Bayesian space-time downscaling fusion model (downscaler) -Derived Estimates of Air Quality for 2011

Bayesian space-time downscaling fusion model (downscaler) -Derived Estimates of Air Quality for 2011

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

Contents

1.0   Introduction

This report describes estimates of daily ozone (maximum 8-hour average) and PM2.5 (24-hour average) concentrations throughout the contiguous United States during the 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 PM2.5.

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

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

---

[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.

a multidisciplinary collaboration that involves the ongoing collection, integration, analysis, interpretation, and dissemination of data from: environmental hazard monitoring activities; human exposure assessment information; and surveillance of noninfectious health conditions. As part of the National EPHT Program efforts, the CDC led the initiative to build the National EPHT Network (http:// www.cdc.gov/nceh/tracking/default.htm). The National EPHT Program, with the EPHT Network as its cornerstone, is the CDC's response to requests calling for improved understanding of how the environment affects human health. The EPHT Network is designed to provide the means to identify, access, and organize hazard, exposure, and health data from a variety of sources and to examine, analyze and interpret those data based on their spatial and temporal characteristics.

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

- Collecting, analyzing and interpreting environmental and health data from both agencies (HHS and EPA).

- Collaborating on emerging information technology practices related to building, supporting, and operating the CDC EPHT Network and the Environmental Information Exchange Network.

- Developing and validating additional environmental public health indicators.

- Sharing reliable environmental and public health data between their respective networks in an efficient and effective manner.

- Consulting and informing each other about dissemination of results obtained through work carried out under the MOU and the associated Interagency Agreement (IAG) between EPA and CDC.

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

---

[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.

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

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

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

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

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

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

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

2.0   Air Quality Data

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

2.1   Introduction to Air Quality Impacts in the United States

2.1.1   The Clean Air Act

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

Under the CAA, the U.S. EPA has established standards or limits for six air pollutants, known as the criteria air pollutants: carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and particulate matter (PM). These standards, called the National Ambient Air Quality Standards (NAAQS), are designed to protect public health and the environment. The CAA established two types of air quality standards. Primary standards set limits to protect public health, including the health of "sensitive" populations such as asthmatics, children, and the elderly. Secondary standards set limits to protect public welfare, including protection against decreased visibility, damage to animals, crops, vegetation, and buildings. The law requires EPA to review periodically these standards. For more specific information on the NAAQS, go to www.epa.gov/air/criteria.html. For general information on the criteria pollutants, go to http://www.epa.gov/air/urbanair/6poll.html.

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 http://www.epa.gov/ozonedesignations/ and http://www.epa.gov/pmdesignations.

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 http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_index.html. The current NAAQS for ozone (last revised in 2008) is a daily maximum 8-hour average of 0.075 parts per million [ppm] (for details, see http://www.epa.gov/air/criteria.html. 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.

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 sources3. Particles less than 10 micrometers in diameter (PM_{10}) pose a health concern because they can be inhaled into and accumulate in the respiratory system. Particles less than 2.5 micrometers in diameter ($PM_{2.5}$) are referred to as "fine" particles. Because of their small size, fine particles can lodge deeply into the lungs. Sources of fine particles include all types of combustion (motor vehicles, power plants, wood burning, etc.) and some industrial processes. Particles with diameters between 2.5 and 10 micrometers ($PM_{10-2.5}$) are referred to as "coarse" or PMc. Sources of PMc include crushing or grinding operations and dust from paved or unpaved roads. The distribution of PM_{10}, $PM_{2.5}$ and PMc varies from the Eastern U.S. to arid western areas.

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

---

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

The current NAAQS for PM$_{2.5}$ (last revised in 2012) includes both a 24-hour standard to protect against short-term effects, and an annual standard to protect against long-term effects. The annual average PM$_{2.5}$ concentration must not exceed 12.0 micrograms per cubic meter (ug/m^3) based on the annual mean concentration averaged over three years, and the 24-hr average concentration must not exceed 35 ug/m^3 based on the 98th percentile 24-hour average concentration averaged over three years. More information is available at http://www.epa.gov/air/criteria.html and http://www.epa.gov/oar/particlepollution/. The standards for PM$_{2.5}$ values are shown in Table 2-1.

Table 2-1. PM$_{2.5}$ Standards

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

2.2   Ambient Air Quality Monitoring in the United States

2.2.1   Monitoring Networks

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

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

Distance to the Nearest Ozone Monitor
- 41 - 10,000 meters
- 10,001 - 25,000 meters
- 25,001 - 50,000 meters
- 50,001 - 75,000 meters
- 75,001 - 100,000 meters
- 100,001 - 150,000 meters
- 150,001 - 333,252 meters

Distance to the Nearest PM2.5 Monitor
- 41 - 10,000 meters
- 10,001 - 25,000 meters
- 25,001 - 50,000 meters
- 50,001 - 75,000 meters
- 75,001 - 100,000 meters
- 100,001 - 150,000 meters
- 150,001 - 333,252 meters
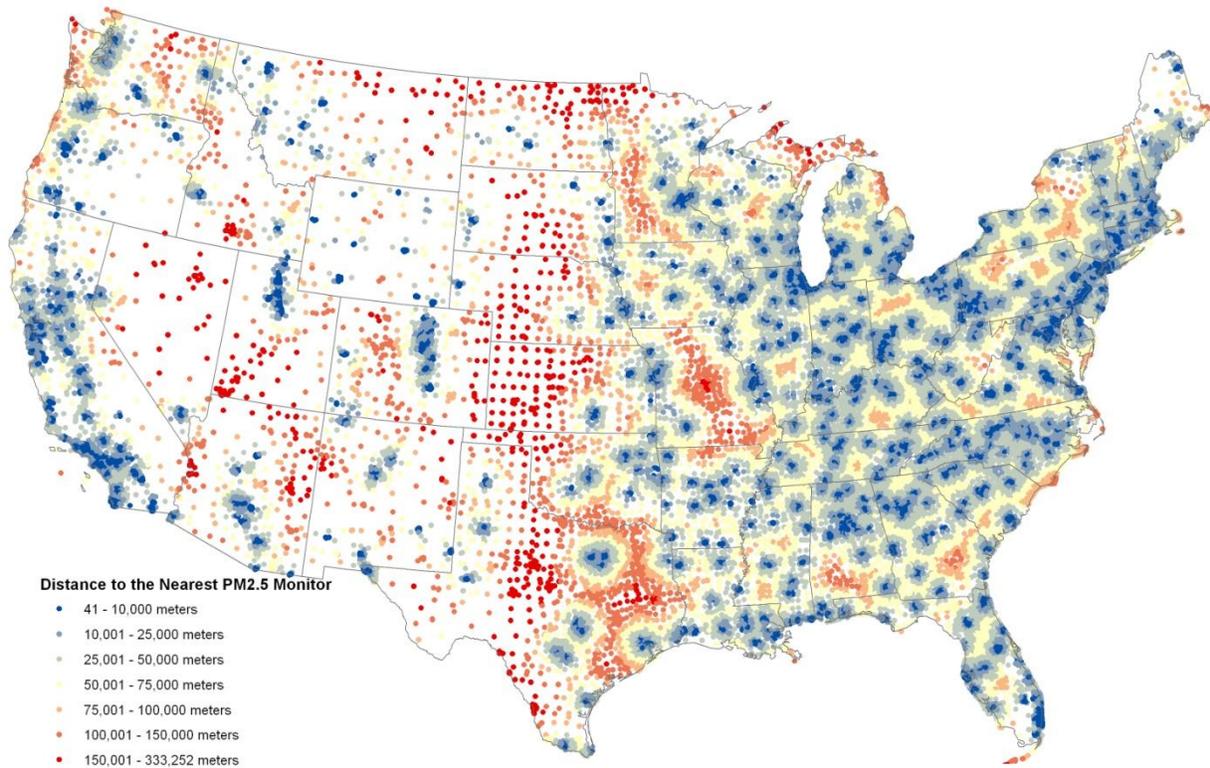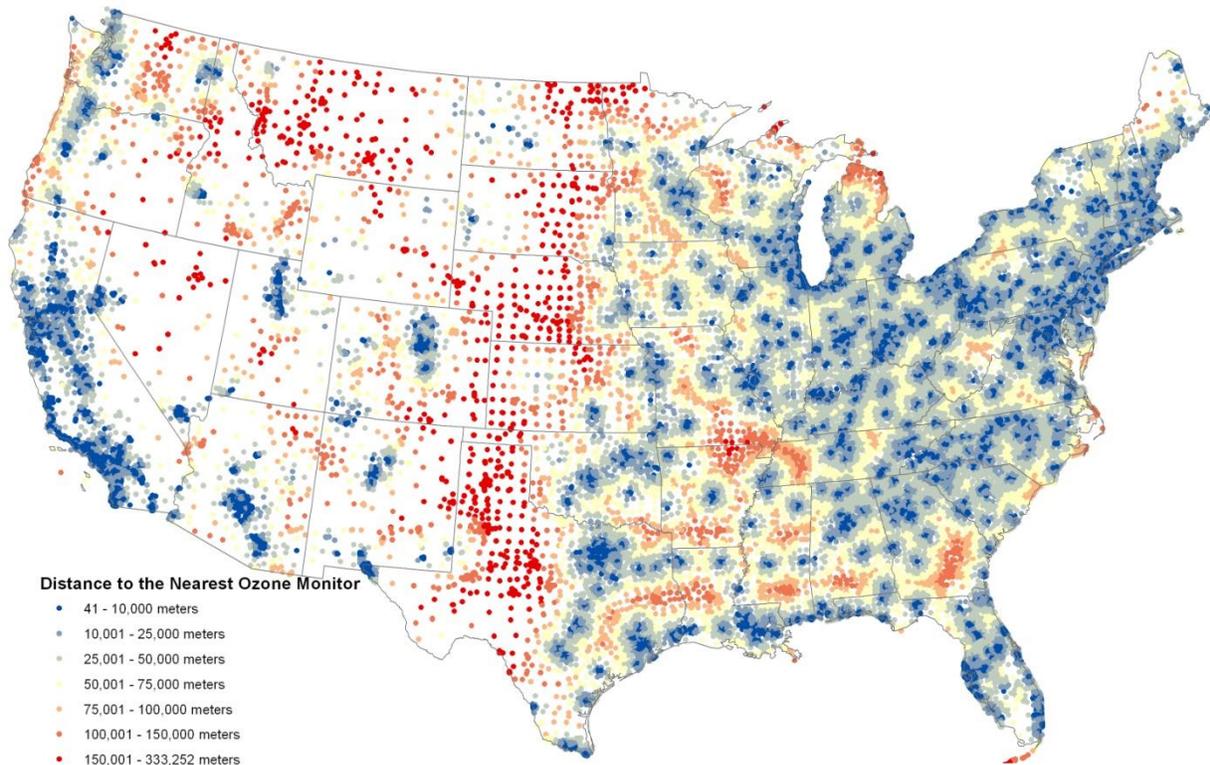
Figure 2-1. Distances from US Census Tract centroids to the nearest monitoring site, 2009.

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 www.epa.gov/airtrends).

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 http://www.epa.gov/ttn/airs/airsaqs/index.htm.

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

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

The advantage of using ambient data from EPA monitoring networks for comparing with health outcomes is that these measurements of pollution concentrations are the best characterization of the concentration of a given pollutant at a given time and location. Furthermore, the data are supported by a comprehensive quality assurance program, ensuring data of known quality. One disadvantage of using the ambient data is that it is usually out of spatial and temporal alignment with health outcomes. This spatial and temporal 'misalignment' between air quality monitoring data and health outcomes is influenced by the following key factors: the living and/or working locations (microenvironments) where a person spends their time not being co-located with an air quality monitor; time(s)/date(s) when a patient experiences a health outcome/symptom (e.g., asthma attack) not coinciding with time(s)/date(s) when an air quality monitor records ambient concentrations of a pollutant high enough to affect the symptom (e.g., asthma attack either during or shortly after a high $PM_{2.5}$ day). To compare/correlate ambient concentrations with acute health effects, daily local air quality data is needed[4]. Spatial gaps exist in the air quality monitoring network, especially in rural areas, since the air quality monitoring network is designed to focus on measurement of pollutant concentrations in high population density areas. Temporal limits also exist. Hourly ozone measurements are aggregated to daily values (the daily max 8-hour average is relevant to the ozone standard). Ozone is typically monitored during the ozone season (the warmer months, approximately April through October). However, year-long

---

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

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, 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 prune to systematic biases since they depend upon so many variables (i.e., metrological models and emission models) and complex chemical and physical process simulations.

Combining monitoring data with air quality models (via fusion or regression) may provide the best results in terms of estimating ambient air concentrations in space and time.   EPA's eVNA[5] is an example of an earlier approach for merging air quality monitor data with CMAQ model predictions. 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.

---

[5] eVNA is described in the "Regulatory Impact Analysis for the Final Clean Air Interstate Rule", EPA-452/R-05-002, March 2005, http://www.epa.gov/cair/pdfs/finaltech08.pdf, Appendix F.

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

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

2.3.1 Rationale for the Air Quality Indicators

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

2.3.2 Air Quality Data Sources

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

2.3.3 Use of Air Quality Indicators for Public Health Practice

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

3.0   Emissions Data

3.1   Introduction to Emissions Data Development

The U.S. Environmental Protection Agency (EPA) developed an air quality modeling platform based primarily on the 2011 National Emissions Inventory (NEI), Version 1 to process year 2011 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/) is used to model ozone (O_3) and particulate matter (PM) for this project. CMAQ requires hourly and gridded emissions of the following inventory pollutants: carbon monoxide (CO),nitrogen oxides (NO_X), volatile organic compounds (VOC), sulfur dioxide (SO_2), ammonia (NH_3), particulate matter less than or equal to 10 microns (PM_{10}), and individual component species for particulate matter less than or equal to 2.5 microns ($PM_{2.5}$). In addition, the CMAQ CB05 with chlorine chemistry used here 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 2011 emission inputs for this study included development of emission inventories for a 2011 model evaluation case, and application of emissions modeling tools to convert the inventories into the format and resolution needed by CMAQ. An evaluation case uses year-specific fire and continuous emission monitoring (CEM) data for electric generating units (EGUs), whereas other types of modeling cases can use averages for these sources. 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 3.5.1 was used to create emissions files for a 12-km national grid. Additional information about SMOKE is available from http://www.smoke-model.org.

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   2011 Emission Inventories and Approaches

This section describes the emissions inventories created for input to SMOKE. The 2011 NEI is the primary basis for the inputs to SMOKE and 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

2011 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 preveny duplication between point and nonpoint source categories such as industrial boilers. Documentation for the 2011 NEI is available at http://www.epa.gov/ttn/chief/net/2011inventory.html#inventorydoc.

EPA used the SMARTFIRE2 system to develop 2011 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 in the 2011NEIv1 were developed using MOVES2010b; however, the 2011 emissions modeling platform used a different version of MOVES, hence forth referred to as "MOVESTier3FRM" that facilitated the representation of the final Tier 3 standards in future years. When given the same inputs, these two versions of MOVES produce similar emissions estimates for the year 2011. Canadian emissions reflect year 2006, as those were the latest available at the time of the modeling. Mexican emissions reflect year 2012 as projected from their 1999 inventory, and offshore emissions reflect year 2008 because 2011 data were not yet available at the time of the modeling.

The methods used to process emissions for this project are very similar to those documented for EPA's Version 6.1, 2011 Emissions Modeling Platform that was also used for the proposed Ozone National Ambient Air Quality Standards (NAAQS). A technical support document (TSD) for this platform is available at EPA's emissions modeling clearinghouse (EMCH): http://www.epa.gov/ttn/chief/emch/index.html#2011 (EPA, 2014a) and includes additional details regarding some aspects of the data preparation and emissions modeling. Electronic copies of the main inventories and ancillary data used for this project are available from the version 6.1 section of the EMCH.

The emissions modeling process, performed using SMOKE v3.5.1 apportions the emissions inventories into the grid cells used by CMAQ and temporalizes the emissions into hourly values. In addition, the pollutants in the inventories (e.g., NOx 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 2011 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 2011 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_2011_emissions_totals_by_sector.xlsx".

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

2011 Platform Sector (Abbrev)	2011 NEI Sector	Description and resolution of the data input to SMOKE
EGU non-peaking units (*ptegu*)	Point	2011 NEI point source EGUs determined to operate as non-peaking units. The 2011NEIv1 emissions are replaced with hourly 2011 CEMS values for NOX and SO2, where the units are matched to the NEI. Annual resolution for non-CEMS sources, hourly for sources matched to CEMS.
EGU peaking units (*ptegu_pk*)	Point	Same as ptegu sector, but limited to EGUs determined to operate as peaking units. All sources in this sector have CEMS data for 2011 and are therefore hourly.
Point source oil and gas (*pt_oilgas*)	Point	2011NEIv1 point sources with oil and gas production emissions processes.
Remaining non-EGU point (*ptnonipm*)	Point	All 2011NEIv1 point source records not matched to the ptegu, ptegu_pk, and pt_oilgas sectors, except for offshore point sources that are in the othpt sector. Includes all aircraft emissions and some rail yard emissions. Annual resolution.
Point source fire (*ptfire*)	Fires	Point source day-specific wildfires and prescribed fires for 2011 computed using SMARTFIRE 2, except for Georgia-submitted emissions. Consistent with 2011NEIv1.
Agricultural (*ag*)	Nonpoint	NH3 emissions from 2011NEIv1 nonpoint livestock and fertilizer application, county and annual resolution.
Area fugitive dust (*afdust*)	Nonpoint	PM10 and PM2.5 from fugitive dust sources in the 2011NEIv1 nonpoint inventory, including building construction, road construction, agricultural dust, and road dust; however, unpaved and paved road dust emissions differ from the NEI in that they do not have a precipitation adjustment. Instead, the emissions modeling adjustment applies a transport fraction and a meteorology-based (precipitation and snow/ice cover) zero-out. County and annual resolution.
Nonpoint source oil and gas (*np_oilgas*)	Nonpoint	2011NEIv1 nonpoint sources from oil and gas-related processes. County and annual resolution.
Residential Wood Combustion (*rwc*)	Nonpoint	2011NEIv1 NEI nonpoint sources with Residential Wood Combustion (RWC) processes. County and annual resolution.
Remaining nonpoint (*nonpt*)	Nonpoint	2011NEIv1 nonpoint sources not included in other platform sectors; county and annual resolution.
C3 commercial marine (*c3marine*)	Nonpoint	Category 3 (C3) CMV emissions projected to 2011 from year 2002 values. These emissions are not from the NEI, but rather were developed for the rule called "Control of Emissions from New Marine

		Compression-Ignition Engines at or Above 30 Liters per Cylinder", usually described as the Emissions Control Area- International Maritime Organization (ECA-IMO) study: http://www.epa.gov/otaq/oceanvessels.htm. (EPA-420-F-10-041, August 2010). U.S. states-only emissions (zero in Midwest); see othpt sector for all non-U.S. emissions. Treated as point sources to reflect shipping lanes, annual resolution.
C1 and C2 marine and locomotive (*c1c2rail*)	Nonpoint	Locomotives and primarily category 1 (C1) and category 2 (C2) commercial marine vessel (CMV) emissions sources from the 2011NEIv1 nonpoint inventory. Midwestern states' CMV emissions, including Class 3 sources, are from a separate year 2010 emissions inventory. County and annual resolution.
Nonroad (*nonroad*)	Nonroad	2011NEIv1 nonroad equipment emissions developed with the National Mobile Inventory Model (NMIM) using NONROAD2008 version NR08a. NMIM was used for all states except California and Texas, which submitted their own emissions to the 2011NEIv1. County and monthly resolution.
Onroad non-refueling (*onroad*)	Onroad	2011 onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles. Includes the following modes: exhaust, extended idle, evaporative, permeation, and brake and tire wear. For all states except California and Texas, based on monthly MOVES emissions tables from MOVESTier3FRM. Texas emissions are from the 2011NEIv1 and are based on MOVES 2010b, and California emissions are based on Emission Factor (EMFAC). MOVES-based emissions computed for each hour and model grid cell using monthly and annual activity data (e.g., VMT, vehicle population).
Onroad Refueling (*onroad_rfl)*	Onroad	2011NEIv1 onroad mobile gasoline and diesel vehicle refueling emissions for all states. Based on MOVESTier3FRM emissions tables. Computed hourly based on temperature and for each model grid cell.
Biogenic (*beis*)	Biogenic	Hour- and grid cell-specific emissions for 2011 generated from the BEIS 3.14 model, including emissions in Canada and Mexico.
Other point sources not from the NEI (*othpt*)	N/A	Point sources from Canada's 2006 inventory and Mexico's 2012 inventory grown from year 1999 (ERG, 2009; Wolf, 2009). Also includes all non-U.S. C3 CMV and U.S. offshore oil production, which are unchanged from the 2008 NEI point source annual emissions.

Other nonpoint and nonroad (*othar*)	N/A	Annual year 2006 Canada (province resolution) and year 2012 (grown from 1999) Mexico Phase III (municipio resolution) nonpoint and nonroad mobile inventories.
Other onroad sources (*othon*)	N/A	Year 2006 Canada (province resolution) and year 2012 (grown from 1999) Mexico Phase III (municipio resolution) onroad mobile inventories, annual resolution.

