

Quality Assurance Enhancements to the SMOKE Modeling System

Marc R. Houyoux and Zachariah Adelman
MCNC, 3021 Cornwallis Road, RTP, NC, 27709
mhouyoux@mcnc.org

ABSTRACT

The Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system has been updated with new quality assurance capabilities. We have added a module to SMOKE that provides a flexible, user-defined approach for generating reports and summaries for quality assurance of emissions data. The new module, called “Smkreport,” also provides a framework for much more easily adding new types of summaries to meet the needs of the modeling community.

Users of Smkreport can control most of the major features of the emissions summaries, such as: pollutants to include, their order, and units; selection of inventory records by region (e.g., state and county codes), model grid cells, and arbitrary combinations of regions and grid cells; and the basis of emission summaries (e.g., state codes, county codes, source category codes, by hour or daily total, grid cell, road class, or any combination). The emission values in the reports can optionally include gridding factors, speciation factors, and temporal factors, which permits the users to make comparisons of summary emissions at various stages of emissions processing.

The Smkreport module has been used to quality assure emissions data processing of the 1996 National Emissions Inventory (NEI) for a yearlong modeling episode. This paper also describes the quality assurance measures used for this processing.

INTRODUCTION

The SMOKE emissions processing system efficiently processes emissions data using matrix-vector multiplication. It performs the core functions of emissions 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 [*U.S. Environmental Protection Agency (EPA), 1994; U.S. EPA, EPA/OMS MOBILE5 Vehicle Emission Modeling Software: MOBILE5b, available at <http://www.epa.gov/oms/m5.htm#m5b>, 1996*], and it also uses a reorganized version of the Urban Airshed Model–Biogenic Emission Inventory System, version 2 (UAM-BEIS2) [*Pierce et al., 1998*]. For mobile sources, SMOKE can process both emissions data and vehicle miles traveled (VMT) activity data. The efficient processing of SMOKE is in more detail by *Houyoux et al. [1998]*.

Several facts about emission inventories give rise to a number of quality assurance (QA) issues. The data available in emission inventories are collected from state and local agencies throughout the country and are combined by U.S. Environmental Protection Agency (EPA) contractors. Although there are guidelines for developing these data, differences in collection methods are present that can significantly affect emission values. Also, the number of records is large (hundreds of thousands for each source category for a national inventory), and because there are many individual data sets that are combined during emissions processing, many problems can be introduced. Ideally, emission processors would be designed to identify all such problems, but it is not always possible to automate all QA steps. Manual review of QA reports for consistency is an important aspect of QA; consequently, a major function of emission processing systems is to provide reports that users can use to identify potential problems.

SMOKE version 1.3 includes a new SMOKE program, *Smkreport* (pronounced “SMOKE report”), which creates user-defined reports to be used for SMOKE quality assurance. User can create customized

reports by mixing a variety of report configuration instructions that Smkreport has been programmed to use. This paper describes this new SMOKE QA module and provides examples of a number of reports that emission modelers may find useful. It also describes a specific application of Smkreport to the National Emission Inventory (NEI) version 3.11, and that effort is described in more detail by *Possiel, et al.* [2001] and *Adelman and Houyoux* [2001].

DESCRIPTION OF REPORTING INPUT REQUIREMENTS

The critical concept that we used to design Smkreport is that emissions modelers should be able to determine for themselves which reports are most useful for their particular modeling efforts. An additional design concept was that the reporting capability be easy to use despite its flexibility. The solution was to design Smkreport around a "report configuration file," which is an ASCII input file that must contain instructions for generating custom reports. While the user is limited to the instructions that have been programmed into Smkreport, the instructions can be combined in many ways to generate hundreds of different reports. Finally, the reports created by Smkreport are formatted so as to be easily imported into other data analysis tools, such as Microsoft[®] Excel[®] or the freely available Java Analysis and Reporting Tool (JART) [MCNC, Java Analysis and Reporting Tool Documentation, available at <http://envpro.ncsc.org/products/jart>, 2000]

Smkreport scans the report configuration (RC) file to determine the other input files it will read. These other files are SMOKE intermediate files that have been created with other SMOKE processors used for the inventory import, gridding, temporal allocation, and chemical speciation. When an emissions modeler runs Smkreport, he or she uses a script that ensures that these other input files are available to Smkreport. We do not describe in more detail here how Smkreport uses these intermediate files, since this process is fairly hidden from view of the user once he or she is familiar with using the SMOKE scripts. For the purposes of this paper, it is simply enough to state that the SMOKE intermediate files for the various steps of emissions processing are available to Smkreport in addition to the RC file.

