



Bayesian space-time downscaling fusion model (downscaler) - Derived Estimates of Air Quality for 2012

EPA-450/R-16-001
March 2016

**Bayesian space-time downscaling fusion model (downscaler) -Derived
Estimates of Air Quality for 2012**

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

Authors:

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

Acknowledgements

The following people served as reviewers of this document and provided valuable comments that were included: Liz Naess (EPA/OAR), Tyler Fox (EPA/OAR), and Dennis Doll (EPA/OAR).

Contents

Contents	1
1.0 Introduction	2
2.0 Air Quality Data.....	5
2.1 Introduction to Air Quality Impacts in the United States.....	5
2.2 Ambient Air Quality Monitoring in the United States	7
2.3 Air Quality Indicators Developed for the EPHT Network	11
3.0 Emissions Data.....	13
3.1 Introduction to Emissions Data Development.....	13
3.2 Emission Inventories and Approaches	13
3.3 Emissions Modeling Summary	31
3.4 Emissions References	65
4.0 CMAQ Air Quality Model Estimates	68
4.1 Introduction to the CMAQ Modeling Platform	68
4.2 CMAQ Model Version, Inputs and Configuration	69
4.3 CMAQ Model Performance Evaluation	74
5.0 Bayesian space-time downscaling fusion model (downscaler) -Derived Air Quality Estimates.....	96
5.1 Introduction	96
5.2 Downscaler Model	96
5.3 Downscaler Concentration Predictions	97
5.4 Downscaler Uncertainties.....	102
5.5 Summary and Conclusions	104
Appendix A - Acronyms	105

1.0 Introduction

This report describes estimates of daily ozone (maximum 8-hour average) and PM_{2.5} (24-hour average) concentrations throughout the contiguous United States during the 2011 calendar year generated by EPA's recently developed data fusion method termed the "downscaler model" (DS). Air quality monitoring data from the State and Local Air Monitoring Stations (SLAMS) and numerical output from the Community Multiscale Air Quality (CMAQ) model were both input to DS to predict concentrations at the 2010 US census tract centroids encompassed by the CMAQ modeling domain. Information on EPA's air quality monitors, CMAQ model, and downscaler model is included to provide the background and context for understanding the data output presented in this report. These estimates are intended for use by statisticians and environmental scientists interested in the daily spatial distribution of ozone and 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 Bayesian¹ 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).

The data contained in this report are an outgrowth of a collaborative research partnership between EPA scientists from the Office of Research and Development's (ORD) National Exposure Research Laboratory (NERL) and personnel from EPA's Office of Air and Radiation's (OAR) Office of Air Quality Planning and Standards (OAQPS). NERL's Human Exposure and Atmospheric Sciences Division (HEASD), Atmospheric Modeling Division (AMD), and Environmental Sciences Division (ESD), in conjunction with OAQPS, work together to provide 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

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.

Program efforts, the CDC led the initiative to build the National EPHT Network (<http://www.cdc.gov/nceh/tracking/default.htm>). The National EPHT Program, with the EPHT Network as its cornerstone, is the CDC's response to requests calling for improved understanding of how the environment affects human health. The EPHT Network is designed to provide the means to identify, access, and organize hazard, exposure, and health data from a variety of sources and to examine, analyze and interpret those data based on their spatial and temporal characteristics.

Since 2002, EPA has collaborated with the CDC on the development of the EPHT Network. On September 30, 2003, the Secretary of Health and Human Services (HHS) and the Administrator of EPA signed a joint Memorandum of Understanding (MOU) with the objective of advancing efforts to achieve mutual environmental public health goals². HHS, acting through the CDC and the Agency for Toxic Substances and Disease Registry (ATSDR), and EPA agreed to expand their cooperative activities in support of the CDC EPHT Network and EPA's Central Data Exchange Node on the Environmental Information Exchange Network in the following areas:

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

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

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

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

contained herein.

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

- Section 2 describes the air quality data obtained from EPA's nationwide monitoring network and the importance of the monitoring data in determining health potential health risks.
- Section 3 details the emissions inventory data, how it is obtained and its role as a key input into the CMAQ air quality computer model.
- Section 4 describes the CMAQ computer model and its role in providing estimates of pollutant concentrations across the U.S. based on 12-km grid cells over the contiguous U.S.
- Section 5 explains the downscaler model used to statistically combine air quality monitoring data and air quality estimates from the CMAQ model to provide daily air quality estimates for the 2010 US census tract centroid locations within the contiguous U.S.
- The appendix provides a description of acronyms used in this report.

2.0 Air Quality Data

To compare health outcomes with air quality measures, it is important to understand the origins of those measures and the methods for obtaining them. This section provides a brief overview of the origins and process of air quality regulation in this country. It provides a detailed discussion of ozone (O₃) 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 <http://www.epa.gov/clean-air-act-overview>.

Under the CAA, the U.S. EPA has established standards or limits for six air pollutants, known as the criteria air pollutants: carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), ozone (O₃), 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 law requires EPA to review periodically these standards. For more specific information on the NAAQS, go to <https://www.epa.gov/criteria-air-pollutants/naaqs-table>. For general information on the criteria pollutants, go to <https://www.epa.gov/criteria-air-pollutants>.

When these standards are not met, the area is designated as a nonattainment area. States must develop state implementation plans (SIPs) that explain the regulations and controls it will use to clean up the nonattainment areas. States with an EPA-approved SIP can request that the area be designated from nonattainment to attainment by providing three consecutive years of data showing NAAQS compliance. The state must also provide a maintenance plan to demonstrate how it will continue to comply with the NAAQS and demonstrate compliance over a 10-year period, and what corrective actions it will take should a NAAQS violation occur after designation. EPA must review and approve the NAAQS compliance data and the maintenance plan before designating the area; thus, a person may live in an area designated as nonattainment even though no NAAQS violation has been observed for quite some time. For more information on designations, go to <https://www.epa.gov/ozone-designations> and <https://www.epa.gov/particulate-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. Toxicological, human exposure, and epidemiological studies were integrated by EPA in “Air Quality Criteria for Ozone and Related Photochemical Oxidants.” It is available at <https://www.epa.gov/naaqs/ozone-o3-air-quality-standards>. The current NAAQS for ozone (last revised in 2015) is a daily maximum 8-hour average of 0.070 parts per million [ppm] (for details, see <https://www.epa.gov/ozone-pollution/setting-and-reviewing-standards-control-ozone-pollution#standards>). The Clean Air Act requires EPA to review the NAAQS at least every five years and revise them as appropriate in accordance with Section 108 and Section 109 of the Act. The standards for ozone are shown in Table 2-1.

Table 2-1. Ozone Standards

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

2.1.3 Particulate Matter

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

As practical convention, PM is divided by sizes into classes with differing health concerns and potential sources⁴. Particles less than 10 micrometers in diameter (PM₁₀) 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 PM₁₀, PM_{2.5} and PMc varies from the Eastern U.S. to arid western areas.

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

⁴ The measure used to classify PM into sizes is the aerodynamic diameter. The measurement instruments used for PM are designed and operated to separate large particles from the smaller particles. For example, the PM_{2.5} instrument only captures and thus measures particles with an aerodynamic diameter less than 2.5 micrometers. The EPA method to measure PMc is designed around taking the mathematical difference between measurements for PM₁₀ 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/m³) based on the annual mean concentration averaged over three years, and the 24-hr average concentration must not exceed 35 ug/m³ based on the 98th percentile 24-hour average concentration averaged over three years. More information is available at <https://www.epa.gov/pm-pollution/setting-and-reviewing-standards-control-particulate-matter-pm-pollution#standards>. The standards for PM_{2.5} are shown in Table 2-2.

Table 2-2. PM_{2.5} Standards

Micrograms Per Cubic Meter: Measurement - (ug/m³)	1997	2006	2012
Annual Average	15.0	15.0	12.0
24-Hour Average	65	35	35

2.2 Ambient Air Quality Monitoring in the United States

2.2.1 *Monitoring Networks*

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

The SLAMS network experienced accelerated growth throughout the 1970s. The networks were further expanded in 1999 based on the establishment of separate NAAQS for fine particles (PM_{2.5}) in 1997. The NAAQS for PM_{2.5} were established based on their link to serious health problems ranging from increased symptoms, hospital admissions, and emergency room visits, to premature death in people with heart or lung disease. While most of the monitors in these networks are located in populated areas of the country, “background” and rural monitors are an important part of these networks. For more information on SLAMS, as well as EPA’s other air monitoring networks go to <http://www3.epa.gov/ttn/amtic/>.

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

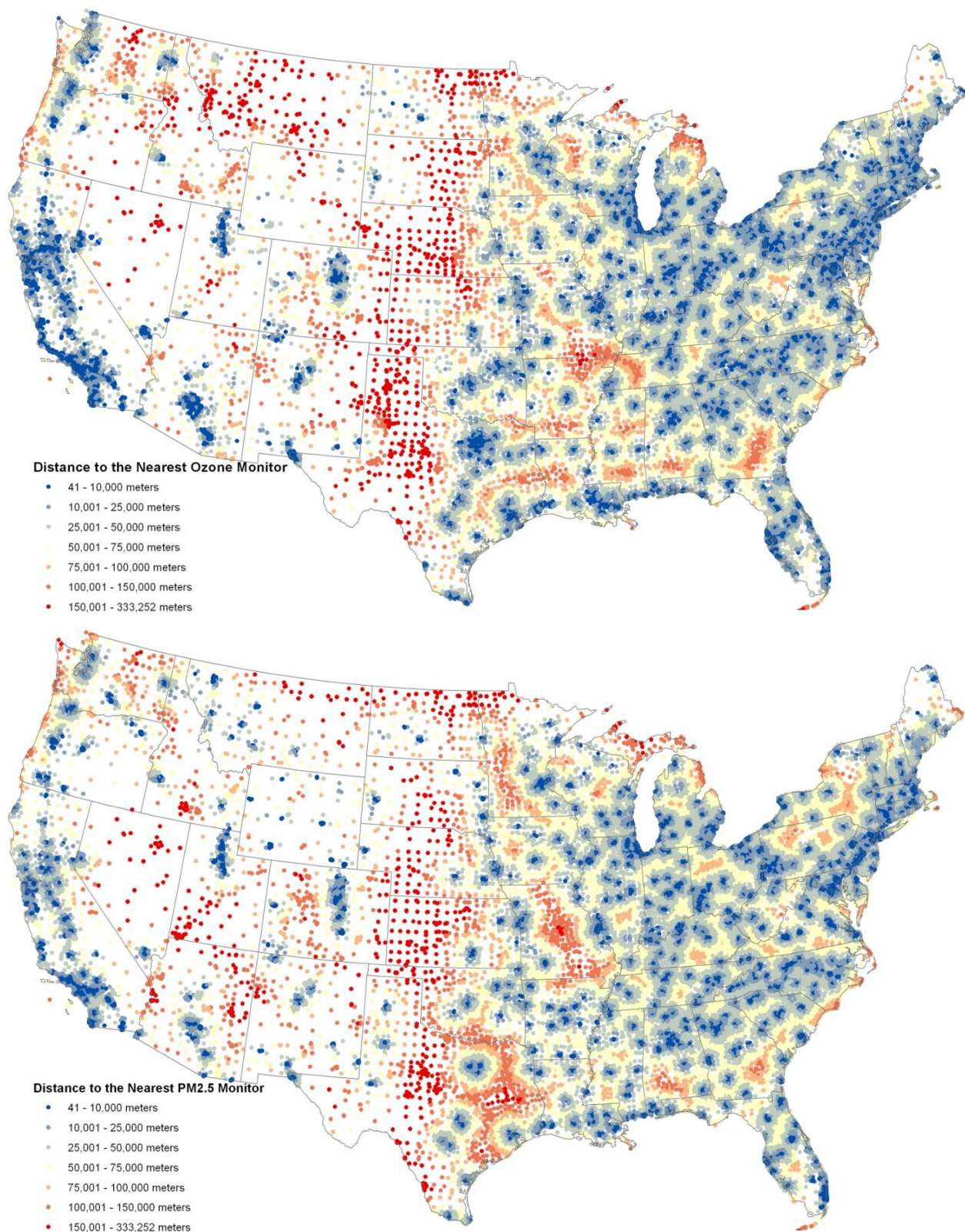


Figure 2-1. Distances from US Census Tract centroids to the nearest monitoring site, 2009.

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

2.2.2 Air Quality System Database

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

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

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

The advantage of using ambient data from EPA monitoring networks for comparing with health outcomes is that these measurements of pollution concentrations are the best characterization of the concentration of a given pollutant at a given time and location. Furthermore, the data are supported by a comprehensive quality assurance program, ensuring data of known quality. One disadvantage of using the ambient data is that it is usually out of spatial and temporal alignment with health outcomes. This spatial and temporal 'misalignment' between air quality monitoring data and health outcomes is influenced by the following key factors: the living and/or working locations (microenvironments) where a person spends their time not being co-located with an air quality monitor; time(s)/date(s) when a patient experiences a health outcome/symptom (e.g., asthma attack) not coinciding with time(s)/date(s) when an air quality monitor records ambient concentrations of a pollutant high enough to affect the symptom (e.g., asthma attack either during or shortly after a high PM_{2.5} day). To compare/correlate ambient concentrations with acute health effects, daily local air quality data is needed⁵. 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. However, over the past several years, continuous monitors, which can automatically collect,

⁵ EPA uses exposure models to evaluate the health risks and environmental effects associated with exposure. These models are limited by the availability of air quality estimates. <http://www.epa.gov/ttn/fera/index.html>.

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

2.2.4 Use of Air Quality Monitoring Data

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

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

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

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

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

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

Combining monitoring data with air quality models (via fusion or regression) may provide the best results in terms of estimating ambient air concentrations in space and time. EPA's eVNA⁶ is an example of an

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

earlier approach for merging air quality monitor data with CMAQ model predictions. The downscaler model attempts to address some of the shortcomings in these earlier attempts to statistically combine monitor and model predicted data, see published paper referenced in section 1 for more information about the downscaler model. As discussed in the next section, there are two methods used in EPHT to provide estimates of ambient concentrations of air pollutants: air quality monitoring data and the downscaler model estimate, which is a statistical ‘combination’ of air quality monitor data and photochemical air quality model predictions (e.g., CMAQ).

2.3 Air Quality Indicators Developed for the EPHT Network

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

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

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

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

Ozone (daily 8-hr period with maximum concentration—ppm—by Federal Reference Method (FRM))
<ul style="list-style-type: none"> • 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)
<ul style="list-style-type: none"> • Average ambient concentrations of particulate matter (< 2.5 microns in diameter) and compared to annual PM_{2.5} NAAQS (by state). • % population exceeding annual PM_{2.5} NAAQS (by state). • % of days with PM_{2.5} concentration over the daily NAAQS (or other relevant benchmarks (by county and MSA) • Number of person-days with PM_{2.5} concentration over the daily NAAQS & other relevant benchmarks (by county and MSA)

2.3.1 Rationale for the Air Quality Indicators

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

2.3.2 Air Quality Data Sources

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

2.3.3 Use of Air Quality Indicators for Public Health Practice

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

3.0 Emissions Data

3.1 Introduction to Emissions Data Development

The U.S. Environmental Protection Agency (EPA) developed an air quality modeling platform based primarily on the 2011 National Emissions Inventory (NEI), Version 2 to process year 2012 emission data for this project. This section provides a summary of the emissions inventory and emissions modeling techniques applied to Criteria Air Pollutants (CAPs) and the following select Hazardous Air Pollutants (HAPs): chlorine (Cl), hydrogen chloride (HCl), benzene, acetaldehyde, formaldehyde and methanol. This section also describes the approach and data used to produce emissions inputs to the air quality model. The air quality modeling, meteorological inputs and boundary conditions are described in a separate section.

The Community Multiscale Air Quality (CMAQ) model (<http://www.epa.gov/AMD/CMAQ/>) was used to model ozone (O₃) and particulate matter (PM) for this project. CMAQ requires hourly and gridded emissions of the following inventory pollutants: carbon monoxide (CO), nitrogen oxides (NO_x), volatile organic compounds (VOC), sulfur dioxide (SO₂), ammonia (NH₃), particulate matter less than or equal to 10 microns (PM₁₀), and individual component species for particulate matter less than or equal to 2.5 microns (PM_{2.5}). In addition, the Carbon bond 2005 (CB05) with chlorine chemistry used here within CMAQ allows for explicit treatment of the VOC HAPs benzene, acetaldehyde, formaldehyde and methanol (BAFM) and includes anthropogenic HAP emissions of HCl and Cl.

The effort to create the 2012 emission inputs for this study included development of emission inventories for input to a 2012 modeling case, along with application of emissions modeling tools to convert the inventories into the format and resolution needed by CMAQ. Year-specific fire and continuous emission monitoring (CEM) data for electric generating units (EGUs) were used. The primary emissions modeling tool used to create the CMAQ model-ready emissions was the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. SMOKE version 3.6.5 was used to create CMAQ-ready emissions files for a 12-km national grid. Additional information about SMOKE is available from <http://cmascenr.org/smoke>.

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

3.2 Emission Inventories and Approaches

This section describes the emissions inventories created for input to SMOKE. The 2011 NEI, version 2 with updates for the year 2012 is the primary basis for the inputs to SMOKE. The NEI includes five main categories of source sectors: 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. Documentation for the 2011 NEI is available at <http://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-documentation>.

Point source data for the year 2012 as submitted to EIS were used for this study. EPA used the SMARTFIRE2 system to develop 2012 fire emissions. SMARTFIRE2 categorizes all fires as either prescribed burning or wildfire categories, and includes improved emission factor estimates for prescribed burning. Onroad mobile source emissions for year 2012 were developed using MOVES2014. Nonroad mobile source emissions for year 2012 were developed using the National Mobile Inventory Model (NMIM). Canadian emissions reflect year 2010 and Mexican emissions reflect year 2008, as those were the latest data available at the time of the modeling.

The methods used to process emissions for this study are very similar to those documented for EPA's Version 6.2, 2011 Emissions Modeling Platform that was also used for the final Ozone National Ambient Air Quality Standards (NAAQS) and for the proposed Cross-State Air Pollution Rule Update. A technical support document (TSD) for this platform is available at EPA's emissions modeling clearinghouse (EMCH): <https://www.epa.gov/air-emissions-modeling/2011-version-62-platform> (EPA, 2015a) and includes additional details regarding the data preparation and emissions modeling.

The emissions modeling process, performed using SMOKE v3.6.5, apportions the emissions inventories into the grid cells used by CMAQ and temporalizes the emissions into hourly values. In addition, the pollutants in the inventories (e.g., NO_x and VOC) are split into the chemical species needed by CMAQ. For the purposes of preparing the CMAQ-ready emissions, the broader NEI emissions inventories are split into emissions modeling "platform" sectors; and biogenic emissions are added along with emissions from other sources other than the NEI, such as the Canadian, Mexican, and offshore inventories. The significance of an emissions sector for the emissions modeling platform is that emissions for that sector are run through all of the SMOKE programs, except the final merge, independently from emissions in the other sectors. The final merge program called Mrggrid combines the sector-specific gridded, speciated and temporalized emissions to create the final CMAQ-ready emissions inputs.

Table 3-1 presents the sectors in the emissions modeling platform used to develop the year 2012 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 2012 emission summaries for the U.S. anthropogenic sectors are shown in Table 3-2 (i.e., biogenic emissions are excluded). Table 3-3 provides a summary of emissions for the anthropogenic sectors containing Canadian, Mexican and offshore sources. State total emissions for each sector are provided in Appendix B, a workbook entitled "Appendix_B_2012_emissions_totals_by_sector.xlsx".

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

2012 Platform Sector (<i>Abbrev</i>)	NEI Category	Description and resolution of the data input to SMOKE
EGUs (<i>ptegu</i>)	Point	2012 point source EGUs submitted to EIS and gapfilled with 2011NEIv2 emissions. Replaced with hourly 2012 CEMS values for NOX and SO2, where the units are matched to the inventory. Annual resolution for non-CEMS sources, hourly for sources matched to CEMS.
Point source oil and gas (<i>pt_oilgas</i>)	Point	2012 point sources with oil and gas production NAICS codes, gapfilled with 2011NEIv2 oil and gas sources.
Remaining non-EGU point (<i>ptnonipm</i>)	Point	2012 point source records not matched to the <i>ptegu</i> or <i>pt_oilgas</i> sectors, except for offshore point sources that are in the <i>othpt</i> sector. Includes all aircraft emissions and some rail yard emissions. Annual resolution.
Point source fire (<i>ptfire</i>)	Fires	Point source day-specific wildfires and prescribed fires for 2012 computed using SMARTFIRE 2. Fires over 20,000 acres on a single day allocated to overlapping grid cells.
Agricultural Fires (<i>agfire</i>)	Nonpoint	Agricultural fires from 2011NEIv2
Agricultural (<i>ag</i>)	Nonpoint	NH3 emissions from 2011NEIv2 nonpoint livestock and fertilizer application, county and annual resolution but temporalized using 2012 meteorological data.
Area fugitive dust (<i>afdust_adj</i>)	Nonpoint	PM10 and PM2.5 from fugitive dust sources in the 2011NEIv2 nonpoint inventory, including building construction, road construction, agricultural dust, and road dust. Unpaved and paved road dust emissions differ from the NEI in that the NEI uses an average meteorological adjustment while the modeling uses gridded, hourly 2012 meteorological data to reduce emissions via a transport fraction based on land use, plus emissions are zeroed out during periods of precipitation and snow/ice cover. County and annual resolution.
Nonpoint source oil and gas (<i>np_oilgas</i>)	Nonpoint	2011NEIv2 nonpoint sources from oil and gas-related processes with updates for five Utah counties. County and annual resolution.
Residential Wood Combustion (<i>rwc</i>)	Nonpoint	2011NEIv2 NEI nonpoint sources for Residential Wood Combustion (RWC) processes but temporalized using 2012 meteorological data. County and annual resolution.
Remaining nonpoint (<i>nonpt</i>)	Nonpoint	2011NEIv2 nonpoint sources not included in other platform sectors; county and annual resolution.
C3 commercial marine (<i>c3marine</i>)	Nonpoint	2011NEIv2 Category 3 (C3) commercial marine vessel (CMV) emissions. County and annual resolution.
C1 and C2 marine and locomotive (<i>c1c2rail</i>)	Nonpoint	2011NEIv2 locomotives and primarily category 1 (C1) and category 2 (C2) commercial marine vessel (CMV) emissions sources. County and annual resolution.

2012 Platform Sector (<i>Abbrev</i>)	NEI Category	Description and resolution of the data input to SMOKE
Nonroad (<i>nonroad</i>)	Nonroad	2012 nonroad equipment emissions developed with the National Mobile Inventory Model (NMIM) using NONROAD2008 version NR08a. NMIM was used for all states except California-submitted 2011 data is used as is and Texas emissions for 2011 are projected to 2012 using data from EPA trends. County and monthly resolution.
Onroad (<i>onroad</i>)	Onroad	2012 onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles. Includes the following modes: exhaust, extended idle, auxiliary power units, evaporative, permeation, refueling and brake and tire wear. For all states except California, based on monthly MOVES emissions tables from MOVES2014. California emissions are based on Emission Factor (EMFAC) and were projected to 2012 using factors specific to each air basin. MOVES-based emissions computed for each hour and model grid cell using monthly and annual activity data (e.g., VMT, vehicle population).
Biogenic (<i>beis</i>)	Biogenic	Hour- and grid cell-specific emissions for 2012 generated from the BEIS 3.61 model, including emissions in Canada and Mexico.
Other area fugitive dust sources	N/A	Area fugitive dust sources from Canada 2010 inventory with transport fraction and snow/ice adjustments based on 2012 meteorological data. Annual and province resolution.
Other point sources not from the NEI (<i>othpt</i>)	N/A	Point sources from Canada's 2010 inventory and Mexico's 2008 inventory. Also includes all non-U.S. and non-Canada C3 CMV and U.S. offshore oil production from 2011NEIv2. Annual resolution.
Other nonpoint and nonroad (<i>othar</i>)	N/A	Year 2010 Canada (province resolution) and year 2008 (municipio resolution) nonpoint and nonroad mobile inventories, annual resolution.
Other onroad sources (<i>othon</i>)	N/A	Year 2010 Canada (province resolution) and year 2008 Mexico (municipio resolution) onroad mobile inventories, annual resolution.

Table 3-2. 2012 Continental United States Emissions by Sector (tons/yr in 48 states + D.C.)

Sector	CO	NH ₃	NO _x	PM ₁₀	PM _{2.5}	SO ₂	VOC
Sector	CO	NH3	NOX	PM10	PM2_5	SO2	VOC
afdust_adj				7,156,383	986,032		
ag		3,515,198					
agfire	956,243	3,321	42,767	140,728	93,959	16,224	74,783
c1c2rail	180,579	511	1,075,217	35,359	33,019	12,609	48,281
nonpt	1,645,989	94,242	720,454	491,825	404,258	275,655	3,672,249
np_oilgas	627,764	0	641,611	16,850	15,395	18,338	2,548,410
nonroad	13,266,383	2,668	1,546,449	154,415	146,950	2,779	1,869,401
onroad	23,707,819	112,443	5,180,893	346,279	175,360	29,309	2,448,957
c3marine	12,532	68	131,382	10,168	9,043	86,373	5,149
ptfire	30,696,364	502,685	361,109	3,071,146	2,602,666	213,027	7,226,100
ptegu	803,584	26,954	1,829,235	226,421	171,851	3,416,894	36,968
ptnonipm	2,226,713	58,339	1,210,541	475,089	312,253	976,937	824,521
pt_oilgas	220,304	5,373	469,889	13,850	13,295	66,731	157,224
rwc	2,517,844	19,693	34,436	381,476	381,252	8,954	442,541
Continental U.S.	76,861,119	4,341,496	13,243,984	12,519,989	5,345,333	5,123,830	19,354,584

Table 3-3. 2012 Non-US Emissions by Sector within Modeling Domain (tons/yr for Canada, Mexico, Offshore)

Sector	CO	NH ₃	NO _x	PM ₁₀	PM _{2.5}	SO ₂	VOC
Canada othafdust				821,220	120,724		
Canada othar	3,023,643	327,439	362,798	159,741	131,777	70,475	889,624
Canada othon	3,039,490	18,700	346,500	17,670	12,245	1,706	178,868
Canada othpt	497,442	13,110	267,633	70,196	29,243	545,999	129,473
Canada Subtotal	6,560,576	359,249	976,931	1,068,827	293,989	618,179	1,197,964
Mexico othar	278,932	163,416	183,435	99,120	50,343	10,707	412,293
Mexico othon	3,369,179	7,997	244,293	2,431	1,628	4,931	320,121
Mexico othpt	153,467	3,715	287,064	55,320	42,222	454,837	53,958
Mexico Subtotal	3,801,578	175,128	714,791	156,871	94,192	470,476	786,373
Offshore to EEZ	175,833	185	902,448	26,320	24,611	139,550	81,825
Non-US SECA C3	17,231	0	202,987	17,253	15,871	127,928	7,314
Total non-U.S.	10,555,218	534,562	2,797,158	1,269,271	428,664	1,356,132	2,073,476

3.2.1 Point Sources (*ptegu*, *pt_oilgas* and *ptnonipm*)

Point sources are sources of emissions for which specific geographic coordinates (e.g., latitude/longitude) are specified, as in the case of an individual facility. A facility may have multiple emission release points, which 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 United States. The offshore oil platform (*othpt* sector) and category 3 CMV emissions outside of state waters (*othpt* sector) are processed by SMOKE as point source inventories and are discussed later in this section. Full documentation for the development of the 2011NEIv2 (EPA, 2015b), is posted at: <http://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-documentation>.

A complete NEI is developed every three years, with 2011 being the most recently finished complete NEI. In years other than the triennial NEI years, including 2012, states are only required to submit emissions for Type A point sources. Type A point sources have the potential to emit more than 2500 tons of CO, NO_x, or SO₂, or more than 250 tons of VOC, PM₁₀, PM_{2.5} or NH₃. Due to this potential incompleteness in the submitted sources for 2012, non-submitted but still operating sources were pulled forward from the 2011NEIv2 inventory.

In preparation for modeling, the complete set of point sources was exported from EIS and split into several sectors for modeling. After moving offshore oil platforms into the *othpt* sector, and dropping sources without specific locations (i.e., the FIPS code ends in 777), initial versions of the other three point source sectors were created from the remaining 2012 point sources. The point sectors are: EGUs (*ptegu*), point source oil and gas extraction-related sources (*pt_oilgas*) and the remaining non-EGUs (*ptnonipm*). The EGU emissions are split out from the other sources to facilitate the use of distinct SMOKE temporal processing and future-year projection techniques. The oil and gas sector emissions (*pt_oilgas*) were processed separately for summary tracking purposes and distinct future-year projection techniques from the remaining non-EGU emissions (*ptnonipm*).

The inventory pollutants processed through SMOKE for both the *ptipm* and *ptnonipm* sectors were: CO, NO_x, VOC, SO₂, NH₃, PM₁₀, and PM_{2.5} and the following HAPs: HCl (pollutant code = 7647010), and Cl (code = 7782505). BAFM from these sectors was not utilized because VOC was speciated without the use (i.e., integration) of VOC HAP pollutants from the inventory.

The *ptnonipm* and *pt_oilgas* sector emissions were provided to SMOKE as annual emissions. For sources in the *ptegu* sector that could be matched to 2012 CEMS data, hourly CEMS NO_x and SO₂ emissions for 2012 from EPA's Acid Rain Program were used rather than NEI emissions. For all other pollutants (e.g., VOC, PM_{2.5}, HCl), annual emissions were used as-is from the NEI, but were allocated to hourly values using heat input from the CEMS data. For the unmatched units in the *ptegu* and *ptegu_pk* sectors, annual emissions were allocated to daily values using IPM region- and pollutant-specific profiles, and similarly, region- and pollutant-specific diurnal profiles were applied to create hourly emissions.

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

- a. CAP and HAP data were provided by States, locals and tribes under the Air Emissions Reporting Rule (AERR)

- 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 provided data for airports and rail yards.
- e. Off-shore platform data were added from the Bureau of Ocean Energy Management (BOEM).

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

- The tribal data, which do not use state/county Federal Information Processing Standards (FIPS) codes in the NEI, but rather use the tribal code, were assigned a state/county FIPS code of 88XXX, where XXX is the 3-digit tribal code in the NEI. This change was made because SMOKE requires all sources to have a state/county FIPS code.
- Sources that did not have specific counties assigned (i.e., the county code ends in 777) were not included in the modeling because it was only possible to know the state in which the sources resided, but no more specific details related to the location of the sources were available.
- Stack parameters for some point sources were defaulted when modeling in SMOKE. SMOKE uses default stack parameters by SCC code to gap fill stack parameters if they are missing in the input inventories. Also, SMOKE adjusts any stack parameter values that are outside of the SMOKE-defined range of acceptable values.

3.2.1.1 EGU sector (ptegu)

The ptegu sector contains emissions from EGUs in the 2012 point source inventory that could be matched to units found in the National Electric Energy Database System (NEEDS) v5.14 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 IPM projects future emissions for the EGUs defined in the NEEDS database. In future year modeling cases, emissions for sources in the ptegu sector are fully replaced with IPM outputs. Sources not matched to units found in NEEDS are placed into the pt_oilgas or ptnonipm sectors and are projected to the future year using projection and control factors appropriate for their source categories. It is important that the matching between the NEI and NEEDS database be as complete as possible because there can be double-counting of emissions in future year modeling scenarios if emissions for units are projected by IPM are not properly matched to the units in the point source inventory.

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

matches from EIS units to the ORIS facility codes and boiler IDs used to reference the CEMS data. Some additional matches were made at the release point level in the emissions modeling platform.

For sources not matched to CEMS data (i.e., “non-CEMS” sources), daily emissions were computed from the annual emissions using average CEMS data profiles specific to fuel type and IPM region and based on input. To allocate emissions to each hour of the day, diurnal profiles were created using average CEMS data for heat input specific to fuel type and IPM region.

3.2.1.2 Non-IPM Sector (*pt_oilgas*)

The *pt_oilgas* sector was separated from the *ptnonipm* sector by selecting sources with specific North American Industry Classification System (NAICS) codes shown in Table 3-4. The emissions and other source characteristics in the *pt_oilgas* sector are submitted by states, while EPA developed a dataset of nonpoint oil and gas emissions for each county in the U.S. with oil and gas activity that was available for states to use. Nonpoint oil and gas emissions can be found in the *np_oilgas* sector. More information on the development of the 2011 oil and gas emissions can be found in Section 3.20 of the 2011NEIv2 TSD.

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

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

3.2.1.3 Non-IPM Sector (*ptnonipm*)

Except for some minor exceptions, the non-IPM (*ptnonipm*) sector contains the point sources that are not in the *ptegu* or *pt_oilgas* sectors. For the most part, the *ptnonipm* sector reflects the non-EGU sources of the NEI point inventory; however, it is likely that some small low-emitting EGUs not matched to the NEEDS database or to CEMS data are in the *ptnonipm* sector. The sector also includes some ethanol plants that have been identified by EPA and require special treatment in the future cases as they are impacted by mobile source rules.

The *ptnonipm* sector contains a small amount of fugitive dust PM emissions from vehicular traffic on paved or unpaved roads at industrial facilities, coal handling at coal mines, and grain elevators. Sources

with state/county FIPS code ending with “777” are in the NEI but are not included in any modeling sectors. These sources typically represent mobile (temporary) asphalt plants that are only reported for some states, and are generally in a fixed location for only a part of the year and are therefore difficult to allocate to specific places and days as is needed for modeling. Therefore, these sources are dropped from the point-based sectors in the modeling platform.

3.2.2 Day-Specific Point Source Fires (*ptfire*)

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

The point source day-specific emission estimates for 2012 fires rely on SMARTFIRE 2 (Sullivan, et al., 2008), which uses the National Oceanic and Atmospheric Administration’s (NOAA’s) Hazard Mapping System (HMS) fire location information as input. Additional inputs include the CONSUMEv3.0 software application (Joint Fire Science Program, 2009) and the Fuel Characteristic Classification System (FCCS) fuel-loading database to estimate fire emissions from wildfires and prescribed burns on a daily basis. The method involves the reconciliation of ICS-209 reports (Incident Status Summary Reports) and GeoMAC Shapefiles with satellite-based fire detections to determine spatial and temporal information about the fires. A functional diagram of the SMARTFIRE 2 process of reconciling fires with ICS-209 reports is available in the documentation (Raffuse, et al., 2007). Once the fire reconciliation process is completed, the emissions are calculated using the U.S. Forest Service’s CONSUMEv3.0 fuel consumption model and the FCCS fuel-loading database in the BlueSky Framework (Ottmar, et. al., 2007).

The SMOKE-ready “ORL” inventory files created from the raw daily fires contain both CAPs and HAPs. The BAFM HAP emissions from the inventory were obtained using VOC speciation profiles (i.e., a “no-integrate noHAP” use case). For the 2012 modeling, fires over 20,000 acres in a single day were split into the respective grid cells that they overlapped. The idea of this was to prevent all emissions from a very large fire going into a single grid cell, when in reality the fire was more dispersed than a single point. The large fires were each projected as a circle over the area centered on the specified latitude and longitude, and then apportioned into the grid cells they overlapped. The area of each of the “subfires” was computed in proportion to the overlap with that grid cell. These “subfires” were given new names that were the same as the original, but with “_a”, “_b”, “_c”, and “_d” appended as needed.