Table 3-2. 2011 Continental United States Emissions by Sector (tons/yr in 48 states + D.C.)

Sector	CO	NH₃	NOₓ	PM₁₀	PM₂.₅	SO₂	VOC
afdust				18,502,317	2,487,403		
ag		3,517,371					
c1c2rail	173,437	481	1,046,095	34,670	32,367	17,651	47,714
nonpt	3,046,375	142,323	832,166	715,709	533,248	392,638	3,792,612
np_oilgas	642,182	0	653,219	21,756	17,200	17,195	2,273,214
nonroad	13,952,389	2,627	1,630,409	162,420	154,660	4,031	2,024,633
onroad_adj	25,230,444	118,130	5,591,695	287,540	207,517	28,475	2,576,504
onroad_rfl							161,415
c3marine	12,425		124,725	4,279	3,909	38,645	4,954
ptfire	22,580,113	362,910	347,103	2,362,132	2,005,142	177,107	5,174,593
ptegu	719,414	21,644	1,925,742	259,011	188,811	4,596,656	32,288
ptegu_pk	8,662	425	21,941	2,159	1,886	28,476	783
ptnonipm	2,565,936	74,841	1,767,748	491,837	338,447	1,071,950	872,433
pt_oilgas	20,579	112	17,026	1,833	1,810	55,142	87,842
rwc	2,578,229	20,343	35,672	389,019	388,288	8,986	446,972
Con.US Total	**71,530,185**	**4,261,207**	**13,993,540**	**23,234,681**	**6,360,688**	**6,436,952**	**17,495,956**

Table 3-3. 2011 Non-US Emissions by Sector within Modeling Domain (tons/yr for Canada, Mexico, Offshore)

Sector	CO	NH₃	NOₓ	PM₁₀	PM₂.₅	SO₂	VOC
Canada othar	2,810,350	386,147	462,996	810,747	248,907	61,179	932,322
Canada othon	3,303,239	17,572	392,209	11,075	7,712	4,046	199,939
Canada othpt	560,661	15,543	369,993	65,782	39,828	825,675	157,170
Canada Subtotal	**6,674,250**	**419,263**	**1,225,198**	**887,604**	**296,447**	**890,900**	**1,289,431**
Mexico othar	439,901	109,861	189,592	69,523	45,923	26,559	499,145
Mexico othon	423,978	3,247	76,880	7,593	6,970	1,413	73,888
Mexico othpt	116,609	0	414,399	137,512	101,884	828,418	83,838
Mexico Subtotal	**980,488**	**113,108**	**680,871**	**214,628**	**154,778**	**856,390**	**656,872**
Offshore to EEZ	130,419	0	610,644	16,961	15,525	133,606	81,286
Non-US c3marine	17,168	0	202,516	17,199	15,823	127,563	7,297
Offshore Subtotal	**147,587**	**0**	**813,159**	**34,160**	**31,348**	**261,168**	**88,583**
2011 TOTAL	**7,802,325**	**532,371**	**2,719,229**	**1,136,392**	**482,573**	**2,008,459**	**2,034,886**

3.2.1 *Point Sources (ptipm 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 (c3marine and othpt sectors) are processed by SMOKE as point source inventories and are discussed later in this section. Full documentation for the development of the 2011 NEI (EPA, 2014b), is posted at:
http://www.epa.gov/ttn/chief/net/2011inventory.html#inventorydoc.

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 four platform point source sectors were created from the remaining 2011NEIv1 point sources. The point sectors are: the EGU sector for non-peaking units (ptegu), the EGU sector for peaking units (ptegu_pk), point source oil and gas extraction -related emissions (pt_oilgas) and the remaining non-EGU sector also called the non-IPM (ptnonipm) sector. 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 EGU sectors are further split into "peaking" (ptegu_pk) and non-peaking units to allow for better analysis of the impact of peaking units. 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, NOX, VOC, SO2, NH3, PM10, and PM2.5 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 2011NEIv1 sources in the ptegu and ptegu_pk sectors that could be matched to 2011 CEMS data, hourly CEMS NO_X and SO_2 emissions for 2011 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 Sources (ptnonipm) emissions were input to SMOKE as annual emissions. The full description of how the 2011 NEI emissions were developed is provided in the NEI documentation, but a brief summary of their development follows:

a. 2011 CAP and HAP data were provided by States, locals and tribes under the Consolidated Emissions Reporting Rule
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 Mineral Management Services (MMS), although in 2011NEIv1, these data were still from 2008 because the 2011 data were not yet available.

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 the3-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 an ancillary file, called the PSTK file, which provides default stack parameters by SCC code to either gap fill stack parameters if they are missing in the NEI or to correct stack parameters if they are outside the ranges specified.

3.2.1.1 EGU sector (ptegu)

The ptegu and ptegu_pk sectors contain emissions from EGUs in the 2011NEIv1 point inventory that could be matched to units found in the NEEDS v5.13 database (http://www.epa.gov/powersectormodeling/BaseCasev513.html). It was necessary to put these EGUs into separate sectors in the platform because IPM projects future emissions for the EGUs defined in the NEEDS database, and emissions for sources in the ptegu and ptegu_pk sectors are replaced with IPM outputs in the future year modeling case. 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 the future year if emissions for units are projected by IPM are not properly matched to the units in the NEI.

Some units in the ptegu and ptegu_pk sectors are matched to CEMS data via ORIS facility codes and boiler ID. For these units, SMOKE replaces the 2011 emissions of NOX and SO2 with the CEMS emissions, thereby ignoring the annual values specified in the NEI. For other pollutants, the hourly CEMS heat input data are used to allocate the NEI annual emissions to hourly values. All stack parameters, stack locations, and SCC codes for these sources come from the NEI. Because these attributes are obtained from the NEI, 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 NEI, the CEMS data for that unit is not used in the modeling platform. However, if the source exists in the NEI and is just not matched to a CEMS unit, the emissions from that source would be modeled using the annual emission value in the NEI 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 modeling platform.

For sources not matched to CEMS data (i.e., "non-CEMS" sources), daily emissions were computed from the NEI annual emissions using average CEMS data profiles specific to fuel type, pollutant, and IPM region. 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. For future-year scenarios, there are no CEMS data available for specific units, but the shape of the CEMS profiles is preserved for sources that are carried into the future year. This method keeps the temporal behavior of the base and future year cases as consistent as possible.

3.2.1.2 EGU Peaking Unit Sector (ptegu_pk)

The ptegu_pk sector includes sources identified by EPA as peaking units. The units were separated into this sector to facilitate analyses of the impact of peaking units. Aside from their inclusion in this sector, in all other ways they are treated in the same way as CEMS sources in the ptegu sector because all of them are matched to CEMS data. To identify units for inclusion in this sector, EPA required them to satisfy two tests: (1) the capacity factor was less than 10% over a 3 year average (2010-2012), and (2) the capacity factor was less than 20% in each of the 3 years. Here, "capacity factor" means either: (1) The ratio of a unit's actual annual electric output (expressed in MWe/hr) to the unit's nameplate capacity (or maximum observed hourly gross load (in MWe/hr) if greater than the nameplate capacity) times 8760 hours; or (2) The ratio of a unit's annual heat input (in million BTUs or equivalent units of measure) to the unit's maximum rated hourly heat input rate (in million BTUs per hour or equivalent units of measure) times 8,760 hours. The list of units in the ptegu_pk sector is provided in the file [ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/reports/2011_emissions/Peakers_CAMD_20 11.080213_NEI_IPM_match.xls](ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/reports/2011_emissions/Peakers_CAMD_2011.080213_NEI_IPM_match.xls)).

3.2.1.3 Non-IPM Sector (pt_oilgas)

The pt_oilgas sector includes sources with the SCCs identified as oil and gas sources. 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. The nonpoint 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 3.21 of the 2011NEIv1 TSD.

3.2.1.4 Non-IPM Sector (ptnonipm)

Except for some minor exceptions, the non-IPM (ptnonipm) sector contains the 2011NEIv1 point sources that are not in the ptegu, ptegu_pk, 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 present in the ptnonipm sector. The sector also includes some ethanol plants that have been identified by EPA but are not in 2011NEIv1. 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. Some point sources in the 2011NEIv1 that are not included in any modeling sectors are sources with state/county FIPS code ending with "777". These sources 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 for modeling. Therefore, these sources are dropped from the point-based sectors in the modeling platform.

Another difference between the 2011NEIv1 data and the modeling platform is the addition of some ethanol production facilities identified by EPA but were not found in the NEI. For some rule development work, EPA developed a list of corn ethanol facilities for 2011. Many of these ethanol facilities were included in the 2011NEIv1, but those that were not matched were added to the ptnonipm sector in a separate inventory data file. Locations and FIPS codes for these ethanol plants were verified using web searches and Google Earth. EPA believes that some of these sources are not included in the NEI as point sources because they do not meet the 100 ton/year potential-to-emit threshold for NEI point sources. In other cases, EPA is following up with states to evaluate whether the state data should include these point sources. Emission rates for the ethanol plants were obtained from EPA's updated spreadsheet model for upstream impacts developed for the Renewable Fuel Standard (RFS2) rule.

3.2.2   *Nonpoint Sources (afdust, ag, nonpt, np_oilgas, rwc)*

Several modeling platform sectors were created from the 2011NEIv1 nonpoint inventory. This section describes the stationary nonpoint sources. Note that locomotives, C1 and C2 CMV, and C3 CMV are also included the 2011NEIv1 nonpoint data category, but are mobile sources and are placed into the c1c2rail and c3marine sectors, respectively. The 2011NEIv1 TSD available from http://www.epa.gov/ttn/chief/net/2011inventory.html includes documentation for the nonpoint sector of the 2011NEIv1.

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.

The emissions modeling sector inventories start with the NEI data. Several source categories were not included in the modeling platform inventories for the following reasons: 1) these sources are only reported by a small number of states or agencies, 2) these sources are 'atypical' and have small emissions, and/or 3) EPA has have other data the Agency believes to be more accurate. The following subsections describe how the remaining sources in the 2011NEIv1 nonpoint inventory were separated into 2011 modeling platform sectors, along with any data that were updated replaced with non-NEI data.

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.

3.2.2.1 Area Fugitive Dust Sector (afdust)

The area-source fugitive dust (afdust) sector contains PM10 and PM2.5 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-based 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), http://www.epa.gov/ttn/chief/conference/ei19/session9/pouliot_pres.pdf, and in "Fugitive Dust Modeling for the 2008 Emissions Modeling Platform" (Adelman, 2012). The purpose of applying the transport fraction and meteorological adjustments is to reduce the overestimation of fugitive dust in the grid modeling as compared to ambient observations. 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.

The dust emissions in the modeling platform are not the same as the 2011NEIv1 emissions because the NEI paved and unpaved road dust emissions include a built-in precipitation reduction that is based on average meteorological data, which is at a coarser temporal and spatial resolution than the modeling platform meteorological adjustment. Due to this, in the platform paved and unpaved road emissions data was used that did not include any precipitation-based reduction. This allows the entire sector to be processed consistently so that the same grid-specific transport fractions and meteorological adjustments can be applied. Where states submitted afdust data, it was assumed that the state-submitted data were not met-adjusted and therefore the meteorological adjustments were still applied. Thus, it is possible that these sources may have been adjusted twice. Even with that possibility, air quality modeling shows that in general, dust is frequently overestimated in the air quality modeling results.

3.2.2.2 Agricultural Ammonia Sector (ag)

The agricultural NH₃ "ag" sector is comprised of livestock and agricultural fertilizer application emissions from the nonpoint sector of the 2011 NEI. The livestock and fertilizer emissions were extracted based on SCC. The "ag" sector includes all of the NH₃ emissions from fertilizer

contained in the NEI. However, the "ag" sector does not include all of the livestock ammonia emissions, as there are also some NH_3 emissions from feedlot livestock in the point source inventory in California (175 tons) and Wisconsin (125 tons). To prevent double-counting, emissions were not included in the nonpoint ag inventory for counties in which they were in the point source inventory.

3.2.2.3 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 2011NEIv1, see Section 3.21 of the 2011NEIv1 TSD. See the pt_oilgas sector of this document for more information on point source oil and gas sources.

3.2.2.4 Residential Wood Combustion Sector (rwc)

The residential wood combustion (rwc) sector includes residential wood burning devices such as fireplaces, fireplaces with inserts (inserts), free standing woodstoves, pellet stoves, outdoor hydronic heaters (also known as outdoor wood boilers), indoor furnaces, and outdoor burning in firepots and chimeneas. Free standing woodstoves and inserts are further differentiated into three categories: 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 2011NEIv1 TSD.

3.2.2.5 Other Nonpoint Sources (nonpt)

Stationary nonpoint sources that were not subdivided into the afdust, ag, np_oilgas, or rwc sectors were assigned to the "nonpt" sector. Locomotives and CMV mobile sources from the 2011NEIv1 nonpoint inventory were assigned to the c1c2rail 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.

Most sources in this sector have annual emissions that are temporally allocated to hourly values using temporal profiles, but the annual agricultural burning estimates are treated as monthly values. The annual values in the 2011NEIv1 were split into monthly emissions by aggregating the data up to monthly values from daily estimates of emissions.

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 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 nonpt 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 fire emissions are consistent with the fires stored in the Events data category of the 2011NEIv1. For more information on the development of the 2011NEIv1 fire inventory, see Section 5.1 of the 2011NEIv1 TSD.

The point source day-specific emission estimates for 2011 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) 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).

SMARTFIRE 2 estimates were used directly for all states except Georgia and Florida. For Georgia, the satellite-derived emissions were removed from the ptfire inventory and replaced with a separate state-supplied ptfire inventory. Adjustments were also made to Florida as described in Section 5.1.4 of the 2011NEIv1 TSD. These changes made the data in the ptfire inventory consistent with the data in the 2011NEIv1.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).

3.2.5   Biogenic Sources (beis)

For CMAQ, biogenic emissions were computed with the BEIS3.14 model within SMOKE using 2011 meteorological data. The BEIS3.14 model creates gridded, hourly, model-species emissions from vegetation and soils). It estimates CO, VOC (most notably isoprene, terpine, and sesquiterpene), and NO emissions for the U.S., Mexico, and Canada. The BEIS3.14 model is described further in http://www.cmascenter.org/conference/2008/slides/pouliot_tale_two_cmas08.ppt. Additional references for this method are provided in (McKenzie, et al., 2007), (Ottmar, et al., 2003), (Ottmar, et al., 2006), and (Anderson et al., 2004). The inputs to BEIS include:

- Temperature data at 2 meters from the CMAQ meteorological input files,
- Land-use data from the Biogenic Emissions Landuse Database, version 3 (BELD3) that provides data on the 230 vegetation classes at 1-km resolution over most of North America.

3.2.6   Mobile Sources (onroad, onroad_rfl, 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).

For the 2011 platform, the 2011 onroad emissions are separated into two sectors: (1) "onroad" and (2) "onroad_rfl". The onroad and onroad_rfl sectors are processed separately to allow for different spatial allocation to be applied to onroad refueling, which is allocated using a gas station surrogate, versus onroad vehicles, which are allocated using surrogates based on roads and population. Except for California and Texas, all onroad and onroad refueling emissions are generated using the SMOKE-MOVES emissions modeling framework that leverages MOVES generated outputs (http://www.epa.gov/otaq/models/moves/index.htm) 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 (including refueling), nonroad

and c1c2rail sources in California were 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 and are treated as point emissions with an elevated release component; all other nonroad emissions are treated as county-specific low-level emissions (i.e., are in model layer 1). The NEI c3marine emissions were replaced with a set of approximately 4-km resolution point source format emissions. These data are used for all states, including California, as well as offshore and international emissions within our air quality modeling domain, and are modeled separately as point sources in the "c3marine" sector.

3.2.7   Onroad non-refueling (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). For the 2011v6.1 platform, the 2011 onroad emissions are separated into two sectors: (1) "onroad" and (2) "onroad_rfl". The onroad and onroad_rfl sectors are processed separately to allow for different spatial allocation to be applied to onroad refueling, which is allocated using a gas station surrogate, versus onroad vehicles, which are allocated using surrogates based on roads and population. Except for California and Texas, all onroad and onroad refueling emissions are generated using the SMOKE-MOVES emissions modeling framework that leverages MOVES generated outputs (http://www.epa.gov/otaq/models/moves/index.htm) and hourly meteorology. All tribal data from the mobile sectors have been dropped because the emissions are small, the emissions could be double-counted with state-provided onroad emissions, all tribal data was developed using the older model MOBILE6, and because spatial surrogate data is not currently available.

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. The "SMOKE-MOVES" integration tool was developed by EPA in 2010 and is in use by states and regional planning organizations for regional air quality modeling of onroad mobile sources. 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 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) or VPOP (vehicle population)) to produce emissions. These calculations were done for every county and grid cell, in the continental U.S. for each hour of the year.

Using SMOKE-MOVES for creating the model-ready emissions requires numerous 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 a list of temperatures and activity data.
5) Run MOVES to create emission factor tables using the MOVESTier3FRM model (specifically, model "Moves 20121002f" with default database "movesdb20121002l_truncatedgfre")
6) Run SMOKE to apply the emission factors to activity data (VMT and VPOP) to calculate emissions
7) Aggregate the results to the county-SCC level for summaries and quality assurance

The California and Texas onroad emissions were created through a hybrid approach of combining state-supplied annual emissions (from the 2011NEIv1) with EPA developed SMOKE-MOVES runs. Through this approach, the platform was able to reflect California's unique rules and Texas' detailed modeling, while leveraging the more detailed SCCs and the highly resolved spatial patterns, temporal patterns, and speciation from SMOKE-MOVES. The basic steps involved in temporally allocating California's and Texas' onroad emissions based on SMOKE-MOVES results were:

1) Run CA/TX using EPA inputs through SMOKE-MOVES to produce hourly 2011 emissions hereafter known as "EPA estimates". These EPA estimates for CA/TX are run in a separate sector called "onroad_catx".

2) Calculate ratios between state-supplied emissions and EPA estimates[6]. For Texas, these ratios were calculated for each county/SCC7 (fuel and vehicle type)/pollutant combination. For California, these were calculated for each county/SCC3 (fuel type)/pollutant combination. These were not calculated at a greater resolution because California's emissions did not provide data for all vehicle types.

3) Create an adjustment factor file (CFPRO) that includes EPA-to-state estimate ratios. For extended idle adjustments, each specific state ratio (county/SCC Group (7 or 3)/pollutant)

---

[6] These ratios were created for all matching pollutants. These ratios were duplicated for all appropriate modeling species. For example, EPA used the NO_X ratio for NO, NO_2, HONO and used the $PM_{2.5}$ ratio for PEC, PNO_3, POC, PSO_4, and PMFINE (For more details on NO_X and PM speciation, see Sections 3.2.3 and **Error! Reference source not ound.**). For VOC model-species, if there was an exact match (e.g., BENZENE), EPA used that HAP pollutant ratio. For other VOC-based model-species that didn't exist in the NEI inventory, EPA used VOC ratios.

was multiplied by the extended idle adjustment factor (see the 2011NEIv1 TSD for details).

4) Rerun CA/TX through SMOKE-MOVES using EPA inputs and the new adjustment factor file.

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

3.2.8 Onroad Refueling (onroad_rfl)

Onroad refueling is modeled very similarly to other onroad emissions, and were generated using MOVESTier3FRM. The onroad_rfl emissions are spatially allocated to gas station locations. Because the refueling emission factors use the same SCCs as the other onroad models, refueling was run in a separate sector from the other onroad mobile sources to allow for the different spatial allocation. To facilitate this, the refueling EFs were separated from the other emission factors into rate-per-distance (RPD) refueling and rate-per-vehicle (RPV) refueling tables[7]. SMOKE-MOVES was run using these EF tables as inputs, and spatially allocated using a gas stations spatial surrogate. Lastly, the SMOKE program Mrggrid combined RPD refueling and RPV refueling into a single onroad_rfl model-ready output for final processing with the other sectors prior to use in CMAQ.

EPA SMOKE-MOVES generated emissions for onroad refueling were used without any adjustments for all states, including California and Texas. These emissions were used instead of state submissions to provide a consistent approach nationwide and also because most states did not submit refueling emissions for diesel fuel. Since the 2011NEIv1 includes the state-submitted emissions, the platform and the NEI refueling emissions in the nonpoint category are inconsistent for states that submitted refueling emissions. For states that didn't submit emissions, the approaches are similar but not identical because of differences in the MOVES version, specifically 2010b for the NEI and Tier3FRM for the modeling platform.

3.2.9 Nonroad Mobile Sources — NMIM-Based (nonroad)

The nonroad equipment emissions are equivalent to the emissions in the nonroad data category of the 2011NEIv1, with the exception that the modeling platform emissions also include monthly totals. All nonroad emissions are compiled at the county/SCC level. NMIM (EPA, 2005) 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 2011NEIv1 nonroad emissions are available in Section 4.5 the 2011NEIv1 TSD.

---

[7] The Moves2smk post-processing script has command line arguments that will either consolidate or split out the refueling EF.

