Bayesian Space-time Downscaling Fusion Model (Downscaler) - Derived Estimates of Air Quality for 2019

EPA-454/R-22-003
October 2022

Bayesian Space-time Downscaling Fusion Model (Downscaler) - Derived Estimates of Air Quality for 2019

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: Liz Naess (EPA/OAR) and David Mintz (EPA/OAR).

Contents

1.0 Introduction

This report describes estimates of daily ozone (maximum 8-hour average) and fine particulate matter ($PM_{2.5}$) (24-hour average) concentrations throughout the contiguous United States during the 2019 calendar year generated by EPA's recently developed data fusion method termed the "downscaler model" (DS). Air quality monitoring data from the State and Local Air Monitoring Stations (SLAMS) and numerical output from the Community Multiscale Air Quality (CMAQ) model were both input to DS to predict concentrations at the 2010 US census tract centroids encompassed by the CMAQ modeling domain. Information on EPA's air quality monitors, CMAQ model, and 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 $PM_{2.5}$.

DS essentially operates by calibrating CMAQ data to the observational data, and then uses the resulting relationship to predict "observed" concentrations at new spatial points in the domain. Although similar in principle to a linear regression, spatial modeling aspects have been incorporated for improving the model fit, and a Bayesian[1] 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/). 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.

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

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

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

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

Although these data have been processed on a computer system at the 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.

[2] The original HHS and EPA MOU is available at 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 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_3) and particulate matter (PM). The EPHT program has focused on these two pollutants, since numerous studies have found them to be most pervasive and harmful to public health and the environment, and there are extensive monitoring and modeling data available.

2.1 Introduction to Air Quality Impacts in the United States

2.1.1 The Clean Air Act

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

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

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

Form of the Standard (parts per million, ppm)	1997	2008	2015
Annual 4th 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.[3] Particles less than 10 micrometers in diameter (PM_{10}) pose a health concern because they can be inhaled into and accumulate in the respiratory system. Particles less than 2.5 micrometers in diameter ($PM_{2.5}$) are referred to as "fine" particles. Because of their small size, fine particles can lodge deeply into the lungs. Sources of fine particles include all types of combustion (motor vehicles, power plants, wood burning, etc.) and some industrial processes. Particles with diameters between 2.5 and 10 micrometers ($PM_{10-2.5}$) are referred to as "coarse" or PMc. Sources of PMc include crushing or grinding operations and dust from paved or unpaved roads. The distribution of PM_{10}, $PM_{2.5}$ and PMc varies from the eastern U.S. to arid western areas.

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

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

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

Table 2-2. $PM_{2.5}$ National Ambient Air Quality Standards

Form of the Standard (micrograms per cubic meter, µg/m^3)	**1997**	**2006**	**2012**
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 ($PM_{2.5}$) in 1997. The NAAQS for $PM_{2.5}$ were established based on their link to serious health problems ranging from increased symptoms, hospital admissions, and emergency room visits, to premature death in people with heart or lung disease. While most of the monitors in these networks are located in populated areas of the country, “background” and rural monitors are an important part of these networks. For more information on SLAMS, as well as EPA’s other air monitoring networks go to https://www.epa.gov/amtic.

In 2019, approximately 40 percent of the U.S. population was living within 10 kilometers of ozone and $PM_{2.5}$ monitoring sites. Highly populated areas in the eastern U.S. and California are well covered by both ozone and $PM_{2.5}$ monitoring network (Figure 2-1).

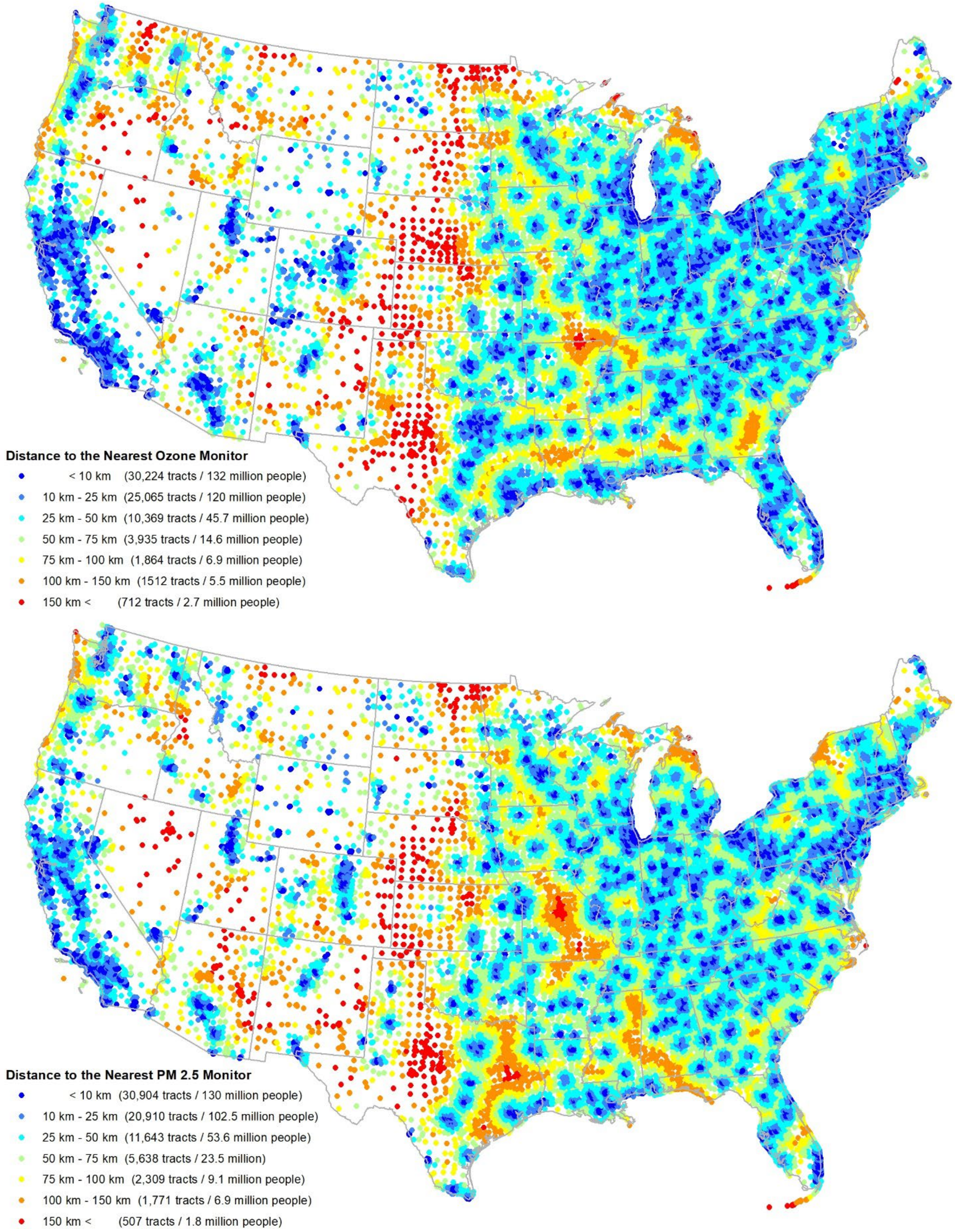


Figure 2-1. Distances from U.S. Census Tract centroids to the nearest monitoring site, 2019.

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 $PM_{2.5}$ day).

To compare/correlate ambient concentrations with acute health effects, daily local air quality data is needed.[4] Spatial gaps exist in the air quality monitoring network, especially in rural areas since the air quality monitoring network is designed to focus on measurement of pollutant concentrations in high population density areas. Temporal limits also exist. Hourly ozone measurements are aggregated to daily values (the daily max 8-hour average is relevant to the ozone standard). Ozone is typically monitored during the ozone season (the warmer months, approximately April through October). However, year-long data is available in many areas and is extremely useful to evaluate whether ozone is a factor in health outcomes during the non-ozone seasons. $PM_{2.5}$ is generally measured year-round. Most Federal Reference Method (FRM) $PM_{2.5}$ monitors collect data one day in every three days, due in part to the time and costs involved in collecting and analyzing the samples. Additionally, continuous monitors have become available which can automatically collect, analyze, and report $PM_{2.5}$ measurements on an hourly basis.

[4] 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.

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., $PM_{2.5}$ monitors do not operate continuously such as recording every third day or ozone monitors operate only certain part of the year) and limited spatial coverage (i.e., most monitors are located in urban areas and rural coverage is limited) hinder the ability of most of the interpolation techniques that use monitoring data alone as the input. If we look at the example of Voronoi Neighbor Averaging (VNA) (referred as the Nearest Neighbor Averaging in most literature), rural estimates would be biased towards the urban estimates. To further explain this point, assume the scenario of two cities with monitors and no monitors in the rural areas between, which is very plausible. Since exposure estimates are guaranteed to be within the range of monitors in VNA, estimates for the rural areas would be higher according to this scenario.

Air quality models may overcome some of the limitations that monitoring networks possess. Models such as 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[5] 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 $PM_{2.5}$ data from EPA. The approach used divides "indicators" into two categories. First, basic air quality measures were developed to compare air quality levels over space and time within a public health context (e.g., using the NAAQS as a benchmark). Next, indicators were developed that mathematically link air quality data to public health tracking data (e.g., daily $PM_{2.5}$ levels and hospitalization data for acute myocardial infarction). Table 2-3 and Table 2-4 describe the issues impacting calculation of basic air quality indicators.

Table 2-3. 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 $PM_{2.5}$ and ozone on health to: Identify populations that may have higher risk of adverse health effects due to $PM_{2.5}$ and ozone, Generate hypothesis for further research, and Provide information to support prevention and pollution control strategies.	Regular discussions have been held on health-air linked indicators and CDC/HFI/EPA convened a workshop January 2008. CDC has collaborated on a health impact assessment (HIA) with Emory University, EPA, and state grantees that can be used to facilitate greater understanding of these linkages.
Produce and disseminate basic indicators and other findings in electronic and print formats to provide the public, environmental health professionals, and policymakers, with current and easy-to-use information about air pollution and the impact on public health.	Templates and "how to" guides for $PM_{2.5}$ and ozone have been developed for routine indicators. Calculation techniques and presentations for the indicators have been developed.

[5] 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)
• Number of days with maximum ozone concentration over the NAAQS (or other relevant benchmarks (by county and MSA) • Number of person-days with maximum 8-hr average ozone concentration over the NAAQS & other relevant benchmarks (by county and MSA)
$PM_{2.5}$ (daily 24-hr integrated samples, ug/$m^{3,}$ by FRM)
• Average ambient concentrations of particulate matter (< 2.5 microns in diameter) and compared to annual $PM_{2.5}$ NAAQS (by state). • Percent of population exceeding annual $PM_{2.5}$ NAAQS (by state). • Percent of days with $PM_{2.5}$ concentration over the daily NAAQS (or other relevant benchmarks (by county and MSA) • Number of person-days with $PM_{2.5}$ concentration over the daily NAAQS & other relevant benchmarks (by county and MSA)

2.3.1 Rationale for the Air Quality Indicators

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

2.3.2 Air Quality Data Sources

The air quality data will be available in the 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 2019. The platform is based on the 2017 National Emissions Inventory (2017 NEI) published in January 2021 (EPA, 2021) along with other data specific to the year 2019. 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 2019 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), benzene, acetaldehyde, formaldehyde, methanol, naphthalene (the last five are also 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) added to CMAQ for the purposes of air quality modeling for the 2017 HAP+CAP platform.

Emissions were prepared for the Community Multiscale Air Quality (CMAQ) model (https://www.epa.gov/cmaq) version 5.3.2[6], which was used to model ozone (O_3) particulate matter (PM), and HAPs. CMAQ requires hourly and gridded emissions of the following inventory pollutants: carbon monoxide (CO), nitrogen oxides (NO_X), volatile organic compounds (VOC), sulfur dioxide (SO_2), ammonia (NH_3), particulate matter less than or equal to 10 microns (PM_{10}), and individual component species for particulate matter less than or equal to 2.5 microns ($PM_{2.5}$). In addition, the Carbon Bond 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 "2019ge", where 2019 is the year modeled, g represents that it was based on the 2017 NEI, and e represents that it was the fifth version of a 2017 NEI-based platform.

The effort to create the emission inputs for this study included development of emission inventories to represent emissions during the year of 2019, along with application of emissions modeling tools to convert the inventories into the format and resolution needed by CMAQ.

The emissions modeling platform includes point sources, nonpoint sources, commercial marine vessels (CMV), onroad and 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

[6] CMAQ version 5.3.2: https://doi.org/10.5281/zenodo.4081737; https://www.epa.gov/cmaq/cmaq-models-0. CMAQ v5.3.2 is also available from the Community Modeling and Analysis System (CMAS) at: http://www.cmascenter.org.

totals and annual temporal resolution. Emissions from Canada and Mexico are used in the CMAQ modeling but 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. Most of the remaining emission inventories were adjusted to represent 2019, primarily using 2017-specific emissions as a starting point.

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 4.8.1 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 3.8, 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 2019 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 (GHRSST) (see https://www.ghrsst.org/) and is given the EPA meteorological case abbreviation "19k." The full case abbreviation includes this suffix following the emissions portion of the case name to fully specify the abbreviation of the case as "2019ge_cb6_19k."

Following the emissions modeling steps to prepare emissions for CMAQ, it was run for the modeling domain covering the Continental United States. CMAQ produced outputs for the overall mass, chemistry and formation for specific hazardous air pollutants (HAPs) formed secondarily in the atmosphere (e.g., formaldehyde, acetaldehyde and acrolein. 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/2019-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 2019 North American Emissions Modeling Platform* (EPA, 2022).

3.2 Emission Inventories and Approaches

This section describes the emissions inventories created for input to SMOKE, which are based on the January 2021 version of the 2017 NEI along with the point source inventory for 2019 and other year 2019-specific data. The NEI includes five main data categories: a) nonpoint (formerly called "stationary area") sources; b) point sources; c) nonroad mobile sources; d) onroad mobile sources; and e) fires. For CAPs, the NEI data are largely compiled from data submitted by state, local and tribal (S/L/T) agencies. HAP emissions data are often augmented by EPA when they are not voluntarily submitted to the NEI by S/L/T agencies. The NEI was compiled using the Emissions Inventory System (EIS). EIS includes hundreds of automated QA checks to improve data quality, and it also supports release point (stack) coordinates separately from facility coordinates. EPA collaboration with S/L/T agencies helped prevent duplication between point and nonpoint source categories such as industrial boilers. The 2017 NEI Technical Support Document describes in detail the development of the 2017 emission inventories and is available at https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-technical-support-document-tsd (EPA, 2021).

The full NEI including all emissions source categories is developed every three years, with 2017 being the most recent year represented wih a full NEI. S/L/T agencies are required to submit large point sources to the NEI in interim years, including the year 2019. Where available, point source data representing 2019 were used for this study. Point sources in the 2017 NEI that did not have data submitted for the year 2019 and that were not marked as closed were pulled forward from the 2017 NEI into the 2019 point source inventory. The SMARTFIRE2 system and the BlueSky Pipeline (https://github.com/pnwairfire/bluesky) emissions modeling system were used to develop year 2019 fire emissions. SMARTFIRE2 categorizes all fires as either prescribed burning or wildfire categories, 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 2019 by running MOVES3 (https://www.epa.gov/moves).

With the exception of onroad, nonroad and fire emissions, Canadian emissions were based on the 2019 inventories developed for EPA's Air Quality Time Series (EQUATES) project (Foley, 2020). For Mexico, year 2016 inventories were projected to 2019. The latest year for which Canada and Mexico inventories were provided was 2016, although the onroad and nonroad emissions were adjusted to represent the year 2019 and some additional adjustments to the Canadian emissions were made for EQUATES.

The emissions modeling process, performed using SMOKE v4.8.1, apportions the emissions inventories into the grid cells used by CMAQ and temporalizes the emissions into hourly values. In addition, the pollutants in the inventories (e.g., NOx, PM and VOC) are split into the chemical species needed by CMAQ. For the purposes of preparing the CMAQ- ready emissions, the NEI emissions inventories by data category are split into emissions modeling platform "sectors"; and emissions from sources other than the NEI are added, such as the Canadian, Mexican, and offshore inventories. Emissions within the emissions modeling platform are separated into sectors for groups of related emissions source categories that are run through all of the 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 uses the inline biogenic emissions option and the CMAQ bidirectional ammonia process for fertilizer emissions.

Table 3-1 presents the sectors in the emissions modeling platform used to develop the year 2019 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_2019_emissions_totals_by_sector.xlsx".

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

Platform Sector: *abbreviation*	NEI Data Category	Description and resolution of the data input to SMOKE
EGU units: ***ptegu***	Point	2019 NEI point source EGUs, replaced with hourly Continuous Emissions Monitoring System (CEMS) values for NO_X and SO_2 , 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 2019 NEI point inventory. Annual resolution for sources not matched to CEMS data, hourly for CEMS sources. EGUs closed in 2019 are not part of the inventory.
Point source oil and gas: ***pt_oilgas***	Point	2019 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. Production-related sources that did not have 2019 data were pulled forward from the 2017 NEI and adjusted to 2019. Annual resolution.
Aircraft and ground support equipment: ***airports***	Point	2017 NEI point source emissions from airports, including aircraft and airport ground support emissions, adjusted to 2019 using Terminal Area Forecast (TAF) data. Airport-specific factors were used where available, state average factors were used for regional airports, and no change was made to military aircraft from 2017. Annual resolution.
Remaining non-EGU point: ***ptnonipm***	Point	All 2019 NEI point source records not matched to the airports, ptegu, or pt_oilgas sectors. Closures were reviewed and implemented based on the most recent submissions to the Emissions Inventory System (EIS). Includes 2017 NEI rail yard emissions, adjusted to 2019 using same projection factors as the rail sector. Annual resolution.
Livestock: ***livestock***	Nonpoint	2017 NEI nonpoint livestock emissions adjusted to 2019 using USDA survey data. Livestock includes ammonia and other pollutants (except $PM_{2.5}$). County and annual resolution.
Agricultural Fertilizer	Nonpoint	2019 agricultural fertilizer ammonia emissions computed inline within CMAQ.
Agricultural fires with point resolution: ***ptagfire***	Nonpoint	Agricultural fire sources for year 2019 were developed by EPA as point and day-specific emissions.[7] Only EPA-developed ag. fire data are used in this study, thus 2017 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.
Area fugitive dust: ***afdust_adj***	Nonpoint	PM_{10} and $PM_{2.5}$ fugitive dust sources from the 2017 NEI nonpoint inventory; including building construction, road construction, agricultural dust, and paved and unpaved road dust; with paved road dust adjusted to 2019 based on vehicle miles traveled (VMT). The emissions modeling system applies a transport fraction reduction and a zero-out based on 2019 gridded hourly meteorology (precipitation and snow/ice cover). Emissions are county and annual resolution.

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

Platform Sector: ***abbreviation***	**NEI Data Category**	**Description and resolution of the data input to SMOKE**
Biogenic: ***beis***	Nonpoint	Year 2019 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 3.7 of the Biogenic Emissions Inventory System (BEIS) was used with Version 5 of the Biogenic Emissions Landuse Database (BELD5).
Category 1, 2 CMV: ***cmv_c1c2***	Nonpoint	2019 Category 1 (C1) and Category 2 (C2), commercial marine vessel (CMV) emissions based on Automatic Identification System (AIS) data. Point and hourly resolution.
Category 3 CMV: ***cmv_c3***	Nonpoint	2019 Category 3 (C3) commercial marine vessel (CMV) emissions based on AIS data. Point and hourly resolution.
Locomotives : ***rail***	Nonpoint	Line haul rail locomotives emissions for year 2017, projected to 2019 using annual energy outlook (AEO) and additional factors supplied by ERTAC. County and annual resolution.
Nonpoint source oil and gas: ***np_oilgas***	Nonpoint	Nonpoint 2017 NEI sources from oil and gas-related processes, projected to 2019 using based on U.S. Energy Information Administration (EIA) and Railroad Commission of Texas (TXRRC) historical production data. County and annual resolution.
Residential Wood Combustion: ***rwc***	Nonpoint	2017 NEI nonpoint sources with residential wood combustion (RWC) processes were used as is, with no projection to represent 2019. County and annual resolution.
Solvents: ***np_solvents***	Nonpoint	Emissions of solvents for the year 2019 (Seltzer, 2021). Includes household cleaners, personal care products, adhesives, architectural and aerosol coatings, printing inks, and pesticides. Annual and county resolution.
Remaining nonpoint: ***nonpt***	Nonpoint	2017 NEI nonpoint sources not included in other platform sectors. No adjustments were made to represent 2019. County and annual resolution.
Nonroad: ***nonroad***	Nonroad	2019 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 and Texas. California submitted their own emissions for the 2017 NEI that were adjusted to 2019 based on interpolations between 2017 and 2023. Texas provided 2017 and 2020 emissions which were interpolated to 2019. County and monthly resolution.
Onroad: ***onroad***	Onroad	Onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles. 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. Activity data were projected from 2017 to 2019 using factors developed using data from Federal Highway Administration and state departments of transportation. MOVES3 was run for 2019 to generate emission factors.
Onroad California: ***onroad_ca_adj***	Onroad	California-provided 2017 CAP and metal HAP onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles based on Emission Factor (EMFAC) 2017, gridded and temporalized based on outputs from MOVES3. Volatile organic compound (VOC) HAP emissions derived from California-provided VOC emissions and MOVES-based speciation. 2019 was interpolated between 2017 and 2023 emissions from EMFAC2017.

Platform Sector: ***abbreviation***	**NEI Data Category**	**Description and resolution of the data input to SMOKE**
Point source prescribed fires: ***ptfire-rx***	Events	Point source day-specific prescribed fires for 2019 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.
Point source wildfires: ***ptfire-wild***	Events	Point source day-specific wildfires for 2019 computed using SMARTFIRE 2 and BlueSky Pipeline.
Non-US. Fires: ***ptfire_othna***	N/A	Point source day-specific wildfires and agricultural fires outside of the U.S. for 2019 from v1.5 of the Fire INventory (FINN) from National Center for Atmospheric Research (NCAR, 2017 and Wiedinmyer, C., 2011) for Canada, Mexico, Caribbean, Central American, and other international fires.
Other Area Fugitive dust sources not from the NEI: ***othafdust***	N/A	Area fugitive dust sources from Canada from EQUATES 2016 with transport fraction and snow/ice adjustments based on 2019 meteorological data. Annual and province resolution.
Other Point Fugitive dust sources not from the NEI: ***othptdust***	N/A	Point source fugitive dust sources from Canada from EQUATES 2016 with transport fraction and snow/ice adjustments based on 2019 meteorological data. Annual and province resolution.
Other point sources not from the NEI: ***othpt***	N/A	Canada and Mexico point source emissions from EQUATES 2016. Canada point sources were provided by ECCCC and Mexico point source emissions for 2016 were provided by SEMARNAT. Mexico sources were projected to 2019 based on national emissions trends from the Community Emissions Data System (CEDS). Annual and monthly resolution.
Other non-NEI nonpoint and nonroad: ***othar***	N/A	For Canada except nonroad, EQUATES 2016. Projected Canada nonroad to 2019 based on US MOVES3 2019/2016 ratios. EQUATES 2016 Mexico (municipio resolution, provided by SEMARNAT) nonpoint and nonroad mobile inventories were projected to 2019 based on national emissions trends from the Community Emissions Data System (CEDS). Annual and monthly resolution.
Other non-NEI onroad sources: ***onroad_can***	N/A	Monthly onroad mobile inventory for Canada from EQUATES 2016 projected to 2019 using US onroad trends. Separate trends applied to refueling (gas/diesel) and non-refueling (gas/diesel and LD/HD). Province resolution.
Other non-NEI onroad sources: ***onroad_mex***	N/A	Monthly onroad mobile inventory from MOVES-Mexico (municipio resolution) for 2017, adjusted to 2019 using interpolation between 2017 and 2020.

Other natural emissions are also merged in with the above sectors, including ocean chlorine and sea salt. The ocean chlorine gas emission estimates are based on the build-up of molecular chlorine (Cl_2) concentrations in oceanic air masses (Bullock and Brehme, 2002).

Data at 12 km resolution were available and were not modified other than the name "CHLORINE" was changed to "CL2" because that is the name required by the CMAQ model.

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/2019-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 2019 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.

Table 3-2. 2019 Continental United States Emissions by Sector (tons/yr in 48 states + D.C.)

Sector	CO	NH3	NOX	PM10	PM2_5	SO2	VOC
afdust_adj				5,391,767	753,763		
airports	494,906	0	135,054	9,899	8,661	16,907	55,888
cmv_c1c2	17,699	60	117,678	3,191	3,092	596	4,264
cmv_c3	9,509	30	95,659	1,683	1,549	3,864	4,418
fertilizer		1,202,914					
livestock		2,602,279					227,985
nonpt	1,927,267	102,898	739,250	572,589	475,154	166,399	818,185
nonroad	10,464,693	1,953	930,665	90,623	85,372	1,168	974,635
np_oilgas	629,304	27	520,213	11,180	11,091	37,872	2,581,373
np_solvents	36	58	34	469	448	5	2,516,324
onroad	16,637,063	102,432	2,780,428	215,217	94,256	16,657	1,162,167
ptegu	479,731	20,734	999,436	116,793	98,046	1,017,892	29,201
ptagfire	687,701	146,655	27,373	104,474	66,053	10,683	99,156
ptfire-rx	9,176,460	150,531	163,988	986,337	842,244	80,673	2,200,144
ptfire-wild	2,121,948	34,930	34,299	220,638	186,981	17,581	502,126
ptnonipm	1,359,439	68,482	855,115	380,052	241,176	500,592	760,023
pt_oilgas	174,426	3,759	354,632	13,514	13,190	36,934	143,335
rail	108,825	339	516,008	15,304	14,821	675	21,940
rwc	2,152,689	16,369	33,925	298,738	297,877	7,937	322,528
beis (not merged)	3,695,221		942,563				25,450,181
TOTAL no beis	46,441,694	4,454,448	8,303,756	8,432,468	3,193,775	1,916,436	12,423,693

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

Sector	CO	NH3	NOX	PM10	PM2_5	SO2	VOC
Canada ag		492,799					105,147
Canada oil and gas 2D	666	7	3,232	185	185	3,933	509,228
Canada othafdust				500,478	77,863		
Canada othptdust				124,644	43,736		
Canada othar	2,180,838	3,818	298,213	222,283	173,889	16,299	721,690
Canada onroad_can	1,587,108	7,125	327,671	24,783	12,312	1,121	132,251
Canada othpt	1,115,145	19,471	650,682	90,031	43,039	989,862	148,178

Sector	CO	NH3	NOX	PM10	PM2_5	SO2	VOC
Canada ptfire_othna	1,503,071	30,374	63,142	210,958	178,315	12,143	439,097
Canada cmv_c1c2	2,811	9	18,432	489	474	60	656
Canada cmv_c3	8,198	22	83,037	1,260	1,159	2,936	3,967
Canada subtotal	**6,397,837**	**553,625**	**1,444,410**	**1,175,112**	**530,971**	**1,026,356**	**2,060,214**
Mexico othar	109,035	115,777	50,564	101,331	32,933	1,576	348,466
Mexico onroad_mex	1,812,455	2,983	447,355	16,112	11,380	6,832	159,395
Mexico othpt	154,181	1,154	172,530	47,798	33,325	345,440	37,488
Mexico ptfire_othna	367,565	7,109	14,666	48,343	41,333	3,026	106,475
Mexico cmv_c1c2	152	1	1,016	27	26	2	45
Mexico cmv_c3	8,091	185	89,440	10,433	9,598	84,305	3,782
Mexico subtotal	**2,451,479**	**127,208**	**775,571**	**224,044**	**128,596**	**441,181**	**655,652**
Offshore cmv_c1c2	4,205	13	26,796	700	678	72	1,007
Offshore cmv_c3	45,473	538	472,445	30,400	27,968	225,335	21,595
Offshore pt_oilgas	51,866	8	49,959	636	635	462	38,803
Can/Mex/offshore total	**8,950,860**	**681,393**	**2,769,180**	**1,430,892**	**688,848**	**1,693,406**	**2,777,271**

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, with 2017 being the most recently finished complete NEI. A comprehensive description about the development of the 2017 NEI is available in the 2017 NEI TSD (EPA, 2021). Point inventories are also available in EIS for intermediate years such as 2019. In the intermediate point inventories, states are required to update larger sources with emissions for the interim year, while sources not updated by states for the interim year are 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 2019 into the Flat File 2010 (FF10) format that is compatible with SMOKE (see https://www.cmascenter.org/smoke/documentation/4.8.1/html/ch08s02s08.html) and was then split into several sectors for modeling. The 20220325 version of the point FF10 file was used for the CMAQ and AERMOD modeling. In the flat file, 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 summary tracking purposes and distinct projection techniques from the remaining non-EGU emissions (ptnonipm).

In some cases, data about facility or unit closures are entered into EIS after the inventory modeling inventory flat were reviewed and implemented based on the most recent submissions to EIS. Prior to processing through SMOKE, submitted closures were reviewed and if closed sources were found in the inventory, those were removed.

For the 2019 platform, an analysis of point source stack parameters (e.g., stack height, diameter, temperature, and velocity) was performed after some specific examples of unrealistic 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 currently available 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_X, VOC, SO_2, NH_3, PM_{10}, and $PM_{2.5}$ 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 2019 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 2019 CEMS data, hourly CEMS NO_X and SO_2 emissions for 2019 from EPA's Acid Rain Program were used rather than annual inventory emissions. For all other pollutants (e.g., VOC, $PM_{2.5}$, HCl), annual emissions were used as-is from the annual inventory but were allocated to hourly values using heat input from the CEMS data. For the unmatched units in the ptegu sector, annual emissions were allocated to daily values using IPM region- and pollutant-specific profiles, and similarly, region- and pollutant-specific diurnal profiles were applied to create hourly emissions.

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

a. CAP and HAP data were provided by States, locals and tribes under the Air Emissions Reporting Rule (AERR) [the reporting size threshold is larger for inventory years between the triennial inventory years of 2011, 2014, 2017, …].
b. EPA corrected known issues and filled PM data gaps.
c. EPA added HAP data from the Toxic Release Inventory (TRI) where corresponding data was not already provided by states/locals.
d. EPA 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 2019 point source inventory that could be matched to units found in the National Electric Energy Database System (NEEDS) v6 (see https://www.epa.gov/power-sector-modeling/national-electric-energy-data-system-needs-v6) that is used by the Integrated Planning Model (IPM) to develop future year EGU emissions. It was necessary to put these EGUs into a separate sector in the platform because EGUs use different temporal profiles than other sources in the point sector and it is useful to segregate these emissions from the rest of the point sources to facilitate summaries of the data. Sources not matched to units found in NEEDS are placed into the pt_oilgas or ptnonipm sectors. For studies with future year cases, the sources in the ptegu sector are fully replaced with the emissions output from IPM. It is therefore important that the matching between the NEI and NEEDS database be as complete as possible because there can be double-counting of emissions in future year modeling scenarios if emissions for units are projected by IPM are not properly matched to the units in the point source inventory.

The 2019 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 described in the list below:

- 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_X and SO_2 in the flat file are replaced with the hourly CEMS emissions in base year modeling. For other pollutants at matched units, the hourly CEMS heat input data are 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 are 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 http://ampd.epa.gov/ampd near the bottom of the "Prepackaged Data" tab. Many smaller emitters in the CEMS program cannot be matched to the NEI due to inconsistencies 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 (Adelman, 2012) 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.

For sources that otherwise would be pulled forward with 2017 emissions values because 2019-specific emissions were not available, projection factors by NAICS and state derived from historical production data from EIA. The factors were applied to those 2017 sources to adjust the emissions to make them more representative of 2019. Texas historical production data by Texas Railroad district (http://webapps.rrc.texas.gov/PDQ/generalReportAction.do) were used to derive and apply district-specific factors instead of state-specific. State (plus TX Railroad Commission district) factors were applied to production-related NAICS. Transportation NAICS were projected using nationally derived production-related factors for oil and gas. All other NAICS were held constant from 2017 NEI. All Tribal data and offshore emissions are held constant from 2017 NEI. More information on the development of the 2017 NEI oil and gas emissions can be found in Section 4.17 of the 2017 NEI TSD.

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

NAICS	NAICS description
2111	Oil and Gas Extraction

NAICS	NAICS description
211111	Crude Petroleum and Natural 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. For 2017, Texas and California submitted aircraft emissions. Additional information about aircraft emission estimates can be found in section 3.2.2 of the 2017 NEI TSD. Data from the 2020 Terminal Area Forecast (TAF) were used to project 2017 NEI emissions to 2019. EPA used airport-specific factors where available. Regional airports were projected using state average factors. Military airports were unchanged from 2017. An update for the 2019 platform was that 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 non-IPM (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 the non-EGU emissions sources and rail yards. However, it is likely 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 2019 NEI point inventory. Unlike in the 2018 platform, instead of removing solvent emissions from the ptnonipm sector, solvent emissions from point sources are instead 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 on a version of the point flat file which included corrections to how the selection was implemented in EIS, updates from the state/local review, and updates specific to ethylene oxide. The np_solvents sector is described in more detail in Section 3.2.3.6.

Emissions from rail yards are included in the ptnonipm sector. Railyards were projected to 2019 from the 2017 NEI railyard inventory using factors derived from the Annual Energy Outlook 2018 (https://www.eia.gov/outlooks/archive/aeo18/).

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 2017 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_{10} and $PM_{2.5}$ emission estimates for nonpoint SCCs identified by EPA staff as dust sources. Categories included in the afdust sector are paved roads, unpaved roads and airstrips, construction (residential, industrial, road and total), agriculture production, and mining and quarrying. It does not include fugitive dust from grain elevators, coal handling at coal mines, or vehicular traffic on paved or unpaved roads at industrial facilities because these are treated as point sources so they are properly located.

The afdust sector is separated from other nonpoint sectors to allow for the application of a "transport fraction," and meteorological/precipitation reductions. These adjustments are 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 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). Both the transport fraction and meteorological adjustments are 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 is accounted for in the subsequent meteorological adjustment.

Paved road dust emissions were projected from the 2017 NEI (January 2021 version) to 2019 based on county-level VMT trends. All other afdust SCCs were held constant from the 2017 NEI. For the data compiled into the 2017 NEI, meteorological adjustments are applied to paved and unpaved road SCCs but not transport adjustments. This is because the modeling platform applies meteorological adjustments and transport adjustments based on unadjusted NEI values. For the 2019 platform, the meteorological adjustments that were applied (to paved and unpaved road SCCs) were backed out in order reapply them in SMOKE. 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 2019 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 SCCs are related to beef and dairy cattle, poultry production and waste, swine production, waste from horses and ponies, and production and waste for sheep, lambs, and goats. The sector does not include quite all of the livestock NH_3 emissions, as there is a very small amount of NH_3 emissions from livestock in the ptnonipm inventory (as point sources). In addition to NH_3, the sector includes livestock emissions from all pollutants other than $PM_{2.5}$. $PM_{2.5}$ from livestock are in the afdust sector.

Agricultural livestock emissions in the 2019 platform were projected from the 2017 NEI (January 2021 version), which is a mix of state-submitted data and EPA estimates. USDA Survey data for 2017 and 2019 was used to create projection factors (https://quickstats.nass.usda.gov/). Livestock emissions utilized improved animal population data. VOC livestock emissions, new for this sector, were estimated by multiplying a national VOC/NH_3 emissions ratio by the county NH_3 emissions. The 2017 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 2017 U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) survey. Details on the approach are provided in Section 4.5 of the 2017 NEI TSD.

For livestock, meteorological-based temporalization is used for month-to-day and day-to-hour temporalization. Monthly profiles for livestock are based on the daily data underlying the EPA estimates from 2014NEIv2.

3.2.3.3 *Agricultural Fertilizer Sector (fertilizer)*

As described in the 2017 NEI TSD, fertilizer emissions for this platform are based on the FEST-C model (https://www.cmascenter.org/fest-c/). The bidirectional version of CMAQ (v5.3) and the Fertilizer Emissions Scenario Tool for CMAQ FEST-C (v1.3) were used within CMAQ to estimate ammonia (NH_3) emissions from agricultural soils. Fertilizer emissions are output from a run of CMAQ in bi-directional mode and summarized for inclusion with the rest of the emissions. The bidirectional version of CMAQ (v5.3) and the Fertilizer Emissions Scenario Tool for CMAQ FEST-C (v1.3) were used to estimate ammonia (NH_3) emissions from agricultural soils. The computed emissions were saved during the CMAQ run for the purposes of summaries and other model runs that did 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. Then CMAQ v5.3 was run with the Surface Tiled Aerosol and Gaseous Exchange (STAGE) deposition option with bidirectional exchange to estimate fertilizer and biogenic NH3 emissions.

The approach to estimate year-specific fertilizer emissions consists of these steps:

- Run FEST-C and CMAQ model with bidirectional ("bidi") NH_3 exchange to produce nitrate (NO_3), Ammonium (NH4+, including Urea), and organic (manure) nitrogen (N) fertilizer usage estimates, and gaseous ammonia NH_3 emission estimates respectively.
- Calculate county-level emission factors as the ratio of bidirectional CMAQ NH_3 fertilizer emissions to FEST-C total N fertilizer application.
- Assign the NH_3 emissions to one SCC: "…Miscellaneous Fertilizers" (2801700099).

For livestock and fertilizer, meteorological-based temporalization is used for month-to-day and day-to-hour temporalization. Monthly profiles for livestock are based on the daily data underlying the EPA estimates from 2014NEIv2. The fertilizer inventory includes monthly emissions from FEST-C and uses the same meteorological-based month-to-hour profiles as livestock in the same way as was done for other recent platforms.

3.2.3.4 *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 2019 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 2017NEI version of the Nonpoint Oil and Gas Emission Estimation Tool (i.e., the “tool”) was used to estimate emissions for 2019. Year 2019 oil and gas activity data was supplied to EPA by Enverus’ activity database (www.enverus.com) and from some state air agencies . The 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 created a CSV-formatted emissions dataset covering all national nonpoint oil and gas emissions. This dataset was then converted to the FF10 format for use in SMOKE modeling. More details on the inputs for and running of the tool with 2017 as an example are provided in the 2017 NEI TSD.

