Data Files Provided with the 2018 Emissions Modeling Platform Data Release

The 2018 emissions modeling platform is available from http://www.epa.gov/ttn/chief/emch/index.html#2011, with summaries and documentation also posted to docket #EPA-HQ-OAR-2013-0809. On the web and FTP site, the 2018-specific files have been added to the previously released 2011 emissions modeling platform files. For more information on the 2011 files, please see the file DetailsAbout2011DataFiles12262013.pdf and docket number #EPA-HQ-OAR-2013-0743. The platform includes the data files that would be needed to run the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (http://smoke-model.org) to prepare emissions input data for air quality modeling of the year 2018. The files in the data release 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.

Because 2018 is a future year, as opposed to an historic year, the 2018 emissions modeling platform data release includes information about how emission estimates for 2018 were developed. For some types of emission sources, models were run with inputs specific to 2018; in other cases adjustments were applied to the 2011 emissions values to account for changing activity levels and regulatory programs coming on-line in the intervening years. Documentation of the 2011 and 2018 modeling platforms can be found in the Technical Support Document Preparation of Emissions Inventories for the Version 6.0, 2011 Emissions Modeling Platform released in January, 2014.

The files for the 2018 emissions modeling platform data release are organized into several directories:

- 2018emissions contains emissions inventories and other SMOKE input files specific to 2018, along with subdirectories that contain state-specific extractions of inventories along with input databases used to create 2018 emissions data (see below for more details);
- reports contains reports and summaries for both 2011 and 2018 including a subdirectory that has state+SCC summaries for both years including SCC descriptions; and
- smoke_scripts_utils contains scripts for running SMOKE and other related utility programs for both 2011 and 2018.

Data in the following two directories are unchanged from the 2011 data release, but are used by the 2018 modeling platform:

- data_common_to_all_years contains many types of ancillary files such as profiles and cross references in ge_dat_for_2011ec.zip, along with ocean chlorine data; and
- spatial_surrogates contains 12km and 36km spatial surrogate data for allocating county level emissions to model grid cells.

Within the dockets, the files are not grouped into directories, but are instead grouped into collections of files addressing a particular topic. On the FTP site, there are several subdirectories of the 2018emissions directory with the following types of data specific to 2018:

- CoST packets contains data used to project emissions of some sectors to 2018 using the Control Strategy Tool (CoST);
- moves eftables contains compressed MOVES emission factor tables used to generate onroad emissions inventory data for 2018;
- moves inputs contains the input databases and other data files needed to run MOVES for 2018
- nonpoint by state contains nonpoint emission inventory data at the county, SCC, pollutant resolution separated into one zip file per state;
- nonroad by state contains 2018 nonroad emission inventory data at the county, SCC, pollutant resolution separated into one zip file per state;
- nonroad NCDs contains the input databases needed to run NMIM to get nonroad emissions for 2018
- onroad by state contains 2018 onroad emissions inventory data created by SMOKE at the county, SCC, pollutant resolution separated into one zip file per state;
- point by state contains 2018 point source emission inventory data for both EGUs and non-EGUs separated into one zip file per state, with additional files also by state and including monthly emissions allocated by SMOKE; and
- power sector modeling contains inputs to and outputs from the Integrated Planning Model used to project emissions of electric generating units (EGUs) for 2018.

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 of the data release described above. 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 2018 case for which data are provided is known as 2018ed_v6_11f and many of the files provided reference this abbreviation. The parts of the abbreviation can be interpreted as follows: 2018 is the year being modeled, "e" stands for evaluation case meaning it uses year-specific data for EGUs and fires, "d" means it is the fourth instance of the modeling platform, "v6" corresponds to the 2011 modeling platform, and "11f" denotes that the meteorological data are for 2011, while f corresponds to the meteorological model version and configuration. Future model runs of 2018 will have different abbreviations. The 2011 case that is comparable to this 2018 case is 2011ed_v6_11f.

The EPA has requested comments on various aspects of the 2018 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 used the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system version 3.5.1 to prepare data for air quality modeling for the 2018ed_v6_11f case. More information about SMOKE including descriptions of input data formats 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 to run smoke for the case 2018ed_v6_11f can be found in the directory ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/smoke_scripts_utils/.

2018 Emission Values: The EPA requests comment on the criteria air pollutant (CAP) 2018 emission projections in the modeling inventories, with the focus on ozone and particulate matter precursors such NO_x, SO_2, $PM_{2.5}$, PM_{10}, VOC, and NH_3. The EPA will also accept comments on 2018 projections of hazardous air pollutants (HAPs), as they are included in the outputs of models used to develop 2018 emission projections, but HAPs are not the focus of this effort. The national inventory files can be found in the compressed zip files 2018ed_v6_11f_inputs.zip and 2018ed_v6_11f_inputs_nonroad.zip within the 2018emissions directory. The national files provided to use as inputs to SMOKE can be quite large because they contain detailed data for the entire United States. To aid in the interpretation of these data for specific states, state-specific extractions of the inventories have been provided in the following subdirectories of 2018emissions: onroad_by_state,

nonroad_by_state, point_by_state, and nonpoint_by_state. The files in the onroad_by_state and nonroad_by_state directories include the monthly emissions for those sectors at the inventory level. The point_by_state directory includes both the inventories input to SMOKE and point-source inventory level data that show the monthly total emissions after they were temporally allocated by SMOKE.