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

Many modeling platform sectors were created from the 2011NEIv2 nonpoint category. This section describes the stationary nonpoint sources. Note that locomotives, C1 and C2 CMV, and C3 CMV are also included the 2011NEIv2 nonpoint data category, but are mobile sources and are placed into the *c1c2rail* and *c3marine* sectors, respectively. The 2011NEIv2 TSD includes documentation for the nonpoint data category of the 2011NEIv2.

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

In the rest of this section, each of the platform sectors into which the 2011 nonpoint NEI was divided is described, along with any changes made to these data for this study.

3.2.3.1 Area Fugitive Dust Sector (*afdust*)

The area-source fugitive dust (*afdust*) sector contains PM₁₀ 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 with a script that applies land use-based gridded transport fractions followed by another script that zeroes out emissions for days on which at least 0.01 inches of precipitation occurs or there is snow cover on the ground. The land use data used to reduce the NEI emissions determines the amount of emissions that are subject to transport. This methodology is discussed in (Pouliot, et al., 2010), https://www3.epa.gov/ttn/chief/conference/ei19/session9/pouliot_pres.pdf, and in “Fugitive Dust Modeling for the 2008 Emissions Modeling Platform” (Adelman, 2012). Both the transport fraction and meteorological adjustments are based on the gridded resolution of the platform (e.g., 12km grid cells); therefore, different emissions will result if the process were applied to different grid resolutions. A limitation of the transport fraction approach is the lack of monthly variability that would be expected with seasonal changes in vegetative cover. While wind speed and direction are not accounted for in the emissions processing, the hourly variability due to soil moisture, snow cover and precipitation is accounted for in the subsequent meteorological adjustment.

3.2.3.2 Agricultural Ammonia Sector (*ag*)

The agricultural NH₃ (*ag*) sector includes livestock and agricultural fertilizer application emissions from the 2011NEIv2 nonpoint inventory. The livestock and fertilizer emissions in this sector are based only on the SCCs starting with 2805 and 2801. The livestock SCCs are related to beef and dairy cattle, poultry production and waste, swine production, waste from horses and ponies, and production and waste for sheep, lambs, and goats. The fertilizer SCCs consist of 15 specific types of ammonia-based fertilizer and one for miscellaneous fertilizers. The “*ag*” sector includes all of the NH₃ emissions from fertilizer from the NEI. However, the “*ag*” sector does not include all of the livestock ammonia emissions, as there are also a small amount of NH₃ emissions from livestock feedlots in the *ptnonipm* inventory (as point sources) in California (175 tons) and Wisconsin (125 tons). The annual agricultural burning estimates are treated as monthly values for modeling. The annual values in the 2011NEIv2 were split into monthly emissions by aggregating the data up to monthly values from daily estimates of emissions.

3.2.3.3 Agricultural fires (*agfire*)

The agricultural fire (*agfire*) sector contains emissions from agricultural fires. These emissions were placed into the sector based on their SCC code. All SCCs starting with 28015 were included. The first three levels of descriptions for these SCCs are: Fires - Agricultural Field Burning; Miscellaneous Area Sources; Agriculture Production - Crops - as nonpoint; Agricultural Field Burning - whole field set on fire. The SCC 2801500000 does not specify the crop type or burn method, while the more specific SCCs

specify field or orchard crops, and in some cases the specific crop being grown. For more information on how emissions for agricultural fires were developed in the 2011NEIv2, see Section 5.2 of the 2011NEIv2 TSD.

3.2.3.4 Nonpoint Oil-gas Sector (*np_oilgas*)

The nonpoint oil and gas (*np_oilgas*) sector contains onshore and offshore oil and gas emissions. EPA estimated emissions for all counties with 2011 oil and gas activity data with the Oil and Gas Tool, and many S/L/T agencies also submitted nonpoint oil and gas data. 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. For more information on the development of the oil and gas emissions in the 2011NEIv2, see Section 3.20 of the 2011NEIv2 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 chimneas. Free standing woodstoves and inserts are further differentiated into three categories: conventional (not EPA certified); EPA certified, catalytic; and EPA certified, noncatalytic. Generally speaking, the conventional units were constructed prior to 1988. Units constructed after 1988 had to meet EPA emission standards and they are either catalytic or non-catalytic. For more information on the development of the residential wood combustion emissions, see Section 3.14 of the 2011NEIv2 TSD.

3.2.3.6 Other Nonpoint Sources (*nonpt*)

Stationary nonpoint sources that were not subdivided into the *afdst*, *ag*, *np_oilgas*, or *rwc* sectors were assigned to the “*nonpt*” sector. The types of sources in the *nonpt* sector include:

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

- miscellaneous area sources such as cremation, hospitals, lamp breakage, and automotive repair shops.

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

3.2.4 Biogenic Sources (beis)

Biogenic emissions were developed using the Biogenic Emission Inventory System, version 3.61 (BEIS3.61) within SMOKE. This was an update from the emissions in the 2011 modeling platform that used BEIS 3.14, and from the 2011NEIv2 that used BEIS 3.60. Like BEIS 3.14, BEIS3.61 creates gridded, hourly, model-species emissions from vegetation and soils. It estimates CO, VOC (most notably isoprene, terpene, and sesquiterpene), and NO emissions for the contiguous U.S. and for portions of Mexico and Canada.

Updates in BEIS3.61 include the incorporation of Version 4 of the Biogenic Emissions Land use Database (BELD4), and the incorporation of a canopy model to estimate leaf-level temperatures (Pouliot and Bash, 2015). BELD4 is based on an updated version of the USDA-USFS Forest Inventory and Analysis (FIA) vegetation speciation based data from 2002 to 2013 from the Forest Inventory and Analysis version 5.1. Canopy coverage is based on the Landsat satellite NLCD product from 2001, since no canopy product was developed for the 2006 NLCD. The FIA includes approximately 250,000 representative plots of species fraction data that are within approximately 75 km of one another in areas identified as forest by the NLCD canopy coverage. The 2006 NLCD provides land cover information with a native data grid spacing of 30 meters. For land areas outside the conterminous United States, 500 meter grid spacing land cover data from the Moderate Resolution Imaging Spectroradiometer (MODIS) is used.

BEIS version 3.61 includes a two-layer canopy model. Layer structure varies with light intensity and solar zenith angle. Both layers of the canopy model 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 3.61

Variable	Description
LAI	leaf-area index
PRSFC	surface pressure
Q2	mixing ratio at 2 m
RC	convective precipitation per met TSTEP
RGRND	solar rad reaching sfc
RN	nonconvective precipitation per met TSTEP
RSTOMI	inverse of bulk stomatal resistance

Variable	Description
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

3.2.5 Mobile Sources (*onroad, onroad_ca_adj, nonroad, c1c2rail, c3marine*)

Onroad mobile sources include emissions 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. The sources are further divided between diesel and gasoline vehicles. The sector characterizes emissions from off-network processes (e.g. starts, hot soak, and extended idle) and on-network processes (i.e., from vehicles moving along the roads).

The 2012 onroad emissions are in the onroad sector. Except for California, all onroad emissions are generated using the SMOKE-MOVES emissions modeling framework that leverages MOVES generated outputs (<http://www.epa.gov/otaq/models/moves/>) and hourly meteorology. All tribal data from the mobile sectors have been dropped because the (1) emissions are small, (2) the emissions could be double-counted with state-provided onroad emissions, (3) all tribal data was developed using the older model MOBILE6, and (4) because spatial surrogate data at the tribal level is not currently available. Emissions for onroad (excluding refueling), nonroad and c1c2rail sources in California are based on data provided by the California Air Resources Board (CARB).

The locomotive and commercial marine vessel (CMV) emissions are divided into two nonroad sectors: “c1c2rail” and “c3marine”. The c1c2rail sector includes all railway and most rail yard emissions as well as the gasoline and diesel-fueled Class 1 and Class 2 CMV emissions. The c3marine sector emissions contain the larger residual fueled ocean-going vessel Class 3 CMV emissions that are within state waters, while those outside of state waters are in the othpt sector. All nonroad emissions are treated as county-specific low-level emissions (i.e., they are released into model layer 1).

3.2.5.1 Onroad (*onroad*)

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

To develop the onroad mobile source emissions for the continental U.S., EPA used a modeling framework that took into account the temperature sensitivity of the on-road emissions. Specifically, EPA used MOVES inputs for representative counties, vehicle miles traveled (VMT) and vehicle population (VPOP) 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 information available from meteorology modeling used for air quality modeling.

SMOKE-MOVES requires that emission rate “lookup” tables be generated by MOVES which 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 2012-specific emission factors by temperature and speed for a series of “representative counties,” to which every other county was mapped. Using the MOVES emission rates, SMOKE selects appropriate emissions rates for each county, hourly temperature, SCC, and speed bin and multiplied the emission rate by activity (VMT (vehicle miles travelled), VPOP (vehicle population)), or HOTELING (hours of extended idle) to produce emissions. These calculations were 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 MOVES 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.

To generate the 2012 onroad mobile source emissions, SMOKE-MOVES was run using 2011 activity data and 2012 emission factors. This decision was made after an analysis of VMT trends from 2011 to 2012 showed that there was very little change between the two years. Onroad emissions for California were created through a hybrid approach of combining state-supplied annual emissions (from the 2011NEIv2) with EPA developed SMOKE-MOVES data. Through this approach, the platform was able to reflect California’s unique rules, while leveraging the more detailed SCCs and the highly resolved spatial patterns, temporal patterns, and speciation from SMOKE-MOVES. The California-provided year 2011 emissions were adjusted to a 2012 inventory using the ratio of 2012 to 2011 factors retrieved from the EMFAC 2011 web data base (<http://www.arb.ca.gov/emfac/>). Factors were applied by air basin, fuel, and pollutant. NH3 used CO ratios (CARB did not provide NH3); VOC and all VOC HAPs used TOG ratios. Some counties are in multiple air basins, but for simplicity, each county was assigned to a single air basin for the projections. There are 15 air basins in California.

3.2.5.2 NMIM-Based Nonroad Mobile Sources (nonroad)

The nonroad equipment emissions were developed by obtaining monthly totals based on a run of NMIM (EPA, 2005) for year 2012. All nonroad emissions are compiled at the county/SCC level. NMIM creates the nonroad emissions on a month-specific basis that accounts for temperature, fuel types, and other variables that vary by month. The nonroad sector includes monthly exhaust, evaporative and refueling emissions from nonroad engines (not including commercial marine, aircraft, and locomotives) that EPA derived from NMIM for all states except California and Texas. Additional details on the development of the 2011NEIv2 nonroad emissions, which provided the starting databases for the 2012 run, are available in Section 4.5 the 2011NEIv2 TSD.

Texas year 2011 nonroad emissions were also submitted to the NEI. The 2011NEIv1 nonroad annual inventory emissions values were converted to monthly values by using EPA's NMIM monthly inventories to compute monthly ratios by county, SCC7, mode, and poll. They were then projected to 2012 using trend lines based on the 2012 NMIM run as compared to EPA 2011 NMIM run.

California year 2011 nonroad emissions were submitted to the 2011NEIv2 and are also documented in a staff report (ARB, 2010a) and are used to represent year 2012 emissions. The nonroad sector emissions in California were developed using a modular approach and include all rulemakings and updates in place by December 2010. These emissions were developed using Version 1 of the CEPAM which supports various California off-road regulations such as in-use diesel retrofits (ARB, 2007), Diesel Risk-Reduction Plan (ARB, 2000) and 2007 State Implementation Plans (SIPS) for the South Coast and San Joaquin Valley air basins (ARB, 2010b).

The CARB-supplied 2011NEIv2 nonroad annual inventory emissions values were converted to monthly values by using the aforementioned EPA NMIM monthly inventories to compute monthly ratios by county, SCC7 (fuel, engine type, and equipment type group), mode, and pollutant. SCC7 ratios were used because the SCCs in the CARB inventory did not align with many of the SCCs in EPA NMIM inventory. By aggregating up to SCC7, the two inventories had a more consistent coverage of sources. Some VOC emissions were added to California to account for situations when VOC HAP emissions were included in the inventory, but there were no VOC emissions. These additional VOC emissions were computed by summing benzene, acetaldehyde, and formaldehyde for the specific sources.

3.2.5.3 Category 1 and 2 Commercial Marine and Locomotive (c1c2rail)

The c1c2rail sector contains locomotive and smaller CMV sources, except for railway maintenance locomotives and C3 CMV sources outside of the Midwest states. The "c1c2" portion of this sector name refers to the Class 1 and 2 CMV emissions, not the railway emissions. Railway maintenance emissions are included in the nonroad sector. The C3 CMV emissions are in the c3marine sector. All emissions in this sector are annual and at the county-SCC resolution. The emissions include the offshore portion of the C1 and C2 commercial marine sources, including fishing vessels and oil rig support vessels in the Gulf of Mexico. Emissions that occur outside of state waters are not assigned to states. The emissions for c1c2rail sector are equivalent to those in the 2011NEIv2 nonpoint inventory. For more information on CMV sources in the NEI, see Section 4.3 of the 2011NEIv2 TSD. For more information on locomotives, see Section 4.4 of the 2011NEIv2 TSD.

3.2.5.4 Category 3 commercial marine (c3marine)

The Category 3 (C3) CMV sources in the c3marine sector of the 2011v6.2 platform run on residual oil and use the SCCs 2280003100 and 2280003200 for port and underway emissions, respectively, and are consistent with the 2011NEIv2. Emissions for this sector use state-submitted values and EPA-developed emissions in areas where states did not submit. This sector only includes emissions in state-waters and the emissions are treated as nonpoint sources. Thus, the c3marine emissions are placed in layer 1 and allocated to grid cells using spatial surrogates. C3 CMV emissions outside of state waters, and non-U.S. emissions farther offshore than U.S. waters are processed in the "othpt" sector. C3 emissions for Canada were provided by Environment Canada for year 2010 and are found in the other inventory.

The EPA-estimated C3 CMV emissions for year 2011 were developed based on a 4-km resolution ASCII raster format dataset that preserves shipping lanes. This dataset has been used since the Emissions Control Area-International Marine Organization (ECA-IMO) project began in 2005, although it was then known as the Sulfur Emissions Control Area (SECA). The ECA-IMO emissions consist of large marine diesel engines (at or above 30 liters/cylinder) that until recently were allowed to meet relatively modest emission requirements and as a result these ships would often burn residual fuel in that region. The emissions in this sector are comprised of primarily foreign-flagged ocean-going vessels, referred to as C3 CMV ships. The c3marine inventory includes these ships in several intra-port modes (i.e., cruising, hoteling, reduced speed zone, maneuvering, and idling) and an underway mode, and includes near-port auxiliary engine emissions.

An overview of the C3 ECA Proposal to the International Maritime Organization (EPA-420-F-10-041, August 2010) project and future-year goals for reduction of NO_x, SO₂, and PM C3 emissions can be found at: <http://www.epa.gov/oms/regs/nonroad/marine/ci/420r09019.pdf>. The resulting ECA-IMO coordinated strategy, including emission standards under the Clean Air Act for new marine diesel engines with per-cylinder displacement at or above 30 liters, and the establishment of Emission Control Areas is available from <https://www.epa.gov/regulations-emissions-vehicles-and-engines/international-standards-reduce-emissions-marine-diesel>. The base year for the ECA inventory is 2002 and consists of these CAPs: PM₁₀, PM_{2.5}, CO, CO₂, NH₃, NO_x, SO_x (assumed to be SO₂), and hydrocarbons (assumed to be VOC). EPA developed regional growth (activity-based) factors that were applied to create the 2012 inventory from the 2002 data. The geographic regions listed in the table are shown in Figure 3-1. * The East Coast and Gulf Coast regions were divided along a line roughly through Key Largo (longitude 80° 26' West). Technically, the EEZ FIPS are not really “FIPS” state-county codes, but are treated as such in the inventory and emissions processing. The Canadian near-shore emissions were assigned to province-level FIPS codes and paired those to region classifications for British Columbia (North Pacific), Ontario (Great Lakes) and Nova Scotia (East Coast).

* Technically, these are not really “FIPS” state-county codes, but are treated as such in the inventory and emissions processing.

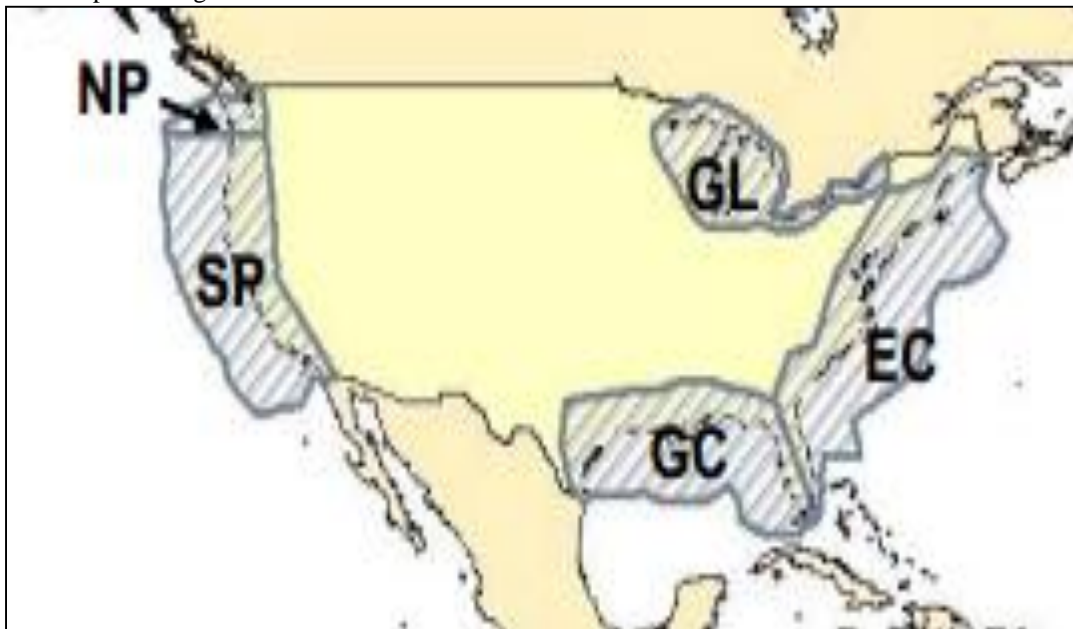


Figure 3-1. Illustration of regional modeling domains in ECA-IMO study

The emissions were converted to SMOKE point source inventory format as described in <https://www3.epa.gov/ttn/chief/conference/ei17/session6/mason.pdf>, which allows for the emissions to be allocated to modeling layers above the surface layer. As described in the paper, the ASCII raster dataset was converted to latitude-longitude, mapped to state/county FIPS codes that extended up to 200 nautical miles (nm) from the coast, assigned stack parameters, and monthly ASCII raster dataset emissions were used to create monthly temporal profiles. All non-US, non-EEZ emissions (i.e., in waters considered outside of the 200 nm EEZ, and hence out of the U.S. and Canadian ECA-IMO controllable domain) were simply assigned a dummy state/county FIPS code=98001, and were projected to year 2011 using the “Outside ECA” factors.

The assignment of U.S. state/county FIPS codes was restricted to state-federal water boundaries data from the Mineral Management Service (MMS) that extend approximately 3 to 10 nautical miles (nm) off shore. Emissions outside the 3 to 10 mile MMS boundary, but within the approximately 200 nm EEZ boundaries in Figure 3-1, were projected to year 2012 using the same regional adjustment factors as the U.S. emissions; however, the state/county FIPS codes were assigned as “EEZ” codes and those emissions processed in the “othpt” sector. Note that state boundaries in the Great Lakes are an exception, extending through the middle of each lake such that all emissions in the Great Lakes are assigned to a U.S. county or Ontario. This holds true for Midwest states and other states such as Pennsylvania and New York. The classification of emissions to U.S. and Canadian FIPS codes was needed to avoid double-counting of C3 CMV U.S. emissions in the Great Lakes because, as discussed in the previous section, all CMV emissions in the Midwest RPO are processed in the “c1c2rail” sector.

The SMOKE-ready data have been cropped from the original ECA-IMO entire northwestern quarter of the globe to cover only the large continental U.S. 36-km “36US1” air quality model domain, the largest domain used by EPA in recent years.

The original ECA-IMO inventory did not delineate between ports and underway emissions (or other C3 modes such as hoteling, maneuvering, reduced-speed zone, and idling). However, a U.S. ports spatial surrogate dataset was used to assign the ECA-IMO emissions to ports and underway SCCs 2280003100 and 2280003200, respectively. This had no effect on temporal allocation or speciation because all C3 CMV emissions, unclassified/total, port and underway, share the same temporal and speciation profiles.

For California, the ECA-IMO 2011 emissions were scaled to match those provided by CARB for year 2011 because CARB has had distinct projection and control approaches for this sector since 2002. These CARB C3 CMV emissions are documented in a staff report available at:

<http://www.arb.ca.gov/regact/2010/offroadlsi10/offroadisor.pdf>. The CMV emissions obtained from the CARB nonroad mobile dataset include the 2011 regulations to reduce emissions from diesel engines on commercial harbor craft operated within California waters and 24 nautical miles of the California shoreline. These emissions were developed using Version 1 of the California Emissions Projection Analysis Model (CEPAM) that supports various California off-road regulations. The locomotive emissions were obtained from the CARB trains dataset “ARMJ_RF#2002_ANNUAL_TRAINS.txt”. Documentation of the CARB offroad mobile methodology, including c1c2rail sector data, is provided at: http://www.arb.ca.gov/msei/categories.htm#offroad_motor_vehicles.

3.2.6 Emissions from Canada, Mexico and Offshore Drilling Platforms (othpt, othar, othon, othafdust)

The emissions from Canada, Mexico, and non-U.S. offshore Class 3 Commercial Marine Vessels (C3 CMV) and drilling platforms are included as part of four emissions modeling sectors: othpt, othar, othon, and othafdust. 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”, and “on” for onroad mobile.

The ECA-IMO-based C3 CMV were processed in the othpt sector. These C3 CMV emissions include those assigned to U.S. federal waters, those assigned to the Exclusive Economic Zone (EEZ; defined as those emissions beyond the U.S. Federal waters approximately 3-10 miles offshore, and extending to about 200 nautical miles from the U.S. coastline), along with any other offshore emissions. These emissions are developed in the same way as the EPA-dataset described in the c3marine sector described above. Emissions in U.S. waters are aggregated into large regions and included in the 2011NEIv2 using special FIPS codes. Because these emissions are treated as point sources, shipping lane routes can be preserved and they may be allocated to air quality model layers higher than layer 1.

For Canadian point sources, 2010 emissions provided by Environment Canada were used. Note that VOC was not provided for Canadian point sources, but any VOC emissions were speciated into CB05 species. Temporal profiles and speciated emissions were also provided. Point sources in Mexico were compiled based on the Inventario Nacional de Emisiones de Mexico, 2008 (ERG, 2014a). The point source emissions in the 2008 inventory 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. Note that there are no explicit HAP emissions in this inventory.

The othpt sector also includes point source offshore oil and gas drilling platforms that are beyond U.S. state-county boundaries in the Gulf of Mexico. For these offshore emissions, data from the 2011NEIv2 were used.

For Canada area, nonroad mobile, and onroad mobile sources, year-2010 emissions provided by Environment Canada were used, including C3 CMV emissions. The Canadian inventory included fugitive dust emissions that do not incorporate either a transportable fraction or meteorological-based adjustments. To properly account for these issues, a separate sector called othafdust was created and modeled using the same adjustments as are done for U.S. sources. Updated Shapefiles for creating spatial surrogates for Canada were also provided.

Area and nonroad mobile sources in Mexico were compiled from the Inventario Nacional de Emisiones de Mexico, 2008 (ERG, 2014a). The 2008 emissions were quality assured for completeness, SCC assignments were made when needed, the pollutants expected for the various processes were reviewed, and adjustments were made to ensure that PM₁₀ was greater than or equal to PM_{2.5}. The resulting inventory was written using English units to the nonpoint FF10 format that could be read by SMOKE. Note that unlike the U.S. inventories, there are no explicit HAPs in the nonpoint or nonroad inventories for Canada and Mexico, and therefore all HAPs are created from speciation.

Onroad mobile sources in Mexico were compiled from the Inventario Nacional de Emisiones de Mexico, 2008 (ERG, 2014a). SCCs compatible with the 2011NEIv2 were assigned to the 2008 onroad mobile

source emissions in Mexico, and it was enforced that PM_{10} be greater than or equal to $PM_{2.5}$. Quality assurance of the onroad mobile source emissions data revealed that Baja California, Michoacán, and Nuevo León had significantly high per capita emissions for all pollutants and should be considered to be outliers. The emissions for these states were replaced with values computed based on the average per capita emissions for the remaining states. The data were written using English units to the nonpoint FF10 format that could be read by SMOKE. Note that unlike the U.S. inventories, there are no explicit HAPs in the onroad mobile inventories for Canada and Mexico, and therefore all HAPs are created from speciation.

3.2.7 *SMOKE-ready non-anthropogenic chlorine inventory*

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

3.3 Emissions Modeling Summary

CMAQ 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 CMAQ, it is necessary to “pre-process” the emission inventories for each of 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. Thus, 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 distribution of the sources does not need to be provided as an input.

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

3.3.1 *The SMOKE Modeling System*

SMOKE version 3.6.5 was used to pre-process the raw emissions inventories into emissions inputs for CMAQ. SMOKE executables and source code are available from the Community Multiscale Analysis System (CMAS) Center at <http://www.cmascenter.org>. Additional information about SMOKE is available from <http://www.smoke-model.org>. For sectors that have plume rise, the in-line emissions capability of the air quality models was used, which allows the creation of source-based and two-dimensional gridded emissions files that are much smaller than full three-dimensional gridded emissions files. For quality assurance of the emissions modeling steps, emissions totals by specie for the entire model domain are output as reports that are then compared to reports generated by SMOKE on the input inventories to ensure that mass is not lost or gained during the emissions modeling process.

3.3.2 Key Emissions Modeling Settings

When preparing emissions for the air quality model, emissions for each sector are processed separately through SMOKE, and then the final merge program (Mrggrid) is run to combine the model-ready, sector-specific emissions across sectors. The SMOKE settings in the run scripts and the data in the SMOKE ancillary files control the approaches used by the individual SMOKE programs for each sector. Table 3-6 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.

Finally, the “plume rise” column indicates the sectors for which the “in-line” approach is used. These sectors are the only ones with emissions in aloft layers based on plume rise. The term “in-line” means that the plume rise calculations are done inside of the air quality model instead of being computed by SMOKE. The air quality model computes the plume rise using the stack data and the hourly air quality model inputs found in the SMOKE output files for each model-ready emissions sector. The height of the plume rise determines the model layer into which the emissions are placed. The othpt sector has only “in-line” emissions, meaning that all of the emissions are treated as elevated sources and there are no emissions for those sectors in the two-dimensional, layer-1 files created by SMOKE. Day-specific point fires are treated separately for CMAQ modeling in that fire plume rise is done within CMAQ itself. After plume rise is applied, there will be emissions in every layer from the ground up to the top of the plume.

Table 3-6. Key emissions modeling steps by sector

Platform sector	Spatial	Speciation	Inventory resolution	Plume rise
afdust	Surrogates	Yes	annual	
ag	Surrogates	Yes	annual	
agfire	Surrogates	Yes	monthly	
beis	Pre-gridded land use	in BEIS	computed hourly	
c1c2rail	Surrogates	Yes	annual	
c3marine	Surrogates	Yes	annual	
nonpt	Surrogates & area-to-point	Yes	annual	
nonroad	Surrogates & area-to-point	Yes	monthly	
np_oilgas	Surrogates	Yes	annual	
onroad	Surrogates	Yes	monthly activity, computed hourly	
othafdust	Surrogates	Yes	annual	
othar	Surrogates	Yes	annual	

Platform sector	Spatial	Speciation	Inventory resolution	Plume rise
othon	Surrogates	Yes	annual	
othpt	Point	Yes	annual	in-line
pt_oilgas	Point	Yes	annual	in-line
ptegu	Point	Yes	daily & hourly	in-line
ptfire	Point	Yes	daily	in-line
ptnonipm	Point	Yes	annual	in-line
rcw	Surrogates	Yes	annual	

SMOKE has the option of grouping sources so that they are treated as a single stack when computing plume rise. For the 2012 modeling case, no grouping was performed because grouping combined with “in-line” processing will not give identical results as “offline” processing (i.e., when SMOKE creates 3-dimensional files). This occurs when stacks with different stack parameters or lat/lons are grouped, thereby changing the parameters of one or more sources. The most straightforward way to get the same results between in-line and offline is to avoid the use of grouping.

3.3.3 Spatial Configuration

For this study, SMOKE was run for the smaller 12-km CONTinental United States “CONUS” modeling domain (12US2) shown in Figure 3-2 and boundary conditions were obtained from a 2011 run of GEOS-Chem. The grid used a Lambert-Conformal projection, with Alpha = 33, Beta = 45 and Gamma = -97, with a center of X = -97 and Y = 40. Later sections provide details on the spatial surrogates and area-to-point data used to accomplish spatial allocation with SMOKE.

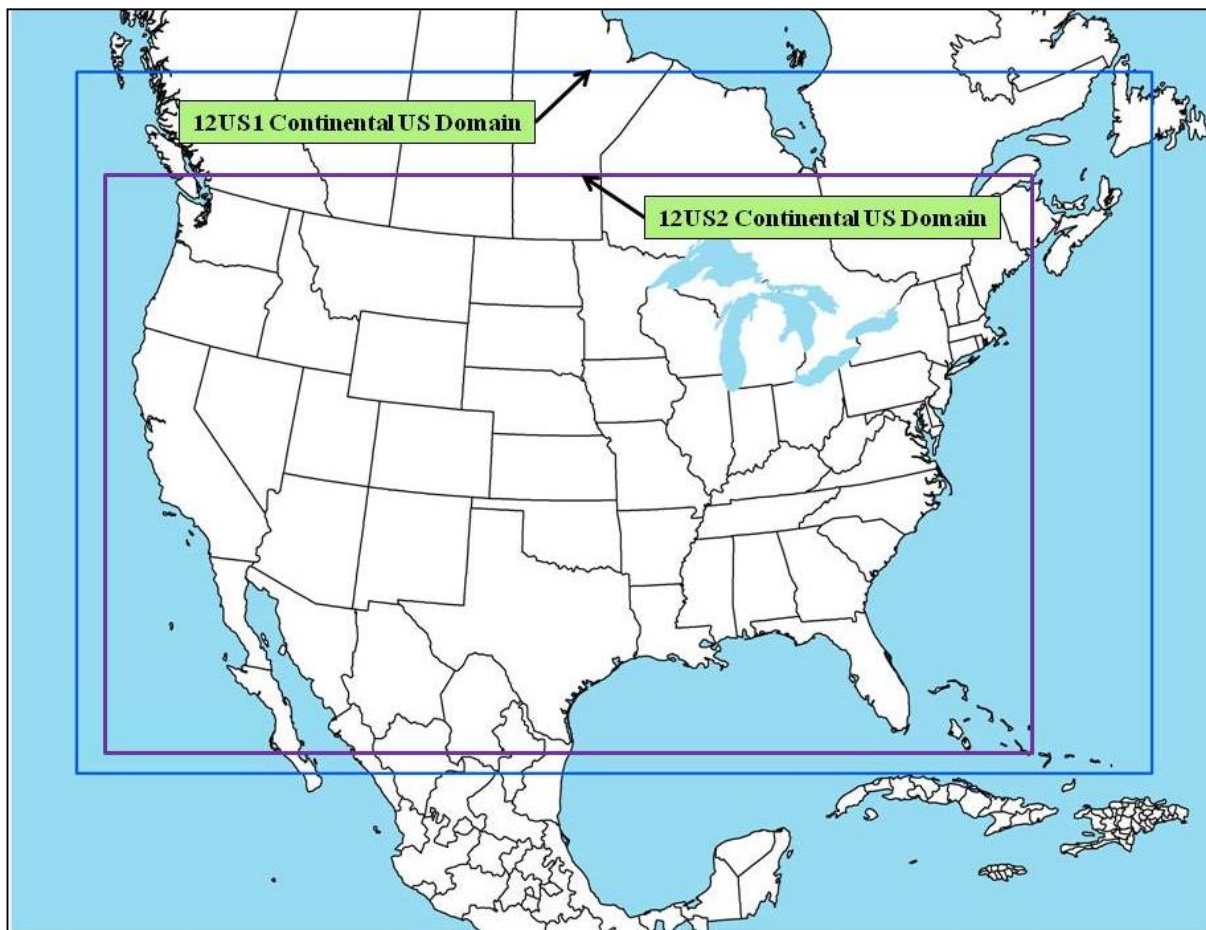


Figure 3-2. CMAQ Modeling Domain

3.3.4 Chemical Speciation Configuration

The emissions modeling step for chemical speciation creates the “model species” needed by the air quality model for a specific chemical mechanism. These model species are either individual chemical compounds or groups of species, called “model species.” The chemical mechanism used for this study is the CB05 mechanism (Yarwood, 2005). The versions of CMAQ used for this study include secondary organic aerosol (SOA) and HONO enhancements. The PM_{2.5} model species are those associated with the CMAQ Aerosol Module, version 6 (AE6). This modeling case uses CB05 with speciation profile mappings that were updated in February, 2015. Table 3-7 lists the model species produced by SMOKE for use in CMAQ. Speciation profiles and cross-references are available in the SMOKE input files for the 2011v6.2 emissions modeling platform.

Table 3-7. Emission model species produced for CB05 with SOA for CMAQ 5.0.2

Inventory Pollutant	Model Species	Model species description
Cl ₂	CL2	Atomic gas-phase chlorine
HCl	HCL	Hydrogen Chloride (hydrochloric acid) gas
CO	CO	Carbon monoxide
NO _x	NO	Nitrogen oxide
	NO2	Nitrogen dioxide
	HONO	Nitrous acid
SO ₂	SO2	Sulfur dioxide
	SULF	Sulfuric acid vapor
NH ₃	NH3	Ammonia
	NH3_FERT	Ammonia from fertilizer
VOC	ALD2	Acetaldehyde
	ALDX	Propionaldehyde and higher aldehydes
	BENZENE	Benzene (not part of CB05)
	CH4	Methane
	ETH	Ethene
	ETHA	Ethane
	ETOH	Ethanol
	FORM	Formaldehyde
	IOLE	Internal olefin carbon bond (R-C=C-R)
	ISOP	Isoprene
	MEOH	Methanol
	NVOL	Non-volatile compounds
	OLE	Terminal olefin carbon bond (R-C=C)
	PAR	Paraffin carbon bond
	SESQ	Sequiterpenes (from biogenics only)
	TERP	Terpenes
	TOL	Toluene and other monoalkyl aromatics
	UNR	Unreactive
	XYL	Xylene and other polyalkyl aromatics
PM ₁₀	PMC	Coarse PM > 2.5 microns and ≤ 10 microns
PM _{2.5}	PEC	Particulate elemental carbon ≤ 2.5 microns
	PNO3	Particulate nitrate ≤ 2.5 microns
	POC	Particulate organic carbon (carbon only) ≤ 2.5 microns
	PSO4	Particulate Sulfate ≤ 2.5 microns
	PMFINE ⁷	Other particulate matter ≤ 2.5 microns
Sea-salt species (non – anthropogenic) ⁸	PCL	Particulate chloride
	PNA	Particulate sodium

It should be noted that the BENZENE model species is not part of CB05 in that the concentrations of BENZENE do not provide any feedback into the chemical reactions (i.e., it is not “inside” the chemical mechanism). Rather, benzene is used as a reactive tracer and as such is impacted by the CB05 chemistry. BENZENE, along with several reactive CB05 species (such as TOL and XYL) plays a role in SOA formation.

