

Processing the National Emissions Inventory 96 (NEI96) version 3.11 with SMOKE

Zachariah Adelman and Marc Houyoux
MCNC–Environmental Programs
P.O. Box 12889, 3021 Cornwallis Rd.
Research Triangle Park, NC 27709-2889
zac@ncsc.org

ABSTRACT

We processed the National Emissions Inventory 96 (NEI96) version 3.11 through version 1 of the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. Using an IBM RS/6000 supercomputer, we prepared emissions for a Community Multiscale Air Quality (CMAQ) modeling system study of ozone and particulate matter over the United States and Canada for the entire year of 1996. For quality assurance, we compared the processed inventory to the NEI96 database totals and performed qualitative spatial analyses using the Package for Analysis and Visualization of Environmental data (PAVE). In addition to summaries of SMOKE performance in processing the NEI96 version 3.11, we present a new technique for executing large SMOKE applications.

INTRODUCTION

The National Emissions Inventory 1996 (NEI96)¹ is the most comprehensive estimate of criteria pollutant emissions in the United States compiled to date.² With version 3.11 released in 2000, the inventory represents a major effort by the individual states to quantify their local and regional contributions to air quality. The U.S. Environmental Protection Agency (EPA) used the NEI96v3.11 emissions inventory as the basis for a Community Multiscale Air Quality Modeling System (CMAQ)³ study of ozone and particulate matter (PM) air quality in the United States for the entire year of 1996.⁴ To prepare an emissions inventory for the CMAQ simulations, we used the Sparse Matrix Operator Kernel Emissions (SMOKE) system version 1⁵ to convert the annual, statewide totals of the NEI96 to temporally, spatially, and chemically allocated emissions estimates. The CMAQ domain covers all of the United States and part of Canada and Mexico with a 36-km resolution grid of 132x90x4 cells. The domain extends vertically to a 10-km model top. All of the SMOKE processing occurred on single processors of an IBM RS/6000 SP running AIX4.3 with 180 four-way 374 MHz Power 3-II nodes and 360 GB of system memory. As this was the first effort to integrate the NEI96v3.11 into SMOKE, we addressed several issues with the inventory before obtaining a data set satisfactory to the scope of the application.⁶ In an effort to strengthen the SMOKE emissions generation process, we developed a new method for executing the various SMOKE programs.

The magnitude of the data requirements for generating a national, year-long emissions inventory forced us to reevaluate the way that we addressed the emissions modeling process with SMOKE. Traditionally, we organized SMOKE emissions modeling by the various source categories. Separate scripts drove the area-, mobile-, point-source, and biogenic emissions processing. Faced with generating an annual inventory for an area covering all of the United States and part of Canada, we realized that our current scripting procedure was not flexible enough to accommodate the large amount of input and output that such a large application required. Integrating multiple temporal profiles throughout the year and applying special temporal profiles to holidays required that we reorganize the SMOKE scripting procedure for this application. In addition, the original technique of categorizing the process by source category required an unnecessary amount of modeler interaction with SMOKE. Reorganizing the procedure by the number of times that each module is executed results in a highly quality-assured process with the flexibility to maintain the data requirements of a large application.

We verified the performance of SMOKE through comparisons between the NEI96v3.11 inventory totals and the quality assurance (QA) reports generated by SMOKE after each major step of the emissions generation process. Checking the state emissions totals generated by SMOKE following the inventory import, spatial allocation, and temporal allocation steps against the NEI96 summaries verified that the NEI96 processed correctly through SMOKE without any unexpected addition or loss of emissions. EPA used additional SMOKE QA reports to address specific inventory issues. We reconciled most of these issues through updates to the NEI96v3.11 and/or revisions to the ancillary SMOKE inputs used for this application.⁴

METHODS

We used an execution-based SMOKE scripting procedure to accommodate this large, data-intensive application. As noted above, viewing SMOKE execution by the number of times that an individual module is implemented, rather than by source category, allows for a more quality-assured emissions generation process. The traditional method of organizing emissions processing by source category leads to redundancy in the modeler's interaction with SMOKE. We based the SMOKE model execution for this application on three primary classes of scripts: yearly, monthly, and daily.

