zones, vertical resolution, meteorology, and seasonality, but also the computational issues such as process management, file management, CPU resources, and disk space resources. Addressing the technical issues is meaningless unless the computational resources are reduced to the extent that seasonal processing can be accomplished in a reasonable amount of time.

The SMRAQ project provides a means to rigorously apply the innovations built into the SMOKE model. As described in this paper, the work performed so far demonstrates that SMOKE will provide great benefits to the project, and it suggests that perhaps such seasonal modeling could not be accomplished using other emissions processing systems. The performance numbers described above assume that the emissions processing can run smoothly from start to finish, with no unforeseen inventory problems. We do not envision emissions processing with any system ever proceeding so smoothly, because inventory problems are more often the rule than the exception. We do believe, however, that SMOKE has significantly improved on our ability to model emissions for seasonal episodes, and it has done so by addressing both the technical issues and computational issues.

## ACKNOWLEDGMENTS

This research described in this document has been funded wholly or in part by US EPA under the cooperative agreement # CR-822074, "Emissions Modeling Research with High Performance Computing." It has been approved for publication by MCNC. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## REFERENCES

1. "The Southern Oxidants Study." URL [http://www2.ncsu.edu/ncsu/CIL/southern\\_oxidants/index.html](http://www2.ncsu.edu/ncsu/CIL/southern_oxidants/index.html)
2. Odman, M.T., R. Mathur, A. F. Hanna, and D. W. Byun, "Applications of the Multiscale Air Quality SIMulation Platform (MAQSIP) to regional modeling. *Proceedings of the 5th International Conference on Atmospheric Sciences and Applications to Air Quality*, Seattle, WA, June 18-20, 1996.
3. Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note NCAR/TN-389+STR, 1994.
4. Coats, C. J., Jr. "High Performance Algorithms in the Sparse Matrix Operator Kernel Emissions Modeling System." *Proceedings of the Ninth Joint Conference on Applications of Air Pollution Meteorology of the American Meteorological Society and the Air and Waste Management Association*, Atlanta, GA, 1996.
5. Coats, C. J., Jr. and M. Houyoux, "Fast Emissions Modeling with the Sparse Matrix Operator Kernel Emissions Modeling System," *The Emissions Inventory: Key to Planning, Permits, Compliance, and Reporting*, Air and Waste Management Association, New Orleans, LA, September 4-6, 1996.
6. Houyoux, M., "Improved Strategies for Mobile Emissions Processing," *The Emissions Inventory: Key to Planning, Permits, Compliance, and Reporting*, Air and Waste Management Association, New Orleans, LA, September 4-6, 1996
7. Coats, C. J., Jr. "Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System." <http://www.iceis.mcnc.org/EDSS/emissions/index.html>. MCNC Environmental Programs, RTP, NC. 1996
8. Radian Corporation, 1993. GEMAP System Documentation. Sacramento, CA. Draft, May 1993.
9. *User's Guide for the Urban Airshed Model Volume IV: User's Manual for the Emission Preprocessor System 2.0*. Document EPA-450/4-90-007D(R). Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, June 1992.
10. Modica, L. G., D. R. Dulleba, R. A. Walters, and J. E. Langstaff, "Flexible Regional Emissions Data System (FREDS) Documentation for the 1985 NAPAP Emissions Inventory." Document EPA-600/9-89-047. Office of Research and Development, US EPA, May 1989.
11. Coats, C. J., Jr. "The EDSS/Models-3 I/O Applications Programming Interface." <http://www.iceis.mcnc.org/EDSS/ioapi/index.html>. MCNC Environmental Programs, RTP, NC. 1995
12. Coats, C. J., Jr., A. F. Hanna, D. Hwang, and D. W. Byun, "Model Engineering Concepts For Air Quality Models In An Integrated Environmental Modeling System." *Transactions of the AWMA Specialty Conference on Regional Photochemical Measurement and Modeling Studies*, San Diego, Nov. 1993.
13. Koerber, M. Volume II: Emissions Modeling - Base Year Version 1.0 (DRAFT), Lake Michigan Air Directors Consortium, August 1996
14. Fine, S. S.; et al. "EDSS Home Page." <http://www.iceis.mcnc.org/EDSS/> MCNC Environmental Programs, RTP, NC 1995.
15. Smith, W. T., C. J. Coats, Jr., A. Eyth, S. S. Fine, D. Hils, and N. Wheeler. "Managing Processing and Data in Seasonal Modeling of Regional Air Quality," this conference.