3.2.3.5 Residential Wood Combustion Sector (rwc)

The residential wood combustion (rwc) sector includes residential wood burning devices such as fireplaces, fireplaces with inserts (inserts), free standing woodstoves, pellet stoves, outdoor hydronic heaters (also known as outdoor wood boilers), indoor furnaces, and outdoor burning in firepots and 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 2019 platform RWC emissions are unchanged from the data in the 2017 NEI. Some improvements to RWC emissions estimates were made for the 2017 NEI and were included in this study. 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 missions 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 4.15 of the 2017 NEI TSD.

3.2.3.6 Solvents (np_solvents)

The np_solvents sector is a diverse collection of emission sources whose 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).

Here, emissions from this sector are derived using the volatile chemical products in python (VCPy) framework (Seltzer et al., 2021). The VCPy framework is based on the principle that the magnitude and speciation of organic emissions from this sector are directly related to (1) the mass of chemical products used, (2) the composition of these products, (3) the physiochemical properties of their constituents that govern volatilization, and (4) the timescale available for these constituents to evaporate. National product usage is preferentially estimated using economic statistics from the U.S. Census Bureau's Annual Survey of Manufacturers (U.S. Census Bureau, 2019), commodity prices from the U.S. Department of Transportation's 2012 Commodity Flow Survey (U.S. Department of Transportation, 2015) and the U.S. Census Bureau's Paint and Allied Products Survey (U.S. Census Bureau, 2011), and producer price indices, which scale commodity prices to target years, are retrieved from the Federal Reserve Bank of St. Louis (U.S. Bureau of Labor Statistics, 2020). In circumstances where the aforementioned datasets are unavailable, default usage estimates are derived using functional solvent usage reported by a business research company (The Freedonia Group, 2016) or in sales reported in a California Air Resources Board (CARB) California-specific survey (CARB, 2019). The composition of products is estimated by generating composites from various CARB surveys from 2007 through 2019, along with profiles reported in the U.S. EPA's SPECIATE database (EPA, 2019). The physiochemical properties of all organic components are generated from the quantitative structure-activity relationship model OPERA (Mansouri et al., 2018) and the characteristic evaporation timescale of each component is estimated using previously published methods (Khare and Gentner, 2018; Weschler and Nazaroff, 2008).

National-level emissions are then allocated to the county-level using several proxies. Most emissions are allocated using population as a spatial surrogate. This includes all cleaners, personal care products, adhesives, architectural coatings, and aerosol coatings. Industrial coatings, allied paint products, printing inks, and dry cleaning emissions are allocated using county-level employment statistics from the U.S. Census Bureau's County Business Patterns (U.S. Census Bureau, 2019) and follow the same mapping scheme used in the U.S. EPA's 2017 National Emissions Inventory (EPA, 2021). Agricultural pesticides are allocated using county-level agricultural pesticide use, as taken from the 2017 NEI.

The version of VCPy used for this platform includes additional product aggregations, variation in the VOC-content of products to reflect state-level area source rules relevant to the solvent sector, and the adoption of an indoor emissions pathway. To compute VCP emissions indoors, each product category is assigned an indoor usage fraction. All coating and industrial products are assigned a 50% indoor emission fraction, all pesticides and automotive aftermarket products are assigned a 0% indoor emission fraction, and all consumer and cleaning products are assigned a 100% indoor emission fraction. The lone exception are daily use personal care products, which are assumed to have a 50% indoor emission fraction. This indoor emission assignment enables the mass transfer coefficient to vary between indoor and outdoor conditions. Typically, the mass transfer coefficent indoors is smaller than the mass transfer coefficient outdoors due to more stagnant atmospheric conditions, and the newest version of the modeling framework reflects these dynamics. Indoor product usage utilizes a mass transfer coefficient of 5 m/hr, and the remaining outdoor portion is assigned a mass transfer coefficient of 30 m/hr (Weschler and Nazaroff, 2008; Khare and Gentner, 2018).

The np_solvents sector also includes emissions from the 2017 NEI not covered by VCPy. This includes some State, Locality, and Tribal emission submissions for other CAPs, such as CO, NOX, and PM2.5. In addition, there are some SCCs not covered by VCPy but included in the np_solvents sector.

Finally, since emissions from solvents occur from both point and nonpoint SCCs, point source subtraction is required to ensure emissions from this sector are not double-counted. Point source subtraction for this sector is performed at the county-level using uncontrolled point source emissions. As such, assumptions related to the control efficiency of the point sources must be made. In most some cases, metadata indicates that the point source emission estimates submitted to the NEI feature 80-90% control efficiencies. Therefore, uncontrolled point source emission calculations are calculated, as necessary, using the submitted point source emissions, engineering judgement, and an assumed control efficiency. If point source subtraction results in negative emissions, emissions will zero out emissions for that source category in that county. The mapping of nonpoint SCCs to point SCCs follows the same crosswalk implemented for the 2020 NEI.

3.2.3.7 Other Nonpoint Sources (nonpt)

The 2019 platform nonpt sector inventory is mostly unchanged from the January 2021 version of the 2017 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 2017 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 five distinct sources of PFC emissions, further distinguished by residential or commercial use. The five sources are: (1) displacement of the vapor within the can; (2) spillage of gasoline while filling the can; (3) spillage of gasoline during transport; (4) emissions due to evaporation (i.e., diurnal emissions); and (5) emissions due to permeation. Note that spillage and vapor displacement associated with using PFCs to refuel nonroad equipment are included in the nonroad inventory.

Volatile chemical product (aka solvent) SCCs were placed into the solvents sector. The EPA incorporated new methods to estimate emissions of VOC and associated HAPs from the solvents sector, for this 2019 modeling platform. The new methods result in improved emissions estimates for the nonpoint (county-wide) solvent emissions. The new emissions method results in improved VOC and HAP estimates for nonpoint categories of coatings, pesticides, adhesives and sealants, oil & gas exploration solvent use, dry cleaning, printing inks, cleaning products, personal care products, and other miscellaneous solvent uses.

3.2.4 Biogenic Sources (beis)

Biogenic emissions were computed based on the 19k version of the 2019 meteorology data used for the air quality modeling and were developed using the Biogenic Emission Inventory System version 3.7 (BEIS3.7) within CMAQ. The BEIS3.7 creates gridded, hourly, model-species emissions from vegetation and soils. It estimates CO, VOC (most notably isoprene, terpene, and sesquiterpene), and NO emissions for the contiguous U.S. and for portions of Mexico and Canada. In the BEIS 3.7 two-layer canopy model, the layer structure varies with light intensity and solar zenith angle (Pouliot and Bash, 2015). Both layers include estimates of sunlit and shaded leaf area based on solar zenith angle and light intensity, direct and diffuse solar radiation, and leaf temperature (Bash et al., 2015). The new algorithm requires additional meteorological variables over previous versions of BEIS. The variables output from the Meteorology-Chemistry Interface Processor (MCIP) that are used to convert WRF outputs to CMAQ inputs are shown in Table 3-5.

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 sfc
RN	nonconvective precipitation per met TSTEP
RSTOMI	inverse of bulk stomatal resistance
SLYTP	soil texture type by USDA category
SOIM1	volumetric soil moisture in top cm
SOIT1	soil temperature in top cm
TEMPG	skin temperature at ground
USTAR	cell averaged friction velocity
RADYNI	inverse of aerodynamic resistance
TEMP2	temperature at 2 m

BEIS3.7 was used in conjunction with Version 5 of the Biogenic Emissions Landuse Database (BELD5). The BELD5 is based on an updated version of the USDA-USFS Forest Inventory and Analysis (FIA) vegetation speciation-based data from 2001 to 2017 from the FIA version 8.0. This same configuration of BEIS3.7 and BELD5 was used to develop the biogenic emissions in the 2017 NEI. Canopy coverage is based on the 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). The FIA includes approximately 250,000 representative plots of species fraction data that are within approximately 75 km of one another in areas identified as forest by the MODIS canopy coverage. For land areas outside the conterminous United States, 500-meter grid spacing land cover data from the MODIS is used. BELD5 also incorporates the following:

- 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/production/files/2019-08/documents/800am_zhang_2_0.pdf)
- 2017 30 meter USDA Cropland Data Layer (CDL) data (http://www.nass.usda.gov/research/Cropland/Release/).

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.

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

Mobile sources are emissions from vehicles that move and include several sectors. Onroad mobile source emissions result from motorized vehicles that are normally operated on public roadways. These include passenger cars, motorcycles, minivans, sport-utility vehicles, light-duty trucks, heavy-duty trucks, and buses. Nonroad mobile source emissions are from vehicles that do not operate on roads such as tractors, construction equipment, lawnmowers, and recreational marine vessels. All nonroad emissions are treated as 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.

3.2.5.1 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 6.5.1 of the 2017 NEI TSD.

For the 2019 modeling platform, VMT were projected from 2017 to 2019 based mostly on Federal Highways administration (FHWA) annual VMT changes at the county level. In a few cases, state Department of Transportation (DOT) data were used instead of FHWA data. Other activity data (i.e., starts, on-network idling, VPOP, and hoteling) are projected by applying a ratio of 2017-based VMT/activity ratios to the 2019 VMT. In addition, a number of states submitted 2017-specific activity data for incorporation into this platform. Finally, a new MOVES run for 2019 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 2019 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, NOx, 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, onroad emissions are generated using the SMOKE-MOVES interface that leverages MOVES generated emission factors (https://www.epa.gov/moves), county and SCC-specific activity data, and hourly meteorological data. SMOKE-MOVES takes into account the temperature sensitivity of the on-road emissions. Specifically, EPA used MOVES inputs for representative counties, VMT, VPOP, starts, 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. Meteorological data were specific to the year 2019.

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 2019-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, and inspection and maintenance (I&M) 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 294 representative counties in the continental U.S.

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, vehicle population (VPOP) is used for many off-network processes, and hoteling hours are used to develop emissions for extended idling of combination long-haul trucks. These calculations are done for every county and grid cell in the continental U.S. for each hour of the year.

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

1) Determine which counties will be used to represent other counties in the MOVES runs.

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

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

- 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; and
- rate-per-hour (RPHO) uses off network idling hours activity data to compute off-network idling emissions for all types of vehicles; and
- rate-per-hour (RPH) uses hoteling hours activity data to compute off-network emissions for idling of long-haul trucks from extended idling and auxiliary power unit process.

The onroad emissions inputs for the 2019 platform are based on the 2017 NEI, described in more detail in Section 6 of the 2017 NEI TSD. These inputs include:

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

Fuel months and other inputs were consistent with those in the 2017 NEI, although age distributions were adjusted to represent the year 2019. A list of states that submitted activity data along with a description of the development of the EPA default activity data sets for VMT, VPOP, and hoteling hours are available in detail in the 2017 NEI TSD and supporting documents. Hoteling hours activity are used to calculate emissions from extended idling and auxiliary power units (APUs) by combination long-haul trucks.

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 8,760 hours in the year 2019; therefore, the maximum number of possible hoteling hours in a particular county is equal to 8,760 * the number of parking spaces in that county. Hoteling hours were capped at that theoretical maximum value for 2019 in all counties, with some exceptions.

Because the truck stop parking space dataset may be incomplete in some areas, and trucks may sometimes idle in areas other than designated spaces, it was assumed that every county has at least 12 parking spaces, even if fewer parking spaces are found in the parking space dataset. Therefore, hoteling hours are never reduced below 105,408 hours for the year in any county. If the unreduced hoteling hours were already below that maximum, the hours were left unchanged; in other words, hoteling activity are never increased as a result of this analysis. Four states requested that no reductions be applied to the hoteling activity based on parking space availability: CO, ME, NJ, and NY. For these states, reductions based on parking space availability were not applied.

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). New Jersey's submittal of hoteling activity specified a 30% APU split, and this was used throughout NJ. For the rest of the country, a 6.9% APU split was used, meaning that during 6.9% of the hoteling hours auxiliary power units are assumed to be running.

The last pieces of activity data needed for SMOKE-MOVES are related to the average speed of vehicles, which affects the selection of MOVES emission factors for on-network emissions. One such dataset is the SPEED inventory read by the SMOKE program Smkinven, which includes a single overall average speed for each county, SCC, and month. The second dataset is the SPDIST dataset read by the SMOKE program Movesmrg, which 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. SMOKE still requires the SPEED dataset exist even when hourly speed data is available, even though only the hourly speed data affects the selection of emission factors. The SPEED and SPDIST datasets are from the 2017 NEI with some of the data carried over from the Coordinating Research Council A-100 study (CRC, 2017).

MOVES3 was run in emission factor mode to create emission factor tables using CB6 speciation for the year 2019, 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 2019. The range of temperatures run along with the average humidities used were specific to the year 2019. The remaining settings for the CDBs are documented in the 2017 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 2019. 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, which uses emission inventory codes (EICs) to characterize the emission processes instead of SCCs. EPA had a 2016v1 platform-based set of emissions for 2023. EPA interpolated between 2017 and 2023 to calculate the 2019 onroad emissions for California. 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. The California inventory had CAPs only and did not have NH3 or refueling. The EPA added NH_3 to the CARB inventory by using the state total NH_3 from MOVES and allocating it at the county level based on CO. Refueling emissions were projected from the 2017 NEI using county total refueling VOC from EQUATES 2017 and the 2019 MOVES3 onroad run for CA. CARB VOCs were speciated to VOC HAPs using MOVES VOC speciation. All other HAPs (e.g., metals and PAHs) are 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 unique rules in California, while leveraging the more detailed SCCs and the highly resolved spatial patterns, temporal patterns, and speciation from SMOKE-MOVES. The basic steps involved in temporally allocating onroad emissions from California based on SMOKE-MOVES results were:

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

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

3.2.5.2 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. Some examples of CMV sources included in cmv_c1c2 are fishing vessels, tug boats, and oil and gas platform support vessels. Table 3-6 shows the number of each type of Category 3 vessel identified as part of the 2020 NEI process. For more information on the CMV sources, see the supplemental documentation for 2020 NEI CMV[8]. C1 and C2 emissions that occur outside of state waters are not assigned to states. All CMV emissions in the cmv_c1c2 sector are treated as hourly gridded point sources with stack parameters that should result in them being placed in layer 1. The C1C2 CMV emissions were computed for 2019 using methods compatible with the 2020 NEI.

[8] https://gaftp.epa.gov/Air/nei/2020/doc/supporting_data/nonpoint/CMV/

Table 3-6. CMV C1C2 Vessels in each Group

Vessel Group	Ship Count
Bulk Carrier	37
Commercial Fishing	1,147
Container Ship	7
Ferry Excursion	441
General Cargo	1,498
Government	1,338
Miscellaneous	1,475
Offshore support	1,149
Reefer	13
Ro Ro	26
Tanker	100
Tug	3,994
Work Boat	77
Total in Inventory:	**11,302**

The cmv_c3 sector contains large engine CMV emissions. 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.[9] 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 the 2019 CMV sources, see the supplemental documentation for the 2020 NEI CMV[10].

The resulting annual point emissions 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 was 539.6 °F, and the velocity was 82.02 ft/s. Hourly emissions were converted to an hourly point 2010 FF10.

The emission factors reflect International Maritime Organization (IMO) Tier 3 NOx regulations that apply to engines installed on ships constructed (i.e., keel is laid) on or after January 1st, 2016.

3.2.5.3 Locomotive (rail)

The rail sector includes all locomotives in the NEI nonpoint data category. This sector excludes railway maintenance locomotives and point source yard locomotives. Railway maintenance emissions are included in the nonroad sector. The point source yard locomotives are included in the ptnonipm sector. Typically, in the NEI, yard locomotive emissions are split between the nonpoint and point categories, but for this study, all yard locomotive emissions are represented as point sources and included in the ptnonipm sector.

[9] https://www.epa.gov/regulations-emissions-vehicles-and-engines/regulations-emissions-marine-vessels.
[10] https://gaftp.epa.gov/Air/nei/2020/doc/supporting_data/nonpoint/CMV.

This study uses 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. More detailed information on the development of the 2017 rail inventory for this study is available in the 2017 NEI TSD. The 2017 inventory was projected to 2019 using activity-based factors. Pollutant-specific factors were applied on top of the activity-based changes for the Class I rail.

3.2.5.4 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, and the Texas Commission on Environmental Quality (TCEQ), which provided their own emissions. California nonroad emissions were provided by the California Air Resources Board (CARB) for the 2017 NEI. The 2019 California nonroad emissions were interpolated from the 2017 NEI and a 2023 projection from the 2016v1 modeling platform, with HAP augmentation. For Texas, the EPA interpolated to 2019 between data provided for 2017 and 2020 and applied HAP augmentation.

The spatial allocation updates to agricultural and construction equipment developed as part of the 2016 platform were carried forward into this platform.

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 and Texas, 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 PM2.5 by speciation profile code for the PM2.5 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 PM2.5 called PM25TOTAL was added to the inventory. As with VOC, PM25_####-PM25TOTAL ratios were calculated and applied to PM2.5 emissions in California and Texas so that PM2.5 emissions in California and Texas 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 that are 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 total CAP emissions are less than $1*10^{-10}$ 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.

3.2.6 Day-Specific Point Source Fires (ptfire)

Wildfire and prescribed burning emissions are contained in the ptfire-rx and ptfire-wild sectors. Both ptfire sectors have 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.

Figure 3-1 shows the processing stream for wildfire and prescribed burn sources. The emissions estimate methodology consists of two tools or systems. The first system is called Satellite Mapping Automated Reanalysis Tool for Fire Incident Reconciliation version 2 (SMARTFIRE2). SMARTFIRE2 is an algorithm and database system that operate within a geographic information system (GIS) framework. 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. 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). In this study, the national and S/L/T fire information is input into SMARTFIRE2 and then all information is merged and associated together based on user-defined weights for each fire information dataset. The output from SMARTFIRE2 is daily acres burned and latitude-longitude coordinates for each fire.

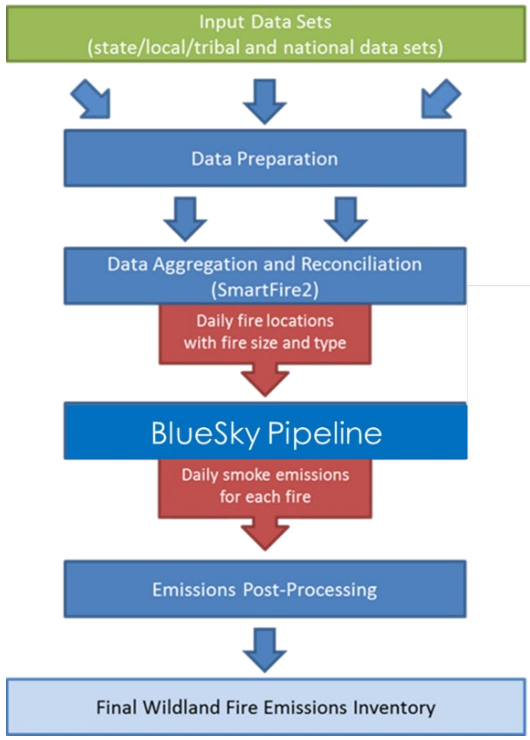


Figure 3-1. Processing flow for fire emission estimates

Inputs to SMARTFIRE2 for 2019 included:

- The National Oceanic and Atmospheric Administration's (NOAA's) Hazard Mapping System (HMS) fire location information
- GeoMAC (Geospatial Multi-Agency Coordination), an online wildfire mapping application designed for fire managers to access maps of current fire locations and perimeters in the United States
- 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)
- Burn scar/fire activity shapefiles for wildfires and some prescribed burns from Monitoring trends in burn severity (MTBS) website (https://www.mtbs.gov/direct-download)
- Prescribed burn activity on federal lands from the Department of Interior (DOI)

- Prescribed burn activity from California Air Resources Board (CARB) specifically from their Prescribed Fire Incident Reporting System (PFIRS)
- Prescribed burn activity from Texas Parks and Wildlife Division (TPWD)
- Active fire perimeters from Bureau of Land Management (BLM)
- Wildfire and prescribed date, location, and locations from a few S/L/T activity submitters (includes Georgia, Florida and Kanas(Flint Hills only))

The second system used to estimate emissions is the BlueSky Modeling Pipeline which 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 and Alaska, where Fuel Characteristic Classification System (FCCS) fuel loading data are available, were processed using the modeling chain described in Figure 3-2. The Fire Emissions Production Simulator (FEPS) (Anderson, 2004) in the BlueSky Pipeline generates all the CAP emission factors for wildland fires used in this 2019 study. HAP emission factors were obtained from Urbanski's (2014) work and applied by region and by fire type.

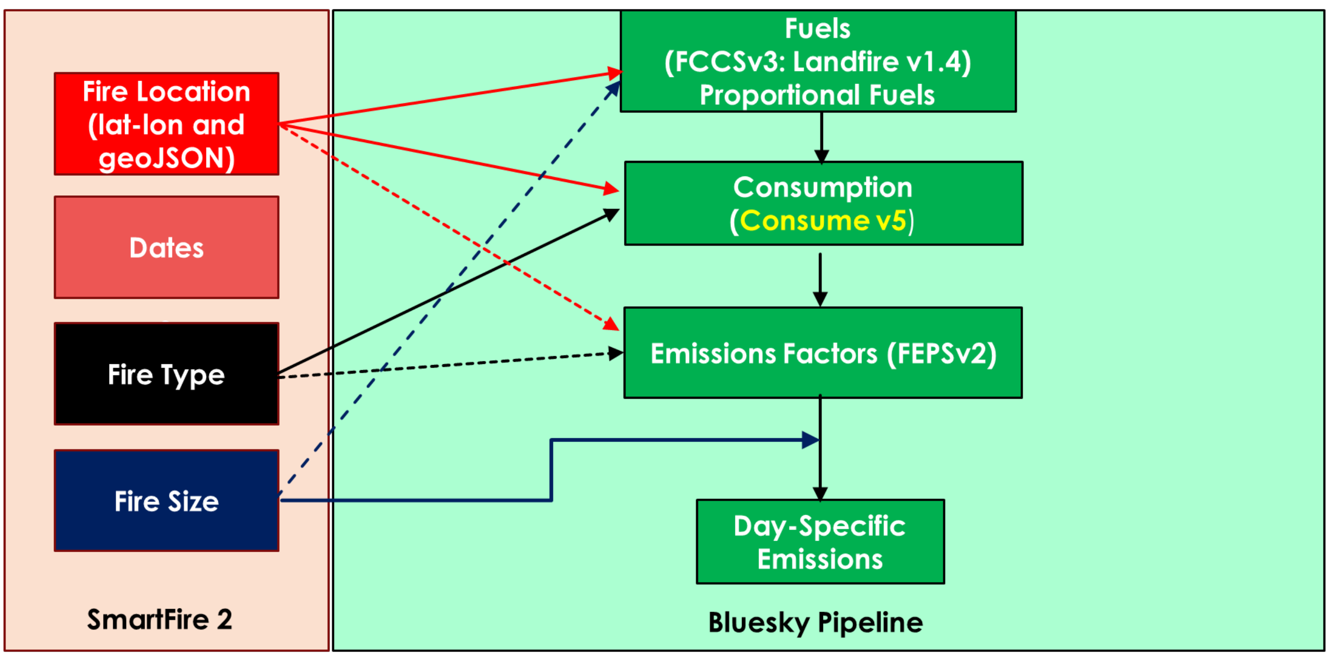


Figure 3-2. BlueSky Pipeline modeling system

The FCCSv3 cross-reference was implemented along with the LANDFIRE (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).

3.2.7 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 2018 in a similar way to the emissions in ptfire. The state of Florida provided their own emissions (separate from the other states) for this study.

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 2019 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 are in the ptfire sector for this 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.

Another feature of the database is that the satellite detections for 2019 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 2018 meteorological simulation (WRF). The location of fire detections was then compared with this daily snow cover file. For any day in which a grid cell has snow cover, that fire detection was excluded. 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 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. Emissions factors were applied to each daily fire to calculate criteria and hazardous pollutant values. These factors vary by crop type.

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 2018 agricultural fire inventories include emissions for HAPs, so HAP integration was used for this study.

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: othpt, othar, othafdust, othptdust, onroad_can, and onroad_mex, canada_ag, and canada_og2D. The "oth" refers to the fact that these emissions are usually "other" than those in the U.S. state-county geographic FIPS, and the remaining characters provide the SMOKE source types: "pt" for point, "ar" for area and nonroad mobile, "afdust" for area fugitive dust (Canada only), and "ptdust" for point fugitive dust (Canada only). The onroad emissions for Canada and Mexico are in the onroad_can and onroad_mex sectors, respectively. Canadian agricultural and low-level (2-D) oil and gas emissions are split into separate sectors from other Canada point sources to reduce the size of the othpt sector.

Emissions in these sectors were taken from the EQUATES 2016 inventories. Environment and Climate Change Canada (ECCC) provided the following inventories for use in EQUATES 2016 and 2017 modeling, which are described in more detail below:

- Agricultural livestock and fertilizer, point source format (canada_ag sector)
- CMV were provided as area sources but converted to point (not currently used)
- Agricultural fugitive dust, point source format (othptdust sector)
- Other area source dust (othafdust sector)
- Onroad (onroad_can sector)
- Nonroad and rail (othar sector)
- Oil and gas surces (low-level in canada_og2D sector, elevated in othpt sector)
- Other area sources (othar sector)
- Airports (othpt sector)
- Other point sources (othpt sector)

Canadian CMV inventories that had been included in this sector in past modeling platforms are no longer needed in the cmv_c1c2 and cmv_c3 sectors.

Temporal profiles, and shapefiles for creating spatial surrogates, were provided by ECCC in a previous Canadian emissions dataset and were reused for this study. Other than the CB6 species of NBAFM present in the speciated point source data, there are no explicit HAP emissions in these Canadian inventories

Canadian point source inventories provided by ECCC for the EQUATES project for 2016 were used as-is for 2019. These inventories include emissions for airports and other point sources. The Canadian point source inventory is pre-speciated for the CB6 chemical mechanism. 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)). The point source emissions were converted to English units and into the FF10 format that could be read by SMOKE, missing stack parameters were gapfilled using SCC-based defaults, and latitude and longitude coordinates were verified and adjusted if they were not consistent with the reported municipality. Only CAPs are covered in the Mexico point source inventory. For this study, Mexico emissions were projected from 2016 to 2019 using projection factors derived from the Community Emissions Data System (CEDS).

Due to the large number of points in the Canada inventories, the agricultural sources were split into a separate sector called canada_ag so that the sources could be placed into layer 1 as plume rise calculations were not needed. Similarly, there were a very large number of Canadian oil and gas point sources, many of which would be appropriate modeled in layer 1. These sources were placed into the canada_og2D sector for layer 1 modeling. Reducing the size of the othpt sector sped up the air quality model run

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 2016 emission inventory. 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 2016 emission inventory. The provided wind erosion emissions were removed. The othptdust emissions have a monthly resolution. 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. The EQUATES 2016 inventory was used as-is with 2018 meteorology applied.

ECCC provided year 2016 Canada province, and in some cases sub-province, resolution emissions from for nonpoint and nonroad sources (othar). The nonroad sources were monthly while the nonpoint and rail emissions were annual. The 2016 Canada nonroad emissions were projected to 2019 using US MOVES-based trends. For Mexico, year 2016 Mexico nonpoint and nonroad inventories at the municipio resolution provided by SEMARNAT were projected to 2019 using projection factors derived from the Community Emissions Data System (CEDS). All Mexico inventories were annual resolution
The onroad emissions for Canada and Mexico are in the onroad_can and onroad_mex sectors, respectively. Emissions for Canada come from the EQUATES 2016 (2016 was the latest year provided by Environment and Climate Change Canada (ECCC)) and were projected from 2016 to 2019 using US MOVES-based trends.

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, 2017). This includes NBAFM plus several other VOC HAPs such as toluene, xylene, ethylbenzene and others. Except for VOC HAPs that are part of the speciation, no other HAPs are included in the Mexico onroad inventory (such as particulate HAPs nor diesel particulate matter). Mexico onroad inventories were generated by MOVES for the years 2017 and 2020, and then interpolated to 2019 for this study.

Annual 2019 wildland fire emissions for Mexico, Canada, Central America, and Caribbean nations are included in the ptfire_othna sector. Canadian fires, along with fires in Mexico, Central America, and the Caribbean, were developed from Fire Inventory from NCAR (FINN) 2017 v1.5 daily fire emissions. For FINN fires, listed vegetation type codes of 1 and 9 are defined as agricultural burning, all other fire detections and assumed to be wildfires. All wildland fires that are not defined as agricultural are assumed to be 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 (Cl_2) concentrations in oceanic air masses (Bullock and Brehme, 2002). Data at 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 last several 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 model requires hourly emissions of specific gas and particle species for the horizontal and vertical grid cells contained within the modeled region (i.e., modeling domain). To provide emissions in the form and format required by the model, it is necessary to "pre-process" the "raw" emissions (i.e., emissions input to SMOKE) for the sectors described above. In brief, the process of emissions modeling transforms the emissions inventories from their original temporal resolution, pollutant resolution, and spatial resolution into the hourly, speciated, gridded resolution required by the air quality model. Emissions modeling includes temporal allocation, spatial allocation, and pollutant speciation. In some cases, emissions modeling also includes the vertical allocation of point sources, but many air quality models also perform this task because it greatly reduces the size of the input emissions files if the vertical layers of the sources are not included.

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

3.3.1 The SMOKE Modeling System

SMOKE version 4.8.1 was used to pre-process the raw emissions inventories into emissions inputs for CMAQ. SMOKE executables and source code are available from the Community Multiscale Analysis System (CMAS) Center at http://www.cmascenter.org. Additional information about SMOKE is available from http://www.smoke-model.org. For sectors that have plume rise, the in-line emissions capability of the air quality models was used, which allows the creation of source-based and two-dimensional gridded emissions files that are much smaller than full three-dimensional gridded emissions files. For quality assurance of the emissions modeling steps, emissions totals by specie for the entire model domain are output as reports that are then compared to reports generated by SMOKE on the input inventories to ensure that mass is not lost or gained during the emissions modeling process.

3.3.2 Key Emissions Modeling Settings

When preparing emissions for the air quality model, emissions for each sector are processed separately through SMOKE, and then the final merge program (Mrggrid) is run to combine the model-ready, sector-specific emissions across sectors. The SMOKE settings in the run scripts and the data in the SMOKE ancillary files control the approaches used by the individual SMOKE programs for each sector. Table 3-7 summarizes the major processing steps of each platform sector. The "Spatial" column shows the spatial approach used: here "point" indicates that SMOKE maps the source from a point location (i.e., latitude and longitude) to a grid cell; "surrogates" indicates that some or all of the sources use spatial surrogates to allocate county emissions to grid cells; and "area-to-point" indicates that some of the sources use the SMOKE area-to-point feature to grid the emissions. The "Speciation" column indicates that all sectors use the SMOKE speciation step, though biogenics speciation is done within the Tmpbeis3 program and not as a separate SMOKE step. The "Inventory resolution" column shows the inventory temporal resolution from which SMOKE needs to calculate hourly emissions. Note that for some sectors (e.g., onroad, beis), there is no input inventory; instead, activity data and emission factors are used in combination with meteorological data to compute hourly emissions.

Table 3-7. Key emissions modeling steps by sector

Platform sector	Spatial	Speciation	Inventory resolution	Plume rise
afdust_adj	Surrogates	Yes	Annual	
airports	Point	Yes	Annual	None
beis	Pre-gridded land use	in BEIS3.7	computed hourly	
fertilizer	Surrogates	No	computed hourly	
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	
np_solvents	Surrogates	Yes	annual	
onroad	Surrogates	Yes	monthly activity, computed hourly	
onroad_ca_adj	Surrogates	Yes	monthly activity, computed hourly	
onroad_can	Surrogates	Yes	monthly	
onroad_mex	Surrogates	Yes	monthly	
othafdust_adj	Surrogates	Yes	annual	
othar	Surrogates	Yes	annual & monthly	
othpt	Point	Yes	annual & monthly	in-line
othptdust_adj	Point	Yes	monthly	None
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
rail	Surrogates	Yes	annual	
rwc	Surrogates	Yes	annual	

The "plume rise" column indicates the sectors for which the "in-line" approach is used. These sectors are the only ones with emissions in aloft layers based on plume rise. The term "in-line" means that the plume rise calculations are done inside of the air quality model instead of being computed by SMOKE. The air quality model computes the plume rise using stack parameters and the hourly emissions in the SMOKE output files for each emissions sector. The height of the plume rise determines the model layer into which the emissions are placed. All of these particular sectors only have "in-line" emissions, meaning that all of the emissions are treated as elevated sources and there are no emissions for those sectors in the two-dimensional, layer-1 files created by SMOKE. 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. The emissions in the airports and othptdust sectors are all low-level emissions, and so in-line emissions files are not created for these two sectors. Instead, all airports and othptdust emissions are output to gridded emissions files, same as if airports and othptdust were area source sectors.

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

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.

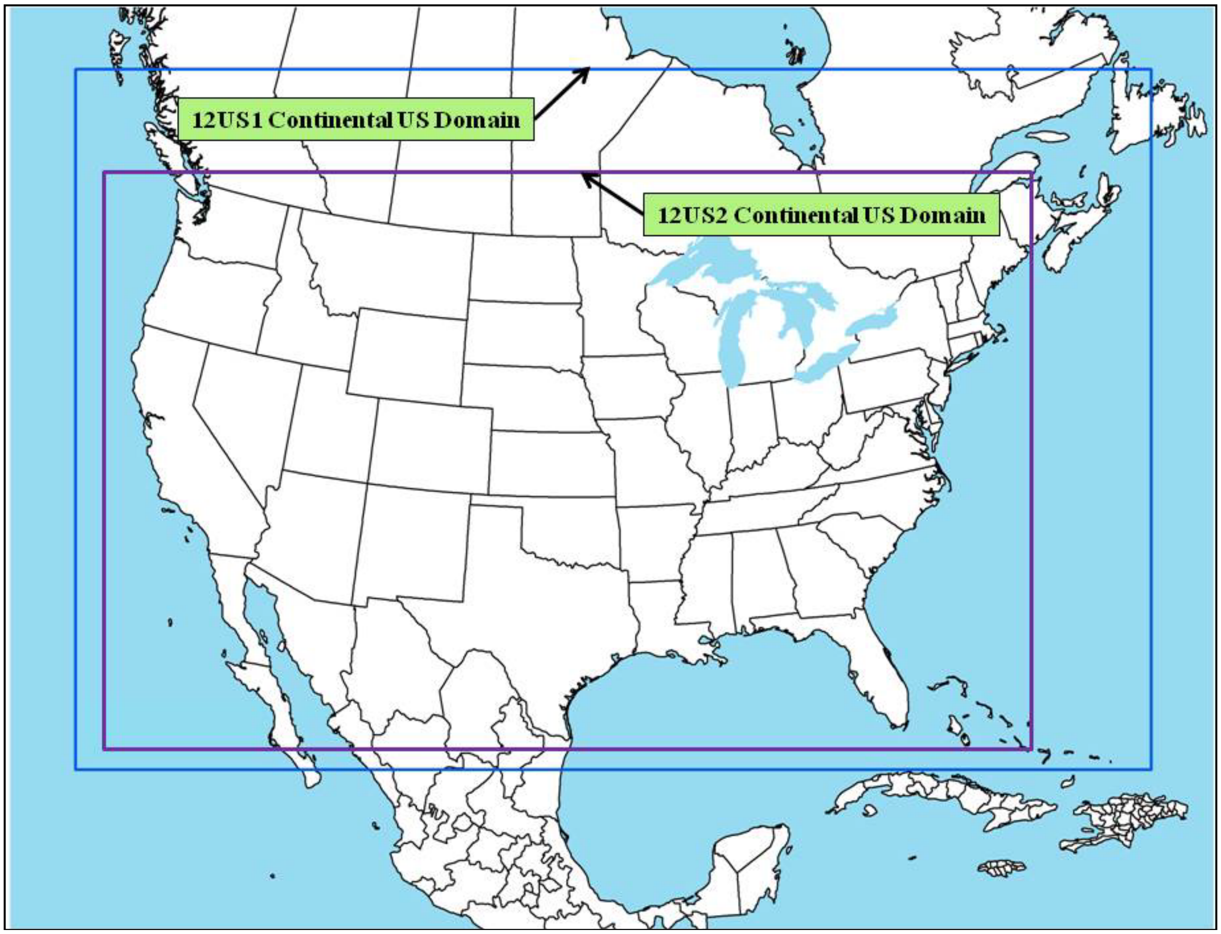


Figure 3-3. CMAQ Modeling Domain

3.3.4 Chemical Speciation Configuration

The emissions modeling step for chemical speciation creates the "model species" needed by the air quality model for a specific chemical mechanism. These model species are either individual chemical compounds (i.e., "explicit species") or groups of species (i.e., "lumped species"). The chemical mechanism used for the platform is the CB6 mechanism (Yarwood, 2010). We used an updated version of CB6 that we refer to as "CB6R3AE7" which includes four new species that were not in the previous version of CB6: AACD, FACD, APIN, and IVOC. This mapping uses a new systematic methodology for mapping low volatility compounds. Compounds with very low vapor pressure are mapped to model species NVOL and intermediate volatility compounds are mapped to a species called IVOC. In previous mappings, some of these low vapor pressure compounds were mapped to CB6 species. The mechanism and mapping are described in more detail in a memorandum describing the mechanism files supplied with the Speciation Tool, the software used to create the CB6 profiles used in SMOKE. It should be noted that the onroad mobile sector does not use this newer mapping because the speciation is done within MOVES and the mapping change was made after MOVES had been run.

This platform generates the $PM_{2.5}$ model species associated with the CMAQ Aerosol Module version 7 (AE7) which has the same $PM_{2.5}$ model species as version 6 (AE6). The AE7 mechanism is built on the AE6 and identical in terms of model species and mechanism definition but requires that alpha-pinene (APIN) be separate from all other monoterpenes (TERP) and not included in TERP to avoid double counting. Table 3-8 lists the model species produced by SMOKE in the platform used for this study.