⁷ For CMAQ 5.0, PMFINE is speciated into a finer set of PM components. Listed in this table are the AE5 species

⁸ These emissions are created outside of SMOKE

The TOG and PM_{2.5} speciation factors that are the basis of the chemical speciation approach were developed from the SPECIATE 4.4 database (<https://www.epa.gov/air-emissions-modeling/speciate-version-45-through-40>), which is EPA's repository of TOG and PM speciation profiles of air pollution sources. However, a few of the profiles used in the v6.2 platform will be published in later versions of the SPECIATE database after the release of this documentation. The SPECIATE database development and maintenance is a collaboration involving EPA's ORD, OTAQ, and the Office of Air Quality Planning and Standards (OAQPS), in cooperation with Environment Canada (EPA, 2006a). 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}.

The speciation of VOC includes HAP emissions from the 2011NEIv2 in the speciation process. Instead of speciating VOC to generate all of the species listed in Table 3-7, emissions of four specific HAPs: benzene, acetaldehyde, formaldehyde and methanol (collectively known as "BAFM") from the NEI were "integrated" with the NEI VOC. The integration process (described in more detail below) 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 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. EPA believes that generally, the HAP emissions from the inventory are more representative of emissions of these compounds than their generation via VOC speciation.

The BAFM HAPs (benzene, acetaldehyde, formaldehyde and methanol) were chosen because, with the exception of BENZENE, they are the only explicit VOC HAPs in the base version of CMAQ 5.0.2 (CAPs only with chlorine chemistry) model. Explicit means that they are not lumped chemical groups like the other CB05 species. These "explicit VOC HAPs" are model species that participate in the modeled chemistry using the CB05 chemical mechanism. The use of these HAP emission estimates along with VOC is called "HAP-CAP integration". BENZENE was chosen because it is a model species in the base version of CMAQ 5.0.2, and there was a desire to keep its emissions consistent between multi-pollutant and base versions of CMAQ.

The integration of HAP VOC with VOC is a feature available in SMOKE for all inventory formats other than PTDAY (the format used for the ptfire sector). SMOKE allows the user to specify both the particular HAPs to integrate via the INVTABLE and the particular sources to integrate via the NHAPEXCLUDE file (which actually provides the sources to be *excluded* from integration⁹). 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 profiles¹⁰. 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

⁹ In SMOKE version 3.6.5, 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 BAFM or VOC, SMOKE will now raise an error.

¹⁰ 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 BAFM.

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. EPA considered CAP-HAP integration for all sectors and developed "integration criteria" for some of them.

The process of partial integration for BAFM is illustrated in Figure 3-3 and means that the BAFM records in the input inventories do not need to be removed from any sources in a partially integrated sector because SMOKE does this automatically using the INVTABLE configuration. For EBAFM integration, this process is identical to that shown in the figure except for the addition of ethanol (E) to the list of subtracted HAP pollutants. For full integration, the process would be very similar except that the NHAPEXCLUDE file would not be used and all sources in the sector would be integrated.

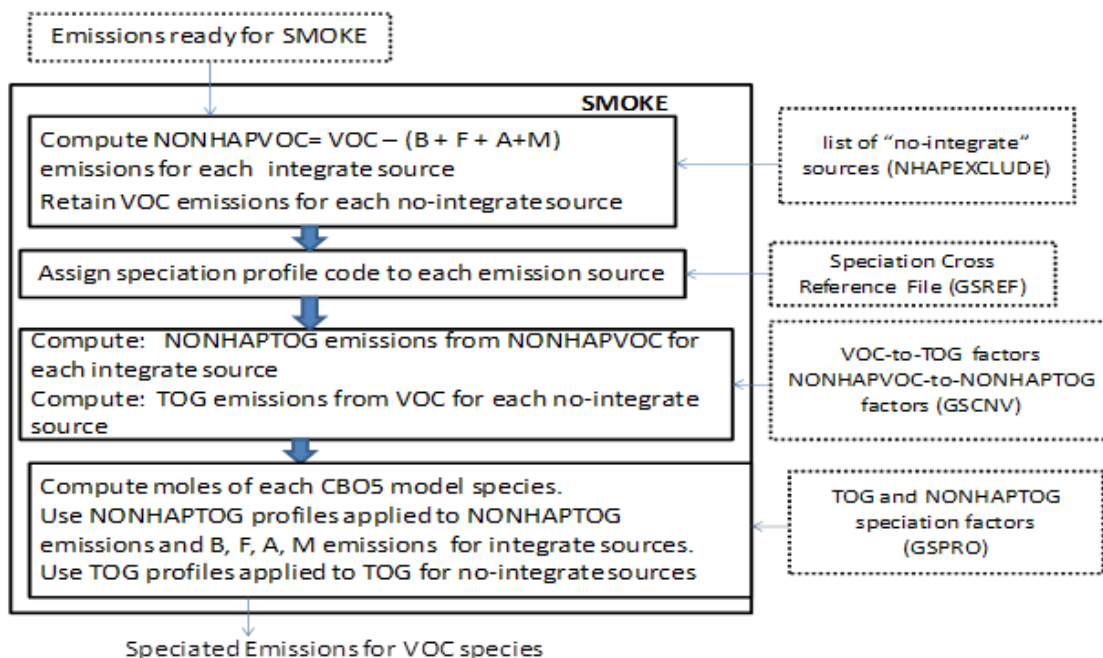


Figure 3-3. Process of integrating BAFM with VOC for use in VOC Speciation

In SMOKE, the INVTABLE allows the user to specify both the particular HAPs to integrate. Two different types of INVTABLE files are included for use with different sectors of the platform. For sectors that had no integration across the entire sector (see Table 3-8), EPA created a "no HAP use" INVTABLE in which the "KEEP" flag is set to "N" for BAFM pollutants. Thus, any BAFM pollutants in the inventory input into SMOKE are automatically dropped. This approach both avoids double-counting of these species and assumes that the VOC speciation is the best available approach for these species for sectors using this approach. The second INVTABLE, used for sectors in which one or more sources are integrated, causes SMOKE to keep the inventory BAFM 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. This type of INVTABLE is further differentiated into a version for those sectors that integrate BAFM and another for those that integrate EBAFM.

Table 3-8. Integration status of benzene, acetaldehyde, formaldehyde and methanol (BAFM) for each platform sector

Platform Sector	Approach for Integrating NEI emissions of Benzene (B), Acetaldehyde (A), Formaldehyde (F), Methanol (M), and Ethanol (E)
ptegu	No integration
ptnonipm	No integration
ptfire	No integration
othar	No integration
othon	No integration
ag	N/A – sector contains no VOC
agfire	Partial integration (BAFM)
afdust	N/A – sector contains no VOC
biog	N/A – sector contains no inventory pollutant "VOC"; but rather specific VOC species
c1c2rail	Partial integration (BAFM)
c3marine	Partial integration (BAFM)
nonpt	Partial integration (BAFM)
nonroad	Partial integration (BAFM)
np_oilgas	Partial integration (BAFM)
othpt	Partial integration (BAFM)
pt_oilgas	Partial integration (BAFM)
rwc	Partial integration (BAFM)
onroad	Full integration (benzene, 1,3 butadiene, formaldehyde, acetaldehyde, naphthalene, acrolein, ethyl benzene, 2,2,4-Trimethylpentane, hexane, propionaldehyde, styrene, toluene, xylene, MTBE)

SMOKE can compute speciation profiles from mixtures of other profiles in user-specified proportions. The combinations are specified in the GSPRO_COMBO ancillary file by pollutant (including pollutant mode, e.g., EXH__VOC), state and county (i.e., state/county FIPS code) and time period (i.e., month). This feature was used to speciate nonroad mobile and gasoline-related 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. SMOKE computes the resultant profile using the fraction of each specific profile assigned by county, month and emission mode.

The GSREF file indicates that a specific source uses a combination file with the profile code "COMBO". Because the GSPRO_COMBO file does not differentiate by SCC and there are various levels of integration across sectors, sector specific GSPRO_COMBO files are used. Different profile combinations are specified by the mode (e.g. exhaust, evaporative, refueling, etc.) by changing the pollutant name (e.g. EXH__NONHAPTOG, EVP__NONHAPTOG, RFL__NONHAPTOG). For the nonpt sector, BAFM integration is used.

Speciation profiles for use with BEIS are not included in SPECIATE. BEIS3.61 includes a species (SESQ) that was mapped to the CMAQ specie SESQT. The profile code associated with BEIS profiles for use with CB05 was "B10C5." For additional sector-specific details on VOC speciation for a variety of sectors, see Section 3.2.1.3 of the 2011v6.2 TSD (EPA, 2015a).

In addition to VOC profiles, the SPECIATE database also contains the PM_{2.5} speciated into both individual chemical compounds (e.g., zinc, potassium, manganese, lead), and into the “simplified” PM_{2.5} components used in the air quality model. For CMAQ 4.7.1 modeling, these “simplified” components (AE5) are all that is needed. Starting with CMAQ 5.0.1, a new thermodynamic equilibrium aerosol modeling tool (ISORROPIA) v2 mechanism was added that needs additional PM components (AE6), which are further subsets of PMFINE (see Table 3-9). The majority of the 2011 platform PM profiles come from the 911XX series which include updated AE6 speciation¹¹.

Table 3-9. PM model species: AE5 versus AE6

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

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

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 NO₂. For the mobile sources except for onroad (including nonroad, c1c2rail, c3marine, othor sectors) and for specific SCCs in othar and ptnonipm, the profile “HONO” splits NO_x into NO, NO₂, 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. MOVES2014 produces speciated

¹¹ The exceptions are 5674 (Marine Vessel – Marine Engine – Heavy Fuel Oil) used for c3marine and 92018 (Draft Cigarette Smoke – Simplified) used in nonpt.

NO, NO₂, and HONO by source, including emission factors for these species in the emission factor tables used by SMOKE-MOVES. Within MOVES, the HONO fraction is a constant 0.008 of NO_x. The NO fraction varies by heavy duty versus light duty, fuel type, and model year and equals 1 – NO – HONO. For more details on the NO_x fractions within MOVES, see <http://www.epa.gov/otaq/models/moves/documents/420r12022.pdf>.

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 hour. This process is typically done by applying temporal profiles to the inventories in this order: monthly, day of the week, and diurnal.

In SMOKE 3.6.5 and in the 2011v6.2 platform, more readable and flexible file formats are used for temporal profiles and cross references. The profiles and cross references were initially created by converting the 2011v6.1 platform temporal profiles into the new formats, and then any specific adjustments for the 2011v6.2 platform were made. 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
ag	Annual	Yes	met-based	all	Yes
agfire	Monthly		week	week	Yes
beis	Hourly		n/a	all	Yes
c1c2rail	Annual	Yes	mwdss	mwdss	
c3marine	Annual	Yes	aveday	aveday	
nonpt	Annual	Yes	week	week	Yes
nonroad	Monthly		mwdss	Mwdss	Yes
np_oilgas	Annual	yes	week	week	Yes
onroad	Annual & monthly ¹		all	all	Yes
onroad_ca_adj	Annual & monthly ¹		all	all	Yes
othafdust_adj	Annual	yes	week	week	
othar	Annual	yes	week	week	
othon	Annual	yes	week	week	
othpt	Annual	yes	mwdss	mwdss	
pt_oilgas	Annual	yes	mwdss	mwdss	Yes
ptegu	Daily & hourly		all	all	Yes
ptnonipm	Annual	yes	mwdss	mwdss	Yes
ptprescfire	Daily		all	all	Yes
ptwildfire	Daily		all	all	Yes
rcw	Annual	no	met-based	all	Yes

1. Note the annual and monthly “inventory” actually refers to the activity data (VMT and VPOP) for onroad. The actual emissions are computed on an hourly basis.

The following values are used in the table: The value “all” means that hourly emissions are computed for every day of the year and that emissions potentially have day-of-year variation. The value “week” means that hourly emissions computed for all days in one “representative” week, representing all weeks for each month. This means emissions have day-of-week variation, but not week-to-week variation within the month. The value “mwdss” means hourly emissions for one representative Monday, representative weekday (Tuesday through Friday), representative Saturday, and representative Sunday for each month. This means emissions have variation between Mondays, other weekdays, Saturdays and Sundays within the month, but not week-to-week variation within the month. The value “aveday” means hourly emissions computed for one representative day of each month, meaning emissions for all days within a month are the same. Special situations with respect to 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, 2012, which is intended to mitigate the effects of initial condition concentrations. The ramp-up period was 10 days (December 22-31, 2011). For most sectors, emissions from December 2012 were used to fill in surrogate emissions for the end of December 2011. In particular, December 2012 emissions

(representative days) were used for December 2011. For biogenic emissions, December 2011 emissions were processed using 2011 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 agfire, nonroad, onroad (for activity data), and ptegu.

3.3.5.1 Standard Temporal Profiles

Some sectors use straightforward temporal profiles not based on meteorology or other factors. For the agfire sector, the emissions were allocated to months by adding up the available values for each day of the month. For all agricultural burning, the diurnal temporal profile used reflected the fact that burning occurs during the daylight. This puts most of the emissions during the work day and suppresses the emissions during the middle of the night. A uniform profile for each day of the week was used for all agricultural burning emissions in all states, except for the following states that for which EPA used state-specific day of week profiles: Arkansas, Kansas, Louisiana, Minnesota, Missouri, Nebraska, Oklahoma, and Texas.

For the c1c2rail and c3marine sectors, emissions are allocated with flat monthly and day of week profiles, and most emissions are also allocated with flat hourly profiles. For the ptwildfire and ptprescfire sectors, the inventories are in the daily point fire format ORL PTDAY, so temporal profiles are only used to go from day-specific to hourly emissions.

For the nonroad sector, while the NEI only stores the annual totals, the modeling platform uses monthly inventories from output from NMIM. For California, a monthly inventory was created from CARB’s annual inventory using EPA-estimated NMIM monthly results to compute monthly ratios by pollutant and SCC7 and these ratios were applied to the CARB inventory to create a monthly inventory.

Updates were made to temporal profiles for the ptnonipm sector in the 2011v6.2 platform based on comments and data review by EPA staff. Temporal profiles for small airports (i.e., non-commercial) were updated to eliminate emissions between 10pm and 6am due to a lack of tower operations. Industrial process that are not likely to shut down on Sundays such as those at cement plants were 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. Some cross reference updates for temporalization of the np_oilgas sector were made in the 2011v6.2 platform based on comments received - np_oilgas sources were assigned to profiles that were 24 hours per day, 7 days a week.

3.3.5.2 Temporal Profiles for EGUs

The 2011NEIv2 annual EGU emissions are allocated to hourly emissions using the following 3-step methodology: annual value to month, month to day, and day to hour. Several updates were made to EGU temporalization in the 2011v6.2 platform. First, the CEMS data were processed using a tool that reviewed the data quality flags that indicate the data were not measured. Unmeasured data can cause erroneously high values to appear in the CEMS data. If the data were not measured at specific hours, and those values were found to be more the 3 times the annual mean for that unit, the data for those hours were replaced with annual mean values (Adelman, et al., 2012). These adjusted CEMS data were then used for the remainder of the temporalization process described below (see Figure 3-4 for an example).

Another update in the 2011v6.2 platform was the incorporation of winter and summer seasons into the development of the diurnal profiles as opposed to using data for the entire year. Analysis of the hourly CEMS data revealed that there were different diurnal patterns in winter versus summer in many areas. Typically a single mid-day peak is visible in the summer, while there are morning and evening peaks in the winter as shown in Figure 3-5. Finally, we identified specific CEMS sources as partial year reporters (e.g., those that only run in the summer) and allocated the difference between the CEMS and the annual emissions to months that did not have any CEMS data available.

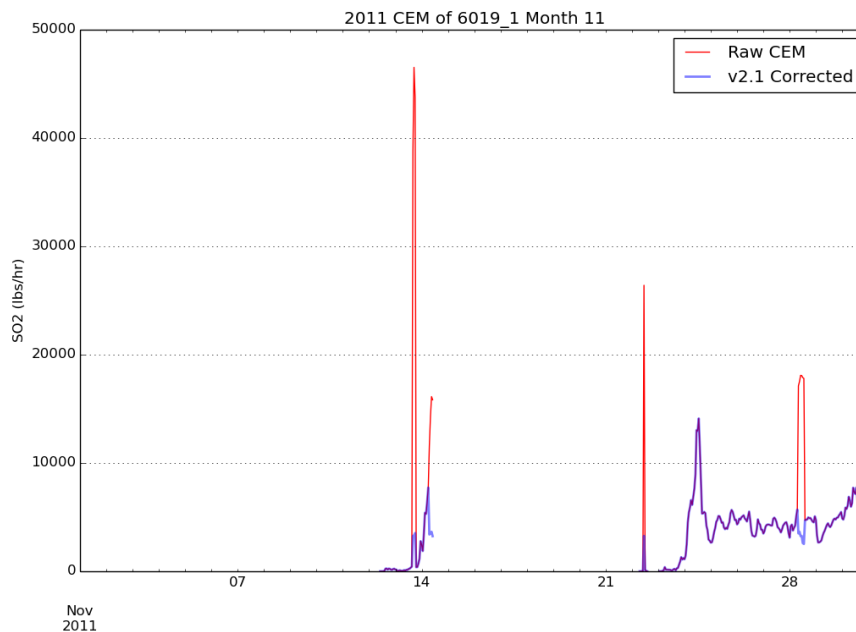


Figure 3-4. Eliminating unmeasured spikes in CEMS data

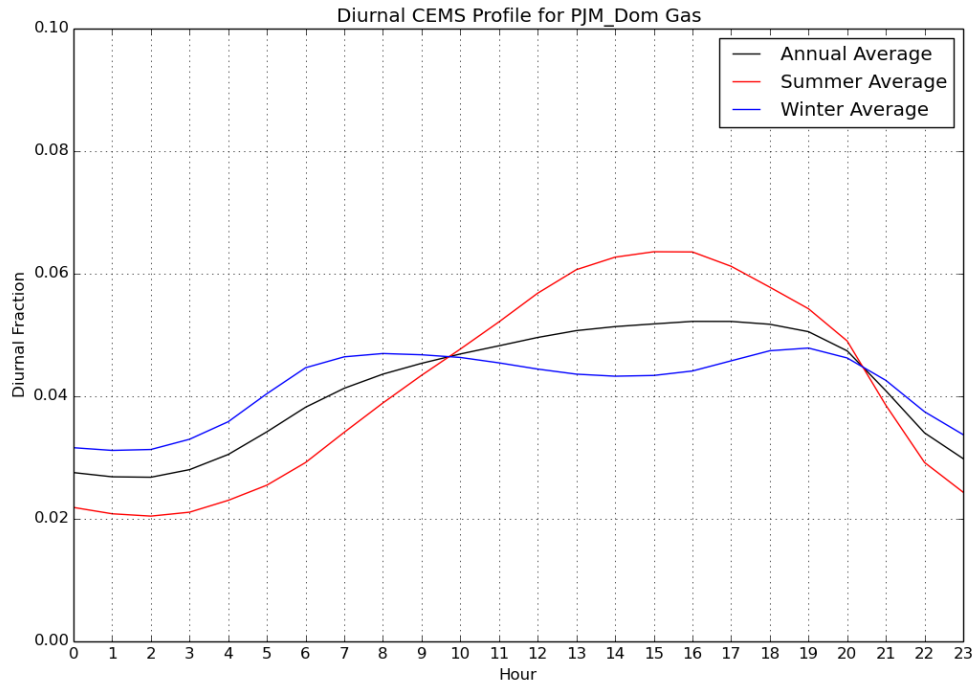


Figure 3-5. Seasonal diurnal profiles for EGU emissions in a Virginia Region

The temporal allocation procedure is differentiated by whether or not the source could be directly matched to a CEMS unit via ORIS facility code and boiler ID. Prior to temporal allocation, as many sources as possible were matched to CEMS data via ORIS facility code and boiler ID. Units were considered matches if the FIPS state/county code matched, the facility name was similar, and the NO_x and SO₂ emissions were similar. EIS stores a base set of previously matched units via alternate facility and unit IDs. Additions to these matches were made for this study. For any units that are matched, the ORIS facility and boiler ID columns of the point FF10 inventory files are filled with the information on the rows for the corresponding NEI unit. Note that for units matched to CEMS data, annual totals of their emissions may be different than the annual values in the inventory because the CEMS data actually replaces the inventory data for the seasons in which the CEMS are operating. If a CEMS-matched unit is determined to be a partial year reporter, as can happen for sources that run CEMS only in the summer, emissions totaling the difference between the annual emissions and the total CEMS emissions are allocated to the non-summer months.

For sources not matched to CEMS units, the allocation of annual emissions to months and then days are done outside of SMOKE and then daily emissions are output to day-specific inventory files. For these units, the allocation of the inventory annual emissions to months is done using average fuel-specific season-to-month factors generated for each of the 64 IPM regions shown in Figure 3-6. These factors are based 2011 CEMS data only. In each region, separate factors were developed for the fuels: coal, natural gas, and “other”, where the types of fuels included in “other” vary by region. Separate profiles were computed for NO_x, SO₂, and heat input. An overall composite profile was also computed and was used when there were no CEMS units with the specified fuel in the region containing the unit. For both CEMS-matched units and units not matched to CEMS, NO_x and SO₂ CEMS data are used to allocate NO_x and

Daily temporal allocation of units matched to CEMS was performed using a procedure similar to the approach to allocate emissions to months in that the CEMS data replaces the inventory data for each pollutant. For units without CEMS data, emissions were allocated from month to day using IPM-region and fuel-specific average month-to-day factors based on the 2011 CEMS data. Separate month-to-day allocation factors were computed for each month of the year using heat input for the fuels coal, natural gas, and “other” in each region. For both CEMS and non-CEMS matched units, NO_x and SO₂ CEMS data are used to allocate NO_x and SO₂ emissions, while CEMS heat input data are used to allocate all other pollutants. An example of month-to-day profiles for gas, coal, and an overall composite for a region in western Texas is shown in Figure 3-7.

[illegible]

45

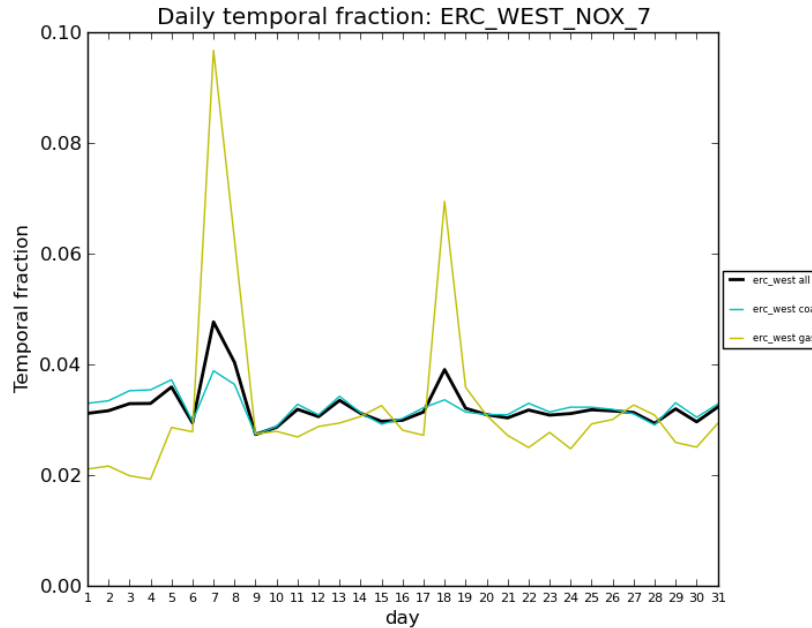


Figure 3-7. Month-to-day profiles for different fuels in a West Texas Region

For the ptfire sector, the inventories are in the daily point fire format ORL PTDAY. The ptfire sector is used in model evaluation cases. The 2007 and earlier platforms had additional regulatory cases that used averaged fires and temporally averaged EGU emissions, but the 2011 platform uses base year-specific (i.e., 2011) data for all cases.

For the nonroad sector, while the NEI only stores the annual totals, the modeling platform uses monthly inventories from output from NMIM. For California, a monthly inventory was created from CARB's annual inventory using EPA-estimated NMIM monthly results to compute monthly ratios by pollutant and SCC7 and these ratios were applied to the CARB inventory to create a monthly inventory.

3.3.5.3 Meteorological-based Temporal Profiles

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

The SMOKE program GenTPRO provides a method for developing meteorology-based temporalization. Currently, the program can utilize three types of temporal algorithms: annual-to-day temporalization for residential wood combustion (RWC), month-to-hour temporalization for agricultural livestock ammonia, and a generic meteorology-based algorithm for other situations. For the 2011 platform, meteorological-based temporalization was used for portions of the rwc sector and for livestock within 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 <http://www.cmascenter.org/smoke/documentation/3.5.1/html/ch05s03s07.html>, respectively.

In the 2011v6.2 platform and in SMOKE 3.6.5, the temporal profile format has been updated. GenTPRO now produces separate files including the monthly temporal profiles (ATPRO_MONTHLY) and day-of-month temporal profiles (ATPRO_DAILY), instead of a single ATPRO_DAILY with day-of-year temporal profiles as it did in SMOKE 3.5. The results are the same either way, so the temporal profiles themselves are effectively the same in 2011v6.2 as they were in 2011v6.0 since the meteorology is the same, but they are formatted differently.

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

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

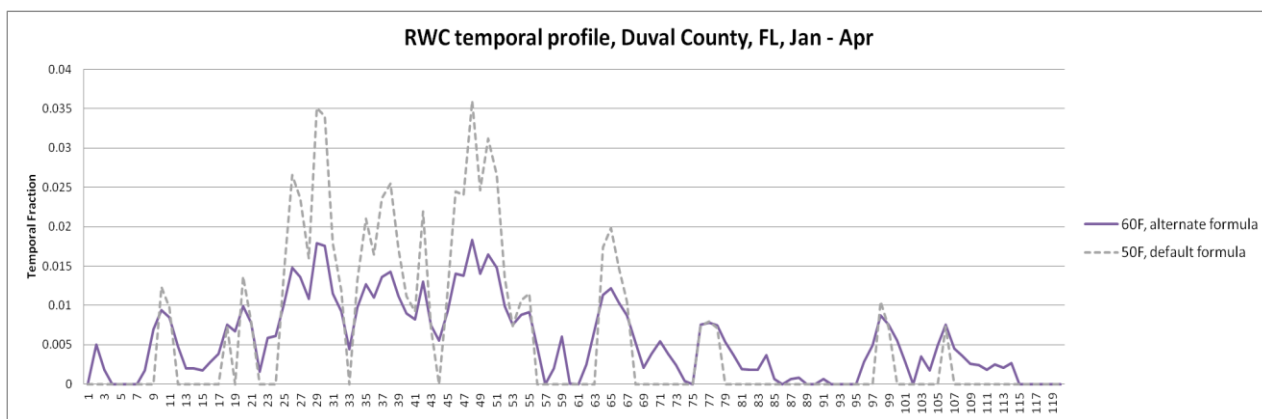


Figure 3-8. Example of RWC temporalization in 2007 using a 50 versus 60 °F threshold

The diurnal profile for used for most RWC sources places more of the RWC emissions in the morning and the evening when people are typically using these sources. This profile is based on a 2004 MANE-VU survey based temporal profiles (see

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

The temporalization for “Outdoor Hydronic Heaters” (i.e., “OHH”, SCC=2104008610) and “Outdoor wood burning device, NEC (fire-pits, chimneas, etc.)” (i.e., “recreational RWC”, SCC=21040087000) were updated because the meteorological-based temporalization used for the rest of the rwc sector did not agree with observations for how these appliances are used. For OHH, the annual-to-month, day-of-week and diurnal profiles were modified based on information in the New York State Energy Research and Development Authority (NYSERDA) “Environmental, Energy Market, and Health Characterization of Wood-Fired Hydronic Heater Technologies, Final Report” (NYSERDA, 2012) as well as a Northeast States for Coordinated Air Use Management (NESCAUM) report “Assessment of Outdoor Wood-fired Boilers” (NESCAUM, 2006). A Minnesota 2008 Residential Fuelwood Assessment Survey of individual household responses (MDNR, 2008) provided additional annual-to-month, day-of-week and diurnal activity information for OHH as well as recreational RWC usage.

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

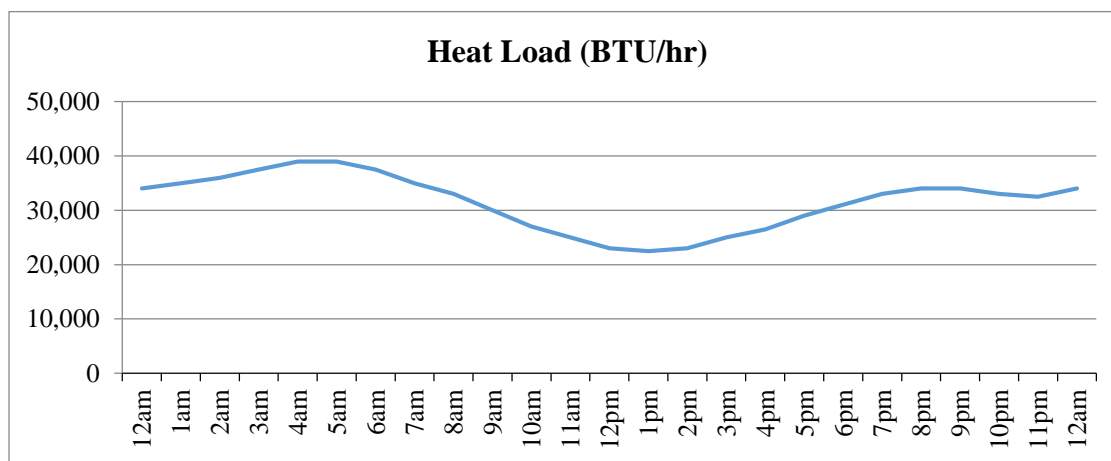


Figure 3-9. Diurnal profile for OHH, based on heat load (BTU/hr)

Monthly Temporal Activity for OHH & Recreational RWC

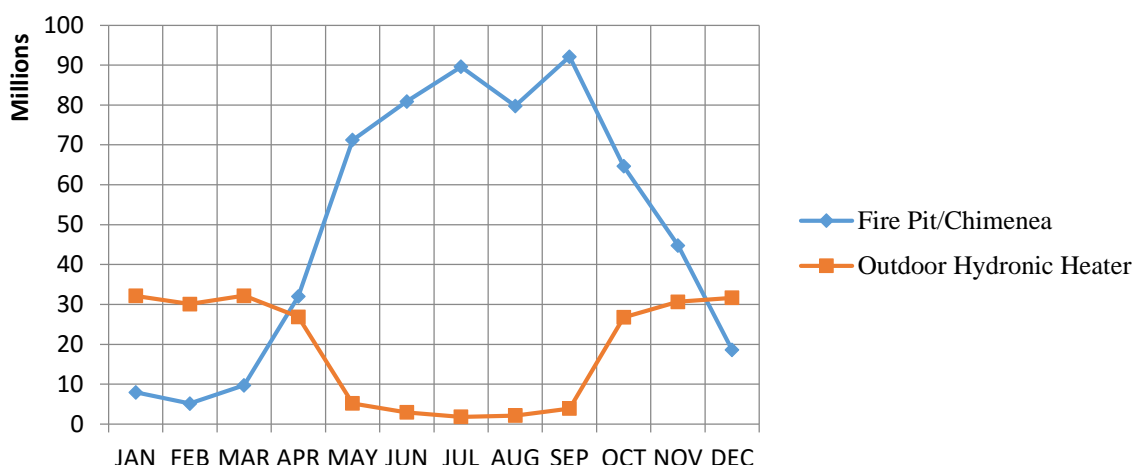


Figure 3-10. Annual-to-month temporal profiles for OHH and recreational RWC

For the agricultural livestock NH_3 algorithm, the GenTPRO algorithm is based on an equation derived by Jesse Bash of EPA ORD based on the Zhu, Henze, et al. (2013) 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:

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

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

where

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

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

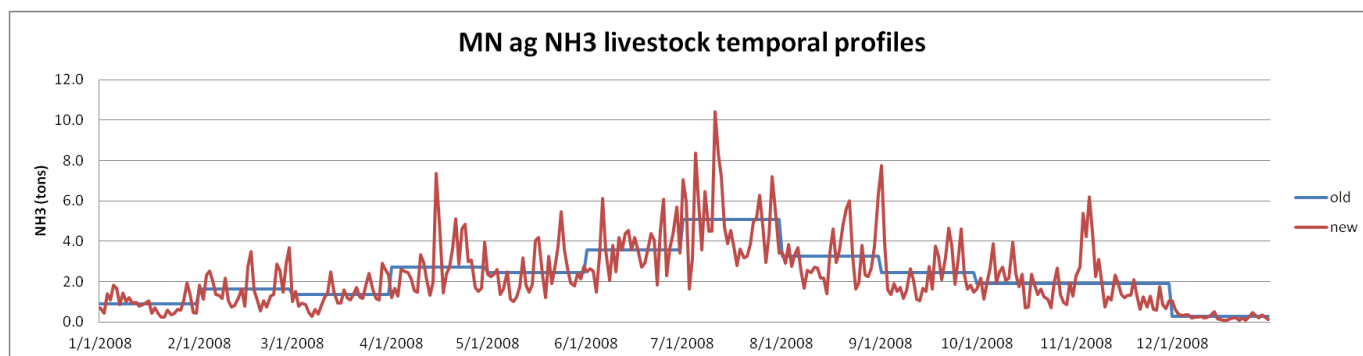


Figure 3-11. Example of new animal NH₃ emissions temporalization approach, summed to daily emissions

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

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

3.3.5.4 Temporal Profiles for Onroad Mobile Sources

For the onroad sector, the temporal distribution of emissions is a combination of more traditional temporal profiles and the influence of meteorology. This section discusses both the meteorological influences and the updates to the diurnal temporal profiles for the 2011v6.2 platform.

Meteorology is not used in the development of the temporal profiles, but rather it impacts the calculation of the hourly emissions through the program Movesmrg. The result is that the emissions vary at the hourly level by grid cell. More specifically, the on-network (RPD) and the off-network parked vehicle (RPV, RPH, and RPP) processes use the gridded meteorology (MCIP) directly. Movesmrg determines the temperature for each hour and grid cell and uses that information to select the appropriate emission factor (EF) for the specified SCC/pollutant/mode combination. In the 2011 platform (and the 2011NEIv2), RPP was updated to use the gridded minimum and maximum temperature for the day. This more spatially resolved temperature range produces more accurate emissions for each grid cell. The combination of these four processes (RPD, RPV, RPH, and RPP) is the total onroad sector emissions. The onroad sector show a strong meteorological influence on their temporal patterns (see the 2011NEIv2 TSD for more details).

Figure 3-12 illustrates the difference between temporalization of the onroad sector used in the 2005 and earlier platforms and the meteorological influence via SMOKE-MOVES. In the plot, the “MOVES” inventory is a monthly inventory that is temporalized by SCC to day-of-week and hour. Similar temporalization is done for the VMT in SMOKE-MOVES, but the meteorologically varying EFs add an additional variation on top of the temporalization. Note, the SMOKE-MOVES run is based on the 2005 platform and previous temporalization of VMT to facilitate the comparison of the results. In the figure, the MOVES emissions have a repeating pattern within the month, while the SMOKE-MOVES shows day-to-day (and hour-to-hour) variability. In addition, the MOVES emissions have an artificial jump between months which is due to the inventory providing new emissions for each month that are then temporalized within the month but not between months. The SMOKE-MOVES emissions have a smoother transition between the months.

