Bayesian Space-time Downscaling Fusion Model (Downscaler) -Derived Estimates of Air Quality for 2015

EPA-454/R-19-009
June 2019

Bayesian Space-time Downscaling Fusion Model (Downscaler) -Derived Estimates of Air Quality for 2015

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Research Triangle Park, NC

Authors:
Adam Reff (EPA/OAR)
Sharon Phillips (EPA/OAR)
Alison Eyth (EPA/OAR)
David Mintz (EPA/OAR)

Acknowledgements
The following people served as reviewers of this document: Liz Naess (EPA/OAR) and Elizabeth Mannshardt (EPA/OAR).

Contents

Contents 1
1.0 Introduction 2
2.0 Air Quality Data 5
2.1 Introduction to Air Quality Impacts in the United States 5
2.2 Ambient Air Quality Monitoring in the United States 7
2.3 Air Quality Indicators Developed for the EPHT Network 11
3.0 Emissions Data 13
3.1 Introduction to Emissions Data Development 13
3.2 Emission Inventories and Approaches 13
3.3 Emissions Modeling Summary 37
3.4 Emissions References 74
4.0 CMAQ Air Quality Model Estimates 78
4.1 Introduction to the CMAQ Modeling Platform 78
4.2 CMAQ Model Version, Inputs and Configuration 79
5.0 Bayesian space-time downscaling fusion model (downscaler) -Derived Air Quality Estimates 108
5.1 Introduction 108
5.2 Downscaler Model 108
5.3 Downscaler Concentration Predictions 109
5.4 Downscaler Uncertainties 114
5.5 Summary and Conclusions 116
Appendix A - Acronyms 117

1.0 Introduction

This report describes estimates of daily ozone (maximum 8-hour average) and PM2.5 (24-hour average) concentrations throughout the contiguous United States during the 2015 calendar year generated by EPA's recently developed data fusion method termed the "downscaler model" (DS). Air quality monitoring data from the State and Local Air Monitoring Stations (SLAMS) and numerical output from the Community Multiscale Air Quality (CMAQ) model were both input to DS to predict concentrations at the 2010 US census tract centroids encompassed by the CMAQ modeling domain. Information on EPA's air quality monitors, CMAQ model, and downscaler model is included to provide the background and context for understanding the data output presented in this report. These estimates are intended for use by statisticians and environmental scientists interested in the daily spatial distribution of ozone and PM2.5.

DS essentially operates by calibrating CMAQ data to the observational data, and then uses the resulting relationship to predict "observed" concentrations at new spatial points in the domain. Although similar in principle to a linear regression, spatial modeling aspects have been incorporated for improving the model fit, and a Bayesian[1] approaching to fitting is used to generate an uncertainty value associated with each concentration prediction. The uncertainties that DS produces are a major distinguishing feature from earlier fusion methods previously used by EPA such as the "Hierarchical Bayesian" (HB) model (McMillan et al, 2009). The term "downscaler" refers to the fact that DS takes grid-averaged data (CMAQ) for input and produces point-based estimates, thus "scaling down" the area of data representation. Although this allows air pollution concentration estimates to be made at points where no observations exist, caution is needed when interpreting any within-gridcell spatial gradients generated by DS since they may not exist in the input datasets. The theory, development, and initial evaluation of DS can be found in the earlier papers of Berrocal, Gelfand, and Holland (2009, 2010, and 2011).

EPA's Office of Air and Radiation's (OAR) Office of Air Quality Planning and Standards (OAQPS) provides air quality monitoring data and model estimates to the Centers for Disease Control and Prevention (CDC) for use in their Environmental Public Health Tracking (EPHT) Network. CDC's EPHT Network supports linkage of air quality data with human health outcome data for use by various public health agencies throughout the U.S. The EPHT Network Program is a multidisciplinary collaboration that involves the ongoing collection, integration, analysis, interpretation, and dissemination of data from: environmental hazard monitoring activities; human exposure assessment information; and surveillance of noninfectious health conditions. As part of the National EPHT Program efforts, the CDC led the initiative to build the National EPHT Network (http:// www.cdc.gov/nceh/tracking/default.htm). The National EPHT Program, with the EPHT Network as its cornerstone, is the CDC's response to requests calling for improved understanding of how the environment affects human health. The EPHT Network is designed to provide the means to identify, access, and organize hazard, exposure, and health

[1] Bayesian statistical modeling refers to methods that are based on Bayes' theorem, and model the world in terms of probabilities based on previously acquired knowledge.

data from a variety of sources and to examine, analyze and interpret those data based on their spatial and temporal characteristics.

Since 2002, EPA has collaborated with the CDC on the development of the EPHT Network. On September 30, 2003, the Secretary of Health and Human Services (HHS) and the Administrator of EPA signed a joint Memorandum of Understanding (MOU) with the objective of advancing efforts to achieve mutual environmental public health goals[2]. HHS, acting through the CDC and the Agency for Toxic Substances and Disease Registry (ATSDR), and EPA agreed to expand their cooperative activities in support of the CDC EPHT Network and EPA's Central Data Exchange Node on the Environmental Information Exchange Network in the following areas:

- Collecting, analyzing and interpreting environmental and health data from both agencies (HHS and EPA).
- Collaborating on emerging information technology practices related to building, supporting, and operating the CDC EPHT Network and the Environmental Information Exchange Network.
- Developing and validating additional environmental public health indicators.
- Sharing reliable environmental and public health data between their respective networks in an efficient and effective manner.
- Consulting and informing each other about dissemination of results obtained through work carried out under the MOU and the associated Interagency Agreement (IAG) between EPA and CDC.

The best available statistical fusion model, air quality data, and CMAQ numerical model output were used to develop the estimates. Fusion results can vary with different inputs and fusion modeling approaches. As new and improved statistical models become available, EPA will provide updates.

Although these data have been processed on a computer system at the Environmental Protection Agency, no warranty expressed or implied is made regarding the accuracy or utility of the data on any other system or for general or scientific purposes, nor shall the act of distribution of the data constitute any such warranty. It is also strongly recommended that careful attention be paid to the contents of the metadata file associated with these data to evaluate data set limitations, restrictions or intended use. The U.S. Environmental Protection Agency shall not be held liable for improper or incorrect use of the data described and/or contained herein.

[2] HHS and EPA agreed to extend the duration of the MOU, effective since 2002 and renewed in 2007, until June 29, 2017. The MOU is available at www.cdc.gov/nceh/tracking/partners/epa_mou_2007.htm.

The four remaining sections and one appendix in the report are as follows:

- Section 2 describes the air quality data obtained from EPA's nationwide monitoring network and the importance of the monitoring data in determining health potential health risks.

- Section 3 details the emissions inventory data, how it is obtained and its role as a key input into the CMAQ air quality computer model.

- Section 4 describes the CMAQ computer model and its role in providing estimates of pollutant concentrations across the U.S. based on 12-km grid cells over the contiguous U.S.

- Section 5 explains the downscaler model used to statistically combine air quality monitoring data and air quality estimates from the CMAQ model to provide daily air quality estimates for the 2010 US census tract centroid locations within the contiguous U.S.

- The appendix provides a description of acronyms used in this report.

2.0 Air Quality Data

To compare health outcomes with air quality measures, it is important to understand the origins of those measures and the methods for obtaining them. This section provides a brief overview of the origins and process of air quality regulation in this country. It provides a detailed discussion of ozone (O_3) and particulate matter (PM). The EPHT program has focused on these two pollutants, since numerous studies have found them to be most pervasive and harmful to public health and the environment, and there are extensive monitoring and modeling data available.

2.1 Introduction to Air Quality Impacts in the United States

2.1.1 The Clean Air Act

In 1970, the Clean Air Act (CAA) was signed into law. Under this law, EPA sets limits on how much of a pollutant can be in the air anywhere in the United States. This ensures that all Americans have the same basic health and environmental protections. The CAA has been amended several times to keep pace with new information. For more information on the CAA, go to https://www.epa.gov/clean-air-act-overview.

Under the CAA, the U.S. EPA has established standards, or limits, for six air pollutants known as the criteria air pollutants: carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and particulate matter (PM). These standards, called the National Ambient Air Quality Standards (NAAQS), are designed to protect public health and the environment. The CAA established two types of air quality standards. Primary standards set limits to protect public health, including the health of "sensitive" populations such as asthmatics, children, and the elderly. Secondary standards set limits to protect public welfare, including protection against decreased visibility, damage to animals, crops, vegetation, and buildings. The CAA requires EPA to review these standards at least every five years. For more specific information on the NAAQS, go to https://www.epa.gov/criteria-air-pollutants/naaqs-table. For general information on the criteria pollutants, go to https://www.epa.gov/criteria-air-pollutants.

When these standards are not met, the area is designated as a nonattainment area. States must develop state implementation plans (SIPs) that explain the regulations and controls it will use to clean up the nonattainment areas. States with an EPA-approved SIP can request that the area be designated from nonattainment to attainment by providing three consecutive years of data showing NAAQS compliance. The state must also provide a maintenance plan to demonstrate how it will continue to comply with the NAAQS and demonstrate compliance over a 10-year period, and what corrective actions it will take should a NAAQS violation occur after designation. EPA must review and approve the NAAQS compliance data and the maintenance plan before designating the area; thus, a person may live in an area designated as nonattainment even though no NAAQS violation has been observed for quite some time. For more information on designations, go to https://www.epa.gov/ozone-designations and https://www.epa.gov/particle-pollution-designations.

2.1.2 Ozone

Ozone is a colorless gas composed of three oxygen atoms. Ground level ozone is formed when pollutants released from cars, power plants, and other sources react in the presence of heat and sunlight. It is the prime ingredient of what is commonly called "smog." When inhaled, ozone can cause acute respiratory

problems, aggravate asthma, cause inflammation of lung tissue, and even temporarily decrease the lung capacity of healthy adults. Repeated exposure may permanently scar lung tissue. EPA's Integrated Science Assessments and Risk and Exposure documents are available at https://www.epa.gov/naaqs/ozone-o3-air-quality-standards. The current NAAQS for ozone (last revised in 2015) is a daily maximum 8-hour average of 0.070 parts per million [ppm] (for details, see https://www.epa.gov/ozone-pollution/setting-and-reviewing-standards-control-ozone-pollution#standards. The Clean Air Act requires EPA to review the NAAQS at least every five years and revise them as appropriate in accordance with Section 108 and Section 109 of the Act. The standards for ozone are shown in Table 2-1.

Table 2-1. Ozone Standards

Parts Per Million: Measurement – (ppm)	**1997**	**2008**	**2015**
4th Highest Daily Max 8-hour average	0.08	0.075	0.070

2.1.3 Particulate Matter

PM air pollution is a complex mixture of small and large particles of varying origin that can contain hundreds of different chemicals, including cancer-causing agents like polycyclic aromatic hydrocarbons (PAH), as well as heavy metals such as arsenic and cadmium. PM air pollution results from direct emissions of particles as well as particles formed through chemical transformations of gaseous air pollutants. The characteristics, sources, and potential health effects of particulate matter depend on its source, the season, and atmospheric conditions.

As practical convention, PM is divided by sizes into classes with differing health concerns and potential sources4. Particles less than 10 micrometers in diameter (PM_{10}) pose a health concern because they can be inhaled into and accumulate in the respiratory system. Particles less than 2.5 micrometers in diameter ($PM_{2.5}$) are referred to as "fine" particles. Because of their small size, fine particles can lodge deeply into the lungs. Sources of fine particles include all types of combustion (motor vehicles, power plants, wood burning, etc.) and some industrial processes. Particles with diameters between 2.5 and 10 micrometers ($PM_{10\text{-}2.5}$) are referred to as "coarse" or PMc. Sources of PMc include crushing or grinding operations and dust from paved or unpaved roads. The distribution of PM_{10}, $PM_{2.5}$ and PMc varies from the Eastern U.S. to arid western areas.

Particle pollution - especially fine particles - contains microscopic solids and liquid droplets that are so small that they can get deep into the lungs and cause serious health problems. Numerous scientific studies have linked particle pollution exposure to a variety of problems, including premature death in people with heart or lung disease, nonfatal heart attacks, irregular heartbeat, aggravated asthma, decreased lung function, and increased respiratory symptoms, such as irritation of airways, coughing or difficulty breathing. Additional information on the health effects of particle pollution and other technical documents related to PM standards are available at https://www.epa.gov/pm-pollution.

[4] The measure used to classify PM into sizes is the aerodynamic diameter. The measurement instruments used for PM are designed and operated to separate large particles from the smaller particles. For example, the $PM_{2.5}$ instrument only captures and thus measures particles with an aerodynamic diameter less than 2.5 micrometers. The EPA method to measure PMc is designed around taking the mathematical difference between measurements for PM_{10} and $PM_{2.5}$.

The current NAAQS for $PM_{2.5}$ (last revised in 2012) includes both a 24-hour standard to protect against short-term effects, and an annual standard to protect against long-term effects. The annual average $PM_{2.5}$ concentration must not exceed 12.0 micrograms per cubic meter (ug/m^3) based on the annual mean concentration averaged over three years, and the 24-hr average concentration must not exceed 35 ug/m^3 based on the 98th percentile 24-hour average concentration averaged over three years. More information is available at https://www.epa.gov/pm-pollution/setting-and-reviewing-standards-control-particulate-matter-pm-pollution#standards. The standards for $PM_{2.5}$ are shown in Table 2-2.

Table 2-2. $PM_{2.5}$ Standards

Micrograms Per Cubic Meter: Measurement - (ug/m^3)	**1997**	**2006**	**2012**
Annual Average	15.0	15.0	12.0
24-Hour Average	65	35	35

2.2 Ambient Air Quality Monitoring in the United States

2.2.1 Monitoring Networks

The Clean Air Act (Section 319) requires establishment of an air quality monitoring system throughout the U.S. The monitoring stations in this network have been called the State and Local Air Monitoring Stations (SLAMS). The SLAMS network consists of approximately 4,000 monitoring sites set up and operated by state and local air pollution agencies according to specifications prescribed by EPA for monitoring methods and network design. All ambient monitoring networks selected for use in SLAMS are tested periodically to assess the quality of the SLAMS data being produced. Measurement accuracy and precision are estimated for both automated and manual methods. The individual results of these tests for each method or analyzer are reported to EPA. Then, EPA calculates quarterly integrated estimates of precision and accuracy for the SLAMS data.

The SLAMS network experienced accelerated growth throughout the 1970s. The networks were further expanded in 1999 based on the establishment of separate NAAQS for fine particles ($PM_{2.5}$) in 1997. The NAAQS for $PM_{2.5}$ were established based on their link to serious health problems ranging from increased symptoms, hospital admissions, and emergency room visits, to premature death in people with heart or lung disease. While most of the monitors in these networks are located in populated areas of the country, "background" and rural monitors are an important part of these networks. For more information on SLAMS, as well as EPA's other air monitoring networks go to https://www.epa.gov/amtic.

In 2009, approximately 43 percent of the US population was living within 10 kilometers of ozone and $PM_{2.5}$ monitoring sites. In terms of US Census Bureau tract locations, 31,341 out of 72,283 census tract centroids were within 10 kilometers of ozone monitoring sites. Highly populated Eastern US and California coasts are well covered by both ozone and $PM_{2.5}$ monitoring network (**Figure 2-1).**

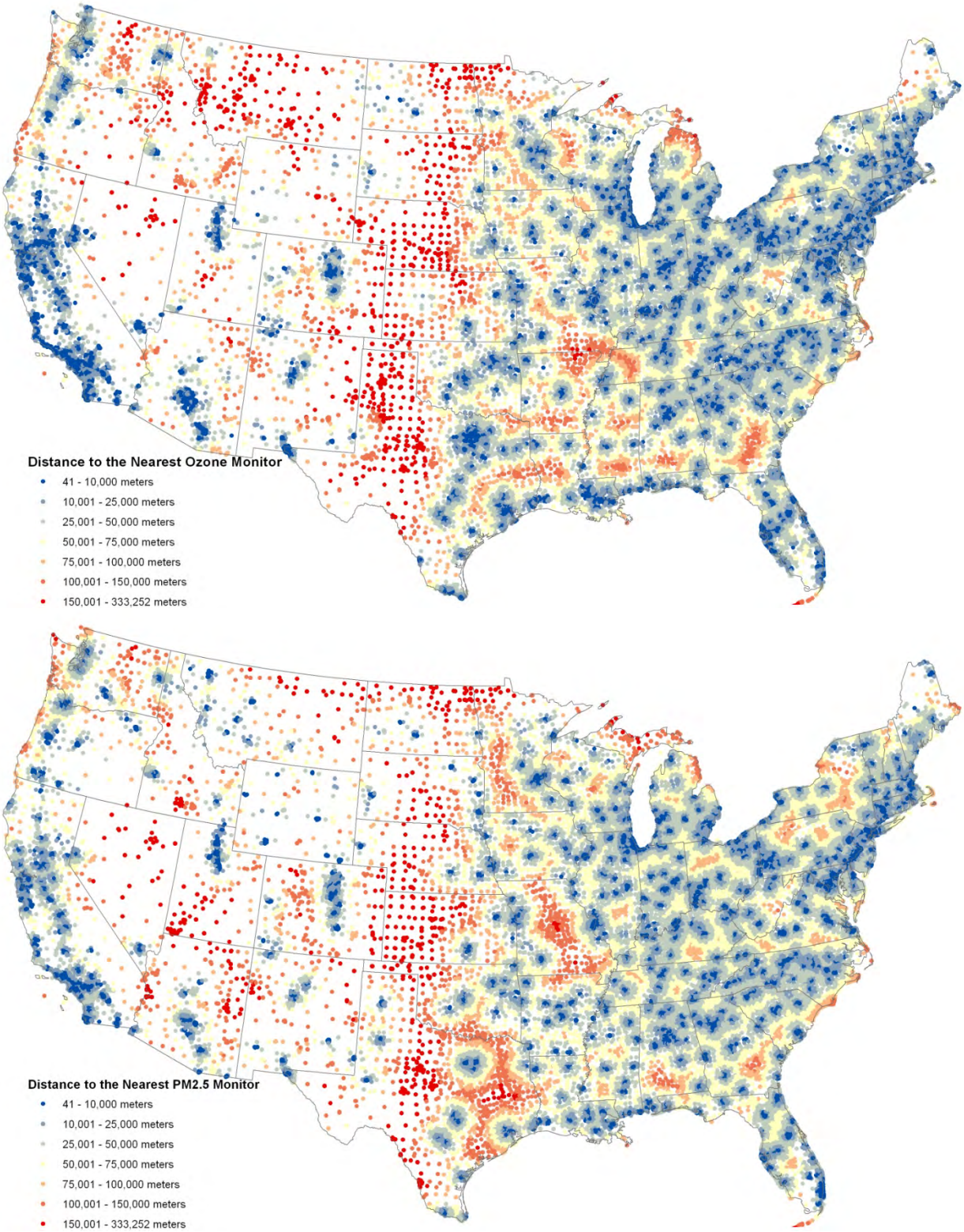


Figure 2-1. Distances from US Census Tract centroids to the nearest monitoring site, 2009.

In summary, state and local agencies and tribes implement a quality-assured monitoring network to measure air quality across the United States. EPA provides guidance to ensure a thorough understanding of the quality of the data produced by these networks. These monitoring data have been used to characterize the status of the nation's air quality and the trends across the U.S. (see https://www.epa.gov/air-trends).

2.2.2 Air Quality System Database

EPA's Air Quality System (AQS) database contains ambient air monitoring data collected by EPA, state, local, and tribal air pollution control agencies from thousands of monitoring stations. AQS also contains meteorological data, descriptive information about each monitoring station (including its geographic location and its operator), and data quality assurance and quality control information. State and local agencies are required to submit their air quality monitoring data into AQS within 90 days following the end of the quarter in which the data were collected. This ensures timely submission of these data for use by state, local, and tribal agencies, EPA, and the public. EPA's Office of Air Quality Planning and Standards and other AQS users rely upon the data in AQS to assess air quality, assist in compliance with the NAAQS, evaluate SIPs, perform modeling for permit review analysis, and perform other air quality management functions. For more details, including how to retrieve data, go to https://www.epa.gov/aqs.

2.2.3 Advantages and Limitations of the Air Quality Monitoring and Reporting System

Air quality data is required to assess public health outcomes that are affected by poor air quality. The challenge is to get surrogates for air quality on time and spatial scales that are useful for Environmental Public Health Tracking activities.

The advantage of using ambient data from EPA monitoring networks for comparing with health outcomes is that these measurements of pollution concentrations are the best characterization of the concentration of a given pollutant at a given time and location. Furthermore, the data are supported by a comprehensive quality assurance program, ensuring data of known quality. One disadvantage of using the ambient data is that it is usually out of spatial and temporal alignment with health outcomes. This spatial and temporal 'misalignment' between air quality monitoring data and health outcomes is influenced by the following key factors: the living and/or working locations (microenvironments) where a person spends their time not being co-located with an air quality monitor; time(s)/date(s) when a patient experiences a health outcome/symptom (e.g., asthma attack) not coinciding with time(s)/date(s) when an air quality monitor records ambient concentrations of a pollutant high enough to affect the symptom (e.g., asthma attack either during or shortly after a high $PM_{2.5}$ day). To compare/correlate ambient concentrations with acute health effects, daily local air quality data is needed[5]. Spatial gaps exist in the air quality monitoring network, especially in rural areas, since the air quality monitoring network is designed to focus on measurement of pollutant concentrations in high population density areas. Temporal limits also exist. Hourly ozone measurements are aggregated to daily values (the daily max 8-hour average is relevant to the ozone standard). Ozone is typically monitored during the ozone season (the warmer months, approximately April through October). However, year-long data is available in many areas and is extremely useful to evaluate whether ozone is a factor in health outcomes during the non-ozone seasons. $PM_{2.5}$ is generally measured year-round. Most Federal Reference Method (FRM) $PM_{2.5}$ monitors collect data one day in every three days, due in part to the time and costs involved in collecting and analyzing the

[5] EPA uses exposure models to evaluate the health risks and environmental effects associated with exposure. These models are limited by the availability of air quality estimates. https://www.epa.gov/technical-air-pollution-resources.

samples. However, over the past several years, continuous monitors, which can automatically collect, analyze, and report $PM_{2.5}$ measurements on an hourly basis, have been introduced. These monitors are available in most of the major metropolitan areas. Some of these continuous monitors have been determined to be equivalent to the FRM monitors for regulatory purposes and are called FEM (Federal Equivalent Methods).

2.2.4 Use of Air Quality Monitoring Data

Air quality monitoring data has been used to provide the information for the following situations:

(1) Assessing effectiveness of SIPs in addressing NAAQS nonattainment areas
(2) Characterizing local, state, and national air quality status and trends
(3) Associating health and environmental damage with air quality levels/concentrations

For the EPHT effort, EPA is providing air quality data to support efforts associated with (2), and (3) above. Data supporting (3) is generated by EPA through the use of its air quality data and its downscaler model.

Most studies that associate air quality with health outcomes use air monitoring as a surrogate for exposure to the air pollutants being investigated. Many studies have used the monitoring networks operated by state and federal agencies. Some studies perform special monitoring that can better represent exposure to the air pollutants: community monitoring, near residences, in-house or work place monitoring, and personal monitoring. For the EPHT program, special monitoring is generally not supported, though it could be used on a case-by-case basis.

From proximity-based exposure estimates to statistical interpolation, many approaches are developed for estimating exposures to air pollutants using ambient monitoring data (Jerrett et al., 2005). Depending upon the approach and the spatial and temporal distribution of ambient monitoring data, exposure estimates to air pollutants may vary greatly in areas further apart from monitors (Bravo et al., 2012). Factors like limited temporal coverage (i.e., $PM_{2.5}$ monitors do not operate continuously such as recording every third day or ozone monitors operate only certain part of the year) and limited spatial coverage (i. e., most monitors are located in urban areas and rural coverage is limited) hinder the ability of most of the interpolation techniques that use monitoring data alone as the input. If we look at the example of Voronoi Neighbor Averaging (VNA) (referred as the Nearest Neighbor Averaging in most literature), rural estimates would be biased towards the urban estimates. To further explain this point, assume the scenario of two cities with monitors and no monitors in the rural areas between, which is very plausible. Since exposure estimates are guaranteed to be within the range of monitors in VNA, estimates for the rural areas would be higher according to this scenario.

Air quality models may overcome some of the limitations that monitoring networks possess. Models such as the Community Multi-Scale Air Quality (CMAQ) modeling systems can estimate concentrations in reasonable temporal and spatial resolutions. However, these sophisticated air quality models are prone to systematic biases since they depend upon so many variables (i.e., metrological models and emission models) and complex chemical and physical process simulations.

Combining monitoring data with air quality models (via fusion or regression) may provide the best results in terms of estimating ambient air concentrations in space and time. EPA's eVNA[6] is an example of an earlier approach for merging air quality monitor data with CMAQ model predictions. The downscaler model attempts to address some of the shortcomings in these earlier attempts to statistically combine monitor and model predicted data, see published paper referenced in section 1 for more information about the downscaler model. As discussed in the next section, there are two methods used in EPHT to provide estimates of ambient concentrations of air pollutants: air quality monitoring data and the downscaler model estimate, which is a statistical 'combination' of air quality monitor data and photochemical air quality model predictions (e.g., CMAQ).

2.3 Air Quality Indicators Developed for the EPHT Network

Air quality indicators have been developed for use in the Environmental Public Health Tracking Network by CDC using the ozone and $PM_{2.5}$ data from EPA. The approach used divides "indicators" into two categories. First, basic air quality measures were developed to compare air quality levels over space and time within a public health context (e.g., using the NAAQS as a benchmark). Next, indicators were developed that mathematically link air quality data to public health tracking data (e.g., daily $PM_{2.5}$ levels and hospitalization data for acute myocardial infarction). Table 2-3 and Table 2-4 describe the issues impacting calculation of basic air quality indicators.

Table 2-2. Public Health Surveillance Goals and Current Status

Goal	Status
Air data sets and metadata required for air quality indicators are available to EPHT state Grantees.	AQS data are available through state agencies and EPA's Air Quality System (AQS). EPA and CDC developed an interagency agreement, where EPA provides air quality data along with statistically combined AQS and Community Multiscale Air Quality (CMAQ) Model data, associated metadata, and technical reports that are delivered to CDC.
Estimate the linkage or association of $PM_{2.5}$ and ozone on health to: Identify populations that may have higher risk of adverse health effects due to $PM_{2.5}$ and ozone, Generate hypothesis for further research, and Provide information to support prevention and pollution control strategies.	Regular discussions have been held on health-air linked indicators and CDC/HFI/EPA convened a workshop January 2008. CDC has collaborated on a health impact assessment (HIA) with Emory University, EPA, and state grantees that can be used to facilitate greater understanding of these linkages.
Produce and disseminate basic indicators and other findings in electronic and print formats to provide the public, environmental health professionals, and policymakers, with current and easy-to-use information about air pollution and the impact on public health.	Templates and "how to" guides for $PM_{2.5}$ and ozone have been developed for routine indicators. Calculation techniques and presentations for the indicators have been developed.

[6] eVNA is described in the "Regulatory Impact Analysis for the Final Clean Air Interstate Rule", EPA-452/R-05-002, March 2005, Appendix F.

Table 2-3. Basic Air Quality Indicators used in EPHT, derived from the EPA data delivered to CDC

Ozone (daily 8-hr period with maximum concentration—ppm—by Federal Reference Method (FRM))
• Number of days with maximum ozone concentration over the NAAQS (or other relevant benchmarks (by county and MSA) • Number of person-days with maximum 8-hr average ozone concentration over the NAAQS & other relevant benchmarks (by county and MSA)
$PM_{2.5}$ (daily 24-hr integrated samples –ug/m^3-by FRM)
• Average ambient concentrations of particulate matter (< 2.5 microns in diameter) and compared to annual $PM_{2.5}$ NAAQS (by state). • % population exceeding annual $PM_{2.5}$ NAAQS (by state). • % of days with $PM_{2.5}$ concentration over the daily NAAQS (or other relevant benchmarks (by county and MSA) • Number of person-days with $PM_{2.5}$ concentration over the daily NAAQS & other relevant benchmarks (by county and MSA)

2.3.1 Rationale for the Air Quality Indicators

The CDC EPHT Network is initially focusing on ozone and $PM_{2.5}$. These air quality indicators are based mainly around the NAAQS health findings and program-based measures (measurement, data and analysis methodologies). The indicators will allow comparisons across space and time for EPHT actions. They are in the context of health-based benchmarks. By bringing population into the measures, they roughly distinguish between potential exposures (at broad scale).

2.3.2 Air Quality Data Sources

The air quality data will be available in the US EPA Air Quality System (AQS) database based on the state/federal air program's data collection and processing. The AQS database contains ambient air pollution data collected by EPA, state, local, and tribal air pollution control agencies from thousands of monitoring stations (SLAMS).

2.3.3 Use of Air Quality Indicators for Public Health Practice

The basic indicators will be used to inform policymakers and the public regarding the degree of hazard within a state and across states (national). For example, the number of days per year that ozone is above the NAAQS can be used to communicate to sensitive populations (such as asthmatics) the number of days that they may be exposed to unhealthy levels of ozone. This is the same level used in the Air Quality Alerts that inform these sensitive populations when and how to reduce their exposure. These indicators, however, are not a surrogate measure of exposure and therefore will not be linked with health data.

3.0 Emissions Data

3.1 Introduction to Emissions Data Development

The U.S. Environmental Protection Agency (EPA) developed an air quality modeling platform based primarily on the 2014 National Emissions Inventory (NEI), Version 2 to process year 2015 emission data for this project. This section provides a summary of the emissions inventory and emissions modeling techniques applied to Criteria Air Pollutants (CAPs) and the following select Hazardous Air Pollutants (HAPs) included in the modeling platform: chlorine (Cl), hydrogen chloride (HCl), benzene, acetaldehyde, formaldehyde, napthalene and methanol. This section also describes the approach and data used to produce emissions inputs to the air quality model. The air quality modeling, meteorological inputs and boundary conditions are described in a separate section.

The Community Multiscale Air Quality (CMAQ) model (https://www.epa.gov/cmaq) was used to model ozone (O_3) and particulate matter (PM) for this project. CMAQ requires hourly and gridded emissions of the following inventory pollutants: carbon monoxide (CO), nitrogen oxides (NO_X), volatile organic compounds (VOC), sulfur dioxide (SO_2), ammonia (NH_3), particulate matter less than or equal to 10 microns (PM_{10}), and individual component species for particulate matter less than or equal to 2.5 microns ($PM_{2.5}$). In addition, the Carbon bond version 6 (CB6) with chlorine chemistry used here within CMAQ allows for explicit treatment of the VOC HAPs naphthalene, benzene, acetaldehyde, formaldehyde and methanol (NBAFM) and includes anthropogenic HAP emissions of HCl and Cl.

The effort to create the 2015 emission inputs for this study included development of emission inventories for input to a 2015 modeling case, along with application of emissions modeling tools to convert the inventories into the format and resolution needed by CMAQ. Year-specific fire and continuous emission monitoring (CEM) data for electric generating units (EGUs) were used. The primary emissions modeling tool used to create the CMAQ model-ready emissions was the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. SMOKE version 4.5 was used to create CMAQ-ready emissions files for a 12-km national grid. Additional information about SMOKE is available from http://www.cmascenter.org/smoke.

This chapter contains two additional sections. Section 3.2 describes the inventories input to SMOKE and the ancillary files used along with the emission inventories. Section 3.3 describes the emissions modeling performed to convert the inventories into the format and resolution needed by CMAQ.

3.2 Emission Inventories and Approaches

This section describes the emissions inventories created for input to SMOKE. The 2014 NEI, version 2 with some updates for 2015 is the primary basis for the inputs to SMOKE. The NEI includes five main data categories: a) nonpoint (formerly called “stationary area”) sources; b) point sources; c) nonroad mobile sources; d) onroad mobile sources; and e) fires. For CAPs, the NEI data are largely compiled from data submitted by state, local and tribal (S/L/T) agencies. HAP emissions data are often augmented by EPA when they are not voluntarily submitted to the NEI by S/L/T agencies. The NEI was compiled using the Emissions Inventory System (EIS). EIS includes hundreds of automated QA checks to improve data

quality, and it also supports release point (stack) coordinates separately from facility coordinates. EPA collaboration with S/L/T agencies helped prevent duplication between point and nonpoint source categories such as industrial boilers. The 2014 NEIv2 Technical Support Document is available at https://www.epa.gov/air-emissions-inventories/2014-national-emissions-inventory-nei-technical-support-document-tsd (EPA, 2018a).

Point source data for the year 2015 as submitted to EIS were used for this study, with emissions for any units not submitted nor marked as closed pulled forward from the 2014NEIv2. EPA used the SMARTFIRE2 system to develop 2015 fire emissions. SMARTFIRE2 categorizes all fires as either prescribed burning or wildfire categories, and includes improved emission factor estimates for prescribed burning. Onroad mobile source emissions for year 2015 were developed using MOVES2014a. Nonroad mobile source emissions were developed by running MOVES2014a for years 2014 and 2016, and then interpolating to 2015. Canadian and Mexican emissions were interpolated to year 2015.

The methods used to process emissions for this study are similar to those documented for EPA's Version 7.1, 2014 Emissions Modeling Platform that was also used for version 2 of the 2014 National Air Toxics Assessment (NATA), with two exceptions. One exception is that many fewer HAPs are included in this platform. Also, many emissions inventories and inputs were updated to the year 2015 for this study. A technical support document (TSD) for the 2014v7.1 platform is available here https://www.epa.gov/air-emissions-modeling/2014-version-71-technical-support-document-tsd (EPA, 2018b) and includes additional details regarding the data preparation and emissions modeling, with the exception of the HAP speciation and any updates specific to 2015.

The emissions modeling process, performed using SMOKE v4.5, apportions the emissions inventories into the grid cells used by CMAQ and temporalizes the emissions into hourly values. In addition, the pollutants in the inventories (e.g., NOx, PM and VOC) are split into the chemical species needed by CMAQ. For the purposes of preparing the CMAQ- ready emissions, the NEI emissions inventories by data category are split into emissions modeling "platform" sectors; and emissions from sources other than the NEI are added, such as the Canadian, Mexican, and offshore inventories. Emissions sectors within the emissions modeling platform are separated out from each other when the emissions for that sector are run through all of the SMOKE programs, except the final merge, independently from emissions in the other sectors. The final merge program called Mrggrid combines the sector-specific gridded, speciated and temporalized emissions to create the final CMAQ-ready emissions inputs. For biogenic emissions, the CMAQ model allows for biogenic emissions to be included in the CMAQ-ready emissions inputs, or for biogenic emissions to be computed within CMAQ itself (the "inline" option). This study uses the inline biogenics option.

Table 3-1 presents the sectors in the emissions modeling platform used to develop the year 2015 emissions for this project. The sector abbreviations are provided in italics; these abbreviations are used in the SMOKE modeling scripts, the inventory file names, and throughout the remainder of this section. Annual 2015 emission summaries for the U.S. anthropogenic sectors are shown in Table 3-2 (i.e., biogenic emissions are excluded). Table 3-3 provides a summary of emissions for the anthropogenic sectors containing Canadian, Mexican and offshore sources. State total emissions for each sector are provided in Appendix B, a workbook entitled "Appendix_B_2015_emissions_totals_by_sector.xlsx".

Table 3-1. Platform Sectors Used in the Emissions Modeling Process

2014 Platform Sector *(Abbrev)*	NEI Category	Description and resolution of the data input to SMOKE
EGUs (*ptegu)*	Point	2015 point source EGUs, replaced with hourly 2015 Continuous Emissions Monitoring System (CEMS) values for NOX and SO2 where the units are matched to the NEI. Emissions for all sources not matched to CEMS data come from 2015 NEI point inventory. Annual resolution for sources not matched to CEMS data, hourly for CEMS sources.
Point source oil and gas (*pt_oilgas)*	Point	2015 NEI point sources that include oil and gas production emissions processes based on facilities with the following NAICS: 211* (Oil and Gas Extraction), 2212* (Natural Gas Distribution), 213111 (Drilling Oil and Gas Wells), 213112 (Support Activities for Oil and Gas Operations), 4861* (Pipeline Transportation of Crude Oil), 4862* (Pipeline Transportation of Natural Gas). Includes U.S. offshore oil production. The portion of the 2015 NEI point inventory oil and gas inventory that was carried forward from 2014NEIv2 (i.e. not updated to 2015 in EIS) was projected to year 2015 estimates. Annual resolution.
Remaining non-EGU point (*ptnonipm*)	Point	All 2015 NEI point source records not matched to the ptegu or pt_oilgas sectors. Includes all aircraft and airport ground support emissions and some rail yard emissions. Annual resolution.
Point source fire (*ptfire*)	Fires	Point source day-specific wildfires and prescribed fires for 2015 computed using SMARTFIRE 2. Fires over 20,000 acres on a single day allocated to overlapping grid cells.
Point Source agricultural fires (ptagfire)	Nonpoint	Agricultural fire sources that were developed by EPA as point and day-specific emissions; they were put into the nonpoint data category of the NEI, but in the platform, they are treated as point sources.
Agricultural (*ag*)	Nonpoint	2014NEIv2 nonpoint livestock and fertilizer application emissions. Livestock includes ammonia and other pollutants (except PM2.5). Fertilizer includes only ammonia. County and annual resolution.
Area fugitive dust (*afdust_adj*)	Nonpoint	PM_{10} and $PM_{2.5}$ fugitive dust sources from the 2014NEIv2 nonpoint inventory; including building construction, road construction, agricultural dust, and road dust. The emissions modeling adjustment applies a transport fraction and a zero-out based on 2015 meteorology (precipitation and snow/ice cover). County and annual resolution.
Biogenic (*beis*)	Nonpoint	Biogenic emissions were left out of the CMAQ-ready merged emissions, in favor of inline biogenics produced during the CMAQ model run itself.
C1 and C2 commercial marine (*cmv_c1c2*)	Nonpoint	2014NEIv2 Category 1 (C1) and Category 2 (C2), commercial marine vessel (CMV) emissions, with SO_2 emissions in the North American Emission Control Area (ECA) reduced by 90% compared to 2014NEIv2. County and annual resolution.

2014 Platform Sector *(Abbrev)*	NEI Category	Description and resolution of the data input to SMOKE
C3 commercial marine (*cmv_c3*)	Nonpoint	Within state and federal waters, 2014NEIv2 Category 3 commercial marine vessel (CMV) emissions, with SO_2 emissions in the North American Emission Control Area (ECA) reduced by 90% compared to 2014NEIv2. Outside of state and federal waters, emissions are based on the Emissions Control Area (ECA) inventory. Point (to allow for plume rise) and annual resolution.
Remaining nonpoint (*nonpt*)	Nonpoint	2014NEIv2 nonpoint sources not included in other platform sectors. County and annual resolution.
Nonpoint source oil and gas (*np_oilgas)*	Nonpoint	2014NEIv2 nonpoint sources from oil and gas-related processes, projected to year 2015 estimates. County and annual resolution. Rail locomotives emissions from the 2014NEIv2. County and annual resolution.
Locomotive (*rail*)	Nonpoint	2014NEIv2 nonpoint sources with residential wood combustion (RWC) processes. County and annual resolution.
Residential Wood Combustion (*rwc*)	Nonpoint	2015 nonroad equipment emissions developed with the MOVES2014a. MOVES was used for all states except California, which submitted their own emissions for the 2014NEIv2 and for the year 2017, from which 2015 estimates were interpolated. County and monthly resolution.
Nonroad (*nonroad*)	Nonroad	2015 onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles. Includes the following modes: exhaust, extended idle, auxiliary power units, evaporative, permeation, refueling, and brake and tire wear. For all states except California, developed using winter and summer MOVES emission factors tables produced by MOVES2014a.
Onroad (*onroad*)	Onroad	California-provided CAP and metal HAP onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles based on Emission Factor (EMFAC), gridded and temporalized using MOVES2014a. Volatile organic compound (VOC) HAP emissions derived from California-provided VOC emissions and MOVES-based speciation. California estimates for 2014 and 2017 were interpolated to 2015 values.
Onroad California (*onroad_ca_adj*)	Onroad	2014NEIv2 nonpoint sources not included in other platform sectors. County and annual resolution.
Onroad Canada (*onroad_can*)	Non-US	Monthly onroad mobile inventory for Canada (province resolution), with year 2015 emissions values interpolated from 2013 and 2025 inventories.
Onroad Mexico (*onroad_mex*)	Non-US	Monthly onroad mobile inventory for Mexico (municipio resolution), with 2015 emissions values interpolated from 2014 and 2018 inventories.

2014 Platform Sector *(Abbrev)*	NEI Category	Description and resolution of the data input to SMOKE
Other area fugitive dust sources	Non-US	Area fugitive dust sources from Canada, with 2015 emissions values interpolated from 2013 and 2025 inventories, and with transport fraction and snow/ice adjustments based on 2015 meteorological data. Annual and province resolution.
Other nonpoint and nonroad (*othar*)	Non-US	Year 2015 Canada (province resolution, interpolated from 2013 and 2025 values) and projected year 2015 Mexico (municipio resolution, interpolated from 2014 and 2018 values) nonpoint and nonroad mobile inventories, annual resolution.
Other point sources not from the NEI (*othpt*)	Non-US	Canada point source emissions for 2015 (interpolated from 2013 and 2025), and Mexico point source emissions for 2015 (interpolated from 2014 and 2018). Annual resolution.
Point fires in Mexico and Canada (ptfire_mxca)	Non-US	Point source day-specific wildfires and prescribed fires for 2015 are computed from SMARTFIRE 2 in Canada and Mexico. Caribbean, Central American, and other international fires are from 2015 v1.5 of the Fire INventory (FINN) from National Center for Atmospheric Research (NCAR) fires (NCAR, 2016 and Wiedinmyer, C., 2011).

Table 3-2. 2015 Continental United States Emissions by Sector (tons/yr in 48 states + D.C.)

Sector	CO	NH_3	NO_x	PM_{10}	$PM_{2.5}$	SO_2	VOC
afdust_adj				6,093,367	857,261		
ag		2,823,395					179,970
cmv_c1c2	47,183	120	260,338	6,493	6,168	345	4,840
cmv_c3	10,885	25	108,268	4,248	3,832	3,883	5,043
nonpt	2,680,775	121,229	758,152	608,827	496,454	162,231	3,672,687
np_oilgas	686,168	15	719,934	17,746	17,480	38,963	3,206,411
nonroad	12,285,118	2,244	1,292,956	131,083	123,997	2,776	1,546,314
onroad	23,064,322	104,472	4,401,420	285,167	144,312	27,173	2,199,205
ptagfire	382,760	53,353	11,971	62,034	43,724	3,719	23,711
ptfire	21,180,425	347,360	275,352	2,142,471	1,815,654	154,996	4,993,305
ptegu	639,943	20,213	1,494,941	180,333	139,355	2,346,129	34,558
ptnonipm	1,953,514	72,943	1,080,957	414,529	270,208	769,257	833,137
pt_oilgas	190,337	1,244	390,734	12,372	11,856	43,422	142,197
rail	118,367	363	672,558	20,728	19,154	700	34,739
rwc	2,098,907	15,331	30,493	314,466	313,945	7,684	338,465
Continental U.S.	**65,338,705**	**738,912**	**11,498,073**	**4,200,496**	**3,406,139**	**3,561,278**	**17,034,613**

Table 3-3. 2015 Non-US Emissions by Sector within Modeling Domain (tons/yr for Canada, Mexico, Offshore)

Sector	CO	NH_3	NO_X	PM_{10}	$PM_{2.5}$	SO_2	VOC
Canada othafdust				2,297,778	449,354		
Canada othar	2,928,791	497,760	609,977	425,349	235,680	33,801	1,135,610
Canada onroad_can	1,978,610	8,272	436,083	26,187	19,376	1,465	172,524
Canada othpt	1,130,185	18,181	602,460	89,118	47,466	892,133	786,582
Canada ptfire_othna	10,277,333	290,735	339,883	1,216,218	1,109,341	85,204	2,853,915
Canada Subtotal	**16,314,918**	**814,948**	**1,988,403**	**4,054,649**	**1,861,217**	**1,012,603**	**4,948,631**
Mexico othar	236,143	203,945	216,175	114,754	53,727	7,661	512,070
Mexico onroad_mex	1,825,267	2,724	437,330	14,935	10,744	6,047	158,562
Mexico othpt	196,410	4,851	456,220	72,957	57,378	509,144	68,615
Mexico ptfire_othna	81,991	1,390	7,168	10,654	8,557	641	28,294
Mexico Subtotal	**2,339,811**	**212,911**	**1,116,894**	**213,300**	**130,406**	**523,494**	**767,541**
Offshore cmv_c1c2	56,393	184	283,431	9,193	8,918	224	5,248
Offshore cmv_c3	77,449	68	854,639	47,136	43,550	254,620	34,059
Offshore pt_oilgas	50,046	15	48,688	668	666	502	48,167
2015 Total non-U.S.	**18,761,168**	**1,028,058**	**3,437,416**	**4,277,811**	**2,001,208**	**1,536,822**	**5,769,587**

3.2.1 Point Sources (ptegu, pt_oilgas and ptnonipm)

Point sources are sources of emissions for which specific geographic coordinates (e.g., latitude/longitude) are specified, as in the case of an individual facility. A facility may have multiple emission release points that may be characterized as units such as boilers, reactors, spray booths, kilns, etc. A unit may have multiple processes (e.g., a boiler that sometimes burns residual oil and sometimes burns natural gas). With a couple of minor exceptions, this section describes only NEI point sources within the contiguous U.S. The offshore oil platform (pt_oilgas sector) and category 3 CMV emissions (cmv_c3 sector) are processed by SMOKE as point source inventories and are discussed later in this section. A complete NEI is developed every three years, with 2014 being the most recently finished complete NEI. A comprehensive description about the development of the 2014NEIv2 is available in the 2014NEIv2 TSD (EPA, 2018a). Point inventories are also available in EIS for intermediate years such as 2015. In this intermediate point inventory, larger sources are updated with emissions for year 2015, while sources not updated by state with 2015 values are either carried forward from 2014NEIv2 or are closed.

