

Appendix D.4a
Emissions Modeling Documentation

Background

The Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system is an emissions modeling system that generates hourly gridded speciated emission inputs of mobile, nonroad mobile, area, point, fire and biogenic emission sources for photochemical grid models. SMOKE is the fastest emissions processing tool currently available to the air quality modeling community. The sparse matrix approach utilized throughout SMOKE permits both rapid and flexible processing of emissions data. The processing is rapid because SMOKE utilizes a series of matrix calculations instead of less efficient algorithms used in previous systems. The processing is flexible because the processing steps of temporal projection, controls, chemical speciation, temporal allocation, and spatial allocation have been separated into independent operations wherever possible. The results from these steps are merged together at a final stage of processing. Each of these emissions processing steps are detailed below.

Temporal Allocation

VISTAS 2002, 2009 and 2018 annual emissions modeling was configured to generate point, area, nonroad mobile, on-road mobile, and biogenic source emissions. In addition, certain subcategories, such as fires and EGUs were maintained in separate source category files in order to allow maximum flexibility in producing alternate strategies. With the exception of biogenic and on-road mobile source emissions that are generated using the BEIS and MOBILE6 modules in SMOKE, pre-computed annual emissions were processed using the month, day, and hour specific temporal profiles of the SMOKE model. Point and biogenic sources were modeled for each day of the annual period while area and nonroad sources were modeled as a block of Thursday, Friday, Saturday, Sunday, Monday, one per month (total of 60 days modeled).

VISTAS based its temporal profiles and source category cross-reference files on the USEPA CAIR/CAMR/CAVR modeling platform with files located on USEPA's CAIR file transfer website (<ftp://www.airmodelingftp.com/>). Modifications were made to reflect State specific profiles or updated state of knowledge application of these profiles. Additionally, EGU CEM-based temporal profiles and onroad emissions modeling were prepared in manners deviating from USEPA's original CAIR platform.

New temporal profiles used by VISTAS during modeling of the 2002, 2009 and 2018 EGU emissions took the place of USEPA provided default temporal profiles that are generally accepted as not accurately depicting temporal distribution of emissions from EGUs in the U.S. (see Attachment 1 by Stella et. al.). VISTAS EGU temporal profiles were developed using hourly CEM data as reported to USEPA's Clean Air Market's Division (CAMD) for the Acid Rain Program.

The work conducted in this process had the main objective of developing temporal profiles for VISTAS EGUs necessary to apply in the generation of SMOKE PTHOUR formatted emissions. Additionally, State-level monthly, day-of-week, and diurnal profiles were developed for application to non-CEM matched units in the VISTAS

emissions inventory. These temporal distributions represent a significant improvement over the USEPA defaults.

On-road mobile modeling in SMOKE was done for selected weeks (seven days) of each month - using these days as a “representative week” of the entire month. This selection allows for the representation of day-of-the-week variability in the on-road motor vehicles, and models a representation of the meteorological variability in each month. The modeled weeks were selected from mid-month, avoiding inclusion of major holidays. Holidays were modeled as the Sunday of the representative week, while the day after a holiday was modeled as a Monday. VISTAS executed sensitivity tests to examine this “representative week” methodology versus an everyday on-road mobile modeling method. VISTAS determined that the use of representative week on-road mobile emissions produced ozone and particulate matter concentrations (and thus regional haze) that were nearly indistinguishable from the “everyday” mobile method. VISTAS determined that the difference in the modeled air quality - resulting from the on-road mobile modeling methods - was insignificant. For more information on this study see Attachment 2 by Abraczinskas et. al.

On-road mobile emissions are represented by the following weeks per month:

January 15-21
February 12-18
March 12-18
April 16-22
May 14-20
June 11-17
July 16-22
August 13-19
September 17-23
October 15-21
November 12-18
December 17-23

Speciation

Speciation is the process of disaggregating inventory pollutants into individual chemical species components or groups of species. The need for speciation is determined by the inventory purpose. Inventory applications that require detailed speciation include photochemical modeling, air toxics inventories, chemical mass balance modeling, and visibility modeling.