- **Yearly** – *run once per model year*. Inventory scripts require one use per model year to import the unprocessed area, point, and biogenic inventories into SMOKE. Spatial allocation and chemical speciation of the area and point data occur in this step as well. The scripts called the Smkreport QA tool during the inventory import to create reports documenting the initial steps of the conversion of raw inventory to model-ready emissions. We also created a holiday script, which needed to be used only once to allocate holiday emissions to Sunday values and diurnal profiles.
- **Monthly** – *run once per model month*. For each month, the monthly scripts import, spatially allocate, and speciate the unprocessed mobile inventory, to allow the inclusion of month-specific mobile emissions that the EPA provided based on average monthly temperatures. This script also creates the monthly temporally allocated emissions for the area, point, and mobile sources. Because the weekly temporal profiles did not vary over the weekdays, we organized the monthly script to run temporal allocation for Saturday, Sunday, Monday, or weekdays. Monday is needed because the resulting hourly emissions include four time zones, so that some Sunday night emissions are included in the Monday file for Western time zones. The monthly script calls Smkreport twice, once during the mobile inventory import and once during the temporal allocation of the area, point, and mobile sources.
- **Daily** – *run once per model day*. Daily scripts create the daily point-source plume rise files, speciate and temporally allocate the biogenic emissions, and create both source-specific and aggregated daily CMAQ-ready files. This script generates QA reports of state or county totals with each iteration of the Smkmerge module.

The setup outlined above is specific to processing the NEI96 for this application. The overall technique of using execution-based scripts, however, is flexible enough to accommodate data in different formats.

Figure 1 illustrates how the original SMOKE version 1 configuration, which is based on source categories, results in different scripts executing the same SMOKE programs. For example, the “area,” “point,” and “mobile” scripts all call Smkinven, Spcmat, Grdmat, Temporal, and several other SMOKE programs. Allowing the scripts to handle the redundant tasks reduces the sources of human error that can be introduced each time a script is executed.

Figure 1. Conceptual model of SMOKE source-based scripting technique.

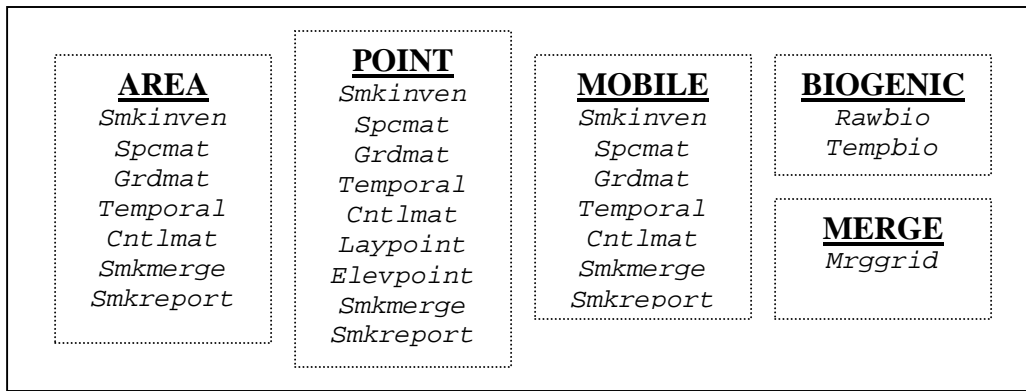


Figure 2. Conceptual model of SMOKE execution-based scripting technique.

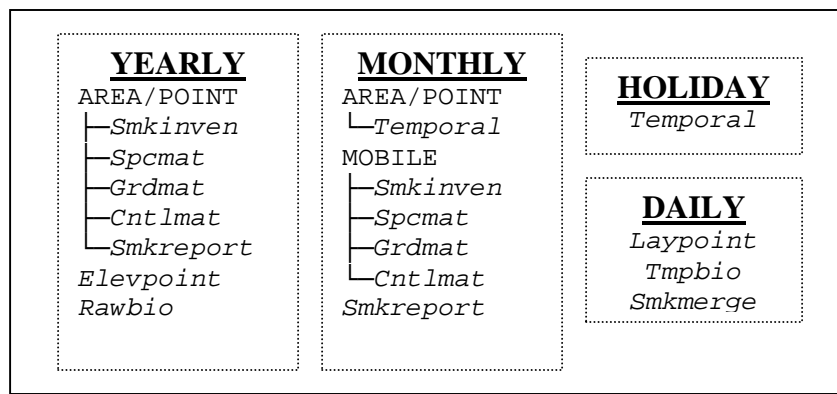


Figure 2 illustrates the execution-based scripting technique that we used for processing the NEI96v3.11. By basing the SMOKE emissions processing on the number of times that each module is executed, we created a system with the flexibility to manage large volumes of input and output data. The execution-based scripts are designed around the data requirements of the SMOKE programs. Figure 2 shows that 7 different SMOKE modules are executed a total of 12 times with the *yearly* script. The *daily* script, on the other hand, implements only three different SMOKE modules, but they are each executed 365 times. The processing arrangement shown in Figure 2 addressed the processing and data management requirements to use SMOKE for an annual run.