In preparation for modeling, the complete set of point sources in the NEI was exported from EIS for the year 2015 into the Flat File 2010 (FF10) format that is compatible with SMOKE (see https://www.cmascenter.org/smoke/documentation/4.5/html/ch08s02s08.html) and was then split into several sectors for modeling. After dropping sources without specific locations (i.e., the FIPS code ends in 777), initial versions of inventories for the other three point source sectors were created from the remaining 2015 point sources. The point sectors are: EGUs (ptegu), point source oil and gas extraction-related sources (pt_oilgas) and the remaining non-EGUs (ptnonipm). The EGU emissions are split out from the other sources to facilitate the use of distinct SMOKE temporal processing and future-year projection techniques. The oil and gas sector emissions (pt_oilgas) were processed separately for summary tracking purposes and distinct projection techniques from the remaining non-EGU emissions (ptnonipm).

The inventory pollutants processed through SMOKE for the ptegu, pt_oilgas, and ptnonipm sectors were: CO, NOX, VOC, SO2, NH3, PM10, and PM2.5 and the following HAPs: HCl (pollutant code = 7647010), and Cl (code = 7782505). NBAFM pollutants from the point sectors were not utilized because VOC was speciated without the use (i.e., integration) of VOC HAP pollutants from the inventory.

The ptnonipm and pt_oilgas sector emissions were provided to SMOKE as annual emissions. For sources in the ptegu sector that could be matched to 2015 CEMS data, hourly CEMS NO_X and SO_2 emissions for 2015 from EPA's Acid Rain Program were used rather than annual inventory emissions. For all other pollutants (e.g., VOC, $PM_{2.5}$, HCl), annual emissions were used as-is from the annual inventory, but were allocated to hourly values using heat input from the CEMS data. For the unmatched units in the ptegu sector, annual emissions were allocated to daily values using IPM region- and pollutant-specific profiles, and similarly, region- and pollutant-specific diurnal profiles were applied to create hourly emissions.

The non-EGU stationary point source (ptnonipm) emissions were input to SMOKE as annual emissions. The full description of how the NEI emissions were developed is provided in the NEI documentation, but a brief summary of their development follows:

a. CAP and HAP data were provided by States, locals and tribes under the Air Emissions Reporting Rule (AERR) [the reporting size threshold is larger for inventory years between the triennial inventory years of 2011, 2014, 2017, …]
b. EPA corrected known issues and filled PM data gaps.

c. EPA added HAP data from the Toxic Release Inventory (TRI) where corresponding data was not already provided by states/locals.
d. EPA stores and applies matches of the point source units to units with CEMS data and also for all EGU units modeled by EPA's Integrated Planning Model (IPM).
e. EPA provided data for airports and rail yards.
f. Off-shore platform data were added from the Bureau of Ocean Energy Management (BOEM).

The changes made to the NEI point sources prior to modeling with SMOKE are as follows:

- The tribal data, which do not use state/county Federal Information Processing Standards (FIPS) codes in the NEI, but rather use the tribal code, were assigned a state/county FIPS code of 88XXX, where XXX is the 3-digit tribal code in the NEI. This change was made because SMOKE requires all sources to have a state/county FIPS code.
- Sources that did not have specific counties assigned (i.e., the county code ends in 777) were not included in the modeling because it was only possible to know the state in which the sources resided, but no more specific details related to the location of the sources were available.
- Stack parameters for point sources missing this information were filled in prior to modeling in SMOKE.

Each of the point sectors is processed separately through SMOKE as described in the following subsections.

3.2.1.1 EGU sector (ptegu)

The ptegu sector contains emissions from EGUs in the 2015 point source inventory that could be matched to units found in the National Electric Energy Database System (NEEDS) v5.16 that is used by the Integrated Planning Model (IPM) to develop future year EGU emissions. It was necessary to put these EGUs into a separate sector in the platform because EGUs use different temporal profiles than other sources in the point sector and it is useful to segregate these emissions from the rest of the point sources to facilitate summaries of the data. Sources not matched to units found in NEEDS are placed into the pt_oilgas or ptnonipm sectors. For studies with future year cases, the sources in the ptegu sector are fully replaced with the emissions output from IPM. It is therefore important that the matching between the NEI and NEEDS database be as complete as possible because there can be double-counting of emissions in future year modeling scenarios if emissions for units are projected by IPM are not properly matched to the units in the point source inventory.

Some units in the ptegu sector are matched to CEMS data via ORIS facility codes and boiler ID. For these units, SMOKE replaces the emissions of NO_X and SO2 with the CEMS emissions, thereby ignoring the annual values specified in the point source inventory. For other pollutants, the hourly CEMS heat input data are used to allocate the ptegu inventory annual emissions to hourly values. All stack parameters, stack locations, and SCC codes for these sources come from the point source inventory. Because these attributes are obtained from the inventory, the chemical speciation of VOC and PM2.5 for the sources is selected based on the SCC or in some cases, based on unit-specific data. If CEMS data exists for a unit, but the unit is not matched to the inventory, the CEMS data for that unit is not used in the modeling platform. However, if the source exists in the inventory and is not matched to a CEMS unit, the emissions from that source would be modeled using the annual emission value in the inventory and would be

allocated to daily values using region-, fuel- and pollutant-specific average profiles. EIS stores many matches from EIS units to the ORIS facility codes and boiler IDs used to reference the CEMS data. Some additional matches were made at the release point level in the emissions modeling platform. This study expanded on the matching effort compared to earlier 2015 emissions studies. For example, in instances where multiple ORIS boiler IDs are matched to a single EIS unit, the EIS unit was split into multiple units in the inventory to allow for a complete allocation of the CEMS data from all of the boilers matched to that unit. This study also used the most recent 2015 CEMS data available at the time the emissions were compiled, published on March 14, 2018.

3.2.1.2 Point Oil and Gas Sector (pt_oilgas)

The pt_oilgas sector was separated from the ptnonipm sector by selecting sources with specific North American Industry Classification System (NAICS) codes shown in Table 3-4. The emissions and other source characteristics in the pt_oilgas sector are submitted by states, while EPA developed a dataset of nonpoint oil and gas emissions for each county in the U.S. with oil and gas activity that was available for states to use. Nonpoint oil and gas emissions can be found in the np_oilgas sector. More information on the development of the 2014 oil and gas emissions can be found in Section 4.16 of the 2014NEIv2 TSD. The pt_oilgas sector includes emissions from offshore oil platforms.

Table 3-4. Point source oil and gas sector NAICS Codes

NAICS	NAICS description
2111	Oil and Gas Extraction
2212	Natural Gas Distribution
4862	Pipeline Transportation of Natural Gas
21111	Oil and Gas Extraction
22121	Natural Gas Distribution
48611	Pipeline Transportation of Crude Oil
48621	Pipeline Transportation of Natural Gas
211111	Crude Petroleum and Natural Gas Extraction
211112	Natural Gas Liquid Extraction
213111	Drilling Oil and Gas Wells
213112	Support Activities for Oil and Gas Operations
221210	Natural Gas Distribution
486110	Pipeline Transportation of Crude Oil
486210	Pipeline Transportation of Natural Gas

The pt_oilgas inventory is a combination of sources with updated year 2015 emissions and sources with emissions carried forward from 2014NEIv2 with no updates. For this study, sources already updated for the year 2015 were used as-is. The emissions carried forward from 2014NEIv2 were projected to 2015. Projection factors for 2015 are based on historical state crude and natural gas production data from the U.S. Energy Information Administration (EIA), which is available at these two links: http://www.eia.gov/dnav/ng/ng_sum_lsum_a_epg0_fgw_mmcf_a.htm; http://www.eia.gov/dnav/pet/pet_crd_crpdn_adc_mbbl_a.htm. Separate factors are calculated for each state, and for sources related to oil production, gas production, or a combination of oil and gas. These factors, which are listed in Table 3-5, were applied to CO, NOx, and VOC emissions only from sources

carried forward from the 2014NEIv2 pt_oilgas inventory. The table does not list every state; emissions in states that do not have projection factors listed were held constant. The complete 2015 pt_oilgas inventory used for this study consists of both sources already updated to 2015 within EIS (used directly), and sources carried forward from 2014NEIv2 (projected to 2015).

Table 3-5. Oil and gas sector 2015 projection factors

State	Oil projection factor	Gas projection factor	"Both" projection factor
Alabama	0.987	0.929	0.958
Alaska	0.973	1.002	0.988
Arizona	0.661	0.896	0.778
Arkansas	0.926	0.900	0.913
California	0.983	0.990	0.987
Colorado	1.284	1.028	1.156
Florida	0.991	1.822	1.407
Illinois	0.997	1.078	1.038
Indiana	0.885	1.096	0.990
Kansas	0.918	0.992	0.955
Kentucky	0.848	1.030	0.939
Louisiana	0.915	0.921	0.918
Maryland	1.000	1.900	1.900
Michigan	0.881	0.936	0.909
Mississippi	1.023	1.069	1.046
Missouri	0.760	0.333	0.547
Montana	0.955	0.984	0.970
Nebraska	0.950	1.144	1.047
Nevada	0.889	1.333	1.111
New Mexico	1.182	1.024	1.103
New York	0.798	0.858	0.828
North Dakota	1.088	1.262	1.175
Ohio	1.788	1.966	1.877
Oklahoma	1.121	1.072	1.097
Oregon	1.000	0.743	0.743
Pennsylvania	1.034	1.130	1.082
South Dakota	0.927	0.948	0.937
Tennessee	0.897	0.808	0.852
Texas	1.089	1.016	1.053
Utah	0.908	0.917	0.913
Virginia	0.786	0.955	0.870
West Virginia	1.087	1.233	1.160
Wyoming	1.136	0.999	1.067

3.2.1.3 Non-IPM Sector (ptnonipm)

Except for some minor exceptions, the non-IPM (ptnonipm) sector contains the point sources that are not in the ptegu or pt_oilgas sectors. For the most part, the ptnonipm sector reflects the non-EGU sources of the 2015 NEI point inventory; however, it is likely that some low-emitting EGUs not matched to the NEEDS database or to CEMS data are in the ptnonipm sector.

The ptnonipm sector contains a small amount of fugitive dust PM emissions from vehicular traffic on paved or unpaved roads at industrial facilities, coal handling at coal mines, and grain elevators. Sources with state/county FIPS code ending with "777" are in the NEI but are not included in any modeling sectors. These sources typically represent mobile (temporary) asphalt plants that are only reported for some states, and are generally in a fixed location for only a part of the year and are therefore difficult to allocate to specific places and days as is needed for modeling. Therefore, these sources are dropped from the point-based sectors in the modeling platform.

3.2.2 Day-Specific Point Source Fires (ptfire)

Wildfire and prescribed burning emissions are contained in the ptfire sector. The ptfire sector has emissions provided at geographic coordinates (point locations) and has daily emissions values. The ptfire sector excludes agricultural burning and other open burning sources that are included in the ptagfire sector. Emissions are day-specific and include satellite-derived latitude/longitude of the fire's origin and other parameters associated with the emissions such as acres burned and fuel load, which allow estimation of plume rise.

The point source day-specific emission estimates for 2015 fires were developed using SMARTFIRE 2 (Sullivan, et al., 2008), which uses the National Oceanic and Atmospheric Administration's (NOAA's) Hazard Mapping System (HMS) fire location information as input. Additional inputs include the CONSUME v4.1 software application (Joint Fire Science Program, 2009) and the Fuel Characteristic Classification System (FCCS) fuel-loading database to estimate fire emissions from wildfires and prescribed burns on a daily basis. The method involves the reconciliation of ICS-209 reports (Incident Status Summary Reports), GeoMAC perimeter Shapefiles, USFS fire information, and USFWS fire information data with satellite-based fire detections to determine spatial and temporal information about the fires. A functional diagram of the SMARTFIRE 2 process of reconciling fires with ICS-209 reports is available in the documentation (Raffuse, et al., 2007). Once the fire reconciliation process is completed, the emissions are calculated using the U.S. Forest Service's CONSUME v4.1 fuel consumption model and the FCCS v2 fuel-loading database in the BlueSky Framework (Ottmar, et. al., 2007).

A difference between the fires for this study and those in the NEI is that the proportion of emissions allocated to flaming versus smoldering SCCs were adjusted. Flaming fractions were calculated for each fire based on the flaming and smoldering consumption divided by the total consumption. Smoldering fractions were calculated by dividing the residual consumption by the total consumption. The fractions were then applied to the 2015 fire emissions to obtain revised emissions for the flaming and smoldering SCCs. The total emissions by state were unchanged, but they were reapportioned to the flaming and smoldering SCCs to facilitate a more realistic plume rise for fires.

Large fires of more than 20,000 acres in a single day were split using GeoMAC (https://www.geomac.gov/) fire shapes, where available, or otherwise using a circle centered on the detect lat/lon based on 12US2 grid cell overlap. The resulting split fires have emissions and area apportioned from the original fire into the grid cells based on fraction of area overlap between the fire shape and the cell. The idea is to prevent all of the emissions from a very large fire from going into a single grid cell,

when in reality the fire emissions were more dispersed than a single point. The area of each of the "subfires" was computed in proportion to the overlap with that grid cell. These "subfires" were given new names that were the same as the original, but with "_a", "_b", "_c", and "_d" appended as needed.

The SMOKE-ready inventory files created from the raw daily fires contain both CAPs and HAPs. The BAFM HAP emissions from the inventory were obtained using VOC speciation profiles (i.e., a "no-integrate noHAP" use case).

3.2.3 Nonpoint Sources (afdust, ag, nonpt, np_oilgas, rwc)

Several modeling platform sectors were created from the 2014NEIv2 nonpoint inventory. This section describes the *stationary* nonpoint sources. Locomotives, C1 and C2 CMV, and C3 CMV are also included the 2014NEIv2 nonpoint data category, but are mobile sources and are described in a later section. The 2014NEIv2 TSD includes documentation for the nonpoint data. The annual emissions from all of the stationary nonpoint sectors were held at 2014 levels for this 2015 study, with the exception of np_oilgas.

The nonpoint tribal-submitted emissions are dropped during spatial processing with SMOKE due to the configuration of the spatial surrogates, which are available by county, but not at the tribal level. In addition, possible double-counting with county-level emissions is prevented. These omissions are not expected to have an impact on the results of the air quality modeling at the 12-km scales used for this platform.

In the rest of this section, each of the platform sectors into which the sources in the nonpoint NEI data category were divided is described, along with any data that were updated or replaced with non-NEI data.

3.2.3.1 Area Fugitive Dust Sector (afdust)

The area-source fugitive dust (afdust) sector contains PM_{10} and $PM_{2.5}$ emission estimates for nonpoint SCCs identified by EPA staff as dust sources. Categories included in the afdust sector are paved roads, unpaved roads and airstrips, construction (residential, industrial, road and total), agriculture production, and mining and quarrying. It does not include fugitive dust from grain elevators, coal handling at coal mines, or vehicular traffic on paved or unpaved roads at industrial facilities because these are treated as point sources so they are properly located.

The afdust sector is separated from other nonpoint sectors to allow for the application of a "transport fraction," and meteorological/precipitation reductions for the year 2015. These adjustments are applied with a script that applies land use-based gridded transport fractions followed by another script that zeroes out emissions for days on which at least 0.01 inches of precipitation occurs or there is snow cover on the ground. The land use data used to reduce the NEI emissions determines the amount of emissions that are subject to transport. This methodology is discussed in (Pouliot, et al., 2010), https://www3.epa.gov/ttn/chief/conference/ei19/session9/pouliot_pres.pdf, and in "Fugitive Dust Modeling for the 2008 Emissions Modeling Platform" (Adelman, 2012). Both the transport fraction and meteorological adjustments are based on the gridded resolution of the platform (e.g., 12km grid cells); therefore, different emissions will result if the process were applied to different grid resolutions. A limitation of the transport fraction approach is the lack of monthly variability that would be expected with seasonal changes in vegetative cover. While wind speed and direction are not accounted for in the emissions processing, the hourly variability due to soil moisture, snow cover and precipitation is accounted for in the subsequent meteorological adjustment.

For the data compiled into the 2014NEIv2, which was also used to represent 2015 in this study, meteorological adjustments are applied to paved and unpaved road SCCs but not transport adjustments. This is because the modeling platform applies meteorological adjustments and transport adjustments based on unadjusted NEI values. For the 2014NEIv2, the meteorological adjustments that were applied (to paved and unpaved road SCCs) had to be backed out in order reapply them in SMOKE. Because it was determined that some counties in the v2 did not have the adjustment applied, their emissions were used as-is. Thus, the FF10 that is run through SMOKE consists of 100% unadjusted emissions, and after SMOKE all afdust sources have both transport and meteorological adjustments applied according to year 2015 meteorology.

For categories other than paved and unpaved roads, where states submitted afdust data, it was assumed that the state-submitted data were not met-adjusted and therefore the meteorological adjustments were applied. Thus, if states submitted data that were met-adjusted for sources other than paved and unpaved roads, these sources would have been adjusted for meteorology twice. Even with that possibility, air quality modeling shows that, in general, dust is frequently overestimated in the air quality modeling results.

3.2.3.2 Agricultural Ammonia Sector (ag)

The agricultural (ag) sector includes livestock and fertilizer application emissions from the 2014NEIv2 nonpoint inventory. The livestock and fertilizer emissions in this sector are based only on the SCCs starting with 2805 and 2801. The livestock SCCs are related to beef and dairy cattle, poultry production and waste, swine production, waste from horses and ponies, and production and waste for sheep, lambs, and goats. The fertilizer SCCs consist of 15 specific types of ammonia-based fertilizer and one for miscellaneous fertilizers. The "ag" sector includes all of the NH_3 emissions from fertilizer from the NEI. However, the "ag" sector does not include all of the livestock NH_3 emissions, as there is a very small amount of NH_3 emissions from livestock in the ptnonipm inventory (as point sources) in California (883 tons; less than 0.5 percent of state total) and Wisconsin (356 tons; about 1 percent of state total). In addition to NH_3, the "ag" sector also includes livestock emissions from all pollutants other than $PM_{2.5}$. $PM_{2.5}$ from livestock are in the afdust sector.

Agricultural emissions in the platform are based on the 2014NEIv2, which is a mix of state-submitted data and EPA estimates. The EPA estimates in 2014NEIv2 were revised from 2014NEIv1, using refined methodologies and/or data for livestock and fertilizer. Livestock emissions utilized improved animal population data. VOC livestock emissions, new for this sector, were estimated by multiplying a national VOC/NH3 emissions ratio by the county NH3 emissions. The 2014NEI approach for livestock utilizes daily emission factors by animal and county from a model developed by Carnegie Mellon University (CMU) (Pinder, 2004, McQuilling, 2015) and 2012 and 2014 U.S. Department of Agriculture (USDA) agricultural census data. Details on the approach are provided in Section 4.5 of 2014NEIv2 TSD.

Annual fertilizer emissions were submitted to the 2014NEI by three states for all or part of the sector as shown in parentheses: California (57 percent), Illinois (100 percent) and Idaho (100 percent). Georgia had previously submitted data in v1 but used the EPA estimates for v2. The EPA estimates employed a methodology that uses the bidirectional (bi-di) version of CMAQ (v5.0.2) and the Fertilizer Emissions Scenario Tool for CMAQ FEST-C (v1.2). The FEST-C and CMAQ simulations were used to directly estimate emission rates based on 2014 inputs. This is a refinement from the earlier estimates that relied on emission factors calculated from a 2011 model simulation applied to 2014 FEST-C county level fertilizer

application estimates. Additionally, revised FEST-C estimates of fertilizer application were reduced for pasture and hay due to estimates of fertilizer use and hay yield being higher than USDA estimates. This resulted in a reduction of NH_3 emissions, primarily in the Southeastern U.S. Section 4.5 of the 2014NEIv2 TSD presents the updated approach.

For livestock and fertilizer, meteorological-based temporalization (described in Section 3.3.5.3) is used for month-to-day and day-to-hour temporalization. Monthly profiles are based on the daily data underlying the EPA estimates. Fertilizer uses different state-specific year-to-month profiles than livestock but uses the same meteorological-based month-to-hour profiles as livestock in the same way as was done for other recent platforms.

3.2.3.3 Agricultural fires (ptagfire)

In the NEI, agricultural fires are stored as county-annual emissions and are part of the nonpoint data category. For this study agricultural fires are modeled as day specific fires derived from satellite data for the year 2015 in a similar way to the emissions in ptfire. State-provided agricultural fire data from the 2014NEIv2 are not used in this study.

Heat flux and acres burned were provided by George Pouliot of EPA's Office of Research and Development. Based on field reconnaissance of J. McCarty (2013, personal communication), a "typical" agricultural field size was assumed for each burn location, which varied by region of the country between 40 and 80 acres. The heat flux calculation for each agricultural fire depends on estimated field size burned and the fuel loading by SCC (tons/acre). The fuel load estimate is also provided in the above spreadsheet. The ptagfire emissions estimated by the EPA are at point source and day-specific resolution. EPA data were developed using a multiple satellite detection database and crop level land use information. For the NEI, these are summed to the county and national level, but because they are computed at this finer temporal resolution, the more detailed data were used for this platform.

The agricultural fires sector includes SCCs starting with '28015'. The first three levels of descriptions for these SCCs are: 1) Fires - Agricultural Field Burning; Miscellaneous Area Sources; 2) Agriculture Production - Crops - as nonpoint; and 3) Agricultural Field Burning - whole field set on fire. The SCC 2801500000 does not specify the crop type or burn method, while the more specific SCCs specify field or orchard crops and, in some cases, the specific crop being grown. New agricultural field burning SCCs were added to the 2014 NEI to account for grass/pasture burning (also known as rangeland burning) which is included the agriculture field burning sector of the NEI.

For this modeling platform, a SMOKE update allows the use of HAP integration for speciation for PTDAY inventories. The 2015 agricultural fire inventories do not include emissions for HAPs, so HAP integration was not used for this study.

3.2.3.4 Nonpoint Oil-gas Sector (np_oilgas)

The nonpoint oil and gas (np_oilgas) sector contains onshore and offshore oil and gas emissions. The EPA estimated emissions for all counties with 2014 oil and gas activity data with the Oil and Gas Tool, and many S/L/T agencies also submitted nonpoint oil and gas data. Where S/L/T submitted nonpoint CAPS but no HAPs, the EPA augmented the HAPs using HAP augmentation factors (county and SCC level) created from the Oil and Gas Tool. The types of sources covered include drill rigs, workover rigs, artificial lift, hydraulic fracturing engines, pneumatic pumps and other devices, storage tanks, flares, truck loading, compressor engines, and dehydrators.

The 2014NEIv2 nonpoint oil and gas inventory was projected to 2015 for this study. The methodology and projection factors for np_oilgas projections were the same as for pt_oilgas, except that 2015 projections were applied to the entire 2014NEIv2 np_oilgas inventory. Projection factors for 2015 are based on the same EIA crude and natural gas production data as the point oil and gas projections discussed in Section 3.2.1.2. Separate factors are calculated for each state, and for sources related to oil production, gas production, or a combination of oil and gas. These factors, which are listed in Table 3-5, were applied to CO, NOx, and VOC emissions from the 2014NEIv2 np_oilgas inventory.

3.2.3.5 Residential Wood Combustion Sector (rwc)

The residential wood combustion (rwc) sector includes residential wood burning devices such as fireplaces, fireplaces with inserts (inserts), free standing woodstoves, pellet stoves, outdoor hydronic heaters (also known as outdoor wood boilers), indoor furnaces, and outdoor burning in firepots and chimneas. Free standing woodstoves and inserts are further differentiated into three categories: 1) conventional (not EPA certified); 2) EPA certified, catalytic; and 3) EPA certified, noncatalytic. Generally speaking, the conventional units were constructed prior to 1988. Units constructed after 1988 have to meet EPA emission standards and they are either catalytic or non-catalytic. As with the other nonpoint categories, a mix of S/L and EPA estimates were used. The EPA's estimates use updated methodologies for activity data and some changes to emission factors. For more information on the development of the residential wood combustion emissions, see Section 4.14 of the 2014NEIv2 TSD.

3.2.3.6 Other Nonpoint Sources (nonpt)

Stationary nonpoint sources that were not subdivided into the afdust, ag, np_oilgas, or rwc sectors were assigned to the "nonpt" sector. Locomotives and CMV mobile sources from the 2014NEIv2 nonpoint inventory are described with the mobile sources. The types of sources in the nonpt sector include:

- stationary source fuel combustion, including industrial, commercial, and residential and orchard heaters;
- chemical manufacturing;
- industrial processes such as commercial cooking, metal production, mineral processes, petroleum refining, wood products, fabricated metals, and refrigeration;
- solvent utilization for surface coatings such as architectural coatings, auto refinishing, traffic marking, textile production, furniture finishing, and coating of paper, plastic, metal, appliances, and motor vehicles;
- solvent utilization for degreasing of furniture, metals, auto repair, electronics, and manufacturing;
- solvent utilization for dry cleaning, graphic arts, plastics, industrial processes, personal care products, household products, adhesives and sealants;
- solvent utilization for asphalt application and roofing, and pesticide application;
- storage and transport of petroleum for uses such as portable gas cans, bulk terminals, gasoline service stations, aviation, and marine vessels;
- storage and transport of chemicals;
- waste disposal, treatment, and recovery via incineration, open burning, landfills, and composting;
- miscellaneous area sources such as cremation, hospitals, lamp breakage, and automotive repair shops.

The nonpt sector includes emission estimates for Portable Fuel Containers (PFCs), also known as "gas cans." The PFC inventory consists of five distinct sources of PFC emissions, further distinguished by residential or commercial use. The five sources are: (1) displacement of the vapor within the can; (2) spillage of gasoline while filling the can; (3) spillage of gasoline during transport; (4) emissions due to evaporation (i.e., diurnal emissions); and (5) emissions due to permeation. Note that spillage and vapor displacement associated with using PFCs to refuel nonroad equipment are included in the nonroad inventory.

3.2.4 Biogenic Sources (beis)

Biogenic emissions were computed based on the same 15j version of the 2015 meteorology data used for the air quality modeling and were developed using the Biogenic Emission Inventory System version 3.61 (BEIS3.61) within SMOKE. The BEIS3.61 creates gridded, hourly, model-species emissions from vegetation and soils. It estimates CO, VOC (most notably isoprene, terpene, and sesquiterpene), and NO emissions for the contiguous U.S. and for portions of Mexico and Canada. In the BEIS 3.61 two-layer canopy model, the layer structure varies with light intensity and solar zenith angle (Pouliot and Bash, 2015). Both layers include estimates of sunlit and shaded leaf area based on solar zenith angle and light intensity, direct and diffuse solar radiation, and leaf temperature (Bash et al., 2015). The new algorithm requires additional meteorological variables over previous versions of BEIS. The variables output from the Meteorology-Chemistry Interface Processor (MCIP) that are used to convert WRF outputs to CMAQ inputs are shown in Table 3-6.

Table 3-6. Meteorological variables required by BEIS 3.61

Variable	Description
LAI	leaf-area index
PRSFC	surface pressure
Q2	mixing ratio at 2 m
RC	convective precipitation per met TSTEP
RGRND	solar rad reaching sfc
RN	nonconvective precipitation per met TSTEP
RSTOMI	inverse of bulk stomatal resistance
SLYTP	soil texture type by USDA category
SOIM1	volumetric soil moisture in top cm
SOIT1	soil temperature in top cm
TEMPG	skin temperature at ground
USTAR	cell averaged friction velocity
RADYNI	inverse of aerodynamic resistance
TEMP2	temperature at 2 m

BEIS3.61 was used in conjunction with Version 4.1 of the Biogenic Emissions Landuse Database (BELD4.1). The BELD version 4.1 is based on an updated version of the USDA-USFS Forest Inventory and Analysis (FIA) vegetation speciation-based data from 2001 to 2014 from the FIA version 5.1. Canopy coverage is based on the Landsat satellite National Land Cover Database (NLCD) product from 2011. The FIA includes approximately 250,000 representative plots of species fraction data that are within approximately 75 km of one another in areas identified as forest by the NLCD canopy coverage. The 2011 NLCD provides land cover information with a native data grid spacing of 30 meters. For land areas

outside the conterminous United States, 500 meter grid spacing land cover data from the Moderate Resolution Imaging Spectroradiometer (MODIS) is used. BELDv4.1 also incorporates the following:

- 30 meter NASA's Shuttle Radar Topography Mission (SRTM) elevation data (http://www2.jpl.nasa.gov/srtm/) to more accurately define the elevation ranges of the vegetation species than in previous versions; and
- 2011 30 meter USDA Cropland Data Layer (CDL) data (http://www.nass.usda.gov/research/Cropland/Release/).

For the 2014NEIv2 and this study, land use changes were made for the states of Florida, Texas and Washington to correct an error with the land use fractions which did not sum to 1; but the version remained named BELD4.1.

Biogenic emissions computed with BEIS version 3.61 were left out of the CMAQ-ready merged emissions, in favor of inline biogenics produced during the CMAQ model run itself.

3.2.5 Mobile Sources (onroad, onroad_ca_adj, nonroad, cmv_c1c2, cmv_c3, rail)

Mobile sources are emissions from vehicles that move and include several sectors. Onroad mobile source emissions result from motorized vehicles that are normally operated on public roadways. These include passenger cars, motorcycles, minivans, sport-utility vehicles, light-duty trucks, heavy-duty trucks, and buses. Nonroad mobile source emissions are from vehicles that do not operate on roads such as tractors, construction equipment, lawnmowers, and recreational marine vessels. All nonroad emissions are treated as county-specific low-level emissions (i.e., they are released into model layer 1).

Commercial marine vessel (CMV) emissions are split into two sectors: emissions from Category 1 and Category 2 vessels are in the cmv_c1c2 sector, and emissions from the larger Category 3 vessels are in the cmv_c3 sector where they are treated as point sources with plume rise. Locomotive emissions are in the rail sector. Having the emissions split into these sectors facilitates separating them in summaries and also allows for the largest vessels to be modeled with plume rise.

3.2.5.1 Onroad (onroad)

Onroad mobile sources include emissions from motorized vehicles that are normally operate on public roadways. These include passenger cars, motorcycles, minivans, sport-utility vehicles, light-duty trucks, heavy-duty trucks, and buses. The sources are further divided between diesel, gasoline, E-85, and compressed natural gas (CNG) vehicles. The sector characterizes emissions from parked vehicle processes (e.g., starts, hot soak, and extended idle) as well as from on-network processes (i.e., from vehicles moving along the roads).

The onroad SCCs in the modeling platform are more finely resolved than those in the NEI, because the NEI SCCs distinguish vehicles and fuels, but in the platform they also distinguish between emissions on roadways, off-network, extended idle, and the various MOVES road-types. For more details on the approach and for a summary of the inputs submitted by states, see the section 6.5.1 of the 2014NEIv2 TSD.

Except for California, onroad emissions are generated using the SMOKE-MOVES interface that leverages MOVES generated emission factors (https://www.epa.gov/moves), county and SCC-specific activity data, and hourly meteorological data. SMOKE-MOVES takes into account the temperature sensitivity of the

on-road emissions. Specifically, EPA used MOVES inputs for representative counties, vehicle miles traveled (VMT), vehicle population (VPOP), and hoteling hours data for all counties, along with tools that integrated the MOVES model with SMOKE. In this way, it was possible to take advantage of the gridded hourly temperature data available from meteorological modeling that are also used for air quality modeling.

SMOKE-MOVES makes use of emission rate "lookup" tables generated by MOVES that differentiate emissions by process (i.e., running, start, vapor venting, etc.), vehicle type, road type, temperature, speed, hour of day, etc. To generate the MOVES emission rates that could be applied across the U.S., EPA used an automated process to run MOVES to produce year 2015-specific emission factors by temperature and speed for a series of "representative counties," to which every other county was mapped. The representative counties for which emission factors are generated are selected according to their state, elevation, fuels, age distribution, ramp fraction, and inspection and maintenance programs. Each county is then mapped to a representative county based on its similarity to the representative county with respect to those attributes. For the 2014v7.1 platform and for this study, there are 303 representative counties, twelve more than the number of representative counties in the 2014v7.0 platform. A detailed discussion of the representative counties is in the 2014NEIv2 TSD, Section 6.8.2.

Once representative counties have been identified, emission factors are generated with MOVES for each representative county and for two "fuel months" – January to represent winter months, and July to represent summer months – due to the different types of fuels used. SMOKE selects the appropriate MOVES emissions rates for each county, hourly temperature, SCC, and speed bin and multiplies the emission rate by appropriate activity data. For on-roadway emissions, vehicle miles travelled (VMT) is the activity data, vehicle population (VPOP) is used for many off-network processes, and hoteling hours are used to develop emissions for extended idling of combination long-haul trucks. These calculations are done for every county and grid cell in the continental U.S. for each hour of the year.

The SMOKE-MOVES process for creating the model-ready emissions consists of the following steps:

1) Determine which counties will be used to represent other counties in the MOVES runs.
2) Determine which months will be used to represent other month's fuel characteristics.
3) Create inputs needed only by MOVES. MOVES requires county-specific information on vehicle populations, age distributions, and inspection-maintenance programs for each of the representative counties.
4) Create inputs needed both by MOVES and by SMOKE, including temperatures and activity data.
5) Run MOVES to create emission factor tables for the temperatures found in each county.
6) Run SMOKE to apply the emission factors to activity data (VMT, VPOP, and HOTELING) to calculate emissions based on the gridded hourly temperatures in the meteorological data.
7) Aggregate the results to the county-SCC level for summaries and quality assurance.

The onroad emissions are processed in four processing streams that are merged together into the onroad sector emissions after each of the four streams have been processed:

- <u>rate-per-distance (RPD)</u> uses VMT as the activity data plus speed and speed profile information to compute on-network emissions from exhaust, evaporative, permeation, refueling, and brake and tire wear processes;

- rate-per-vehicle (RPV) uses VPOP activity data to compute off-network emissions from exhaust, evaporative, permeation, and refueling processes;
- rate-per-profile (RPP) uses VPOP activity data to compute off-network emissions from evaporative fuel vapor venting, including hot soak (immediately after a trip) and diurnal (vehicle parked for a long period) emissions; and
- rate-per-hour (RPH) uses hoteling hours activity data to compute off-network emissions for idling of long-haul trucks from extended idling and auxiliary power unit process.

The onroad emissions inputs for the platform are based on the 2014NEIv2, described in more detail in Section 6 of the 2014NEIv2 TSD. These inputs include:

- MOVES County databases (CDBs) including Low Emission Vehicle (LEV) table
- Representative counties
- Fuel months
- Meteorology
- Activity data (VMT, VPOP, speed, HOTELING)

Representative counties and fuel months are the same as for the 2014NEIv2, while other inputs were updated for the year 2015. The activity data was projected from 2014 to 2015 using the following procedure. First, VMT was projected using factors calculated from FHWA VM-2 data (https://www.fhwa.dot.gov/policyinformation/statistics/2014/vm2.cfm, https://www.fhwa.dot.gov/policyinformation/statistics/2015/vm2.cfm). Year-to-year projection factors were calculated by state, with separate factors for urban and rural road types, and then applied to the 2014NEIv2 VMT. In some states, a single state-wide projection factor for all road types was computed in states with large differences in how activity is split between urban and rural road types in the FHWA data compared to the 2014NEIv2 VMT dataset. States for which a single projection factor was applied state-wide are: Alaska, Georgia, Indiana, Louisiana, Maine, Massachusetts, Nebraska, New Mexico, New York, North Dakota, Tennessee, Virginia, and West Virginia. There are two other exceptions: In Texas and Utah, a single state-wide projection factor was calculated based on state-wide VMT totals provided by each state's Department of Transportation[7]. Once the VMT data were finalized for 2015, VPOP activity for 2015 was calculated by applying VMT/VPOP ratios based on 2014NEIv2 to the projected 2015 VMT for each county, fuel, and vehicle type. Hoteling hours activity for 2015 was calculated in a similar manner, by applying 2014NEIv2-based VMT/hoteling ratios to the projected 2015 VMT, but only for VMT from long-haul combination trucks on restricted roads.

An additional step was taken for the refueling emissions. Colorado submitted point emissions for refueling for some counties[8]. For these counties, the EPA zeroed out the onroad estimates of refueling (i.e., SCCs =220xxxxx62) so that the states' point emissions would take precedence. The onroad refueling emissions were zeroed out using the adjustment factor file (CFPRO) and Movesmrg.

---

[7] Sources of Texas data: https://ftp.dot.state.tx.us/pub/txdot-info/trf/crash_statistics/2014/01.pdf, https://ftp.dot.state.tx.us/pub/txdot-info/trf/crash_statistics/2015/01.pdf
Sources of Utah data: https://www.udot.utah.gov/main/uconowner.gf?n=32396326443209656, https://www.udot.utah.gov/main/uconowner.gf?n=27035817009129993

[8] There were 52 counties in Colorado that had point emissions for refueling. Outside Colorado, it was determined that refueling emissions in the 2014 NEIv2 point did not significantly duplicate the refueling emissions in onroad.

For more detailed information on the methods used to develop the 2014 onroad mobile source emissions and the input data sets, see Section 6.6 of the 2014NEIv1 TSD.

California is the only state agency for which submitted onroad emissions were used in the 2014 NEI v2 and 2014v7.1 platform. California uses their own emission model, EMFAC, which uses emission inventory codes (EICs) to characterize the emission processes instead of SCCs. The EPA and California worked together to develop a code mapping to better match EMFAC's EICs to EPA MOVES' detailed set of SCCs that distinguish between off-network and on-network and brake and tire wear emissions. This detail is needed for modeling but not for the NEI. This code mapping is provided in "2014v1_EICtoEPA_SCCmapping.xlsx." California then provided their CAP and HAP emissions by county using EPA SCCs after applying the mapping. There was one change made after the mapping: the vehicle/fuel type combination gas intercity buses (first 6 digits of the SCC = 220141), that is not generated using MOVES, was changed to gasoline single unit short-haul trucks (220152) for consistency with the modeling inventory. California provided EMFAC2014-based onroad emissions inventories for 2014 and 2017; emissions inventories from those two years were interpolated to 2015 values for this study.

The California onroad mobile source emissions were created through a hybrid approach of combining state-supplied annual emissions with EPA-developed SMOKE-MOVES runs. Through this approach, the platform was able to reflect the unique rules in California, while leveraging the more detailed SCCs and the highly resolved spatial patterns, temporal patterns, and speciation from SMOKE-MOVES. The basic steps involved in temporally allocating onroad emissions from California based on SMOKE-MOVES results were:

1) Run CA using EPA inputs through SMOKE-MOVES to produce hourly 2015 emissions hereafter known as "EPA estimates." These EPA estimates for CA are run in a separate sector called "onroad_ca."
2) Calculate ratios between state-supplied emissions and EPA estimates. The ratios were calculated for each county/SCC/pollutant combination based on the interpolated 2015 California onroad emissions inventory. Unlike in previous platforms, the California data separated off and on-network emissions and extended idling. However, the on-network did not provide specific road types, and California's emissions did not include information for vehicles fueled by E-85, so these differentiations were obtained using MOVES.
3) Create an adjustment factor file (CFPRO) that includes EPA-to-state estimate ratios.
4) Rerun CA through SMOKE-MOVES using EPA inputs and the new adjustment factor file.

Through this process, adjusted model-ready files were created that sum to annual totals from California, but have the temporal and spatial patterns reflecting the highly resolved meteorology and SMOKE-MOVES. After adjusting the emissions, this sector is called "onroad_ca_adj." Note that in emission summaries, the emissions from the "onroad" and "onroad_ca_adj" sectors are summed and designated as the emissions for the onroad sector.

3.2.5.2 MOVES-based Nonroad Mobile Sources (nonroad)

The nonroad equipment emissions in the platform and the NEI result primarily from running the MOVES2014a model, which incorporates the NONROAD2008 model. MOVES2014a replaces NMIM, which was used for 2011 and earlier NEIs. MOVES2014a provides a complete set of HAPs and

incorporates updated nonroad emission factors for HAPs. MOVES2014a was used for all states other than California, which uses their own model. Additional details on the development of the 2014NEIv2 nonroad emissions are available in Section 5 the 2014NEIv2 TSD. A separate MOVES2014a run was performed for the year 2016, and the 2014 and 2016 nonroad emissions were interpolated to 2015 values for this study. Interpolations for earlier 2015 studies were based on a version of the 2016 nonroad inventory that was not properly run for the year 2016. For this study, a corrected 2016 nonroad inventory was developed, allowing for a more accurate interpolation and representation of nonroad emissions for 2015.

The magnitude of the annual emissions in the nonroad inventory used here are similar to the emissions in the nonroad data category of the 2014NEIv2. However, the platform has monthly emission totals, which are provided by MOVES2014a and contain additional pollutants used in the emissions modeling. The emissions in the modeling platform include NONHAPTOG and ETHANOL, which are not included in the NEI. NONHAPTOG is the difference between total organic gases (TOG) and explicit species that are estimated separately such as benzene, toluene, styrene, ethanol, and numerous other compounds and are integrated into the chemical speciation process. MOVES2014a provides estimates of NONHAPTOG along with the speciation profile code for the NONHAPTOG emission source. This is accomplished by using NONHAPTOG#### as the pollutant code in the FF10 inventory file, where #### is a speciation profile code. Since speciation profiles are applied by SCC and pollutant, no changes to SMOKE were needed in order to use the FF10 with this profile information. This approach is not used for California, because their model provides VOC and traditional speciation is performed in SMOKE instead.

Nonroad emissions for California submitted to NEI were developed using the California Emissions Projection Analysis Model (CEPAM) that supports various California off-road regulations. Documentation of the CARB offroad mobile methodology, including CMV sector data, is provided at: http://www.arb.ca.gov/msei/categories.htm#offroad_motor_vehicles. The CARB-supplied nonroad annual inventory emissions values were temporalized to monthly values using monthly temporal profiles applied in SMOKE by SCC. Some VOC emissions were added to California to account for situations when VOC HAP emissions were included in the inventory, but VOC emissions were either less than the sum of the VOC HAP emissions, or were missing entirely. These additional VOC emissions were computed by summing benzene, acetaldehyde, formaldehyde, and naphthalene for the specific sources. California nonroad inventories were available for years 2014 and 2017; for this study emissions were interpolated between 2014 and 2017 to estimate 2015 values.

3.2.5.3 Locomotive (rail)

The rail sector includes all locomotives in the NEI nonpoint data category. This sector excludes railway maintenance locomotives and point source yard locomotives. Railway maintenance emissions are included in the nonroad sector. The point source yard locomotives are included in the ptnonipm sector. The nonpoint rail data are a mix of S/L and EPA data. EPA estimates cover only SCCs 2285002006 and 2285002007. Revised and/or new data were provided by some states for the 2014NEIv2. The EPA data were completely replaced from the v1 estimates, which had been carried forward from the 2011 NEI. The updated EPA data were developed by the Eastern Regional Technical Advisory Committee's (ERTAC) rail group. The group coordinated with the Federal Rail Administration to collect link-based activity data and apply the equipment-specific emission factors appropriate. For more information on locomotive sources in the NEI, see Section 4.20 of the 2014NEIv2 TSD. For this 2015 study, rail emissions from the 2014NEIv2 were used as-is.

3.2.5.4 Category 1, 2, and 3 commercial marine vessels (cmv_c1c2 and cmv_3)

The cmv_c1c2 sector contains Category 1 and 2 CMV emissions from the 2014 NEIv2. Category 1 and 2 vessels use diesel fuel. All emissions in this sector are annual and at county-SCC resolution; however, in the NEI they are provided at the sub-county level (port or underway shape ids) and by SCC and emission type (e.g., hoteling, maneuvering). This sub-county data in the NEI are used to create spatial surrogates. For more information on CMV sources in the NEI, see Section 4.19 of the 2014NEIv2 TSD. C1 and C2 emissions that occur outside of state waters are not assigned to states. All CMV emissions in the cmv_c1c2 sector are treated as nonpoint sources and are placed in layer 1 and allocated to grid cells using spatial surrogates. For this 2015 study, cmv_c1c2 emissions from the 2014NEIv2 were used as-is, with the exception of SO_2 emissions, which were reduced by 90% from 2014NEIv2 levels in accordance with ECA-IMO emissions standards for 2015. It should be noted that this reduction was not appropriate for C1 and C2 ships, because those ships use diesel fuel and not residual fuel; however, since SO_2 emissions levels for C1 and C2 ships are already low in 2014NEIv2, this further reduction had a small impact.

The Category 3 CMV vessels in the cmv_c3 sector use residual oil. The cmv_c3 sector uses 2014NEIv2 emissions not only in state waters, but also in Federal Waters (FIPS codes beginning with 85), which is a change from the 2014v7.0 platform. SO_2 emissions in the cmv_c3 sector were reduced by 90% from 2014NEIv2 levels within state and federal waters, in accordance with ECA-IMO emissions standards for 2015. Emissions from the Emissions Control Area-International Marine Organization (ECA-IMO)-based C3 CMV are used for waters not covered by the NEI (FIPS code 98001). The C3 CMV emissions are treated as point sources, which allows for them to have plume rise when modeled by SMOKE and CMAQ. The ECA-IMO inventory is also used for allocating the county-level NEI emissions to geographic locations.

The ECA-IMO dataset has been used since the Emissions Control Area-International Marine Organization (ECA-IMO) project began in 2005, although it was then known as the Sulfur Emissions Control Area (SECA). The ECA-IMO emissions consist of large marine diesel engines (at or above 30 liters/cylinder) that until recently were allowed to meet relatively modest emission requirements and as a result these ships would often burn residual fuel in that region. The ECA-IMO dataset was developed based on a 4-km resolution ASCII raster format dataset that preserves shipping lanes and extends within and beyond the federal waters.

The emissions in the cmv_c3 sector are comprised of primarily foreign-flagged ocean-going vessels, referred to as C3 CMV ships. The C3 portion of the CMV inventory includes these ships in several intra-port modes (i.e., cruising, hoteling, reduced speed zone, maneuvering, and idling) and an underway mode, and includes near-port auxiliary engine emissions.