California year 2011 nonroad emissions were submitted to the 2011NEIv1 and are also documented in a staff report (ARB, 2010a). 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 2011NEIv1 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.

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.

3.2.10 Nonroad Mobile Sources: Commercial Marine C1, C2, 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 starting point for the c1c2rail sector is the 2011NEIv1 nonpoint inventory for all but specific Midwest states, which are instead derived from the Great Lakes 2010 CMV inventory. The modeling platform emissions for the c1c2rail SCCs were extracted from the NEI nonpoint inventory. For more information on CMV sources in the NEI, see Section 4.3 of the 2011NEIv1 TSD. For more information on locomotives, see Section 4.4 of the 2011NEIv1 TSD.

The difference between the 2011NEIv1 and the modeling platform for this sector is due to the availability of alternative data from the Midwest RPO. Year-2010 emissions were received from the Lake Michigan Air Directors Consortium for tug boats, Great Lakes vessels ("Lakers") and inland waterways for states within the Midwest RPO and Minnesota, hereafter simply referred to as "MWRPO" (http://www.ladco.org/). The states in the MWRPO are: Illinois, Indiana, Michigan, Minnesota, Ohio and Wisconsin. These MWRPO CMV emissions include coverage for bordering states/counties along the inland waterways such as the Mississippi and Ohio rivers in Iowa, Missouri, Kentucky, West Virginia, Pennsylvania and New York. The LADCO 2010 inventory was used to replace EPA-estimated CMV emissions in the MWRPO states, but was

not used to replace the 2011NEIv1 emissions in the bordering non-MWRPO states.

Some modifications to the MWRPO CMV data were made prior to SMOKE processing:

- Emissions provided at the level of NEI Shape IDs were aggregated to county-level.
- The 2011NEIv1 was used to determine which counties had ports; for those counties that had ports, 90% of emissions in the MWPRO inventory were assigned as underway (SCC=2280002200) and 10% were assigned as port emissions (SCC=2280002100).
- Emissions were converted to short tons and PM2.5 was added by assuming it is equal to 92% of PM10 at the suggestion of the MWRPO.
- Tugs were assigned a unique SCC (2280002021) to allow for unique spatial allocation (see Section 3.4.1).
- Tugs were assigned from MWRPO total to counties based on 2011NEIv1 county-level activity information for tug vessels.

Because the Great Lakes vessels include all CMV activity on the Great Lakes, EPA-estimated C3 CMV (c3marine) sector emissions (discussed in the following section) in the MWRPO states were removed to avoid potential double-counting of C3 CMV with the LADCO inventory in the MWRPO states.

3.2.11 Nonroad mobile sources: C3 commercial marine (c3marine)

The U.S. C3 CMV inventory was developed based on a 4-km resolution ASCII raster format dataset used since the Emissions Control Area-International Marine Organization (ECA-IMO) project began in 2005, then known as the Sulfur Emissions Control Area (SECA). The ECA-IMO data are used instead of the 2011NEIv1 data for the modeling platform because accompanying estimates of emission projections for future years are available. In addition, the inventory preserves shipping lanes in federal waters while these are not stored within the NEI data. Keeping the sources in this sector separate from smaller CMV sources allows for the emissions to be elevated above the surface layer within the AQ model. The ECA-IMO data are used for all states with C3 CMV emissions. For the MWPRO states, the ECA-IMO C3 CMV emissions in the Great Lakes are assumed to be misclassified as C3 vessels for which emissions are included in the c1c2rail sector as part of the LADCO inventory, therefore the ECA-IMO emissions are not included in the c3marine sector.

The development of this ECA-IMO-based C3 CMV inventory is discussed below; however, all non-U.S. emissions (Canadian emissions and emissions farther offshore than U.S. waters) are processed in the "othpt" sector. This splitting of the C3 CMV emissions from the farther offshore emissions allows for easier summaries of U.S.-only and state or county total emissions.

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 often burn residual fuel. 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, SO_2, 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 at: http://www.epa.gov/oms/oceanvessels.htm.

The ECA-IMO emissions data were converted to SMOKE point-source ORL input format as described in http://www.epa.gov/ttn/chief/conference/ei17/session6/mason.pdf. 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. Counties were assigned as extending up to 200nm from the coast because this was the distance to the edge of the U.S. Exclusive Economic Zone (EEZ), a distance that defines the outer limits of ECA-IMO controls for these vessels.

The base year ECA inventory is 2002 and consists of these CAPs: PM_{10}, $PM_{2.5}$, CO, CO_2, NH_3, NO_X, SO_X (assumed to be SO_2), and hydrocarbons (assumed to be VOC). EPA developed regional growth (activity-based) factors that were applied to create the 2011 inventory from the 2002 data. These growth factors are provided in Table 3-4. The emissions were converted to SMOKE point source inventory format, allowing for the emissions to be allocated to modeling layers above the surface layer. 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 in Table 3-4. 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.

For California, the ECA-IMO 2011 emissions were scaled by county 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.

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). 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).

Table 3-4. Growth factors to project the 2002 ECA inventory to 2011

Region	EEZ FIPS	NO$_X$	PM$_{10}$	PM$_{2.5}$	VOC (HC)	CO	SO$_2$
East Coast (EC)	85004	1.301	0.500	0.496	1.501	1.501	0.536
Gulf Coast (GC)	85003	1.114	0.428	0.423	1.288	1.288	0.461
North Pacific (NP)	85001	1.183	0.467	0.458	1.353	1.353	0.524
South Pacific (SP)	85002	1.367	0.525	0.521	1.565	1.562	0.611
Great Lakes (GL)	n/a	1.072	0.394	0.390	1.177	1.176	0.415
Outside ECA	98001	1.341	1.457	1.457	1.457	1.457	1.457

* 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 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, were projected to year 2011 using the same regional adjustment factors as the U.S. emissions; however, the state/county FIPS codes were assigned as "EEZ" codes and these 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 MWRPO states and other states such as Pennsylvania and New York. The classification of emissions to U.S. and Canadian FIPS codes is 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 emissions were converted to SMOKE point source inventory format, allowing for the emissions to be allocated to modeling layers above the surface layer. 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 in Table 3-4. 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[8].

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.

3.2.12 *Emissions from Canada, Mexico and Offshore Drilling Platforms (othpt, othar, othon)*

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 three emissions modeling sectors: othpt, othar, and othon. The "oth" refers to the fact that these emissions are usually "other" than those in the U.S. state-county geographic FIPS, and the third and fourth 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 emissions for non-U.S. states are processed in the othpt sector. These C3 CMV emissions include those assigned to Canada, those assigned to the Exclusive Economic Zone (EEZ, defined as those emissions just beyond U.S. waters approximately 3-10 miles offshore, extending to about 200 nautical miles from the U.S. coastline), and all other offshore emissions –far offshore and non-U.S. These emissions are included in the othpt sector for simplicity of creating U.S.-only emissions summaries. Otherwise, these emissions are developed in the same way as the U.S. C3 CMV emissions in the c3marine sector.

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, the 2008 NEI version 3 point source inventory data were used because the 2011 data were not yet available. This is consistent with the 2011NEIv1. Updated offshore oil and gas drilling emissions are expected to be incorporated into version 2 of the 2011 NEI. The 2008-based offshore emission sources were provided by the Mineral Management Services (MMS).

For Canada, year-2006 Canadian emissions were the latest available at the time the modeling was performed. These were the starting point with the addition of several modifications to these inventories. The SCCs in these inventories were changed to the generic 39999999 and the industrial code information was removed to preserve confidentiality. The Canadian point sources were split into three inventory files:

---

[8] The extent of the "36US1" domain is similar to the full geographic region shown in Figure 3-1. Note that this domain is not specifically used in this 2011 platform, although spatial surrogates that can be used with it are provided.

- ptinv_canada_point_2006_orl_13aug2013_v3_orl.txt: contains point sources for all pollutants except VOC;

- ptinv_canada_point_cb5_2006_orl_13aug2013_v1_orl.txt: contains VOC emissions split into CB05 species;

- ptinv_canada_point_uog_2006_orl_02mar2009_v0_orl.txt: contains oil and gas-related sources.

The year-2006 nonpoint emissions provided by Canada were unchanged from EPA 2007 platform. Inventory files were provided for area fugitive dust, agricultural, commercial marine, railroad, nonroad, aircraft, and other area sources. Canadian onroad emissions are also unchanged from the EPA 2007 platform.

For Mexico, point, nonpoint, and onroad emissions for year 2012 are projections of their 1999 inventory originally developed by Eastern Research Group Inc., (ERG, 2006; ERG, 2009; Wolf, 2009) as part of a partnership between Mexico's Secretariat of the Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales-SEMARNAT) and National Institute of Ecology (Instituto Nacional de Ecología-INE), the U.S. EPA, the Western Governors' Association (WGA), and the North American Commission for Environmental Cooperation (CEC). This inventory includes emissions from all states in Mexico. A background on the development of year-2012 Mexico emissions from the 1999 inventory is available at: http://www.wrapair.org/forums/ef/inventories/MNEI/index.html.

3.2.13 *SMOKE-ready non-anthropogenic chlorine inventory*

The ocean chlorine gas emission estimates are based on the build-up of molecular chlorine (Cl2) concentrations in oceanic air masses (Bullock and Brehme, 2002). Data at 36 km and 12 km resolution were available and were not modified other than the name "CHLORINE" was changed to "CL2" because that is the name required by the CMAQ model. The same data was used as in the CAP and HAP 2002-based Platform was used. See ftp://ftp.epa.gov/EmisInventory/2002v3CAPHAP/ documentation for additional details.

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 the model, it is necessary to "pre-process" the emission inventories (i.e., emissions input to SMOKE) for the sectors as described above. In brief, the process of emissions modeling transforms the emissions inventories from their original temporal resolution, pollutant resolution, and spatial resolution into the hourly, speciated, gridded resolution required by the air quality model. Emissions modeling includes temporal allocation, spatial allocation, and pollutant speciation. In some cases, emissions modeling also includes the vertical allocation of point sources, but many air quality models also perform this task because it greatly reduces the size of the input emissions files if the vertical 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 even 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 3.5.1 was used to pre-process the raw emissions inventories into emissions inputs for CMAQ. SMOKE executables and source code are available from the Community Multiscale Analysis System (CMAS) Center at http://www.cmascenter.org. Additional information about SMOKE is available from http://www.smoke-model.org. For sectors that have plume rise, the in-line emissions capability of the air quality models was used, which allows the creation of source-based and two-dimensional gridded emissions files that are much smaller than full three-dimensional gridded emissions files. For quality assurance of the emissions modeling steps, emissions totals by specie for the entire model domain are output as reports that are then compared to reports generated by SMOKE on the input inventories to ensure that mass is not lost or gained during the emissions modeling process.

3.3.2 Key Emissions Modeling Settings

When preparing emissions for the air quality model, emissions for each sector are processed separately through SMOKE, and then the final merge program (Mrggrid) is run to combine the model-ready, sector-specific emissions across sectors. The SMOKE settings in the run scripts and the data in the SMOKE ancillary files control the approaches used by the individual SMOKE programs for each sector. Table 3-5 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 c3marine, othpt, and ptfire sectors are the only sectors with only "in-line" emissions, meaning that all of the emissions are placed in aloft layers and there are no emissions for those sectors in the two-dimensional, layer-1 files created by SMOKE.

Table 3-5. Key emissions modeling steps by sector

Platform sector	Spatial	Speciation	Inventory resolution	Plume rise
Afdust	Surrogates	Yes	annual	
Ag	Surrogates	Yes	annual (some monthly)	
Beis	Pre-gridded land use	in BEIS3.14	computed hourly	
c1c2rail	Surrogates	Yes	annual	
c3marine	Point	Yes	annual	in-line
Nonpt	Surrogates & area-to-point	Yes	annual (some monthly)	
Nonroad	Surrogates & area-to-point	Yes	monthly	
np_oilgas	Surrogates	Yes	annual	
Onroad	Surrogates	Yes	monthly activity, computed hourly	
onroad_rfl	Surrogates	Yes	monthly activity, computed hourly	
Othar	Surrogates	Yes	annual	
Othon	Surrogates	Yes	annual	
Othpt	Point	Yes	annual	in-line
pt_oilgas	Point	Yes	annual	in-line
ptegu	Point	Yes	daily & hourly	in-line
ptegu_pk	Point	Yes	daily & hourly	in-line
ptfire	Point	Yes	daily	in-line
ptnonipm	Point	Yes	annual	in-line
rwc	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 2011 platform, 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.
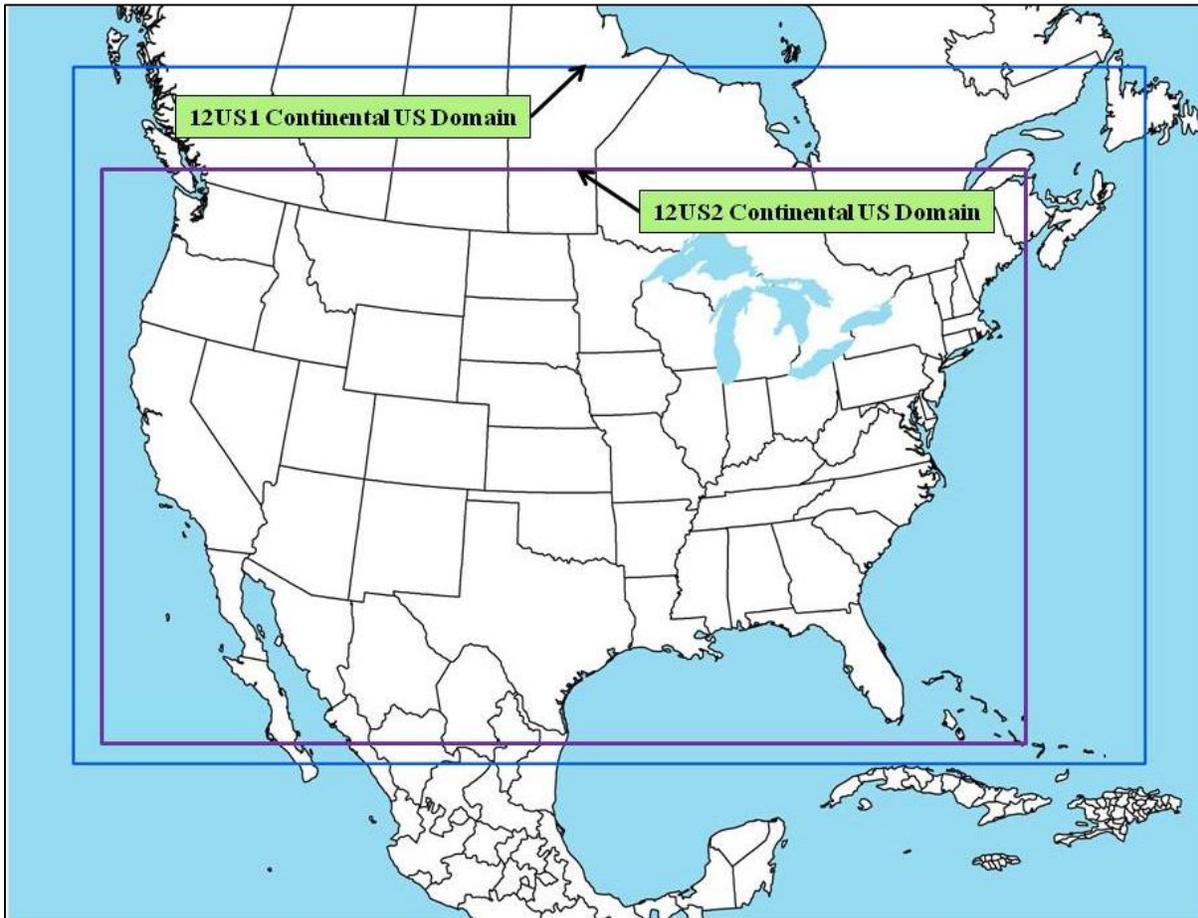
Figure 3-2. CMAQ Modeling Domain

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 the 2011 platform is the CB05 mechanism (Yarwood, 2005). The same base chemical mechanism is used within both CMAQ and CAM$_X$, but the implementation differs slightly between the two models. The specific versions of CMAQ and CAMx used in applications of this platform include secondary organic aerosol (SOA) and HONO enhancements.

From the perspective of emissions preparation, the CB05 with SOA mechanism is the same as was used in the 2007 platform. Table 3-6 lists the model species produced by SMOKE for use in CMAQ and CAM$_X$. 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.

The TOG and PM$_{2.5}$ speciation factors that are the basis of the chemical speciation approach were developed from the SPECIATE 4.3 database (http://www.epa.gov/ttn/chief/software/speciate), 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 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$_{2.5}$.

Speciation profiles and cross-references for 2011v6 platform are available in spreadsheet form from ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/reports/speciation_profiles/. The profiles are in the Excel files "gspro_2011.xlsx" and "gspro_combo_2011.xlsx, gsref_2011.xlsx". The cross reference information is in "gsref_2011.xlsx. A spreadsheet showing emission totals for each speciation profile for the 2011ed case by modeling sector is available in the file "2011ed_speciation_profile_CAPs_feb112014.xlsx". Note that the emissions totals differ slightly from the 2011ef case, as do some of the VOC to TOG conversion factors. However, the reports still convey the relative importance of each speciation profile in terms of emissions affected.

Table 3-6. Emission model species produced for CB05 with SOA for CMAQ 5.0.1

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
VOC	ALD2	Acetaldehyde
	ALDX	Propionaldehyde and higher aldehydes
	BENZENE	Benzene (not part of CB05)
	CH4	Methane9
	ETH	Ethene
	ETHA	Ethane
	ETOH	Ethanol
	FORM	Formaldehyde
	IOLE	Internal olefin carbon bond (R-C=C-R)
	ISOP	Isoprene
	MEOH	Methanol
	OLE	Terminal olefin carbon bond (R-C=C)
	PAR	Paraffin carbon bond
	TOL	Toluene and other monoalkyl aromatics
	XYL	Xylene and other polyalkyl aromatics
VOC species from the biogenics model that do not map to model species above	SESQ	Sesquiterpenes
	TERP	Terpenes
PM_{10}	PMC	Coarse PM > 2.5 microns and ≤ 10 microns
$PM_{2.5}$[10]	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	Other particulate matter ≤ 2.5 microns
Sea-salt species (non – anthropogenic) [11]	PCL	Particulate chloride
	PNA	Particulate sodium

*The same species names are used for the CAM_X model with exceptions as follows:
1. CL2 is not used in CAM_X
2. CAM_X particulate sodium is NA (in CMAQ it is PNA)
3. CAM_X uses different names for species that are both in CBO5 and SOA for the following: TOLA=TOL, XYLA=XYL, ISP=ISOP, TRP=TERP. They are duplicate species in CAM_X that are used in the SOA chemistry. CMAQ uses the same names in CB05 and SOA for these species.
4. CAM_X uses a different name for sesquiterpenes: CMAQ SESQ = CAM_X SQT
5. CAM_X particulate species have different names for organic carbon, coarse particulate matter and other particulate mass: CMAQ uses POC, PMC, PMFINE, and PMOTHR, while CAM_X uses POA, CPRM, FCRS, and FPRM, respectively.

---

[9] Technically, CH4 is not a VOC but part of TOG. Although emissions of CH4 are derived, the AQ models do not use

The speciation of VOC includes HAP emissions from the 2011NEIv1 in the speciation process. Instead of speciating VOC to generate all of the species listed in Table 3-6, 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 NEI 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 5.0.1 (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 5.0.1, and there was a desire to keep its emissions consistent between multi-pollutant and base versions of CMAQ.

For specific sources, especially within the onroad and onroad_rfl sectors, the integration included ethanol. To differentiate when a source was integrating BAFM versus EBAFM (ethanol in addition to BAFM), the speciation profiles that do not include ethanol are referred to as an "E-profile" (to be used when the ethanol comes from the inventory pollutant). For example, use E10 headspace gasoline evaporative speciation profile 8763 when ethanol is speciated from VOC, but use 8763E when ethanol is obtained directly from the inventory.

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 integration[12]). For the "integrated" sources, SMOKE subtracts the "integrated" HAPs from the VOC (at the source level) to compute emissions for the new pollutant "NONHAPVOC." The user provides NONHAPVOC-to-NONHAPTOG factors and NONHAPTOG speciation profiles[13]. 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

---

these emissions because the anthropogenic emissions are dwarfed by the CH4 already in the atmosphere.

[10] For CMAQ 5.0, PM$_{2.5}$ is speciated into a finer set of PM components. Listed in this table are the AE5 species

[11] These emissions are created outside of SMOKE

[12] In SMOKE version 3.5.1, 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.

[13] 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.

emissions, then the sector is considered fully integrated and does not need a NHAPEXCLUDE file. If on the other hand, certain sources do not have the necessary HAPs, then an NHAPEXCLUDE file must be provided based on the evaluation of each source's pollutant mix. EPA considered CAP-HAP integration for all sectors and developed "integration criteria" for some of them.

The process of partial integration for BAFM 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.

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 **Error! Reference ource not found.**7), 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, such as the onroad and onroad_rfl sectors.