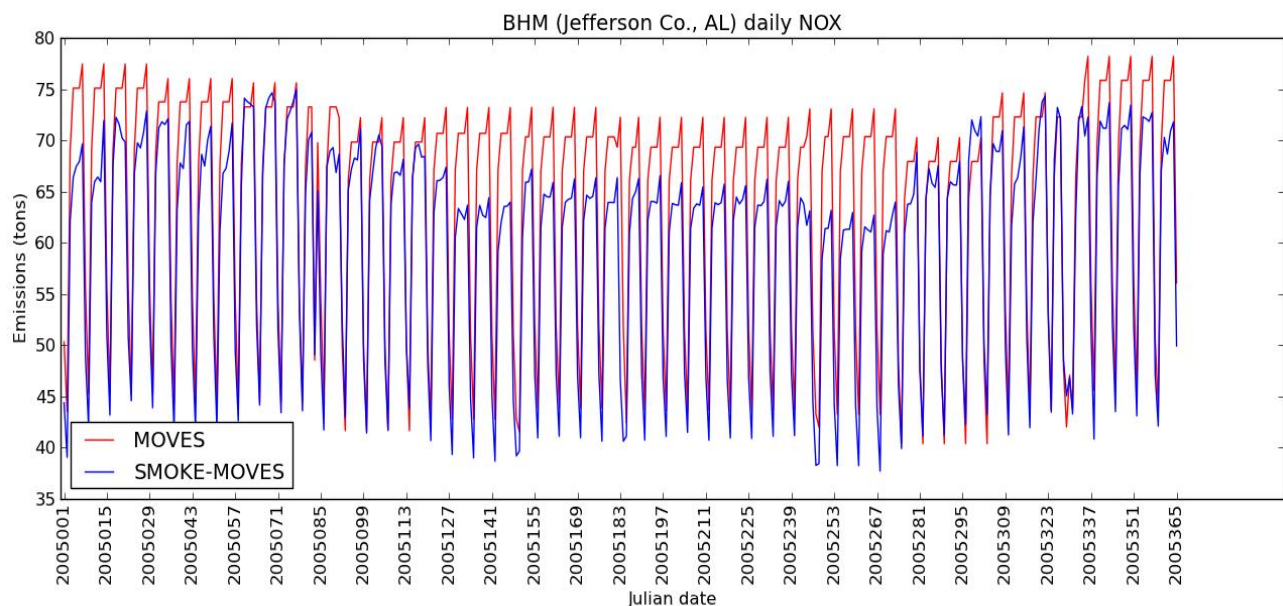


Figure 3-12. Example of SMOKE-MOVES temporal variability of NO_x emissions

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 monthly and was temporalized to days of the week and then to hourly VMT through temporal profiles. The RPD processes require a speed profile (SPDPRO) that consists of vehicle speed by hour for a typical weekday and weekend day. Unlike other sectors, the temporal profiles and SPDPRO will impact not only the distribution of emissions through time but also the total emissions. Because SMOKE-MOVES (for RPD) calculates emissions from VMT, speed and meteorology, if one shifted the VMT or speed to different hours, it would align with different temperatures and hence different EF. In other words, two SMOKE-MOVES runs with identical annual VMT, meteorology, and MOVES EF, will have different total emissions if the temporalization of VMT changes. For RPH, the HOTELING inventory is monthly and was temporalized to days of the week and to 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.

In previous platforms, the diurnal profile for VMT varied by road type but not by vehicle type and these profiles were used throughout the nation. Diurnal profiles that could differentiate by vehicle type as well as by road type and would potentially vary over geography were desired. In the development of the 2011v6.0 platform, the EPA updated these profiles to include information submitted by states in their MOVES county databases (CDBs). The development of the 2011NEIv2 and the 2011v6.2 platform provided an opportunity to update these diurnal profile with new information submitted by states, to supplement the data with additional sources, and to refine the methodology.

States submitted MOVES county databases (CDBs) that included information on the distribution of VMT by hour of day and by day of week (see the 2011NEIv2 TSD for details on the submittal process for onroad). EPA mined the state submitted MOVES CDBs for non-default diurnal profiles. Further QA was done to remove duplicates and profiles that were missing two or more hours. If they were missing a single hour, the missing hour could be calculated by subtracting all other hours' fractions from 1. The list of potential diurnal profiles was then analyzed to see whether the profiles varied by vehicle type, road type, weekday vs. weekend, and by county within a state.

For the MOVES diurnal profiles, EPA only considered the state profiles that varied significantly by both vehicle and road types. Only those profiles that passed these criteria were used in that state or used in developing default temporal profiles. The Vehicle Travel Information System (VTRIS) is a repository for reported traffic count data to the Federal Highway Administration (FHWA). EPA used 2012 VTRIS data to create additional temporal profiles for states that did not submit temporal information in their CDBs or where those profiles did not pass the variance criteria. The VTRIS data were used to create state specific diurnal profiles by HPMS vehicle and road type. EPA created distinct diurnal profiles for weekdays, Saturday and Sunday along with day of the week profiles. Note that the day of the week profiles (ie. Monday vs Tuesday vs etc) are only from the VTRIS data. The MOVES CDBs only have weekday vs weekend profiles so they were not included in calculating a new national default day of the week profile

EPA attempted to maximize the use of state and/or county specific diurnal profiles (either from MOVES or VTRIS). Where there was no MOVES or VTRIS data, then a new default profile would be used (see below for description of new profiles). This analysis was done separately for weekdays and for weekends, therefore some areas had submitted profiles for weekdays but defaults for weekends. The result was a set of profiles that varied geographically depending on the source of the profile and the characteristics of the profiles (see Figure 3-13).

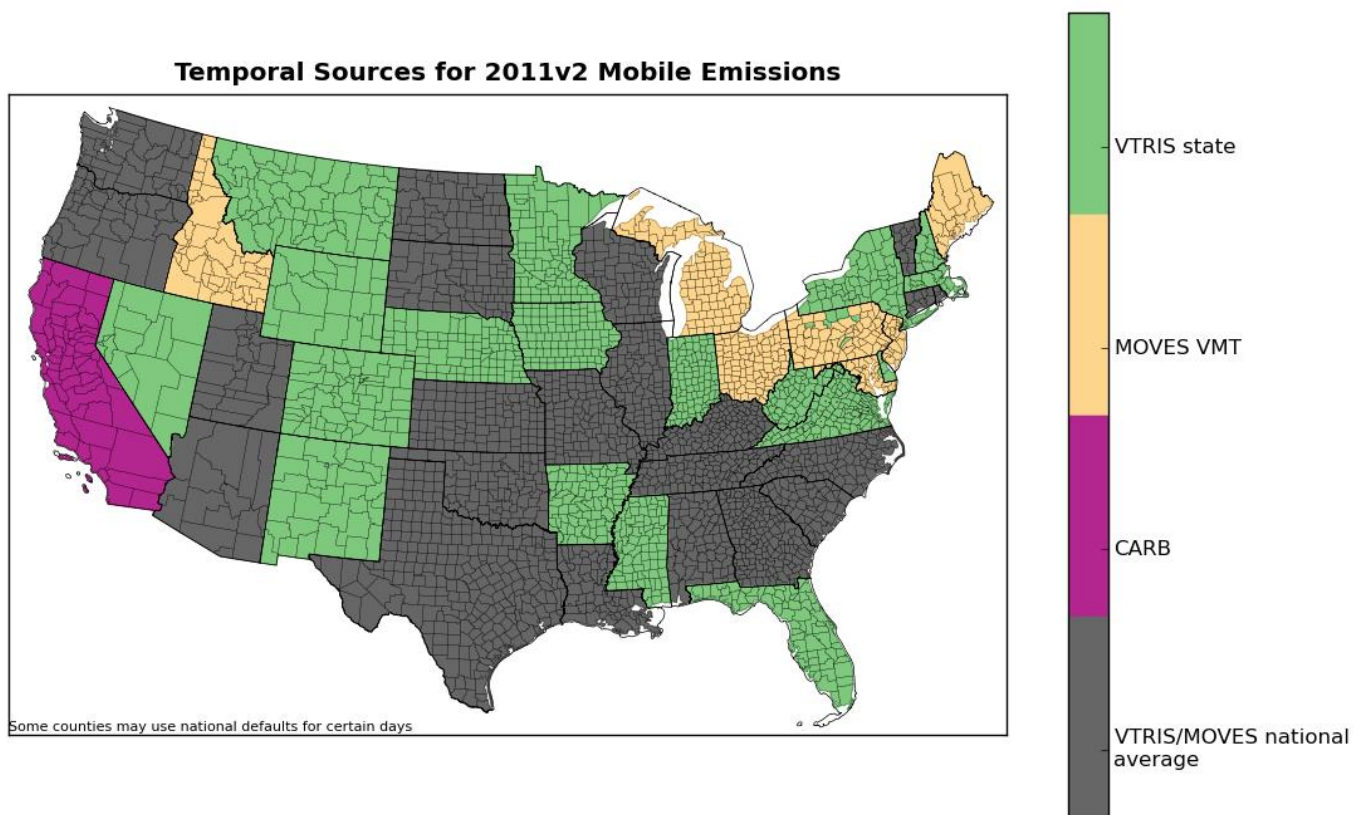


Figure 3-13. Use of submitted versus new national default profiles

A new set of diurnal profiles was developed from the submitted profiles that varied by both vehicle type and road type. For the purposes of constructing the national default diurnal profiles, EPA created individual profiles for each state (averaging over the counties within) to create a single profile by state, vehicle type, road type, and the day (i.e. weekday vs Saturday vs Sunday). The source of the underlying profiles was either MOVES or VTRIS data. The states individual profiles were averaged together to create a new default profile. Figure 3-14 shows two new national default profiles for light duty gas vehicles (LDGV, SCC6 220121) and combination long-haul diesel trucks (HHDDV, SCC6 220262) on restricted urban roadways (interstates and freeways). The blue lines indicate the weekday profile, the green the Saturday profile, and the red the Sunday profile. In comparison, the new default profiles for weekdays places more LDGV VMT (upper plot) in the rush hours while placing HHDDV VMT (lower plot) predominately in the middle of the day with a longer tail into the evening hours and early morning. In addition to creating diurnal profiles,

EPA also developed day of week profiles using the VTRIS data. The creation of the state and national profiles was similar to that of the diurnal profiles (described above). Figure 3-15 shows a set of national default profiles for rural restricted roads (top plot) and urban unrestricted roads (lower plot). Each vehicle type is a different color on the plots.

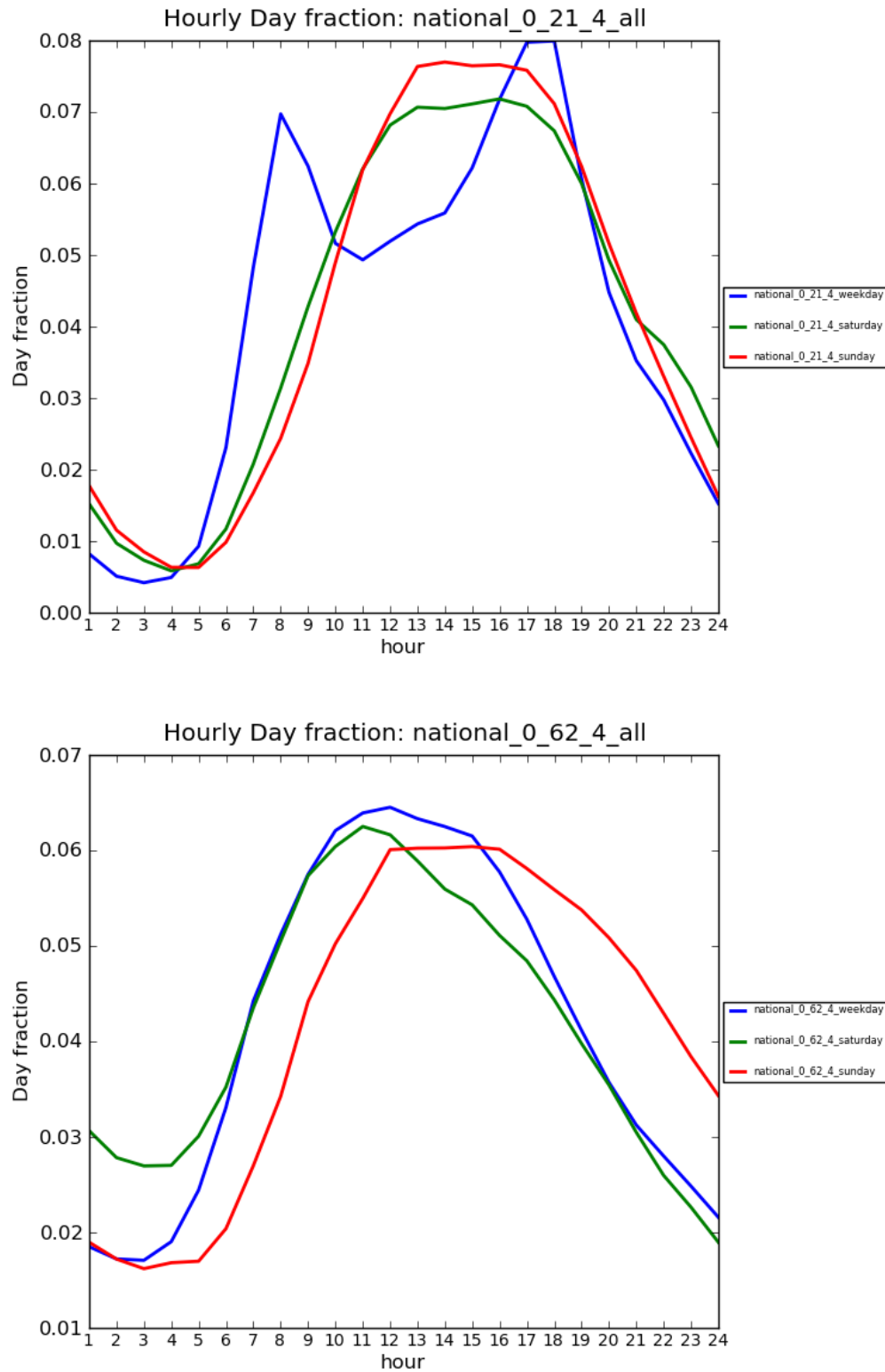


Figure 3-14. Updated national default profiles for LDGV vs. HHDDV, urban restricted

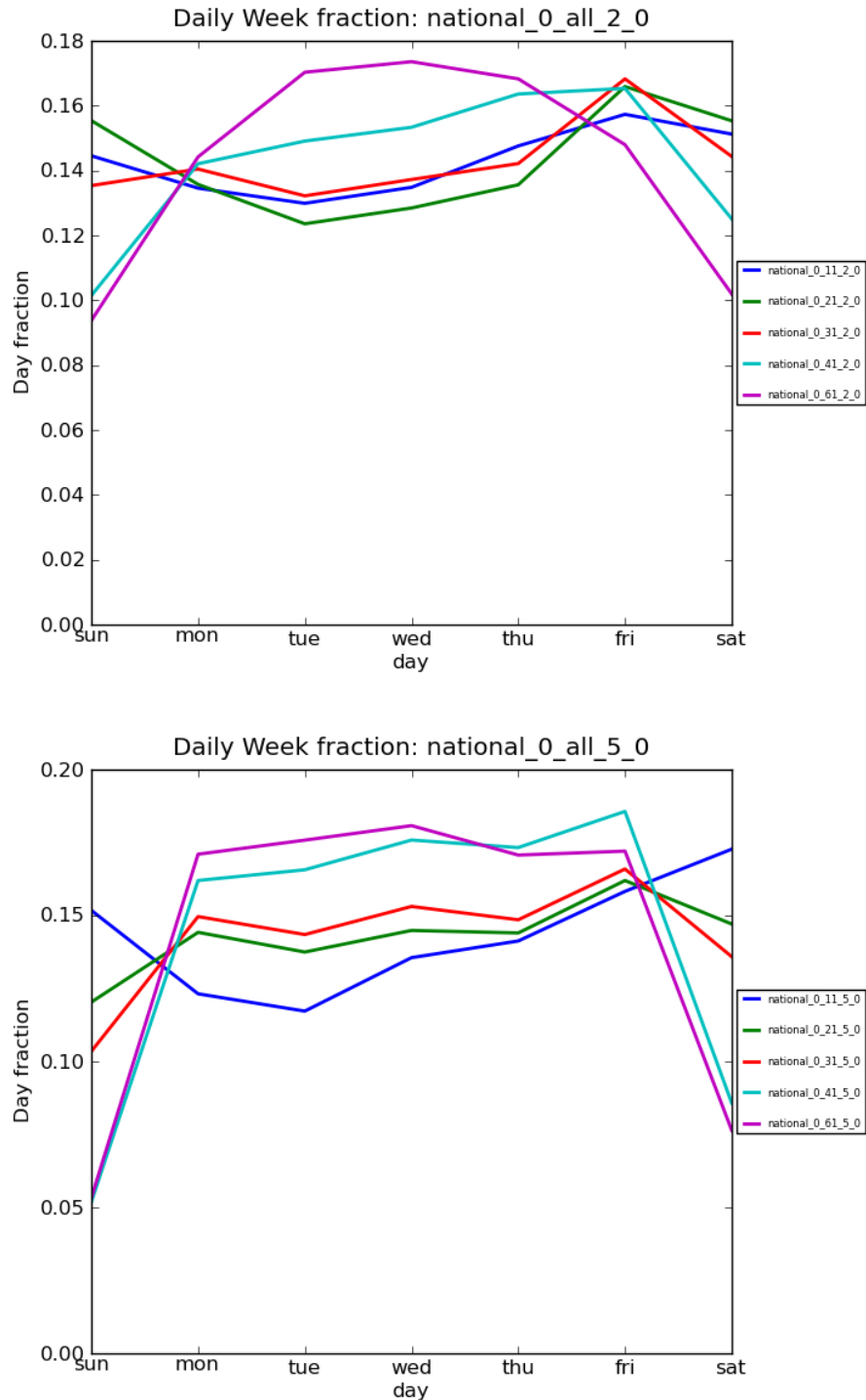


Figure 3-15. Updated national default profiles for day of week (rural restricted top and urban restricted bottom)

In addition to creating diurnal profiles for VMT, EPA developed a national profile for hoteling. EPA averaged all the combination long-haul truck profiles on restricted roads (urban and rural) for weekdays to create a single national restricted profile (blue line in Figure 3-16). This was then inverted to create a

profile for hoteling (green line in Figure 3-16). This single national profile was used for hoteling irrespective of location.

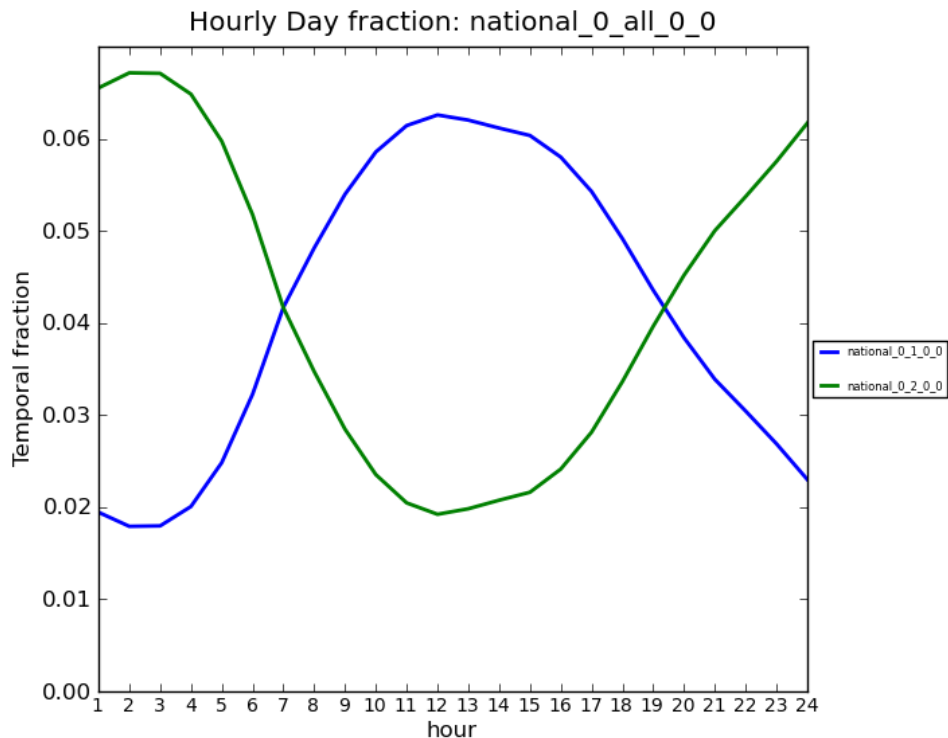


Figure 3-16. Combination long-haul truck restricted and hoteling profile

For California, CARB supplied diurnal profiles that varied by vehicle type, day of the week (Monday, Tuesday-Thursday, Saturday, and Sunday), and air basin. These CARB specific profiles were used in developing EPA estimates for California. Although EPA adjusted the total emissions to match California’s submittal to the 2011NEIv2, 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 the 2011NEIv2 TSD.

3.3.6 Vertical Allocation of Emissions

Table 3-5 specifies the sectors for which plume rise is calculated. If there is no plume rise for a sector, the emissions are placed into layer 1 of the air quality model. Vertical plume rise was performed in-line within CMAQ for all of the SMOKE point-source sectors (i.e., ptipm, ptnonipm, ptfire, othpt, and c3marine). 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. 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 source is slightly different than the one used in SMOKE and this can result in slightly different placement of point sources near grid cell boundaries.

For point sources, the stack parameters are used as inputs to the Briggs algorithm, but point fires do not have stack parameters. However, the ptfire inventory does contain data on the acres burned (acres per day)

and fuel consumption (tons fuel per acre) for each day. CMAQ uses these additional parameters to estimate the plume rise of emissions into layers above the surface model layer. Specifically, these data are used to calculate heat flux, which is then used to estimate plume rise. In addition to the acres burned and fuel consumption, heat content of the fuel is needed to compute heat flux. The heat content was assumed to be 8000 Btu/lb of fuel for all fires because specific data on the fuels were unavailable in the inventory. The plume rise algorithm applied to the fires is a modification of the Briggs algorithm with a stack height of zero.

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

3.3.7 Emissions Modeling Spatial Allocation

The methods used to perform spatial allocation are summarized in this section. For the modeling platform, spatial factors are typically applied by county and SCC. Spatial allocation was performed for a national 12-km domain. To accomplish this, SMOKE used national 12-km spatial surrogates and a SMOKE area-to-point data file. For the U.S., EPA updated surrogates to use circa 2010-2011 data wherever possible. For Mexico, updated spatial surrogates were used as described below. For Canada surrogates provided by Environment Canada were used and are unchanged from the 2007 platform. The U.S., Mexican, and Canadian 12-km surrogates cover the entire CONUS domain 12US1 shown in Figure 3-2.

The details regarding how the 2011v6.2 platform surrogates were created are available from ftp://newftp.epa.gov/Air/emismod/2011/v2platform/spatial_surrogates/ in the files *US_SpatialSurrogate_Workbook_v072115.xlsx* and *US_SpatialSurrogate_Documentation_v070115.pdf*, and *SurrogateTools_Scripts_2014.zip* available. The remainder of this subsection provides further detail on the origin of the data used for the spatial surrogates and the area-to-point data.

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. As described in Section 3.4.2, 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 the 2011v6.2 platform, but they are sometimes used to gapfill other surrogates, or as an input for merging two surrogates to create a new surrogate that is used.

Many surrogates use circa 2010-based data, including 2010 census data at the block group level, 2010 American Community Survey Data for heating fuels, 2010 TIGER/Line data for railroads and roads, the 2006 National Land Cover Database, 2011 gas station and dry cleaner data, and the 2012 National Transportation Atlas Data for rail-lines, ports and navigable waterways. Surrogates for ports (801) and shipping lanes (802) were developed based on the 2011NEIv2 shapefiles: *Ports_032310_wrf* and *ShippingLanes_111309FINAL_wrf*, but also included shipping lane data in the Great Lakes and support vessel activity data in the Gulf of Mexico. The creation of surrogates and shapefiles for the U.S. was generated via the Surrogate Tool. The tool is available from <http://cmascenter.org> and documentation for it is available at <https://www.cmascenter.org/sa-tools/documentation/4.2/html/srgtool/>

Table 3-12. U.S. Surrogates available for the 2011 modeling platform

Code	Surrogate Description	Code	Surrogate Description
N/A	Area-to-point approach (see 3.3.1.2)	507	Heavy Light Construction Industrial Land
100	Population	510	Commercial plus Industrial
110	<i>Housing</i>	515	Commercial plus Institutional Land
120	<i>Urban Population</i>	520	Commercial plus Industrial plus Institutional
130	Rural Population	525	<i>Golf Courses + Institutional + Industrial + Commercial</i>
137	<i>Housing Change</i>	526	Residential Non-Institutional
140	Housing Change and Population	527	Single Family Residential
150	Residential Heating - Natural Gas	530	<i>Residential - High Density</i>
160	<i>Residential Heating – Wood</i>	535	Residential + Commercial + Industrial + Institutional + Government
165	0.5 Residential Heating - Wood plus 0.5 Low Intensity Residential	540	Retail Trade
170	Residential Heating - Distillate Oil	545	Personal Repair
180	Residential Heating – Coal	550	<i>Retail Trade plus Personal Repair</i>
190	Residential Heating - LP Gas	555	Professional/Technical plus General Government
200	Urban Primary Road Miles	560	Hospitals
205	Extended Idle Locations	565	<i>Medical Offices/Clinics</i>
210	Rural Primary Road Miles	570	<i>Heavy and High Tech Industrial</i>
220	<i>Urban Secondary Road Miles</i>	575	Light and High Tech Industrial
221	Urban Unrestricted Roads	580	Food, Drug, Chemical Industrial
230	<i>Rural Secondary Road Miles</i>	585	Metals and Minerals Industrial
231	Rural Unrestricted Roads	590	Heavy Industrial
240	Total Road Miles	595	Light Industrial
250	Urban Primary plus Rural Primary	596	<i>Industrial plus Institutional plus Hospitals</i>
255	<i>0.75 Total Roadway Miles plus 0.25 Population</i>	600	Gas Stations
256	Off-Network Short-Haul Trucks	650	Refineries and Tank Farms
257	Off-Network Long-Haul Trucks	675	Refineries and Tank Farms and Gas Stations
258	Intercity Bus Terminals	680	Oil & Gas Wells, IHS Energy, Inc. and USGS (see updated surrogates in Table 3-19)
259	Transit Bus Terminals	700	Airport Areas
260	Total Railroad Miles	710	<i>Airport Points</i>
261	NTAD Total Railroad Density	720	<i>Military Airports</i>
270	<i>Class 1 Railroad Miles</i>	800	<i>Marine Ports</i>
271	NTAD Class 1, 2, 3 Railroad Density	801	NEI Ports
280	Class 2 and 3 Railroad Miles	802	<i>NEI Shipping Lanes</i>
300	Low Intensity Residential	806	Offshore Shipping NEI NOx
310	Total Agriculture	807	<i>Navigable Waterway Miles</i>
312	Orchards/Vineyards	808	<i>Gulf Tug Zone Area</i>
320	Forest Land	810	<i>Navigable Waterway Activity</i>
330	Strip Mines/Quarries	812	<i>Midwest Shipping Lanes</i>
340	<i>Land</i>	820	Ports NEI NOx
350	Water	850	Golf Courses

Code	Surrogate Description	Code	Surrogate Description
400	Rural Land Area	860	Mines
500	Commercial Land	870	Wastewater Treatment Facilities
505	Industrial Land	880	Drycleaners
506	Education	890	Commercial Timber

For the onroad sector, the on-network (RPD) emissions were spatially allocated to roadways, and the off-network (RPP and RPV) emissions were allocated according to the mapping in Table 3-13. The refueling emissions were spatially allocated to gas station locations (surrogate 600). On-network (i.e., on-roadway) mobile source emissions were assigned to the following surrogates: rural restricted access to rural primary road miles (210), rural unrestricted access to 231, urban restricted access to urban primary road miles (200), and urban unrestricted access to 221. In the 2011v6.2 platform, emissions from the extended (i.e., overnight) idling of trucks were assigned to a new surrogate 205 that is based on locations of overnight truck parking spaces.

Table 3-13. Off-Network Mobile Source Surrogates

Source type	Source Type name	Surrogate ID
11	Motorcycle	535
21	Passenger Car	535
31	Passenger Truck	535
32	Light Commercial Truck	510
41	Intercity Bus	258
42	Transit Bus	259
43	School Bus	506
51	Refuse Truck	507
52	Single Unit Short-haul Truck	256
53	Single Unit Long-haul Truck	257
54	Motor Home	526
61	Combination Short-haul Truck	256
62	Combination Long-haul Truck	257

For the oil and gas sources in the np_oilgas sector, the spatial surrogates were updated to those shown in Table 3-14 using 2011 data consistent with what was used to develop the 2011NEI nonpoint oil and gas emissions. Note that the “Oil & Gas Wells, IHS Energy, Inc. and USGS” (680) is older and based on circa-2005 data. These surrogates were based on the same GIS data of well locations and related attributes as was used to develop the 2011NEIv2 data for the oil and gas sector. The data sources include Drilling Info (DI) Desktop’s HPDI database (Drilling Info, 2012) aggregated to grid cell levels, along with data from Oil and Gas Commission (OGC) websites. Well completion data from HPDI was supplemented by implementing the methodology for counting oil and gas well completions developed for the U.S. National Greenhouse Gas Inventory. Under that methodology, both completion date and date of first production from HPDI were used to identify wells completed during 2011. In total, over 1.08 million unique well locations were compiled from the various data sources. The well locations cover 33 states and 1,193 counties (ERG, 2014b). Although basically the same surrogates were used, some minor updates to the oil and gas surrogates were made in the 2011v6.2 platform to correct some mis-located emissions.

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

Surrogate Code	Surrogate Description
681	Spud count - Oil Wells
682	Spud count - Horizontally-drilled wells
683	Produced Water at all wells
684	Completions at Gas and CBM Wells
685	Completions at Oil Wells
686	Completions at all wells
687	Feet drilled at all wells
688	Spud count - Gas and CBM Wells
689	Gas production at all wells
692	Spud count - All Wells
693	Well count - all wells
694	Oil production at oil wells
695	Well count - oil wells
697	Oil production at Gas and CBM Wells
698	Well counts - Gas and CBM Wells

Some spatial surrogate cross reference updates were made between the 2011v6.1 platform and the 2011v6.2 platform aside from the reworking of the onroad mobile source surrogates described above. These updates included the following:

- Nonroad SCCs using spatial surrogate 525 (50% commercial + industrial + institutional, 50% golf courses) were changed to 520 (100% commercial + industrial + institutional). The golf course surrogate 850, upon which 525 is partially based, is incomplete and subject to hot spots;
- Some nonroad SCCs for commercial equipment in New York County had assignments updated to surrogate 340;
- Commercial lawn and garden equipment was updated to use surrogate 520; and
- Some county-specific assignments for RWC were updated to use surrogate 300.

Not all of the available surrogates are used to spatially allocate sources in the modeling platform; that is, some surrogates shown in Table 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.

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

There are numerous airport-related emission sources in the NEI, such as aircraft, airport ground support equipment, and jet refueling. The modeling platform includes the aircraft and airport ground support

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

3.3.7.3 Surrogates for Canada and Mexico Emission Inventories

The surrogates for Canada to spatially allocate the 2010 Canadian emissions have been updated in the 2011v6.2 platform. The spatial surrogate data came from Environment Canada, along with cross references. The surrogates they provided were outputs from the Surrogate Tool (previously referenced). The Canadian surrogates used for this platform are listed in Table 3-15. The leading “9” was added to the surrogate codes to avoid duplicate surrogate numbers with U.S. surrogates. Surrogates for Mexico are circa 1999 and 2000 and were based on data obtained from the Sistema Municipal de Bases de Datos (SIMBAD) de INEGI and the Bases de datos del Censo Economico 1999. Most of the CAPs allocated to the Mexico and Canada surrogates are shown in Table 3-16. The entries in this table are for the other sector except for the “MEX Total Road Miles” and the “CAN traffic” rows, which are for the other sector.