An overview of the C3 ECA Proposal to the International Maritime Organization (EPA-420-F-10-041, August 2010) project and future-year goals for reduction of NO_X, SO_2, and PM C3 emissions can be found at: https://www.epa.gov/fuels-registration-reporting-and-compliance-help/guidance-documents-marine-fuel. The resulting ECA-IMO coordinated strategy, including emission standards under the Clean Air Act for new marine diesel engines with per-cylinder displacement at or above 30 liters, and the establishment of Emission Control Areas is available from https://www.epa.gov/regulations-emissions-vehicles-and-engines/international-standards-reduce-emissions-marine-diesel. The base year for the ECA inventory is 2002 and consists of these CAPs: PM_{10}, $PM_{2.5}$, CO, CO_2, NH_3, NO_X, SO_X (assumed to be SO_2), and hydrocarbons (assumed to be VOC). EPA developed regional growth (activity-based) factors that were applied to create the 2011 inventory from the 2002 data. This 2011 ECA-IMO inventory is still

in use outside of federal waters, but is only used to spatially allocate emissions from the 2014NEIv2 within state and federal waters. The geographic regions that are considered part of federal waters are shown in Figure 3-1. The East Coast and Gulf Coast regions were divided along a line roughly through Key Largo (longitude 80° 26' West). Technically, the EEZ FIPS are not really "FIPS" state-county codes, but are treated as such in the inventory and emissions processing.

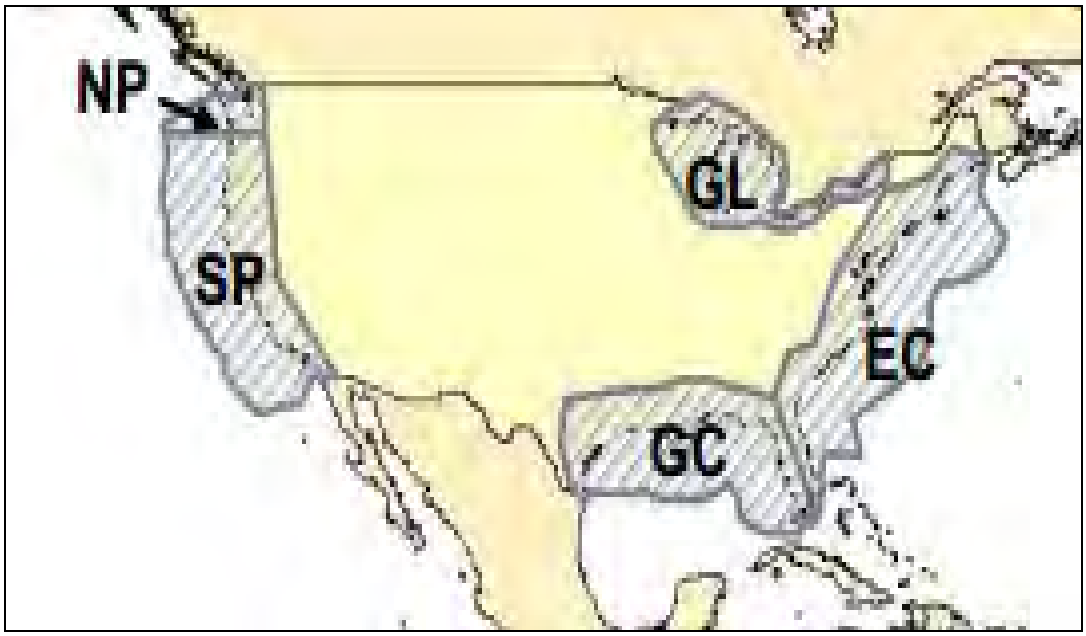


Figure 3-1. Illustration of regional modeling domains in ECA-IMO study

3.2.6 Emissions from Canada, Mexico (othpt, othar, othafdust, onroad_can, onroad_mex, ptfire_mxca)

The emissions from Canada and Mexico are included as part of the emissions modeling sectors: othpt, othar, othafdust, onroad_can, and onroad_mex. The "oth" refers to the fact that these emissions are usually "other" than those in the U.S. state-county geographic FIPS, and the remaining characters provide the SMOKE source types: "pt" for point, "ar" for area and nonroad mobile, "afdust" for area fugitive dust (Canada only). The onroad emissions for Canada and Mexico are in the onroad_can and onroad_mex sectors, respectively.

For Canadian point sources, 2013 and 2025 emissions provided by Environment Canada were interpolated to year 2015 for facilities included in both the 2013 and 2025 datasets. Sources that were only in the 2013 dataset and not in 2025 (i.e. closures) were omitted from the 2015 dataset. Sources that were only in the 2025 dataset and not in 2013 (i.e. newly opened facilities) were included in the 2015 inventory with emissions set to 2025 values, except for the Bonnybrook Energy Centre facility in Alberta, which as of 2018 has not opened and thus was left out of the 2015 inventory. These Canadian point source inventories included VOC emissions with CB6 speciation, although the CB6 VOCs differed slightly from the version of CB6 in CMAQ. Environment Canada also provided total unspeciated VOC, which was added to the inventory as VOC_INV and was speciated for ACET, CH4 and CB6-CMAQ species not covered in the CB6-speciated inventory (XYLMN, NAPH and SOAALK). Airport emissions were provided by month. Temporal profiles were provided for all source categories. Other than the CB6 species of NBAFM present in the speciated NPRI data, there are no explicit HAP emissions in this

inventory.

Point sources in Mexico were compiled based on inventories projected from the the Inventario Nacional de Emisiones de Mexico, 2008 (ERG, 2017). The point source emissions were converted to English units and into the FF10 format that could be read by SMOKE, missing stack parameters were gapfilled using SCC-based defaults, and latitude and longitude coordinates were verified and adjusted if they were not consistent with the reported municipality. Mexican point inventories were projected from 2008 to the years 2014 and 2018, and then those emissions values were interpolated to the year 2015 for this study. Only CAPs are covered in the Mexico point source inventory.

For Canadian area and nonroad sources, year-2013 and year-2025 emissions provided by Environment Canada were interpolated to year 2015, including CMV emissions. The Canadian inventory included fugitive dust emissions that do not incorporate either a transportable fraction or meteorological-based adjustments. To properly account for this, a separate sector called othafdust was created and modeled using the same adjustments as are done for U.S. sources. Updated Shapefiles used for creating spatial surrogates for Canada were also provided. For Canada nonroad mobile sources, the provided 2013 and 2025 monthly emissions were interpolated to 2015.

For Mexican area and nonroad sources, emission projections based on Mexico's 2008 inventory were used for area, point and nonroad sources (ERG, 2017). The resulting inventory was written using English units to the nonpoint FF10 format that could be read by SMOKE. Note that unlike the U.S. inventories, there are no explicit HAPs in the nonpoint or nonroad inventories for Canada and Mexico and, therefore, all HAPs are created from speciation. Similar to the point inventories, Mexican area and nonroad inventories were projected from 2008 to the years 2014 and 2018, and then emissions values were interpolated to year 2015 values for this study.

For Canada onroad emissions, month-specific year-2013 and year-2025 emissions provided by Environment Canada were interpolated to year 2015. There are no explicit HAPs in the onroad inventories for Canada and, therefore, NBAFM HAPs are created from speciation. For Mexico onroad emissions, a version of the MOVES model for Mexico was run that provided the same VOC HAPs and speciated VOCs as for the U.S. MOVES model (ERG, 2016a). This includes NBAFM plus several other VOC HAPs such as toluene, xylene, ethylbenzene and others. Except for VOC HAPs that are part of the speciation, no other HAPs are included in the Mexico onroad inventory (such as particulate HAPs nor diesel particulate matter). Mexico onroad inventories were generated by MOVES for the years 2014 and 2017, and then emissions values were interpolated to the year 2015 for this study.

Annual 2015 wildland emissions for Mexico, Canada, Central America, and Caribbean nations are included in the ptfire_othna sector. Mexico and Canada emissions are calculated in SMARTFIRE2 (Sullivan, et al., 2008) using NOAA's Hazard Mapping System (HMS) satellite data reconciled with Canadian Wildland Fire Information Systems polygons in Canada and MODIS Collection 6 hotspot data in Mexico. The SMARTFIRE2 output was further processed through the BlueSky Framework (BSF) similar to the US ptfire sector. The wildland fire emissions for all other regions in ptfire_othna were developed from Fire Inventory from NCAR (FINN) 2015 v1.5 daily fire emissions. For FINN fires, listed vegetation type codes of 1 and 9 are defined as agricultural burning, all other fire detections and assumed to be wildfires. All wildland fires that are not defined as agricultural are assumed to be wild fires rather than prescribed. FINN fire detects less than 50 square meters (0.012 acres) are removed from the inventory. The locations of FINN fires are geocoded from latitude and longitude to FIPS code.

3.2.7 SMOKE-ready non-anthropogenic chlorine inventory

The ocean chlorine gas emission estimates are based on the build-up of molecular chlorine (Cl2) concentrations in oceanic air masses (Bullock and Brehme, 2002). Data at 36 km and 12 km resolution were available and were not modified other than the name "CHLORINE" was changed to "CL2" because that is the name required by the CMAQ model.

3.3 Emissions Modeling Summary

The CMAQ model requires hourly emissions of specific gas and particle species for the horizontal and vertical grid cells contained within the modeled region (i.e., modeling domain). To provide emissions in the form and format required by the model, it is necessary to "pre-process" the "raw" emissions (i.e., emissions input to SMOKE) for the sectors described above. In brief, the process of emissions modeling transforms the emissions inventories from their original temporal resolution, pollutant resolution, and spatial resolution into the hourly, speciated, gridded resolution required by the air quality model. Emissions modeling includes temporal allocation, spatial allocation, and pollutant speciation. In some cases, emissions modeling also includes the vertical allocation of point sources, but many air quality models also perform this task because it greatly reduces the size of the input emissions files if the vertical layers of the sources are not included.

As previously discussed, the temporal resolutions of the emissions inventories input to SMOKE vary across sectors and may be hourly, daily, monthly, or annual total emissions. The spatial resolution, may be individual point sources, county/province/municipio totals, or gridded emissions and varies by sector. This section provides some basic information about the tools and data files used for emissions modeling as part of the modeling platform.

3.3.1 The SMOKE Modeling System

SMOKE version 4.5 was used to pre-process the raw emissions inventories into emissions inputs for CMAQ. SMOKE executables and source code are available from the Community Multiscale Analysis System (CMAS) Center at http://www.cmascenter.org. Additional information about SMOKE is available from http://www.smoke-model.org. For sectors that have plume rise, the in-line emissions capability of the air quality models was used, which allows the creation of source-based and two-dimensional gridded emissions files that are much smaller than full three-dimensional gridded emissions files. For quality assurance of the emissions modeling steps, emissions totals by specie for the entire model domain are output as reports that are then compared to reports generated by SMOKE on the input inventories to ensure that mass is not lost or gained during the emissions modeling process.

3.3.2 Key Emissions Modeling Settings

When preparing emissions for the air quality model, emissions for each sector are processed separately through SMOKE, and then the final merge program (Mrggrid) is run to combine the model-ready, sector-specific emissions across sectors. The SMOKE settings in the run scripts and the data in the SMOKE ancillary files control the approaches used by the individual SMOKE programs for each sector. Table 3-7 summarizes the major processing steps of each platform sector. The "Spatial" column shows the spatial approach used: here "point" indicates that SMOKE maps the source from a point location (i.e., latitude and longitude) to a grid cell; "surrogates" indicates that some or all of the sources use spatial surrogates to allocate county emissions to grid cells; and "area-to-point" indicates that some of the sources use the

SMOKE area-to-point feature to grid the emissions. The "Speciation" column indicates that all sectors use the SMOKE speciation step, though biogenics speciation is done within the Tmpbeis3 program and not as a separate SMOKE step. The "Inventory resolution" column shows the inventory temporal resolution from which SMOKE needs to calculate hourly emissions. Note that for some sectors (e.g., onroad, beis), there is no input inventory; instead, activity data and emission factors are used in combination with meteorological data to compute hourly emissions.

Table 3-7. Key emissions modeling steps by sector

Platform sector	Spatial	Speciation	Inventory resolution	Plume rise
afdust_adj	Surrogates	Yes	annual	
ag	Surrogates	Yes	annual	
beis	Pre-gridded land use	in BEIS3.61	computed hourly	
cmv_c1c2	Surrogates	Yes	annual	
cmv_c3	Point	Yes	annual	in-line
nonpt	Surrogates & area-to-point	Yes	annual	
nonroad	Surrogates & area-to-point	Yes	monthly	
np_oilgas	Surrogates	Yes	annual	
onroad	Surrogates	Yes	monthly activity, computed hourly	
onroad_ca_adj	Surrogates	Yes	monthly activity, computed hourly	
onroad_can	Surrogates	Yes	monthly	
onroad_mex	Surrogates	Yes	monthly	
othafdust	Surrogates	Yes	annual	
othar	Surrogates	Yes	annual & monthly	
othpt	Point	Yes	annual & monthly	in-line
ptagfire	Point	Yes	daily	in-line
pt_oilgas	Point	Yes	annual	in-line
ptegu	Point	Yes	daily & hourly	in-line
ptfire	Point	Yes	daily	in-line
ptfire_othna	Point	Yes	daily	in-line
ptnonipm	Point	Yes	annual	in-line
rail	Surrogates	Yes	annual	
rwc	Surrogates	Yes	annual	

Biogenic emissions can be modeled two different ways in the CMAQ model. The BEIS model in SMOKE can produce gridded biogenic emissions that are then included in the gridded CMAQ-ready emissions inputs, or alternatively, CMAQ can be configured to create "in-line" biogenic emissions within CMAQ itself. For this study, the in-line biogenic emissions option was used, and so biogenic emissions from BEIS were not included in the gridded CMAQ-ready emissions.

The “plume rise” column indicates the sectors for which the “in-line” approach is used. These sectors are the only ones with emissions in aloft layers based on plume rise. The term “in-line” means that the plume rise calculations are done inside of the air quality model instead of being computed by SMOKE. The air quality model computes the plume rise using stack parameters and the hourly emissions in the SMOKE output files for each emissions sector. The height of the plume rise determines the model layer into which the emissions are placed. The othpt sector has only “in-line” emissions, meaning that all of the emissions are treated as elevated sources and there are no emissions for those sectors in the two-dimensional, layer-1 files created by SMOKE. Other inline-only sectors are: cmv_c3, ptegu, ptfire, ptfire_othna, ptagfire. Day-specific point fire emissions are treated differently in CMAQ. After plume rise is applied, there are emissions in every layer from the ground up to the top of the plume.

SMOKE has the option of grouping sources so that they are treated as a single stack when computing plume rise. For this modeling case, no grouping was performed because grouping combined with “in-line” processing will not give identical results as “offline” processing (i.e., when SMOKE creates 3-dimensional files). This occurs when stacks with different stack parameters or lat/lons are grouped, thereby changing the parameters of one or more sources. The most straightforward way to get the same results between in-line and offline is to avoid the use of grouping.

3.3.3 Spatial Configuration

For this study, SMOKE was run for the larger 12-km CONtinental United States "CONUS" modeling domain (12US1) shown in Figure 3-2. The grid used a Lambert-Conformal projection, with Alpha = 33, Beta = 45 and Gamma = -97, with a center of X = -97 and Y = 40. Later sections provide details on the spatial surrogates and area-to-point data used to accomplish spatial allocation with SMOKE.

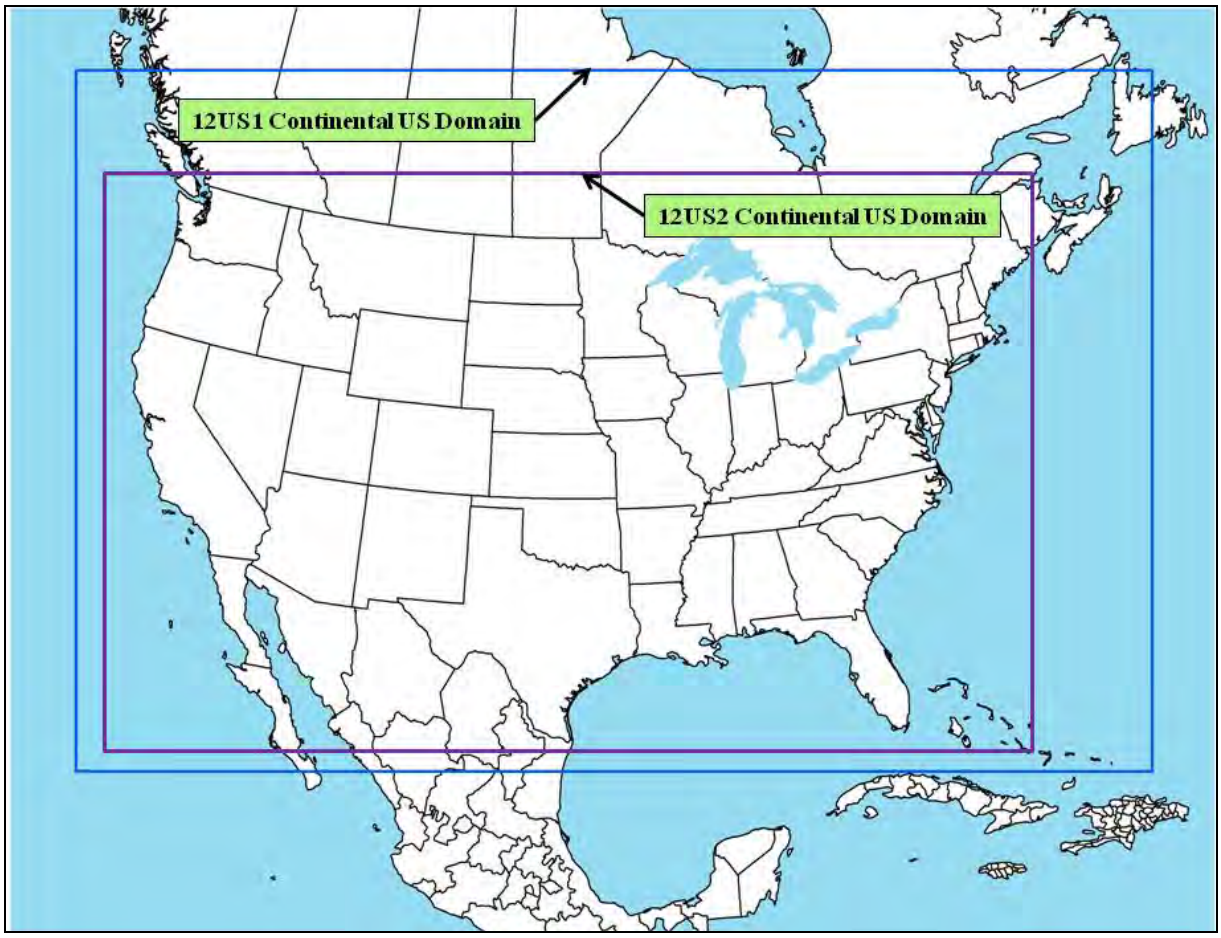


Figure 3-2. CMAQ Modeling Domain

3.3.4 Chemical Speciation Configuration

The emissions modeling step for chemical speciation creates the "model species" needed by the air quality model for a specific chemical mechanism. These model species are either individual chemical compounds (i.e., "explicit species") or groups of species (i.e., "lumped species"). The chemical mechanism used for the 2014 platform is the CB6 mechanism (Yarwood, 2010). We used a specific version of CB6 that we refer to as "CMAQ CB6" that breaks out naphthalene from XYL as an explicit model species, resulting in model species NAPH and XYLMN instead of XYL and uses SOAALK. This platform generates the $PM_{2.5}$ model species associated with the CMAQ Aerosol Module version 6 (AE6). Table 3-8 lists the model species produced by SMOKE in the platform used for this study.

Table 3-8. Emission model species produced for CB6 for CMAQ

Inventory Pollutant	Model Species	Model species description
Cl_2	CL2	Atomic gas-phase chlorine
HCl	HCL	Hydrogen Chloride (hydrochloric acid) gas
CO	CO	Carbon monoxide
NO_X	NO	Nitrogen oxide
	NO2	Nitrogen dioxide
	HONO	Nitrous acid
SO_2	SO2	Sulfur dioxide
	SULF	Sulfuric acid vapor
NH_3	NH3	Ammonia
	NH3_FERT	Ammonia from fertilizer
VOC	ACET	Acetone
	ALD2	Acetaldehyde
	ALDX	Propionaldehyde and higher aldehydes
	BENZ	Benzene (not part of CB05)
	CH4	Methane
	ETH	Ethene
	ETHA	Ethane
	ETHY	Ethyne
	ETOH	Ethanol
	FORM	Formaldehyde
	IOLE	Internal olefin carbon bond (R-C=C-R)
	ISOP	Isoprene
	KET	Ketone Groups
	MEOH	Methanol
	NAPH	Naphthalene
	NVOL	Non-volatile compounds
	OLE	Terminal olefin carbon bond (R-C=C)
	PAR	Paraffin carbon bond
	PRPA	Propane
	SESQ	Sequiterpenes (from biogenics only)
	SOAALK	Secondary Organic Aerosol (SOA) tracer
	TERP	Terpenes (from biogenics only)
	TOL	Toluene and other monoalkyl aromatics
	UNR	Unreactive
	XYLMN	Xylene and other polyalkyl aromatics, minus naphthalene
Naphthalene	NAPH	Naphthalene from inventory
Benzene	BENZ	Benzene from the inventory
Acetaldehyde	ALD2	Acetaldehyde from inventory
Formaldehyde	FORM	Formaldehyde from inventory
Methanol	MEOH	Methanol from inventory
PM_{10}	PMC	Coarse PM > 2.5 microns and ≤ 10 microns
$PM_{2.5}$	PEC	Particulate elemental carbon ≤ 2.5 microns
	PNO3	Particulate nitrate ≤ 2.5 microns
	POC	Particulate organic carbon (carbon only) ≤ 2.5 microns
	PSO4	Particulate Sulfate ≤ 2.5 microns
	PAL	Aluminum

Inventory Pollutant	Model Species	Model species description
	PCA	Calcium
	PCL	Chloride
	PFE	Iron
	PK	Potassium
	PH2O	Water
	PMG	Magnesium
	PMN	Manganese
	PMOTHR	$PM_{2.5}$ not in other AE6 species
	PNA	Sodium
	PNCOM	Non-carbon organic matter
	PNH4	Ammonium
	PSI	Silica
	PTI	Titanium
Sea-salt species (non – anthropogenic) [9]	PCL	Particulate chloride
	PNA	Particulate sodium

The TOG and $PM_{2.5}$ speciation factors that are the basis of the chemical speciation approach were developed from the SPECIATE 4.5 database (https://www.epa.gov/air-emissions-modeling/speciate), which is the EPA's repository of TOG and PM speciation profiles of air pollution sources. The SPECIATE database development and maintenance is a collaboration involving the EPA's Office of Research and Development (ORD), Office of Transportation and Air Quality (OTAQ), and the Office of Air Quality Planning and Standards (OAQPS), in cooperation with Environment Canada (EPA, 2016). The SPECIATE database contains speciation profiles for TOG, speciated into individual chemical compounds, VOC-to-TOG conversion factors associated with the TOG profiles, and speciation profiles for $PM_{2.5}$.

Some key features and updates to speciation from previous platforms include the following (the subsections below contain more details on the specific changes):

- VOC speciation profile cross reference assignments for point and nonpoint oil and gas sources were updated to (1) make corrections to the 2011v6.3 cross references, (2) use new and revised profiles that were added to SPECIATE4.5 and (3) account for the portion of VOC estimated to come from flares, based on data from the Oil and Gas estimation tool used to estimate emissions for the NEI. The new/revised profiles included oil and gas operations in specific regions of the country and a national profile for natural gas flares;
- the Western Regional Air Partnership (WRAP) speciation profiles used for the np_oilgas sector are the SPECIATE4.5 revised versions (profiles with "_R" in the profile code);
- the VOC speciation process for nonroad mobile has been updated - profiles are now assigned within MOVES2014a which outputs the emissions with those assignments; also the nonroad profiles themselves were updated;
- VOC and PM speciation for onroad mobile sources occurs within MOVES2014a except for brake and tirewear PM speciation which occurs in SMOKE;

[9] These emissions are created outside of SMOKE

- speciation for onroad mobile sources in Mexico is done within MOVES and is more consistent with that used in the United States;
- the PM speciation profile for C3 ships in the US and Canada was updated to a new profile, 5675AE6; and
- As with previous platforms, some Canadian point source inventories are provided from Environment Canada as pre-speciated emissions; however for the 2013 and 2025 inventories, not all CB6-CMAQ species were provided; missing species were supplemented by speciating VOC which was provided separately.

Speciation profiles and cross-references for this study platform are available in the SMOKE input files for the 2014v7.1 platform. Emissions of VOC and $PM_{2.5}$ emissions by county, sector and profile for all sectors other than onroad mobile can be found in the sector summaries for the case. Totals of each model species by state and sector can be found in the Appendix B state-sector totals workbook for this case.

The speciation of VOC includes HAP emissions from the 2014NEIv2 in the speciation process. Instead of speciating VOC to generate all of the species listed in Table 3-8, emissions of five specific HAPs: naphthalene, benzene, acetaldehyde, formaldehyde and methanol (collectively known as "NBAFM") from the NEI were "integrated" with the NEI VOC. The integration combines these HAPs with the VOC in a way that does not double count emissions and uses the HAP inventory directly in the speciation process. The basic process is to subtract the specified HAPs emissions mass from the VOC emissions mass, and to then use a special "integrated" profile to speciate the remainder of VOC to the model species excluding the specific HAPs. The EPA believes that the HAP emissions in the NEI are often more representative of emissions than HAP emissions generated via VOC speciation, although this varies by sector.

The NBAFM HAPs were chosen for integration because they are the only explicit VOC HAPs in the CMAQ version 5.2. Explicit means that they are not lumped chemical groups like PAR, IOLE and several other CB6 model species. These "explicit VOC HAPs" are model species that participate in the modeled chemistry using the CB6 chemical mechanism. The use of inventory HAP emissions along with VOC is called "HAP-CAP integration."

The integration of HAP VOC with VOC is a feature available in SMOKE for all inventory formats, including PTDAY (the format used for the ptfire and ptagfire sectors). The ability to use integration with the PTDAY format was made available in the version of SMOKE used for the v7.1 platform, but this new feature is not used for this particular study because the ptfire and ptagfire inventories for 2015 do not include HAPs. SMOKE allows the user to specify both the particular HAPs to integrate via the INVTABLE. This is done by setting the "VOC or TOG component" field to "V" for all HAP pollutants chosen for integration. SMOKE allows the user to also choose the particular sources to integrate via the NHAPEXCLUDE file (which actually provides the sources to be *excluded* from integration[10]). For the "integrated" sources, SMOKE subtracts the "integrated" HAPs from the VOC (at the source level) to compute emissions for the new pollutant "NONHAPVOC." The user provides NONHAPVOC-to-

10 Since SMOKE version 3.7, the options to specify sources for integration are expanded so that a user can specify the particular sources to include or exclude from integration, and there are settings to include or exclude all sources within a sector. In addition, the error checking is significantly stricter for integrated sources. If a source is supposed to be integrated, but it is missing NBAFM or VOC, SMOKE will now raise an error.

NONHAPTOG factors and NONHAPTOG speciation profiles[11]. SMOKE computes NONHAPTOG and then applies the speciation profiles to allocate the NONHAPTOG to the other air quality model VOC species not including the integrated HAPs. After determining if a sector is to be integrated, if all sources have the appropriate HAP emissions, then the sector is considered fully integrated and does not need a NHAPEXCLUDE file. If, on the other hand, certain sources do not have the necessary HAPs, then an NHAPEXCLUDE file must be provided based on the evaluation of each source's pollutant mix. The EPA considered CAP-HAP integration for all sectors in determining whether sectors would have full, no or partial integration (see Figure 3-3). For sectors with partial integration, all sources are integrated other than those that have either the sum of NBAFM > VOC or the sum of NBAFM = 0.

In this platform, we create NBAFM species from the no-integrate source VOC emissions using speciation profiles. Figure 3-3 illustrates the integrate and no-integrate processes for U.S. Sources. Since Canada and Mexico inventories do not contain HAPs, we use the approach of generating the HAPs via speciation, except for Mexico onroad mobile sources where emissions for integrate HAPs were available.

It should be noted that even though NBAFM were removed from the SPECIATE profiles used to create the GSPRO for both the NONHAPTOG and no-integrate TOG profiles, there still may be small fractions for "BENZ", "FORM", "ALD2", and "MEOH" present. This is because these model species may have come from species in SPECIATE that are mixtures. The quantity of these model species is expected to be very small compared to the BAFM in the NEI. There are no NONHAPTOG profiles that produce "NAPH."

In SMOKE, the INVTABLE allows the user to specify the particular HAPs to integrate. Two different INVTABLE files are used for different sectors of the platform. For sectors that had no integration across the entire sector (see Table 3-9), EPA created a "no HAP use" INVTABLE in which the "KEEP" flag is set to "N" for NBAFM pollutants. Thus, any NBAFM pollutants in the inventory input into SMOKE are automatically dropped. This approach both avoids double-counting of these species and assumes that the VOC speciation is the best available approach for these species for sectors using this approach. The second INVTABLE, used for sectors in which one or more sources are integrated, causes SMOKE to keep the inventory NBAFM pollutants and indicates that they are to be integrated with VOC. This is done by setting the "VOC or TOG component" field to "V" for all four HAP pollutants. Note for the onroad sector, "full integration" includes the integration of benzene, 1,3 butadiene, formaldehyde, acetaldehyde, naphthalene, acrolein, ethyl benzene, 2,2,4-Trimethylpentane, hexane, propionaldehyde, styrene, toluene, xylene, and MTBE.

11 These ratios and profiles are typically generated from the Speciation Tool when it is run with integration of a specified list of pollutants, for example NBAFM.

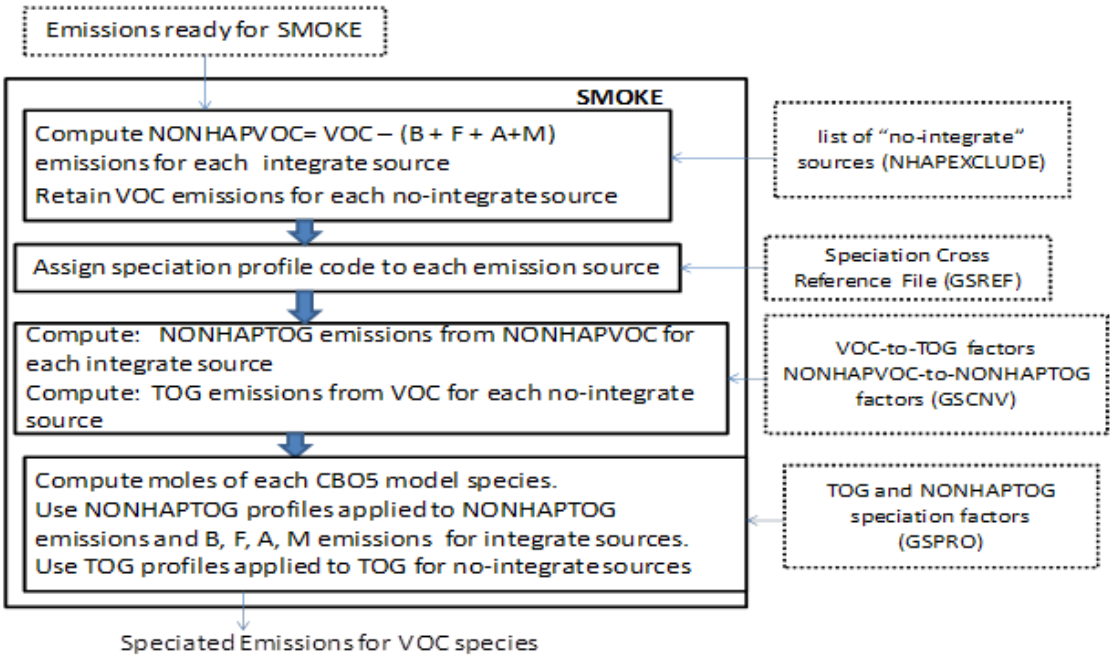


Figure 3-3. Process of integrating BAFM with VOC for use in VOC Speciation

Table 3-9. Integration status of naphthalene, benzene, acetaldehyde, formaldehyde and methanol (NBAFM) for each platform sector

Platform Sector	Approach for Integrating NEI emissions of Naphthalene (N), Benzene (B), Acetaldehyde (A), Formaldehyde (F) and Methanol (M)
ptegu	No integration, create NBAFM from VOC speciation
ptnonipm	No integration, create NBAFM from VOC speciation
ptfire	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
ptfire_othna	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
ptagfire	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
ag	Partial integration (NBAFM)
afdust	N/A – sector contains no VOC
beis	N/A – sector contains no inventory pollutant "VOC"; but rather specific VOC species
cmv_c1c2	Full integration (NBAFM)
cmv_c3	Full integration (NBAFM)
rail	Partial integration (NBAFM)
nonpt	Partial integration (NBAFM)
nonroad	Full integration (NBAFM in California, internal to MOVES elsewhere)
np_oilgas	Partial integration (NBAFM)
othpt	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
pt_oilgas	No integration, create NBAFM from VOC speciation
rwc	Partial integration (NBAFM)
onroad	Full integration (internal to MOVES); however, MOVES2014a speciation was CB6-CAMx, not CB6-CMAQ, so post-SMOKE emissions were converted to CB6-CMAQ

Platform Sector	Approach for Integrating NEI emissions of Naphthalene (N), Benzene (B), Acetaldehyde (A), Formaldehyde (F) and Methanol (M)
onroad_can	No integration, no NBAFM in inventory, create NBAFM from speciation
onroad_mex	Full integration (internal to MOVES-Mexico); however, MOVES-MEXICO speciation was CB6-CAMx, not CB6-CMAQ, so post-SMOKE emissions were converted to CB6-CMAQ
othafdust	N/A – sector contains no VOC
othar	No integration, no NBAFM in inventory, create NBAFM from VOC speciation

Integration for the mobile sources estimated from MOVES (onroad and nonroad sectors, other than for California) is done differently. Briefly there are three major differences: 1) for these sources integration is done using more than just NBAFM, 2) all sources from the MOVES model are integrated and 3) integration is done fully or partially within MOVES. For onroad mobile, speciation is done fully within MOVES2014a such that the MOVES model outputs emission factors for individual VOC model species along with the HAPs. This requires MOVES to be run for a specific chemical mechanism. MOVES was run for the CB6-CAMx mechanism rather than CB6-CMAQ, so post-SMOKE onroad emissions were converted to CB6-CMAQ. For nonroad mobile, speciation is partially done within MOVES such that it does not need to be run for a specific chemical mechanism. For nonroad, MOVES outputs emissions of HAPs and NONHAPTOG split by speciation profile. Taking into account that integrated species were subtracted out by MOVES already, the appropriate speciation profiles are then applied in SMOKE to get the VOC model species. HAP integration for nonroad uses the same additional HAPs and ethanol as for onroad.

In previous platforms, the GSPRO_COMBO feature was used to speciate nonroad mobile and gasoline-related stationary sources that use fuels with varying ethanol content. In these cases, the speciation profiles require different combinations of gasoline profiles, e.g. E0 and E10 profiles. Since the ethanol content varies spatially (e.g., by state or county), temporally (e.g., by month), and by modeling year (future years have more ethanol), the GSPRO_COMBO feature allows combinations to be specified at various levels for different years. For the 2014v7.1 platform, GSPRO_COMBO is still used for nonroad sources in California and for certain gasoline-related stationary sources nationwide. GSPRO_COMBO is also no longer needed for nonroad sources outside of California because nonroad emissions within MOVES have the speciation profiles built into the results, so there is no need to assign them via the GSREF or GSPRO_COMBO feature. In Canada and Mexico, only E0 speciation profiles are used, but the GSPRO_COMBO feature is still used for inventories where VOC emissions are not explicitly defined by mode (e.g. exhaust versus evaporative). Here, the GSPRO_COMBO specifies a mix of exhaust and evaporative speciation profiles. This is no longer necessary for Canadian mobile sources, whose inventories now include the mode in the pollutant, or for Mexico onroad sources, where VOC speciation is calculated by the MOVES model. The GSPRO_COMBO is still used for Mexican nonroad sources which do not have modes in the inventory.

A new method to combine multiple profiles is available in SMOKE4.5. It allows multiple profiles to be combined by pollutant, state and county (i.e., state/county FIPS code) and SCC. This was used specifically for the oil and gas sectors (pt_oilgas and np_oilgas) because SCCs include both controlled and uncontrolled oil and gas operations which use different profiles.

Speciation profiles for use with BEIS are not included in SPECIATE. BEIS3.61 includes a species (SESQ) that was mapped to the CMAQ specie SESQT. The profile code associated with BEIS profiles for use with

CB6 was "B10C6." For additional sector-specific details on VOC speciation for a variety of sectors, see Section 3.2.1.3 of the 2011v6.3 TSD (EPA, 2017b).

In addition to VOC profiles, the SPECIATE database also contains the $PM_{2.5}$ speciated into both individual chemical compounds (e.g., zinc, potassium, manganese, lead), and into the "simplified" $PM_{2.5}$ components used in the air quality model. For CMAQ 4.7.1 modeling, these "simplified" components (AE5) are all that is needed. Starting with CMAQ 5.0.1, a new thermodynamic equilibrium aerosol modeling tool (ISORROPIA) v2 mechanism was added that needs additional PM components (AE6), which are further subsets of PMFINE (see Table 3-10). The majority of the 2014 platform PM profiles come from the 911XX series which include updated AE6 speciation[12].

Table 3-10. PM model species: AE5 versus AE6

Species name	Species description	AE5	AE6
POC	organic carbon	Y	Y
PEC	elemental carbon	Y	Y
PSO4	Sulfate	Y	Y
PNO3	Nitrate	Y	Y
PMFINE	unspeciated $PM_{2.5}$	Y	N
PNH4	Ammonium	N	Y
PNCOM	non-carbon organic matter	N	Y
PFE	Iron	N	Y
PAL	Aluminum	N	Y
PSI	Silica	N	Y
PTI	Titanium	N	Y
PCA	Calcium	N	Y
PMG	Magnesium	N	Y
PK	Potassium	N	Y
PMN	Manganese	N	Y
PNA	Sodium	N	Y
PCL	Chloride	N	Y
PH2O	Water	N	Y
PMOTHR	$PM_{2.5}$ not in other AE6 species	N	Y

Unlike other sectors, the onroad sector has pre-speciated PM. This speciated PM comes from the MOVES model and is processed through the SMOKE-MOVES system. Unfortunately, the MOVES speciated PM does not map one-to-one to the AE5 speciation (nor the AE6 speciation) needed for CMAQ modeling. For additional details on PM speciation, see Section 3.2.2 of the 2011v6.2 platform TSD (EPA, 2015a).

[12] The exceptions are: 5675AE6 (Marine Vessel – Marine Engine – Heavy Fuel Oil) used for cmv_c3, replacing profile 5674 from previous 2015 studies; 92018 (Draft Cigarette Smoke – Simplified) used in nonpt; and 95475 (Composite - Refinery Fuel Gas and Natural Gas Combustion), which in this platform replaces 91112 and is used for sources across multiple sectors.

NO_X can be speciated into NO, NO2, and/or HONO. For the non-mobile sources, EPA used a single profile "NHONO" to split NO_X into NO and NO_2. For the mobile sources except for onroad (including nonroad, cmv_c1c2, cmv_c3, rail, onroad_can, onroad_mex sectors) and for specific SCCs in othar and ptnonipm, the profile "HONO" splits NO_X into NO, NO_2, and HONO. Table 3-11 gives the split factor for these two profiles. The onroad sector does not use the "HONO" profile to speciate NO_X. MOVES2014 produces speciated NO, NO_2, and HONO by source, including emission factors for these species in the emission factor tables used by SMOKE-MOVES. Within MOVES, the HONO fraction is a constant 0.008 of NO_X. The NO fraction varies by heavy duty versus light duty, fuel type, and model year and equals 1 – NO – HONO. For more details on the NO_X fractions within MOVES, see https://www.epa.gov/moves/moves-onroad-technical-reports#moves2014.

Table 3-11. NO_X speciation profiles

Profile	pollutant	species	split factor
HONO	NOX	NO2	0.092
HONO	NOX	NO	0.9
HONO	NOX	HONO	0.008
NHONO	NOX	NO2	0.1
NHONO	NOX	NO	0.9

Additional details on speciation for onroad, nonroad, and oil and gas sources, and new PM profiles used are discussed in Sections 3.2.1.4, 3.2.1.5, and 3.2.2 of the 2014v7.1 TSD (EPA, 2018b).

3.3.5 Temporal Processing Configuration

Temporal allocation (i.e., temporalization) is the process of distributing aggregated emissions to a finer temporal resolution, thereby converting annual emissions to hourly emissions. While the total emissions are important, the timing of the occurrence of emissions is also essential for accurately simulating ozone, PM, and other pollutant concentrations in the atmosphere. Many emissions inventories are annual or monthly in nature. Temporalization takes these aggregated emissions and, if needed, distributes them to the month, and then distributes the monthly emissions to the day and the daily emissions to the hours of each day. This process is typically done by applying temporal profiles to the inventories in this order: monthly, day of the week, and diurnal.

The temporal factors applied to the inventory are selected using some combination of country, state, county, SCC, and pollutant. Table 3-12 summarizes the temporal aspects of emissions modeling by comparing the key approaches used for temporal processing across the sectors. In the table, "Daily temporal approach" refers to the temporal approach for getting daily emissions from the inventory using the SMOKE Temporal program. The values given are the values of the SMOKE L_TYPE setting. The "Merge processing approach" refers to the days used to represent other days in the month for the merge step. If this is not "all," then the SMOKE merge step runs only for representative days, which could include holidays as indicated by the right-most column. The values given are those used for the SMOKE M_TYPE setting (see below for more information).

Table 3-12. Temporal Settings Used for the Platform Sectors in SMOKE

Platform sector short name	Inventory resolutions	Monthly profiles used?	Daily temporal approach	Merge processing approach	Process Holidays as separate days
afdust_adj	Annual	Yes	week	all	Yes
ag	Monthly		met-based	all	No
beis	Hourly		n/a	all	No
cmv_c1c2	Annual	Yes	aveday	aveday	No
cmv_c3	Annual	Yes	aveday	aveday	No
nonpt	Annual	Yes	week	week	Yes
nonroad	Monthly		mwdss	mwdss	Yes
np_oilgas	Annual	Yes	week	week	Yes
onroad	Annual & monthly[1]		all	all	Yes
onroad_ca_adj	Annual & monthly[1]		all	all	Yes
othafdust_adj	Annual	Yes	week	week	No
othar	Annual & monthly	Yes	week	week	No
onroad_can	Monthly		week	week	No
onroad_mex	Monthly		week	week	No
othpt	Annual & monthly	Yes	mwdss	mwdss	No
ptagfire	Daily		all	all	No
pt_oilgas	Annual	Yes	mwdss	mwdss	Yes
ptegu	Annual & hourly	Yes[2]	all	all	No
ptnonipm	Annual	Yes	mwdss	mwdss	Yes
ptfire	Daily		all	all	No
ptfire_othna	Daily		all	all	No
rail	Annual	Yes	aveday	aveday	No
rwc	Annual	No[3]	met-based	all	No[3]

1. Note the annual and monthly "inventory" actually refers to the activity data (VMT and VPOP) for onroad. The actual emissions are computed on an hourly basis.
2. Only units that do not have matching hourly CEMs data use monthly temporal profiles.
3. Except for 2 SCCs that do not use met-based temporalization.

The following values are used in Table 3-12: The value "all" means that hourly emissions are computed for every day of the year and that emissions potentially have day-of-year variation. The value "week" means that hourly emissions computed for all days in one "representative" week, representing all weeks for each month. This means emissions have day-of-week variation, but not week-to-week variation within the month. The value "mwdss" means hourly emissions for one representative Monday, representative weekday (Tuesday through Friday), representative Saturday, and representative Sunday for each month. This means emissions have variation between Mondays, other weekdays, Saturdays and Sundays within the month, but not week-to-week variation within the month. The value "aveday" means hourly emissions computed for one representative day of each month, meaning emissions for all days within a month are the same. Special situations with respect to temporalization are described in the following subsections.

In addition to the resolution, temporal processing includes a ramp-up period for several days prior to January 1, 2015, which is intended to mitigate the effects of initial condition concentrations. The ramp-up period was 10 days (December 22-31, 2014). For all anthropogenic sectors, emissions from December 2015 were used to fill in surrogate emissions for the end of December 2014. In particular, December 2015 emissions (representative days) were used for December 2014. For biogenic emissions, December 2014 emissions were processed using 2014 meteorology.

The Flat File 2010 format (FF10) inventory format for SMOKE provides a more consolidated format for monthly, daily, and hourly emissions inventories than prior formats supported. Previously, processing monthly inventory data required the use of 12 separate inventory files. With the FF10 format, a single inventory file can contain emissions for all 12 months and the annual emissions in a single record. This helps simplify the management of numerous inventories. Similarly, daily and hourly FF10 inventories contain individual records with data for all days in a month and all hours in a day, respectively.

SMOKE prevents the application of temporal profiles on top of the "native" resolution of the inventory. For example, a monthly inventory should not have annual-to-month temporalization applied to it; rather, it should only have month-to-day and diurnal temporalization. This becomes particularly important when specific sectors have a mix of annual, monthly, daily, and/or hourly inventories. The flags that control temporalization for a mixed set of inventories are discussed in the SMOKE documentation. The modeling platform sectors that make use of monthly values in the FF10 files are ag, nonroad, onroad (for activity data), onroad_can, onroad_mex, othar, othpt, and ptegu.

3.3.5.1 Standard Temporal Profiles

Some sectors use straightforward temporal profiles not based on meteorology or other factors. For the ptfire, ptagfire, and ptfire_othna sectors, the inventories are in the daily point fire format, so temporal profiles are only used to go from day-specific to hourly emissions. For all agricultural burning, the diurnal temporal profile used reflected the fact that burning occurs during the daylight. This puts most of the emissions during the work day and suppresses the emissions during the middle of the night. This diurnal profile was used for each day of the week for all agricultural burning emissions in all states.