The execution-based scripting system has the flexibility to allow for different combinations of implementing the SMOKE modules. For instance, the NEI96 processing required that the import of the point and area inventories occur in the *yearly* script because the area and point inventories are yearly averages. As the mobile inventory consists of monthly averages, we imported the mobile inventories with the *monthly* script. Had the NEI96 point-source inventory been a monthly inventory, we easily could have imported the monthly point inventories into SMOKE with the *monthly* script. The same pattern applies to any part of the SMOKE processing. If we had chosen to use a single weekday temporal profile for the entire year, we would have executed the Temporal module once through the *yearly* script. Alternatively, if we had chosen to have high-resolution, day-specific temporal profiles, we would have executed Temporal 365 times through the *daily* script. In addition to the *yearly*, *monthly*, and *daily* scripts, we created a *holiday* script to generate special Sunday temporal profiles for holiday emissions. We define the days that get the holiday status in this *holiday* script. The *daily* script checks for day-specific temporal profiles during the creation of the daily, model-ready emissions, overriding the month-specific temporal profile if one exists. Table 1 summarizes the NEI96 SMOKE scripting process.

Table 1. NEI96 SMOKE scripting description.

Script	Frequency	SMOKE description
<i>Yearly</i>	1x/year	SMKINVEN, SPCMAT, and GRDMAT for area and point RAWBIO for biogenic SMKREPORT for area and point
<i>Monthly</i>	1x/month	SMKINVEN, SPCMAT, and GRDMAT for mobile TEMPORAL for area, point, and mobile SMKREPORT for mobile inventory import and area, point, and mobile temporal allocation
<i>Holiday</i>	1x/year	TEMPORAL for area, point, and mobile for the predefined holidays
<i>Daily</i>	1x/day	LAYPOINT for point TMPBIO for biogenic SMKMERGE for area, mobile, point, and all sources

We compared the state totals reported by SMOKE following each major processing step against the NEI96v3.11 inventory totals to confirm the performance of SMOKE. We expect zero difference between the inventory totals and SMOKE totals following the inventory import and spatial allocation steps. The processed inventory will differ from the NEI96 totals following the temporal allocation step, but averaged over the entire year, we expect the difference to be zero. We used linear regression plots for qualitative display of the SMOKE processing results and normalized differences for quantitative analysis.^{4,7}

RESULTS

The new scripting convention increases the flexibility of the SMOKE emissions process for accommodating large amounts of input and output data. With the scripts organized by the number of SMOKE program executions needed to generate an entire emissions inventory, the intermediate and final output files are managed more easily than with the old system in which the scripts were organized by source category. Where the old scripting system required a single script to be invoked several times to process one source category and then a separate script invoked to merge all of the source categories together, at a minimum the new convention requires the execution of four scripts to process an entire inventory. There is, however, a considerable amount of setup needed to prepare the execution-based scripts. Monthly “assigns” files set month-specific dates and mobile inventory parameters in addition to setting up the SMOKE run-time environment. A “calendar” script defines the first and last Julian dates of each month and a “holidays” file defines the Julian dates of the holidays. These files are each referenced by the SMOKE scripts at different points in the emissions generation process. We confirmed the performance of SMOKE under the new scripting convention through comparisons between SMOKE state-totals and the NEI96v3.11 inventory totals.

Normalized differences between the SMOKE-generated state totals following each step in the emissions generation process and the NEI96 state totals confirmed that SMOKE correctly processed the inventory. Average normalized differences of zero ensured that SMOKE correctly imported the inventory and allocated the emissions in space and time in accordance with the spatial and temporal allocation files, respectively.^{4,7}

Tables 2 through 6 summarize the SMOKE performance and disk usage statistics on the IBM RS/6000 SP for generating an annual, continental emissions inventory with the NEI96v3.11. These tables present statistics for SMOKE with and without the use of the Smkreport program. As Smkreport provides customized output, the amount of time that it takes to execute and the total size of the output will vary. The statistics given in Tables 2 through 6 represent SMOKE performance with Smkreport configured for this application. Choosing a less detailed report configuration would decrease the run

times and total output file sizes. More detailed report configurations would increase the SMOKE run times and total output.