Table 3-8. Emission model species produced for CB6R3AE7 for CMAQ

Inventory Pollutant	Model Species	Model species description
Cl_2	CL2	Atomic gas-phase chlorine
HCl	HCL	Hydrogen Chloride (hydrochloric acid) gas
CO	CO	Carbon monoxide
NO_X	NO	Nitrogen oxide
NO_X	NO2	Nitrogen dioxide
NO_X	HONO	Nitrous acid
SO_2	SO2	Sulfur dioxide
SO_2	SULF	Sulfuric acid vapor
NH_3	NH3	Ammonia
NH_3	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 (not part of CB05)
VOC	CH4	Methane
VOC	ETH	Ethene
VOC	ETHA	Ethane
VOC	ETHY	Ethyne
VOC	ETOH	Ethanol
VOC	FACD	Formic acid
VOC	FORM	Formaldehyde
VOC	IOLE	Internal olefin carbon bond (R-C=C-R)
VOC	ISOP	Isoprene
VOC	IVOC	Intermediate volatility organic compounds
VOC	KET	Ketone Groups
VOC	MEOH	Methanol
VOC	NAPH	Naphthalene
VOC	NVOL	Non-volatile compounds
VOC	OLE	Terminal olefin carbon bond (R-C=C)
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)
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

Inventory Pollutant	Model Species	Model species description
Methanol	MEOH	Methanol from inventory
PM_{10}	PMC	Coarse PM > 2.5 microns and ≤ 10 microns
$PM_{2.5}$	PEC	Particulate elemental carbon ≤ 2.5 microns
$PM_{2.5}$	PNO3	Particulate nitrate ≤ 2.5 microns
$PM_{2.5}$	POC	Particulate organic carbon (carbon only) ≤ 2.5 microns
$PM_{2.5}$	PSO4	Particulate Sulfate ≤ 2.5 microns
$PM_{2.5}$	PAL	Aluminum
$PM_{2.5}$	PCA	Calcium
$PM_{2.5}$	PCL	Chloride
$PM_{2.5}$	PFE	Iron
$PM_{2.5}$	PK	Potassium
$PM_{2.5}$	PH2O	Water
$PM_{2.5}$	PMG	Magnesium
$PM_{2.5}$	PMN	Manganese
$PM_{2.5}$	PMOTHR	$PM_{2.5}$ not in other AE6 species
$PM_{2.5}$	PNA	Sodium
$PM_{2.5}$	PNA	Sodium
$PM_{2.5}$	PNCOM	Non-carbon organic matter
$PM_{2.5}$	PNH4	Ammonium
$PM_{2.5}$	PSI	Silica
$PM_{2.5}$	PTI	Titanium

The TOG and $PM_{2.5}$ speciation factors that are the basis of the chemical speciation approach were developed from a draft version of the SPECIATE 5.2 database (https://www.epa.gov/air-emissions-modeling/speciate), which is the EPA's repository of TOG and PM speciation profiles of air pollution sources. The SPECIATE database development and maintenance is a collaboration involving the EPA's Office of Research and Development (ORD), Office of Transportation and Air Quality (OTAQ), and the Office of Air Quality Planning and Standards (OAQPS). The SPECIATE database contains speciation profiles for TOG, speciated into individual chemical compounds, VOC-to-TOG conversion factors associated with the TOG profiles, and speciation profiles for $PM_{2.5}$.

Some key features of the speciation approach for this study are listed here and described further in the subsections below:

- Use of the CBR3AE7 mechanism, as described earlier
- Non-methane organic gases (NMOG), which are total organic gases with methane subtracted from it, is included as a pollutant in the emissions output files to assist with the use of these data with future versions of the CMAQ model.
- Several new VOC and $PM_{2.5}$ profiles slated for the final version of SPECIATE 5.2 were used.
- $PM_{2.5}$ speciation process for nonroad mobile use profiles assigned within MOVES3 (which outputs the emissions with those assignments).

- As with previous platforms, some Canadian point source inventories are provided from Environment Canada as pre-speciated emissions, and not all CB6 species were provided; missing species were supplemented by speciating total VOC.

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

For onroad mobile sources, speciation is done in MOVES, to allow for profiles that vary by model year, which is not part of the SCC code, to be used. Therefore, cross-references or emissions summaries by profile for onroad mobile sources are not provided. These profiles are documented in a MOVES technical report on speciation (EPA, 2020).

A number of speciation profiles for VOC and $PM_{2.5}$ that had been added in SPECIATE 5.1 (EPA, 2020) and 5.2 were used. In addition, we profile assignments were updated to incorporate data provided by states or to correct errors in previous assignments.

For PM2.5 the following profile updates were made for the 2017 platform:

- Corrected the wildfire and prescribed fire profile due to error in compositing (the previous profile included creosote in the average)
- Updated the profile for aircraft
- Corrected several profile assignments for the petroleum industry

For PM2.5 the following profile and cross-reference updates were made for the 2018 platform:

- Corrected the speciation profile assignment for several SCCs which should have been mapped to the Heat Treating speciation profile for PM2.5 according to comments in the cross-reference file.
- Updated the profile for sugar cane burning in the ptagfire sector.
- Updated the wildfire and prescribed fire profiles.
- Updated SCC 30400740 to use the Natural Gas combustion profile (95475).

For PM2.5, the following speciation profile and cross-reference updates were made for the 2019 platform:

- Updated the speciation profile assignments for two pulp and paper SCCs, changing from the overall default profile to the wood products drying profile (91128).
- Changed SCC 31000208 from the surface coating profile (91129) to the petroleum industry average profile (91145).
- Assignments for new PM2.5 SCCs in the 2019 point inventory were included.

For VOC the following profile updates were made for the 2017 platform:

- Volatile consumer products - recent methods to estimate emissions of Volatile Organic Compounds (VOC) and associated Hazardous Air Pollutants (HAPs) from Volatile Chemical Products (VCPs) aka solvents were used in this modeling platform. These methods result in improved emissions estimates for the nonpoint (county-wide) solvent emissions. This emissions method results in improved VOC and HAP estimates for nonpoint categories of coatings, pesticides, adhesives and sealants, oil & gas exploration solvent use, dry cleaning, printing inks, cleaning products, personal care products, and other miscellaneous solvent uses. See section 3.2.1 for more details.

- Oil and Gas – used additional region-specific profiles or updated assignments
  - Used county-specific profiles gas for several Wyoming counties developed from data provided by the Wyoming DEQ
  - Used Willison Basin gas composition data, separate profiles for the Montana and North Dakota portions of the basin, based on data developed by the Western Regional Air Partnership WRAP
  - Used Central Montana Uplift area gas composition data, based on data developed by the WRAP
  - Updated Uinta basin profile assignments (based on data provided by Utah)
  - Used Utah and Wyoming oil and gas produced water pond profiles
  - Updated profile assignments (by county and SCC) for nonpoint oil and gas sources that account for the portion of VOC estimated to come from flares. These were updated using results from the Oil and Gas estimation tool run that was used for the 2017 NEI
  - Updated profile assignment for miscellaneous engines to use internal combustion engine natural gas profile
- Commercial Marine vessel – changed profile assignment to an existing Pre-Tier 1 nonroad diesel profile because the previous profile was missing key species (aldehydes)
- Livestock – updated profile assignments
- Agricultural burning – updated profiles for rice straw and wheat straw burning, and used new sugar cane burning profile

For VOC the following cross reference updates were made for the 2018 platform:

- Changed all 8746 to G8746 (Profile name: Rice Straw and Wheat Straw Burning Composite of G4420 and G4421)
- Changed 2104008230/330 from 1084 to 4642 to match all other RWC SCCs
- For solvents, updated all speciation profiles for SCCs in the VCPy inventory
- Changed 2680001000 from 0000 to G95241TOG
- Uinta Basin oil/gas profiles:
  - Replaced profile 95417 with either UTUBOGC (2310010300, 2310011500, 2310111401, 2310010700, 2310010400, 31000107) or UTUBOGD (other SCCs)
  - Replaced profile 95418 with UTUBOGF
  - Replaced profile 95419 with UTUBOGE
- PA gas profiles: Replaced all 8949 with PAGAS01 (FIPS 42059 only), PAGAS02 (FIPS 42019 only), PAGAS03 (FIPS 42125 only). To do this, we first replaced existing county-specific 8949 profiles with the new PAGAS profiles for these three counties in the ERG COMBO GSREF. This covered 5 SCCs. For all SCCs other than those 5 SCCs, where the national profile assignment is 8949. New county-specific profile assignments were made to the appropriate PAGAS profile for each of the three counties, added to the Ramboll basin specific GSREF (since that GSREF is also county-specific and not combo).
- Colorado 2310030300: Set Archuleta/La Plata to SUIROGWT (counties are in Southern Ute reservation), rest of Colorado to DJTFLR95
- Colorado 2310030220: Set to DJTFLR95 (formerly FLR99)
- Colorado 2310021010: Set Archuleta/La Plata to SUIROGCT (counties are in Southern Ute reservation), rest of Colorado to 95398
- Changed 2310000551 (CBM produced water) to a new profile, CBMPWWY. Speciation Tool inputs for this profile tool run by GDIT. Documentation: Profiles are means from WY tests in SPECIATE, newly composited. Reference: https://doi.org/10.1016/j.scitotenv.2017.11.161 Reference: Lyman, Seth N.; Mansfield, Marc L.; Tran, Huy N. Q.; Evans, Jordan D.; Jones, Colleen; O'Neil, Trevor; Bowers, Ric; Smith, Ann; and Keslar, Cara, "Emissions of organic compounds from produced water ponds I: Characteristics and speciation" (2018). Chemistry and Biochemistry Faculty Presentations. Paper 154. Contact: Art Diem and Jeff Vukovich of the EPA's Office of Air Quality Planning and Standards (OAQPS)
- Assignments for new VOC SCCs in the 2018 point inventory were included along with changes to VOC profiles for 16 point SCCs.

For VOC, the following speciation profile and cross-reference updates were made for the 2019 platform:

- Speciation profiles were regenerated using version 5.2 of the SPECIATE database, and with the latest version of the Speciation Tool which includes greater number precision. SPECIATE 5.2 includes several new speciation profiles for solvents, and the cross-reference was updated to use those profiles.
- The definition of the model species SOAALK was changed. Compared to the 2017 and 2018 platforms, the SOAALK emissions are now generally lower.
- Updated the speciation profile assignments for pulp and paper. Two pulp and paper SCCs were updated from the overall default profile to the pulp and paper industry composite profile (95326), and four other SCCs were updated from profile 95326 to the pulp and paper playwood veneer dryer profile (1189).
- For oil and gas, the portion of emissions for SCC 2310010200 which was speciated using profile 2487 was changed to profile 95247.
- All emissions which were previously speciated with profile 1011 were changed to profile 95404. This affects SCCs associated with oil production fugitive leaks and venting.
- All emissions which were previously speciated with profile 1207 were changed to profile 95782. This affects produced water from oil and gas production.
- Assignments for new VOC SCCs in the 2019 point inventory were included.

The speciation of VOC includes HAP emissions from the emissions inventories in the speciation process. Instead of speciating VOC to generate all of the species listed in Table 3-7, emissions of five specific HAPs: naphthalene, benzene, acetaldehyde, formaldehyde, and methanol (collectively known as "NBAFM") from the NEI were "integrated" with the NEI VOC. The integration combines these HAPs with the VOC in a way that does not double count emissions and uses the HAP inventory directly in the speciation process. The basic process is to subtract the specified HAPs emissions mass from the VOC emissions mass, and to then use a special "integrated" profile to speciate the remainder of VOC to the model species excluding the specific HAPs. The EPA believes that the HAP emissions in the NEI are often more representative of emissions than HAP emissions generated via VOC speciation, although this varies by sector.

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

The integration of HAP VOC with VOC is a feature available in SMOKE for all inventory formats, including PTDAY (the format used for the ptfire and ptagfire sectors). The ability to use integration with the PTDAY format was made available in the version of SMOKE used for the v7.1 platform, and this new feature is used for this particular study because the ptfire and ptagfire inventories for this study include HAPs. SMOKE allows the user to specify both the particular HAPs to integrate via the INVTABLE. This is done by setting the "VOC or TOG component" field to "V" for all HAP pollutants chosen for integration. SMOKE allows the user to also choose the particular sources to integrate via the NHAPEXCLUDE file (which actually provides the sources to be *excluded* from integration[11]). For the "integrated" sources, SMOKE subtracts the "integrated" HAPs from the VOC (at the source level) to compute emissions for the new pollutant "NONHAPVOC." The user provides NONHAPVOC-to-NONHAPTOG factors and NONHAPTOG speciation profiles.[12] SMOKE computes NONHAPTOG and then applies the speciation profiles to allocate the NONHAPTOG to the other air quality model VOC species not including the integrated HAPs. After determining if a sector is to be integrated, if all sources have the appropriate HAP emissions, then the sector is considered fully integrated and does not need a NHAPEXCLUDE file. If, on the other hand, certain sources do not have the necessary HAPs, then an NHAPEXCLUDE file must be provided based on the evaluation of each source's pollutant mix. The EPA considered CAP-HAP integration for all sectors in determining whether sectors would have full, no, or partial integration (see Figure 3-4). For sectors with partial integration, all sources are integrated other than those that have either the sum of NBAFM > VOC or the sum of NBAFM = 0.

Figure 3-4 illustrates the integrate and no-integrate processes for U.S. Sources. Since Canada and Mexico inventories do not contain HAPs, we use the approach of generating the HAPs via speciation, except for Mexico onroad mobile sources where emissions for integrate HAPs were available.

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

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

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

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

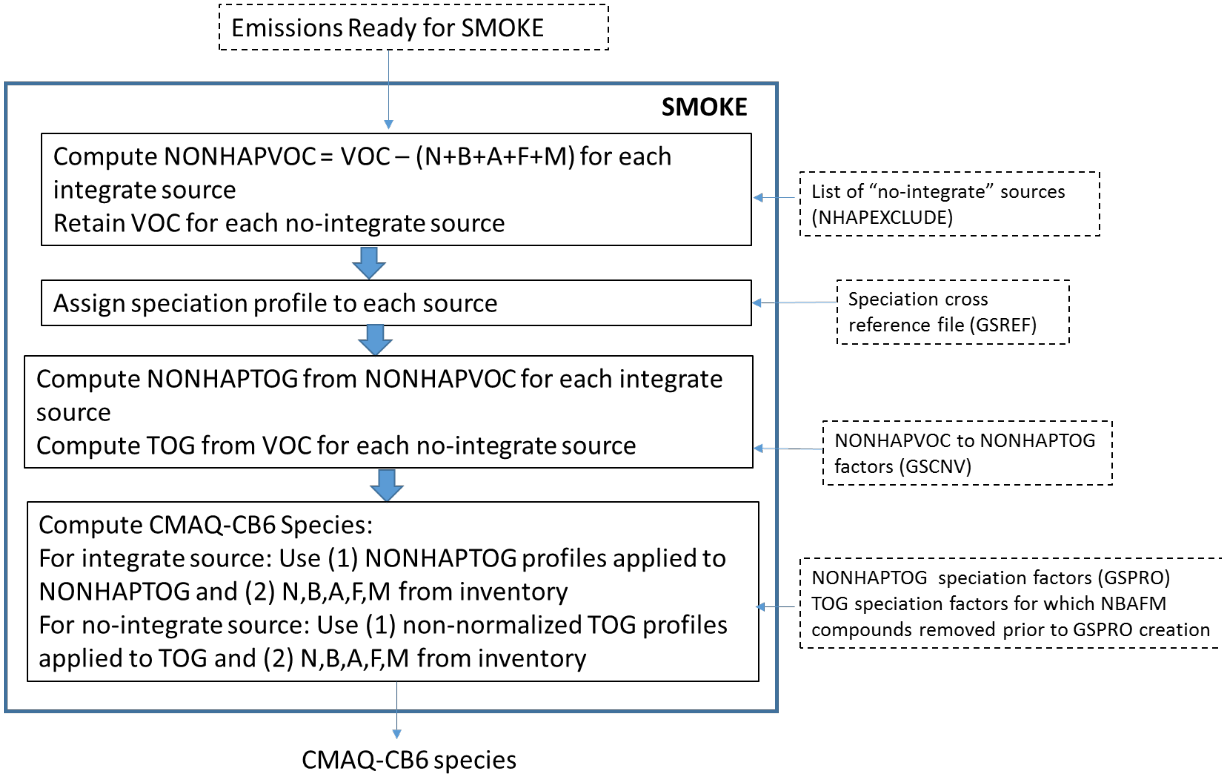


Figure 3-4. Process of integrating NBAFM with VOC for use in VOC Speciation

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

Platform Sector	Approach for Integrating NEI emissions of Naphthalene (N), Benzene (B), Acetaldehyde (A), Formaldehyde (F) and Methanol (M)
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	Partial integration (NBAFM)
nonpt	Partial integration (NBAFM)
nonroad	Full integration (internal to MOVES)
np_oilgas	Partial integration (NBAFM)
np_solvents	Partial integration (NBAFM)
onroad	Full integration (internal to MOVES)
onroad_can	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
onroad_mex	Full integration (internal to MOVES-Mexico); however, MOVES-MEXICO speciation was older CB6, so post-SMOKE emissions were converted to CB6R3AE6
othafdust	N/A – sector contains no VOC
othar	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
othpt	No integration, no NBAFM in inventory, create NBAFM from VOC speciation
othptdust	N/A – sector contains no VOC

Platform Sector	Approach for Integrating NEI emissions of Naphthalene (N), Benzene (B), Acetaldehyde (A), Formaldehyde (F) and Methanol (M)
pt_oilgas	No integration, use NBAFM in inventory
ptagfire	Full integration (NBAFM)
ptegu	No integration, use NBAFM in inventory
ptfire-rx	Partial 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)
rwc	Full integration (NBAFM)

Integration for the mobile sources estimated from MOVES (onroad and nonroad sectors, other than for California) is done differently. Briefly there are three major differences: 1) for these sources integration is done using more than just NBAFM, 2) all sources from the MOVES model are integrated, and 3) integration is done fully or partially within MOVES. For onroad mobile, VOC speciation is done fully within MOVES such that the MOVES model outputs emission factors for individual VOC model species along with the HAPs. This requires MOVES to be run for a specific chemical mechanism. MOVES was run for the CB6R3AE7 mechanism, so no additional species needed to be added after SMOKE-MOVES processing. For nonroad mobile, speciation is partially done within MOVES such that MOVES does not need to be run for a specific chemical mechanism. For nonroad, MOVES outputs emissions of HAPs and NONHAPTOG were split by speciation profile. Taking into account that integrated species were subtracted out by MOVES already, the appropriate speciation profiles are then applied within SMOKE to get the VOC model species. HAP integration for nonroad uses the same additional HAPs and ethanol as for onroad.

In previous platforms, the GSPRO_COMBO feature was used to speciate nonroad mobile and gasoline-related stationary sources that use fuels with varying ethanol content. In these cases, the speciation profiles require different combinations of gasoline profiles (e.g., E0 and E10 profiles). Since the ethanol content varies spatially (e.g., by state or county), temporally (e.g., by month), and by modeling year (future years have more ethanol), the GSPRO_COMBO feature allows combinations to be specified at various levels for different years. For the 2017 platform, GSPRO_COMBO is still used for certain gasoline-related stationary sources nationwide. GSPRO_COMBO is no longer needed for nonroad sources because nonroad emissions within MOVES have the speciation profiles built into the results, so there is no need to assign them via the GSREF or GSPRO_COMBO feature.

In Canada, ECCC provided estimates of ethanol mixes by Canadian province. These estimates were used to develop a GSPRO_COMBO for Canadian gasoline onroad emissions. For example, a province where the average ethanol mix is 6% would have 60% E10 speciation and 40% E0 speciation. A 10% ethanol mix would imply 100% E10 speciation. In Mexico, only E0 speciation profiles are used, but the GSPRO_COMBO feature is still used in Mexico for inventories where VOC emissions are not explicitly defined by mode (e.g., exhaust versus evaporative). Here, the GSPRO_COMBO specifies a mix of exhaust and evaporative speciation profiles. Using the GSPRO_COMBO to split total VOC into exhaust and evaporative components is no longer necessary for Canadian mobile sources, whose inventories include the mode in the pollutant, or for Mexico onroad sources, where VOC speciation is calculated by the MOVES model. The GSPRO_COMBO is still used for Mexican nonroad sources which do not have modes in the inventory.

The newer combo method for combining profiles that is available in SMOKE4.5 and later versions is used in this platform and allows multiple profiles to be combined by pollutant, state and county (i.e., state/county FIPS code), and SCC. This is used specifically for the np_oilgas sector because SCCs include both controlled and uncontrolled oil and gas operations which use different profiles. The underlying data which defines the profile splits for oil and gas was updated for this study.

Speciation profiles for use with BEIS are not included in SPECIATE. BEIS includes a species Sequiterpenes (SESQ) that was mapped to the CMAQ specie SESQT. The profile code associated with BEIS profiles for use with CB6 was "BC6E7."

NO_X can be speciated into NO, NO2, and/or HONO. For the non-mobile sources, EPA used a single profile "NHONO" to split NO_X into NO and NO_2. For the mobile sources except for onroad (including nonroad, cmv_c1c2, cmv_c3, rail, onroad_can, onroad_mex sectors) and for specific SCCs in othar and ptnonipm, the profile "HONO" splits NO_X into NO, NO_2, and HONO. Table 3-10 gives the split factor for these two profiles. The onroad sector does not use the "HONO" profile to speciate NO_X. MOVES produces speciated NO, NO_2, and HONO by source, including emission factors for these species in the emission factor tables used by SMOKE-MOVES. Within MOVES, the HONO fraction is a constant 0.008 of NO_X. The NO fraction varies by heavy duty versus light duty, fuel type, and model year, and equals 1 – NO – HONO. For more details on the NO_X fractions within MOVES for onroad, see Exhaust Emission Rates for Heavy-Duty Onroad Vehicles in MOVES3 and Exhaust Emission Rates for Light-Duty Onroad Vehicles in MOVES3.

Table 3-10. NO_X speciation profiles

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

3.3.5 Temporal Processing Configuration

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

The temporal factors applied to the inventory are selected using some combination of country, state, county, SCC, and pollutant. Table 3-11 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 3-11. 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
nonpt	Annual	Yes	week	week	Yes
nonroad	Monthly		mwdss	mwdss	Yes
np_oilgas	Annual	Yes	aveday	aveday	No
onroad	Annual & monthly[1]		all	all	Yes
onroad_ca_adj	Annual & monthly[1]		all	all	Yes
othafdust_adj	Annual	Yes	week	all	No
othar	Annual & monthly	Yes	week	week	No
onroad_can	Monthly		week	week	No
onroad_mex	Monthly		week	week	No
othpt	Annual & monthly	Yes	mwdss	mwdss	No
othptdust_adj	Monthly		week	all	No
pt_oilgas	Annual	Yes	mwdss	mwdss	Yes
ptegu	Annual & hourly	Yes[2]	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
rwc	Annual	No[3]	met-based[3]	all	No[3]
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 above 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 temporalization are described in the following subsections.

In addition to the resolution, temporal processing includes a ramp-up period for several days prior to January 1, 2019, which is intended to mitigate the effects of initial condition concentrations. The ramp-up period was 10 days (December 22-31, 2018). For anthropogenic sectors, emissions from December 2019 were used to fill in surrogate emissions for the end of December 2018. For biogenic emissions, December 2018 emissions were computed using year 2018 meteorology.

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

SMOKE prevents the application of temporal profiles on top of the “native” resolution of the inventory. For example, a monthly inventory should not have annual-to-month temporalization applied to it; rather, it should only have month-to-day and diurnal temporalization. This becomes particularly important when specific sectors have a mix of annual, monthly, daily, and/or hourly inventories. The flags that control temporalization for a mixed set of inventories are discussed in the SMOKE documentation. The modeling platform sectors that make use of monthly values in the FF10 files are nonroad, onroad (for activity data), onroad_can, onroad_mex, othar, othpt, and othptdust. 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, ptagfire, and ptfire_othna sectors, the inventories are in the daily point fire format, so temporal profiles are only used to go from day-specific to hourly emissions. For all agricultural burning, the diurnal temporal profile used reflected the fact that burning occurs during the daylight. This puts most of the emissions during the 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.

An update in this 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.

A monthly temporal profile for freight rail was developed from AAR data for the year 2016 (https://www.aar.org/data-center/rail-traffic-data/) and continues to be used in this study, Monthly Rail Traffic Data, Total Carloads & Intermodal. Passenger trains use a flat monthly profile. Monthly passenger miles data are available; however, it is not known if there is a correlation between passenger miles and actual rail emissions. This is because passenger trains often operate on a fixed schedule, independent of actual passenger traffic. So, it was decided to not apply a monthly profile to passenger train emissions. All sources in the rail sector use a flat profile for both day-of-week and hour-of-day temporalization.

For the ptfire and ptagfire sectors, the inventories are in the daily point fire format, so temporal profiles are only used to go from day-specific to hourly emissions. For ptfire, state-specific hourly profiles were used, with distinct profiles for prescribed fires and wildfires. For ptagfire, the diurnal temporal profile used reflected the fact that burning occurs during the daylight hours. Additional details on these profiles are available in the 2016v1 platform TSD (EPA, 2020b).

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.

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

Monthly temporalization of np_oilgas emissions is based primarily on 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 (OG17_0001 through OG17_6272), 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 in those four states were assigned flat monthly profiles where there was not already a profile assignment in the ERG dataset.

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

3.3.5.2 Temporal Profiles for EGUs

Electric generating unit (EGU) sources matched to ORIS units were temporally allocated to hourly emissions needed for modeling using the hourly CEMS data. Those hourly data were processed through v2.1 of the CEMCorrect tool to mitigate the impact of unmeasured values in the data. An example of before and after the application of the tool is shown in Figure 3-5.

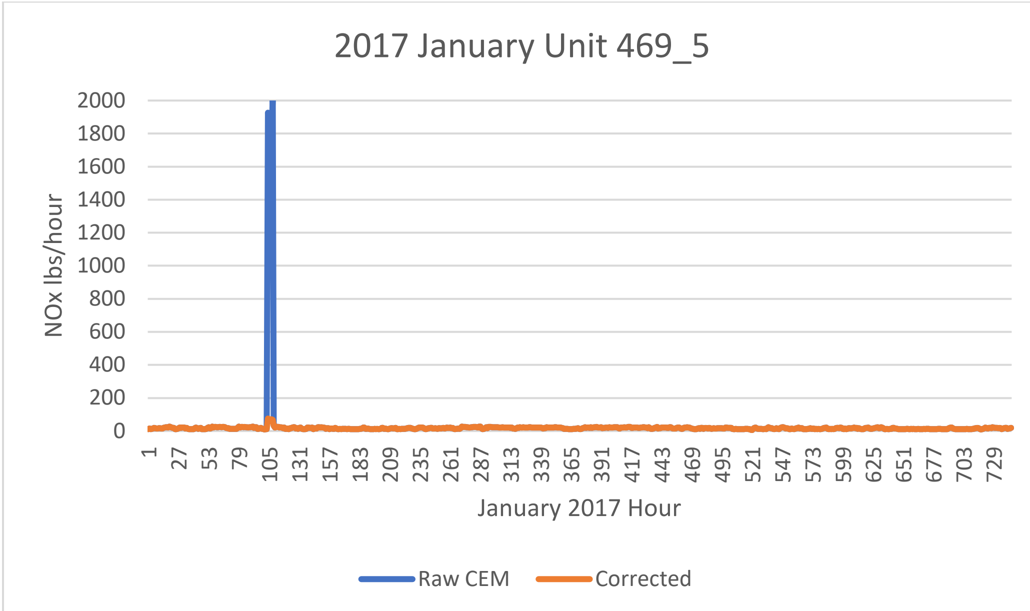


Figure 3-5. Eliminating unmeasured spikes in CEMS data

The region, fuel, and type (peaking or non-peaking) must be identified for each input EGU with CEMS data that are used for generating profiles. The identification of peaking units was done using hourly heat input data from the 2018 base year and the two previous years (2016 and 2017). 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 (2017, 2018, and 2019) and a 3-year average capacity factor of less than 0.1.

Equation 1. Annual unit power output

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

Equation 2. Unit capacity factor

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

Input regions were determined from one of the eight EGU modeling regions based on multi-jurisdictional planning organization (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 as shown in Figure 3-6. 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 is a possible region, fuel, and type group maximum of 64 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 2019 CEM 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 is 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 a greater 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 CEM data in that region. Figure 3-7 shows peaking and non-peaking daily temporal profiles for the gas fuel type in the LADCO region. Figure 3-8 shows the diurnal profiles for the coal fuel type in the Mid-Atlantic/Northeast Visibility Union (MANE VU) region.

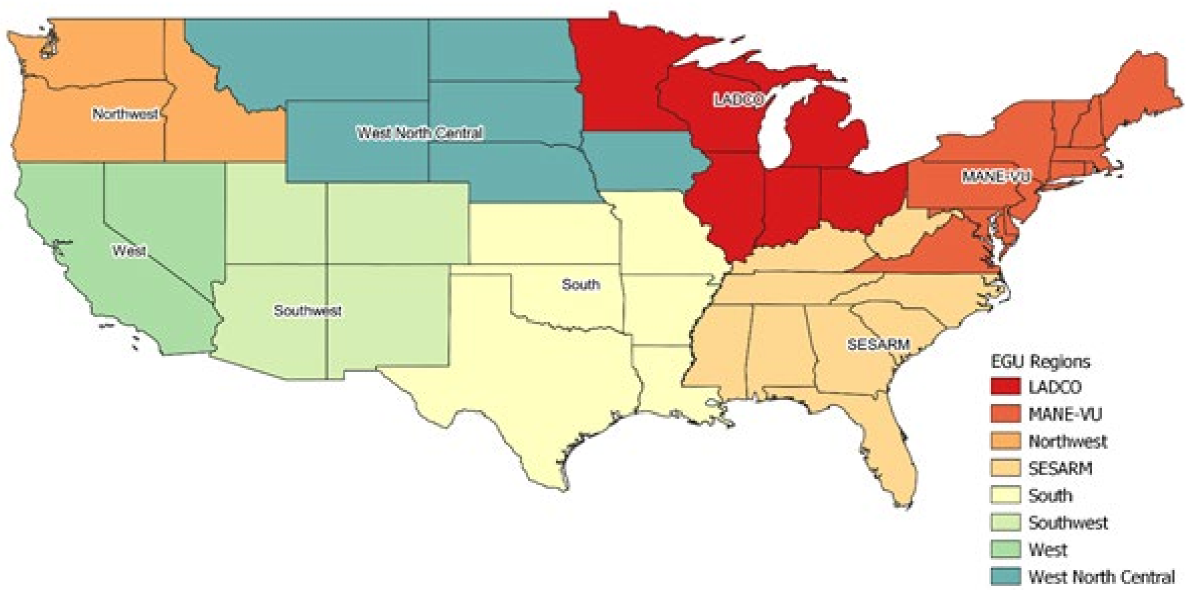


Figure 3-6. Small EGU Temporal Profile Regions

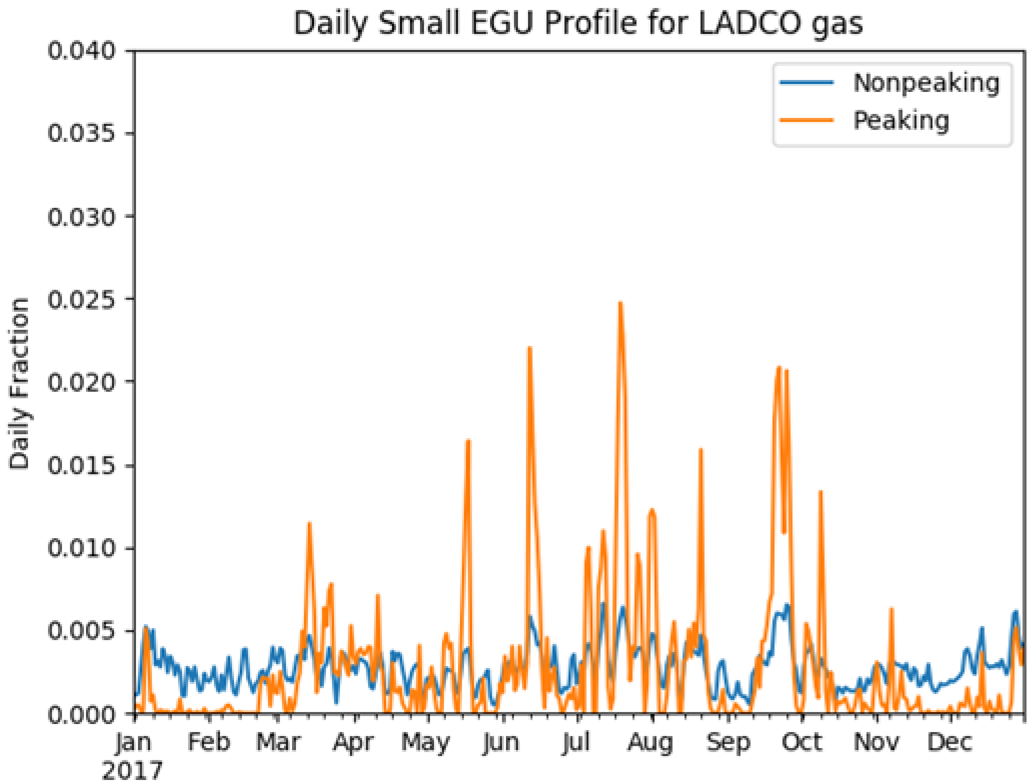


Figure 3-7. Example Daily Temporal Profiles for the LADCO region and Gas Fuel Type

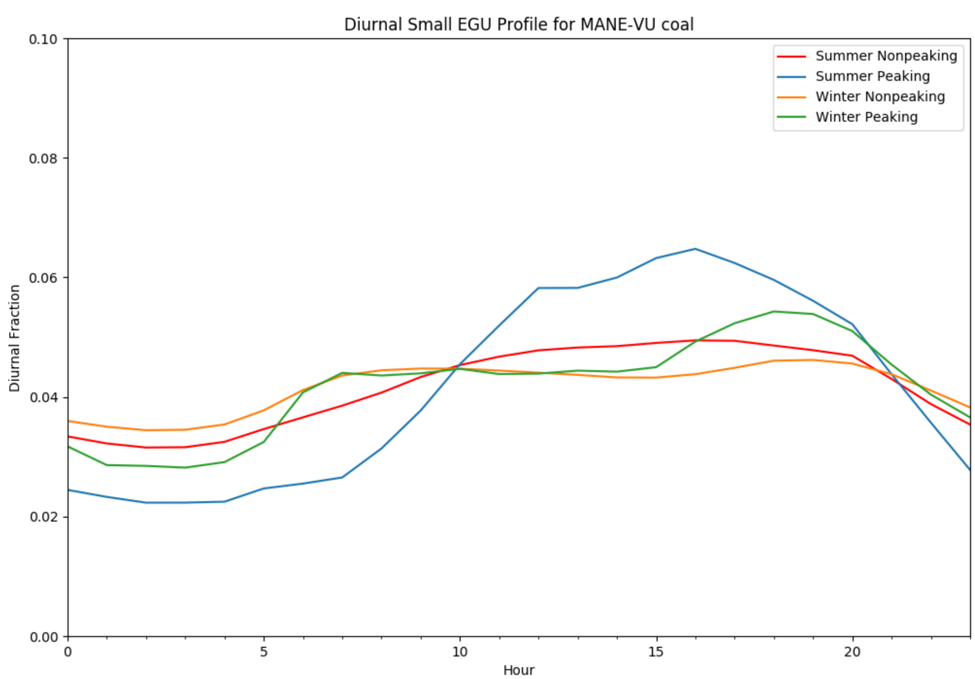


Figure 3-8. Example Diurnal Profile for MANE-VU Region and Coal Fuel Type

SMOKE uses a cross-reference file to select a monthly, daily, and diurnal profile for each source. For this platform, the temporal profiles were assigned in the cross-reference at the unit level to EGU sources without hourly CEM 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. As with the input unit the regions are assigned using the state from the unit FIPS. The fuel is 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, NO_X, $PM_{2.5}$, and SO_2 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 221112. 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. Municipal waste combustors (MWC) and cogeneration units were identified using the NEEDS primary fuel type and cogeneration flag, respectively, from the NEEDS v6 database.

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 temporalization are: (1) a meteorological dataset consistent with that used by the AQ model is available (e.g., outputs from WRF); (2) the meteorological model data are highly resolved in terms of spatial resolution; and (3) the meteorological variables vary at hourly resolution and can therefore be translated into hour-specific temporalization.

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

GenTPRO reads in gridded meteorological data (output from MCIP) along with spatial surrogates and uses the specified algorithm to produce a new temporal profile that can be input into SMOKE. The meteorological variables and the resolution of the generated temporal profile (hourly, daily, etc.) depend on the selected algorithm and the run parameters. For more details on the development of these algorithms and running GenTPRO, see the GenTPRO documentation and the SMOKE documentation at http://www.cmascenter.org/smoke/documentation/3.1/GenTPRO_TechnicalSummary_Aug2012_Final.pdf and https://www.cmascenter.org/smoke/documentation/4.6/html/ch05s03s05.html, respectively.

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

Figure 3-9 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.

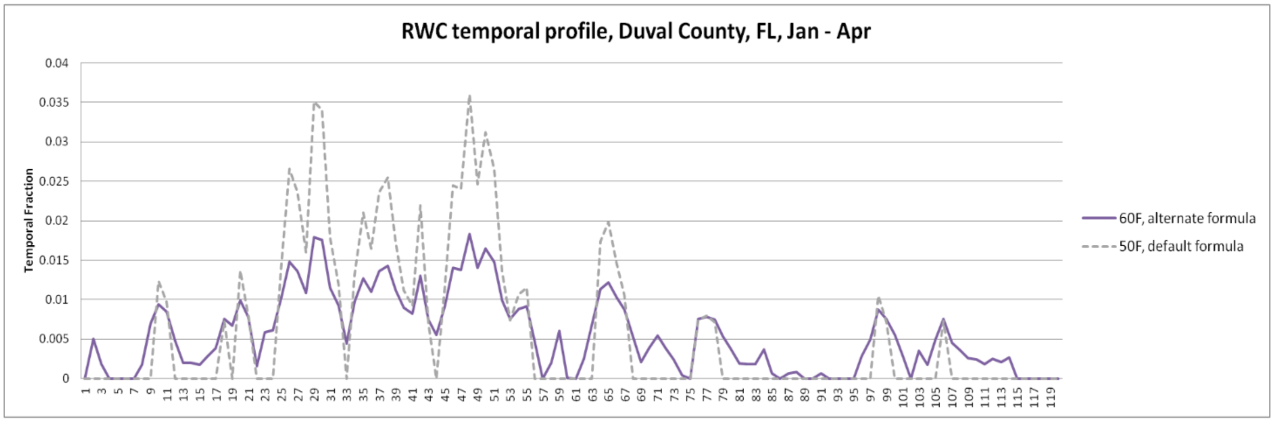


Figure 3-9. Example of RWC temporalization using a 50 ˚F versus 60˚F threshold

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). This profile was created by averaging three indoor and three RWC outdoor temporal profiles from counties in Delaware and aggregating them into a single RWC diurnal profile. This new profile was compared to a concentration-based analysis of aethalometer measurements in Rochester, NY (Wang *et al.* 2011) for various seasons and day of the week and found that the new RWC profile generally tracked the concentration based temporal patterns.