For the cmv sectors, emissions are allocated with flat day of week and flat hourly profiles. Updated monthly profiles were developed for the LADCO states using link-level NO_X emissions for ship traffic provided by LADCO. These data were based on activities reported by ship AIS (transponder) devices. Monthly NOx emissions were normalized to create temporal profiles for each lake. For the port SCCs, an in-port profile was developed as the average of the maneuvering and hoteling emissions. The cruising emissions were used for the underway SCCs. As some of the lakes did not include complete data for the in-port sources (Ontario, Canada, St. Claire), a hybrid profile was created as an average of the in-port NOx emissions for Lakes Michigan, Huron, Superior, and Erie. A resulting 22 profiles were developed and applied to C1, C2 and C3 ships based county and SCC (i.e., port versus underway). Only new monthly profiles were developed from these data because the weekly and diurnal variation were deemed to be comparable to the existing EPA profiles. For non-LADCO areas, C1 and C2 monthly profiles are flat and C3 monthly profiles are highest (but not significantly different from the rest of the year) in the summer.

For the rail sector, new monthly profiles were developed for the 2014 platform. Monthly temporalization for rail freight emissions is based on AAR Rail Traffic Data, Total Carloads and Intermodal, for 2014. For passenger trains, monthly temporalization is based on rail passenger miles data for 2014 from the

Bureau of Transportation Statistics. Rail emissions are allocated with flat day of week profiles, and most emissions are allocated with flat hourly profiles.

For the ptfire and ptagfire sectors, the inventories are in the daily point fire format, so temporal profiles are only used to go from day-specific to hourly emissions. For ptfire, state-specific hourly profiles were used, with distinct profiles for prescribed fires and wildfires. For ptagfire, the diurnal temporal profile used reflected the fact that burning occurs during the daylight hours. Additional details on these profiles are available in the 2014v7.1 TSD.

For the nonroad sector, while the NEI only stores the annual totals, the modeling platform uses monthly inventories from output from MOVES. For California, CARB's annual inventory was temporalized to monthly using monthly temporal profiles applied in SMOKE by SCC. This is an improvement over the 2011 platform, which applied monthly temporalization in California at the broader SCC7 level.

Diurnal, weekly, and monthly temporal profiles for aviation-related sources were updated in the 2014v7.0 platform based on aviation metrics. Details on these new profiles are available in the 2014v7.0 TSD. Temporal profiles for small airports (i.e., non-commercial) do not have any emissions between 10pm and 6am due to a lack of tower operations. Industrial processes that are not likely to shut down on Sundays such as those at cement plants are assigned to other more realistic profiles that included emissions on Sundays. This also affected emissions on holidays because Sunday emissions are also used on holidays.

For oil and gas sources, monthly oil and gas temporal profiles by county and SCC from the 2014v7.1 platform were not used for this study. The underlying data for those temporal profiles is too specific to the year 2014 to be used for any other year such as 2015. Instead, oil and gas sources use a flat monthly profile for 2015. Weekly and diurnal profiles are flat and are based on comments received on a version of the 2011 platform.

For agricultural livestock, annual-to-month profiles were developed based on daily emissions data output from the CMU model by state and SCC. These profiles were used to temporally allocate 2014NEIv2 ag livestock emissions to monthly emissions, which are further temporally allocated to hours as described below in section 3.3.5.3.

3.3.5.2 Temporal Profiles for EGUs

The 2015 annual EGU emissions not matched to CEMS sources use region/fuel specific profiles based on average hourly emissions for each respective region and fuel type. Peaking units were removed during the averaging to minimize the spikes generated by those units. The non-matched units are allocated to hourly emissions using a three-step process: annual value to monthly value, monthly to daily, and daily to hourly. Prior to temporal allocation or the calculation of average profiles, the CEMS data were processed using a tool that reviewed the data quality flags that indicate the data were not measured because unmeasured data can cause erroneously high values in the CEMS data. If the data were not measured at specific hours, and those values were found to be more than three times the annual mean for that unit, the data for those hours were replaced with annual mean values (Adelman et al., 2012). These adjusted CEMS data were then used for the remainder of the temporalization process described below (see Figure 3-4 for an example).

Winter and summer seasons are included in the development of the diurnal profiles as opposed to using data for the entire year because analysis of the hourly CEMS data revealed that there were different

diurnal patterns in winter versus summer in many areas. For the purposes of diurnal temporal allocation of EGU emissions winter is defined as January through April and October through December, while summer is defined as May through September. Typically, a single mid-day peak is visible in the summer, while there are morning and evening peaks in the winter, an example of which is shown Figure 3-5.

The temporal allocation procedure is differentiated by whether or not the source could be directly matched to a CEMS unit via ORIS facility code and boiler ID. Note that for units matched to CEMS data, annual totals of their emissions may be different than the annual values in NEI because the CEMS data actually replaces the inventory data for the seasons in which the CEMS are operating. If a CEMS-matched unit is determined to be a partial year reporter, as can happen for sources that run CEMS only in the summer, emissions totaling the difference between the annual emissions and the total CEMS emissions are allocated to the non-summer months.

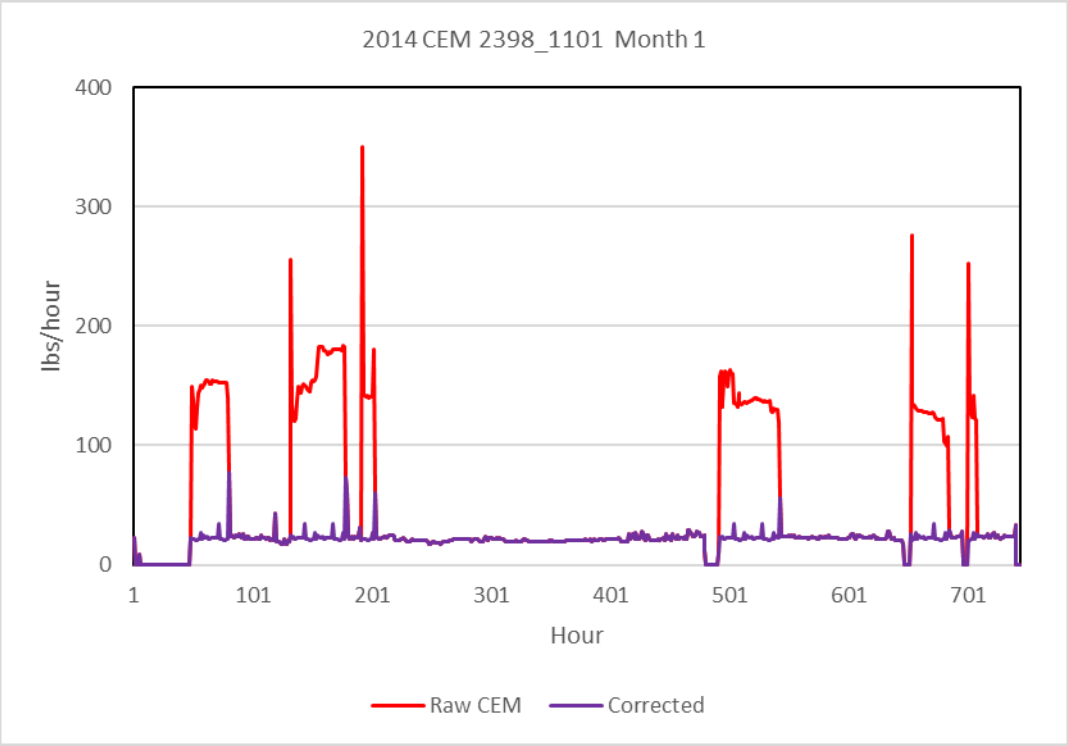


Figure 3-4. Eliminating unmeasured spikes in CEMS data

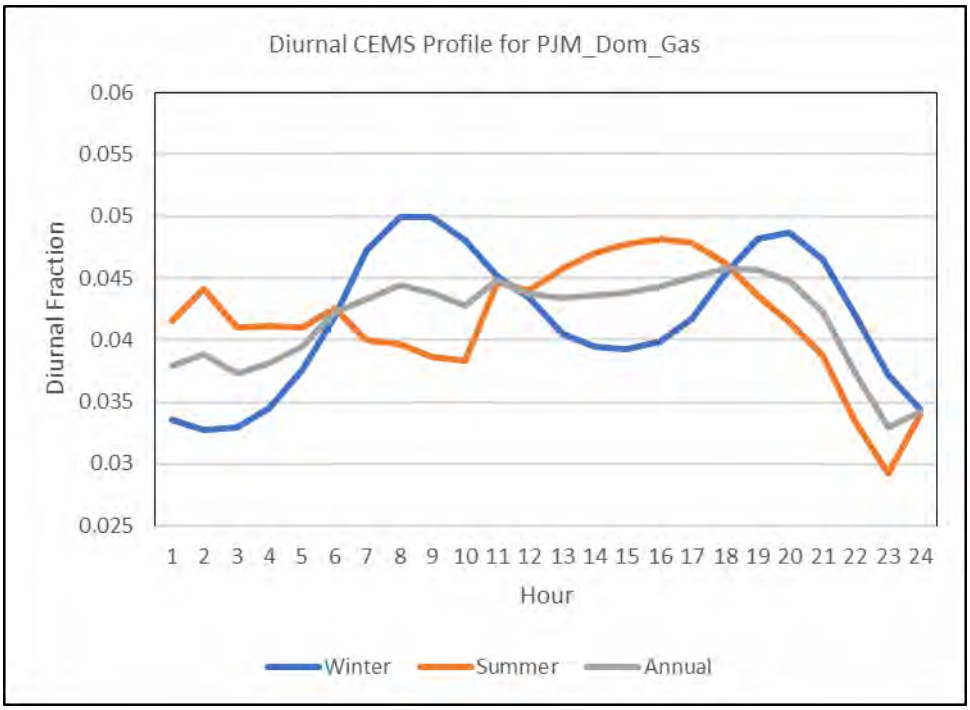


Figure 3-5. Seasonal diurnal profiles for EGU emissions in a Virginia Region

For sources not matched to CEMS units, temporal profiles are calculated that are used by SMOKE to allocate the annual emissions to hourly values. For these units, the allocation of the inventory annual emissions to months is done using average fuel-specific annual-to-month factors generated for each of the 64 IPM regions shown in in Figure 3-6. These factors are based on 2015 CEMS data only. In each region, separate factors were developed for the fuels: coal, natural gas, and "other," where the types of fuels included in "other" vary by region. Separate profiles were computed for NO_X, SO_2, and heat input. An overall composite profile was also computed and used when there were no CEMS units with the specified fuel in the region containing the unit. For both CEMS-matched units and units not matched to CEMS, NO_X and SO_2 CEMS data are used to allocate NO_X and SO_2 emissions to monthly emissions, respectively, while heat input data are used allocate emissions of all pollutants from monthly to daily emissions.

Daily temporal allocation of units matched to CEMS was performed using a procedure similar to the approach to allocate emissions to months where the hourly CEMS data replaces the inventory data for each pollutant. For the CEMS matched units, NO_X and SO_2 CEMS data are used to replace and temporally allocate NO_X and SO_2 emissions, while CEMS heat input data are used to allocate all other pollutants. For units without CEMS data emissions were allocated from month to day using IPM-region and fuel-specific average month-to-day factors based on the 2015 CEMS hourly heat data. Separate month-to-day allocation factors were computed for each month of the year using heat input for the fuel types coal, natural gas, and "other" in each respective region. An example of month-to-day profiles for gas, coal, and an overall composite for a region in western Texas is shown in Figure 3-7.

For units matched to CEMS data, hourly emissions use the hourly CEMS values for NO_X and SO_2, while other pollutants are allocated according to heat input values. For units not matched to CEMS data, temporal profiles from days to hours are computed based on the season-, region- and fuel-specific average day-to-hour factors derived from the CEMS data for those fuels and regions using the appropriate subset

of data. For the unmatched units, CEMS heat input data are used to allocate *all* pollutants (including NO_X and SO_2) because the heat input data was generally found to be more complete than the pollutant-specific data. SMOKE then allocates the daily emissions data to hours using the temporal profiles obtained from the CEMS data for the analysis base year (i.e., 2015 in this case).

Certain sources without CEMS data that typically run at a constant hourly rate, such as specific municipal waste combustors (MWCs) and cogeneration facilities (cogens), were assigned a flat temporal profile by source. The emissions for these sources have an equal value for each day of the year.

All 2015 CEMS data used for this study, whether directly for CEMS matched units, or indirectly for the calculation of monthly and daily temporal profiles for units without CEMS matches, were based on the version of the 2015 CEMS that was published on March 14, 2018.

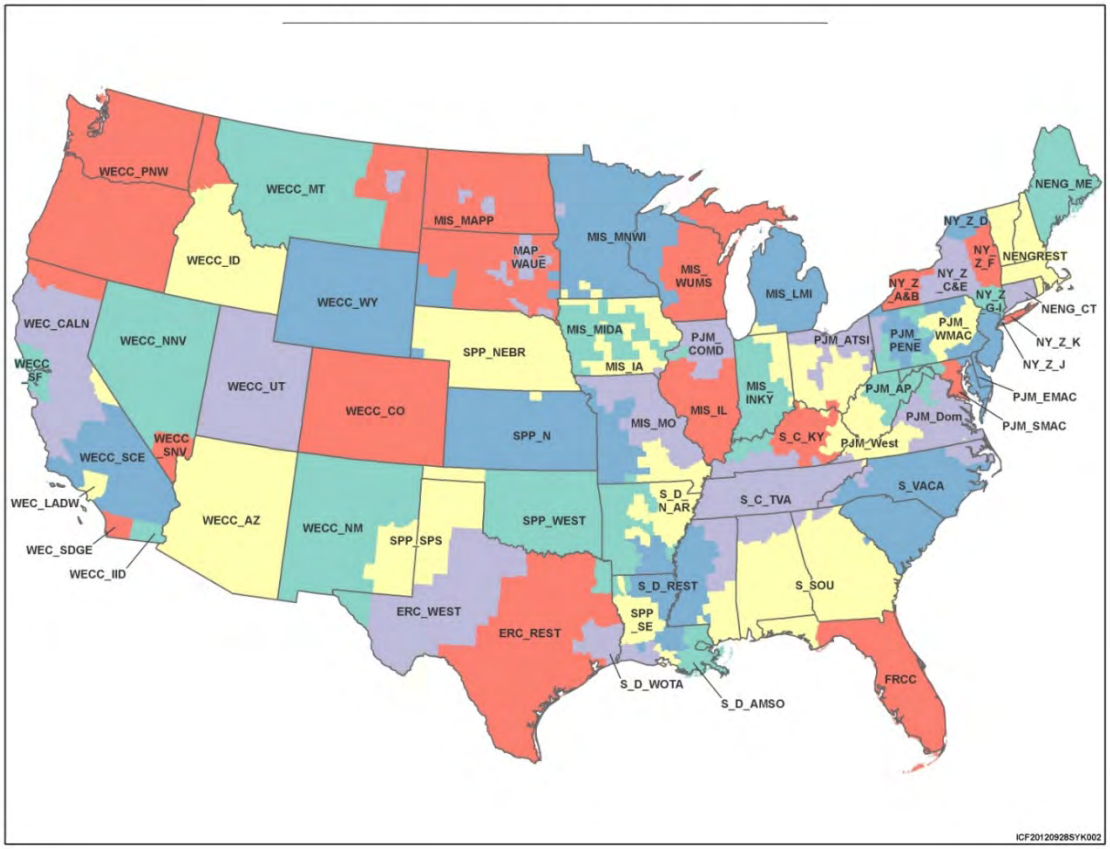


Figure 3-6. IPM Regions for EPA Base Case v5.13

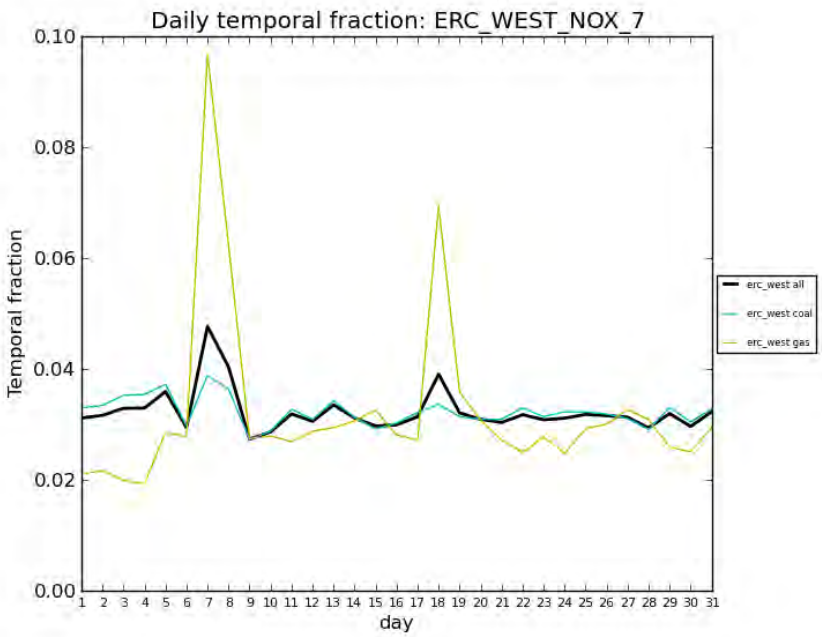


Figure 3-7. Month-to-day profiles for different fuels in a West Texas Region

3.3.5.3 Meteorological-based Temporal Profiles

There are many factors that impact the timing of when emissions occur, and for some sectors this includes meteorology. The benefits of utilizing meteorology as method for temporalization are: (1) a meteorological dataset consistent with that used by the AQ model is available (e.g., outputs from WRF); (2) the meteorological model data are highly resolved in terms of spatial resolution; and (3) the meteorological variables vary at hourly resolution and can therefore be translated into hour-specific temporalization.

The SMOKE program GenTPRO provides a method for developing meteorology-based temporalization. Currently, the program can utilize three types of temporal algorithms: annual-to-day temporalization for residential wood combustion (RWC), month-to-hour temporalization for agricultural livestock ammonia, and a generic meteorology-based algorithm for other situations. For this platform, meteorological-based temporalization was used for portions of the rwc sector and for the entirety of the ag sector.

GenTPRO reads in gridded meteorological data (output from MCIP) along with spatial surrogates and uses the specified algorithm to produce a new temporal profile that can be input into SMOKE. The meteorological variables and the resolution of the generated temporal profile (hourly, daily, etc.) depend on the selected algorithm and the run parameters. For more details on the development of these algorithms and running GenTPRO, see the GenTPRO documentation and the SMOKE documentation at http://www.cmascenter.org/smoke/documentation/3.1/GenTPRO_TechnicalSummary_Aug2012_Final.pdf and https://www.cmascenter.org/smoke/documentation/4.5/html/ch05s03s05.html, respectively.

As of the 2011v6.2 platform and in SMOKE 3.6.5, the temporal profile format was updated to support more flexibility in profile application. The corresponding version of GenTPRO produces separate files including the monthly temporal profiles (ATPRO_MONTHLY) and day-of-month temporal profiles

(ATPRO_DAILY), instead of a single ATPRO_DAILY with day-of-year temporal profiles as it did in SMOKE 3.5. The new and old temporal allocation results are equivalent when given the same inputs.

For the RWC algorithm, GenTPRO uses the daily minimum temperature to determine the temporal allocation of emissions to days. GenTPRO was used to create an annual-to-day temporal profile for the RWC sources. These generated profiles distribute annual RWC emissions to the coldest days of the year. On days where the minimum temperature does not drop below a user-defined threshold, RWC emissions for most sources in the sector are zero. Conversely, the program temporally allocates the largest percentage of emissions to the coldest days. Similar to other temporal allocation profiles, the total annual emissions do not change, only the distribution of the emissions within the year is affected. The temperature threshold for rwc emissions was 50 ˚F for most of the country, and 60 ˚F for the following states: Alabama, Arizona, California, Florida, Georgia, Louisiana, Mississippi, South Carolina, and Texas.

Figure 3-8 illustrates the impact of changing the temperature threshold for a warm climate county. The plot shows the temporal fraction by day for Duval County, Florida for the first four months of 2007. The default 50 ˚F threshold creates large spikes on a few days, while the 60 ˚F threshold dampens these spikes and distributes a small amount of emissions to the days that have a minimum temperature between 50 and 60 ˚F.

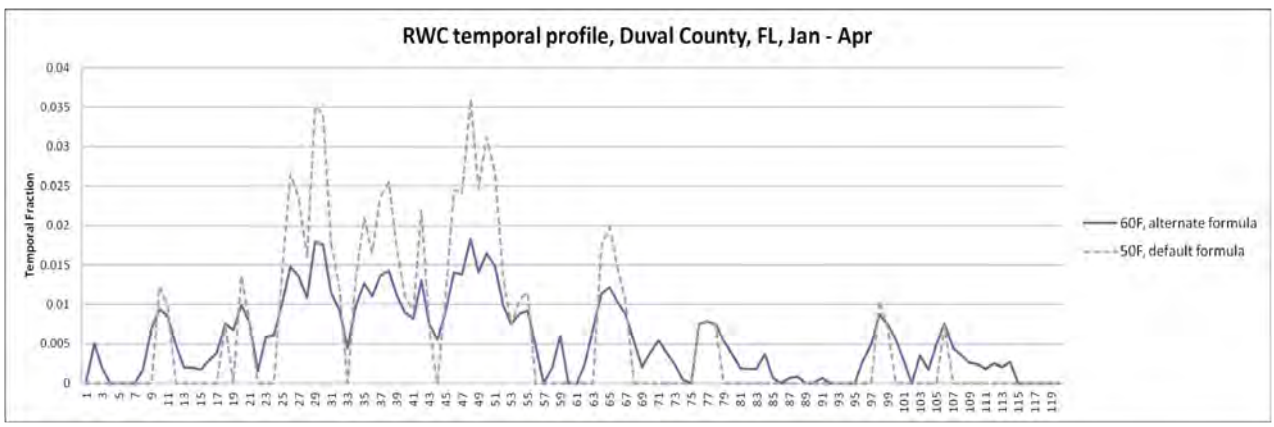


Figure 3-8. Example of RWC temporalization in 2007 using a 50 versus 60 ˚F threshold

The diurnal profile for used for most RWC sources places more of the RWC emissions in the morning and the evening when people are typically using these sources. This profile is based on a 2004 MANE-VU survey based temporal profiles (see http://www.marama.org/publications_folder/ResWoodCombustion/Final_report.pdf). This profile was created by averaging three indoor and three RWC outdoor temporal profiles from counties in Delaware and aggregating them into a single RWC diurnal profile. This new profile was compared to a concentration based analysis of aethalometer measurements in Rochester, NY (Wang *et al.* 2011) for various seasons and day of the week and found that the new RWC profile generally tracked the concentration based temporal patterns.

The temporalization for "Outdoor Hydronic Heaters" (i.e.,"OHH", SCC=2104008610) and "Outdoor wood burning device, NEC (fire-pits, chimneas, etc.)" (i.e., "recreational RWC", SCC=21040087000) were updated because the meteorological-based temporalization used for the rest of the rwc sector did not agree with observations for how these appliances are used. For OHH, the annual-to-month, day-of-week

and diurnal profiles were modified based on information in the New York State Energy Research and Development Authority (NYSERDA) "Environmental, Energy Market, and Health Characterization of Wood-Fired Hydronic Heater Technologies, Final Report" (NYSERDA, 2012) as well as a Northeast States for Coordinated Air Use Management (NESCAUM) report "Assessment of Outdoor Wood-fired Boilers" (NESCAUM, 2006). A Minnesota 2008 Residential Fuelwood Assessment Survey of individual household responses (MDNR, 2008) provided additional annual-to-month, day-of-week and diurnal activity information for OHH as well as recreational RWC usage.

The diurnal profile for OHH, shown in Figure 3-9 is based on a conventional single-stage heat load unit burning red oak in Syracuse, New York. The NESCAUM report describes how for individual units, OHH are highly variable day-to-day but that in the aggregate, these emissions have no day-of-week variation. In contrast, the day-of-week profile for recreational RWC follows a typical "recreational" profile with emissions peaked on weekends. Annual-to-month temporalization for OHH as well as recreational RWC were computed from the MN DNR survey (MDNR, 2008) and are illustrated in Figure 3-10. OHH emissions still exhibit strong seasonal variability, but do not drop to zero because many units operate year-round for water and pool heating. In contrast to all other RWC appliances, recreational RWC emissions are used far more frequently during the warm season.

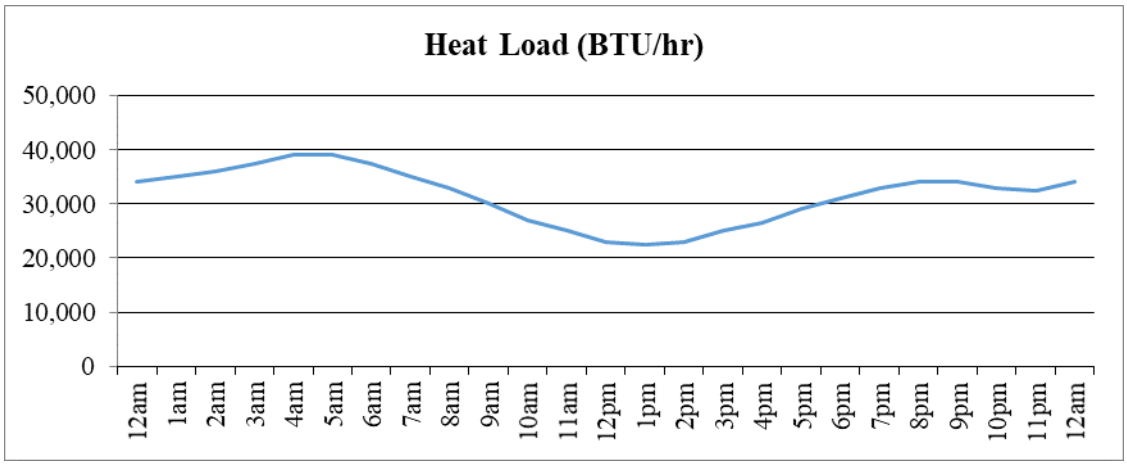


Figure 3-9. Diurnal profile for OHH, based on heat load (BTU/hr)

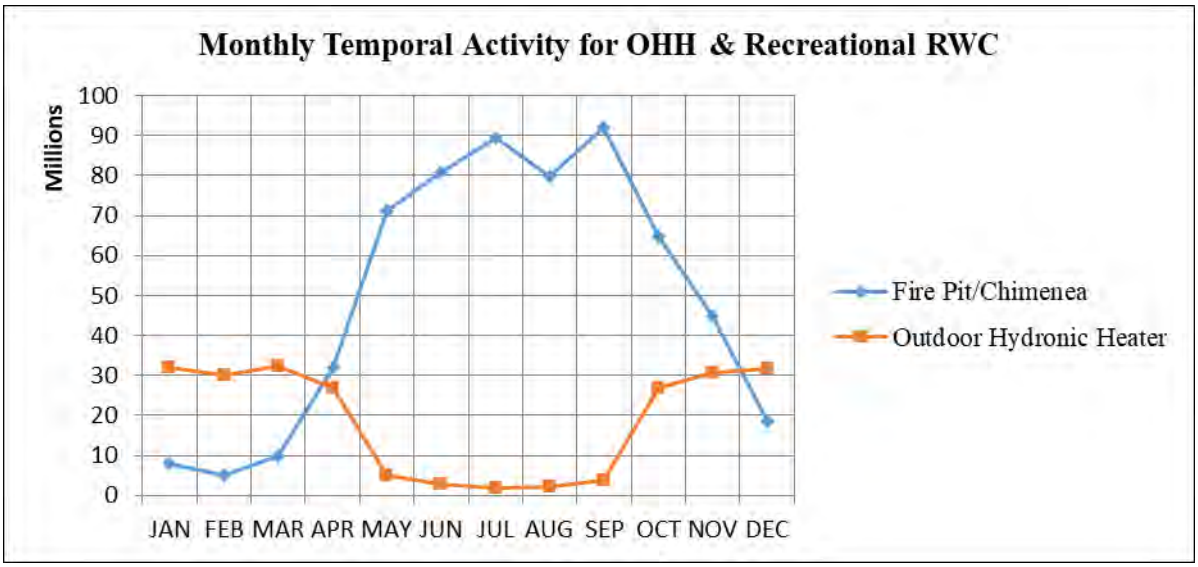


Figure 3-10. Annual-to-month temporal profiles for OHH and recreational RWC

For the ag sector, agricultural GenTPRO temporalization was applied to both livestock and fertilizer emissions, and to all pollutants within the ag sector, not just NH_3. This is a change from the 2014v7.0 modeling platform, in which agricultural GenTPRO temporalization was only applied to livestock NH_3 sources. The GenTPRO algorithm is based on an equation derived by Jesse Bash of EPA ORD based on the Zhu, Henze, et al. (2014) empirical equation. This equation is based on observations from the TES satellite instrument with the GEOS-Chem model and its adjoint to estimate diurnal NH3 emission variations from livestock as a function of ambient temperature, aerodynamic resistance, and wind speed. The equations are:

$$E_{i,h} = [161500/T_{i,h} \text{ x } e^{(-1380/T_{i,h})}] \text{ x } AR_{i,h}$$

$$PE_{i,h} = E_{i,h} / \text{Sum}(E_{i,h})$$

where

- $PE_{i,h}$ = Percentage of emissions in county *i* on hour *h*
- $E_{i,h}$ = Emission rate in county *i* on hour *h*
- $T_{i,h}$ = Ambient temperature (Kelvin) in county *i* on hour *h*
- $V_{i,h}$ = Wind speed (meter/sec) in county *i* (minimum wind speed is 0.1 meter/sec)
- $AR_{i,h}$ = Aerodynamic resistance in county *i*

GenTPRO was run using the "BASH_NH3" profile method to create month-to-hour temporal profiles for these sources. Because these profiles distribute to the hour based on monthly emissions, the monthly emissions are obtained from a monthly inventory, or from an annual inventory that has been temporalized to the month. Figure 3-11 compares the daily emissions for Minnesota from the "old" approach (uniform monthly profile) with the "new" approach (GenTPRO generated month-to-hour profiles). Although the GenTPRO profiles show daily (and hourly variability), the monthly total emissions are the same between the two approaches.

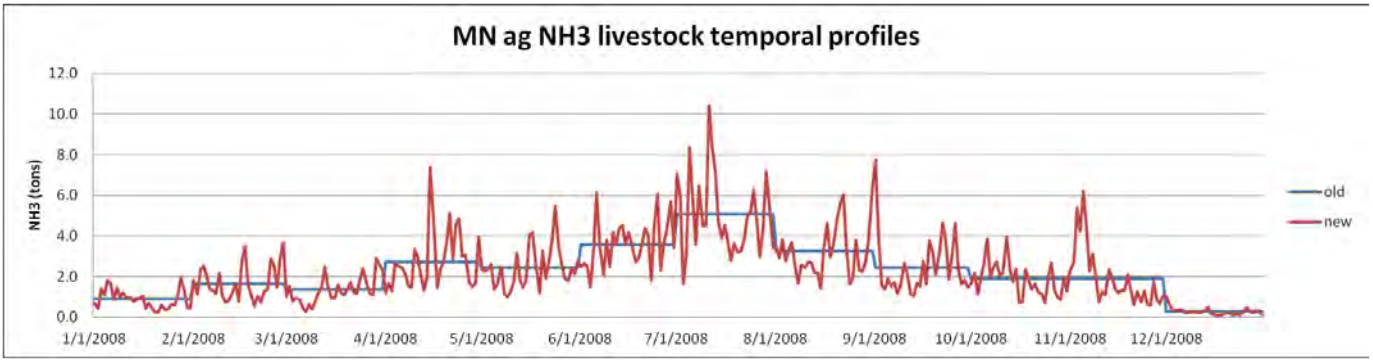


Figure 3-11. Example of animal NH_3 emissions temporalization approaches, summed to daily emissions

For the afdust sector, meteorology is not used in the development of the temporal profiles, but it is used to reduce the total emissions based on meteorological conditions. These adjustments are applied through sector-specific scripts, beginning with the application of land use-based gridded transport fractions and then subsequent zero-outs for hours during which precipitation occurs or there is snow cover on the ground. The land use data used to reduce the NEI emissions explains the amount of emissions that are subject to transport. This methodology is discussed in Pouliot, et al., 2010, and in "Fugitive Dust Modeling for the 2008 Emissions Modeling Platform" (Adelman, 2012). The precipitation adjustment is applied to remove all emissions for days where measureable rain occurs. Therefore, the afdust emissions vary day-to-day based on the precipitation and/or snow cover for that grid cell and day. Both the transport fraction and meteorological adjustments are based on the gridded resolution of the platform; therefore, somewhat different emissions will result from different grid resolutions. Application of the transport fraction and meteorological adjustments prevents the overestimation of fugitive dust impacts in the grid modeling as compared to ambient samples.

Biogenic emissions in the beis sector vary by every day of the year because they are developed using meteorological data including temperature, surface pressure, and radiation/cloud data. The emissions are computed using appropriate emission factors according to the vegetation in each model grid cell, while taking the meteorological data into account.

3.3.5.4 Temporal Profiles for Onroad Mobile Sources

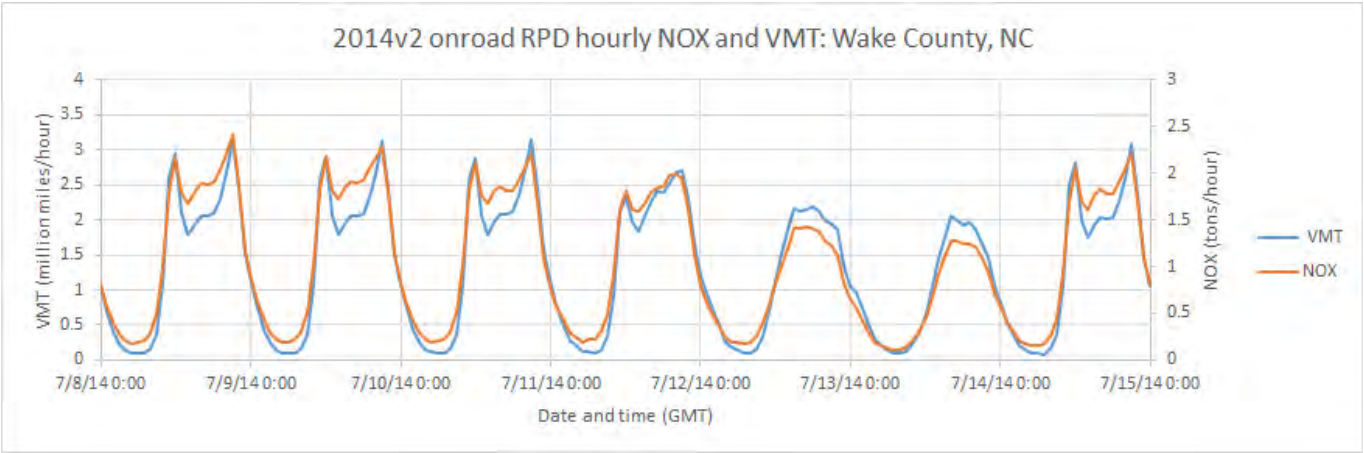
For the onroad sector, the temporal distribution of emissions is a combination of more traditional temporal profiles and the influence of meteorology. This section discusses both the meteorological influences and the updates to the diurnal temporal profiles for this platform in addition to the 2014v7.1 platform.

Meteorology is not used in the development of the temporal profiles, but rather it impacts the calculation of the hourly emissions through the program Movesmrg. The result is that the emissions vary at the hourly level by grid cell. More specifically, the on-network (RPD) and the off-network parked vehicle (RPV, RPH, and RPP) processes use the gridded meteorology (MCIP) directly. Movesmrg determines the temperature for each hour and grid cell and uses that information to select the appropriate emission factor (EF) for the specified SCC/pollutant/mode combination. In the 2014 platform (and the 2014NEIv2), RPP was updated to use the gridded minimum and maximum temperature for the day. This more spatially resolved temperature range produces more accurate emissions for each grid cell. The combination of these four processes (RPD, RPV, RPH, and RPP) is the total onroad sector emissions.

The onroad sector show a strong meteorological influence on their temporal patterns (see the 2014NEIv2 TSD for more details).

Figure 3-12 illustrates the difference between temporalization of the onroad sector and the meteorological influence via SMOKE-MOVES. Similar temporalization is done for the VMT in SMOKE-MOVES, but the meteorologically varying emission factors add variation on top of the temporalization.

Figure 3-12. Example of SMOKE-MOVES temporal variability of NO_X emissions versus activity



For the onroad sector, the "inventories" referred to in Table 3-12 actually consist of activity data, not emissions. For RPP and RPV processes, the VPOP inventory is annual and does not need temporalization. For RPD, the VMT inventory is annual for some sources and monthly for other sources, depending on the source of the data. Sources without monthly VMT were temporalized from annual to month through temporal profiles. VMT was also temporalized from month to day of the week, and then to hourly through temporal profiles. The RPD processes require a speed profile (SPDPRO) that consists of vehicle speed by hour for a typical weekday and weekend day. Unlike other sectors, the temporal profiles and SPDPRO will impact not only the distribution of emissions through time but also the total emissions. Because SMOKE-MOVES (for RPD) calculates emissions from VMT, speed and meteorology, if one shifted the VMT or speed to different hours, it would align with different temperatures and hence different emission factors. In other words, two SMOKE-MOVES runs with identical annual VMT, meteorology, and MOVES emission factors, will have different total emissions if the temporalization of VMT changes. For RPH, the HOTELING inventory is annual and was temporalized to month, day of the week, and hour of the day through temporal profiles. This is an analogous process to RPD except that speed is not included in the calculation of RPH.

New VMT day-of-week and hour-of-day temporal profiles were developed as part of the effort to update the inputs to MOVES and SMOKE-MOVES under CRC A-100 (Coordinating Research Council, 2017). CRC A-100 data includes profiles by region or county, road type, and broad vehicle category. There are three vehicle categories: passenger vehicles (11/21/31), commercial trucks (32/52), and combination trucks (53/61/62). CRC A-100 does not cover buses, refuse trucks, or motor homes, so those vehicle types were mapped to other vehicle types for which CRC A-100 did provide profiles, as follows: 1) Intercity/transit buses were mapped to commercial trucks; 2) Motor homes were mapped to passenger vehicles for day-of-week and commercial trucks for hour-of-day; 3) School buses and refuse trucks were mapped to commercial trucks for hour-of-day and use a new custom day-of-week profile called

LOWSATSUN that has a very low weekend allocation, since school buses and refuse trucks operate primarily on business days. In addition to temporal profiles, CRC A-100 data was also used to develop the average hourly speed data (SPDPRO) used by SMOKE-MOVES. In areas where CRC A-100 data does not exist, hourly speed data is based on MOVES county databases.

The CRC A-100 dataset includes temporal profiles for individual counties, Metropolitan Statistical Areas (MSAs), and entire regions (e.g. West, South). Counties without temporal profiles specific to itself, or to its MSA, are assigned to regional temporal profiles. Temporal profiles also vary between MOVES road types, and there are distinct hour-of-day profiles for each day of the week. Plots of hour-of-day profiles for passenger vehicles in Fulton County, GA, are shown in Figure 3-13. Separate plots are shown for Monday, Friday, Saturday, and Sunday, and each line corresponds to a particular MOVES road type (e.g. road type 2 = rural restricted). Figure 3-14 shows which counties have temporal profiles specific to that county, and which counties use regional average profiles.

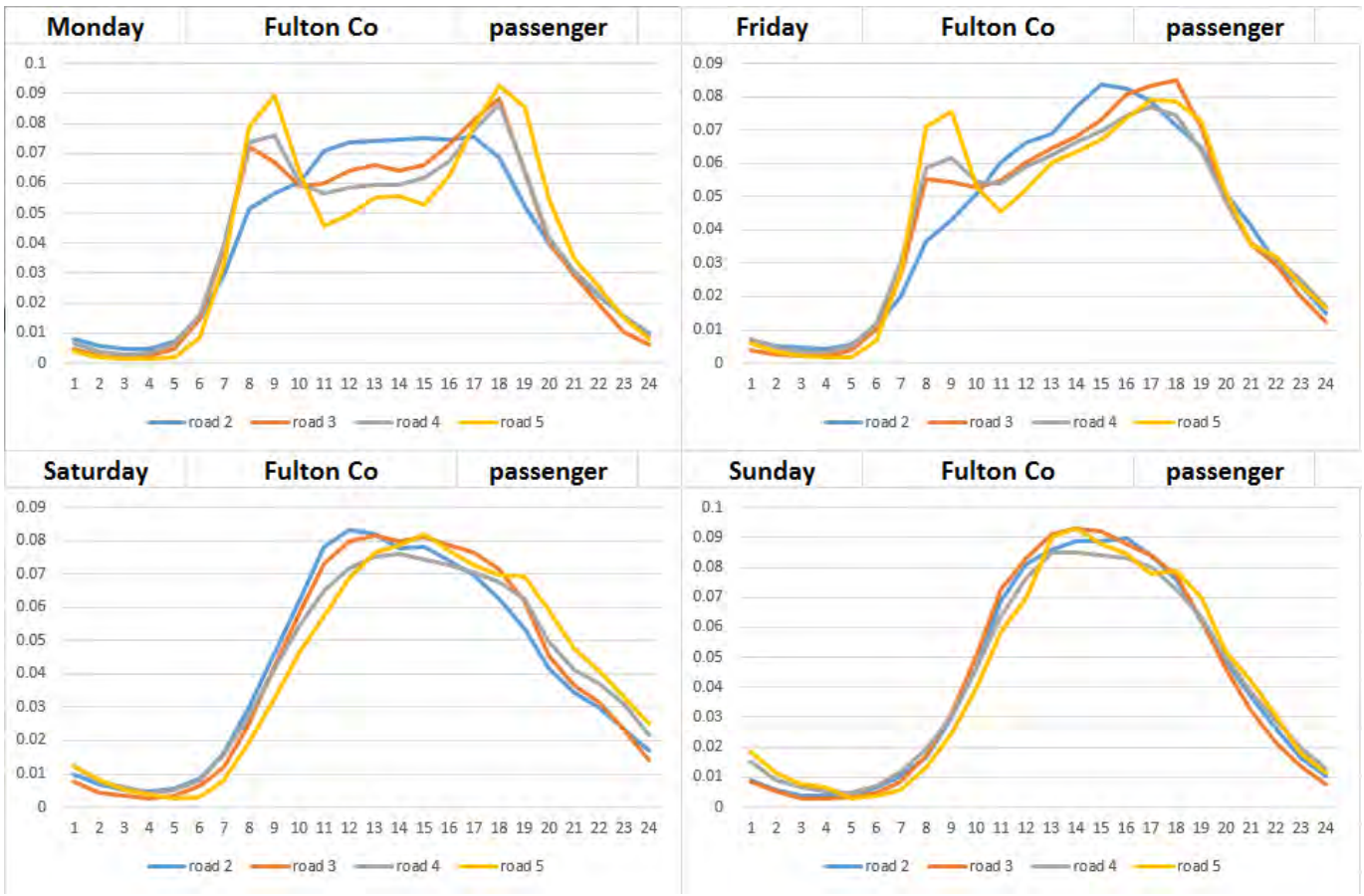


Figure 3-13. Sample onroad diurnal profiles for Fulton County, GA

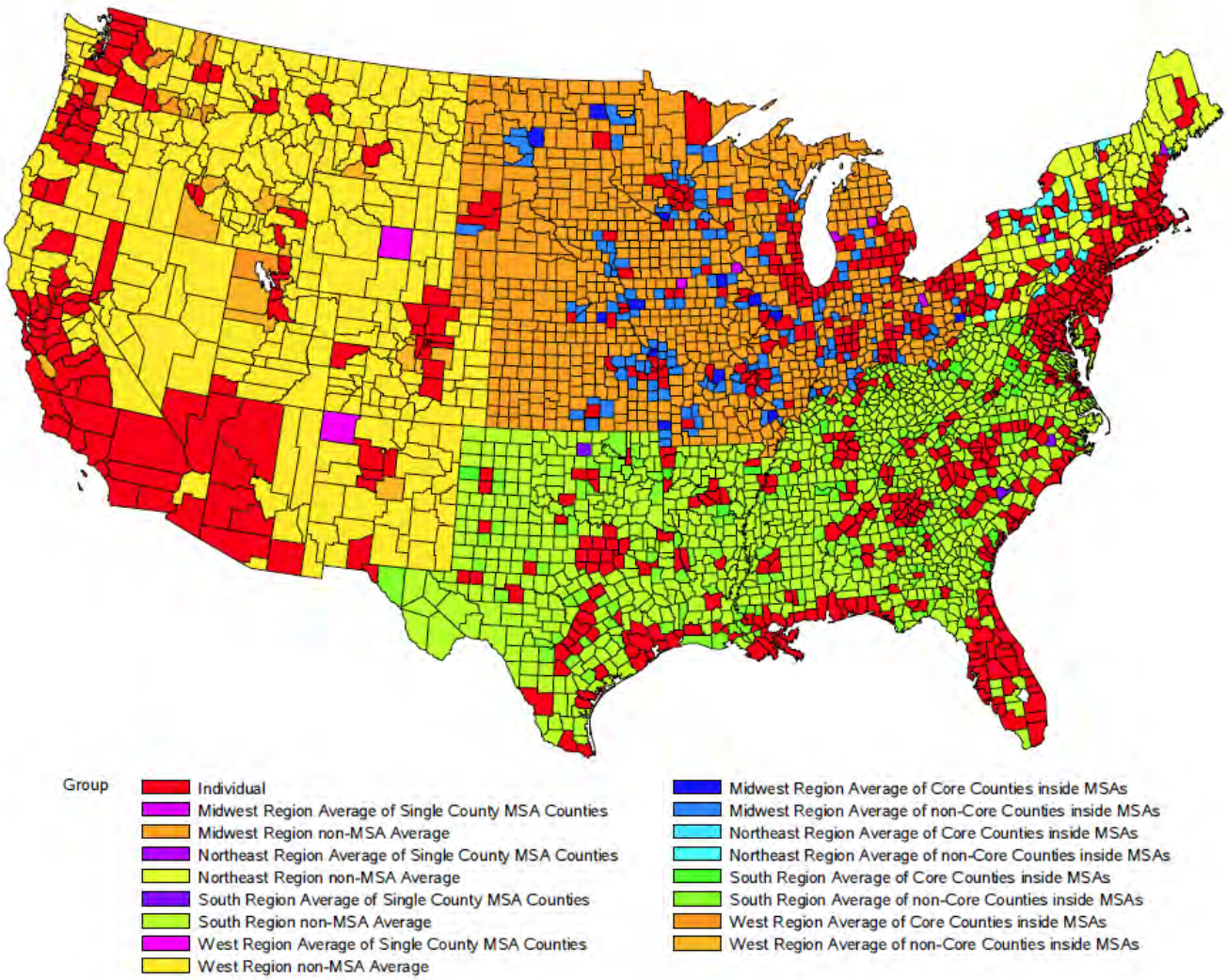


Figure 3-14. Counties for which MOVES Speeds and Temporal Profiles could be Populated

For hoteling, day-of-week profiles are the same as non-hoteling for combination trucks, while hour-of-day non-hoteling profiles for combination trucks were inverted to create new hoteling profiles that peak overnight instead of during the day.

