



# Bayesian Space-time Downscaling Fusion Model (Downscaler) - Derived Estimates of Air Quality for 2020



EPA-454/R-23-004  
December 2023

Bayesian Space-time Downscaling Fusion Model (Downscaler) - Derived Estimates of Air  
Quality for 2020

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)  
Janice Godfrey (EPA/OAR)  
Jeff Vukovich (EPA/OAR)  
David Mintz (EPA/OAR)

**Acknowledgements**

The following people served as reviewers of this document: Caroline Farkas (EPA/OAR) and David Mintz (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 .....	15
3.3 Emissions Modeling Summary .....	50
3.4 Emissions References .....	95
4.0 CMAQ Air Quality Model Estimates .....	100
4.1 Introduction to the CMAQ Modeling Platform .....	100
4.2 CMAQ Model Version, Inputs and Configuration .....	101
5.0 Bayesian space-time downscaling fusion model (downscaler) -Derived Air Quality Estimates...	132
5.1 Introduction .....	132
5.2 Downscaler Model .....	132
5.3 Downscaler Concentration Predictions .....	133
5.4 Downscaler Uncertainties .....	137
5.5 Summary and Conclusions .....	140
Appendix A - Acronyms .....	141
Appendix B – Emissions Totals by Sector .....	143

## 1.0 Introduction

This report describes estimates of daily ozone (maximum 8-hour average) and fine particulate matter (PM<sub>2.5</sub>) (24-hour average) concentrations throughout the contiguous United States during the 2020 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 and 2020 US census tract centroids encompassed by the CMAQ modeling domain. Information on EPA's air quality monitors, CMAQ model, and DS 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 PM<sub>2.5</sub>.

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<sup>1</sup> approach 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](https://www.cdc.gov/nceh/tracking/) (<https://www.cdc.gov/nceh/tracking/>). 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 data from a variety of sources and to examine, analyze and interpret those data based on their spatial and temporal characteristics.

---

<sup>1</sup> 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.

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.<sup>2</sup> 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 EPA, 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 EPA shall not be held liable for improper or incorrect use of the data described and/or contained herein.

The four remaining sections and 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 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.

---

<sup>2</sup> The original HHS and EPA MOU is available at [https://www.cdc.gov/nceh/tracking/pdfs/epa\\_mou\\_2007.pdf](https://www.cdc.gov/nceh/tracking/pdfs/epa_mou_2007.pdf).

- 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 and 2020 U.S. census tract centroid locations within the contiguous U.S.
- Appendix A provides a description of acronyms used in this report.
- Appendix B is a separate spreadsheet that shows emissions totals for the modeling domain and for each emissions modeling sector (see Section 3 for more details).



## 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<sub>3</sub>) 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 EPA has established standards, or limits, for six air pollutants known as the criteria air pollutants: carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO<sub>2</sub>), sulfur dioxide (SO<sub>2</sub>), ozone (O<sub>3</sub>), 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 ozone designations, go to <https://www.epa.gov/ozone-designations> and for PM designations, go to <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 CAA 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 National Ambient Air Quality Standards**

<b>Form of the Standard (parts per million, ppm)</b>	<b>1997</b>	<b>2008</b>	<b>2015</b>
Annual 4 <sup>th</sup> highest daily max 8-hour average, averaged over three years	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 sources.<sup>3</sup> Particles less than 10 micrometers in diameter (PM<sub>10</sub>) 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<sub>2.5</sub>) 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<sub>10-2.5</sub>) 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<sub>10</sub>, PM<sub>2.5</sub> 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>.

---

<sup>3</sup> 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<sub>2.5</sub> 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<sub>10</sub> and PM<sub>2.5</sub>.

The current NAAQS for PM<sub>2.5</sub> (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<sub>2.5</sub> concentration must not exceed 12.0 micrograms per cubic meter (ug/m<sup>3</sup>) based on the annual mean concentration averaged over three years, and the 24-hr average concentration must not exceed 35 ug/m<sup>3</sup> based on the 98<sup>th</sup> 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<sub>2.5</sub> are shown in Table 2-2.

**Table 2-2. PM<sub>2.5</sub> National Ambient Air Quality Standards**

<b>Form of the Standard (micrograms per cubic meter, µg/m<sup>3</sup>)</b>	<b>1997</b>	<b>2006</b>	<b>2012</b>
Annual mean of 24-hour averages, averaged over 3 years	15.0	15.0	12.0
98th percentile of 24-hour averages, averaged over 3 years	65	35	35

## 2.2 Ambient Air Quality Monitoring in the United States

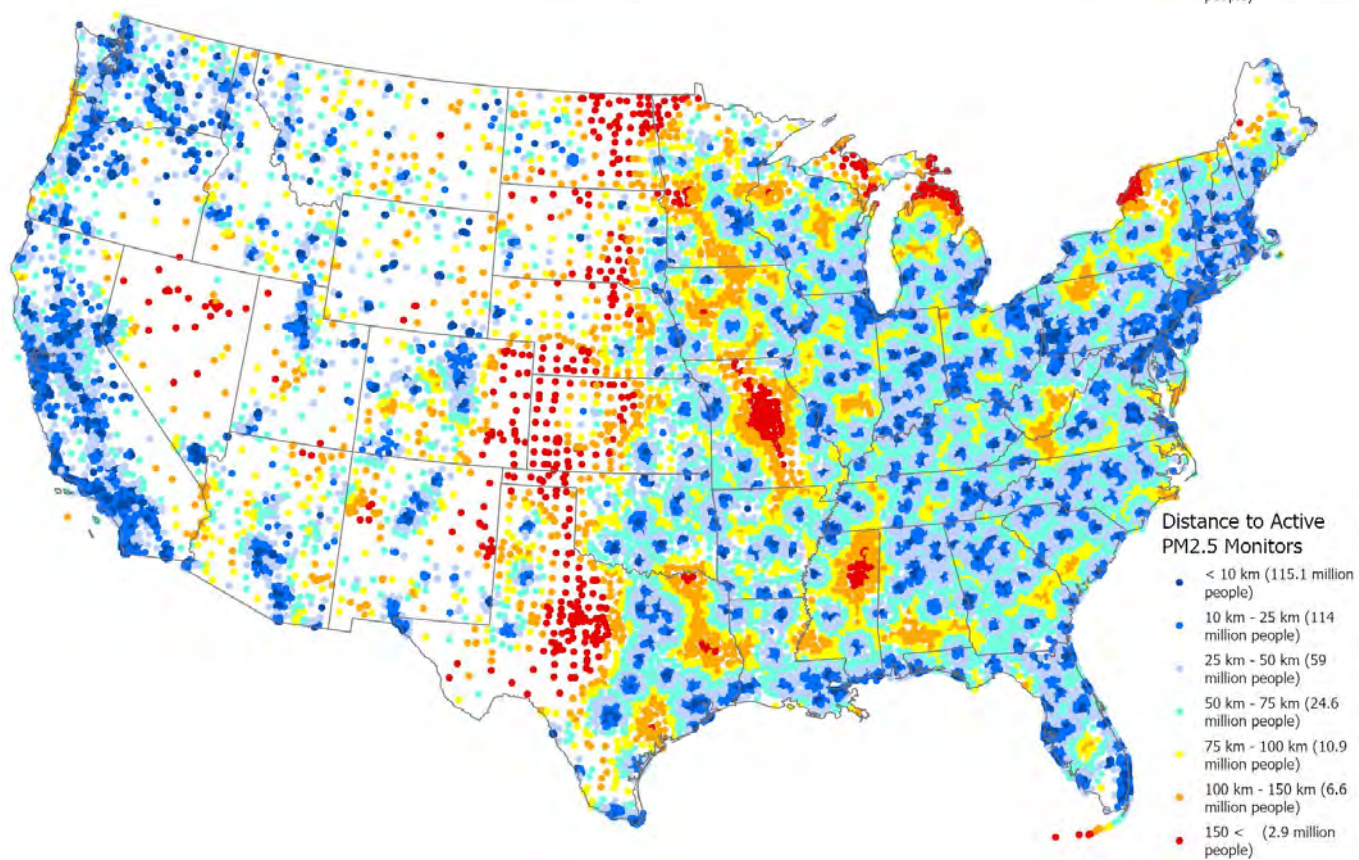
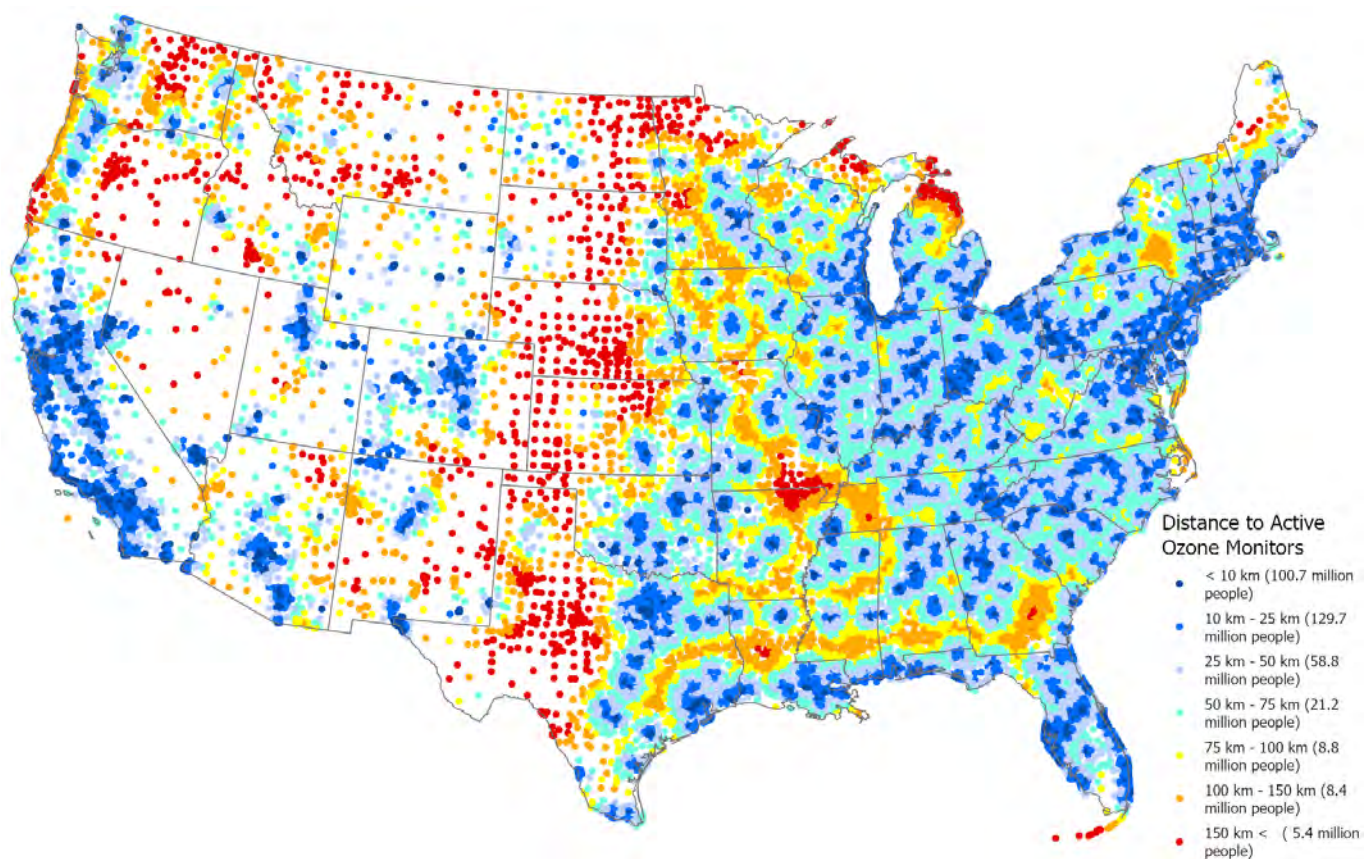
### 2.2.1 Monitoring Networks

The CAA (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<sub>2.5</sub>) in 1997. The NAAQS for PM<sub>2.5</sub> 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 2023, approximately 35 percent of the U.S. population was living within 10 kilometers of ozone and PM<sub>2.5</sub> monitoring sites. Highly populated areas in the eastern U.S. and California are well covered by both ozone and PM<sub>2.5</sub> monitoring network (Figure 2-1).





## **Figure 2-1. Distances from U.S. Census Tract centroids to the nearest monitoring site, 2023.**

In summary, state and local agencies and tribes implement a quality-assured monitoring network to measure air quality across the U.S. The 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 OAQPS 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 EPHT activities.

The advantage of using ambient data from EPA monitoring networks for comparison 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<sub>2.5</sub> day).

To compare/correlate ambient concentrations with acute health effects, daily local air quality data is needed.<sup>4</sup> 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<sub>2.5</sub> is generally measured year-round. Most Federal Reference Method (FRM) PM<sub>2.5</sub> monitors collect data one day in every three days, due in part to the time and costs

---

<sup>4</sup> 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>.

involved in collecting and analyzing the samples. Additionally, continuous monitors have become available which can automatically collect, analyze, and report PM<sub>2.5</sub> measurements on an hourly basis. 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 Federal Equivalent Methods (FEM).

#### ***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 workplace 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<sub>2.5</sub> 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 CMAQ 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<sup>5</sup> is an example of an earlier approach for merging air quality monitor data with CMAQ model predictions. DS 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 DS. 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<sub>2.5</sub> 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<sub>2.5</sub> 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 2-3. 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.	Data are available through state agencies and EPA's AQS. EPA and CDC developed an interagency agreement, where EPA provides air quality data along with statistically combined AQS and CMAQ data, associated metadata, and technical reports that are delivered to CDC.
Estimate the linkage or association of PM <sub>2.5</sub> and ozone on health to: Identify populations that may have higher risk of adverse health effects due to PM <sub>2.5</sub> 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 <sub>2.5</sub> and ozone have been developed for routine indicators. Calculation techniques and presentations for the indicators have been developed.

<sup>5</sup> 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 2-4. 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 FRM)
<ul style="list-style-type: none"> <li>• Number of days with maximum ozone concentration over the NAAQS (or other relevant benchmarks (by county and MSA)</li> <li>• Number of person-days with maximum 8-hr average ozone concentration over the NAAQS &amp; other relevant benchmarks (by county and MSA)</li> </ul>
PM <sub>2.5</sub> (daily 24-hr integrated samples, ug/m <sup>3</sup> , by FRM)
<ul style="list-style-type: none"> <li>• Average ambient concentrations of particulate matter (&lt; 2.5 microns in diameter) and compared to annual PM<sub>2.5</sub> NAAQS (by state).</li> <li>• Percent of population exceeding annual PM<sub>2.5</sub> NAAQS (by state).</li> <li>• Percent of days with PM<sub>2.5</sub> concentration over the daily NAAQS (or other relevant benchmarks (by county and MSA)</li> <li>• Number of person-days with PM<sub>2.5</sub> concentration over the daily NAAQS &amp; other relevant benchmarks (by county and MSA)</li> </ul>

### ***2.3.1 Rationale for the Air Quality Indicators***

The CDC EPHT Network is initially focusing on ozone and PM<sub>2.5</sub>. 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 EPA's 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 for air toxics and criteria air pollutants that represents the year 2020. The platform is based on the 2020 National Emissions Inventory (2020 NEI) published in April 2023 (EPA, 2023) along with other data specific to the year 2020. The air quality modeling platform consists of all the emissions inventories and ancillary data files used for emissions modeling, as well as the meteorological, initial condition, and boundary condition files needed to run the air quality model. This document focuses on the emissions modeling component of the 2020 modeling platform, including the emission inventories, the ancillary data files, and the approaches used to transform inventories for use in air quality modeling.

The modeling platform includes all criteria air pollutants and precursors (CAPs), two groups of hazardous air pollutants (HAPs) and diesel particulate matter. The first group of HAPs are those explicitly used by the chemical mechanism in the Community Multiscale Air Quality (CMAQ) model (Appel, 2018) for ozone/particulate matter (PM): chlorine (Cl), hydrogen chloride (HCl), naphthalene, benzene, acetaldehyde, formaldehyde, and methanol (the last five are abbreviated as NBAFM in subsequent sections of the document). The second group of HAPs consists of 52 HAPs or HAP groups (such as polycyclic aromatic hydrocarbon groups) that are included in CMAQ for the purposes of air quality modeling for a HAP+CAP platform.

Emissions were prepared for the Community Multiscale Air Quality (CMAQ) model (<https://www.epa.gov/cmaq>) version 5.4,<sup>6</sup> which was used to model ozone (O<sub>3</sub>) particulate matter (PM), and HAPs. CMAQ requires hourly and gridded emissions of the following inventory pollutants: carbon monoxide (CO), nitrogen oxides (NO<sub>x</sub>), volatile organic compounds (VOC), sulfur dioxide (SO<sub>2</sub>), ammonia (NH<sub>3</sub>), particulate matter less than or equal to 10 microns (PM<sub>10</sub>), and individual component species for particulate matter less than or equal to 2.5 microns (PM<sub>2.5</sub>). In addition, the Carbon Bond mechanism version 6 (CB6) with chlorine chemistry within CMAQ allows for explicit treatment of the VOC HAPs naphthalene, benzene, acetaldehyde, formaldehyde and methanol (NBAFM), includes anthropogenic HAP emissions of HCl and Cl, and can model additional HAPs as described in Section 3. The short abbreviation for the modeling case name was “2020ha2”, where 2020 is the year modeled, ‘h’ represents that it was based on the 2020 NEI, and ‘a’ represents that it was the first version of a 2020 NEI-based platform. The additional ‘2’ after the ‘ha’ is related to a second run of the 2020ha case with an updated version of some spatial surrogates.

Emissions were also prepared for an air dispersion modeling system: American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD) (EPA, 2018). AERMOD was run for 2020 for all NEI HAPs (about 130 more than covered by CMAQ) across all 50 states, Puerto Rico and the Virgin Islands in a similar way as was done for the 2018 version of AirToxScreen (EPA, 2022a). This TSD focuses on the CMAQ aspects of the 2020 modeling platform from which ozone and PM data were developed for the Centers for Disease Control and Prevention.

---

<sup>6</sup> CMAQ version 5.4: <https://zenodo.org/record/7218076>. CMAQ is also available from the Community Modeling and Analysis System (CMAS) Center at: <http://www.cmascenter.org>.

The effort to create the emission inputs for this study included development of emission inventories to represent emissions during the year of 2020, along with application of emissions modeling tools to convert the inventories into the format and resolution needed by CMAQ and AERMOD.

The emissions modeling platform includes point sources, nonpoint sources, onroad mobile sources, nonroad mobile sources, biogenic emissions and fires for the U.S., Canada, and Mexico. Some platform categories use more disaggregated data than are made available in the NEI. For example, in the platform, onroad mobile source emissions are represented as hourly emissions by vehicle type, fuel type process and road type while the NEI emissions are aggregated to vehicle type/fuel type totals and annual temporal resolution. Emissions used in the CMAQ modeling from Canada are provided by Environment and Climate Change Canada (ECC) and Mexico are mostly provided by SEMARNAT and are not part of the NEI. Year-specific emissions were used for fires, biogenic sources, fertilizer, point sources, and onroad and nonroad mobile sources. Where available, continuous emission monitoring system (CEMS) data were used for electric generating unit (EGU) emissions.

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 4.9 was used to create CMAQ-ready emissions files for a 12-km grid covering the continental U.S. Additional information about SMOKE is available from <http://www.cmascenter.org/smoke>.

The gridded meteorological model used to provide input data for the emissions modeling was developed using the Weather Research and Forecasting Model (WRF, <https://ral.ucar.edu/solutions/products/weather-research-and-forecasting-model-wrf>) version 4.1.1, Advanced Research WRF core (Skamarock, et al., 2008). The WRF Model is a mesoscale numerical weather prediction system developed for both operational forecasting and atmospheric research applications. The WRF was run for 2020 over a domain covering the continental U.S. at a 12km resolution with 35 vertical layers. The run for this platform included high resolution sea surface temperature data from the Group for High Resolution Sea Surface Temperature (GHR SST) (see <https://www.ghrsst.org/>) and is given the EPA meteorological case abbreviation “20k.” The full case abbreviation includes this suffix following the emissions portion of the case name to fully specify the abbreviation of the case as “2020ha2\_cb6\_20k.”

Following the emissions modeling steps to prepare emissions for CMAQ and AERMOD, both models were run for each of the four modeling domains. CMAQ outputs provide the overall mass, chemistry and formation for specific hazardous air pollutants (HAPs) formed secondarily in the atmosphere (e.g., formaldehyde, acetaldehyde, and acrolein), whereas AERMOD provides spatial granularity and more detailed source attribution. CMAQ also provided the biogenic and fire concentrations, as these sources are not run in AERMOD. Special steps were taken to estimate secondary HAPs, fire and biogenic emissions in these areas. The outputs from CMAQ and AERMOD were combined to provide spatially refined concentration estimates for HAPs, from which estimates of cancer and non-cancer risk were derived. Information about the emissions and associated data files for this platform are available from this section of the air emissions modeling website <https://www.epa.gov/air-emissions-modeling/2020-emissions-modeling-platform>.

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. Additional details on the development of the emissions inputs to CMAQ are provided in the publication *Technical Support*

*Document (TSD): Preparation of Emissions Inventories for the 2020 North American Emissions Modeling Platform* (EPA, 2023).

### 3.2 Emission Inventories and Approaches

This section describes the emissions inventories created for input to SMOKE, which are based on the April 2023 version of the 2020 NEI. The NEI includes five main data categories: a) nonpoint 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 collects and stores facility inventory and emissions data for the NEI and 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 2020 NEI Technical Support Document describes in detail the development of the 2020 emission inventories and is available at <https://www.epa.gov/air-emissions-inventories/2020-national-emissions-inventory-nei-technical-support-document-tsd> (EPA, 2023).

A full set of emissions for all source categories is developed every three years, with 2020 being the most recent year represented with a full “triennial” NEI. S/L/T agencies are required to submit all applicable point sources to the NEI in triennial years, including the year 2020. Because all applicable point sources were submitted for 2020, it was not necessary to pull forward unsubmitted sources from another NEI year, as was done for interim years such as 2018 and 2019. The SMARTFIRE2 system and the BlueSky Pipeline (<https://github.com/pnwairfire/bluesky>) emissions modeling system were used to develop year 2020 fire emissions. SMARTFIRE2 categorizes all fires as either prescribed burning or wildfire, and the BlueSky Pipeline system includes fuel loading, consumption and emission factor estimates for both types of fires. Onroad and nonroad mobile source emissions were developed for this project for the year 2020 by running MOVES3 (<https://www.epa.gov/moves>).

With the exception of onroad and fire emissions, Canadian emissions were provided by Environment Canada and Climate Change (ECCC) for the year 2020. For Mexico, inventories from the 2019 emissions modeling platform (EPA, 2022b) were used as the starting point. Adjustments were made to the Canadian and Mexican emissions also include additional adjustments to account for the impacts of the COVID pandemic.

The emissions modeling process was performed using SMOKE v4.9. Through this process, the emissions inventories were apportioned into the grid cells used by CMAQ and temporally allocated into hourly values. In addition, the pollutants in the inventories (e.g., NO<sub>x</sub>, PM and VOC) were split into the chemical species needed by CMAQ. For the purposes of preparing the CMAQ- ready emissions, the NEI emissions inventories by data category were split into emissions modeling platform “sectors”; and emissions from sources other than the NEI were added, such as the Canadian, Mexican, and offshore inventories. Emissions within the emissions modeling platform were separated into sectors for groups of related emissions source categories that are run through all of the appropriate SMOKE programs, except the final merge, independently from emissions categories in the other sectors. The final merge program called Mrggrid combines low-level sector-specific gridded, speciated and temporalized emissions to create the final CMAQ-ready emissions inputs. For biogenic and fertilizer emissions, the CMAQ model allows for these emissions to be included in the CMAQ-ready emissions inputs, or to be computed within

CMAQ itself (the “inline” option). This study used the option to compute biogenic emissions within the model and the CMAQ bidirectional ammonia process to compute the fertilizer emissions.

Table 3-1 presents the sectors in the emissions modeling platform used to develop the year 2020 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 emission summaries for the U.S. sectors are shown in Table 3-2. 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\_20202\_emissions\_totals\_by\_sector.xlsx”.

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

<b>Platform Sector: <i>abbreviation</i></b>	<b>NEI Data Category</b>	<b>Description and resolution of the data input to SMOKE</b>
<b>EGU units: <i>Ptegu</i></b>	Point	2020 NEI point source EGUs, replaced with hourly Continuous Emissions Monitoring System (CEMS) values for NO <sub>x</sub> and SO <sub>2</sub> , and the remaining pollutants temporally allocated according to CEMS heat input where the units are matched to the NEI. Emissions for all sources not matched to CEMS data come from 2020 NEI point inventory. Annual resolution for sources not matched to CEMS data, hourly for CEMS sources. EGUs closed in 2020 are not part of the inventory.
<b>Point source oil and gas: <i>pt_oilgas</i></b>	Point	2020 NEI point sources that include oil and gas production emissions processes for facilities with North American Industry Classification System (NAICS) codes related to Oil and Gas Extraction, Natural Gas Distribution, Drilling Oil and Gas Wells, Support Activities for Oil and Gas Operations, Pipeline Transportation of Crude Oil, and Pipeline Transportation of Natural Gas. Includes U.S. offshore oil production.
<b>Aircraft and ground support equipment: <i>airports</i></b>	Point	2020 NEI point source emissions from airports, including aircraft and airport ground support emissions. Annual resolution.
<b>Remaining non-EGU point: <i>Ptnonipm</i></b>	Point	All 2020 NEI point source records not matched to the airports, ptegu, or pt_oilgas sectors. Includes 2020 NEI rail yard emissions. Annual resolution.
<b>Livestock: <i>Livestock</i></b>	Nonpoint	2020 NEI nonpoint livestock emissions. Livestock includes ammonia and other pollutants (except PM <sub>2.5</sub> ). County and annual resolution.
<b>Agricultural Fertilizer: <i>fertilizer</i></b>	Nonpoint	2020 agricultural fertilizer ammonia emissions computed inline within CMAQ.
<b>Area fugitive dust: <i>afdust_adj</i></b>	Nonpoint	PM <sub>10</sub> and PM <sub>2.5</sub> fugitive dust sources from the 2020 NEI nonpoint inventory; including building construction, road construction, agricultural dust, and paved and unpaved road dust. The emissions modeling system applies a transport fraction reduction and a zero-out based on 2020 gridded hourly meteorology (precipitation and snow/ice cover). Emissions are county and annual resolution.

<b>Platform Sector: <i>abbreviation</i></b>	<b>NEI Data Category</b>	<b>Description and resolution of the data input to SMOKE</b>
<b>Biogenic:</b> <i>beis</i>	Nonpoint	Year 2020 emissions from biogenic sources. These were left out of the CMAQ-ready merged emissions, in favor of inline biogenic emissions produced during the CMAQ model run itself. Version 4 of the Biogenic Emissions Inventory System (BEIS) was used with Version 6 of the Biogenic Emissions Landuse Database (BELD6). Therefore, the biogenic emissions used here are similar to the 2020 NEI biogenic emissions, but not exactly the same.
<b>Category 1, 2 CMV:</b> <i>cmv_c1c2</i>	Nonpoint	2020 NEI Category 1 (C1) and Category 2 (C2), commercial marine vessel (CMV) emissions based on Automatic Identification System (AIS) data. Point and hourly resolution.
<b>Category 3 CMV:</b> <i>cmv_c3</i>	Nonpoint	2020 NEI Category 3 (C3) commercial marine vessel (CMV) emissions based on AIS data. Point and hourly resolution.
<b>Locomotives :</b> <i>Rail</i>	Nonpoint	Line haul rail locomotives emissions from 2020 NEI. County and annual resolution.
<b>Nonpoint source oil and gas:</b> <i>np_oilgas</i>	Nonpoint	Nonpoint 2020 NEI sources from oil and gas-related processes. County and annual resolution.
<b>Residential Wood Combustion:</b> <i>rwc</i>	Nonpoint	2020 NEI nonpoint sources with residential wood combustion (RWC) processes. County and annual resolution.
<b>Solvents:</b> <i>np_solvents</i>	Nonpoint	Emissions of solvents from the 2020 NEI (Seltzer, 2021). Includes household cleaners, personal care products, adhesives, architectural and aerosol coatings, printing inks, and pesticides. Annual and county resolution.
<b>Remaining nonpoint:</b> <i>nonpt</i>	Nonpoint	2020 NEI nonpoint sources not included in other platform sectors. County and annual resolution.
<b>Nonroad:</b> <i>nonroad</i>	Nonroad	2020 NEI nonroad equipment emissions developed with MOVES3, including the updates made to spatial apportionment that were developed with the 2016v1 platform. MOVES3 was used for all states except California, which submitted their own emissions for the 2020 NEI. County and monthly resolution.
<b>Onroad:</b> <i>onroad</i>	Onroad	Onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles from 2020 NEI. Includes the following emission processes: exhaust, extended idle, auxiliary power units, evaporative, permeation, refueling, vehicle starts, off network idling, long-haul truck hoteling, and brake and tire wear. MOVES3 was run for 2020 to generate emission factors.
<b>Onroad California:</b> <i>onroad_ca_adj</i>	Onroad	California-provided 2020 CAP and HAP (VOCs and metals) onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles based on Emission Factor (EMFAC), gridded and temporalized based on outputs from MOVES3. Polycyclic aromatic hydrocarbon (PAH) emissions are based on MOVES3.

<b>Platform Sector: <i>abbreviation</i></b>	<b>NEI Data Category</b>	<b>Description and resolution of the data input to SMOKE</b>
<b>Point source agricultural fires:</b> <i>ptagfire</i>	Nonpoint	Agricultural fire sources for 2020 developed by EPA as point and day-specific emissions. <sup>7</sup> Only EPA-developed ag. fire data are used in this study, thus 2020 NEI state submissions are not included. Agricultural fires are in the nonpoint data category of the NEI, but in the modeling platform, they are treated as day-specific point sources. Updated HAP-augmentation factors were applied.
<b>Point source prescribed fires:</b> <i>ptfire-rx</i>	Nonpoint	Point source day-specific prescribed fires for 2020 NEI computed using SMARTFIRE 2 and BlueSky Pipeline. The ptfire emissions were run as two separate sectors: ptfire-rx (prescribed, including Flint Hills / grasslands) and ptfire-wild.
<b>Point source wildfires:</b> <i>ptfire-wild</i>	Nonpoint	Point source day-specific wildfires for 2020 NEI computed using SMARTFIRE 2 and BlueSky Pipeline.
<b>Non-US. Fires:</b> <i>ptfire_othna</i>	N/A	Point source day-specific wildfires and agricultural fires outside of the U.S. for 2020. Canadian fires for May through December are provided by ECCC. All other fire emissions, including Canadian emissions from January through April, as well as Mexico, Caribbean, Central American, and other international fires, are from v2.5 of the Fire INventory (FINN) from National Center for Atmospheric Research (Wiedinmyer, C., 2023).
<b>Canada Area Fugitive dust sources:</b> <i>canada_afdust</i>	N/A	Area fugitive dust sources from ECCC for 2020 with transport fraction and snow/ice adjustments based on 2020 meteorological data. Annual and province resolution.
<b>Canada Point Fugitive dust sources:</b> <i>canada_ptdust</i>	N/A	2020 point source fugitive dust sources from ECCC with transport fraction and snow/ice adjustments based on 2020 meteorological data. Monthly and province resolution.
<b>Canada and Mexico stationary point sources:</b> <i>canmex_point</i>	N/A	Canada and Mexico point source emissions not included in other sectors. Canada point sources for 2020 were provided by ECCC and Mexico point source emissions for 2016 were provided by SEMARNAT. Mexico sources were projected from 2019ge (EPA, 2022b) with COVID adjustments applied. Canada monthly temporalization adjusted for COVID. Annual and monthly resolution.
<b>Canada and Mexico agricultural sources:</b> <i>canmex_ag</i>		Canada and Mexico agricultural emissions. Canada point sources for 2020 were provided by ECCC and Mexico emissions for 2016 were provided by SEMARNAT and adjusted to 2019. COVID adjustments were not applied to the ag sector. Annual resolution.
<b>Canada low-level oil and gas sources:</b> <i>canada_og2D</i>		2020 Canada emissions from upstream oil and gas. This sector contains the portion of oil and gas emissions which are not subject to plume rise. The rest of the 2020 Canada oil and gas emissions are in the canmex_point sector. Provided by ECCC with COVID-adjusted monthly temporalization. Monthly resolution.

<sup>7</sup> Only EPA-developed agricultural fire data were included in this study; data submitted by states to the NEI were excluded.



<b>Platform Sector: <i>abbreviation</i></b>	<b>NEI Data Category</b>	<b>Description and resolution of the data input to SMOKE</b>
<b>Canada and Mexico nonpoint and nonroad sources:</b> <i>canmex_area</i>	N/A	2020 Canada and Mexico nonpoint source emissions not included in other sectors. Canada: ECCC provided a 2020 inventory and surrogates. Mexico: applied COVID adjustments to 2019ge. Monthly temporalization adjusted for COVID.
<b>Canada onroad sources:</b> <i>canada_onroad</i>	N/A	Canada onroad emissions. 2020 Canada inventory provided by ECCC and processed using updated surrogates. COVID impacts applied to monthly profiles (not to annual totals). Province and monthly resolution.
<b>Mexico onroad sources:</b> <i>mexico_onroad</i>	N/A	Mexico onroad emissions. 2020 MOVES-Mexico with COVID adjustments applied. Municipio and monthly resolution.

Ocean chlorine emissions were also merged in with the above sectors. The ocean chlorine gas emission estimates are based on the build-up of molecular chlorine (Cl<sub>2</sub>) concentrations in oceanic air masses (Bullock and Brehme, 2002). Ocean chlorine data at 12 km resolution were available from earlier studies and were not modified other than the name “CHLORINE” was changed to “CL2” because that is the name required by the CMAQ model.

The emission inventories in SMOKE input formats for the platform are available from EPA’s Air Emissions Modeling website: <https://www.epa.gov/air-emissions-modeling/2020-emissions-modeling-platform>. The platform informational text file indicates the particular zipped files associated with each platform sector. Some emissions data summaries are available with the data files for the 2020 platform. The types of reports include state summaries of inventory pollutants and model species by modeling platform sector and county annual totals by modeling platform sector. Summaries of the emissions in the Contiguous U.S. and emissions within the 12-km domain but outside of the U.S. are shown in Table 3-2. 2020 Contiguous United States Emissions by Sector (tons/yr in 48 states + D.C.) Table 3-2 and Table 3-3, respectively.

**Table 3-2. 2020 Contiguous United States Emissions by Sector (tons/yr in 48 states + D.C.)**

<b>Sector</b>	<b>CO</b>	<b>NH3</b>	<b>NOX</b>	<b>PM10</b>	<b>PM2_5</b>	<b>SO2</b>	<b>VOC</b>
afdust_adj				5,513,981	765,892		
airports	324,335	0	81,729	8,295	7,334	8,889	48,680
cmv_c1c2	17,242	57	113,213	3,051	2,956	571	3,973
cmv_c3	9,216	29	91,850	1,640	1,508	3,690	4,233
fertilizer		1,401,045					
livestock		2,693,568					215,483
nonpt	2,199,000	145,244	739,200	724,647	634,164	107,619	1,007,035
nonroad	11,005,619	1,980	866,081	85,040	79,961	990	977,863
np_oilgas	621,795	16	571,317	10,541	10,453	135,998	2,583,242
np_solvents							2,586,519
onroad	14,063,910	89,328	2,327,115	188,720	78,626	9,785	1,030,292
ptegu	400,900	21,491	847,682	101,118	86,781	820,839	25,466
ptagfire	664,858	140,954	28,037	102,245	66,604	11,025	107,166
ptfire-rx	7,181,506	114,977	140,674	794,163	681,777	64,751	1,654,719
ptfire-wild	18,664,856	306,009	239,530	1,885,536	1,597,986	135,617	4,399,094
ptnonipm	1,157,963	63,289	769,850	343,959	222,800	443,029	705,590
pt_oilgas	171,082	8,264	330,517	12,668	12,168	35,130	196,102
rail	92,100	282	422,975	10,819	10,459	351	17,492
rcw	2,955,189	22,735	44,869	450,864	448,073	12,019	455,660
beis	3,265,206		980,749				28,254,267
<b>CONUS no beis</b>	<b>59,529,571</b>	<b>5,009,270</b>	<b>7,614,637</b>	<b>10,237,288</b>	<b>4,707,543</b>	<b>1,790,303</b>	<b>16,018,609</b>
<b>CONUS + beis</b>	<b>62,794,777</b>	<b>5,009,270</b>	<b>8,595,386</b>	<b>10,237,288</b>	<b>4,707,543</b>	<b>1,790,303</b>	<b>44,272,876</b>

**Table 3-3. Non-US Emissions by Sector within the 12US1 Modeling Domain (tons/yr for Canada, Mexico, Offshore)**

<b>Sector</b>	<b>CO</b>	<b>NH3</b>	<b>NOX</b>	<b>PM10</b>	<b>PM2_5</b>	<b>SO2</b>	<b>VOC</b>
Canada ag		495,216		6,567	1,876		124,394
Canada oil and gas 2D		8					318,720
Canada afdust				799,628	154,654		
Canada ptdust				2,791	361		
Canada area	2,020,228	5,987	321,437	184,241	135,848	14,263	709,347
Canada onroad	1,622,797	6,848	354,849	24,288	13,272	830	115,863
Canada point	1,011,453	18,160	549,975	111,671	41,376	499,692	146,194
Canada fires	654,404	8,746	10,058	118,455	102,005	5,444	215,854
Canada cmv_c1c2	2,596	8	16,691	441	428	60	580
Canada cmv_c3	7,160	19	71,623	1,051	967	2,167	3,497
Mexico ag		115,994		66,380	14,465		0
Mexico area	115,014	81	55,083	29,228	16,992	1,586	278,327
Mexico onroad	1,241,148	2,130	311,807	11,557	8,144	4,888	110,159
Mexico point	124,965	949	144,798	39,649	27,670	293,438	29,882
Mexico fires	211,379	3,612	13,079	24,985	21,413	2,000	109,543



Sector	CO	NH3	NOX	PM10	PM2_5	SO2	VOC
Mexico cmv_c1c2	118	0	766	20	19	2	32
Mexico cmv_c3	7,375	72	79,149	4,088	3,761	10,888	3,442
Offshore cmv_c1c2	3,647	11	23,290	610	591	64	885
Offshore cmv_c3	43,133	254	434,674	14,334	13,187	36,361	20,624
Offshore pt_oilgas	52,008	8	50,096	638	637	463	38,910
<b>Can/Mex/offshore total</b>	<b>7,117,423</b>	<b>658,106</b>	<b>2,437,376</b>	<b>1,440,620</b>	<b>557,665</b>	<b>872,147</b>	<b>2,226,254</b>

### 3.2.1 Point Sources (*ptegu*, *pt\_oilgas*, *ptnonipm*, and *airports*)

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 CMV emissions (*cmv\_c1c2* and *cmv\_c3* sectors) are processed by SMOKE as point source inventories and are discussed later in this section. A complete NEI is developed every three years. At the time of this writing, 2020 is the most recently finished complete NEI. A comprehensive description about the development of the 2020 NEI is available in the 2020 NEI TSD (EPA, 2023). Point inventories are also available in EIS for non-triennial NEI years such as 2019 and 2021. In the interim year point inventories, states are required to update larger sources with the emissions that occurred in that year, while sources not updated by states for the interim year were either carried forward from the most recent triennial NEI or marked as closed and removed.

In preparation for modeling, the complete set of point sources in the NEI was exported from EIS for the year 2020 into the Flat File 2010 (FF10) format that is compatible with SMOKE (see <https://cmascenter.org/smoke/documentation/4.9/html/ch06s02s08.html>) and was then split into several sectors for modeling. For both flat files, sources without specific locations (i.e., the FIPS code ends in 777) were dropped and inventories for the other point source sectors were created from the remaining point sources. The point sectors are: EGUs (*ptegu*), point source oil and gas extraction-related sources (*pt\_oilgas*), airport emissions (*airports*), and the remaining non-EGUs (*ptnonipm*). The EGU emissions were 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*) and airport emissions (*airports*) were processed separately for the purposes of developing emissions summaries and due to distinct projection techniques from the remaining non-EGU emissions (*ptnonipm*), although this study does not include emissions projected to other years.

In some cases, data about facility or unit closures are entered into EIS after the inventory modeling inventory flat files have been extracted. EIS. Prior to processing through SMOKE, submitted facility and unit closures were reviewed and where closed sources were found in the inventory, those were removed.

For the 2020 platform, an analysis of point source stack parameters (e.g., stack height, diameter, temperature, and velocity) was performed due to the presence of unrealistic and repeated stack parameters as default values were noticed. The defaulted values were noticed in data submissions for the states of Illinois, Louisiana, Michigan, Pennsylvania, Texas, and Wisconsin. Where these defaults were detected and deemed to be unreasonable for the specific process, the affected stack parameters were replaced by

values from the PSTK file that is input to SMOKE. PSTK contains default stack parameters by source classification code (SCC). These updates impacted the ptnonipm and pt\_oilgas inventories.

The inventory pollutants processed through SMOKE for input to CMAQ for the ptegu, pt\_oilgas, ptnonipm, and airports sectors included: CO, NO<sub>x</sub>, VOC, SO<sub>2</sub>, NH<sub>3</sub>, PM<sub>10</sub>, and PM<sub>2.5</sub> and the following HAPs: HCl (pollutant code = 7647010), Cl (code = 7782505), and several dozen other HAPs listed in Section 3. NBAFM pollutants from the point sectors were utilized. For AERMOD, additional HAPs were included as described in the 2020 AirToxScreen TSD.

The ptnonipm, pt\_oilgas, and airports sector emissions were provided to SMOKE as annual emissions. For sources in the ptegu sector that could be matched to 2020 CEMS data, hourly CEMS NO<sub>x</sub> and SO<sub>2</sub> emissions for 2020 from EPA's Acid Rain Program were used rather than annual inventory emissions. For all other pollutants (e.g., VOC, PM<sub>2.5</sub>, 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 used as inputs to SMOKE as annual emissions. The full description of how the NEI emissions were developed is provided in the NEI documentation - 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, 2020, ...].
- 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 stored and applied 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. Data for airports and rail yards were incorporated.
- 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.

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 2020 point source inventory that could be matched to units found in the National Electric Energy Database System (NEEDS) v6 that is used by the Integrated Planning Model (IPM) to develop projected 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 were placed into the pt\_oilgas or ptnonipm sectors. For studies that include analytic years, 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 analytic year modeling scenarios if emissions for units projected by IPM are not properly matched to the units in the base year point source inventory.

The 2020 ptegu emissions inventory is a subset of the point source flat file exported from the Emissions Inventory System (EIS). In the point source flat file, emission records for sources that have been matched to the NEEDS database have a value filled into the ipm\_yn column based on the matches stored within EIS. Thus, unit-level emissions were split into a separate EGU flat file for units that have a populated (non-null) ipm\_yn field. A populated ipm\_yn field indicates that a match was found for the EIS unit in the NEEDS v6 database. Updates were made to the flat file output from EIS as follows:

- ORIS facility and unit identifiers were updated based on additional matches in a cross-platform spreadsheet, based on state comments, and using the EIS alternate identifiers table as described later in this section.

Some units in the ptegu sector are matched to Continuous Emissions Monitoring System (CEMS) data via Office of Regulatory Information System (ORIS) facility codes and boiler IDs. For the matched units, the annual emissions of NO<sub>x</sub> and SO<sub>2</sub> in the flat file were replaced with the hourly CEMS emissions in base year modeling. For other pollutants at matched units, the hourly CEMS heat input data were used to allocate the NEI annual emissions to hourly values. All stack parameters, stack locations, and Source Classification Codes (SCC) for these sources come from the flat file. If CEMS data exists for a unit, but the unit is not matched to the NEI, the CEMS data for that unit were not used in the modeling platform. However, if the source exists in the NEI and is not matched to a CEMS unit, the emissions from that source are still modeled using the annual emission value in the NEI temporally allocated to hourly values.

EIS stores many matches from NEI units to the ORIS facility codes and boiler IDs used to reference the CEMS data. In the flat file, emission records for point sources matched to CEMS data have values filled into the ORIS\_FACILITY\_CODE and ORIS\_BOILER\_ID columns. The CEMS data are available at <https://campd.epa.gov/data>. Many smaller emitters in the CEMS program cannot be matched to the NEI due to differences in the way a unit is defined between the NEI and CEMS datasets, or due to uncertainties in source identification such as inconsistent plant names in the two data systems. In addition, the NEEDS database of units modeled by IPM includes many smaller emitting EGUs that do not have CEMS. Therefore, there will be more units in the ptegu sector than have CEMS data.

Matches from the NEI to ORIS codes and the NEEDS database were improved in the platform where applicable. In some cases, NEI units in EIS match to many CAMD units. In these cases, a new entry was made in the flat file with a “\_M\_” in the ipm\_yn field of the flat file to indicate that there are “multiple” ORIS IDs that match that unit. This helps facilitate appropriate temporal allocation of the emissions by SMOKE. Temporal allocation for EGUs is discussed in more detail in the Ancillary Data section below.

The EGU flat file was split into two flat files: those that have unit-level matches to CEMS data using the oris\_facility\_code and oris\_boiler\_id fields and those that do not so that different temporal profiles could be applied. In addition, the hourly CEMS data were processed through v2.1 of the CEMCorrect tool to mitigate the impact of unmeasured values in the data.

### 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. The pt\_oilgas sector includes emissions from offshore oil platforms. Where available, the point source emissions submitted as part of the 2020 NEI process were used. More information on the development of the 2020 NEI oil and gas emissions can be found in Section 13 of the 2020 NEI TSD.

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

<b>NAICS</b>	<b>NAICS description</b>
2111	Oil and Gas Extraction
211112	Natural Gas Liquid Extraction
21112	Crude Petroleum Extraction
211120	Crude Petroleum Extraction
21113	Natural Gas Extraction
211130	Natural Gas Extraction
213111	Drilling Oil and Gas Wells
213112	Support Activities for Oil and Gas Operations
2212	Natural Gas Distribution
22121	Natural Gas Distribution
221210	Natural Gas Distribution
237120	Oil and Gas Pipeline and Related Structures Construction
4861	Pipeline Transportation of Crude Oil
48611	Pipeline Transportation of Crude Oil
486110	Pipeline Transportation of Crude Oil
4862	Pipeline Transportation of Natural Gas
48621	Pipeline Transportation of Natural Gas
486210	Pipeline Transportation of Natural Gas

### **3.2.1.3 Airports Sector (*airports*)**

Emissions at airports were separated from other sources in the point inventory based on sources that have the facility source type of 100 (airports). The airports sector includes all aircraft types used for public, private, and military purposes and aircraft ground support equipment. The Federal Aviation Administration's (FAA) Aviation Environmental Design Tool (AEDT) is used to estimate emissions for this sector. Additional information about aircraft emission estimates can be found in section 3 of the 2020 NEI TSD. EPA used airport-specific factors where available. Airport emissions were spread out into multiple 12km grid cells when the airport runways were determined to overlap multiple grid cells. Otherwise, airport emissions for a specific airport are confined to one air quality model grid cell.

### **3.2.1.4 Non-IPM Sector (*ptnonipm*)**

With some exceptions, the *ptnonipm* sector contains the point sources that are not in the *ptegu*, *pt\_oilgas*, or airports sectors. For the most part, the *ptnonipm* sector reflects non-EGU emissions sources and rail yards. However, it is possible that some low-emitting EGUs not matched to units 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.

The *ptnonipm* sources (i.e., not EGUs and non -oil and gas sources) were used as-is from the 2020 NEI point inventory. Solvent emissions from point sources were removed from the *np\_solvents* sector to prevent double-counting, so that all point sources can be retained in the modeling as point sources rather than as area sources. The modeling was based the point flat file exported from EIS on January 28, 2023 with edits made through April 14, 2023 that included corrections to how the selection was implemented in EIS, updates from the state/local review, and updates specific to ethylene oxide.