The temporalization for "Outdoor Hydronic Heaters" (i.e.,"OHH", SCC=2104008610), and "Outdoor wood burning device, NEC (fire-pits, chimeneas, etc.)" (i.e., "recreational RWC", SCC=21040087000) do not use the same meteorological-based temporalization as the rest of the rwc sector, because meteorological-based temporalization 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 (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.

The 2017 NEI includes two new hydronic heater SCCs, 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.

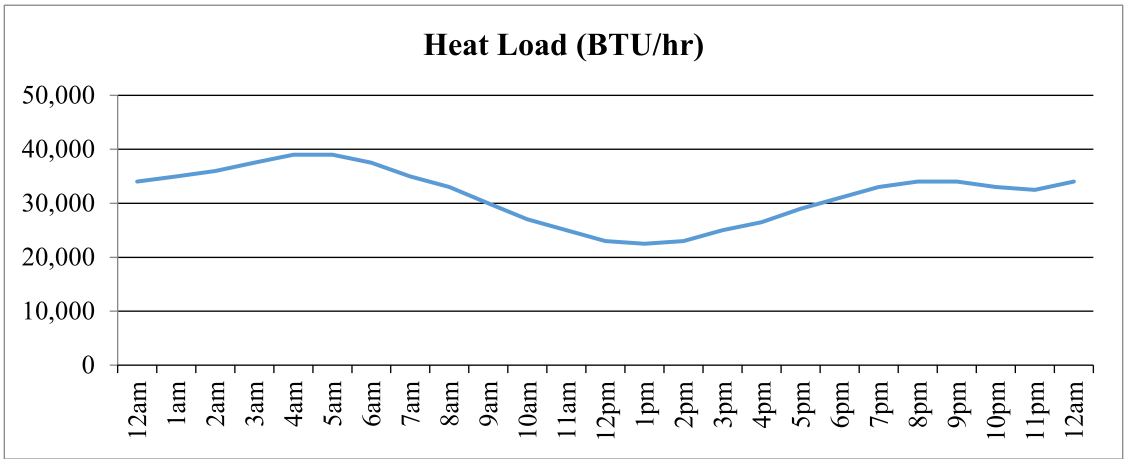


Figure 3-10. Diurnal profile for OHH, based on heat load (BTU/hr)

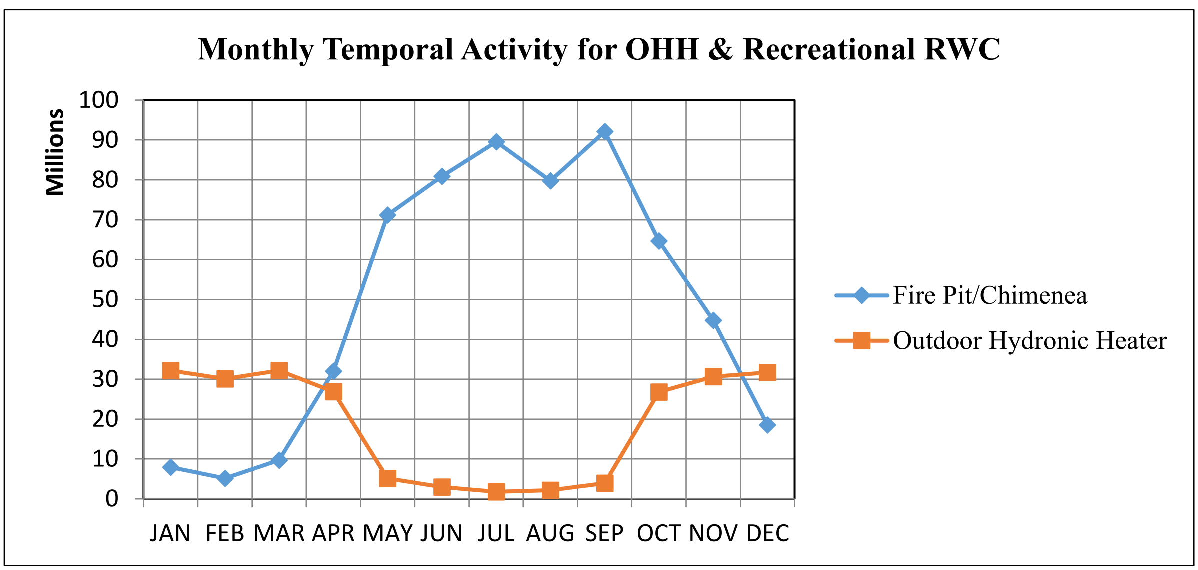


Figure 3--11. Annual-to-month temporal profiles for OHH and recreational RWC

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

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

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

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*

GenTPRO was run using the "BASH_NH3" profile method to create month-to-hour temporal profiles for these sources. Because these profiles distribute to the hour based on monthly emissions, the monthly emissions are obtained from a monthly inventory, or from an annual inventory that has been temporalized to the month. Figure 3-12 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.

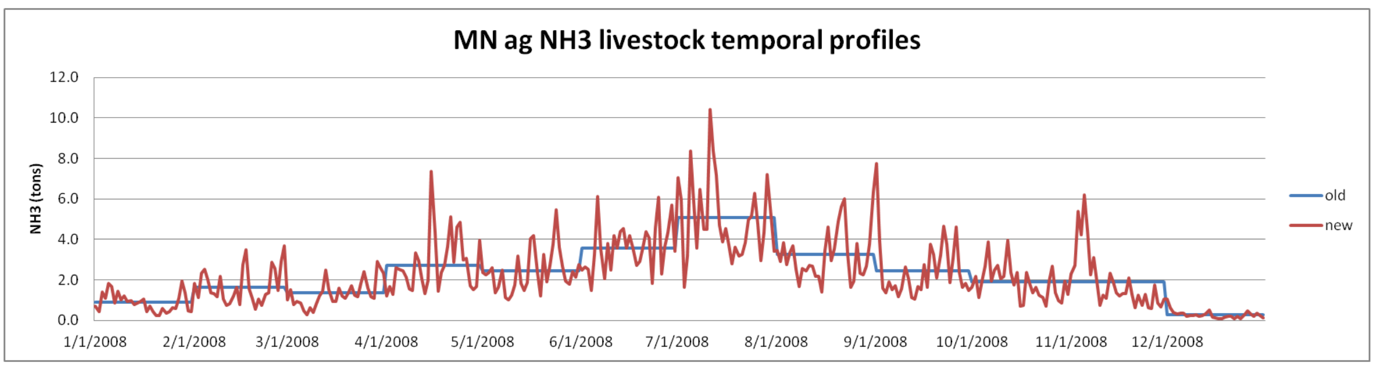


Figure 3--12. Example of animal NH_3 emissions temporalization approaches, summed to daily emissions

For the afdust sector, meteorology is not used in the development of the temporal profiles, but it is used to reduce the total emissions based on meteorological conditions. These adjustments are applied through sector-specific scripts, beginning with the application of land use-based gridded transport fractions and then subsequent zero-outs for hours during which precipitation occurs or there is snow cover on the ground. The land use data used to reduce the NEI emissions explains the amount of emissions that are subject to transport. This methodology is discussed in Pouliot, et al., 2010, and in "Fugitive Dust Modeling for the 2008 Emissions Modeling Platform" (Adelman, 2012). The precipitation adjustment is applied to remove all emissions for days where 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.

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

3.3.5.4 Temporal Profiles for Onroad Mobile Sources

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

Meteorology is not used in the development of the temporal profiles, but rather it impacts the calculation of the hourly emissions through the program Movesmrg. The result is that the emissions vary at the hourly level by grid cell. More specifically, the on-network (RPD) and the off-network parked vehicle (RPV, RPH, and RPP) processes use the gridded meteorology (MCIP) directly. Movesmrg determines the temperature for each hour and grid cell and uses that information to select the appropriate emission factor (EF) for the specified SCC/pollutant/mode combination. RPP uses the gridded minimum and maximum temperature for the day. The combination of these four processes (RPD, RPV, RPH, and RPP) is the total onroad sector emissions. The onroad sector shows a strong meteorological influence on its temporal patterns.

Figure 3-13 illustrates the difference between temporalization of the onroad sector and the meteorological influence via SMOKE-MOVES. Similar temporalization is done for the VMT in SMOKE-MOVES, but the meteorologically varying emission factors add variation on top of the temporalization.

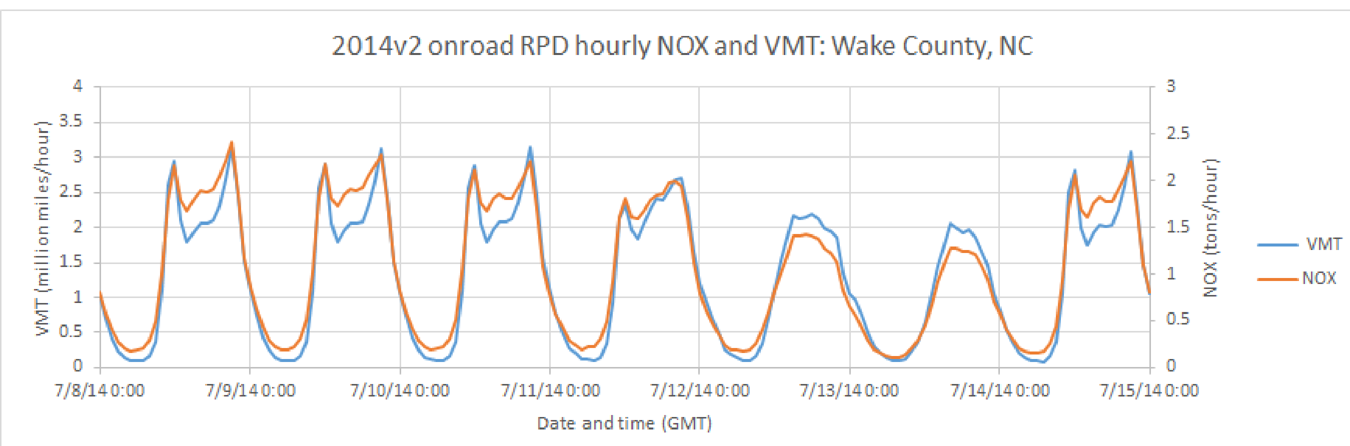


Figure 3-13. Example of SMOKE-MOVES temporal variability of NOx emissions versus activity

For the onroad sector, the "inventories" referred to in Table 3-11 actually consist of activity data, not emissions. For RPP and RPV processes, the VPOP inventory is annual and does not need temporalization. For RPD, the VMT inventory is annual for some sources and monthly for other sources, depending on the source of the data. Sources without monthly VMT were temporalized from annual to month through temporal profiles. VMT data were also temporalized from month to day-of-the-week, and then to hourly through temporal profiles. The RPD processes use an average speed distribution (SPDIST) that 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. Unlike other sectors, the temporal profiles and SPDIST will impact not only the distribution of emissions through time but also the total emissions. Because SMOKE-MOVES (for RPD) calculates emissions from VMT, speed and meteorology, if one shifted the VMT or speed to different hours, it would align with different temperatures and hence different emission factors. In other words, two SMOKE-MOVES runs with identical annual VMT, meteorology, and MOVES emission factors, will have different total emissions if the temporalization of VMT changes. For RPH, the HOTELING inventory is annual and was temporalized to month, day of the week, and hour of the day through temporal profiles. This is an analogous process to RPD except that speed is not included in the calculation of RPH.

For this study, the temporal profiles for the onroad sector come from the 2017 NEI-based platform, which is a compilation of state/local-provided data and nationally available datasets. VMT day-of-week and hour-of-day temporal profiles were developed for counties across the continental U.S. as part of the effort to update the inputs to MOVES and SMOKE-MOVES under CRC A-100 (Coordinating Research Council, 2017). CRC A-100 data includes profiles by region or county, road type, and broad vehicle category. There are three vehicle categories: passenger vehicles (source types 11, 21, and 31), commercial trucks (source types 32, 52, and 53), and combination trucks (source types 61 and 62). CRC A-100 does not cover buses, refuse trucks, or motor homes, so those vehicle types were mapped to other vehicle types for which CRC A-100 did provide profiles, as follows: 1) Intercity/transit buses were mapped to commercial trucks; 2) Motor homes were mapped to passenger vehicles for day-of-week and commercial trucks for hour-of-day; 3) School buses and refuse trucks were mapped to commercial trucks for hour-of-day and use a new custom day-of-week profile called LOWSATSUN that has a very low weekend allocation, since school buses and refuse trucks operate primarily on business days. In addition to temporal profiles, CRC A-100 data were also used to develop the average speed distribution data (SPDIST) used by SMOKE-MOVES. In areas where state-provided data and CRC A-100 data does not exist, hourly speed data is based on MOVES county databases.

The CRC A-100 dataset includes temporal profiles for individual counties, Metropolitan Statistical Areas (MSAs), and entire regions (e.g., West, South). Counties without temporal profiles specific to itself, or to its MSA, are assigned to regional temporal profiles. Temporal profiles also vary between MOVES road types, and there are distinct hour-of-day profiles for each day of the week. Plots of hour-of-day profiles for passenger vehicles in Fulton County, GA, are shown in Figure 3-14. Separate plots are shown for Monday, Friday, Saturday, and Sunday, and each line corresponds to a particular MOVES road type (e.g. road type 2 = rural restricted). Figure 3-15 shows which counties have temporal profiles specific to that county, and which counties use regional average profiles in the CRC A-100 data.

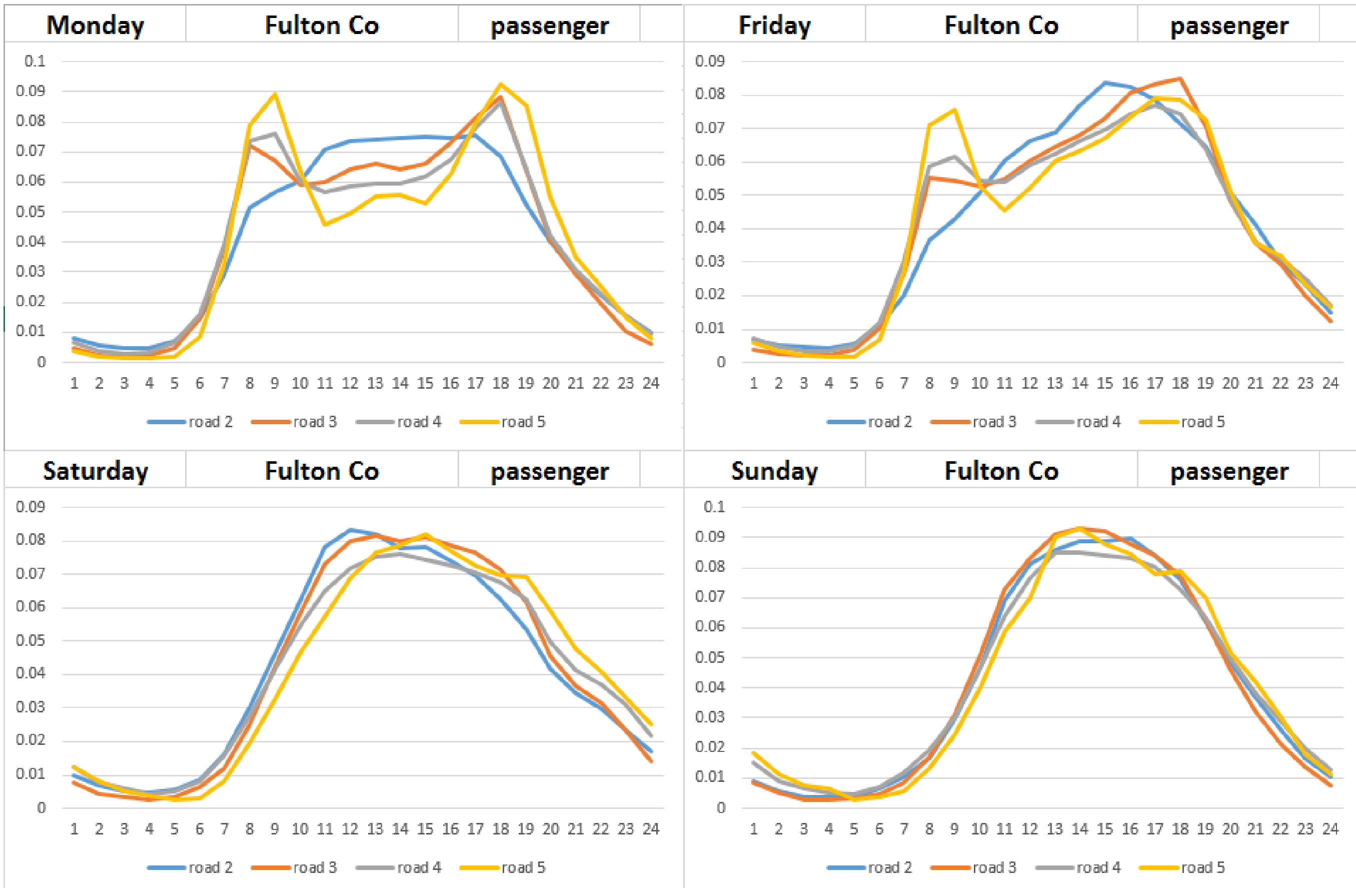


Figure 3-14. Sample onroad diurnal profiles for Fulton County, GA

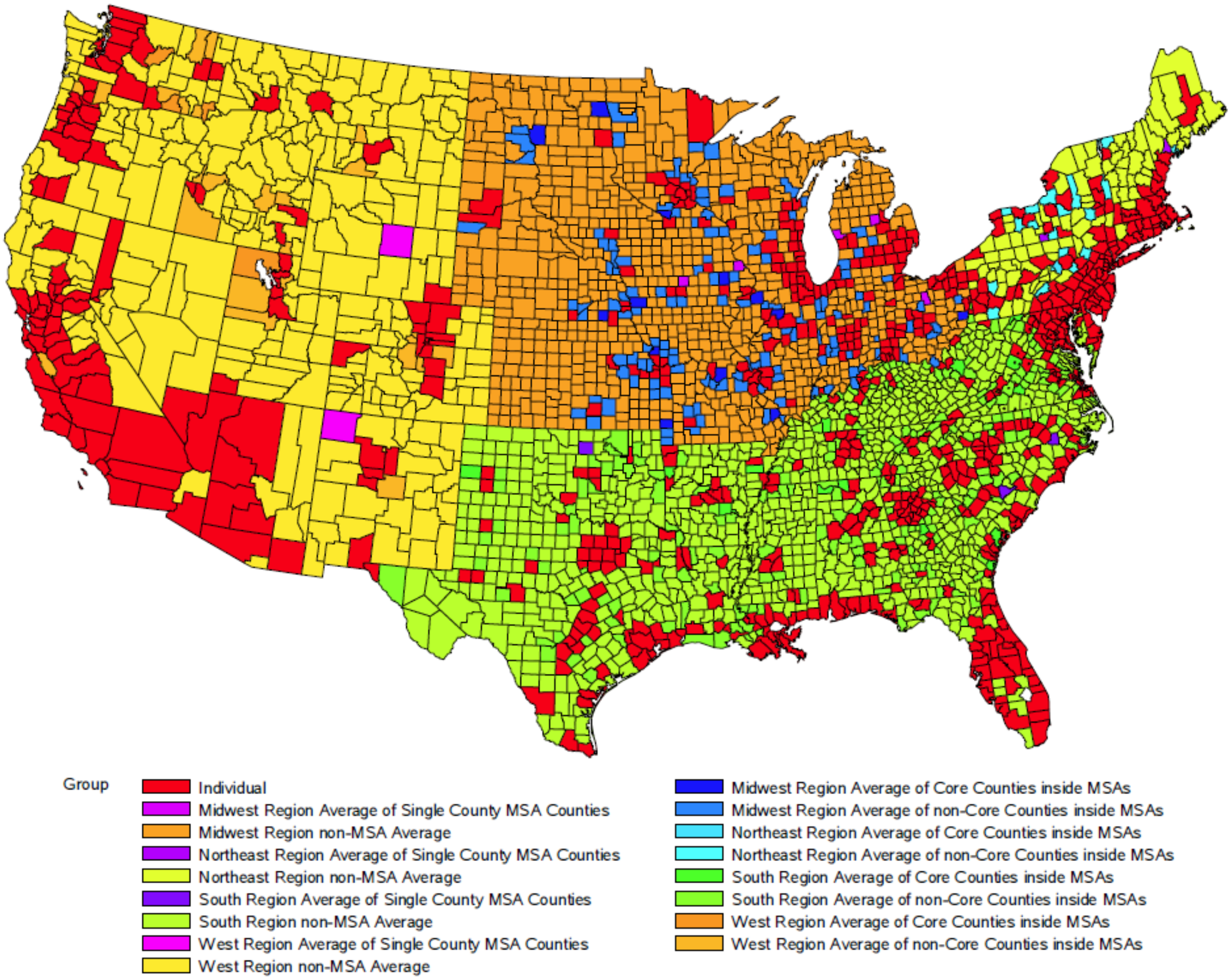


Figure 3-15. Counties for which MOVES Speeds and Temporal Profiles could be Populated

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

State/local-provided data for the 2017 NEI were accepted for use in the 2017 NEI if they were deemed to be at least as credible as the CRC A-100 data. The 2017 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 CRC A-100 temporal profiles were used, except in California. All California temporal profiles were carried over from the 2014v7.1 platform, although California hoteling uses CRC A-100-based profiles just like the rest of the country, since CARB didn't have a hoteling-specific profile. Monthly profiles in all states (national profiles by broad vehicle type) were also carried over from 2014v1 and applied directly to the VMT. For California, CARB supplied diurnal profiles that varied by vehicle type, day of the week,[13] and air basin. These CARB-specific

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

profiles were used in developing EPA estimates for California. Although the EPA adjusted the total emissions to match interpolated 2016 levels based on California's submitted inventories for 2014 and 2017, the temporalization of these emissions took into account both the state-specific VMT profiles and the SMOKE-MOVES process of incorporating meteorology. For more details on the adjustments to California's onroad emissions, see section 3.2.5.1.

3.3.6 Vertical Allocation of Emissions

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

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

CMAQ uses the Briggs algorithm to determine the plume top and bottom, and then computes the plumes' distributions into the vertical layers that the plumes intersect. The pressure difference across each layer divided by the pressure difference across the entire plume is used as a weighting factor to assign the emissions to layers. This approach gives plume fractions by layer and source. Note that the implementation of fire plume rise in CMAQ differs from the implementation of plume rise in SMOKE 4.8. 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 a national 12-km domain. To accomplish this, SMOKE used national 12-km spatial surrogates and a SMOKE area-to-point data file. For the U.S., EPA updated surrogates to use circa 2016-2017 data wherever possible. For Mexico, updated spatial surrogates were used as described below. 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.

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 3-12 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, or as an input for merging two surrogates to create a new surrogate that is used.

Many of the surrogates currently in use were developed for use in the 2014v7.0 platform using recently available data sets (Adelman, 2016). They include the 2011 National Land Cover Database is used including various development density levels such as open, low, medium high and various combinations of these. These landuse surrogates largely replaced the FEMA category surrogates that were used in the 2011 platform. Additionally, onroad surrogates were developed using average annual daily traffic counts from the highway monitoring performance system (HPMS). Onroad surrogates for this platform do not distinguish between urban and rural road types, which prevents issues in areas where there are inconsistent urban and rural definitions between MOVES and the surrogate data.

Recent surrogate updates include:

- A public school surrogate (508) was developed for off-network school buses.
- Oil and gas surrogates were updated to represent 2019, including a new surrogate for total gas produced (689)
- Corrections were made to the rail surrogates (261/271).
- The transit bus terminal surrogate was re-gapfilled with the NLCD medium+high surrogate (306)
- Some gridding cross reference corrections / updates were made including the use of NLCD medium+high surrogate instead of intercity bus terminals for off network emissions from other buses.
- The 500 series surrogates are no longer used and SCCs that used them (e.g., cigarette smoke, accidental releases) were remapped to NLCD surrogates.
- Onroad surrogates were generated to incorporate 2017 Average Annual Daily Traffic (AADT);

Surrogates for the U.S. were generated using the Surrogate Tool to drive the Spatial Allocator with some surrogates were developed directly within ArcGIS or using the Surrogate Tools DB. 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, and the tool and documentation for the Surrogate Tools DB is available from https://www.cmascenter.org/surrogate_tools_db/. The file Surrogate_specifications_2019_platform_US_Can_Mex.xlsx documents the configuration for generating the surrogates.

Table 3-12. 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)	650	Refineries and Tank Farms
100	Population	670	Spud Count - CBM Wells
110	*Housing*	671	Spud Count - Gas Wells

Code	Surrogate Description	Code	Surrogate Description
150	Residential Heating - Natural Gas	*672*	*Gas Production at Oil Wells*
170	Residential Heating - Distillate Oil	*673*	*Oil Production at CBM Wells*
180	Residential Heating – Coal	674	Unconventional Well Completion Counts
190	Residential Heating - LP Gas	*676*	*Well Count - All Producing*
205	Extended Idle Locations	*677*	*Well Count - All Exploratory*
239	Total Road AADT	678	Completions at Gas Wells
240	Total Road Miles	679	Completions at CBM Wells
242	All Restricted AADT	681	Spud Count - Oil Wells
244	All Unrestricted AADT	*683*	*Produced Water at All Wells*
258	*Intercity Bus Terminals*	6831	Produced Water at CBM Wells
259	Transit Bus Terminals	6832	Produced Water at Gas Wells
260	*Total Railroad Miles*	6833	Produced Water at Oil Wells
261	NTAD Total Railroad Density	685	Completions at Oil Wells
271	NTAD Class 1 2 3 Railroad Density	*686*	*Completions at All Wells*
300	NLCD Low Intensity Development	687	Feet Drilled at All Wells
304	NLCD Open + Low	689	Gas Produced - Total
305	NLCD Low + Med	691	Well Counts - CBM Wells
306	NLCD Med + High	*692*	*Spud Count - All Wells*
307	NLCD All Development	*693*	*Well Count - All Wells*
308	NLCD Low + Med + High	694	Oil Production at Oil Wells
309	NLCD Open + Low + Med	695	Well Count - Oil Wells
310	NLCD Total Agriculture	696	Gas Production at Gas Wells
319	NLCD Crop Land	*697*	*Oil Production at Gas Wells*
320	NLCD Forest Land	698	Well Count - Gas Wells
321	NLCD Recreational Land	699	Gas Production at CBM Wells
340	*NLCD Land*	711	Airport Areas
350	NLCD Water	801	Port Areas
500	*Commercial Land*	*805*	*Offshore Shipping Area*
505	*Industrial Land*	*806*	*Offshore Shipping NEI2014 Activity*
506	*Education*	*807*	*Navigable Waterway Miles*
508	Public Schools	*808*	*2013 Shipping Density*
510	*Commercial plus Industrial*	*820*	*Ports NEI2014 Activity*
535	*Residential + Commercial + Industrial + Institutional + Government*	850	Golf Courses
560	*Hospital (COM6)*	860	Mines

For the onroad sector, the on-network (RPD) emissions were allocated differently from the off-network (RPP, starts, ONI, and RPV). On-network used average annual daily traffic (AADT) data and off network used land use surrogates as shown in Table 3-13. Emissions from the extended (i.e., overnight) idling of trucks were assigned to surrogate 205 that is based on locations of overnight truck parking spaces. The underlying data in this surrogate was updated for use in the 2016 platform to include additional data sources and corrections based on comments received.

Table 3-13. 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

Source type	Source Type name	**Surrogate ID**	**Description**
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 3-14 using 2019 data consistent with what was used to develop the 2017 nonpoint oil and gas emissions. The exploration and production of oil and gas has 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 2019 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 2019.

The primary activity data source used for the development of the oil and gas spatial surrogates was data from Drilling Info (DI) Desktop's HPDI database (Drilling Info, 2019[14]). This database contains well-level location, production, and exploration statistics at the monthly level. Due to a proprietary agreement with DI Desktop, individual well locations and ancillary production cannot be made publicly available, but aggregated statistics are allowed. These data were supplemented with data from state Oil and Gas Commission (OGC) websites (Alaska, Arizona, Idaho, Illinois, Indiana, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Nevada, Oregon and Pennsylvania, 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) was downloaded and used. Under that methodology, both completion date and date of first production from HPDI were used to identify wells completed during 2019.

The spatial surrogates, numbered 670 through 699 and also 6831, 6832, and 6833, were processed at 12km resolution and gapfilled with the Surrogate Tool. The surrogates were first 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 part 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.

Table 3-14. Spatial Surrogates for Oil and Gas Sources

Surrogate Code	Surrogate Description
670	Spud Count - CBM Wells
671	Spud Count - Gas Wells

[14] https://www.enverus.com/drillinginfo-and-rigdata/

Surrogate Code	Surrogate Description
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

Not all of the available surrogates are used to spatially allocate sources in the modeling platform; that is, some surrogates shown in Table 3-12 were not assigned to any SCCs, although many of the "unused" surrogates are actually used to "gap fill" other surrogates that are used. When the source data for a surrogate has no values for a particular county, gap filling is used to provide values for the surrogate in those counties to ensure that no emissions are dropped when the spatial surrogates are applied to the emission inventories. The U.S. CAP emissions allocated to the various spatial surrogates are shown in Table 3-15.

Table 3-15. Selected 2019 CAP emissions by sector for U.S. Surrogates (CONUS domain totals, tons)

Sector	ID	Description	NH_3	NO_X	$PM_{2.5}$	SO_2	VOC
afdust	240	Total Road Miles	0	0	315,096	0	0
afdust	304	NLCD Open + Low	0	0	842,116	0	0
afdust	306	NLCD Med + High	0	0	52,278	0	0

Sector	ID	Description	NH_3	NO_X	$PM_{2.5}$	SO_2	VOC
afdust	308	NLCD Low + Med + High	0	0	117,047	0	0
afdust	310	NLCD Total Agriculture	0	0	791,881	0	0
livestock	310	NLCD Total Agriculture	2,602,279	0	0	0	227,985
nonpt	100	Population	34,304	0	0	0	208
nonpt	150	Residential Heating - Natural Gas	33,550	204,371	4,041	1,365	12,055
nonpt	170	Residential Heating - Distillate Oil	1,531	30,031	3,284	11,510	1,039
nonpt	180	Residential Heating - Coal	1	3	1	3	3
nonpt	190	Residential Heating - LP Gas	98	31,061	163	712	1,181
nonpt	239	Total Road AADT	0	22	541	0	297,798
nonpt	240	Total Road Miles	0	0	0	0	0
nonpt	244	All Unrestricted AADT	0	0	0	0	101,255
nonpt	271	NTAD Class 1 2 3 Railroad Density	0	0	0	0	2,203
nonpt	300	NLCD Low Intensity Development	4,823	19,093	94,548	2,882	72,599
nonpt	304	NLCD Open + Low	0	0	0	0	0
nonpt	306	NLCD Med + High	23,668	272,514	245,871	131,592	112,049
nonpt	307	NLCD All Development	85	25,798	110,610	8,169	69,262
nonpt	308	NLCD Low + Med + High	884	156,033	15,683	10,076	10,037
nonpt	310	NLCD Total Agriculture	0	0	38	0	0
nonpt	319	NLCD Crop Land	0	0	97	72	299
nonpt	320	NLCD Forest Land	3,953	68	273	0	279
nonpt	650	Refineries and Tank Farms	0	16	0	0	106,401
nonpt	711	Airport Areas	0	0	0	0	596
nonpt	801	Port Areas	0	0	0	0	6,730
nonroad	261	NTAD Total Railroad Density	3	1,807	184	1	356
nonroad	304	NLCD Open + Low	4	1,620	136	3	2,423
nonroad	305	NLCD Low + Med	96	14,661	3,879	77	105,749
nonroad	306	NLCD Med + High	335	160,244	9,947	265	89,188
nonroad	307	NLCD All Development	101	29,155	15,414	73	170,963
nonroad	308	NLCD Low + Med + High	565	262,271	21,894	231	45,811
nonroad	309	NLCD Open + Low + Med	122	21,080	1,240	99	46,015
nonroad	310	NLCD Total Agriculture	421	305,710	21,805	183	32,213
nonroad	320	NLCD Forest Land	15	3,281	522	8	3,675
nonroad	321	NLCD Recreational Land	83	13,038	5,523	57	203,840
nonroad	350	NLCD Water	192	113,237	4,467	160	268,186
nonroad	850	Golf Courses	13	2,087	119	10	5,738
nonroad	860	Mines	2	2,467	240	1	462
np_oilgas	670	Spud Count - CBM Wells	0	0	0	0	89
np_oilgas	671	Spud Count - Gas Wells	0	0	0	0	3,284
np_oilgas	674	Unconventional Well Completion Counts	27	20,730	496	26	870
np_oilgas	678	Completions at Gas Wells	0	7,874	193	2,968	20,964
np_oilgas	679	Completions at CBM Wells	0	9	0	400	1,159
np_oilgas	681	Spud Count - Oil Wells	0	0	0	0	38,758

Sector	ID	Description	NH_3	NO_X	$PM_{2.5}$	SO_2	VOC
np_oilgas	685	Completions at Oil Wells	0	377	0	1,361	42,017
np_oilgas	687	Feet Drilled at All Wells	0	75,545	1,995	107	3,318
np_oilgas	689	Gas Produced - Total	0	460	55	4	68,697
np_oilgas	691	Well Counts - CBM Wells	0	29,113	521	11	30,841
np_oilgas	694	Oil Production at Oil Wells	0	3,695	0	31,403	1,052,276
np_oilgas	695	Well Count - Oil Wells	0	129,122	3,032	1,465	676,769
np_oilgas	696	Gas Production at Gas Wells	0	211	0	1	50,268
np_oilgas	698	Well Count - Gas Wells	0	253,031	4,794	127	498,114
np_oilgas	699	Gas Production at CBM Wells	0	47	5	0	5,190
np_oilgas	6831	Produced water at CBM wells	0	0	0	0	3,695
np_oilgas	6832	Produced water at gas wells	0	0	0	0	38,515
np_oilgas	6833	Produced water at oil wells	0	0	0	0	46,549
np_solvents	100	Population	0	0	0	0	1,372,923
np_solvents	240	Total Road Miles	0	0	0	0	48,397
np_solvents	306	NLCD Med + High	33	27	300	1	409,967
np_solvents	307	NLCD All Development	24	6	19	5	527,883
np_solvents	308	NLCD Low + Med + High	0	0	129	0	7,970
np_solvents	310	NLCD Total Agriculture	0	0	0	0	149,185
onroad	205	Extended Idle Locations	342	34,291	780	18	4,084
onroad	242	All Restricted AADT	34,157	925,436	25,956	5,041	134,605
onroad	244	All Unrestricted AADT	63,200	1,496,382	55,842	10,503	374,603
onroad	259	Transit Bus Terminals	15	2,037	50	1	439
onroad	304	NLCD Open + Low	0	687	20	0	4,124
onroad	306	NLCD Med + High	921	95,906	3,529	76	20,258
onroad	307	NLCD All Development	3,576	204,352	7,433	960	594,388
onroad	308	NLCD Low + Med + High	204	19,149	569	56	29,121
onroad	508	Public Schools	16	2,189	81	1	544
rail	261	NTAD Total Railroad Density	14	35,834	1,061	31	1,822
rail	271	NTAD Class 1 2 3 Railroad Density	324	480,174	13,760	644	20,118
rwc	300	NLCD Low Intensity Development	16,369	33,925	297,877	7,937	322,528

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

There are numerous airport-related emission sources in the NEI, such as aircraft, airport ground support equipment, and jet refueling. The modeling platform includes the aircraft and airport ground support equipment emissions as point sources. For the modeling platform, EPA used the SMOKE "area-to-point" (ARTOPNT) approach for only jet refueling in the nonpt sector. The following SCCs use this approach: 2501080050 and 2501080100 (petroleum storage at airports), and 2810040000 (aircraft/rocket engine firing and testing). The ARTOPNT approach is described in detail in the 2002 platform documentation: https://www.epa.gov/sites/production/files/2020-10/documents/emissions_tsd_vol1_02-28-08.pdf.

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 2015 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 3-16. The population surrogate was updated for Mexico for the 2014v7.1 platform. Surrogate code 11, which uses 2015 population data at 1 km resolution, replaces the previous population surrogate code 10. The other surrogates for Mexico are circa 1999 and 2000 and were based on data obtained from the Sistema Municpal de Bases de Datos (SIMBAD) de INEGI and the Bases de datos del Censo Economico 1999. Most of the CAPs allocated to the Mexico and Canada surrogates are shown in Table 3-17. The entries in Table 3-17 are for the othar, othafdust, onroad_can, and onroad_mex sectors.