The CRC A-100 temporal profiles were used in the entire contiguous United States, except in California. All California temporal profiles were carried over from the 2014v7.1 platform, although California hoteling uses CRC A-100-based profiles just like the rest of the country, since CARB didn't have a hoteling-specific profile. Monthly profiles in all states (national profiles by broad vehicle type) were also carried over from 2014v1 and applied directly to the VMT. For California, CARB supplied diurnal profiles that varied by vehicle type, day of the week[13], and air basin. These CARB-specific profiles were used in developing EPA estimates for California. Although the EPA adjusted the total emissions to match interpolated 2015 levels based on California's submitted inventories for 2014 and 2017, the

13 California's diurnal profiles varied within the week. Monday, Friday, Saturday, and Sunday had unique profiles and Tuesday, Wednesday, Thursday had the same profile.

temporalization of these emissions took into account both the state-specific VMT profiles and the SMOKE-MOVES process of incorporating meteorology. For more details on the adjustments to California's onroad emissions, see the 2014v7.1 TSD.

3.3.6 Vertical Allocation of Emissions

Table 3-7 specifies the sectors for which plume rise is calculated. If there is no plume rise for a sector, the emissions are placed into layer 1 of the air quality model. Vertical plume rise was performed in-line within CMAQ for all of the SMOKE point-source sectors (i.e., ptegu, ptnonipm, ptfire, ptagfire, ptfire_othna, othpt, and cmv_c3). The in-line plume rise computed within CMAQ is nearly identical to the plume rise that would be calculated within SMOKE using the Laypoint program. The selection of point sources for plume rise is pre-determined in SMOKE using the Elevpoint program. The calculation is done in conjunction with the CMAQ model time steps with interpolated meteorological data and is therefore more temporally resolved than when it is done in SMOKE. Also, the calculation of the location of the point sources is slightly different than the one used in SMOKE and this can result in slightly different placement of point sources near grid cell boundaries.

For point sources, the stack parameters are used as inputs to the Briggs algorithm, but point fires do not have stack parameters. However, the ptfire, ptagfire, and ptfire_othna inventories do contain data on the acres burned (acres per day) and fuel consumption (tons fuel per acre) for each day. CMAQ uses these additional parameters to estimate the plume rise of emissions into layers above the surface model layer. Specifically, these data are used to calculate heat flux, which is then used to estimate plume rise. In addition to the acres burned and fuel consumption, heat content of the fuel is needed to compute heat flux. The heat content was assumed to be 8000 Btu/lb of fuel for all fires because specific data on the fuels were unavailable in the inventory. The plume rise algorithm applied to the fires is a modification of the Briggs algorithm with a stack height of zero.

CMAQ uses the Briggs algorithm to determine the plume top and bottom, and then computes the plumes' distributions into the vertical layers that the plumes intersect. The pressure difference across each layer divided by the pressure difference across the entire plume is used as a weighting factor to assign the emissions to layers. This approach gives plume fractions by layer and source.

3.3.6 Vertical Allocation of Emissions

Table 3-5 specifies the sectors for which plume rise is calculated. If there is no plume rise for a sector, the emissions are placed into layer 1 of the air quality model. Vertical plume rise was performed in-line within CMAQ for all of the SMOKE point-source sectors (i.e., ptegu, ptnonipm, ptfire, ptagfire, othpt, and cmv_c3). The in-line plume rise computed within CMAQ is nearly identical to the plume rise that would be calculated within SMOKE using the Laypoint program. The selection of point sources for plume rise is pre-determined in SMOKE using the Elevpoint program. The calculation is done in conjunction with the CMAQ model time steps with interpolated meteorological data and is therefore more temporally resolved than when it is done in SMOKE. Also, the calculation of the location of the point sources is slightly different than the one used in SMOKE and this can result in slightly different placement of point sources near grid cell boundaries.

For point sources, the stack parameters are used as inputs to the Briggs algorithm, but point fires do not have stack parameters. However, the ptfire inventory does contain data on the acres burned (acres per day) and fuel consumption (tons fuel per acre) for each day. CMAQ uses these additional parameters to estimate the plume rise of emissions into layers above the surface model layer. Specifically, these data are

used to calculate heat flux, which is then used to estimate plume rise. In addition to the acres burned and fuel consumption, heat content of the fuel is needed to compute heat flux. The heat content was assumed to be 8000 Btu/lb of fuel for all fires because specific data on the fuels were unavailable in the inventory. The plume rise algorithm applied to the fires is a modification of the Briggs algorithm with a stack height of zero.

CMAQ uses the Briggs algorithm to determine the plume top and bottom, and then computes the plumes' distributions into the vertical layers that the plumes intersect. The pressure difference across each layer divided by the pressure difference across the entire plume is used as a weighting factor to assign the emissions to layers. This approach gives plume fractions by layer and source.

3.3.7 Emissions Modeling Spatial Allocation

The methods used to perform spatial allocation are summarized in this section. For the modeling platform, spatial factors are typically applied by county and SCC. Spatial allocation was performed for a national 12-km domain. To accomplish this, SMOKE used national 12-km spatial surrogates and a SMOKE area-to-point data file. For the U.S., EPA updated surrogates to use circa 2010-2014 data wherever possible. For Mexico, updated spatial surrogates were used as described below. For Canada, shapefiles for generating new surrogates were provided by Environment Canada for use with their 2013 and 2025 inventories. The U.S., Mexican, and Canadian 12-km surrogates cover the entire CONUS domain 12US1 shown in Figure 3-2.

3.3.7.1 Surrogates for U.S. Emissions

There are more than 100 spatial surrogates available for spatially allocating U.S. county-level emissions to the 12-km grid cells used by the air quality model. Note that an area-to-point approach overrides the use of surrogates for a limited set of sources. Table 3-13 lists the codes and descriptions of the surrogates. Surrogate names and codes listed in *italics* are not directly assigned to any sources for the 2014v7.1 platform, but they are sometimes used to gapfill other surrogates, or as an input for merging two surrogates to create a new surrogate that is used.

Many surrogates were updated or newly developed for use in the 2014v7.0 platform (Adelman, 2016). They include the use of the 2011 National Land Cover Database (the previous platform used 2006) and development of various development density levels such as open, low, medium high and various combinations of these. These landuse surrogates largely replaced the FEMA category surrogates that were used in the 2011 platform. Additionally, onroad surrogates were developed using average annual daily traffic counts from the highway monitoring performance system (HPMS). Previously, the "activity" for the onroad surrogates was length of road miles.

Several surrogates were updated or developed as new surrogates for the 2014v7.1 platform:

- C1/C2 ships at ports uses a surrogate based on 2014 NEI ports activity data based on use of the 2014NEIv1 (surrogate 820); previously, just the port shapes (801) were used.
- C1/C2 ships underway uses a 2013-shipping density surrogate (surrogate 808); previously Offshore Shipping NEI2014 Activity (806) was used.
- Oil and gas surrogates were updated to correct errors found after they were used for 2014v7.0.

- Onroad surrogates that do not distinguish between urban and rural road types, correcting the issue arising in some counties due to the inconsistent urban and rural definitions between MOVES and the surrogate data.
- Correction was made to the water surrogate to gap fill missing counties using 2006 NLCD

The surrogates for the U.S. were mostly generated using the Surrogate Tool to drive the Spatial Allocator, but a few surrogates were developed directly within ArcGIS or using scripts that manipulate spatial data in PostgreSQL . The tool and documentation for the Surrogate Tool is available at https://www.cmascenter.org/sa-tools/documentation/4.2/SurrogateToolUserGuide_4_2.pdf.

Table 3-13. U.S. Surrogates available for the 2014v7.1 modeling platform

Code	Surrogate Description	Code	Surrogate Description
N/A	Area-to-point approach (see 3.6.2)	505	Industrial Land
100	Population	506	Education
110	*Housing*	*507*	*Heavy Light Construction Industrial Land*
131	*urban Housing*	*510*	*Commercial plus Industrial*
132	*Suburban Housing*	*515*	*Commercial plus Institutional Land*
134	*Rural Housing*	*520*	*Commercial plus Industrial plus Institutional*
137	*Housing Change*	*525*	*Golf Courses plus Institutional plus Industrial plus Commercial*
140	*Housing Change and Population*	*526*	*Residential - Non-Institutional*
150	Residential Heating - Natural Gas	*527*	*Single Family Residential*
160	*Residential Heating - Wood*	535	Residential + Commercial + Industrial + Institutional + Government
170	Residential Heating - Distillate Oil	*540*	*Retail Trade (COM1)*
180	Residential Heating - Coal	*545*	*Personal Repair (COM3)*
190	Residential Heating - LP Gas	*555*	*Professional/Technical (COM4) plus General Government (GOV1)*
201	*Urban Restricted Road Miles*	560	Hospital (COM6)
202	Urban Restricted AADT	*575*	*Light and High Tech Industrial (IND2 + IND5)*
205	Extended Idle Locations	*580*	*Food Drug Chemical Industrial (IND3)*
211	*Rural Restricted Road Miles*	*585*	*Metals and Minerals Industrial (IND4)*
212	*Rural Restricted AADT*	*590*	*Heavy Industrial (IND1)*
221	*Urban Unrestricted Road Miles*	*595*	*Light Industrial (IND2)*
222	*Urban Unrestricted AADT*	*596*	*Industrial plus Institutional plus Hospitals*
231	*Rural Unrestricted Road Miles*	650	Refineries and Tank Farms
232	*Rural Unrestricted AADT*	670	Spud Count - CBM Wells
239	Total Road AADT	671	Spud Count - Gas Wells
240	Total Road Miles	672	Gas Production at Oil Wells
241	*Total Restricted Road Miles*	673	Oil Production at CBM Wells
242	All Restricted AADT	674	Unconventional Well Completion Counts
243	*Total Unrestricted Road Miles*	*676*	*Well Count - All Producing*
244	All Unrestricted AADT	*677*	*Well Count - All Exploratory*
258	Intercity Bus Terminals	678	Completions at Gas Wells
259	Transit Bus Terminals	679	Completions at CBM Wells
260	*Total Railroad Miles*	681	Spud Count - Oil Wells
261	NTAD Total Railroad Density	683	Produced Water at All Wells

Code	Surrogate Description	Code	Surrogate Description
271	NTAD Class 1 2 3 Railroad Density	685	Completions at Oil Wells
272	*NTAD Amtrak Railroad Density*	*686*	*Completions at All Wells*
273	*NTAD Commuter Railroad Density*	687	Feet Drilled at All Wells
275	*ERTAC Rail Yards*	691	Well Counts - CBM Wells
280	*Class 2 and 3 Railroad Miles*	692	Spud Count - All Wells
300	NLCD Low Intensity Development	693	Well Count - All Wells
301	*NLCD Med Intensity Development*	694	Oil Production at Oil Wells
302	*NLCD High Intensity Development*	695	Well Count - Oil Wells
303	*NLCD Open Space*	696	Gas Production at Gas Wells
304	NLCD Open + Low	697	Oil Production at Gas Wells
305	NLCD Low + Med	698	Well Count - Gas Wells
306	NLCD Med + High	699	Gas Production at CBM Wells
307	NLCD All Development	*710*	*Airport Points*
308	NLCD Low + Med + High	711	Airport Areas
309	NLCD Open + Low + Med	801	Port Areas
310	NLCD Total Agriculture	*805*	*Offshore Shipping Area*
318	*NLCD Pasture Land*	*806*	*Offshore Shipping NEI2014 Activity*
319	NLCD Crop Land	*807*	*Navigable Waterway Miles*
320	NLCD Forest Land	808	2013 Shipping Density
321	NLCD Recreational Land	820	Ports NEI2014 Activity
340	*NLCD Land*	850	Golf Courses
350	NLCD Water	860	Mines
500	*Commercial Land*	*890*	*Commercial Timber*

For the onroad sector, the on-network (RPD) emissions were allocated differently from the off-network (RPP and RPV). On-network used average annual daily traffic (AADT) data and off network used land use surrogates as shown in Table 3-14. Emissions from the extended (i.e., overnight) idling of trucks were assigned to surrogate 205 that is based on locations of overnight truck parking spaces. The underlying data in this surrogate was updated for use in the 2014v7.0 platform to include additional data sources and corrections based on comments received.

Table 3-14. Off-Network Mobile Source Surrogates

Source type	Source Type name	**Surrogate ID**	**Description**
11	Motorcycle	307	NLCD All Development
21	Passenger Car	307	NLCD All Development
31	Passenger Truck	307	NLCD All Development
32	Light Commercial Truck	308	NLCD Low + Med + High
41	Intercity Bus	258	Intercity Bus Terminals
42	Transit Bus	259	Transit Bus Terminals
43	School Bus	506	Education
51	Refuse Truck	306	NLCD Med + High
52	Single Unit Short-haul Truck	306	NLCD Med + High
53	Single Unit Long-haul Truck	306	NLCD Med + High
54	Motor Home	304	NLCD Open + Low
61	Combination Short-haul Truck	306	NLCD Med + High

Source type	Source Type name	**Surrogate ID**	**Description**
62	Combination Long-haul Truck	306	NLCD Med + High

For the oil and gas sources in the np_oilgas sector, the spatial surrogates were updated to those shown in Table 3-15 using 2014 data consistent with what was used to develop the 2014NEI nonpoint oil and gas emissions. The primary activity data source used for the development of the oil and gas spatial surrogates was data from Drilling Info (DI) Desktop's HPDI database (Drilling Info, 2015). This database contains well-level location, production, and exploration statistics at the monthly level. Due to a proprietary agreement with DI Desktop, individual well locations and ancillary production cannot be made publicly available, but aggregated statistics are allowed. These data were supplemented with data from state Oil and Gas Commission (OGC) websites (Illinois, Idaho, Indiana, Kentucky, Missouri, Nevada, Oregon and Pennsylvania, Tennessee). In many cases, the correct surrogate parameter was not available (e.g., feet drilled), but an alternative surrogate parameter was available (e.g., number of spudded wells) and downloaded. Under that methodology, both completion date and date of first production from HPDI were used to identify wells completed during 2011. In total, over 1.43 million unique wells were compiled from the above data sources. The wells cover 34 states and 1,158 counties. (ERG, 2016b). Corrections to these data were made for the 2014v7.1 platform after errors were discovered in some counties.

Table 3-15. Spatial Surrogates for Oil and Gas Sources

Surrogate Code	Surrogate Description
670	Spud Count - CBM Wells
671	Spud Count - Gas Wells
672	Gas Production at Oil Wells
673	Oil Production at CBM Wells
674	Unconventional Well Completion Counts
676	Well Count - All Producing
677	Well Count - All Exploratory
678	Completions at Gas Wells
679	Completions at CBM Wells
681	Spud Count - Oil Wells
683	Produced Water at All Wells
685	Completions at Oil Wells
686	Completions at All Wells
687	Feet Drilled at All Wells
691	Well Counts - CBM Wells
692	Spud Count - All Wells
693	Well Count - All Wells
694	Oil Production at Oil Wells
695	Well Count - Oil Wells
696	Gas Production at Gas Wells
697	Oil Production at Gas Wells
698	Well Count - Gas Wells

699	Gas Production at CBM Wells

Not all of the available surrogates are used to spatially allocate sources in the modeling platform; that is, some surrogates shown in Table 3-13 were not assigned to any SCCs, although many of the "unused" surrogates are actually used to "gap fill" other surrogates that are used. When the source data for a surrogate has no values for a particular county, gap filling is used to provide values for the surrogate in those counties to ensure that no emissions are dropped when the spatial surrogates are applied to the emission inventories. The U.S. CAP emissions allocated to the various spatial surrogates are shown in Table 3-16.

Table 3-16. Selected 2015 CAP emissions by sector for U.S. Surrogates (CONUS domain totals)

Sector	ID	Description	NH3	NOX	PM2_5	SO2	VOC
afdust	240	Total Road Miles			283,210		
afdust	304	NLCD Open + Low			1,053,145		
afdust	306	NLCD Med + High			43,636		
afdust	308	NLCD Low + Med + High			122,943		
afdust	310	NLCD Total Agriculture			987,447		
ag	310	NLCD Total Agriculture	2,823,395				179,970
cmv_c1c2	808	2013 Shipping Density	293	520,571	14,357	421	9,117
cmv_c1c2	820	Ports NEI2014 Activity	11	23,201	729	148	972
nonpt	100	Population	32,842	0	0	0	1,222,980
nonpt	150	Residential Heating - Natural Gas	47,819	227,291	3,837	1,494	13,756
nonpt	170	Residential Heating - Distillate Oil	1,861	35,101	3,978	56,026	1,241
nonpt	180	Residential Heating - Coal	20	101	53	1,086	111
nonpt	190	Residential Heating - LP Gas	121	34,432	183	762	1,332
nonpt	239	Total Road AADT	0	25	551	0	274,177
nonpt	240	Total Road Miles	0	0	0	0	34,027
nonpt	242	All Restricted AADT	0	0	0	0	5,451
nonpt	244	All Unrestricted AADT	0	0	0	0	95,292
nonpt	271	NTAD Class 1 2 3 Railroad Density	0	0	0	0	2,252
nonpt	300	NLCD Low Intensity Development	5,184	27,632	103,906	3,720	74,580
nonpt	304	NLCD Open + Low	0	0	0	0	0
nonpt	306	NLCD Med + High	28,046	200,320	238,731	65,131	948,148
nonpt	307	NLCD All Development	24	46,331	126,722	14,185	596,598
nonpt	308	NLCD Low + Med + High	1,166	185,948	16,915	19,736	65,608
nonpt	310	NLCD Total Agriculture	0	0	37	0	204,819
nonpt	319	NLCD Crop Land	0	0	95	71	293
nonpt	320	NLCD Forest Land	4,143	378	1,289	9	474
nonpt	505	Industrial Land	0	0	0	0	174
nonpt	535	Residential + Commercial + Industrial + Institutional + Government	5	2	130	0	39
nonpt	560	Hospital (COM6)	0	0	0	0	0
nonpt	650	Refineries and Tank Farms	0	22	0	0	98,989

Sector	ID	Description	NH3	NOX	PM2_5	SO2	VOC
nonpt	711	Airport Areas	0	0	0	0	282
nonpt	801	Port Areas	0	0	0	0	8,059
nonroad	261	NTAD Total Railroad Density	3	2,479	259	3	479
nonroad	304	NLCD Open + Low	4	2,114	184	5	3,075
nonroad	305	NLCD Low + Med	112	21,204	4,599	150	143,054
nonroad	306	NLCD Med + High	345	224,494	14,465	477	119,772
nonroad	307	NLCD All Development	103	35,119	15,498	128	169,155
nonroad	308	NLCD Low + Med + High	673	416,045	34,379	689	66,352
nonroad	309	NLCD Open + Low + Med	112	21,564	1,251	150	43,651
nonroad	310	NLCD Total Agriculture	484	392,828	29,323	526	45,663
nonroad	320	NLCD Forest Land	19	7,543	1,244	20	8,368
nonroad	321	NLCD Recreational Land	159	21,746	14,325	232	526,330
nonroad	350	NLCD Water	215	143,011	8,069	376	414,255
nonroad	850	Golf Courses	13	2,098	116	17	5,628
nonroad	860	Mines	2	2,711	284	3	532
np_oilgas	670	Spud Count - CBM Wells	0	0	0	0	179
np_oilgas	671	Spud Count - Gas Wells	0	0	0	0	10,213
np_oilgas	672	Gas Production at Oil Wells	0	3,114	0	21,703	132,924
np_oilgas	673	Oil Production at CBM Wells	0	60	0	0	3,510
np_oilgas	674	Unconventional Well Completion Counts	0	49,995	1,793	237	3,633
np_oilgas	678	Completions at Gas Wells	0	3,598	26	6,768	71,380
np_oilgas	679	Completions at CBM Wells	0	13	0	483	1,581
np_oilgas	681	Spud Count - Oil Wells	0	0	0	0	71,799
np_oilgas	683	Produced Water at All Wells	0	12	0	0	96,489
np_oilgas	685	Completions at Oil Wells	0	3,526	129	2,266	55,417
np_oilgas	687	Feet Drilled at All Wells	0	119,951	3,995	449	9,569
np_oilgas	691	Well Counts - CBM Wells	0	32,515	483	12	27,146
np_oilgas	692	Spud Count - All Wells	0	9,020	255	113	366
np_oilgas	693	Well Count - All Wells	0	0	0	0	191
np_oilgas	694	Oil Production at Oil Wells	0	5,446	0	6,337	1,148,869
np_oilgas	695	Well Count - Oil Wells	0	121,851	2,892	80	452,987
np_oilgas	696	Gas Production at Gas Wells	0	48,679	2,123	163	56,273
np_oilgas	697	Oil Production at Gas Wells	0	1,405	0	25	379,201
np_oilgas	698	Well Count - Gas Wells	15	318,258	5,457	299	679,839
np_oilgas	699	Gas Production at CBM Wells	0	2,489	325	26	4,837
onroad	205	Extended Idle Locations	509	182,233	2,501	73	35,634
onroad	239	Total Road AADT					6,780
onroad	242	All Restricted AADT	36,812	1,414,639	45,066	8,378	226,757
onroad	244	All Unrestricted AADT	67,151	2,138,188	82,736	17,676	590,047
onroad	258	Intercity Bus Terminals		153	2	0	35
onroad	259	Transit Bus Terminals		100	4	0	222
onroad	304	NLCD Open + Low		779	19	1	2,595

Sector	ID	Description	NH3	NOX	PM2_5	SO2	VOC
onroad	306	NLCD Med + High		15,884	317	18	18,741
onroad	307	NLCD All Development		608,367	12,902	965	1,253,173
onroad	308	NLCD Low + Med + High		40,355	744	62	64,388
onroad	506	Education		545	21	1	835
rail	261	NTAD Total Railroad Density	4	15,222	368	286	873
rail	271	NTAD Class 1 2 3 Railroad Density	359	657,335	18,786	415	33,866
rwc	300	NLCD Low Intensity Development	15,331	30,493	313,945	7,684	338,465

3.3.7.2 Allocation Method for Airport-Related Sources in the U.S.

There are numerous airport-related emission sources in the NEI, such as aircraft, airport ground support equipment, and jet refueling. The modeling platform includes the aircraft and airport ground support equipment emissions as point sources. For the modeling platform, EPA used the SMOKE "area-to-point" approach for only jet refueling in the nonpt sector. The following SCCs use this approach: 2501080050 and 2501080100 (petroleum storage at airports), and 2810040000 (aircraft/rocket engine firing and testing). The ARTOPNT approach is described in detail in the 2002 platform documentation: https://www3.epa.gov/scram001/reports/Emissions%20TSD%20Vol1_02-28-08.pdf. The ARTOPNT file that lists the nonpoint sources to locate using point data were unchanged from the 2005-based platform.

3.3.7.3 Surrogates for Canada and Mexico Emission Inventories

The surrogates for Canada to spatially allocate the Canadian emissions are based on the 2013 Canadian inventories and associated data. The spatial surrogate data came from Environment Canada, along with cross references. The shapefiles they provided were used in the Surrogate Tool (previously referenced) to create spatial surrogates. The Canadian surrogates used for this platform are listed in Table 3-17. The population surrogate was updated for Mexico for the 2014v7.1 platform. Surrogate code 11, which uses 2015 population data at 1 km resolution, replaces the previous population surrogate code 10. The other surrogates for Mexico are circa 1999 and 2000 and were based on data obtained from the Sistema Municpal de Bases de Datos (SIMBAD) de INEGI and the Bases de datos del Censo Economico 1999. Most of the CAPs allocated to the Mexico and Canada surrogates are shown in Table 3-18. The entries in Table 3-18 are for the othar, othafdust, onroad_can, and onroad_mex sectors.

Table 3-17 Canadian Spatial Surrogates

Code	Canadian Surrogate Description	Code	Description
100	Population	941	PAVED ROADS
101	total dwelling	942	UNPAVED ROADS
106	ALL_INDUST	945	Commercial Marine Vessels
113	Forestry and logging	948	Forest
115	Agriculture and forestry activities	950	Combination of Forest and Dwelling
200	Urban Primary Road Miles	955	UNPAVED_ROADS_AND_TRAILS
210	Rural Primary Road Miles	960	TOTBEEF
212	Mining except oil and gas	965	TOTBEEF_CD
220	Urban Secondary Road Miles	966	TOTPOUL_CD
221	Total Mining	967	TOTSWIN_CD
222	Utilities	968	TOTFERT_CD

Code	Canadian Surrogate Description	Code	Description
230	Rural Secondary Road Miles	970	TOTPOUL
240	Total Road Miles	980	TOTSWIN
308	Food manufacturing	990	TOTFERT
321	Wood product manufacturing	996	urban_area
323	Printing and related support activities	1211	Oil and Gas Extraction
324	Petroleum and coal products manufacturing	1212	OilSands
326	Plastics and rubber products manufacturing	1251	OFFR_TOTFERT
327	Non-metallic mineral product manufacturing	1252	OFFR_MINES
331	Primary Metal Manufacturing	1253	OFFR Other Construction not Urban
412	Petroleum product wholesaler-distributors	1254	OFFR Commercial Services
416	Building material and supplies wholesaler-distributors	1255	OFFR Oil Sands Mines
448	clothing and clothing accessories stores	1256	OFFR Wood industries CANVEC
562	Waste management and remediation services	1257	OFFR Unpaved Roads Rural
921	Commercial Fuel Combustion	1258	OFFR_Utilities
923	TOTAL INSTITUTIONAL AND GOVERNEMNT	1259	OFFR total dwelling
924	Primary Industry	1260	OFFR_water
925	Manufacturing and Assembly	1261	OFFR_ALL_INDUST
926	Distribtution and Retail (no petroleum)	1262	OFFR Oil and Gas Extraction
927	Commercial Services	1263	OFFR_ALLROADS
931	OTHERJET	1264	OFFR_OTHERJET
932	CANRAIL	1265	OFFR_CANRAIL

Table 3-18. CAPs Allocated to Mexican and Canadian Spatial Surrogates in 2015

Code	Mexican or Canadian Surrogate Description	NH_3	NO_X	PM_{2_5}	SO_2	VOC
11	MEX 2015 Population	26,089	119,206	4,128	473	142,715
14	MEX Residential Heating - Wood	0	1,323	16,963	203	116,625
16	MEX Residential Heating - Distillate Oil	0	13	0	4	0
20	MEX Residential Heating - LP Gas	0	5,649	171	0	96
22	MEX Total Road Miles	2,725	360,388	10,170	5,886	73,886
24	MEX Total Railroads Miles	0	22,751	508	199	887
26	MEX Total Agriculture	177,847	135,558	28,722	6,492	10,886
32	MEX Commercial Land	0	75	1,634	0	23,657
34	MEX Industrial Land	4	1,109	1,975	0	120,470
36	MEX Commercial plus Industrial Land	0	2,123	30	5	98,045

Code	Mexican or Canadian Surrogate Description	NH_3	NO_X	PM_{2_5}	SO_2	VOC
38	MEX Commercial plus Institutional Land	3	1,699	76	3	49
40	MEX Residential (RES1-4)+Comercial+Industrial+Institutional+Government	0	4	11	0	76,212
42	MEX Personal Repair (COM3)	0	0	0	0	5,773
44	MEX Airports Area	0	3,410	97	441	1,166
50	MEX Mobile sources - Border Crossing	5	146	1	3	267
100	CAN Population	738	65	757	13	341
101	CAN total dwelling	408	35,050	2,572	4,715	144,742
106	CAN ALL_INDUST	0	0	11,874	0	70
113	CAN Forestry and logging	496	2,718	0	144	7,429
115	CAN Agriculture and forestry activities	51	593	2,936	13	1,715
200	CAN Urban Primary Road Miles	1,903	86,881	3,720	299	11,467
210	CAN Rural Primary Road Miles	768	52,938	2,049	125	5,000
212	CAN Mining except oil and gas	0	0	3,522	0	0
220	CAN Urban Secondary Road Miles	3,560	132,864	7,157	636	28,328
221	CAN Total Mining	0	0	57,248	0	0
222	CAN Utilities	81	9,310	55,508	3,166	218
230	CAN Rural Secondary Road Miles	1,998	91,918	3,867	328	13,083
240	CAN Total Road Miles	45	71,550	2,600	77	114,728
308	CAN Food manufacturing	0	0	11,383	0	6,107
321	CAN Wood product manufacturing	292	1,921	0	151	8,039
323	CAN Printing and related support activities	0	0	0	0	11,824
324	CAN Petroleum and coal products manufacturing	0	1,067	1,328	419	6,397
326	CAN Plastics and rubber products manufacturing	0	0	0	0	23,116
327	CAN Non-metallic mineral product manufacturing	0	0	6,841	0	0
331	CAN Primary Metal Manufacturing	0	157	5,652	51	74
412	CAN Petroleum product wholesaler-distributors	0	0	0	0	40,328
448	CAN clothing and clothing accessories stores	0	0	0	0	116
562	CAN Waste management and remediation services	223	1,670	2,313	2,328	16,570
921	CAN Commercial Fuel Combustion	200	25,117	2,323	4,840	1,182
923	CAN TOTAL INSTITUTIONAL AND GOVERNEMNT	0	0	0	0	14,202
924	CAN Primary Industry	0	0	0	0	37,207

Code	Mexican or Canadian Surrogate Description	NH_3	NO_X	PM_{2_5}	SO_2	VOC
925	CAN Manufacturing and Assembly	0	0	0	0	71,905
926	CAN Distribtution and Retail (no petroleum)	0	0	0	0	7,144
927	CAN Commercial Services	0	0	0	0	31,421
932	CAN CANRAIL	55	107,033	2,529	380	5,381
941	CAN PAVED ROADS	0	0	311,668	0	0
945	CAN Commercial Marine Vessels	231	167,861	6,648	4,170	15,027
948	CAN Forest	0	20	7	0	229
950	CAN Combination of Forest and Dwelling	1,800	19,969	164,497	2,842	232,985
955	CAN UNPAVED_ROADS_AND_TRAILS	0	0	467,403	0	0
960	CAN TOTBEEF	0	0	1,241	0	264,882
965	CAN TOTBEEF_CD	280,635	0	0	0	0
966	CAN TOTPOUL_CD	23,918	0	0	0	0
967	CAN TOTSWIN_CD	68,018	0	0	0	0
968	CAN TOTFERT_CD	120,197	0	0	0	0
970	CAN TOTPOUL	0	0	182	0	243
980	CAN TOTSWIN	0	0	757	0	2,590
990	CAN TOTFERT	0	3,910	380,135	9,537	152
996	CAN urban_area	0	0	1,295	0	0
1211	CAN Oil and Gas Extraction	2	33	236,452	150	932
1212	CAN OilSands	143	2,267	0	675	1,858
1251	CAN OFFR_TOTFERT	110	110,079	8,076	80	10,776
1252	CAN OFFR_MINES	43	39,469	3,362	32	4,182
1253	CAN OFFR Other Construction not Urban	27	22,461	3,798	20	9,636
1254	CAN OFFR Commercial Services	35	17,166	2,181	29	23,255
1255	CAN OFFR Oil Sands Mines	0	0	0	0	0
1256	CAN OFFR Wood industries CANVEC	14	11,227	1,102	10	1,988
1257	CAN OFFR Unpaved Roads Rural	34	9,881	1,739	29	68,512
1258	CAN OFFR_Utilities	17	8,353	527	14	10,462
1259	CAN OFFR total dwelling	18	5,297	1,432	15	35,438
1260	CAN OFFR_water	9	2,246	343	13	20,736
1261	CAN OFFR_ALL_INDUST	4	4,040	262	3	874
1262	CAN OFFR Oil and Gas Extraction	1	992	56	1	153
1263	CAN OFFR_ALLROADS	2	1,039	75	1	518
1264	CAN OFFR_OTHERJET	1	805	70	1	71
1265	CAN OFFR_CANRAIL	0	80	8	0	14

3.4 Emissions References

Adelman, Z. 2012. Memorandum: Fugitive Dust Modeling for the 2008 Emissions Modeling Platform. UNC Institute for the Environment, Chapel Hill, NC. September, 28, 2012

Adelman, Z., M. Omary, Q. He, J. Zhao and D. Yang, J. Boylan, 2012. "A Detailed Approach for Improving Continuous Emissions Monitoring Data for Regulatory Air Quality Modeling." Presented at the 2012 International Emission Inventory Conference, Tampa, Florida. Available from https://www3.epa.gov/ttn/chief/conference/ei20/index.html#ses-5.

Anderson, G.K.; Sandberg, D.V; Norheim, R.A., 2004. Fire Emission Production Simulator (FEPS) User's Guide. Available at http://www.fs.fed.us/pnw/fera/feps/FEPS_users_guide.pdf.

ARB, 2000. "Risk Reduction Plan to Reduce Particulate Matter Emissions from Diesel-Fueled Engines and Vehicles". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. October, 2000. Available at: http://www.arb.ca.gov/diesel/documents/rrpFinal.pdf.

ARB, 2007. "Proposed Regulation for In-Use Off-Road Diesel Vehicles". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. April, 2007. Available at: http://www.arb.ca.gov/regact/2007/ordiesl07/isor.pdf

ARB, 2010a. "Proposed Amendments to the Regulation for In-Use Off-Road Diesel-Fueled Fleets and the Off-Road Large Spark-Ignition Fleet Requirements". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. October, 2010. Available at: http://www.arb.ca.gov/regact/2010/offroadlsi10/offroadisor.pdf.

ARB, 2010b. "Estimate of Premature Deaths Associated with Fine Particle Pollution (PM2.5) in California Using a U.S. Environmental Protection Agency Methodology". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. August, 2010. Available at: http://www.arb.ca.gov/research/health/pm-mort/pm-report_2010.pdf.Adelman, Z. 2012. *Memorandum: Fugitive Dust Modeling for the 2008 Emissions Modeling Platform*. UNC Institute for the Environment, Chapel Hill, NC. September, 28, 2012.

Bash, J.O., Baker, K.R., Beaver, M.R., Park, J.-H., Goldstein, A.H., 2015. Evaluation of improved land use and canopy representation in BEIS with biogenic VOC measurements in California (in preparation)

Bullock Jr., R, and K. A. Brehme (2002) "Atmospheric mercury simulation using the CMAQ model: formulation description and analysis of wet deposition results." Atmospheric Environment 36, pp 2135–2146.

Environ Corp. 2008. Emission Profiles for EPA SPECIATE Database, Part 2: EPAct Fuels (Evaporative Emissions). Prepared for U. S. EPA, Office of Transportation and Air Quality, September 30, 2008.

EPA, 2005. *EPA's National Inventory Model (NMIM), A Consolidated Emissions Modeling System for MOBILE6 and NONROAD*, U.S. Environmental Protection Agency, Office of Transportation and Air Quality, Ann Arbor, MI 48105, EPA420-R-05-024, December 2005. Available at HYPERLINK "https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P10023FZ.pdf"https://nepis.epa.gov/Exe/ZyPDF.c

gi?Dockey=P10023FZ.pdf.

EPA 2006a. SPECIATE 4.0, Speciation Database Development Document, Final Report, U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Research Triangle Park, NC 27711, EPA600-R-06-161, February 2006. Available at https://www.epa.gov/air-emissions-modeling/speciate.

EPA, 2015a. 2011 Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 6.2, 2011 Emissions Modeling Platform. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at https://www.epa.gov/air-emissions-modeling/2011-version-62-platform.

EPA, 2016b. SPECIATE Version 4.5 Database Development Documentation, U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Research Triangle Park, NC 27711, EPA/600/R-16/294, September 2016. Available at https://www.epa.gov/sites/production/files/2016-09/documents/speciate_4.5.pdf

EPA, 2018a. 2014 National Emissions Inventory, version 2 Technical Support Document. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at https://www.epa.gov/air-emissions-inventories/2014-national-emissions-inventory-nei-technical-support-document-tsd.

EPA, 2018b. Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 7.1, 2014 Emissions Modeling Platform for the 2014 National Air Toxics Assessment. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at https://www.epa.gov/air-emissions-modeling/2014-version-71-technical-support-document-tsd.

ERG, 2014a. Develop Mexico Future Year Emissions Final Report. Available at ftp://newftp.epa.gov/air/emismod/2011/v2platform/2011emissions/Mexico_Emissions_WA%204-09_final_report_121814.pdf.

ERG, 2014b. "Technical Memorandum: Modeling Allocation Factors for the 2011 NEI".

ERG, 2016a. "Technical Report: Development of Mexico Emission Inventories for the 2014 Modeling Platform."

ERG, 2016b. "Technical Memorandum: Modeling Allocation Factors for the 2014 Oil and Gas Nonpoint Tool."

Frost & Sullivan, 2010. "Project: Market Research and Report on North American Residential Wood Heaters, Fireplaces, and Hearth Heating Products Market (P.O. # PO1-IMP403-F&S). Final Report April 26, 2010". Prepared by Frost & Sullivan, Mountain View, CA 94041.

Joint Fire Science Program, 2009. Consume 3.0--a software tool for computing fuel consumption. Fire Science Brief. 66, June 2009. Consume 3.0 is available at: http://www.fs.fed.us/pnw/fera/research/smoke/consume/index.shtml

McCarty, J.L., Korontzi, S., Jutice, C.O., and T. Loboda. 2009. The spatial and temporal distribution of crop residue burning in the contiguous United States. Science of the Total Environment, 407 (21): 5701-5712.

McKenzie, D.; Raymond, C.L.; Kellogg, L.-K.B.; Norheim, R.A; Andreu, A.G.; Bayard, A.C.; Kopper, K.E.; Elman. E. 2007. Mapping fuels at multiple scales: landscape application of the Fuel Characteristic Classification System. Canadian Journal of Forest Research. 37:2421-2437.

McQuilling, A. M. & Adams, P. J. Semi-empirical process-based models for ammonia emissions from beef, swine, and poultry operations in the United States. Atmos. Environ. 120, 127–136 (2015).

NCAR, 2016. FIRE EMISSION FACTORS AND EMISSION INVENTORIES, FINN Data. downloaded 2014 SAPRC99 version from https://www.fs.fed.us/pnw/fera/research/smoke/consume/index.shtml.

NYSERDA, 2012; "Environmental, Energy Market, and Health Characterization of Wood-Fired Hydronic Heater Technologies, Final Report". New York State Energy Research and Development Authority (NYSERDA). Available from: http://www.nyserda.ny.gov/Publications/Case-Studies/-/media/Files/Publications/Research/Environmental/Wood-Fired-Hydronic-Heater-Tech.ashx.

Ottmar, R.D.; Sandberg, D.V.; Riccardi, C.L.; Prichard, S.J. 2007. An Overview of the Fuel Characteristic Classification System – Quantifying, Classifying, and Creating Fuelbeds for Resource Planning. Canadian Journal of Forest Research. 37(12): 2383-2393. FCCS is available at: http://www.fs.fed.us/pnw/fera/fccs/index.shtml

Pinder, R., Strader, R., Davidson, C. & Adams, P. A temporally and spatially resolved ammonia emission inventory for dairy cows in the United States. Atmos. Environ. 38.23, 3747–3756 (2004). 2. Pinder, R., Pekney, N., Davidson, C. & Adams, P. A process-based model of ammonia emissions from dairy cows: improved temporal and spatial resolution. Atmos. Environ. 38.9, 1357–1365 (2004).

Pinder, R., Pekney, N., Davidson, C. & Adams, P. A process-based model of ammonia emissions from dairy cows: improved temporal and spatial resolution. Atmos. Environ. 38.9, 1357–1365 (2004).

Pouliot, G. and J. Bash, 2015. Updates to Version 3.61 of the Biogenic Emission Inventory System (BEIS). Presented at Air and Waste Management Association conference, Raleigh, NC, 2015.

Pouliot, G., H. Simon, P. Bhave, D. Tong, D. Mobley, T. Pace, and T. Pierce . (2010) "Assessing the Anthropogenic Fugitive Dust Emission Inventory and Temporal Allocation Using an Updated Speciation of Particulate Matter." International Emission Inventory Conference, San Antonio, TX. Available at http://www.epa.gov/ttn/chief/conference/ei19/session9/pouliot.pdf

Raffuse, S., D. Sullivan, L. Chinkin, S. Larkin, R. Solomon, A. Soja, 2007. Integration of Satellite-Detected and Incident Command Reported Wildfire Information into BlueSky, June 27, 2007. Available at: https://www.airfire.org/smartfire.

Reichle, L.,R. Cook, C. Yanca, D. Sonntag, 2015. "Development of organic gas exhaust speciation profiles for nonroad spark-ignition and compression-ignition engines and equipment", Journal of the Air & Waste Management Association, 65:10, 1185-1193, DOI: 10.1080/10962247.2015.1020118.

Sullivan D.C., Raffuse S.M., Pryden D.A., Craig K.J., Reid S.B., Wheeler N.J.M., Chinkin L.R., Larkin N.K., Solomon R., and Strand T. (2008) Development and applications of systems for modeling emissions and smoke from fires: the BlueSky smoke modeling framework and SMARTFIRE: 17th International Emissions Inventory Conference, Portland, OR, June 2-5.

Wang, Y., P. Hopke, O. V. Rattigan, X. Xia, D. C. Chalupa, M. J. Utell. (2011) "Characterization of Residential Wood Combustion Particles Using the Two-Wavelength Aethalometer", Environ. Sci. Technol., 45 (17), pp 7387–7393

Wiedinmyer, C., S.K. Akagi, R.J. Yokelson, L.K. Emmons, J.A. Al-Saadi[3], J. J. Orlando[1], and A. J. Soja. (2011) "The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning", Geosci. Model Dev., 4, 625-641. http://www.geosci-model-dev.net/4/625/2011/ doi:10.5194/gmd-4-625-2011

Yarwood, G., S. Rao, M. Yocke, and G. Whitten, 2005: Updates to the Carbon Bond Chemical Mechanism: CB05. Final Report to the US EPA, RT-0400675. Available at http://www.camx.com/publ/pdfs/CB05_Final_Report_120805.pdf

4.0 CMAQ Air Quality Model Estimates

4.1 Introduction to the CMAQ Modeling Platform

The Clean Air Act (CAA) provides a mandate to assess and manage air pollution levels to protect human health and the environment. EPA has established National Ambient Air Quality Standards (NAAQS), requiring the development of effective emissions control strategies for such pollutants as ozone and particulate matter. Air quality models are used to develop these emission control strategies to achieve the objectives of the CAA.

Historically, air quality models have addressed individual pollutant issues separately. However, many of the same precursor chemicals are involved in both ozone and aerosol (particulate matter) chemistry; therefore, the chemical transformation pathways are dependent. Thus, modeled abatement strategies of pollutant precursors, such as volatile organic compounds (VOC) and NOx to reduce ozone levels, may exacerbate other air pollutants such as particulate matter. To meet the need to address the complex relationships between pollutants, EPA developed the Community Multiscale Air Quality (CMAQ) modeling system[13.] The primary goals for CMAQ are to:

- Improve the environmental management community's ability to evaluate the impact of air quality management practices for multiple pollutants at multiple scales.
- Improve the scientist's ability to better probe, understand, and simulate chemical and physical interactions in the atmosphere.

The CMAQ modeling system brings together key physical and chemical functions associated with the dispersion and transformations of air pollution at various scales. It was designed to approach air quality as a whole by including state-of-the-science capabilities for modeling multiple air quality issues, including tropospheric ozone, fine particles, toxics, acid deposition, and visibility degradation. CMAQ relies on emission estimates from various sources, including the U.S. EPA Office of Air Quality Planning and Standards' current emission inventories, observed emission from major utility stacks, and model estimates of natural emissions from biogenic and agricultural sources. CMAQ also relies on meteorological predictions that include assimilation of meteorological observations as constraints. Emissions and meteorology data are fed into CMAQ and run through various algorithms that simulate the physical and chemical processes in the atmosphere to provide estimated concentrations of the pollutants. Traditionally, the model has been used to predict air quality across a regional or national domain and then to simulate the effects of various changes in emission levels for policymaking purposes. For health studies, the model can also be used to provide supplemental information about air quality in areas where no monitors exist.

CMAQ was also designed to have multi-scale capabilities so that separate models were not needed for

[13] Byun, D.W., and K. L. Schere, 2006: Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. Applied Mechanics Reviews, Volume 59, Number 2 (March 2006), pp. 51-77.

urban and regional scale air quality modeling. The CMAQ simulation performed for this 2015 assessment used a single domain that covers the entire continental U.S. (CONUS) and large portions of Canada and Mexico using 12 km by 12 km horizontal grid spacing. Currently, 12 km x 12 km resolution is sufficient as the highest resolution for most regional-scale air quality model applications and assessments.[14] With the temporal flexibility of the model, simulations can be performed to evaluate longer term (annual to multi-year) pollutant climatologies as well as short-term (weeks to months) transport from localized sources. By making CMAQ a modeling system that addresses multiple pollutants and different temporal and spatial scales, CMAQ has a "one atmosphere" perspective that combines the efforts of the scientific community. Improvements will be made to the CMAQ modeling system as the scientific community further develops the state-of-the-science.

For more information on CMAQ, go to https://www.epa.gov/cmaq or http://www.cmascenter.org.