Table 3-7. 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
ptegu_pk	No integration
ptnonipm	No integration
ptfire	No integration
othar	No integration
othon	No integration
ag	N/A – sector contains no VOC
afdust	N/A – sector contains no VOC
biog	N/A – sector contains no inventory pollutant "VOC"; but rather specific VOC species
nonpt	Partial integration (BAFM and EBAFM)
np_oilgas	Partial integration (BAFM)
pt_oilgas	Partial integration (BAFM)
rwc	Partial integration (BAFM)
nonroad	Partial integration (BAFM)

Platform Sector	Approach for Integrating NEI emissions of Benzene (B), Acetaldehyde (A), Formaldehyde (F), Methanol (M), and Ethanol (E)
c1c2rail	Partial integration (BAFM)
othpt	Partial integration (BAFM)
c3marine	Full integration (BAFM)
onroad	Full integration (EBAFM and BAFM)
onroad_rfl	Full integration (EBAFM and BAFM)

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 onroad and 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. For the onroad and onroad_rfl sectors, the GSPRO_COMBO uses E-profiles (i.e. there is EBAFM integration). 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, a combination of BAFM and EBAFM integration is used. Due to the lack of SCC-specificity in the GSPRO_COMBO, the only way to differentiate the sources that should use BAFM integrated profiles versus E-profiles is by changing the pollutant name. For example, EPA changed the pollutant name for the PFC future year inventory so the integration would use EVP__NONHAPVOC to correctly select the E-profile combinations, while other sources used NONHAPVOC to select the typical BAFM profiles.

Speciation profiles for use with BEIS are not included in SPECIATE. The 2010 Platform uses BEIS3.14 and includes a species (SESQ) that was not in BEIS3.13 (the version used for the 2002 Platform). This species was mapped to the CMAQ specie SESQT. The profile code associated with BEIS3.14 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 TSD Preparation of Emission Inventories for the Version 6.1 Emissions Modeling Platform (EPA, 2014a).

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. For CMAQ 5.0.1, there is a new thermodynamic equilibrium aerosol modeling tool (ISORROPIA) v2 mechanism that needs additional PM components (AE6), which are further subsets of PMFINE (see Table 3-8). EPA

speciated PM$_{2.5}$ so that it included both AE5 and AE6 PM model species without causing any double counting. Therefore, emissions from this platform can be used with either CMAQ 4.7.1 or CMAQ 5.0.1. The majority of the 2011 platform PM profiles come from the 911XX series which include updated AE6 speciation[14].

Table 3-8. 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 1-to-1 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 Version 6.1 platform TSD (EPA, 2014a).

NO$_X$ can be speciated into NO, NO2, and/or HONO. For the non-mobile sources, EPA used a single profile "NHONO" to split NO$_X$ into NO and NO$_2$. For the mobile sources except for onroad (including nonroad, c1c2rail, c3marine, othon sectors) and for specific SCCs in othar and ptnonipm, the profile "HONO" splits NO$_X$ into NO, NO$_2$, and HONO. The onroad sector does not use the "HONO" profile to speciate NO$_X$. MOVES2010b produces speciated NO, NO$_2$, and HONO by source, including emission factors for these species in the emission factor tables used by SMOKE-MOVES. Within MOVES, the HONO fraction is a constant 0.008 of NO$_X$. The

---

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

NO fraction varies by heavy duty versus light duty, fuel type, and model year. The NO_2 fraction = 1 – NO – HONO. For more details on the NO_X fractions within MOVES, see http://www.epa.gov/otaq/models/moves/documents/420r12022.pdf. HONO is not calculated directly by the Tier 3 proposal version of MOVES. For these EF tables, the calculation of HONO and the NO_2 fraction are calculated externally by the moves2smk script[15]. The SMOKE-MOVES system then models these species directly without further speciation.

3.3.4   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.

The temporal profiles and associated cross references used to create the hourly emissions inputs for the 2011 air quality modeling platform were similar to those used for the 2007 platform. The temporal factors applied to the inventory are selected using some combination of country, state, county, SCC, and pollutant. Table 3-9 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 3-9. Temporal Settings Used for the Platform Sectors in SMOKE

Platform Sector	Inventory resolutions	Monthly profiles used?	Daily temporal approach	Merge processing approach	Process Holidays as separate days
ptegu	Daily & hourly		all	all	Yes
ptegu_pk	Daily & hourly		all	all	Yes
ptnonipm	Annual	yes	mwdss	mwdss	Yes
pt_oilgas	Annual	yes	mwdss	mwdss	Yes
ptfire	Daily		all	all	Yes
othpt	Annual	yes	mwdss	mwdss	
nonroad	Monthly		mwdss	mwdss	Yes

---

[15] A specific version of the moves2smk script was developed to do this calculation of HONO. The typical version assumes that HONO was calculated directly by MOVES2010b.

Platform Sector	Inventory resolutions	Monthly profiles used?	Daily temporal approach	Merge processing approach	Process Holidays as separate days
othar	Annual	yes	week	week	
c1c2rail	Annual	yes	mwdss	mwdss	
c3marine	Annual	yes	aveday	aveday	
onroad	Annual & monthly[1]		all	all	Yes
onroad_rfl	Annual & monthly[2]		all	all	Yes
othon	Annual	yes	week	week	
nonpt	Annual & monthly	yes	all	all	Yes
np_oilgas	Annual	yes	mwdss	mwdss	Yes
rwc	Annual	no	met-based	All	Yes
ag	Annual	yes	all	all	Yes
afdust_adj	Annual	yes	week	all	Yes
beis	Hourly		n/a	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.
2. Note the annual and monthly "inventory" actually refers to the activity data (VMT and VPOP) for onroad_rfl. The actual emissions are computed on an hourly basis.

The following values are used in the table: The value "all" means that hourly emissions 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, 2011, which is intended to mitigate the effects of initial condition concentrations. The ramp-up period was 10 days (December 22-31, 2010). For most sectors, emissions from December 2011 were used to fill in surrogate emissions for the end of December 2010. In particular, December 2011 emissions (representative days) were used for December 2010. For biogenic emissions, December 2010 emissions were processed using 2010 meteorology.

The Flat File 2010 format (FF10) inventory format for SMOKE provides a more consolidated format for monthly, daily, and hourly emissions inventories than previous formats supported. Previously, to process 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 3.5.1 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 (e.g. the nonpt sector). The flags that control temporalization for a mixed set of inventories are discussed in the SMOKE documentation. The modeling platform sectors that make use of monthly values in the FF10 files are nonroad, onroad, and the agricultural burning inventory within the nonpt sector.

The 2011NEIv1 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. 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. EIS stores a base set of previously matched units via alternate facility and unit IDs. For any units not yet matched, reports were generated by unit to identify potential matches with the NEI. The reports included FIPS state/county code, facility name, and NOx and SO_2 emissions. Units were considered matches if the FIPS state/county code matched, the facility name was similar, and the NOx and SO_2 emissions were similar.

For sources not matched to CEMS measurements, the first two steps of the allocation are done outside of SMOKE. For sources in the ptegu and ptegu_pk sectors that are matched to CEMS data, annual totals of the emissions may be different than the annual values in 2011NEIv1 because the CEMS data actually replaces the inventory data. All units in the ptegu_pk sector with non-zero emissions for 2011 were matched to CEMS data.

For units not matched to CEMS data, 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 3-3. 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 and SO_2, and heat input. An overall composite profile was also computed and was used in a few cases in which the fuel-specific profile was too irregular, or there were no CEMS units with the specified fuel in the region containing the unit. For both CEMS and non-CEMS matched units, NO_x and SO_2 CEMS data are used to allocate NO_x and SO_2 emissions, while CEMS heat input data is used to allocate all other pollutants.

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 nonpt sector, most the inventories are annual except for the agricultural burning (SCC

2801500000) inventory which was 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 hours - see Figure 3-4 (McCarty et al., 2009). This puts most of the emissions during the work day and suppresses the emissions during the middle of the night. A uniform profile was used for each day of the week for all agricultural burning emissions in all states except for the following, for which state-specific day of week profiles were used: Arkansas, Kansas, Louisiana, Minnesota, Missouri, Nebraska, Oklahoma, and Texas.

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

Figure 3-4. Agricultural burning diurnal temporal profile

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   *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 is 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.

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 3-5 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.
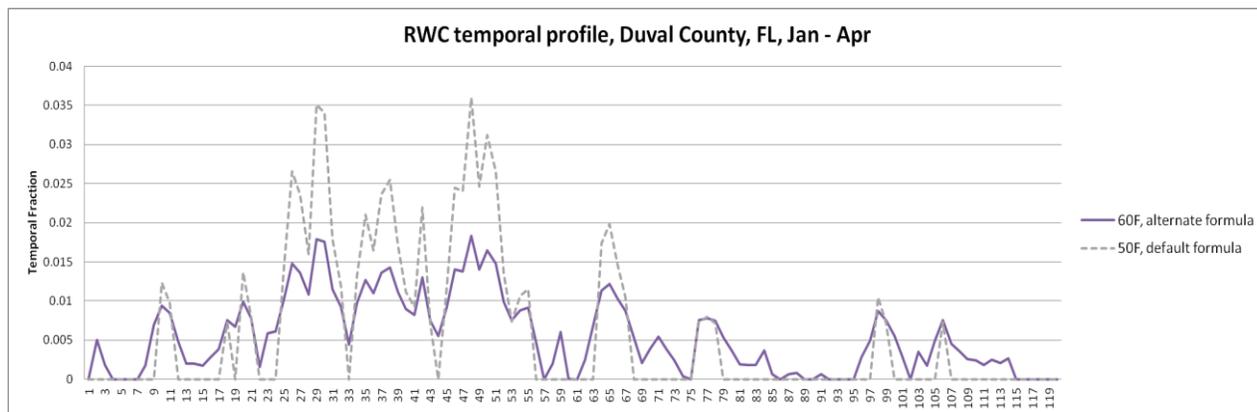

Figure 3-5. Example of RWC temporalization in 2007 using a 50 versus 60 ˚F threshold

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. Additional details regarding temporalization of RWC sources can be found in Section 3.3.3 of the 2011v6.1 Platform TSD.

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 NH3 emission variations from livestock as a function of ambient temperature, aerodynamic resistance, and wind speed. The equations are:

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

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

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 3-6 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.
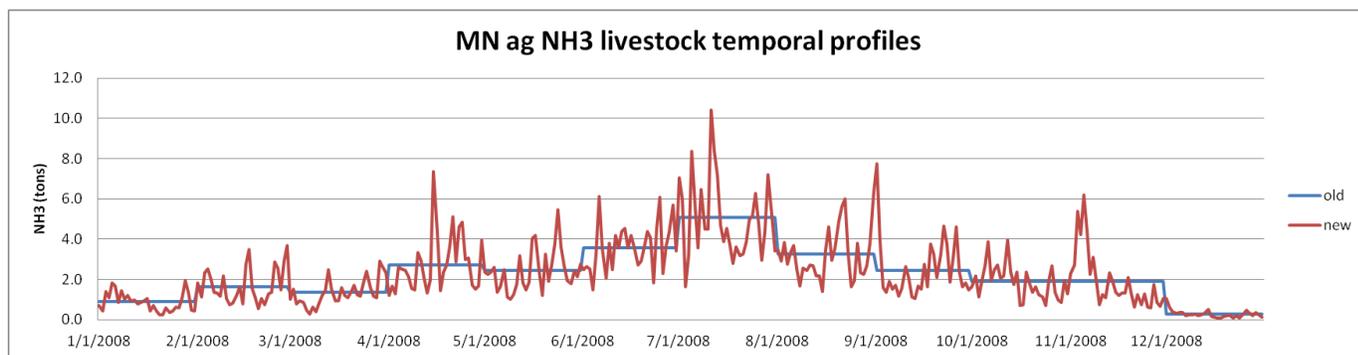

Figure 3-6. Example of new animal NH_3 emissions temporalization approach, summed to daily emissions

For the onroad and onroad_rfl sectors, 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 2011

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 (RPV 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 previous platform, RPP used county level minimum and maximum temperature ranges for the day to determine the appropriate EF. This potentially overestimated the temperature range for any particular grid cell, which would result in increased emissions for vapor-venting. In the 2011 platform (and the 2011NEIv1), 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 three processes (RPD, RPV, and RPP) is the total onroad sector emissions, while the combination of the two processes (RPD, RPV) for the refueling mode only is the total onroad_rfl sector emissions. Both sectors show a strong meteorological influence on their temporal patterns (see the 2011NEIv1 TSD for more details).

Figure 3-7 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.
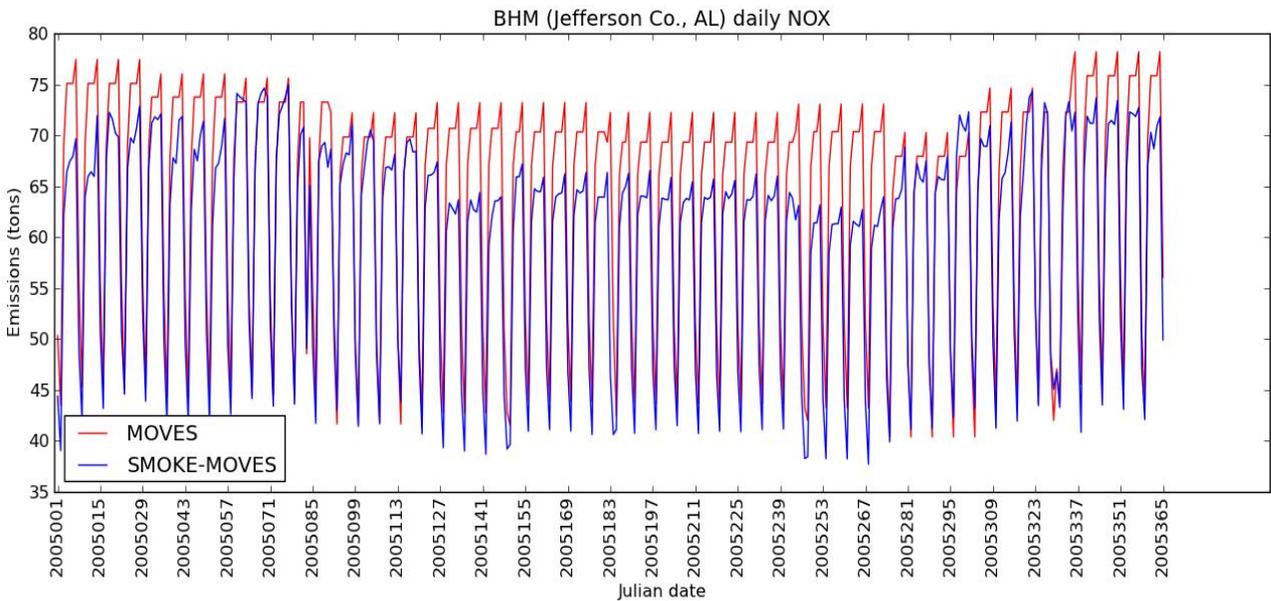
Figure 3-7. Example of SMOKE-MOVES temporal variability of NOx emissions

For the onroad and onroad_rfl sectors, the "inventories" actually consist of activity data. 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' process 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.

In previous platforms, the diurnal profile for VMT[16] varied by road type but not by vehicle type. These profiles were used throughout the nation. EPA wanted to create new diurnal profiles that could differentiate by vehicle type as well as by road type and would potentially vary over geography. The 2011NEIv1 process provided an opportunity to update the diurnal profile with information submitted by states. States submitted MOVES county databases (CDBs) that included information on the distribution of VMT by hour of day and by day of week[17] (see the 2011NEIv1 TSD for details on the submittal process for onroad). EPA decided not to update the day of week profile because MOVES only differentiated weekday versus weekend while the default SMOKE profiles differentiated each of the 7 days. EPA mined the state submitted MOVES CDBs for non-default diurnal profiles[18]. The list of potential diurnal profiles was then

---

[16] These same profiles were used for onroad emissions in the 2005 platform.

[17] The MOVES tables are the hourvmtfraction and the dayvmtfraction.

[18] 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.

analyzed to see whether the profiles varied by vehicle type, road type, weekday vs. weekend, and by county within a state.

EPA attempted to maximize the use of state and/or county specific diurnal profiles. If a specific state or county's profiles varied by vehicle type or/and road type, then the submitted profile was used. If the profile had less variability than the old SMOKE defaults (i.e. neither varied by vehicle type nor road type), 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 whether or not the profile was submitted and the characteristics of the profiles.

A new set of diurnal profiles was developed from the submitted profiles that varied by both vehicle type and road type. Before developing the national profiles, there needs to be a mapping between MOVES road types and SMOKE road types (i.e., the last three digits of the SCC) and between MOVES source types and SMOKE vehicle types. The mapping between road types is relatively straight forward. Basically the road types are consolidated into 4 types in MOVES, therefore the new profiles will not differentiate at the level of the SMOKE road type. For example, the SMOKE "urban interstate" (SCCLAST3=230) will have the same profile as the SMOKE "urban other freeways and expressways" (SCCLAST3=250). The mapping between MOVES source type and SMOKE vehicle type is more complicated; it is a many-to-many mapping. Figure 3-8 Illustrates the difference between the profiles for the light-duty gas vehicles versus the heavy-duty diesel vehicles. For additional details on the updated onroad mobile temporal profiles, see Section 3.5.5 of the 2011v6.1 Platform TSD.

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, http://www.epa.gov/ttn/chief/conference/ei19/session9/pouliot_pres.pdf, 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.
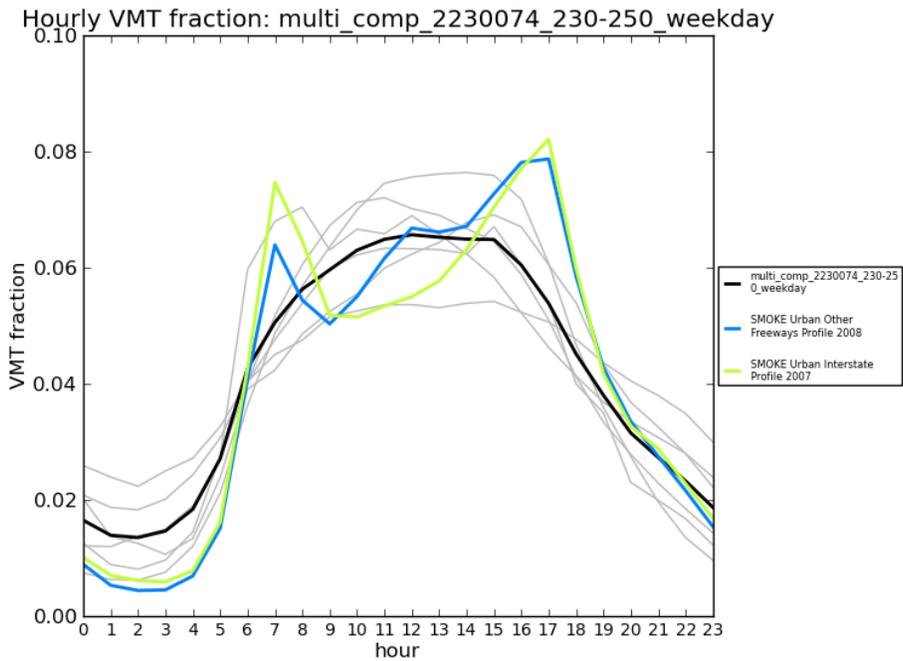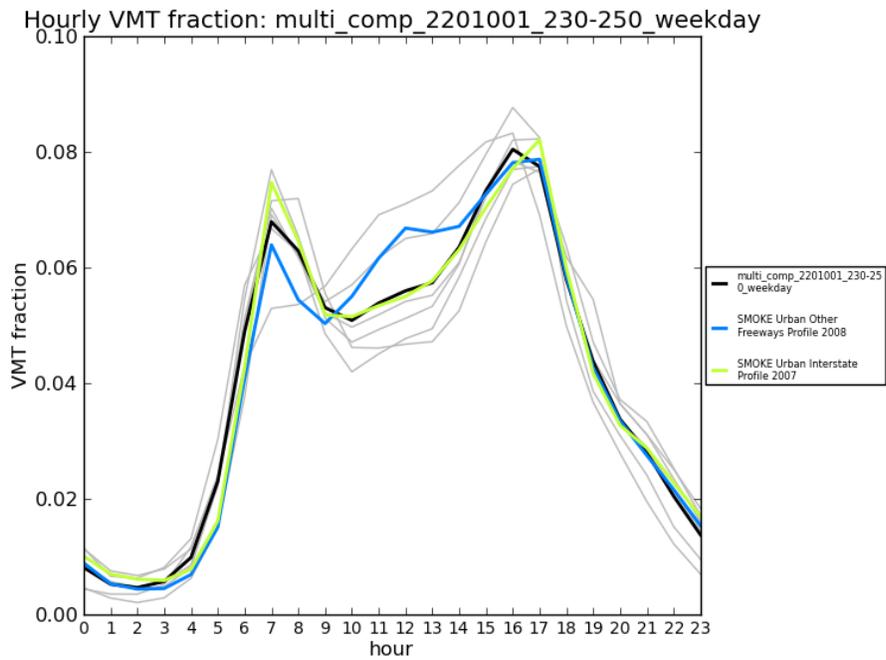
Figure 3-8. Updated national default profiles for LDGV (top) vs. HHDDV (bottom), urban restricted weekday

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 the national 12-km domain "12US2" (see Figure 3-2). 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 remainder of this subsection provides further detail on the origin of the data used for the spatial surrogates and the area-to-point data.