The remainder of this section provides a synopsis of the Smkreport instructions that can be included in the RC file, which has three major types of instructions: (1) report-control instructions, (2) group-definition instructions, and (3) report-creation instructions. More details on the input file formats are also available on line at http://envpro.ncsc.org/products/smoke/version1/smoke_files.htm#QA_other.

Report-control Instructions

Table 1 summarizes the control instructions. The SMK_SOURCE instruction should appear only once in the RC file, and it must appear first. It sets the source category to area, mobile, or point sources (biogenic sources can not be quality assured using Smkreport). Since SMK_SOURCE can only be used once, this means that Smkreport can be run only for one source category at a time.

The other report-control instructions are like switches, and they can be used to change report attributes at many places in the RC file. If these instructions are used, they must be positioned between the report-creation instructions, which are initiated by the /CREATE REPORT/ instruction (described below). When a report -control instruction appears in the RC file, all reports after it are affected by the control until the switch is applied again.

The /NEWFILE/ instruction simply indicates that subsequent reports should appear in a new output file. This instruction permits users to arrange reports in any manner they would like. For example, users could choose to place each report in a different output file (the best choice if the reports are to be imported to another program), to group similar reports in separate output files, or put all reports in a single output file.

When the “/OZONE SEASON/ Y” instruction appears in the RC file, Smkreport is instructed to use the ozone season emissions in generating the reports. “/OZONE SEASON/ N” causes Smkreport to switch back to using annual average emissions, which is also the default reporting behavior. These instructions only affect the reports that include emissions from the inventory before temporal allocation, since the SMOKE hourly emissions files have already been created with or without the ozone-season emissions.

Lastly, the “/REPORT TIME/ <value>” instruction permits users to select the “end of the day” when reporting daily total emissions summed from hourly emissions. This can be useful when the time zone label on the hours in the SMOKE hourly intermediate files is not consistent with the location of the emissions sources themselves. For example, the SMOKE intermediate files may need to be in Greenwich Mean Time (GMT) to meet the air-quality model input requirements, but the emissions sources themselves may be in Pacific Standard Time (PST). By using the instruction “/REPORT TIME/ 80000”, this 8-hour difference in the “end of the day” can be considered when creating the daily emissions totals.

Table 1. Control instructions for the Smkreport report configuration input file.

Instruction	Status	Description
SMK_SOURCE <value>	Required	Sets the source category for running Smkreport, which can be run for only one source category at a time. The <value> should be A for area sources, M for mobile sources, and P for point sources. Smkreport does not work with biogenic sources.
/NEWFILE/	Optional	Controls the number and names of output files from Smkreport, and it allows the user to group multiple reports into the same Smkreport output file. The packet is optional because all Smkreport outputs can be written to a single file using a default file name. It must appear before or between /CREATE REPORT/ instructions.
/O3SEASON/ <value>	Optional	Controls whether the ozone-season emissions or the annual average emissions are read from the SMOKE inventory files for use in the reports. The <value> can be Y or N to indicate use of ozone season emissions or use of annual average emissions, respectively. It only affect reports that do not use the TEMPORAL instruction.
/REPORT TIME/ <value>	Optional	Permits changing the hour value for which Smkreport outputs daily total hourly emissions. It only applies when the TEMPORAL report-creation instruction is used in a report without using the BY HOUR instruction. The content of the <value> is hours as HHMMSS (e.g. 130000 is 1 PM).

Group-definition Instructions

Group-definition instructions permit users to create a subset of records from the emission inventory for reference later in creating reports. The group definitions do not generate a report themselves, but rather they create an item for use by a report creation instruction. The current version of Smkreport supports region groups and subgrids, as summarized in Table 2. The syntax for defining both types of groups uses the INCLUDE and EXCLUDE instructions, which are switches that describe how other lines of the group definition are interpreted. If neither the INCLUDE or EXCLUDE instructions appear in the group definition, Smkreport assumes INCLUDE.

Region groups are groups of country, state, and county Federal Implementation and Planning Standards (FIPS) codes. These are useful when users need to create a detailed report with a large number

of report records, but want the detailed information for only a subset of the inventory. The region group would allow the detailed report to be a more manageable size. The region groups are defined starting with the /DEFINE GROUP REGION/ instruction and ending with an /END/ instruction, and the lines between these instructions can contain INCLUDE, EXCLUDE, and state or county FIPS codes. There is no limit to the number of counties or states that can be referenced in a region group.