Emissions from rail yards are included in the *ptnonipm* sector. Railyards are from the 2020 NEI railyard inventory. Additional information about railyard estimates can be found in section 3 of the 2020 NEI TSD.

### **3.2.3 Nonpoint Sources (*afdust*, *ag*, *nonpt*, *np\_oilgas*, *rwc*)**

This section describes the *stationary* nonpoint sources in the NEI nonpoint data category. Locomotives, C1 and C2 CMV, and C3 CMV are included in the NEI nonpoint data category but are mobile sources that are described in Section 2.4. The 2020 NEI TSD includes documentation for the nonpoint data.

Nonpoint tribal emissions submitted to the NEI are dropped during spatial processing with SMOKE due to the configuration of the spatial surrogates. Part of the reason for this is to prevent possible double-counting with county-level emissions and also because spatial surrogates for tribal data are not currently available. These omissions are not expected to have an impact on the results of the air quality modeling at the 12-km resolution used for this platform.

The following subsections describe how the sources in the NEI nonpoint inventory were separated into modeling platform sectors, along with any data that were updated (replaced) with non-NEI data.

### **3.2.3.1      *Area Fugitive Dust Sector (afdust)***

The area-source fugitive dust (afdust) sector contains PM<sub>10</sub> and PM<sub>2.5</sub> emission estimates for nonpoint SCCs identified by EPA 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 was separated from other nonpoint sectors to allow for the application of a “transport fraction,” and meteorological/precipitation reductions. These adjustments were applied using a script that applies land use-based gridded transport fractions based on landscape roughness, 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 were 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). Both the transport fraction and meteorological adjustments were based on the gridded resolution of the platform (i.e., 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 were accounted for in the subsequent meteorological adjustment.

Paved road dust emissions were from the 2020 NEI. For the fugitive dust emissions compiled into the 2020 NEI, meteorological adjustments were 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 2020 platform, the meteorological adjustments that were applied in the NEI to paved and unpaved road SCCs were backed out and reapplied in SMOKE at an hourly resolution for each grid cell. 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 2020 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 Livestock Sector (livestock)***

The livestock emissions in this sector are based only on the SCCs starting with 2805. The livestock emissions 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 sector does not include quite all of the livestock NH<sub>3</sub> emissions, as there is a very small amount of NH<sub>3</sub> emissions from livestock

in the point source inventory (as point sources). In addition to  $\text{NH}_3$ , the sector includes livestock emissions from all pollutants other than  $\text{PM}_{2.5}$ .  $\text{PM}_{2.5}$  from livestock are in the dust sector.

Agricultural livestock emissions in the 2020 platform were from the 2020 NEI, which is a mix of state-submitted data and EPA estimates. Livestock emissions utilized improved animal population data. VOC livestock emissions, new for this sector, were estimated by multiplying a national VOC/ $\text{NH}_3$  emissions ratio by the county  $\text{NH}_3$  emissions. The 2020 NEI 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 2020 U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) survey. Details on the approach are provided in Section 10 of the 2020 NEI TSD.

### **3.2.3.3 Agricultural Fertilizer Sector (fertilizer)**

As described in the 2020 NEI TSD, fertilizer emissions for 2020 were based on the FEST-C model. As described in the 2020 NEI TSD, fertilizer emissions for 2020 were based on the FEST-C model (<https://www.cmascenter.org/fest-c/>). Unlike most of the other emissions input to the CMAQ model, fertilizer emissions are computed during a run of CMAQ in bi-directional mode and are output during the model run. The bidirectional version of CMAQ (v5.3) and the Fertilizer Emissions Scenario Tool for CMAQ FEST-C (v1.3) were used to estimate ammonia ( $\text{NH}_3$ ) emissions from agricultural soils. The computed emissions were saved during the CMAQ run so they can be included in emissions summaries and in other model runs that do not use the bidirectional method.

FEST-C is the software program that processes land use and agricultural activity data to develop inputs for the CMAQ model when run with bidirectional exchange. FEST-C reads land use data from the Biogenic Emissions Landuse Dataset (BELD), meteorological variables from the Weather Research and Forecasting (WRF) model, and nitrogen deposition data from a previous or historical average CMAQ simulation. FEST-C, then uses the Environmental Policy Integrated Climate (EPIC) modeling system (<https://epicapex.tamu.edu/epic/>) to simulate the agricultural practices and soil biogeochemistry and provides information regarding fertilizer timing, composition, application method and amount.

An iterative calculation was applied to estimate fertilizer emissions. First, fertilizer application by crop type was estimated using FEST-C modeled data. To develop the NEI emissions, CMAQ v5.4 was run with the Surface Tiled Aerosol and Gaseous Exchange (STAGE) deposition option along with bidirectional exchange to estimate fertilizer and biogenic  $\text{NH}_3$  emissions. However, for this study, the M3DRY option was used to develop the fertilizer emissions.

The following activity parameters were input into the EPIC model:

- Grid cell meteorological variables from WRF
- Initial soil profiles/soil selection
- Presence of 21 major crops: irrigated and rain fed hay, alfalfa, grass, barley, beans, grain corn, silage corn, cotton, oats, peanuts, potatoes, rice, rye, grain sorghum, silage sorghum, soybeans, spring wheat, winter wheat, canola, and other crops (e.g., lettuce, tomatoes, etc.)
- Fertilizer sales to establish the type/composition of nutrients applied
- Management scenarios for the 10 USDA production regions. These include irrigation, tile drainage, intervals between forage harvest, fertilizer application method (injected versus surface applied), and equipment commonly used in these production regions.



The WRF meteorological model was used to provide grid cell meteorological parameters for year 2020 using a national 12-km rectangular grid covering the continental U.S. Initial soil nutrient and pH conditions in EPIC were based on the 1992 USDA Soil Conservation Service (CSC) Soils-5 survey. The EPIC model then was run for 25 years using current fertilization and agricultural cropping techniques to estimate soil nutrient content and pH for the 2017 EPIC/WRF/CMAQ simulation.

The presence of crops in each model grid cell was determined using USDA Census of Agriculture data (2012) and USGS National Land Cover data (2011). These two data sources were used to compute the fraction of agricultural land in a model grid cell and the mix of crops grown on that land.

Fertilizer sales data and the 6-month period in which they were sold were extracted from the 2014 Association of American Plant Food Control Officials (AAPFCO, <http://www.aapfco.org/publications.html>). AAPFCO data were used to identify the composition (e.g., urea, nitrate, organic) of the fertilizer used, and the amount applied is estimated using the modeled crop demand. These data were useful in making a reasonable assignment of what kind of fertilizer is being applied to which crops.

Management activity data refers to data used to estimate representative crop management schemes. The USDA Agricultural Resource Management Survey (ARMS, [https://www.nass.usda.gov/Surveys/Guide\\_to\\_NASS\\_Surveys/Ag\\_Resource\\_Management/](https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Ag_Resource_Management/)) was used to provide management activity data. These data cover 10 USDA production regions and provide management schemes for irrigated and rain fed hay, alfalfa, grass, barley, beans, grain corn, silage corn, cotton, oats, peanuts, potatoes, rice, rye, grain sorghum, silage sorghum, soybeans, spring wheat, winter wheat, canola, and other crops (e.g., lettuce, tomatoes, etc.).

#### **3.2.3.4      *Nonpoint Oil-gas Sector (np\_oilgas)***

The nonpoint oil and gas (np\_oilgas) sector includes onshore and offshore oil and gas emissions. The EPA estimated emissions for all counties with 2020 oil and gas activity data with 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. Because of the importance of emissions from this sector, special consideration is given to the speciation, spatial allocation, and monthly temporalization of nonpoint oil and gas emissions, instead of relying on older, more generalized profiles.

The 2020 NEI version of the Nonpoint Oil and Gas Emission Estimation Tool (i.e., the “NEI oil and gas tool”) was used to estimate 2020. Year 2020 oil and gas activity data obtained from Enverus’ activity database ([www.enverus.com](http://www.enverus.com)) and supplied by some state air agencies. The NEI oil and gas tool is an Access database that utilizes county-level activity data (e.g., oil production and well counts), operational characteristics (types and sizes of equipment), and emission factors to estimate emissions. The tool was used to create a CSV-formatted emissions dataset covering all national nonpoint oil and gas emissions. This dataset was converted to the FF10 format for use in SMOKE modeling. More details on the inputs for and running of the tool for 2020 are provided in the 2020 NEI TSD.

A new source was added to the oil and gas sector for the 2020 NEI. Pipeline Blowdowns and Pigging (SCC= 2310021801) emissions were estimated using US EPA Greenhouse Gas Reporting Program (GHGRP) data. These Pipeline Blowdowns and Pigging emissions included county-level estimates of



VOC, benzene, toluene, ethylbenzene, and xylene (BTEX). These emissions estimates were calculated outside of the Oil and Gas Tool and submitted to EIS separately from the Oil and Gas Tool emissions. These emissions were considered EPA default emissions and SLTs had the opportunity to submit their own Pipeline Blowdowns and Pigging (e.g., Utah) emissions and/or accept/omit these emissions using the Nonpoint Survey. Unfortunately, these EPA default Pipeline Blowdowns and Pigging emissions did not get into the 2020 NEI release for the states that accepted these emissions due to EIS tagging issues. These emissions were included in this 2020 Emissions Modeling Platform.

Lastly, EPA and the state of New Mexico worked together to exercise the point source subtraction step in the Oil and Gas Tool during the 2020 NEI development period. This point source subtraction step was used for New Mexico because additional oil and gas point sources were submitted by New Mexico that were the same processes that are estimated in the Oil and Gas Tool (non-point sources). This point source subtraction step is a process used to eliminate possible double counting of sources in the Oil and Gas Tool that are already defined in the point source inventory. Unfortunately, the resulting non-point emissions from the point source subtraction step for New Mexico did not get into the 2020 NEI release due to EIS tagging issues. New Mexico non-point oil and gas emissions are overestimated in the 2020 NEI as a result. This overestimation was corrected for this 2020 Emissions Modeling Platform.

#### **3.2.3.5 Residential Wood Combustion Sector (*rw*)**

The residential wood combustion (*rw*) 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 chimeneas. 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.

The 2020 platform RWC emissions are unchanged from the data in the 2020 NEI and include some improvements to RWC emissions estimates developed as part of the 2020 NEI process. The EPA, along with the Commission on Environmental Cooperation (CEC), the Northeast States for Coordinated Air Use Management (NESCAUM), and Abt Associates, conducted a national survey of wood-burning activity in 2018. The results of this survey were used to estimate county-level burning activity data. The activity data for RWC processes is the amount of wood burned in each county, which is based on data from the CEC survey on the fraction of homes in each county that use each wood-burning appliance and the average amount of wood burned in each appliance. These assumptions are used with the number of occupied homes in each county to estimate the total amount of wood burned in each county, in cords for cordwood appliances and tons for pellet appliances. Cords of wood are converted to tons using county-level density factors from the U.S. Forest Service. RWC emissions were calculated by multiplying the tons of wood burned by emissions factors. For more information on the development of the residential wood combustion emissions, see Section 27 of the 2020 NEI TSD.

#### **3.2.3.6 Solvents (*np\_solvents*)**

The *np\_solvents* sector is a diverse collection of emission sources for which emissions are driven by evaporation. Included in this sector are everyday items, such as cleaners, personal care products, adhesives, architectural and aerosol coatings, printing inks, and pesticides. These sources exclusively emit

organic gases and feature origins spanning residential, commercial, institutional, and industrial settings. The organic gases that evaporate from these sources often fulfill other functions than acting as a traditional solvent (e.g., propellants, fragrances, emollients). For this reason, the solvents sector is often referred to as “volatile chemical products.” Emissions from this sector for the 2020 modeling platform are unchanged from the 2020 NEI, and users should review [Section 32 of the 2020 NEI TSD](#) for additional information on the construction of emissions estimates for solvents in the 2020 NEI.

### **3.2.3.7 Other Nonpoint Sources (*nonpt*)**

The 2020 platform nonpt sector inventory is unchanged from the April 2023 version of the 2020 NEI. Stationary nonpoint sources that were not subdivided into the afdust, livestock, fertilizer, np\_oilgas, rwc or np\_solvents sectors were assigned to the “nonpt” sector. Locomotives and CMV mobile sources from the 2020 NEI 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;
- 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; and
- 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 three distinct sources of PFC emissions, further distinguished by residential or commercial use. The three sources are: (1) displacement of the vapor within the can; (2) emissions due to evaporation (i.e., diurnal emissions); and (3) 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 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 low-level emissions (i.e., they are released into model layer 1) and most nonroad emission are represented as county totals. Note that rail yard and airport emissions are part of the NEI point data category.

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 ocean-going Category 3

vessels are in the cmv\_c3 sector. Both CMV sectors 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 CMV to be modeled with plume rise. In addition, CMV emissions are treated as hourly point source emissions in the modeling platform, although they are part of the NEI nonpoint data category.

#### **3.2.4.1 Onroad (onroad)**

Onroad mobile source include emissions from motorized vehicles operating 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 by the fuel they use, including 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 as they move along the roads). For more details on the approach and for a summary of the MOVES inputs submitted by states, see section 5 of the 2020 NEI TSD.

For the 2020 modeling platform activity data (i.e., VMT, VPOP, starts, on-network idling, and hoteling) were based on state submitted CDBs, as well as data from Federal Highways administration (FHWA) annual VMT at the county level. A new MOVES run for 2020 was done using MOVES3.

Except for California, all onroad emissions are generated using the SMOKE-MOVES emissions modeling framework that leverages MOVES-generated emission factors (<https://www.epa.gov/moves>), county and SCC-specific activity data, and hourly 2020 meteorological data. Specifically, EPA used MOVES3 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. The onroad source classification codes (SCCs) in the modeling platform are more finely resolved than those in the National Emissions Inventory (NEI). The NEI SCCs distinguish vehicles and fuels. The SCCs used in the model platform also distinguish between emissions processes (i.e., off-network, on-network, and extended idle), and road types.

MOVES3 includes the following updates from MOVES2014b:

- Updated emission rates:
  - Updated heavy-duty (HD) diesel running emission rates based on manufacturer in-use testing data from hundreds of HD trucks
  - Updated HD gasoline and compressed natural gas (CNG) trucks
  - Updated light-duty (LD) emission rates for hydrocarbons (HC), CO, NO<sub>x</sub>, and PM
- Includes updated fuel information
- Incorporates HD Phase 2 Greenhouse Gas (GHG) rule, allowing for finer distinctions among HD vehicles
- Accounts for glider vehicles that incorporate older engines into new vehicle chassis
- Accounts for off-network idling – emissions beyond the idling that is already considered in the MOVES drive cycle
- Includes revisions to inputs for hoteling
- Adds starts as a separate type of rate and activity data

Except for California, all onroad emissions were computed with SMOKE-MOVES by multiplying specific types of vehicle activity data by the appropriate emission factors. SMOKE-MOVES was run for specific modeling grids. Emissions for the contiguous U.S. states and Washington, D.C., were computed for a grid covering those areas.

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 2020-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 this study, there are 254 representative counties in the continental U.S. and a total of 292 including the non-CONUS areas.

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 then multiplies the emission rate by appropriate activity data. For on-roadway emissions, vehicle miles travelled (VMT) is the activity data; off-network processes use vehicle population (VPOP), vehicle starts, and hours of off-network idling (ONI); 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, STARTS, off-network idling, 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 were processed in six processing streams that were then merged together into the onroad sector emissions after each of the six streams have been processed:

- rate-per-distance (RPD) 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 (RPS) uses STARTS activity data to compute off-network emissions from vehicles starts;
- 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;
- 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; and
- rate-per-hour off-network idling (RPHO) uses off network idling hours activity data to compute off-network idling emissions for all types of vehicles.

The onroad emissions inputs to MOVES for the 2020 platform are based on the 2020 NEI, described in more detail in Section 5 of the 2020 NEI TSD. These inputs include:

- Key parameters in the MOVES County databases (CDBs) including Low Emission Vehicle (LEV) table
- Fuel months
- Activity data (e.g., VMT, VPOP, speed, HOTELING)

Fuel months, age distributions, and other inputs were consistent with those used to compute the 2020 NEI. Activity data submitted by states and development of the EPA default activity data sets for VMT, VPOP, and hoteling hours are described in detail in the 2020 NEI TSD and supporting documents. Hoteling hours activity were used to calculate emissions from extended idling and auxiliary power units (APUs) by combination long-haul trucks.

SMOKE-MOVES uses vehicle miles traveled (VMT), vehicle population (VPOP), vehicle starts, hours of off-network idling (ONI), and hours of hoteling, to calculate emissions. These datasets are collectively known as “activity data”. For each of these activity datasets, first a national dataset was developed; this national dataset is called the “EPA default” dataset. The default dataset started with the 2020 NEI activity data, which was supplemented with data submitted by state and local agencies. EPA default activity was used for California, but the emissions were scaled to California-supplied values during the emissions processing. States that submitted activity data and development of the EPA default activity data sets for VMT, VPOP, and hoteling hours are described in detail in the 2020 NEI TSD (EPA, 2023) and supporting documents.

In SMOKE 4.7, SMOKE-MOVES was updated to use speed distributions similarly to how they are used when running MOVES in inventory mode. This new speed distribution file, called SPDIST, specifies the amount of time spent in each MOVES speed bin for each county, vehicle (aka source) type, road type, weekday/weekend, and hour of day. This file contains the same information at the same resolution as the Speed Distribution table used by MOVES but is reformatted for SMOKE. Using the SPDIST file results in a SMOKE emissions calculation that is more consistent with MOVES than the old hourly speed profile (SPDPRO) approach, because emission factors from all speed bins can be used, rather than interpolating between the two bins surrounding the single average speed value for each hour as is done with the SPDPRO approach.

For the 2020 NEI, to more accurately reflect the variation of average speeds from month to month throughout the year 2020, month-specific SPDIST files were generated. Speed data from the Streetlight dataset were used to generate hourly speed profiles by county, SCC, and month. The SPDIST files for 2020 NEI are based on a combination of the Streetlight project data and 2020 NEI MOVES CDBs. More information can be found in the 2020 NEI TSD (EPA, 2023) and supporting documents.

Hoteling hours were capped by county at a theoretical maximum and any excess hours of the maximum were reduced. For calculating reductions, a dataset of truck stop parking space availability was used, which includes a total number of parking spaces per county. This same dataset is used to develop the spatial surrogate for allocating county-total hoteling emissions to model grid cells. The parking space dataset includes several recent updates based on new truck stops opening and other new information. There are 8,784 hours in the year 2020; therefore, the maximum number of possible hoteling hours in a particular county is equal to 8,784 \* the number of parking spaces in that county. Hoteling hours were capped at that theoretical maximum value for 2020 in all counties. The final step related to hoteling activity is to split county totals into separate values for extended idling (SCC 2202620153) and Auxiliary Power Units (APUs) (SCC 2202620191). For 2020 modeling with MOVES3, a 7.2% APU split is used nationwide, meaning that during 7.2% of the hoteling hours auxiliary power units are assumed to be running.

Onroad “start” emissions are the instantaneous exhaust emissions that occur at the engine start (e.g., due to the fuel rich conditions in the cylinder to initiate combustion) as well as the additional running exhaust emissions that occur because the engine and emission control systems have not yet stabilized at the running operating temperature. Operationally, start emissions are defined as the difference in emissions between an exhaust emissions test with an ambient temperature start and the same test with the engine and emission control systems already at operating temperature. As such, the units for start emission rates are instantaneous grams/start.

MOVES3 uses vehicle population information to sort the vehicle population into source bins defined by vehicle source type, fuel type (gas, diesel, etc.), regulatory class, model year and age. The model uses default data from instrumented vehicles (or user-provided values) to estimate the number of starts for each source bin and to allocate them among eight operating mode bins defined by the amount of time parked (“soak time”) prior to the start. Thus, MOVES3 accounts for different amounts of cooling of the engine and emission control systems. Each source bin and operating mode has an associated g/start emission rate. Start emissions are also adjusted to account for fuel characteristics, LD inspection and maintenance programs, and ambient temperatures.

After creating VMT inputs for SMOKE-MOVES, Off-network idle (ONI) activity data were also needed. ONI is defined in MOVES as time during which a vehicle engine is running idle and the vehicle is somewhere other than on the road, such as in a parking lot, a driveway, or at the side of the road. This engine activity contributes to total mobile source emissions but does not take place on the road network. Examples of ONI activity include:

- light duty passenger vehicles idling while waiting to pick up children at school or to pick up passengers at the airport or train station,
- single unit and combination trucks idling while loading or unloading cargo or making deliveries, and
- vehicles idling at drive-through restaurants.

Note that ONI does not include idling that occurs on the road, such as idling at traffic signals, stop signs, and in traffic—these emissions are included as part of the running and crankcase running exhaust processes on the other road types. ONI also does not include long-duration idling by long-haul combination trucks (hoteling/extended idle), as that type of long duration idling is accounted for in other MOVES processes.

ONI activity hours were calculated based on VMT. For each representative county, the ratio of ONI hours to onroad VMT (on all road types) was calculated using the MOVES ONI Tool by source type, fuel type, and month. These ratios are then multiplied by each county's total VMT (aggregated by source type, fuel type, and month) to get hours of ONI activity.

MOVES3 was run in emission rate mode to create emission factor tables for 2020, for all representative counties and fuel months. The county databases used to run MOVES to develop the emission factor tables included the state-specific control measures such as the California LEV program, and fuels represented the year 2020. The range of temperatures run along with the average humidities used were specific to the year 2020. The remaining settings for the CDBs are documented in the 2020 NEI TSD. To create the emission factors, MOVES was run separately for each representative county and fuel month for each temperature bin needed for the calendar year 2020. The MOVES results were post-processed into CSV-formatted emission factor tables that can be read by SMOKE-MOVES.

The county databases CDBs used to run MOVES to develop the emission factor tables were those used for the 2020 NEI and therefore included any updated data provided and accepted for the 2020 NEI process. The 2020 NEI development included an extensive review of the various tables including speed distributions were performed. Each county in the continental U.S. was classified according to its state, altitude (high or low), fuel region, the presence of inspection and maintenance programs, the mean light-duty age, and the fraction of ramps. A binning algorithm was executed to identify “like counties. The result was 254 representative counties for CONUS.

Age distributions are a key input to MOVES in determining emission rates. The age distributions for 2020 were updated based on vehicle registration data obtained from IHS Markit, subject to reductions for older vehicles. One of the findings of CRC project A-115 is that IHS data contain higher vehicle populations than state agency analyses of the same Department of Motor Vehicles data, and the discrepancies tend to increase with increasing vehicle age (i.e., there are more older vehicles in the IHS data) and appropriate decreases in older vehicles were applied when the age distributions were computed for 2020.

To create the emission factors, MOVES was run separately for each representative county and fuel month and for each temperature bin needed for calendar year 2020. The CDBs used to run MOVES include the state-specific control measures such as the California low emission vehicle (LEV) program. In addition, the range of temperatures run along with the average humidities used were specific to the year 2020. The MOVES results were post-processed into CSV-formatted emission factor tables that can be read by SMOKE-MOVES.

California uses their own emission model, EMFAC, to develop onroad emissions inventories and provides those inventories to EPA. EMFAC 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. California submitted onroad emissions for the 2020 NEI, and these emissions were used for 2020

modeling. The California inventory had CAPs and select HAPs, but did not have NH<sub>3</sub> or refueling emissions. The EPA added NH<sub>3</sub> to the CARB inventory by using the state total NH<sub>3</sub> from MOVES and allocating it at the county level based on CO. Refueling emissions were taken from MOVES in California. HAP emissions for VOCs and metals as provided by California were used, while other HAPs (e.g., PAHs) were from MOVES.

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 California-developed emissions, 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 emissions hereafter known as “EPA estimates.” These EPA estimates for CA were 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 California onroad emissions inventory. The 2020 California data did not separate off and on-network emissions or extended idling, and also 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 were summed and designated as the emissions for the onroad sector.

#### **3.2.4.2 Category 1, 2, and 3 commercial marine vessels (cmv\_c1c2 and cmv\_3)**

The cmv\_c1c2 sector contains Category 1 and 2 CMV emissions. 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 (i.e., port shape ids) and by SCC and emission type (e.g., hoteling, maneuvering). For more information on CMV sources, see Section 11 of the 2020 NEI TSD and the supplemental documentation.<sup>8</sup> C1 and C2 emissions that occur outside of state waters are not assigned to states. For this modeling platform, all CMV emissions in the cmv\_c1c2 sector are treated as hourly gridded point sources with stack parameters that should result in them being placed in layer 1.

Sulfur dioxide (SO<sub>2</sub>) emissions reflect rules that reduced sulfur emissions for CMV that took effect in the year 2015. The cmv\_c1c2 inventory sector contains small to medium-size engine CMV emissions. Category 1 and Category 2 (C1C2) marine diesel engines typically range in size from about 700 to 11,000 hp. These engines are used to provide propulsion power on many kinds of vessels including tugboats, towboats, supply vessels, fishing vessels, and other commercial vessels in and around ports. They are also

---

<sup>8</sup> [https://gaftp.epa.gov/Air/nei/2020/doc/supporting\\_data/nonpoint/CMV/](https://gaftp.epa.gov/Air/nei/2020/doc/supporting_data/nonpoint/CMV/).



used as stand-alone generators for auxiliary electrical power on many types of vessels. Category 1 represents engines up to 7 liters per cylinder displacement. Category 2 includes engines from 7 to 30 liters per cylinder.

The cmv\_c1c2 inventory sector contains sources that traverse state and federal waters along with emissions from surrounding areas of Canada, Mexico, and international waters. The cmv\_c1c2 sources are modeled as point sources but using plume rise parameters that cause the emissions to be released in the ground layer of the air quality model.

The cmv\_c1c2 sources within state waters are identified in the inventory with the Federal Information Processing Standard (FIPS) county code for the state and county in which the vessel is registered. The cmv\_c1c2 sources that operate outside of state waters but within the Emissions Control Area (ECA) are encoded with a state FIPS code of 85. The ECA areas include parts of the Gulf of Mexico, and parts of the Atlantic and Pacific coasts.

Category 1 and 2 CMV emissions were developed for the 2020 NEI. The emissions were developed based signals from Automated Identification System (AIS) transmitters. AIS is a tracking system used by vessels to enhance navigation and avoid collision with other AIS transmitting vessels. The USEPA Office of Transportation and Air Quality received AIS data from the U.S. Coast Guard (USCG) to quantify all ship activity which occurred between January 1 and December 31, 2020. To ensure coverage for all of the areas needed by the NEI, the requested and provided AIS data extend beyond 200 nautical miles from the U.S. coast. The area covered by the NEI is roughly equivalent to the border of the U.S Exclusive Economic Zone and the North American ECA, although some non-ECA activity are captured as well. Two types of AIS data were received: satellite (S-AIS) and terrestrial (T-AIS).

The AIS data were compiled into five-minute intervals by the USCG, providing a reasonably refined assessment of a vessel's movement. For example, using a five-minute average, a vessel traveling at 25 knots would be captured every two nautical miles that the vessel travels. For slower moving vessels, the distance between transmissions would be less. The ability to track vessel movements through AIS data and link them to attribute data, has allowed for the development of an inventory of very accurate emission estimates. These AIS data were used to define the locations of individual vessel movements, estimate hours of operation, and quantify propulsion engine loads. The compiled AIS data also included the vessel's International Marine Organization (IMO) number and Maritime Mobile Service Identifier (MMSI); which allowed each vessel to be matched to their characteristics obtained from the Clarksons ship registry (Clarksons, 2021).

The engine bore and stroke data were used to calculate cylinder volume. Any vessel that had a calculated cylinder volume greater than 30 liters was incorporated into the USEPA's new Category 3 Commercial Marine Vessel (C3CMV) model. The remaining records were assumed to represent Category 1 and 2 (C1C2) or non-ship activity. The C1C2 AIS data were quality assured including the removal of duplicate messages, signals from pleasure craft, and signals that were not from CMV vessels (e.g., buoys, helicopters, and vessels that are not self-propelled).

The emissions were calculated for each time interval between consecutive AIS messages for each vessel and allocated to the location of the message following to the interval. Emissions were calculated according to Equation 3-1.

$$Emissions_{interval} = Time (hr)_{interval} \times Power(kW) \times EF(\frac{g}{kWh}) \times LLAF \quad 3-1$$

Power is calculated for the propulsive (main), auxiliary, and auxiliary boiler engines for each interval and emission factor (EF) reflects the assigned emission factors for each engine, as described below. LLAF represents the low load adjustment factor, a unitless factor which reflects increasing propulsive emissions during low load operations. Time indicates the activity duration time between consecutive intervals. 11,302 vessels were directly identified by their ship and cargo number. The remaining group of miscellaneous ships represent 13 percent of the AIS vessels (excluding recreational vessels) for which a specific vessel type could not be assigned.

Next, vessels were identified in order determine their vessel type, and thus their vessel group, power rating, and engine tier information which are required for the emissions calculations. See the 2020 NEI documentation for more details on this process. Following the identification, 108 different vessel types were matched to the C1C2 vessels. Vessel attribute data was not available for all these vessel types, so the vessel types were aggregated into 13 different vessel groups for which surrogate data were available. The cmv\_c3 sector contains large engine CMV emissions.

The final components of the emissions computation equation are the emission factors and the low load adjustment factor. The emission factors used in this inventory take into consideration the EPA's marine vessel fuel regulations as well as exhaust standards that are based on the year that the vessel was manufactured to determine the appropriate regulatory tier. Emission factors in g/kWhr by tier for NO<sub>x</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, CO, CO<sub>2</sub>, SO<sub>2</sub> and VOC were developed using Tables 3-7 through 3-10 in USEPA's (2008) Regulatory Impact Analysis on engines less than 30 liters per cylinder. To compile these emissions factors, population-weighted average emission factors were calculated per tier based on C1C2 population distributions grouped by engine displacement. Boiler emission factors were obtained from an earlier Swedish Environmental Protection Agency study (Swedish EPA, 2004). If the year of manufacture was unknown then it was assumed that the vessel was Tier 0, such that actual emissions may be less than those estimated in this inventory. Without more specific data, the magnitude of this emissions difference cannot be estimated.

Propulsive emissions from low-load operations were adjusted to account for elevated emission rates associated with activities outside the engines' optimal operating range. The emission factor adjustments were applied by load and pollutant, based on the data compiled for the Port Everglades 2015 Emission Inventory.<sup>9</sup> Hazardous air pollutants and ammonia were added to the inventory according to multiplicative factors applied either to VOC or PM<sub>2.5</sub>.

The stack parameters used for cmv\_c1c2 are a stack height of 1 ft, stack diameter of 1 ft, stack temperature of 70°F, and a stack velocity of 0.1 ft/s. These parameters force emissions into layer 1.

For more information on the C1C2 CMV emission computations for 2020, see the supporting documentation for the 2020 NEI. The cmv\_c1c2 emissions were aggregated to total hourly values in each

---

<sup>9</sup> USEPA. EPA and Port Everglades Partnership: Emission Inventories and Reduction Strategies. US Environmental Protection Agency, Office of Transportation and Air Quality, June 2018. <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100UKV8.pdf>.

grid cell and run through SMOKE as point sources. SMOKE requires an annual inventory file to go along with the hourly data and this file was generated for 2020.

The cmv\_c3 sector contains large engine CMV emissions. Category 3 (C3) marine diesel engines at or above 30 liters per cylinder. Category 3 (C3) marine diesel engines are those at or above 30 liters per cylinder, typically these are the largest engines rated at 3,000 to 100,000 hp. C3 engines are typically used for propulsion on ocean-going vessels including container ships, oil tankers, bulk carriers, and cruise ships. Emissions control technologies for C3 CMV sources are limited due to the nature of the residual fuel used by these vessels.<sup>10</sup> The cmv\_c3 sector contains sources that traverse state and federal waters; along with sources in waters not covered by the NEI in surrounding areas of Canada, Mexico, and international waters. For more information on CMV sources in the 2020 NEI, see Section 11 of the 2020 NEI TSD and the supplemental documentation for 2020 NEI CMV.

The process for computing the C3 CMV emissions was similar to that used for C1C2 CMV described above. The 2020 CMV C3 NEI data were computed based on the AIS data from the USGS for the year of 2020. The AIS data were coupled with ship registry data that contained engine parameters, vessel power parameters, and other factors such as tonnage and year of manufacture which helped to separate the C3 vessels from the C1C2 vessels. Where specific ship parameters were not available, they were gap-filled. The types of vessels that remain in the C3 data set include bulk carrier, chemical tanker, liquefied gas tanker, oil tanker, other tanker, container ship, cruise, ferry, general cargo, fishing, refrigerated vessel, roll-on/roll-off, tug, and yacht.

Prior to use, the AIS data were reviewed - data deemed to be erroneous were removed, and data found to be at intervals greater than 5 minutes were interpolated to ensure that each ship had data every five minutes. The five-minute average data provide a reasonably refined assessment of a vessel's movement. For example, using a five-minute average, a vessel traveling at 25 knots would be captured every two nautical miles that the vessel travels. For slower moving vessels, the distance between transmissions would be less.

Emissions were computed according to a computed power need (kW) multiplied by the time (hr) and by an engine-specific emission factor (g/kWh) and finally by a low load adjustment factor that reflects increasing propulsive emissions during low load operations. The resulting emissions were available at 5-minute intervals. Code was developed to aggregate these emissions to modeling grid cells and up to hourly levels so that the emissions data could be input to SMOKE for emissions modeling with SMOKE. Within SMOKE, the data were speciated into the pollutants needed by the air quality model but since the data were already in the form of point sources at the center of each grid cell, and they were already hourly, no other processing was needed within SMOKE. SMOKE requires an annual inventory file to go along with the hourly data, so this file was also generated for 2020.

On January 1st, 2015, the ECA initiated a fuel sulfur standard which regulated large marine vessels to use fuel with 1,000 ppm sulfur or less. These standards are reflected in the cmv\_c3 inventories.

The resulting point emissions centered on each grid cell were converted to an annual point 2010 flat file format (FF10). A set of standard stack parameters were assigned to each release point in the cmv\_c3 inventory. The assigned stack height was 65.62 ft, the stack diameter was 2.625 ft, the stack temperature

---

<sup>10</sup> <https://www.epa.gov/regulations-emissions-vehicles-and-engines/regulations-emissions-marine-vessels>.

was 539.6 °F, and the velocity was 82.02 ft/s. Emissions were computed for each grid cell needed for modeling.

#### **3.2.4.3 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. The rail sector includes all locomotives in the NEI nonpoint data category including line haul locomotives on Class 1, 2, and 3 railroads along with emissions from commuter rail lines and Amtrak. The rail 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 rail emissions for the 2020 platform use the 2020 NEI. The 2020 NEI is based on methods developed during the 2017 rail inventory developed for the 2017 NEI by the Lake Michigan Air Directors Consortium (LADCO) and the State of Illinois with support from various other states. Class I railroad emissions are based on confidential link-level line-haul activity GIS data layer maintained by the Federal Railroad Administration (FRA). In addition, the Association of American Railroads (AAR) provided national emission tier fleet mix information. Class II and III railroad emissions are based on a comprehensive nationwide GIS database of locations where short line and regional railroads operate. Passenger rail (Amtrak) emissions follow a similar procedure as Class II and III, except using a database of Amtrak rail lines. Yard locomotive emissions are based on a combination of yard data provided by individual rail companies, and by using Google Earth and other tools to identify rail yard locations for rail companies which did not provide yard data. Information on specific yards were combined with fuel use data and emission factors to create an emissions inventory for rail yards. Pollutant-specific factors were applied on top of the activity-based changes for the Class I rail. More detailed information on the development of the 2020 NEI rail inventory for this study is available in the 2020 NEI TSD and in the [Rail 2020 National Emissions Inventory](#) supplementary document on the 2020 NEI supporting data FTP site.

#### **3.2.4.4 MOVES-based Nonroad Mobile Sources (nonroad)**

The mobile nonroad equipment sector includes all mobile source emissions that do not operate on roads, excluding commercial marine vehicles, railways, and aircraft. Types of nonroad equipment include recreational vehicles, pleasure craft, and construction, agricultural, mining, and lawn and garden equipment. Nonroad equipment emissions were computed by running MOVES3 which incorporates the NONROAD model. MOVES3 incorporated updated nonroad engine population growth rates, nonroad Tier 4 engine emission rates, and sulfur levels of nonroad diesel fuels. MOVES provides a complete set of HAPs and incorporates updated nonroad emission factors for HAPs. MOVES3 was used for all states other than California, which uses their own model. California nonroad emissions were provided by the California Air Resources Board (CARB) for the 2020 NEI. CARB emissions were used in California for all pollutants except PAHs, which were taken from MOVES.

MOVES creates a monthly emissions inventory for criteria air pollutants (CAPs) and a full set of HAPs, plus additional pollutants such as NONHAPTOG and ETHANOL, which are not part of the NEI but are used for speciation. MOVES provides estimates of NONHAPTOG along with the speciation profile code for the NONHAPTOG emission source. This was accomplished by using NHTOG#### as the pollutant code in the Flat File 2010 (FF10) inventory file that can be read into SMOKE, where #### is a speciation profile code. For California, NHTOG####-VOC and HAP-VOC ratios from MOVES-based emissions were applied to VOC emissions so that VOC emissions can be speciated consistently with other states.

MOVES also provides estimates of PM<sub>2.5</sub> by speciation profile code for the PM<sub>2.5</sub> emission source, using PM25\_#### as the pollutant code in the FF10 inventory file, where #### is a speciation profile code. To facilitate calculation of PMC within SMOKE, and to help create emissions summaries, an additional pollutant representing total PM<sub>2.5</sub> called PM25TOTAL was added to the inventory. As with VOC, PM25\_####-PM25TOTAL ratios were calculated and applied to PM<sub>2.5</sub> emissions in California so that PM<sub>2.5</sub> emissions in California can be speciated consistently with other states.

MOVES3 outputs emissions data in county-specific databases, and a post-processing script converts the data into FF10 format. Additional post-processing steps were performed as follows:

- County-specific FF10s were combined into a single FF10 file.
- Emissions were aggregated from the more detailed SCCs modeled in MOVES to the SCCs modeled in SMOKE. A list of the aggregated SMOKE SCCs is in Appendix A of the 2016v1 platform nonroad specification sheet (NEIC, 2019).
- To reduce the size of the inventory, HAPs not needed for air quality modeling, such as dioxins and furans, were removed from the inventory.
- To reduce the size of the inventory further, all emissions for sources (identified by county/SCC) for which CAP emissions totaling less than  $1 \times 10^{-10}$  were removed from the inventory. The MOVES model attributes a very tiny amount of emissions to sources that are actually zero, for example, snowmobile emissions in Florida. Removing these sources from the inventory reduces the total size of the inventory by about 7%.
- Gas and particulate components of HAPs that come out of MOVES separately, such as naphthalene, were combined.
- VOC was renamed VOC\_INV so that SMOKE does not speciate both VOC and NONHAPTOG, which would result in a double count.
- PM25TOTAL, referenced above, was also created at this stage of the process.
- Emissions for airport ground support vehicles (SCCs ending in -8005), and oil field equipment (SCCs ending in -10010), were removed from the inventory at this stage, to prevent a double count with the airports and np\_oilgas sectors, respectively.
- California emissions from MOVES were deleted and replaced with the CARB-supplied emissions.

California nonroad emissions were provided by CARB for the 2020 NEI. All California nonroad inventories were annual, with monthly temporalization applied in SMOKE. Emissions for oil field equipment (SCCs ending in -10010) were removed from the California inventory in order to prevent a double count with the np\_oilgas sector. VOC HAPs from California were incorporated into speciation similarly to VOC HAPs from MOVES elsewhere, e.g. model species BENZ is equal to HAP emissions for benzene as submitted by CARB. VOC and PM<sub>2.5</sub> emissions were allocated to speciation profiles. Ratios of VOC (PM<sub>2.5</sub>) by speciation profile to total VOC (PM<sub>2.5</sub>) were calculated by county and SCC from the MOVES run in California, and then applied CARB-provided VOC (PM<sub>2.5</sub>) in the inventory so that California nonroad emissions could be speciated consistently with the rest of the country.

For more information on the nonroad sector in the 2020 NEI see [Section 4 of the 2020 NEI TSD](#).

### **3.2.5 Day-Specific Point Source Fires (*ptfire*)**

Multiple types of fires are represented in the modeling platform. These include wild and prescribed fires that are grouped into the *ptfire-wild* and *ptfire-rx* sectors, respectively, and agricultural fires that comprise the *ptagfire* sector. All *ptfire* and *ptagfire* fires are in the United States. Fires outside of the United States are described in the *ptfire\_othna* sector later in this document.

Wildfire and prescribed burning emissions are contained in the *ptfire-wild* and *ptfire-rx* sectors, respectively. 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 *ptfire-rx* and *ptfire-wild* inventories include separate SCCs for the flaming and smoldering combustion phases for wildfire and prescribed burns. Note that prescribed grassland fires or Flint Hills, Kansas have their own SCC (2811021000) in the inventory. These wild grassland fires were assigned the standard wildfire SCCs.

Inputs to SMARTFIRE2 for 2020 include:

- The National Oceanic and Atmospheric Administration's (NOAA's) Hazard Mapping System (HMS) fire location information
- National Incident Feature Services (NIFS) (formerly GeoMAC) wildland fire perimeter polygons
- The Incident Status Summary, also known as the "ICS-209", used for reporting specific information on fire incidents of significance
- Hazardous fuel treatment reduction polygons for prescribed burns from the Forest Service Activity Tracking System (FACTS)
- Fire activity on federal lands from the United States Fish and Wildlife Service (USFWS) and other Department of Interior agencies
- Wildfire and prescribed date, location, and locations from S/L/T activity 2020 NEI submitters (includes Alaska, Arizona, California, Delaware, Georgia, Florida, Iowa, Idaho, Kansas (Flint Hills only), Louisiana, Maine, Massachusetts, Montana, New Jersey, North Carolina, Nevada (Washoe Co.), Oklahoma, Oregon, Rhode Island, South Carolina, Texas, Utah, Virginia, Washington, and Wyoming)

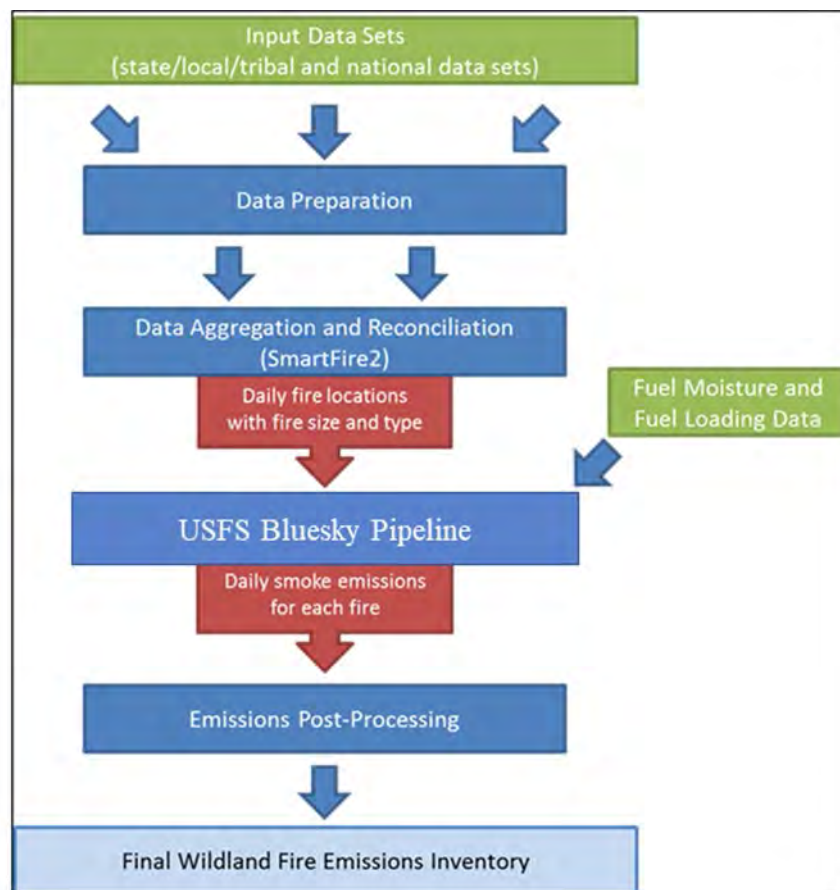
The national and S/L/T data mentioned earlier were used to estimate daily wildfire and prescribed burn emissions from flaming combustion and smoldering combustion phases for the 2020 inventory. Flaming combustion is more complete combustion than smoldering and is more prevalent with fuels that have a high surface-to-volume ratio, a low bulk density, and low moisture content. Smoldering combustion occurs without a flame, is a less complete burn, and produces some pollutants, such as PM<sub>2.5</sub>, VOCs, and CO, at higher rates than flaming combustion. Smoldering combustion is more prevalent with fuels that have low surface-to-volume ratios, high bulk density, and high moisture content. Models sometimes differentiate between smoldering emissions that are lofted with a smoke plume and those that remain near



the ground (residual emissions), but for the purposes of the inventory the residual smoldering emissions were allocated to smoldering SCCs.

Figure 3-1 is a schematic of the data processing stream for the inventory of wildfire and prescribed burn sources. The ptfire-rx and ptfire-wild inventory sources were estimated using Satellite Mapping Automated Reanalysis Tool for Fire Incident Reconciliation version 2 (SMARTFIRE2) and Blue Sky Pipeline. SMARTFIRE2 is an algorithm and database system that operate within a geographic information system (GIS). SMARTFIRE2 combines multiple sources of fire information and reconciles them into a unified GIS database. It reconciles fire data from space-borne sensors and ground-based reports, thus drawing on the strengths of both data types while avoiding double-counting of fire events. At its core, SMARTFIRE2 is an association engine that links reports covering the same fire in any number of multiple databases. In this process, all input information is preserved, and no attempt is made to reconcile conflicting or potentially contradictory information (for example, the existence of a fire in one database but not another).

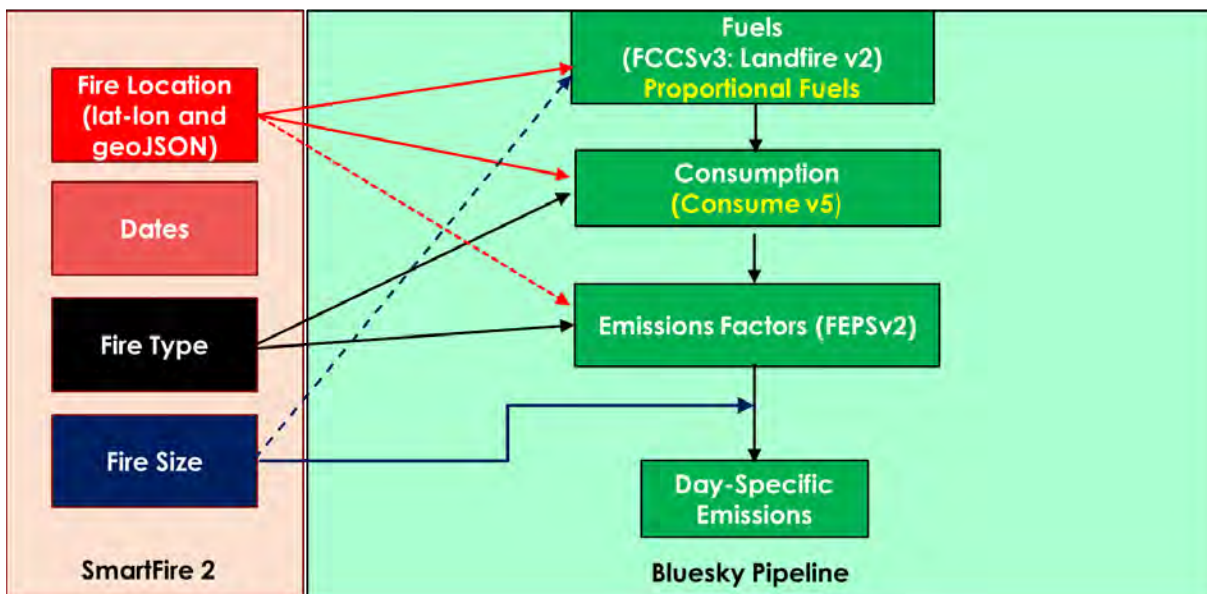
For the 2020 platform, the national and S/L/T fire information was input into SMARTFIRE2 and then merged and associated based on user-defined weights for each fire information dataset. The output from SMARTFIRE2 was daily acres burned by fire type, and latitude-longitude coordinates for each fire. The fire type assignments were made using the fire information datasets. If the only information for a fire was a satellite detect for fire activity, then the flow described in Figure 3-1 was used to make fire type assignment by state and by month in conjunction with the default fire type assignments.