Additional documentation on the 2011 spatial surrogates is available at ftp://ftp.epa.gov/EmisInventory/2011v6/v1platform/reports/spatial_surrogates/ in the files US_SpatialSurrogate_Documentation_v091113.pdf and US_SpatialSurrogate_Workbook_v093013.xlsx. The spatial cross reference file is in gsref_2011.xlsx. Plots of the spatial surrogates are available in all_surrogate_maps_2011platform_12US1_v2.pdf. Note that these are plots of the surrogate fractions summed by grid cell, so grid cells that overlap multiple counties can show values

greater than one. These maps are only to give an idea of the spatial distribution of the surrogates. Allocations of CAP emissions to each of the surrogate codes is given in 2011ed_spatial_surrogate_CAPs_feb112014.xlsx. One noteworthy change between the 2011v6.0 and 2011v6.1 platforms is the update of the oil and gas surrogates.

3.3.7.1 Surrogates for U.S. Emissions

There are more than 70 spatial surrogates available for spatially allocating U.S. county-level emissions to the 12-km grid cells used by the air quality model. An area-to-point approach overrides the use of surrogates for some sources. Table 3-10 lists the codes and descriptions of the surrogates. The surrogates in bold have been updated with 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 2011NEIv1 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 and documentation for it is available at http://www.ie.unc.edu/cempd/projects/mims/spatial/ and http://www.cmascenter.org/help/documentation.cfm?MODEL=spatial_allocator& VERSION=3.6&temp_id=99999.

Table 3-10. 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)	520	Commercial plus Industrial plus Institutional
100	Population	525	Golf Courses + Institutional +Industrial + Commercial
110	Housing	527	Single Family Residential
120	Urban Population	530	Residential - High Density
130	Rural Population	535	Residential + Commercial + Industrial + Institutional + Government
137	Housing Change	540	Retail Trade
140	Housing Change and Population	545	Personal Repair
150	Residential Heating - Natural Gas	550	Retail Trade plus Personal Repair
160	Residential Heating – Wood	555	Professional/Technical plus General Government
165	0.5 Residential Heating - Wood plus 0.5 Low Intensity Residential	560	Hospital
170	Residential Heating - Distillate Oil	565	Medical Office/Clinic
180	Residential Heating – Coal	570	Heavy and High Tech Industrial
190	Residential Heating - LP Gas	575	Light and High Tech Industrial
200	Urban Primary Road Miles	580	Food, Drug, Chemical Industrial
210	Rural Primary Road Miles	585	Metals and Minerals Industrial
220	Urban Secondary Road Miles	590	Heavy Industrial
230	Rural Secondary Road Miles	595	Light Industrial
240	Total Road Miles	596	Industrial plus Institutional plus Hospitals

Code	Surrogate Description	Code	Surrogate Description
250	Urban Primary plus Rural Primary	600	Gas Stations
255	0.75 Total Roadway Miles plus 0.25 Population	650	Refineries and Tank Farms
260	Total Railroad Miles	675	Refineries and Tank Farms and Gas Stations
270	Class 1 Railroad Miles	680	Oil & Gas Wells, IHS Energy, Inc. and USGS
261	NTAD Total Railroad Density	700	Airport Areas
271	NTAD Class 1, 2, 3 Railroad Density	710	Airport Points
280	Class 2 and 3 Railroad Miles	720	Military Airports
300	Low Intensity Residential	800	Marine Ports
310	Total Agriculture	801	NEI Ports
312	Orchards/Vineyards	802	NEI Shipping Lanes
320	Forest Land	807	Navigable Waterway Miles
330	Strip Mines/Quarries	808	Gulf Tug Zone Area
340	Land	810	Navigable Waterway Activity
350	Water	812	Midwest Shipping Lanes
400	Rural Land Area	850	Golf Courses
500	Commercial Land	860	Mines
505	Industrial Land	870	Wastewater Treatment Facilities
510	Commercial plus Industrial	880	Drycleaners
515	Commercial plus Institutional Land	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 to population. For the onroad_rfl sector, the emissions were spatially allocated to gas station locations. For the oil and gas sources in the np_oilgas sector, the spatial surrogates were updated to those shown in Table 3-11 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 2011NEIv1 data for the oil and gas sector. The data sources included 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, 2014).

Table 3-11. 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

Not all of the available surrogates are used to spatially allocate sources in the modeling platform; that is, some surrogates shown in Table 3-10 were not assigned to any SCCs, although many of the "unused" surrogates are actually used to "gap fill" other surrogates that are assigned. 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. Additional information on U.S. Spatial Surrogates, including total CAP emissions for each spatial surrogate, is available in Section 3.4.1 of the 2011v6.1 Platform TSD (EPA, 2014a).

3.3.7.3 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 emissions as point sources. For the modeling platform, EPA used the SMOKE "area-to-point" approach for only airport ground support equipment (nonroad sector), and jet refueling (nonpt sector). The approach is described in detail in the 2002 platform documentation: http://www.epa.gov/scram001/reports/Emissions%20TSD%20Vol1_02-28-08.pdf.The ARTOPNT file that lists the nonpoint sources to locate using point data was unchanged from the 2005-based platform.

3.3.7.4 Surrogates for Canada and Mexico Emission Inventories

The surrogates for Canada to spatially allocate the 2006 Canadian emissions are unchanged from the 2007 platform. The spatial surrogate data came from Environment Canada, along with cross references. Over 100 surrogates were provided and were outputs from the Surrogate Tool (previously referenced), although only about 40 were used in the modeling platform. Per Environment Canada, the surrogates are based on 2001 Canadian census data.

The 2011 platform uses about 20 surrogates for Mexico became. The surrogates are circa 1999 and 2000 and were based on data obtained from the Sistema Municpal de Bases de Datos (SIMBAD) de INEGI and the Bases de datos del Censo Economico 1999. Most of the CAPs allocated to the Mexico and Canada surrogates are shown in Table 3-12. The entries in this table are for the othar sector except for the MEX Total Road Miles and The CAN traffic rows, which

are for the othon sector.

Table 3-12. CAPs Allocated to Mexican and Canadian Spatial Surrogates

Srg code	Description	NH$_3$	NO$_X$	PM$_{2\_5}$	SO$_2$	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

Srg code	Description	NH₃	NOₓ	PM 2 5	SO₂	VOC
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
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

Srg code	Description	NH$_3$	NO$_X$	PM$_{2.5}$	SO$_2$	VOC
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

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

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

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

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

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

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.

ERG, 2006. *Mexico National Emissions Inventory, 1999: Final*, prepared by Eastern Research Group for Secratariat of the Environment and Natural Resources and the National Institute of Ecology, Mexico, October 11, 2006. Available at:

http://www.epa.gov/ttn/chief/net/mexico/1999_mexico_nei_final_report.pdf

ERG, 2009. *Development of Mexico National Emissions Inventory Projecions for 2008, 2012, and 2023.* Available at: http://www.azdeq.gov/environ/air/plan/download/dmneip.pdf

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

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 http://www.epa.gov/otaq/models/nmim/420r05024.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 http://www.epa.gov/ttn/chief/software/speciate/speciate4/documentation/speciatedoc_1206.pdf.

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: http://www.epa.gov/ttn/chief/net/2008inventory.html#inventorydoc.

EPA, 2014a. 2011 Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 6.1, 2011 Emissions Modeling Platform. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at http://www.epa.gov/ttn/chief/emch/index.html#2011.

EPA, 2014b. 2011 National Emissions Inventory, version 1 Technical Support Document. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Available at http://www.epa.gov/ttn/chief/net/2011inventory.html#inventorydoc.

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

Kochera, A., 1997. "Residential Use of Fireplaces," Housing Economics, March 1997, 10-11. Also see: http://www.epa.gov/ttnchie1/conference/ei10/area/houck.pdf.

LADCO, 2012. "Regional Air Quality Analyses for Ozone, $PM_{2.5}$, and Regional Haze: Base C Emissions Inventory (September 12, 2011)". Lake Michigan Air Directors Consortium,

Rosemont, IL 60018. Available at:
http://www.ladco.org/tech/emis/basecv8/Base_C_Emissions_Documentation_Sept_12.pdf

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. Oak Ridge National Laboratory, 2009. Analysis of Fuel Ethanol Transportation Activity and Potential Distribution Constraints. U.S. Department of Energy, March 2009. Docket No. EPA–HQ–OAR–2010–0133.

Ottmar, R.D.; Sandberg, D.V.; Bluhm, A. 2003. Biomass consumption and carbon pools. Poster. In: Galley, K.E.M., Klinger, R.C.; Sugihara, N.G. (eds.) Proceedings of Fire Ecology, Prevention, and Management. Misc. Pub. 13, Tallahassee, FL: Tall Timbers Research Station.

Ottmar, R.D.; Prichard, S.J.; Vihnanek, R.E.; Sandberg, D.V. 2006. Modification and validation of fuel consumption models for shrub and forested lands in the Southwest, Pacific Northwest, Rockes, Midwest, Southeast, and Alaska. Final report, JFSP Project 98-1-9-06.

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

Raffuse, S., N. Larkin, P. Lahm, Y. Du, 2012. Development of Version 2 of the Wildland Fire Portion of the [2011] National Emissions Inventory. International Emission Inventory Conference, Tampa, FL. Available at: http://www.epa.gov/ttn/chief/conference/ei20/session2/sraffuse.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: http://getbluesky.org/smartfire/docs.cfm

Russell, A.G. and G.R. Cass, 1986. *Verification of a Mathematical Model for Aerosol Nitrate and Nitric Acid Formation and Its Use for Control Measure Evaluation*, Atmospheric Environment, 20: 2011-2025.

SESARM, 2012a. "Development of the 2007 Base Year and Typical Year Fire Emission Inventory for the Southeastern States", Air Resources Managers, Inc., Fire Methodology, AMEC Environment and Infrastructure, Inc. AMEC Project No.: 6066090326, April,

2012

SESARM, 2012b. "Area and Nonroad 2007 Base Year Inventories. Revised Final Report", Contract No. S-2009-06-01, Prepared by Transystems Corporation, January 2012. .

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. Available at: http://www.epa.gov/ttn/chief/conferences.html

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

Wolf, et al, 2009. *Developing Mexico National Emissions Inventory Projections for the Future Years of 2008, 2012, and 2030*. Available at: http://www.epa.gov/ttnchie1/conference/ei18/session2/wolf.pdf

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.

Zue, Henze, et al, 2013. "Constraining U.S. Ammonia Emissions using TES Remote Sensing Observations and the GEOS-Chem adjoint model", Journal of Geophysical Research: Atmospheres, 118: 1-14.

4.0   CMAQ Air Quality Model Estimates

4.1   Introduction to the CMAQ Modeling Platform

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

Historically, air quality models have addressed individual pollutant issues separately. However, many of the same precursor chemicals are involved in both ozone and aerosol (particulate matter) chemistry; therefore, the chemical transformation pathways are dependent. Thus, modeled abatement strategies of pollutant precursors, such as volatile organic compounds (VOC) and NOx to reduce ozone levels, may exacerbate other air pollutants such as particulate matter. To meet the need to address the complex relationships between pollutants, EPA developed the Community Multiscale Air Quality (CMAQ) modeling system[19.] 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

---

[19] 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.

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 nested 12 km x 12 km domain covered the Eastern or Western United States. The CMAQ simulation performed for this 2011 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.[20] 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 http://www.epa.gov/asmdnerl/Research/RIA/cmaq.html 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 degredation. 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 PM2.5 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

---

[20] U.S. EPA (2014), Draft Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5, and Regional Haze, pp 214. http://www.epa.gov/ttn/scram/guidance/guide/Draft_O3-PM-RH_Modeling_Guidance-2014.pdf.

summarized in Simon et al., 2012[21].

4.2   CMAQ Model Version, Inputs and Configuration

This section describes the air quality modeling platform used for the 2011 CMAQ simulation. A modeling platform is a structured system of connected modeling-related tools and data that provide a consistent and transparent basis for assessing the air quality response to changes in emissions and/or meteorology. A platform typically consists of a specific air quality model, emissions estimates, a set of meteorological inputs, and estimates of "boundary conditions" representing pollutant transport from source areas outside the region modeled. We used the CMAQ model as part of the 2011 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 2011 CMAQ simulation along with the results of a model performance evaluation in which the 2011 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 2011 analysis employed CMAQ version 5.0.2[22] which reflects updates to version 5.0.1 which include several changes to the science algorithms to improve the underlying science. The CMAQ simulation included parameterizations to estimate the vertical distribution nitrogen oxide emissions generated due to lightning as well as to estimate bi-directional ammonia flux. The CMAQ model version 5.0 was most recently peer-reviewed in June of 2011 for the U.S. EPA.[23] 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

[21] 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.

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

[23] 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, http://www.epa.gov/AMD/Reviews/2011_CMAQ_Review_FinalReport.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.

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

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)

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

The 12 km CMAQ modeling domain was modeled for the entire year of 2011. The 2011 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

2011 Emissions: The emissions inventories used in the 2011 air quality modeling are described in Section 3, above.

Meteorological Input Data: The gridded meteorological data for the entire year of 2011 at the 12 km continental United States scale domain was derived from version 3.4[24] of the Weather Research and Forecasting Model (WRF), Advanced Research WRF (ARW) core.[25] 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 2011 WRF simulation included the physics options of the Pleim-Xiu land surface model (LSM), Asymmetric Convective Model version 2 planetary boundary layer (PBL) scheme, Morrison double moment microphysics, Kain- Fritsch cumulus parameterization scheme utilizing the moisture-advection trigger[26] and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme.[27] In addition, the Group for High Resolution Sea Surface Temperatures (GHRSST)[28] 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.[29]

The WRF meteorological outputs were processed using the Meteorology-Chemistry Interface

[24] Version 3.4 was the current version of WRF at the time the 2011 meteorological model simulation was performed.
[25] 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.
[26] Ma, L-M. and Tan, Z-M, 2009. Improving the behavior of the Cumulus Parameterization for Tropical Cyclone Prediction: Convection Trigger. Atmospheric Research 92 Issue 2, 190-211.
http://www.sciencedirect.com/science/article/pii/S0169809508002585
[27] Gilliam, R.C., Pleim, J.E., 2010. Performance Assessment of New Land Surface and Planetary Boundary Layer Physics in the WRF-ARW. Journal of Applied Meteorology and Climatology 49, 760-774.
[28] Stammer, D., F.J. Wentz, and C.L. Gentemann, 2003, Validation of Microwave Sea Surface Temperature Measurements for Climate Purposes, J. Climate, 16, 73-87.
[29] 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.

Processor (MCIP) package[30], version 4.1.3, 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 2011 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 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 2011 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

[30] 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.

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
	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-CHEM[31] 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 2011 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.[32] A GEOS-Chem evaluation was conducted for the purpose of validating the 2011 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.[33] More information is available about the GEOS-CHEM

---

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

[32] 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.

[33] 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.

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 2011 simulation using state/local monitoring sites data in order to estimate the ability of the CMAQ modeling system to replicate the 2011 base year concentrations for the 12 km continental U.S. domain.

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

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

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

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

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

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

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

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

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

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 2011 in the continental U.S. were included in the evaluation and were taken from the 2011 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. domain[34]: 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 predicted by the 2011 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 over predicted with the greatest over prediction in Central U.S. (NMB is 15.7 percent) at AQS and CASTNet sites. However, 8-hour ozone is slightly underpredicted in the West at CASTNet sites (NMB is -4.5 percent). Ozone performance in spring shows better performance with slight under predictions in most of the subregions except in the Southeast (slight over prediction of 5.4 at AQS and 0.8 at CASTNet sites). In the winter, when concentrations are generally low, the model slightly under predicts 8-hour ozone with the exception of the West at AQS sites (NMB is 6.3). In the fall, when concentrations are also relatively low, ozone is slightly over predicted; with NMBs less than 11 percent in each subregion.

Model bias at individual sites during the ozone season is similar to that seen on a subregional basis for the summer. Figure 4-2 shows the mean bias for 8-hour daily maximum ozone greater than 60 ppb is generally ±4 ppb across the AQS and CASTNet sites. Likewise, the information in Figure 4-4 indicates that the bias for days with observed 8-hour daily maximum ozone greater than 60 ppb is within ± 20 percent at the vast majority of monitoring sites across the U.S. domain. The exceptions are sites along the California coast, St. Paul, WI, and Cleveland, OH. At these sites observed concentrations greater than 60 ppb are generally predicted in the range of ±20 to 60 percent. Looking at the map of bias, Figure 4-4 indicates that the higher or lower bias at these sites is not evident at other sites in these same areas. This suggests that the under prediction

---

[34] 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.

at these sites is likely due to very local features (e.g., meteorology and/or emissions) and not indicative of a systematic problem in the modeling platform. Model error, as seen from Figures 4-3 and 4-5, is 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, Salem, OR, and North Dakota.

Table 4-4. Summary of CMAQ 2011 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	7,445	-2.8	5.4	-9.1	17.6
		Spring	14,741	-0.5	5.5	-1.2	12.9
		Summer	15,835	1.7	7.4	3.5	15.4
		Fall	12,697	6.7	3.4	10.9	20.9
	CASTNet	Winter	1,188	-3.1	4.8	-9.2	14.1
		Spring	1,160	-1.3	5.2	-2.9	11.6
		Summer	1,217	1.5	6.4	3.2	14.0
		Fall	1,295	3.2	5.7	9.5	17.1
Midwest	AQS	Winter	2,894	-1.6	5.0	-5.7	17.6
		Spring	11,329	0.2	5.8	0.5	13.5
		Summer	15,887	2.3	7.8	4.7	15.6
		Fall	9,482	2.9	6.6	6.1	14.2
	CASTNet	Winter	1,015	-1.6	4.6	-4.8	14.2
		Spring	962	-1.0	4.8	-2.2	10.6
		Summer	905	0.0	6.0	0.1	11.9
		Fall	979	2.2	5.2	6.1	14.2
Central States	AQS	Winter	11,632	-0.2	5.4	-0.7	16.7
		Spring	16,667	1.8	6.2	-3.7	12.9
		Summer	17,459	7.6	11.3	15.7	23.5
		Fall	15,496	1.7	6.4	3.9	14.6
	CASTNet	Winter	660	-2.8	5.7	-7.6	15.6
		Spring	634	-1.8	6.2	-3.7	12.9
		Summer	645	1.7	7.5	3.3	14.7
		Fall	612	-0.2	4.82	-0.6	11.0
Southeast	AQS	Winter	5,741	-0.1	5.2	-0.4	14.1
		Spring	19,791	2.5	6.4	5.4	13.8