Subgrids are groups of grid cells. Smkreport supports any groups of grid cell, whether they are rectangular, contiguous non-rectangular regions, or noncontiguous cells or cell clusters. These groups can be referenced when creating reports to subset the inventory and grid cells to specific regions. For example, a “nonattainment regions” subgrid could be created that includes all of the ozone nonattainment regions in a state, by grid cell. The subgrids are defined starting with the /DEFINE SUBGRID/ instruction and ending with an /END/ instruction, and the lines between these instructions can contain INCLUDE, EXCLUDE, single x-y coordinates, and ranges of x-y coordinates such as “(1,5) TO (5,5)”. There is no limit to the number of grid cells that can be referenced in a region group.

Table 2. Group definition instructions for the Smkreport report configuration input file.

Instruction	Status	Description
/DEFINE GROUP REGION/ <label>	Optional	Allows users to group region codes (country, state, and county codes) together and name the group for later reference within a /CREATE REPORT/ instruction. The <label> is the name of the group.
/DEFINE SUBGRID/ <label>	Optional	Allows users to grid cells together and name the group for later reference within a /CREATE REPORT/ instruction. Grid cells need not form a rectangular shape or even be in a contiguous section of the domain. The <label> is the name of the sugrid.
INCLUDE	Optional	Used with either of the group instructions above to include region codes or cells in the group.
EXCLUDE	Optional	Used with either of the group instructions above to exclude region codes or cells from the group.
/END/	Optional	Used to close a /DEFINE GROUP REGION/ or /DEFINE SUBGRID/ instruction.

Report-creation Instructions

Report-creation instructions actually cause Smkreport to create output files. As shown in Table 3, all reporting-creating instructions require the /CREATE REPORT/ instruction to initiate the report-creation section, and require the /END/ instruction to complete the section. At least one BY instruction must appear between these two instructions, but all other report-creation instructions are optional. In addition, except the TITLE and BY instructions, the instructions should be used no more than once per /CREATE REPORT/ instruction.

The BY instruction has a number of options that let the user control which emissions properties Smkreport should used to sum the emissions or activity data. These include BY CELL (sum by grid cell), BY COUNTRY (sum by country and optionally include the country name), BY STATE (sum by state and optionally include the state name), BY SCC10 (sum by 10-digit source category code (SCC)), etc. Table 3 list all of the BY options and describes these in more detail. Users can use any or all BY instructions that are appropriate for a given source category to achieve as fine or as coarse a report as they like. For example, using both the BY CELL and BY SCC10 instructions would create a report with data reported for the intersection of each SCC in the inventory and cell in the domain.

The SELECT instruction gives users several options to create a report for a subset of the inventory. With the SELECT DATA instruction, users can request that specific pollutants, activities, or model species are included in the report. This instruction also provides a mechanism for summing model-species emissions back to their original pollutant total, so that pollutants can be compared before and after chemical speciation. With the SELECT GROUP REGION instruction, users can reference a group (previously defined in the group definition section of the RC file) using a label. Users can also define an “inline” group, in which the country, state, or county code is specified as the label. The data output in the report is limited to the region codes in the referenced group. Finally, with the SELECT SUBGRID instruction, users can reference a previously defined subgrid using the label, or set the label to a range of cells. In this last case, the data output in the report is limited to the cells in the referenced subgrid.

Other report-creation instructions include TITLE, to imbed one or more titles in the report; GRIDDING, to output emission totals after the gridding step; TEMPORAL, to output emission totals after the temporal allocation step; SPECIATION, to output emissions after the chemical speciation step; NUMBER, to set scientific or standard notation and the number of decimal places to include in the data totals; and UNITS, to control the output units.

Table 3. Report-creation instructions for the Smkreport report configuration input file.