**Figure 3-1. Processing flow for fire emission estimates**

The second system used to estimate emissions is the BlueSky Modeling Pipeline. The framework supports the calculation of fuel loading and consumption, and emissions using various models depending on the available inputs as well as the desired results. The contiguous United States, where Fuel Characteristic Classification System (FCCS) fuel loading data are available, were processed using the modeling chain described in Figure 3-2. **Error! Reference source not found..** The Fire Emissions Production Simulator (FEPS) (Anderson, 2004) in the BlueSky Pipeline generates all the CAP emission factors for wildland fires used in the 2020 study. HAP emission factors were obtained from Urbanski's (2014) work and applied by region and by fire type.



**Figure 3-2. BlueSky Pipeline modeling system**

The FCCSv3 cross-reference was implemented along with the LANDFIREv1 (at 200 meter resolution) to provide better fuel bed information for the BlueSky Pipeline (BSP). The LANDFIREv2 was aggregated from the native resolution and projection to 200 meter using a nearest-neighbor methodology. Aggregation and reprojection was required for the proper function on BSP.

The final products from this process are annual and daily FF10-formatted emissions inventories. These SMOKE-ready inventory files contain both CAPs and HAPs. The BAFM HAP emissions from the inventory were used directly in modeling and were not overwritten with VOC speciation profiles (i.e., an “integrate HAP” use case).

### 3.2.6 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 2020 in a similar way to the emissions in *ptfire*.

Daily year-specific agricultural burning emissions are derived from HMS fire activity data, which contains the date and location of remote-sensed anomalies. The activity is filtered using the 2020 USDA

cropland data layer (CDL). Satellite fire detects over agricultural lands are assumed to be agricultural burns and assigned a crop type. Detects that are not over agricultural lands are output to a separate file for use in the ptfire sector. Each detect is assigned an average size of between 40 and 80 acres based on crop type. Grassland/pasture fires were moved to the ptfire sectors for this 2020 modeling platform. Depending on their origin, grassland fires are in both ptfire-rx and ptfire-wild sectors because both fire types do involve grassy fuels.

The point source agricultural fire (ptagfire) inventory sector contains daily agricultural burning emissions. Daily fire activity was derived from the NOAA Hazard Mapping System (HMS) fire activity data. 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.

Another feature of the ptagfire database is that the satellite detections for 2020 were filtered out to exclude areas covered by snow during the winter months. To do this, the daily snow cover fraction per grid cell was extracted from a 2020 meteorological Weather Research Forecast (WRF) model simulation. The locations of fire detections were then compared with this daily snow cover file. For any day in which a grid cell had snow cover, the fire detections in that grid cell on that day were excluded from the inventory. Due to the inconsistent reporting of fire detections from the Visible Infrared Imaging Radiometer Suite (VIIRS) platform, any fire detections in the HMS dataset that were flagged as VIIRS or Suomi National Polar-orbiting Partnership satellite were excluded. In addition, certain crop types (corn and soybeans) have been excluded from these specific midwestern states: Iowa, Kansas, Indiana, Illinois, Michigan, Missouri, Minnesota, Wisconsin, and Ohio. The reason for these crop types being excluded is because states have indicated that these crop types are not burned.

Heat flux for plume rise was calculated using the size and assumed fuel loading of each daily agricultural fire. This information is needed for a plume rise calculation within a chemical transport modeling system.

The daily agricultural and open burning emissions were converted from a tabular format into the SMOKE-ready daily point flat file format. The daily emissions were also aggregated into annual values by location and converted into the annual point flat file format.

For this modeling platform, a SMOKE update allows the use of HAP integration for speciation for PTDAY inventories. The 2020 agricultural fire inventories include emissions for HAPs, so HAP integration was used for this study.

### **3.2.7 Biogenic Sources (*beis*)**

Biogenic emissions were computed based on the 2020 meteorology data used for the 2020 NEI and were developed using the Biogenic Emission Inventory System version 4 (BEIS4) within CMAQ. BEIS4 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 BEIS4 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). BEIS4 computes the seasonality of emissions using the 1-meter soil

temperature (SOIT2) instead of the BIOSEASON file, and canopy temperature and radiation environments are now modeled using the driving meteorological model's (WRF) representation of leaf-area index (LAI) rather than the estimated LAI values from BELD data alone. See [these CMAQ Release Notes](#) for technical information on BEIS4: <https://github.com/USEPA/CMAQ/wiki/CMAQ-Release-Notes:-Emissions-Updates:-BEIS-Biogenic-Emissions>. The variables output from the Meteorology-Chemistry Interface Processor (MCIP) that are used to convert WRF outputs to CMAQ inputs are shown in Table 3-5.

**Table 3-5. Meteorological variables required by BEIS 3.7**

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 surface
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
WSAT_PX	soil saturation from (Pleim-Xiu Land Surface Model) PX-LSM

The Biogenic Emissions Landcover Database version 6 (BELD6) was used as the input gridded land use information in generating 2020 NEI estimates. BELD version 5 (BELD5) was used to generate 2017 NEI estimates. There are now two different BELD6 datasets that are input into BEIS4. The gridded landuse and the other is the gridded dry leaf biomass (grams/m<sup>2</sup>) values for various vegetation types. The BELD6 includes the following datasets:

- High resolution tree species and biomass data from Wilson et al. 2013a, and Wilson et al. 2013b for which species names were changed from non-specific common names to scientific names
- Tree species biogenic volatile organic carbon (BVOC) emission factors for tree species were taken from the NCAR Enclosure database (Wiedinmyer, 2001)
  - <https://www.sciencedirect.com/science/article/pii/S1352231001004290>
- Agricultural land use from [US Department of Agriculture \(USDA\) crop data layer](#)
- Global Moderate Resolution Imaging Spectroradiometer (MODIS) 20 category data with enhanced lakes and Fraction of Photosynthetically Active Radiation (FPAR) for vegetation coverage from [National Center for Atmospheric Research \(NCAR\)](#)

- Canadian BELD land use, updates to Version 4 of the Biogenic Emissions Landuse Database (BELD4) for Canada and Impacts on Biogenic VOC Emissions ([https://www.epa.gov/sites/default/files/2019-08/documents/800am\\_zhang\\_2\\_0.pdf](https://www.epa.gov/sites/default/files/2019-08/documents/800am_zhang_2_0.pdf)).

Bug fixes included in BEIS4 included the following:

- Solar radiation attenuation in the shaded portion of the canopy was using the direct beam photosynthetically active radiation (PAR) when the diffuse beam PAR attenuation coefficient should have been used.
  - This update had little impact on the total emissions but did result in slightly higher emissions in the morning and evening transition periods for isoprene, methanol and Methylbutenol (MBO).
- The fraction of solar radiation in the sunlit and shaded canopy layers, SOLSUN and SOLSHADE respectively were estimated using a planar surface. These should have been estimated based on the PAR intercepted by a hemispheric surface rather than a plane.
  - This update can result in an earlier peak in leaf temperature, approximately up to an hour.
- The quantum yield for isoprene emissions (ALPHA) was updated to the mean value in [Niinemets et al. 2010a](#) and the integration coefficient (CL) was updated to yield 1 when PAR = 1000 following [Niinemets et al 2010b](#).
  - This updated resulted in a slight reduction in isoprene, methanol, and MBO emissions.

Biogenic emissions computed with BEIS were used to review and prepare summaries, but were left out of the CMAQ-ready merged emissions in favor of inline biogenics produced during the CMAQ model run itself using the same algorithm described above but with finer time steps within the air quality model. Biogenic emissions computed with BEIS to review and prepare summaries, but they were left out of the CMAQ-ready merged emissions. Instead, the biogenic emissions are produced inline during the CMAQ model run which uses the same algorithm described above, but with finer time steps within the air quality model.

### ***3.2.8 Emissions from Canada, Mexico (othpt, othar, othafdust, othptdust, onroad\_can, onroad\_mex, ptfire\_othna)***

The emissions from Canada and Mexico are included as part of the emissions modeling sectors: canmex\_point, canmex\_area, canada\_afdust, canada\_ptdust, canada\_onroad, mexico\_onroad, canmex\_ag, and canada\_og2D. These sector names are new to 2020 platform, but the general organization of these sectors is unchanged from the 2019 platform, except for agricultural emissions in Canada and Mexico. The canmex\_ag sector is processed as a separate sector for reporting and tracking purposes, and unlike in other recent emissions platforms, the Canada ag sources are area sources in this platform rather than pre-gridded point sources. As in prior platforms, Fugitive dust emissions in Canada are represented as both area sources (canada\_afdust sector, formerly “othafdust”) and point sources (canada\_ptdust sector, formerly “othptdust”). Due to the large number of individual points, low-level oil and gas emissions in Canada are processed separately from the canmex\_point sector to reduce the number of individual points to track within CMAQ, and also to reduce the size of the model-ready emissions files.

Emissions in these sectors were taken from the 2020 inventories. Environment and Climate Change Canada (ECCC) provided the following inventories for use in the 2020 modeling. The sectors in which they were incorporated are listed and the inventories are described in more detail below:

- Agricultural livestock and fertilizer, area source format (canmex\_ag sector)
- Surface-level oil and gas emissions in Canada (canada\_og2D sector)
- Agricultural fugitive dust, point source format (canada\_ptdust sector)
- Other area source dust (canada\_afdust sector)
- Onroad (canada\_onroad sector)
- Nonroad and rail (canmex\_area sector)
- Airports (canmex\_point sector)
- Other area sources (canmex\_area sector)
- Other point sources (canmex\_point sector)

The 2020 NEI CMV included coastal waters of Canada and Mexico with emissions derived from AIS data. These NEI emissions were used for all areas of Canada and Mexico and are included in the cmv\_c1c2 and cmv\_c3 sectors. Both the C1C2 and C3 emissions were developed in a point source format with point locations at the center of the 12km grid cells.

Other than the CB6 species of NBAFM present in the speciated point source data, there are no explicit HAP emissions in these Canadian inventories. In addition to emissions inventories, the ECCC 2020 dataset also included shapefiles for creating spatial surrogates. These surrogates were used for this study.

Canadian point source inventories provided by ECCC for the 2020 NEI were adjusted for the impacts of COVID. These inventories include emissions for airports and other point sources. The Canadian point source inventory is pre-speciated for the CB6 chemical mechanism. Annual emissions provided by ECCC already reflected pandemic effects, but the monthly distributions of emissions did not. To account for pandemic effects, monthly emissions in Canada were redistributed using data from the CONFORM dataset (<https://permalink.aeris-data.fr/CONFORM>), which provides country-specific adjustment factors to account for pandemic effects for each month in 2020. Monthly temporal profiles were calculated from the CONFORM dataset as ratios of monthly totals versus annual totals for several different categories (aviation, energy, industry, public and commercial, residential, and transport) and applied to the annual emissions provided by ECCC, with each SCC mapped to a CONFORM category. Annual emissions totals in Canada were not changed as part of this process, only the distribution to months.

Point sources in Mexico were compiled based on inventories projected from the Inventario Nacional de Emisiones de Mexico, 2016 (Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT)), projected to 2019 as part of the 2019 emissions modeling platform, and then projected to 2020 to include COVID pandemic effects. 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, latitude and longitude coordinates were verified and adjusted if they were not consistent with the reported municipality and were additionally adjusted for COVID. Only CAPs are covered in the Mexico point source inventory. The CONFORM dataset was used to apply pandemic adjustments to emissions in Mexico, except that unlike in Canada, annual emissions as well as monthly temporal profiles were adjusted. First, monthly emissions totals for the unadjusted 2019 inventory were calculated using existing temporal profiles. Then, a 2019-to-2020 scaling factor was calculated for each month using data from the CONFORM dataset, and for each emissions category in the CONFORM dataset (energy, industry, public

and commercial, residential, and transport). These scaling factors were applied to the 2019 monthly Mexico emissions, and a new annual total for 2020 was calculated from the adjusted monthly totals.

Fugitive dust sources of particulate matter emissions excluding land tilling from agricultural activities, were provided by Environment and Climate Change Canada (ECCC) as part of their 2020 emission inventory. This inventory no longer contains agricultural dust. Different source categories were provided as gridded point sources and area (nonpoint) source inventories. Gridded point source emissions resulting from land tilling due to agricultural activities were provided as part of the ECCC 2020 emission inventory. The provided wind erosion emissions were removed. Both the `canada_afdust` and `canada_ptdust` emissions have a COVID-adjusted monthly resolution based on the CONFORM dataset categories of industry and transport, following a similar process as the `canmex_point` sector. A transport fraction adjustment that reduces dust emissions based on land cover types was applied to both point and nonpoint dust emissions, along with a meteorology-based (precipitation and snow/ice cover) zero-out of emissions when the ground is snow covered or wet.

Agricultural emissions from Canada and Mexico, excluding fugitive dust, are included in the `canmex_ag` sector. Canadian agricultural emissions were provided by Environment and Climate Change Canada (ECCC) as part of their 2020 emission inventory. Unlike in recent platforms, Canadian agricultural were not represented as point sources, instead they were represented as area sources and gridded using spatial surrogates. In Mexico, agricultural sources are based on the 2019ge Mexico nonpoint inventory at the municipio resolution. The 2019 inventory was based on a projection of 2016 inventories provided by SEMARNAT. COVID pandemic adjustments were not applied to the agricultural sector.

Canadian point source inventories provided by ECCC for the 2020 NEI included oil and gas emissions. A very large number of these oil and gas point sources are surface level emissions, appropriate to be modeled in layer 1. Reducing the size of the `canmex_point` sector improves air quality model run time because plume rise calculations are needed for fewer sources, so these surface level oil and gas sources were placed into the `canada_og2D` sector for layer 1 modeling. These emissions include COVID-adjusted monthly data based on the CONFORM dataset industry sector.

ECCC provided year 2020 Canada province, and in some cases sub-province, resolution emissions from for nonpoint and nonroad sources (`canmex_area`). The nonroad sources were monthly while the nonpoint and rail emissions were annual. Annual emissions provided by ECCC already reflected pandemic effects, but monthly distributions of emissions did not. Following a similar process as the `canmex_point` sector, monthly emissions in Canada were redistributed using data from the CONFORM dataset to reflect pandemic effects. The CONFORM categories used for the Canada monthly COVID adjustments were energy, industry, public and commercial, residential, and transport.

For Mexico, 2019ge Mexico nonpoint and nonroad inventories at the municipio resolution (which were based on a projection of 2016 inventories provided by SEMARNAT) were projected to 2020 to include COVID pandemic effects using a process similar to the one described for the `canmex_point` sector. The CONFORM categories used for the projection and monthly distribution included: industry, public and commercial, residential, and transport.

The onroad emissions for Canada and Mexico are in the `canada_onroad` and `mexico_onroad` sectors, respectively. Emissions for Canada are new for 2020. In Canada, COVID impacts were applied to the

monthly profiles (not to the annual totals) using the CONFORM dataset emissions from the transport category.

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). Emissions from MOVES-Mexico for the year 2020 did not include any COVID pandemic effects, so monthly and annual emissions were adjusted using the monthly CONFORM adjustment factors for Mexico transport.

Annual 2020 wildland fire emissions for Mexico, Canada, Central America, and Caribbean nations are included in the ptfire\_othna sector. Canadian fires from May-December were provided by ECCC and are based on their Firework system (<https://weather.gc.ca/firework/>). Canadian fires for the non-summer months along with fires in Mexico, Central America, and the Caribbean, were developed from the Fire Inventory from NCAR (FINN) v2.5 daily fire emissions for 2020 (Wiedenmyer, 2023). 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 wildfires rather than prescribed. FINN fire detects of 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.

### ***3.2.9 Ocean Chlorine, Ocean Sea Salt, and Volcanic Mercury***

The ocean chlorine gas emission estimates are based on the build-up of molecular chlorine ( $\text{Cl}_2$ ) 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 model-species name “CHLORINE” was changed to “CL2” to support CMAQ modeling.

For mercury, the volcanic mercury emissions that were used in the recent modeling platforms were not included in this study. The emissions were originally developed for a 2002 multipollutant modeling platform with coordination and data from Christian Seigneur and Jerry Lin for 2001 (Seigneur et. al, 2004 and Seigneur et. al, 2001). ). The volcanic emissions from the most recent eruption were not included in the because they have diminished by the year 2019. Thus no volcanic emissions were included.

Because of mercury bidirectional flux within the latest version of CMAQ, no other natural mercury emissions are included in the emissions merge step.

## **3.3 Emissions Modeling Summary**

The CMAQ and CAMx air quality models require 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 and vertical resolution required by the air quality model. Emissions modeling includes temporal allocation, spatial allocation, and pollutant speciation. Emissions modeling sometimes includes the vertical allocation (i.e., plume rise)



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.

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; totals by county (U.S.), province (Canada), or municipio (Mexico); or gridded emissions. 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 4.9 was used to process the raw emissions inventories into emissions inputs for each modeling sector into a format compatible with 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 plume rise capability allows for the use of 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 2-D gridded 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 3-6 summarizes the major processing steps of each platform sector with the columns as follows.

The “Spatial” column shows the spatial approach used: “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.

Finally, 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. In all of the “in-line” sectors, all sources are output by SMOKE into point source files which are subject to plume rise calculations in the air quality model. In other words, no emissions are output to layer 1 gridded emissions files from those sectors as has been done in past platforms. The air quality

model computes the plume rise using stack parameters, the Briggs algorithm, and the hourly emissions in the SMOKE output files for each emissions sector. The height of the plume rise determines the model layers into which the emissions are placed. The plume top and bottom are computed, along with 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. 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.

**Table 3-6. Key emissions modeling steps by sector**

<b>Platform sector</b>	<b>Spatial</b>	<b>Speciation</b>	<b>Inventory resolution</b>	<b>Plume rise</b>
afdust_adj	Surrogates	Yes	Annual	
airports	Point	Yes	Annual	None
beis	Pre-gridded land use	in BEIS4	computed hourly in CMAQ	
fertilizer	EPIC	No	computed hourly in CMAQ	
livestock	Surrogates	Yes	Annual	
cmv_c1c2	Point	Yes	hourly	in-line
cmv_c3	Point	Yes	hourly	in-line
nonpt	Surrogates & area-to-point	Yes	Annual	
nonroad	Surrogates	Yes	monthly	
np_oilgas	Surrogates	Yes	Annual	
onroad	Surrogates	Yes	monthly activity, computed hourly	
onroad_ca_adj	Surrogates	Yes	monthly activity, computed hourly	
canada_onroad	Surrogates	Yes	monthly	
mexico_onroad	Surrogates	Yes	monthly	
canada_afdust	Surrogates	Yes	annual & monthly	
canmex_area	Surrogates	Yes	monthly	
canmex_point	Point	Yes	monthly	in-line
canada_ptdust	Point	Yes	annual	None
canada_og2D	Point	Yes	monthly	None
canmex_ag	Surrogates	Yes	annual	
ptagfire	Point	Yes	daily	in-line
pt_oilgas	Point	Yes	annual	in-line
ptegu	Point	Yes	daily & hourly	in-line
ptfire-rx	Point	Yes	daily	in-line
ptfire-wild	Point	Yes	daily	in-line
ptfire_othna	Point	Yes	daily	in-line
ptnonipm	Point	Yes	annual	in-line

<b>Platform sector</b>	<b>Spatial</b>	<b>Speciation</b>	<b>Inventory resolution</b>	<b>Plume rise</b>
rail	Surrogates	Yes	annual	
rwc	Surrogates	Yes	annual	
np_solvents	Surrogates	Yes	annual	

Note that SMOKE has the option of grouping sources so that they are treated as a single stack when computing plume rise. For the modeling cases discussed in this document, 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 latitude and longitudes 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 stack grouping.

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.

### ***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 3-3, but the air quality model was run on the smaller 12-km domain (12US2). 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. Later sections provide details on the spatial surrogates and area-to-point data used to accomplish spatial allocation with SMOKE.



**Figure 3-3. CMAQ Modeling Domain**

### ***3.3.4 Chemical Speciation Configuration***

Chemical speciation involves the process of translating emissions from the inventory into the chemical mechanism-specific “model species” needed by an air quality model. Using the CB6R5\_AE7 chemical mechanism as an example, which is the mechanism utilized by the 2020 NEI modeling platform, these model species either represent explicit chemical compounds (e.g., acetone, benzene, ethanol) or groups of species (i.e., “lumped species;” e.g., PAR, OLE, KET). This chemical mechanism is an updated version of the CB6R3\_AE7 chemical mechanism and features new reaction rates for some chemical reactions (Yarwood et al., 2020). CMAQ’s Aerosol Module version 7 (AE7) is an updated version of the AE6 aerosol module, with alpha-pinene made an explicit emitted species. Table 3-7 lists the model species produced by SMOKE in the platform used for this study.

**Table 3-7. Emission model species produced for CB6R3AE7 for CMAQ**

<b>Inventory Pollutant</b>	<b>Model Species</b>	<b>Model species description</b>
Cl <sub>2</sub>	CL2	Atomic gas-phase chlorine
HCl	HCL	Hydrogen Chloride (hydrochloric acid) gas
CO	CO	Carbon monoxide
NO <sub>x</sub>	NO	Nitrogen oxide
NO <sub>x</sub>	NO2	Nitrogen dioxide
NO <sub>x</sub>	HONO	Nitrous acid
SO <sub>2</sub>	SO2	Sulfur dioxide
SO <sub>2</sub>	SULF	Sulfuric acid vapor
NH <sub>3</sub>	NH3	Ammonia
NH <sub>3</sub>	NH3_FERT	Ammonia from fertilizer
VOC	AACD	Acetic acid
VOC	ACET	Acetone
VOC	ALD2	Acetaldehyde
VOC	ALDX	Propionaldehyde and higher aldehydes
VOC	APIN	Alpha pinene
VOC	BENZ	Benzene
VOC	CAT1	Methyl-catechols
VOC	CH4	Methane
VOC	CRES	Cresols
VOC	CRON	Nitro-cresols
VOC	ETH	Ethene
VOC	ETHA	Ethane
VOC	ETHY	Ethyne
VOC	ETOH	Ethanol
VOC	FACD	Formic acid
VOC	FORM	Formaldehyde
VOC	GLY	Glyoxal
VOC	GLYD	Glycolaldehyde
VOC	IOLE	Internal olefin carbon bond (R-C=C-R)
VOC	ISOP	Isoprene
VOC	ISPD	Isoprene Product
VOC	IVOC	Intermediate volatility organic compounds
VOC	KET	Ketone Groups
VOC	MEOH	Methanol
VOC	MGLY	Methylglyoxal
VOC	NAPH	Naphthalene
VOC	NVOL	Non-volatile compounds
VOC	OLE	Terminal olefin carbon bond (R-C=C)
VOC	PACD	Peroxyacetic and higher peroxy-carboxylic acids
VOC	PAR	Paraffin carbon bond
VOC	PRPA	Propane
VOC	SESQ	Sesquiterpenes (from biogenics only)
VOC	SOAALK	Secondary Organic Aerosol (SOA) tracer
VOC	TERP	Terpenes (from biogenics only)

Inventory Pollutant	Model Species	Model species description
VOC	TOL	Toluene and other monoalkyl aromatics
VOC	UNR	Unreactive
VOC	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 <sub>10</sub>	PMC	Coarse PM > 2.5 microns and ≤ 10 microns
PM <sub>2.5</sub>	PEC	Particulate elemental carbon ≤ 2.5 microns
PM <sub>2.5</sub>	PNO3	Particulate nitrate ≤ 2.5 microns
PM <sub>2.5</sub>	POC	Particulate organic carbon (carbon only) ≤ 2.5 microns
PM <sub>2.5</sub>	PSO4	Particulate Sulfate ≤ 2.5 microns
PM <sub>2.5</sub>	PAL	Aluminum
PM <sub>2.5</sub>	PCA	Calcium
PM <sub>2.5</sub>	PCL	Chloride
PM <sub>2.5</sub>	PFE	Iron
PM <sub>2.5</sub>	PK	Potassium
PM <sub>2.5</sub>	PH2O	Water
PM <sub>2.5</sub>	PMG	Magnesium
PM <sub>2.5</sub>	PMN	Manganese
PM <sub>2.5</sub>	PMOTHR	PM <sub>2.5</sub> not in other AE6 species
PM <sub>2.5</sub>	PNA	Sodium
PM <sub>2.5</sub>	PNCOM	Non-carbon organic matter
PM <sub>2.5</sub>	PNH4	Ammonium
PM <sub>2.5</sub>	PSI	Silica
PM <sub>2.5</sub>	PTI	Titanium

The TOG and PM<sub>2.5</sub> profiles used to speciate emissions are part of the SPECIATE v5.2 database (<https://www.epa.gov/air-emissions-modeling/speciate>). The SPECIATE database is developed and maintained by 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). These profiles are processed using the EPA's S2S-Tool (<https://github.com/USEPA/S2S-Tool>) to generate the GSPRO and GSCNV files needed by SMOKE. As with previous platforms, some Canadian point source inventories are provided from Environment Canada as pre-specified emissions.

Speciation profiles (GSPRO files) and cross-references (GSREF files) for this study platform are available in the SMOKE input files for the platform. Emissions of VOC and PM<sub>2.5</sub> emissions by county, sector, and profile for all sectors other than onroad mobile can be found in the sector summaries. Total emissions for each model species by state and sector can be found in the state-sector totals workbook.

The following updates to profile assignments were made to this modeling platform and vary from prior years:

- For PM2.5:
  - The profile for grass fires was updated to profile 95809.
  - The profile for hydrogen boilers was updated to a gas combustion profile.
  - Assignments for new PM2.5 SCCs in the 2020 point and nonpoint inventories were included.
- For VOC:
  - The profile for wildfires and prescribed fires was updated to profile 95861.
  - Assignments for new VOC SCCs in the 2020 point and nonpoint inventories were included (e.g., agricultural silage and asphalt paving).
  - Several point and nonpoint SCCs which were previously assigned the overall average profile were reassigned to more appropriate profiles.

The base emissions inventory for this modeling platform includes total VOC and individual HAP emissions. Often, individual HAPs are components of VOC (HAP-VOC), and these HAP-VOCs are included (“integrated”) in the speciation process. This HAP integration is performed in a way to ensure double counting of emitted mass does not occur and requires specific data processing by the S2S-Tool and user input in SMOKE.

To incorporate HAP emissions from the base inventory into the modeling platform, one of two methods are performed. (1) Integrate, HAP-use is a method where the mass of integrated HAP-VOCs is summed and subtracted from VOC, and the residual mass (NONHAPVOC) is speciated using a renormalized speciation profile that does not include the integrated HAP-VOCs (they are subtracted from the profile and then the profile is renormalized to 100%). (2) No-Integrate, HAP-use is a method where the mass of VOC is speciated using a speciation profile that does not include the integrated HAP-VOCs (they are subtracted from the profile and the profile is not renormalized to 100%). In this scenario, the HAP-VOC and VOC portions of the inventory are difficult to harmonize, and it is assumed that the proportions of HAPs from these sources are adequately captured in the speciation profile used to speciate the VOC emissions (which is why there is no renormalization). In addition, HAPs can be introduced into a modeling platform using speciation profiles. In this scenario, HAP-VOC emissions are “generated” through VOC speciation and are not incorporated from the base inventory. This method is called “Criteria” speciation. The integration methods used for each platform sector are shown in Table 3-8.

**Table 3-8. Integration status for each platform sector**

<b>Platform Sector</b>	<b>Approach for Integrating NEI emissions of Naphthalene (N), Benzene (B), Acetaldehyde (A), Formaldehyde (F) and Methanol (M)</b>
afdust	N/A – sector contains no VOC
airports	No integration, use NBAFM in inventory
beis	N/A – sector contains no inventory pollutant “VOC”; but rather specific VOC species
cmv_c1c2	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
cmv_c3	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
fertilizer	N/A – sector contains no VOC
livestock	Full integration (NBAFM)
nonpt	Partial integration (NBAFM)
nonroad	Full integration (internal to MOVES)
np_oilgas	Partial integration (NBAFM)
onroad	Full integration (internal to MOVES)
canada_onroad	No integration, no NBAFM in inventory, create NBAFM from VOC speciation



<b>Platform Sector</b>	<b>Approach for Integrating NEI emissions of Naphthalene (N), Benzene (B), Acetaldehyde (A), Formaldehyde (F) and Methanol (M)</b>
mexico_onroad	Full integration (internal to MOVES-Mexico); however, MOVES-MEXICO speciation was older CB6, so post-SMOKE emissions were converted to CB6R3AE6
canada_afdust	N/A – sector contains no VOC
canmex_area	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
canmex_point	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
canada_ptdust	N/A – sector contains no VOC
canada_og2D	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
canmex_ag	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
pt_oilgas	No integration, use NBAFM in inventory
ptagfire	Full integration (NBAFM)
ptegu	No integration, use NBAFM in inventory
ptfire-rx	Full integration (NBAFM)
ptfire-wild	Partial integration (NBAFM)
ptfire_othna	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
ptnonipm	No integration, use NBAFM in inventory
rail	Full integration (NBAFM)
rcw	Full integration (NBAFM)
np_solvents	Partial integration (NBAFM)

The HAPs integrated from the base inventory into the modeling platform are sector and chemical mechanism specific. In recent years, CB6R3\_AE7 has been the primary chemical mechanism used at the EPA. Within that mechanism, naphthalene (NAPH), benzene (BENZ), acetaldehyde (ALD2), formaldehyde (FORM), and methanol (MEOH) are explicit HAP-VOCs, and these compounds are collectively referred to as NBAFM. Since NBAFM are explicitly modeled in CB6R3\_AE7, these species have become the default collection of integrated HAP species at the EPA. MOVES, the EPA's mobile emissions model, features additional species that are explicitly modeled (e.g., ethanol). These species (are also incorporated directly into modeling platforms if they are explicit in CB6R3\_AE7. To incorporate these species, additional files from the S2S-Tool are required. For California, speciation of NONHAPTOG is performed on CARB's VOC submissions using the county-specific speciation profile assignments generated by MOVES in California.

Several sectors require VOC speciation to occur at the county-level and consistent speciation profiles cannot be applied across the nation. To accomplish this, the GSREF\_COMBO functionality within SMOKE is leveraged. A GSREF\_COMBO allows profiles to be "blended" at the county/SCC-level using proportions included in the input file. These variable VOC speciation methods are applied in the oil and gas sector and for various mobile emissions sources. In both the np\_oilgas and pt\_oilgas sector, VOC speciation profiles are weighted to reflect region-specific application of controls, differences in gas composition, and variable sources of emissions (e.g., varying proportions of emissions from associated gas, condensate tanks, crude oil tanks, dehydrators, liquids unloading and well completions). The Nonpoint Oil and Gas Emissions Estimation Tool generates an intermediate file that provides SCC and county-specific emissions proportions, which are subsequently incorporated into the modeling platform.

For onroad and nonroad mobile sources, the VOC speciation weighting factors vary for each SCC, representative county, emissions mode (e.g., exhaust, evaporative), month for start exhaust, and season. To generate onroad emissions and perform the subsequent speciation, SMOKE-MOVES is first run to estimate emissions and both the MEPROC and INVTABLE files are used to control which pollutants are

processed and eventually integrated. Next, a MOVES post-processing tool is used to generate the needed GSREF\_COMBO data/files. While similar in nature and outcome, the post-processing tools/scripts used for onroad and nonroad are different. This script allows speciation to occur outside of MOVES, which better supports processing of onroad emissions for chemical mechanisms other than CB6, without having to rerun the MOVES model. From there, the NONHAPTOG emission factor tables produced by MOVES are speciated within SMOKE using the GSREF\_COMBO file and the NONHAPTOG GSPRO files generated by the S2S-Tool. For further details on speciation methods involving MOVES can be found in the [associated technical report](#).

In Canada, a GSPRO\_COMBO file is used to generate speciated gasoline emissions that account for various ethanol mixes. In Mexico, onroad emissions are pre-speciated from the MOVES-Mexico model, thus eliminating the need for a GSPRO\_COMBO file. For both Canada and Mexico, nonroad VOC emissions are not defined by mode (e.g., exhaust versus evaporative), which necessitates the need for a GSPRO\_COMBO file that splits total VOC into exhaust and evaporative components. In addition, MOVES-Mexico uses an older version of MOVES that is hardcoded for an older version of the CB6 chemical mechanism (“CB6-CAMx”). This version does not generate the model species XYLMN or SOAALK, so additional post-processing is performed to generate those emissions:

- $XYLMN = XYL[1] - 0.966 * NAPHTHALENE[1]$
- $PAR = PAR[1] - 0.00001 * NAPHTHALENE[1]$   
 $SOAALK = 0.108 * PAR[1]$

Unlike VOC speciation, PM<sub>2.5</sub> speciation does not integrate species from the base inventory. Except for mobile sources, speciation is performed within SMOKE, using SPECIATE profiles that were post-processed using the S2S-Tool. In this modeling platform, onroad PM<sub>2.5</sub> speciation is performed within MOVES, meaning that the model generates emissions factor tables that include total PM<sub>2.5</sub> and each of its components (e.g., POC, PEC, PFE, etc.). Nonroad PM<sub>2.5</sub> speciation is also performed within MOVES, but the output is not speciated emissions. Rather, MOVES outputs emissions of PM<sub>2.5</sub> for each relevant speciation profile. Small adjustments to the methods were needed to accommodate the reporting by California. Since California does not provide speciated PM<sub>2.5</sub> emissions, total PM<sub>2.5</sub> emissions for onroad and nonroad sources in California were speciated using the profile proportions estimated by MOVES in California. Finally, onroad brake and tire wear PM<sub>2.5</sub> emissions were speciated in the *moves2smk* postprocessor using the SPECIATE profiles 95462 and 95460, respectively.

Diesel PM emissions are explicitly included in the NEI using the pollutant names DIESEL-PM10 and DIESEL-PM25 for select mobile sources whose engines burn diesel or residual-oil fuels. This includes sources in onroad, nonroad, point airport ground support equipment, point locomotives, nonpoint locomotives, and all PM from diesel or residual oil fueled nonpoint CMV. These emissions are equal to their primary PM10-PRI and PM25-PRI counterparts, are exclusively from exhaust (i.e., do not include brake/tire wear), and are exclusively used in toxics modeling. Diesel PM is then speciated in SMOKE using the same speciation profiles and methods as primary PM, except that diesel PM is mapped to model species that feature “DIESEL\_PM” in their species name.

In the NEI, NO<sub>x</sub> emissions are inventoried on a NO<sub>2</sub> weighted basis, but must be speciated into NO, NO<sub>2</sub>, and HONO. Table 3-9 provides the NO<sub>x</sub> speciation profiles used in EPA’s modeling platforms. The only difference between the two profiles is the allocation of some NO<sub>2</sub> mass to HONO in the “HONO” profile. HONO emissions from mobile sources have been identified in tunnel studies and its inclusion in emissions inventories is important for urban chemistry. Here, a HONO to NO<sub>x</sub> ratio of 0.008 was selected

(Sarwar, 2008). In this modeling platform, all non-mobile sources use the “NHONO” profile, all non-onroad mobile sources (including nonroad, cmv, and rail) use the “HONO” profile, and all onroad NO<sub>x</sub> speciation occurs within MOVES. For further details on NO<sub>x</sub> speciation within MOVES, please see the [associated technical report](#).

**Table 3-9. NO<sub>x</sub> 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

### 3.3.5 Temporal Processing Configuration

Temporal allocation is the process of distributing aggregated emissions to a finer temporal resolution, thereby converting annual emissions to hourly emissions as is required by CMAQ. 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. Temporal allocation takes these aggregated emissions and distributes the 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, with monthly and day-of-week profiles applied only if the inventory is not already at that level of detail.

The temporal factors applied to the inventory were selected using some combination of country, state, county, SCC, and pollutant. Table 3-10 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 3-10. 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
airports	Annual	Yes	week	week	Yes
beis	Hourly		n/a	all	No
cmv_c1c2	Annual & hourly		All	all	No
cmv_c3	Annual & hourly		All	all	No
fertilizer	Monthly		met-based	All	Yes
livestock	Annual	Yes	met-based	All	Yes

<b>Platform sector short name</b>	<b>Inventory resolutions</b>	<b>Monthly profiles used?</b>	<b>Daily temporal approach</b>	<b>Merge processing approach</b>	<b>Process holidays as separate days</b>
nonpt	Annual	Yes	week	week	Yes
nonroad	Monthly		mwdss	mwdss	Yes
np_oilgas	Annual	Yes	aveday	aveday	No
onroad	Annual & monthly <sup>1</sup>		all	all	Yes
onroad_ca_adj	Annual & monthly <sup>1</sup>		all	all	Yes
canada_afdust	Annual & monthly	Yes	week	all	No
canmex_area	Monthly		week	week	No
canada_onroad	Monthly		week	week	No
mexico_onroad	Monthly		week	week	No
canmex_point	Monthly	Yes	mwdss	mwdss	No
canada_ptdust	Annual	Yes	week	all	No
canmex_ag	Annual	Yes	mwdss	mwdss	No
canada_og2D	Monthly		mwdss	mwdss	No
pt_oilgas	Annual	Yes	mwdss	mwdss	Yes
ptegu	Annual & hourly	Yes <sup>2</sup>	all	All	No
ptnonipm	Annual	Yes	mwdss	mwdss	Yes
ptagfire	Daily		all	all	No
ptfire-rx	Daily		all	all	No
ptfire-wild	Daily		all	all	No
ptfire_othna	Daily		all	all	No
rail	Annual	Yes	aveday	aveday	No
rcw	Annual	No <sup>3</sup>	met-based <sup>3</sup>	all	No <sup>3</sup>
np_solvents	Annual	Yes	aveday	aveday	No

1. Note the annual and monthly “inventory” actually refers to the activity data (VMT, VPOP, starts) 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 the table. 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 temporal allocation 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, 2020, which is intended to mitigate the effects of initial condition concentrations. The ramp-up period was 10 days (December 22-31, 2019). For all anthropogenic sectors, emissions from December 2020 were used to fill in surrogate emissions for the end of December 2019. For biogenic emissions, December 2019 emissions were computed using year 2019 meteorology.

The FF10 inventory format for SMOKE provides a consolidated format for monthly, daily, and hourly emissions inventories. 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 temporal allocation applied to it; rather, it should only have month-to-day and diurnal temporal allocation. This becomes particularly important when specific sectors have a mix of annual, monthly, daily, and/or hourly inventories. The flags that control temporal allocation 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 nonroad, onroad (for activity data), and all Canada and Mexico inventories except for agriculture. Commercial marine vessels in cmv\_c3 and cmv\_c1c2 use hourly data in the FF10 files.

### ***3.3.5.1 Standard Temporal Profiles***

Some sectors use straightforward temporal profiles not based on meteorology or other factors. For the ptfire, ptgfire, 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 workday 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.

Most temporal profiles in ptnonipm result in primarily constant emissions for each day of the year, although some have lower emissions on Sundays. An update in the 2018 platform was an analysis of monthly temporal profiles for non-EGU point sources in the ptnonipm sector. A number of profiles were found to be not quite flat over the months but were so close to flat that the difference was not meaningful. These profiles were replaced in the cross reference to point instead to the flat monthly profile. The codes for the profiles that were replaced were: 202, 214, 220, 221, 222, 223, 227, 257, 263, 264, 265, 266, 267, 269, 271, 272, 279, 280, 295, 302, 303, 304, 305, 306, 309, 310, 327, 329, 332, and 333.

Monthly temporalization of np\_oilgas emissions is based primarily on year-specific monthly factors from the Oil and Gas Tool (OGT). Factors were specific to each county and SCC. For use in SMOKE, each unique set of factors was assigned a label (OG20M\_0001 through OG20M\_6306), and then a SMOKE-formatted ATPRO\_MONTHLY and an ATREF were developed. This dataset of monthly temporal factors included profiles for all counties and SCCs in the Oil and Gas Tool inventory. Because we are using non-tool datasets in some states, this monthly temporalization dataset did not cover all counties and SCCs in the entire inventory used for this study. To fill in the gaps in those states, state average monthly profiles for oil, natural gas, and combination sources were calculated from Energy Information Administration (EIA) data and assigned to each county/SCC combination not already covered by the

OGT monthly temporal profile dataset. Coal bed methane (CBM) and natural gas liquid sources were assigned flat monthly profiles where there was not already a profile assignment in the ERG dataset.

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 explain 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 hours where measurable rain occurs, or where there is snow cover. Therefore, the afdust emissions vary day-to-day based on the precipitation and/or snow cover for each grid cell and hour. 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 from the BEIS model vary each 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.

For the cmv sectors, most areas use hourly emission inventories derived from the 5-minute AIS data. In some areas where AIS data are not available, such as in Canada between the St. Lawrence Seaway and the Great Lakes and in the southern Caribbean, the flat temporal profiles are used for hourly and day-of-week values. Most regions without AIS data also use a flat monthly profile, with some offshore areas using an average monthly profile derived from the 2008 ECA inventory monthly values. These areas without AIS data also use flat day of week and hour of day profiles.

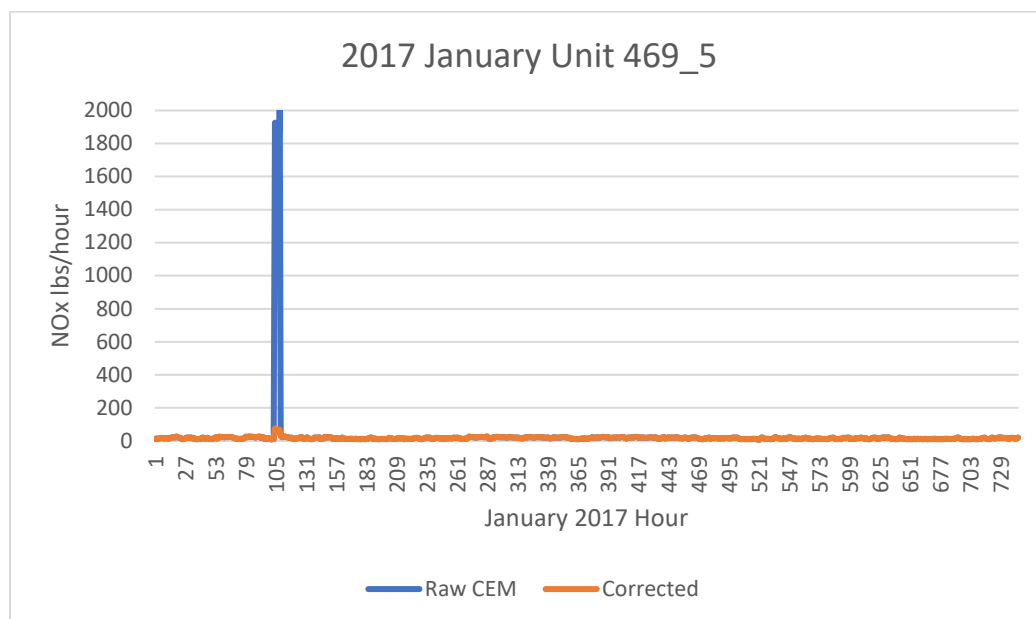
For the rail sector, monthly profiles from the 2016 platform were used. Monthly temporal allocation for rail freight emissions is based on AAR Rail Traffic Data, Total Carloads and Intermodal, for 2016. For passenger trains, monthly temporal allocation is flat for all months. Rail passenger miles data is available by month but it is not known how closely rail emissions track with passenger activity since passenger trains run on a fixed schedule regardless of how many passengers are aboard, and so a flat profile is chosen for passenger trains. Rail emissions are allocated with flat day of week profiles, and most emissions are allocated with flat hourly profiles.

For the ptfire sectors, the inventories are in the daily point fire format FF10 PTDAY, so temporal profiles are only used to go from day-specific to hourly emissions. Separate hourly profiles for prescribed and wildfires were used. For ptfire, state-specific hourly profiles were used, with distinct profiles for prescribed fires and wildfires. The wildfire diurnal profiles are similar but vary according to the average meteorological conditions in each state. 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 workday 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.

### 3.3.5.2 Temporal Profiles for EGUs

Electric generating unit (EGU) sources matched to ORIS units were temporally allocated to hourly emissions needed for modeling using the hourly CEMS data for units that could be matched to the CEMS emissions. Those hourly data were processed through v2.1 of the CEMCorrect tool to mitigate the impact of unmeasured values in the data.

The temporal allocation procedure for EGUs in the base year is differentiated by whether or not the unit could be directly matched to a unit with CEMS data via its ORIS facility code and boiler ID. Note that for units matched to CEMS data, annual totals of their emissions input to CMAQ may be different than the values in the annual inventory because the CEMS data replaces the NO<sub>x</sub> and SO<sub>2</sub> annual 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. Prior to use of the CEMS data in SMOKE it is processed through the CEMCorrect tool. The CEMCorrect tool identifies hours for which the data were not measured as indicated by the data quality flags in the CEMS data files. Unmeasured data can be filled in with maximum values and thereby cause erroneously high values in the CEMS data. When data were flagged as unmeasured and the 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 temporal allocation process described below (see Figure 3-4 for an example).



**Figure 3-4. Eliminating unmeasured spikes in CEMS data**

The region, fuel, and type (peaking or non-peaking) must be identified for each input EGU with CEMS data so the data can be used to generate profiles. The identification of peaking units was done using hourly heat input data from the 2020 base year and the two previous years (2018 and 2019). The heat input was summed for each year. Equation 1 shows how the annual heat input value is converted from heat units (BTU/year) to power units (MW) using the NEEDS v6 derived unit-level heat rate (BTU/kWh). In equation 2 a capacity factor is calculated by dividing the annual unit MW value by the NEEDS v6 unit



capacity value (MW) multiplied by the hours in the year. A peaking unit was defined as any unit that had a maximum capacity factor of less than 0.2 for every year (2018, 2019, and 2020) and a 3-year average capacity factor of less than 0.1.

