EPA United States Environmental Protection Agency

Bayesian space-time downscaling fusion model (downscaler) -Derived Estimates of Air Quality for 2010

Bayesian space-time downscaling fusion model (downscaler) -Derived Estimates of Air Quality for 2010

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 2010 calendar year generated by EPA's recently developed data fusion method termed the "downscaler model" (DS). Air quality monitoring data from the National Air Monitoring Stations/State and Local Air Monitoring Stations (NAMS/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.

---

[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.

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

---

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

The best available statistical fusion model, air quality data, and CMAQ numerical model output were used to develop the 2010 estimates. Fusion results can vary with different inputs and fusion modeling approaches. As new and improved statistical models become available, EPA will provide updates.

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

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 non- attainment 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 (as of October 2008) NAAQS for ozone 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 $PM_{2.5}$ 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 NAAQSs for fine particles ($PM_{2.5}$) in 1997. The NAAQSs 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).**
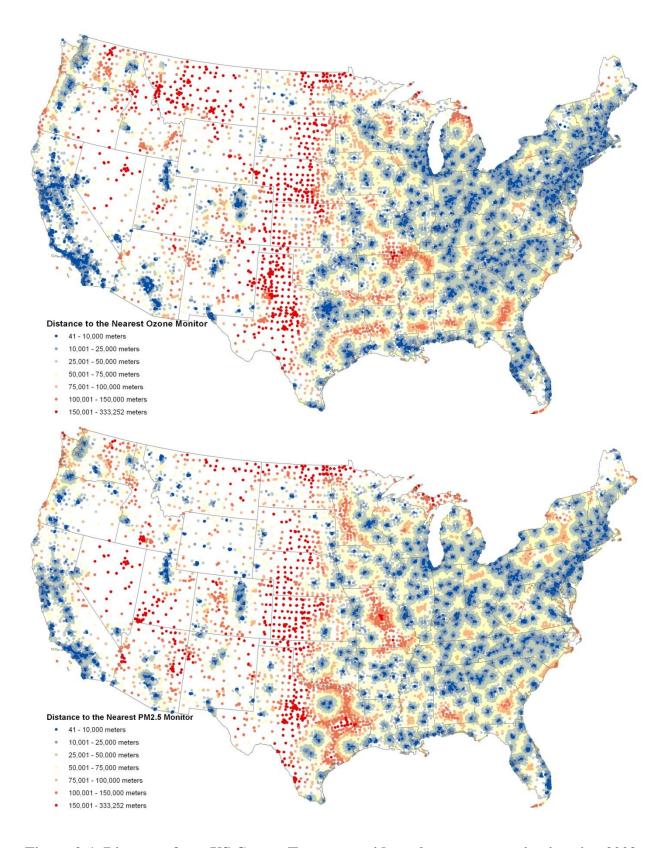
Figure 2-1. Distances from US Census Tract centroids to the nearest monitoring site, 2009.

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 users can 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 and NAMS).

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 2008 National Emissions Inventory (NEI), Version 2 to process year 2010 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 to10 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 2010 emission inputs for this study included development of emission inventories for a 2010 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 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 3.1 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   2010 Emission Inventories and Approaches

This section describes the emissions inventories created for input to SMOKE. The 2008 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 provided by S/L/T agencies because they are a voluntarily

submitted to the NEI. The 2008 NEI was compiled using the Emissions Inventory System (EIS). EIS includes hundreds of automated QA checks to help improve data quality, and also supports release point (stack) coordinates separately from facility coordinates. Improved EPA collaboration with S/L/T agencies prevented duplication between point and nonpoint source categories such as industrial boilers. Documentation for the 2008 NEI is available at http://www.epa.gov/ttn/chief/net/2008inventory.html#inventorydoc.