Instruction	Status	Description
/CREATE REPORT/	Required	Causes Smkreport to generate a single report. The output file that this report is written to depends on the previous /NEWFILE/ instruction in the RC file. It must be used with at least some of the instructions in the remainder of this table. More than one of these instructions can be used in a RC file.
TITLE	Optional	Allows users to add their own title information to the reports. Can be used any number of times in a single report.
GRIDDING	Optional	Indicates that the report should include emissions or activities after the gridding processing step has been applied.
TEMPORAL	Optional	Indicates that the report should include emissions after the temporal allocation processing step has been applied.
SPECIATION <option>	Optional	Indicates that the report should include emissions after the chemical speciation processing step has been applied. <option> can be “MASS” or “MOLE”, depending on the units desired for speciated emission in the report.
BY <option>	Required	Selects the resolution at which the emission summaries will be reported. For each BY option selected, one or more columns will be created in the report. Any combination of BY instructions are permitted that make sense for a given source category. Multiple BY instructions are applied together, providing finer resolution for the emissions summaries.
where <option> is:		
CELL	Optional	Report includes emissions totals by grid cell.
COUNTRY [NAME]	Optional	Report includes emissions totals by country code and optionally country name.
COUNTY [NAME]	Optional	Report includes emissions totals by county code and optionally county name.
ELEVSTAT	Optional	For point sources, report includes elevated source status (whether the source is elevated, elevated plume-in-grid, or layer-1).

Instruction	Status	Description
HOUR	Optional	Report includes emissions totals by hour
ROADCLASS	Optional	Report includes emissions totals by road class
SCC10	Optional	Report includes emissions totals by 10-digit SCC
SOURCE [STACKPARAM] [NAME]	Optional	Report includes emissions totals by inventory source, and for point sources, can include plant name and stack parameters.
SRGCODE	Optional	For area and mobile sources, report includes emissions totals by surrogate code applied during the gridding phase of emissions processing.
STATE [NAME]	Optional	Report includes emissions totals by state code and optionally state name.
SELECT <option>	Optional	Selects a specific part of the emissions data and/or inventory. Multiple SELECT statements can be included, but only one using each of the options.
DATA	Optional	Specifies emissions pollutants, model species, and/or activities to include in the report. The default is that all applicable pollutants and species will be included in the report.
GROUP REGION <label>	Optional	Selects a part of the emission inventory to include in the report, based on a label assigned to a group defined with the /DEFINE GROUP REGION/ instruction.
SUBGRID <name>	Optional	Selects a part of the emission inventory to include in the report, based on a label assigned to a subgrid defined with the /DEFINE SUBGRID/ instruction.
NUMBER	Optional	Sets the column widths and number of significant figures for the data in the reports.
UNITS	Optional	Sets the units of the data in the reports.
/END/	Optional	Used to close a /CREATE REPORT/ instruction.

Example Report Configuration Instructions

The following are some examples of report configuration instructions. These examples are not a complete list of all possible examples; we intend simply that these examples provide clarification to the instructions described above, and perhaps that they provide templates to readers for their emissions processing and QA efforts. We do not include the SMK_SOURCE in these examples, and unless otherwise noted, the examples can be used for area, mobile, or point sources.

Example 1: Region group example

```

/DEFINE GROUP/ NC without Triangle
INCLUDE
37000    ## NC state code
EXCLUDE
37063    ## Durham Co
37135    ## Orange Co
37183    ## Wake Co
/END/

```

This group is labeled “NC without Triangle” and it includes all counties in North Carolina but not Durham, Orange, or Wake counties (collectively known locally as “the triangle”). Note that the double pound sign can be used as a comment identifier in the middle of a line.

Example 2: Subgrid example

```
# The cells of this subgrid are based on the U.S. 36-km Lambert
# unified grid
/DEFINE SUBGRID/ Southeastern US
INCLUDE
(75,8) TO (121,46)
/END/
```

This example includes grid cells in the (x,y) range from (75,8) to (121,46). Selecting the emissions only within the grid cells would be impossible based on the state and county codes alone, because the cells cut across several states and counties in the domain. Note that comments that have own their own line in the file need only to have a single pound sign to start the comment. This example is somewhat confusing because the grid on which the subgrid is based is not included as part of the example. When SMOKE gridding is run, the intermediate files are created that contain the grid information, and these files are read as input to Smkreport. For clarity, Figure 1 shows the grid on which this example is based.

Example 3: State and county report with daily totals

```
/NEWFILE/ /storage/mydir/smoke/report/myreport.txt
/CREATE REPORT/
TEMPORAL
BY STATE NAME
BY COUNTY NAME
/END/
```

This example will create a report called “myreport.txt” in the directory /storage/mydir/smoke/report. The report will include data from the hourly emissions SMOKE intermediate file. Daily emissions totals will be output to the report, since this is the default program behavior when the BY HOUR instruction is not given. Smkreport will sum the emissions and report them by county, because county is the finest resolution specified with the BY instruction in this example. The county name will be included because the “NAME” option is used with the BY COUNTY instruction. The state code and state name will also be included in the report because of the BY STATE NAME instruction.