Table 3-15. Canadian Spatial Surrogates

Code	Canadian Surrogate Description	Code	Description
9100	Population	92424	BARLEY
9101	total dwelling	92425	BUCWHT
9103	rural dwelling	92426	CANARY
9106	ALL_INDUST	92427	CANOLA
9111	Farms	92428	CHICPEA
9113	Forestry and logging	92429	CORNGR
9211	Oil and Gas Extraction	92425	BUCWHT
9212	Mining except oil and gas	92430	CORNSI
9221	Total Mining	92431	DFPEAS
9222	Utilities	92432	FLAXSD
9233	Total Land Development	92433	FORAGE
9308	Food manufacturing	92434	LENTIL
9321	Wood product manufacturing	92435	MUSTSD
9323	Printing and related support activities	92436	MXDGRN
9324	Petroleum and coal products manufacturing	92437	OATS
9327	Non-metallic mineral product manufacturing	92438	ODFBNS
9331	Primary Metal Manufacturing	92439	OTTAME
9412	Petroleum product wholesaler-distributors	92440	POTATS
9416	Building material and supplies wholesaler-distributors	92441	RYEFAL
9447	Gasoline stations	92442	RYESPG
9448	clothing and clothing accessories stores	92443	SOYBNS
9481	Air transportation	92444	SUGARB
9482	Rail transportation	92445	SUNFLS

Code	Canadian Surrogate Description	Code	Description
9562	Waste management and remediation services	92446	TOBACO
9921	Commercial Fuel Combustion	92447	TRITCL
9924	Primary Industry	92448	WHITBN
9925	Manufacturing and Assembly	92449	WHTDUR
9932	CANRAIL	92450	WHTSPG
9941	PAVED ROADS	92451	WHTWIN
9942	UNPAVED ROADS	92452	BEANS
9945	Commercial Marine Vessels	92453	CARROT
9946	Construction and mining	92454	GRPEAS
9948	Forest	92455	OTHVEG
9950	Combination of Forest and Dwelling	92456	SWCORN
9955	UNPAVED_ROADS_AND_TRAILS	92457	TOMATO
9960	TOTBEEF	92430	CORNSI
9970	TOTPOUL	92431	DFPEAS
9980	TOTSWIN	92432	FLAXSD
9990	TOTFERT	92433	FORAGE
9996	urban_area	92434	LENTIL
9997	CHBOISQC	92435	MUSTSD
91201	traffic_bcw	92436	MXDGRN
92401	BULLS	92437	OATS
92402	BFCOWS	92438	ODFBNS
92403	BFHEIF	92439	OTTAME
92404	CALFU1	92440	POTATS
92405	FDHEIF	92441	RYEFAL
92406	STEERS	92442	RYESPG
92407	MLKCOW	92443	SOYBNS
92408	MLKHEIF	92444	SUGARB
92409	MBULLS	92445	SUNFLS
92410	MCALFU1	92446	TOBACO
92412	BROILER	92447	TRITCL
92413	LAYHEN	92448	WHITBN
92414	TURKEY	92449	WHTDUR
92416	BOARS	92450	WHTSPG
92417	GRWPIG	92451	WHTWIN
92418	NURPIG	92452	BEANS
92419	SOWS	92453	CARROT
92421	IMPAST	92454	GRPEAS
92422	UNIMPAST	92455	OTHVEG
92423	ALFALFA	92456	SWCORN
		92457	TOMATO

Table 3-16. CAPs Allocated to Mexican and Canadian Spatial Surrogates

Code	Mexican or Canadian Surrogate Description	NH₃	NO_x	PM_{2.5}	SO₂	VOC
22	MEX Total Road Miles	15,965	370,867	34,396	13,713	375,276
10	MEX Population	0	0	0	0	431,231
12	MEX Housing	0	161,013	17,483	2,123	452,685
14	MEX Residential Heating - Wood	0	20,093	211,525	2,859	380,572
16	MEX Residential Heating - Distillate Oil	0	38	0	11	2
20	MEX Residential Heating - LP Gas	0	25,303	787	63	614
22	MEX Total Road Miles	0	0	0	0	3,513
24	MEX Total Railroads Miles	0	74,969	1,669	663	2,824
26	MEX Total Agriculture	679,212	164,144	72,372	2,127	43,958
28	MEX Forest Land	0	16,224	67,683	660	79,018
32	MEX Commercial Land	0	125,211	7,726	0	286,982
34	MEX Industrial Land	0	45,831	5,684	59,201	133,440
36	MEX Commercial plus Industrial Land	0	0	0	0	332,495
38	MEX Commercial plus Institutional Land	0	6,400	216	84	28,293
40	Residential (RES1-4)+Commercial+Industrial+ Institutional+ Government	0	8	20	0	241,710
42	MEX Personal Repair (COM3)	0	0	0	0	33,616
44	MEX Airports Area	0	14,639	0	1,149	6,857
46	MEX Marine Ports	0	124,951	2,991	1,482	1,099
48	Brick Kilns - Mexico	0	776	6,691	0	10,244
50	Mobile sources - Border Crossing - Mexico	0	454	0	0	2,668
9100	CAN Population	603	0	276	0	304
9101	CAN total dwelling	643	46,256	12,783	14,698	32,944
9106	CAN ALL_INDUST	133	21,526	381	3,921	2
9113	CAN Forestry and logging	1,582	8,561	28,622	1,809	36,114
9115	CAN Agriculture and forestry activities	160	239,553	25,318	9,092	26,526
9116	CAN Total Resources	0	17	0	0	5
9212	CAN Mining except oil and gas	0	0	5,391	0	0
9221	CAN Total Mining	42	2,292	45,374	728	26
9222	CAN Utilities	189	14,882	369	1,124	255
9233	CAN Total Land Development	17	20,789	1,928	981	2,551
9308	CAN Food manufacturing	0	0	0	0	4,535
9323	CAN Printing and related support activities	0	0	0	0	25,203
9324	CAN Petroleum and coal products manufacturing	0	0	2,402	0	0
9327	CAN Non-metallic mineral product manufacturing	0	238	7,708	2,941	1,218
9331	CAN Primary Metal Manufacturing	0	98	5,062	12	6
9412	CAN Petroleum product wholesaler-distributors	0	0	0	0	70,125

Code	Mexican or Canadian Surrogate Description	NH₃	NO_x	PM_{2.5}	SO₂	VOC
9416	CAN Building material and supplies wholesaler-distributors	2	0	1,461	3,259	560
9448	CAN clothing and clothing accessories stores	0	0	0	0	328
9562	CAN Waste management and remediation services	165	893	1,596	1,998	16,551
9921	CAN Commercial Fuel Combustion	494	33,816	2,750	35,471	850
9924	CAN Primary Industry	0	0	0	0	219,282
9925	CAN Manufacturing and Assembly	0	0	0	0	139,227
9931	CAN OTHERJET	9	14,388	548	1,139	7,629
9932	CAN CANRAIL	109	122,694	4,093	5,737	3,304
9942	CAN UNPAVED ROADS	40	3,462	3,499	48	152,674
9945	CAN Commercial Marine Vessels	28	45,454	6,404	14,325	61,139
9946	CAN Construction and mining	247	156,770	10,070	5,667	17,180
9947	CAN Agriculture Construction and mining	19	37,452	536	26	32,683
9950	CAN Intersection of Forest and Housing	1,053	11,700	120,045	1,671	173,130
9960	CAN TOTBEEF	176,156	0	7,420	0	317,394
9970	CAN TOTPOUL	74,204	0	2	0	264
9980	CAN TOTSWIN	122,094	0	996	0	3,186
9990	CAN TOTFERT	178,791	0	9,279	0	0
9991	CAN traffic	22,294	550,896	10,888	5,548	285,104
9994	CAN ALLROADS	0	0	55,468	0	0
9995	CAN 30UNPAVED_70trail	0	0	106,707	0	0
9996	CAN urban_area	0	0	284	0	0

3.4 Emissions References

- Adelman, Z. 2012. Memorandum: Fugitive Dust Modeling for the 2008 Emissions Modeling Platform. UNC Institute for the Environment, Chapel Hill, NC. September, 28, 2012
- Adelman, Z., M. Omary, Q. He, J. Zhao and D. Yang, J. Boylan, 2012. "A Detailed Approach for Improving Continuous Emissions Monitoring Data for Regulatory Air Quality Modeling." Presented at the 2012 International Emission Inventory Conference, Tampa, Florida. Available from <https://www3.epa.gov/ttn/chief/conference/ei20/index.html#ses-5>.
- Anderson, G.K.; Sandberg, D.V; Norheim, R.A., 2004. Fire Emission Production Simulator (FEPS) User's Guide. Available at http://www.fs.fed.us/pnw/fera/feps/FEPS_users_guide.pdf
- ARB, 2000. "Risk Reduction Plan to Reduce Particulate Matter Emissions from Diesel-Fueled Engines and Vehicles". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. October, 2000. Available at: <http://www.arb.ca.gov/diesel/documents/rrpFinal.pdf>.
- ARB, 2007. "Proposed Regulation for In-Use Off-Road Diesel Vehicles". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. April, 2007. Available at: <http://www.arb.ca.gov/regact/2007/ordiesl07/isor.pdf>
- ARB, 2010a. "Proposed Amendments to the Regulation for In-Use Off-Road Diesel-Fueled Fleets and the Off-Road Large Spark-Ignition Fleet Requirements". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. October, 2010. Available at: <http://www.arb.ca.gov/regact/2010/offroadlsi10/offroadisor.pdf>.
- ARB, 2010b. "Estimate of Premature Deaths Associated with Fine Particle Pollution (PM_{2.5}) in California Using a U.S. Environmental Protection Agency Methodology". California Environmental Protection Agency Air Resources Board, Mobile Source Control Division, Sacramento, CA. August, 2010. Available at: http://www.arb.ca.gov/research/health/pm-mort/pm-report_2010.pdf.
- Adelman, Z. 2012. *Memorandum: Fugitive Dust Modeling for the 2008 Emissions Modeling Platform*. UNC Institute for the Environment, Chapel Hill, NC. September, 28, 2012.
- Bash, J.O., Baker, K.R., Beaver, M.R., Park, J.-H., Goldstein, A.H., 2015. Evaluation of improved land use and canopy representation in BEIS with biogenic VOC measurements in California (in preparation)
- Bullock Jr., R, and K. A. Brehme (2002) "Atmospheric mercury simulation using the CMAQ model: formulation description and analysis of wet deposition results." *Atmospheric Environment* 36, pp 2135–2146.
- Environ Corp. 2008. Emission Profiles for EPA SPECIATE Database, Part 2: EPAct Fuels (Evaporative Emissions). Prepared for U. S. EPA, Office of Transportation and Air Quality, September 30, 2008.
- EPA, 2005. *EPA's National Inventory Model (NMIM), A Consolidated Emissions Modeling System for MOBILE6 and NONROAD*, U.S. Environmental Protection Agency, Office of Transportation and Air Quality, Assessment and Standards Division. Ann Arbor, MI 48105, EPA420-R-05-024, December 2005. Available at <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockkey=P10023FZ.pdf>.

- EPA 2006a. SPECIATE 4.0, Speciation Database Development Document, Final Report, U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Research Triangle Park, NC 27711, EPA600-R-06-161, February 2006. Available at <https://www.epa.gov/air-emissions-modeling/speciate-version-45-through-40>.
- EPA, 2012a. 2008 National Emissions Inventory, version 2 Technical Support Document. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at : <https://www.epa.gov/air-emissions-inventories/2008-national-emissions-inventory-nei-documentation-draft>.
- EPA, 2015a. 2011 Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 6.2, 2011 Emissions Modeling Platform. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at <https://www.epa.gov/air-emissions-modeling/2011-version-62-platform>.
- EPA, 2015b. 2011 National Emissions Inventory, version 2 Technical Support Document. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at <https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-technical-support-document>.
- ERG, 2014a. Develop Mexico Future Year Emissions Final Report. Available at ftp://newftp.epa.gov/air/emismod/2011/v2platform/2011emissions/Mexico_Emissions_WA%204-09_final_report_121814.pdf
- ERG, 2014b. “Technical Memorandum: Modeling Allocation Factors for the 2011 NEI”.
- Frost & Sullivan, 2010. “Project: Market Research and Report on North American Residential Wood Heaters, Fireplaces, and Hearth Heating Products Market (P.O. # PO1-IMP403-F&S). Final Report April 26, 2010”. Prepared by Frost & Sullivan, Mountain View, CA 94041.
- Joint Fire Science Program, 2009. Consume 3.0--a software tool for computing fuel consumption. Fire Science Brief. 66, June 2009. Consume 3.0 is available at: <http://www.fs.fed.us/pnw/fera/research/smoke/consume/index.shtml>
- McCarty, J.L., Korontzi, S., Jutice, C.O., and T. Loboda. 2009. The spatial and temporal distribution of crop residue burning in the contiguous United States. *Science of the Total Environment*, 407 (21): 5701-5712.
- McKenzie, D.; Raymond, C.L.; Kellogg, L.-K.B.; Norheim, R.A; Andreu, A.G.; Bayard, A.C.; Kopper, K.E.; Elman. E. 2007. Mapping fuels at multiple scales: landscape application of the Fuel Characteristic Classification System. *Canadian Journal of Forest Research*. 37:2421-2437.
- NYSERDA, 2012; “Environmental, Energy Market, and Health Characterization of Wood-Fired Hydronic Heater Technologies, Final Report”. New York State Energy Research and Development Authority (NYSERDA). Available from: <http://www.nyserdera.ny.gov/Publications/Case-Studies/-/media/Files/Publications/Research/Environmental/Wood-Fired-Hydronic-Heater-Tech.ashx>.
- Ottmar, R.D.; Sandberg, D.V.; Riccardi, C.L.; Prichard, S.J. 2007. An Overview of the Fuel Characteristic Classification System – Quantifying, Classifying, and Creating Fuelbeds for Resource Planning. *Canadian Journal of Forest Research*. 37(12): 2383-2393. FCCS is available at: <http://www.fs.fed.us/pnw/fera/fccs/index.shtml>

- 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. Bhawe, D. Tong, D. Mobley, T. Pace, and T. Pierce . (2010) “Assessing the Anthropogenic Fugitive Dust Emission Inventory and Temporal Allocation Using an Updated Speciation of Particulate Matter.” International Emission Inventory Conference, San Antonio, TX. Available at <http://www.epa.gov/ttn/chief/conference/ei19/session9/pouliot.pdf>
- Raffuse, S., D. Sullivan, L. Chinkin, S. Larkin, R. Solomon, A. Soja, 2007. Integration of Satellite-Detected and Incident Command Reported Wildfire Information into BlueSky, June 27, 2007. Available at: <https://www.airfire.org/smartfire>.
- 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://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf
- Sullivan D.C., Raffuse S.M., Pryden D.A., Craig K.J., Reid S.B., Wheeler N.J.M., Chinkin L.R., Larkin N.K., Solomon R., and Strand T. (2008) Development and applications of systems for modeling emissions and smoke from fires: the BlueSky smoke modeling framework and SMARTFIRE: 17th International Emissions Inventory Conference, Portland, OR, June 2-5.
- Wang, Y., P. Hopke, O. V. Rattigan, X. Xia, D. C. Chalupa, M. J. Utell. (2011) “Characterization of Residential Wood Combustion Particles Using the Two-Wavelength Aethalometer”, Environ. Sci. Technol., 45 (17), pp 7387–7393
- Yarwood, G., S. Rao, M. Yocke, and G. Whitten, 2005: Updates to the Carbon Bond Chemical Mechanism: CB05. Final Report to the US EPA, RT-0400675. Available at http://www.camx.com/publ/pdfs/CB05_Final_Report_120805.pdf.

4.0 CMAQ Air Quality Model Estimates

4.1 Introduction to the CMAQ Modeling Platform

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

Historically, air quality models have addressed individual pollutant issues separately. However, many of the same precursor chemicals are involved in both ozone and aerosol (particulate matter) chemistry; therefore, the chemical transformation pathways are dependent. Thus, modeled abatement strategies of pollutant precursors, such as volatile organic compounds (VOC) and NO_x 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 system¹¹. 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 grid spatial resolutions in past annual CMAQ runs have been 36 km x 36 km per grid for the "parent" domain, and nested within that domain are 12 km x 12 km grid resolution domains. The parent domain typically covered the continental United States, and the

¹¹ 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.

nested 12 km x 12 km domain covered the Eastern or Western United States. The CMAQ simulation performed for this 2012 assessment used a single domain that covers the entire continental U.S. (CONUS) and large portions of Canada and Mexico using 12 km by 12 km horizontal grid spacing. Currently, 12 km x 12 km resolution is sufficient as the highest resolution for most regional-scale air quality model applications and assessments.¹² With the temporal flexibility of the model, simulations can be performed to evaluate longer term (annual to multi-year) pollutant climatologies as well as short-term (weeks to months) transport from localized sources. By making CMAQ a modeling system that addresses multiple pollutants and different temporal and spatial scales, CMAQ has a “one atmosphere” perspective that combines the efforts of the scientific community. Improvements will be made to the CMAQ modeling system as the scientific community further develops the state-of-the-science.

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

4.1.1 Advantages and Limitations of the CMAQ Air Quality Model

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

An uncertainty of using the CMAQ model includes structural uncertainties, representation of physical and chemical processes in the model. These consist of: choice of chemical mechanism used to characterize reactions in the atmosphere, choice of land surface model and choice of planetary boundary layer. Another uncertainty in the CMAQ model is based on parametric uncertainties, which includes uncertainties in the model inputs: hourly meteorological fields, hourly 3-D gridded emissions, initial conditions, and boundary conditions. Uncertainties due to initial conditions are minimized by using a 10 day ramp-up period from which model results are not used in the aggregation and analysis of model outputs. Evaluations of models against observed pollutant concentrations build confidence that the model performs with reasonable accuracy despite the uncertainties listed above. A detailed model evaluation for ozone and PM_{2.5} species provided in Section 4.3 shows generally acceptable model performance which is equivalent or better than typical state-of-the-science regional modeling simulations as summarized in Simon et al., 2012¹³.

4.2 CMAQ Model Version, Inputs and Configuration

This section describes the air quality modeling platform used for the 2012 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

¹² U.S. EPA (2014), Draft Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, PM_{2.5}, and Regional Haze, pp 214. http://www.epa.gov/ttn/scram/guidance/guide/Draft_O3-PM-RH_Modeling_Guidance-2014.pdf.

¹³ 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.

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 model as part of the 2012 Platform to provide a national scale air quality modeling analysis. The CMAQ model simulates the multiple physical and chemical processes involved in the formation, transport, and destruction of ozone and fine particulate matter (PM_{2.5}).

This section provides a description of each of the main components of the 2012 CMAQ simulation along with the results of a model performance evaluation in which the 2012 model predictions are compared to corresponding measured concentrations.

4.2.1 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 pollutants in the atmosphere. This 2012 analysis employed CMAQ version 5.0.2¹⁴ which reflects updates to version 5.0.1 which include several changes to the science algorithms to improve the underlying science. The CMAQ model version 5.0 was most recently peer-reviewed in June of 2011 for the U.S. EPA.¹⁵ The model enhancements in version 5.0.2 include:

1. SOA yield update
2. Gas-phase chemistry
3. Sulfate inhibition effect in aqueous chemistry
4. CSQY_DATA files
5. WRF land use options
6. Ammonia bidirectional exchange and dry deposition change
7. M3DRY backward compatibility with MCIP for wind staggering
8. Vertical advection time step
9. Aerosol updates
10. ACONC bug fix

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 model extends vertically from the surface to 50 millibars (approximately 17,600 meters) using a sigma-pressure coordinate system. Air quality conditions at the outer boundary of the 12 km domain were taken from a global model. Table 4-1

¹⁴ CMAQ version 5.0.2 model code is available from the Community Modeling and Analysis System (CMAS) at: <http://www.cmascenter.org>.

¹⁵ Brown, N.J., Allen, D.T., Amar, P., Kallos, G., McNider, R., Russell, A.G., Stockwell, W.R. (September 2011). Final Report: Fourth Peer Review of the CMAQ Model, https://www.epa.gov/sites/production/files/2016-11/documents/cmaq_fifth_review_final_report_2015.pdf. CMAQ version 5.0 was released on February, 2012. It is available from the Community Modeling and Analysis System (CMAS) as well as previous peer-review reports at: <http://www.cmascenter.org>.

provides some basic geographic information regarding the 12 km CMAQ domain.

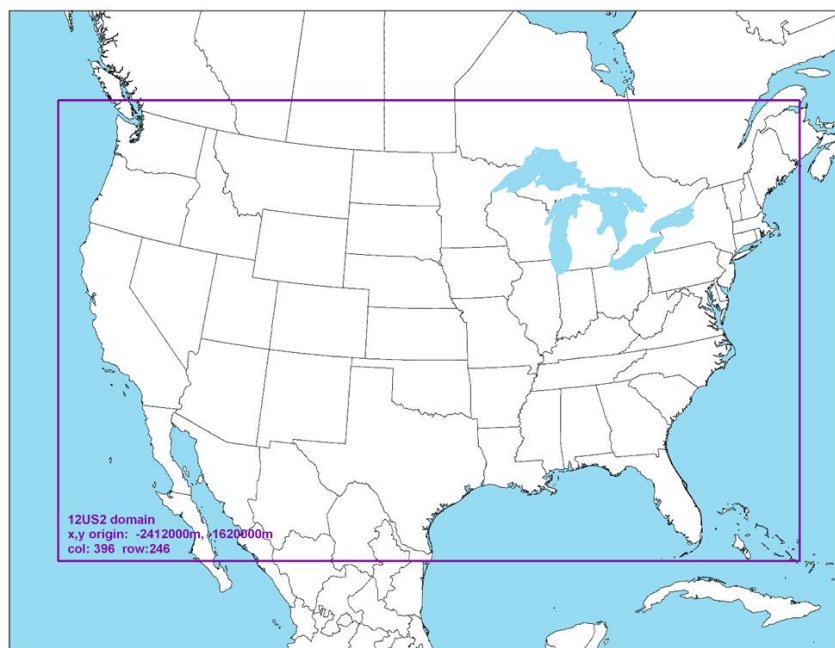


Figure 4-1. Map of the CMAQ Modeling Domain. The purple box denotes the 12 km national modeling domain. (Same as Figure 3-3.)

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

National 12 km CMAQ Modeling Configuration	
Map Projection	Lambert Conformal Projection
Grid Resolution	12 km
Coordinate Center	97 W, 40 N
True Latitudes	33 and 45 N
Dimensions	396 x 246 x 25
Vertical Extent	25 Layers: Surface to 50 mb level (see Table 4-2)

4.2.3 Modeling Period / Ozone Episodes

The 12 km CMAQ modeling domain was modeled for the entire year of 2012. The 2012 annual simulation was performed in two half-year segments (i.e., January through June, and July through December) for each emissions scenario. With this approach to segmenting an annual simulation we were able to reduce the overall throughput time for an annual simulation. The annual simulation included a “ramp-up” period, comprised of 10 days before the beginning of each half-year segment, 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

2012 Emissions: The emissions inventories used in the 2012 air quality modeling are described in Section 3, above.

Meteorological Input Data: The gridded meteorological data for the entire year of 2012 at the 12 km continental United States scale domain was derived from version 3.6.1¹⁶ of the Weather Research and Forecasting Model (WRF), Advanced Research WRF (ARW) core.¹⁷ 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 2012 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 trigger¹⁸ and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme.¹⁹ In addition, the Group for High Resolution Sea Surface Temperatures (GHRST)²⁰ 1km SST data was used for SST information to provide more resolved information compared to the more coarse data in the NAM analysis. Landuse and land cover data are based on the National Land Cover Database 2006.²¹

The WRF meteorological outputs were processed using the Meteorology-Chemistry Interface Processor (MCIP) package²², version 4.2, to derive the specific inputs to CMAQ: horizontal wind components (i.e., speed and direction), temperature, moisture, and its related speciated components was conducted for vertical diffusion rates, and rainfall rates for each grid cell in each vertical layer. The WRF simulation used the same CMAQ map projection, a Lambert Conformal projection centered at (-97, 40) with true latitudes at 33 and 45 degrees north. The 12 km WRF domain consisted of 396 by 246 grid cells and 35 vertical layers up to 50 mb. Table 4-2 shows the vertical layer structure used in WRF and the layer collapsing approach to generate the CMAQ meteorological inputs. CMAQ resolved the vertical atmosphere with 25 layers, preserving greater resolution in the PBL.

In terms of the 2012 WRF meteorological model performance evaluation, a combination of qualitative and quantitative analyses was used to assess the adequacy of the WRF simulated fields. The qualitative aspects involved comparisons of the model-estimated synoptic patterns against observed patterns from

¹⁶ Version 3.6.1 was the current version of WRF at the time the 2012 meteorological model simulation was performed.

¹⁷ 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.

¹⁸ 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>

¹⁹ 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.

²⁰ 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.

²¹ Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J., 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States, PE&RS, Vol. 77(9):858-864.

²² Otte T.L., Pleim, J.E., 2010. The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through v3.4.1. Geoscientific Model Development 3, 243-256.

historical weather chart archives. Additionally, the evaluations compared spatial patterns of monthly average rainfall and monthly maximum planetary boundary layer (PBL) heights. The statistical portion of the evaluation examined the model bias and error for temperature, water vapor mixing ratio, solar radiation, and wind fields. These statistical values were calculated on a monthly basis.

Table 4-2. Vertical layer structure for 2012 WRF and CMAQ simulations (heights are layer top).

CMAQ Layers	WRF Layers	Sigma P	Pressure (mb)	Approximate Height (m)
25	35	0	5000	17,556
	34	0.05	9750	14,780
24	33	0.1	14500	12,822
	32	0.15	19250	11,282
23	31	0.2	24000	10,002
	30	0.25	28750	8,901
22	29	0.3	33500	7,932
	28	0.35	38250	7,064
21	27	0.4	43000	6,275
	26	0.45	47750	5,553
20	25	0.5	52500	4,885
	24	0.55	57250	4,264
19	23	0.6	62000	3,683
18	22	0.65	66750	3,136
17	21	0.7	71500	2,619
16	20	0.74	75300	2,226
15	19	0.77	78150	1,941
14	18	0.8	81000	1,665
13	17	0.82	82900	1,485
12	16	0.84	84800	1,308
11	15	0.86	86700	1,134
10	14	0.88	88600	964
9	13	0.9	90500	797
	12	0.91	91450	714
8	11	0.92	92400	632
	10	0.93	93350	551
7	9	0.94	94300	470
	8	0.95	95250	390
6	7	0.96	96200	311
5	6	0.97	97150	232
4	5	0.98	98100	154
3	4	0.985	98575	115

CMAQ Layers	WRF Layers	Sigma P	Pressure (mb)	Approximate Height (m)
	3	0.99	99050	77
2	2	0.995	99525	38
1	1	0.9975	99763	19
0	0	1	100000	0

Initial and Boundary Conditions: The lateral boundary and initial species concentrations are provided by a three-dimensional global atmospheric chemistry model, the GEOS-CHEM²³ model (standard version 8-03-02 with 8-02-03 chemistry). The global GEOS-CHEM model simulates atmospheric chemical and physical processes driven by assimilated meteorological observations from the NASA's Goddard Earth Observing System (GEOS-5). This model was run for 2012 with a grid resolution of 2.0 degrees x 2.5 degrees (latitude-longitude). The predictions were processed using the GEOS-2-CMAQ tool and used to provide one-way dynamic boundary conditions at one-hour intervals.²⁴ A GEOS-Chem evaluation was conducted for the purpose of validating the 2012 GEOS-Chem simulation for predicting selected measurements relevant to their use as boundary conditions for CMAQ. This evaluation included using satellite retrievals paired with GEOS-Chem grid cells.²⁵ More information is available about the GEOS-CHEM model and other applications using this tool at: <http://gmao.gsfc.nasa.gov/GEOS/> and <http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-5>.

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 2012 simulation using state/local monitoring sites data in order to estimate the ability of the CMAQ modeling system to replicate the 2012 base year concentrations for the 12 km continental U.S. domain.

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

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

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

²³ Yantosca, B., 2004. GEOS-CHEMv7-01-02 User's Guide, Atmospheric Chemistry Modeling Group, Harvard University, Cambridge, MA, October 15, 2004.

²⁴ Akhtar, F., Henderson, B., Appel, W., Napelenok, S., Hutzell, B., Pye, H., Foley, K., 2012. Multiyear Boundary Conditions for CMAQ 5.0 from GEOS-Chem with Secondary Organic Aerosol Extensions, 11th Annual Community Modeling and Analysis System conference, Chapel Hill, NC, October 2012.

²⁵ Henderson, B.H., Akhtar, F., Pye, H.O.T., Napelenok, S.L., and Hutzell, W.T. (2014) A database and tool for boundary conditions for regional air quality modeling: description and evaluation, Geosci. Model Dev., 7, 339-360.

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

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

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

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

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

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

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

Evaluation for 8-hour Daily Maximum Ozone: The operational model performance evaluation for eight-hour daily maximum ozone was conducted using the statistics defined above. Ozone measurements for 2012 in the continental U.S. were included in the evaluation and were taken from the 2012 State/local monitoring site data in the EPA Air Quality System (AQS) and the Clean Air Status and Trends Network (CASTNet). 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 the following geographic groupings in the 12-km continental U.S. domain²⁶: five large subregions: Midwest, Northeast, Southeast, Central and Western U.S.

The 8-hour ozone model performance bias and error statistics for each subregion 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). Spatial plots of the mean bias, mean error, normalized mean bias and error for individual monitors are shown in Figures 4-2 through 4-5. The statistics shown in these two figures were calculated over the ozone season, May through September, using data pairs on days with observed 8-hour ozone of greater than or equal to 60 ppb.

In general, the model performance statistics indicate that the 8-hour daily maximum ozone concentrations

²⁶ The subregions are defined by States where: Midwest is IL, IN, MI, OH, and WI; Northeast is CT, DE, MA, MD, ME, NH, NJ, NY, PA, RI, and VT; Southeast is AL, FL, GA, KY, MS, NC, SC, TN, VA, and WV; Central is AR, IA, KS, LA, MN, MO, NE, OK, and TX; West is AK, CA, OR, WA, AZ, NM, CO, UT, WY, SD, ND, MT, ID, and NV.

predicted by the 2012 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 in the summer is slightly under predicted with the greatest under prediction in the Midwest U.S. at AQS and CASTNet sites (NMB is -6.7 percent and -10.3 percent, respectively). However, 8-hour ozone is slightly over predicted in Central U.S. at AQS sites (NMB is 4.0 percent) and in the Southeast at AQS and CASTNet sites (NMB is 7.2 percent and 2.0 percent, respectively). Ozone performance in spring shows better performance with a mix of slight under and over predictions in the U.S. at AQS and CASTNet sites. In the winter, when concentrations are generally low, the model slightly over predicts 8-hour maximum ozone with the exception of the Northeast at CASNet sites (NMB is -1.2 percent). In the fall, when concentrations are also relatively low, ozone is slightly over predicted (with NMBs less than 11 percent in each subregion); except for under prediction at AQS sites in the Central and the Western U.S (NMB is -2.7 percent and -1.0 percent, respectively).

Model bias at individual sites during the ozone season is similar to that seen on a subregional basis for the summer. Figure 4-2 shows the mean bias for 8-hour daily maximum ozone greater than 60 ppb is generally ± 6 ppb across the AQS and CASTNet sites. For some areas in the Northeast, Midwest, Central and California mean bias is under predicted by 8 to 12 ppb. Likewise, the information in Figure 4-4 indicates that the bias for days with observed 8-hour daily maximum ozone greater than 60 ppb is within ± 20 percent at the vast majority of monitoring sites across the U.S. domain. Model error, as seen from Figures 4-3 and 4-5, is generally 14 ppb and 20 percent or less at most of the sites across the U.S. modeling domain. Somewhat greater error is evident at sites in several areas most notably along portions of the California coastline, Northeast coastline, Great Lakes coastline, Seattle, WA, and North Dakota.

Table 4-4. Summary of CMAQ 2012 8-Hour Daily Maximum Ozone Model Performance Statistics by Subregion, by Season and Monitoring Network.

Subregion	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
Northeast	AQS	Winter	11,007	0.7	4.3	2.4	14.9
		Spring	16,122	1.2	5.7	2.6	12.6
		Summer	17,696	-0.3	7.4	-0.6	14.9
		Fall	14,473	2.5	6.2	7.5	18.7
	CASTNet	Winter	1,248	-0.4	3.7	-1.2	11.5
		Spring	1,309	0.1	5.0	0.2	10.8
		Summer	1,238	-1.4	6.6	-2.9	13.9
		Fall	1,106	2.1	5.6	6.2	16.3

Subregion	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
Midwest	AQS	Winter	4,192	2.5	4.6	9.6	17.5
		Spring	12,777	0.7	5.6	1.5	11.6
		Summer	16,772	-3.7	7.5	-6.7	13.6
		Fall	10,055	0.3	6.0	0.8	16.9
	CASTNet	Winter	1,037	1.8	4.0	6.2	13.5
		Spring	992	-0.8	5.0	-1.5	10.2
		Summer	952	-5.8	7.4	-10.3	13.1
		Fall	980	0.0	5.1	14.2	0.9
Central States	AQS	Winter	12,455	2.0	5.1	6.1	16.0
		Spring	16,394	2.7	6.8	5.7	14.5
		Summer	17,814	2.0	9.7	4.0	19.7
		Fall	16,391	0.7	6.5	1.6	15.9
	CASTNet	Winter	655	0.2	3.9	0.5	11.2
		Spring	688	-0.8	5.7	-1.6	11.6
		Summer	645	-3.3	7.7	-6.3	14.5
		Fall	652	-1.1	5.8	-2.7	13.9
Southeast	AQS	Winter	7,623	2.5	5.4	7.0	15.3
		Spring	21,864	3.4	6.5	7.2	13.9
		Summer	24,106	4.2	8.6	9.0	18.8
		Fall	18,706	3.7	6.7	9.6	17.3
	CASTNet	Winter	1,904	1.7	4.4	4.9	12.5
		Spring	1,901	0.8	5.3	1.6	10.9
		Summer	1,864	1.0	7.2	2.0	14.5
		Fall	1,832	1.5	5.4	3.9	13.6
West	AQS	Winter	29,291	2.7	5.5	7.8	16.0
		Spring	32,626	-0.7	5.5	-1.5	10.8
		Summer	35,804	-0.9	7.4	-1.7	14.0
		Fall	32,898	1.3	6.3	3.1	14.8
	CASTNet	Winter	1,634	0.0	4.3	0.0	10.3
		Spring	1,675	-3.5	5.7	-6.4	10.3
		Summer	1,639	-3.4	7.4	-6.0	13.0
		Fall	1,599	-0.4	4.9	-1.0	10.8

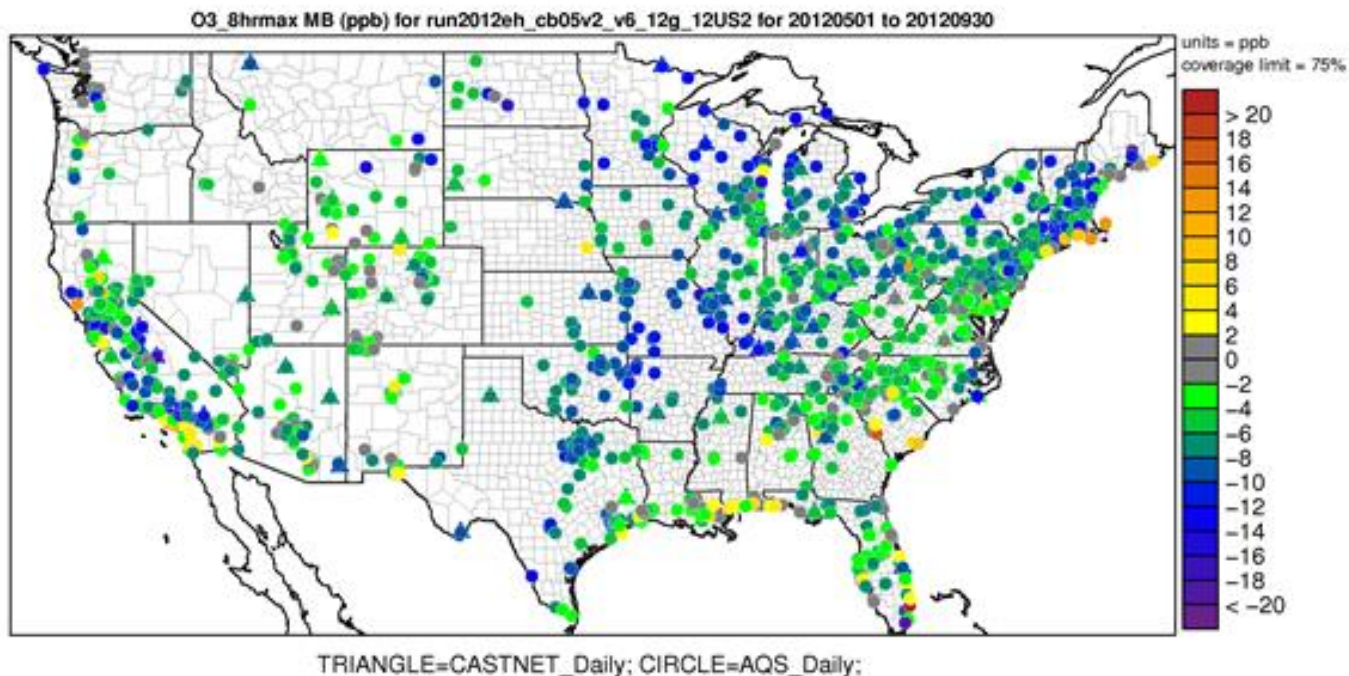


Figure 4-2. Mean Bias (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2012 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

