**Summary Data Files Provided with the 2011 Emissions Modeling Platform Data Release**
**Updated 12/26/2013**

The 2011 emissions modeling platform is available from http://www.epa.gov/ttn/chief/emch/index.html#2011. It includes the data files that would be needed to run the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system to prepare emissions input data for air quality modeling of the year 2011. These files include emission inventories, data to spatially allocate the emissions to model grid cells, data to temporally allocate the emissions to hours, and data to speciate the inventory pollutants into the chemical species used by the air quality model. The emission inventories are largely consistent with the 2011 National Emissions Inventory, Version 1, with some specific exceptions. These exceptions include different onroad emissions due to the use of MOVES 2010B in the NEI versus the Tier 3 NPRM version in the 2011 emissions modeling platform; some differences in mobile source emissions in California and Texas; differences in commercial marine emissions in the Great Lakes states and in cases where states submitted data; and the addition of some ethanol plants. A spreadsheet that quantifies these differences is included in the data release.

Documentation of, and supporting data files for, the 2011NEIv1 can be found under the 2011 NEI Version 1 Documentation section of the 2011 NEI web page found here: http://www.epa.gov/ttn/chief/net/2011inventory.html. These include input databases and other input files used to develop onroad and nonroad emissions for the modeling platform. Although different versions of MOVES were used for the modeling platform and the NEI, the MOVES inputs were identical for the two applications.

The files for the release are organized into several directories and subdirectories:

- 2011emissions contains compressed emissions inventory files for 2011
- 2011emissions/moves_eftables contains compressed MOVES emission factor tables used to generate onroad emissions inventory data
- 2011emissions/onroad_by_state contains onroad emissions data at the county, SCC, pollutant resolution separated into one zip file per state
- data_common_to_all_years contains many types of ancillary files in the ge_data_for_2011ec.zip along with ocean chlorine data
- reports contains reports and summaries
- smoke_scripts_utils contains scripts for running SMOKE and other related utility programs
- spatial_surrogates contains 12km and 36km spatial surrogate data for allocating county level emissions to model grid cells

To aid in the review of the data, they have been summarized and reorganized into a series of reports with the goal of making the data easier to understand. The summaries and reports can be found in the reports subdirectory. Some of the summaries have data in terms of emissions modeling sectors, which represent emissions for specific categories of sources. The modeling sectors are defined as follows:

- ag - agricultural ammonia emissions
- afdust/afdust_adj - area fugitive dust emissions (PM only); afdust_adj are the emissions after meteorological and land use adjustments
- biogenics - emissions from natural sources
- c1c2rail - C1 and C2 commerical marine emissions plus railroad emissions
- c3marine - C3 marine (large ocean going vessel ) emissions
- nonpt - nonpoint (county-level) emissions not included in other sectors
- nonroad - mobile source emissions from off-road equipment
- onroad - mobile source emissions on roads; onroad_RPD, onroad_RPV, and onroad_RPP are specific subcategories based on the type of activity data and emission factors used
- onroad_rfl - refueling emissions from onroad vehicles
- othar - Non-U.S. area sources
- othon - Non-U.S. onroad sources
- othpt - Non-U.S. point sources
- ptfire - Point source wild and prescribed fire emissions
- ptegu_pk - Emissions from specific EGU peaking units
- ptegu - Emissions from EGUs not specifically designated as peaking
- ptnonipm – U.S. Point source emissions not included in ptegu_pk, ptegu, or pt_oilgas
- pt_oilgas - oil and gas emissions from point sources
- np_oilgas - oil and gas emissions from nonpoint sources
- rwc - residential wood combustion emissions

The EPA uses unique abbreviations for each modeling case. The case for which data is provided is known as 2011ec_v6_11f and many of the files provided reference this abbreviation. The parts of the abbreviation can be translated as follows: 2011 is the year being modeled, e stands for evaluation case meaning it uses year-specific data for EGUs and fires, c means it is the third instance of the 2011 modeling platform (earlier instances were based on the draft NEI), v6 corresponds to the 2011 modeling platform, and 11f is used to denote that the meteorological data is for 2011, while f corresponds to the meteorological model version and configuration. Future model runs of 2011 will have different abbreviations.

The EPA has requested comments on various aspects of the emissions modeling platform. These aspects are described below, along with suggestions regarding which files to use to support the review of the specific aspects of the platform.

**Emissions Modeling Methods:** EPA uses the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system version 3.5 to prepare data for air quality modeling. More information is available from http://cmascenter.org/smoke/. The specific methods and settings for SMOKE are embodied in the scripts used to run each sector. The scripts are in the directory ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/smoke_scripts_utils/.