Depending on the purpose of a particular emissions inventory, the inventory may include TOG, NO_x, sulfur oxides (SO_x), CO, total suspended particulate matter (TSP), particulate matter less than 10 micrometers in aerodynamic diameter (PM₁₀), or ammonia (NH₃). However, modeling inventories may require these emissions to be expressed in terms of other pollutants. Additionally, for some models, NO_x emissions may need to be specified as NO and NO₂. Also, PM may need to be separated into

various fractions, such as PM10 and PM less than 2.5 micrometers in aerodynamic diameter (PM2.5).

SMOKE was configured to speciate the emissions estimates according to the requirements of the Carbon Bond Mechanism version four (CBM-IV, CB-IV or CB4). The SMOKE model reformats the emissions estimates for use in CMAQ modeling based on source category code (SCC) and speciation profile cross-reference files. The speciation profiles and source category cross-references use in VISTAS modeling are based on USEPA's CAIR/CAVR/CAMR modeling platform with files located on USEPA's CAIR file transfer website (<ftp://www.airmodelingftp.com/>). Minor modifications were made to reflect State specific profiles or updated state of knowledge application of these profiles. One major change made in the VISTAS modeling was the modification of coal combustion cross-reference from speciation profile "NCOAL" to profile "22001."

Spatial Allocation

Because air quality modeling strives to replicate the actual physical and chemical processes that occur in an inventory domain, it is important that the physical location of emissions be determined as accurately as possible. In an ideal situation, the physical location of all emissions would be known exactly. In reality, however, the spatial allocation of emissions in a modeling inventory only approximates the actual location of emissions.

Gridding surrogates are used to spatially allocate emission sources from a coarse geographic area to finer grid cells used for modeling. There can be hundreds of unique source categories in an emissions inventory, which is typically developed for counties, states, or other areas. The exact location of most major emission sources is known and their geographic coordinates are usually contained in the inventory. These usually are referred to as major point sources and include electric utilities and major industrial facilities. However, other emission sources are estimated for the entire county or other area as an aggregate since the exact locations of each source are not included in the modeling inventory. Surrogates are human activities or land use information that are used to represent a more precise location of emission source category groups. A gridded surrogate ratio is the ratio of the amount of a surrogate in a modeling grid cell to the total amount of that surrogate in a county. Grid cell emissions are calculated by multiplying the cell's gridded surrogate ratio by the county emissions.

These surrogates and their associated SCC cross-references were originally developed by USEPA (<http://www.epa.gov/ttn/chief/emch/spatial/newsurrogate.html>) and converted to the gridded domain definitions of the VISTAS model requirements.

Development of Gridded Surrogate Files

The general process for creating the SMOKE-ready gridded surrogate files from the ArcGIS shape files is as follows:

1. Overlay the grid on the surrogates. Generate the grid polygons (36/12km) with the specifications of the VISTAS domain and spatially overlay (intersect) the grid onto the surrogate area polygons or points. The resulting geodatabase contains, for each surrogate, the county FIPS code, the grid column and row number, and the amount (area, miles or count) of the county's portion of the surrogate in that cell.
2. Extract and convert each geodatabase table to a useful dataset. Each table contains the gridded area, miles or count in each county for a specific surrogate. The variables include FIPS code, column number, row number and area, miles or count.
3. Calculate surrogate ratios. Surrogate ratios are calculated for each surrogate using a series of program files. The programs sum the surrogates for each county and calculate each the ratio by dividing the county cell surrogate value by the total county surrogate value. Combination surrogates where both are of the same type (i.e., Heavy and High Tech Industrial are both area) were summed prior to calculating the ratio. Combination surrogates with unlike data (i.e., 3/4 Roadway Miles plus 1/4 Population are line and area data) were summed after calculating the ratios and then normalized. The surrogate cross-reference code was also assigned here.
4. Gap-fill surrogates for counties missing data. There will be many instances where inventory emissions exist for a particular county but there is no data, for that county, for the surrogate assigned. For example, a county with class 1 locomotive emissions may not have data for the class 1 railroad surrogate. In this case we have selected to incorporate, within the assigned surrogate, a different source of data (a different surrogate) for that particular county. We incorporate secondary surrogates even if there is no emission source that requires it for that particular county. We denote this process as "gap-filling." All surrogates resulting from the gap-filling process have ratios for all counties.