2010-specific data submitted by S/L/T agencies was used for point sources that were large enough to be required to submit emissions to the NEI every year. These are sources with the potential to emit 2500 tons per year of SO_2, NO_x, or CO, or 250 tons per year of the other CAPs besides lead. 2010 continuous emissions monitoring (CEM) data was used where it was available. For fires, EPA used the SMARTFIRE2 system to develop 2010 emissions. SMARTFIRE2 was the first system to categorize all fires as either prescribed burning or wildfire categories, and it also includes improved emission factor estimates for prescribed burning. 2010-specific data were also developed for onroad, nonroad, and large commercial marine sources. Some data obtained from regional planning organizations (RPOs) was substituted for NEI data in cases for which the RPO data was more recently collected. California-provided mobile source emissions were used. Canadian emissions reflect year 2006, Mexico emissions reflect year 2008 as projected from the 1999 inventory, and offshore emissions reflect year 2008.

The methods used to process emissions for this project are very similar to those documented for EPA's Version 5, 2007 Emissions Modeling Platform. 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#pmnaaqs. Electronic copies of inventories similar to those used for this project are available in the same section of the EMCH.

The emissions modeling process, performed using SMOKE v3.1 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., NO_x and VOC) are split into the chemical species needed by CMAQ. For the purposes of preparing the CMAQ- ready emissions, the broader NEI emissions inventories are split into emissions modeling "platform" sectors; and biogenic emissions are added along with emissions from other sources other than the NEI, such as the Canadian, Mexican, and offshore inventories. The significance of an emissions sector for the emissions modeling platform is that emissions for that sector are run through all of the SMOKE programs, except the final merge, independently from emissions in the other sectors. The final merge program called Mrggrid combines the sector-specific gridded, speciated and temporalized emissions to create the final CMAQ-ready emissions inputs.

Table 3-1 presents the sectors in the emissions modeling platform used to develop 2010 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 2010 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_2010_emissions_totals_by_sector.xlsx".

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

2010 Platform Sector (*Abbrev*)	2010 NEI Sector	Description and resolution of the data input to SMOKE
IPM (*ptipm*)	Point	2010 NEI point source EGUs that can be mapped to the Integrated Planning Model (IPM) model. NEI values are replaced with year 2010 hourly continuous emission monitoring (CEM) NO_X and SO_2 emissions, where available. Other pollutants are scaled from 2008 NEI using heat input.
Point non-IPM (*ptnonipm*)	Point	A mix of 2008 NEI point source emissions with some 2010 records where data was provided by states and locals plus 2006 WRAP oil and gas data; these are emissions not matched to the ptipm sector, annual resolution, including all aircraft emissions
Point source fire (*ptfire*)	Fires	Point source day-specific wildfires and prescribed fires for 2010.
Agricultural (*ag*)	Nonpoint	2008 NEI nonpoint NH_3 emissions from livestock and fertilizer application; county and annual resolution with some 2007 monthly resolution data provided by the Midwest.
Area fugitive dust (*afdust*)	Nonpoint	2008 NEI nonpoint PM_{10} and $PM_{2.5}$ from fugitive dust sources (e.g., building construction, road construction, paved roads, unpaved roads, agricultural dust), county and annual resolution. A land use-based transport fraction and 2010-based precipitation zero-out is applied.
Remaining nonpoint (*nonpt*)	Nonpoint	Primarily 2008 NEI nonpoint for sources not included in other sectors, plus 2006 WRAP oil and gas data, county and annual resolution.
Nonroad (*nonroad*)	Nonroad	Year 2010 monthly nonroad emissions from the National Mobile Inventory Model (NMIM) plus California-provided data; county and annual resolution.
C1 and C2 marine and locomotive (*c1c2rail*)	Nonroad	Year 2008 non-rail maintenance locomotives, and category 1 and category 2 commercial marine vessel (CMV) emissions sources; county and annual resolution; year 2010 for California.
C3 commercial marine (*c3marine*)	Nonroad	Non-NEI, year 2010 category 3 (C3) CMV emissions projected from year 2002. Developed for the rule "Control of Emissions from New Marine Compression-Ignition Engines at or Above 30 Liters per Cylinder", also known as the Emissions Control Area-International