**Emission Values:** The values in the emission inventories are available in national files in the 2011emis directory in Flat Files 2010 (FF10) formats defined for nonpoint and point sources at http://www.cmascenter.org/smoke/documentation/3.5/html/ch08s02s04.html#d0e40589 and http://www.cmascenter.org/smoke/documentation/3.5/html/ch08s02s10.html#d0e44884, respectively . The ANN_VALUE column has annual emissions. Some emission inventories (e.g., nonroad) may also have values filled in to the monthly value columns (e.g., JAN_VALUE, FEB_VALUE, …, DEC_VALUE. Note that the FF10 files in the emissions modeling platform can be quite large because they contain a lot of detailed data for the entire United States. The data has also been summarized at the state and county level to aid with the review. State total emissions for each modeling sector for can be found in the summary report 2011ec_v6_11f_state_sector_totals.xlsx. County total annual and monthly emissions for each modeling sector can be found the reports 2011ec_county_monthly_report_CAPs.xlsx (for criteria pollutants), 2011ec_county_monthly_report_AE5.xlsx (for speciated PM with PM2.5 split into 6 components), and 2011ec_county_monthly_report_AE6.xlsx (for speciated PM with the PMFINE component of PM2.5 split into 14 components).

**Note**: After the original publication of the data, a double counting error was found in the area fugitive dust (afdust) emissions for some states. This was corrected in the 2011ed case, and updated versions of any affected reports were posted. The new reports are: 2011ed_v6_11f_state_sector_totals.xlsx, 2011ed_county_monthly_report_CAPs.xlsx, 2011ed_county_monthly_report_AE5.xlsx, 2011ed_county_monthly_report_AE6.xlsx and 2011NEIv1 versus 2011ed differences.xlsx that quantifies the differences between the 2011NEIv1 data and the 2011ed emissions modeling data.

For point sectors, most nonpoint emissions, and for nonroad emissions, custom extractions of the data can be obtained from the NEI inventory data page at http://www.epa.gov/ttn/chief/net/2011inventory.html. Because the onroad emissions are not the same in the NEI as in the modeling platform due to different versions of MOVES being used, special state-specific onroad files in FF10 format were developed and can be found in ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/2011emissions/onroad_by_state/.

**CEMS data as it differs from NEI values:** A summary of the differences between the CEMS and NEI values are important because the CEMS data will actually be passed through to the air quality model via SMOKE, while the NEI would report another value if they differed. It is best to have the NEI and CEMS data to match as closely as possible, except in the case of partial year reporters. It would be helpful to have partial year reporters flagged in the modeling platform. The CEMS values as compared to the NEI values for 2011 can be found in this report: 2011_EGUs_NEI_CEMS.xlsx. Note that the format of the CEMS emissions files found in 2011ec_v6_11f_cem.zip is described in http://www.cmascenter.org/smoke/documentation/3.5/html/ch08s02s09.html#d0e43955. The CEMS data is used to substitute emissions in the NEI according to the ORIS facility code and boiler ID as specified in the FF10 files and in the stack parameter report file described below.

**Stack Parameters and Locations**: Stack parameters and locations can be found in the FF10 files in the STKHGT, STKDIAM, STKTEMP, STKFLOW, STKVEL, LATITUDE, LONGITUDE, and LL_DATUM columns, but can also be found in the stack parameter report file 2011ec_stack_parameter_report.xlsx.

**Existing Control Techniques**: All control techniques for a unit independent of pollutants can be found in the stack parameter summary above. The point source flat files specify the control techniques specific to each pollutant in the CONTROL_IDS and CONTROL_MEASURES columns with levels of reduction in the ANN_PCT_RED column. Control measures for each stack are also given in 2011ec_stack_parameter_report.xlsx.

**Boiler Design Capacity**: The boiler design capacities can be found in the stack parameter report and also in the flat files in the DESIGN_CAPACITY and DESIGN_CAPACITY_UNITS columns.

**Temporal allocation:** The annual emissions in the modeling inventories are allocated into hourly values for air quality modeling. Emission totals by county, sector, and month can be found in the county-month-sector reports 2011ec_county_monthly_report_CAPs.xlsx, 2011ec_county_monthly_report_AE5.xlsx, and 2011ec_county_monthly_report_AE6.xlsx. For more details about the temporal profiles used for modeling down to the day and hourly level, the SMOKE profiles have been summarized into spreadsheets for easier review. The temporal profiles for most sectors are in tpro_2011.xlsx, while the remaining profiles are in tpro_onroad_2011.xlsx and tpro_rwc_2011.xlsx. Note that these files include pointers to the relevant file formats as defined in the SMOKE User's Manual. The temporal cross reference data that maps the emission sources to temporal profiles by SCC, county, and sometimes other factors can be found in tref_2011.xlsx. Note that additional county-specific temporal profile data could be added if it was made available to EPA. The SMOKE-ready temporal data is contained in the zip file ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/data_common_to_all_years/ge_dat_for_2011ec.zip. Please provide any alternative temporal profiles or cross reference information provided in a format compatible with SMOKE.