**Equation 1. Annual unit power output**

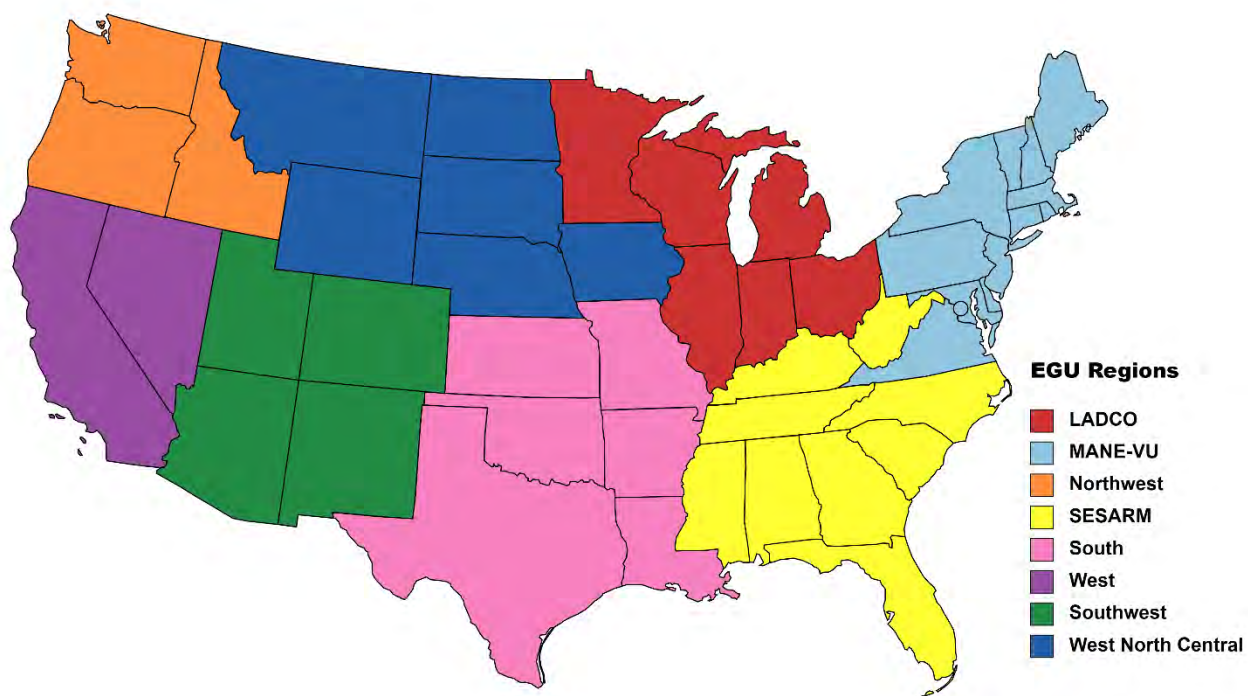
$$\text{Annual Unit Output (MW)} = \frac{\sum_{t=0}^{8760} \text{Hourly HI (BTU)} * 1000 \left( \frac{\text{MW}}{\text{kW}} \right)}{\text{NEEDS Heat Rate} \left( \frac{\text{BTU}}{\text{kWh}} \right)}$$

**Equation 2. Unit capacity factor**

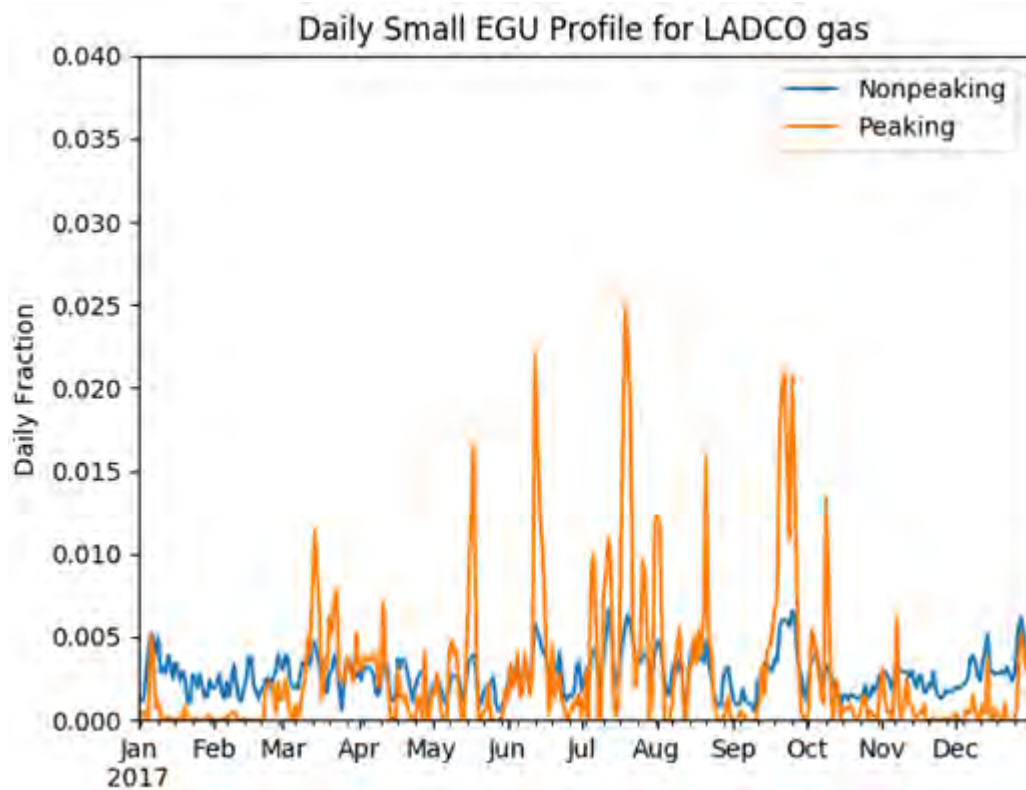
$$\text{Capacity Factor} = \frac{\text{Annual Unit Output (MW)}}{\text{NEEDS Unit Capacity} \left( \frac{\text{MW}}{\text{h}} \right) * 8760 \text{ (h)}}$$

Input regions were determined from one of the eight EGU modeling regions based on MJO and climate regions. Regions were used to group units with similar climate-based load demands. Region assignment is made on a state level, where all units within a state were assigned to the appropriate region. Unit fuel assignments were made using the primary NEEDS v6 fuel. Units fueled by bituminous, subbituminous, or lignite are assigned to the coal fuel type. Natural gas units were assigned to the gas fuel type. Distillate and residual fuel oil were assigned to the oil fuel type. Units with any other primary fuel were assigned the “other” fuel type. Figure 3-5 shows the regions used to generate the profiles. Unit fuel assignments were made using the primary NEEDS v6 fuel. Units fueled by bituminous, subbituminous, or lignite are assigned to the coal fuel type. Natural gas units were assigned to the gas fuel type. Distillate and residual fuel oil were assigned to the oil fuel type. Units with any other primary fuel were assigned the “other” fuel type. Currently there are 64 profiles based on 8 regions, 4 fuels, and two types (peaking and non-peaking).

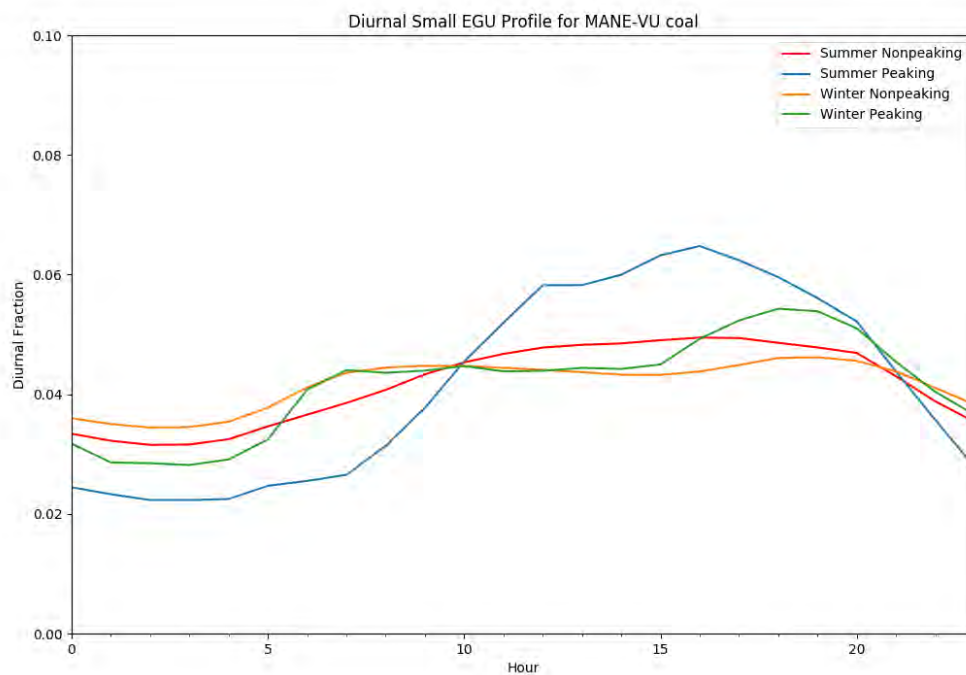
The daily and diurnal profiles were calculated for each region, fuel, and peaking type group from the year 2020 CEMS heat input values. The heat input values were summed for each input group to the annual level at each level of temporal resolution: monthly, month-of-day, and diurnal. The sum by temporal resolution value was then divided by the sum of annual heat input in that group to get a set of temporalization factors. Diurnal factors were created for both the summer and winter seasons to account for the variation in hourly load demands between the seasons. For example, the sum of all hour 1 heat input values in the group was divided by the sum of all heat inputs over all hours to get the hour 1 factor. Each grouping contained 12 monthly factors, up to 31 daily factors per month, and two sets of 24 hourly factors. The profiles were weighted by unit size where the units with more heat input have more influence on the shape of the profile. Composite profiles were created for each region and type across all fuels as a way to provide profiles for a fuel type that does not have hourly CEMS data in that region. Figure 3-6 shows peaking and non-peaking daily temporal profiles for the gas fuel type in the LADCO region. Figure 3-7 shows the diurnal profiles for the coal fuel type in the Mid-Atlantic/Northeast Visibility Union (MANE VU) region.



**Figure 3-5. Small EGU Temporal Profile Regions**



**Figure 3-6. Example Daily Temporal Profiles for the LADCO region and Gas Fuel Type**



**Figure 3-7. Example Diurnal Profile for MANE-VU Region and Coal Fuel Type**

SMOKE uses a cross-reference file to select a monthly, daily, and diurnal profile for each source. For the 2020 platform, the temporal profiles were assigned in the cross-reference at the unit level to EGU sources without hourly CEMS data. An inventory of all EGU sources without CEMS data was used to identify the region, fuel type, and type (peaking/non-peaking) of each source. The region used to select the temporal profile is assigned based on the state from the unit FIPS. The fuel was assigned by SCC to one of the four fuel types: coal, gas, oil, and other. A fuel type unit assignment is made by summing the VOC, NOX, PM2.5, and SO2 for all SCCs in the unit. The SCC that contributed the highest total emissions to the unit for selected pollutants was used to assign the unit fuel type. Peaking units were identified as any unit with an oil, gas, or oil fuel type with a NAICS of 22111 or 22112. Some units may be assigned to a fuel type within a region that does not have an available input unit with a matching fuel type in that region. These units without an available profile for their group were assigned to use the regional composite profile. MWC and cogen units were identified using the NEEDS primary fuel type and cogeneration flag, respectively, from the NEEDS v6 database. Assignments for each unit needed a profile were made using the regions shown in Figure 3-5.

### ***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 a method for temporal allocation 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 temporal allocation.

The SMOKE program Gentpro provides a method for developing meteorology-based temporal allocation. Currently, the program can utilize three types of temporal algorithms: annual-to-day temporal allocation for residential wood combustion (RWC); month-to-hour temporal allocation for agricultural livestock NH<sub>3</sub>; and a generic meteorology-based algorithm for other situations. Meteorological-based temporal allocation was used for portions of the rwc sector and for all agricultural sources. For 2020, some new temporal profiles were introduced for livestock that differ by animal type and county.

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](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.

For the RWC sector, two different algorithms for calculating temporal allocation are used. For most SCCs in the sector, in which wood burning is more prominent on colder days, Gentpro was used to compute annual to day-of-year temporal profiles based on the daily minimum temperature. These 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

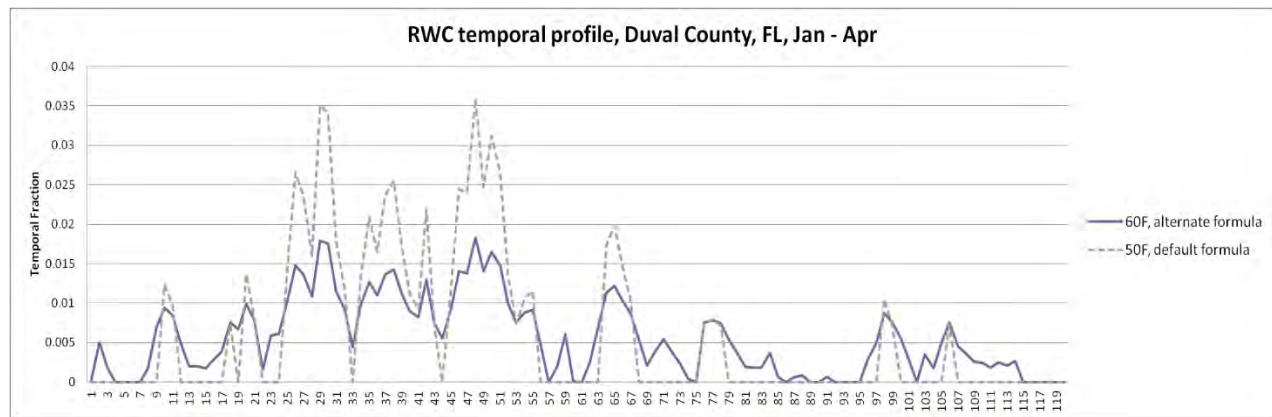
of the country, and 60 °F for the following states: Alabama, Arizona, California, Florida, Georgia, Louisiana, Mississippi, South Carolina, and Texas. The algorithm is as follows:

If  $T_d \geq T_t$ : no emissions that day  
 If  $T_d < T_t$ : daily factor =  $0.79 \cdot (T_t - T_d)$

where ( $T_d$  = minimum daily temperature;  $T_t$  = threshold temperature, which is 60 degrees F in southern states and 50 degrees F elsewhere).

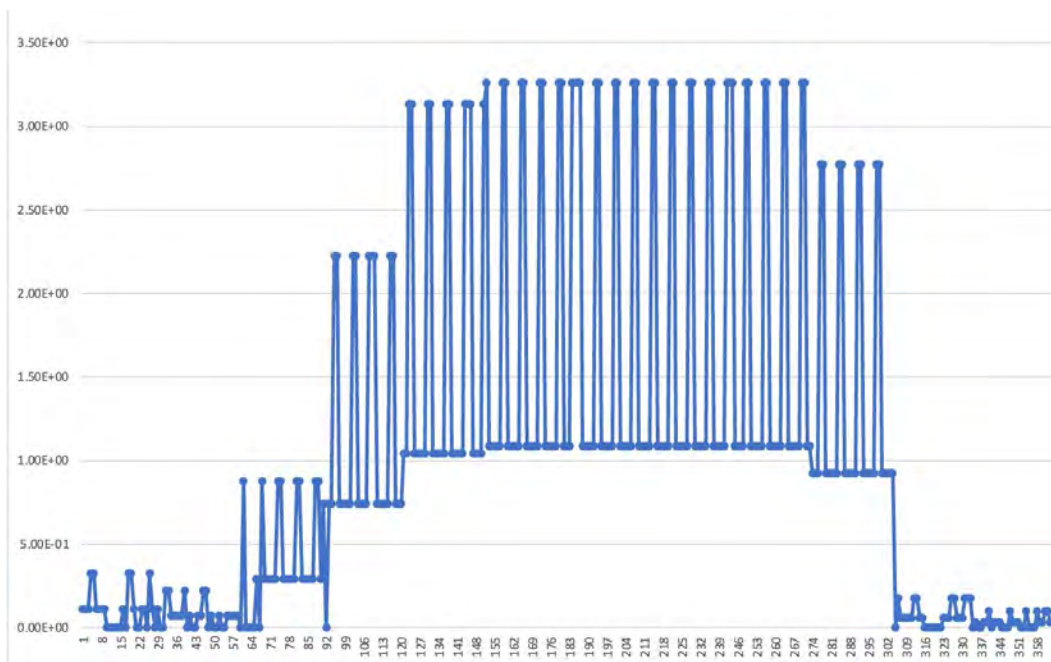
Once computed, the factors were normalized to sum to 1 to ensure that the total annual emissions are unchanged (or minimally changed) during the temporal allocation process.

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 using a 50 °F versus 60°F threshold

For the 2020 emissions modeling platform, a separate algorithm is used to determine temporal allocation of recreational wood burning, e.g. fire pits (SCC 2104008700) and is applied by Gentpro. Recreational wood burning depends on both minimum and maximum daily temperatures by county, and also uses a day-of-week temporal profile (61500) in which emissions are much higher on weekends than on weekdays. According to the recreational wood burning algorithm, only days in which the temperature falls within a range of 50°F and 80°F at some point during the day receive emissions. On days when the maximum temperature is less than 50°F or the minimum temperature is above 80°F, the daily temporal factor is zero. For all other days, the day-of-week profile 61500 is applied, which has 33% of the emissions on each weekend day and lower emissions on weekdays. An example is shown in Figure 3-9. As a result of applying this algorithm, northern states have more recreational wood burning in summer months while southern states show a flatter pattern with emissions distributed more evenly throughout the months.



**Figure 3-9.** Example of RWC temporalization using a 50 °F 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](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 temporal profiles for hydronic heaters” (i.e., SCCs=2104008610 [outdoor], 2104008620 [indoor], and 2104008620 [pellet-fired]) are not based on temperature data, because the meteorologically based temporal allocation used for the rest of the rwc sector did not agree with observations for how these appliances are used.

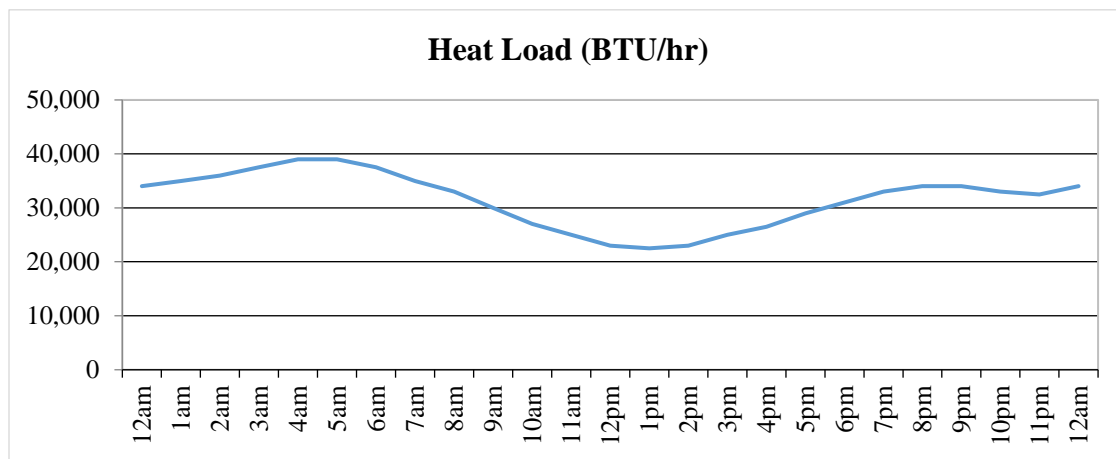
For hydronic heaters, the annual-to-month, day-of-week and diurnal profiles were modified based on information in the New York State Energy Research and Development Authority’s (NYSERDA) “Environmental, Energy Market, and Health Characterization of Wood-Fired Hydronic Heater Technologies, Final Report” (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 3-10 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.

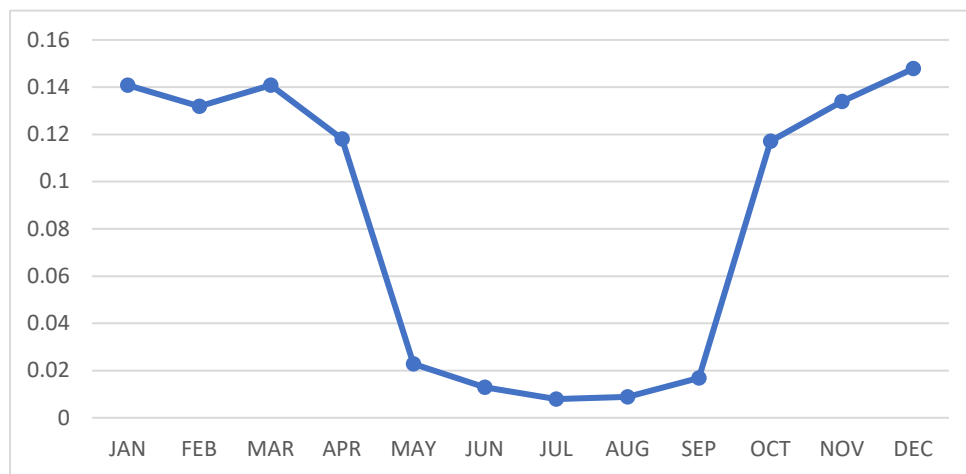


In contrast, the day-of-week profile for recreational RWC follows a typical “recreational” profile with emissions peaked on weekends. Annual-to-month temporalization for OHH as well as recreational RWC were computed from the MN DNR survey (MDNR, 2008) and are illustrated in Figure 3-11. OHH emissions 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.

Annual-to-month temporal allocation for OHH was computed from the MDNR 2008 survey and is illustrated in Figure 3-10. There are two types of hydronic heaters 2104008620 (indoor hydronic heaters) and 2104008630 (pellet-fired hydronic heaters). Both of these SCCs use the same monthly, weekly, and diurnal temporal profiles as OHHs as is shown in Figure 3-11.



**Figure 3-10.** Diurnal profile for OHH, based on heat load (BTU/hr)



**Figure 3--11.** Annual-to-month temporal profiles for Outdoor Hydronic Heaters

For the ag sector, agricultural GenTPRO temporal allocation was applied to livestock emissions and to all pollutants within the sector, not just NH<sub>3</sub>. 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 NH<sub>3</sub> emission variations from livestock as a function of ambient temperature, aerodynamic resistance, and wind speed. The equations are:

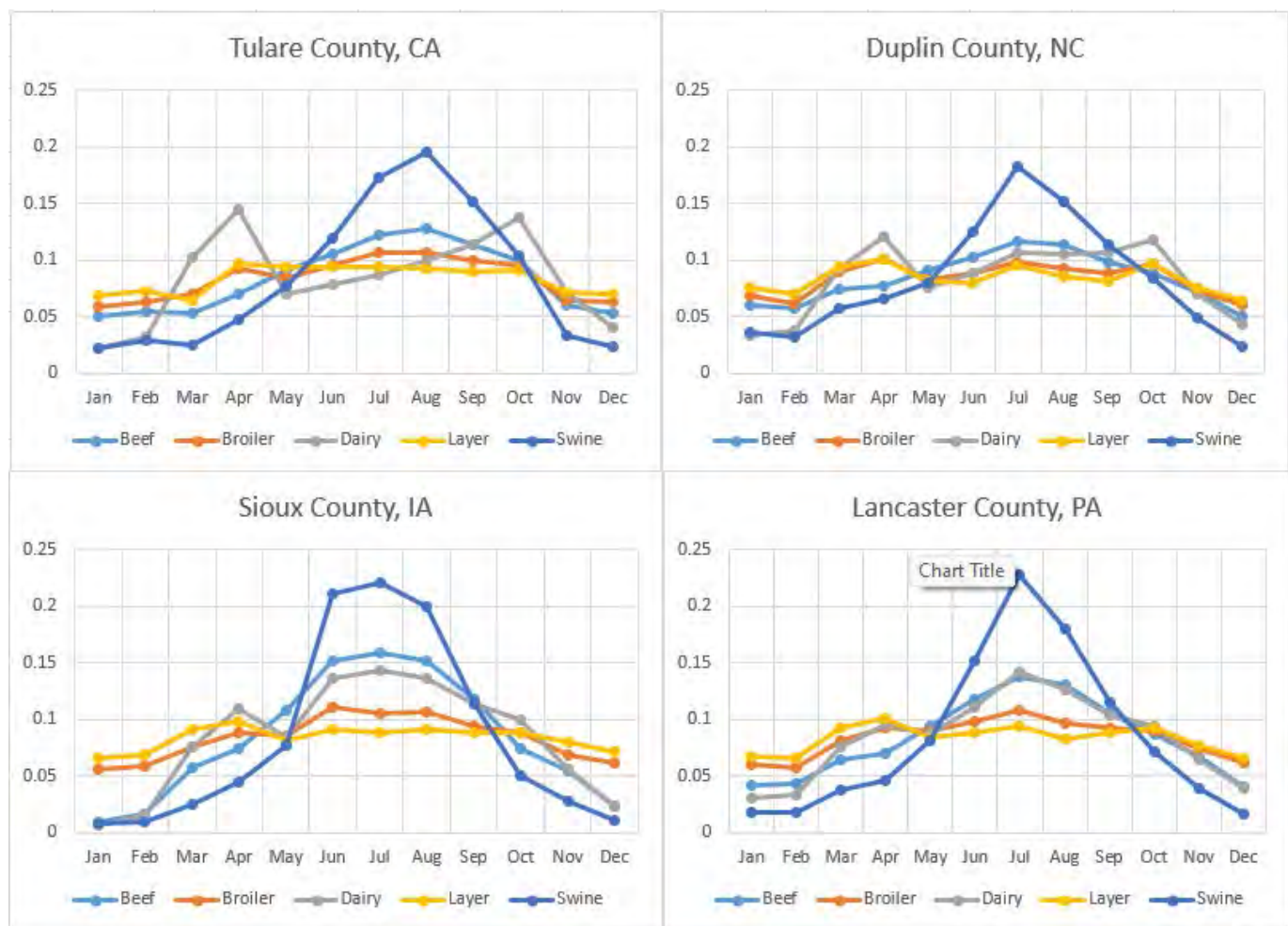
$$E_{i,h} = [161500/T_{i,h} \times e^{(-1380/T_{i,h})}] \times AR_{i,h}$$

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

where

- $PE_{i,h}$  = Percentage of emissions in county  $i$  in hour  $h$
- $E_{i,h}$  = Emission rate in county  $i$  in hour  $h$
- $T_{i,h}$  = Ambient temperature (Kelvin) in county  $i$  in 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  $i$

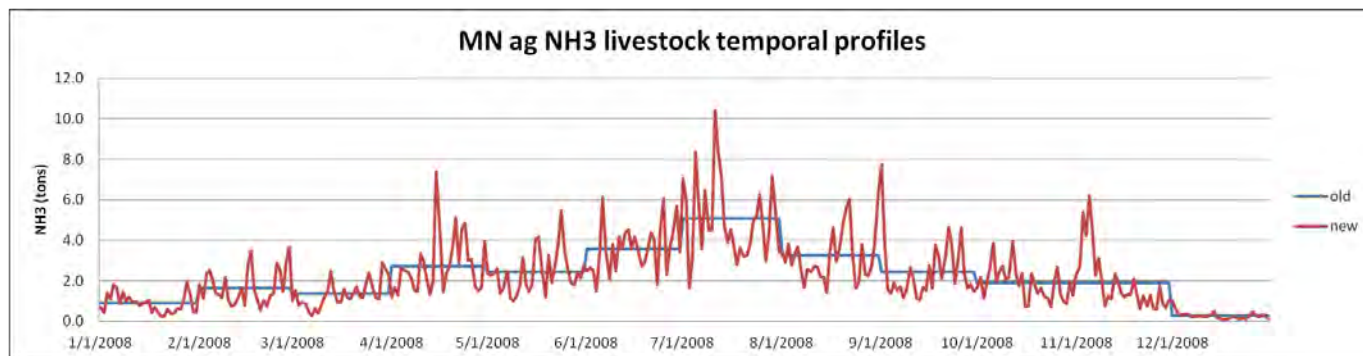
Some examples plots of the profiles by animal type in different parts of the country are shown in Figure 3-12.



**Figure 3-12. Examples of livestock temporal profiles in several parts of the country**

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 were obtained from a monthly inventory, or from an annual inventory that has been temporalized to the month. Figure 3-13 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 3-13. Example of animal NH<sub>3</sub> 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 measurable 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.

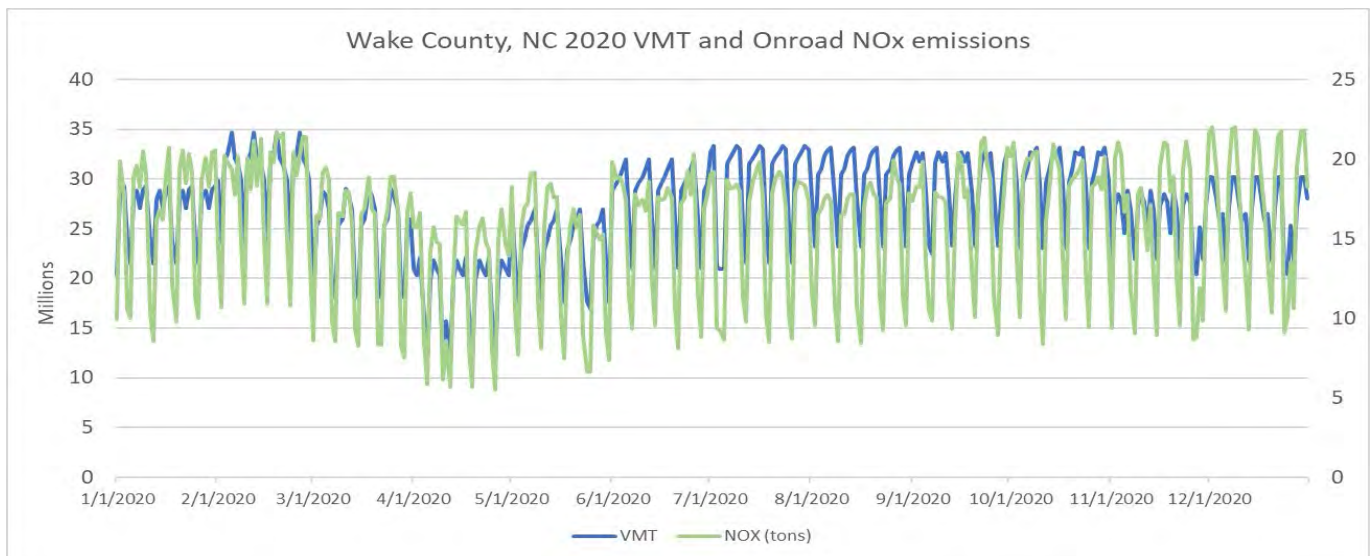
#### **3.3.5.4 Temporal Profiles for Onroad Mobile Sources**

For the onroad sector, the temporal distribution of emissions is a combination of traditional temporal profiles and the influence of meteorology. For the 2020 NEI EPA purchased county-level telematics data from StreetLight for characterization of vehicle speed profiles and VMT temporal distributions for 2020. Temporal profiles for speeds by road type were obtained by month, day of week, and hour. Vehicle types included personal, commercial medium-duty, and commercial heavy-duty. This section will discuss both the meteorological influence and the development of the temporal profiles for this platform.

The “inventories” for onroad consist of activity data for the onroad sector, not emissions. VMT is the activity data used for on-network rate-per-distance (RPD) processes. For the off-network emissions from the rate-per-profile (RPP) and rate-per-vehicle (RPV) processes, the VPOP activity data are annual and do not need temporal allocation. For rate-per-hour (RPH) processes that result from hoteling of combination trucks, the HOTELING inventory is annual and was temporalized to month, day of the week, and hour of the day through temporal profiles. Day-of-week and hour-of-day temporal profiles are also used to temporalize the starts activity used for rate-per-start (RPS) processes, and the off-network idling (ONI)

hours activity used for rate-per-hour-ONI (RPHO) processes. The inventories for starts and ONI activity contain monthly activity so that monthly temporal profiles are not needed.

For on-roadway RPD processes, the VMT activity data are 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 also use hourly speed distributions (SPDIST). For onroad, the temporal profiles and SPDIST will impact not only the distribution of emissions through time but also the total emissions. SMOKE-MOVES calculates emissions for RPD processed based on the VMT, speed and meteorology. Thus, if the VMT or speed data were shifted 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 temporal allocation of VMT changes. Figure 3-14 illustrates the temporal allocation of the onroad activity data (i.e., VMT) and the pattern of the emissions that result after running SMOKE-MOVES. In this figure, it can be seen that the meteorologically varying emission factors add variation on top of the temporal allocation of the activity data.



**Figure 3-14. Example temporal variability of VMT compared to onroad NO<sub>x</sub> emissions**

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 and stationary vehicle (RPV, RPH, RPHO, RPS, and RPP) processes use the gridded meteorology (MCIP) either directly or indirectly. For RPD, RPV, RPH, RPHO, and RPS, Movesmrg determines the temperature for each hour and grid cell and uses that information to select the appropriate emission factor for the specified SCC/pollutant/mode combination. For RPP, instead of reading gridded hourly meteorology, Movesmrg reads gridded daily minimum and maximum temperatures. The total of the emissions from the combination of these six processes (RPD, RPV, RPH, RPHO, RPS, and RPP) comprise the onroad sector emissions. In summary, the temporal patterns of emissions in the onroad sector are influenced by meteorology.

Day-of-week and hour-of-day temporal profiles for VMT were developed for use in the 2020 NEI using data acquired from StreetLight. Data were provided for three vehicle categories: passenger vehicles (11/21/31), commercial trucks (32/52), and combination trucks (53/61/62). StreetLight data did not cover buses, refuse trucks, or motor homes, so those vehicle types were mapped to other vehicle types as follows: 1) other/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. In addition to temporal profiles, StreetLight data were also used to develop the hourly speed distributions (SPDIST) used by SMOKE-MOVES.

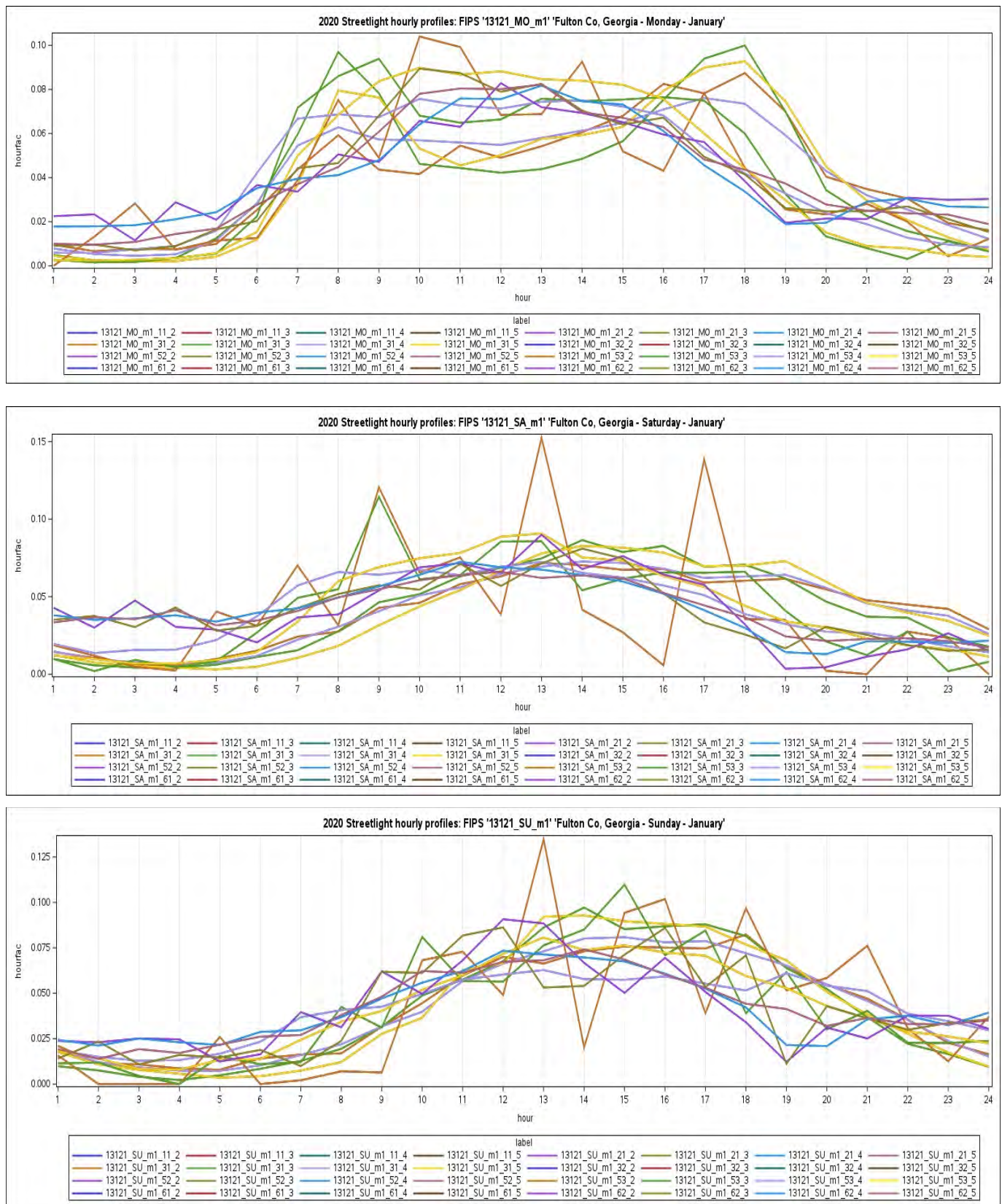
The StreetLight dataset includes temporal profiles for individual counties. Temporal profiles also vary by each of the MOVES road types, and there are distinct hour-of-day profiles for each day of the week. Plots of hour-of-day profiles for all vehicles and road types in Fulton County, GA, are shown in Figure 3-15. Separate plots are shown for Monday, Saturday, and Sunday in January 2020, and each line corresponds to a particular MOVES road type (i.e., road type 2 = rural restricted, 3 = rural unrestricted, 4 = urban restricted, and 5 = urban unrestricted) and vehicle type (as described in the previous paragraph). In the pre-pandemic profiles shown in this figure, there are bimodal peaks for light-duty vehicles on Monday, but there is only a single peak on the weekend days.

State/local-provided data for the 2020 NEI were accepted for use in the 2020 NEI if they were deemed to be at least as credible as the StreetLight data (i.e., reflected the effects of COVID). The 2020 NEI TSD includes more details on which data were used for which counties. In areas of the contiguous United States where state/local-provided data were not provided or deemed unacceptable, the StreetLight temporal profiles were used, including in California. The StreetLight temporal profiles were used in areas of the contiguous United States that did not submit temporal profiles of sufficient detail for the 2020 NEI. For this platform, the data selection hierarchy favored local input data over EPA-developed information, with the exception of the three MOVES tables `hourVMTFraction`, `dayVMTFraction`, and `avgSpeedDistribution` where county-level, telematics-based EPA Defaults were adopted for the NEI universally due to unique activity patterns by month during 2020.

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.

Temporal profiles for RPHO are based on the same temporal profiles as the on-network processes in RPD, but since the on-network profiles are road-type-specific and ONI is not road-type-specific, the RPHO profiles were assigned to use rural unrestricted profiles for counties considered "rural" and urban unrestricted profiles for counties considered "urban". RPS uses the same day-of-week profiles as on-network processes in RPD, but uses a separate set of diurnal temporal profiles specifically for starts activity. For starts, there are two hour-of-day temporal profiles for each source type, one for weekdays and one for weekends. The starts diurnal temporal profiles are applied nationally and are based on the default starts-hour-fraction tables from MOVES.

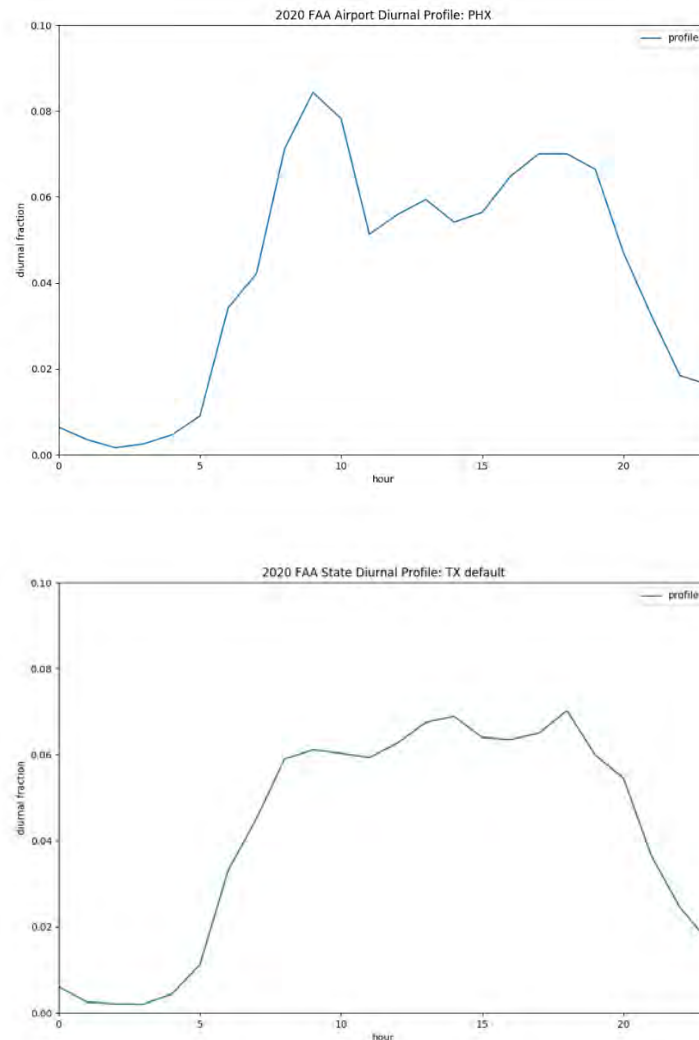




**Figure 3-15. Sample onroad diurnal profiles for Fulton County, GA**

#### 3.3.5.4 Airport Temporal Profiles

Airport temporal profiles were updated to 2020-specific temporal profiles for all airports other than Alaska seaplanes (which are not in the CMAQ modeling domain). Hourly airport operations data were obtained from the Aviation System Performance Metrics (ASPM) Airport Analysis website (<https://aspm.faa.gov/apm/sys/AnalysisAP.asp>). A report of 2020 hourly Departures and Arrivals for Metric Computation by airport was generated. An overview of the ASPM metrics is at [http://aspmhelp.faa.gov/index.php/Aviation\\_Performance\\_Metrics\\_%28APM%29](http://aspmhelp.faa.gov/index.php/Aviation_Performance_Metrics_%28APM%29). Figure 3-16 shows examples of diurnal airport profiles for Phoenix airport (PHX) and the default diurnal profile for Texas.



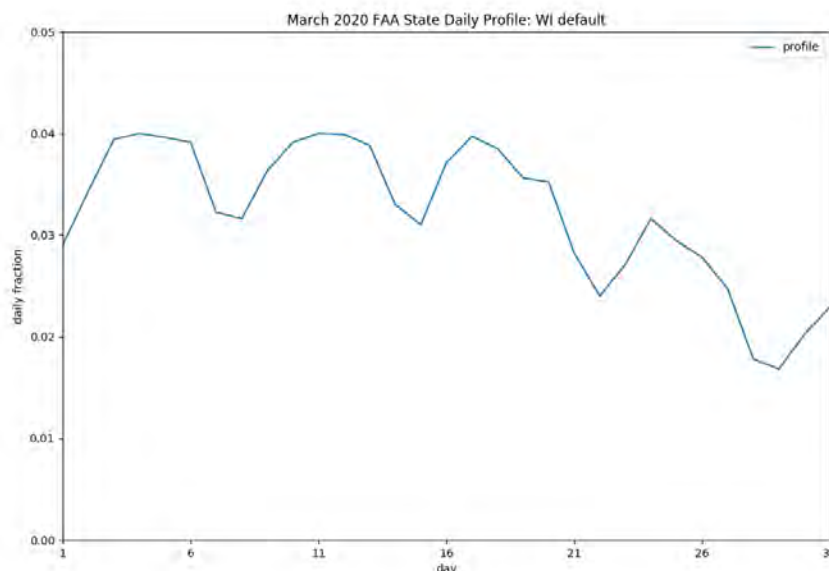
**Figure 3-16. 2020 Airport Diurnal Profiles for PHX and state of Texas**

Month-to-day and Annual-to month temporal profiles were developed based on a separate query of the 2020 Aviation System Performance Metrics (ASPM) Airport Analysis (<https://aspm.faa.gov/apm/sys/AnalysisAP.asp>). A report of all airport operations (takeoffs and landings) by day for 2020 was generated. Day-of-month profiles were derived directly from the daily airport operations report. An example is shown for Wisconsin in Figure 3-17 while Figure 3-18 shows the pre-

pandemic day of week profile. The prepandemic annual-to-month profile is shown in Figure 3-19. The 2020 airport data were summed to create the example annual-to-month temporal profiles shown in Figure 3-20.

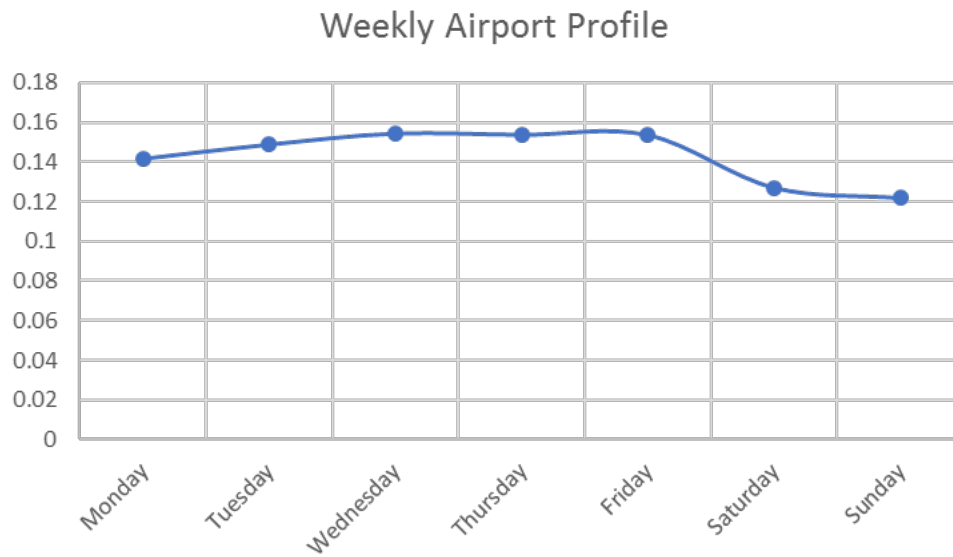
For 2020, all airport SCCs (i.e., 2275\*, 2265008005, 2267008005, 2268008005 and 2270008005) were assigned to individual commercial airports where a match could be made between the inventory facility and the FAA identifier in the ASPM derived data. State average profiles were calculated as the average of the temporal fractions for all airports within a state. The state average profiles were assigned by state to all airports in the inventory that did not have an airport specific match in the ASPM data. Package processing hubs at the Memphis (MEM), Indianapolis (IND), Louisville (SDF), and Chicago Rockford (RFD) airports produced peaks in the average state profiles at times not typical for activity in smaller commercial airports. These packaging hubs were removed from the state averages. Airports that required state-defaults in states lacking ASPM data use national average profiles calculated from the average of the state temporal profiles.

Alaska seaplanes, which are outside the CONUS domain use the monthly profile in Figure 3-21. These were assigned based on the facility ID.