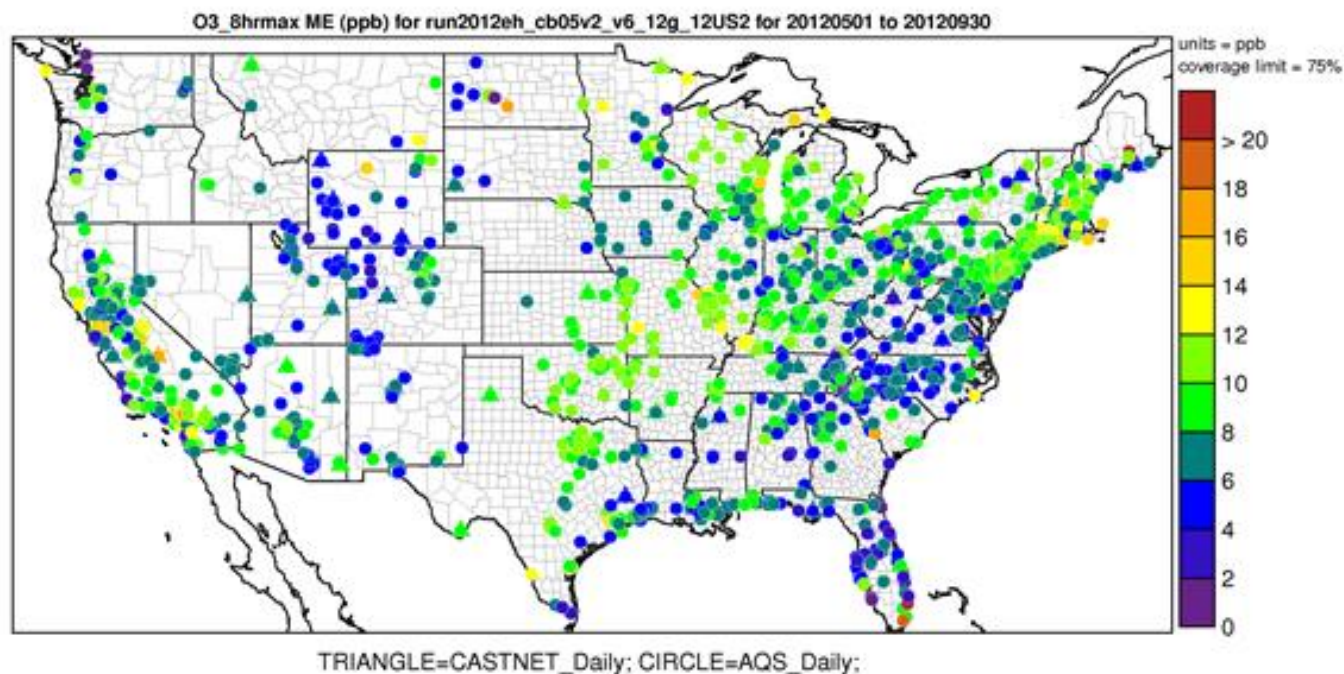


Figure 4-4. Mean Error (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2012 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

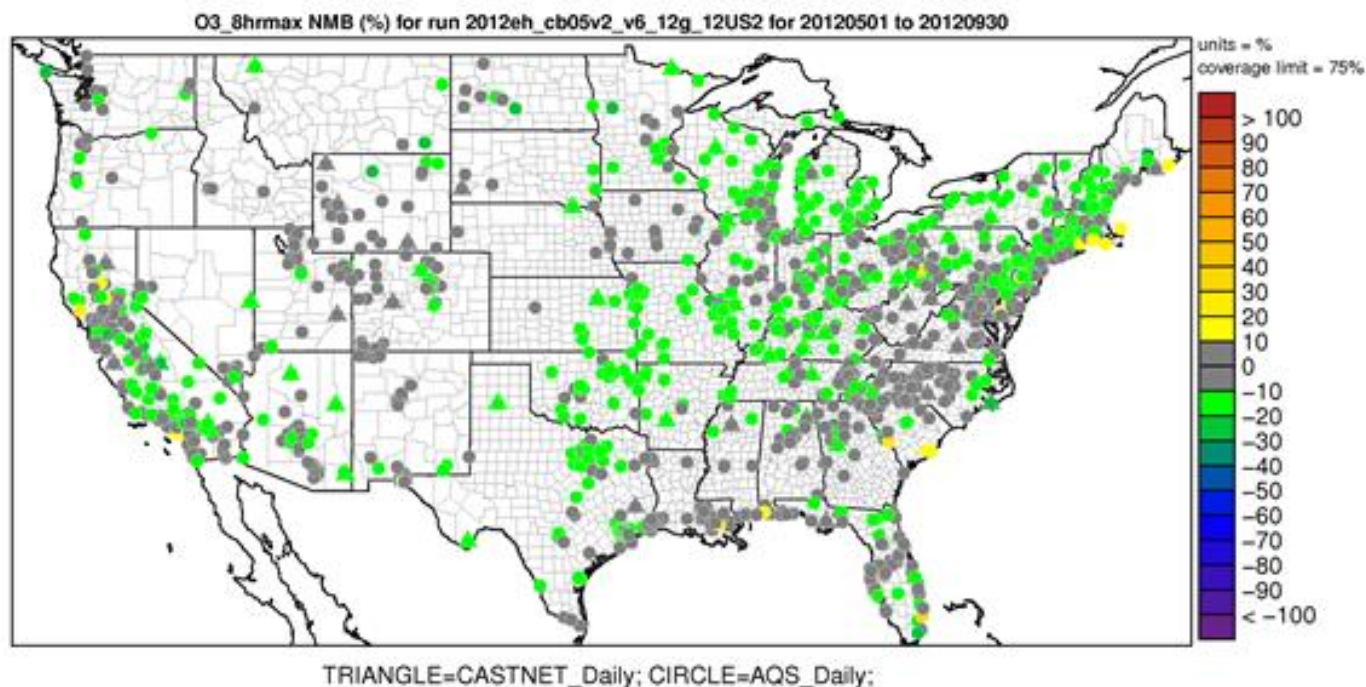


Figure 4-3. Normalized Mean Bias (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2012 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

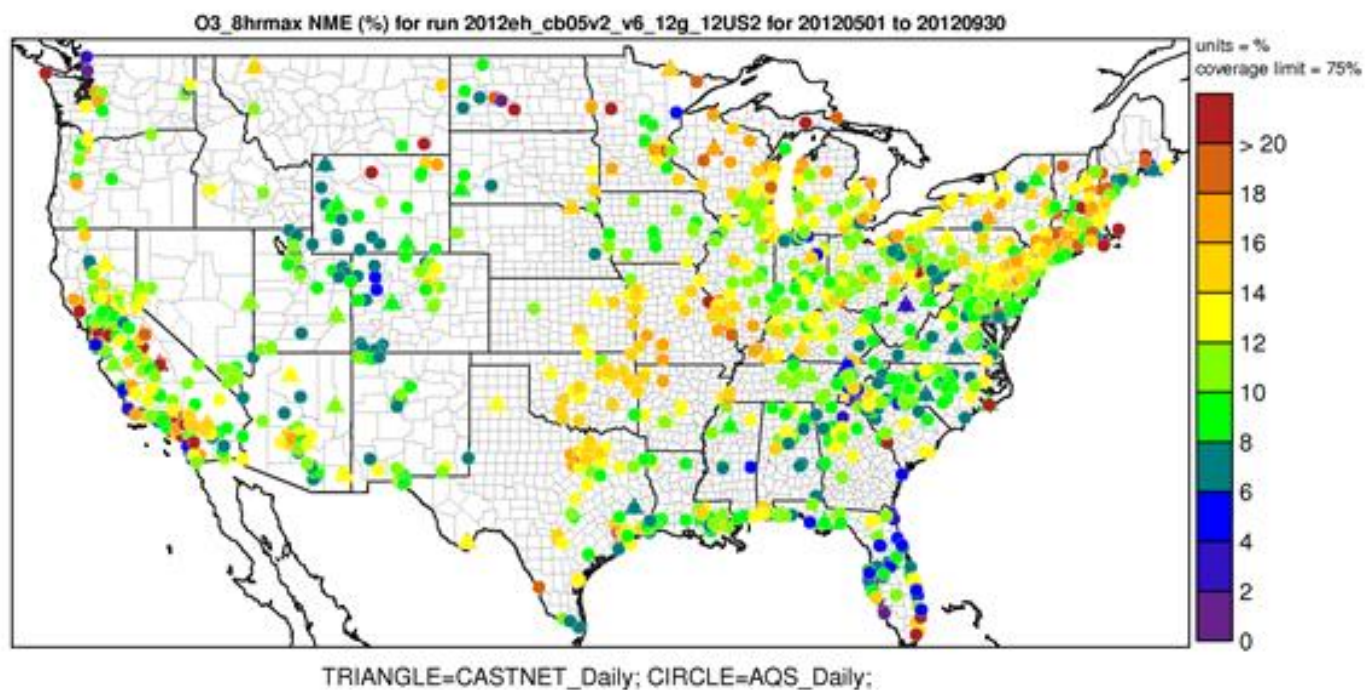


Figure 4-5. Normalized Mean Error (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2012 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

Evaluation for Annual PM_{2.5} components: The PM evaluation focuses on PM_{2.5} components including sulfate (SO₄), nitrate (NO₃), total nitrate (TNO₃ = NO₃ + HNO₃), ammonium (NH₄), elemental carbon (EC), and organic carbon (OC). The bias and error performance statistics were calculated on an annual basis for each subregion (Table 4-5). PM_{2.5} measurements for 2012 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. For brevity, Table 4-5 provides annual model performance statistics for the PM_{2.5} component species for the five sub-regions in the 12 km continental U.S. domain defined above (Northeast, Midwest, Southeast, Central, and West). 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 4-6 through 4-29.

As indicated by the statistics in Table 4-5, annual average sulfate is consistently under predicted at CSN, IMPROVE, and CASTNet monitoring sites across the modeling domain, with MB values ranging from 0.0 to -0.6 µgm⁻³ and NMB values ranging from near negligible to -30.3 percent. Sulfate performance shows moderate error, ranging from 24 to 51.4 percent. Figures 4-6 through 4-9, suggest spatial patterns vary by region. The model bias for most of the Northeast, Southeast, Central and Southwest states are within ±30 percent. The model bias appears to be slightly greater in the Northwest with over predictions up to 80 percent at individual monitors. Model error also shows a spatial trend by region, where much of the Eastern states are 20 to 40 percent, the Western and Central U.S. states are 30 to 80 percent.

Annual average nitrate is under predicted at the urban CSN monitoring sites in the Northeast, Midwest, and Southeast (NMB in the range of 7.7 to 19.9 percent), except in the Central and Western U.S. where nitrate is under predicted (NMB is -7.4 percent and -36.5 percent, respectively). At IMPROVE rural sites, annual average nitrate is over predicted at all subregions, except in the Western U.S. where nitrate is under predicted by -25.8 percent. Model performance of total nitrate at sub-urban CASTNet monitoring sites also shows an over prediction across all subregions (NMB in the range of 3.7 to 19 percent), except in the Central and Western U.S. (NMB on average is underpredicted by 23 percent). Model error for nitrate is somewhat greater for each subregion as compared to sulfate. Model bias at individual sites indicates mainly over prediction of greater than 20 percent at most monitoring sites in the Eastern half of the U.S. as indicated in Figure 4-12. The exception to this is in the Southern Florida and the Southwest of the modeling domain where there appears to be a greater number of sites with under prediction of nitrate of 20 to 80 percent. Model error for annual nitrate, as shown in Figures 4-11 and 4-13, is least at sites in portions of the Midwest and extending eastward to the Northeast corridor. Nitrate concentrations are typically higher in these areas than in other portions of the modeling domain.

Annual average ammonium model performance as indicated in Table 4-5 has a tendency for the model to under predict across the CASTNet sites (ranging from 7 to -40 percent). Ammonium performance across the urban CSN sites shows an over prediction (ranging from 7 to 28 percent) in the Northeast, Midwest, Southeast, and Central U.S., and an under prediction of approximately 25 percent in the Western U.S. There is not a large variation from subregion to subregion or at urban versus rural sites in the error statistics for ammonium. The spatial variation of ammonium across the majority of individual

monitoring sites in the Eastern U.S. shows bias within ± 30 percent. A larger bias is seen in the Southwest, bias on average 60-70 percent.

Annual average elemental carbon is over predicted in all subregions at urban and rural sites. Similar to ammonium error there is not a large variation 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 3 to 56 percent), except in the Western U.S. where the bias is approximately -2 percent. The model over predicted annual average organic carbon in all subregions at urban CSN sites. Similar to ammonium and 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 2012 Annual PM Species Model Performance Statistics by Subregion, by Monitoring Network.

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
Sulfate	CSN	Northeast	2,054	-0.1	0.6	-5.3	29.7
		Midwest	1,747	-0.3	0.7	-12.5	30.9
		Southeast	1,813	-0.1	0.6	-7.2	29.5
		Central	1,150	-0.2	0.6	-10.8	32.2
		West	1,624	0.0	0.4	-5.3	44.6
	IMPROVE	Northeast	1,824	0.2	0.4	14.7	39.5
		Midwest	442	-0.1	0.5	-6.1	34.4
		Southeast	1,750	-0.2	0.6	-11.4	31.9
		Central	2,253	-0.2	0.5	-13.7	34.4
		West	9,636	0.0	0.3	-3.7	51.4
	CASTNet	Northeast	788	-0.3	0.4	-15.6	24.0
		Midwest	592	-0.5	0.6	-24.2	26.3
		Southeast	1,134	-0.6	0.6	-26.3	28.7
		Central	386	-0.5	0.6	-30.3	33.6
		West	1,039	-0.1	0.3	-21.7	40.4
Nitrate	CSN	Northeast	2,054	0.1	0.7	7.7	62.4
		Midwest	1,747	0.2	0.7	11.0	51.2
		Southeast	1,813	0.1	0.4	19.9	85.0
		Central	1,150	-0.1	0.6	-7.4	60.0
		West	1,624	-0.5	0.9	-36.5	68.0

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
	IMPROVE	Northeast	1,824	0.2	0.4	67.6	118.0
		Midwest	442	0.2	0.5	38.5	74.2
		Southeast	1,749	0.1	0.4	20.2	106.0
		Central	2,252	0.1	0.5	14.9	73.7
		West	9,603	-0.1	0.2	-25.8	89.6
Total Nitrate ($\text{NO}_3 + \text{HNO}_3$)	CASTNet	Northeast	788	0.3	0.5	19.0	36.6
		Midwest	592	0.7	0.1	7.7	27.4
		Southeast	1,134	0.1	0.6	3.7	41.8
		Central	386	-0.3	0.6	-17.4	38.2
		West	1,038	-0.2	0.4	-27.6	45.7
Ammonium	CSN	Northeast	2,054	0.1	0.3	18.3	43.7
		Midwest	1,747	0.1	0.3	7.2	37.8
		Southeast	1,813	0.1	0.3	27.9	52.0
		Central	1,150	0.0	0.3	6.7	50.1
		West	1,624	-0.2	0.3	-25.4	75.1
	CASTNet	Northeast	788	0.0	0.2	-6.9	26.6
		Midwest	592	-0.1	0.3	-7.1	27.0
		Southeast	1,134	-0.1	0.2	-17.3	32.1
		Central	386	-0.1	0.2	-14.0	36.5
		West	1,039	-0.1	0.1	-39.9	56.5
Elemental Carbon	CSN	Northeast	1,979	0.3	0.5	67.8	84.6
		Midwest	1,668	0.3	0.4	56.2	71.0
		Southeast	1,775	0.3	0.4	46.4	71.1
		Central	1,144	0.5	0.5	94.1	106.0
		West	1,582	0.6	0.7	99.0	117.0
	IMPROVE	Northeast	1,929	0.2	0.2	67.8	84.6
		Midwest	473	0.1	0.1	40.5	63.7
		Southeast	1,806	0.1	0.2	40.1	64.1
		Central	2,304	0.1	0.2	38.9	61.7
		West	10,366	0.1	0.2	70.9	108.0

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
Organic Carbon	CSN	Northeast	1,966	1.2	1.6	75.3	95.1
		Midwest	1,629	0.6	1.0	38.3	60.0
		Southeast	1,775	0.6	1.2	32.6	61.6
		Central	1,139	0.8	1.2	52.5	78.4
		West	1,574	1.8	2.2	101.0	123.0
	IMPROVE	Northeast	1,927	0.5	0.8	55.6	86.8
		Midwest	472	0.4	0.7	39.4	76.8
		Southeast	1,817	0.1	0.8	9.3	60.8
		Central	2,307	0.0	0.5	3.0	53.2
		West	10,124	0.0	0.6	-1.8	66.2

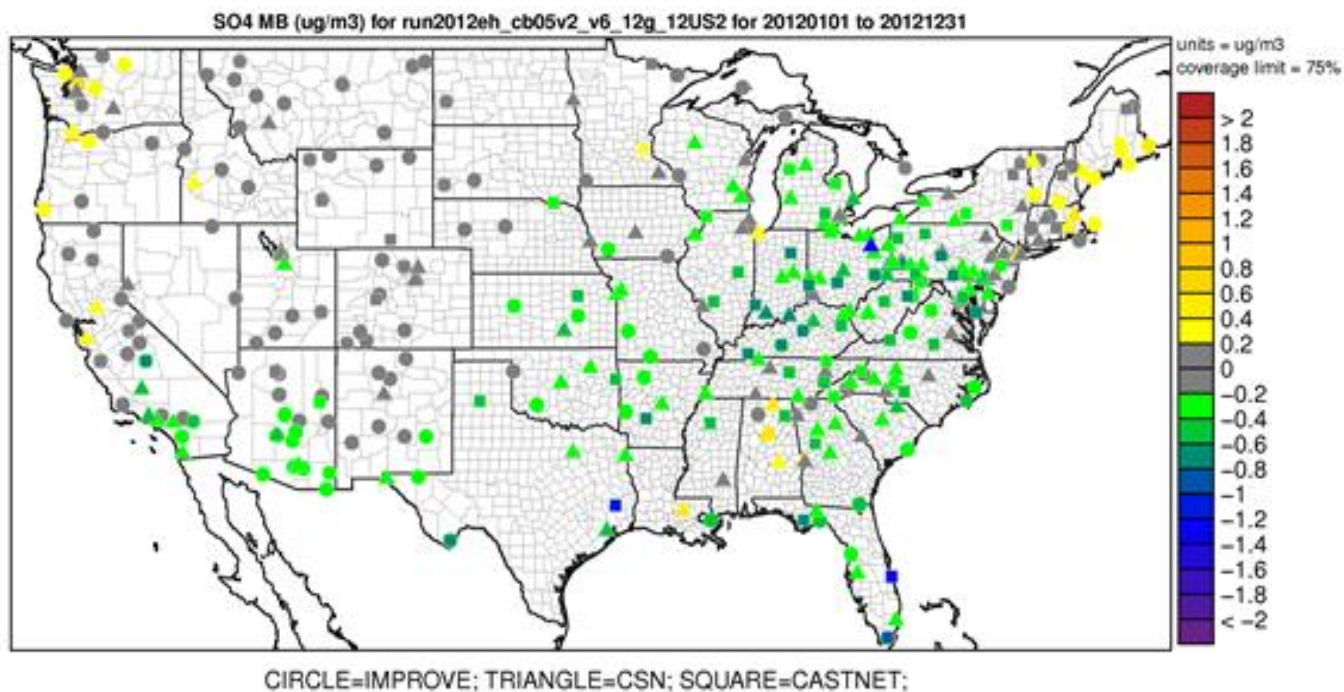


Figure 4-6. Mean Bias ($\mu\text{g}\cdot\text{m}^{-3}$) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

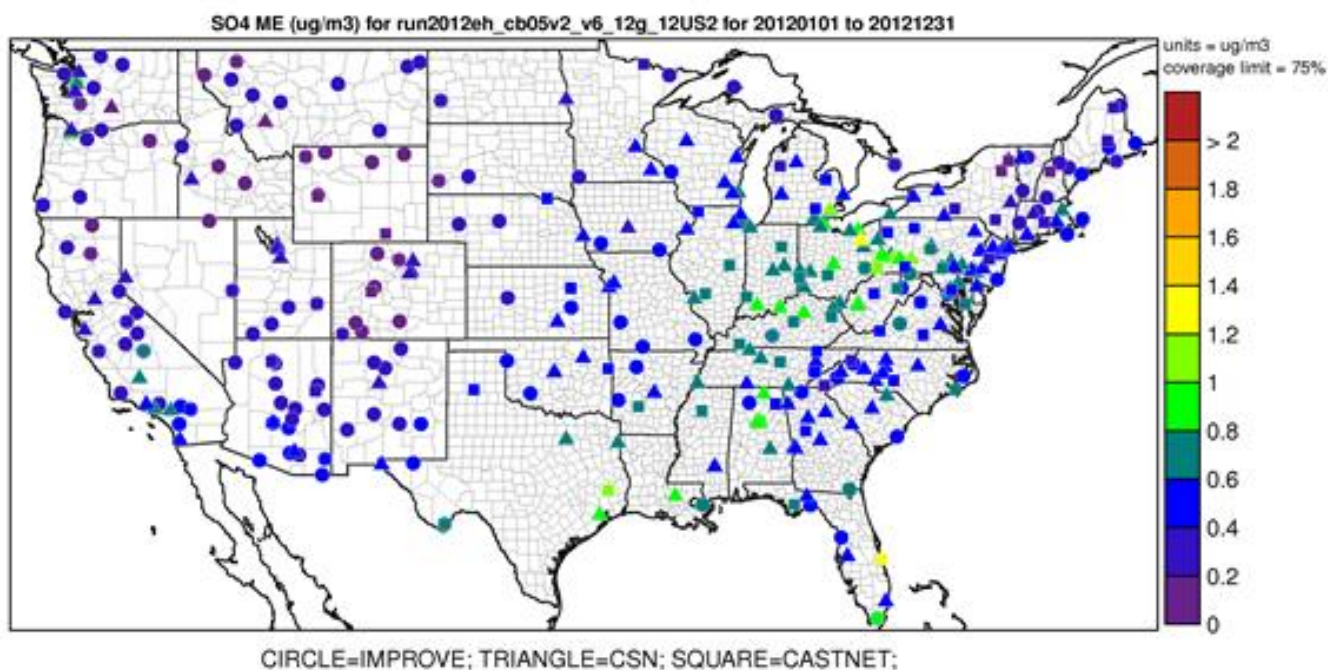


Figure 4-7. Mean Error ($\mu\text{g}\cdot\text{m}^{-3}$) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

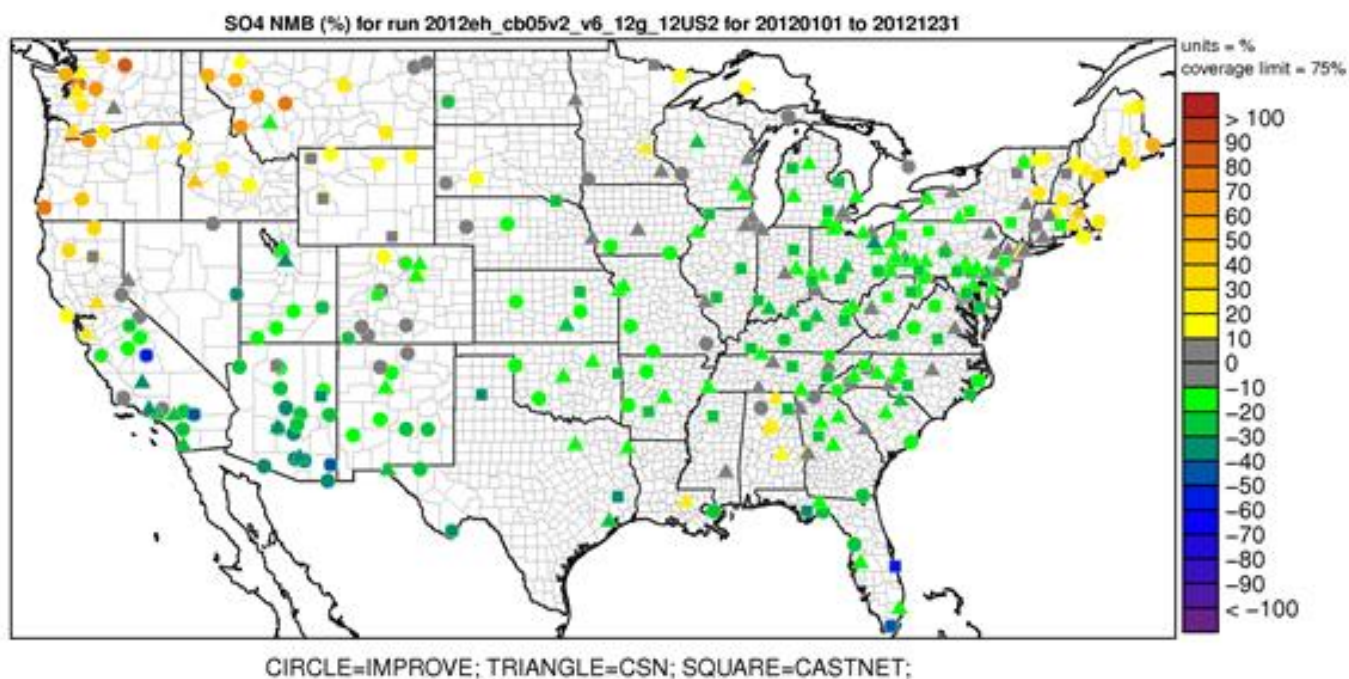


Figure 4-8. Normalized Mean Bias (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

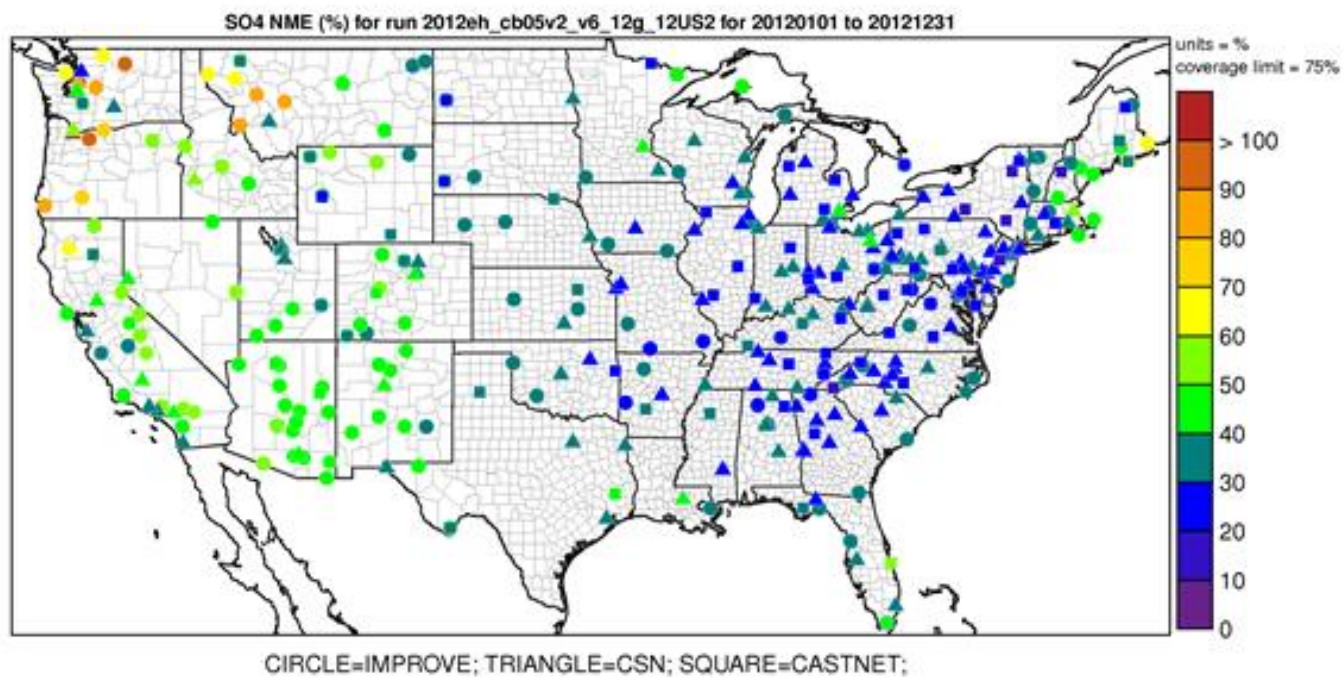


Figure 4-9. Normalized Mean Error (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

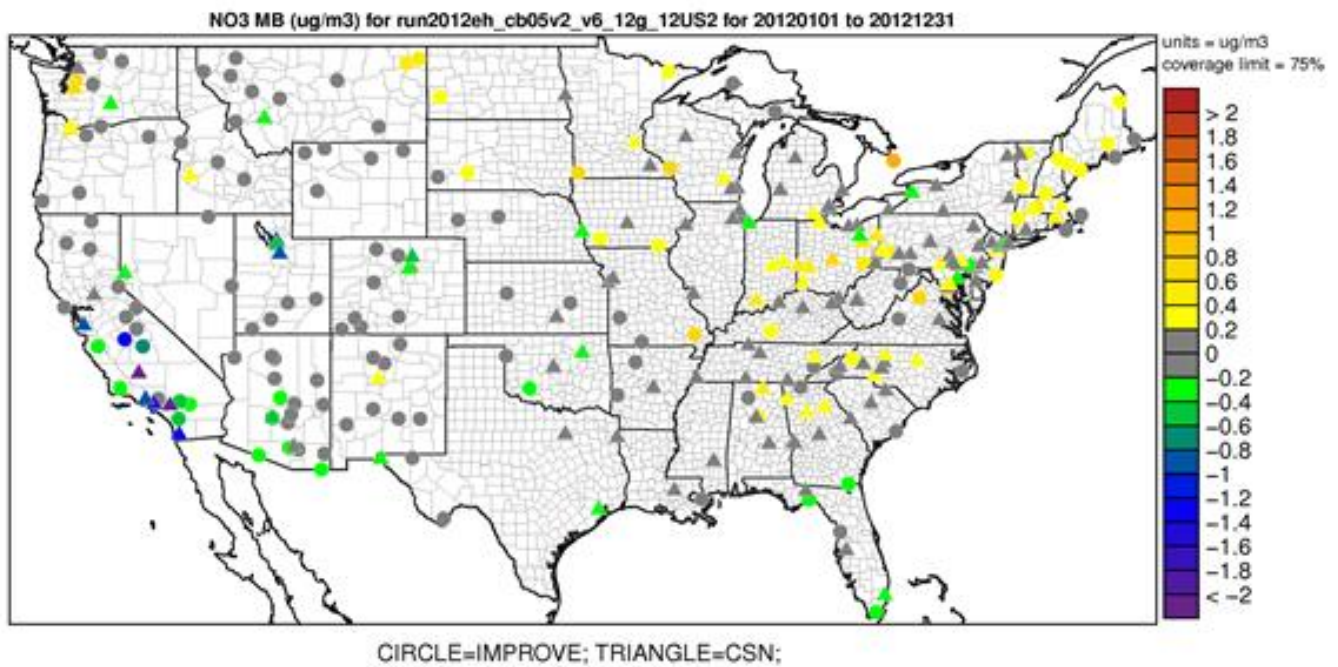


Figure 4-10. Mean Bias ($\mu\text{g m}^{-3}$) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

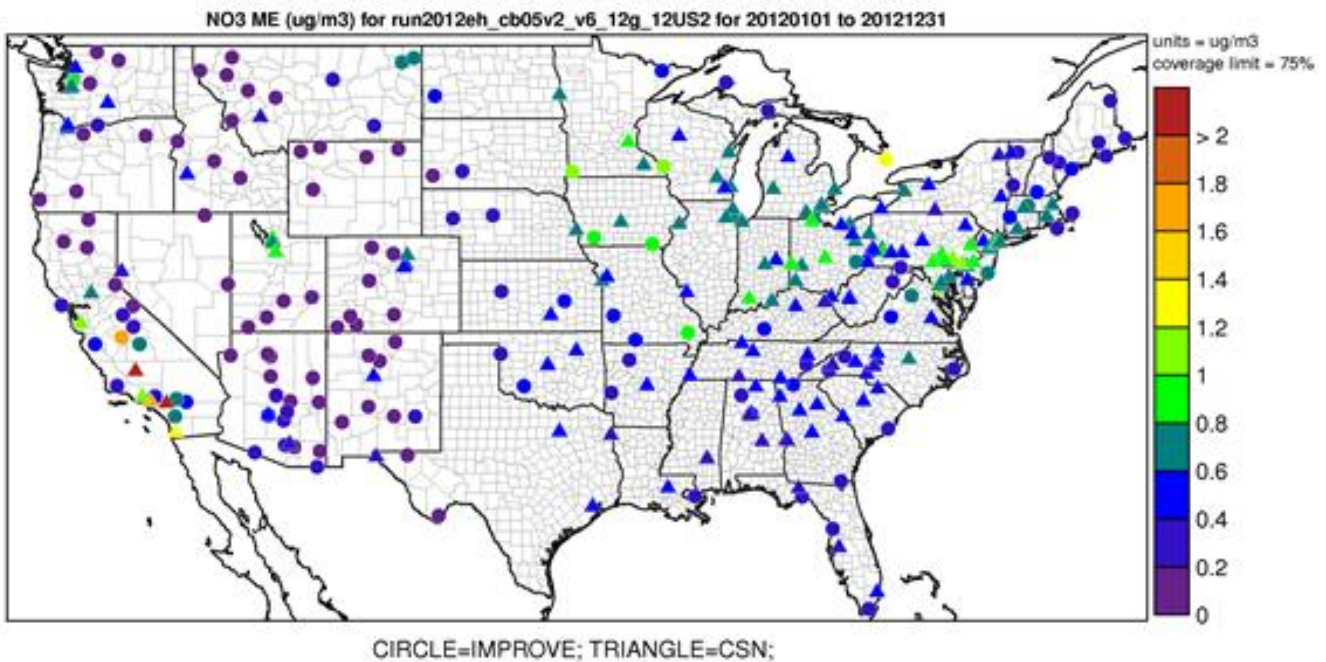


Figure 4-11. Mean Error ($\mu\text{g m}^{-3}$) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

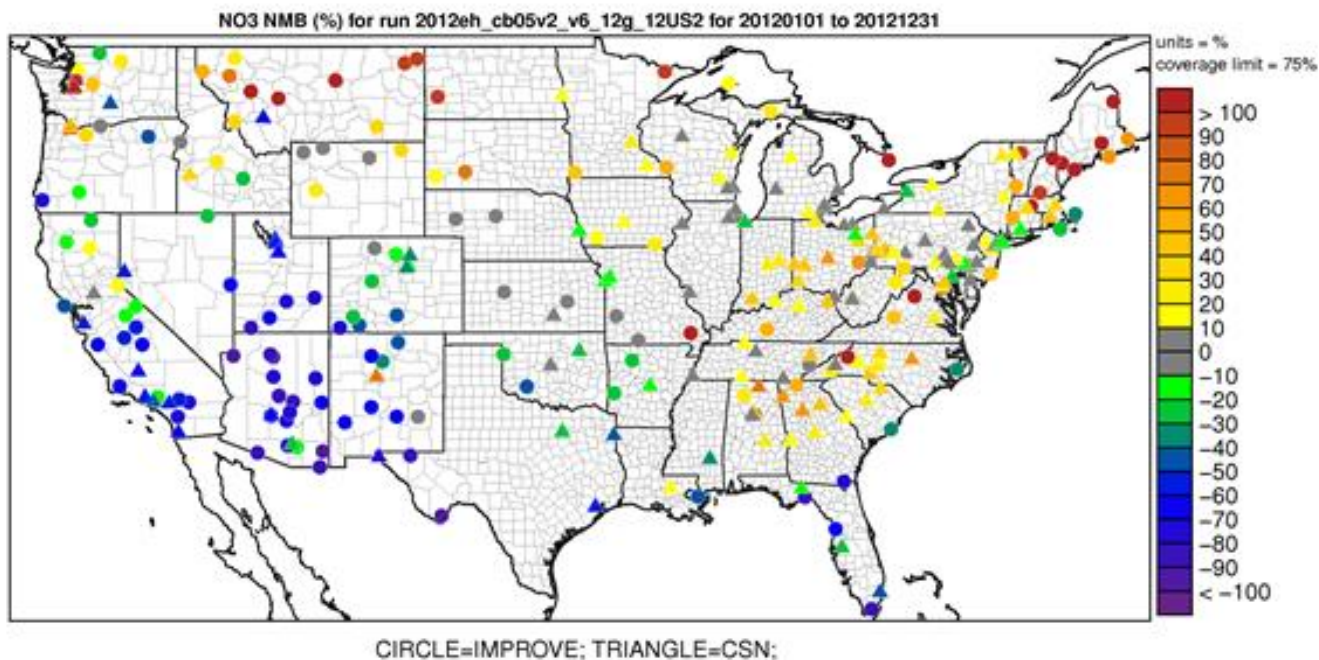


Figure 4-12. Normalized Mean Bias (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

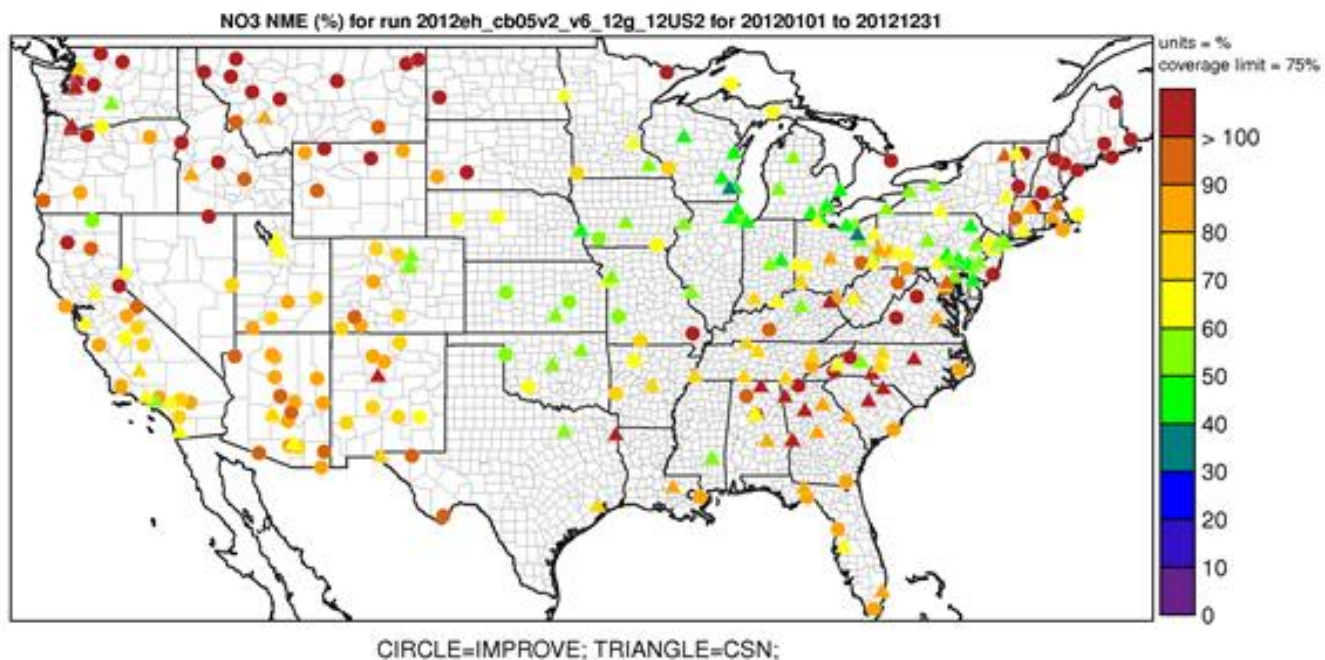


Figure 4-13. Normalized Mean Error (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.