Example 4: SCC and county report, after gridding

```
/CREATE REPORT/
GRIDDING
BY SCC10
BY COUNTY
NUMBER F12.4
/END/
```

In this example, the emissions in the output report will be summed by 10-digit SCC and county. The county name will not be included in the report because the NAME option has been left off of the BY COUNTY instruction. The emissions in the report will be adjusted by the gridding factors applied by SMOKE, although the report will not include grid-cell summaries, because no BY CELL instruction appears in the report instructions. Based on the NUMBER F12.4 instruction, the emission values in the report will include 4 decimal places, and the columns will be a minimum of 12-digits wide. Smkreport will adjust the column width to use wider columns if the summed numbers required greater than 12 digits.

Example 5: Using a region group and reporting by surrogate type

```
/CREATE REPORT/  
TITLE North Carolina area-source emissions without RTP counties  
SELECT GROUP REGION NC without Triangle  
BY COUNTY NAME  
BY SRGCODE  
/END/
```

In this example, a title is given using the TITLE instruction. Based on the BY instructions, the report will include emissions totals by county and gridding surrogate type, and will include the county name. Only counties in the “NC without Triangle” group will be included in the report, regardless of what other counties may have data in the emission inventory. Note that the title assumes that the report is for area sources, but in fact, the other instructions could be used for area or mobile sources, which both use gridding surrogates.

Example 6: Using a subgrid and reporting emissions by hour and SCC

```
/CREATE REPORT/  
SELECT SUBGRID Southeastern US  
BY HOUR  
BY SCC  
UNITS g/hr  
/END/
```

In this example, the output data are limited to the grid cells in the Southeastern U.S. subgrid, defined in example 2. Because of the BY instructions, the emissions are reported by SCC for each hour, and the UNITS instruction modifies the emissions values to appear as grams per hour (the default is tons per hour in this case). The BY HOUR instruction implies the TEMPORAL instruction, and the SELECT SUBGRID instruction implies the GRIDDING instruction. So, Smkreport will create this report as if the TEMPORAL and GRIDDING instructions appeared with the other report instructions.

Example 7: Using an in-line subgrid and reporting emissions by cell, source, and elevated status

```
/CREATE REPORT/  
SELECT SUBGRID (100,23) TO (101,24)  
BY CELL  
BY SOURCE STACKPARAM NAME  
BY ELEVSTAT  
/END/
```

This example is only valid for point sources because of the STACKPARAM option on the BY SOURCE instruction and the BY ELEVSTAT instruction. The report created by this example will contain source-by-source emissions for all point sources in cells (100,23), (100,24), (101, 23), and (101,24). The emissions will be accompanied by the plant name (because of the NAME option) and the stack parameters (because of the STACKPARAM option). The finest resolution BY option is BY SOURCE, so Smkreport will sum the emissions by source, and the BY CELL and BY ELEVSTAT instructions will cause the grid cell and elevated status to be included in the report as attributes for the sources.

Example 8: Reporting emissions before speciation and summed after speciation

```
/CREATE REPORT/  
BY STATE NAME  
BY COUNTY NAME  
SPECIATION  
SELECT DATA VOC S-VOC  
/END/
```

In this last example, the inventory VOC emissions will be reported both before and after chemical speciation by SMOKE. The “SELECT DATA VOC S-VOC” instruction indicates that the output data should include the inventory VOC as well as the summed VOC (S-VOC) emissions. The emission totals will be by county, and the state name and county name will be included in the report as well.

DESCRIPTION OF OUTPUT REPORTS

The specific format of the Smkreport output reports depend on the input instructions, but they do follow a general pattern, which we will describe here. Multiple reports can appear in the same file, and this depends on the arrangement of /NEWFILE/ and /CREATE REPORT/ instructions in the RC file. A new output report file is created for each /NEWFILE/ instruction, and another report is put in that file for each /CREATE REPORT/ instruction. The general template of a Smkreport output file is the following:

```
<REPORT 1>
  <User Title(s), one per line>
  <Automatic Titles, one per line>
  <Column headers>
  <Units, for columns that need them>
  <Data values for all rows>
#####
<REPORT2, if any>
  <User Title(s), one per line>
  <Automatic Titles, one per line>
  <Column headers>
  <Units, for columns that need them>
  <Line of minus signs>
  <Data values for all rows>
#####
<Additional reports, if any>
#####
METADATA
  <Input files>
  <Echoed group definitions from REPCONFIG file>
  <Echoed report instructions from REPCONFIG file>
<End of file>
```