Table 3-16. Canadian Spatial Surrogates

Code	Canadian Surrogate Description	Code	Description
100	Population	921	Commercial Fuel Combustion
101	total dwelling	923	TOTAL INSTITUTIONAL AND GOVERNEMNT
104	capped total dwelling	924	Primary Industry
106	ALL_INDUST	925	Manufacturing and Assembly
113	Forestry and logging	926	Distribution and Retail (no petroleum)
200	Urban Primary Road Miles	927	Commercial Services
210	Rural Primary Road Miles	932	CANRAIL
211	Oil and Gas Extraction	940	PAVED ROADS NEW
212	Mining except oil and gas	946	Construction and mining
220	Urban Secondary Road Miles	948	Forest
221	Total Mining	951	Wood Consumption Percentage
222	Utilities	955	UNPAVED_ROADS_AND_TRAILS
230	Rural Secondary Road Miles	960	TOTBEEF
233	Total Land Development	970	TOTPOUL
240	capped population	980	TOTSWIN
308	Food manufacturing	990	TOTFERT
321	Wood product manufacturing	996	urban_area
323	Printing and related support activities	1251	OFFR_TOTFERT
324	Petroleum and coal products manufacturing	1252	OFFR_MINES
326	Plastics and rubber products manufacturing	1253	OFFR Other Construction not Urban
327	Non-metallic mineral product manufacturing	1254	OFFR Commercial Services
331	Primary Metal Manufacturing	1255	OFFR Oil Sands Mines
350	Water	1256	OFFR Wood industries CANVEC
412	Petroleum product wholesaler-distributors	1257	OFFR UNPAVED ROADS RURAL
448	clothing and clothing accessories stores	1258	OFFR_Utilities

Code	Canadian Surrogate Description	Code	Description
482	Rail transportation	1259	OFFR total dwelling
562	Waste management and remediation services	1260	OFFR_water
901	AIRPORT	1261	OFFR_ALL_INDUST
902	Military LTO	1262	OFFR Oil and Gas Extraction
903	Commercial LTO	1263	OFFR_ALLROADS
904	General Aviation LTO	1265	OFFR_CANRAIL
945	Commercial Marine Vessels	9450	Commercial Marine Vessel Ports

Table 3-17. 2018 CAPs Allocated to Mexican and Canadian Spatial Surrogates for 12US1 (tons)

Code	Mexican or Canadian Surrogate Description	NH_3	NO_X	$PM_{2.5}$	SO_2	VOC
11	MEX 2015 Population	0	85,151	465	178	223,016
14	MEX Residential Heating - Wood	0	2,333	6,512	190	17,541
16	MEX Residential Heating - Distillate Oil	1	28	0	0	1
22	MEX Total Road Miles	2,983	359,738	12,086	6,647	71,102
24	MEX Total Railroads Miles	0	19,518	408	185	732
26	MEX Total Agriculture	115,677	18,186	16,239	473	3,847
32	MEX Commercial Land	0	75	1,631	0	27,763
34	MEX Industrial Land	92	2,036	1,257	6	34,866
36	MEX Commercial plus Industrial Land	5	7,049	305	15	101,386
40	MEX Residential (RES1-4)+Comercial+Industrial+Institutional +Government	0	15	61	2	21,041
42	MEX Personal Repair (COM3)	0	0	0	0	5,130
44	MEX Airports Area	0	3,420	48	241	1,294
48	MEX Brick Kilns	0	266	5,297	470	130
50	MEX Mobile sources - Border Crossing	3	58	2	0	45
100	CAN Population	776	51	604	14	219
101	CAN total dwelling	0	0	0	0	147,322
104	CAN capped total dwelling	347	30,961	2,282	2,473	1,614
106	CAN ALL_INDUST			583		
113	CAN Forestry and logging	117	1,392	7,471	29	3,938
200	CAN Urban Primary Road Miles	1,572	66,863	2,365	232	7,134
210	CAN Rural Primary Road Miles	628	39,000	1,297	98	3,023
211	CAN Oil and Gas Extraction	1	39	424	41	1,657
212	CAN Mining except oil and gas	0	0	3,051	0	0
220	CAN Urban Secondary Road Miles	2,951	105,604	4,764	482	18,852
221	CAN Total Mining	0	0	13,221	0	0

Code	Mexican or Canadian Surrogate Description	NH_3	NO_X	$PM_{2.5}$	SO_2	VOC
222	CAN Utilities	55	3,344	2,859	453	63
230	CAN Rural Secondary Road Miles	1,652	71,811	2,515	257	8,255
240	CAN capped population	327	44,524	1,385	52	95,157
308	CAN Food manufacturing	0	0	18,982	0	17,387
321	CAN Wood product manufacturing	785	4,450	1,543	328	15,455
323	CAN Printing and related support activities	0	0	0	0	11,693
324	CAN Petroleum and coal products manufacturing	0	1,179	1,599	462	9,154
326	CAN Plastics and rubber products manufacturing	0	0	0	0	24,027
327	CAN Non-metallic mineral product manufacturing	0	0	6,449	0	0
331	CAN Primary Metal Manufacturing	0	156	5,561	29	72
412	CAN Petroleum product wholesaler-distributors	0	0	0	0	43,724
448	CAN clothing and clothing accessories stores	0	0	0	0	140
482	CAN Rail transportation	1	4,062	88	1	256
562	CAN Waste management and remediation services	240	1,924	2,620	2,483	9,199
901	CAN AIRPORT	0	93	9	0	9
921	CAN Commercial Fuel Combustion	187	24,020	2,379	1,414	1,224
923	CAN TOTAL INSTITUTIONAL AND GOVERNEMNT	0	0	0	0	14,458
924	CAN Primary Industry	0	0	0	0	38,858
925	CAN Manufacturing and Assembly	0	0	0	0	69,488
926	CAN Distribution and Retail (no petroleum)	0	0	0	0	7,285
927	CAN Commercial Services	0	0	0	0	31,311
932	CAN CANRAIL	48	83,844	1,662	43	3,559
940	CAN PAVED ROADS NEW			27,751		
946	CAN Construction and mining	0	0	0	0	9,850
951	CAN Wood Consumption Percentage	957	10,634	107,554	1,519	152,072
955	CAN UNPAVED_ROADS_AND_TRAILS			383,147		
990	CAN TOTFERT	48	4,047	265	6,827	152
996	CAN urban_area	0	0	2,994	0	0
1251	CAN OFFR_TOTFERT	73	50,505	3,421	48	4,585
1252	CAN OFFR_MINES	1	563	38	1	82
1253	CAN OFFR Other Construction not Urban	68	28,864	3,601	41	10,186
1254	CAN OFFR Commercial Services	43	14,335	2,235	26	35,468
1255	OFFR Oil Sands Mines	0	0	0	0	0

Code	Mexican or Canadian Surrogate Description	NH_3	NO_X	$PM_{2.5}$	SO_2	VOC
1256	CAN OFFR Wood industries CANVEC	7	2,103	214	4	853
1257	CAN OFFR UNPAVED ROADS RURAL	24	10,272	574	14	24,333
1258	CAN OFFR_Utilities	8	3,727	176	5	824
1259	CAN OFFR total dwelling	17	5,844	588	10	12,557
1260	CAN OFFR_water	19	5,604	264	22	20,367
1261	CAN OFFR_ALL_INDUST	4	5,065	145	2	1,068
1262	CAN OFFR Oil and Gas Extraction	1	570	42	0	168
1263	CAN OFFR_ALLROADS	3	1,361	129	2	438
1265	CAN OFFR_CANRAIL	0	489	15	0	37

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.

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

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

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.

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

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

California Air Resources Board (CARB): Final 2015 Consumer & Commercial Product Survey Data Summaries, 2019.

CRC, 2017. Improvement of Default Inputs for MOVES and SMOKE-MOVES: Final Report. Coordinating Research Council Report No. A-100. Available from http://crcsite.wpengine.com/wp-content/uploads/2019/05/ERG_FinalReport_CRCA100_28Feb2017.pdf

U.S. Environmental Protection Agency: 2017 National Emissions Inventory (NEI), Research Triangle Park, NC, 2020b. https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-data.

EPA, 2020. SPECIATE Version 5.1 Database Development Documentation, U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC 27711, EPA/600/R-20/189, June 2020. Available at https://www.epa.gov/sites/production/files/2020-07/documents/speciate_5.1.pdf.

EPA, 2020. Speciation of Total Organic Gas and Particulate Matter Emissions from On-road Vehicles in MOVES3. Office of Transportation and Air Quality. Report # EPA-420-R-20-021, November, 2020. Available from https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1010THD.pdf.

EPA, 2019. Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 7.1, 2016 Emissions Modeling Platform for the 2014 National Air Toxics Assessment. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at https://www.epa.gov/air-emissions-modeling/2016-version-71-technical-support-document.

EPA, 2021. January 2021 release of the 2017 National Emissions Inventory (NEI) Technical Support Document (TSD). Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-technical-support-document-tsd.

EPA, 2020b. Technical Support Document (TSD) Preparation of Emissions Inventories for the 2016v1 North American Emissions Modeling Platform Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available from https://www.epa.gov/air-emissions-modeling/2016v1-platform.

EPA, 2022. Technical Support Document (TSD) Preparation of Emissions Inventories for the 2019 North American Emissions Modeling Platform Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available from https://www.epa.gov/air-emissions-modeling/2019-emissions-modeling-platform.

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

ERG, 2017. "Technical Report: Development of Mexico Emission Inventories for the 2014 Modeling Platform." https://gaftp.epa.gov/Air/emismod/2014/v2/2014fd/emissions/EPA%205-18%20Report_Clean%20Final_01042017.pdf

The Freedonia Group: Solvents, Industry Study #3429, 2016.

Foley, 2020. "EQUATES: EPA's Air QUAlity TimE Series Project". Presented at the 19th Annual CMAS Conference. Available from: https://www.cmascenter.org/conference/2020/slides/KFoley_EQUATES_CMAS_2020.pdf

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.

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.

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

NEIC, 2019; 2016v1 platform nonroad specification sheet, Dallas Texas, 2019. https://www.epa.gov/air-emissions-inventories/2019-international-emissions-inventory-conference-collaborative.

NESCAUM, 2006; "Assessment of Outdoor Wood-fired Boilers". Northeast States for Coordinated Air Use Management (NESCAUM). Available from https://www.nescaum.org/documents/assessment-of-outdoor-wood-fired-boilers/2006-1031-owb-report_revised-june2006-appendix.pdf/.

NCAR, 2019. FIRE EMISSION FACTORS AND EMISSION INVENTORIES, FINN Data. downloaded from https://www.fs.fed.us/pnw/fera/research/smoke/consume/index.shtml.

NYSERDA, 2012; "Environmental, Energy Market, and Health Characterization of Wood-Fired Hydronic Heater Technologies, Final Report". New York State Energy Research and Development Authority (NYSERDA)..

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

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

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

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

Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), Gobierno De Mexico, 2019. Inventarios Nacionales de Emisiones en el ano 2016. Documentation available from SEMARNAT: https://www.gob.mx/semarnat/documentos/documentos-del-inventario-nacional-de-emisiones.

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. 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.

Urbanski S.P. (2014) Wildland fire emissions, carbon, and climate: emissions factors. Forest Ecology and Management, 317, 51-60.

U.S. Bureau of Labor Statistics, Producer Price Index by Industry, retrieved from FRED, Federal Reserve Bank of St. Louis, 2020. https://fred.stlouisfed.org/categories/31

U.S. Census Bureau: 2019 Annual Survey of Manufacturers (ASM), Washington D.C., USA, 2021.

U.S. Census Bureau, Economy Wide Statistics Division: County Business Patterns, 2019. https://www.census.gov/programs-surveys/cbp/data/datasets.html.

U.S. Census Bureau: Paint and Allied Products - 2010, MA325F(10), 2011.

U.S. Department of Agriculture, 2017 National Agricultural Statistics Service Survey, 2017. https://www.nass.usda.gov/Publications/AgCensus/2017/index.php

U.S. Department of Transportation and the U.S. Department of Commerce, 2012 Commodity Flow Survey, EC12TCF-US, 2015. https://www.census.gov/library/publications/2015/econ/ec12tcf-us.html.

U.S. Energy Information Administration: The Distribution of U.S. Oil and Natural Gas Wells by Production Rate, Washington, DC, 2019. 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.

Weschler, C. J., and Nazaroff, W. W.: Semivolatile organic compounds in indoor environments, Atmos Environ, 42, 9018-9040, 2008.

Wiedinmyer, C., S.K. Akagi, R.J. Yokelson, L.K. Emmons, J.A. Al-Saadi, J. J. Orlando, 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., Jung, J., Whitten, G., Heo, G., Mellberg, J., and Estes, M.: Updates to the Carbon Bond Mechanism for Version 6 (CB6), 9th Annual CMAS Conference, Chapel Hill, NC, October 11-13, 2010.

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 NOx to reduce ozone levels, may exacerbate other air pollutants such as particulate matter. To meet the need to address the complex relationships between pollutants, EPA developed the Community Multiscale Air Quality (CMAQ) modeling system.[15] 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 2019 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

---

[15] 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.[16] 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 $PM_{2.5}$ species provided in Section 4.3 shows generally acceptable model performance which is equivalent or better than typical state-of-the-science regional modeling simulations as summarized in Simon et al., 2012.[17]

4.2 CMAQ Model Version, Inputs and Configuration

This section describes the air quality modeling platform used for the 2019 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 2019 Platform to provide a national

[16] U.S. EPA (2018), Modeling Guidance for Demonstrating Air Quality Goals for Ozone, PM2.5, and Regional Haze, pp 205. https://www3.epa.gov/ttn/scram/guidance/guide/O3-PM-RH-Modeling_Guidance-2018.pdf.
[17] 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 $PM_{2.5}$.

This section provides a description of each of the main components of the 2019 CMAQ simulation along with the results of a model performance evaluation in which the 2019 model predictions are compared to corresponding measured ambient concentrations.

4.2.1 CMAQ Model Version

CMAQ is a non-proprietary computer model that simulates the formation and fate of photochemical oxidants, including $PM_{2.5}$ and ozone, for given input sets of meteorological conditions and emissions. As mentioned previously, CMAQ includes numerous science modules that simulate the emission, production, decay, deposition and transport of organic and inorganic gas-phase and particle pollutants in the atmosphere. This 2019 analysis employed CMAQ version 5.3.2.[18] The 2019 CMAQ run included CB6r3 chemistry[19], AERO7 aerosol module[20] with non-volatile Primary Organic Aerosol (POA), and updated halogen chemistry[21]. The CMAQ community model versions 5.0.2 and 5.1 were most recently peer-reviewed in September of 2016 for the U.S. EPA.[22]

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 2019 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 2019 simulation. Air quality conditions at the outer boundary of the 12-km domain were taken from the northern hemispheric CMAQ model (discussed in Section 4.2.4).

[18] CMAQ version 5.3.2: https://doi.org/10.5281/zenodo.4081737; https://www.epa.gov/cmaq/cmaq-models-0. CMAQ v5.3.2 is also available from the Community Modeling and Analysis System (CMAS) at: http://www.cmascenter.org.
[19] 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.
[20] 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.
[21] 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.
[22] Moran, M.D., Astitha, M., Barsanti, K.C., Brown, N.J., Kaduwela, A., McKeen, S.A., Pickering, K.E. (September 28, 2015). Final Report: Fifth Peer Review of the CMAQ Model, https://www.epa.gov/sites/production/files/2016-11/documents/cmaq_fifth_review_final_report_2015.pdf. This peer review was focused on CMAQ v5.0.2, which was released in May, 2014, as well as CMAQ v5.1, which was released in October 2015. It is available from the Community Modeling and Analysis System (CMAS) as well as previous peer-review reports at: http://www.cmascenter.org.

Table 4-1. Geographic Information for 2019 12-km Modeling Domain

National 12 km CMAQ Modeling Configuration	
Map Projection	Lambert Conformal Projection
Grid Resolution	12 km
Coordinate Center	97 W, 40 N
True Latitudes	33 and 45 N
Dimensions	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 2019 CMAQ simulation (heights are layer top).

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

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

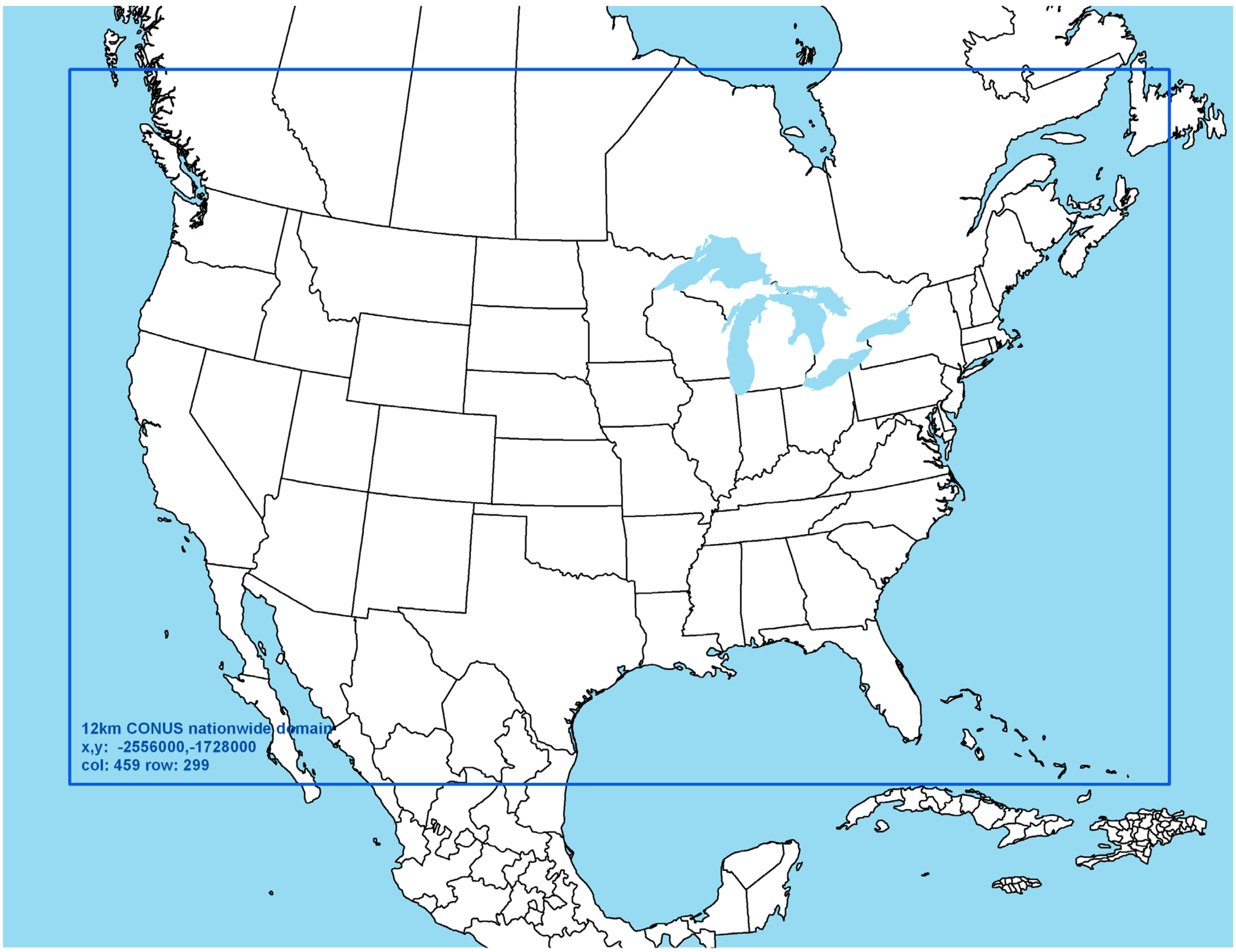


Figure 4-1. Map of the 2019 CMAQ Modeling Domain. The blue box denotes the 12-km national modeling domain.

4.2.3 Modeling Period / Ozone Episodes

The 12-km CMAQ modeling domain was modeled for the entire year of 2019. The annual simulation included a spin-up period, comprised of 10 days before the beginning of the simulation, to mitigate the effects of initial concentrations. All 365 model days were used in the annual average levels of $PM_{2.5}$. For the 8-hour ozone, we used modeling results from the period between May 1 and September 30. This 153-day period generally conforms to the ozone season across most parts of the U.S. and contains the majority of days that observed high ozone concentrations.

4.2.4 Model Inputs: Emissions, Meteorology and Boundary Conditions

2019 Emissions: The emissions inventories used in the 2019 air quality modeling are described in Section 3, above.

2019 Meteorological Input Data: The gridded meteorological data for the entire year of 2019 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.[23] 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)[24] 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 2019 WRF meteorology simulated was based on 2011 National Land Cover Database (NLCD).[25] 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[26] and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme.[27] In addition, the Group for High Resolution Sea Surface Temperatures (GHRSST)[28,29] 1-km SST data was used for SST information to provide more resolved information compared to the more coarse data in the NAM analysis.

---

[23] 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.
[24] North American Model Analysis-Only, http://nomads.ncdc.noaa.gov/data.php; download from ftp://nomads.ncdc.noaa.gov/NAM/analysis_only/.
[25] National Land Cover Database 2011, http://www.mrlc.gov/nlcd2011.php.
[26] 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.
[27] 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.
[28] 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.
[29] Global High-Resolution SST (GHRSST) 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.

2019 Initial and Boundary Conditions: The 2019 annual lateral boundary and initial species concentrations were provided using an Hemispheric CMAQv5.3.2 (H-CMAQ) simulation. The CMAQ model simulation covered the Northern Hemisphere at 108km horizontal resolution with a 44-layer vertical structure reaching 50 hPa.[30] Simulations use CB6r3 chemistry[31], AERO7 aerosols[32], and updated halogen chemistry[33]. Anthropogenic emissions were modeled using representative day emissions, created by averaging prior emissions on a day-of-week basis by month. Anthropogenic emissions and LNOx emissions are prior emissions scaled using inverse estimates based on OMI NO2 satellite observations. The inverse modeling system for estimating emissions is described in East et al. (2022, in prep).[34] The 2019 prior simulation was initialized with a 1-year spin-up period not considered in the analyses. Biomass burning emissions were 2019 FINN database[35]. Soil NOx emissions used were 2018 CAMS v2.1 emissions with canopy reduction factor[36]. Meteorology used in this H-CMAQ run was 2019 WRF v4.1.1.[37]

4.3 CMAQ Model Performance Evaluation

An operational model performance evaluation for ozone and $PM_{2.5}$ and its related speciated components was conducted for the 2019 simulation using state/local monitoring sites data in order to estimate the ability of the CMAQ modeling system to replicate the 2019 base year concentrations for the 12-km continental U.S. domain.

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

---

[30] Mathur, R., Xing, J., Gilliam, R., Sarwar, G., Hogrefe, C., Pleim, J., Pouliot, G., Roselle, S., Spero, T. L., Wong, D. C., and Young, J.: Extending the Community Multiscale Air Quality (CMAQ) modeling system to hemispheric scales: overview of process considerations and initial applications, Atmos Chem Phys, 17, 12449-12474, 10.5194/acp-17-12449-2017, 2017.
[31] 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.
[32] 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.
[33] 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.
[34] East, J.D., Henderson, B. H., Napelenok, S. L., Koplitz, S. N., Sarwar, G., Gilliam, R., Lenzen, A., Tong, D., Pierce, R. B., Garcia-Menendez, F. Comparing OMI and TROPOMI NO2 data assimilation for estimating NOx emissions. In preparation. 2022.
[35] Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geosci. Model Dev., 4, 625-641, 10.5194/gmd-4-625-2011, 2011.
[36] Simpson, D.: Soil N emissions for 2000-present. (D81.3.6.1.) [dataset], 2018.
[37] Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., Liu, Z. Q., Snyder, C., Chen, F., Barlage, M. J., Yu, W., and Duda, M. G.: THE WEATHER RESEARCH AND FORECASTING MODEL Overview, System Efforts, and Future Directions, B Am Meteorol Soc, 98, 1717-1737, 10.1175/Bams-D-15-00308.1, 2017.

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:

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

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

$$\mathrm{ME} = \frac{1}{n}\sum_1^n |P - O|$$

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:

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

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

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

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[38] 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[39,40] as were originally identified in Karl and Koss (1984).[41]

---

[38] 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.

[39] 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.

[40] 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.

[41] Karl, T. R. and Koss, W. J., 1984: "Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983." Historical Climatology Series 4-3, National Climatic Data Center, Asheville, NC, 38 pp.

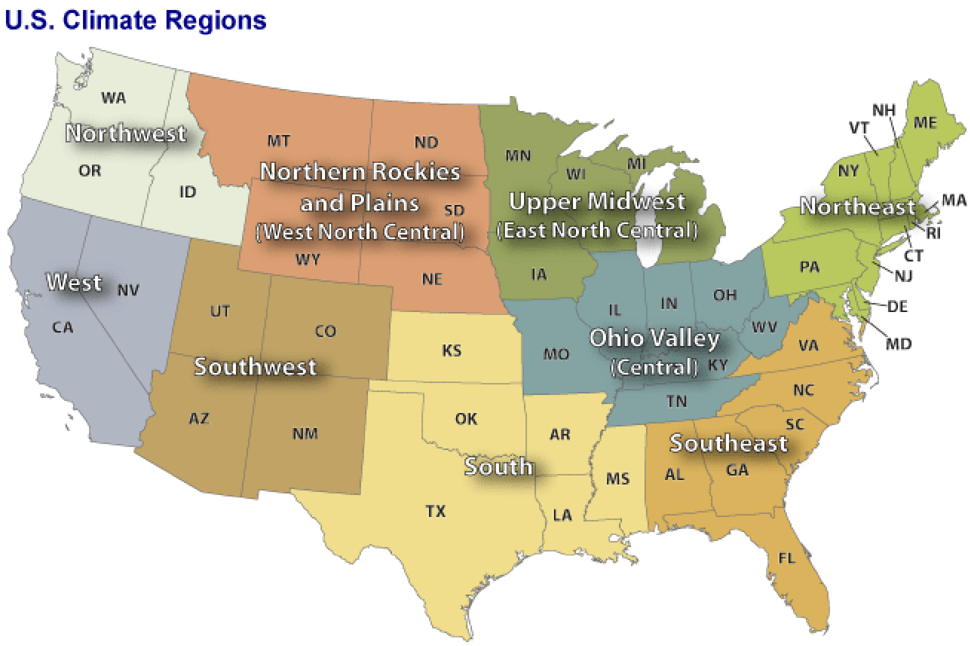


Figure 4-2. NOAA Nine Climate Regions (source: http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php#references)

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

Evaluation for 8-hour Daily Maximum Ozone: The operational model performance evaluation for eight-hour daily maximum ozone was conducted using the statistics defined above. Ozone measurements in the continental U.S. were included in the evaluation and were taken from the 2019 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 2019 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 0.6 to 8.5 percent) except in the Southwest, Northern Rockies, West. and in the 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 1.4 to 13.8 percent) except in Southwest and West as well as in the Ohio Valley, Southeast at CASTNet sites only. However, 8-hour ozone concentrations in the winter and spring are under predicted at AQS and CASTNet sites in all NOAA climate regions (with NMBs less than approximately 20 percent in each subregion) except in the winter at AQS sites in the Ohio Valley and Upper Midwest (slight overprediction of NMB ranging between 3.9 and 4.5 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 ±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 ± 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 2019 8-Hour Daily Maximum Ozone Model Performance Statistics by NOAA climate region, by Season and Monitoring Network.

Climate region	Monitor		No. of	MB	ME	NMB	NME
Northeast	AQS	Winter	10,839	-0.5	4.1	-1.6	12.5
		Spring	16,565	-5.0	6.6	-11.4	15.0
		Summer	16,721	1.6	5.7	3.7	13.1
		Fall	13,991	1.5	4.6	4.4	13.4
	CASTNet	Winter	1,267	-2.0	4.0	-5.8	11.2
		Spring	1,287	-6.3	7.2	-13.9	15.7
		Summer	1,229	0.8	5.5	1.9	13.2
		Fall	1,221	0.9	4.5	2.7	12.9
Ohio Valley	AQS	Winter	5,901	1.4	4.6	4.5	15.2
		Spring	20,951	-3.8	6.1	-8.6	13.8
		Summer	20,720	1.0	4.6	2.3	12.3
		Fall	15,591	1.1	4.4	2.8	11.8
	CASTNet	Winter	1,574	-0.1	4.5	-0.2	13.9
		Spring	1,602	-5.2	7.1	-11.1	15.3
		Summer	1,584	0.2	5.1	0.6	11.5
		Fall	1,537	-0.6	4.4	-1.7	11.5
Upper Midwest	AQS	Winter	2,042	1.3	4.1	3.9	12.5
		Spring	8,297	-4.1	5.8	-9.6	13.4
		Summer	8,858	1.2	5.7	2,8	13.8
		Fall	6,225	4.2	5.5	13.8	18.0

Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
	CASTNet	Winter	418	-0.9	3.6	-2.5	10.5
		Spring	449	-6.6	7.2	-14.8	16.2
		Summer	438	0.4	5.6	1.1	14.3
		Fall	433	3.4	4.7	11.3	15.6
Southeast	AQS	Winter	6,698	-2.2	4.7	-6.3	12.9
		Spring	15,470	-4.2	6.6	-9.2	14.3
		Summer	14,708	3.4	6.3	8.5	15.7
		Fall	12,167	0.9	4.5	2.3	11.6
	CASTNet	Winter	891	-3.4	5.4	-9.2	14.6
		Spring	958	-5.9	7.3	-12.5	15.4
		Summer	938	1.9	5.7	4.6	14.0
		Fall	970	-3.0	6.4	-7.2	15.2
South	AQS	Winter	10,537	-0.9	5.3	-2.9	16.5
		Spring	12,498	-1.7	7.5	-3.8	17.2
		Summer	12,313	2.5	6.5	6.1	16.1
		Fall	11,844	0.7	5.6	1.9	14.8
	CASTNet	Winter	502	-1.3	4.6	-3.7	13.4
		Spring	523	-3.2	7.4	-7.0	16.5
		Summer	507	0.6	6.1	1.4	14.8
		Fall	515	0.5	4.7	1.4	12.7
Southwest	AQS	Winter	9,586	-2.4	5.9	-6.2	15.3
		Spring	10,650	-7.3	8.1	-14.3	15.8
		Summer	10,695	-3.1	6.1	-5.7	11.2
		Fall	10,550	-1.4	4.8	-3.2	11.0
	CASTNet	Winter	793	-5.1	6.6	-11.5	14.8
		Spring	795	-9.0	9.4	-16.8	17.6
		Summer	805	-2.0	5.0	-3.8	9.6
		Fall	792	-2.5	4.3	-5.4	9.3
Northern Rockies	AQS	Winter	4,524	-0.7	4.8	-1.9	12.5
		Spring	4,850	-5.8	6.9	-12.4	14.8
		Summer	4,891	-0.6	4.9	-1.4	10.6
		Fall	4,745	2.0	4.4	5.5	12.2

Climate region	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
	CASTNet	Winter	695	-1.3	5.5	-3.3	13.9
		Spring	702	-6.8	7.4	-14.1	15.4
		Summer	688	-0.6	4.6	-1.4	10.0
		Fall	705	1.3	4.5	3.5	11.9
Northwest	AQS	Winter	644	-0.8	5.1	-2.6	16.1
		Spring	1,337	-4.6	7.9	-11.6	19.8
		Summer	2,411	1.3	5.9	3.5	16.2
		Fall	1,204	3.7	6.0	12.4	20.0
	CASTNet	Winter	83	-0.6	3.9	-1.8	11.4
		Spring	85	-6.4	6.9	-15.0	16.1
		Summer	88	-1.7	4.1	-4.1	9.7
		Fall	90	2.8	4.4	8.9	14.2
West	AQS	Winter	14,058	-2.2	5.2	-6.5	15.4
		Spring	16,167	-6.8	7.7	-14.9	17.0
		Summer	17,106	-3.7	7.0	-7.3	14.0
		Fall	15,390	-3.9	6.7	-8.6	14.9
	CASTNet	Winter	536	-3.4	5.4	-8.5	13.6
		Spring	565	-7.9	8.5	-16.3	17.5
		Summer	626	-8.0	8.9	-14.0	15.7
		Fall	602	-5.8	7.2	-11.9	14.8

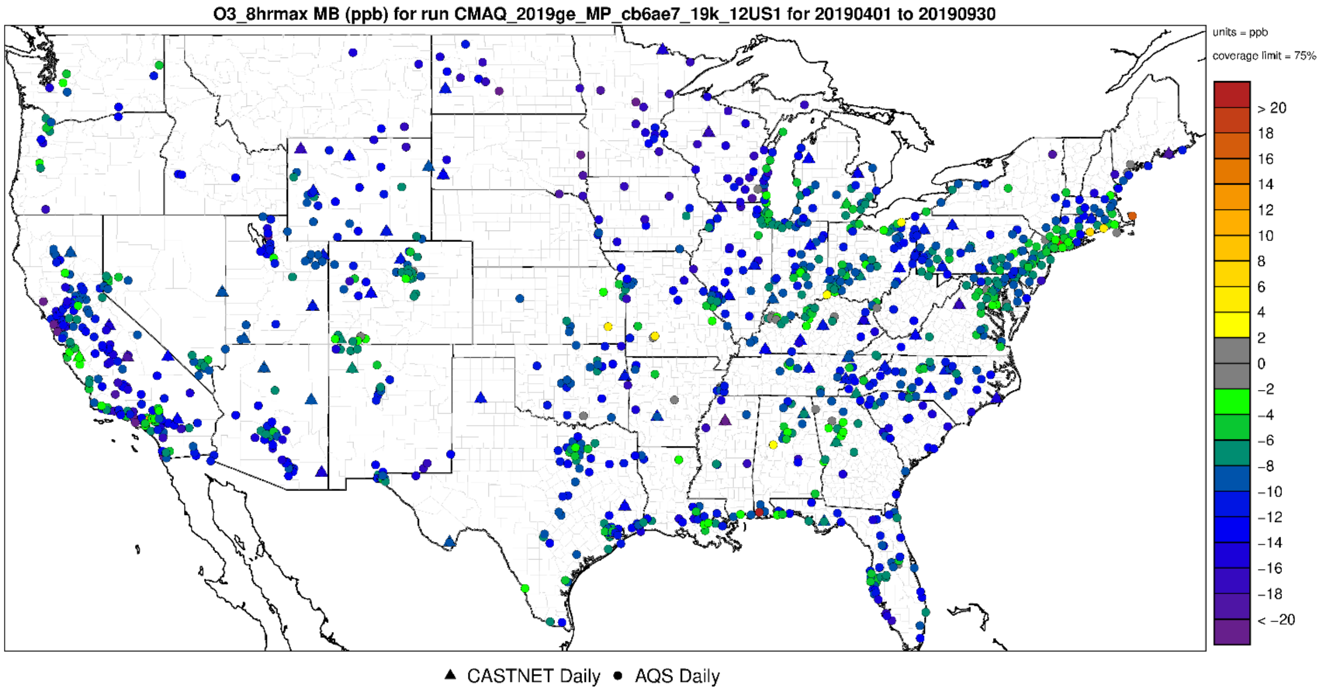


Figure 4-3. Mean Bias (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period April-September 2019 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

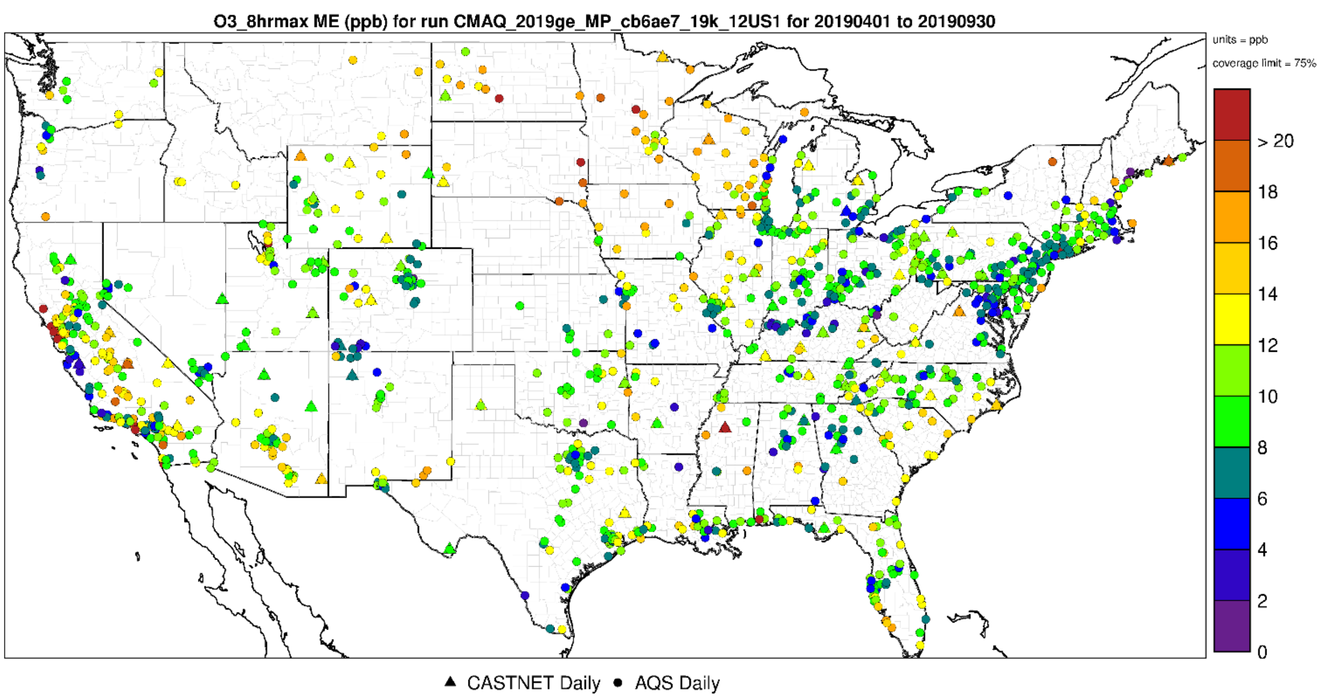


Figure 4-4. Mean Error (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period April-September 2019 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