Subregion	Monitor Network	Season	No. of Obs	MB (ppb)	ME (ppb)	NMB (%)	NME (%)
		Summer	22,093	6.7	10.2	13.5	20.7
		Fall	17,050	3.6	6.7	9.0	16.9
	CASTNet	Winter	1,756	-0.9	4.6	-2.5	12.7
		Spring	1,824	0.4	5.6	0.8	11.9
		Summer	1,777	4.4	8.0	9.0	16.1
		Fall	1,827	1.9	5.5	4.8	13.7
	AQS	Winter	25,735	2.2	6.7	6.3	19.4
		Spring	29,951	-1.8	5.9	-3.6	12.0
		Summer	32,485	1.0	7.6	1.9	14.8
		Fall	28,901	3.4	7.0	8.0	16.5
West							
	CASTNet	Winter	1,614	-1.6	5.6	-3.8	12.9
		Spring	1,649	-4.5	6.5	-8.4	12.2
		Summer	1,654	-2.6	7.2	-4.5	12.7
		Fall	1,609	0.6	5.6	1.3	11.9
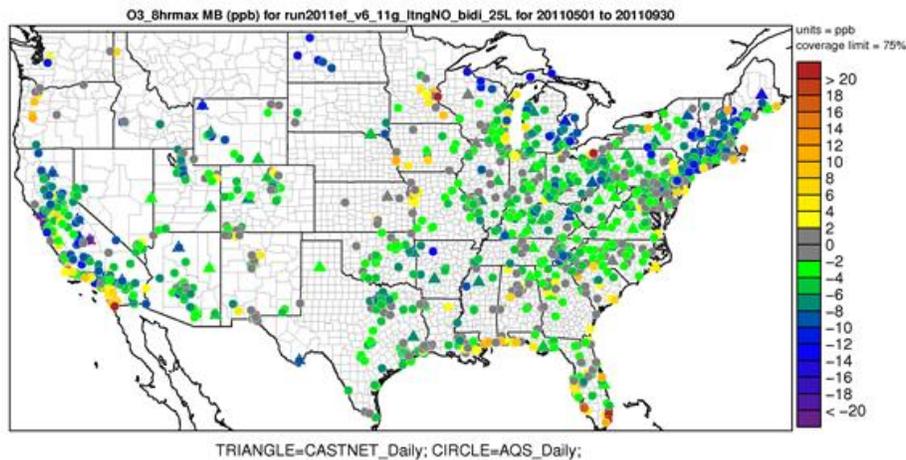
Figure 4-2. Mean Bias (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2011 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.
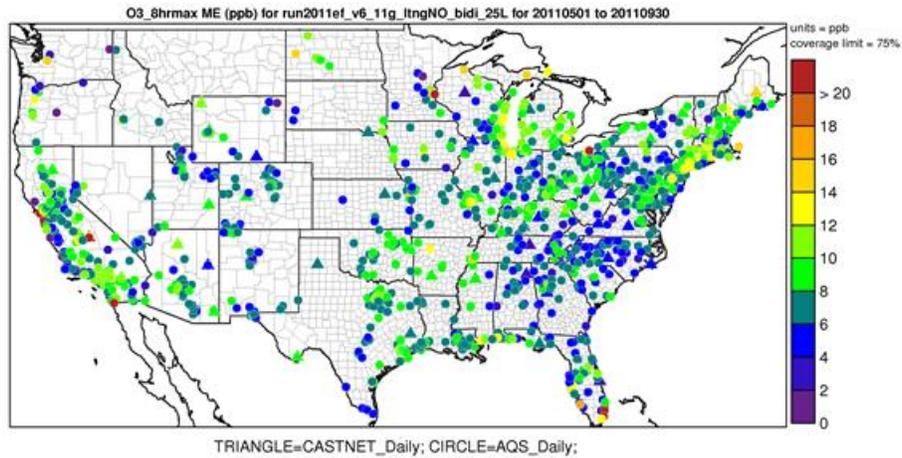
Figure 4-4. Mean Error (ppb) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2011 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.
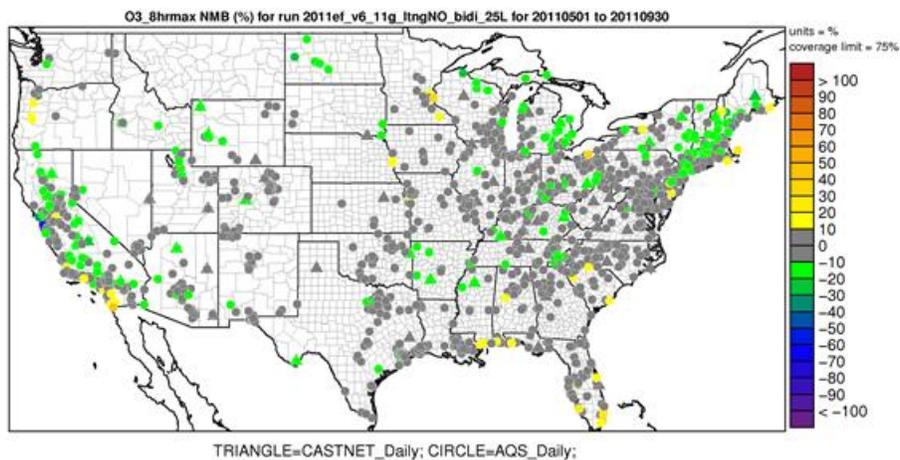
Figure 4-3. Normalized Mean Bias (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2011 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

77

O3_8hrmax NME (%) for run 2011ef_v6_11g_ltngNO_bidi_25L for 20110501 to 20110930

units = %
coverage limit = 75%

TRIANGLE=CASTNET_Daily; CIRCLE=AQS_Daily;
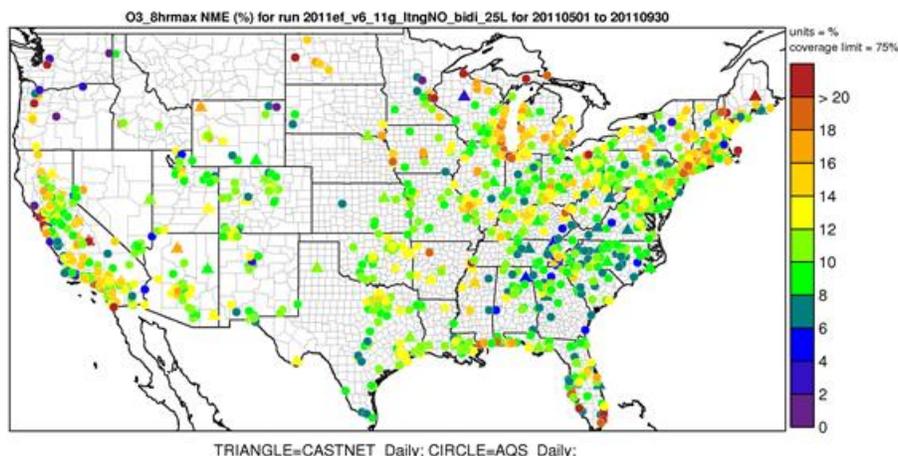
Figure 4-5. Normalized Mean Error (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2011 at AQS and CASTNet monitoring sites in the continental U.S. modeling domain.

Evaluation for Annual PM$_{2.5}$ components: The PM evaluation focuses on PM$_{2.5}$ components including sulfate (SO$_4$), nitrate (NO$_3$), total nitrate (TNO$_3$ = NO$_3$ + HNO$_3$), ammonium (NH$_4$), elemental carbon (EC), and organic carbon (OC). The bias and error performance statistics were calculated on an annual basis for each subregion (Table 4-5). PM$_{2.5}$ measurements for 2011 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 MBB values ranging from -0.1 to -0.7 μgm^{-3} and NMB values ranging from near negligible to -31 percent. Sulfate performance shows moderate error, ranging from 23 to 47 percent. Figures 4-6 through 4-9, suggest spatial patterns vary by region. The model bias for most of the Southeast, Central and Southwest states are within 0 to -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 most of the subregions (NMB in the range of near negligible to -41 percent), except in the Southeast where nitrate is over predicted on average by 17.5 percent. At IMPROVE rural sites, annual

average nitrate is over predicted at all subregions, except in the Central and Western U.S. where nitrate is under predicted. Model performance of total nitrate at sub-urban CASTNet monitoring sites also shows an over prediction across all subregions (NMB in the range of near negligible to 27 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 near negligible to -34 percent). Ammonium performance across the urban CSN sites shows an over prediction (ranging from 5 to 24 percent) in the Northeast, Midwest, and Southeast, and an under prediction (ranging from 4 to 29 percent in the Central and 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 4 to 39 percent), except in the Central and Southeast U.S. where the bias is 2 to 19 percent respectively. 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 2011 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	2,898	-0.1	0.7	-4.5	30.8
		Midwest	2,448	-0.3	0.8	-13.1	33.1
		Southeast	2,584	-0.4	0.8	-14.1	30.5
		Central	1,643	-0.2	0.7	-10.8	34.3
		West	2,544	-0.1	0.4	-8.4	45.5

Pollutant	Monitor Network	Subregion	No. of Obs	MB (μgm^{-3})	ME (μgm^{-3})	NMB (%)	NME (%)
	IMPROVE	Northeast	1,840	0.1	0.5	4.4	35.9
		Midwest	464	-0.2	0.6	-12.0	32.9
		Southeast	1,675	-0.4	0.7	-16.4	31.5
		Central	2,203	-0.3	0.5	-16.5	34.6
		West	9,518	0.0	0.3	-2.1	46.7
	CASTNet	Northeast	746	-0.3	0.5	-15.3	22.7
		Midwest	566	-0.6	0.7	-25.5	27.8
		Southeast	1,094	-0.7	.08	-26.6	29.0
		Central	369	-0.6	0.6	-31.0	34.0
		West	1,028	-0.1	0.2	-20.3	36.5
Nitrate	CSN	Northeast	2,898	0.4	0.8	40.4	73.4
		Midwest	2,448	0.3	0.8	47.4	6.8
		Southeast	2,584	0.3	0.6	56.5	99.8
		Central	1,643	0.0	.5	-3.6	53.1
		West	2,544	-0.5	1.2	-29.0	67.0
	IMPROVE	Northeast	1,839	0.4	0.5	113.0	155.0
		Midwest	464	0.1	0.5	11.7	64.1
		Southeast	1,675	0.2	0.4	42.6	112.0
		Central	1,643	0.0	0.7	-0.2	47.1
		West	9,484	0.0	0.2	-13.2	97.5
Total Nitrate (NO_3 + HNO_3)	CASTNet	Northeast	746	0.4	0.6	26.5	39.7
		Midwest	566	0.0	0.6	1.6	23.5
		Southeast	1,094	0.0	0.6	0.7	39.1
		Central	369	-0.4	0.6	-20.6	34.3
		West	1,028	-0.2	0.4	-26.2	47.4
Ammonium	CSN	Northeast	2,898	0.2	0.4	24.3	46.8
		Midwest	2,448	0.1	.4	5.3	38.7
		Southeast	2,584	0.1	0.3	14.5	44.6
		Central	1,643	0.0	0.4	-4.0	42.1

Pollutant	Monitor Network	Subregion	No. of Obs	MB (µgm⁻³)	ME (µgm⁻³)	NMB (%)	NME (%)
		West	2,544	-0.2	0.4	-29.2	67.5
	CASTNet	Northeast	746	0.0	0.2	0.9	27.5
		Midwest	566	-0.2	0.3	-15.9	27.1
		Southeast	1,094	-0.1	0.2	-14.9	27.0
		Central	369	0.1	0.2	-20.3	34.5
		West	1,028	-0.1	0.1	-33.6	54.5
	CSN	Northeast	2,750	0.5	0.6	73.2	90.6
		Midwest	2,426	0.5	.6	80.3	91.4
		Southeast	2,513	0.3	0.5	54.5	74.2
		Central	1,625	0.6	0.7	121.0	129.0
		West	2,510	0.7	0.8	92.8	111.0
Elemental Carbon							
	IMPROVE	Northeast	1,916	0.2	0.2	70.9	87.0
		Midwest	474	0.1	0.1	40.6	57.8
		Southeast	1,784	0.1	0.2	26.6	55.7
		Central	2,331	0.1	0.2	44.2	68.5
		West	10,058	0.1	0.2	80.2	117.0
	CSN	Northeast	2,735	0.4	0.9	38.9	84.1
		Midwest	2,412	0.8	1.1	53.9	70.9
		Southeast	2,500	0.5	1.1	25.2	54.1
		Central	1,622	1.2	1.6	71.5	95.8
		West	2,493	1.9	2.4	93.7	118.0
Organic Carbon							
	IMPROVE	Northeast	1,911	0.4	0.9	38.9	84.1
		Midwest	473	0.1	0.7	12.3	61.2
		Southeast	1,785	-0.3	0.9	-18.7	49.4
		Central	2,332	0.0	0.7	-2.2	55.4
		West	9,921	0.0	0.5	4.3	73.5
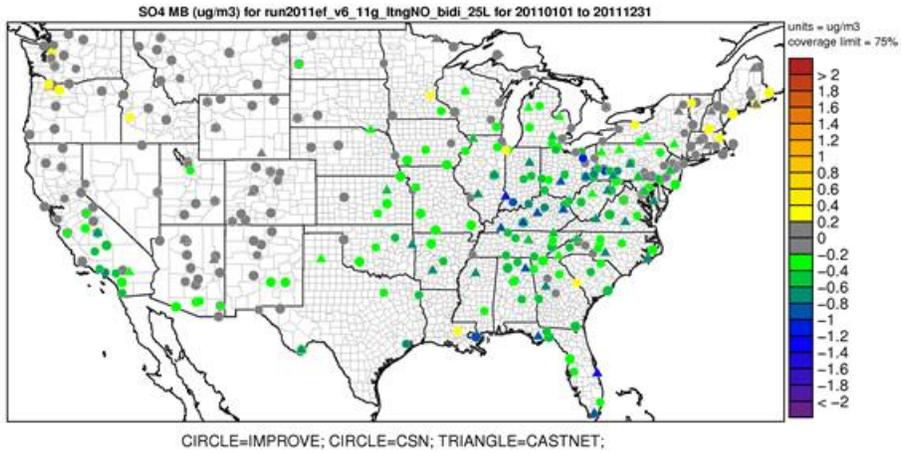
**Figure 4-6. Mean Bias (μgm⁻³) of annual sulfate at monitoring sites in the continental U.S. modeling domain.**
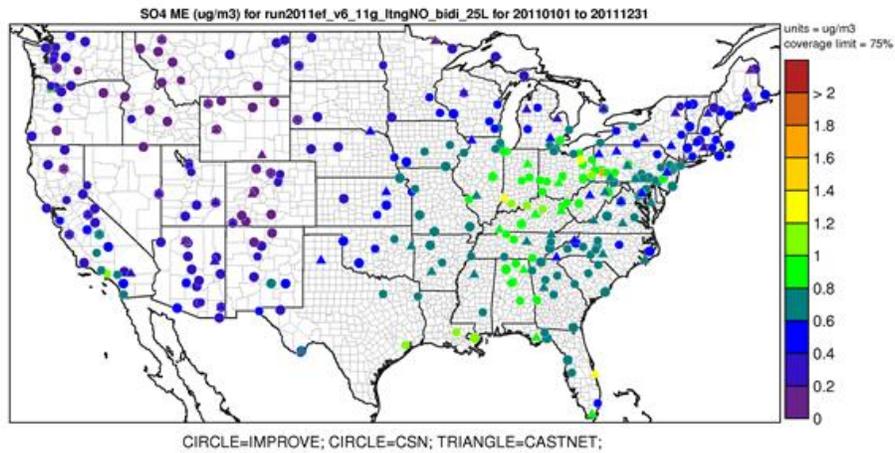
Figure 4-7. Mean Error (μgm⁻³) of annual sulfate at monitoring sites in the continental U.S. modeling domain.
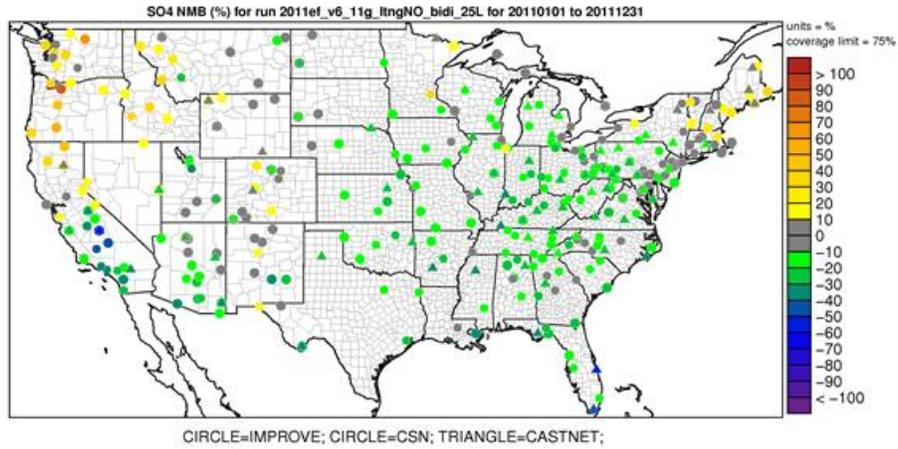
**Figure 4-8. Normalized Mean Bias (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.**
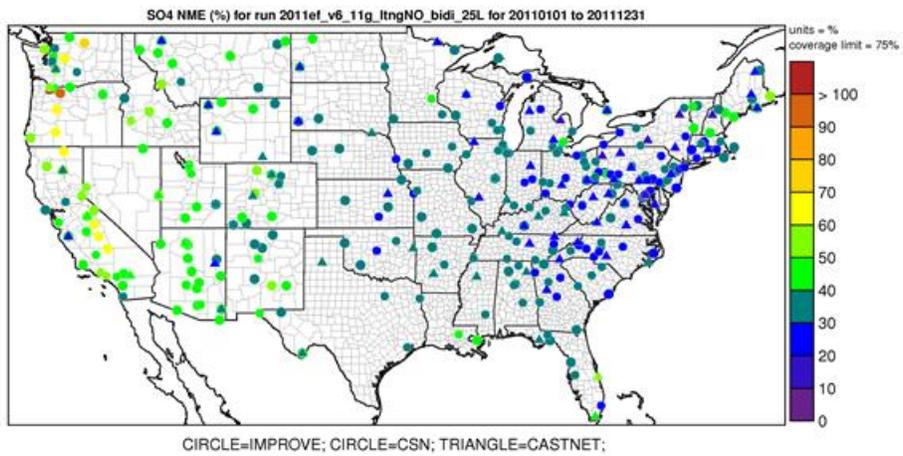

Figure 4-9. Normalized Mean Error (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.
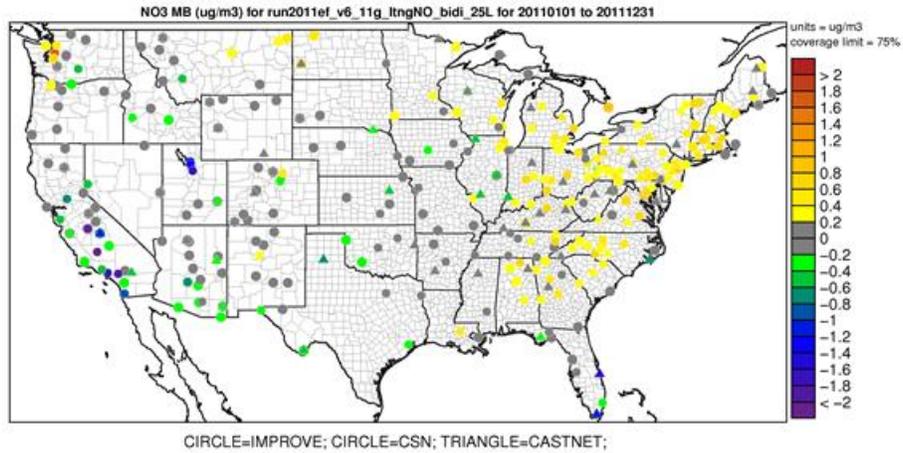
Figure 4-10. Mean Bias (µgm⁻³) of annual nitrate at monitoring sites in the continental U.S. modeling domain.
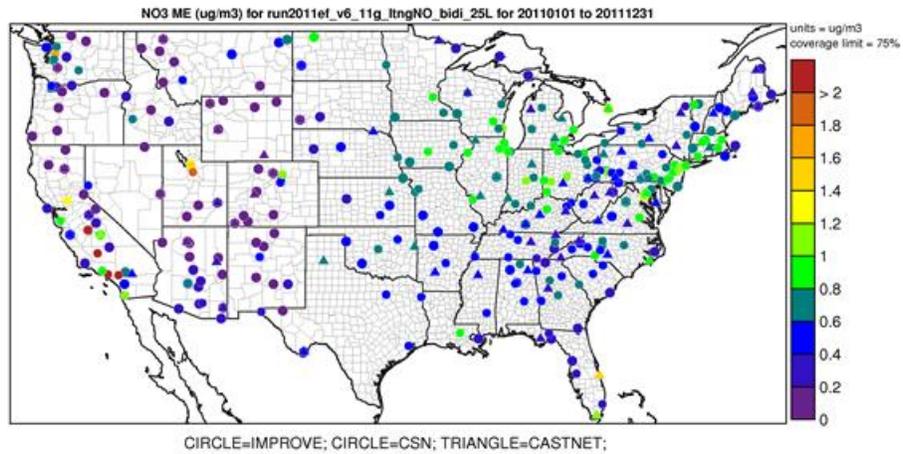
Figure 4-11. Mean Error (µgm⁻³) of annual nitrate at monitoring sites in the continental U.S. modeling domain.
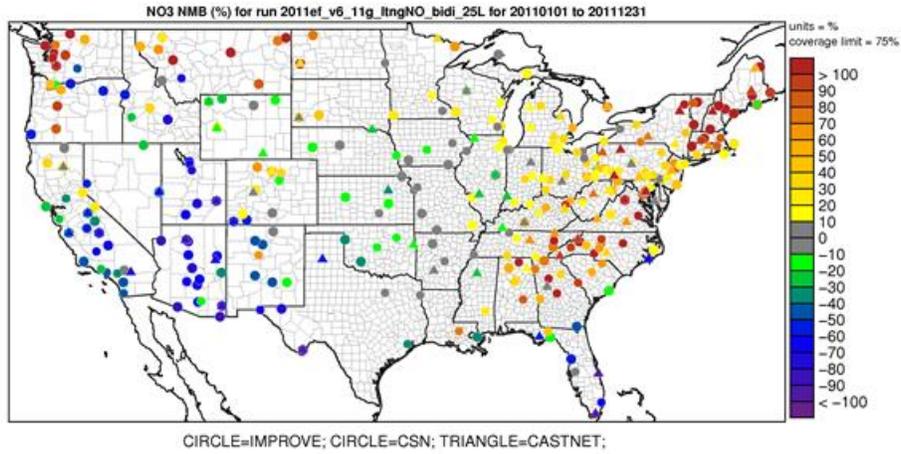
Figure 4-12. Normalized Mean Bias (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.
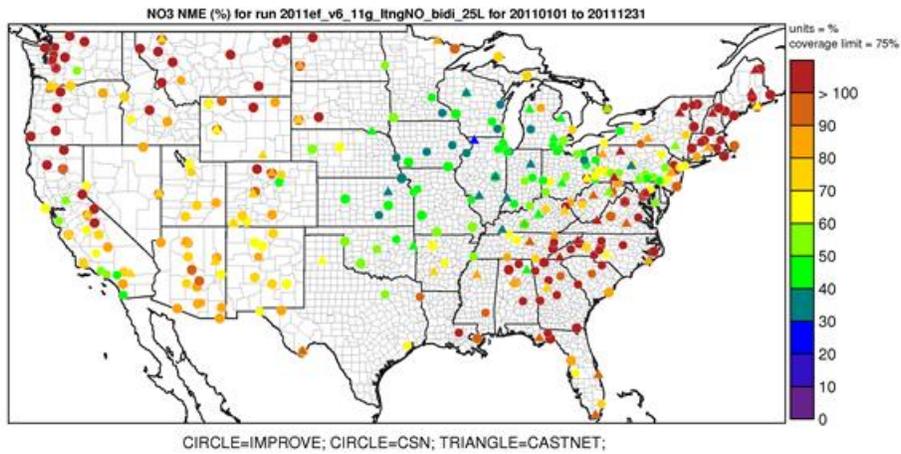
Figure 4-13. Normalized Mean Error (%) of annual nitrate at monitoring sites in the continental U.S. modeling domain.
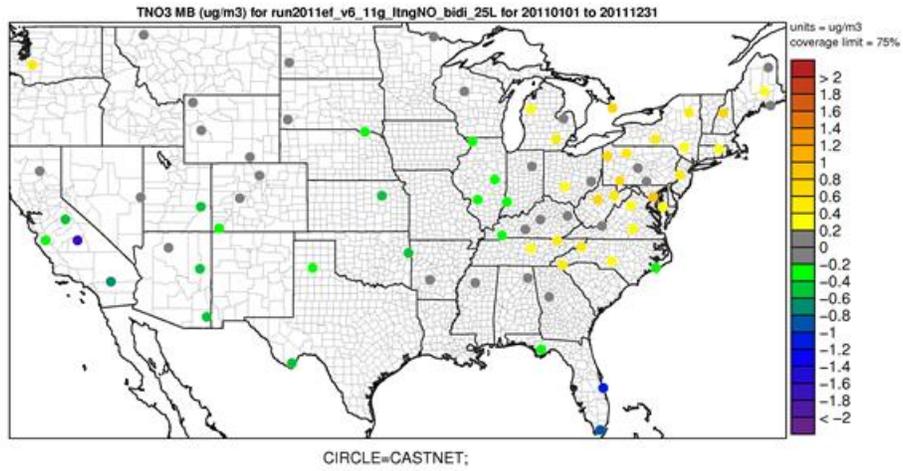
Figure 4-14. Mean Bias (μgm⁻³) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.
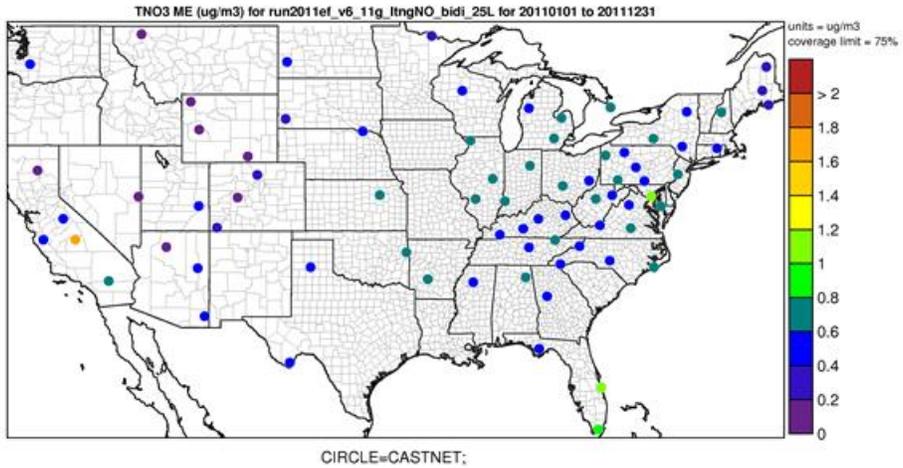