Table 2 presents the SMOKE output file sizes for a single execution of the various SMOKE programs used in this application. The output file sizes from the time-dependent programs will vary slightly for different model months from the July numbers listed in Table 2.

Table 2. Single-day file sizes (MB) for U.S. NEI96; July 3, 1996. All columns are for one copy except as indicated.

Source	Inven	Gridmat	Temporal ^a	Laypoint ^b	Spcmat	Smkrmerge ^b	Total
Area	76.4	15.4	296.5	n/a	30.3 (x2)	25.0	473.9
Point	106.6	1.6	318.0	159.0	33.4 (x2)	99.8	751.8
Mobile	32.2	20.7	238.6	n/a	25.1 (x2)	25.0	366.7
Biogenic	1.1 (x2)	n/a	n/a	n/a	n/a	8.0 (x2)	18.2
Merged	n/a	n/a	n/a	n/a	n/a	n/a	104.6
Total	217.4	37.7	853.1	159.0	177.6	165.8	

^a64 copies (4/month + 16 holidays)

^b365 copies

Table 3 lists the QA report file sizes for each source category. There are two types of report output by SMOKE: those generated by the Smkreport program and those generated by the Smkmerge program.⁷ The reports generated by Smkreport are customizable, so their total file sizes are application specific. The reports generated by Smkmerge are either state-totals only, or state- and county-totals. The format of these reports is less flexible than those generated by Smkreport, but their total file sizes will also vary depending on the extent of the information they contain. For example, Smkmerge reports can be generated following each step in the emissions generation process. Reports of state-totals following the inventory import step will be slightly smaller than reports of state-totals that assimilate the temporal, spatial, and chemical allocation steps.

Table 3. Annual SMOKE run Smkreport and Smkmerge report file sizes for January (MB).

Source	Inventory	Temporalization	Merge	Total
Area/non-road	1.9	64.8	46.5	113.2
Mobile	1.3	64.8	46.5	112.6
Point	4.4	106.4	46.5	157.3
Biogenic	n/a	n/a	37.2	37.2
Merged	n/a	n/a	46.5	46.5
Total	7.6	236.0	223.2	

Table 4 lists SMOKE performance statistics on the IBM RS/6000 SP by source category. These statistics are also application-specific in that they are coupled to the size of the domain, the number of output species, the resolution of the temporal allocation, and the content of the Smkreport outputs. The memory usage listed in Table 4 is the maximum memory usage reached by that SMOKE program.

Table 4. Annual SMOKE run performance statistics for a single execution of each module on an IBM RS/6000 SP.

SMOKE module		CPU time (s)	Elapsed wall clock time (mm:ss)	Memory usage (KB)
Area	SMKINVEN	380.4	6:41	373148
	GRDMAT	24.7	0:29	323596
	SPCMAT	62.7	1:52	61348
	TEMPORAL	44.6	2:49	399904
	SMKREPORT _{inv}	264.7	4:27	225116
	SMKREPORT _{tmp}	496.8	8:58	253304
Mobile	SMKINVEN	179.0	3:15	330452
	GRDMAT	34.6	0:49	410952
	SPCMAT	91.0	2:14	59944
	TEMPORAL	88.0	3:21	338116
	SMKREPORT _{inv}	204.9	3:27	158684
	SMKREPORT _{tmp}	574.9	10:19	184528
Point	SMKINVEN	552.0	10:21	530684
	GRDMAT	2.7	0:05	12744
	SPCMAT	113.8	3:19	74292
	TEMPORAL	56.7	3:36	437816
	LAYPOINT	146.9	2:40	157372
	SMKREPORT _{inv}	134.5	2:18	61128
	SMKREPORT _{tmp}	302.0	5:49	95880
	RAWBIO	9.2	0:09	2420
	TMPBIO	2.2	0:10	5952
	SMKMERGE	230.7	5:57	364432
Total		3997	82:38	n/a

Table 5 lists timing statistics for the various execution-based scripts used to process the NEI96. As QA reporting takes place in each of these scripts, the run times will vary depending on the user-defined contents of the reports. The numbers listed in Table 5 reflect the report configuration used in this application.

Table 5. SMOKE performance statistics for a single execution of each script on an IBM RS/6000 SP; total times reflect the time required to complete all of the annual SMOKE run.