		Maritime Organization (ECA-IMO) study: www.epa.gov/otaq/ oceanvessels.htm. (EPA-420-F-10-041, August 2010). Annual resolution, treated as point sources.
Onroad (*onroad*)	Onroad	Year 2010 gridded hourly emissions from onroad mobile gasoline and diesel vehicles from parking lots and moving vehicles including exhaust, evaporative, permeation, and brake and tire wear. Generated using MOVES 2010b emission factors, 2010 VMT and vehicle population data, and 2010 gridded met. data. In California, adjusted to match CA-provided emissions.
Onroad Refueling (*onroad_rfl)*	Onroad	Year 2010 gridded hourly refueling emissions from onroad mobile gasoline and diesel vehicles from parking lots and moving vehicles. Generated using MOVES 2010b, emission actors, 2010 VMT and vehicle population data, and 2010 gridded met. data. Spatially allocated to gasoline station locations.
Biogenic (*beis*)	Biogenic	Hour- and grid cell-specific emissions for 2010 generated from the BEIS 3.14 model, including emissions in Canada and Mexico.
Other point sources (*othpt*)	N/A	Point sources not from the NEI, including Canada's 2006 inventory and a 2008 projection of Mexico's Phase III 1999 inventory; annual resolution. Also includes 2008 offshore oil point source emissions for the U.S. from the 2008 NEI.
Other nonpoint and nonroad (*othar*)	N/A	Nonpoint and nonroad sources not from the NEI, including annual 2006 Canada sources at province resolution and a 2008 projection of annual 1999 Mexico sources at municipio resolution.
Other onroad sources (*othon*)	N/A	Onroad sources not from the NEI, including annual 2006 Canada sources at province resolution and a 2008 projection of 1999 Mexico sources at municipio resolution.

Table 3-2. 2010 Continental United States Emissions by Sector (tons/yr in 48 states + D.C.)

Sector	CO	NH₃	NOₓ	PM₁₀	PM₂.₅	SO₂	VOC
afdust				6,211,274	874,619		
ag		3,595,429					
c1c2rail	216,862	559	1,321,691	43,248	40,467	48,112	60,353
nonpt	4,336,565	155,317	1,230,624	767,225	676,243	402,633	6,456,455
nonroad	14,497,993	2,025	1,720,692	169,509	160,950	16,426	2,177,810
onroad (inc. rfl)	28,044,484	119,366	5,665,524	276,929	197,366	36,764	2,747,883
ptfire	13,170,780	216,518	197,824	1,356,023	1,149,172	104,516	3,112,451
ptipm	732,585	24,191	2,092,202	303,455	212,912	5,279,849	34,383
ptnonipm	2,684,750	75,288	1,928,047	532,559	378,408	1,372,508	1,116,417
c3marine	15,296		155,779	5,132	4,674	46,151	5,934
Con.US Total	63,699,315	4,188,694	14,312,384	9,665,354	3,694,812	7,306,960	15,711,685

Table 3-3. 2010 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	3,746,954	537,911	719,026	1,422,503	393,804	97,714	1,267,370
Canada othon	4,514,632	21,814	537,706	15,004	10,634	5,430	277,915
Canada othpt	1,148,014	21,138	861,258	117,254	68,114	1,762,350	425,792
Canada Subtotal	9,409,600	580,863	2,117,990	1,554,761	472,553	1,865,493	1,971,077
Mexico othar	477,952	132,913	199,048	88,354	56,824	56,418	510,965
Mexico othon	659,796	2,972	93,849	7,937	7,349	5,740	96,253
Mexico othpt	101,309	0	344,896	122,654	90,304	740,235	78,465
Mexico Subtotal	1,239,057	135,885	637,793	218,944	154,477	802,393	685,684
Offshore othpt	82,133	0	74,277	780	769	1,021	60,756
Canada c3marine	13,930		157,046	4,708	4,283	38,030	5,919
Offshore c3marine	83,610		960,546	49,509	45,403	380,001	35,509
2010 TOTAL	179,673	0	1,191,869	54,997	50,455	419,052	102,184

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 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). The point sources used for this study include a limited set of emissions data for 2010 collected via the NEI process, with 2008 NEI data for any sources that did not report in 2010. Note that only large sources are required to report annually as opposed to triennially. This section describes NEI point sources within the contiguous United States. The offshore oil (othpt sector), fires (ptfire) and category 3 CMV emissions (c3marine sector) are point source formatted inventories discussed later in this section. Full documentation for the development of the 2008 NEI (EPA, 2012), is posted at: http://www.epa.gov/ttn/chief/net/2008inventory.html#inventorydoc.