Figure 4-15. Mean Error (μgm⁻³) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.
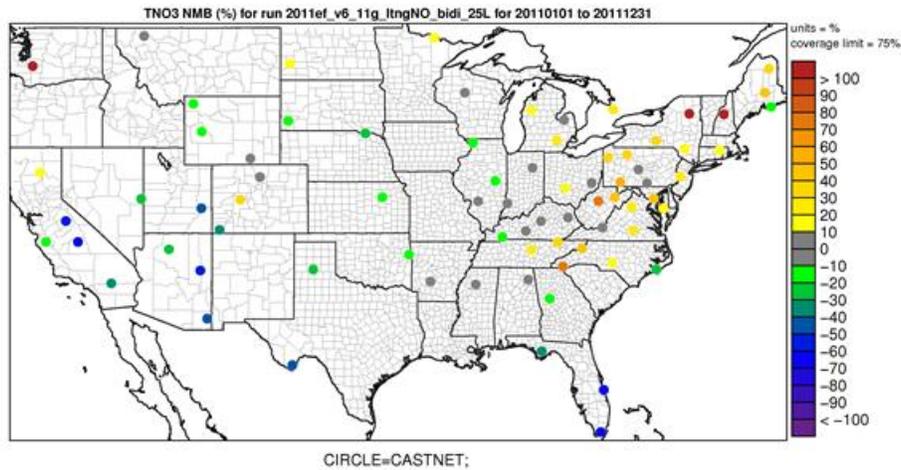
Figure 4-16. Normalized Mean Bias (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.
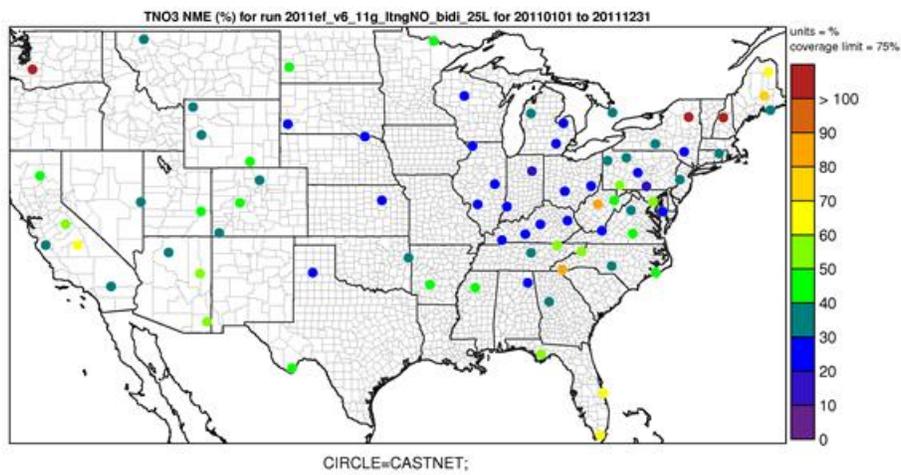

Figure 4-17. Normalized Mean Error (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.
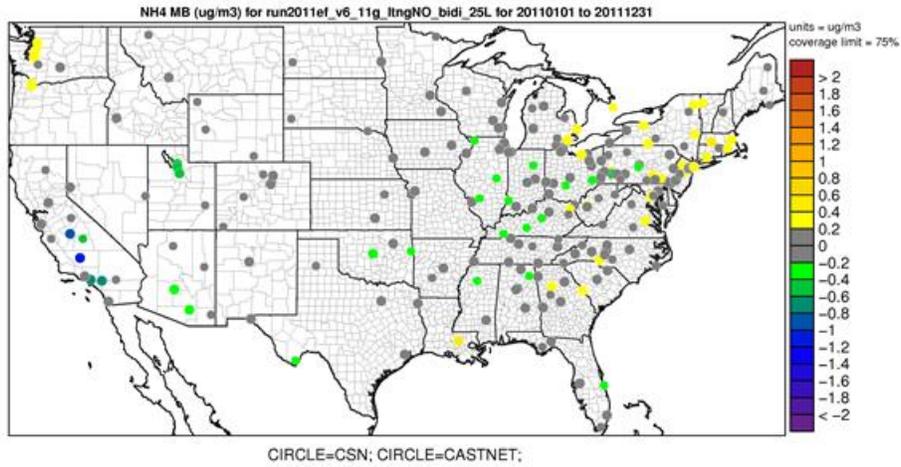
Figure 4-18. Mean Bias (**μgm⁻³**) of annual *ammonium* at monitoring sites in the continental U.S. modeling domain.
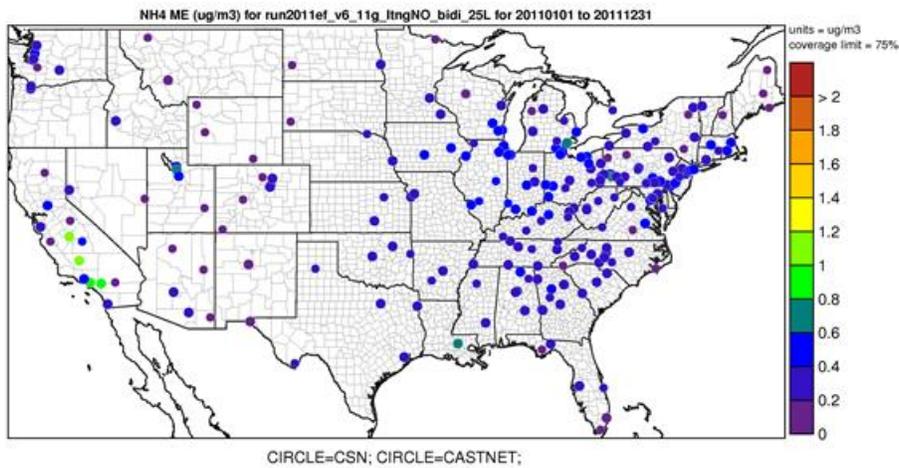
Figure 4-19. Mean Error (**μgm⁻³**) of annual ammonium at monitoring sites in the continental U.S. modeling domain.
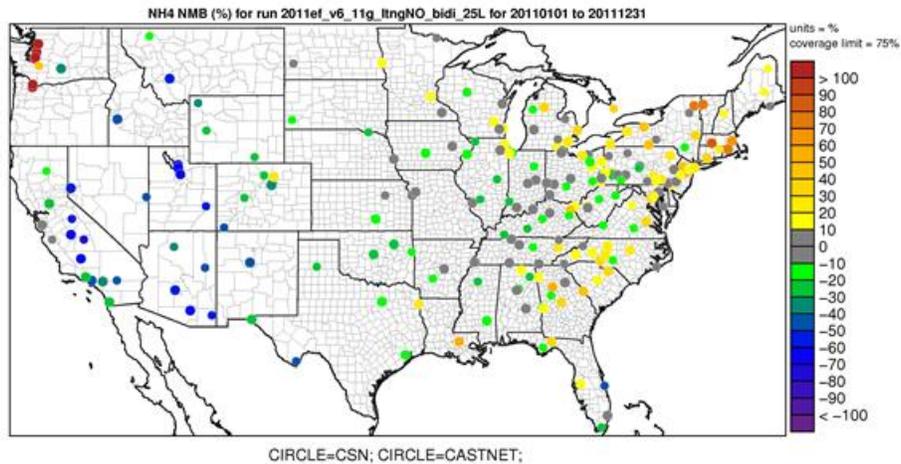
Figure 4-20. Normalized Mean Bias (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.
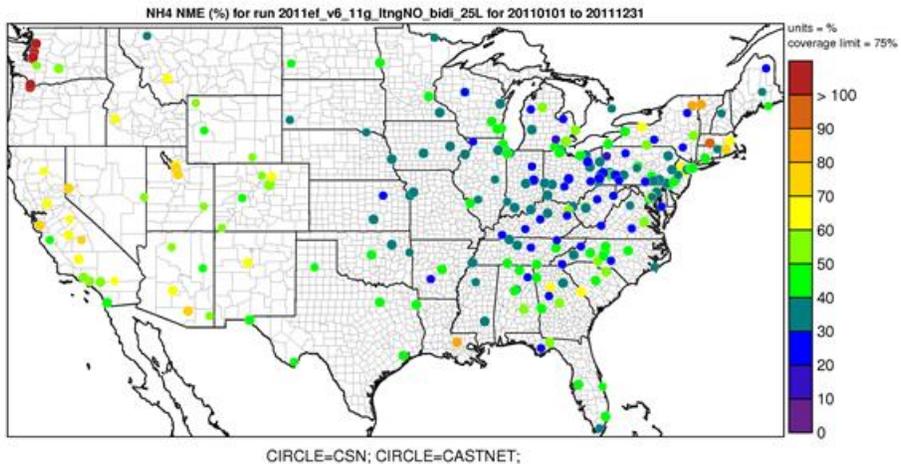
**Figure 4-21. Normalized Mean Error (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.**
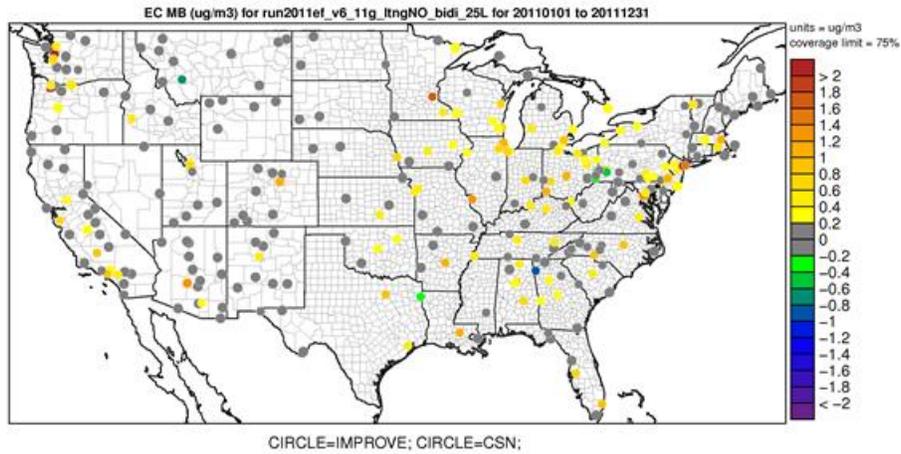
Figure 4-22. Mean Bias (μgm⁻³) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.
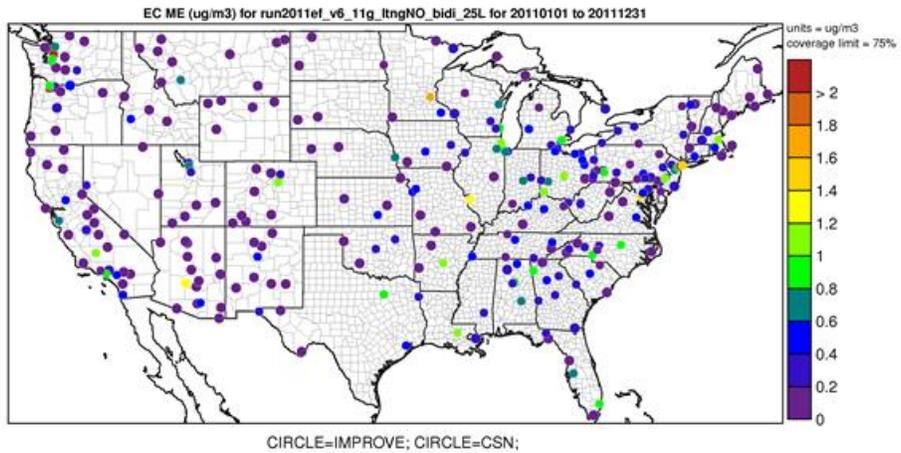

Figure 4-23. Mean Error (μgm⁻³) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.
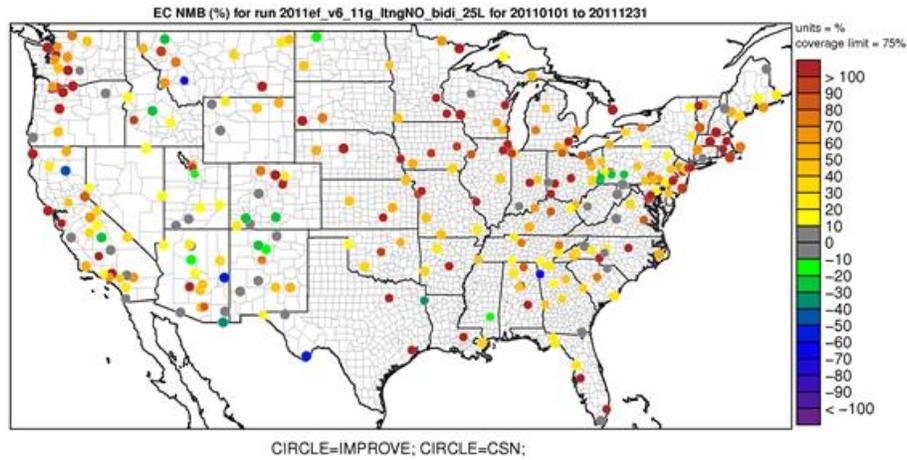
Figure 4-24. Normalized Mean Bias (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.
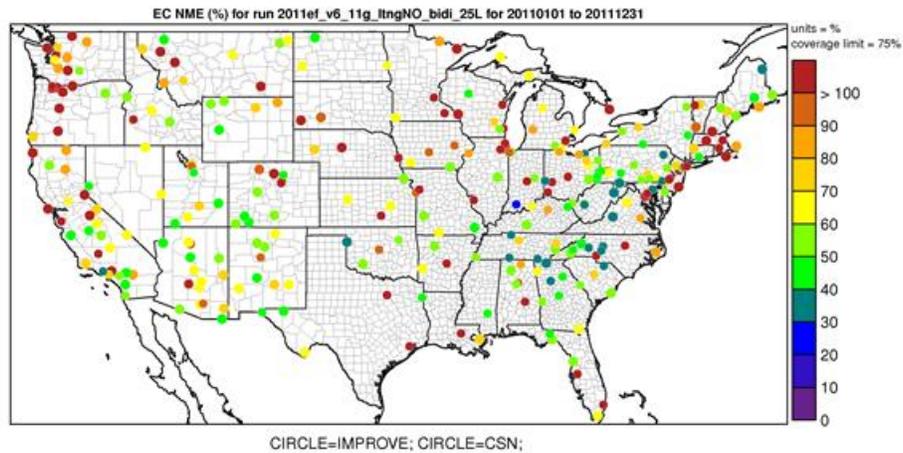
Figure 4-25. Normalized Mean Error (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.
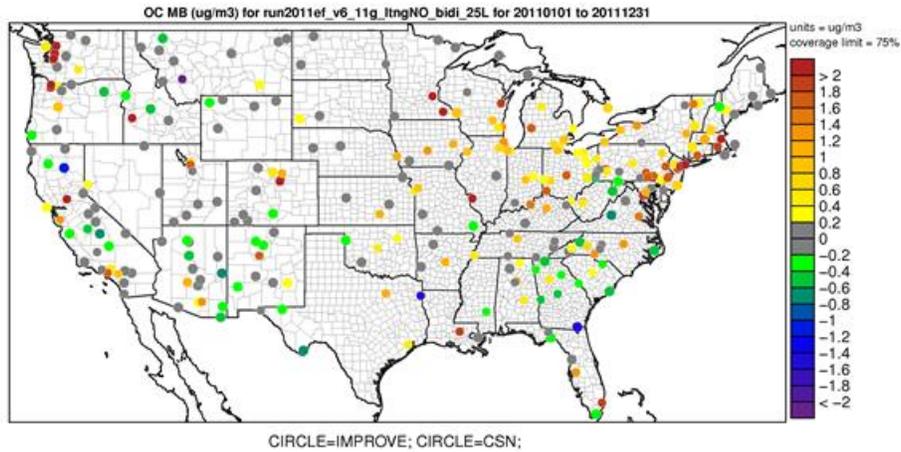
Figure 4-26. Mean Bias (μgm⁻³) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.
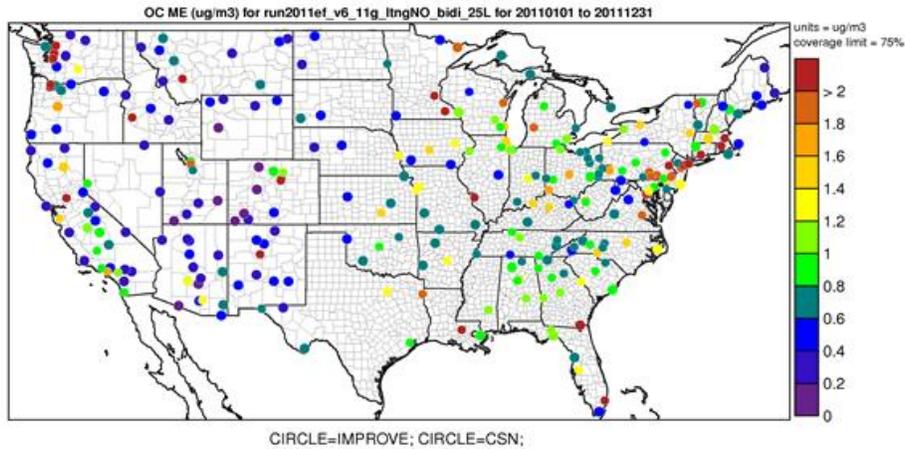
Figure 4-27. Mean Error (μgm⁻³) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.
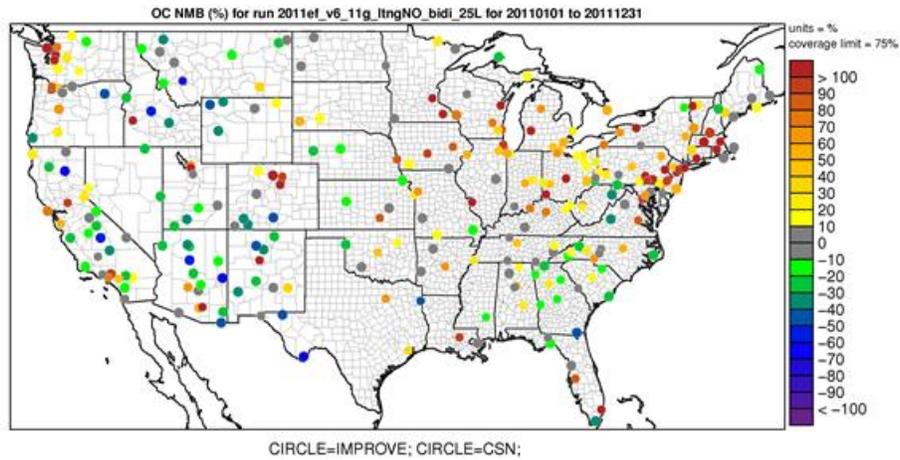
Figure 4-28. Normalized Mean Bias (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.
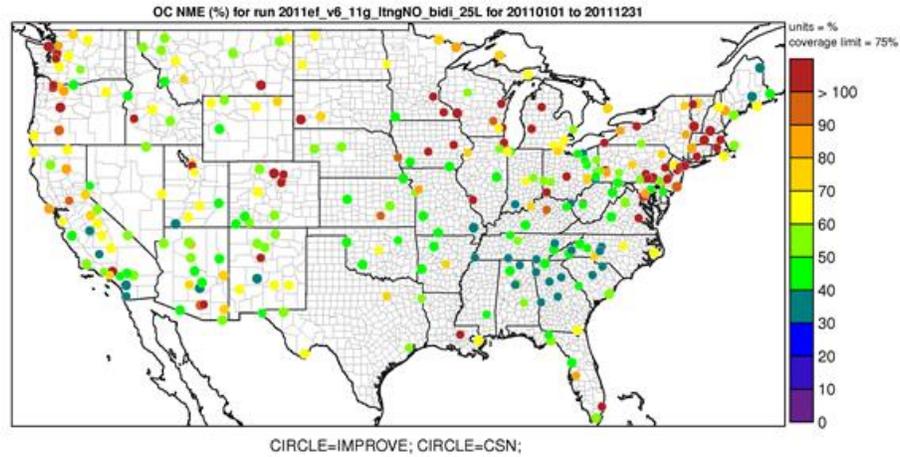