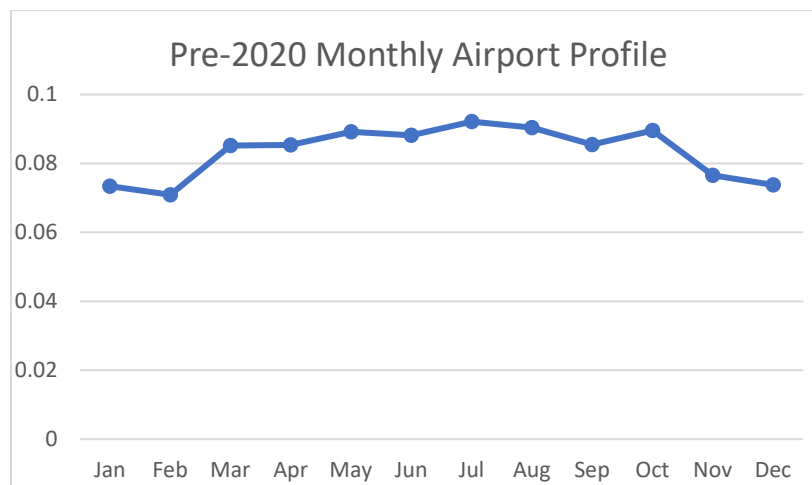


**Figure 3-17. 2020 Wisconsin month-to-day profile for airport emissions**

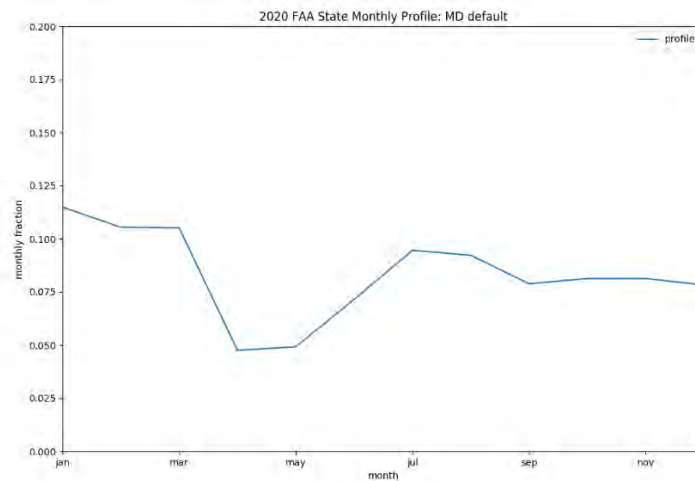
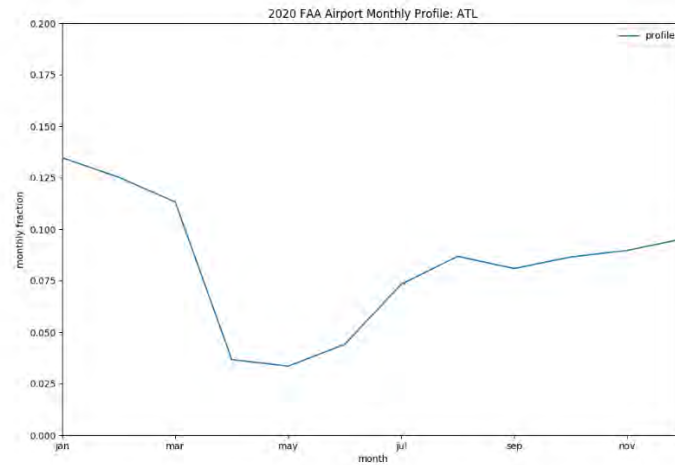




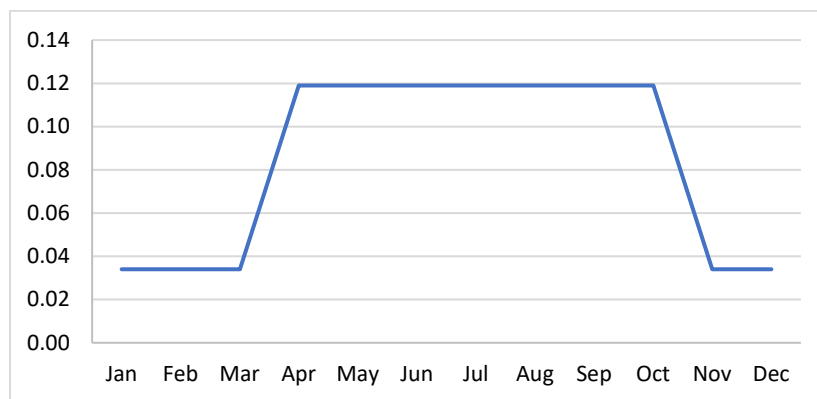
**Figure 3-18. Prepandemic weekly profile for airport emissions**



**Figure 3-19. Pre-pandemic monthly profile for airport emissions**



**Figure 3-20. 2020 Monthly airport profiles for ATL and state of Maryland**

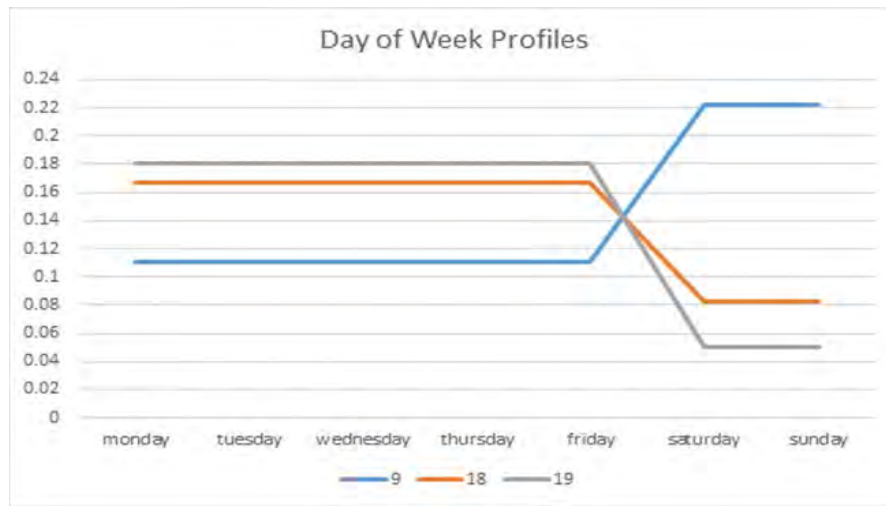


**Figure 3-21. Alaska Seaplane Profile**

### 3.3.5.5 Nonroad Temporal Profiles

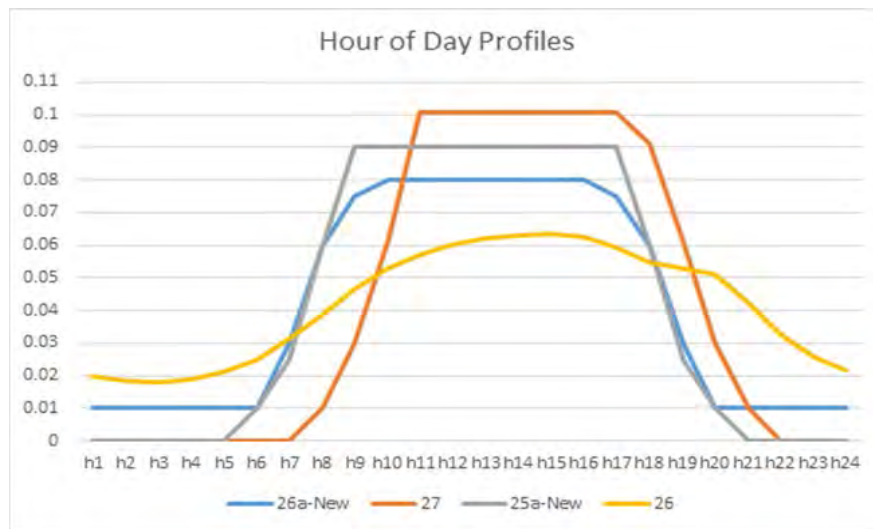
For nonroad mobile sources, temporal allocation is performed differently for different SCCs. Beginning with the final 2011 platform, improvements to temporal allocation of nonroad mobile sources were made to make the temporal profiles more realistically reflect real-world practices. The specific updates were made for agricultural sources (e.g., tractors), construction, and commercial residential lawn and garden sources.

Figure 3-22 shows two previously existing temporal profiles (9 and 18) and a newer temporal profile (19) which has lower emissions on weekends. In this platform, construction and commercial lawn and garden sources use the new profile 19 which has lower emissions on weekends. Residential lawn and garden sources continue to use profile 9 and agricultural sources continue to use profile 18.



**Figure 3-22. Example Nonroad Day-of-week Temporal Profiles**

Figure 3-23 shows the previously existing temporal profiles 26 and 27 along with newer temporal profiles (25a and 26a) which have lower emissions overnight. In this platform, construction sources use profile 26a. Commercial lawn and garden and agriculture sources use the profiles 26a and 25a, respectively. Residential lawn and garden sources use profile 27.



**Figure 3-23. Example Nonroad Diurnal Temporal Profiles**

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.

### **3.3.6 Vertical Allocation of Emissions**

Table 3-6 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, pt\_oilgas, ptfire-rx, ptfire-wild, pttagfire, 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 traditional stack parameters. However, the ptfire-rx, ptfire-wild, pttagfire, 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. Note that the implementation of fire plume rise in CMAQ differs from the implementation of plume rise in SMOKE. This study uses CMAQ to compute the fire plume rise.

### **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 each of the modeling grids shown in Section 3.1. To accomplish this, SMOKE used national 12-km spatial surrogates and a SMOKE area-to-point data file. For the U.S., the EPA updated surrogates to use circa 2020 data. The U.S., Mexican, and Canadian 12-km surrogates cover the entire CONUS domain. For Canada, shapefiles for generating new surrogates were provided by ECCC for use with their 2015 inventories. The U.S., Mexican, and Canadian 12-km surrogates cover the entire CONUS domain 12US1 shown in Figure 3-3. While highlights of information are provided below, the file *Surrogate\_specifications\_2020\_platform\_US\_Can\_Mex.xlsx* documents the complete configuration for generating the surrogates and can be referenced for more details.

### 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 3-11 lists the codes and descriptions of the surrogates. Surrogate names and codes listed in *italics* are not directly assigned to any sources for this platform, but they are sometimes used to gapfill other surrogates. When the source data for a surrogate have no values for a particular county, gap filling is used to provide values for the spatial surrogate in those counties to ensure that no emissions are dropped when the spatial surrogates are applied to the emission inventories.

The surrogates for the platform are based on a variety of geospatial data sources, including the American Community Survey (ACS) for census-related data, the National Land Cover Database (NLCD) Onroad surrogates are based on average annual daily traffic counts (AADT) from the highway monitoring performance system (HPMS).

Surrogate updates for this platform include:

- County boundaries used for all surrogates were updated to use the 2020 TIGER boundaries.
- Oil and gas surrogates were updated to represent 2020.
- ACS-based surrogates were updated to use the 2020 ACS
- Updated surrogates for residential wood combustion were developed based on ACS data
- NLCD-based surrogates were updated to use NLCD 2019.
- Animal specific livestock waste surrogates were derived from National Pollutant Discharge Elimination System (NPDES) animal operation water permits and Food and Agriculture Organization (FAO) gridded livestock count data
- New surrogates for fuel stations, asphalt surfaces, and unpaved roads were created using data from the OpenStreetMap database
- Gravel and lead mines were split out to their own surrogates from the more general United States Geological Survey mining surrogate

Surrogates for the U.S. were generated using the Surrogate Tools DB with the Java-based Surrogate tools used to perform gapfilling and normalization where needed. The tool and documentation for the original Surrogate Tool are available at [https://www.cmascenter.org/sa-tools/documentation/4.2/SurrogateToolUserGuide\\_4\\_2.pdf](https://www.cmascenter.org/sa-tools/documentation/4.2/SurrogateToolUserGuide_4_2.pdf), and the tool and documentation for the Surrogate Tools DB is available from [https://www.cmascenter.org/surrogate\\_tools\\_db/](https://www.cmascenter.org/surrogate_tools_db/). The file Surrogate\_specifications\_2020\_platform\_US\_Can\_Mex.xlsx documents the configuration for generating the surrogates

**Table 3-11. U.S. Surrogates available for the 2019 modeling platform**

Code	Surrogate Description	Code	Surrogate Description
N/A	Area-to-point approach (see 3.6.2)	672	<i>Gas production - oil wells</i>
100	Population	674	Unconventional Well Completion Counts
<i>110</i>	<i>Housing</i>	676	<i>Well count - all producing</i>
135	Detached Housing	677	<i>Well count - all exploratory</i>
136	Single and Dual Unit Housing	678	Completions at Gas Wells
150	Residential Heating - Natural Gas	679	Completions at CBM Wells

Code	Surrogate Description	Code	Surrogate Description
170	Residential Heating - Distillate Oil	681	Spud Count - Oil Wells
180	Residential Heating - Coal	683	Produced Water at All Wells
190	Residential Heating - LP Gas	6831	Produced water at CBM wells
205	Extended Idle Locations	6832	Produced water at gas wells
239	Total Road AADT	6833	Produced water at oil wells
240	Total Road Miles	685	Completions at Oil Wells
242	All Restricted AADT	686	<i>Completions - all wells</i>
244	All Unrestricted AADT	687	Feet Drilled at All Wells
258	<i>Intercity Bus Terminals</i>	689	Gas Produced - Total
259	Transit Bus Terminals	691	Well Counts - CBM Wells
261	NTAD Total Railroad Density	692	Spud Count - All Wells
271	NTAD Class 1 2 3 Railroad Density	693	Well Count - All Wells
300	NLCD Low Intensity Development	694	Oil Production at Oil Wells
304	NLCD Open + Low	695	Well Count - Oil Wells
305	NLCD Low + Med	696	Gas Production at Gas Wells
306	NLCD Med + High	697	Oil production - gas wells
307	NLCD All Development	698	Well Count - Gas Wells
308	NLCD Low + Med + High	699	Gas Production at CBM Wells
309	NLCD Open + Low + Med	711	Airport Areas
310	NLCD Total Agriculture	801	Port Areas
319	NLCD Crop Land	850	Golf Courses
320	NLCD Forest Land	860	Mines
321	NLCD Recreational Land	861	Sand and Gravel Mines
340	<i>NLCD Land</i>	862	<i>Lead Mines</i>
350	NLCD Water	863	Crushed Stone Mines
401	<i>FAO 2010 Cattle</i>	900	OSM Fuel
402	<i>FAO 2010 Pig</i>	901	OSM Asphalt Surfaces
403	<i>FAO 2010 Chicken</i>	902	OSM Unpaved Roads
404	<i>FAO 2010 Goat</i>	4011	FAO 2010 Large Cattle Operations
405	FAO 2010 Horse	4012	NPDES 2020 Beef Cattle
406	FAO 2010 Sheep	4013	NPDES 2020 Dairy Cattle
508	Public Schools	4021	NPDES 2020 Swine
650	Refineries and Tank Farms	4031	NPDES 2020 Chicken
670	Spud Count - CBM Wells	4041	NPDES 2020 Goat
671	Spud Count - Gas Wells	4071	NPDES 2020 Turkey

For the onroad sector, the on-network (RPD) emissions were spatially allocated differently from other off-network processes (i.e. RPV, RPP, RPHO, RPS, RPH). Surrogates for on-network processes are based on AADT data and off network processes (including the off-network idling included in RPHO) are based on land use surrogates as shown in Table 3-12. Emissions from the extended (i.e., overnight) idling of trucks were assigned to surrogate 205, which is based on locations of overnight truck parking spaces. The underlying data for this surrogate were updated during the development of the 2016 platforms to include additional data sources and corrections based on comments received and these updates were carried into this platform

**Table 3-12. 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	Other Bus	306	NLCD Med + High
42	Transit Bus	259	Transit Bus Terminals
43	School Bus	508	Public Schools
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
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 3-13 using 2020 data consistent with what was used to develop the nonpoint oil and gas emissions.

The exploration and production of oil and gas have increased in terms of quantities and locations over the last seven years, primarily through the use of new technologies, such as hydraulic fracturing. Census-tract, 2-km, and 4-km sub-county Shapefiles were developed, from which the 2020 oil and gas surrogates were generated. All spatial surrogates for np\_oilgas are developed based on known locations of oil and gas activity for year 2020.

The primary activity data source used for the development of the oil and gas spatial surrogates was data from ENVERUS [formerly Drilling Info (DI) Desktop's HPDI] database (ENVERUS, 2021). This database contains well-level location, production, and exploration statistics at the monthly level. Due to a proprietary agreement with ENVERUS, 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 (Alaska, Arizona, Idaho, Illinois, Indiana, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Nevada, Oregon, Pennsylvania, and Tennessee). In cases when the desired surrogate parameter was not available (e.g., feet drilled), data for an alternative surrogate parameter (e.g., number of spudded wells) were downloaded and used. Under that methodology, both completion date and date of first production from HPDI were used to identify wells completed during 2020.

The spatial surrogates, numbered 670 through 699 and also 6831, 6832, and 6833, were gapfilled using fallback surrogates. For each surrogate, the last two fallbacks were surrogate 693 (Well Count – All Wells) and 304 (NLCD Open + Low). Where appropriate, other surrogates were also parts of the gapfilling procedure. For example, surrogate 670 (Spud Count – CBM Wells) was first gapfilled with 692 (Spud Count – All Wells), and then 693 and finally 304. All gapfilling was performed with the Surrogate Tool.

The U.S. CAP emissions (i.e., NH<sub>3</sub>, NO<sub>x</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, and VOC) allocated to the various spatial surrogates are shown in Table 3-14.



**Table 3-13. Spatial Surrogates for Oil and Gas Sources**

<b>Surrogate Code</b>	<b>Surrogate Description</b>
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
689	Gas Produced – Total
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
6831	Produced water at CBM wells
6832	Produced water at gas wells
6833	Produced water at oil wells

**Table 3-14. Selected 2019 CAP emissions by sector for U.S. Surrogates (12US1, tons)**

<b>Sector</b>	<b>ID</b>	<b>Description</b>	<b>NH3</b>	<b>NOX</b>	<b>PM2_5</b>	<b>SO2</b>	<b>VOC</b>
afdust	240	Total Road Miles	0	0	333,425	0	0
afdust	306	NLCD Med + High	0	0	41,167	0	0
afdust	308	NLCD Low + Med + High	0	0	122,726	0	0
afdust	310	NLCD Total Agriculture	0	0	502,702	0	0
afdust	861	Sand and Gravel Mines	0	0	271	0	0
afdust	863	Crushed Stone Mines	0	0	291	0	0

Sector	ID	Description	NH3	NOX	PM2_5	SO2	VOC
afdust	902	OSM Unpaved Roads	0	0	960,028	0	0
afdust	4012	NPDES 2020 Beef Cattle	0	0	191,878	0	0
afdust	4013	NPDES 2020 Dairy Cattle	0	0	15,033	0	0
afdust	4021	NPDES 2020 Swine	0	0	658	0	0
afdust	4031	NPDES 2020 Chicken	0	0	5,069	0	0
afdust	4071	NPDES 2020 Turkey	0	0	1,959	0	0
fertilizer	310	NLCD Total Agriculture	1,832,594	0	0	0	0
livestock	405	FAO 2010 Horse	31,969	0	0	0	2,558
livestock	406	FAO 2010 Sheep	19,235	0	0	0	1,539
livestock	4012	NPDES 2020 Beef Cattle	702,119	0	0	0	56,170
livestock	4013	NPDES 2020 Dairy Cattle	572,321	0	0	0	45,786
livestock	4021	NPDES 2020 Swine	838,696	0	0	0	67,096
livestock	4031	NPDES 2020 Chicken	426,996	0	0	0	34,157
livestock	4041	NPDES 2020 Goat	19,231	0	0	0	1,538
livestock	4071	NPDES 2020 Turkey	83,001	0	0	0	6,640
nonpt	100	Population	454	0	0	0	36
nonpt	135	Detached Housing	0	16,359	81,108	2,724	18,946
nonpt	150	Residential Heating - Natural Gas	44,524	214,626	2,669	1,436	12,680
nonpt	170	Residential Heating - Distillate Oil	1,499	25,521	3,165	624	1,086
nonpt	180	Residential Heating - Coal	0	2	1	7	2
nonpt	190	Residential Heating - LP Gas	127	36,460	150	164	1,435
nonpt	239	Total Road AADT	0	0	0	0	6,536
nonpt	244	All Unrestricted AADT	0	0	0	0	90,591
nonpt	271	NTAD Class 1 2 3 Railroad Density	0	0	0	0	2,074
nonpt	300	NLCD Low Intensity Development	2,860	3,417	17,009	400	26,432
nonpt	306	NLCD Med + High	17,840	251,201	383,854	85,559	132,283
nonpt	307	NLCD All Development	76,463	28,172	126,918	10,917	81,342
nonpt	308	NLCD Low + Med + High	961	162,993	18,656	5,676	10,691
nonpt	310	NLCD Total Agriculture	517	311	504	31	440
nonpt	319	NLCD Crop Land	0	0	95	70	292
nonpt	320	NLCD Forest Land	0	11	31	0	44
nonpt	650	Refineries and Tank Farms	0	0	0	0	90,120
nonpt	711	Airport Areas	0	0	0	0	367
nonpt	801	Port Areas	0	0	0	0	2,351
nonpt	900	OSM Fuel	0	0	0	0	209,354
nonpt	4011	FAO 2010 Large Cattle Operations	0	0	0	0	295,993
nonroad	136	Single and Dual Unit Housing	99	14,706	2,913	47	91,098
nonroad	261	NTAD Total Railroad Density	3	1,664	168	1	341
nonroad	304	NLCD Open + Low	6	1,695	155	4	5,964
nonroad	305	NLCD Low + Med	5	837	1,014	2	21,724
nonroad	306	NLCD Med + High	366	160,863	9,452	257	98,364
nonroad	307	NLCD All Development	112	29,888	16,088	52	183,321

Sector	ID	Description	NH3	NOX	PM2_5	SO2	VOC
nonroad	308	NLCD Low + Med + High	585	242,493	20,187	235	45,408
nonroad	309	NLCD Open + Low + Med	133	21,682	1,301	64	50,114
nonroad	310	NLCD Total Agriculture	358	257,080	18,310	166	28,069
nonroad	320	NLCD Forest Land	15	2,439	438	7	3,532
nonroad	321	NLCD Recreational Land	80	12,898	5,082	35	185,202
nonroad	350	NLCD Water	203	115,290	4,502	112	258,398
nonroad	850	Golf Courses	13	2,108	122	6	5,875
nonroad	860	Mines	2	2,439	231	1	452
np_oilgas	670	Spud Count - CBM Wells	0	0	0	0	35
np_oilgas	671	Spud Count - Gas Wells	0	0	0	0	3,544
np_oilgas	674	Unconventional Well Completion Counts	16	23,908	540	37	1,222
np_oilgas	678	Completions at Gas Wells	0	5,343	121	2,770	10,821
np_oilgas	679	Completions at CBM Wells	0	4	0	169	489
np_oilgas	681	Spud Count - Oil Wells	0	0	0	0	22,225
np_oilgas	683	Produced Water at All Wells	0	41	0	0	807
np_oilgas	685	Completions at Oil Wells	0	217	0	651	14,055
np_oilgas	687	Feet Drilled at All Wells	0	35,527	733	48	2,426
np_oilgas	689	Gas Produced - Total	0	485	29	2	44,237
np_oilgas	691	Well Counts - CBM Wells	0	19,267	307	7	17,464
np_oilgas	692	Spud Count – All Wells	0	589	34	1	74
np_oilgas	693	Well Count – All Wells	0	0	0	0	2
np_oilgas	694	Oil Production at Oil Wells	0	3,060	4	26,019	805,474
np_oilgas	695	Well Count - Oil Wells	0	159,345	4,270	104,961	619,896
np_oilgas	696	Gas Production at Gas Wells	0	42,067	228	1,120	400,686
np_oilgas	697	Oil Production at Gas Wells	0	261	0	0	43,524
np_oilgas	698	Well Count - Gas Wells	0	281,181	4,185	212	504,727
np_oilgas	699	Gas Production at CBM Wells	0	22	2	0	5,334
np_oilgas	6831	Produced water at CBM wells	0	0	0	0	875
np_oilgas	6832	Produced water at gas wells	0	0	0	0	15,908
np_oilgas	6833	Produced water at oil wells	0	0	0	0	69,418
np_solvents	100	Population	0	0	0	0	1,405,567
np_solvents	240	Total Road Miles/	0	0	0	0	41,753
np_solvents	306	NLCD Med + High	0	0	0	0	418,876
np_solvents	307	NLCD All Development	0	0	0	0	197,955
np_solvents	308	NLCD Low + Med + High	0	0	0	0	30,003
np_solvents	310	NLCD Total Agriculture	0	0	0	0	152,587
np_solvents	901	OSM Asphalt Surfaces	0	0	0	0	339,778
onroad	205	Extended Idle Locations	290	33,058	750	14	3,717
onroad	242	All Restricted AADT	29,464	783,301	20,867	3,049	103,641
onroad	244	All Unrestricted AADT	54,906	1,215,064	45,715	6,043	303,973
onroad	259	Transit Bus Terminals	42	1,539	37	1	476
onroad	304	NLCD Open + Low		510	13	0	2,811

Sector	ID	Description	NH3	NOX	PM2_5	SO2	VOC
onroad	306	NLCD Med + High	914	91,100	2,823	67	26,456
onroad	307	NLCD All Development	3,519	182,771	7,802	578	559,726
onroad	308	NLCD Low + Med + High	179	18,151	535	32	29,126
onroad	508	Public Schools	13	1,589	72	1	440
rail	261	NTAD Total Railroad Density	13	22,177	599	16	1,015
rail	271	NTAD Class 1 2 3 Railroad Density	269	400,799	9,861	336	16,478
rwc	135	Detached Housing	7,054	13,004	132,683	3,635	124,847
rwc	136	Single and Dual Unit Housing	15,681	31,864	315,389	8,383	330,813

### 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, the 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: [http://www3.epa.gov/scram001/reports/Emissions%20TSD%20Vol1\\_02-28-08.pdf](http://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 was 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 2020 Canadian inventories and associated data. The spatial surrogate data came from ECCC, 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 3-15. The population surrogate was updated for Mexico is based on the 2015 GPW v4 (see <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/sets/browse>). The other surrogates for Mexico are circa 1999 and 2000 and were based on data obtained from the Sistema Municipal 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 3-16.

**Table 3-15. Canadian Spatial Surrogates**

Code	Canadian Surrogate Description	Code	Description
100	Population	925	Manufacturing and Assembly
101	total dwelling	926	Distribution and Retail (no petroleum)
102	urban dwelling	927	Commercial Services
103	rural dwelling	933	Rail-Passenger
104	capped total dwelling	934	Rail-Freight
105	capped meat cooking dwelling	935	Rail-Yard
106	ALL_INDUST	940	PAVED ROADS NEW
113	Forestry and logging	945	Commercial Marine Vessels
116	Total Resources	946	Construction and mining
200	Urban Primary Road Miles	948	Forest

Code	Canadian Surrogate Description	Code	Description
210	Rural Primary Road Miles	949	<i>Combination of Dwelling</i>
211	<i>Oil and Gas Extraction</i>	951	Wood Consumption Percentage
212	Mining except oil and gas	952	<i>Residential Fuel Wood Combustion (PIRD)</i>
220	Urban Secondary Road Miles	955	UNPAVED_ROADS_AND_TRAILS
221	Total Mining	960	TOTBEEF
222	Utilities	961	80110_Broilers
230	Rural Secondary Road Miles	962	80111_Cattle_dairy_and_Heifer
233	<i>Total Land Development</i>	963	80112_Cattle_non-Dairy
240	capped population	964	80113_Laying_hens_and_Pullets
308	Food manufacturing	965	80114_Horses
321	Wood product manufacturing	966	80115_Sheep_and_Lamb
323	Printing and related support activities	967	80116_Swine
324	Petroleum and coal products manufacturing	968	80117_Turkeys
326	Plastics and rubber products manufacturing	969	80118_Goat
327	Non-metallic mineral product manufacturing	970	TOTPOUL
331	Primary Metal Manufacturing	971	80119_Buffalo
340	<i>Construction - Oil and Gas</i>	972	80120_Llama_and_Alpacas
350	<i>Water</i>	973	80121_Deer
412	Petroleum product wholesaler-distributors	974	80122_Elk
448	clothing and clothing accessories stores	975	80123_Wild boars
562	Waste management and remediation services	976	80124_Rabbit
601	SCL:12003 Petroleum Liquids Transportation (PIRD)	977	80125_Mink
602	SCL:12007 Oil Sands In-Situ Extraction and Processing (PIRD)	978	80126_Fox
603	SCL:12010 Light Medium Crude Oil Production (PIRD)	980	TOTSWIN
604	SCL:12011 Well Drilling (PIRD)	981	Harvest_Annual
605	SCL:12012 Well Servicing (PIRD)	982	<i>Harvest_Perennial</i>
606	SCL:12013 Well Testing (PIRD)	983	Synthfert_Annual
607	SCL:12014 Natural Gas Production (PIRD)	984	<i>Synthfert_Perennial</i>
608	SCL:12015 Natural Gas Processing (PIRD)	985	Tillage_Annual
609	<i>SCL:12016 Heavy Crude Oil Cold Production (PIRD)</i>	990	TOTFERT
610	<i>SCL:12018 Disposal and Waste Treatment (PIRD)</i>	996	urban_area
611	SCL:12019 Accidents and Equipment Failures (PIRD)	1251	OFFR_TOTFERT
612	SCL:12020 Natural Gas Transmission and Storage (PIRD)	1252	OFFR_MINES

Code	Canadian Surrogate Description	Code	Description
651	MEIT C1C2 Anchored	1253	OFFR Other Construction not Urban
652	MEIT C1C2 Underway	1254	OFFR Commercial Services
653	MEIT C1C2 Berthed	1255	OFFR Oil Sands Mines
661	MEIT C3 Anchored	1256	OFFR Wood industries CANVEC
662	MEIT C3 Underway	1257	OFFR UNPAVED ROADS RURAL
663	MEIT C3 Berthed	1258	OFFR_Uilities
901	AIRPORT	1259	OFFR total dwelling
902	Military LTO	1260	OFFR_water
903	Commercial LTO	1261	OFFR_ALL_INDUST
904	General Aviation LTO	1262	OFFR Oil and Gas Extraction
905	Air Taxi LTO	1263	OFFR_ALLROADS
921	Commercial Fuel Combustion	1264	OFFR_AIRPORT
923	TOTAL INSTITUTIONAL AND GOVERNEMNT	1265	OFFR_RAILWAY
924	Primary Industry		

**Table 3-16. 2018 CAPs Allocated to Mexican and Canadian Spatial Surrogates for 12US1 (tons)**

Code	Mexican or Canadian Surrogate Description	NH <sub>3</sub>	NO <sub>x</sub>	PM <sub>2.5</sub>	SO <sub>2</sub>	VOC
11	MEX 2015 Population	0	60,516	330	133	167,796
14	MEX Residential Heating - Wood	0	2,468	6,890	201	18,559
16	MEX Residential Heating - Distillate Oil	1	31	0	0	1
22	MEX Total Road Miles	2,130	249,454	8,629	4,749	48,885
24	MEX Total Railroads Miles	0	21,516	450	204	806
26	MEX Total Agriculture	115,677	20,235	16,414	527	3,658
32	MEX Commercial Land	0	59	1,287	0	21,908
34	MEX Industrial Land	72	1,598	927	5	24,672
36	MEX Commercial plus Industrial Land	5	6,830	324	14	79,869
40	MEX Residential (RES1-4)+Comercial+Industrial+Institutional +Government	0	13	48	1	16,400
42	MEX Personal Repair (COM3)	0	0	0	0	4,049
44	MEX Airports Area	0	3,805	53	268	1,440
48	MEX Brick Kilns	0	210	4,180	371	102
50	MEX Mobile sources - Border Crossing	3	64	2	0	50
100	CAN Population	698	56	221	16	3,798
101	CAN total dwelling	0	0	0	0	105,422
104	CAN Capped Total Dwelling	321	32,970	2,486	2,030	1,688
106	CAN ALL_INDUST	0	0	543	0	0
113	CAN Forestry and logging	83	627	2,934	15	2,717

<b>Code</b>	<b>Mexican or Canadian Surrogate Description</b>	<b>NH<sub>3</sub></b>	<b>NO<sub>x</sub></b>	<b>PM<sub>2.5</sub></b>	<b>SO<sub>2</sub></b>	<b>VOC</b>
200	CAN Urban Primary Road Miles	1,527	75,221	2,659	176	7,124
210	CAN Rural Primary Road Miles	584	40,602	1,405	74	2,880
212	CAN Mining except oil and gas	0	0	1,618	0	0
220	CAN Urban Secondary Road Miles	2,866	119,406	5,355	357	18,967
221	CAN Total Mining	0	0	12,266	0	0
222	CAN Utilities	0	2,562	2,504	32	110
230	CAN Rural Secondary Road Miles	1,545	74,760	2,682	187	7,677
240	CAN Total Road Miles	330	44,970	1,181	38	79,357
308	CAN Food manufacturing	0	0	17,591	0	5,104
321	CAN Wood product manufacturing	517	1,700	578	207	8,374
323	CAN Printing and related support activities	0	0	0	0	18,212
324	CAN Petroleum and coal products manufacturing	0	920	1,285	384	5,820
326	CAN Plastics and rubber products manufacturing	0	0	0	0	21,854
327	CAN Non-metallic mineral product manufacturing	0	0	6,686	0	0
331	CAN Primary Metal Manufacturing	0	112	3,880	21	45
412	CAN Petroleum product wholesaler-distributors	0	0	0	0	36,768
448	CAN clothing and clothing accessories stores	0	0	0	0	177
562	CAN Waste management and remediation services	2,656	1,259	2,401	2,119	16,006
601	CAN SCL:12003 Petroleum Liquids Transportation (PIRD)	0	0	12	163	6,141
602	CAN SCL:12007 Oil Sands In-Situ Extraction and Processing (PIRD)	0	0	0	0	108
603	CAN SCL:12010 Light Medium Crude Oil Production (PIRD)	0	0	0	0	2
604	CAN SCL:12011 Well Drilling (PIRD)	0	0	0	563	594
605	CAN SCL:12012 Well Servicing (PIRD)	0	0	0	62	65
606	CAN SCL:12013 Well Testing (PIRD)	0	0	0	0	0
607	CAN SCL:12014 Natural Gas Production (PIRD)	0	31	1	0	215
608	CAN SCL:12015 Natural Gas Processing (PIRD)	0	0	0	0	0
611	CAN SCL:12019 Accidents and Equipment Failures (PIRD)	0	0	0	0	99,936
612	CAN SCL:12020 Natural Gas Transmission and Storage (PIRD)	1	800	55	11	408
901	CAN Airport	0	99	9	0	10



<b>Code</b>	<b>Mexican or Canadian Surrogate Description</b>	<b>NH<sub>3</sub></b>	<b>NO<sub>x</sub></b>	<b>PM<sub>2.5</sub></b>	<b>SO<sub>2</sub></b>	<b>VOC</b>
921	CAN Commercial Fuel Combustion	195	22,375	2,452	449	969
923	CAN TOTAL INSTITUTIONAL AND GOVERNMENT	0	0	0	0	14,276
924	CAN Primary Industry	0	0	0	0	31,784
925	CAN Manufacturing and Assembly	0	0	0	0	64,541
926	CAN Distribution and Retail (no petroleum)	0	0	0	0	6,633
927	CAN Commercial Services	0	0	0	0	30,243
933	CAN Rail-Passenger	1	3,038	60	1	121
934	CAN Rail-Freight	49	77,610	1,537	43	3,430
935	CAN Rail-Yard	1	4,587	95	1	279
940	CAN Paved Roads New			24,023		
946	CAN Construction and Mining	42	2,675	149	257	38
951	CAN Wood Consumption Percentage	1,119	12,431	75,655	1,776	105,563
955	CAN UNPAVED_ROADS_AND_TRAILS	0	0	403,589	0	00
961	CAN 80110_Broilers	12,630	0	115	0	12,787
962	CAN 80111_Cattle_dairy_and_Heifer	57,942	0	276	0	40,516
963	CAN 80112_Cattle_non-Dairy	164,849	0	884	0	42,876
964	CAN 80113_Laying_hens_and_Pullets	9,451	0	40	0	10,596
965	CAN 80114_Horses	2,937	0	19	0	1,321
966	CAN 80115_Sheep_and_Lamb	2,122	0	6	0	170
967	CAN 80116_Swine	59,569	0	824	0	9,949
968	CAN 80117_Turkeys	4,877	0	41	0	4,509
969	CAN 80118_Goat	1,680	0	2	0	135
971	CAN 80119_Buffalo	2,092	0	6	0	517
972	CAN 80120_Llama_and_Alpacas	110	0	0	0	0
973	CAN 80121_Deer	18	0	0	0	0
974	CAN 80122_Elk	18	0	0	0	0
975	CAN 80123_Wild boars	34	0	0	0	0
976	CAN 80124_Rabbit	73	0	0	0	1
977	CAN 80125_Mink	284	0	0	0	951
978	CAN 80126_Fox	4	0	0	0	3
981	CAN Harvest_Annual	0	0	24,807	0	0
983	CAN Synthfert_Annual	177,194	3,616	2,117	5,933	132
985	CAN Tillage_Annual	0	0	106,732	0	0
996	CAN urban_area	0	0	3,423	0	0
1251	CAN OFFR_TOTFERT	83	63,804	4,510	57	6,290
1252	CAN OFFR_MINES	1	585	42	1	81
1253	CAN OFFR Other Construction not Urban	66	38,916	4,649	44	10,239

<b>Code</b>	<b>Mexican or Canadian Surrogate Description</b>	<b>NH<sub>3</sub></b>	<b>NO<sub>x</sub></b>	<b>PM<sub>2.5</sub></b>	<b>SO<sub>2</sub></b>	<b>VOC</b>
1254	CAN OFFR Commercial Services	44	16,547	2,478	38	37,831
1255	CAN OFFR Oil Sands Mines	0	0	0	0	0
1256	CAN OFFR Wood industries CANVEC	9	3,343	272	6	922
1257	CAN OFFR Unpaved Roads Rural	23	10,032	626	20	26,879
1258	CAN OFFR_Uilities	7	3,988	205	6	829
1259	CAN OFFR total dwelling	17	6,202	598	14	12,332
1260	CAN OFFR_water	16	4,665	355	24	24,371
1261	CAN OFFR_ALL_INDUST	3	4,781	168	2	842
1262	CAN OFFR Oil and Gas Extraction	1	400	32	0	120
1263	CAN OFFR_ALLROADS	3	1,811	182	2	463
1265	CAN OFFR_CANRAIL	0	65	6	0	12

### 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. 2016. *2014 Emissions Modeling Platform Spatial Surrogate Documentation*. UNC Institute for the Environment, Chapel Hill, NC. October 1, 2016. Available at [https://gaftp.epa.gov/Air/emismod/2014/v1/spatial\\_surrogates/](https://gaftp.epa.gov/Air/emismod/2014/v1/spatial_surrogates/).
- 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 <http://www.epa.gov/ttn/chief/conference/ei20/index.html#ses-5>.
- Appel, K.W., Napelenok, S., Hogrefe, C., Pouliot, G., Foley, K.M., Roselle, S.J., Pleim, J.E., Bash, J., Pye, H.O.T., Heath, N., Murphy, B., Mathur, R., 2018. Overview and evaluation of the Community Multiscale Air Quality Model (CMAQ) modeling system version 5.2. In Mensink C., Kallos G. (eds), *Air Pollution Modeling and its Application XXV*. ITM 2016. Springer Proceedings in Complexity. Springer, Cham. Available at [https://doi.org/10.1007/978-3-319-57645-9\\_11](https://doi.org/10.1007/978-3-319-57645-9_11).
- Bash, J.O., Baker, K.R., Beaver, M.R., Park, J.-H., Goldstein, A.H., 2016. Evaluation of improved land use and canopy representation in BEIS with biogenic VOC measurements in California. Available from <http://www.geosci-model-dev.net/9/2191/2016/>.
- 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. Available at [https://doi.org/10.1016/S1352-2310\(02\)00220-0](https://doi.org/10.1016/S1352-2310(02)00220-0).
- California Air Resources Board (CARB): Final 2015 Consumer & Commercial Product Survey Data Summaries, 2019.
- Coordinating Research Council (CRC), 2017. Report A-100. Improvement of Default Inputs for MOVES and SMOKE-MOVES. Final Report. February 2017. Available at [http://crcsite.wpengine.com/wp-content/uploads/2019/05/ERG\\_FinalReport\\_CRCA100\\_28Feb2017.pdf](http://crcsite.wpengine.com/wp-content/uploads/2019/05/ERG_FinalReport_CRCA100_28Feb2017.pdf).
- Coordinating Research Council (CRC), 2019. Report A-115. Developing Improved Vehicle Population Inputs for the 2017 National Emissions Inventory. Final Report. April 2019. Available at [http://crcsite.wpengine.com/wp-content/uploads/2019/05/CRC-Project-A-115-Final-Report\\_20190411.pdf](http://crcsite.wpengine.com/wp-content/uploads/2019/05/CRC-Project-A-115-Final-Report_20190411.pdf).
- Drillinginfo, Inc. 2017. “DI Desktop Database powered by HPDI.” Currently available from <https://www.enverus.com/>.

- England, G., Watson, J., Chow, J., Zielenska, B., Chang, M., Loos, K., Hidy, G., 2007. "Dilution-Based Emissions Sampling from Stationary Sources: Part 2-- Gas-Fired Combustors Compared with Other Fuel-Fired Systems," Journal of the Air & Waste Management Association, 57:1, 65-78, DOI: 10.1080/10473289.2007.10465291. Available at <https://www.tandfonline.com/doi/abs/10.1080/10473289.2007.10465291>.
- EPA, 2007a. Control of Hazardous Air Pollutants from Mobile Sources Regulatory Impact Analysis. EPA420-R-07-002. EPA Office of Transportation and Air Quality (OTAQ) Assessment and Standards Division, Ann Arbor, MI. Available online at <https://nepis.epa.gov/Exe/ZyPdf.cgi?Dockey=P1004LNN.PDF>.
- EPA, 2015b. Draft Report Speciation Profiles and Toxic Emission Factors for Nonroad Engines. EPA-420-R-14-028. Available at [https://cfpub.epa.gov/si/si\\_public\\_record\\_Report.cfm?dirEntryId=309339&CFID=83476290&CF\\_TOKEN=35281617](https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=309339&CFID=83476290&CF_TOKEN=35281617).
- EPA, 2015c. Speciation of Total Organic Gas and Particulate Matter Emissions from On-road Vehicles in MOVES2014. EPA-420-R-15-022. Available at <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100NOJG.pdf>.
- EPA, 2016. 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](https://www.epa.gov/sites/production/files/2016-09/documents/speciate_4.5.pdf).
- EPA, 2018. AERMOD Model Formulation and Evaluation Document. EPA-454/R-18-003. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711. Available at [https://www3.epa.gov/ttn/scram/models/aermod/aermod\\_mfed.pdf](https://www3.epa.gov/ttn/scram/models/aermod/aermod_mfed.pdf).
- EPA, 2019. Final Report, SPECIATE Version 5.0, Database Development Documentation, Research Triangle Park, NC, EPA/600/R-19/988. . Available at <https://www.epa.gov/air-emissions-modeling/speciate-51-and-50-addendum-and-final-report>.
- EPA and National Emissions Inventory Collaborative (NEIC), 2019. Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 7.2 North American Emissions Modeling Platform. Available at <https://www.epa.gov/air-emissions-modeling/2016-version-72-technical-support-document>.
- EPA, 2020. Population and Activity of Onroad Vehicles in MOVES3. EPA-420-R-20-023. Office of Transportation and Air Quality. US Environmental Protection Agency. Ann Arbor, MI. November 2020. Available under the MOVES3 section at <https://www.epa.gov/moves/moves-technical-reports>.
- EPA, 2020b. Technical Support document: "Development of Mercury Speciation Factors for EPA's Air Emissions Modeling Programs, April 2020". US EPA Office of Air Quality Planning and Standards.
- EPA, 2021. 2017 National Emission Inventory: January 2021 Updated Release, Technical Support Document. U.S. Environmental Protection Agency, OAQPS, Research Triangle Park, NC 27711. Available at: <https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-technical-support-document-tsd>.

- EPA, 2021. 2017 National Emissions Inventory (NEI) data, Research Triangle Park, NC, January 2021. <https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-data>.
- EPA and NEIC, 2021. Technical Support Document (TSD) Preparation of Emissions Inventories for the 2016v1 North American Emissions Modeling Platform. Available at: <https://www.epa.gov/air-emissions-modeling/2016-version-1-technical-support-document>.
- EPA, 2022a. Technical Support Document EPA's Air Toxics Screening Assessment – 2018 AirToxScreen TSD. Available at: <https://www.epa.gov/AirToxScreen/2018-airtoxscreen-technical-support-document>.
- EPA, 2022b. Technical Support Document: Preparation of Emissions Inventories for the 2019 North American Emissions Modeling Platform. Available at: <https://www.epa.gov/air-emissions-modeling/2019-emissions-modeling-platform-technical-support-document>.
- EPA, 2023. 2020 National Emission Inventory Technical Support Document. U.S. Environmental Protection Agency, OAQPS, Research Triangle Park, NC 27711. Available at: <https://www.epa.gov/air-emissions-inventories/2020-national-emissions-inventory-nei-technical-support-document-tsd>.
- ERG, 2016b. "Technical Memorandum: Modeling Allocation Factors for the 2014 Oil and Gas Nonpoint Tool." Available at [https://gaftp.epa.gov/air/emismod/2014/v1/spatial\\_surrogates/oil\\_and\\_gas/](https://gaftp.epa.gov/air/emismod/2014/v1/spatial_surrogates/oil_and_gas/).
- ERG, 2017. "Technical Report: Development of Mexico Emission Inventories for the 2014 Modeling Platform." Available at [https://gaftp.epa.gov/air/emismod/2016/v1/reports/EPA%205-18%20Report\\_Clean%20Final\\_01042017.pdf](https://gaftp.epa.gov/air/emismod/2016/v1/reports/EPA%205-18%20Report_Clean%20Final_01042017.pdf).
- ERG, 2018. Technical Report: "2016 Nonpoint Oil and Gas Emission Estimation Tool Version 1.0". Available at [https://gaftp.epa.gov/air/emismod/2016/v1/reports/2016%20Nonpoint%20Oil%20and%20Gas%20Emission%20Estimation%20Tool%20V1\\_0%20December\\_2018.pdf](https://gaftp.epa.gov/air/emismod/2016/v1/reports/2016%20Nonpoint%20Oil%20and%20Gas%20Emission%20Estimation%20Tool%20V1_0%20December_2018.pdf).
- The Freedonia Group: Solvents, Industry Study #3429, 2016.
- [Khare, P., and Gentner, D. R.: Considering the future of anthropogenic gas-phase organic compound emissions and the increasing influence of non-combustion sources on urban air quality, Atmos Chem Phys, 18, 5391-5413, 10.5194/acp-18-5391-2018, 2018.](#)
- Luecken D., Yarwood G, Hutzell WT, 2019. Multipollutant modeling of ozone, reactive nitrogen and HAPs across the continental US with CMAQ-CB6. Atmospheric environment. 2019 Mar 15;201:62-72.
- Mansouri, K., Grulke, C. M., Judson, R. S., and Williams, A. J.: OPERA models for predicting physicochemical properties and environmental fate endpoints, J Cheminformatics, 10, 10.1186/s13321-018-0263-1, 2018.
- 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. Available at <https://doi.org/10.1016/j.scitotenv.2009.07.009>.
- MDNR, 2008. "A Minnesota 2008 Residential Fuelwood Assessment Survey of individual household responses". Minnesota Department of Natural Resources. Available from [http://files.dnr.state.mn.us/forestry/um/residentialfuelwoodassessment07\\_08.pdf](http://files.dnr.state.mn.us/forestry/um/residentialfuelwoodassessment07_08.pdf).

- NCAR, 2016. FIRE EMISSION FACTORS AND EMISSION INVENTORIES, FINN Data. downloaded 2014 SAPRC99 version from <http://bai.acom.ucar.edu/Data/fire/>.
- NEIC, 2019. Specification sheets for the 2016v1 platform. Available from <http://views.cira.colostate.edu/wiki/wiki/10202>.
- NESCAUM, 2006. “Assessment of Outdoor Wood-fired Boilers”. Northeast States for Coordinated Air Use Management (NESCAUM) report. Available from [http://www.nescaum.org/documents/assessment-of-outdoor-wood-fired-boilers/2006-1031-owb-report\\_revised-june2006-appendix.pdf](http://www.nescaum.org/documents/assessment-of-outdoor-wood-fired-boilers/2006-1031-owb-report_revised-june2006-appendix.pdf).
- 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>.
- Pouliot, G., H. Simon, P. Bhawe, 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://www3.epa.gov/ttn/chief/conference/ei19/session9/pouliot\\_pres.pdf](http://www3.epa.gov/ttn/chief/conference/ei19/session9/pouliot_pres.pdf).
- 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, Rao V, McCarty JL, Soja A. Development of the crop residue and rangeland burning in the 2014 National Emissions Inventory using information from multiple sources. Journal of the Air & Waste Management Association. 2017 Apr 27;67(5):613-22.
- [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. Available at <https://doi.org/10.1080/10962247.2015.1020118>.](#)
- [Reff, A., Bhawe, P., Simon, H., Pace, T., Pouliot, G., Mobley, J., Houyoux, M. “Emissions Inventory of PM<sub>2.5</sub> Trace Elements across the United States”, Environmental Science & Technology 2009 43 \(15\), 5790-5796, DOI: 10.1021/es802930x. Available at <https://doi.org/10.1021/es802930x>.](#)
- Sarwar, G., S. Roselle, R. Mathur, W. Appel, R. Dennis, “A Comparison of CMAQ HONO predictions with observations from the Northeast Oxidant and Particle Study”, Atmospheric Environment 42 (2008) 5760–5770). Available at <https://doi.org/10.1016/j.atmosenv.2007.12.065>.
- Schauer, J., G. Lough, M. Shafer, W. Christensen, M. Arndt, J. DeMinter, J. Park, “Characterization of Metals Emitted from Motor Vehicles,” Health Effects Institute, Research Report 133, March 2006. Available at <https://www.healtheffects.org/publication/characterization-metals-emitted-motor-vehicles>.
- Seltzer, K. M., Pennington, E., Rao, V., Murphy, B. N., Strum, M., Isaacs, K. K., and Pye, H. O. T., 2021: “Reactive organic carbon emissions from volatile chemical products”, Atmos. Chem. Phys. 21, 5079–5100, 2021. <https://doi.org/10.5194/acp-21-5079-2021> and <https://acp.copernicus.org/articles/21/5079/2021/>.