**Table 1: Performance comparison between EMS-95 and SMOKE using OTAG 1990 inventory**

	CPU time [min]		Wall clock [min]		File Sizes [MB]			
	EMS-95	SMOKE	EMS-95	SMOKE	EMS-95		SMOKE	
Point:								
1 per scenario:	110	3.8	181	3.8	Input: 77	Input: 77	Intmdt: 33	Intmdt: 33
1 per day:	37	3.2	88	3.8	Intmdt: 145	Intmdt: 59	Output: 126	Output: 126
5 day case:	295	20	619	23	Total: 2,855	Total: 1,035		
Area:								
1 per scenario:	158	3.1	343	3.1	Input: 64	Input: 64	Intmdt: 61	Intmdt: 61
1 per day:	53	3.9	98	3.9	Intmdt: 382	Intmdt: 68	Output: 45	Output: 45
5 day case:	423	23	832	22	Total: 3,419	Total: 690		
Mobile:								
1 per scenario:	95	1.2	221	1.2	Input: 10	Input: 10	Intmdt: 104	Intmdt: 104
1 per day:	256	4.4	350	5.5	Intmdt: 1,548	Intmdt: 127	Output: 45	Output: 45
5 day case:	1,374	23	1,973	29	Total: 2,758	Total: 974		

**Table 2: Performance of SMOKE for SMRAQ**

	CPU time [min]		Wall clock [min]		File Sizes [MB]	
Point:						
1 per scenario:		3.8		3.8	Input: 77	Input: 77
1 per day:		10		11	Intmdt: 33	Intmdt: 33
5 day case:		54		59	Intmdt: 178	Intmdt: 178
					Output: 39	Output: 39
					Total: 1,195	Total: 1,195
Area:						
1 per scenario:		2.2		2.2	Input: 54	Input: 54
1 per day:		2.0		2.0	Intmdt: 36	Intmdt: 36
5 day case:		12		12	Intmdt: 68	Intmdt: 68
					Output: 3	Output: 3
					Total: 445	Total: 445
Mobile:						
1 per scenario:		37		43	Input: 4	Input: 4
1 per day:		3.6		4.2	Intmdt: 102	Intmdt: 102
5 day case:		55		64	Intmdt: 127	Intmdt: 127
					Output: 3	Output: 3
					Total: 756	Total: 756
Biogenics:						
1 per scenario:		0.10		0.10	Input: 5	Input: 5
1 per day:		0.23		0.30	Intmdt: 0.2	Intmdt: 0.2
5 day case:		1.3		1.4	Intmdt: 0	Intmdt: 0
					Output: 1.1	Output: 1.1
					Total: 11	Total: 11