Figure 4-29. Normalized Mean Bias (%) of annual organic carbon at monitoring sites in the continental U.S. modeling domain.

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. 2011 census tract centroids for the year 2011.

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 $\sim\beta_0(s,t)$, $\beta_1(t)$, $\sim x(s,t)$) 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 2011 US census tract centroids across the continental U.S. using 2011 measurement and CMAQ data as inputs. For ozone, the concentration unit is the daily maximum 8-hour average in ppb and for $PM_{2.5}$ the concentration unit is the 24-hour average in $\mu g/m^3$.

5.3.1   *Summary of 8-hour Ozone Results*

Figure 5-1 summarizes the AQS, CMAQ and DS ozone data over the year 2011. 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 2011 over half of the US Census tracts (42455 out of 72283) experienced at least one day with an ozone value above the NAAQS of 75 ppb.
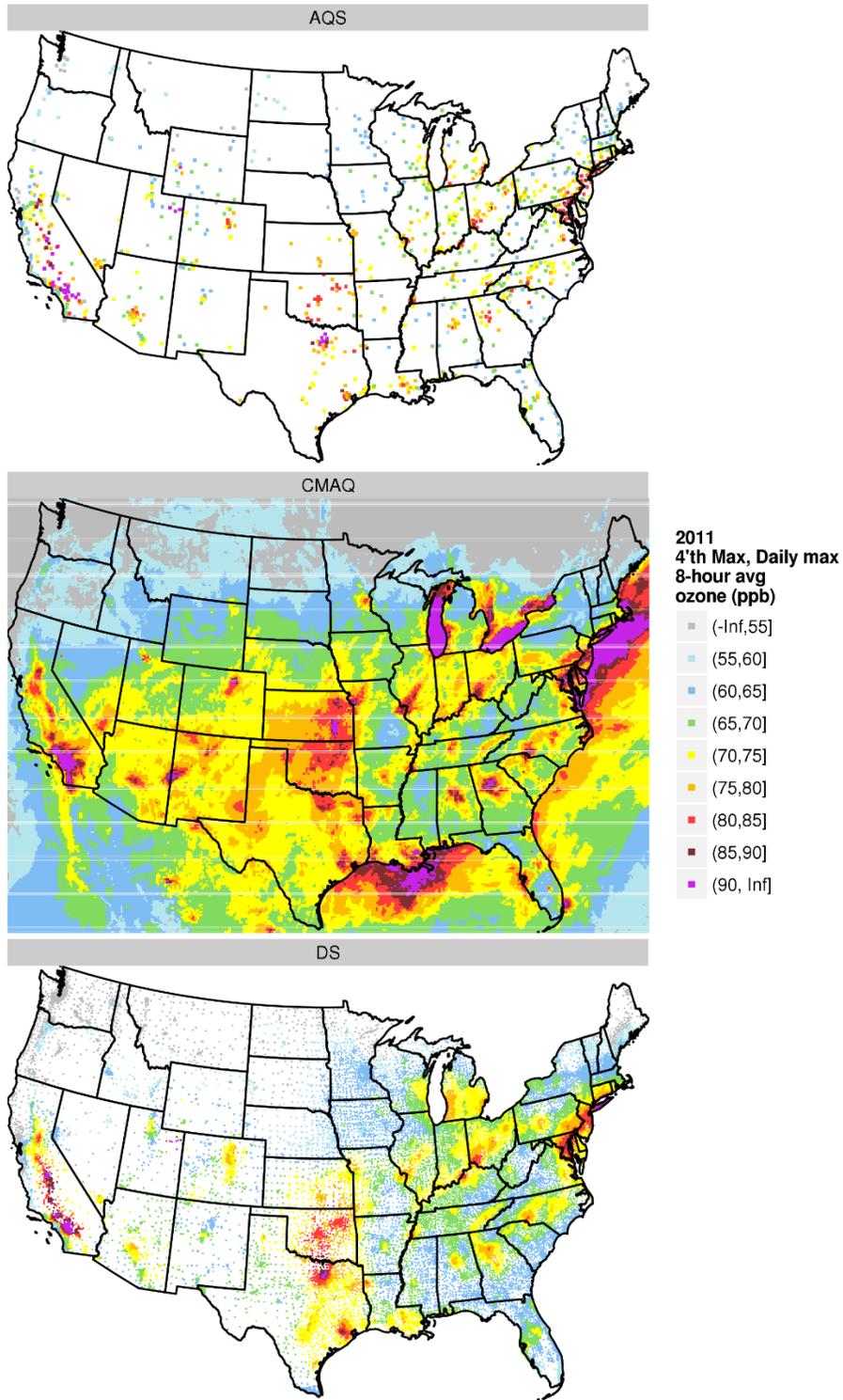
Figure 5-1. Annual 4th max (daily max 8-hour ozone concentrations) derived from AQS, CMAQ and DS data.

5.3.2 Summary of PM$_{2.5}$ Results

Figures 5-2 and 5-3 summarize the AQS, CMAQ and DS PM2.5 data over the year 2011. Figure 5-2 shows annual means and Figure 5-3 shows 98th percentiles of 24-hour PM2.5 concentrations for AQS observations, CMAQ model predictions and DS model results. The DS model estimated that for 2011 about 30% of the US Census tracts (21336 out of 72283) experienced at least one day with a PM2.5 value above the 24-hour NAAQS of 35 ug/m3.
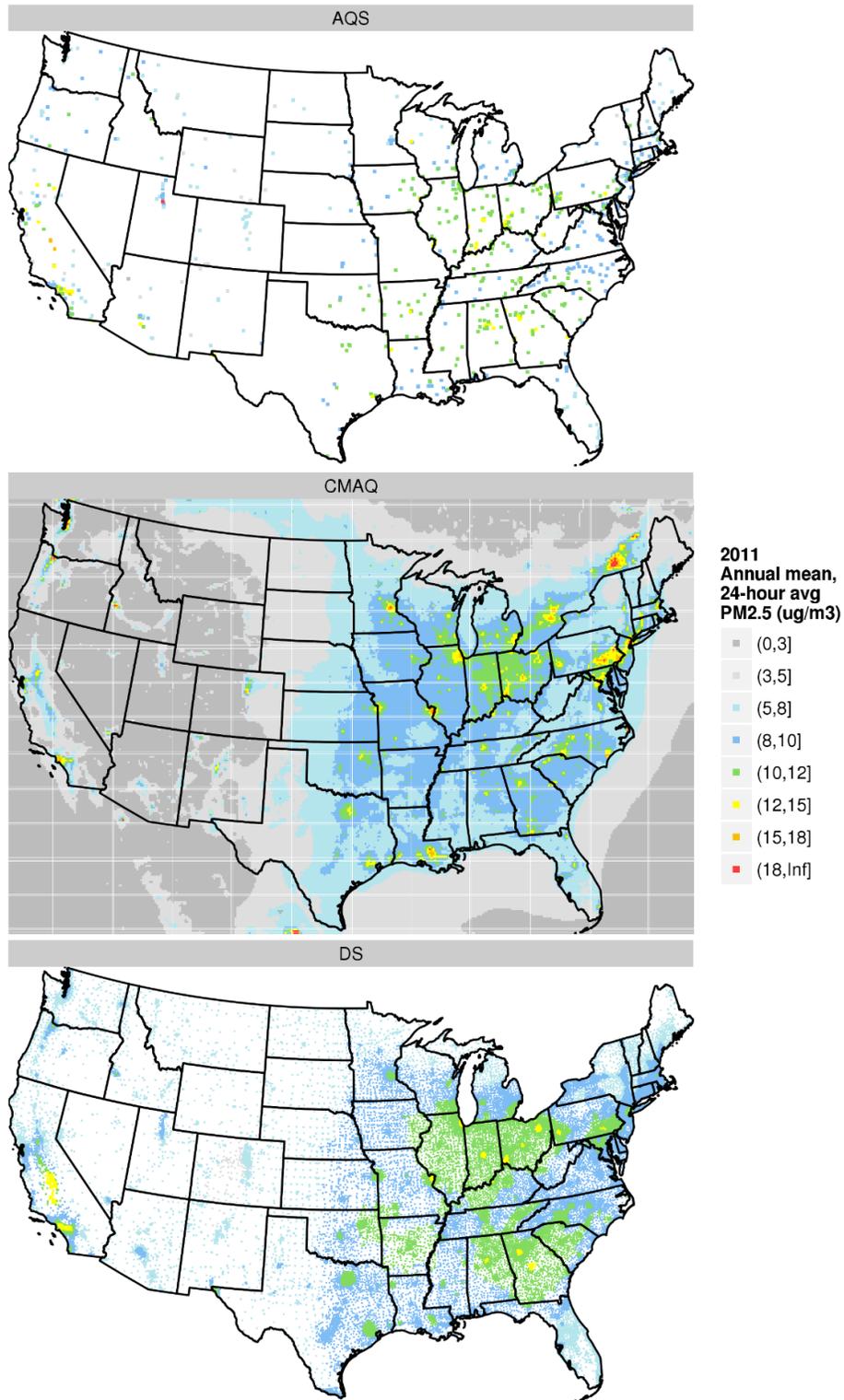
**Figure 5-2. Annual mean PM2.5 concentrations derived from AQS, CMAQ and DS data.**
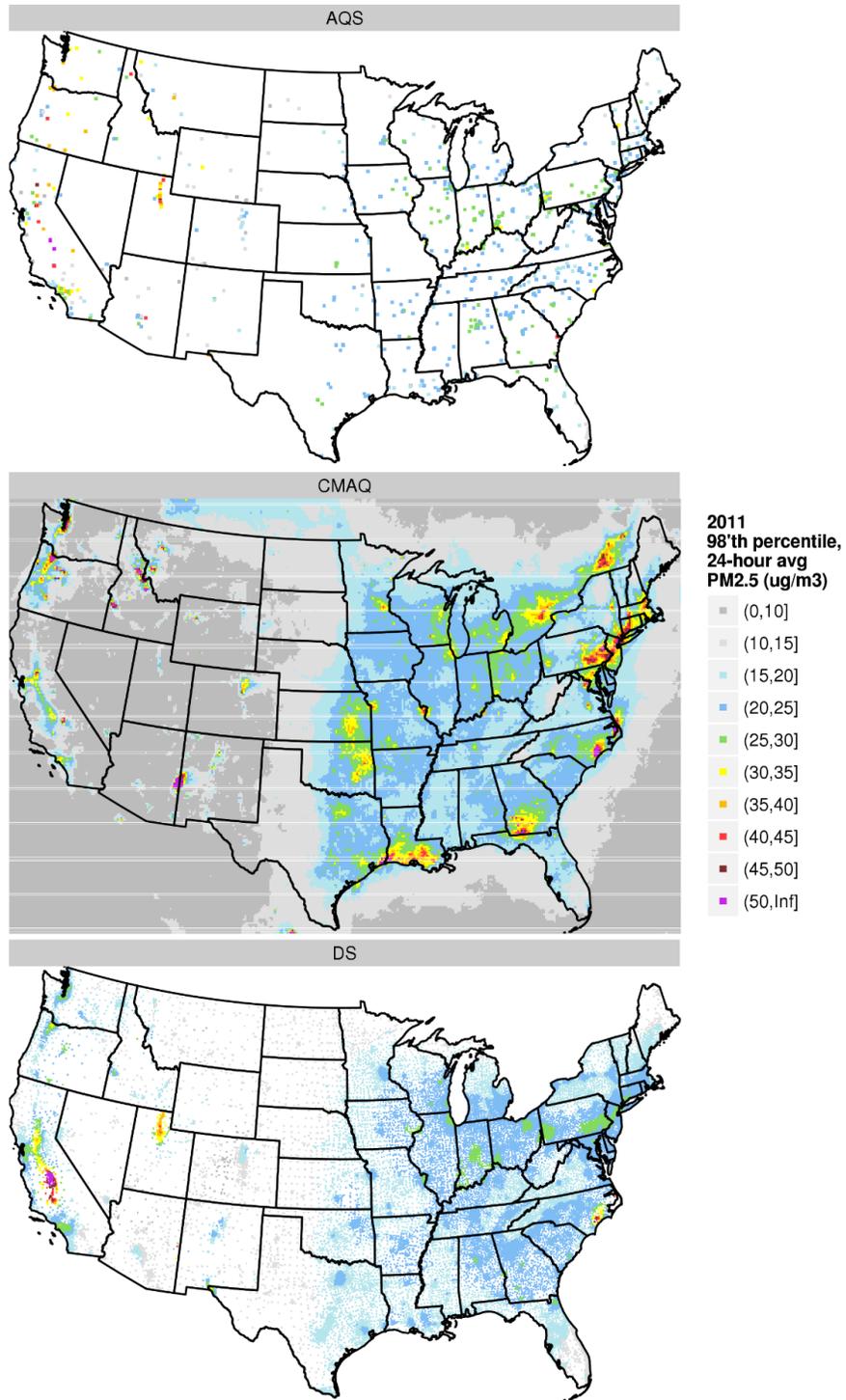
Figure 5-3. 98th percentile 24-hour average PM2.5 concentrations derived from AQS, CMAQ and DS data.

5.4   Downscaler Uncertainties

5.4.1   Standard Errors

As mentioned above, the DS model works by drawing random samples from built-in distributions during its parameter estimation. The standard errors associated with each of these populations provide a measure of uncertainty associated with each concentration prediction. Figure 5-4 shows the percent errors resulting from dividing the DS standard errors by the associated DS prediction. The black dots on the maps show the location of EPA sampling network monitors whose data was input to DS via the AQS datasets (Chapter 2). The maps show that, in general, errors are relatively smaller in regions with more densely situation monitors (ie the eastern US), and larger in regions with more sparse monitoring networks (ie western states). These standard errors could potentially be used to estimate the probability of an exceedance for a given point estimate of a pollutant concentration.
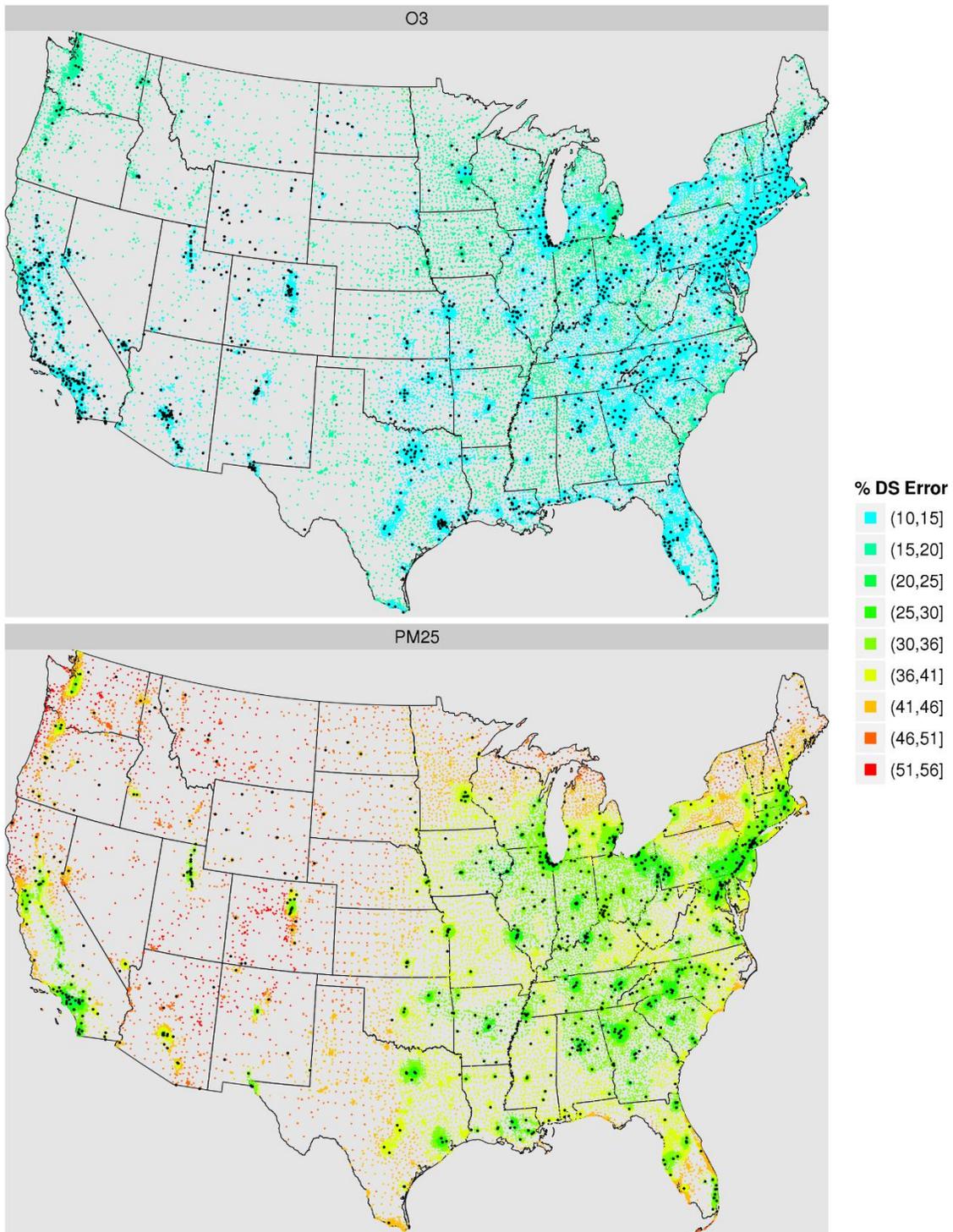
Figure 5-4. Annual mean relative errors (standard errors divided by predictions) from the DS 2011 runs. The black dots show the locations of monitors that generated the AQS data used as input to the DS model.

5.4.2   Cross Validation

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

Pollutant	# Monitors	Mean Bias	RMSE	Mean Coverage
PM2.5	830	3.933e-2	2.82	0.96
O3	1290	1.096e-2	4.87	0.95

Table 5-1. Cross-validation statistics associated with the 2011 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 2011 U.S. census tract centroids by utilizing monitoring data and CMAQ output for 2011. 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