- Skamarock, W., J. Klemp, J. Dudhia, D. Gill, D. Barker, M. Duda, X. Huang, W. Wang, J. Powers, 2008. A Description of the Advanced Research WRF Version 3. NCAR Technical Note. National Center for Atmospheric Research, Mesoscale and Microscale Meteorology Division, Boulder, CO. June 2008. Available at: [http://www2.mmm.ucar.edu/wrf/users/docs/arw\\_v3\\_bw.pdf](http://www2.mmm.ucar.edu/wrf/users/docs/arw_v3_bw.pdf).
- Swedish Environmental Protection Agency, 2004. Swedish Methodology for Environmental Data; Methodology for Calculating Emissions from Ships: 1. Update of Emission Factors.
- U.S. Bureau of Labor and Statistics, 2020. Producer Price Index by Industry, retrieved from FRED, Federal Reserve Bank of St. Louis, available at: <https://fred.stlouisfed.org/categories/31>, access date: 21 August 2020.
- U.S. Census Bureau: Paint and Allied Products - 2010, MA325F(10), 2011. <https://www.census.gov/data/tables/time-785/series/econ/cir/ma325f.html>.
- U.S. Census Bureau, Economy Wide Statistics Division: County Business Patterns, 2018. <https://www.census.gov/programs-surveys/cbp/data/datasets.html>.
- U.S. Department of Transportation and the U.S. Department of Commerce, 2015. 2012 Commodity Flow Survey, EC12TCF-US. <https://www.census.gov/library/publications/2015/econ/ec12tcf-us.html>.
- U.S. Energy Information Administration, 2019. The Distribution of U.S. Oil and Natural Gas Wells by Production Rate, Washington, DC. <https://www.eia.gov/petroleum/wells/>.
- 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. Available at <https://doi.org/10.1021/es2013984>.
- Weschler, C. J., and Nazaroff, W. W.: Semivolatile organic compounds in indoor environments, Atmos Environ, 42, 9018-9040, 2008.
- Wiedinmyer, C., Y. Kimura, E. C. McDonald-Buller, L. K. Emmons, R. R. Buchholz, W. Tang, K. Seto, M. B. Joseph, K. C. Barsanti, A. G. Carlton, and R. Yokelson, Volume 16, issue 13, GMD, 16, 3873–3891, 2023. <https://gmd.copernicus.org/articles/16/3873/2023/>.
- Wiedinmyer, C., S.K. Akagi, R.J. Yokelson, L.K. Emmons, J.A. Al-Saadi<sup>3</sup>, J. J. Orlando<sup>1</sup>, 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., R. Beardsley, Y. Shi, and B. Czader: Revision 5 of the Carbon Bond 6 Mechanism (CB6r5). Presented at the Annual CMAS Conference, Chapel Hill, NC, 2020.
- Zhu, Henze, et al, 2013. “Constraining U.S. Ammonia Emissions using TES Remote Sensing Observations and the GEOS-Chem adjoint model”, Journal of Geophysical Research: Atmospheres, 118: 1-14. Available at <https://doi.org/10.1002/jgrd.50166>.



## 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 VOC and NO<sub>x</sub> 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.<sup>11</sup> 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 urban and regional scale air quality modeling. The CMAQ simulation performed for this 2020 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

---

<sup>11</sup> 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.



as the highest resolution for most regional-scale air quality model applications and assessments.<sup>12</sup> 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 include 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<sub>2.5</sub> 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.<sup>13</sup>

## **4.2 CMAQ Model Version, Inputs and Configuration**

This section describes the air quality modeling platform used for the 2020 CMAQ simulation. A modeling platform is a structured system of connected modeling-related tools and data that provide a consistent and 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 2020 Platform to provide a national

---

<sup>12</sup> U.S. EPA (2018), Modeling Guidance for Demonstrating Air Quality Goals for Ozone, PM<sub>2.5</sub>, and Regional Haze, pp 205. [https://www3.epa.gov/ttn/scram/guidance/guide/O3-PM-RH-Modeling\\_Guidance-2018.pdf](https://www3.epa.gov/ttn/scram/guidance/guide/O3-PM-RH-Modeling_Guidance-2018.pdf).

<sup>13</sup> 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.

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 PM<sub>2.5</sub>.

This section provides a description of each of the main components of the 2020 CMAQ simulation along with the results of a model performance evaluation in which the 2020 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<sub>2.5</sub> 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 2020 analysis employed CMAQ version 5.4.<sup>14</sup> The 2020 CMAQ run included CB6r5 chemistry<sup>15,16</sup>, AERO7 aerosol module<sup>17</sup> with non-volatile Primary Organic Aerosol (POA), and updated halogen chemistry<sup>18</sup>. The CMAQ community model versions 5.2 and 5.3 were most recently peer-reviewed in May of 2019 for the U.S. EPA.<sup>19</sup>

#### **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 2020 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 459 by 299 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 2020 simulation. Air quality

---

<sup>14</sup> CMAQ version 5.4: United States Environmental Protection Agency. (2022). CMAQ (Version 5.4) [Software]. Available from <https://doi.org/10.5281/zenodo.7218076>; <https://www.epa.gov/cmaq>. CMAQ v5.4 is also available from the Community Modeling and Analysis System (CMAS) at: <http://www.cmascenter.org>.

<sup>15</sup> Luecken, D. J., Yarwood, G., and Hutzell, W. T.: Multipollutant modeling of ozone, reactive nitrogen and HAPs across the continental US with CMAQ-CB6, *Atmos Environ*, 201, 62-72, 10.1016/j.atmosenv.2018.11.060, 2019.

<sup>16</sup> Yarwood, G., Beardsley, R., Shi, Y., Czader, B.: Revision 5 of the Carbon Bond 6 Mechanism (CB6r5), CMAS 2020, October 27, 2020.

[https://www.cmascenter.org/conference/2020/slides/BeardsleyR\\_CMAS2020\\_CarbonBond6\\_Revision5\\_clean.pdf](https://www.cmascenter.org/conference/2020/slides/BeardsleyR_CMAS2020_CarbonBond6_Revision5_clean.pdf)

<sup>17</sup> Xu, L., Pye, H. O. T., He, J., Chen, Y. L., Murphy, B. N., and Ng, N. L.: Experimental and model estimates of the contributions from biogenic monoterpenes and sesquiterpenes to secondary organic aerosol in the southeastern United States, *Atmos Chem Phys*, 18, 12613-12637, 10.5194/acp-18-12613-2018, 2018.

<sup>18</sup> Kang, D.; Willison, J.; Sarwar, G.; Madden, M.; Hogrefe, C.; Mathur, R.; Gantt, B.; and Saiz-Lopez, A.: Improving the Characterization of Natural Emissions in CMAQ, Environmental Manager, A&WMA, October 2021.

<sup>19</sup> Barsanti, K.C., Pickering, K.E., Pour-Biazar, A., Saylor, R.D., Stroud, C.A., (June 19, 2019). Final Report: Sixth Peer Review of the Community Multiscale Air Quality (CMAQ) Modeling System, [https://www.epa.gov/sites/default/files/2019-08/documents/sixth\\_cmaq\\_peer\\_review\\_comment\\_report\\_6.19.19.pdf](https://www.epa.gov/sites/default/files/2019-08/documents/sixth_cmaq_peer_review_comment_report_6.19.19.pdf).

This peer review was focused on CMAQv5.2, which was released in June of 2017, as well as CMAQ v5.3, which was released in August of 2019. It is available from the Community Modeling and Analysis System (CMAS) as well as previous peer-review reports at: <http://www.cmascenter.org>.

conditions at the outer boundary of the 12-km domain were taken from the GEOS-Chem global model (discussed in Section 4.2.4).

**Table 4-1. Geographic Information for 2020 12-km Modeling Domain**

<b>National 12 km CMAQ Modeling Configuration</b>	
Map Projection	Lambert Conformal Projection
Grid Resolution	12 km
Coordinate Center	97 W, 40 N
True Latitudes	33 and 45 N
Dimensions	459 x 299 x 35
Vertical Extent	35 Layers: Surface to 50 mb level (see Table 4-2)

**Table 4-2. Vertical layer structure for 2020 CMAQ simulation (heights are layer top).**

<b>Vertical Layers</b>	<b>Sigma P</b>	<b>Pressure (mb)</b>	<b>Approximate Height (m)</b>
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 2020 CMAQ Modeling Domain. The blue 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 2020. 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<sub>2.5</sub>. 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

**2020 Emissions:** The emissions inventories used in the 2020 air quality modeling are described in Section 3, above.

**2020 Meteorological Input Data:** The gridded meteorological data for the entire year of 2020 at the 12-km continental United States scale domain was derived from the publicly available version 4.1.1 of the Weather Research and Forecasting Model (WRF), Advanced Research WRF (ARW) core.<sup>20</sup> 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 12US WRF model was initialized using the 12-km North American Model (12NAM)<sup>21</sup> analysis product provided by National Climatic Data Center (NCDC). Where 12NAM data was unavailable, the 40-km Eta Data Assimilation System (EDAS) analysis (ds609.2) from the National Center for Atmospheric Research (NCAR) was used. Analysis nudging for temperature, wind, and moisture was applied above the boundary layer only. The model simulations were conducted continuously. The 'ipxwrf' program was used to initialize deep soil moisture at the start of the run using a 10-day spin-up period. The 2020 WRF meteorology simulated was based on 2011 National Land Cover Database (NLCD).<sup>22</sup> The WRF simulation included the physics options of the Pleim-Xiu land surface model (LSM), Asymmetric Convective Model version 2 planetary boundary layer (PBL) scheme, Morrison double moment microphysics, Kain-Fritsch cumulus parameterization scheme utilizing the moisture-advection trigger<sup>23</sup> and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme.<sup>24</sup> In addition, the Group for High Resolution Sea Surface Temperatures (GHR SST)<sup>25,26</sup> 1-km SST data was used for SST information to provide more resolved information compared to the more coarse data in the NAM analysis.

---

<sup>20</sup> 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.

<sup>21</sup> North American Model Analysis-Only, <http://nomads.ncdc.noaa.gov/data.php>; download from [ftp://nomads.ncdc.noaa.gov/NAM/analysis\\_only/](ftp://nomads.ncdc.noaa.gov/NAM/analysis_only/).

<sup>22</sup> National Land Cover Database 2011, <http://www.mrlc.gov/nlcd2011.php>.

<sup>23</sup> 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>.

<sup>24</sup> 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.

<sup>25</sup> 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.

<sup>26</sup> Global High-Resolution SST (GHR SST) analysis, <https://www.ghrsst.org/>.



Additionally, the hybrid-vertical coordinate system was employed, where the model is terrain-following (Eta) near the surface and isobaric aloft, reducing the influence of surface features on upper-level dynamics.

*2020 Initial and Boundary Conditions:* The 2020 annual lateral boundary and initial species concentrations were provided using a global 3-D GEOS-Chem (Goddard Earth Observing System) v14.0.1 simulation (of the National Aeronautics and Space Administration (NASA) Global Modeling Assimilation Office) utilizing standard options and full atmospheric chemistry.<sup>27</sup> The GEOS-Chem simulation was performed at 2 x 2.5-degree horizontal resolution with a 72-layer vertical structure (36 layers in troposphere, 120-meter first layer). Simulation used full chemistry with online strat, non-local planetary boundary layer and simple secondary organic aerosols and updated methane, lightning, and other parameters for 2020. Emissions included online Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 2.1<sup>28</sup>, online DUST module, and online sea salt module. Global Fire Emissions Database (GFED)<sup>29</sup> were monthly mean. Anthropogenic emissions included fugitive, combustion, and industrial dust.<sup>30</sup> Marine emissions were based on Community Emissions Data System (CEDS) version 2 including shipping vessels.<sup>31</sup> Aircraft Emissions Inventory Code (AEIC)<sup>32</sup> monthly aircraft input data. In addition, CEDS and AEIC was scaled by Covid-19 adjustment Factors for eMissions (CONFORM) dataset.<sup>33</sup> Meteorology used in this 2020 GEOS-Chem run was from Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA2)<sup>34</sup> meteorology at 2 x 2.5-degree.

### 4.3 CMAQ Model Performance Evaluation

An operational model performance evaluation for ozone and PM<sub>2.5</sub> and its related speciated components was conducted for the 2020 simulation using state/local monitoring sites data in order to estimate the ability of the CMAQ modeling system to replicate the 2020 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

---

<sup>27</sup> GEOS-Chem, <https://geoschem.github.io/index.html>

<sup>28</sup> Guenther, A.B., Jiang, X., Heald, C.L., Sakulyanontvittaya, T., Duhl, T., Emmons, L.K., and Wang, X. The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, 2012, GMD, Volume 5, Issue 6, 1471-1492.

<sup>29</sup> <https://www.globalfiredata.org/>

<sup>30</sup> Philip, S., Martin, R.V., Snider, G., Weagle, C.L., van Donkelaar, A., Brauer, M., Henze, D.K., Klimont, Z., Venkataraman, C., Guttikunda, S.K., and Zhang, Q., April 2017. "Anthropogenic fugitive, combustion and industrial dust is a significant, underrepresented fine particulate matter source in global atmospheric models." *Environmental Research Letters*; Bristol, Vol. 12, Iss. 4. Doi:10.1088/1748-9326/aa65a4.

<sup>31</sup> A Community Emissions Data System (CEDS) for Historical emissions, <https://www.pnnl.gov/projects/ceds>

<sup>32</sup> Simone, N.W., Stettler, M.E.J., Barrett, S.R.H., 2013. Rapid estimation of global civil aviation emissions with uncertainty quantification, *Transportation Research Part D: Transport and Environment*, Volume 25, 33-41, ISSN 1361-9209, <https://doi.org/10.1016/j.trd.2013.07.001>.

<sup>33</sup> Doumbia, T., Granier, C., Elguindi, N., Bouarar, I., Darras, S., Brasseur, G., Gaubert, B., Liu, Y., Shi, X., Stavrakou, T., Tilmes, S., Lacey, F., Deroubaix, A., and Wang, T., 2021: Changes in global air pollutant emissions during the COVID-19 pandemic: a dataset for atmospheric modeling, *Earth Syst. Sci. Data*, 13, 4191–4206, <https://doi.org/10.5194/essd-13-4191-2021>.

<sup>34</sup> Global Modeling and Assimilation Office (GMAO). Inst3\_3d\_asm\_Cp; MERRA-2 IAU State Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4, Greenbelt, MD, USA: Goddard Space Flight Center (GSFC DAAC), 2015. Doi: 10.5067/VJAFPLICSIV.

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:

$$MB = \frac{1}{n} \sum_1^n (P - O), \text{ where } P = \text{predicted and } O = \text{observed concentrations.}$$

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:

$$NMB = \frac{\sum_1^n (P - O)}{\sum_1^n (O)} * 100, \text{ where } P = \text{predicted concentrations and } O = \text{observed}$$

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

$$NME = \frac{\sum_1^n |P - O|}{\sum_1^n (O)} * 100$$

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<sup>35</sup> 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<sup>36,37</sup> as were originally identified in Karl and Koss (1984).<sup>38</sup>

---

<sup>35</sup> 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>.

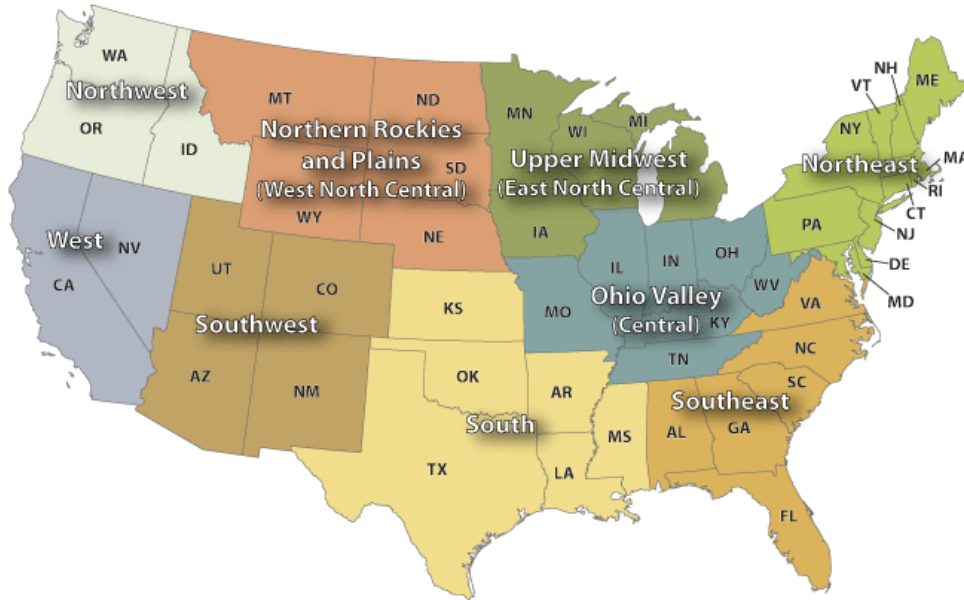
<sup>36</sup> The nine climate regions are defined by States where: Northeast includes CT, DE, ME, MA, MD, NH, NJ, NY, PA, RI, and VT; Ohio Valley includes IL, IN, KY, MO, OH, TN, and WV; Upper Midwest includes IA, MI, MN, and WI; Southeast includes AL, FL, GA, NC, SC, and VA; South includes AR, KS, 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.

<sup>37</sup> 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.

<sup>38</sup> 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.



### U.S. Climate Regions



**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<sub>2.5</sub> 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 in the continental U.S. were included in the evaluation and were taken from the 2020 state/local monitoring site data in 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, April 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 2020 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 and CASTNet sites in the summer is over predicted at all climate regions (NMB ranging between 0.0 to 25.6 percent) except in the Southwest and in the Northern Rockies, West and Northwest at CASTNet sites only where there is a slight under prediction. Likewise, 8-hour ozone at the AQS

and CASTNet sites in the fall is typically over predicted across the contiguous U.S. (NMB ranging between 0.0 to 21.9 percent) except in the West as well as in the Southeast and West at CASTNet sites only. In the winter, 8-hour ozone is overpredicted in all climate regions at AQS and CASTNet sites (NMB ranging between 0.3 to 20.2 percent). In the Spring, 8-hour ozone concentrations are over predicted at AQS and CASTNet sites in all NOAA climate regions (with NMBs less than approximately 20 percent in each subregion) except at AQS sites in the Southwest, Northwest and West (slight under prediction of NMB ranging between -0.8 and -3.9 percent) and at CASTNet sites in the Northeast, Southwest, Northern Rockies, Northwest, and West (NMB ranging between -0.3 and -5.7 percent).

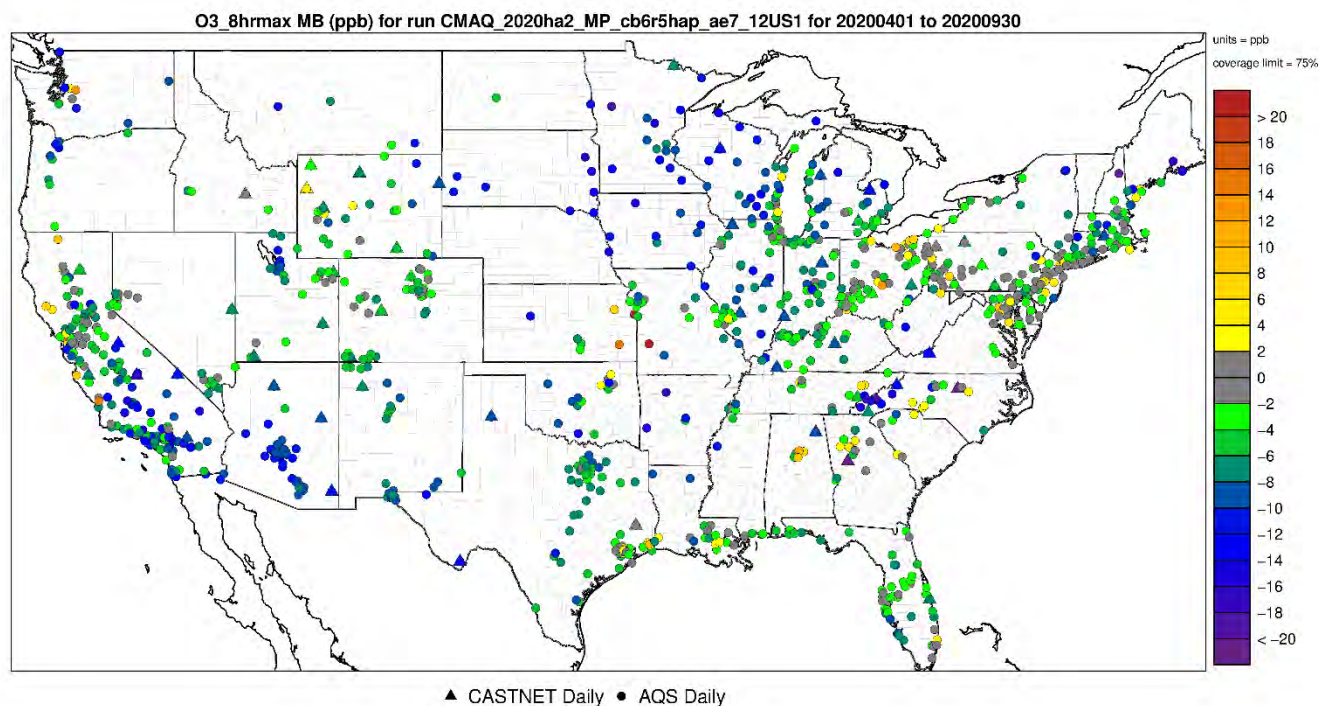
Model bias at individual sites during the ozone season is similar to that seen on a subregional basis for the summer. Figure 4-3 shows the mean bias for 8-hour daily maximum ozone greater than 60 ppb is generally  $\pm 15$  ppb across the AQS and CASTNet sites. Likewise, the information in Figure 4-5 indicates that the normalized mean bias for days with observed 8-hour daily maximum ozone greater than 60 ppb is within  $\pm 20$  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 16 ppb and 30 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 central California, Northern Rockies, Upper Midwest, and Southeast.

**Table 4-4. Summary of CMAQ 2020 8-Hour Daily Maximum Ozone Model Performance Statistics by NOAA climate region, by Season and Monitoring Network.**

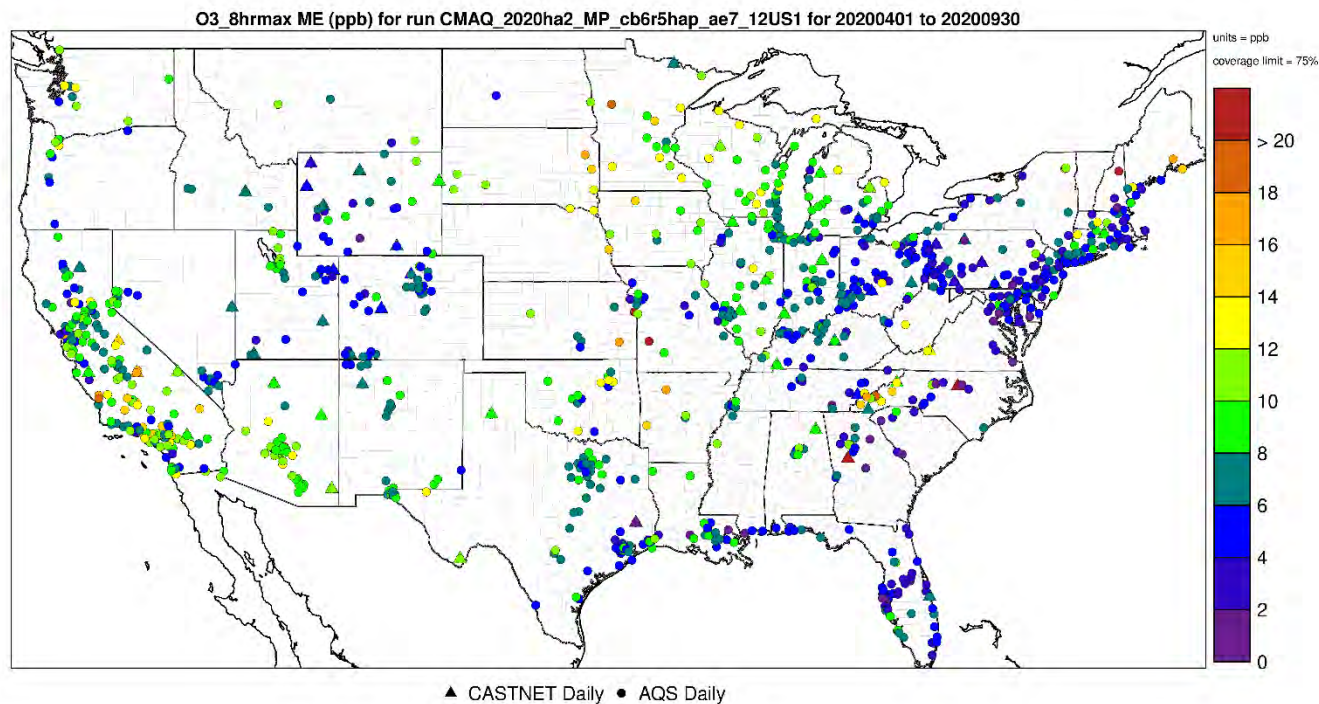
Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
Northeast	AQS	Winter	11,255	3.6	4.9	11.3	15.4
		Spring	16,442	0.8	4.2	2.0	10.1
		Summer	16,412	4.6	6.4	10.9	15.1
		Fall	13,609	4.4	6.1	13.7	18.9
	CASTNet	Winter	1,240	2.5	3.8	7.4	11.1
		Spring	1,267	-0.1	4.0	-0.3	9.2
		Summer	1,234	3.6	5.6	8.8	13.7
		Fall	1,241	3.5	5.6	10.5	16.8
Ohio Valley	AQS	Winter	5,808	5.9	6.8	20.2	23.1
		Spring	20,625	2.8	5.0	6.8	12.4
		Summer	20,549	4.9	7.0	10.9	15.6
		Fall	15,292	5.8	6.7	17.5	20.4
	CASTNet	Winter	1,582	4.3	5.7	13.4	17.6
		Spring	1,630	1.1	4.5	2.6	10.8
		Summer	1,635	4.0	6.6	9.3	15.5
		Fall	1,602	3.7	5.8	11.0	17.2
Upper Midwest	AQS	Winter	1,829	4.5	5.2	13.8	15.8
		Spring	8,092	2.2	5.1	5.6	12.8

Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
		Summer	8,726	2.2	5.9	5.0	13.6
		Fall	6,245	4.7	6.0	15.0	18.8
		CASTNet Winter	445	3.5	4.3	10.5	12.8
		Spring	452	0.2	4.3	0.5	10.4
		Summer	451	0.9	4.7	2.2	11.7
		Fall	448	3.4	5.4	10.9	17.3
Southeast	AQS	Winter	7,187	2.8	4.3	7.7	12.0
		Spring	15,229	2.3	4.8	5.5	11.4
		Summer	14,850	8.9	9.5	25.6	27.3
		Fall	11,938	7.0	7.6	21.9	23.7
	CASTNet	Winter	1,049	1.5	4.1	4.2	11.3
		Spring	1,092	0.2	4.5	0.4	10.4
		Summer	1,055	6.7	8.4	18.6	23.3
		Fall	1,077	4.1	5.8	12.2	17.1
South	AQS	Winter	10,415	3.5	5.7	10.7	17.4
		Spring	12,445	3.5	6.1	8.4	14.7
		Summer	12,307	6.8	8.8	17.1	22.3
		Fall	11,773	5.3	6.9	14.9	19.3
	CASTNet	Winter	520	2.7	4.8	8.0	13.9
		Spring	481	1.4	5.3	3.4	12.5
		Summer	511	5.1	8.1	13.0	20.5
		Fall	525	4.3	6.0	12.1	17.0
Southwest	AQS	Winter	10,182	2.1	4.7	5.4	12.2
		Spring	10,884	-1.9	4.9	-3.9	9.8
		Summer	11,039	-2.2	5.5	-4.0	10.2
		Fall	10,736	0.0	4.9	0.0	10.9
	CASTNet	Winter	979	0.5	3.3	8.0	1.3
		Spring	977	-1.9	3.7	-3.8	7.4
		Summer	990	-0.8	4.2	-1.6	8.0
		Fall	920	-0.4	3.9	-0.8	8.5
Northern Rockies	AQS	Winter	4,383	2.9	4.5	7.6	12.0
		Spring	4,876	0.4	5.1	0.8	11.7
		Summer	4,672	0.3	4.7	0.6	10.2

Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
		Fall	4,428	1.7	4.7	4.7	12.7
	CASTNet	Winter	608	1.5	3.7	3.6	9.2
		Spring	629	-2.1	4.5	-4.4	9.6
		Summer	625	-0.1	4.1	-0.1	8.6
		Fall	578	1.3	4.3	3.3	10.5
Northwest	AQS	Winter	669	1.8	4.8	5.6	14.6
		Spring	1,319	-0.3	4.7	-0.8	12.0
		Summer	2,409	0.9	4.2	2.5	11.4
		Fall	1,129	4.0	6.5	11.7	18.9
	CASTNet	Winter	201	3.5	4.4	9.9	12.3
		Spring	182	-0.3	3.7	-0.6	8.9
		Summer	182	-0.9	4.1	-2.1	9.4
		Fall	202	2.7	5.3	7.4	14.5
West	AQS	Winter	14,257	1.5	4.9	4.3	13.9
		Spring	16,605	-0.4	4.9	-0.9	11.0
		Summer	17,005	0.0	6.9	0.1	13.8
		Fall	15,610	-0.9	7.5	-2.0	15.9
	CASTNet	Winter	614	0.1	3.9	0.3	9.4
		Spring	631	-2.7	5.1	-5.7	10.8
		Summer	638	-5.7	7.8	-9.9	13.7
		Fall	615	-3.3	5.9	-6.7	12.0

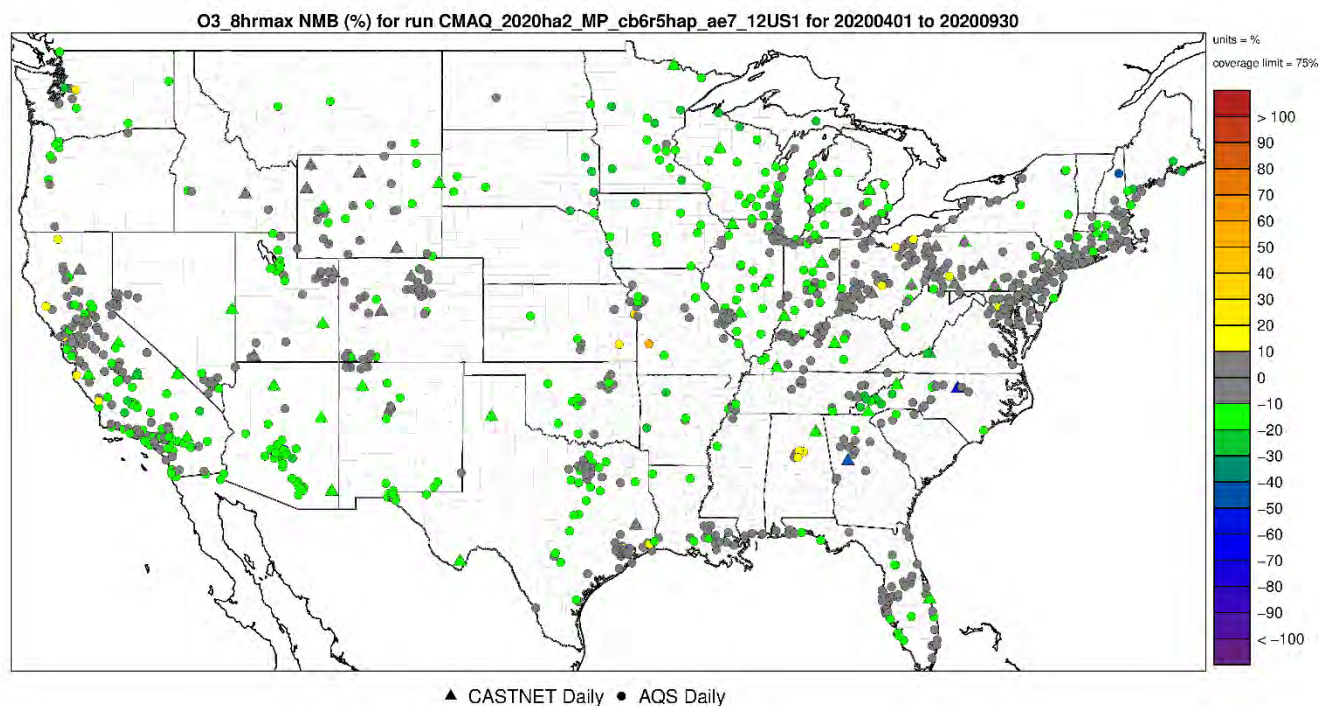


**Figure 4-3. Mean Bias (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period April-September 2020 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.**

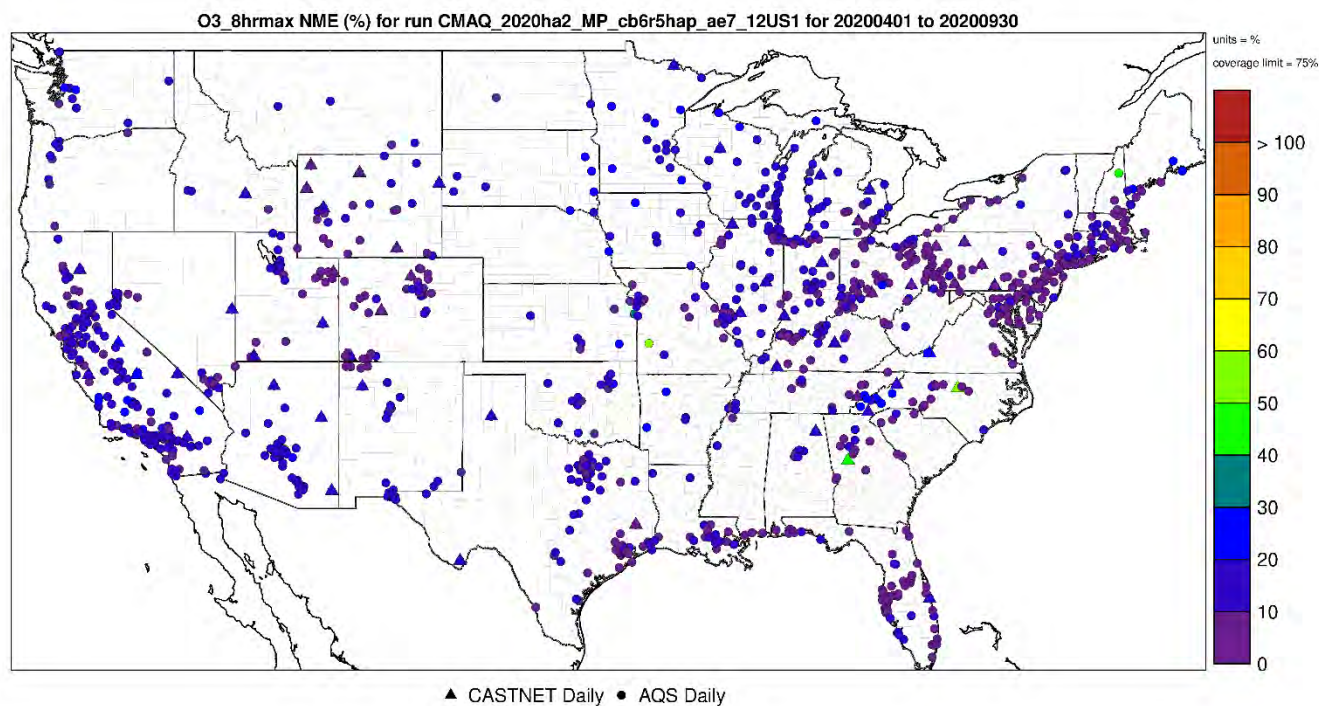


**Figure 4-4. Mean Error (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period April-September 2020 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.**





**Figure 4-5. Normalized Mean Bias (%) of 8-hour daily maximum ozone greater than 60 ppb over the period April-September 2020 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 April-September 2020 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.**

Evaluation for Annual PM<sub>2.5</sub> components: The PM evaluation focuses on PM<sub>2.5</sub> components including sulfate (SO<sub>4</sub>), nitrate (NO<sub>3</sub>), total nitrate (TNO<sub>3</sub> = NO<sub>3</sub> + HNO<sub>3</sub>), ammonium (NH<sub>4</sub>), 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<sub>2.5</sub> measurements for 2020 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<sub>2.5</sub> 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<sub>2.5</sub> 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 CASTNet, IMPROVE, and CSN monitoring sites across the 12-km modeling domain (with MB values ranging from -0.0 to -0.5 µgm<sup>-3</sup>) except at IMPROVE and CSN sites in the Northwest (over prediction, 0.1 to 0.4 µgm<sup>-3</sup>, respectively). Sulfate performance shows moderate error in the eastern subregions (average of approximately 30-50 percent) while Western subregions show slightly larger error (ranging from 30 to 80 percent). Figures 4-7 through 4-10, suggest spatial patterns vary by region. The model bias for most of the Northeast, Southeast, Ohio Valley, and Southwest states are under predicted within ±40 percent. The model bias appears to be greater in the Northwest with predictions up to approximately 60-80 percent at individual monitors. Model error also shows a spatial trend by region, where much of the Eastern states are 30 to 50 percent, the Western and Central U.S. states are 40 to 100 percent.

Annual average nitrate is under predicted at the rural IMPROVE monitoring sites at all NOAA climate subregions (NMB averaging of -40 percent), except in the Northeast, Ohio Valley, Southeast and Northwest where nitrate is over predicted (between 4 to 83 percent). At CSN urban sites, annual average nitrate is under predicted at all subregions, except in the Northeast (29.7 percent), Southeast (69.9 percent) and Northwest (64.4 percent) where nitrate is over predicted. Likewise, model performance of total nitrate at sub-urban CASTNet monitoring sites shows an under prediction at all subregions (NMB in the range of -10.4 to -53.3 percent), except in the Northeast (21.7 percent), Ohio Valley (3.2 percent) and Northwest (46.7 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 over prediction of greater than 10 percent at monitoring sites along the upper Northeast, and Northwest coastline as well as in the South and Southeast as indicated in Figure 4-13. The exception to this is in the Southwest, Northern Rockies 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.

Annual average ammonium model performance as indicated in Table 4-5 has a tendency for the model to under predict across CASTNet sites (ranging from -18 to -72 percent). Ammonium performance across the urban CSN sites shows an under prediction in all NOAA climate subregions (ranging from -4.4 to -66.8 percent), except over predictions in the Northeast (19.5 percent), Upper Midwest (3.5 percent), South (4.0 percent), and Northwest (of 41.7 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 Northeast and in the Northern Rockies, (over prediction bias on average 80 to 100 percent). The urban monitoring sites exhibit slightly larger errors than at rural sites for ammonium.

Annual average elemental carbon is under predicted in all of the nine climate regions at urban and rural sites (biases between -19.8 to 53.8 percent) except at urban Northwest sites (over prediction ranging 10.8 percent). There is not a large variation in error statistics from subregion to subregion or at urban versus rural sites.

Similar to elemental carbon, annual average organic carbon is under predicted in all of the nine climate regions at urban and rural sites (biases between -4.7 to 67.2 percent) except at urban Northwest sites (over prediction ranging 36.5 percent). Likewise, error model performance does not show a large variation from subregion to subregion or at urban versus rural sites.

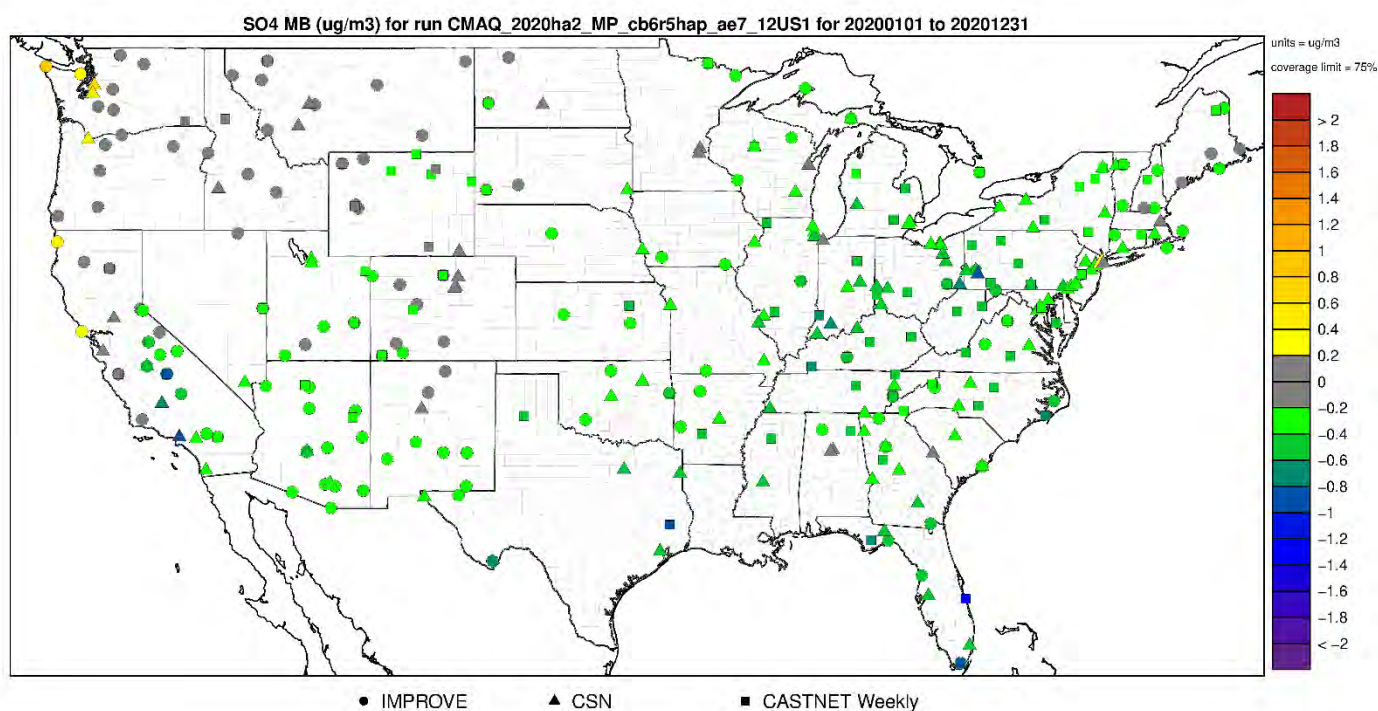


**Table 4-5. Summary of CMAQ 2020 Annual PM Species Model Performance Statistics by NOAA Climate region, by Monitoring Network.**

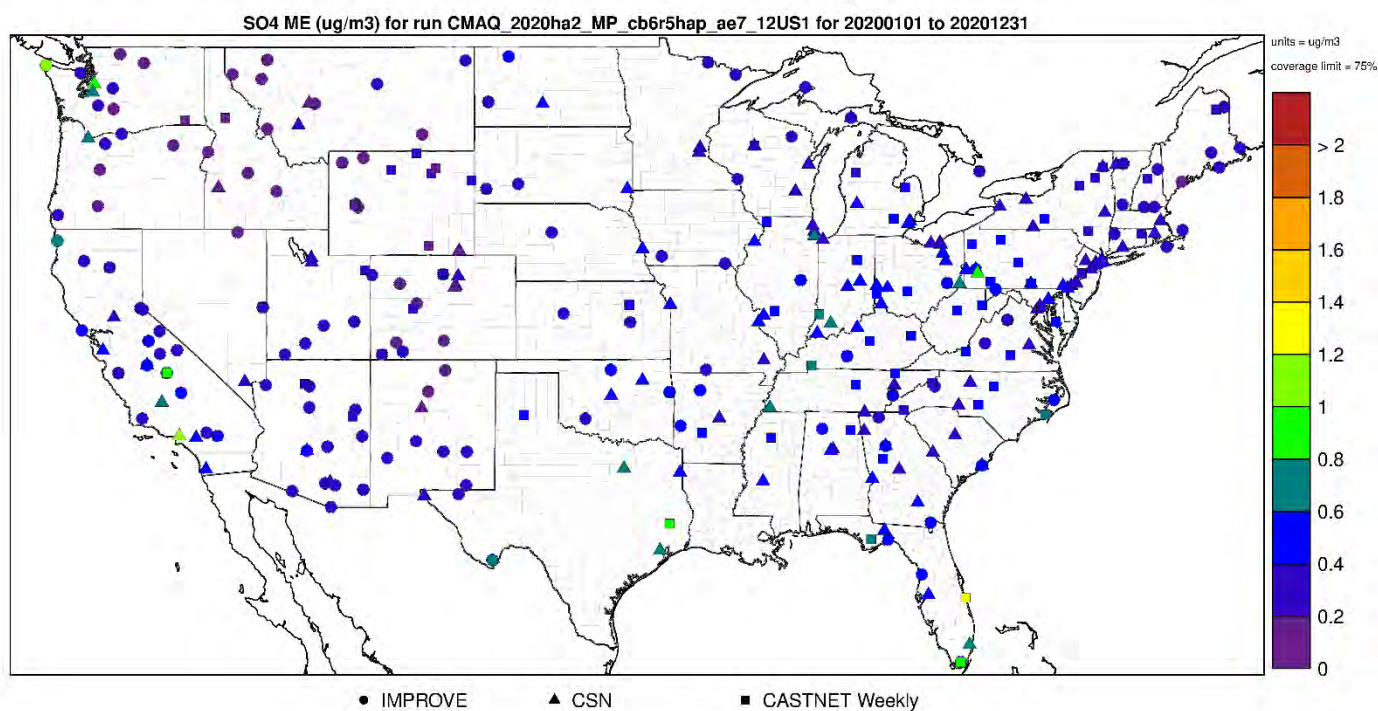
Pollutant	Monitor Network	Subregion	No. of Obs	MB ( $\mu\text{gm}^{-3}$ )	ME ( $\mu\text{gm}^{-3}$ )	NMB (%)	NME (%)
Sulfate	CSN	Northeast	2,666	-0.3	0.4	-36.9	46.9
		Ohio Valley	1,852	-0.4	0.5	-36.1	42.8
		Upper Midwest	1,009	-0.3	0.4	-32.8	44.8
		Southeast	1,718	-0.4	0.4	-31.1	44.8
		South	1,203	-0.4	0.5	-34.6	45.0
		Southwest	1,031	-0.3	0.3	-51.1	55.5
		Northern Rockies	647	-0.2	0.3	-28.2	51.0
		Northwest	556	0.4	0.5	72.8	>100
		West	1,074	-0.4	0.6	-38.1	56.1
	IMPROVE	Northeast	1,899	-0.3	0.3	-43.1	47.7
		Ohio Valley	851	-0.4	0.4	-45.9	49.3
		Upper Midwest	941	-0.2	0.3	-39.2	45.4
		Southeast	1,466	-0.4	0.4	-42.5	50.2
		South	1,082	-0.3	0.4	-41.0	48.0
		Southwest	3,828	-0.2	0.2	-56.1	59.0
		Northern Rockies	2,012	-0.1	0.2	-28.2	52.9
		Northwest	1,867	0.1	0.3	38.8	>100
		West	2,488	-0.2	0.4	-38.3	71.0
	CASTNet	Northeast	891	-0.4	0.4	-51.2	51.8
		Ohio Valley	894	-0.5	0.5	-49.5	49.5
		Upper Midwest	248	-0.4	0.4	-47.5	47.9
		Southeast	647	-0.5	0.5	-57.4	54.7
		South	352	-0.5	0.5	-51.1	51.2
		Southwest	451	-0.3	0.3	-63.3	63.4
		Northern Rockies	544	-0.2	0.2	-50.2	52.1
		Northwest	56	-0.0	0.1	-16.0	39.6
		West	298	-0.4	0.4	-60.6	66.0
Nitrate	CSN	Northeast	2,665	0.2	0.5	29.7	66.1
		Ohio Valley	1,851	-0.0	0.5	-2.5	43.7
		Upper Midwest	1,008	-0.0	0.5	-1.1	38.4
		Southeast	1,720	0.3	0.4	69.9	>100
		South	1,200	-0.0	0.4	-5.7	67.5
		Southwest	1,032	-0.3	0.6	-40.5	71.9
		Northern Rockies	645	-0.2	0.4	-26.8	50.3
		Northwest	556	0.5	0.9	64.4	>100
		West	1,072	-1.1	1.4	-48.8	63.1
	IMPROVE	Northeast	1,899	0.3	0.4	83.1	>100