For each surrogate, we assign a secondary or tertiary surrogate where needed for gap filling. For the class 1 railroad surrogate example mentioned above, we chose total railroads as the secondary surrogate. The secondary or tertiary surrogate chosen would be the same across all counties and apply to all SCCs that use the particular primary surrogate. We pull in and substitute the secondary surrogate for counties where the primary surrogate is missing. Tertiary surrogates will then be assigned to those counties that are still without surrogates.

For identified counties having no values for each surrogate, we assign the data based on the appropriate secondary or tertiary surrogate to these counties. A check to see that surrogate ratios for each county sum to approximately 1.00 is also performed in our surrogate development. Ratios will not always sum exactly to 1.00 due to rounding. However, SMOKE will normalize surrogates greater than 1.00.

5. Create SMOKE-formatted spatial surrogate files. The resulting data from the previous steps is then reconfigured into SMOKE-ready format and used in the spatial allocation process.

Treatment of Large Fire Plume Rise

Wildfire, agricultural, and prescribed burn emissions were handled separately from the standard area source input files. We used day specific or monthly estimates of fire emissions from VISTAS, which include burn acreage and biomass loading information for the VISTAS states. Depending on the completeness and quality of the data received, VISTAS-specific calculations were made to calculate spatial and temporal distributions of the fire emissions, rather than relying on standard distribution profiles. We calculated vertical distribution of the fire emissions, based on fire size and biomass involvement. SMOKE v2.1 can model fire plume rise when provided with the following variables:

PTOP – Top of the fire plume profile (meters above ground level)

PBOT – Bottom of the fire plume profile (meters above ground level)

Lay1 – The percent of the emissions entrained in the first modeling layer

For those fires as having the necessary data elements to site these files with distinct time and space coordinates, these variables were prepared and included in the modeling files used to process this emission source type.

The WRAP Fire Emissions Joint Forum Emissions Inventory Report (see http://www.wrapair.org/forums/fejfd/documents/WRAP_2002_PhII_EI_Report_20050722.pdf) has documented an approach for calculating these plume descriptors. In this method, the fires are assigned to one of 5 size categories, based on the total burn acreage, and the biomass fuel loading. These categories are then used to calculate representative hourly plume profiles. These profiles are then used by SMOKE 2.1 to distribute the vertical emissions for the fires. To successfully model fires as elevated point sources, the data included both the day or days on which the fire occurs, and a spatial identifier of the fire location.

Quality Assurance

The Quality Assurance (QA) is one of the most important steps in performing an air quality modeling study. Because emissions inventory development is tedious, time consuming and involves complex manipulation of many different types of large data sets, errors are frequently made and, if rigorous QA measures are not in place, these errors may remain undetected.

A number of QA files were prepared and used to check for gross errors in the emissions inputs. Importing the model-ready emissions into PAVE and looking at both the spatial and temporal distribution of the emission provides insight into the quality and accuracy of the emissions inputs. Some of the additional steps for checking the emissions are summarized in the bulleted list below.

- Visualizing the model-ready emissions with the scale of the plots set to a very low value, we can determine whether there are areas omitted from the raw inventory or if emissions sources are erroneously located in water cells.
- Spot-check the holiday emissions files to confirm that they are temporally allocated like Sundays.
- Producing pie charts emission summaries that highlight the contribution of each emissions source component (e.g. nonroad mobile).
- Normalizing the emissions by population for each state will illustrate where the inventories may be deficient and provide a reality check of the inventories.
- Spot check vertical allocation of point sources using PAVE.

State inventory summaries prepared prior to the emissions processing were compared against SMOKE output report totals generated after each major step of the emissions generation process.

For speciation, a comparison of the inventory state totals versus the same state totals with the speciation matrix applied was completed.

For checking the vertical allocation of the emissions, reports by source, hour, and layer for randomly selected states in the domain were created. These reports were created for a representative weekday in each of the episodes for each of these selected states.

The quantitative QA analyses often reveal significant deficiencies in the input data or the model setup. It may become necessary to tailor these procedures to track down the source of each major problem. As such, we will only outline the basic quantitative QA steps that were performed in an attempt to reveal the underlying problems with the inventories or processing.