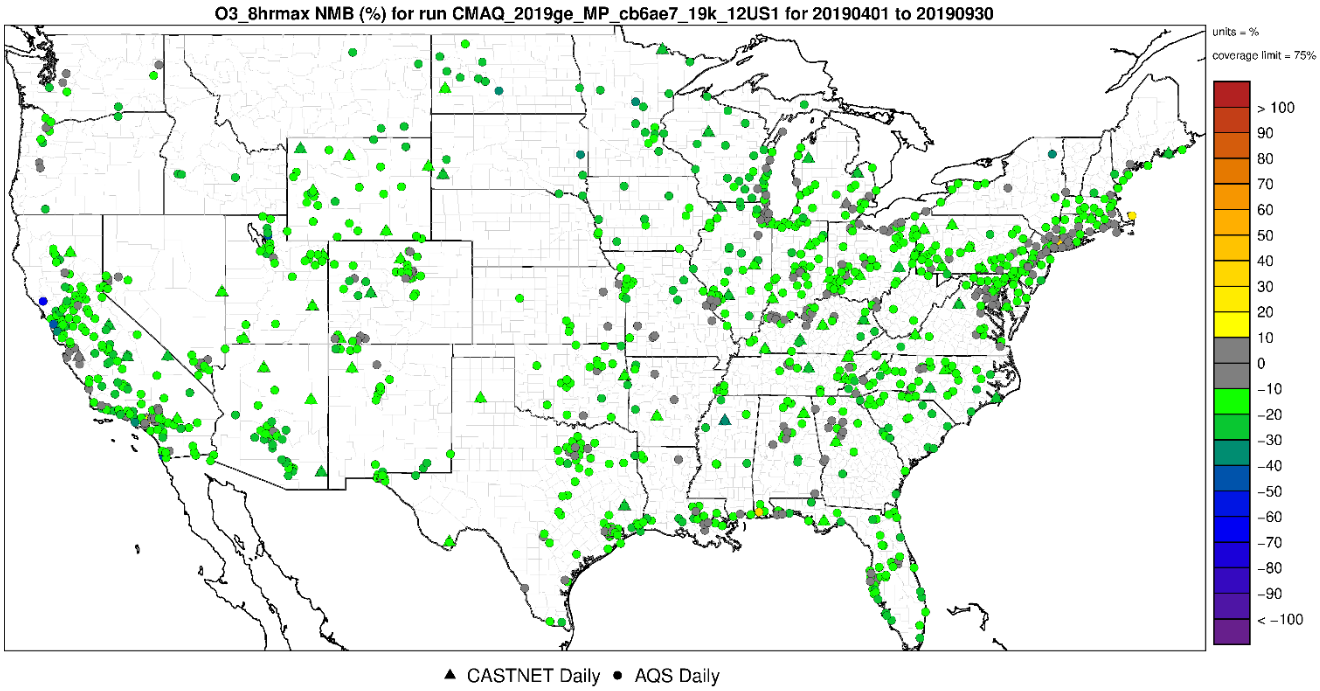


Figure 4-5. Normalized Mean Bias (%) of 8-hour daily maximum ozone greater than 60 ppb over the period April-September 2019 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

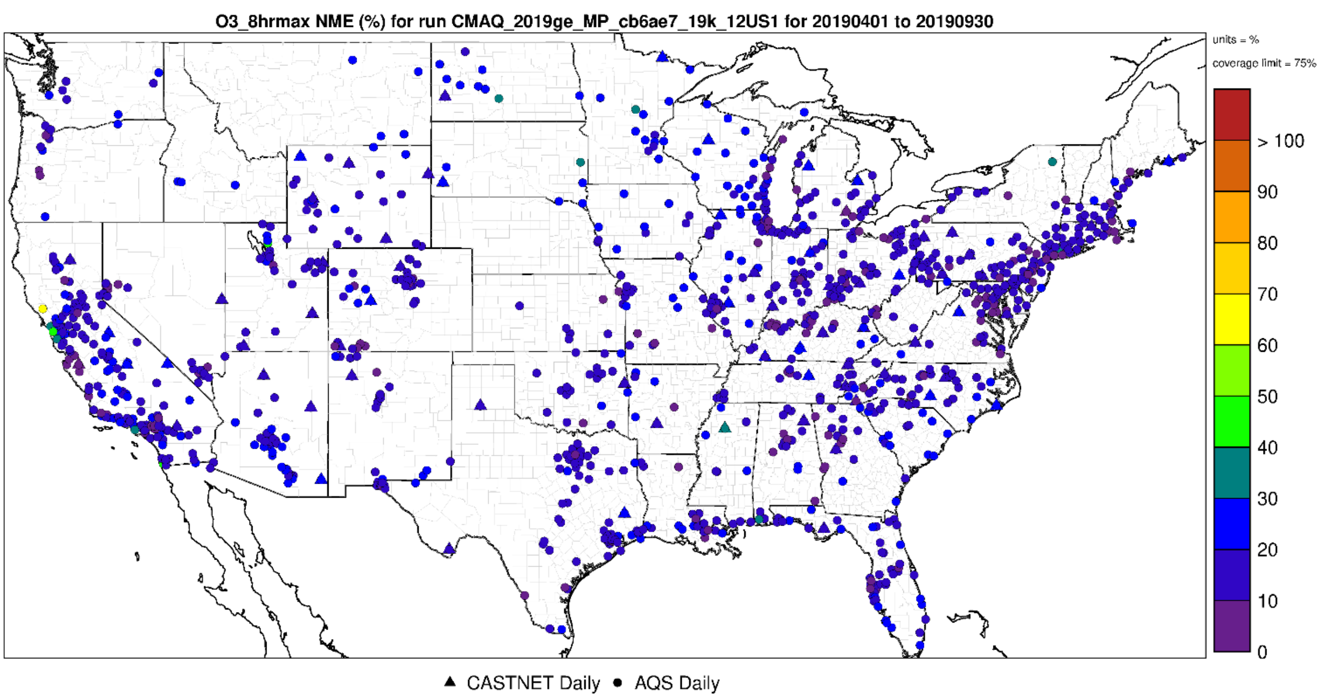


Figure 4-6. Normalized Mean Error (%) of 8-hour daily maximum ozone greater than 60 ppb over the period April-September 2019 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

Evaluation for Annual $PM_{2.5}$ components: The PM evaluation focuses on $PM_{2.5}$ components including sulfate (SO_4), nitrate (NO_3), total nitrate ($TNO_3 = NO_3 + HNO_3$), ammonium (NH_4), elemental carbon (EC), and organic carbon (OC). The bias and error performance statistics were calculated on an annual basis for each of the nine NOAA climate subregions defined above (provided in Table 4-5). $PM_{2.5}$ measurements for 2019 were obtained from the following networks for model evaluation: Chemical Speciation Network (CSN, 24-hour average), Interagency Monitoring of Protected Visual Environments (IMPROVE, 24-hour average, and Clean Air Status and Trends Network (CASTNet, weekly average). For $PM_{2.5}$ species that are measured by more than one network, we calculated separate sets of statistics for each network by subregion. In addition to the tabular summaries of bias and error statistics, annual spatial maps which show the mean bias, mean error, normalized mean bias and normalized mean error by site for each $PM_{2.5}$ species are provided in Figures 4-7 through 4-30.

As indicated by the statistics in Table 4-5, annual average sulfate is consistently under predicted at CASTNet, IMPROVE, and CSN monitoring sites across the 12-km modeling domain (with MB values ranging from -0.0 to -0.4 μgm^{-3}) except at CSN and IMPROVE sites in the Upper Midwest, Northern Rockies and Northwest and at IMPROVE sites in the Northeast, as well as at CASTNet sites in the Northwest (MB approximately 0.1 μgm^{-3}). Sulfate performance shows moderate error in the eastern subregions (average of approximately 30 percent) while Western subregions show slightly larger error (ranging from 30 to 60 percent). Figures 4-7 through 4-10, suggest spatial patterns vary by region. The model bias for most of the Northeast, Southeast, Central and Southwest states are within ±30 percent. The model bias appears to be slightly greater in the Northwest and in the Upper Midwest with over predictions up to approximately 60 percent at individual monitors. Model error also shows a spatial trend by region, where much of the Eastern states are 10 to 40 percent, the Western and Central U.S. states are 20 to 100 percent.

Annual average nitrate is under predicted at the urban CSN monitoring sites at all NOAA climate subregions (NMB in the range of -1 to -34 percent), except in the Northwest where nitrate is over predicted (62.3 percent). At IMPROVE rural sites, annual average nitrate is under predicted at all subregions, except in the Northeast (18.6 percent) and Northeast (14.0 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). 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 Southeast as indicated in Figure 4-13. The exception to this is in the Ohio Valley, South, 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 -22.3 to -59.2 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 in the Northwest (over prediction of 28.7 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 -9.2 to 50.5 percent) except in the Northwest (over prediction ranging 19.0 to 41.7 percent). There is not a large variation in error statistics from subregion to subregion or at urban versus rural sites.

Annual average organic carbon is over predicted across most subregions in rural IMPROVE areas (NMB ranging from 4.1to 65.5 percent), except in the Southwest, Northern Rockies and West where the NMB ranges from -16.0 to -25.4 percent. The model over predicted annual average organic carbon in all subregions at urban CSN sites except in the Northern Rockies (NMB approximately -31 percent). Similar to elemental carbon, error model performance does not show a large variation from subregion to subregion or at urban versus rural sites.

Table 4-5. Summary of CMAQ 2019 Annual PM Species Model Performance Statistics by NOAA Climate region, by Monitoring Network.

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
Sulfate	CSN	Northeast	3,118	0.0	0.3	-5.8	35.2
		Ohio Valley	2,175	-0.1	0.4	-10.8	33.3
		Upper Midwest	1,156	0.1	0.3	12.4	38.8
		Southeast	2,134	-0.1	0.3	-13.8	33.8
		South	1,334	-0.2	0.5	-15.4	38.5
		Southwest	1,111	-0.1	0.2	-8.8	41.4
		Northern Rockies	633	0.1	0.3	17.6	46.8
		Northwest	776	0.1	0.1	22.3	49.7
		West	1,484	-0.3	0.5	-32.1	50.8
	IMPROVE	Northeast	1,888	0.0	0.3	3.3	13.8
		Ohio Valley	925	-0.2	0.3	-16.2	32.9
		Upper Midwest	933	0.0	0.2	6.5	37.2
		Southeast	1,500	-0.2	0.4	-22.8	35.4
		South	948	-0.2	0.4	-18.9	37.1
		Southwest	3,599	-0.0	0.2	-6.9	43.9
		Northern Rockies	2,152	0.0	0.2	11.0	42.0
		Northwest	1,841	0.1	0.2	35.7	62.7
		West	2,432	-0.0	0.2	-6.0	51.7
	CASTNet	Northeast	876	-0.1	0.2	-14.2	25.3
		Ohio Valley	909	-0.3	0.3	-24.4	26.5
		Upper Midwest	248	-0.0	0.2	-2.8	23.2
		Southeast	587	-0.4	0.4	-33.5	36.1
		South	381	-0.3	0.4	-27.7	30.9
		Southwest	426	-0.0	0.1	-8.1	36.0
		Northern Rockies	497	-0.0	0.1	-1.1	30.1
		Northwest	49	0.1	0.1	20.7	38.5
		West	270	-0.2	0.3	-26.9	47.3
Nitrate	CSN	Northeast	3,117	-0.3	0.5	-28.7	47.2
		Ohio Valley	2,175	-0.5	0.7	-34.9	46.9
		Upper Midwest	1,156	-0.6	0.8	-36.0	47.8
		Southeast	2,150	-0.0	0.3	-6.6	70.0
		South	1,334	-0.3	0.4	-40.7	62.9
		Southwest	1,110	-0.8	0.9	-64.9	70.4
		Northern Rockies	632	-0.5	0.6	-49.2	56.6
		Northwest	776	0.5	1.1	62.3	>100
		West	1,483	-1.2	1.3	-63.1	66.5
	IMPROVE	Northeast	1,888	0.1	0.2	18.6	72.1

Pollutant	Monitor Network	Subregion	No. of Obs	MB (µgm^{-3})	ME (µgm^{-3})	NMB (%)	NME (%)
		Ohio Valley	925	-0.3	0.4	-41.8	61.9
		Upper Midwest	931	-0.3	0.4	-42.5	52.0
		Southeast	1,500	-0.1	0.2	-26.5	73.3
		South	948	-0.2	0.3	-41.7	63.8
		Southwest	3,597	-0.2	0.2	-78.9	83.8
		Northern Rockies	2,151	-0.2	0.2	-58.6	77.7
		Northwest	1,833	0.0	0.2	14.0	>100
		West	2,431	-0.2	0.2	-47.1	67.7
Total Nitrate (NO_3+HNO_3)	CASTNet	Northeast	875	-0.1	0.4	-10.4	31.1
		Ohio Valley	909	-0.3	0.6	-19.1	32.1
		Upper Midwest	248	-0.4	0.5	-26.7	36.0
		Southeast	586	-0.2	0.4	-24.4	43.7
		South	381	-0.3	0.5	-27.8	38.1
		Southwest	426	-0.3	0.3	-41.4	44.9
		Northern Rockies	497	-0.2	0.2	33.0	39.7
		Northwest	--	--	--	--	--
		West	270	-0.5	0.5	-53.3	55.1
Ammonium	CSN	Northeast	3,118	-0.0	0.2	-4.4	47.1
		Ohio Valley	2,175	-0.1	0.3	-16.0	42.4
		Upper Midwest	1,156	-0.0	0.2	-4.6	44.1
		Southeast	2,130	-0.1	0.2	-21.3	59.3
		South	1,333	-0.1	0.2	-27.9	53.3
		Southwest	1,111	-0.2	0.3	-58.7	68.3
		Northern Rockies	632	-0.0	0.2	-0.8	52.1
		Northwest	774	0.1	0.3	28.7	>100
		West	1,482	-0.4	0.5	-66.8	73.8
	CASTNet	Northeast	876	-0.1	0.1	-22.3	33.2
		Ohio Valley	909	-0.2	0.2	-35.0	37.4
		Upper Midwest	248	-0.2	0.2	-30.4	34.8
		Southeast	587	-0.1	0.1	-33.6	42.1
		South	381	-0.1	0.2	-28.6	38.7
		Southwest	426	-0.1	0.1	-49.3	54.1
		Northern Rockies	497	-0.1	0.1	-41.2	45.7
		Northwest	49	-0.1	0.1	-53.9	58.2
		West	270	-0.1	0.1	-59.2	67.4
Elemental Carbon	CSN	Northeast	3,114	-0.2	0.3	-22.6	42.7
		Ohio Valley	2,179	-0.2	0.3	-33.7	43.0

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
		Upper Midwest	1,271	-0.1	0.2	-19.7	41.2
		Southeast	1,777	-0.3	0.3	-39.8	48.0
		South	1,163	-0.2	0.2	-26.6	38.6
		Southwest	1,112	-0.2	0.3	-22.9	44.6
		Northern Rockies	584	-0.2	0.2	-50.5	62.9
		Northwest	774	0.2	0.7	19.0	77.0
		West	1,253	-0.3	0.4	-40.3	44.8
	IMPROVE	Northeast	1,804	-0.0	0.1	-9.2	39.5
		Ohio Valley	925	-0.1	0.1	-31.8	40.6
		Upper Midwest	1,032	-0.1	0.1	-30.2	44.1
		Southeast	1,623	-0.1	0.2	-39.1	47.3
		South	935	-0.1	0.1	-36.3	44.8
		Southwest	3,537	-0.1	0.1	-43.6	58.2
		Northern Rockies	2,177	-0.0	0.1	-29.4	57.4
		Northwest	1,742	0.1	0.1	41.7	98.9
		West	2,242	-0.1	0.1	-35.0	52.4
Organic Carbon	CSN	Northeast	3,113	0.9	1.2	52.8	71.4
		Ohio Valley	2,179	0.4	0.8	23.0	47.6
		Upper Midwest	1,272	0.4	0.9	29.5	56.9
		Southeast	1,777	0.7	1.0	31.1	47.8
		South	1,163	0.6	1.1	32.6	58.8
		Southwest	1,112	0.4	0.9	26.4	58.7
		Northern Rockies	584	-0.4	0.7	-31.2	54.6
		Northwest	774	2.8	3.0	>100	>100
		West	1,253	0.2	0.9	8.2	40.5
	IMPROVE	Northeast	1,816	0.4	0.6	38.3	64.8
		Ohio Valley	925	0.3	0.7	28.6	54.6
		Upper Midwest	1,053	0.0	0.4	4.1	48.1
		Southeast	1,642	0.4	0.7	27.0	52.9
		South	939	0.2	0.6	17.6	51.2
		Southwest	3,588	-0.1	0.4	-16.0	59.3
		Northern Rockies	2,243	-0.1	0.4	-20.2	54.2
		Northwest	1,817	0.4	0.7	65.5	>100
		West	2,323	-0.2	0.4	-25.4	46.3

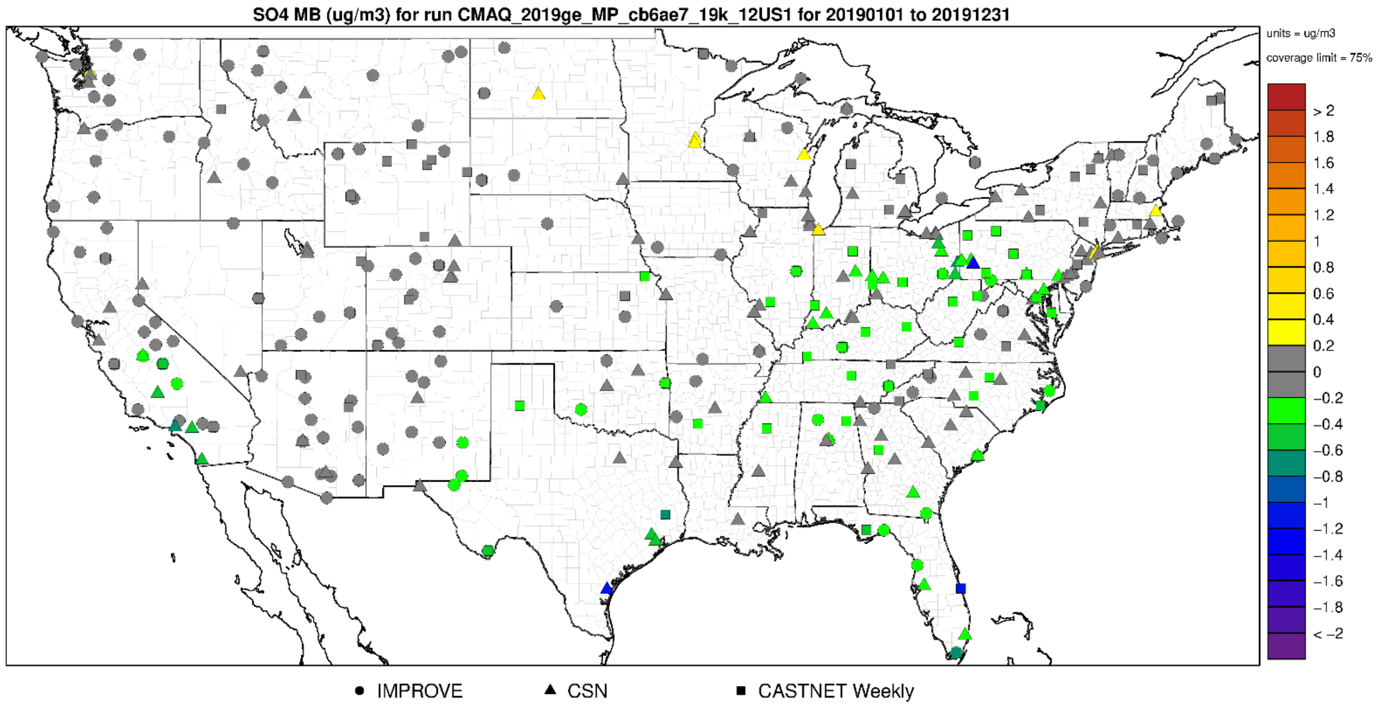


Figure 4-7. Mean Bias (µgm^{-3}) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

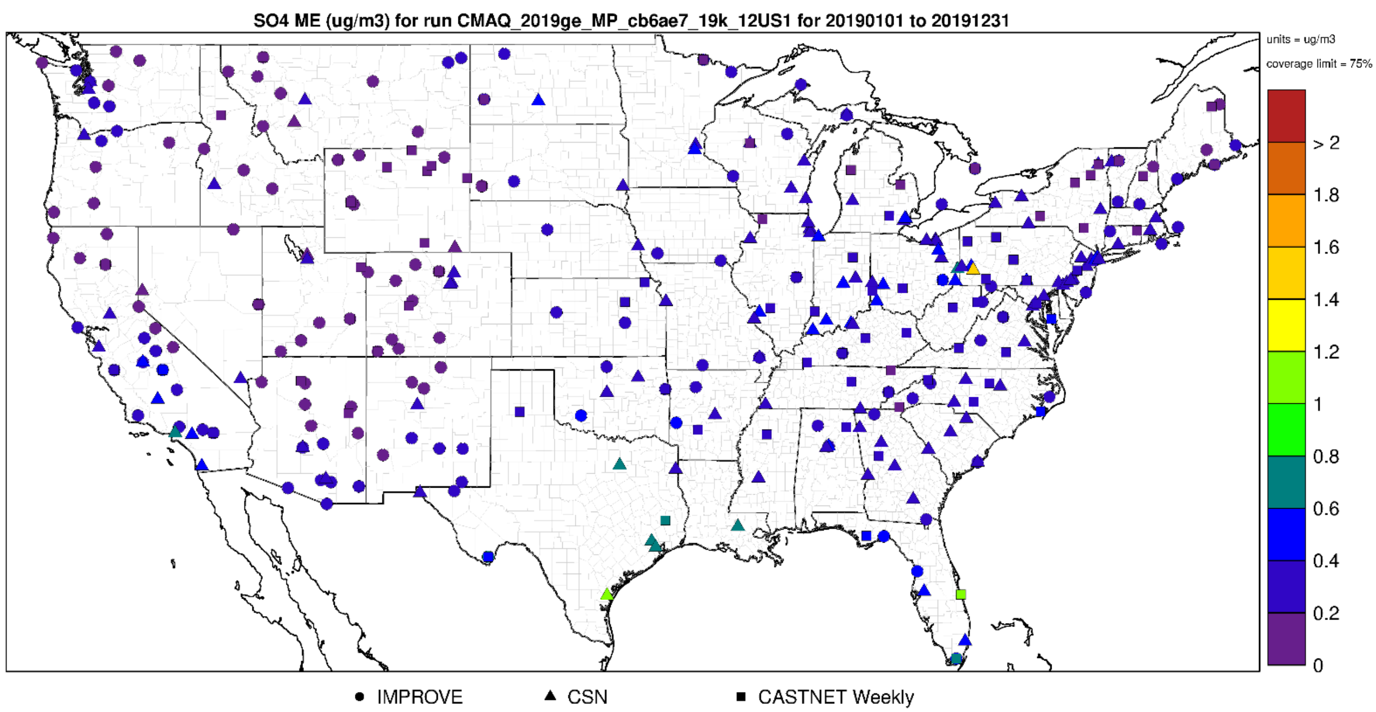


Figure 4-8. Mean Error (µgm^{-3}) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

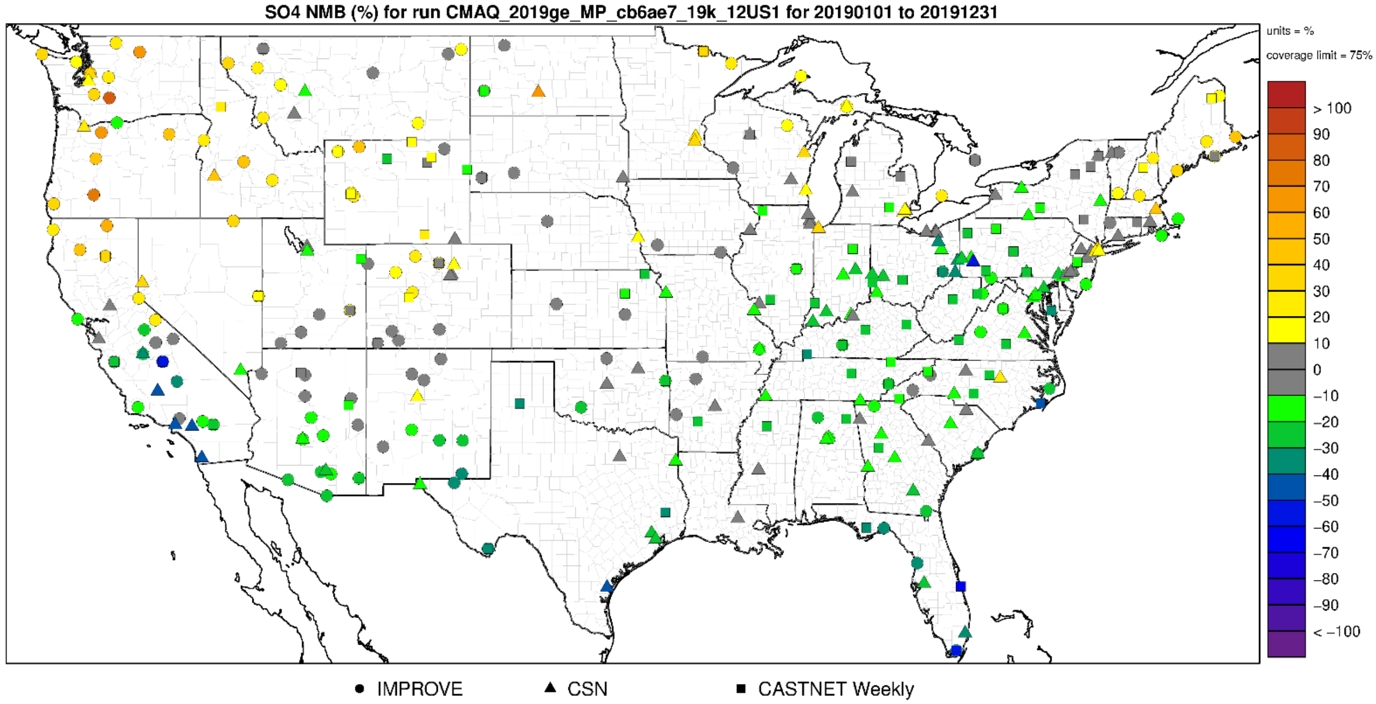


Figure 4-9. Normalized Mean Bias (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

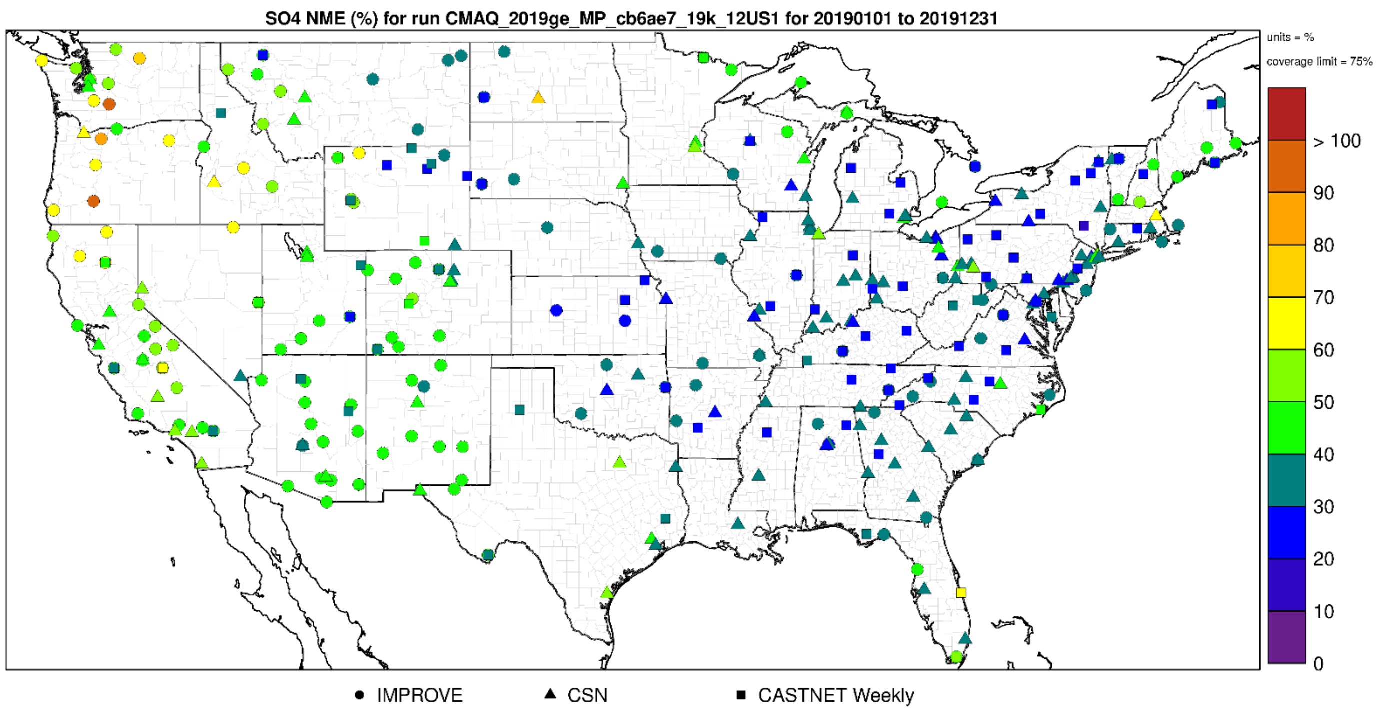


Figure 4-10. Normalized Mean Error (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

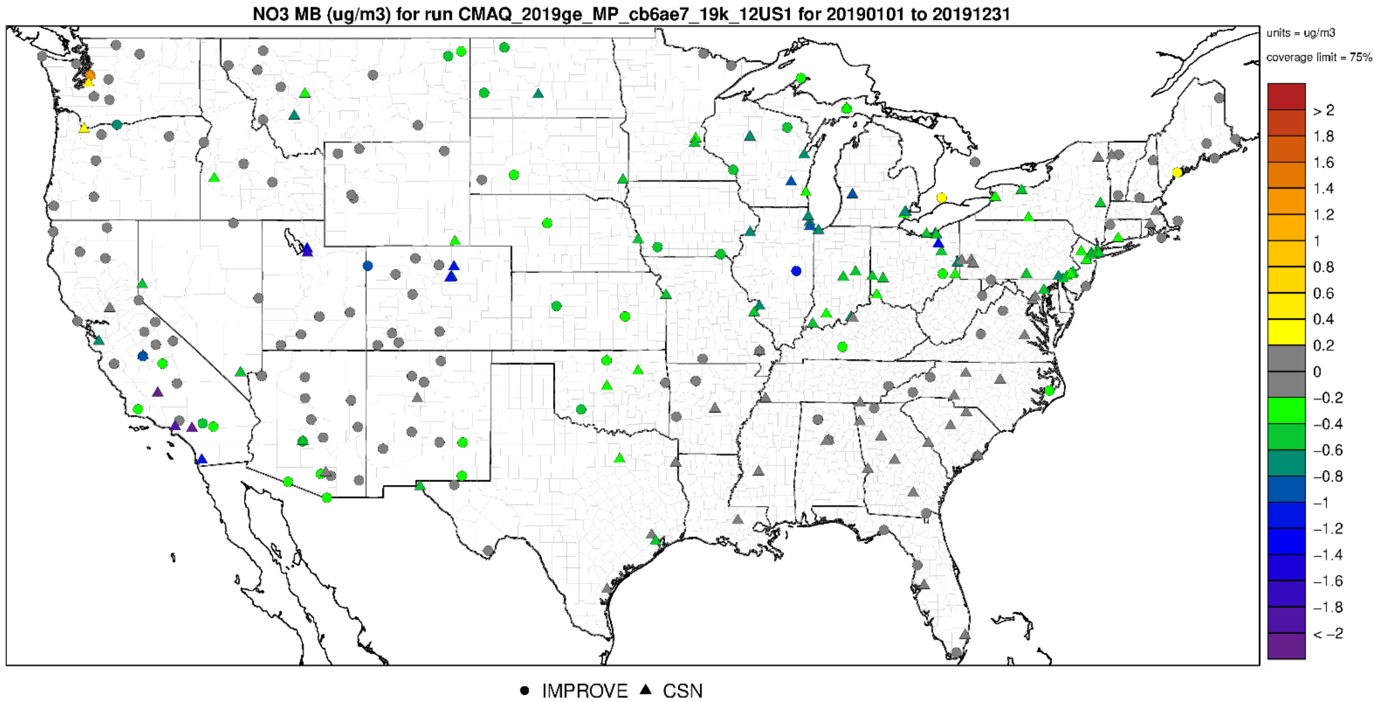


Figure 4-11. Mean Bias (µgm^{-3}) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

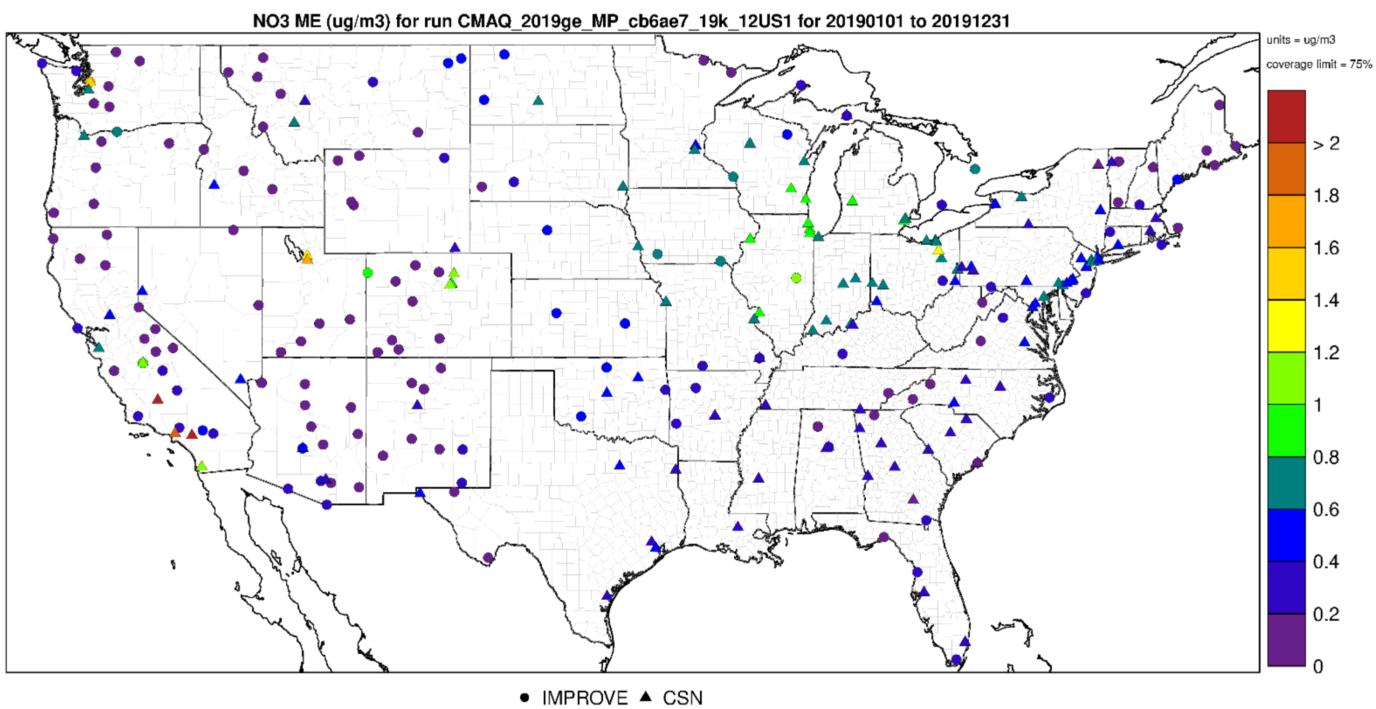


Figure 4-12. Mean Error (µgm^{-3}) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

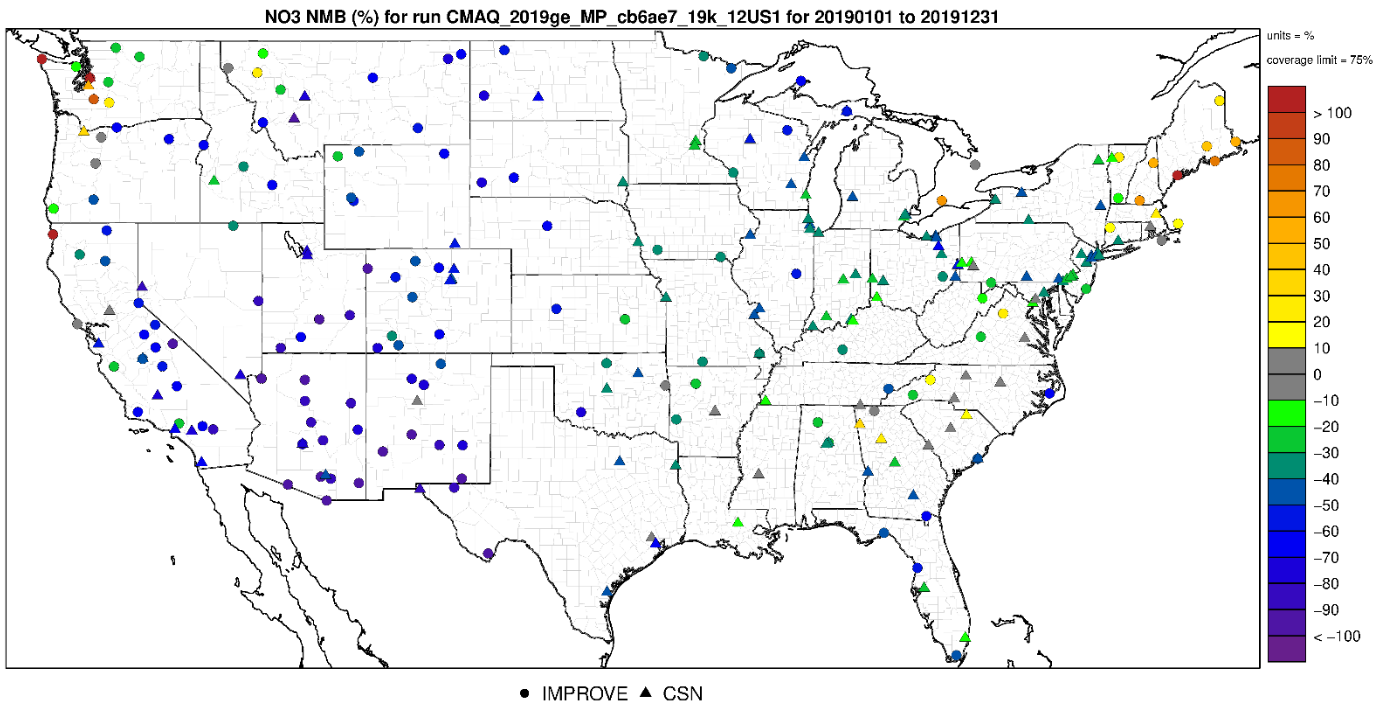


Figure 4-13. Normalized Mean Bias (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