The first section of each report is a title section, which starts with user-defined titles from the TITLE instructions. These titles appear in the same order in the output file as they are listed in the input file. Automatic titles are also included with the report to indicate the source category; the year of the inventory (base and future); and whether the emissions include gridding factors, speciation factors, ozone season emissions, and/or temporally allocated emissions.

All columns have labels headers and some columns have units. With some exceptions, Smkreport writes one column to the report for each BY instruct used to create that report. The exceptions are (1) the BY SOURCE instruction, which has one column for each source characteristic, and additional columns if the STACKPARAM or NAME options are used; and (2) the BY COUNTRY, BY STATE, and BY COUNTY instructions, which share a column for the region code, but may add one column each for the names. All data values are written in their own columns. The report will include all data values in the inventory unless a SELECT DATA statement is specified, in which case the report will include only columns for the specific data values. The order of the columns is fixed, as described in more detail in the SMOKE user manual.

Although not complete in Smkreport at the time this paper was written, the output reports will include metadata at the end of the report file, which will have all relevant input files, group definitions, and report instructions for the report(s) in the output file. This feature will be very useful for reminding users which data were used to create the report.

PROCESSING AND QA OF THE 1996 NEI

We used SMOKE v1.3 for a project to process the 1996 National Emission Inventory (NEI), version 3.11, for a U.S. national 36-km grid (Figure 1) for all of 1996. Because the scope of the processing was large, we needed to have quality assurance that could check to ensure that the new version of SMOKE was working with the NEI, and that would permit refined quality assurance if problems were found during the processing. Smkreport was developed, tested, and applied as part of the NEI processing.

For this project, we worked closely with the U.S. EPA Air Quality Modeling Group (AQM), who provided us most of the data, and who assisted in reviewing the quality assurance reports created by SMOKE. Collectively, we used reports from Smkreport to quantitatively evaluate the emissions processing for the project. The reports allowed us to check SMOKE emissions at each stage of the processing, to ensure emissions were not being dropped, and to confirm that any changes in emissions from one step to the next made sense. We initially configured Smkreport and the SMOKE scripts to create a large number of reports, too numerous to mention here. We found, however, that after checking the SMOKE reports that summed emissions SCC and state totals at each stage of the SMOKE processing, the remaining reports were less useful unless a specific problem was identified in the state or SCC reports. In other words, the detailed reports were most useful only when a less detailed report suggests a possible problem with the inventory or emissions processing. We subsequently modified our processing for the remainder of the project to use a more limited number of report to speed up processing, and created more detailed reports as needed.

EPA and we used many of these reports to compare back with the NEI 1996 totals created outside of SMOKE. The state totals after each stage of the SMOKE processing were particularly useful for quality assuring a national inventory. We plotted emissions from each SMOKE processing stage on a state-by-state basis to compare to state total emissions before the data were read into SMOKE. Figure 2 provides an example of one of these plots for area-source VOC emissions. In the figure, the inventory import and gridding symbols match with the dashed line, which represents the emissions data before they are imported into SMOKE. The daily totals of the temporally allocated emissions do not fall exactly on this line, because the purpose of the temporal allocation is to change the emissions from the daily average values. As shown in the figure, the amount that the temporally allocated emissions change from the inventory depends on the day of the week and on the state. However, the amount of the deviation is consistent with what we expect based on the temporal profiles. The degree of the change differs from state to state because SMOKE assigns the weekly temporal profiles using the SCC, and each state has different proportions of emissions from each SCC.

FUTURE QUALITY ASSURANCE

We plan to implement several natural extension to the Smkreport capabilities that we have described in this document. Additional group types will be permitted, such as SCC groups, plant groups, stack groups, and hour groups. Additional BY types will also be allowed, such as parts of SCCs, plant, stack, vehicle type, and Standard Industrial Code (SIC). Groups will also be available for reference in a BY instruction, so that emission totals can be created for the group, instead of only using groups to select records from the inventory.