In the Flat File 2010 (FF10) format files, the annual emissions values are located in the ANN_VALUE column. 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 all of the emission values in the flat files are in units of tons. The EPA requests comment on both the annual and monthly emissions values, where applicable. The Flat File 2010 (FF10) formats are defined for nonpoint (including nonroad) 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.

Summaries of emissions by state and county are provided to aid in the review of emissions values. State total emissions for each modeling sector for can be found in the summary report 2018ed_v6_11f_state_sector_totals.xlsx. County total annual and monthly emissions for all modeling sectors can be found the reports 2018ed_county_monthly_report_CAPs.xlsx (for criteria pollutants), 2018ed_county_monthly_report_AE5.xlsx (for speciated PM with PM2.5 split into 6 components), and 2018ed_county_monthly_report_AE6.xlsx (for speciated PM with the PMFINE component of PM2.5 split into 14 components). State and SCC level emission summaries that include SCC descriptions are also provided in reports/State-SCC-summaries. An overall comparison of 2011 and 2018 emissions by state can be found the in report 2011ed_2018ed_comparison_6jan2014.xlsx.

Emission Values for EGUs: The EPA requests comment on the IPM version 5.13 input assumptions, NEEDS database, 2018 unit-level parsed files, 2018 flat file inputs and outputs, and cross references for matching between IPM and the NEI. The EPA also requests comment on the specific units that are expected to be used as peaking units in the future year and on the nature of the expected 2018 emissions from those units. The following files related to this part of the request for comment are provided in the power_sector_modeling subdirectory:

- NEEDS_v5.13.xlsx: National Electric Energy Data System (NEEDS) database input to IPM version 5.13 that contains the generation unit records used to construct the "model" plants that represent existing and planned/committed units in EPA modeling applications of IPM, along with basic geographic, operating, air emissions, and other data on these generating units;
- NEEDS_513_User_Guide.pdf: User's guide for NEEDS v5.13;
- Web-Ready_Parsed_File_EPA5-13_Base_Case_2018.xlsx: A file showing the IPM results approximated down to the generating unit level

- FlatFile_Inputs.xlsx: A spreadsheet of lookup table data used by IPM v5.13 to convert the state level power estimates to the detailed flat file information needed by SMOKE; and
- FlatFile_Methodology.pdf: Documentation of how the SMOKE flat file is created by IPM v5.13.
- FlatFile_EPA513_BC_7c_2018_20131108_(To EPA 11-11-13).txt: The emission inventory flat file produced by IPM v5.13 that can be input to SMOKE for air quality modeling;
- FlatFile_EPA513_BC_7c_2018_20131108.xlsx: The emission inventory flat file produced by IPM v5.13 loaded into a spreadsheet to make it easier to sort and filter the data;

Model inputs and activity data used to develop mobile source emission inventories: The EPA requests comment on the mobile source model input data used to develop the projected future mobile source emission inventories. These include both the databases used to create emission factors and the vehicle miles traveled and vehicle population activity data used to compute the emissions. Of particular interest are county total vehicle miles traveled, the mixture of vehicle types in 2018, and changes to the inspection and maintenance programs. Alternative activity data may be provided in the form of MOVES county databases or in SMOKE FF10 activity data format. Other alternative MOVES input data should be provided as MOVES databases for each affected county. The moves emission factors used to compute the 2018 emissions can be found in the moves_eftables subdirectory of 2018emissions, while the inputs to MOVES can be found in the moves_inputs subdirectory. The national onroad activity datasets can be found in the file 2018ed_v6_11f_inputs.zip. The nonroad NMIM inputs can be found in the nonroad_NCD subdirectory.

Projection data and methods. The EPA seeks comment on the data used to project point and nonpoint source emissions from 2011 to 2018, and on the methods and assumptions used to implement the projections. In this context, nonpoint source emissions are inclusive of commercial marine vessel, railroad, and other nonpoint emissions. In particular, EPA seeks comment on its assumptions regarding the manner in which specific consent decrees and state- or locality-specific control programs will be implemented. Summaries are provided in the reports 2018ed_CoST_ptnonipm_facility_summary.xlsx and 2018ed_CoST_sector_summaries.xlsx to illustrate EPA's current assumptions regarding the implementation of consent decrees and other programs. The detailed data used to project emissions of various sectors with the Control Strategy Tool can be found in the file 2018ed_CoST_packets.zip and additional information including the projection data can be found in 2018_Projections_with_CoST_23dec13.xlsx.