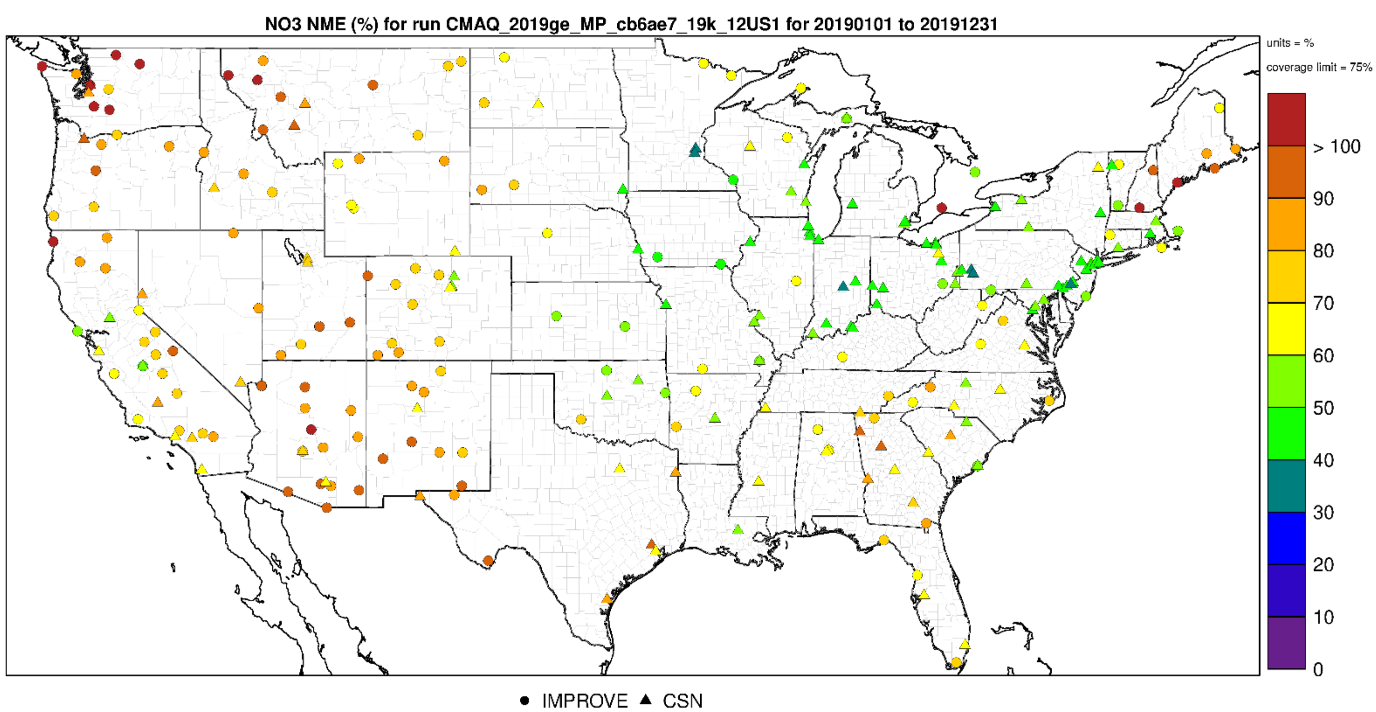


Figure 4-14. Normalized Mean Error (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

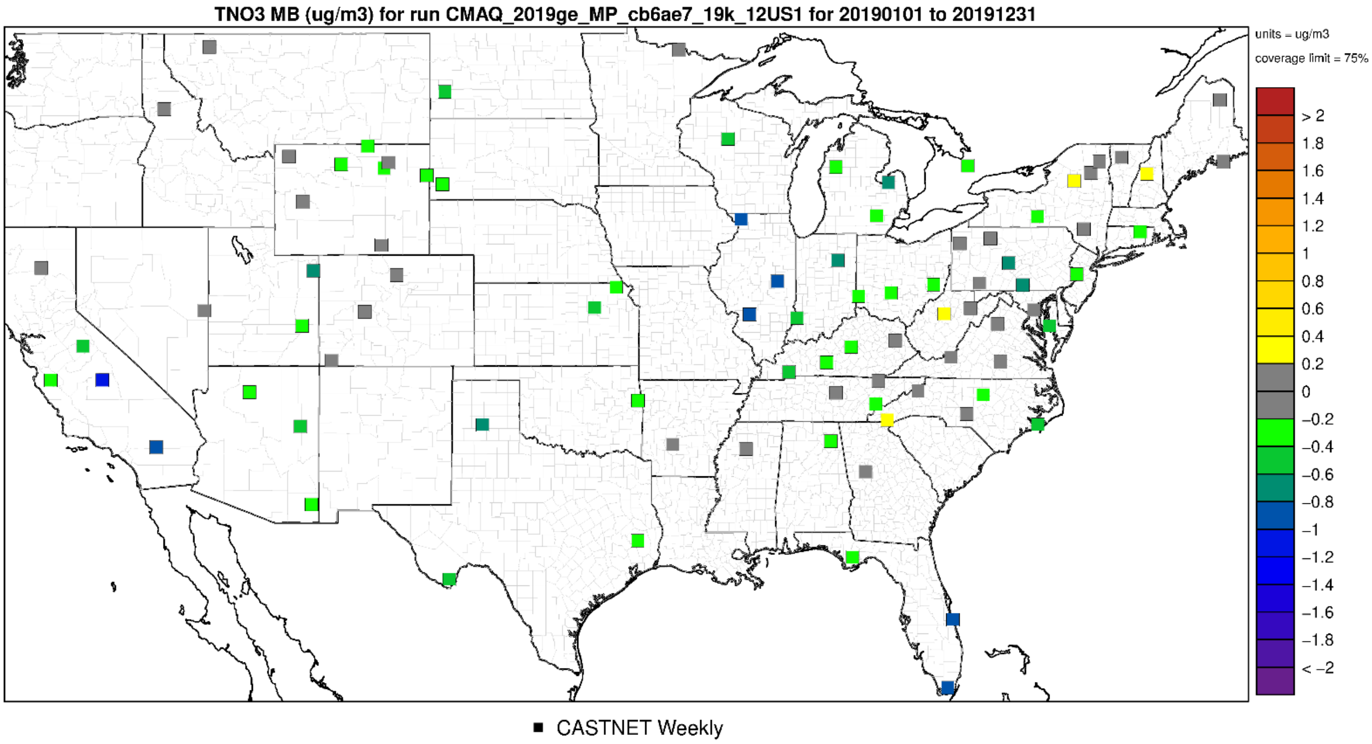


Figure 4-15. Mean Bias (µgm^{-3}) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

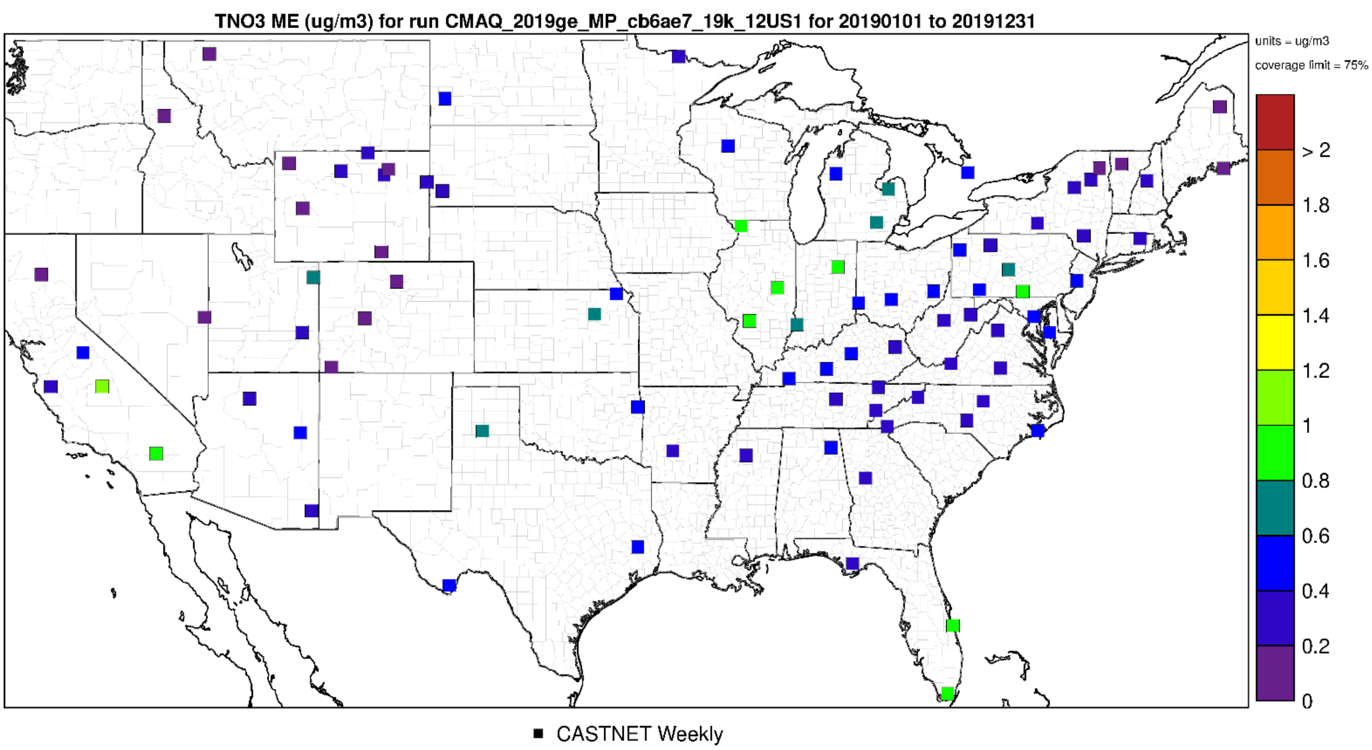


Figure 4-16. Mean Error (µgm^{-3}) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

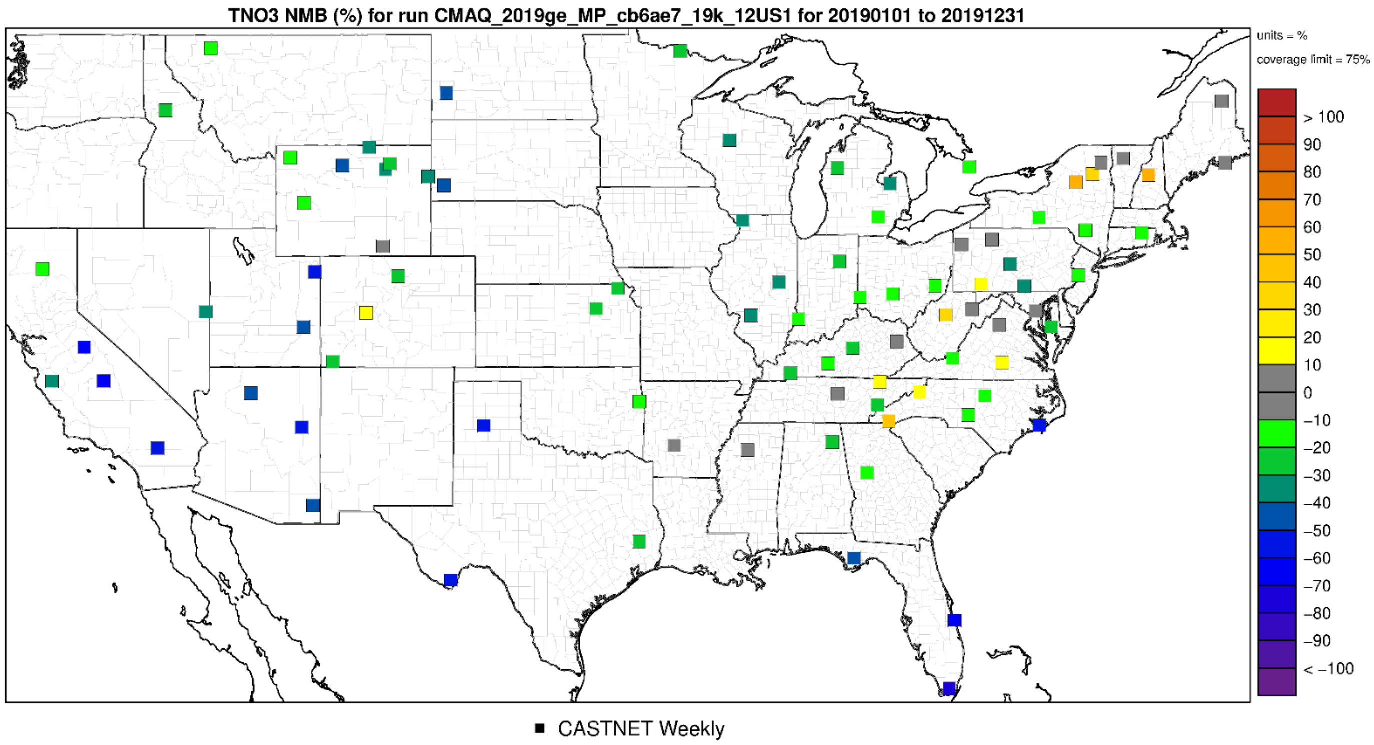


Figure 4-17. Normalized Mean Bias (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

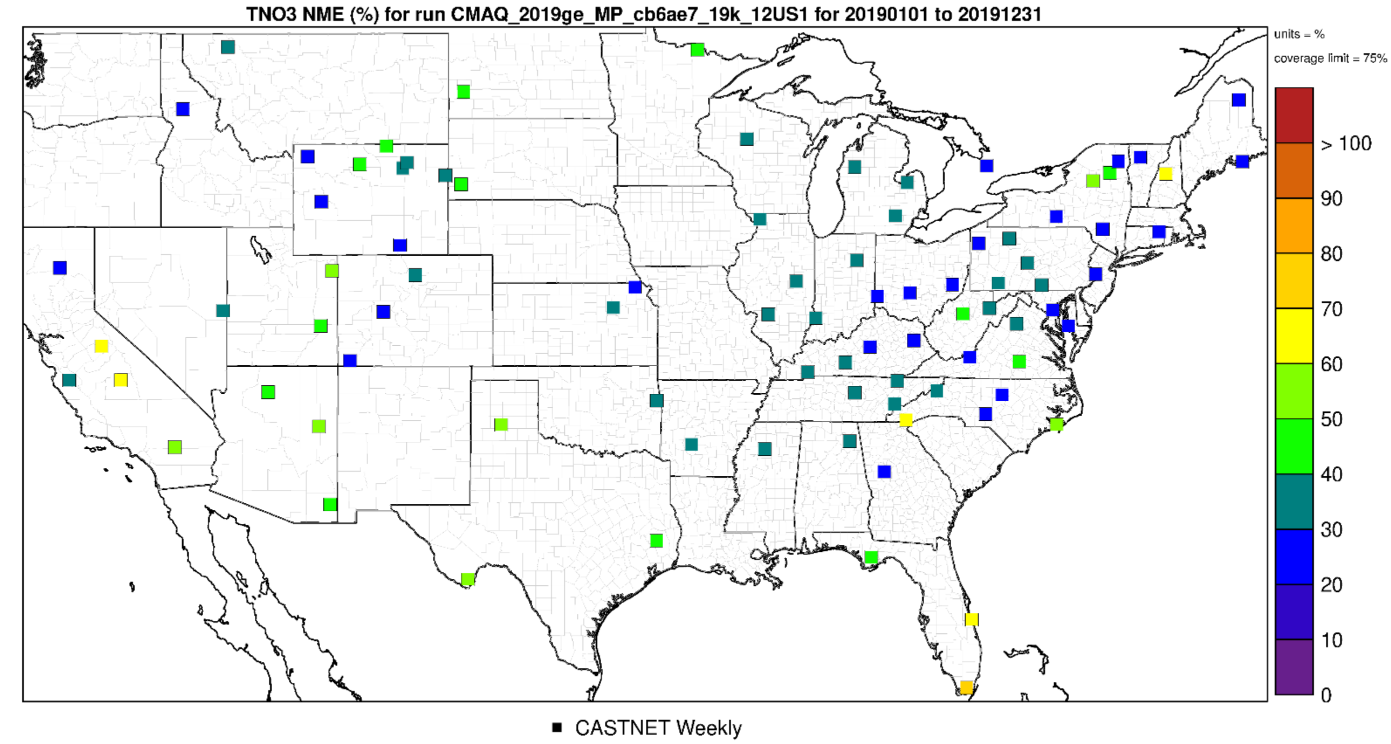


Figure 4-18. Normalized Mean Error (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

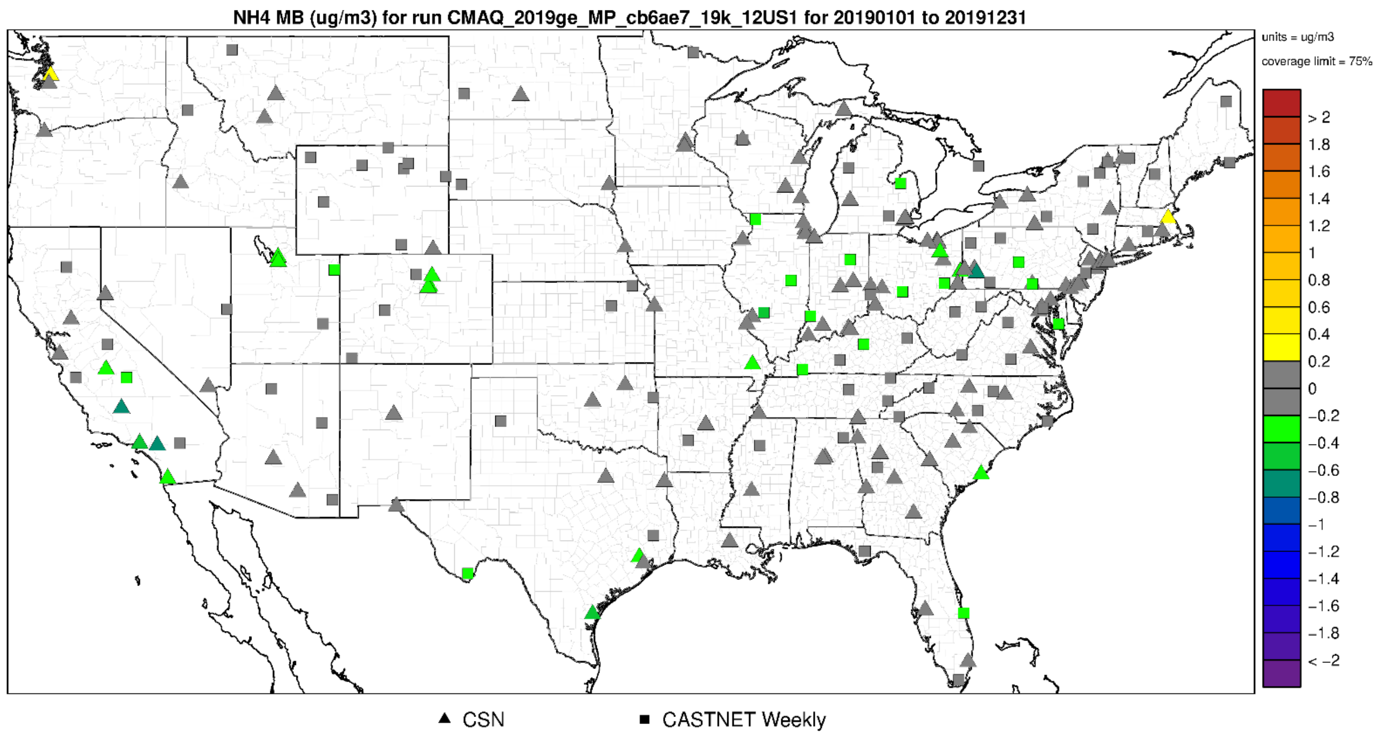


Figure 4-19. Mean Bias (µgm^{-3}) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

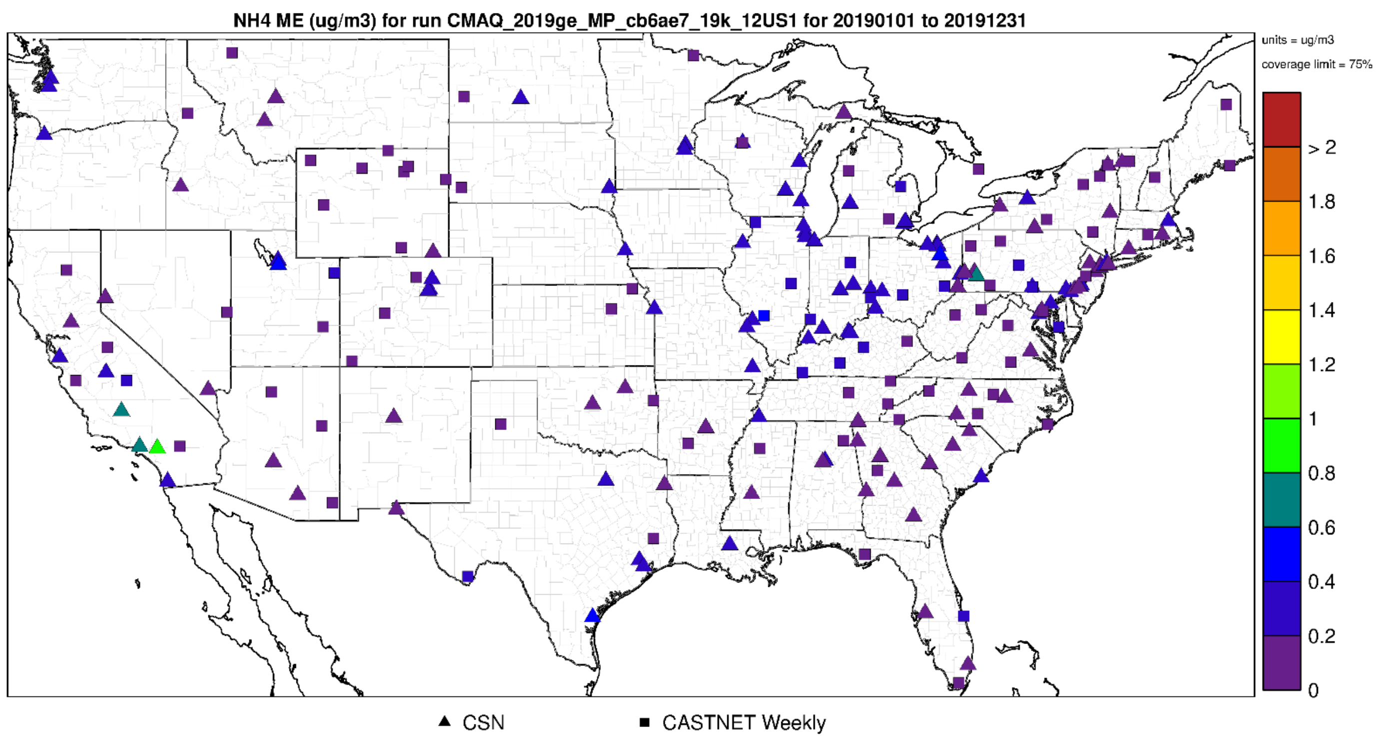


Figure 4-20. Mean Error (µgm^{-3}) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

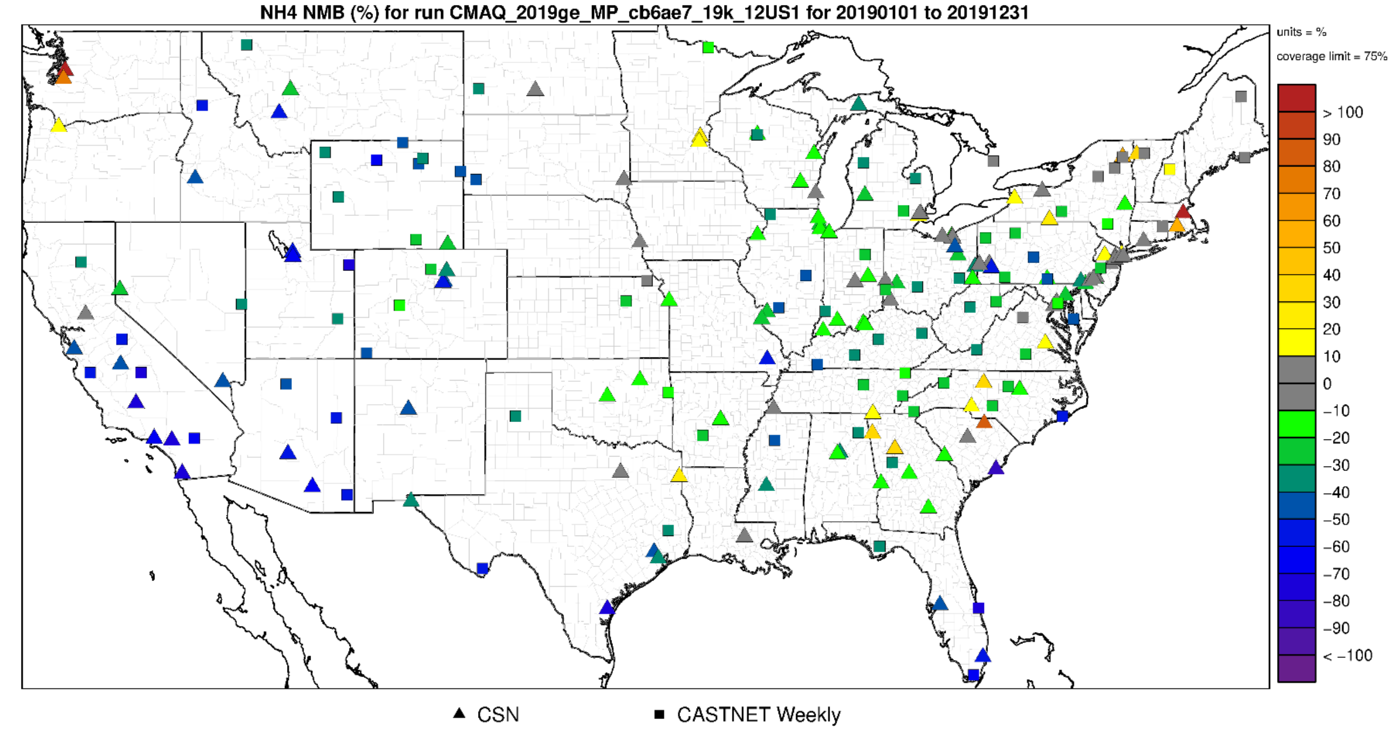


Figure 4-21. Normalized Mean Bias (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

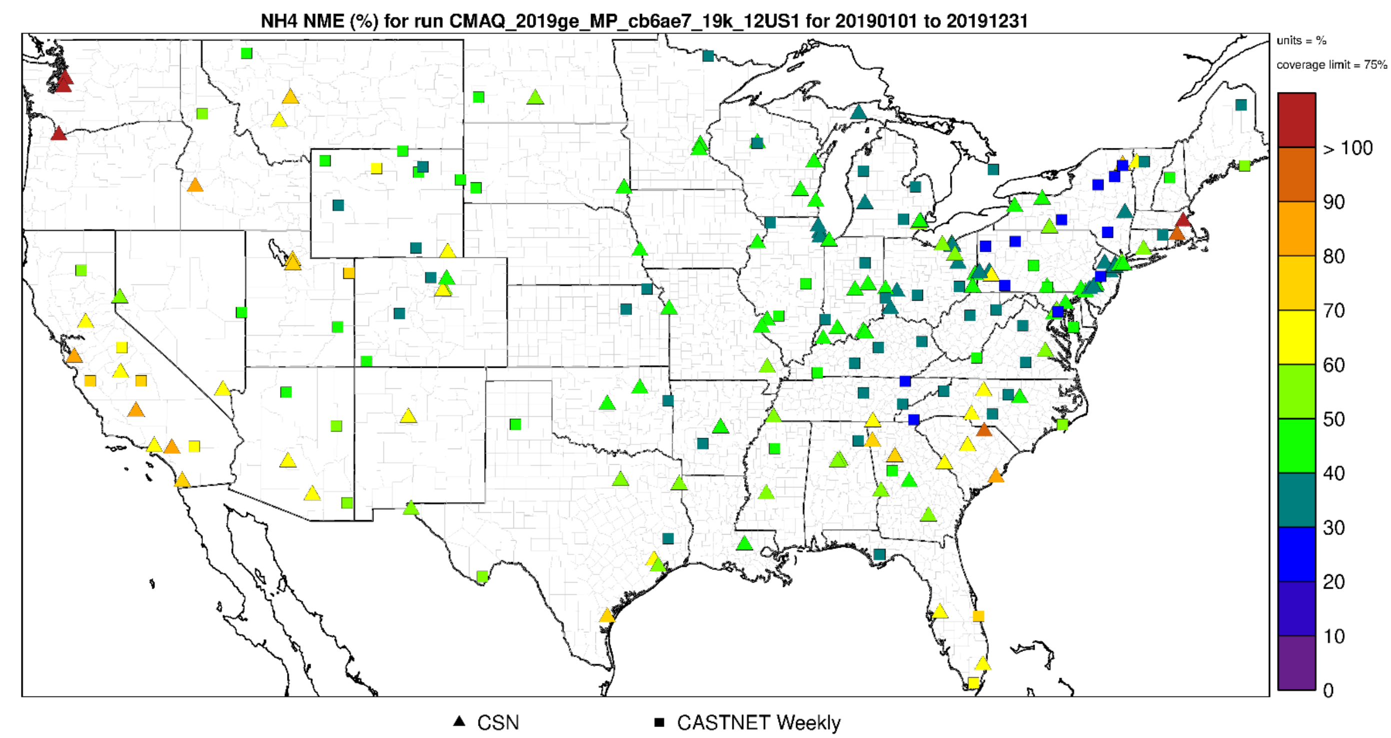


Figure 4-22. Normalized Mean Error (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

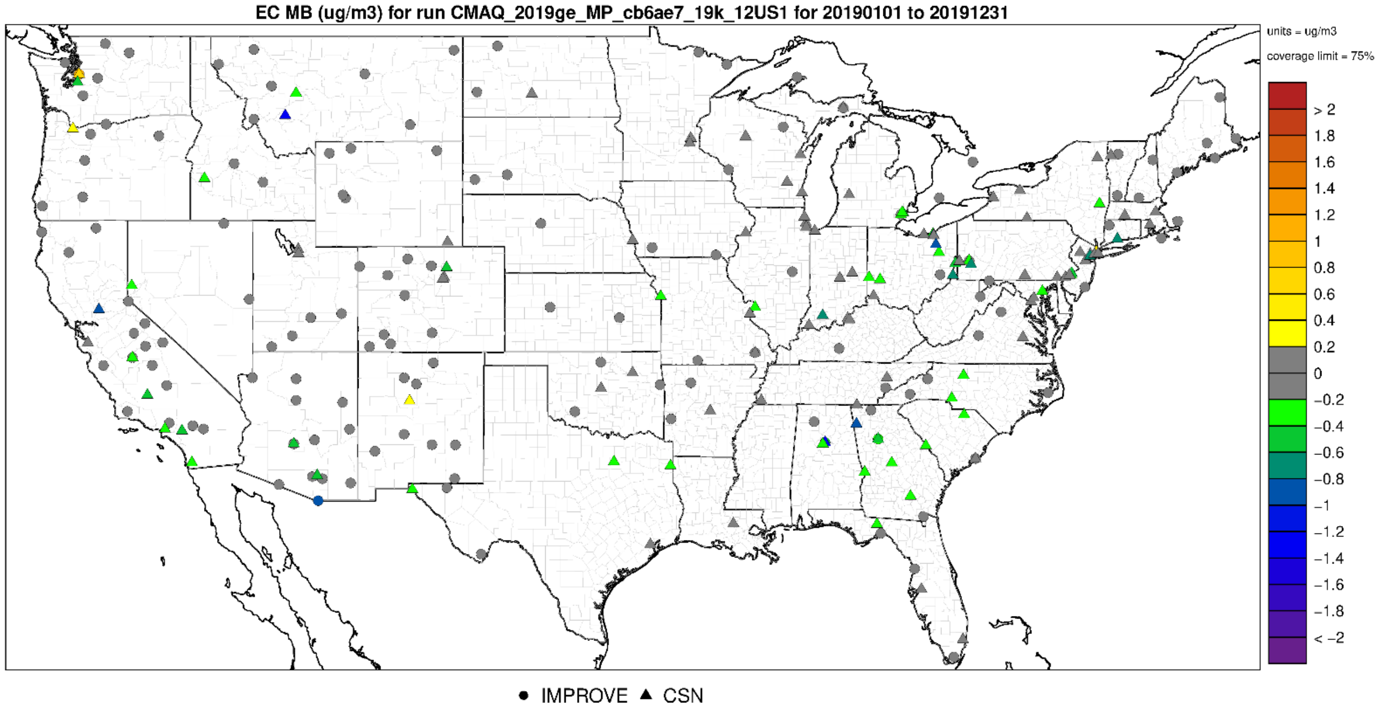


Figure 4-23. Mean Bias (µgm^{-3}) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

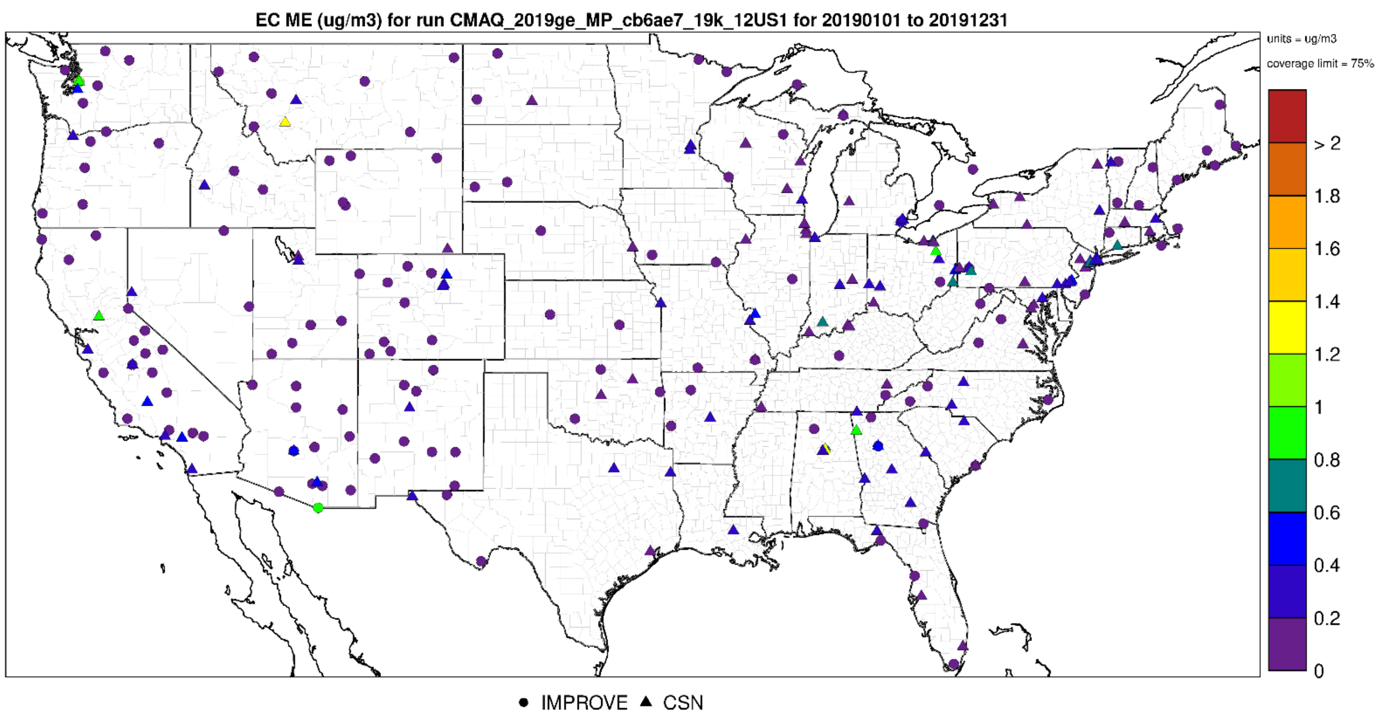


Figure 4-24. Mean Error (µgm^{-3}) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

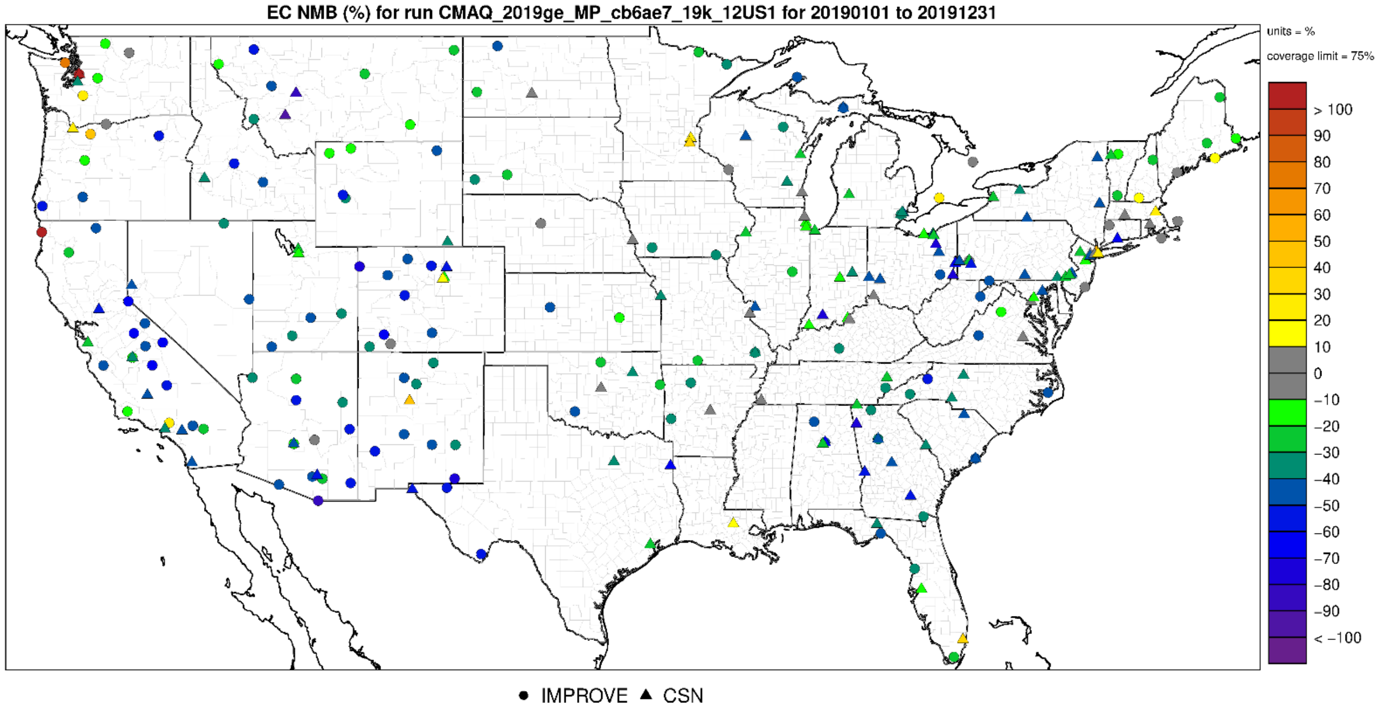


Figure 4-25. Normalized Mean Bias (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