Following are some of the reports that were generated to review the processed emissions:

- State and county totals from inventory for each source category
 - (example provided in Table 1 below for Area sources)
- State and county totals after spatial allocation for each source category
- State and county totals by day after temporal allocation for each source category for representative days
- State and county totals by model species after chemical speciation for each source category
- State and county model-ready totals (after spatial allocation, temporal allocation, and chemical speciation) for each source category and for all source categories combined
- If elevated source selection is chosen by user, the report indicating which sources have been selected as elevated and plume-in-grid will be included.
- Totals by source category code (SCC) from the inventory for area, mobile, and point sources
- Totals by state and SCC from the inventory for area, mobile, and point sources
- Totals by county and SCC from the inventory for area, mobile, and point sources

- Totals by SCC and spatial surrogates code for area and mobile sources
- Totals by speciation profile code for area, mobile, and point sources
- Totals by speciation profile code and SCC for area, mobile, and point sources
- Totals by monthly temporal profile code for area, mobile, and point sources
- Totals by monthly temporal profile code and SCC for area, mobile, and point sources
- Totals by weekly temporal profile code for area, mobile, and point sources
- Totals by weekly temporal profile code and SCC for area, mobile, and point sources
- Totals by diurnal temporal profile code for area, mobile, and point sources
- Totals by diurnal temporal profile code and SCC for area, mobile, and point sources
- PAVE plots of gridded inventory pollutants for all pollutants for area, mobile, and point sources

Additional State QA Procedures

Once the on-road mobile SMOKE outputs were acquired by GA EPD, a number of metrics were generated to further QA and summarize the emissions. Those included:

- County emissions totals, bar charts to visually examine whether the counties with the highest emissions are consistent with what was expected from county VMT distribution.
- PAVE plots were generated to check to ensure emissions were showing up in all counties in GA. Scale was lowered to make sure no emissions were omitted.
- PAVE plots were animated over a 24-hr period to ensure diurnal changes in emissions were as they should be.
- PAVE plots were visually inspected to make sure emissions were highest/lowest in logical places.

The following QA checks were performed both together and separate for EGU and non-EGU point sources:

- Data product summaries and raw NIF 3.0 data files were examined.
- County emissions totals were examined to assure the counties with the highest emissions were consistent with what was expected.
- PAVE plots were generated to check to ensure emissions were showing up in all counties in NC. Scale was lowered to make sure no emissions were omitted.
- PAVE plots were animated over a 24-hr period to ensure diurnal changes in emissions were as they should be.
- PAVE plots were visually inspected to make sure emissions were highest/lowest in logical places.
- Errors detected in earlier model runs were rechecked with each successive model run to assure their correction was carried forward in subsequent runs.
- NIF files were examined to identify problems with latitude and longitude, as well as, stack parameters.
- Parsed files were examined (i.e., Excel spreadsheets that provide unit-level results derived from the model plant projections obtained by the IPM) for accuracy.
- Facility level emission summaries for 2018 were examined for both the base case and CAIR case to ensure that emissions were consistent and that there were no missing sources.
- Emissions and controls for Georgia Power Company were compared to requirements in Georgia's Multipollutant Rule. These controls varied substantially from the IPM results. As a result, GA EPD replaced the IPM emission projections for 2009 and 2018 with projections and controls that were consistent with the requirements of Georgia's Multipollutant Rule.
- Ensured that stack parameters were modified appropriately and where necessary at facilities where new controls are scheduled to be installed.
- Input files were examined to assure there were no double counted facilities (example would be if a facility was known by two different names and counted under each).

Typical fire emissions SMOKE output in the VISTAS states were acquired by GA EPD from Alpine Geophysics. The plots and summary reports for these area source fire emissions were spot-checked for QA and included:

- County emissions totals and county emissions by SCC code, were visually examine whether the counties with the highest emissions are consistent with what was expected.
- PAVE plots were generated to check to ensure emissions were showing up in all counties in GA. Scale was lowered to make sure no emissions were omitted.
- PAVE plots were animated over a 24-hr period to ensure diurnal changes in emissions were as they should be.
- PAVE plots were visually inspected to make sure emissions were highest/lowest in logical places.