Although we do not have specific plans to do so at the time this paper was written, we would like to update the Java Analysis and Report Tool (JART) to provide a better interface for further manipulation

of the Smkreport output files. Currently, JART can read in the Smkreport reports and allow users to rearrange columns, sort records, select records using formulas, create simple charts, and print the reports. We would like to update JART to provide even more capabilities for using the Smkreport report, such as better graphical analysis, formulas to combine into new columns, merging columns and/or rows from multiple reports, comparing reports, and driving JART from a script.

Lastly, a large missing piece from SMOKE's quality assurance is the ability to compare emission inventory data with the Photochemical Assessment Monitoring Stations (PAMS) data. Techniques have been available for years to compare ratios of emission inventory data with ratios of the PAMS data. Methods are also available to compare the fractions of chemical compounds collected at certain PAMS sites with the fraction of those compounds in the emission inventory before the chemical compounds are aggregated into model species by the emissions data processor. We added this capability to the SMOKE prototype for internal usage, but we have not updated the latest version of SMOKE with this capability. We would like to do so in the future.

CONCLUSIONS

The Smkreport program in SMOKE is a flexible method to create customized reports of emissions data for quality assurance. This update to SMOKE has greatly increased the available options of SMOKE users for performing quality assurance on emissions data and processing. The tool is user-friendly in the sense that its inputs are based on terminology (the input "instructions") with which emissions modelers are already familiar. As a result, report customization is easier to learn and understand than writing a new report with a programming language such as FORTRAN or SAS®. Although users are limited to the types of customizations that Smkreport has been programmed to recognize, the number of options already available are extensive enough to be quite useful. We have had a lot of positive feedback from the user community on Smkreport, and we have found it very valuable in quality assuring emissions processing of the 1996 NEI.

REFERENCES

- Adelman, Z.; Houyoux, M. "Processing the NET96 version 3.11 Inventory with SMOKE", In *10th Annual Emission Inventory Conference "One Atmosphere, One Inventory, Many Challenges"*; U.S. EPA Office of Air Quality Planning and Standards: Denver, CO, May 1-3, 2001.
- Houyoux, M. R.; Vukovich, J. M.; Coats, C. J. Jr.; Wheeler, N. M.; Kasibhatla, P. S. "Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project", *J. Geophys. Res.*, 2000, 105, D7, pp 9079-9090.
- 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, pp 25,611-25,629.
- Possiel, N.; Stella, G. S.; Ryan, R. R.; Pace, T.; Benjey, W.; Beidler, A.; Kinnee, E.; Houyoux, M.; Adelman, Z. "Development of an Emissions Inventory for Annual Nationwide Models-3/CMAQ Simulations of Ozone and Aerosols"; *10th Annual Emission Inventory Conference "One Atmosphere, One Inventory, Many Challenges"*, U.S. EPA Office of Air Quality Planning and Standards, Denver, CO, May 1-3, 2001.
- U.S. Environmental Protection Agency (EPA), "User's guide to MOBILE5 (Mobile Source Emission Factor Model)", Rep. EPA-AA-AQAB-94-01, 174 pp., Office of Air and Radiat./Office of Mobile Sources, Ann Arbor, Mich., May 1994.

FIGURES

Figure 1. U.S. National unified grid using Lambert projection.