4.1.1 Advantages and Limitations of the CMAQ Air Quality Model

An advantage of using the CMAQ model output for characterizing air quality for use in comparing with health outcomes is that it provides a complete spatial and temporal coverage across the U.S. CMAQ is a three-dimensional Eulerian photochemical air quality model that simulates the numerous physical and chemical processes involved in the formation, transport, and destruction of ozone, particulate matter and air toxics for given input sets of initial and boundary conditions, meteorological conditions and emissions. The CMAQ model includes state-of-the-science capabilities for conducting urban to regional scale simulations of multiple air quality issues, including tropospheric ozone, fine particles, toxics, acid deposition and visibility degradation. However, CMAQ is resource intensive, requiring significant data inputs and computing resources.

An uncertainty of using the CMAQ model includes structural uncertainties, representation of physical and chemical processes in the model. These consist of: choice of chemical mechanism used to characterize reactions in the atmosphere, choice of land surface model and choice of planetary boundary layer. Another uncertainty in the CMAQ model is based on parametric uncertainties, which includes uncertainties in the model inputs: hourly meteorological fields, hourly 3-D gridded emissions, initial conditions, and boundary conditions. Uncertainties due to initial conditions are minimized by using a 10-day ramp-up period from which model results are not used in the aggregation and analysis of model outputs. Evaluations of models against observed pollutant concentrations build confidence that the model performs with reasonable accuracy despite the uncertainties listed above. A detailed model evaluation for ozone and $PM_{2.5}$ species provided in Section 4.3 shows generally acceptable model performance which is equivalent or better than typical state-of-the-science regional modeling simulations as summarized in Simon et al., 2012[15].

4.2 CMAQ Model Version, Inputs and Configuration

This section describes the air quality modeling platform used for the 2015 CMAQ simulation. A modeling platform is a structured system of connected modeling-related tools and data that provide a consistent and

[14] U.S. EPA (2014), Draft Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5, and Regional Haze, pp 214. https://www3.epa.gov/ttn/scram/guidance/guide/Draft_O3-PM-RH_Modeling_Guidance-2014.pdf .
[15] Simon, H., Baker, K.R., and Phillips, S. (2012) Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012. *Atmospheric Environment* **61**, 124-139.

transparent basis for assessing the air quality response to changes in emissions and/or meteorology. A platform typically consists of a specific air quality model, emissions estimates, a set of meteorological inputs, and estimates of boundary conditions representing pollutant transport from source areas outside the region modeled. We used the CMAQ modeling system as part of the 2015 Platform to provide a national scale air quality modeling analysis. The CMAQ model simulates the multiple physical and chemical processes involved in the formation, transport, and destruction of ozone and fine particulate matter ($PM_{2.5}$).

This section provides a description of each of the main components of the 2015 CMAQ simulation along with the results of a model performance evaluation in which the 2015 model predictions are compared to corresponding measured ambient concentrations.

4.2.1 CMAQ Model Version

CMAQ is a non-proprietary computer model that simulates the formation and fate of photochemical oxidants, including $PM_{2.5}$ and ozone, for given input sets of meteorological conditions and emissions. As mentioned previously, CMAQ includes numerous science modules that simulate the emission, production, decay, deposition and transport of organic and inorganic gas-phase and particle pollutants in the atmosphere. This 2015 analysis employed CMAQ version 5.2.1.[16] The 2015 CMAQ run included bi-directional ammonia (NH_3) air-surface exchange, CB6r3 chemical mechanism, AERO6 aerosol module with non-volatile Primary Organic Aerosol (POA). The CMAQ community model versions 5.0.2 and 5.1 were most recently peer-reviewed in September of 2015 for the U.S. EPA.[17]

4.2.2 Model Domain and Grid Resolution

The CMAQ modeling analyses were performed for a domain covering the continental United States, as shown in Figure 4-1. This single domain covers the entire continental U.S. (CONUS) and large portions of Canada and Mexico using 12 km by 12 km horizontal grid spacing. The 2015 simulation used a Lambert Conformal map projection centered at (-97, 40) with true latitudes at 33 and 45 degrees north. The 12 km CMAQ domain consisted of 396 by 246 grid cells and 35 vertical layers. Table 4-1 provides some basic geographic information regarding the 12 km CMAQ domain. The model extends vertically from the surface to 50 millibars (approximately 17,600 meters) using a sigma-pressure coordinate system. Table 4-2 shows the vertical layer structure used in the 2015 simulation. Air quality conditions at the outer boundary of the 12-km domain were taken from the northern hemispheric CMAQ model (discussed in Section 4.2.4).

---

[16] CMAQ version 5.2.1: doi:10.5281; https://zenodo.org/record/1212601. Model code for CMAQ v5.2.1 is also available from the Community Modeling and Analysis System (CMAS) at: http://www.cmascenter.org.

[17] Moran, M.D., Astitha, M., Barsanti, K.C., Brown, N.J., Kaduwela, A., McKeen, S.A., Pickering, K.E. (September 28, 2015). Final Report: Fifth Peer Review of the CMAQ Model, https://www.epa.gov/sites/production/files/2016-11/documents/cmaq_fifth_review_final_report_2015.pdf. This peer review was focused on CMAQ v5.0.2, which was released in May, 2014, as well as CMAQ v5.1, which was released in October 2015. It is available from the Community Modeling and Analysis System (CMAS) as well as previous peer-review reports at: http://www.cmascenter.org.

Table 4-1. Geographic Information for 2015 12-km Modeling Domain

National 12 km CMAQ Modeling Configuration	
Map Projection	Lambert Conformal Projection
Grid Resolution	12 km
Coordinate Center	97 W, 40 N
True Latitudes	33 and 45 N
Dimensions	396 x 246 x 35
Vertical Extent	35 Layers: Surface to 50 mb level (see Table 4-2)

Table 4-2. Vertical layer structure for 2015 CMAQ simulation (heights are layer top).

Vertical Layers	Sigma P	Pressure (mb)	Approximate Height (m)
35	0.0000	50.00	17,556
34	0.0500	97.50	14,780
33	0.1000	145.00	12,822
32	0.1500	192.50	11,282
31	0.2000	240.00	10,002
30	0.2500	287.50	8,901
29	0.3000	335.00	7,932
28	0.3500	382.50	7,064
27	0.4000	430.00	6,275
26	0.4500	477.50	5,553
25	0.5000	525.00	4,885
24	0.5500	572.50	4,264
23	0.6000	620.00	3,683
22	0.6500	667.50	3,136
21	0.7000	715.00	2,619
20	0.7400	753.00	2,226
19	0.7700	781.50	1,941
18	0.8000	810.00	1,665
17	0.8200	829.00	1,485
16	0.8400	848.00	1,308
15	0.8600	867.00	1,134
14	0.8800	886.00	964
13	0.9000	905.00	797
12	0.9100	914.50	714
11	0.9200	924.00	632
10	0.9300	933.50	551
9	0.9400	943.00	470

Vertical Layers	Sigma P	Pressure (mb)	Approximate Height (m)
8	0.9500	952.50	390
7	0.9600	962.00	311
6	0.9700	971.50	232
5	0.9800	981.00	154
4	0.9850	985.75	115
3	0.9900	990.50	77
2	0.9950	995.25	38
1	0.9975	997.63	19
0	1.0000	1000.00	0

Figure 4-1. Map of the 2015 CMAQ Modeling Domain. The purple box denotes the 12-km national modeling domain.

4.2.3 Modeling Period / Ozone Episodes

The 12-km CMAQ modeling domain was modeled for the entire year of 2015. The annual simulation included a spin-up period, comprised of 10 days before the beginning of the simulation, to mitigate the effects of initial concentrations. All 365 model days were used in the annual average levels of $PM_{2.5}$. For the 8-hour ozone, we used modeling results from the period between May 1 and September 30. This 153-day period generally conforms to the ozone season across most parts of the U.S. and contains the majority of days that observed high ozone concentrations.

4.2.4 Model Inputs: Emissions, Meteorology and Boundary Conditions

2015 Emissions: The emissions inventories used in the 2015 air quality modeling are described in Section 3, above.

Meteorological Input Data: The gridded meteorological data for the entire year of 2015 at the 12 km continental United States scale domain was derived from the publicly available version 3.8[18] of the Weather Research and Forecasting Model (WRF), Advanced Research WRF (ARW) core.[19] The WRF Model is a state-of-the-science mesoscale numerical weather prediction system developed for both operational forecasting and atmospheric research applications (http://wrf-model.org). The 2015 WRF meteorology simulated for 2015 with 2011 National Land Cover Database (NLCD)[20] and based on blended 3-hourly reanalysis fields (combination of 6-hour *(*Meteorological Assimilation Data Ingest System*)* MADIS[21] data and intermediate North American Mesoscale Model[22] (NAM) 3-hour forecast) organized into 12km NAM Data Assimilation System (NDAS) fields up to 50 hPa. The WRF simulation included the physics options of the Pleim-Xiu land surface model (LSM) with NLCD woody wetlands lad use category recognized, Asymmetric Convective Model version 2 planetary boundary layer (PBL) scheme, Morrison double moment microphysics, Kain- Fritsch cumulus parameterization scheme utilizing the moisture-advection trigger[23] and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme.[24] In addition, the Group for High Resolution Sea Surface Temperatures (GHRSST)[25,26] 1km SST data was used for SST information to provide more resolved information compared to the more coarse data in the NAM analysis.

Initial and Boundary Conditions: The lateral boundary and initial species concentrations were provided by a northern hemispheric application of a CMAQ modeling platform to the year 2015. The hemispheric-scale platform uses a polar stereographic projection at 108 km resolution to completely and continuously cover the northern hemisphere for 2015 with meteorology, emissions, and atmospheric processing of pollutants. Meteorology is provided by Weather Research and Forecasting model (WRF v3.8) using 44 non-hydrostatic sigma-pressure layers between the surface and 50 hPa (~20 km asl). Emissions were provided by the emissions modeling platform (v7.1) combining EDGAR-HTAP (v2)[27], Chinese emissions

---

[18] Version 3.6.1 was the current version of WRF at the time the 2013 meteorological model simulation was performed.
[19] Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X., Wang, W., Powers, J.G., 2008. A Description of the Advanced Research WRF Version 3.
[20] National Land Cover Database 2011, http://www.mrlc.gov/nlcd2011.php
[21] Meteorological Assimilation Data Ingest System. http://madis.noaa.gov/.
[22] North American Model Analysis-Only, http://nomads.ncdc.noaa.gov/data.php; download from ftp://nomads.ncdc.noaa.gov/NAM/analysis_only/.
[23] Ma, L-M. and Tan, Z-M, 2009. Improving the behavior of the Cumulus Parameterization for Tropical Cyclone Prediction: Convection Trigger. Atmospheric Research 92 Issue 2, 190-211. http://www.sciencedirect.com/science/article/pii/S0169809508002585
[24] Gilliam, R.C., Pleim, J.E., 2010. Performance Assessment of New Land Surface and Planetary Boundary Layer Physics in the WRF-ARW. Journal of Applied Meteorology and Climatology 49, 760-774.
[25] Stammer, D., F.J. Wentz, and C.L. Gentemann, 2003, Validation of Microwave Sea Surface Temperature Measurements for Climate Purposes, J. Climate, 16, 73-87.
[26] Global High Resolution SST (GHRSST) analysis, https://www.ghrsst.org/.
[27] Janssens-Maenhout, G., Dentener, F., Van Aardenne, J., Monni, S., Pagliari, V., Orlandini, L., Klimont, Z., Kurokawa, J., Akimoto, H., Ohara, T., others, 2012. EDGAR-HTAP: a harmonized gridded air pollution emission dataset based on national inventories. European Commission Publications Office, Ispra (Italy). JRC68434, EUR report No EUR 25, 299–2012.

provided by Tsinghua University, and the EPA 2016 national modeling platform (alpha, 2016fe), climatological lightning, and natural emissions as processed by GEOS-CHEM[28] (soil NOx and biogenic VOC). The atmospheric processing (transformation and fate) was simulated by CMAQ (v5.2.1, doi:10.5281/zenodo.1212601) using the Carbon Bond (cb6r3) with linearized halogen chemistry and the aerosol model with non-volatile primary organic carbon (AE6nvPOA). The CMAQ model also included the on-line windblown dust emission sources (excluding agricultural land), which are not always included in the regional platform but are important for large-scale transport of dust. The simulation uses 8-months spin-up from 2015-05-01 0Z to 2015-12-22 0Z as a surrogate for the 2014 spin-up year, for which we did not have hemispheric emissions or meteorology. Evaluation against ozonesondes and CASTNet ozone monitors show best performance in summer for the hemispheric platform.

4.3 CMAQ Model Performance Evaluation

An operational model performance evaluation for ozone and $PM_{2.5}$ and its related speciated components was conducted for the 2015 simulation using state/local monitoring sites data in order to estimate the ability of the CMAQ modeling system to replicate the 2015 base year concentrations for the 12 km continental U.S. domain.

There are various statistical metrics available and used by the science community for model performance evaluation. For a robust evaluation, the principal evaluation statistics used to evaluate CMAQ performance were two bias metrics, mean bias and normalized mean bias; and two error metrics, mean error and normalized mean error.

Mean bias (MB) is used as average of the difference (predicted – observed) divided by the total number of replicates (*n*). Mean bias is defined as:

MB = $\frac{1}{n}\sum_{1}^{n}(P - O)$, where P = predicted and O = observed concentrations.

Mean error (ME) calculates the absolute value of the difference (predicted - observed) divided by the total number of replicates (*n*). Mean error is defined as:

ME = $\frac{1}{n}\sum_{1}^{n}|P - O|$

Normalized mean bias (NMB) is used as a normalization to facilitate a range of concentration magnitudes. This statistic averages the difference (model - observed) over the sum of observed values. NMB is a useful model performance indicator because it avoids overinflating the observed range of values, especially at low concentrations. Normalized mean bias is defined as:

[28] Yantosca, B., 2004. GEOS-CHEMv7-01-02 User's Guide, Atmospheric Chemistry Modeling Group, Harvard University, Cambridge, MA, October 15, 2004.

$$\text{NMB} = \frac{\sum_{1}^{n}(P - O)}{\sum_{1}^{n}(O)} * 100$$, where P = predicted concentrations and O = observed

Normalized mean error (NME) is also similar to NMB, where the performance statistic is used as a normalization of the mean error. NME calculates the absolute value of the difference (model - observed) over the sum of observed values. Normalized mean error is defined as

$$\text{NME} = \frac{\sum_{1}^{n}|P - O|}{\sum_{1}^{n}(O)} * 100$$

The performance statistics were calculated using predicted and observed data that were paired in time and space on an 8-hour basis. Statistics were generated for each of the nine National Oceanic and Atmospheric Administration (NOAA) climate regions[29] of the 12-km U.S. modeling domain (Figure 4-2). The regions include the Northeast, Ohio Valley, Upper Midwest, Southeast, South, Southwest, Northern Rockies, Northwest and West[30,31] as were originally identified in Karl and Koss (1984)[32].

[29] NOAA, National Centers for Environmental Information scientists have identified nine climatically consistent regions within the contiguous U.S., http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php.
[30] The nine climate regions are defined by States where: Northeast includes CT, DE, ME, MA, MD, NH, NJ, NY, PA, RI, and VT; Ohio Valley includes IL, IN, KY, MO, OH, TN, and WV; Upper Midwest includes IA, MI, MN, and WI; Southeast includes AL, FL, GA, NC, SC, and VA; South includes AR, KS, LA, MS, OK, and TX; Southwest includes AZ, CO, NM, and UT; Northern Rockies includes MT, NE, ND, SD, WY; Northwest includes ID, OR, and WA; and West includes CA and NV.
[31] Note most monitoring sites in the West region are located in California (see Figure 4-2), therefore statistics for the West will be mostly representative of California ozone air quality.
[32] Karl, T. R. and Koss, W. J., 1984: "Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983." Historical Climatology Series 4-3, National Climatic Data Center, Asheville, NC, 38 pp.

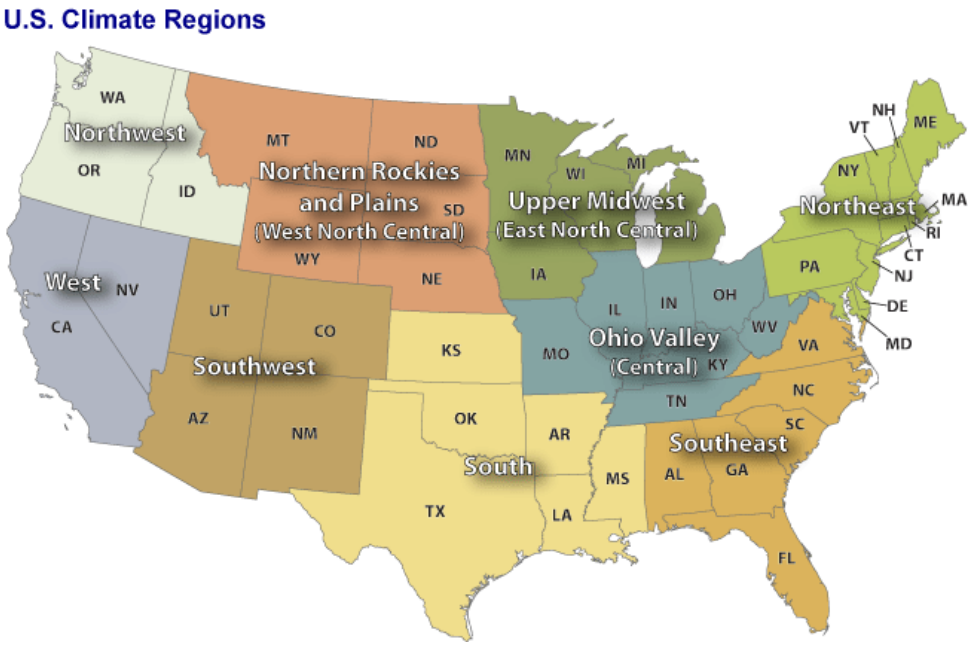


Figure 4-2. NOAA Nine Climate Regions (source: http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php#references)

In addition to the performance statistics, regional maps which show the MB, ME, NMB, and NME were prepared for the ozone season, May through September, at individual monitoring sites as well as on an annual basis for $PM_{2.5}$ and its component species.

Evaluation for 8-hour Daily Maximum Ozone: The operational model performance evaluation for eight-hour daily maximum ozone was conducted using the statistics defined above. Ozone measurements for 2015 in the continental U.S. were included in the evaluation and were taken from the 2015 State/local monitoring site data in the EPA Air Quality System (AQS) and the Clean Air Status and Trends Network (CASTNet).
The 8-hour ozone model performance bias and error statistics for each of the nine NOAA climate regions and each season are provided in Table 4-4. Seasons were defined as: winter (December-January-February), spring (March-April-May), summer (June, July, August), and fall (September-October-November). In some instances, observational data were excluded from the analysis and model evaluation based on a completeness criterion of 75 percent. Spatial plots of the MB, ME, NMB and NME for individual monitors are shown in Figures 4-3 through 4-6, respectively. The statistics shown in these two figures were calculated over the ozone season, May through September, using data pairs on days with observed 8-hour ozone of greater than or equal to 60 ppb.

In general, the model performance statistics indicate that the 8-hour daily maximum ozone concentrations predicted by the 2015 CMAQ simulation closely reflect the corresponding 8-hour observed ozone concentrations in space and time in each subregion of the 12-km modeling domain. As indicated by the statistics in Table 4-4, bias and error for 8-hour daily maximum ozone are relatively low in each subregion, not only in the summer when concentrations are highest, but also during other times of the year. Generally, 8-hour ozone at the AQS sites in the summer and fall is over predicted with the greatest over prediction in the South, Southeast and Ohio Valley (NMB ranging between 5 to 20 percent). Likewise, 8-hour ozone at the CASTNet sites in the summer and fall is typically over predicted except in the West,

Southwest and Northern Rockies where the bias shows an under prediction (NMB ranging from -1% to -20%). 8-hour ozone is under predicted at AQS and CASTNet sites in all of the climate regions in the winter and spring (with NMBs less than approximately 20 percent in each subregion).

Model bias at individual sites during the ozone season is similar to that seen on a subregional basis for the summer. Figure 4-2 shows the mean bias for 8-hour daily maximum ozone greater than 60 ppb is generally ±10 ppb across the AQS and CASTNet sites. Likewise, the information in Figure 4-4 indicates that the bias for days with observed 8-hour daily maximum ozone greater than 60 ppb is within ± 20 percent at the vast majority of monitoring sites across the U.S. domain. Model error, as seen from Figures 4-4 and 4-6, is generally 2 to 10 ppb and 20 percent or less at most of the sites across the U.S. modeling domain. Somewhat greater error is evident at sites in several areas most notably in the West, Northern Rockies, Northeast, Upper Midwest, Southeast, along portions of the Gulf Coast, and Great Lakes coastline.

Table 4-4. Summary of CMAQ 2015 8-Hour Daily Maximum Ozone Model Performance Statistics by NOAA climate region, by Season and Monitoring Network.

Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
Northeast	AQS	Winter	11,096	-6.2	7.8	-19.9	24.8
		Spring	15,455	-5.4	7.5	-12.0	16.7
		Summer	16,586	6.9	8.3	15.6	19.0
		Fall	13,816	2.8	5.2	7.7	14.7
	CASTNet	Winter	1,253	-6.5	8.3	-19.9	25.2
		Spring	1,297	-7.0	8.2	-15.2	17.8
		Summer	1,293	5.2	6.8	12.7	16.6
		Fall	1,332	2.0	4.7	5.7	13.2
Ohio Valley	AQS	Winter	4,031	-3.7	6.0	-12.6	20.2
		Spring	15,603	-2.2	6.3	-4.9	14.2
		Summer	19,303	8.6	9.6	19.6	22.0
		Fall	12,883	4.6	6.2	12.0	16.1
	CASTNet	Winter	1,539	-3.3	5.9	-10.3	18.2
		Spring	1,533	-4.3	6.6	-9.2	14.3
		Summer	1,573	6.8	8.1	15.8	18.9
		Fall	1,554	2.5	5.0	6.6	13.0
Upper Midwest	AQS	Winter	1,453	-7.4	8.8	-24.6	29.1
		Spring	6,926	-3.8	7.1	-8.6	16.0
		Summer	9,623	5.2	7.2	12.5	17.3
		Fall	5,990	3.8	5.3	10.3	14.4
	CASTNet	Winter	409	-9.1	10.0	-27.7	30.4

Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
		Spring	438	-7.1	8.6	-15.7	19.2
		Summer	438	3.2	5.9	7.9	14.7
		Fall	436	2.0	4.3	5.9	12.6
Southeast	AQS	Winter	7,129	-0.9	4.7	-2.7	13.3
		Spring	14,854	-0.9	5.7	-2.0	13.2
		Summer	16,160	7.4	8.2	18.2	20.3
		Fall	13,061	5.7	7.0	16.0	19.5
	CASTNet	Winter	910	-3.4	5.9	16.7	9.3
		Spring	981	-4.4	6.5	-9.6	14.1
		Summer	937	6.1	7.1	15.1	17.7
		Fall	966	3.4	5.6	9.2	15.4
South	AQS	Winter	11,126	-0.9	4.9	-2.9	15.2
		Spring	13,128	1.8	6.7	4.6	16.9
		Summer	13,014	7.0	8.4	17.0	20.6
		Fall	12,557	3.2	5.3	8.0	13.5
	CASTNet	Winter	479	-1.3	4.5	-3.8	12.9
		Spring	528	1.2	6.2	-2.8	14.6
		Summer	444	3.9	6.1	9.3	14.6
		Fall	527	2.4	4.4	6.1	11.1
Southwest	AQS	Winter	9,191	-2.1	5.4	-6.0	15.2
		Spring	10,835	-4.8	6.2	-9.2	11.8
		Summer	11,400	0.1	5.9	0.2	11.1
		Fall	11,022	2.0	4.8	4.8	11.4
	CASTNet	Winter	763	-6.0	6.9	-14.1	16.1
		Spring	770	-6.5	7.1	-12.0	13.2
		Summer	784	-0.8	5.2	-1.6	10.1
		Fall	802	-0.8	3.7	-1.7	8.3
Northern Rockies	AQS	Winter	4,672	-7.0	8.1	-19.8	23.1
		Spring	5,141	-3.9	6.7	-8.5	14.6
		Summer	5,070	2.3	6.1	5.0	13.0
		Fall	4,857	2.1	5.0	5.6	13.6
	CASTNet	Winter	746	-7.4	8.5	-20.1	23.0
		Spring	791	-5.7	7.0	-11.9	14.7

Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
		Summer	764	-0.3	5.0	-0.6	10.4
		Fall	773	0.2	4.7	0.4	11.9
Northwest	AQS	Winter	592	-3.3	6.2	-11.1	20.7
		Spring	1,242	-2.3	7.8	-5.6	18.8
		Summer	2,557	2.2	7.1	5.0	16.3
		Fall	1,145	3.5	6.3	10.3	18.7
	CASTNet	Winter	--	--	--	--	--
		Spring	--	--	--	--	--
		Summer	--	--	--	--	--
		Fall	--	--	--	--	--
West	AQS	Winter	13,524	0.6	5.6	1.8	17.0
		Spring	16,705	-5.5	7.3	-10.9	14.5
		Summer	17,998	2.1	7.9	4.1	15.7
		Fall	16,560	0.7	5.6	1.5	12.7
	CASTNet	Winter	518	-0.8	4.4	-2.2	11.7
		Spring	535	-7.6	8.0	-14.3	15.1
		Summer	529	-3.7	6.8	-6.5	11.9
		Fall	513	-2.2	4.8	-4.7	10.1

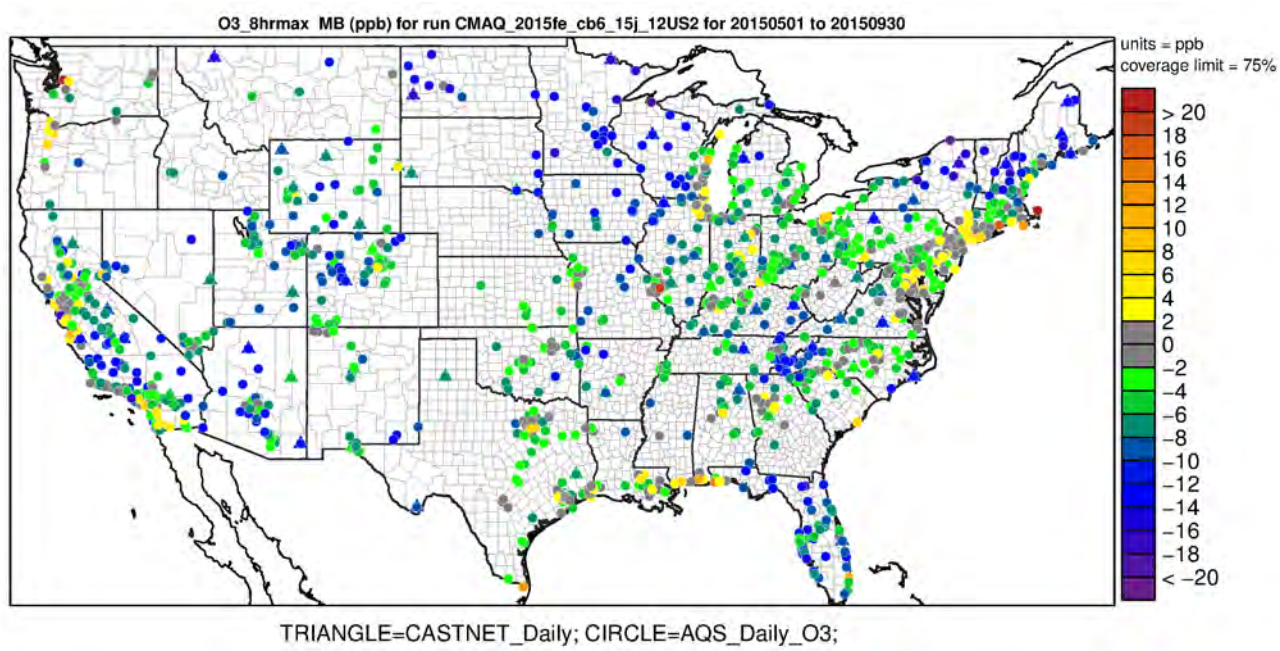


Figure 4-3. Mean Bias (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period

May-September 2015 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

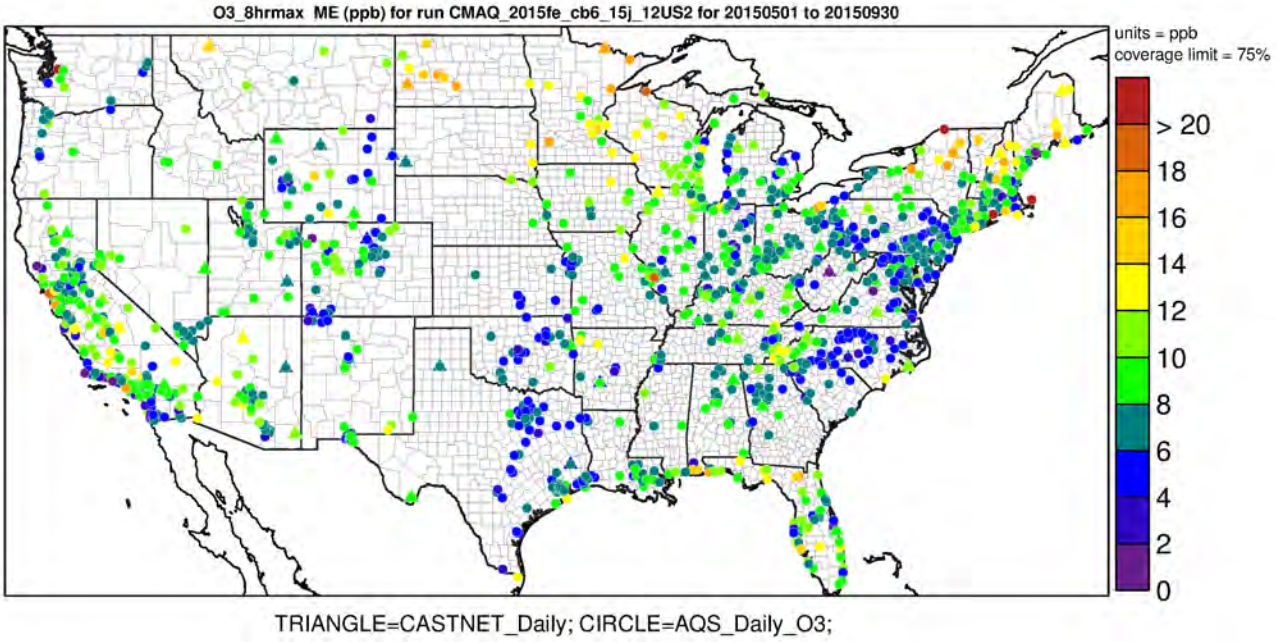


Figure 4-4. Mean Error (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2015 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

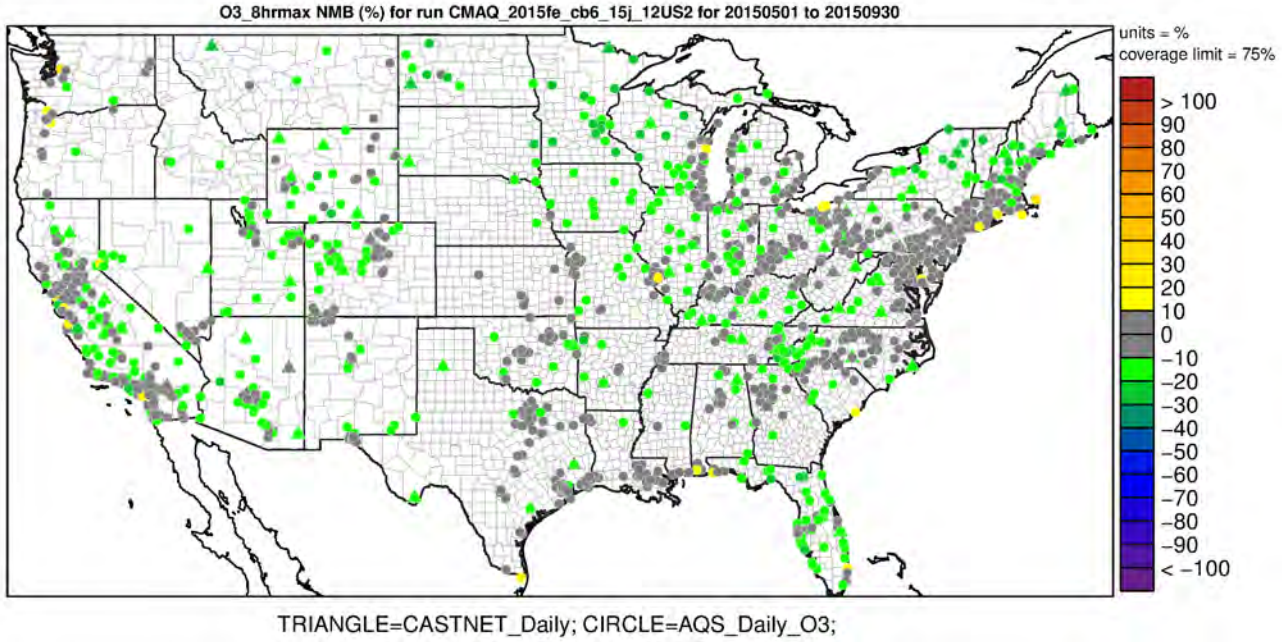


Figure 4-5. Normalized Mean Bias (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2015 at AQS and CASTNet monitoring sites in the continental U.S.

modeling domain.

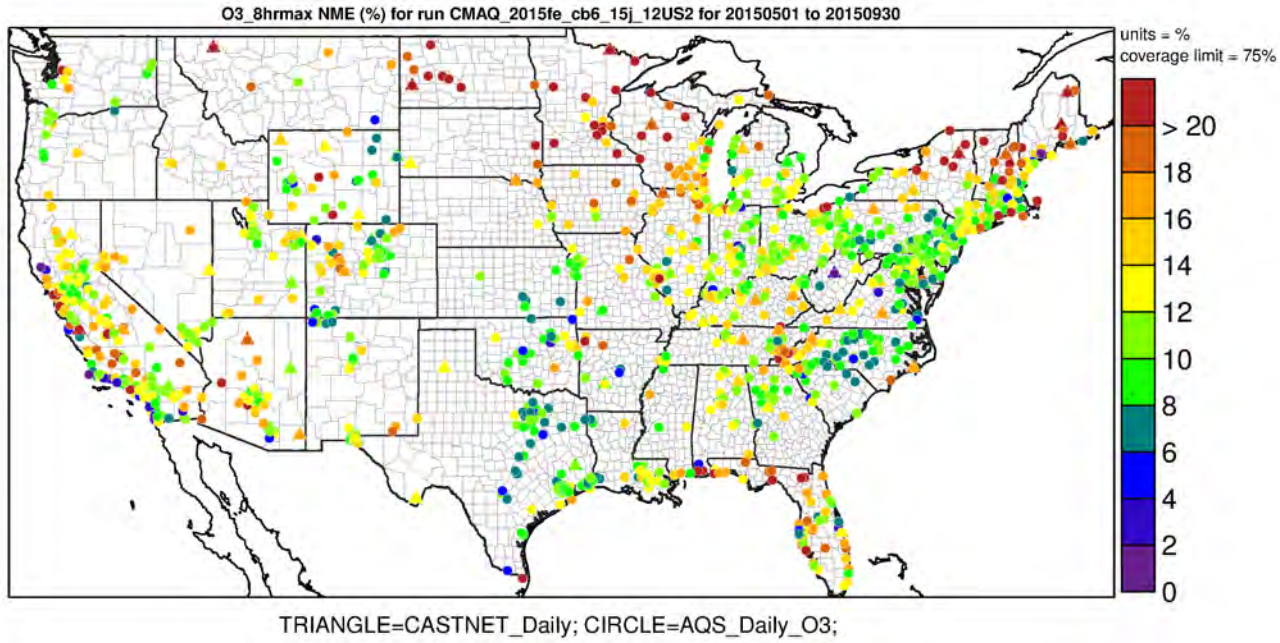


Figure 4-6. Normalized Mean Error (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2015 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

Evaluation for Annual $PM_{2.5}$ components: The PM evaluation focuses on $PM_{2.5}$ components including sulfate (SO_4), nitrate (NO_3), total nitrate ($TNO_3 = NO_3 + HNO_3$), ammonium (NH_4), elemental carbon (EC), and organic carbon (OC). The bias and error performance statistics were calculated on an annual basis for each of the nine NOAA climate subregions defined above (provided in Table 4-5). $PM_{2.5}$ measurements for 2015 were obtained from the following networks for model evaluation: Chemical Speciation Network (CSN, 24-hour average), Interagency Monitoring of Protected Visual Environments (IMPROVE, 24-hour average, and Clean Air Status and Trends Network (CASTNet, weekly average). For $PM_{2.5}$ species that are measured by more than one network, we calculated separate sets of statistics for each network by subregion. In addition to the tabular summaries of bias and error statistics, annual spatial maps which show the mean bias, mean error, normalized mean bias and normalized mean error by site for each $PM_{2.5}$ species are provided in Figures 4-7 through 4-30.

As indicated by the statistics in Table 4-5, annual average sulfate is consistently under predicted at CSN, IMPROVE, and CASTNet monitoring sites across the 12-km modeling domain (with MB values ranging from 0.0 to -0.4 μgm^{-3} and NMB values ranging from near negligible to -28 percent) except at CSN sites in the Southeast and Southwest as well as IMPROVE sites in the Upper Midwest, Southwest, Northern Rockies, Northwest and West. Sulfate performance shows moderate error in the eastern subregions (ranging from 23 to 38 percent) while Western subregions show slightly larger error (ranging from 36 to

85 percent). Figures 4-7 through 4-10, suggest spatial patterns vary by region. The model bias for most of the Northeast, Southeast, Central and Southwest states are within ±30 percent. The model bias appears to be slightly greater in the Northwest with over predictions up to 80 percent at individual monitors. Model error also shows a spatial trend by region, where much of the Eastern states are 10 to 40 percent, the Western and Central U.S. states are 30 to 80 percent.

Annual average nitrate is under predicted at the urban CSN monitoring sites in the Upper Midwest, Southwest, Northern Rockies, and West (NMB in the range of -5 to -45 percent), except in the Northeast, Ohio Valley, Southeast, South and Northwest where nitrate is over predicted (NMB in the range of 3 percent to greater than 100 percent). At IMPROVE rural sites, annual average nitrate is over predicted at all subregions, except in the Southwest and West where nitrate is under predicted by 14 to 37 percent, respectively. Model performance of total nitrate at sub-urban CASTNet monitoring sites shows an over prediction in the Northeast, Ohio Valley, South, and Southeast (NMB in the range of 5 to 23 percent), except in the Upper Midwest, Southwest, Northern Rockies and Western U.S. (NMB in the range of -1 to -37 percent). Model error for nitrate and total nitrate is somewhat greater for each of the nine NOAA climate subregions as compared to sulfate. Model bias at individual sites indicates mainly over prediction of greater than 20 percent at most monitoring sites in the Eastern half of the U.S. as indicated in Figure 4-13.nnThe exception to this is in the Southern Florida, Ohio Valley, Southwest, and Western U.S. of the modeling domain where there appears to be a greater number of sites with under prediction of nitrate of 10 to 80 percent. Model error for annual nitrate, as shown in Figures 4-12 and 4-15, is least at sites in portions of the Ohio Valley and Upper Midwest.

Annual average ammonium model performance as indicated in Table 4-5 has a tendency for the model to under predict across the CASTNet sites (ranging from -11 to -53 percent). Ammonium performance across the urban CSN sites shows an under prediction in four of the climate subregions (ranging from -1 to -56 percent), except in the Northeast, Ohio Valley, Northwest, Southeast, South, and Northern Rockies (over prediction of NMB 1 to 56 percent). The spatial variation of ammonium across the majority of individual monitoring sites in the Eastern U.S. shows bias within ±50 percent (Figures 4-19 and 4-21). A larger bias is seen in the Southeast and in the Northern Rockies, (over prediction bias on average 80 percent). The urban monitoring sites exhibit larger errors than at rural sites for ammonium.

Annual average elemental carbon is over predicted in all of the nine climate regions at urban and rural sites. There is not a large variation in error statistics from subregion to subregion or at urban versus rural sites.

Annual average organic carbon is over predicted across most subregions in rural IMPROVE areas (NMB ranging from 6 to 78 percent), except in the Southwest where the NMB is -6 percent. The model over predicted annual average organic carbon in all subregions at urban CSN sites except in the Ohio Valley, Southwest, Northern Rockies and Western U.S. (NMB ranges from -1 to -32 percent). Similar to elemental carbon, error model performance does not show a large variation from subregion to subregion or at urban versus rural sites.

Table 4-5. Summary of CMAQ 2015 Annual PM Species Model Performance Statistics by NOAA Climate region, by Monitoring Network.