After removing offshore oil platforms into the othpt sector, two platform sectors were created from the remaining point sources: the EGU sector – also called the IPM sector (i.e., ptipm) and the non-EGU sector – also called the non-IPM sector (i.e., ptnonipm). This split facilitates the use of different SMOKE temporal processing and future-year projection techniques for each of these sectors. 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. Integration is discussed in detail in Section 3.3.4).

In the 2010 model evaluation case used in this study, for ptipm sector sources with CEM data that could be matched to the NEI, 2010 hourly SO_2 and NO_x emissions were used alongside annual emissions of all other pollutants allocated down to each hour. The hourly electric generating unit (EGU) emissions were obtained for SO_2 and NO_x emissions and heat input from EPA's Acid Rain Program. This data also contained heat input, which was used to allocate the annual emissions for other pollutants (e.g., VOC, $PM_{2.5}$, HCl) to hourly values. For unmatched EGU units, annual emissions were temporalized to days using multi-year averages and to hours using state-specific averages for that year.

The Non-EGU Stationary Point Sources (ptnonipm) emissions were input to SMOKE as annual emissions. The full description of how the 2008 NEI emissions were developed is provided in the NEI documentation, but a summary of their development follows:

a. 2008 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 was added from Mineral Management Services (MMS).

Note that some sources were large enough to require emissions to be reported to the NEI for 2009. The 2009 emissions were used, where 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 the state/county FIPS code.
- 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.
- Replaced stack parameters with values from the 2008 NEI where 2008 values were determined to be more realistic.
- Replaced facility emissions with 2008 NEI values where the 2010 NEI contained questionable values.

3.2.1.1 IPM Sector (ptipm)

The ptipm sector contains emissions from EGUs in the 2010 NEI point inventory that could be matched to the units found in the NEEDS database, version 4.10 (http://www.epa.gov/airmarkets/progsregs/epa-ipm/ index.html). IPM provides future year emission inventories for the universe of EGUs contained in the NEEDS database. As described below, matching with NEEDS was done (1) to provide consistency between the 2010 EGU sources and future year EGU emissions for sources which are forecasted by IPM, and (2) to avoid double counting when projecting point source emissions.

The 2010 NEI point source inventory contains emissions estimates for both EGU and non-EGU sources. When future years are modeled, IPM is used to predict the future year emissions for the EGU sources. The remaining non-EGU point sources are projected by applying projection and control factors to the base year emissions. It was therefore necessary to identify and separate into two sectors: (1) sources that are projected via IPM (i.e., the "ptipm" sector) and (2) sources that are not (i.e., "the "ptnonipm" sector). The two sectors are modeled separately in the base year as well as the future years.

A primary reason the ptipm sources were separated from the other point sources was due to the difference in the temporal resolution of the data input to SMOKE. The ptipm sector uses the available hourly CEM data via a method first implemented in the 2002 platform and still used for the 2010 platform. Hourly CEM data for 2010 were obtained from the CAMD Data and Maps website[3]. For sources and pollutants with CEM data, the actual year 2010 hourly CEM data were used. The SMOKE modeling system matches the ORIS Facility and Boiler IDs in the NEI SMOKE-ready file to the same fields in the CEM data, thereby allowing the hourly SO_2 and NO_x CEM emissions to be read directly from the CEM data file. The heat input from the hourly CEM data was used to allocate the NEI annual values to hourly values for all other pollutants from

CEM sources, because CEMs are not used to measure emissions of these pollutants.

For this project, the point source inventory was reviewed to determine whether additional matches needed to be made. Newly identified matches for CEM and NEEDS IDs were loaded into the Emissions Inventory System (EIS) so they could then be written into the modeling files. Some matches were made outside of EIS when IDs were not mapped one to one between the systems.