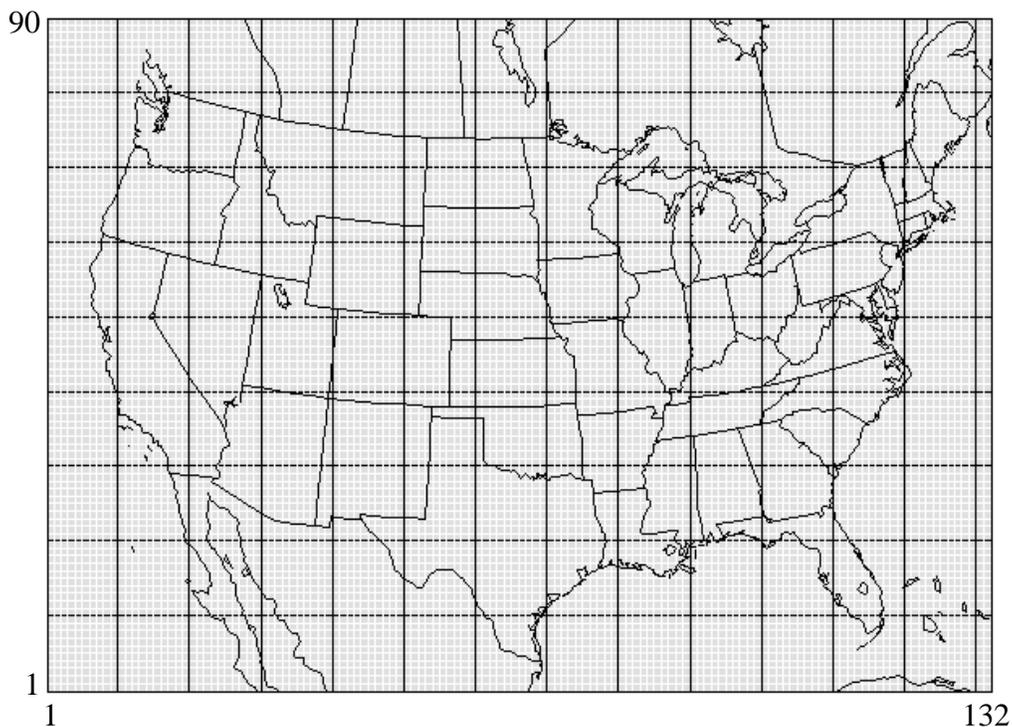
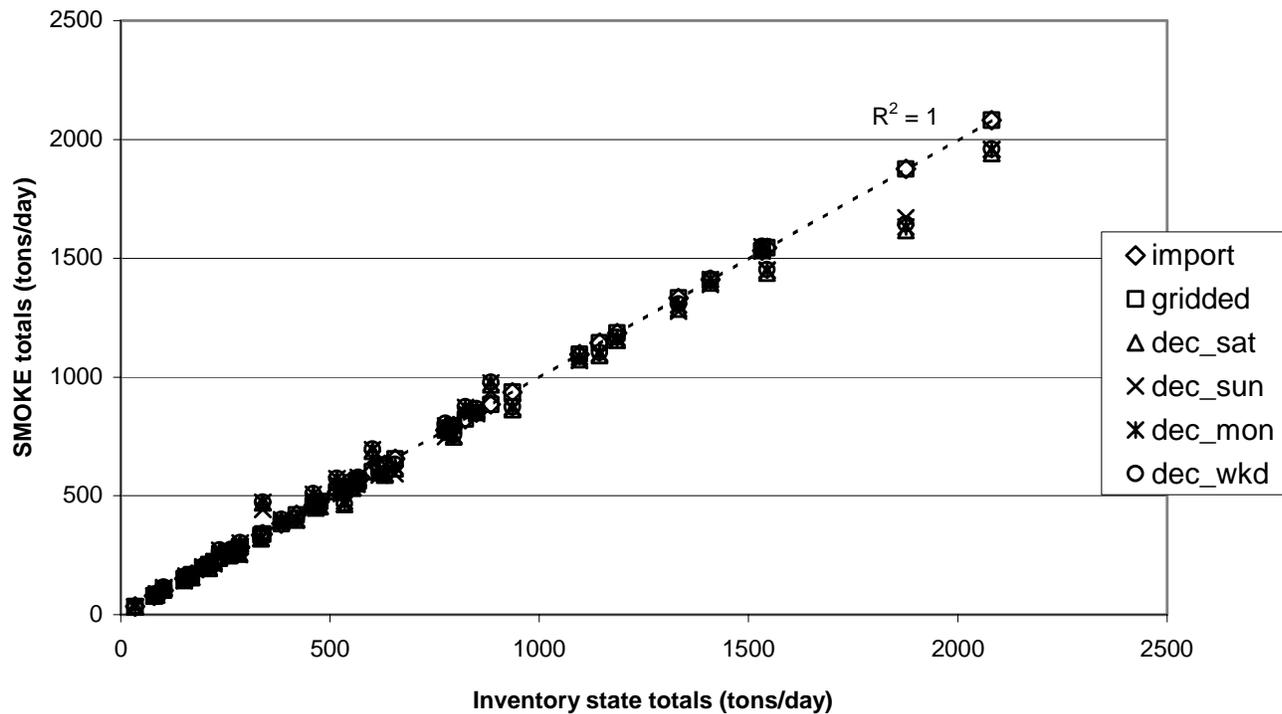


Figure 2: Example of state totals “linear regression” for comparing SMOKE processing to original inventory totals, using area-source VOC emissions for December.



KEYWORDS

SMOKE

quality assurance

emissions processing

National Emissions Inventory