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
Sulfate	CSN	Northeast	2,982	0.0	0.5	-0.4	34.9
		Ohio Valley	2,288	-0.1	0.6	-6.4	33.0
		Upper Midwest	1,238	0.0	0.4	-1.3	31.0
		Southeast	1,994	0.1	0.5	5.7	36.2
		South	1,168	-0.2	0.5	-12.2	35.7
		Southwest	996	0.0	0.3	2.0	44.4
		Northern Rockies	585	0.0	0.3	-1.9	36.5
		Northwest	--	--	--	--	--
		West	1,163	-0.1	0.4	-15.0	43.0
	IMPROVE	Northeast	1,755	0.0	0.3	-0.1	31.1
		Ohio Valley	861	-0.1	0.5	-8.5	31.0
		Upper Midwest	940	0.0	0.3	0.9	33.2
		Southeast	1,334	-0.1	0.4	-8.3	31.4
		South	1,130	-0.2	0.5	-18.1	38.3
		Southwest	3,694	0.0	0.3	7.2	49.1
		Northern Rockies	2,130	0.1	0.2	25.3	53.7
		Northwest	1,819	0.2	0.3	73.1	84.8
		West	2,444	0.1	0.3	10.2	53.6
	CASTNet	Northeast	916	-0.2	0.3	-17.7	23.0
		Ohio Valley	867	-0.3	0.4	-19.6	23.9
		Upper Midwest	293	-0.2	0.3	-15.5	25.7
		Southeast	630	-0.3	0.4	-23.5	27.6
		South	387	-0.4	0.5	-27.5	29.9
		Southwest	436	0.0	0.2	-2.1	35.8
		Northern Rockies	566	0.0	0.2	-0.6	36.3
		Northwest	1,819	0.4	0.5	66.9	85.4
		West	296	-0.1	0.3	-18.3	43.9
Nitrate	CSN	Northeast	2,983	0.3	0.7	22.5	56.4
		Ohio Valley	2,164	0.0	0.7	3.0	52.7
		Upper Midwest	1,121	-0.1	0.8	-4.8	47.6
		Southeast	1,999	0.4	0.6	82.0	>100.0
		South	1,168	0.0	0.5	5.5	70.0
		Southwest	996	-0.4	0.6	-44.8	71.8
		Northern Rockies	586	-0.1	0.5	-12.3	59.9
		Northwest	636	0.8	1.2	>100.0	>100.0
		West	1,163	-0.9	1.3	-36.1	51.7

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
	IMPROVE	Northeast	1,754	0.2	0.3	70.1	>100.0
		Ohio Valley	861	0.2	0.5	32.8	81.1
		Upper Midwest	940	0.1	0.4	10.7	58.4
		Southeast	1,333	0.2	0.4	63.4	>100.0
		South	1,130	0.1	0.4	18.6	89.3
		Southwest	3,684	-0.1	0.1	-37.3	84.5
		Northern Rockies	2,125	0.0	0.2	19.8	98.2
		Northwest	1,799	0.2	0.3	76.4	>100.0
		West	2,437	-0.1	0.3	-13.9	68.0
Total Nitrate (NO_3+HNO_3)	CASTNet	Northeast	916	0.3	0.5	22.9	37.0
		Ohio Valley	867	0.3	0.6	13.7	34.8
		Upper Midwest	293	0.0	0.5	-0.7	33.2
		Southeast	630	0.2	0.5	13.8	46.3
		South	387	0.1	0.5	4.5	35.1
		Southwest	436	-0.2	0.3	-21.6	39.6
		Northern Rockies	566	-0.1	0.2	-16.4	34.5
		Northwest	--	--	--	--	--
		West	296	-0.5	0.6	-37.3	42.7
Ammonium	CSN	Northeast	2,983	0.1	0.3	23.2	55.9
		Ohio Valley	2,164	0.0	0.4	3.1	49.5
		Upper Midwest	1,121	0.0	0.3	-0.7	45.1
		Southeast	1,993	0.1	0.2	25.9	68.0
		South	1,168	0.0	0.3	3.8	59.7
		Southwest	996	-0.2	0.3	-55.9	72.4
		Northern Rockies	586	0.0	0.2	1.1	58.6
		Northwest	636	0.1	0.3	56.0	>100.0
		West	1,163	-0.4	0.5	-47.4	62.0
	CASTNet	Northeast	916	-0.1	0.1	-11.0	25.0
		Ohio Valley	867	-0.1	0.2	-13.8	26.9
		Upper Midwest	293	-0.1	0.2	-21.1	32.2
		Southeast	630	-0.1	0.1	-15.2	31.3
		South	387	-0.1	0.2	-14.0	37.6
		Southwest	436	-0.1	0.1	-46.8	56.0
		Northern Rockies	566	-0.1	0.1	-35.6	53.8
		Northwest	--	--	--	--	--
		West	296	-0.1	0.2	-52.6	61.6
Elemental	CSN	Northeast	2,940	0.2	0.4	28.2	57.6

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
Carbon		Ohio Valley	2,150	0.1	0.3	18.4	49.6
		Upper Midwest	1,109	0.2	0.3	49.3	64.8
		Southeast	1,476	0.2	0.3	33.2	62.6
		South	1,155	0.2	0.2	34.9	53.5
		Southwest	742	0.3	0.4	51.5	74.1
		Northern Rockies	562	0.1	0.2	48.8	90.2
		Northwest	520	1.2	1.3	>100.0	>100.0
		West	996	0.1	0.3	20.1	48.5
	IMPROVE	Northeast	1,785	0.1	0.2	67.8	80.8
		Ohio Valley	886	0.1	0.1	38.5	64.5
		Upper Midwest	998	0.1	0.1	49.9	67.9
		Southeast	1,498	0.1	0.2	32.7	59.0
		South	1,125	0.1	0.2	73.2	100.0
		Southwest	3,673	0.0	0.1	24.9	65.4
		Northern Rockies	2,229	0.1	0.2	>100.0	>100.0
		Northwest	1,815	0.4	0.4	>100.0	>100.0
		West	2,446	0.1	0.2	74.6	>100.0
Organic Carbon	CSN	Northeast	2,940	0.4	1.3	15.4	47.2
		Ohio Valley	2,150	-0.1	1.0	-4.2	36.4
		Upper Midwest	1,109	0.3	1.0	10.7	43.4
		Southeast	1,476	0.0	1.2	0.7	42.0
		South	1,155	-0.2	1.0	34.7	79.6
		Southwest	742	0.0	1.0	-1.0	45.6
		Northern Rockies	562	-0.7	1.2	-31.6	55.1
		Northwest	520	1.8	2.7	56.9	86.9
		West	996	-0.2	1.0	-7.3	32.6
	IMPROVE	Northeast	1,786	0.5	0.8	48.8	70.6
		Ohio Valley	884	0.3	0.8	30.5	62.5
		Upper Midwest	998	0.4	0.7	34.8	60.8
		Southeast	1,498	0.3	0.9	21.6	63.3
		South	1,123	0.4	0.9	34.7	79.6
		Southwest	3,667	0.0	0.3	-6.2	49.1
		Northern Rockies	2,214	0.2	0.9	12.2	72.0
		Northwest	1,785	0.9	1.6	77.9	>100.0
		West	2,442	0.1	0.8	6.4	69.7

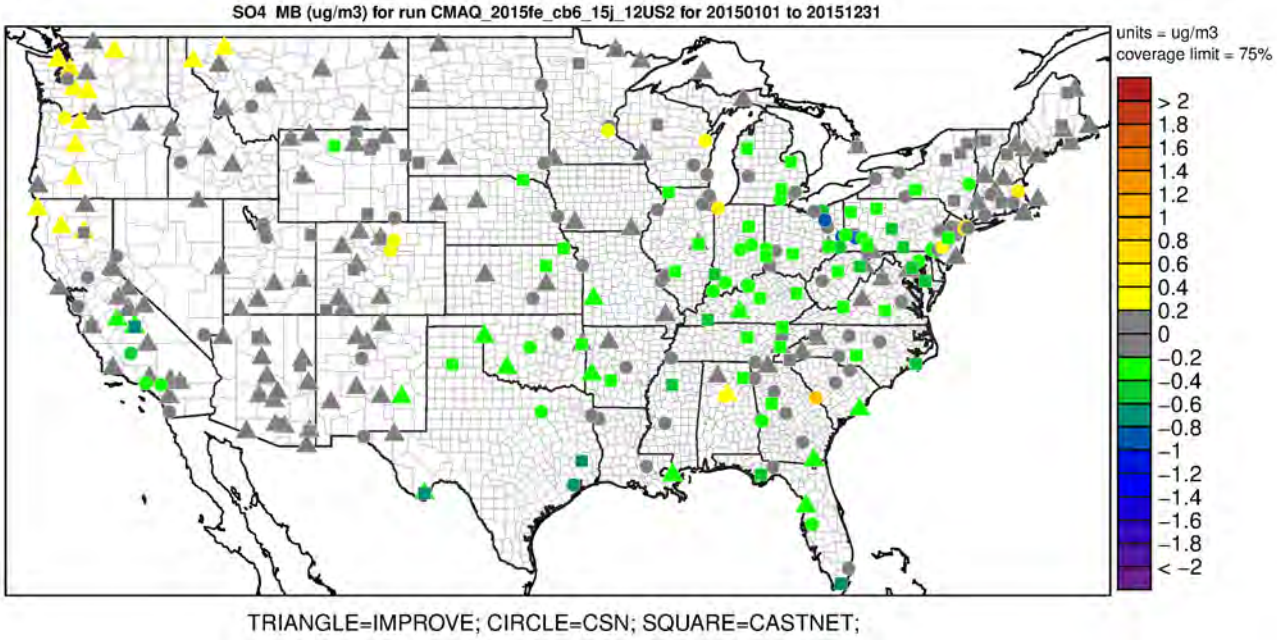


Figure 4-7. Mean Bias (µgm^{-3}) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

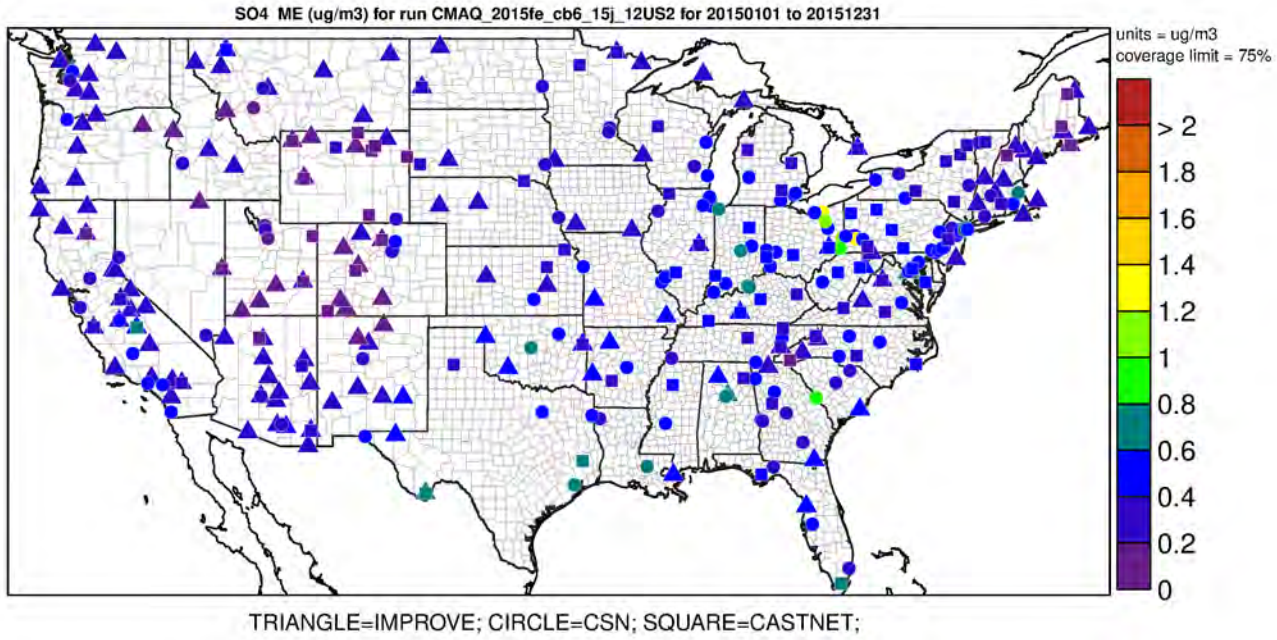


Figure 4-8. Mean Error (µgm^{-3}) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

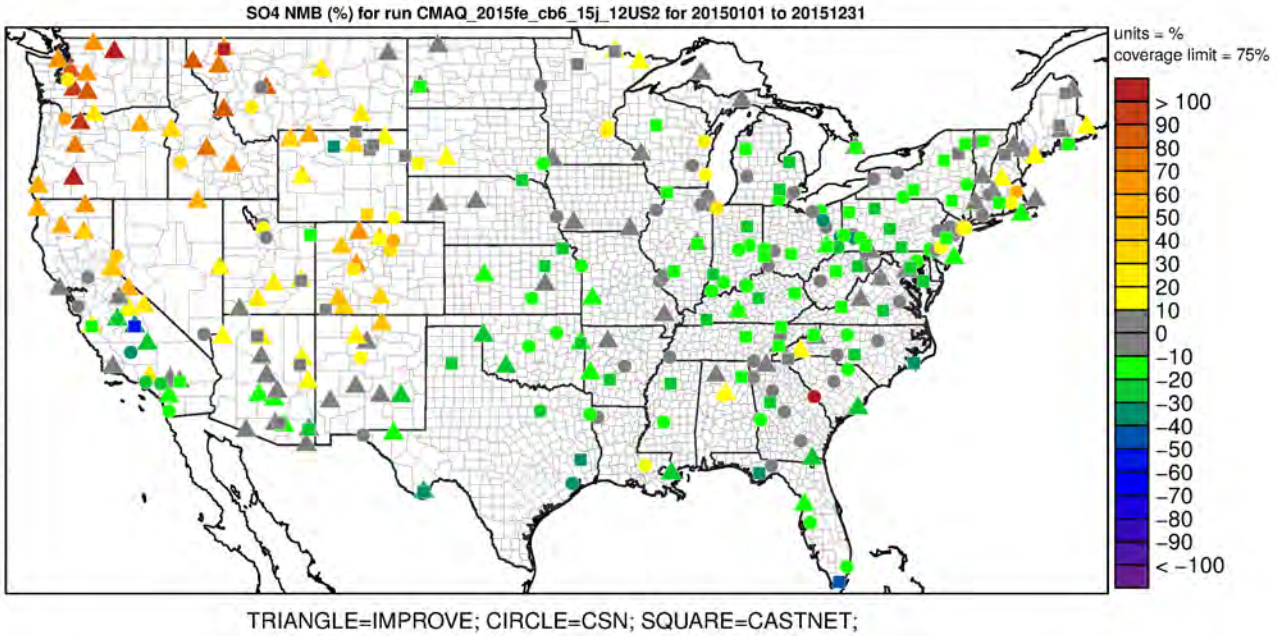


Figure 4-9. Normalized Mean Bias (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

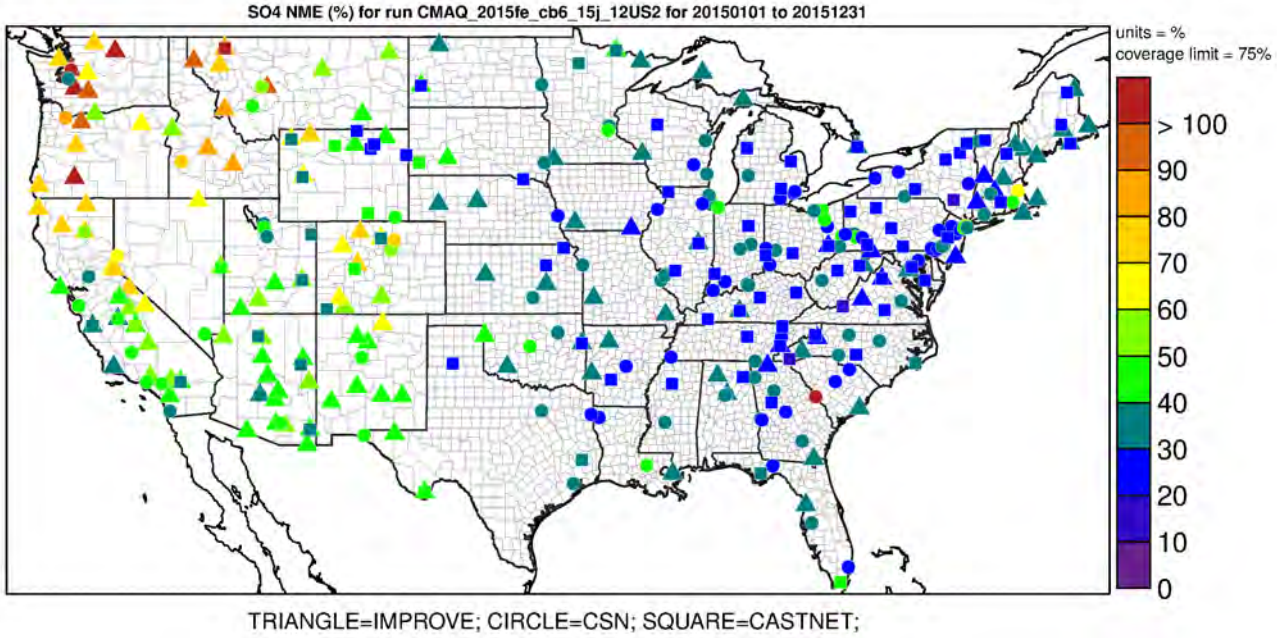


Figure 4-10. Normalized Mean Error (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

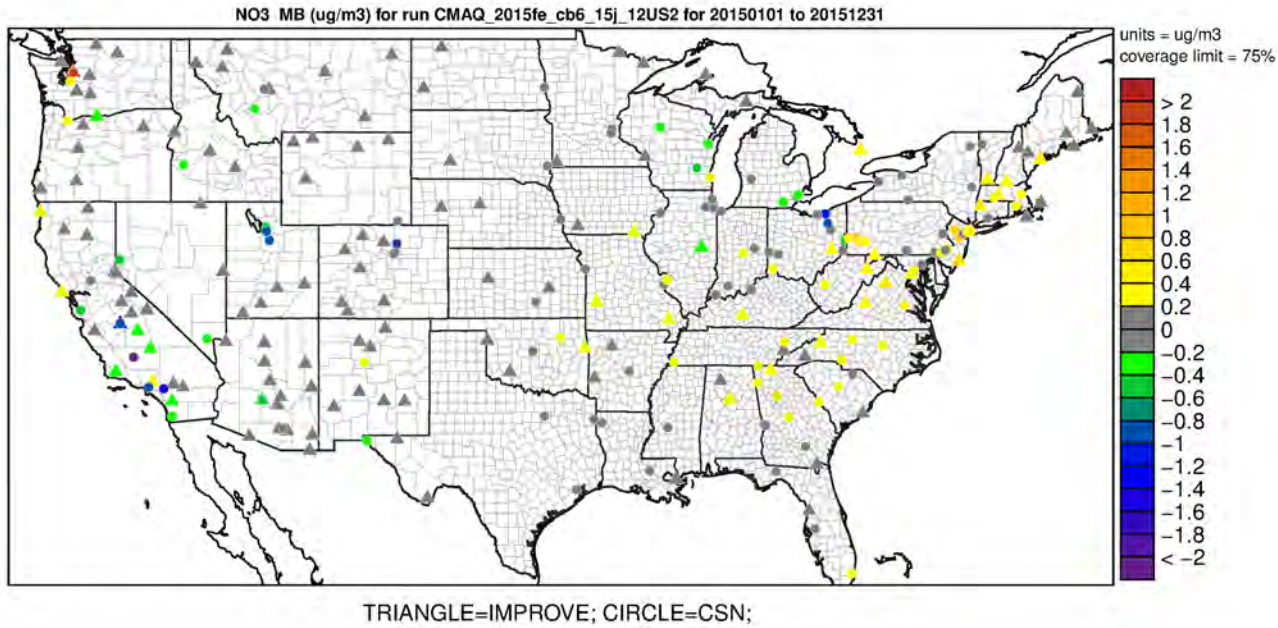


Figure 4-11. Mean Bias (µgm^{-3}) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

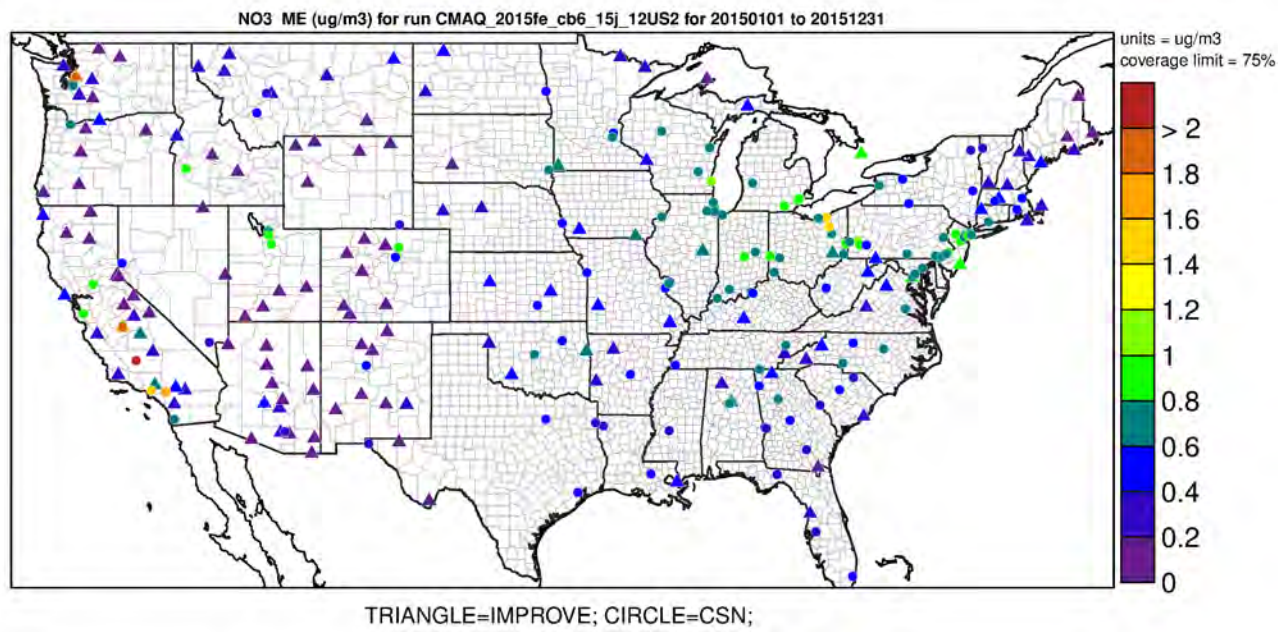


Figure 4-12. Mean Error (µgm^{-3}) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

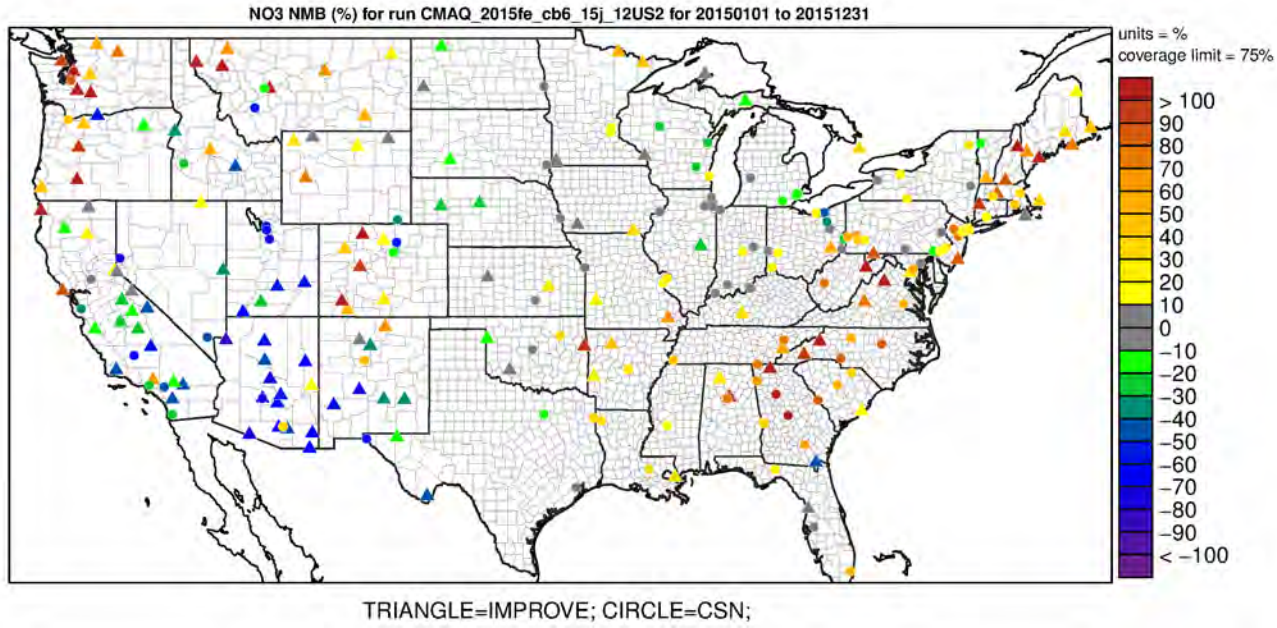


Figure 4-13. Normalized Mean Bias (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

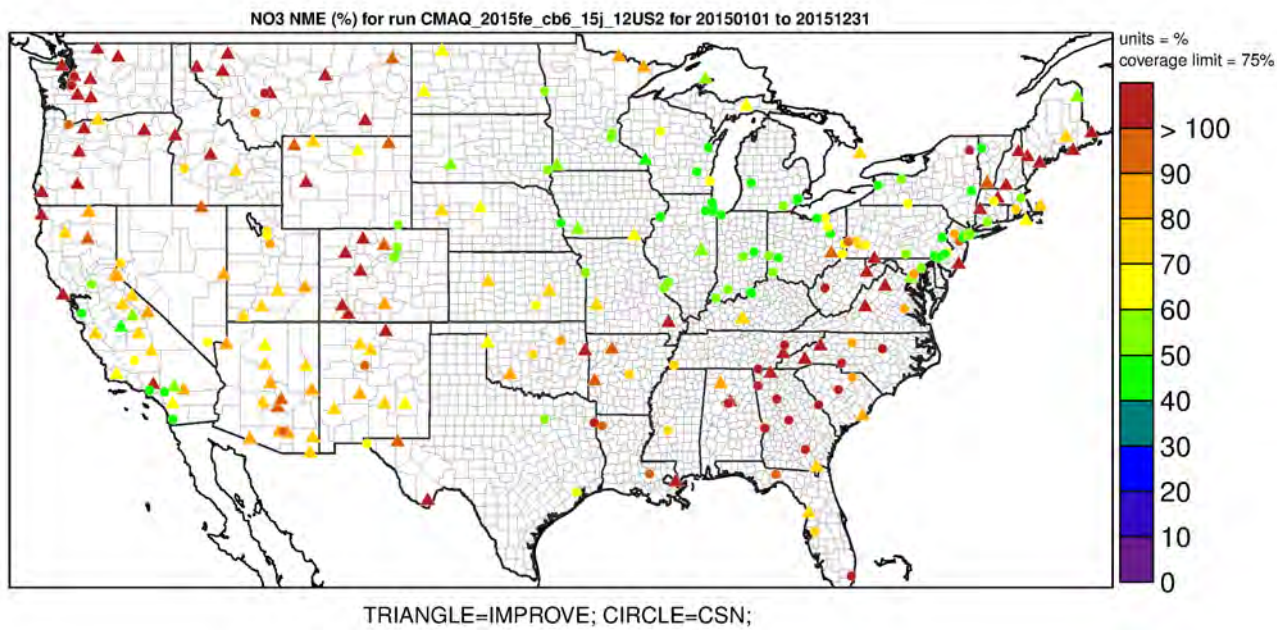


Figure 4-14. Normalized Mean Error (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

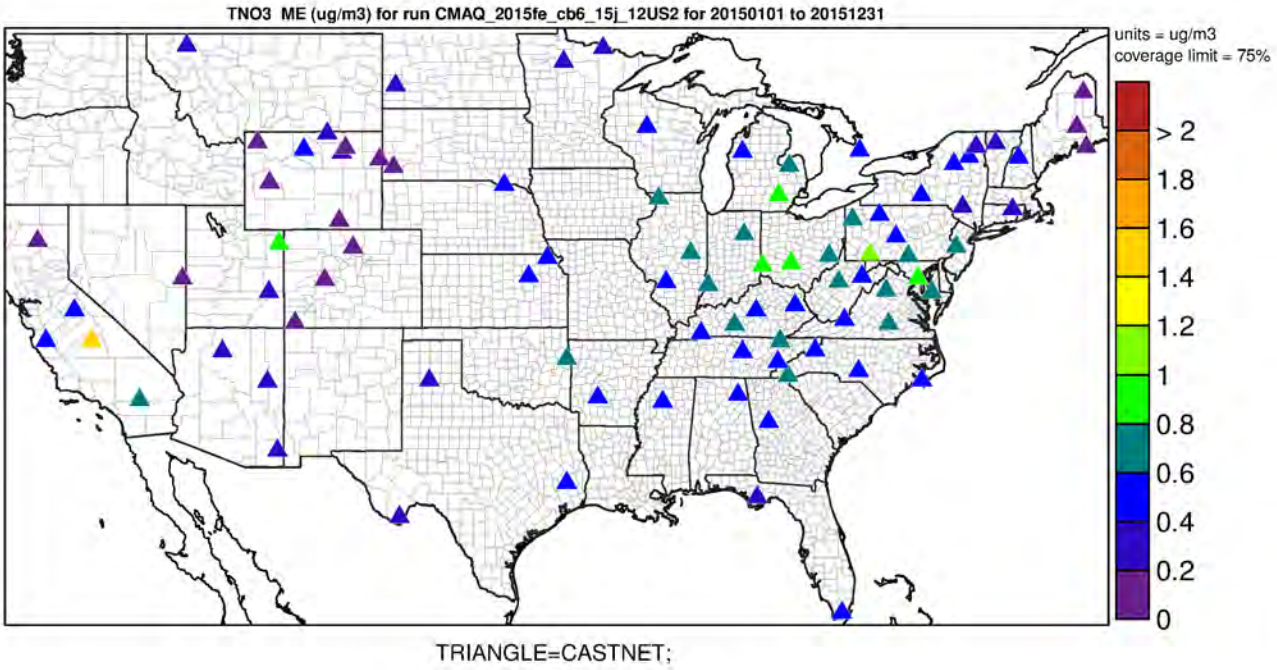


Figure 4-15. Mean Bias (μgm^{-3}) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

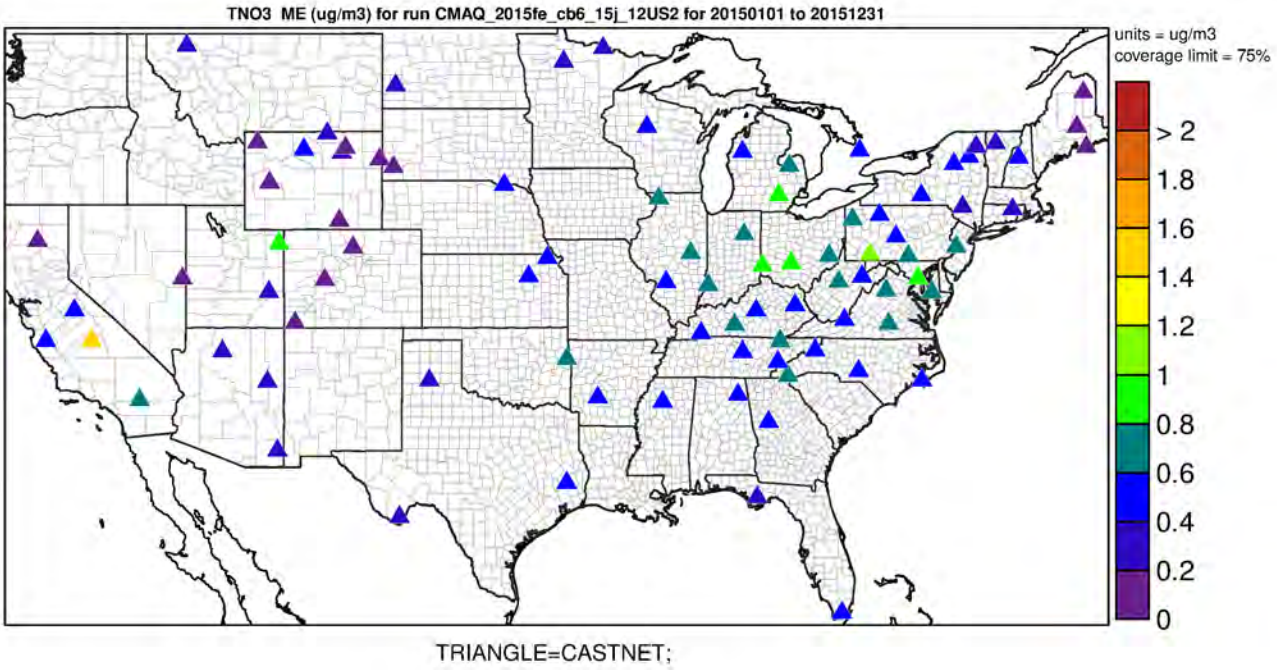


Figure 4-16. Mean Error (μgm^{-3}) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

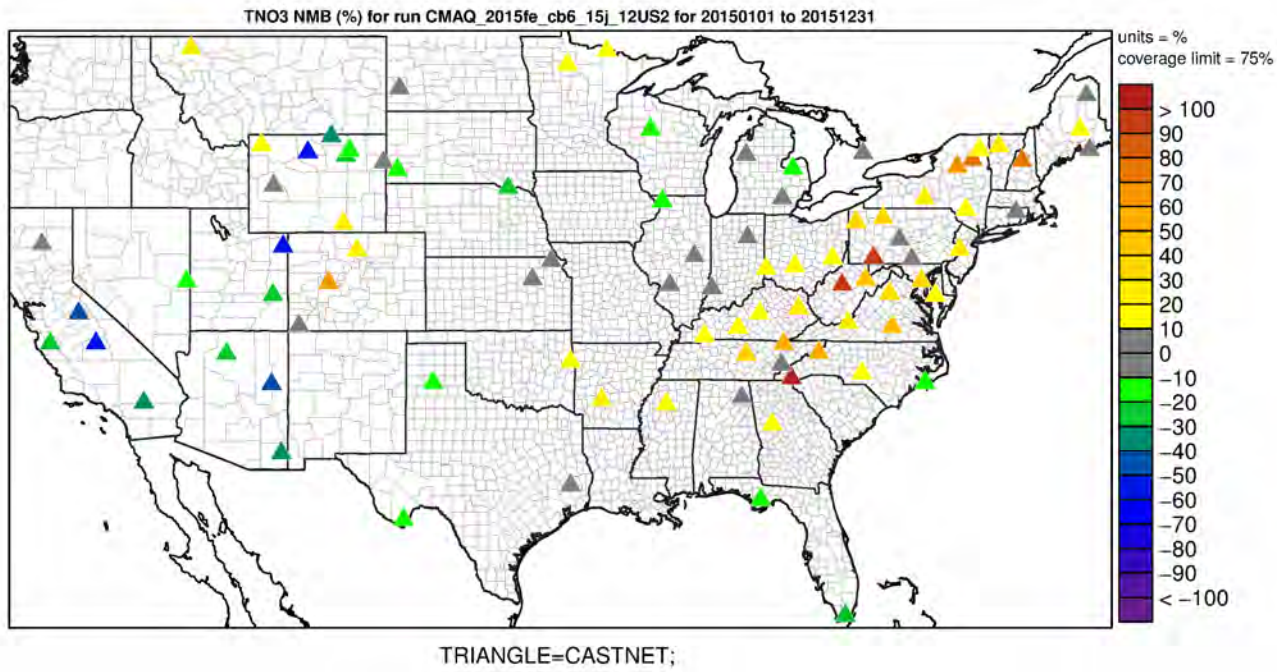


Figure 4-17. Normalized Mean Bias (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

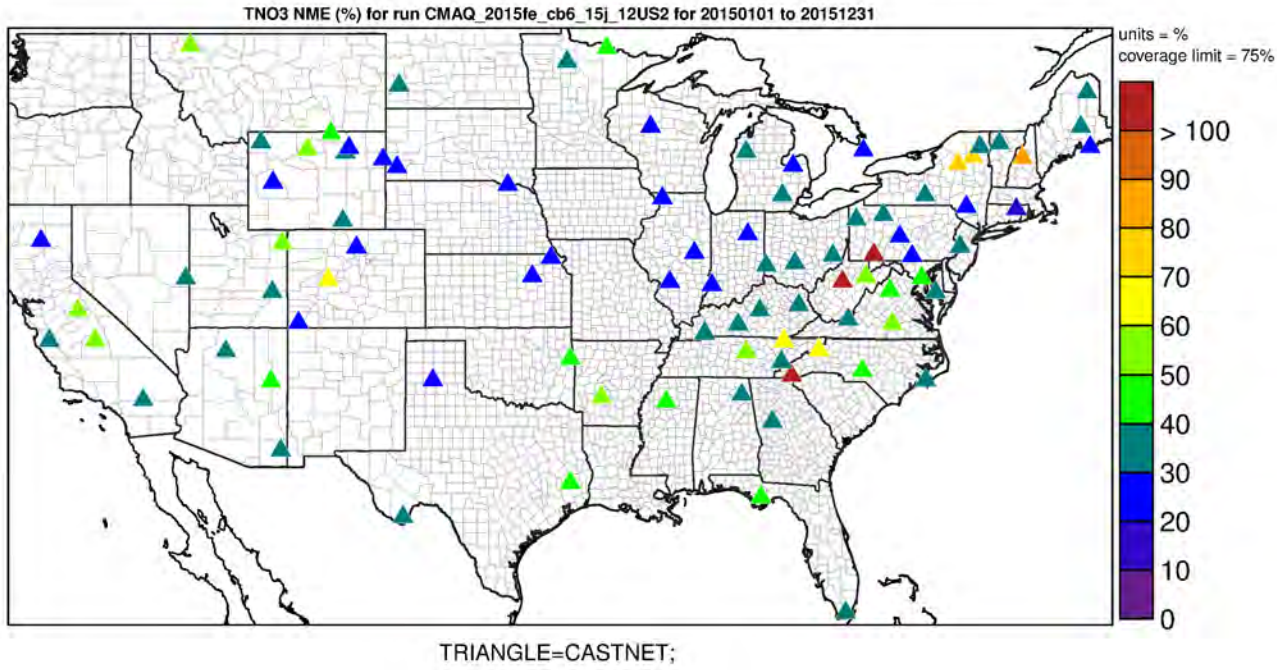


Figure 4-18. Normalized Mean Error (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

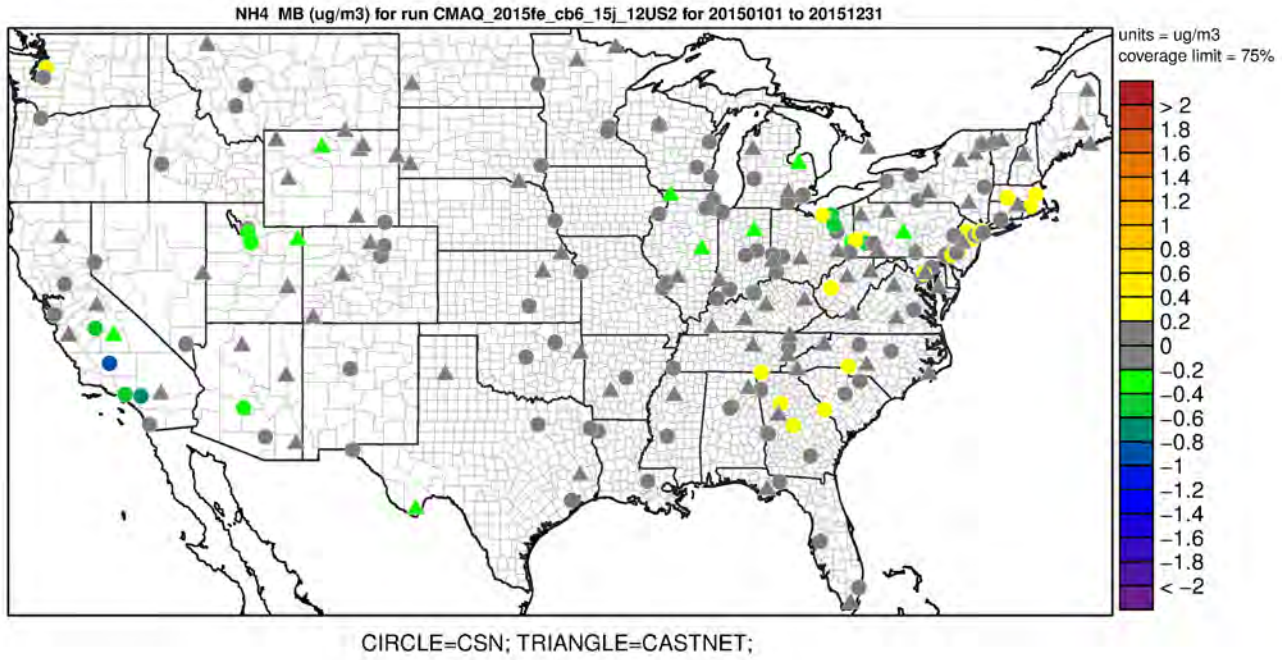


Figure 4-19. Mean Bias (μgm^{-3}) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

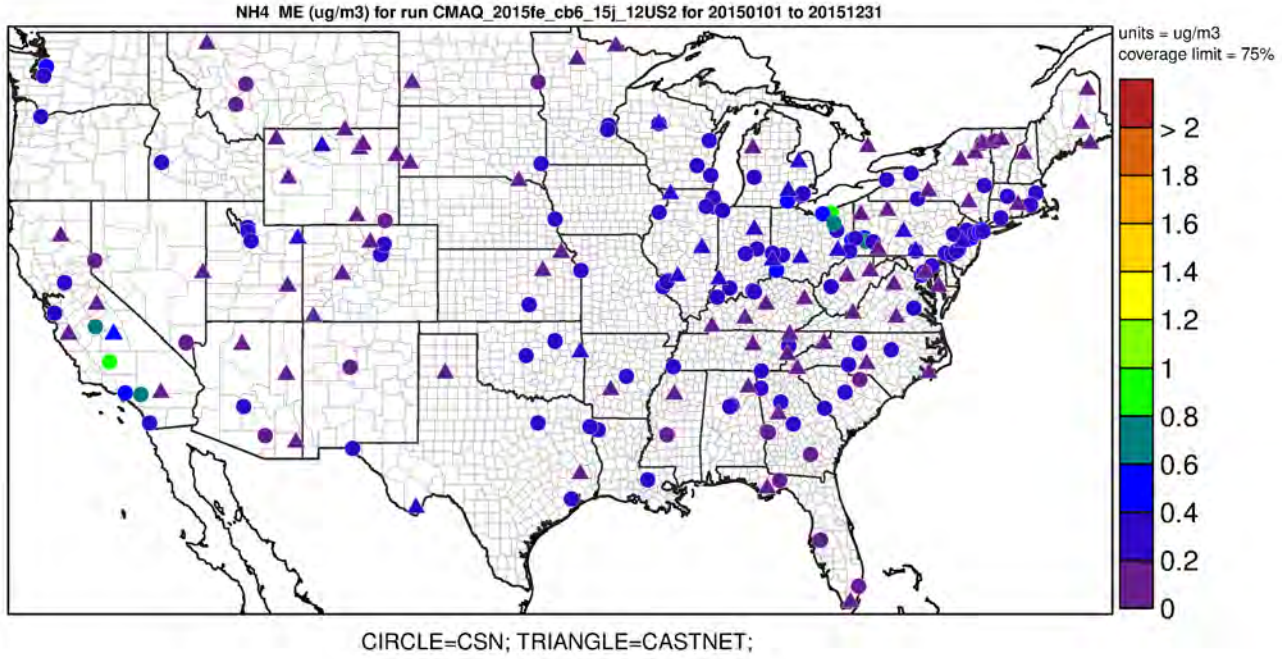


Figure 4-20. Mean Error (μgm^{-3}) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

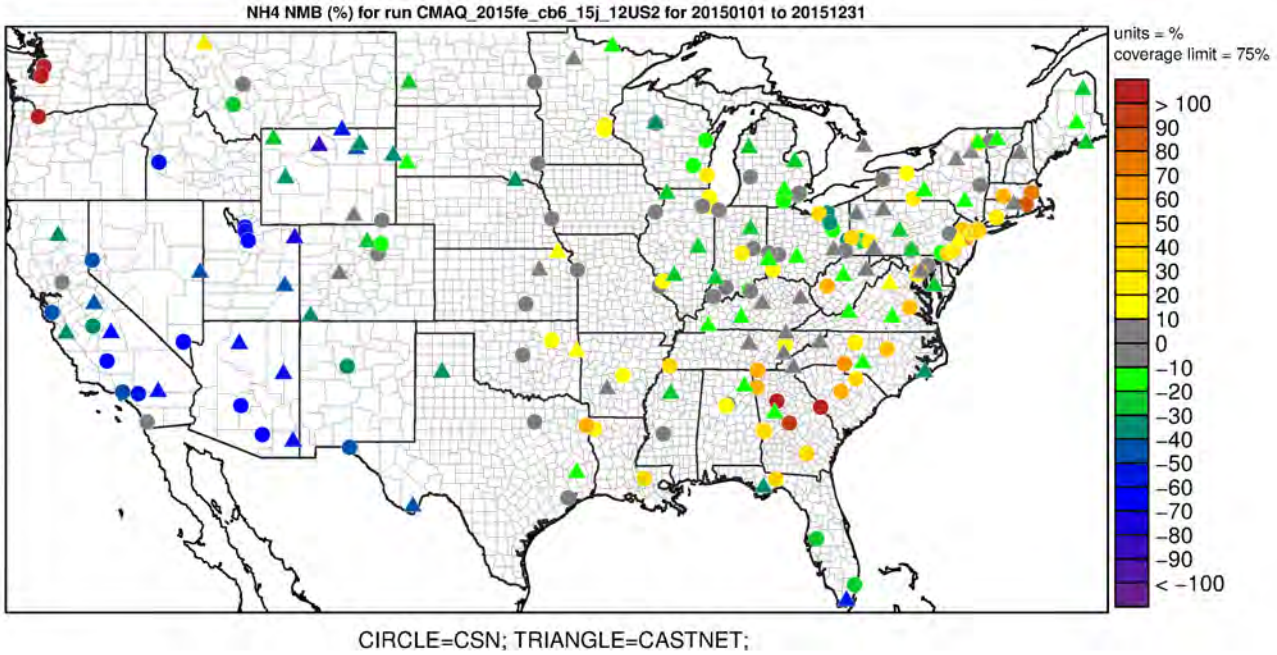


Figure 4-21. Normalized Mean Bias (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

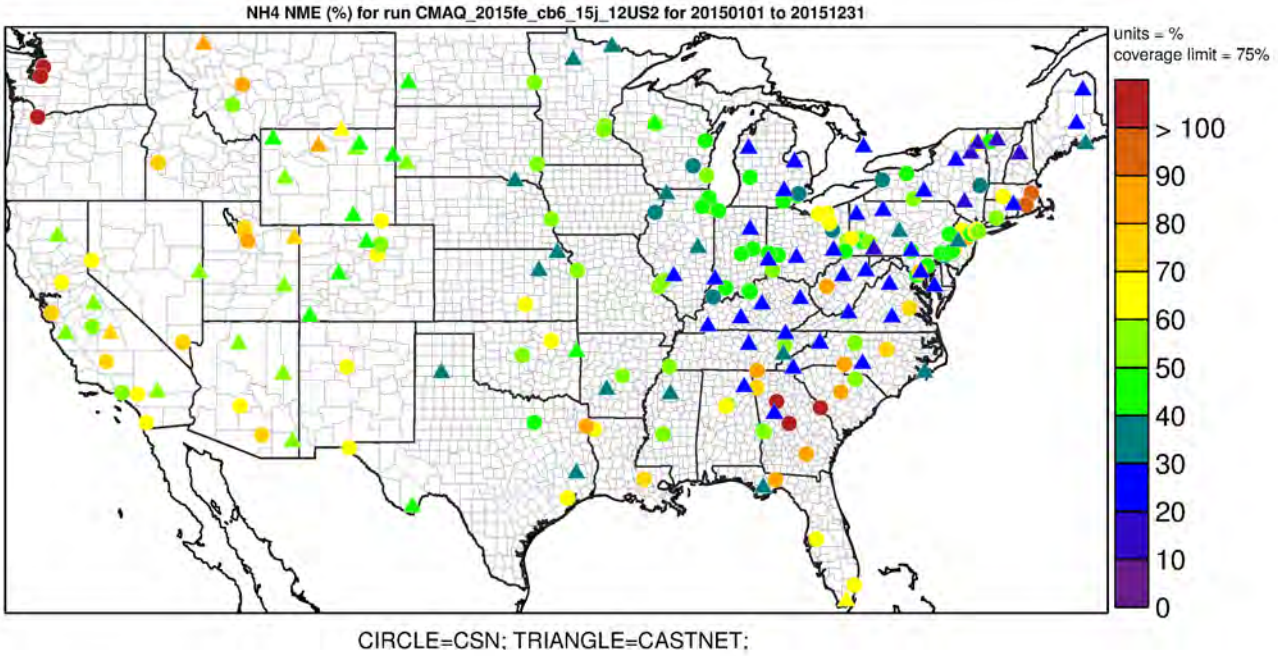


Figure 4-22. Normalized Mean Error (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

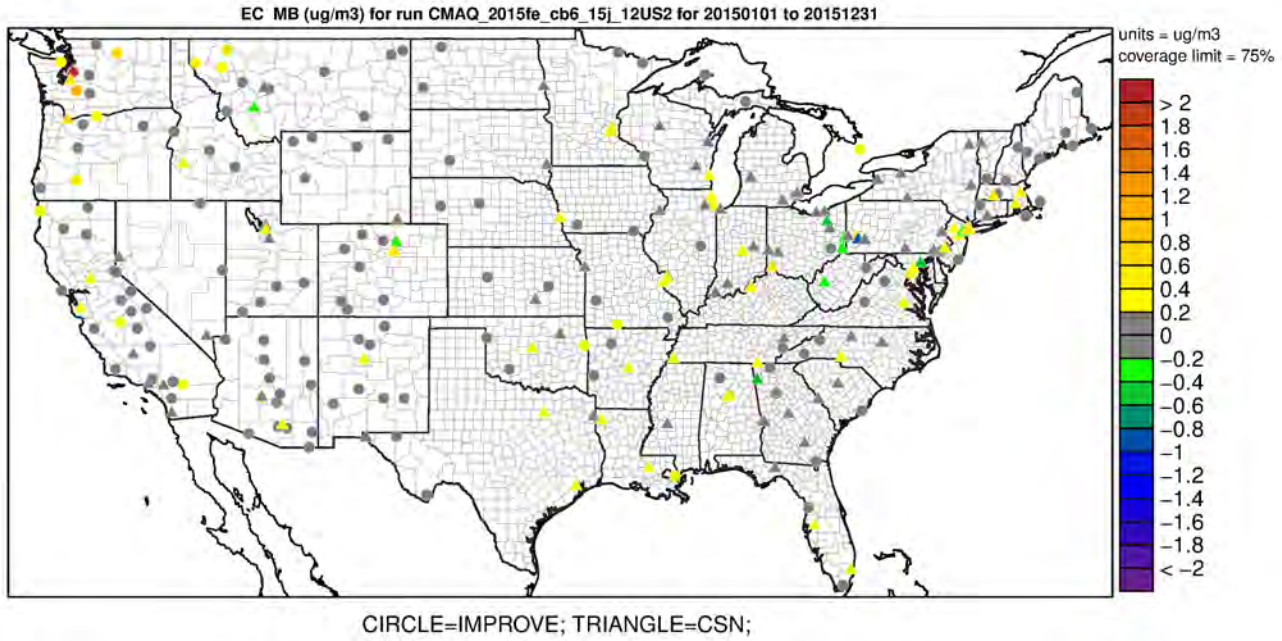


Figure 4-23. Mean Bias (μgm^{-3}) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