Existing 2011 Control Techniques: All control techniques for a unit independent of pollutants can be found in the stack parameter report 2011ec_stack_parameter_report.xlsx. 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. In addition to the flat files, control measures used for each stack in 2011 are listed in 2011ec_stack_parameter_report.xlsx in the reports subdirectory.

Temporal allocation: The annual emissions in the modeling inventories are allocated into hourly values for air quality modeling. Except for electric generating units, factors for the temporal allocation of emissions to hours are the same in 2011 as in 2018. Emission totals by county, sector, and month can be found in the county-month-sector reports 2018ed_county_monthly_report_CAPs.xlsx, 2018ed_county_monthly_report_AE5.xlsx, and 2018ed_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, with all available in the reports subdirectory. 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. Additional county-specific temporal profile data could be added to the temporal cross reference if such data was made available to EPA. The SMOKE-ready temporal data are 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 profile and temporal cross reference data in a format compatible with SMOKE as specified in http://www.cmascenter.org/smoke/documentation/3.5/html/ch08s03.html.

The file 2018ed_v6_11f_cem.zip contains hourly CEMS-style emissions for the ptegu and ptegu_pk sectors. These data are based on the hourly temporal patterns derived from the 2011 hourly CEMS data available on EPA's Clean Air Markets Program Data website (ampd.epa.gov). Seasonal (winter/summer) to hour temporal factors based on 2011 CEM sources were applied to the seasonal 2018 IPM source emissions, where the ORIS Facility and Boiler IDs could be matched. Note that the 2018emissions/point_by_state directory includes files that show the monthly values for point sources after they were temporally allocated by SMOKE. The nonroad and onroad files given in 2018emissions/nonroad_by_state and 2018emissions/onroad_by_state directories also include monthly emissions as they are seen by SMOKE split into state-specific files for the nonroad and onroad sectors, respectively.

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 the reports directory 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. The same spatial surrogate data are used for both 2011 and 2018. 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 in the reports directory. 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 in a format compatible with SMOKE.

Chemical Speciation: Prior to air quality modeling, 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 the application of speciation profiles. The speciation profiles in the 2018 emissions modeling platform are consistent with version 4.3 of the SPECIATE database. Some updates to speciation data are made for the 2018 case as compared to 2011. The report 2011ed_2018ed_speciation_differences.xlsx highlights the differences. The EPA requests comment on the speciation profiles used in the 2018 modeling platform, as well as any information that could help improve the speciation of oil and gas emissions in both the eastern and western United States in 2018. Oil and gas speciation information, along with VOC to TOG adjustment factors that are used to compute methane emissions, would be of the most use at the county or oil/gas basin level of detail and also for each distinct process at oil and gas drilling/production facilities (e.g., glycol dehydrators).

Speciated emissions by state and sector are shown in the report 2018ed_v6_11f_state_sector_totals.xlsx in the reports directory. Speciated PM emissions by county, sector, and month are shown in the county-sector-month spreadsheet 2018ed_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. Some speciation profiles and related information were updated for the 2018 case and can be found in the file ge_dat_for_2018ed.zip under 2018emissions. 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:

- 2011ec_stack_parameter_report.xlsx : Stack parameter and location, 2011 control device, and boiler capacity information for point sources in the 2011ec and 2011ed cases
- 2011ed_2018ed_comparison_6jan2014.xlsx: Comparison of 2011 and 2018 CAPs
- 2011ed_2018ed_speciation_differences.xlsx: Differences in the speciation data used in the 2011ed vs 2018ed cases, along with the updated 2018 data
- 2018ed_CoST_ptnonipm_facility_summary.xls: Summary of ptnonipm control strategy tool information at the facility level

- 2018ed_CoST_sector_summaries.xls: Summary of control strategy tool information by sector at the national, state, state+SCC, and county levels
- 2018ed_county_monthly_report_CAPs.xlsx: Monthly and annual total emissions for each county and modeling sector for CAPs
- 2018ed_county_monthly_report_AE5.xlsx: Monthly and annual total emissions for each county and modeling sector for AE5 PM species
- 2018ed_county_monthly_report_AE6.xlsx: Monthly and annual total emissions for each county and modeling sector for AE6 PM species
- 2018ed_Projections_with_CoST_23dec13.xlsx: Descriptions of the projection packets used with CoST to produce the 2018ed inventories, in addition to the data from the packets as tabs
- 2018ed_v6_11f_state_sector_totals.xlsx: State total inventory pollutant and speciated emissions for each modeling sector
- 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
- State-SCC-summaries: contains state-SCC summaries with descriptions for point, nonpoint, onroad, and nonroad emissions and also for onroad activity data
- 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