Pollutant	Monitor Network	Subregion	No. of Obs	MB ( $\mu\text{g m}^{-3}$ )	ME ( $\mu\text{g m}^{-3}$ )	NMB (%)	NME (%)
		Ohio Valley	851	0.0	0.4	3.8	58.4
		Upper Midwest	941	-0.0	0.4	-3.8	47.5
		Southeast	1,466	0.1	0.3	44.2	>100
		South	1,081	-0.1	0.3	-19.4	67.4
		Southwest	3,826	-0.2	0.2	-69.6	83.4
		Northern Rockies	2,011	-0.1	0.2	-30.9	73.9
		Northwest	1,856	0.0	-0.0	15.2	>100
		West	2,486	-0.2	0.3	-45.5	69.8
Total Nitrate ( $\text{NO}_3+\text{HNO}_3$ )	CASTNet	Northeast	891	0.2	0.4	21.7	35.8
		Ohio Valley	894	0.0	0.4	3.2	24.5
		Upper Midwest	248	-0.1	0.3	-4.3	20.5
		Southeast	647	-0.0	0.4	-2.5	45.7
		South	352	-0.2	0.4	-14.8	29.7
		Southwest	451	-0.2	0.3	-28.7	36.7
		Northern Rockies	544	-0.1	0.2	-24.8	35.1
		Northwest	56	0.1	0.2	46.7	56.0
Ammonium	CSN	Northeast	2,664	0.1	0.2	19.5	66.8
		Ohio Valley	1,851	-0.0	0.2	-3.9	45.6
		Upper Midwest	1,009	0.0	0.2	3.5	47.2
		Southeast	2,130	-0.1	0.2	-21.3	59.3
		South	1,718	0.0	0.2	4.0	80.2
		Southwest	1,203	-0.0	0.2	-12.5	62.2
		Northern Rockies	645	-0.0	0.2	-2.7	62.0
		Northwest	555	0.1	0.3	41.7	>100
		West	1,072	-0.4	0.5	-52.5	71.9
	CASTNet	Northeast	891	-0.1	0.2	-18.2	47.5
		Ohio Valley	894	-0.1	0.2	-29.3	40.3
		Upper Midwest	248	-0.1	0.2	-28.4	38.6
		Southeast	587	-0.1	0.1	-33.6	42.1
		South	647	-0.1	0.2	-35.0	58.6
		Southwest	352	-0.1	0.2	-31.5	49.8
		Northern Rockies	544	-0.1	0.1	-58.3	61.1
		Northwest	56	-0.0	0.1	-28.3	50.0
		West	298	-0.2	0.2	-71.9	79.6
Elemental Carbon	CSN	Northeast	2,614	-0.1	0.3	-19.8	44.7
		Ohio Valley	896	-0.1	0.1	-45.0	50.5

Pollutant	Monitor Network	Subregion	No. of Obs	MB ( $\mu\text{g m}^{-3}$ )	ME ( $\mu\text{g m}^{-3}$ )	NMB (%)	NME (%)
		Upper Midwest	1,090	-0.1	0.2	-22.6	45.4
		Southeast	1,572	-0.1	0.3	-40.9	53.3
		South	1,140	-0.2	0.3	-39.9	47.4
		Southwest	1,112	-0.2	0.3	-22.9	44.6
		Northern Rockies	565	-0.2	0.3	-53.8	62.4
		Northwest	538	0.1	0.7	10.8	68.0
		West	2,354	-0.2	0.2	-49.6	63.5
		IMPROVE Northeast	1,777	0.0	0.1	0.0	51.7
		Ohio Valley	1,786	-0.3	0.3	-40.2	47.8
		Upper Midwest	1,041	-0.1	0.1	-41.5	52.8
		Southeast	1,625	-0.3	0.3	-42.8	50.6
		South	1,049	-0.1	0.1	-52.2	55.0
		Southwest	3,537	-0.1	0.1	-43.6	58.2
		Northern Rockies	2,076	-0.1	0.1	-49.5	65.5
		Northwest	1,796	-0.2	0.3	-51.3	85.4
		West	2,354	-0.2	0.2	-49.6	63.5
Organic Carbon	CSN	Northeast	2,614	-0.1	0.9	-4.7	51.4
		Ohio Valley	1,786	-0.6	0.8	-29.4	40.8
		Upper Midwest	1,041	-0.4	0.5	-44.6	54.0
		Southeast	1,591	-0.3	0.7	-22.4	60.2
		South	1,144	-0.7	0.9	-35.4	48.1
		Southwest	1,014	-0.6	1.2	-30.4	57.6
		Northern Rockies	564	-0.8	1.0	-57.9	65.8
		Northwest	538	1.0	2.5	36.5	90.9
		West	1,041	-1.9	2.3	-46.2	55.6
	IMPROVE	Northeast	1,788	-0.2	0.5	-17.0	50.0
		Ohio Valley	899	-0.5	0.6	-38.7	49.8
		Upper Midwest	1,090	-0.4	0.7	-26.9	45.3
		Southeast	1,631	-0.3	0.8	-14.7	41.6
		South	1,062	-0.6	0.6	-52.7	56.2
		Southwest	3,824	-0.7	0.7	-67.2	72.0
		Northern Rockies	2,101	-0.6	0.7	-58.6	74.2
		Northwest	1,826	-0.7	1.4	-44.2	87.0
		West	2,397	-1.4	1.7	-64.0	77.6

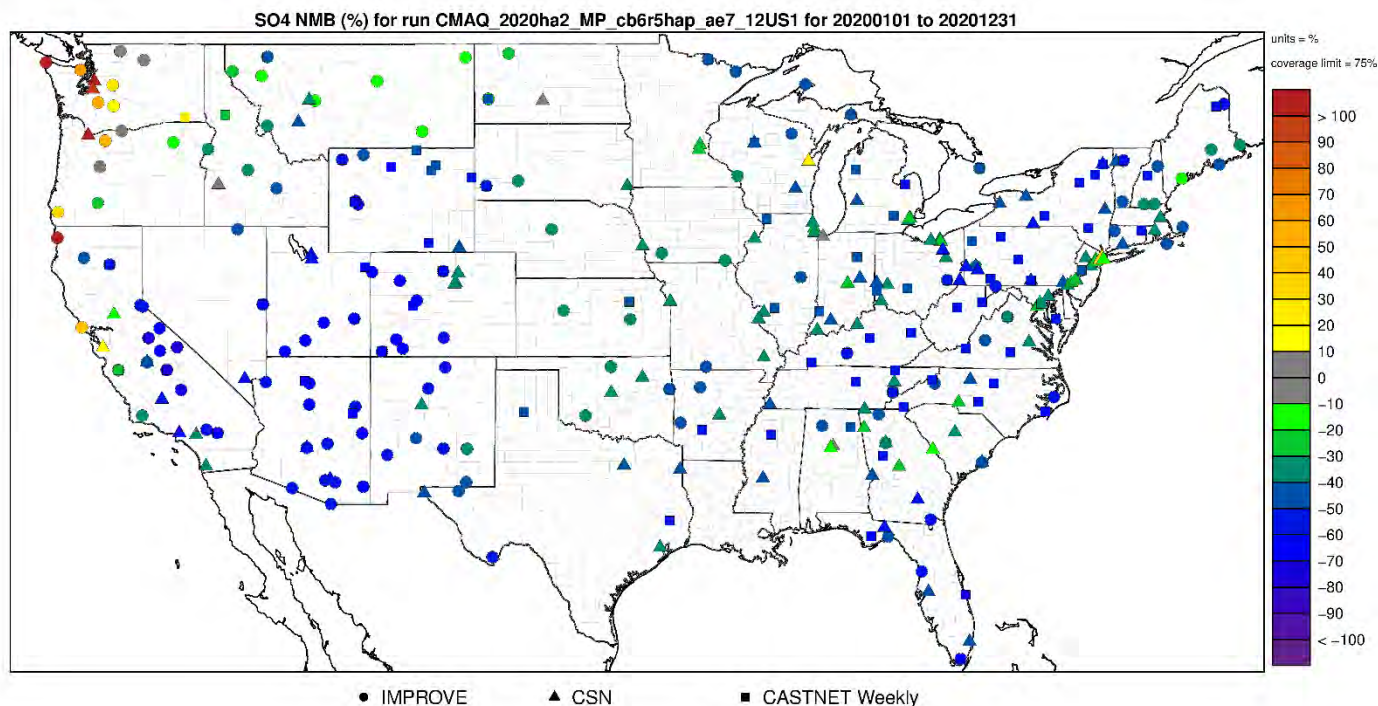


**Figure 4-7. Mean Bias ( $\mu\text{g}\text{m}^{-3}$ ) of annual sulfate at monitoring sites in the continental U.S. modeling domain.**

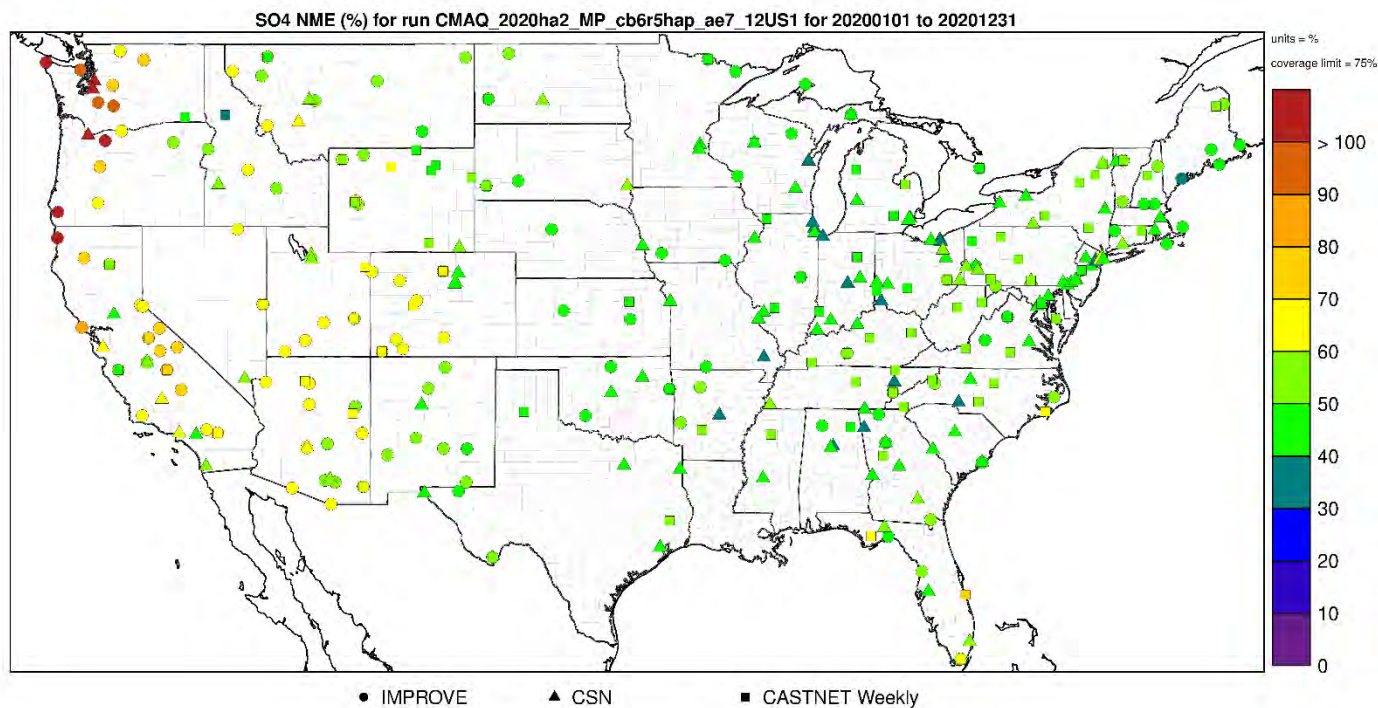


**Figure 4-8. Mean Error ( $\mu\text{g}\text{m}^{-3}$ ) of annual sulfate at monitoring sites in the continental U.S. modeling domain.**

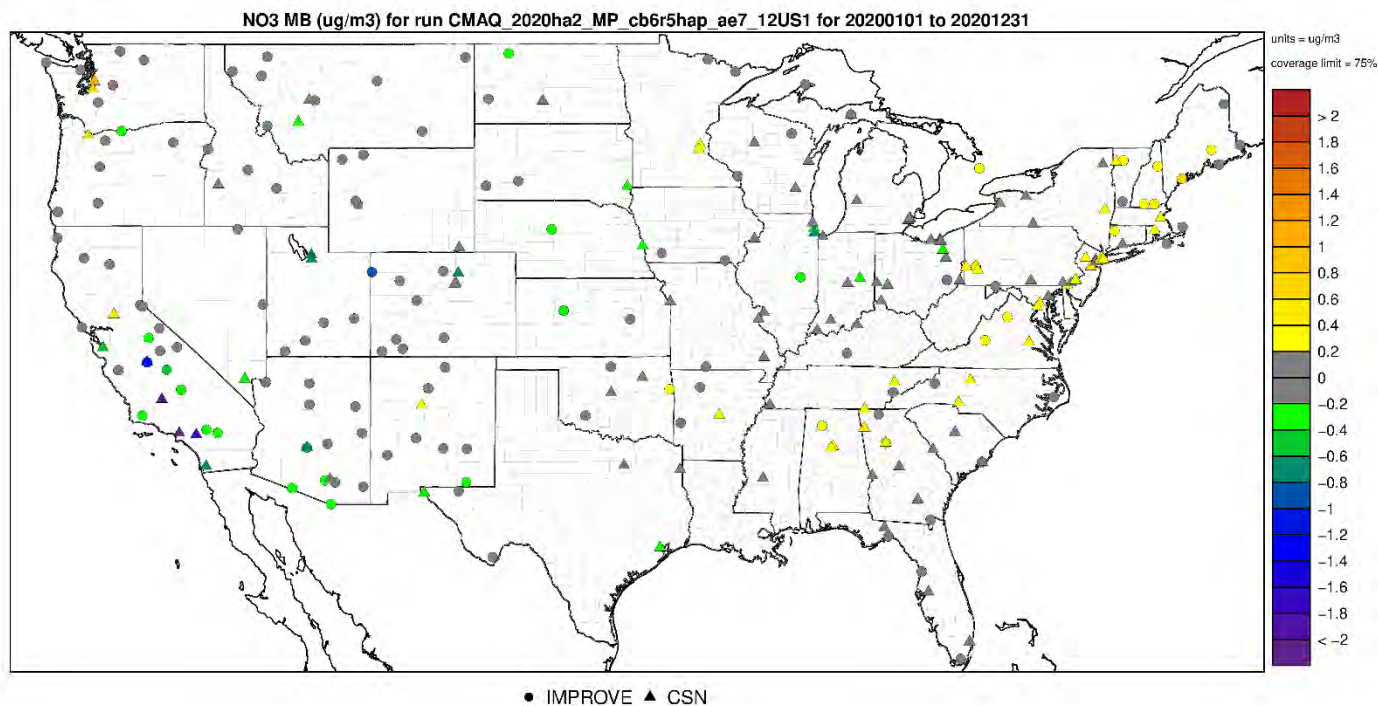




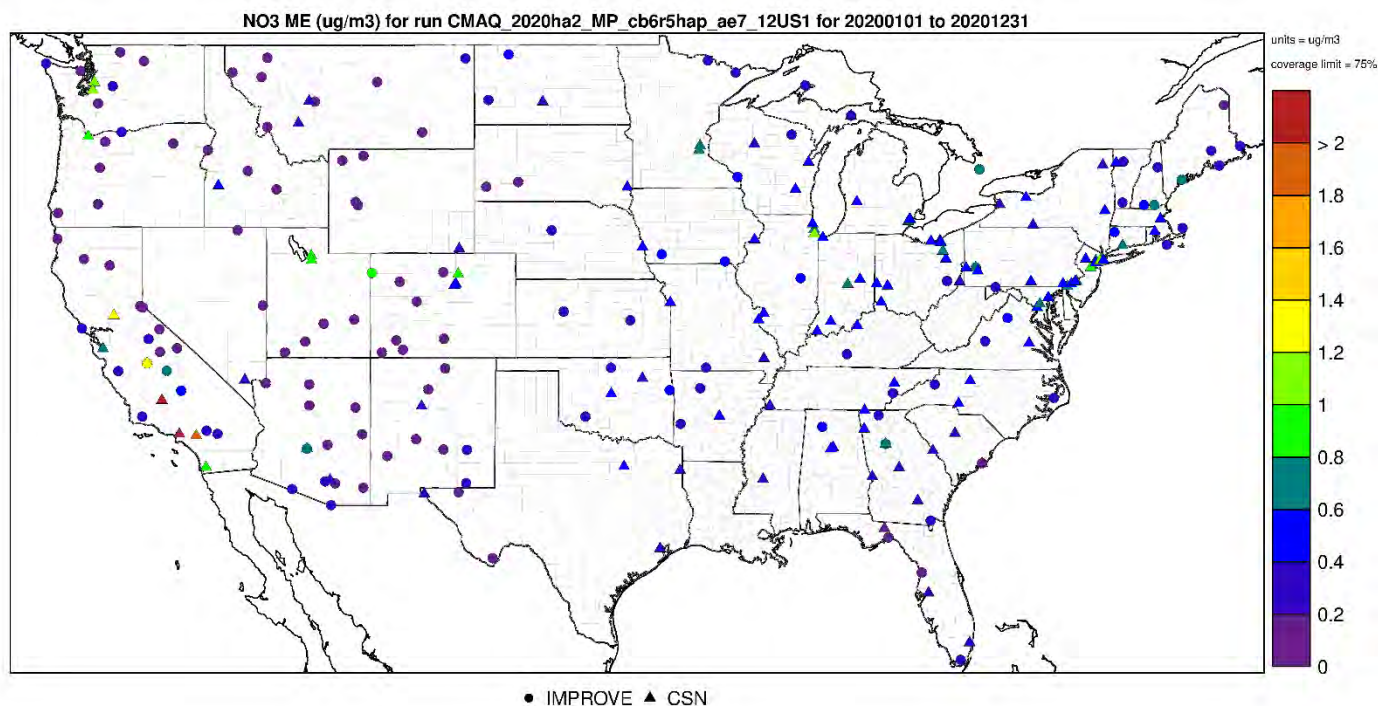
**Figure 4-9. Normalized Mean Bias (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.**



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

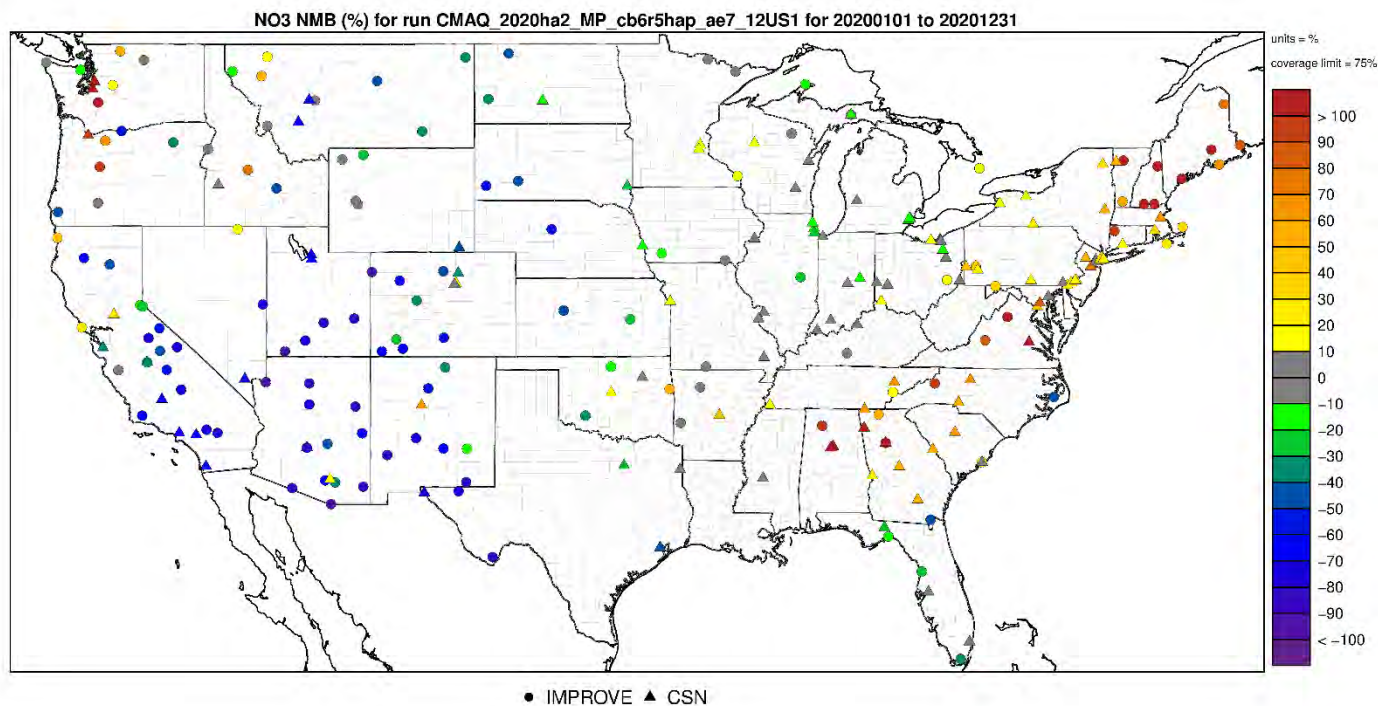


**Figure 4-11. Mean Bias ( $\mu\text{g}\text{m}^{-3}$ ) of annual nitrate at monitoring sites in the continental U.S. modeling domain.**

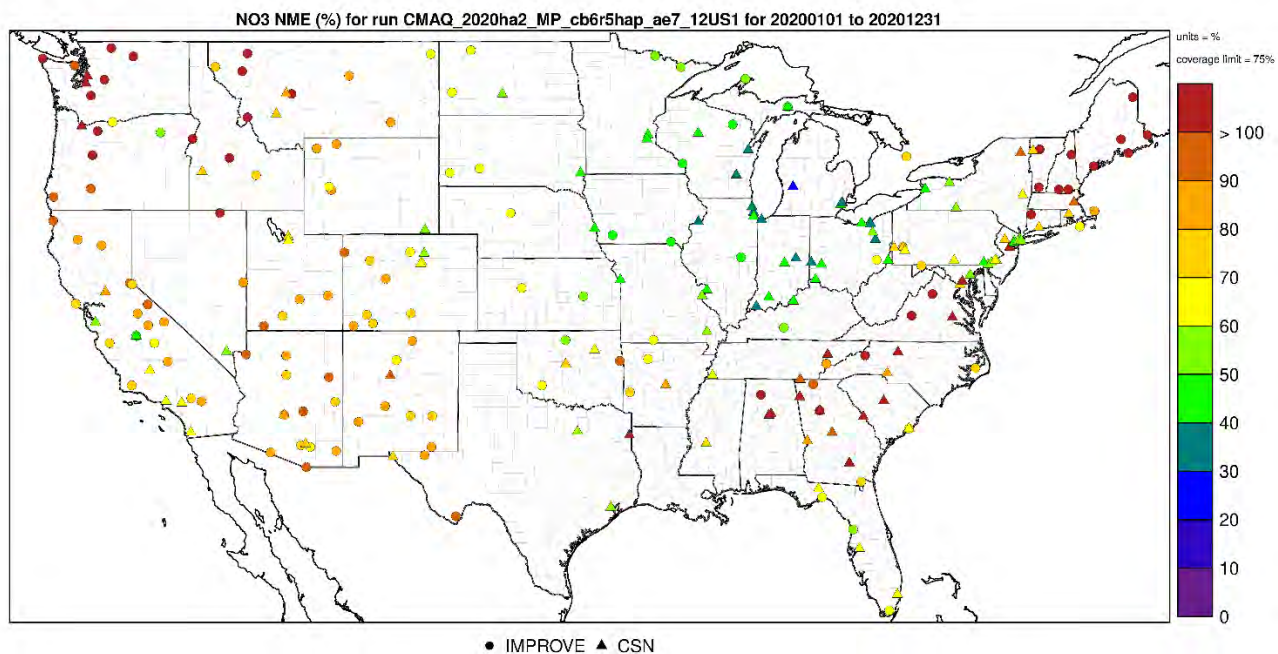


**Figure 4-12. Mean Error ( $\mu\text{g}\text{m}^{-3}$ ) of annual nitrate at monitoring sites in the continental U.S. modeling domain.**



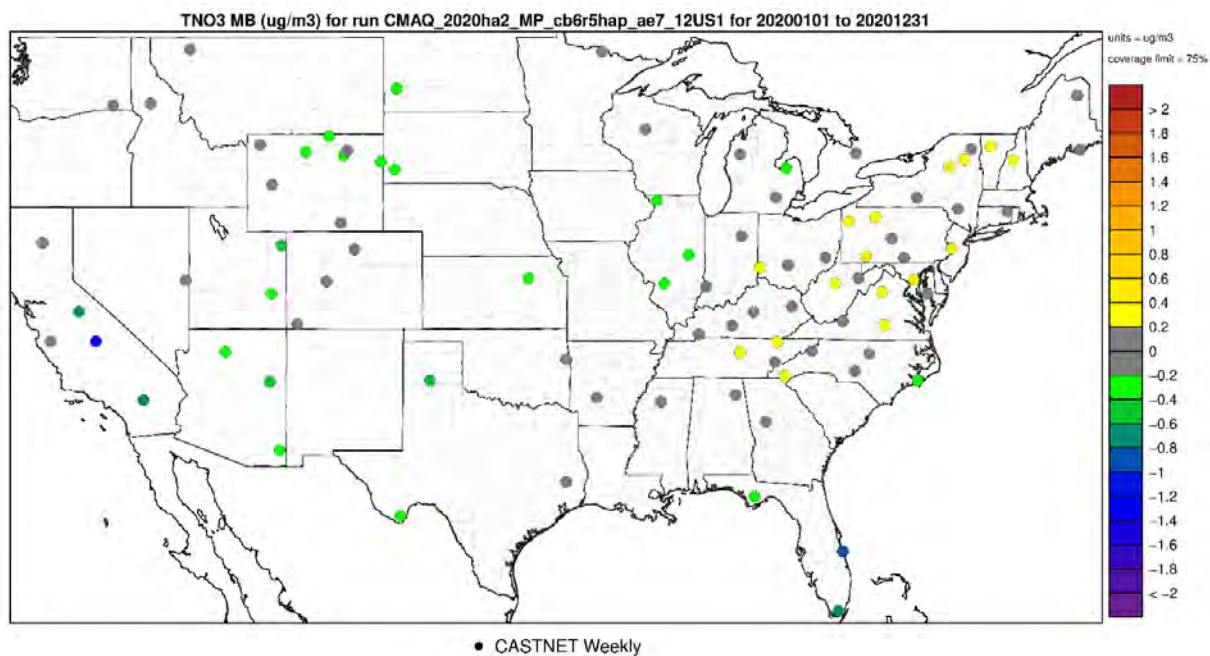


**Figure 4-13. Normalized Mean Bias (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.**

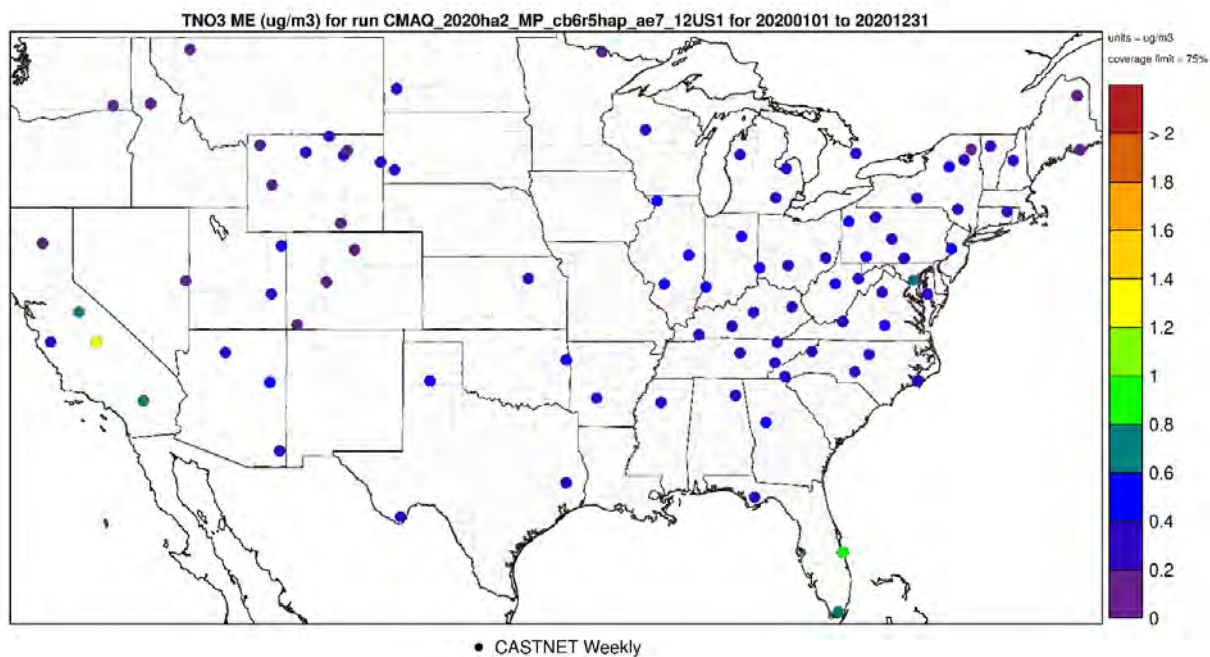


**Figure 4-14. Normalized Mean Error (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.**

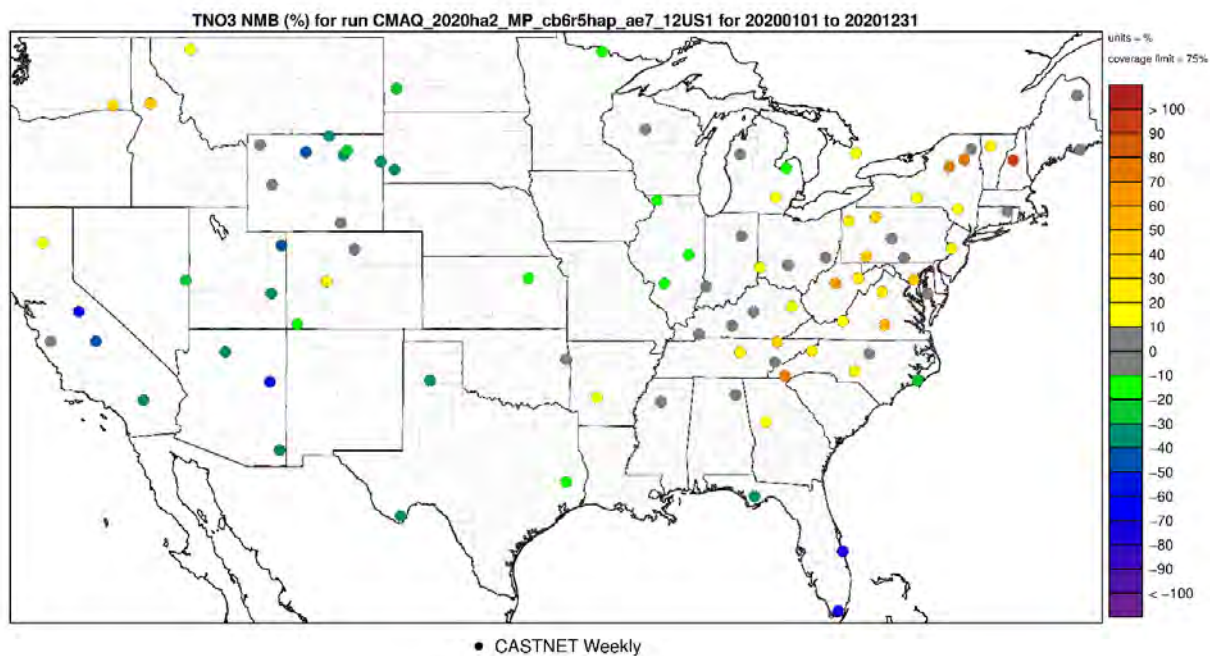




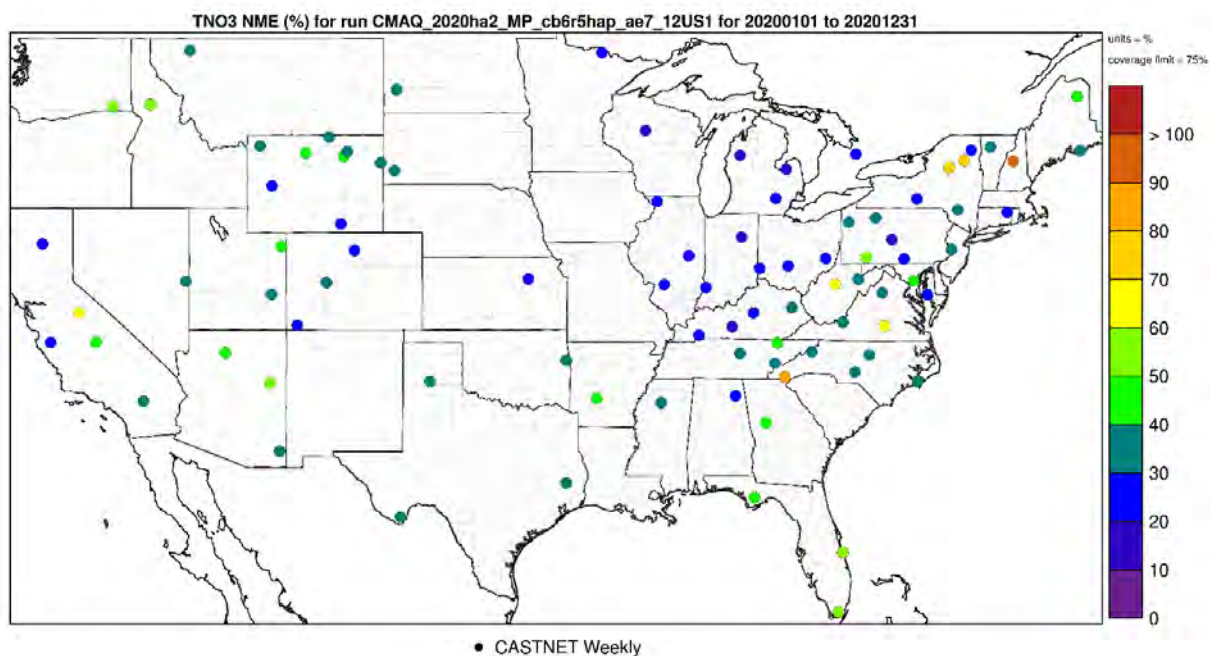
**Figure 4-15. Mean Bias ( $\mu\text{g}\text{m}^{-3}$ ) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.**



**Figure 4-16. Mean Error ( $\mu\text{g}\text{m}^{-3}$ ) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.**

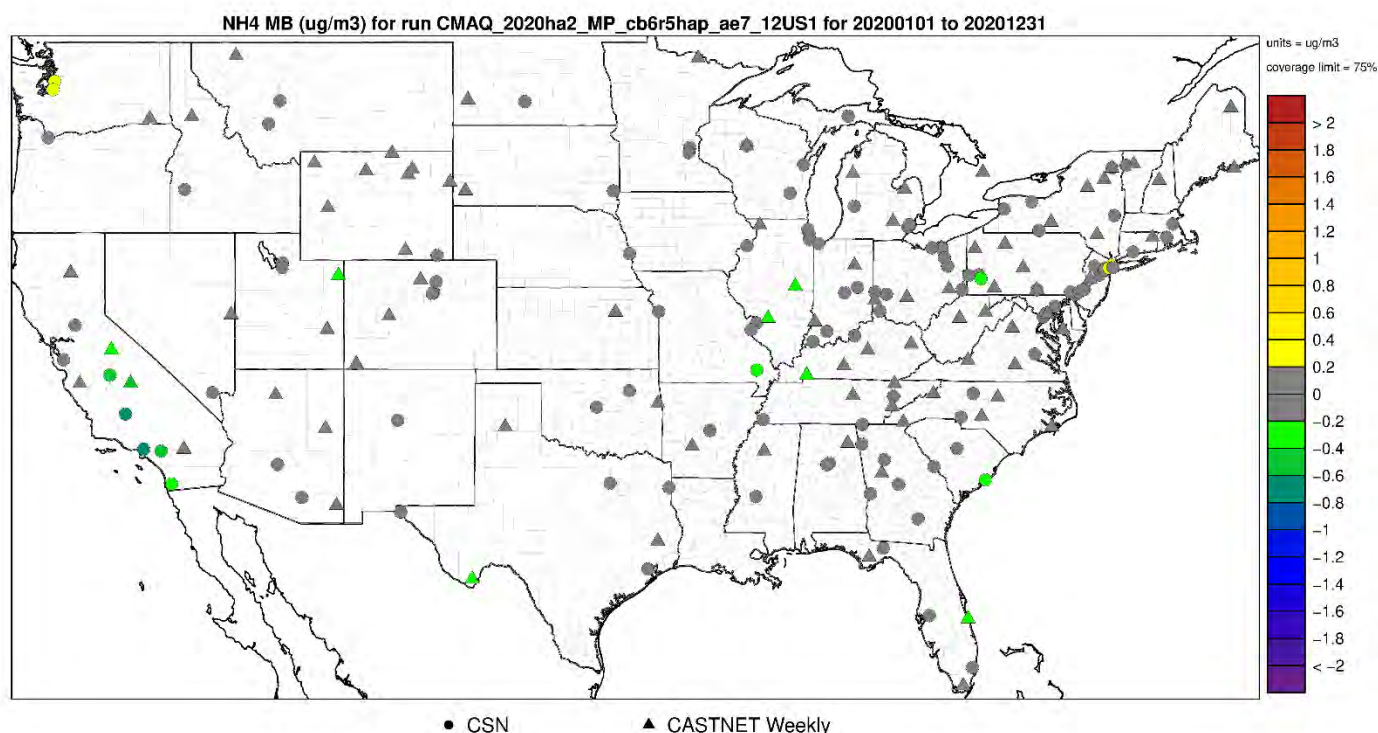


**Figure 4-17. Normalized Mean Bias (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.**

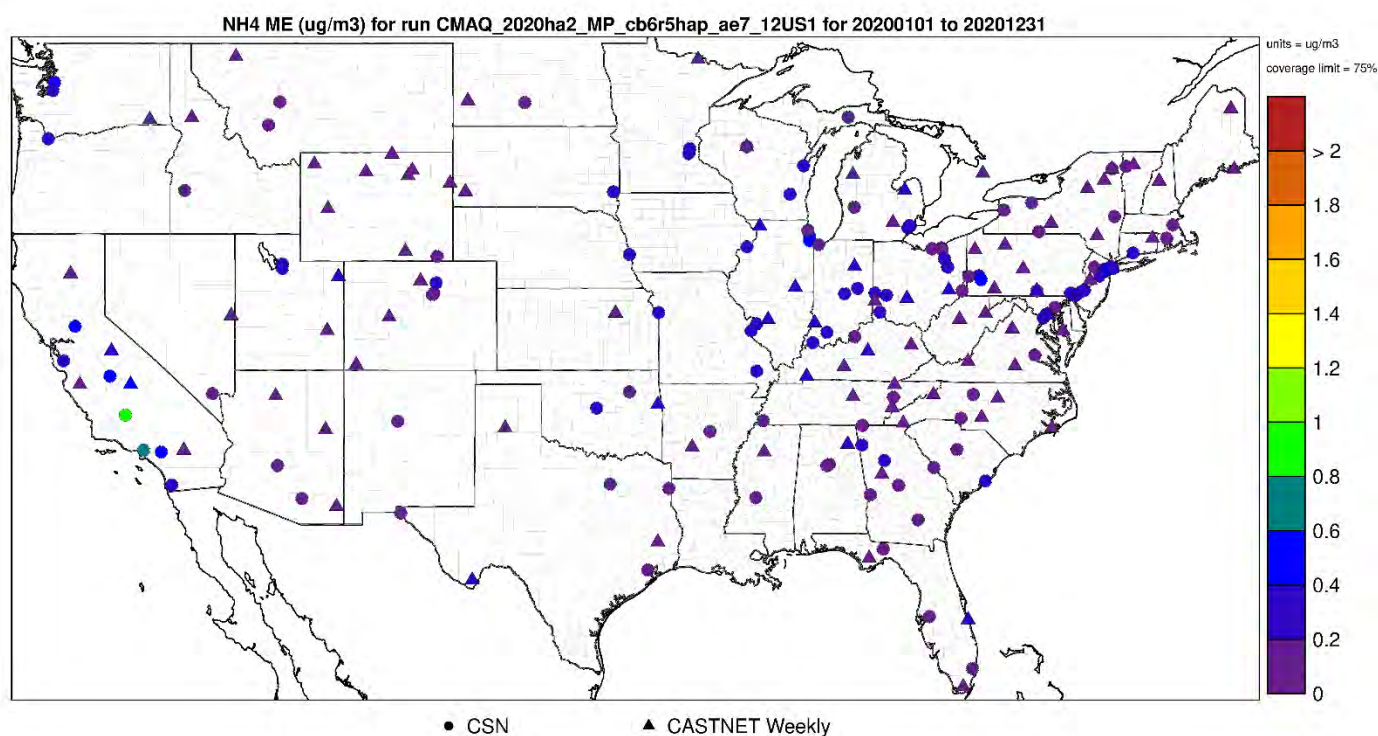


**Figure 4-18. Normalized Mean Error (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.**

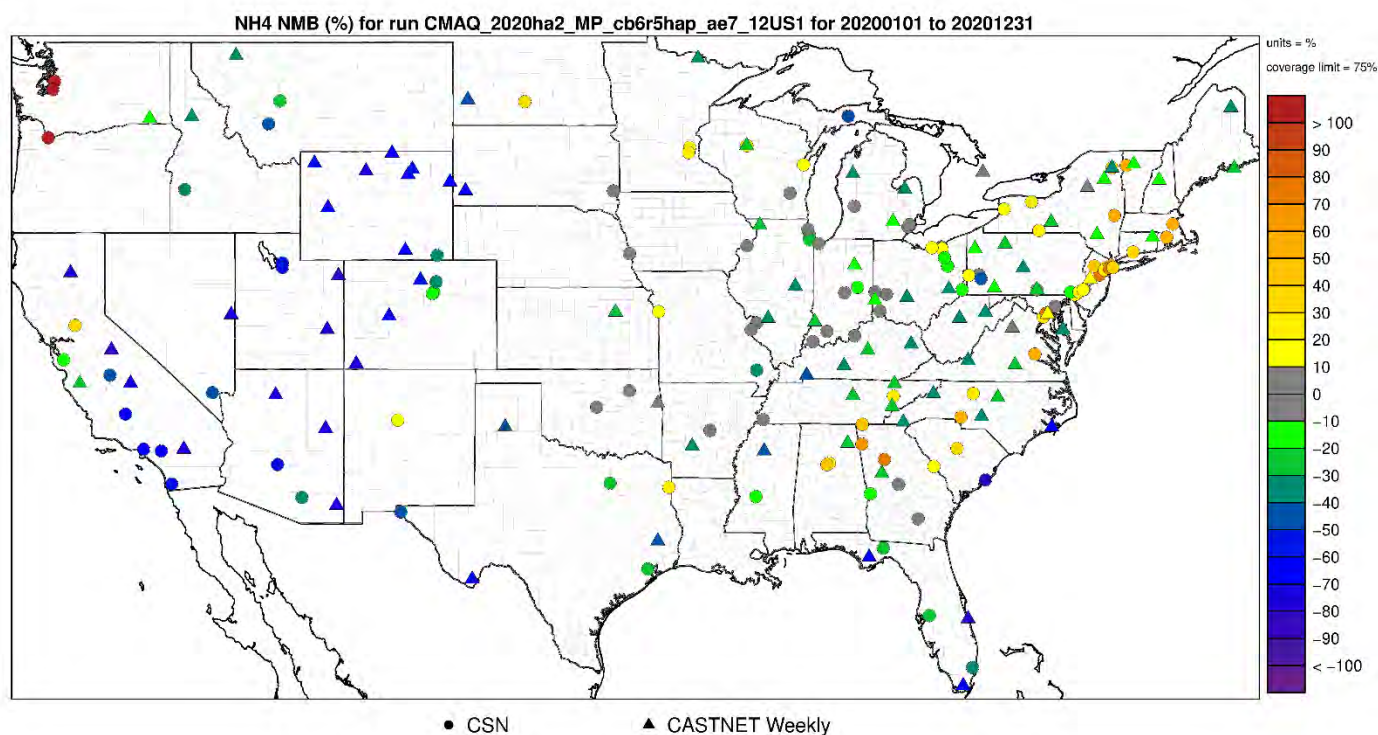




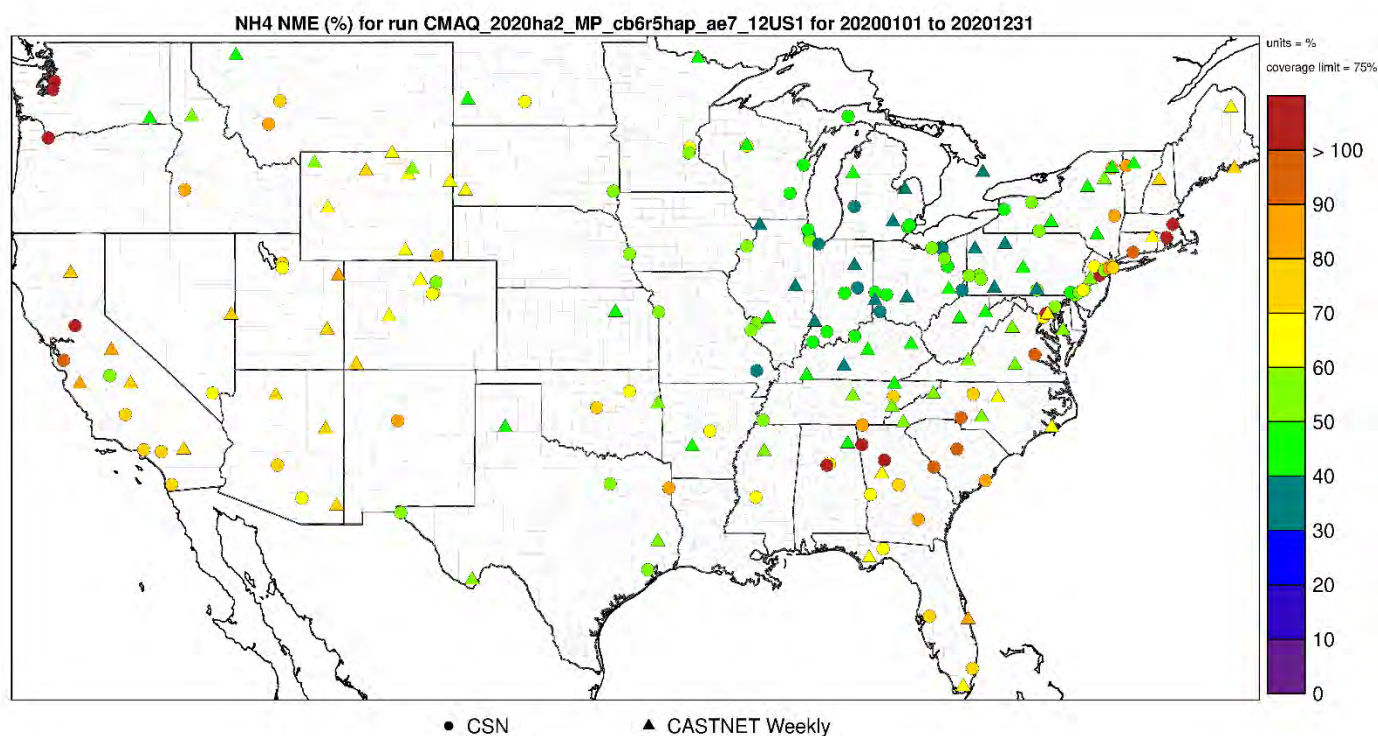
**Figure 4-19. Mean Bias ( $\mu\text{g}\cdot\text{m}^{-3}$ ) of annual ammonium at monitoring sites in the continental U.S. modeling domain.**