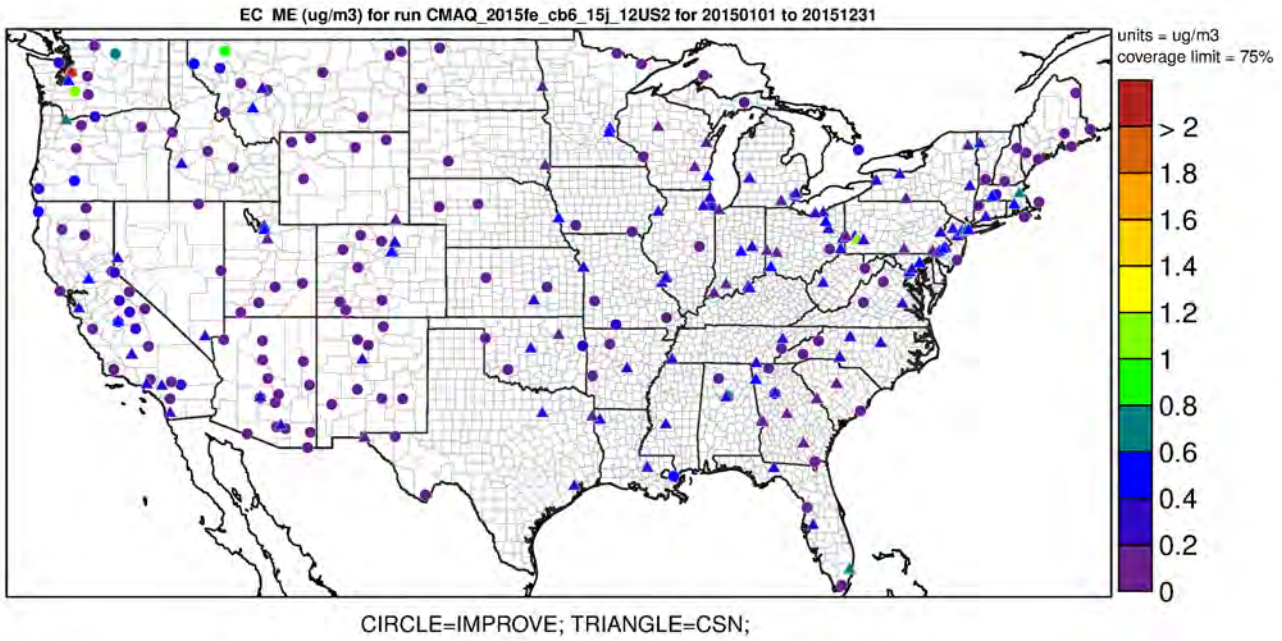


Figure 4-24. Mean Error (μgm^{-3}) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

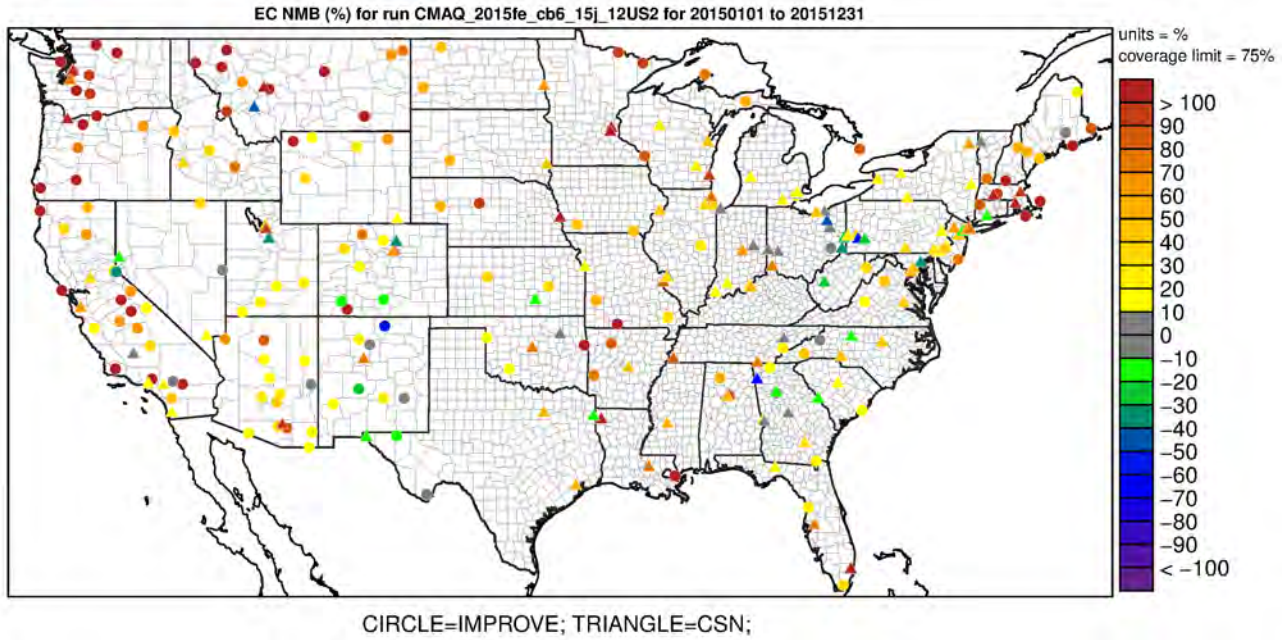


Figure 4-25. Normalized Mean Bias (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

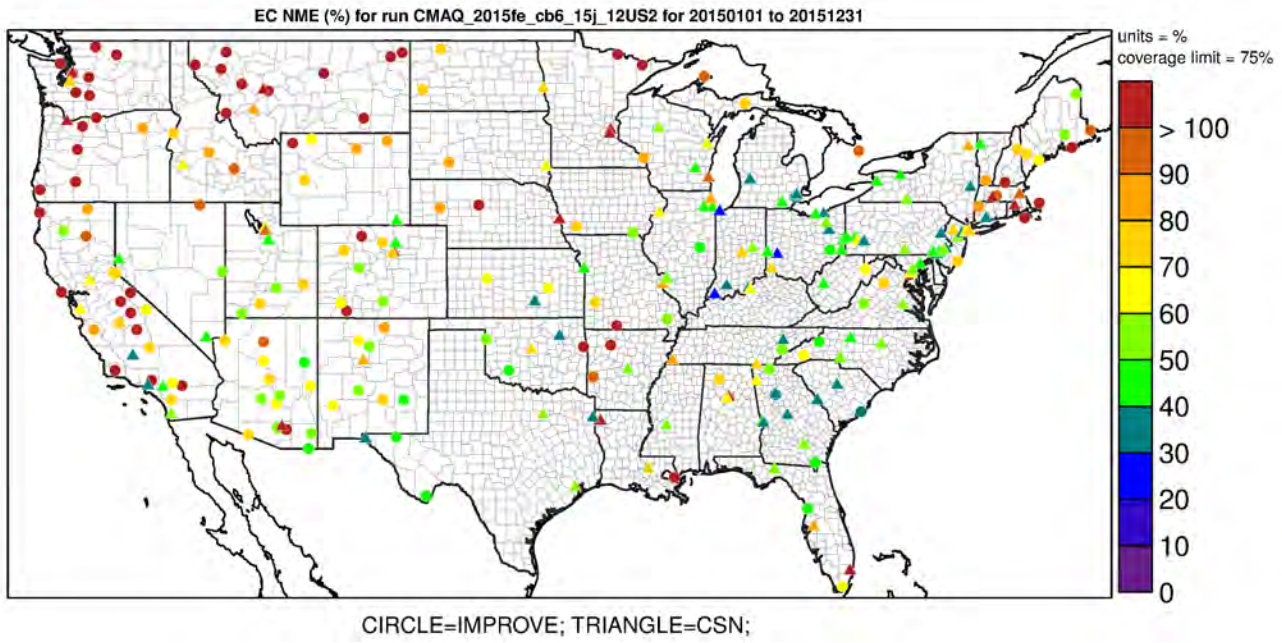


Figure 4-26. Normalized Mean Error (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

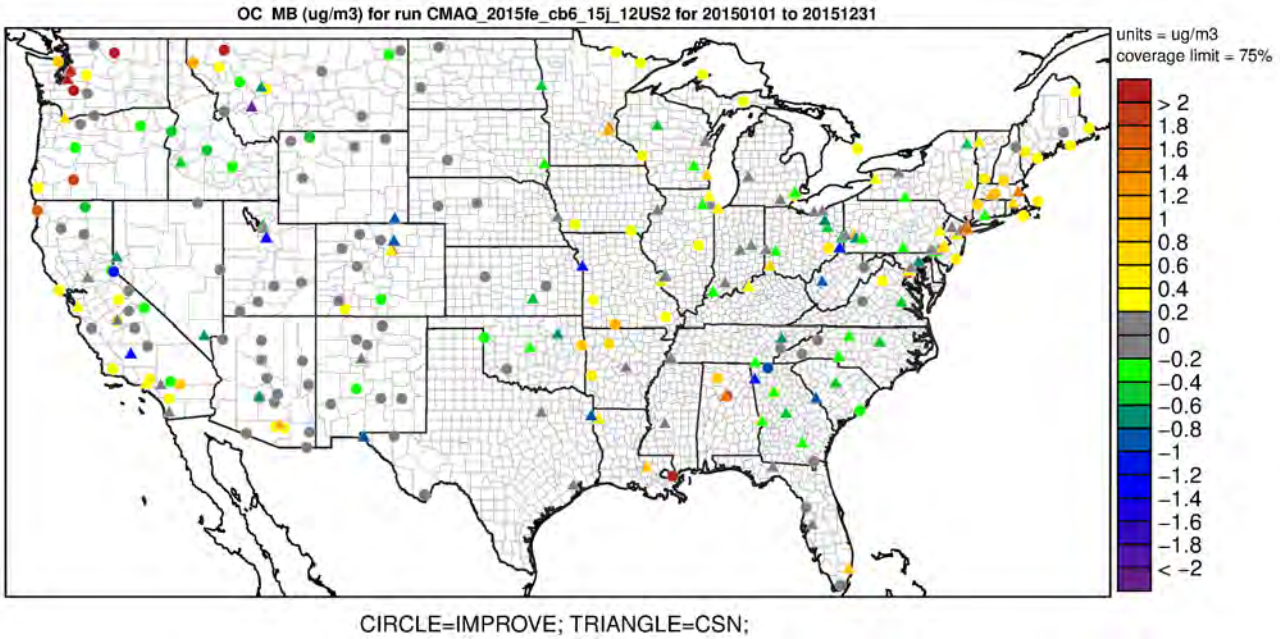


Figure 4-27. Mean Bias (μgm^{-3}) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

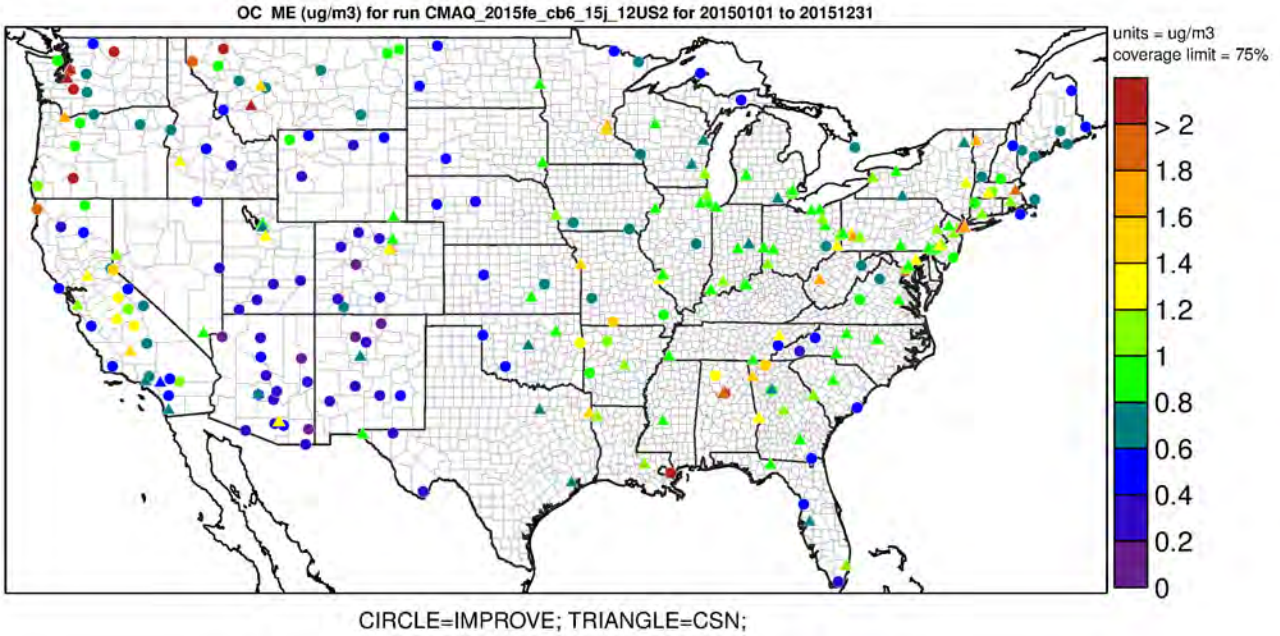


Figure 4-28. Mean Error (μgm^{-3}) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

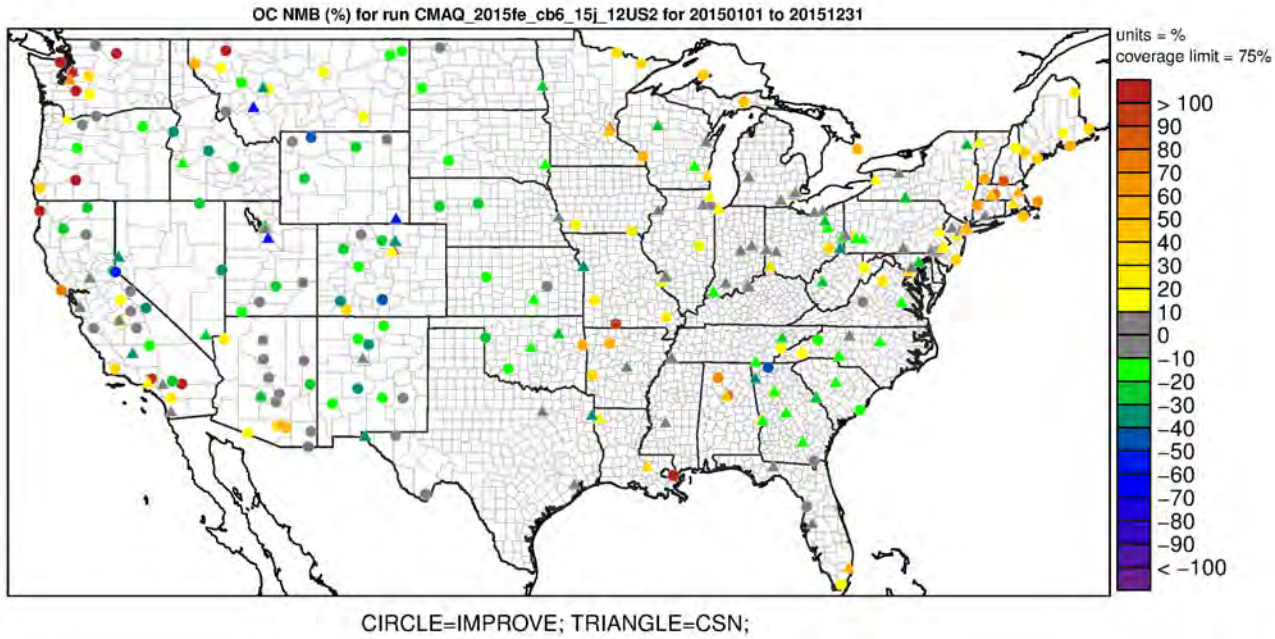


Figure 4-29. Normalized Mean Bias (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

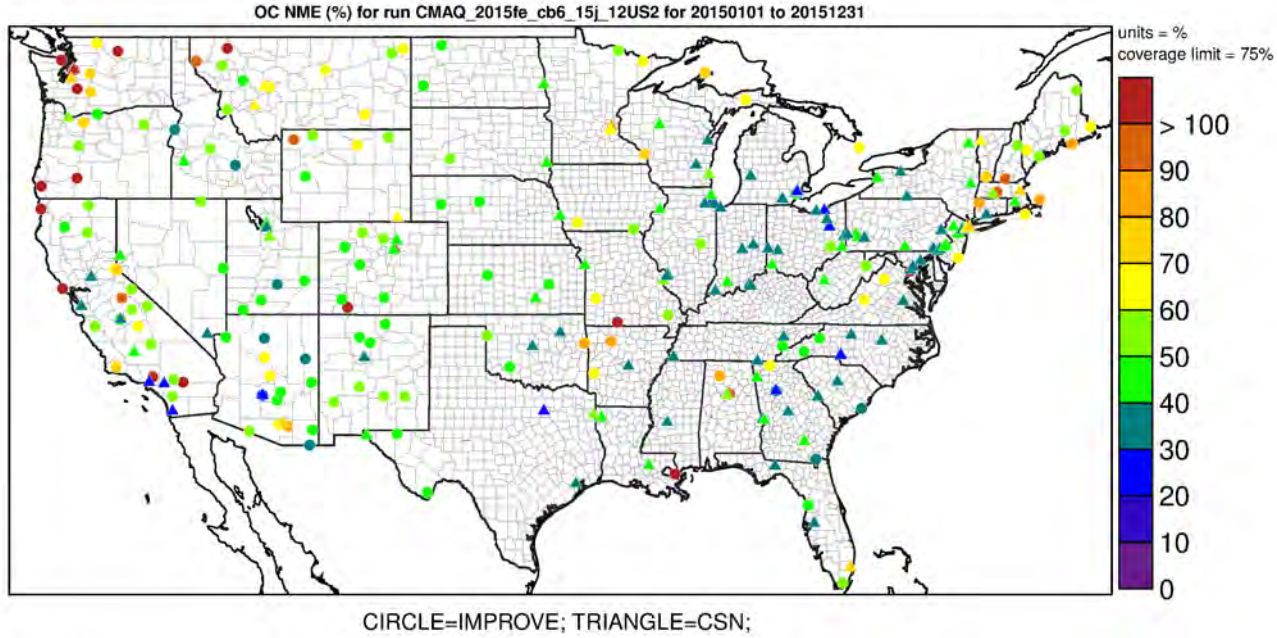


Figure 4-30. Normalized Mean Bias (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

5.0 Bayesian space-time downscaling fusion model (downscaler) - Derived Air Quality Estimates

5.1 Introduction

The need for greater spatial coverage of air pollution concentration estimates has grown in recent years as epidemiology and exposure studies that link air pollution concentrations to health effects have become more robust and as regulatory needs have increased. Direct measurement of concentrations is the ideal way of generating such data, but prohibitive logistics and costs limit the possible spatial coverage and temporal resolution of such a database. Numerical methods that extend the spatial coverage of existing air pollution networks with a high degree of confidence are thus a topic of current investigation by researchers. The downscaler model (DS) is the result of the latest research efforts by EPA for performing such predictions. DS utilizes both monitoring and CMAQ data as inputs and attempts to take advantage of the measurement data's accuracy and CMAQ's spatial coverage to produce new spatial predictions. This chapter describes methods and results of the DS application that accompany this report, which utilized ozone and $PM_{2.5}$ data from AQS and CMAQ to produce predictions to continental U.S. 2010 census tract centroids for the year 2015.

5.2 Downscaler Model

DS develops a relationship between observed and modeled concentrations, and then uses that relationship to spatially predict what measurements would be at new locations in the spatial domain based on the input data. This process is separately applied for each time step (daily in this work) of data, and for each of the pollutants under study (ozone and $PM_{2.5}$). In its most general form, the model can be expressed in an equation similar to that of linear regression:

$$Y(s) = \tilde{\beta}_0(s) + \beta_1 \tilde{x}(s) + \varepsilon(s) \qquad \text{(Equation 1)}$$

Where:

Y(s) is the observed concentration at point *s*. Note that *Y(s)* could be expressed as $Y_t(s)$, where *t* indicates the model being fit at time *t* (in this case, *t=1,…,365* would represent day of the year.)
$\tilde{x}(s)$ is the point-level regressor based on the CMAQ concentration at point *s*. This value is a weighted average of both the gridcell containing the monitor and neighboring gridcells.
$\tilde{\beta}_0(s)$ is the intercept, where $\tilde{\beta}_0(s) = \beta_0 + \beta_0(s)$ is composed of both a global component β_0 and a local component $\beta_0(s)$ that is modeled as a mean-zero Gaussian Process with exponential decay
β_1 is the global slope; local components of the slope are contained in the $\tilde{x}(s)$ term.
$\varepsilon(s)$ is the model error.

DS has additional properties that differentiate it from linear regression:

1) Rather than just finding a single optimal solution to Equation 1, DS uses a Bayesian approach so that uncertainties can be generated along with each concentration prediction. This involves drawing random

samples of model parameters from built-in "prior" distributions and assessing their fit on the data on the order of thousands of times. After each iteration, properties of the prior distributions are adjusted to try to improve the fit of the next iteration. The resulting collection of $\tilde{\beta}_0$ and β_1 values at each space-time point are the "posterior" distributions, and the means and standard distributions of these are used to predict concentrations and associated uncertainties at new spatial points.

2) The model is "hierarchical" in structure, meaning that the top-level parameters in Equation 1 (ie $\tilde{\beta}_0(s)$, β_1, $\tilde{x}(s)$) are actually defined in terms of further parameters and sub-parameters in the DS code. For example, the overall slope and intercept is defined to be the sum of a global (one value for the entire spatial domain) and local (values specific to each spatial point) component. This gives more flexibility in fitting a model to the data to optimize the fit (i.e. minimize $\varepsilon(s)$).

Further information about the development and inner workings of the current version of DS can be found in Berrocal, Gelfand and Holland (2012)33 and references therein. The DS outputs that accompany this report are described below, along with some additional analyses that include assessing the accuracy of the DS predictions. Results are then summarized, and caveats are provided for interpreting them in the context of air quality management activities.

5.3 Downscaler Concentration Predictions

In this application, DS was used to predict daily concentration and associated uncertainty values at the 2010 US census tract centroids across the continental U.S. using 2015 measurement and CMAQ data as inputs. For ozone, the concentration unit is the daily maximum 8-hour average in ppb and for $PM_{2.5}$ the concentration unit is the 24-hour average in $\mu g/m^3$.

5.3.1 Summary of 8-hour Ozone Results

Figure 5-1 summarizes the AQS, CMAQ and DS ozone data over the year 2015. It shows the 4th max daily maximum 8-hour average ozone for AQS observations, CMAQ model predictions and DS model results. The DS model estimated that for 2015, about 29% of the US Census tracts (20972 out of 72283) experienced at least one day with an ozone value above the NAAQS of 75 ppb.

33 Berrocal, V., Gelfand, A., and D. Holland. Space-Time Data Fusion Under Error in Computer Model Output: An Application to Modeling Air Quality. *Biometrics.* 2012. September ; 68(3): 837–848. doi:10.1111/j.1541-0420.2011.01725.x

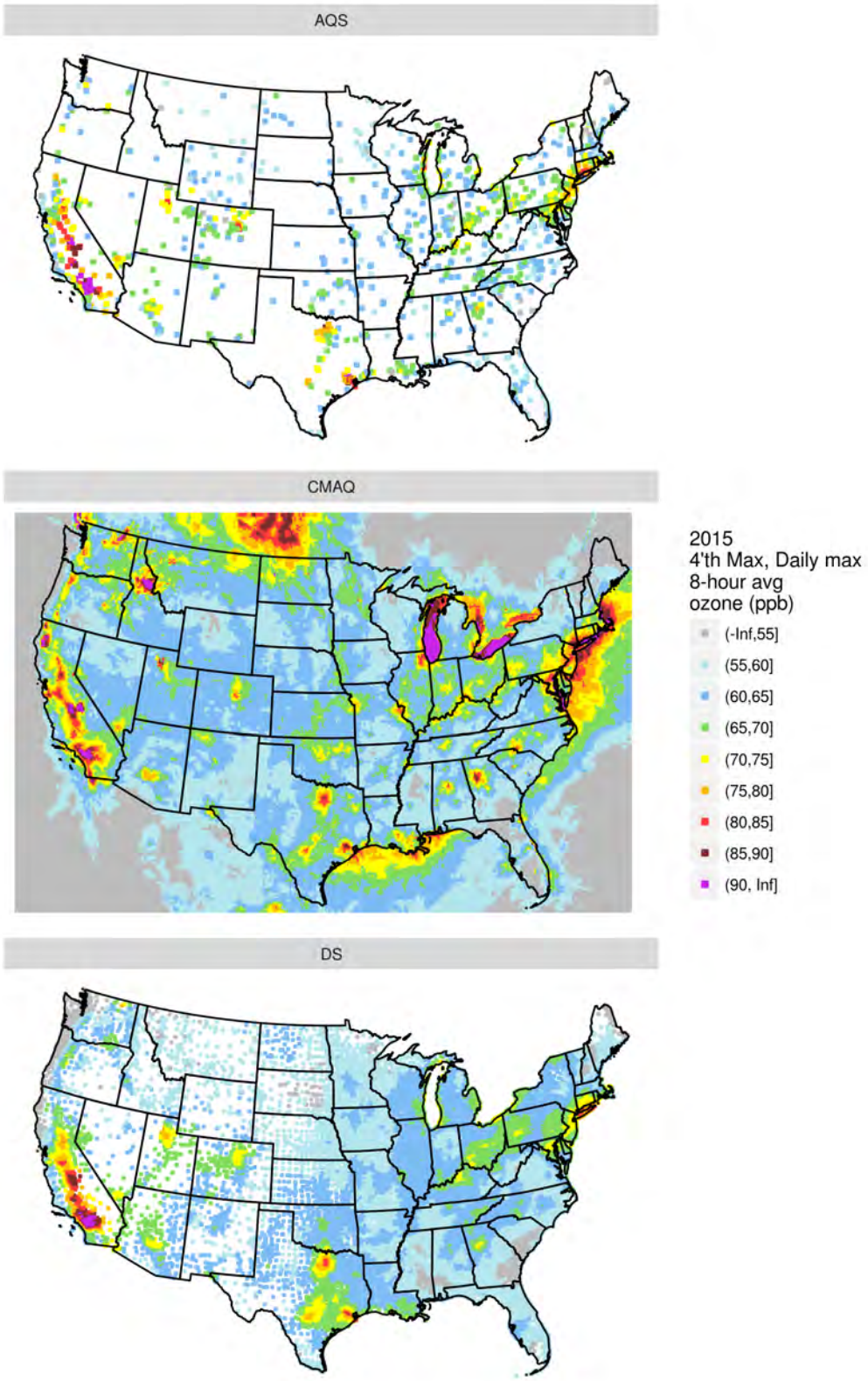


Figure 5-1. Annual 4th max (daily max 8-hour ozone concentrations) derived from AQS, CMAQ and DS data.

5.3.2 Summary of $PM_{2.5}$ Results

Figures 5-2 and 5-3 summarize the AQS, CMAQ and DS PM2.5 data over the year 2015. Figure 5-2 shows annual means and Figure 5-3 shows 98thpercentiles of 24-hour PM2.5 concentrations for AQS observations, CMAQ model predictions and DS model results. The DS model estimated that for 2015 about 28% of the US Census tracts (20061 out of 72283) experienced at least one day with a PM2.5 value above the 24-hour NAAQS of 35 $\mu g/m^3$.

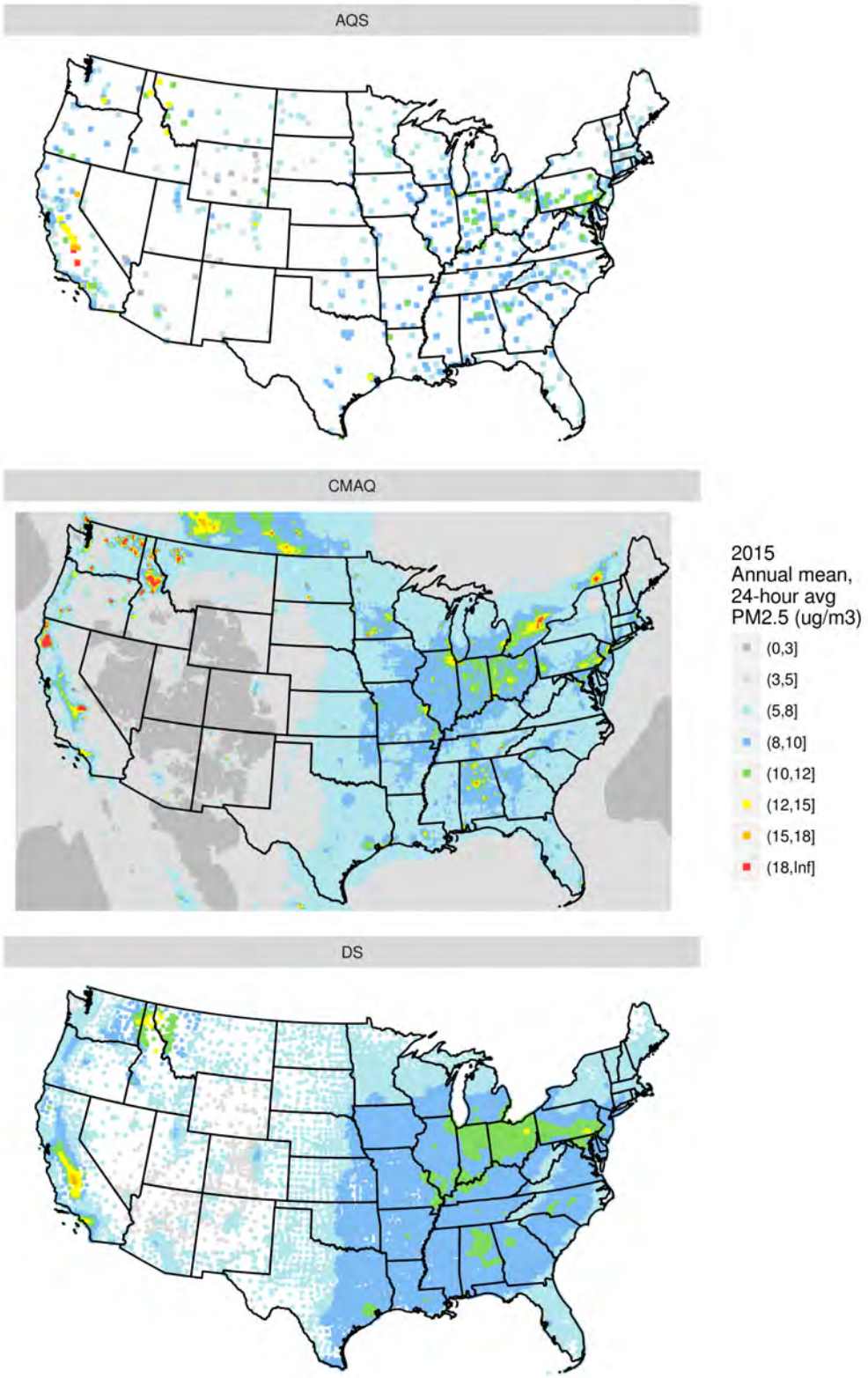


Figure 5-2. Annual mean $PM_{2.5}$ concentrations derived from AQS, CMAQ and DS data.

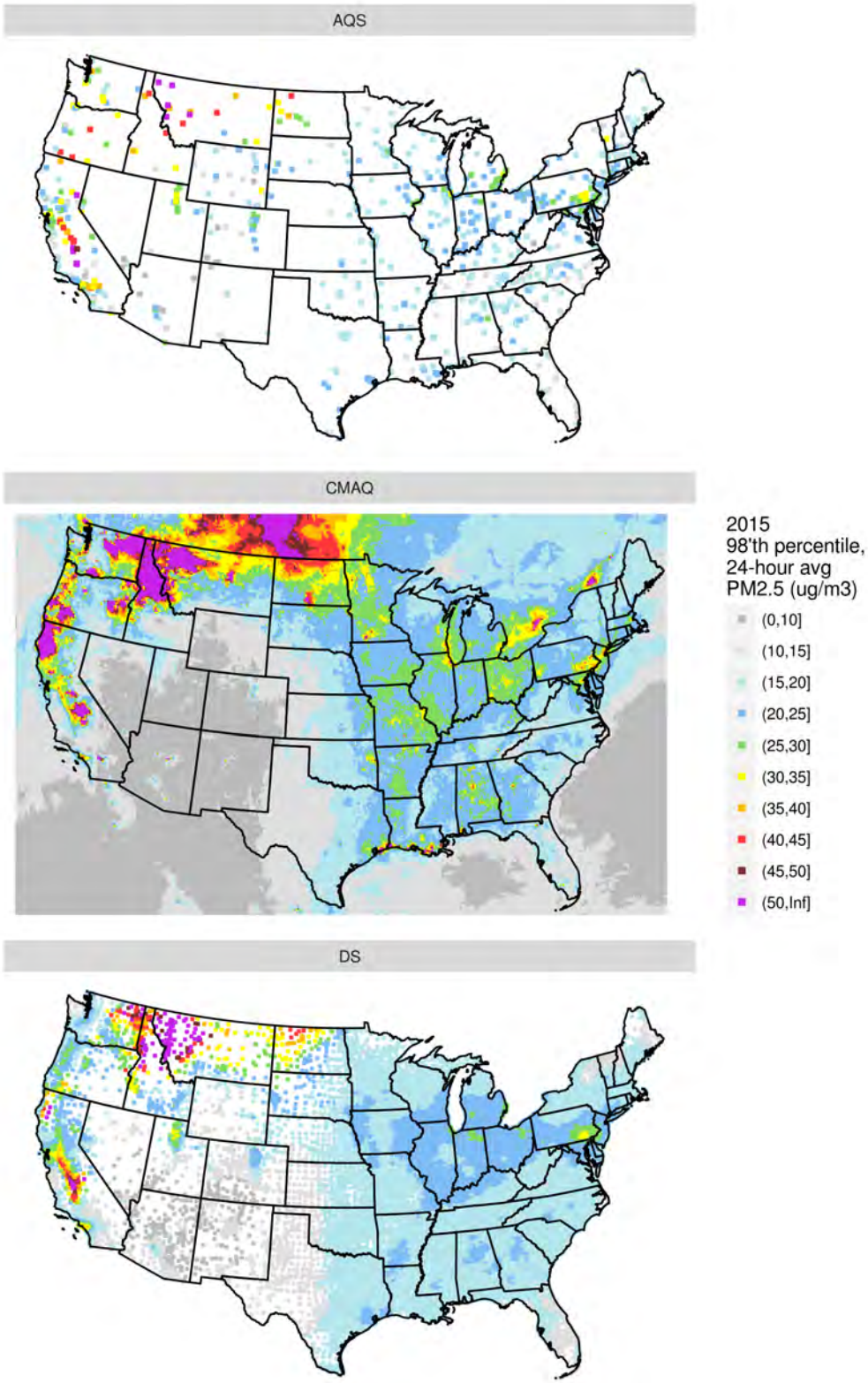


Figure 5-3. 98th percentile 24-hour average $PM_{2.5}$ concentrations derived from AQS, CMAQ and DS data.

5.4 Downscaler Uncertainties

5.4.1 Standard Errors

As mentioned above, the DS model works by drawing random samples from built-in distributions during its parameter estimation. The standard errors associated with each of these populations provide a measure of uncertainty associated with each concentration prediction. Figure 5-4 shows the percent errors resulting from dividing the DS standard errors by the associated DS prediction. The black dots on the maps show the location of EPA sampling network monitors whose data was input to DS via the AQS datasets (Chapter 2). The maps show that, in general, errors are relatively smaller in regions with more densely situation monitors (ie the eastern US), and larger in regions with more sparse monitoring networks (ie western states). These standard errors could potentially be used to estimate the probability of an exceedance for a given point estimate of a pollutant concentration.

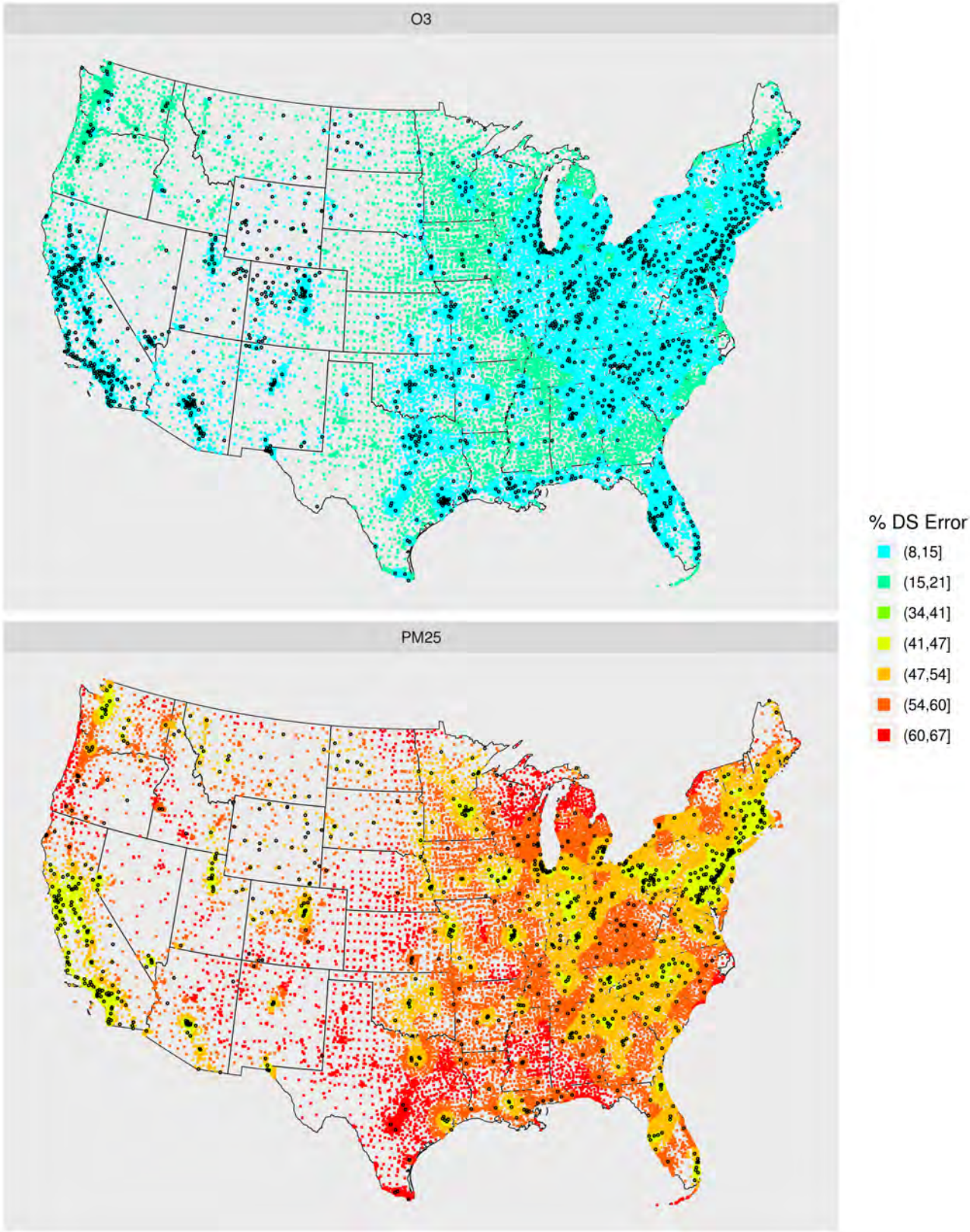


Figure 5-4. Annual mean relative errors (standard errors divided by predictions) from the DS 2015 runs. The black dots show the locations of monitors that generated the AQS data used as input to the DS model.

5.4.2 Cross Validation

To check the quality of its spatial predictions, DS can be set to perform "cross-validation" (CV), which involves leaving a subset of AQS data out of the model run and predicting the concentrations of those left out points. The predicted values are then compared to the actual left-out values to generate statistics that provide an indicator of the predictive ability. In the DS runs associated with this report, 10% of the data was chosen randomly by the DS model to be used for the CV process. The resulting CV statistics are shown below in Table 5-1.

Pollutant	**# Monitors**	**Mean Bias**	**RMSE**	**Mean Coverage**
PM2.5	947	0.28	3.5	0.96
O3	1247	-0.0017	4.4	0.96

Table 5-1. Cross-validation statistics associated with the 2015 DS runs.

The statistics indicated by the columns of Table 5-1 are as follows:

- Mean Bias: The bias of each prediction is the DS prediction minus the AQS value. This column is the mean of all biases across the CV cases.

- Root Mean Squared Error (RMSE): The bias is squared for each CV prediction, then the square root of the mean of all squared biases across all CV predictions is obtained.

- Mean Coverage: A value of 1 is assigned if the measured AQS value lies in the 95% confidence interval of the DS prediction (the DS prediction +/- the DS standard error), and 0 otherwise. This column is the mean of all those 0's and 1's.

5.5 Summary and Conclusions

The results presented in this report are from an application of the DS fusion model for characterizing national air quality for Ozone and $PM_{2.5}$. DS provided spatial predictions of daily ozone and $PM_{2.5}$ at 2010 U.S. census tract centroids by utilizing monitoring data and CMAQ output for 2015. Large-scale spatial and temporal patterns of concentration predictions are generally consistent with those seen in ambient monitoring data. Both Ozone and $PM_{2.5}$ were predicted with lower error in the eastern versus the western U.S., presumably due to the greater monitoring density in the east.

An additional caution that warrants mentioning is related to the capability of DS to provide predictions at multiple spatial points within a single CMAQ gridcell. Care needs to be taken not to over-interpret any within-gridcell gradients that might be produced by a user. Fine-scale emission sources in CMAQ are diluted into the gridcell averages, but a given source within a gridcell might or might not affect every spatial point contained therein equally. Therefore DS-generated fine-scale gradients are not expected to represent actual fine-scale atmospheric concentration gradients, unless possibly where multiple monitors are present in the gridcell.

Appendix A - Acronyms

Acronyms

ARW	Advanced Research WRF core model
BEIS	Biogenic Emissions Inventory System
BlueSky	Emissions modeling framework
CAIR	Clean Air Interstate Rule
CAMD	EPA's Clean Air Markets Division
CAP	Criteria Air Pollutant
CAR	Conditional Auto Regressive spatial covariance structure (model)
CARB	California Air Resources Board
CEM	Continuous Emissions Monitoring
CHIEF	Clearinghouse for Inventories and Emissions Factors
CMAQ	Community Multiscale Air Quality model
CMV	Commercial marine vessel
CO	Carbon monoxide
CSN	Chemical Speciation Network
DQO	Data Quality Objectives
EGU	Electric Generating Units
Emission Inventory	Listing of elements contributing to atmospheric release of pollutant substances
EPA	Environmental Protection Agency
EMFAC	Emission Factor (California's onroad mobile model)
FAA	Federal Aviation Administration
FDDA	Four Dimensional Data Assimilation
FIPS	Federal Information Processing Standards
HAP	Hazardous Air Pollutant
HMS	Hazard Mapping System
ICS-209	Incident Status Summary form
IPM	Integrated Planning Model
ITN	Itinerant
LSM	Land Surface Model
MOBILE	OTAQ's model for estimation of onroad mobile emissions factors
MODIS	Moderate Resolution Imaging Spectroradiometer
MOVES	Motor Vehicle Emission Simulator
NEEDS	National Electric Energy Database System
NEI	National Emission Inventory
NERL	National Exposure Research Laboratory
NESHAP	National Emission Standards for Hazardous Air Pollutants
NH	Ammonia
NMIM	National Mobile Inventory Model
NONROAD	OTAQ's model for estimation of nonroad mobile emissions
NO	Nitrogen oxides
OAQPS	EPA's Office of Air Quality Planning and Standards

OAR	EPA's Office of Air and Radiation
ORD	EPA's Office of Research and Development
ORIS	Office of Regulatory Information Systems (code) - is a 4 or 5 digit number assigned by the Department of Energy's (DOE) Energy Information Agency (EIA) to facilities that generate electricity
ORL	One Record per Line
OTAQ	EPA's Office of Transportation and Air Quality
PAH	Polycyclic Aromatic Hydrocarbon
PFC	Portable Fuel Container
$PM_{2.5}$	Particulate matter less than or equal to 2.5 microns
PM_{10}	Particulate matter less than or equal to 10 microns
PMc	Particulate matter greater than 2.5 microns and less than 10 microns
Prescribed Fire	Intentionally set fire to clear vegetation
RIA	Regulatory Impact Analysis
RPO	Regional Planning Organization
RRTM	Rapid Radiative Transfer Model
SCC	Source Classification Code
SMARTFIRE	Satellite Mapping Automatic Reanalysis Tool for Fire Incident Reconciliation
SMOKE	Sparse Matrix Operator Kernel Emissions
TCEQ	Texas Commission on Environmental Quality
TSD	Technical support document
VOC	Volatile organic compounds
VMT	Vehicle miles traveled
Wildfire	Uncontrolled forest fire
WRAP	Western Regional Air Partnership
WRF	Weather Research and Forecasting Model

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Air Quality Assessment Division Research Triangle Park, NC	Publication No. EPA-454/R-19-009 June 2019