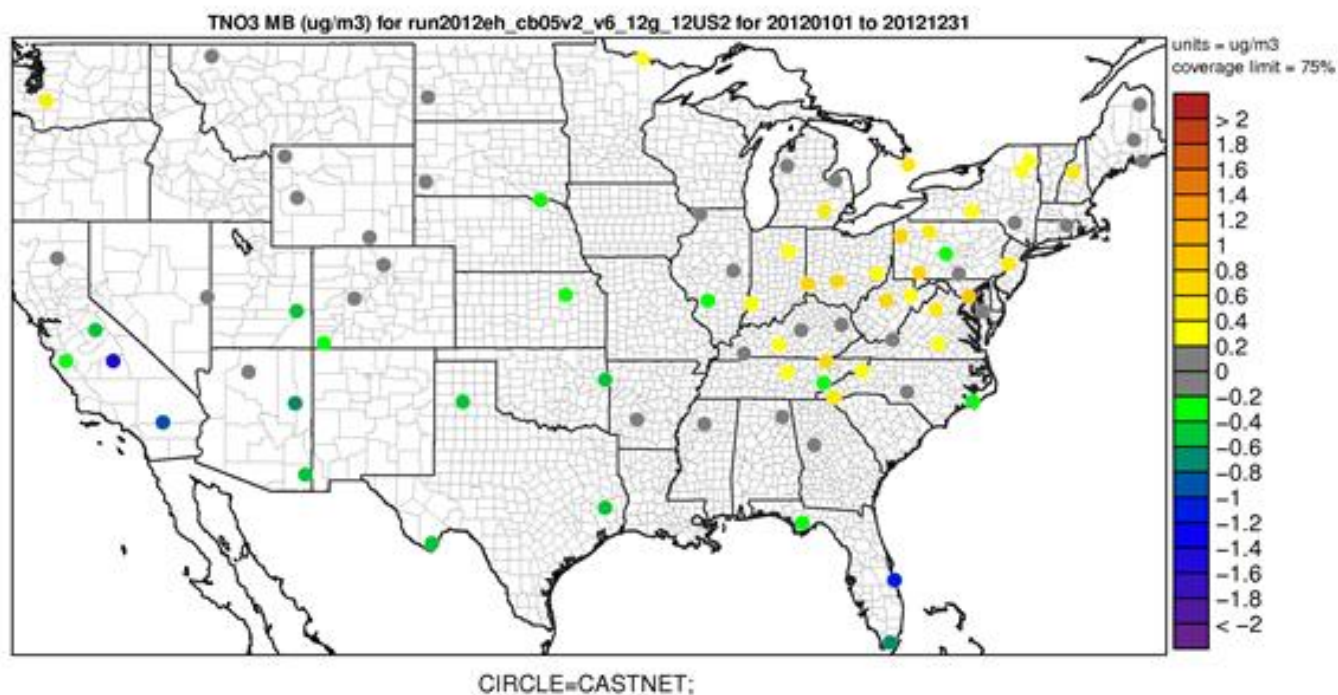


Figure 4-14. Mean Bias ($\mu\text{g}\text{m}^{-3}$) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

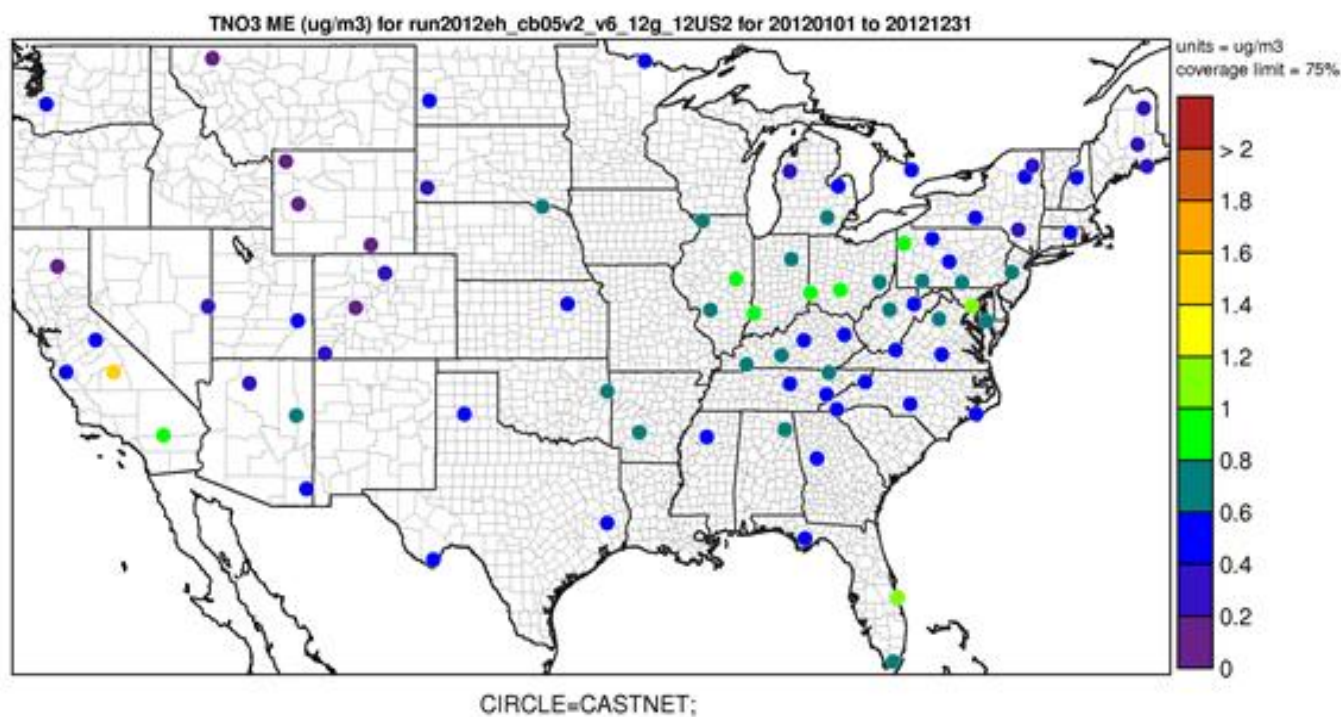


Figure 4-15. Mean Error ($\mu\text{g}\text{m}^{-3}$) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

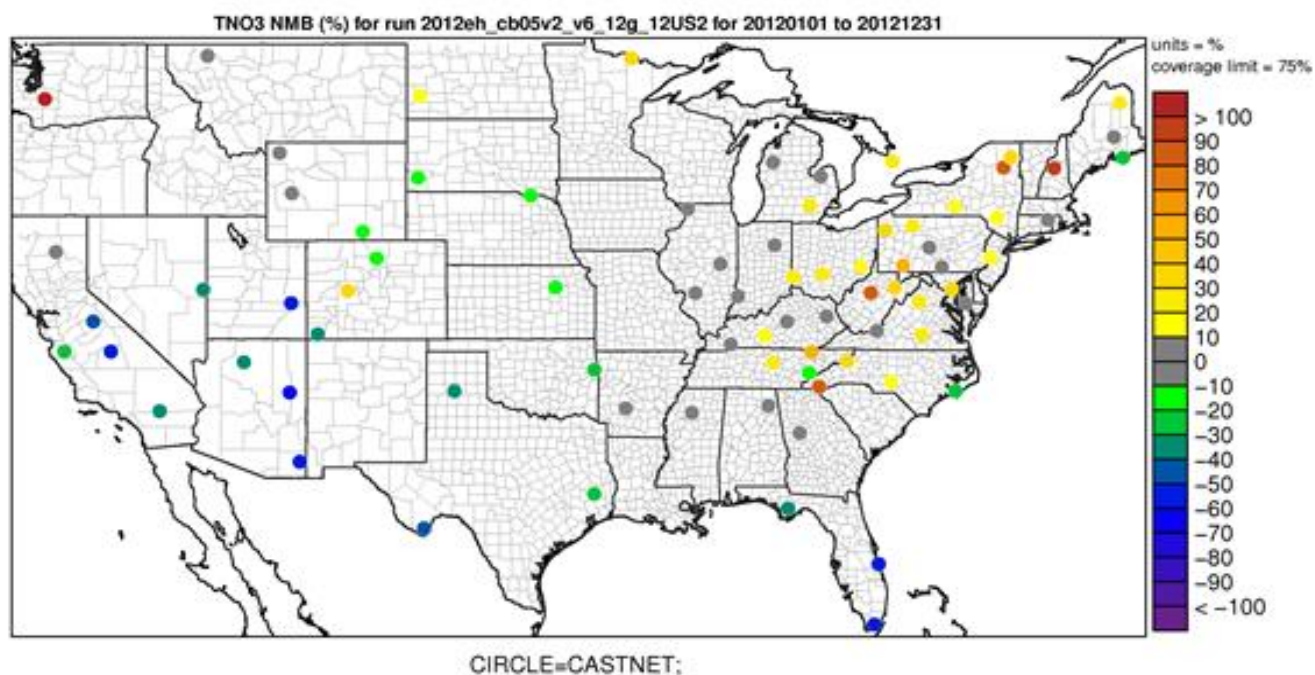


Figure 4-16. Normalized Mean Bias (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

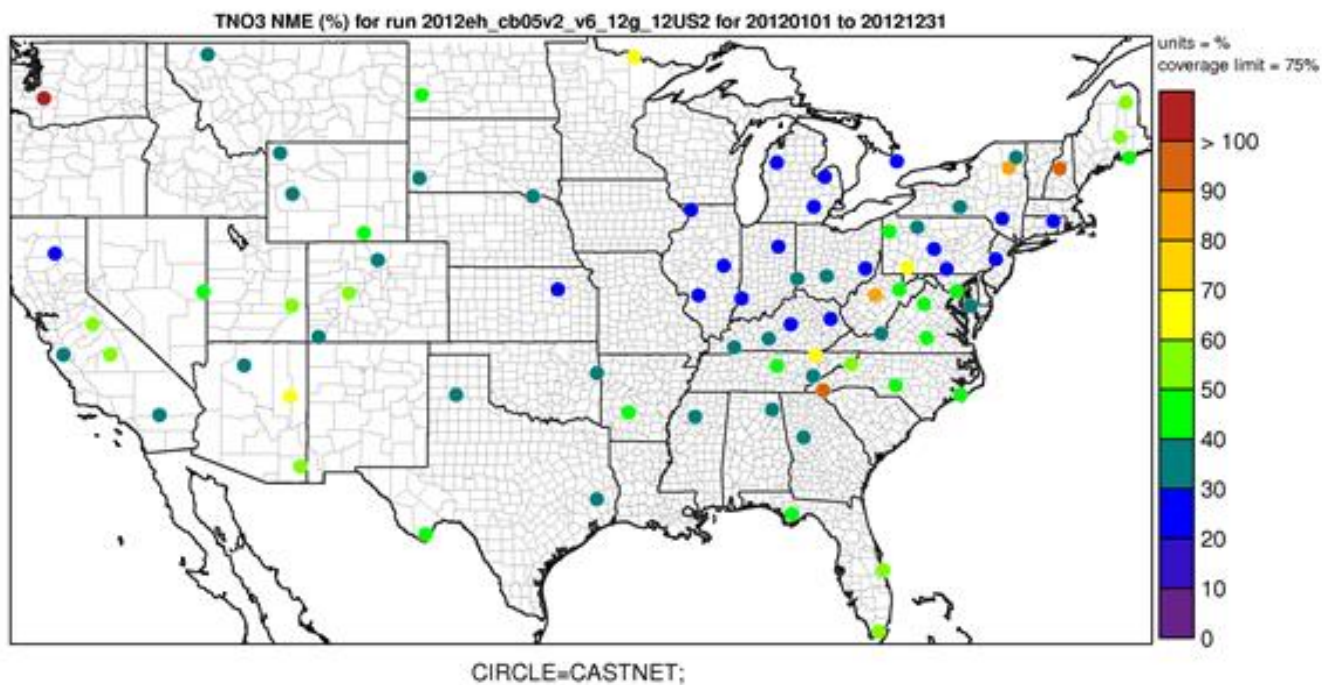


Figure 4-17. Normalized Mean Error (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.

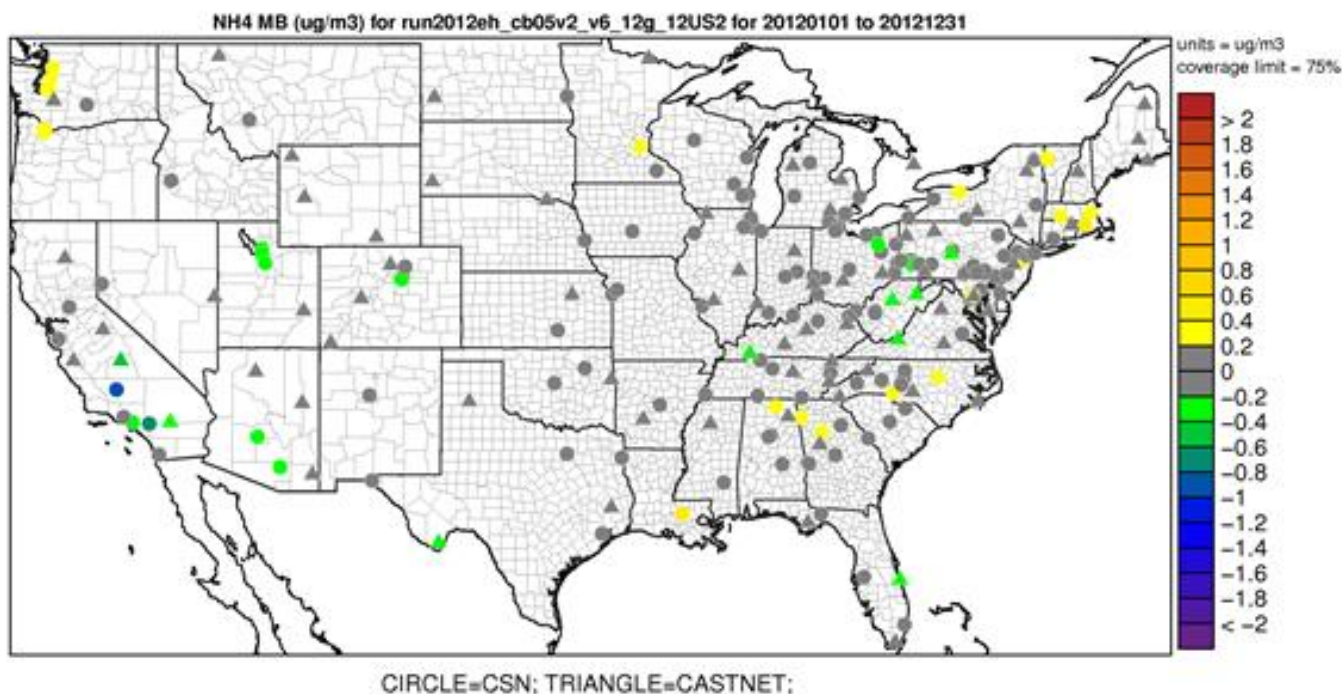


Figure 4-18. Mean Bias ($\mu\text{g m}^{-3}$) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

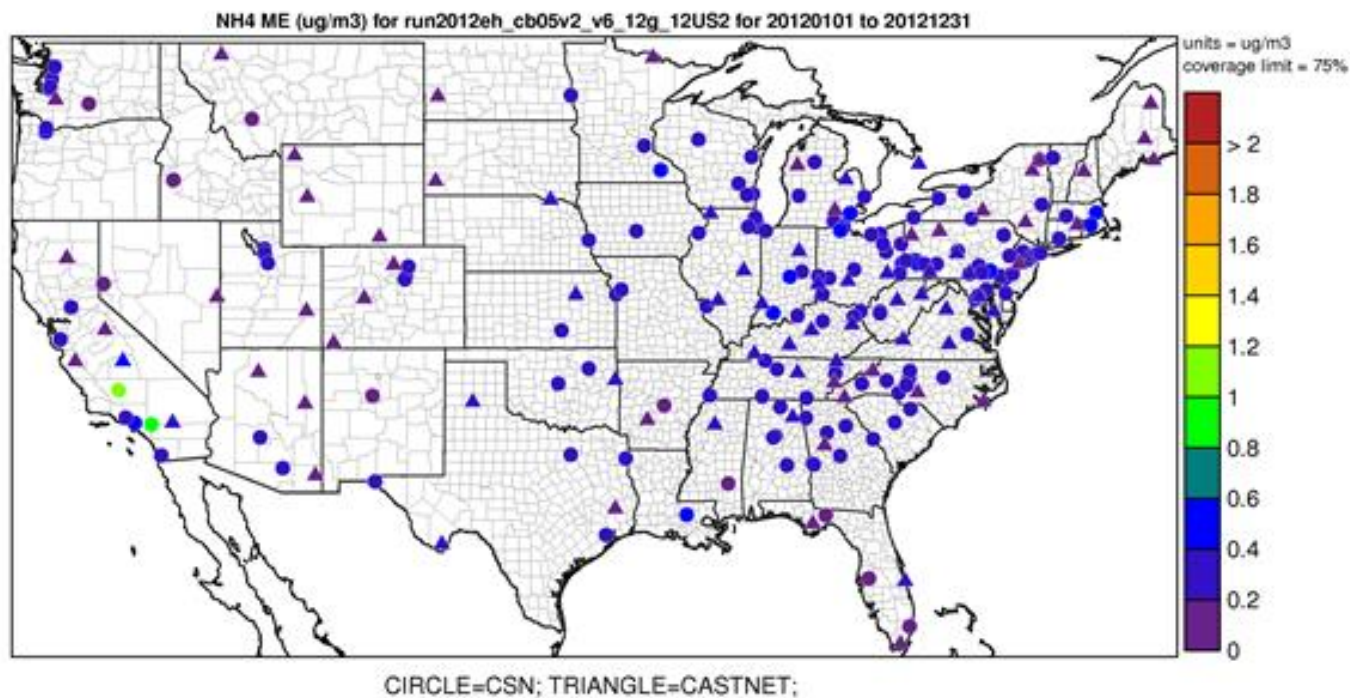


Figure 4-19. Mean Error ($\mu\text{g m}^{-3}$) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

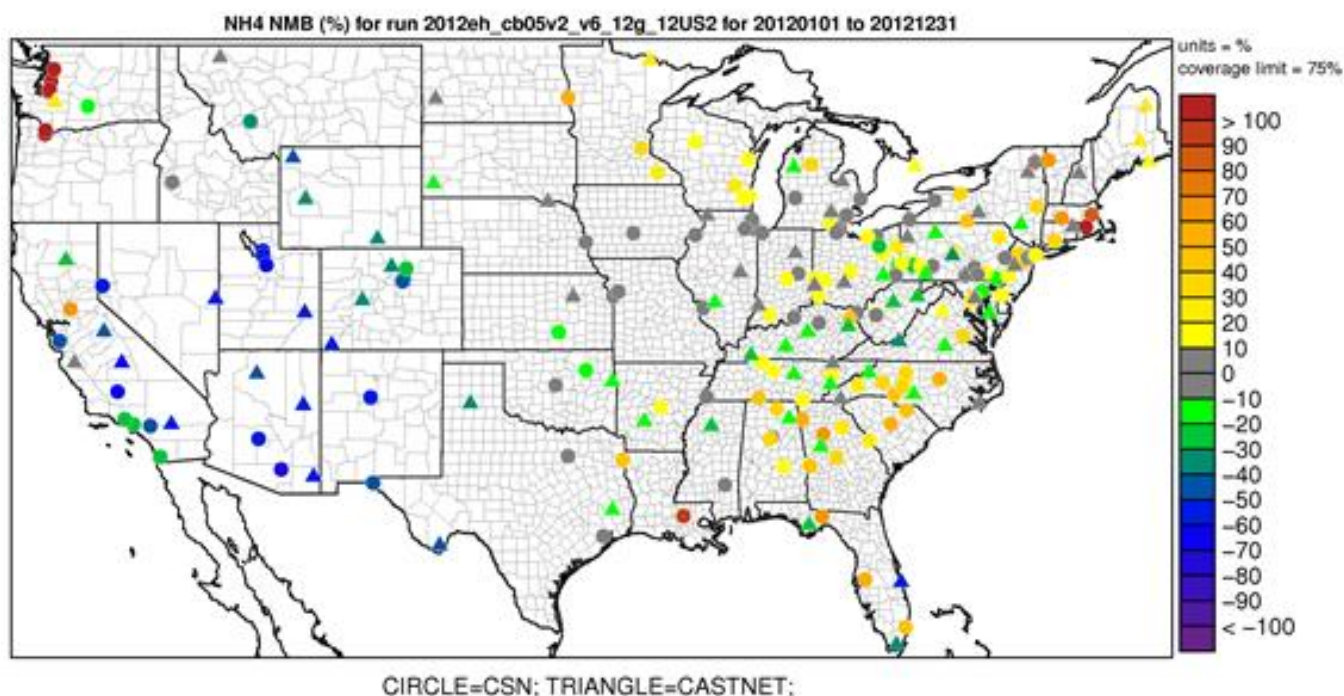


Figure 4-20. Normalized Mean Bias (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

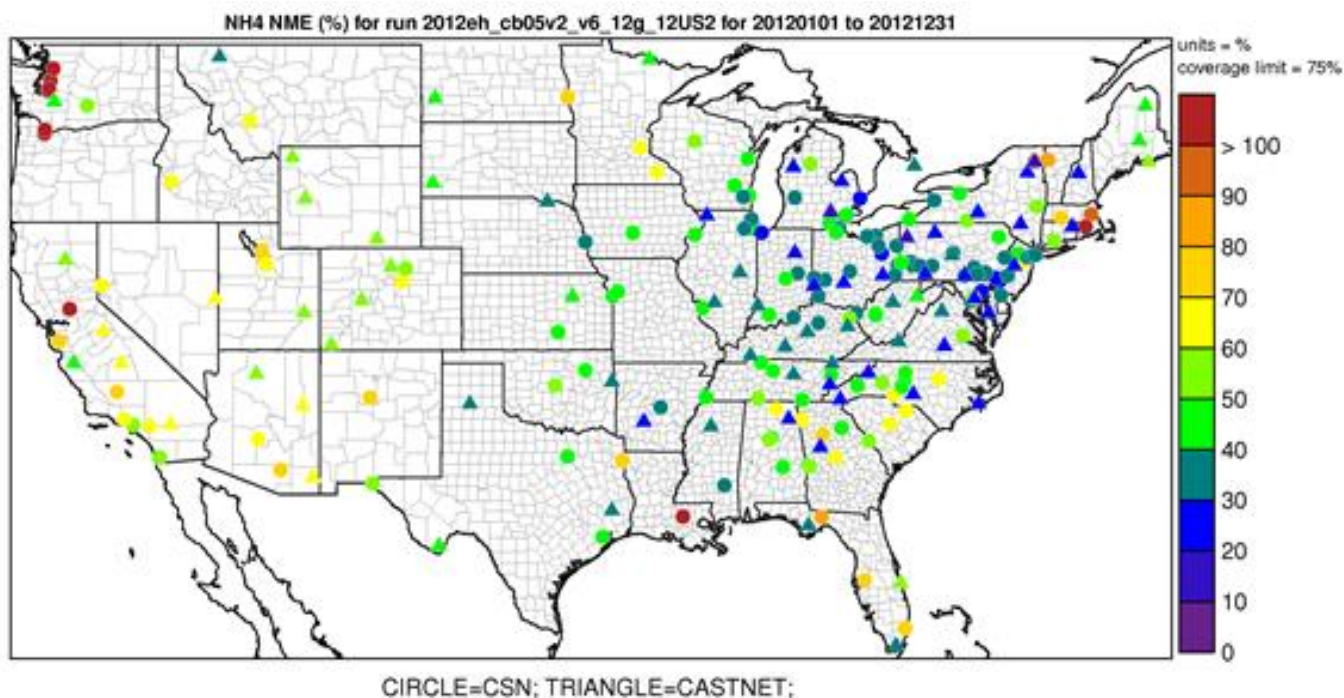


Figure 4-21. Normalized Mean Error (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.

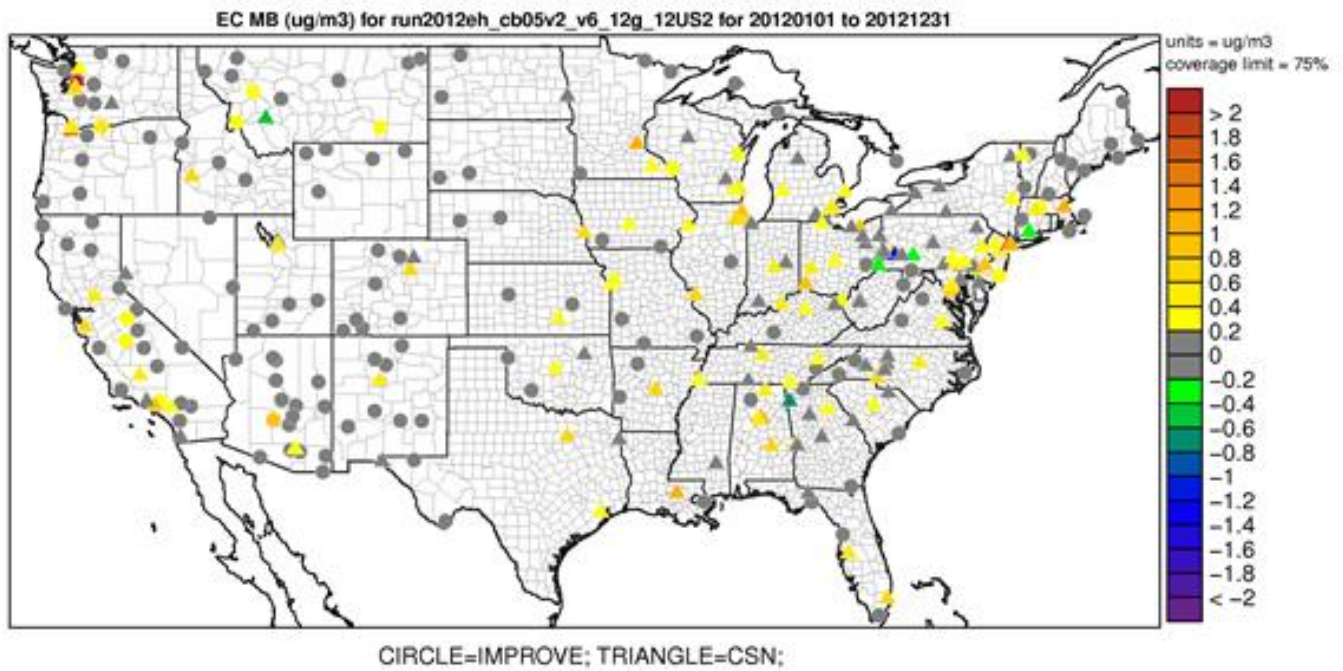


Figure 4-22. Mean Bias ($\mu\text{g}\text{m}^{-3}$) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

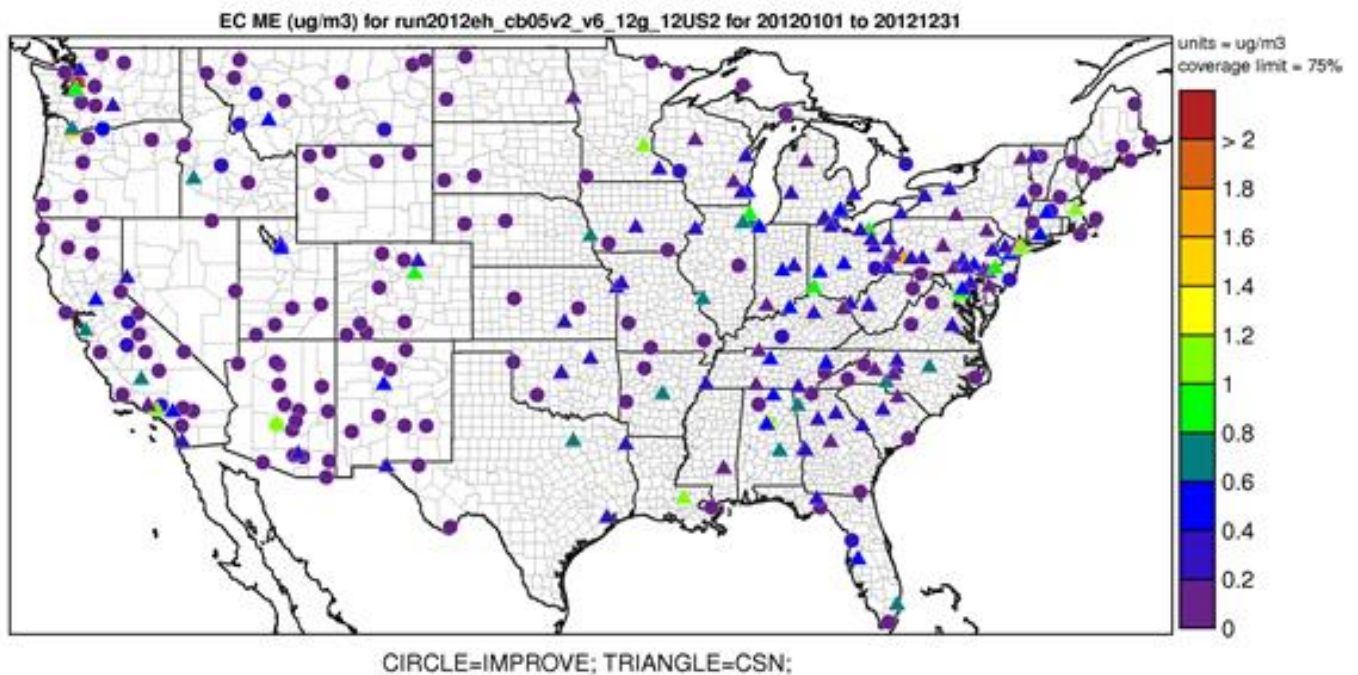


Figure 4-23. Mean Error ($\mu\text{g}\text{m}^{-3}$) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

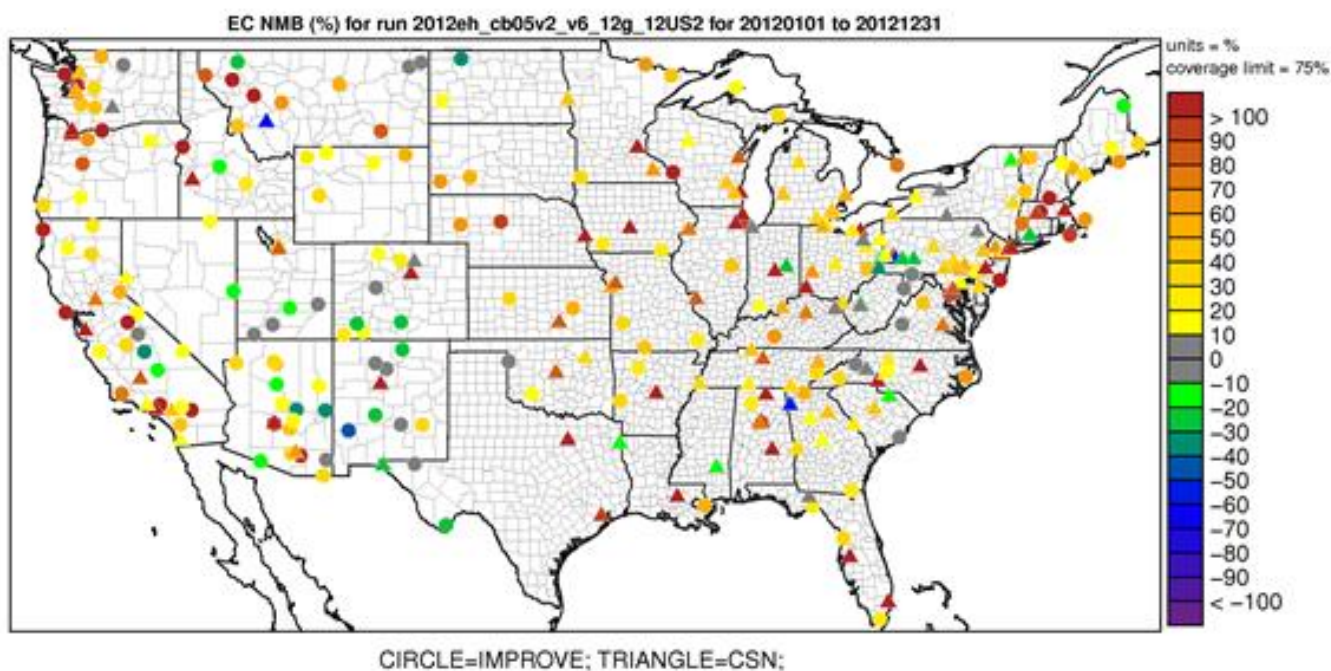


Figure 4-24. Normalized Mean Bias (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