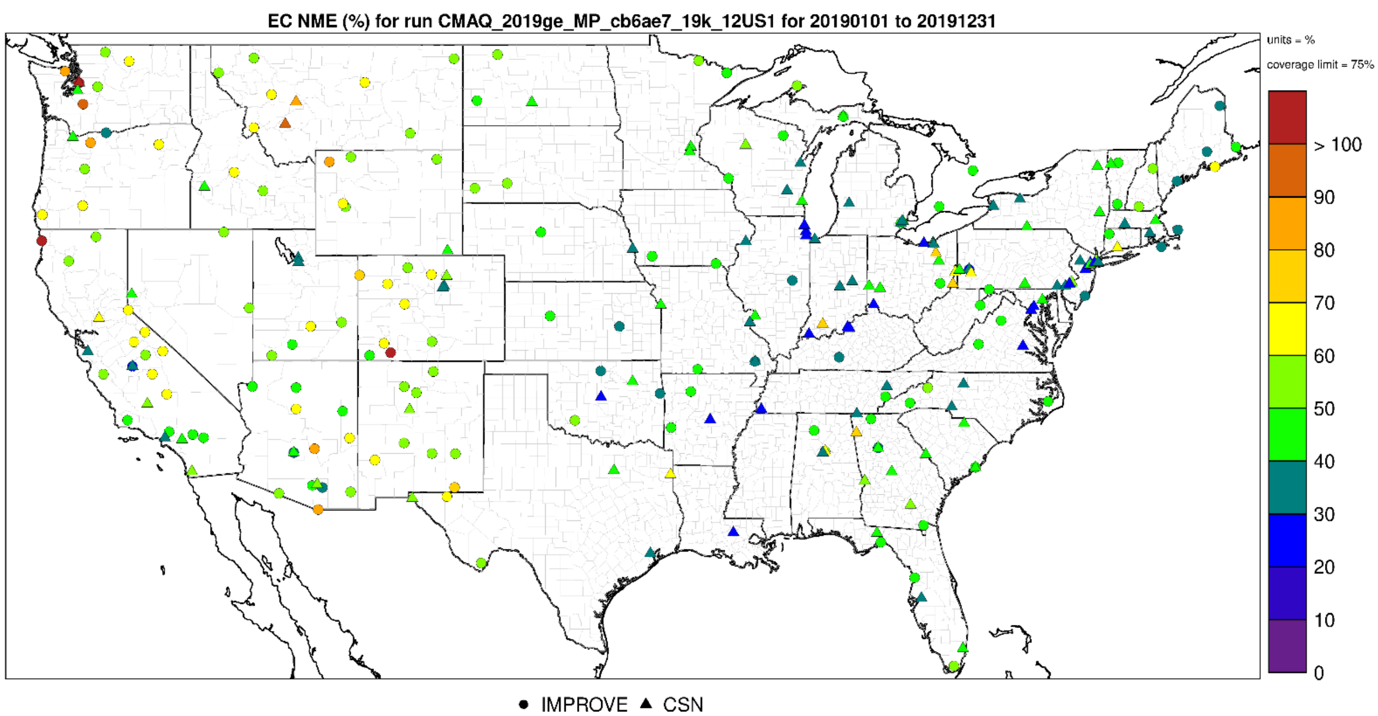


Figure 4-26. Normalized Mean Error (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

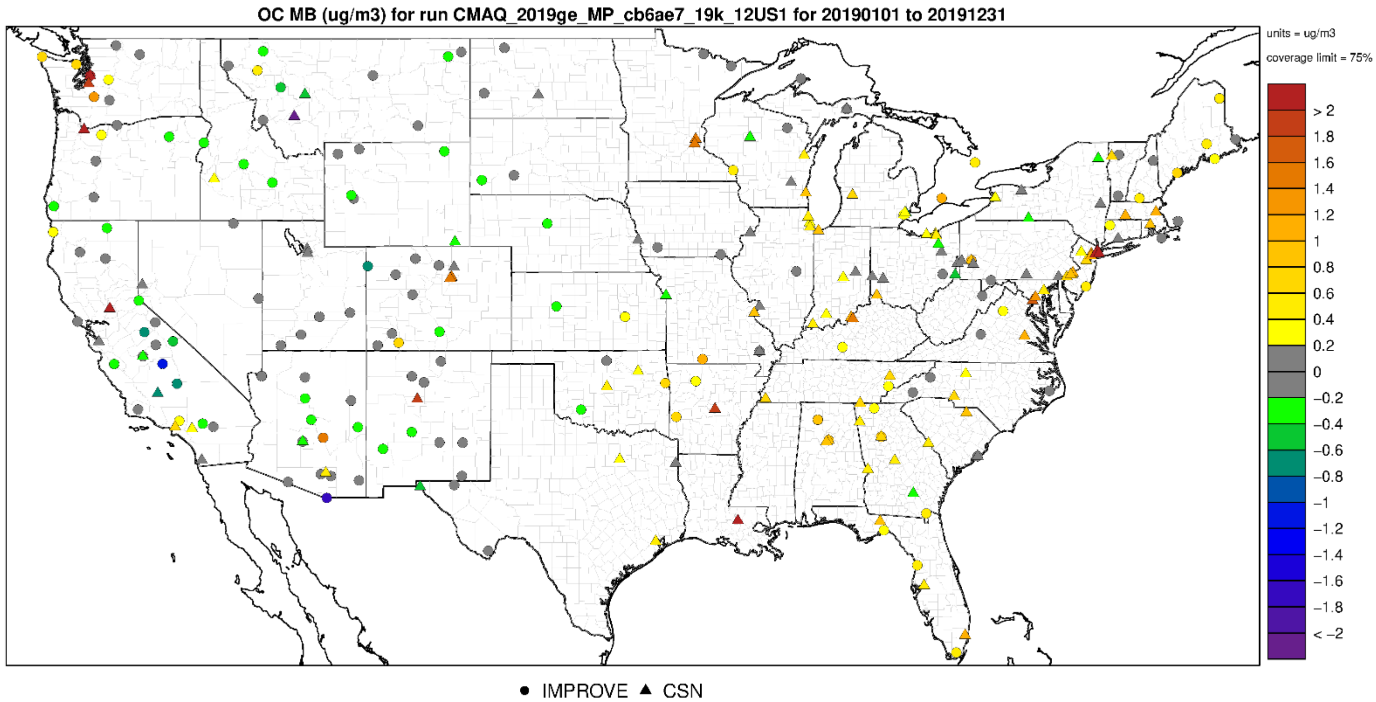


Figure 4-27. Mean Bias (µgm^{-3}) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

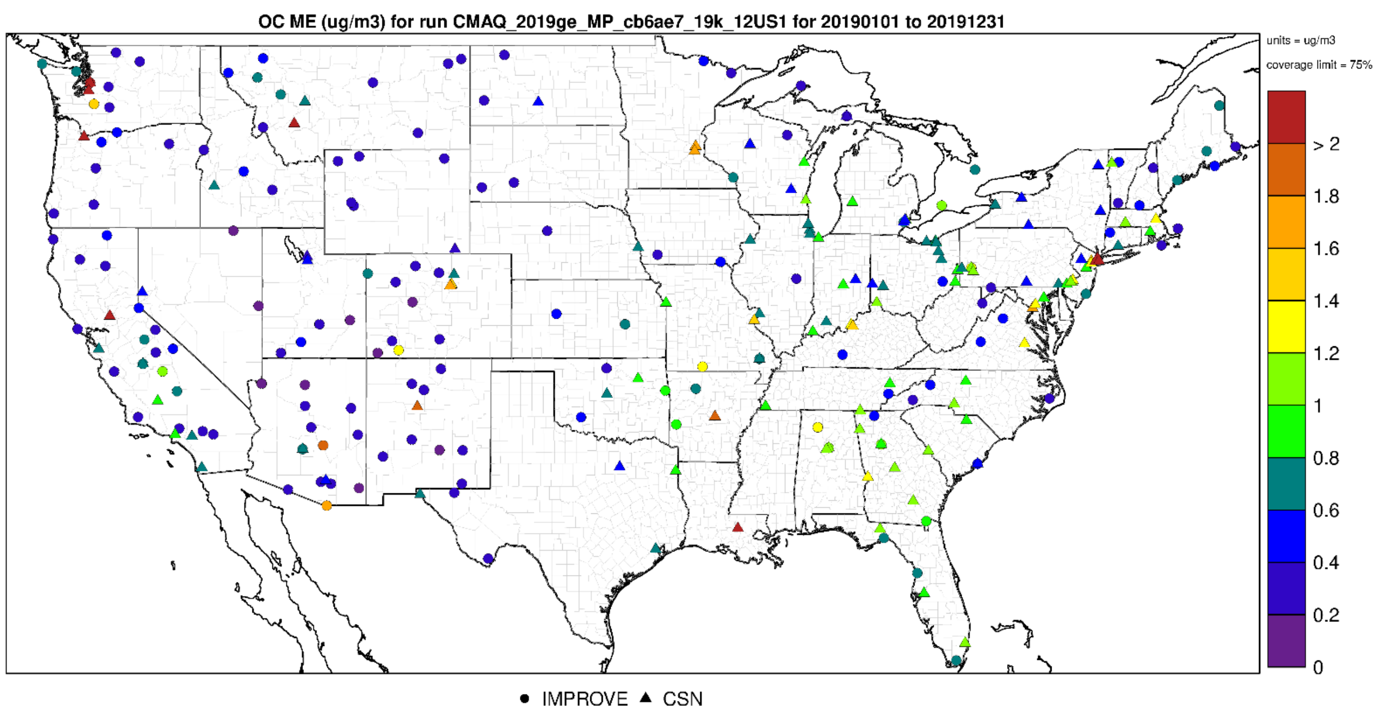


Figure 4-28. Mean Error (µgm^{-3}) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

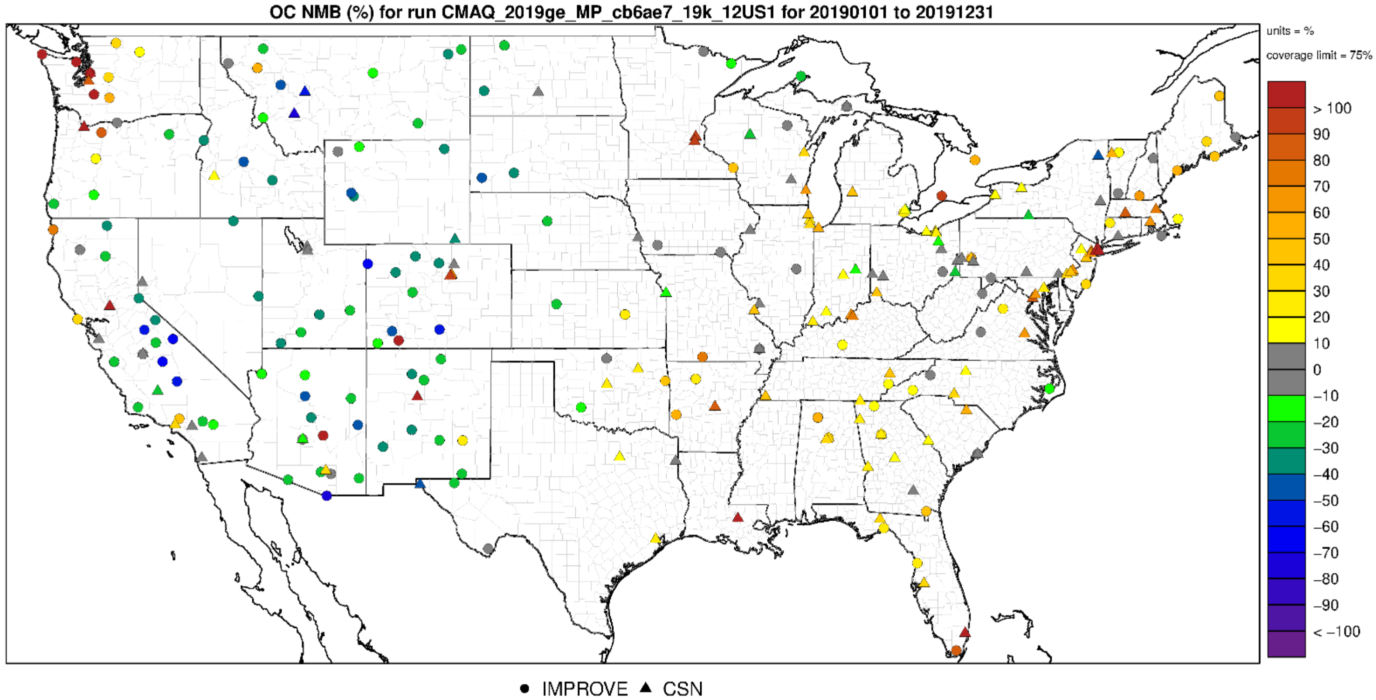


Figure 4-29. Normalized Mean Bias (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

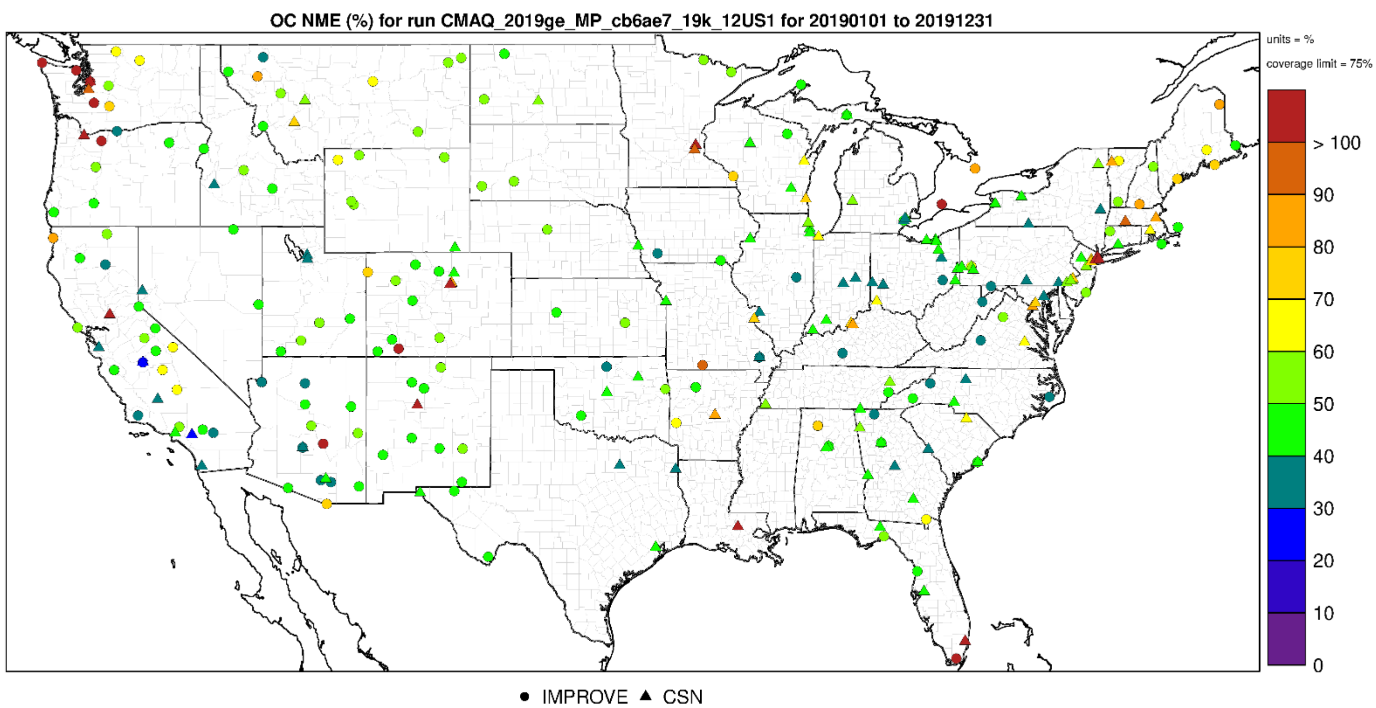


Figure 4-30. Normalized Mean Error (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

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 PM2.5 data from AQS and CMAQ to produce predictions to continental U.S. 2010 census tract centroids for the year 2019.

5.2 Downscaler Model

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

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

Where:

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

DS has additional properties that differentiate it from linear regression:

1) Rather than just finding a single optimal solution to Equation 1, DS uses a Bayesian approach so that uncertainties can be generated along with each concentration prediction. This involves drawing random samples of model parameters from built-in "prior" distributions and assessing their fit on the data on the order of thousands of times. After each iteration, properties of the prior distributions are adjusted to try to improve

the fit of the next iteration. The resulting collection of $\widetilde{\beta}_0$ and β_1 values at each space-time point are the "posterior" distributions, and the means and standard distributions of these are used to predict concentrations and associated uncertainties at new spatial points.

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

Further information about the development and inner workings of the current version of DS can be found in Berrocal, Gelfand and Holland (2012)[42] and references therein. The DS outputs that accompany this report are described below, along with some additional analyses that include assessing the accuracy of the DS predictions. Results are then summarized, and caveats are provided for interpreting them in the context of air quality management activities.

5.3 Downscaler Concentration Predictions

In this application, DS was used to predict daily concentration and associated uncertainty values at the 2010 US census tract centroids across the continental U.S. using 2019 measurement and CMAQ data as inputs. For ozone, the concentration unit is the daily maximum 8-hour average in ppb and for PM2.5 the concentration unit is the 24-hour average in µg/m3.

5.3.1 Summary of 8-hour Ozone Results

Figure 5-1 summarizes the AQS, CMAQ and DS ozone data over the year 2019. 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 2019, about 33% of the US Census tracts (23786 out of 72283) experienced at least one day with an ozone value above the NAAQS of 70 ppb.

[42] 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.

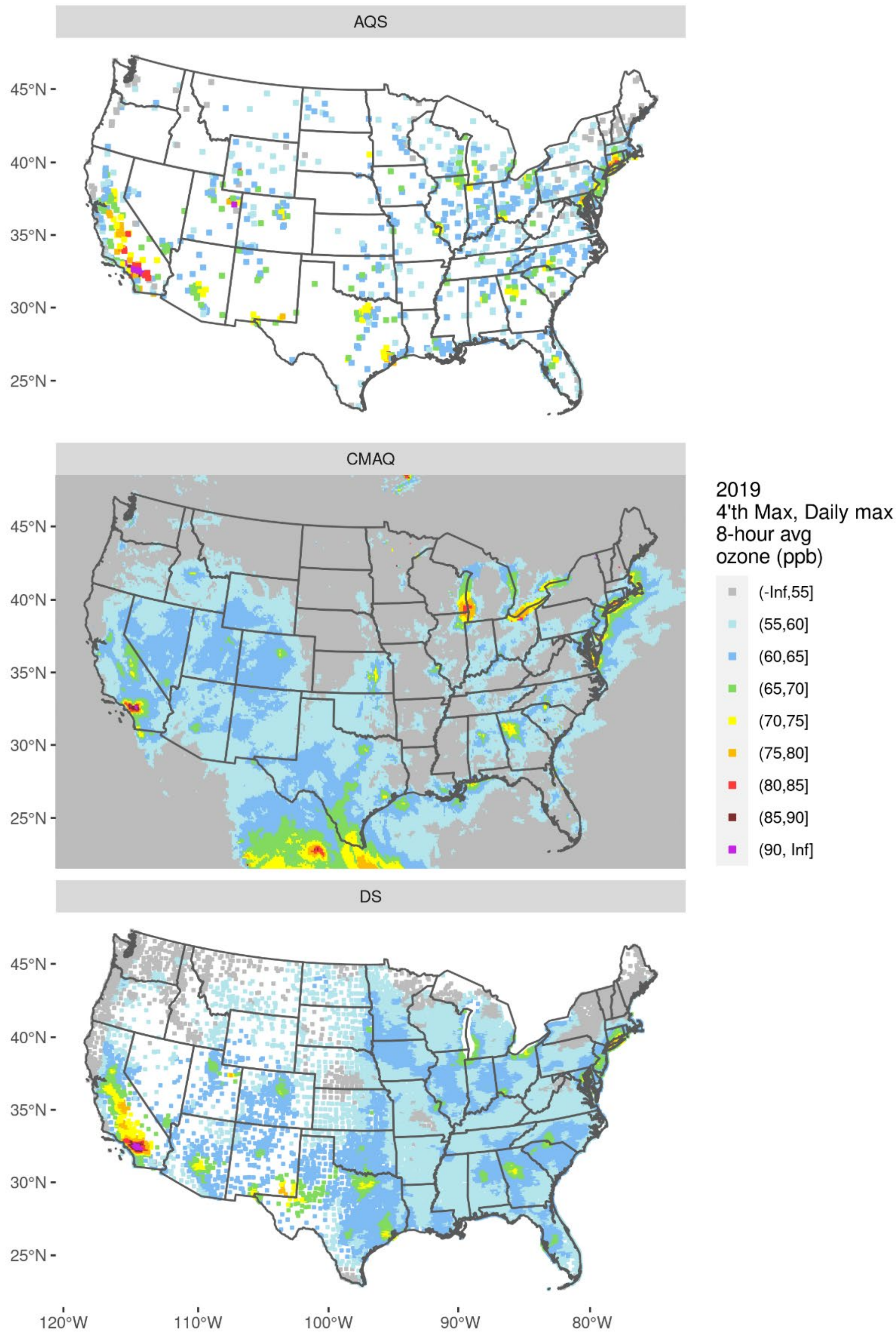

AQS
CMAQ
DS
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
2019
4'th Max, Daily max
8-hour avg
ozone (ppb)
(-Inf,55]
(55,60]
(60,65]
(65,70]
(70,75]
(75,80]
(80,85]
(85,90]
(90, Inf]

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

5.3.2 Summary of $PM_{2.5}$ Results

Figures 5-2 and 5-3 summarize the AQS, CMAQ and DS PM2.5 data over the year 2019. Figure 5-2 shows annual means and Figure 5-3 shows 98th percentiles of 24-hour PM2.5 concentrations for AQS observations, CMAQ model predictions and DS model results. The DS model estimated that for 2019 about 12% of the US Census tracts (8449 out of 72283) experienced at least one day with a PM2.5 value above the 24-hour NAAQS of 35 µg/m3.

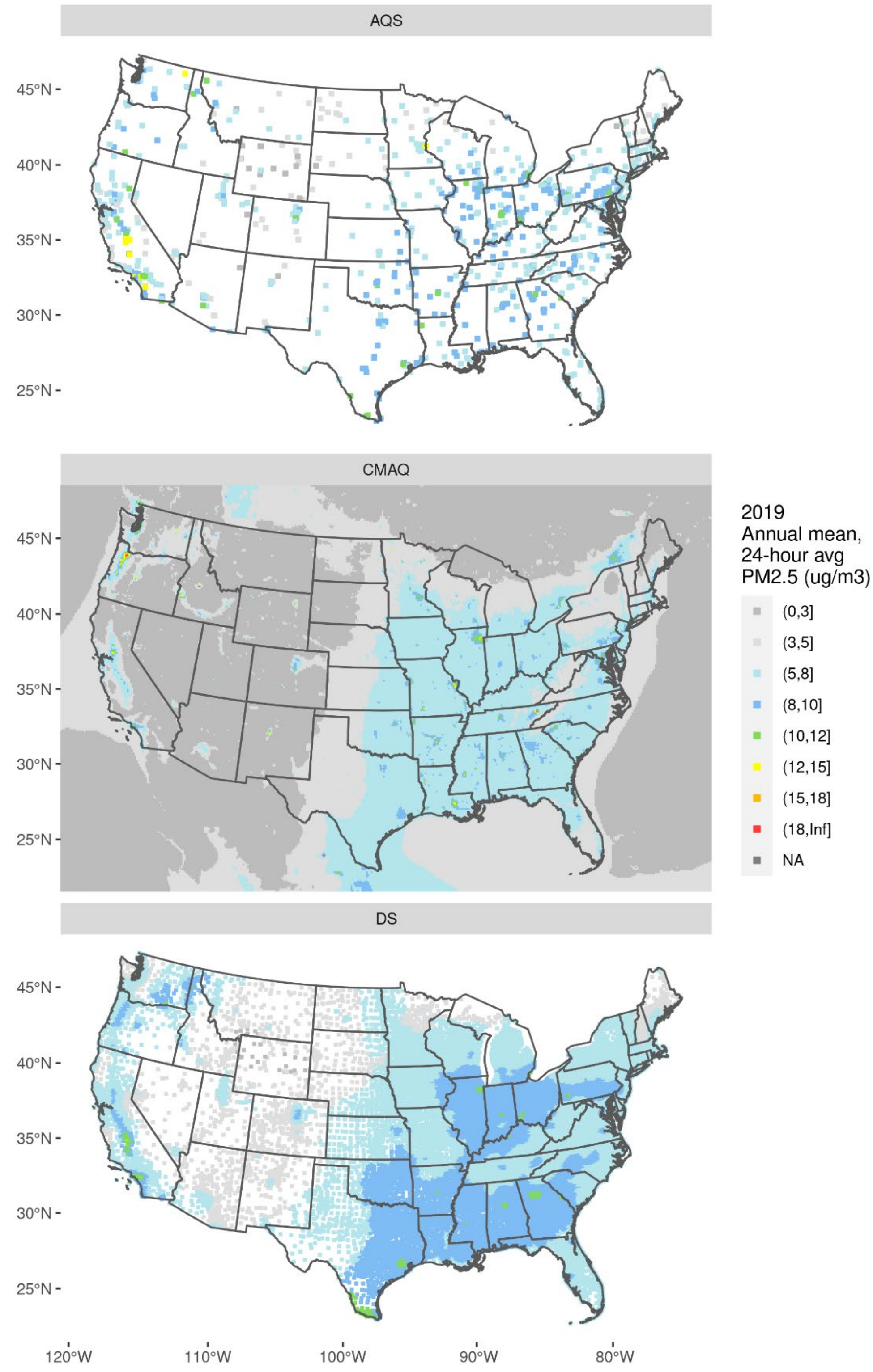


Figure 5-2. Annual mean $PM_{2.5}$ concentrations derived from AQS, CMAQ and DS data.

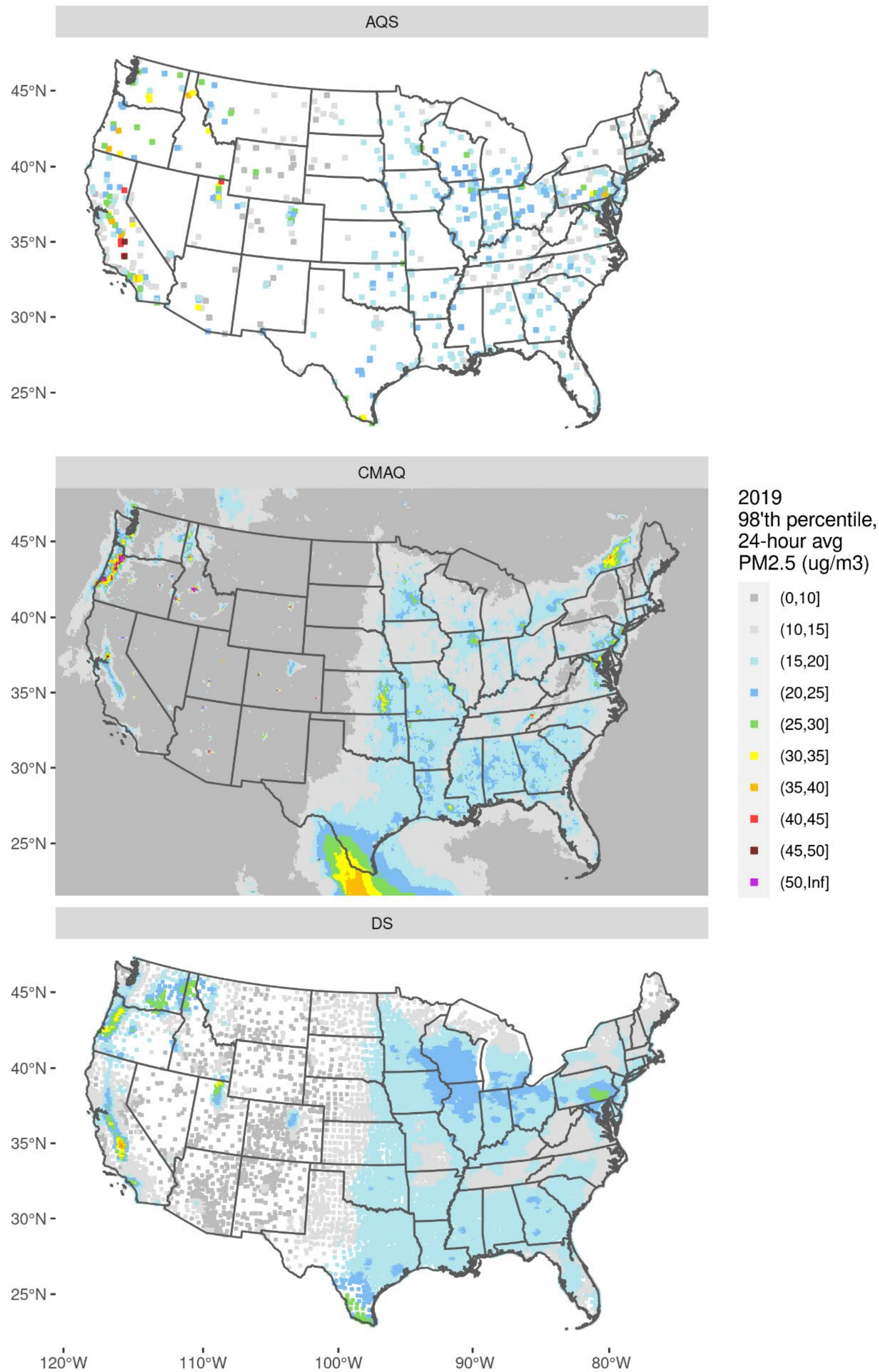
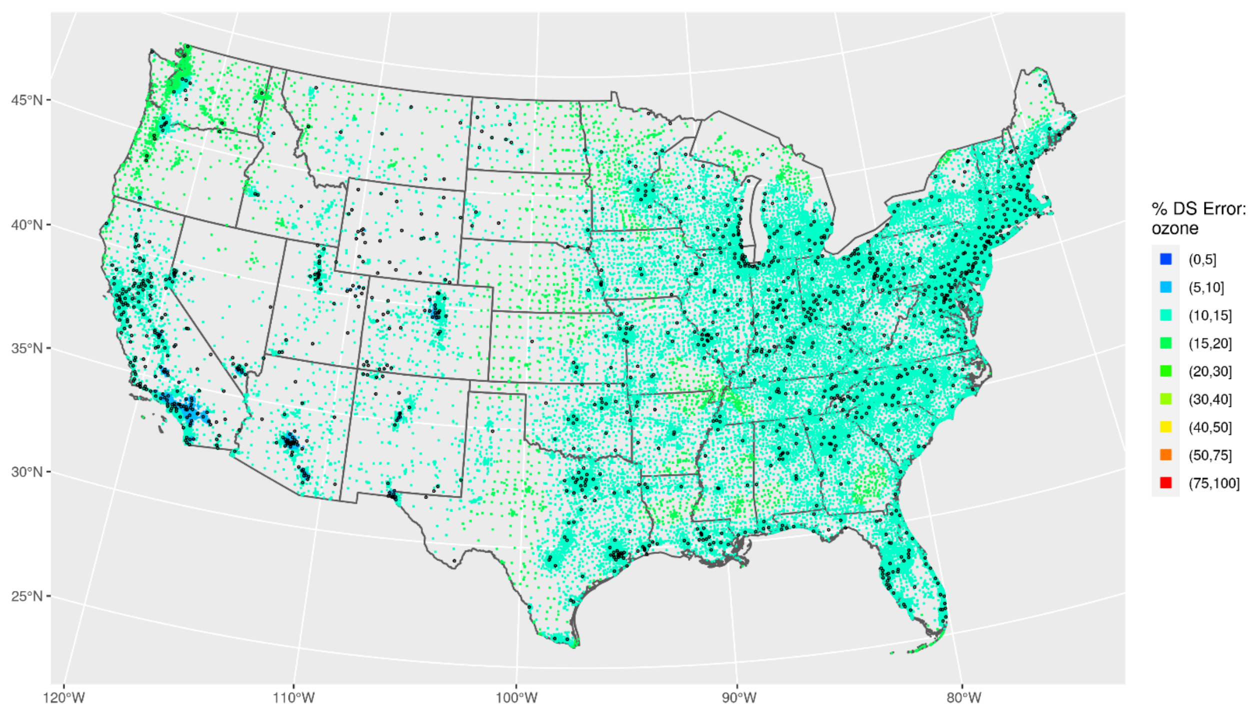

AQS
CMAQ
DS
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
2019
98'th percentile,
24-hour avg
PM2.5 (ug/m3)
(0,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
(40,45]
(45,50]
(50,Inf]

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

5.4 Downscaler Uncertainties

5.4.1 Standard Errors

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



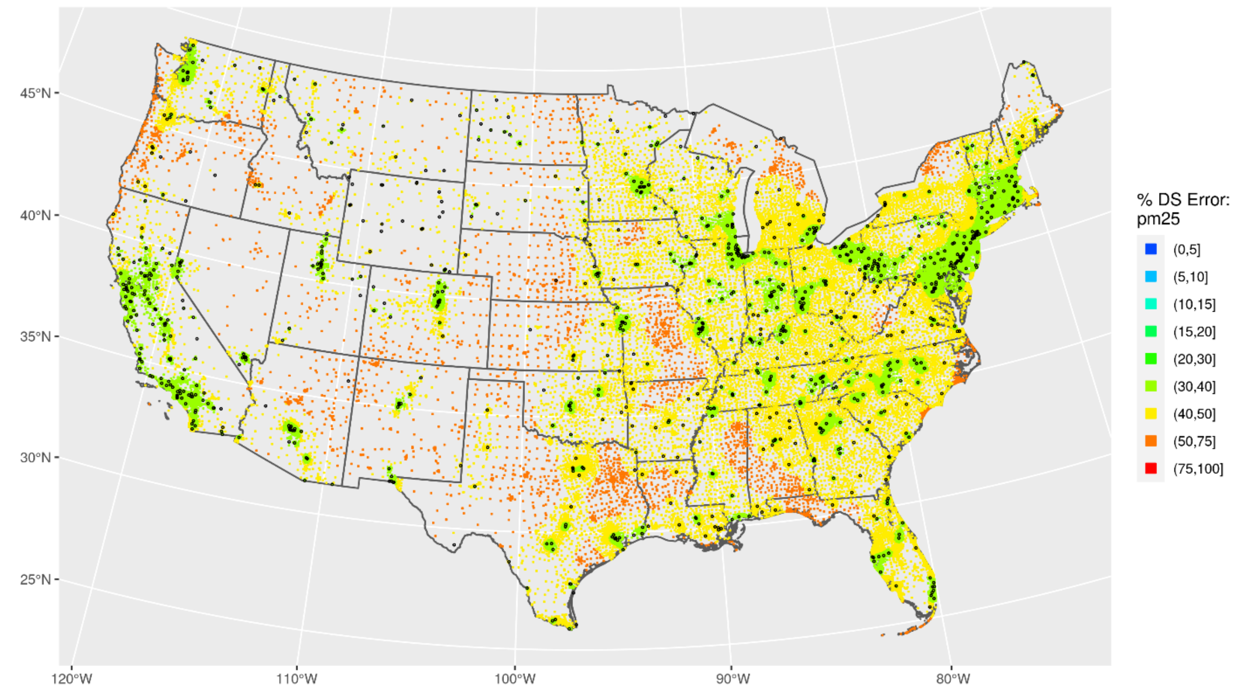


Figure 5-4. Annual mean relative errors (standard errors divided by predictions) from the DS 2019 runs. The black dots show the locations of monitors that generated the AQS data used as input to the DS model.

5.4.2 Cross Validation

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

Pollutant	# Monitors	Mean Bias	RMSE	Mean Coverage
$PM_{2.5}$	964	0.22	2.52	0.95
O_3	1231	-0.01	4.25	0.96

Table 5-1. Cross-validation statistics associated with the 2019 DS runs.

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

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

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

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

5.5 Summary and Conclusions

The results presented in this report are from an application of the DS fusion model for characterizing national air quality for Ozone and $PM_{2.5}$. DS provided spatial predictions of daily ozone and $PM_{2.5}$ at 2010 U.S. census tract centroids by utilizing monitoring data and CMAQ output for 2019. Large-scale spatial and temporal patterns of concentration predictions are generally consistent with those seen in ambient monitoring data. Both Ozone and $PM_{2.5}$ were predicted with lower error in the eastern versus the western U.S., presumably due to the greater monitoring density in the east.

An additional caution that warrants mentioning is related to the capability of DS to provide predictions at multiple spatial points within a single CMAQ 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
$PM_{2.5}$	Particulate matter less than or equal to 2.5 microns
PM_{10}	Particulate matter less than or equal to 10 microns
PMc	Particulate matter greater than 2.5 microns and less than 10 microns
Prescribed Fire	Intentionally set fire to clear vegetation
RIA	Regulatory Impact Analysis
RPO	Regional Planning Organization
RRTM	Rapid Radiative Transfer Model
SCC	Source Classification Code
SMARTFIRE	Satellite Mapping Automatic Reanalysis Tool for Fire Incident Reconciliation
SMOKE	Sparse Matrix Operator Kernel Emissions
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_2018_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-22-003 October 2022