**Figure 4-20. Mean Error ( $\mu\text{g}\cdot\text{m}^{-3}$ ) of annual ammonium at monitoring sites in the continental U.S. modeling domain.**

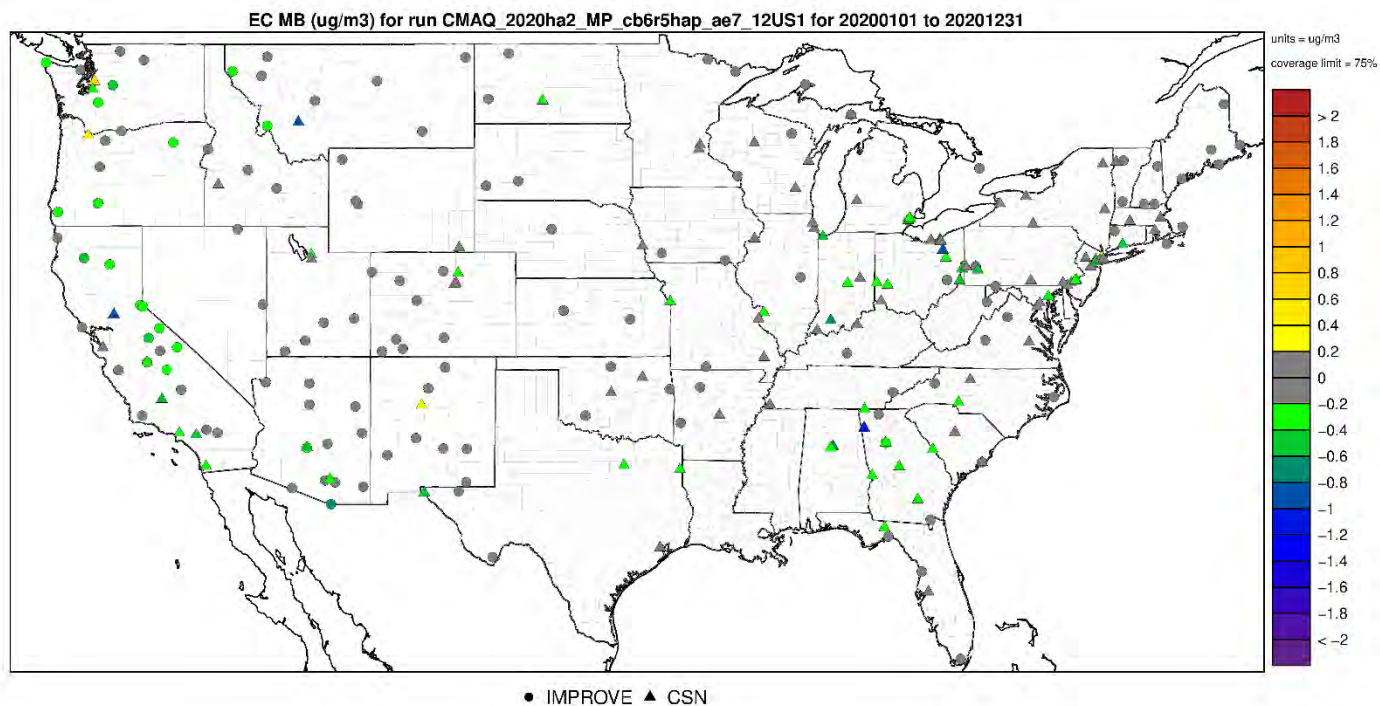


**Figure 4-21. Normalized Mean Bias (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.**

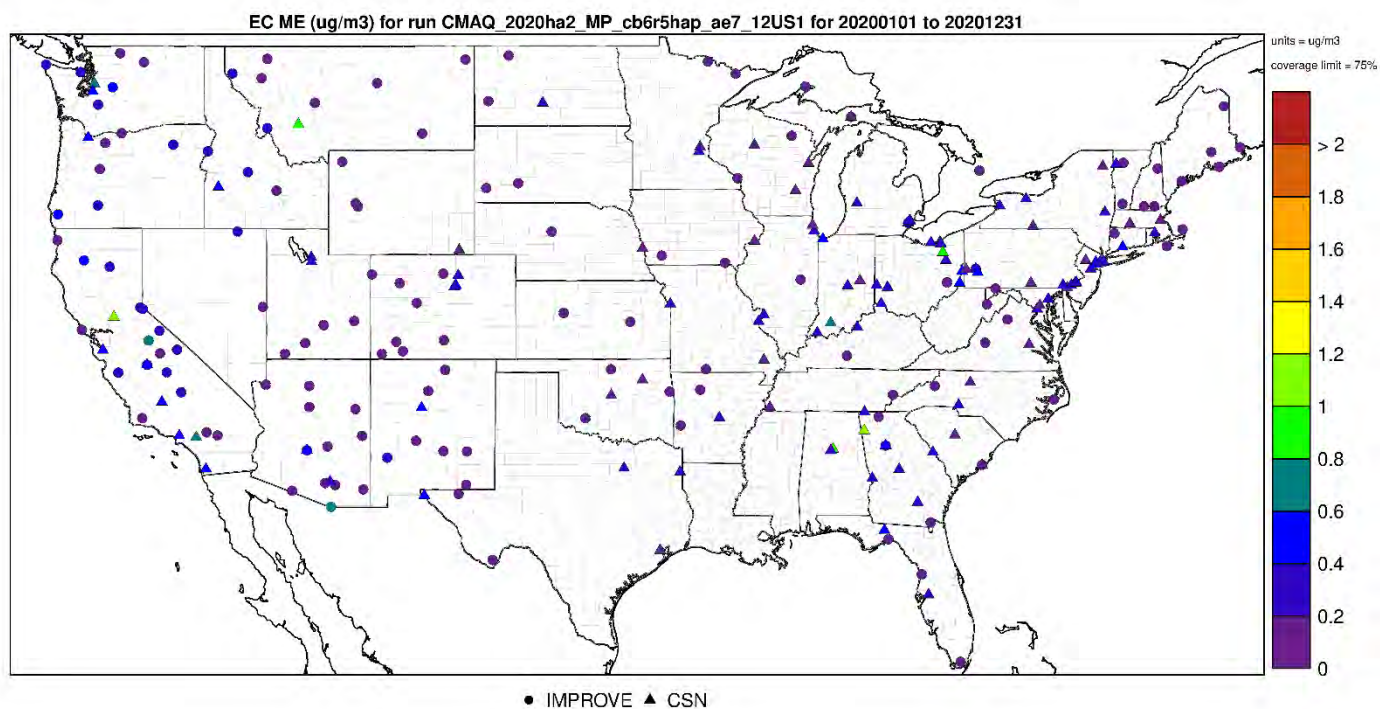


**Figure 4-22. Normalized Mean Error (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.**

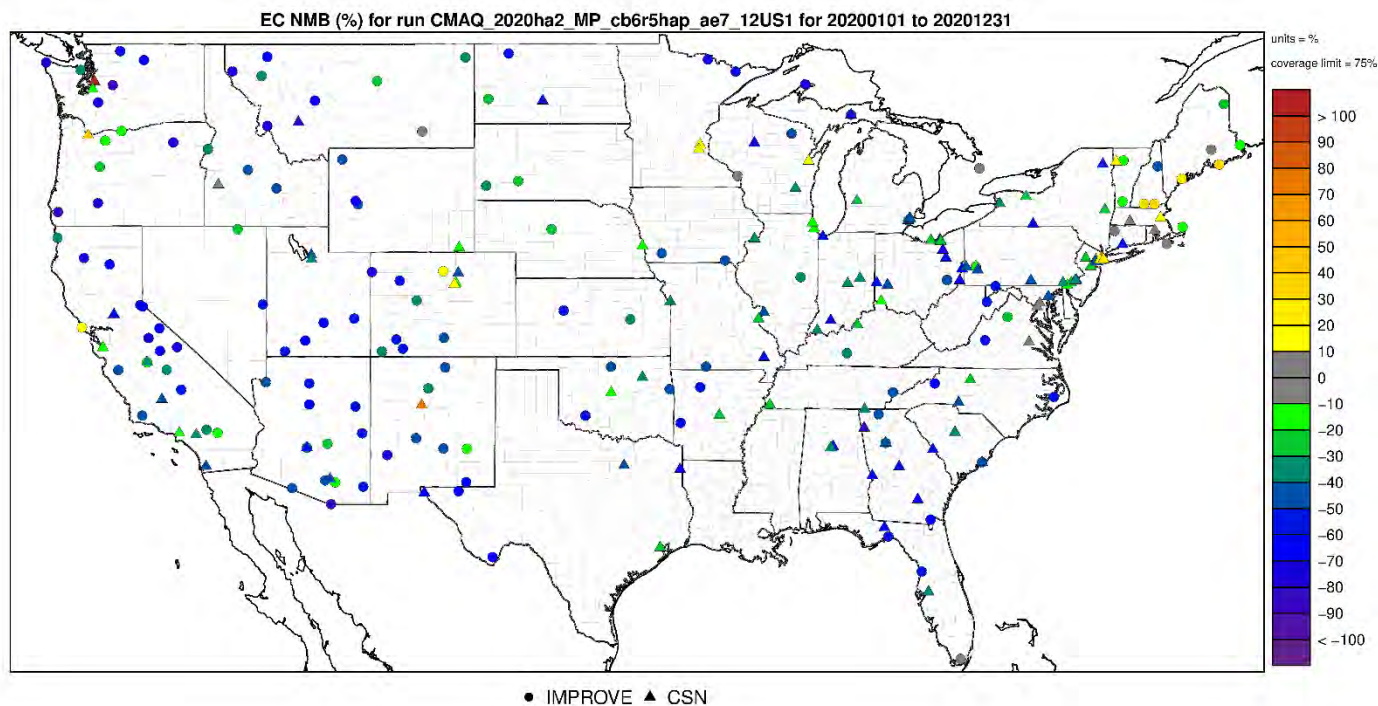




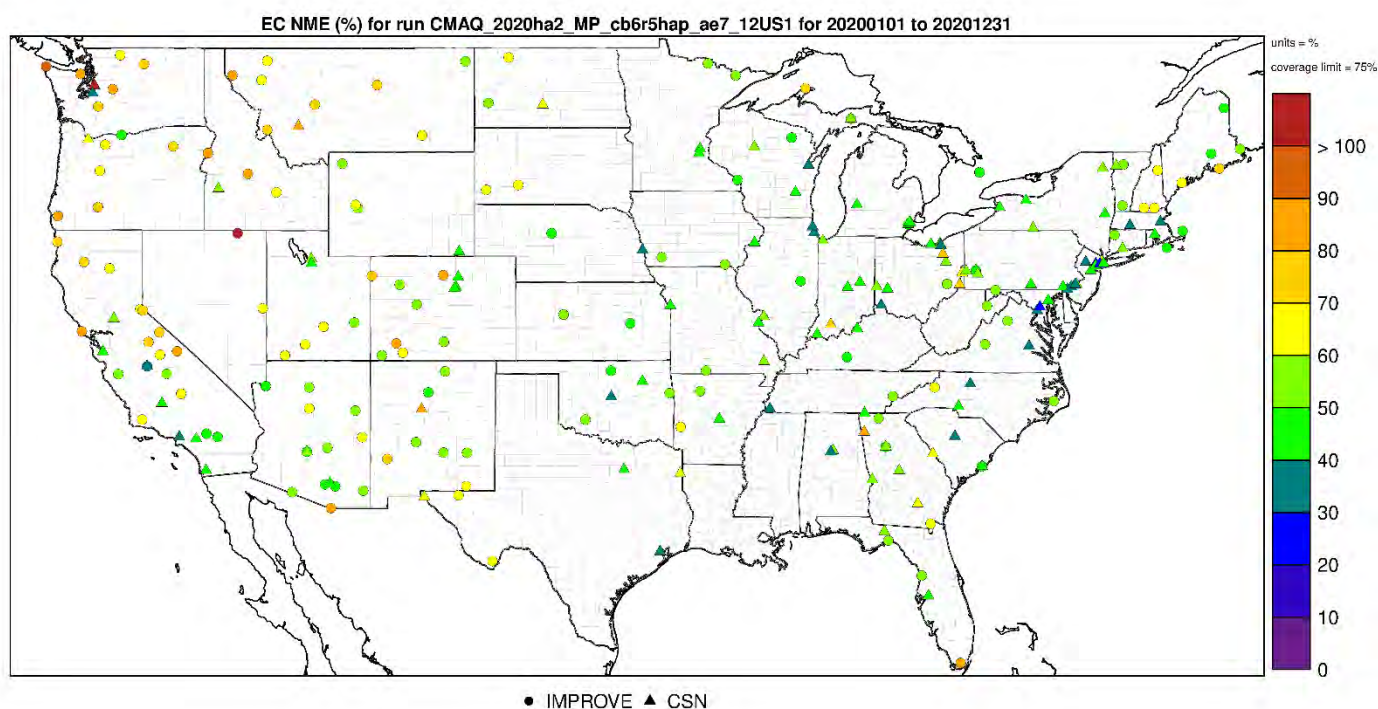
**Figure 4-23. Mean Bias ( $\mu\text{g}\text{m}^{-3}$ ) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.**



**Figure 4-24. Mean Error ( $\mu\text{g}\text{m}^{-3}$ ) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.**

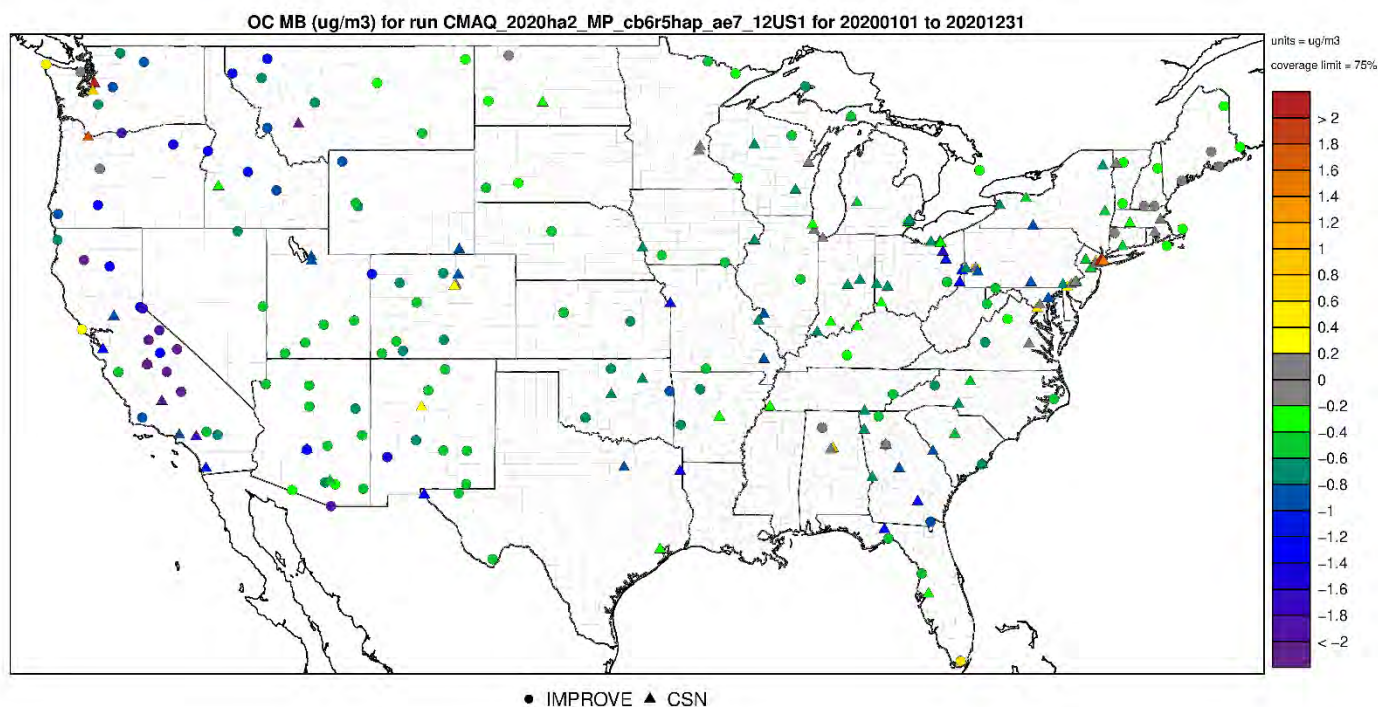


**Figure 4-25. Normalized Mean Bias (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.**

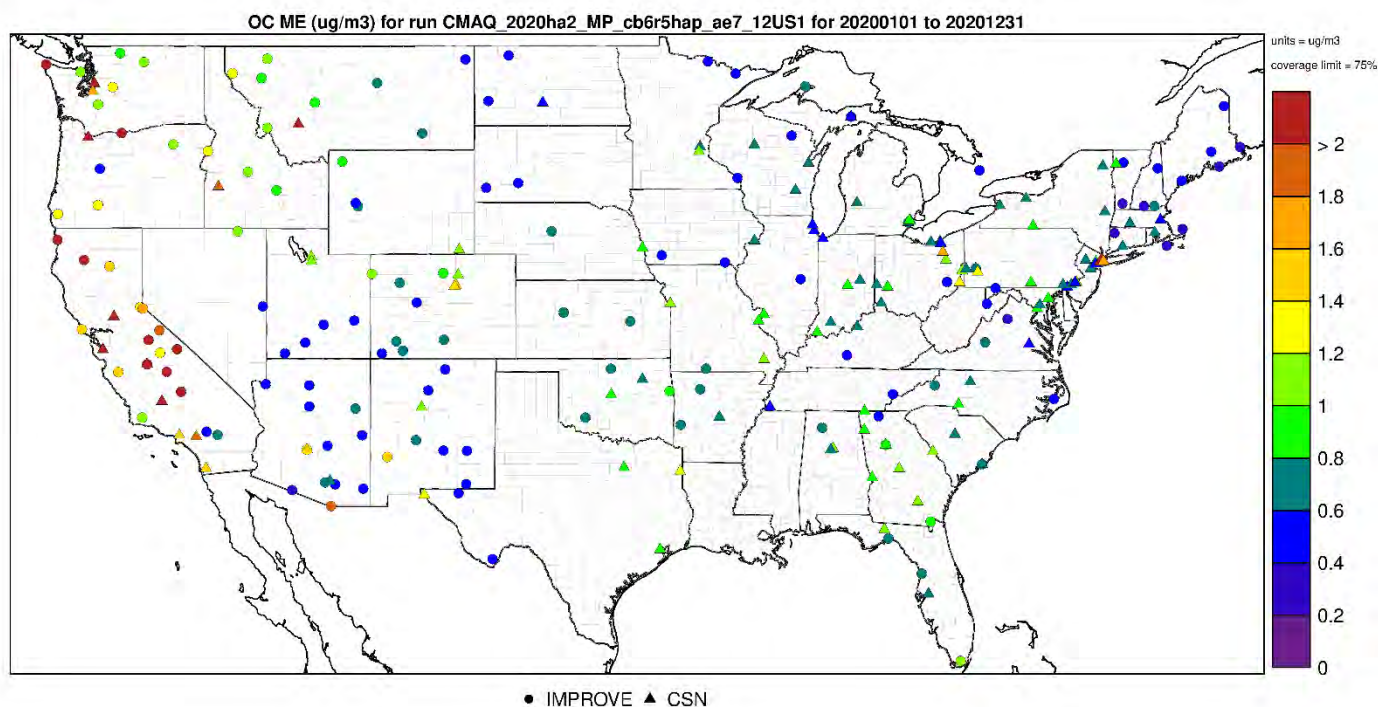


**Figure 4-26. Normalized Mean Error (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.**



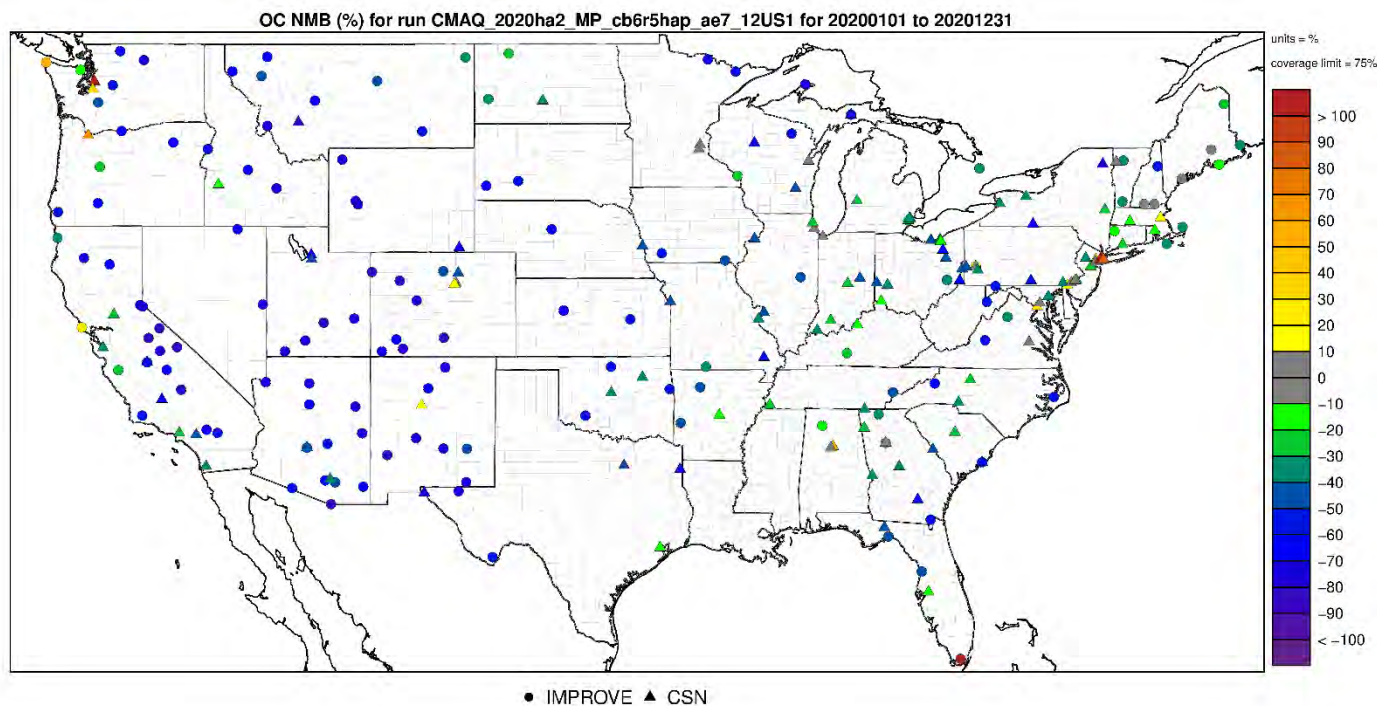


**Figure 4-27. Mean Bias ( $\mu\text{g}\text{m}^{-3}$ ) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.**

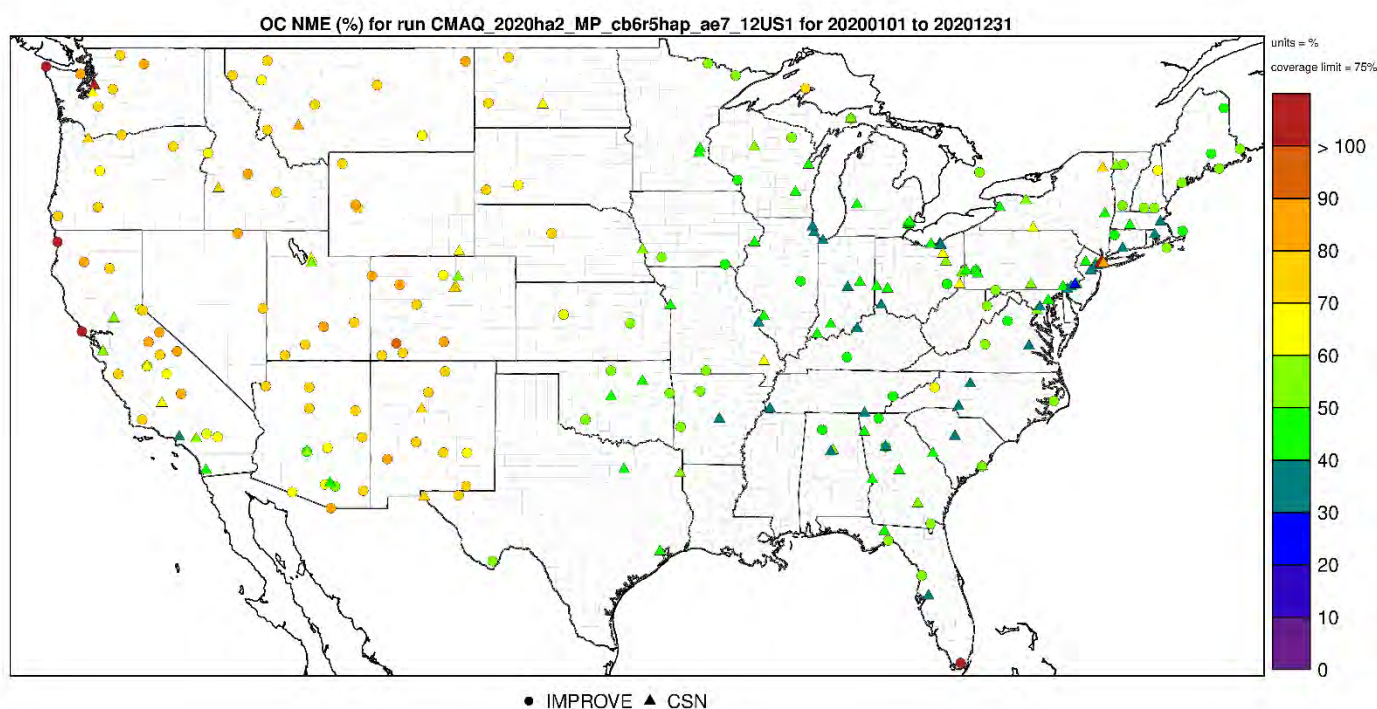


**Figure 4-28. Mean Error ( $\mu\text{g}\text{m}^{-3}$ ) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.**





**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 Error (%) 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 accuracy and CMAQs 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<sub>2.5</sub> data from AQS and CMAQ to produce predictions to continental U.S. 2020 census tract centroids for the year 2020.

### 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<sub>2.5</sub>). In its most general form, the model can be expressed in an equation similar to that of linear regression:

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

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,  $t=1, \dots, 365$  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  $\beta_0$  and a local component  $\beta_0(s)$  that is modeled as a mean-zero Gaussian Process with exponential decay

$\beta_1$  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  $\beta_1$  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 (i.e.,  $\tilde{\beta}_0(s)$ ,  $\beta_1$ ,  $\tilde{x}(s)$ ) 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)<sup>39</sup> 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

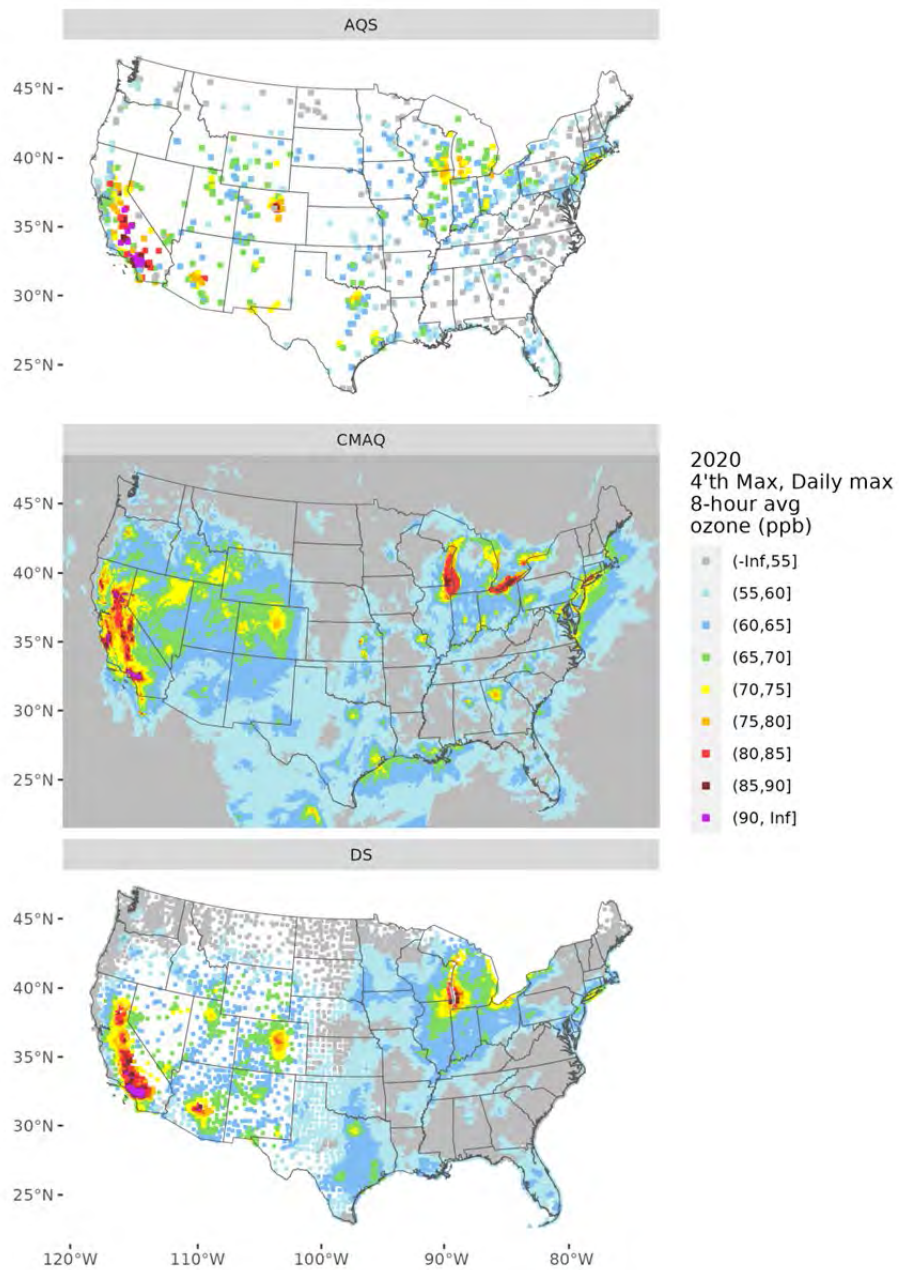
In this application, DS was used to predict daily concentration and associated uncertainty values at the 2020 US census tract centroids across the continental U.S. using measurement and CMAQ data as inputs. For ozone, the concentration unit is the daily maximum 8-hour average in ppb and for PM<sub>2.5</sub> the concentration unit is the 24-hour average in g/m.

#### 5.3.1 Summary of 8-hour Ozone Results

Figure 5-1 summarizes the AQS, CMAQ and DS ozone data over the year 2020. 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 2020, about 35% of the US Census tracts (29542 out of 83776) experienced at least one day with an ozone value above the NAAQS of 70 ppb.

---

<sup>39</sup> 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.

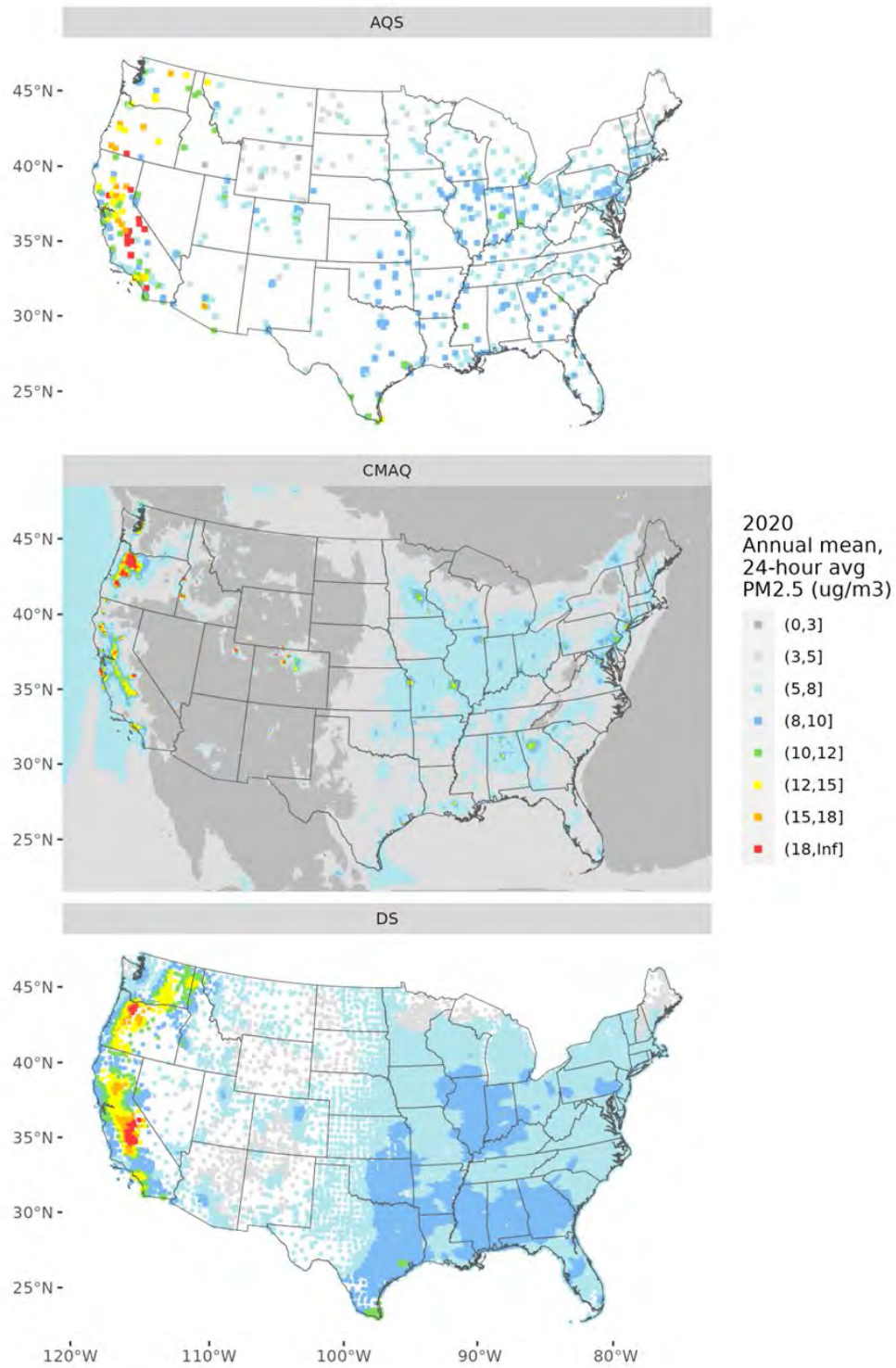


**Figure 5-1. Annual 4<sup>th</sup> max (daily max 8-hour ozone concentrations) derived from AQS, CMAQ and DS data.**

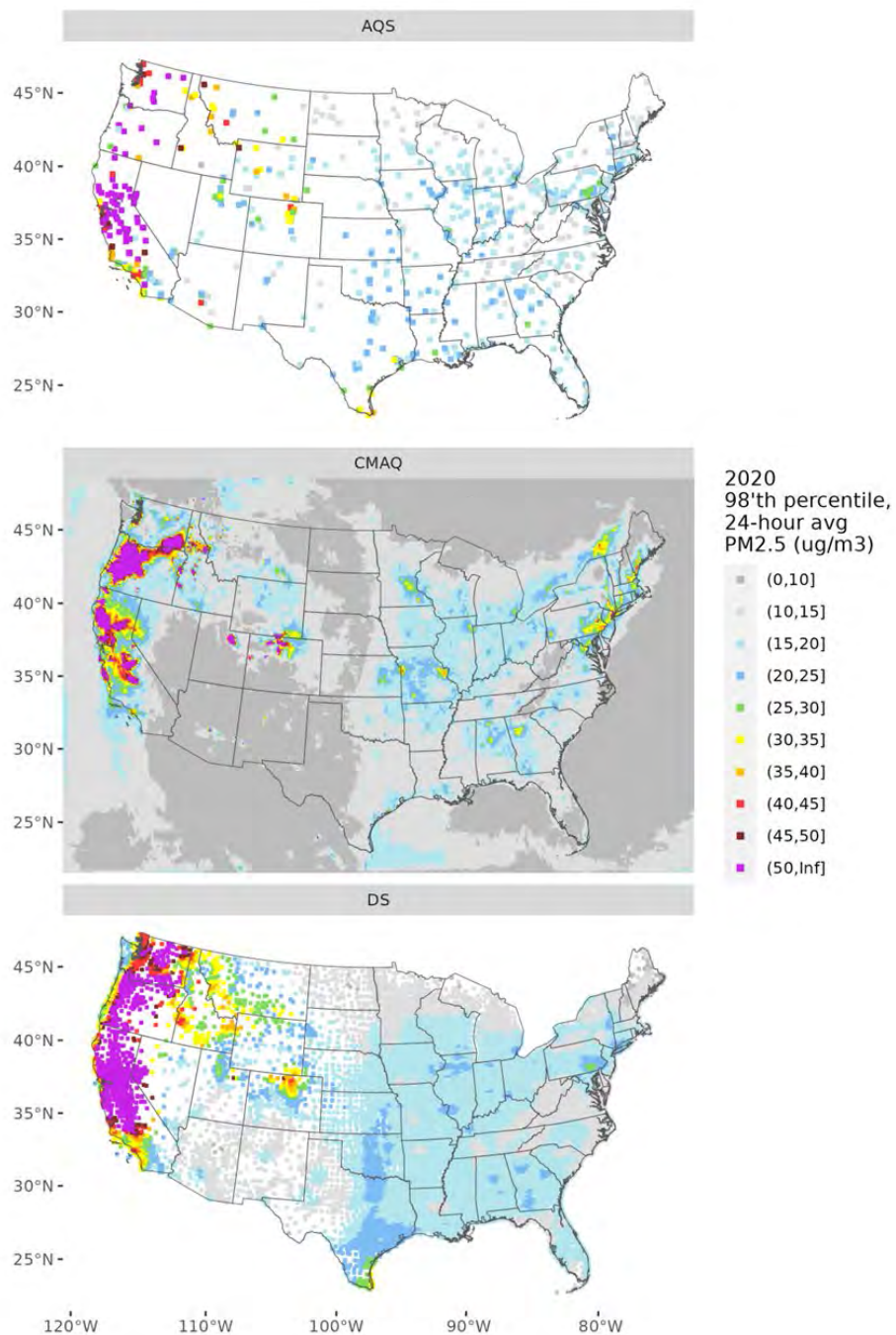
### **5.3.2 Summary of $PM_{2.5}$ Results**

Figures 5-2 and 5-3 summarize the AQS, CMAQ and DS  $PM_{2.5}$  data over the year 2020. Figure 5-2 shows annual means and Figure 5-3 shows 98th percentiles of 24-hour  $PM_{2.5}$  concentrations for AQS observations, CMAQ model predictions and DS model results. The DS model estimated that for 2020 about 40% of the US Census tracts (33298 out of 83776 experienced at least one day with a  $PM_{2.5}$  value above the 24-hour NAAQS of 35 g/m.





**Figure 5-2. Annual mean PM<sub>2.5</sub> concentrations derived from AQS, CMAQ and DS data.**



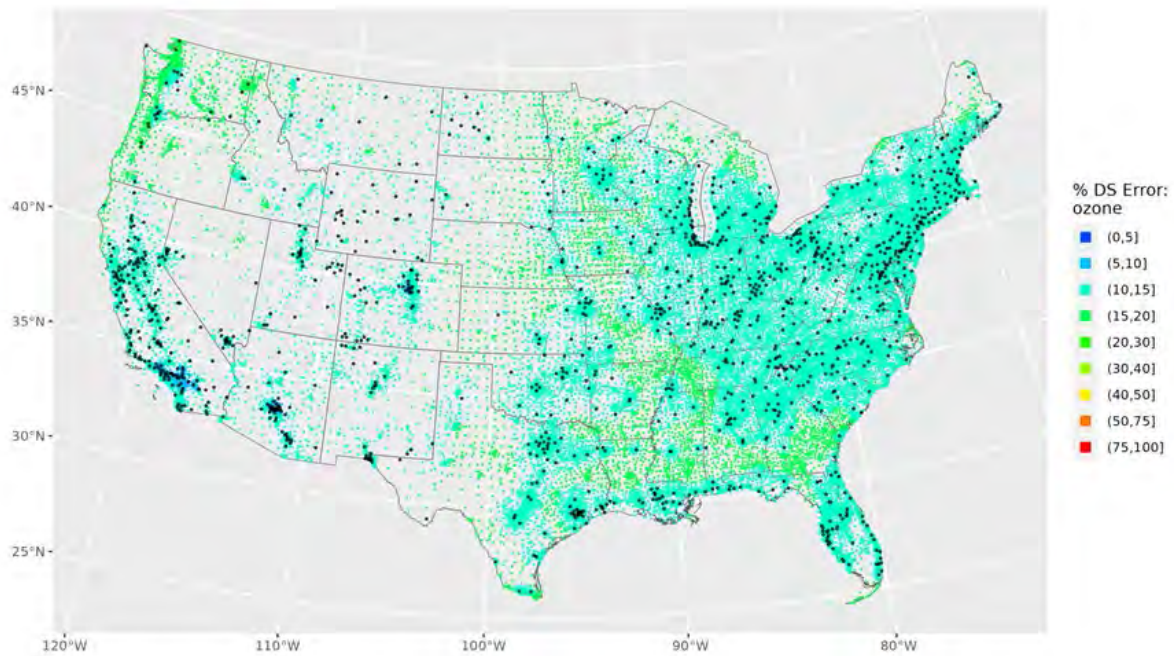
**Figure 5-3. 98<sup>th</sup> percentile 24-hour average PM<sub>2.5</sub> concentrations derived from AQS, CMAQ and DS data.**

#### 5.4 Downscaler Uncertainties

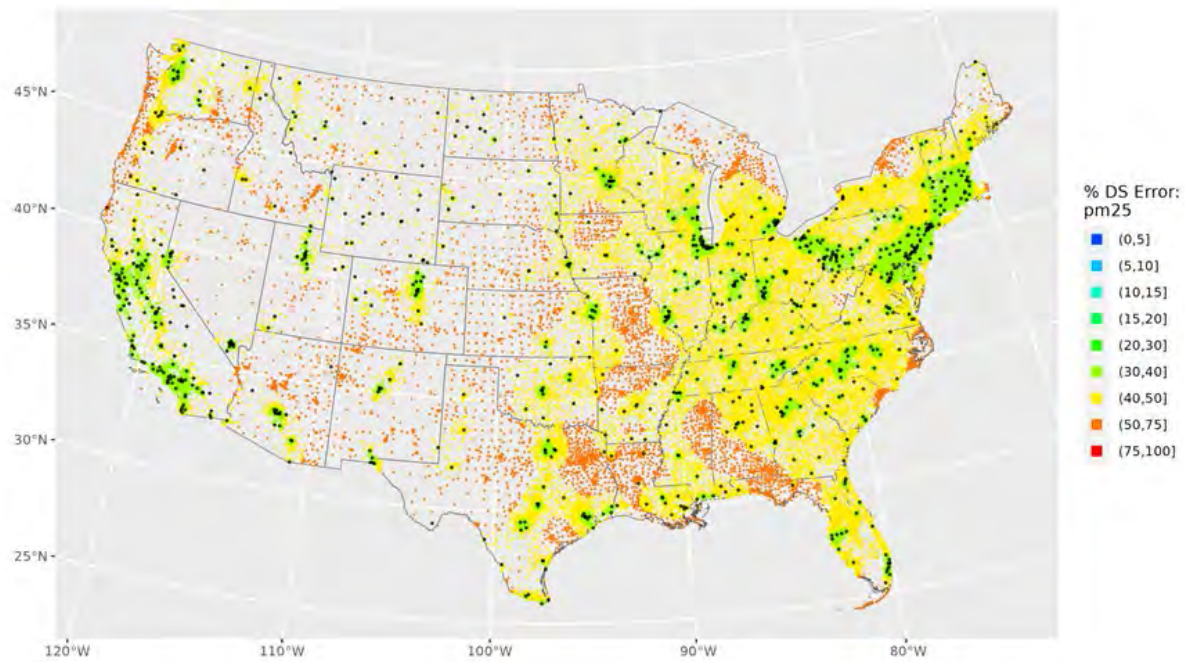


### 5.4.1 Standard Errors

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. Figures 5-4 and 5-5 show 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 situated 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 2020 runs for ozone. The black dots show the locations of monitors that generated the AQS data used as input to the DS model*



*Figure 5-5: Annual mean relative errors (standard errors divided by predictions) from the DS 2020 runs for PM<sub>2.5</sub>. 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	Monitor Count	Mean Bias	RMSE	Mean Coverage
PM	943	0.146	4.987	0.953
O3	1224	0.018	4.221	0.962

**Table 5-1. Cross-validation statistics associated with the 2020 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<sub>2.5</sub>. DS provided spatial predictions of daily ozone and PM<sub>2.5</sub> at 2020 U.S. census tract centroids by utilizing monitoring data and CMAQ output for 2020. Large-scale spatial and temporal patterns of concentration predictions are generally consistent with those seen in ambient monitoring data. Both ozone and PM<sub>2.5</sub> 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 grid cell. Care needs to be taken not to over-interpret any within-grid cell gradients that might be produced by a user. Fine-scale emission sources in CMAQ are diluted into the grid cell averages, but a given source within a grid cell 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 grid cell.

## Appendix A - Acronyms

### Acronyms

ARW	Advanced Research WRF core model
BEIS	Biogenic Emissions Inventory System
BlueSky	Emissions modeling framework
BSP	BlueSky Pipeline modeling system
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
HC	Hydrocarbon
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 <sub>2.5</sub>	Particulate matter less than or equal to 2.5 microns
PM <sub>10</sub>	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
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

## **Appendix B – Emissions Totals by Sector**

Please see the independent spreadsheet [AppendixB\\_2020\\_emissions\\_totals\\_by\\_sector.xlsx](#) that provides inventory and speciation emissions totals for each emissions modeling sector.

---

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-23-004  
December 2023

---