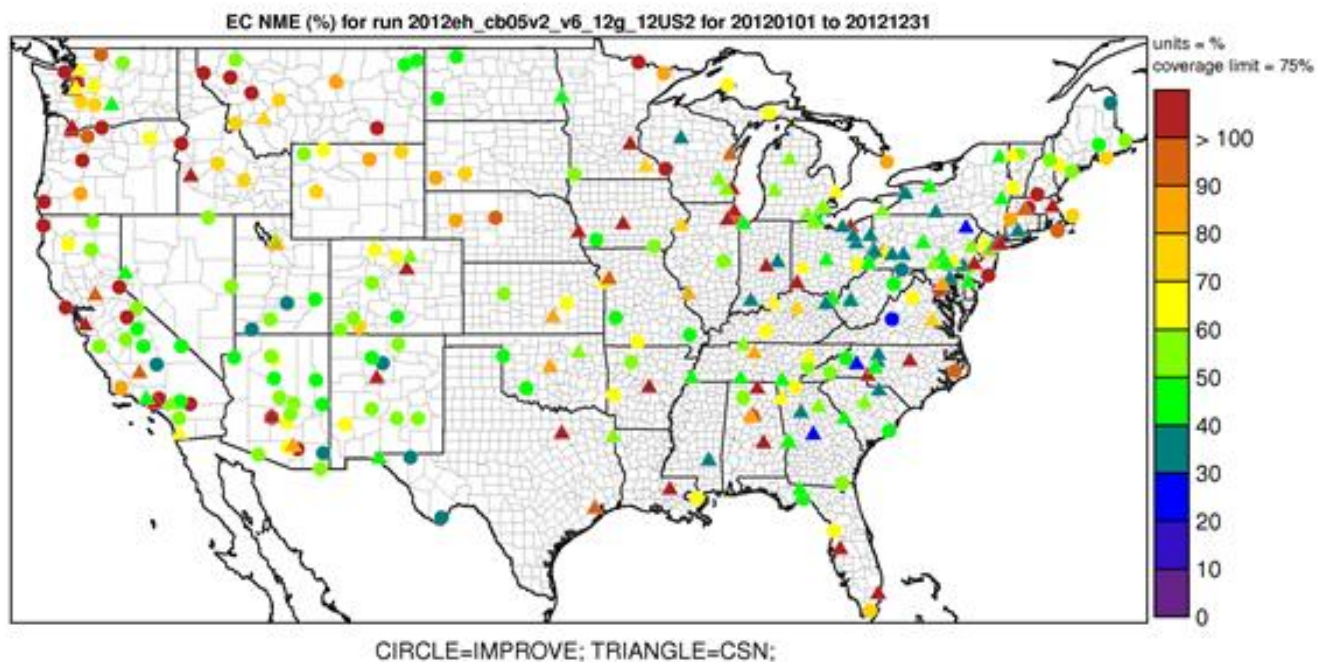


Figure 4-25. Normalized Mean Error (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

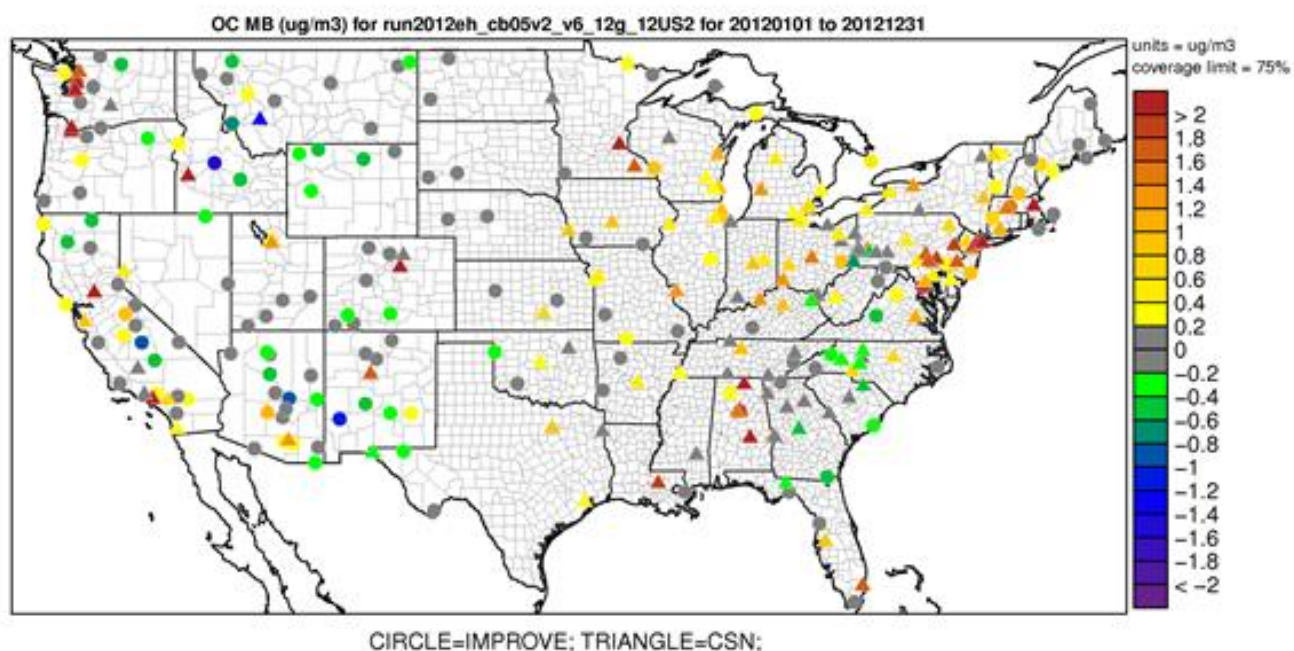


Figure 4-26. Mean Bias ($\mu\text{g}/\text{m}^3$) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

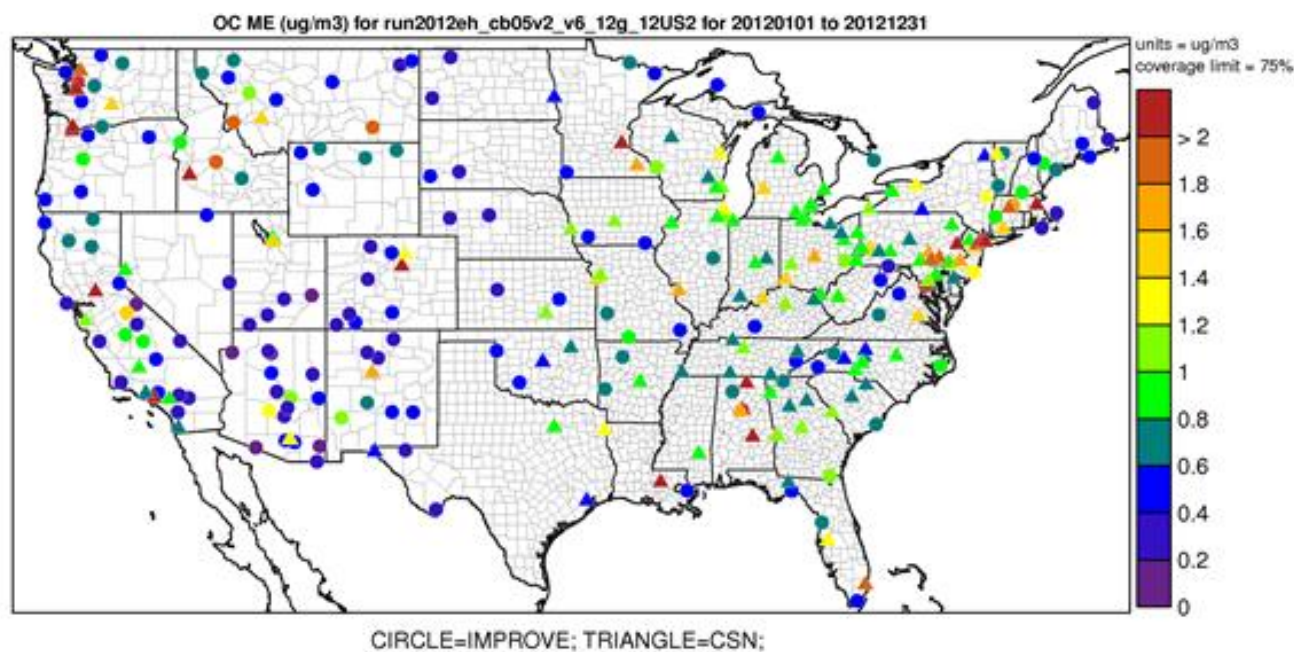


Figure 4-27. Mean Error ($\mu\text{g}/\text{m}^3$) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

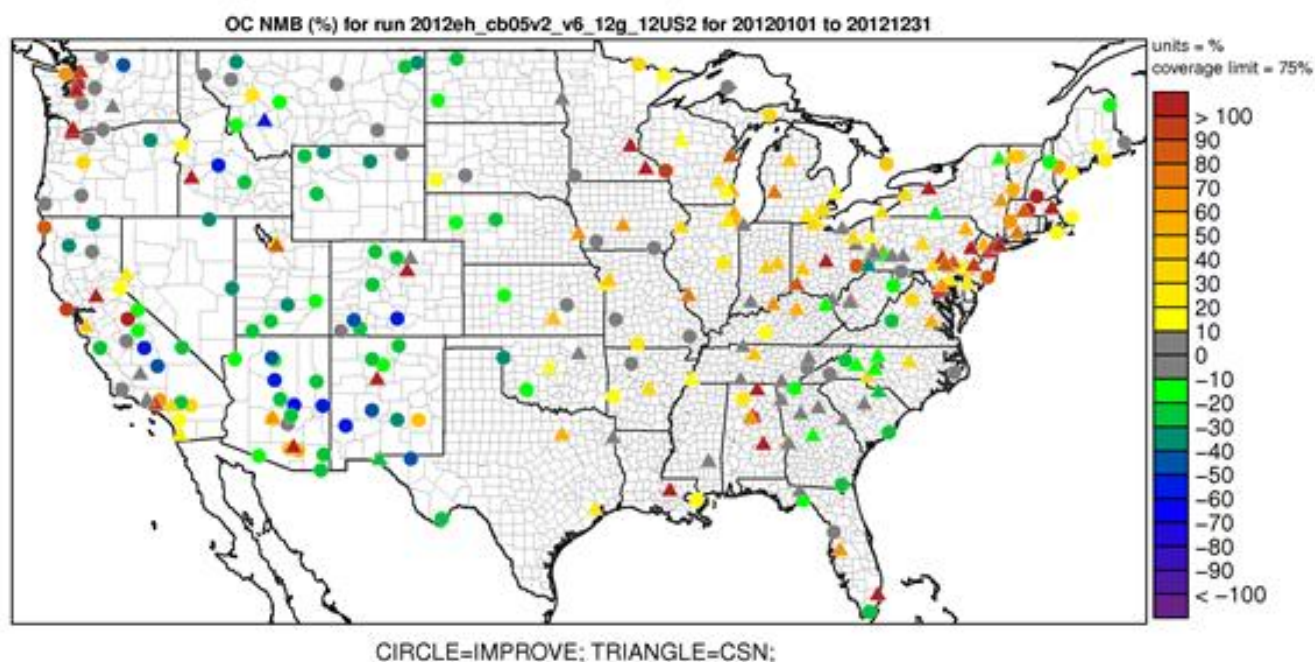


Figure 4-28. Normalized Mean Bias (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

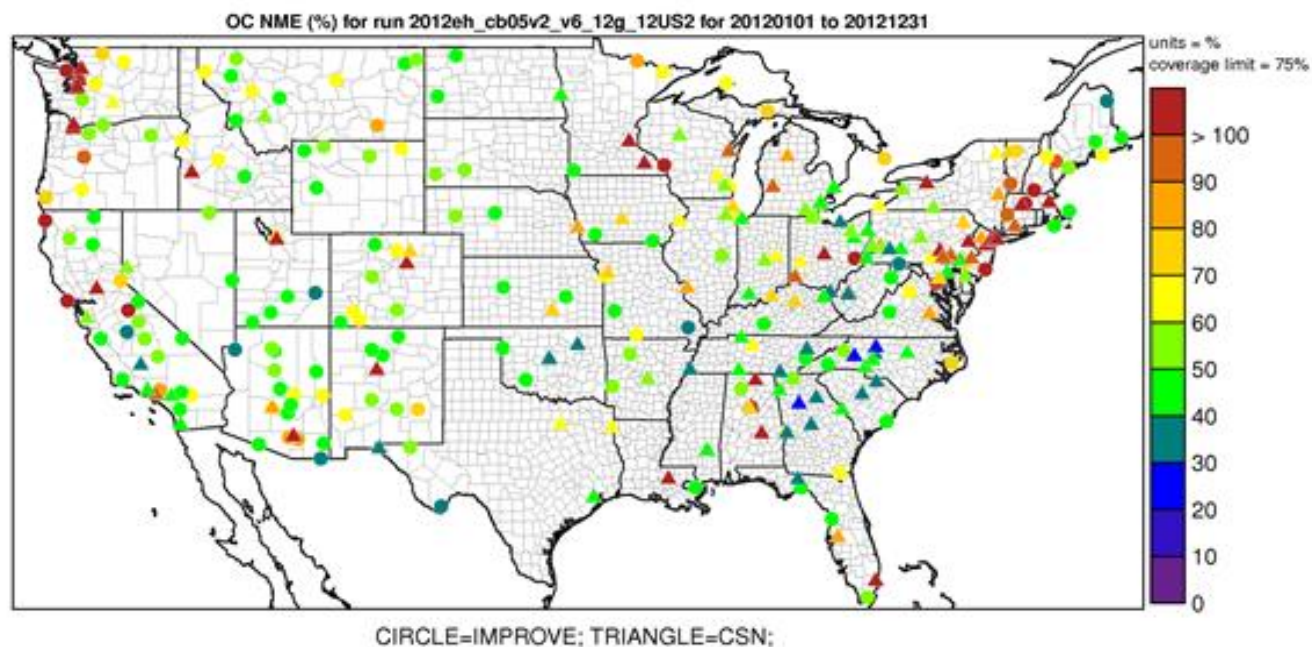


Figure 4-29. Normalized Mean Bias (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

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

5.1 Introduction

The need for greater spatial coverage of air pollution concentration estimates has grown in recent years as epidemiology and exposure studies that link air pollution concentrations to health effects have become more robust and as regulatory needs have increased. Direct measurement of concentrations is the ideal way of generating such data, but prohibitive logistics and costs limit the possible spatial coverage and temporal resolution of such a database. Numerical methods that extend the spatial coverage of existing air pollution networks with a high degree of confidence are thus a topic of current investigation by researchers. The downscaler model (DS) is the result of the latest research efforts by EPA for performing such predictions. DS utilizes both monitoring and CMAQ data as inputs, and attempts to take advantage of the measurement data's accuracy and CMAQ's spatial coverage to produce new spatial predictions. This chapter describes methods and results of the DS application that accompany this report, which utilized ozone and PM_{2.5} data from AQS and CMAQ to produce predictions to continental U.S. 2010 census tract centroids for the year 2012.

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:

$$Y(s, t) = \sim\beta_0(s, t) + \beta_1(s, t) * \sim x(s, t) + \varepsilon(s, t) \quad (\text{Equation 1})$$

Where:

$Y(s, t)$ is the observed concentration at point s and time t .

$\sim x(s, t)$ is the CMAQ concentration at time t . This value is a weighted average of both the gridcell containing the monitor and neighboring gridcells.

$\sim\beta_0(s, t)$ is the intercept, and is composed of both a global and a local component.

$\beta_1(t)$ is the global slope; local components of the slope are contained in the $\sim x(s, t)$ term.

$\varepsilon(s, t)$ 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 $\sim\beta_0$ 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 "heirarchical" in structure, meaning that the top level parameters in Equation 1 (ie $\sim\beta_0(s,t)$, $\beta_I(t)$, $\sim x(s,t)$) 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 $\varepsilon(s,t)$).

Further information about the development and inner workings of the current version of DS can be found in Berrocal, Gelfand and Holland (2011) 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 2012 US census tract centroids across the continental U.S. using 2012 measurement and CMAQ data as inputs. For ozone, the concentration unit is the daily maximum 8-hour average in ppb and for PM_{2.5} the concentration unit is the 24-hour average in $\mu\text{g}/\text{m}^3$.

5.3.1 *Summary of 8-hour Ozone Results*

Figure 5-1 summarizes the AQS, CMAQ and DS ozone data over the year 2012. 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 2012, about 75% of the US Census tracts (54633 out of 72283) experienced at least one day with an ozone value above the NAAQS of 75 ppb.

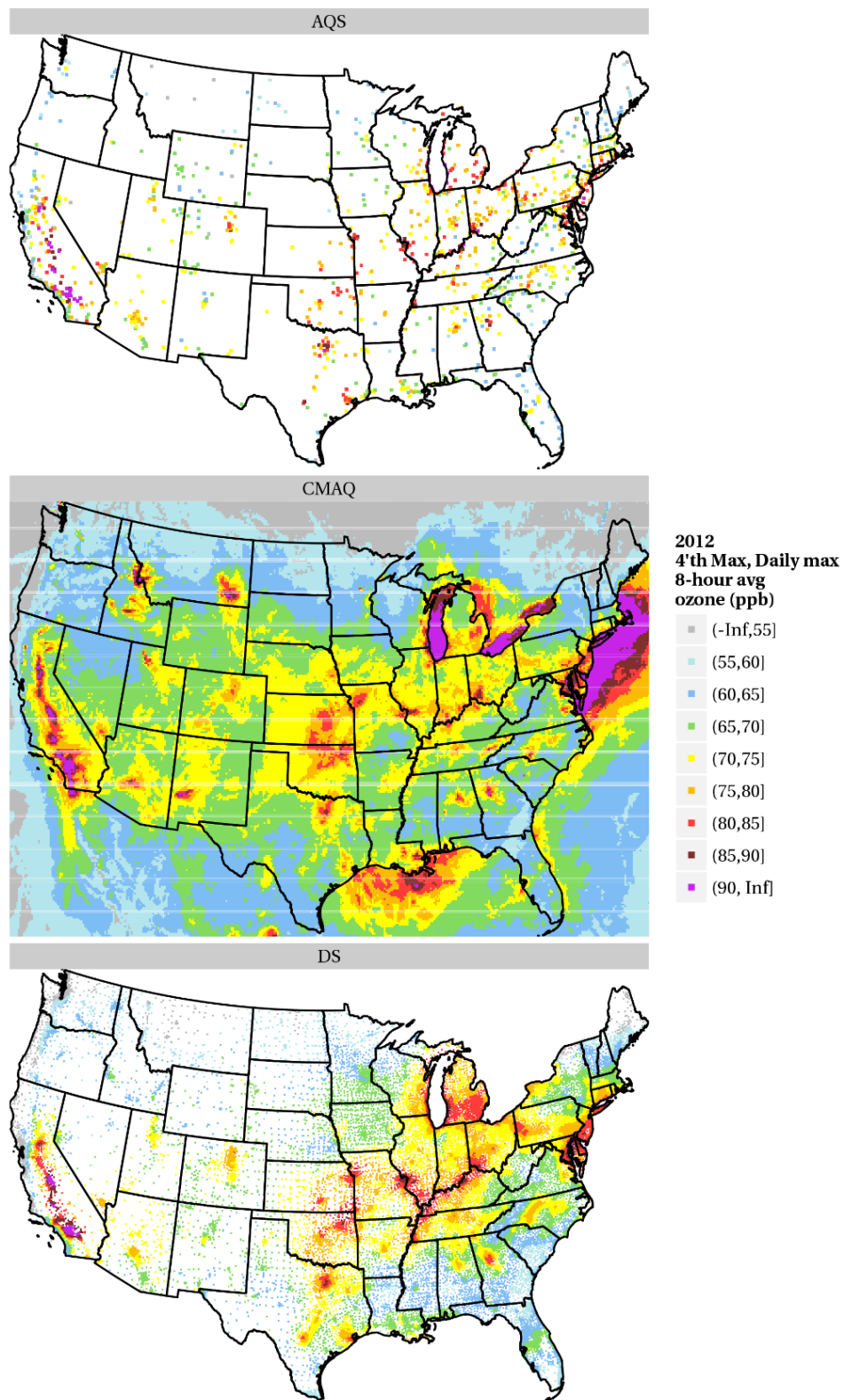


Figure 5-2. Annual 4th max (daily max 8-hour ozone concentrations) derived from AQS, CMAQ and DS data.

5.3.2 Summary of PM_{2.5} Results

Figures 5-2 and 5-3 summarize the AQS, CMAQ and DS PM_{2.5} data over the year 2012. Figure 5-2 shows annual means and Figure 5-3 shows 98'th percentiles of 24-hour PM_{2.5} concentrations for AQS observations, CMAQ model predictions and DS model results. The DS model estimated that for 2012 about 39% of the US Census tracts (28402 out of 72283) experienced at least one day with a PM_{2.5} value above the 24-hour NAAQS of 35 ug/m³.

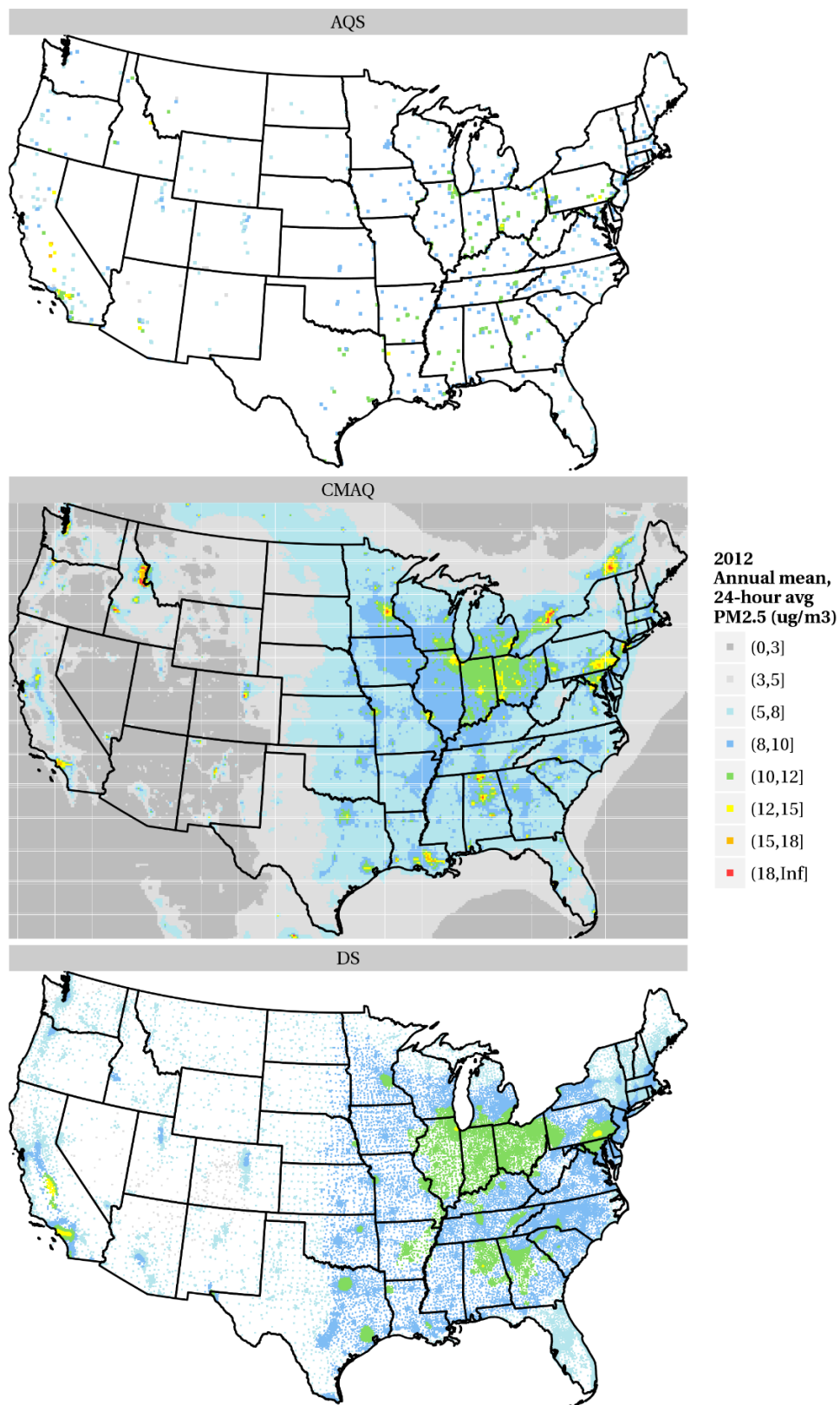


Figure 5-2. Annual mean PM_{2.5} concentrations derived from AQS, CMAQ and DS data.

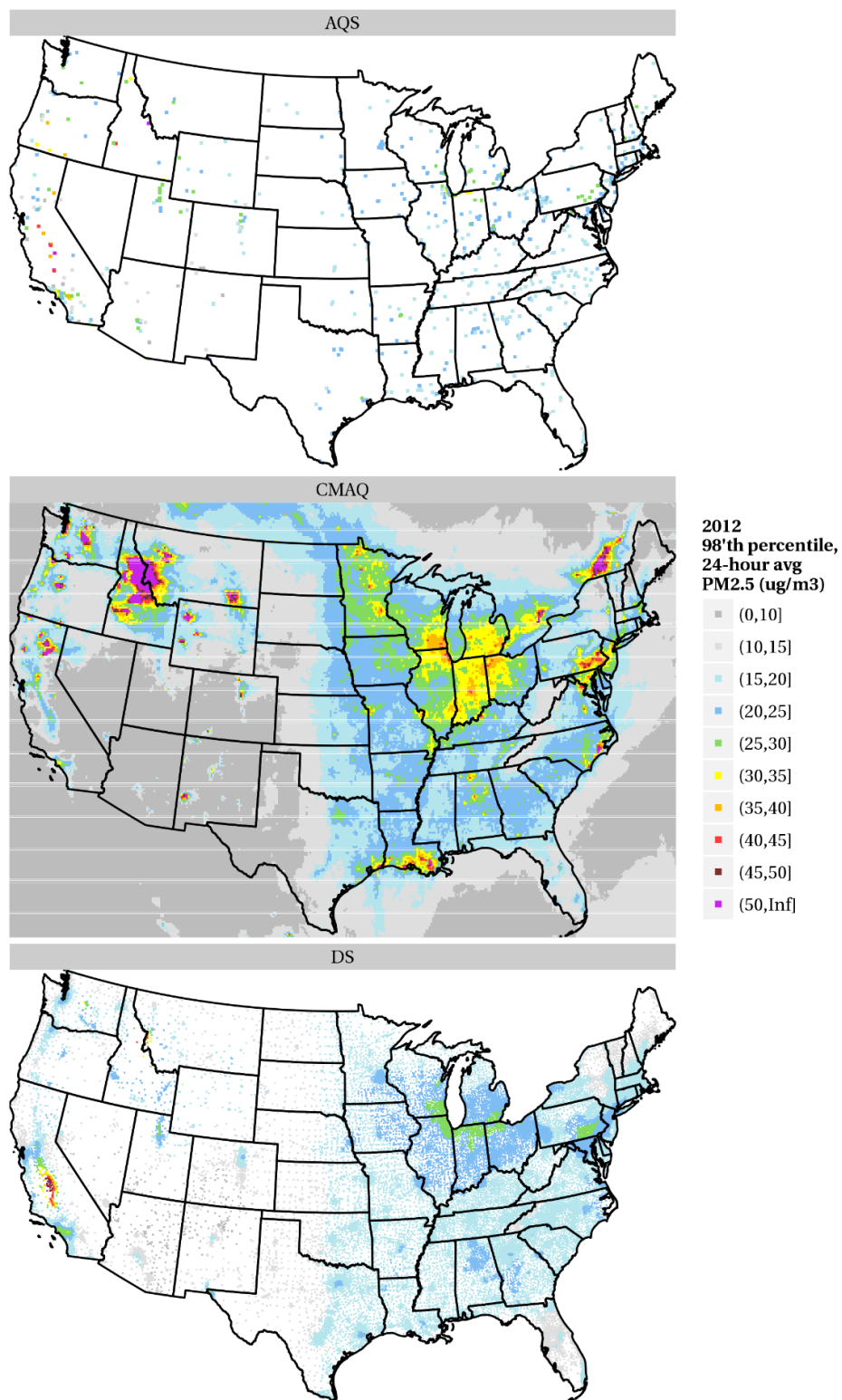


Figure 5-3. 98th percentile 24-hour average PM_{2.5} concentrations derived from AQS, CMAQ and DS data.

5.4 Downscaler Uncertainties

5.4.1 *Standard Errors*

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

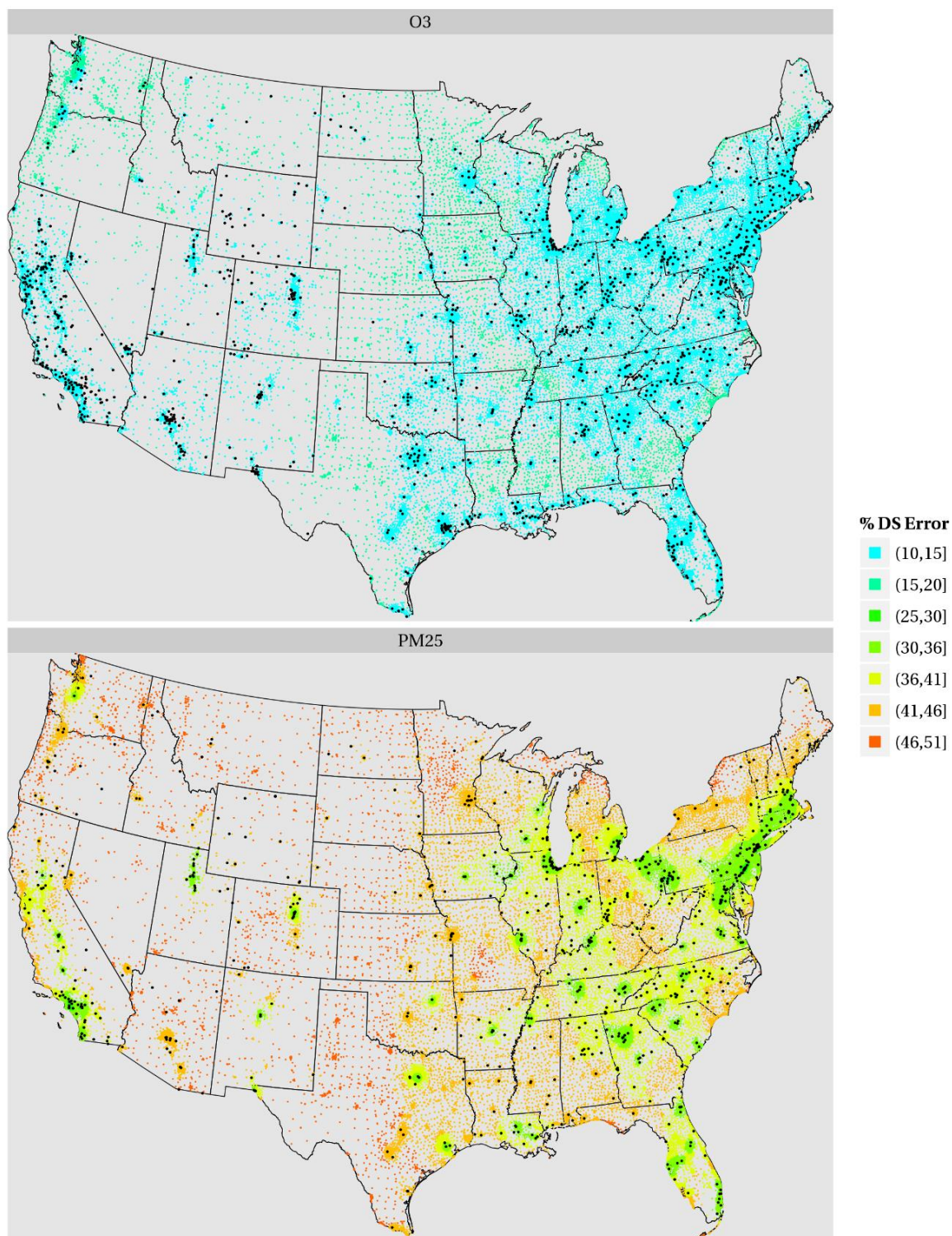


Figure 5-4. Annual mean relative errors (standard errors divided by predictions) from the DS 2012 runs. The black dots show the locations of monitors that generated the AQS data used as input to the DS model.

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}	815	0.17	2.93	0.96
O ₃	1311	-0.01	4.58	0.95

Table 5-1. Cross-validation statistics associated with the 2012 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 2012 U.S. census tract centroids by utilizing monitoring data and CMAQ output for 2012. Large-scale spatial and temporal patterns of concentration predictions are generally consistent with those seen in ambient monitoring data. Both ozone and PM_{2.5} were predicted with lower error in the eastern versus the western U.S., presumably due to the greater monitoring density in the east.

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

Appendix A - Acronyms

Acronyms

ARW	Advanced Research WRF core model
BEIS	Biogenic Emissions Inventory System
BlueSky	Emissions modeling framework
CAIR	Clean Air Interstate Rule
CAMD	EPA's Clean Air Markets Division
CAP	Criteria Air Pollutant
CAR	Conditional Auto Regressive spatial covariance structure (model)
CARB	California Air Resources Board
CEM	Continuous Emissions Monitoring
CHIEF	Clearinghouse for Inventories and Emissions Factors
CMAQ	Community Multiscale Air Quality model
CMV	Commercial marine vessel
CO	Carbon monoxide
CSN	Chemical Speciation Network
DQO	Data Quality Objectives
EGU	Electric Generating Units
Emission Inventory	Listing of elements contributing to atmospheric release of pollutant substances
EPA	Environmental Protection Agency
EMFAC	Emission Factor (California's onroad mobile model)
FAA	Federal Aviation Administration
FDDA	Four Dimensional Data Assimilation
FIPS	Federal Information Processing Standards
HAP	Hazardous Air Pollutant
HMS	Hazard Mapping System
ICS-209	Incident Status Summary form
IPM	Integrated Planning Model
ITN	Itinerant
LSM	Land Surface Model
MOBILE	OTAQ's model for estimation of onroad mobile emissions factors
MODIS	Moderate Resolution Imaging Spectroradiometer
MOVES	Motor Vehicle Emission Simulator
NEEDS	National Electric Energy Database System
NEI	National Emission Inventory
NERL	National Exposure Research Laboratory
NESHAP	National Emission Standards for Hazardous Air Pollutants
NH	Ammonia
NMIM	National Mobile Inventory Model
NONROAD	OTAQ's model for estimation of nonroad mobile emissions
NO	Nitrogen oxides
OAQPS	EPA's Office of Air Quality Planning and Standards
OAR	EPA's Office of Air and Radiation
ORD	EPA's Office of Research and Development
ORIS	Office of Regulatory Information Systems (code) - is a 4 or 5 digit

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

United States	Office of Air Quality Planning and	Publication No. EPA-
Environmental Protection	Standards	450R16001
Agency	Air Quality Assessment Division	March 2016
	Research Triangle Park, NC	