Script	CPU time (hh:mm:ss)	Elapsed wall clock time (hh:mm:ss)
Inventory	00:25:45	00:29:41
Monthly	01:56:10	02:29:13
Holiday	06:56:49	09:17:51
Daily	00:06:20	00:08:47
Task 2 Total	79:17:52	103:00:55

Table 6 compares the SMOKE run times with and without the Smkreport program. The run times with Smkreport will vary depending on the configuration that the SMOKE modeler chooses. The inventory-based reports are less time consuming than the temporal reports, and in this application were also more useful. The user must determine what the ideal Smkreport configuration is for his or her application to maximize the amount of usable information in the QA reports without compromising the overall performance of SMOKE. We recommend using the daily reporting sparingly as the overall report generating process tends to be slow and the information provided by SMOKE can be redundant.

Table 6. SMOKE run times with and without the generation of QA reports.

Script	Elapsed wall clock time with SMKREPORT (hh:mm:ss)*	Elapsed wall clock time without SMKREPORT (hh:mm:ss)
Inventory	00:29:41	00:10:12
Monthly	02:29:13	00:37:44
Task 2 Total	02:58:54	00:47:56

*These times will vary depending on the number and content of the reports.

The total estimated disk usage for SMOKE outputs for this application is 211.7 GB. The total estimated disk usage for the SMOKE QA reports for this application is 5.8 GB. The total elapsed wall clock time with and without full QA reporting required for completing this application was 103 hours and 73 hours, respectively. The time required for QA reporting will vary depending on the number and content of the reports.

DISCUSSION

In the process of generating a CMAQ emissions inventory with the NEI96v3.11 we developed a SMOKE scripting convention with the flexibility to handle the data requirements of large applications. Comparisons between the SMOKE QA reports and the NEI96v3.11 inventory totals confirmed that the inventory import and spatial and temporal allocation of the emissions completed as expected. The lack of Carbon Bond 4 (CB-IV) speciated NEI96 totals inhibited our ability to check the SMOKE processing beyond the chemical allocation step. In addition to checking the chemical allocation of the emissions, QA steps involving higher resolution NEI96 totals (e.g., county-level or SCC totals) should be performed to verify that the raw inventory numbers do not contain localized erroneous information that is lost in the state-totals. We are confident that SMOKE performs as expected using the execution-based scripting technique.

CONCLUSIONS

The execution-based scripting technique developed for processing the NEI96v3.11 increases the flexibility of implementing the various SMOKE programs for an annual emissions run. The new scripts simplify the management of the SMOKE data requirements and minimize the amount of modeler interaction with SMOKE. Combined with the new SMOKE QA tool, Smkreport, these scripts give emissions modelers the flexibility to generate massive emissions data sets with an almost unlimited combination of QA reports for troubleshooting problems with inventories.⁷

REFERENCES

- ¹TTN CHIEF Web Site, National Emission Inventory Data, <http://www.epa.gov/ttn/chief/net/index.html>.
- ²TTN CHIEF Web Site, Emission Models Clearinghouse, <http://www.epa.gov/ttn/chief/emch/index.html>.

³“Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System”, Byun, D.W., Ching, J.K.S., Eds.; U.S. Environmental Protection Agency, Research Triangle Park, NC, 1999.

⁴Houyoux, M.; Adelman, Z. “1996 Nationwide CMAQ Base Year Emissions, Quality Assurance Document”, Prepared for the U.S. Environmental Protection Agency under Task Order 4TAG68002724 by MCNC, Research Triangle Park, NC. 2001.

⁵Houyoux, M.R.; Vukovich, J.M. “Updates to the Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System and Integration with Models-3”, Presented at The Emission Inventory: Regional Strategies for the Future, Air & Waste Management Association, Raleigh, NC, October 1999.

⁶Possiel, N.; Ryan, 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”, Presented at the 10th Emission Inventory Conference of the U.S. Environmental Protection Agency, Denver, CO, May, 2001.

⁷Houyoux, M. “Quality Assurance Enhancements to the SMOKE Modeling System”, Presented at the 10th Emission Inventory Conference of the U.S. Environmental Protection Agency, Denver, CO, May, 2001.

KEYWORDS

SMOKE

Emission inventories

NEI96

CMAQ

ACKNOWLEDGMENT

This project was funded by the U.S. Environmental Protection Agency under contract number 4TAG68002724 to MCNC.