**Spatial Surrogates:** County-level emissions in the modeling inventories are allocated onto air quality model grid cells with spatial surrogates. Pictures of the spatial surrogates are provided in all_surrogate_maps_2011platform_12US1_v2.pdf. Documentation of recent updates to surrogates can be found in US SpatialSurrogate_Documentation_v091113.pdf with full specifications of surrogates provided in US_SpatialSurrogate_Workbook_v093013.xlsx. When reviewing the pictures of the surrogates, it is important to keep in mind that these are not pictures of where emissions are, but instead show how any emissions would be allocated within each county if that spatial surrogate was used. Spatial surrogates should typically sum to one in each county. The plots can show values higher than one in cases where there are multiple grid cells intersecting a county. The assignment of spatial surrogates by SCC is shown in the file gref_2011.xlsx. The actual spatial surrogates can be found in these files in the files in the directory

ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/spatial_surrogates/. The formats for the spatial surrogate files are in the SMOKE User's Manual on this page: http://www.cmascenter.org/smoke/documentation/3.5/html/ch08s04s02.html. Please provide any alternative spatial surrogates or cross-reference data should in a format compatible with SMOKE.

**Chemical Speciation:** The inventory pollutants such as $NO_x$, $PM_{2.5}$, and VOC need to be allocated into the chemical species used by the air quality model through applying speciation profiles. Speciated emissions by state and sector are shown in the file 2011ec_v6_11f_state_sector_totals.xlsx. Speciated PM emissions by county, sector, and month are shown in the county-sector-month spreadsheet 2011ec_county_monthly_report_AE5.xlsx. The speciation profiles themselves are shown in gspro_2011.xlsx. In some cases, combinations of profiles are used as shown in gspro_combo_2011.xlsx. The assignment of the sources to speciation profiles by SCC is shown in the file gsref_2011.xlsx. The formats for speciation data are provided in the SMOKE User's Manual on this page: http://www.cmascenter.org/smoke/documentation/3.5/html/ch08s05.html. Please provide any alternative speciation data in a format compatible with SMOKE.

For more information on the files provided with the data release, please contact Alison Eyth at eyth.alison@epa.gov, or (919) 541-2478.

**Alphabetic list and contents of provided summary files:**

- 2011_EGUs_NEI_CEMS.xlsx: A comparison of the NEI and CEMS data for EGUs
- 2011ed_county_monthly_report_CAPs.xlsx: Monthly and annual total emissions for each county and modeling sector for CAPs
- 2011ed_county_monthly_report_AE5.xlsx: Monthly and annual total emissions for each county and modeling sector for AE5 PM species
- 2011ed_county_monthly_report_AE6.xlsx: Monthly and annual total emissions for each county and modeling sector for AE6 PM species
- 2011ec_stack_parameter_report.xlsx : Stack parameter and location, control device, and boiler capacity information for point sources
- 2011ec_v6_11f_state_sector_totals.xlsx: State total inventory pollutant and speciated emissions for each modeling sector for the original 2011ec case
- 2011ed_v6_11f_state_sector_totals.xlsx: State total inventory pollutant and speciated emissions for each modeling sector for the updated 2011ed case that removed double counting of area fugitive dust in some states
- 2011NEIv1 versus 2011ed differences.xlsx: Differences between the 2011NEI v1 and the 2011ed modeling platform emission inventories
- all_surrogate_maps_2011platform_12US1_v2.pdf: Maps of spatial surrogates used to allocate county-level emissions to modeling grid cells
- gref_2011.xlsx: Spatial surrogate cross reference: Gridding cross-reference to assign spatial surrogates to sources

- gspro_2011.xlsx: Chemical speciation profiles used to allocate inventory pollutants to model species
- gspro_combo_2011.xlsx: Chemical speciation profiles that are used as combinations of other profiles
- gsref_2011.xlsx: Cross reference used to determine which speciation profiles are used for each emissions source
- spro_2011.xlsx: Cross reference used to determine which spatial surrogate to use for each emissions source
- tpro_2011.xlsx: Temporal profiles that allocate annual emissions to months, days, and hours (for sectors other than nonpt, onroad, rwc, and ag)
- tpro_onroad_2011.xlsx: Temporal profiles that allocate annual vehicle miles traveled to months, days, and hours for the onroad sector
- tpro_rwc_2011.xlsx: Temporal profiles that allocate annual emissions to months, days, and hours for the rwc sector
- tref_2011.xlsx: Cross reference used to determine the temporal profiles used for each emissions source
- US_SpatialSurrogate_Documentation_v091113.pdf: Documentation of how the spatial surrogates were developed
- US_SpatialSurrogate_Workbook_v093013.xlsx: Documentation of how the spatial surrogates were developed in a spreadsheet format