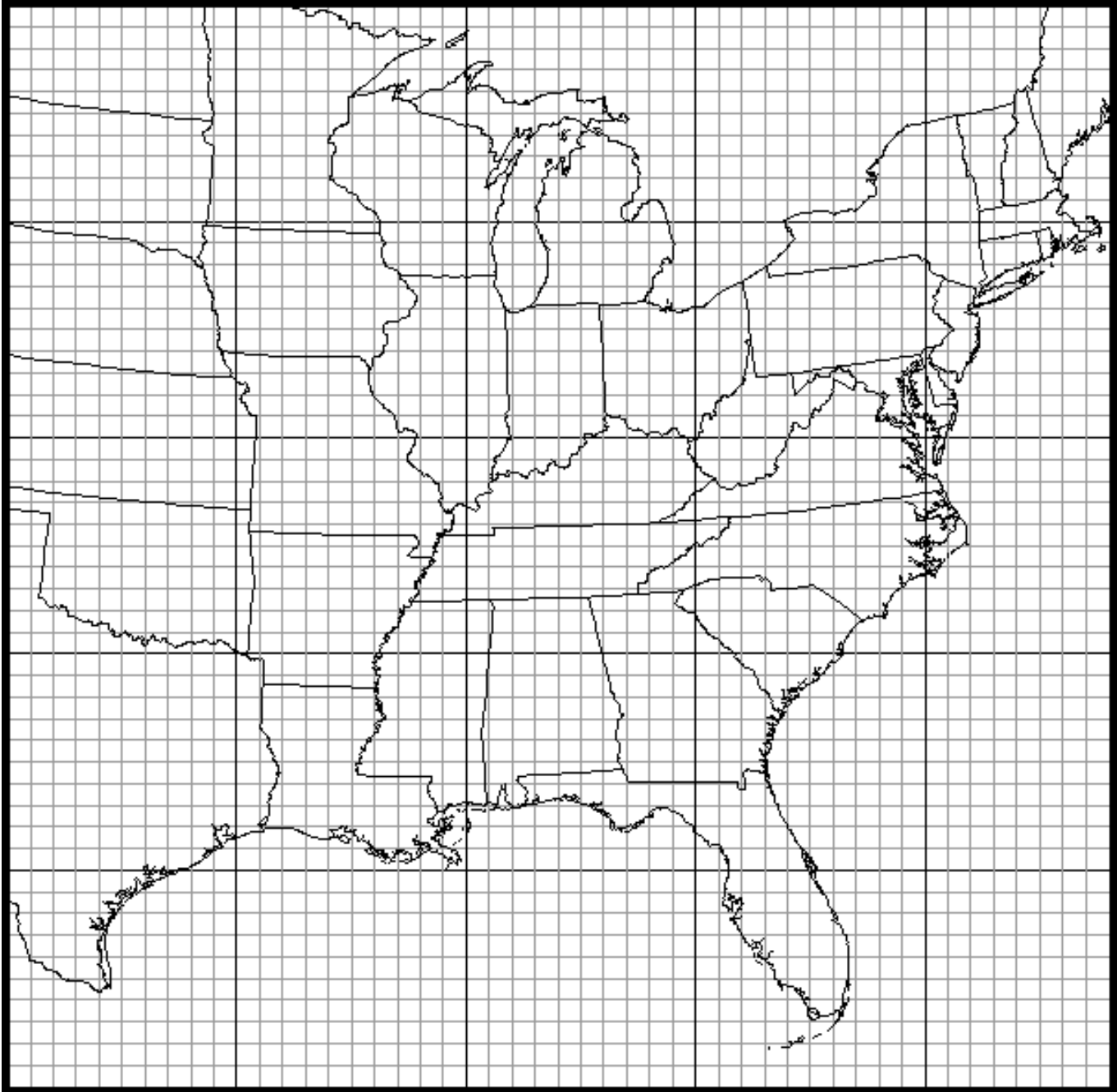
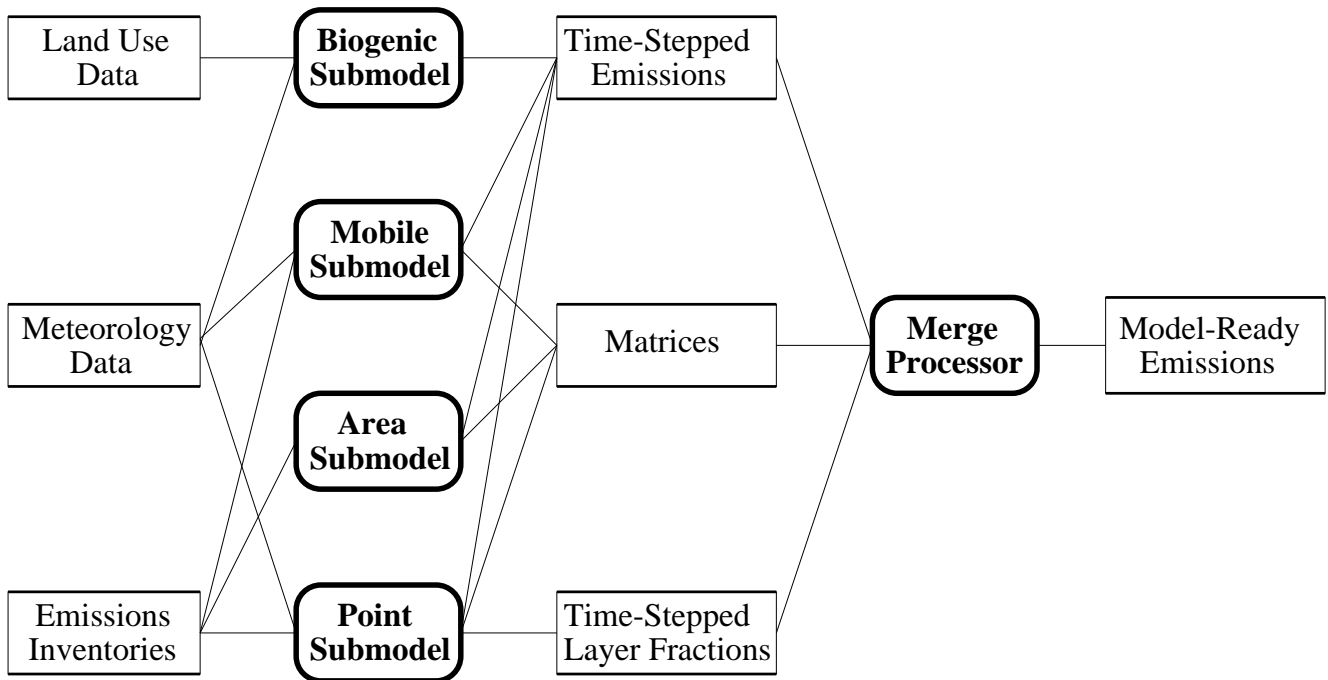
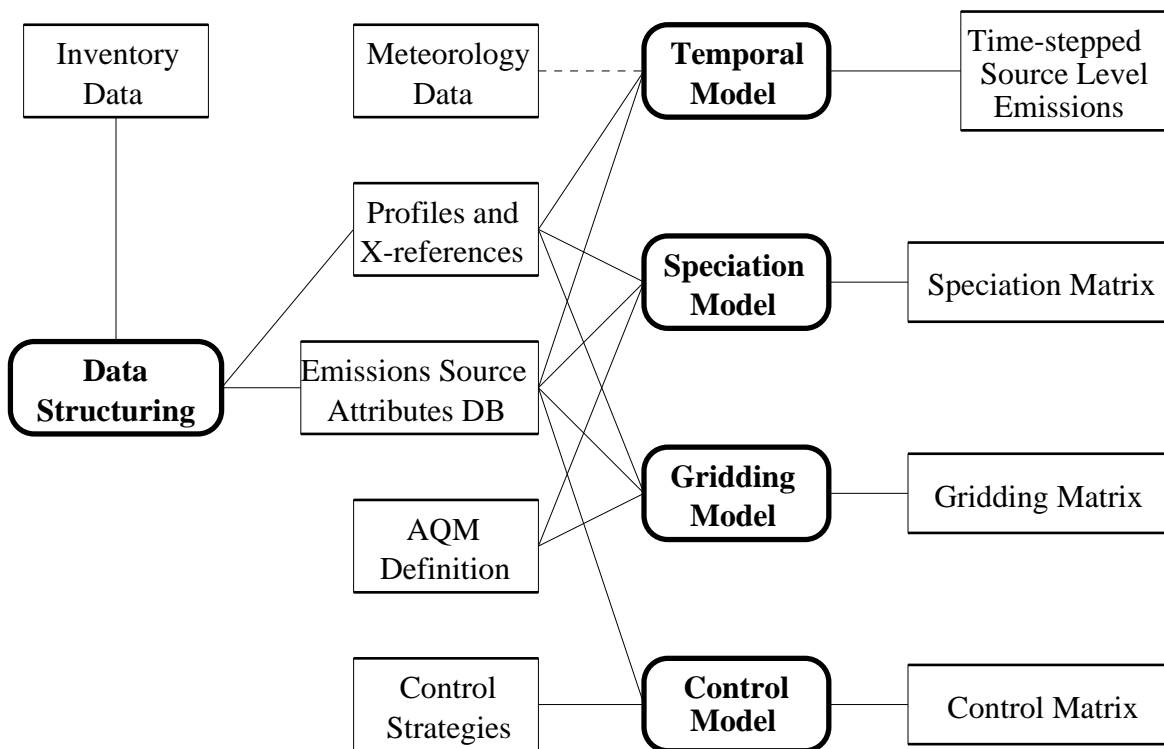


Figure 1: SMRAQ modeling domain



**Figure 2: SMOKE submodels and dataflows**



**Figure 3: SMOKE submodel structure.**