Emissions were scaled from 2008 levels to 2010 levels where possible based on CEM data. For sources not matching the CEM data ("non-CEM" sources), daily emissions were computed from the NEI annual emissions using a structured query language (SQL) program and state-average CEM data. To allocate annual emissions to each month, state-specific, three-year averages of 2008-2010 CEM data were created. These average annual- to-month factors were assigned to non-CEM sources by state. To allocate the monthly emissions to each day, the 2010 CEM data were used to compute state-specific month- to-day factors, which were then averaged across all units in each state. The resulting daily emissions were input into SMOKE. The daily-to-hourly allocation was performed in SMOKE using diurnal profiles. The development of these diurnal ptipm-specific profiles, considered ancillary data for SMOKE, is described in a later section.

3.2.1.2 Non-IPM Sector (ptnonipm)

The non-IPM (ptnonipm) sector contains all NEI point sources not included in the IPM (ptipm) sector except for the offshore oil and day-specific fire emissions. For the most part, the ptnonipm sector reflects the non-EGU component of the NEI point inventory; however, as previously discussed, it is likely that some small low-emitting EGUs that are not reflected in the CEMs database are present in the ptnonipm sector. The ptnonipm sector contains a small amount of fugitive dust PM emissions from vehicular traffic on paved or unpaved roads at industrial facilities or coal handling at coal mines. In previous versions of the platform, these emissions were reduced prior to input to SMOKE. However, in this platform the reduction is not made because of a new methodology used to reduce PM dust based on gridded meteorological data.

For some geographic areas, some of the sources in the ptnonipm sector belong to source categories that are contained in other sectors. This occurs in the inventory when states, tribes or local programs report certain inventory emissions as point sources because they have specific geographic coordinates for these sources. They may use point source SCCs (8-digit) or they may use non- point, onroad or nonroad (10-digit) SCCs. In the 2008 NEI, examples of these types of sources include: aircraft and ground support emissions, livestock (i.e., cattle feedlots) in California, and rail yards.

Some adjustments were made to the point inventory prior to its processing with SMOKE. These include:

- Removing sources with state county codes ending in '777'. These are used for 'portable' point sources like asphalt plants.
- Removing sources with SCCs not typically used for modeling.
- Adjusting latitude-longitude coordinates for sources identified to be substantially outside

the county in which they reside.

- Removing offshore oil records as reflected by FIPS=85000 because these sources are processed in the othpt sector.
- Added 2008 ethanol facilities provided by EPA's OTAQ that were not already included in the 2008 NEI.
- Correcting stack parameters for some units with missing or invalid parameter assignments.
- Adding South Dakota emissions because they did not submit to the 2008 NEI.
- Adding MeadWestVaco facility in Covington, VA because it was missing in the 2008 NEI.
- Adding oil and gas emissions that were not otherwise included in the NEI from the Western Regional Air Partnership (WRAP) RPO created year 2006 "Phase III" oil and gas inventory project.
- Removing onroad refueling emissions that some states included in the point sector because these are modeled nationwide using MOVES2010b.

3.2.2 Nonpoint Sources (afdust, ag, nonpt)

The nonpoint emissions sources used in this study are primarily from the 2008 NEI. Documentation for the 2008 NEI is available at http://www.epa.gov/ttn/chief/net/2008inventory.html#inventorydoc. Prior to processing with SMOKE, the nonpoint portion of the 2008 NEI was divided into the following sectors for which the data is processed in consistent ways: area fugitive dust (afdust), agricultural ammonia (ag), and the other nonpoint sources (nonpt). This section describes stationary nonpoint sources only. Class 1 & Class 2 (c1c2) and Class 3 (c3) commercial marine vessels and locomotives are also in the 2008 NEI nonpoint data category, but these sources are included in the mobile source portion of this documentation. Nonpoint tribal-submitted emissions were removed to prevent possible double counting with county-level emissions. Because the tribal nonpoint emissions are small, these omissions should not impact results at the 12-km scale used for modeling. This omission also eliminated the need to develop costly spatial surrogate data to allocate tribal data to grid cells during the SMOKE processing. Some specific types of nonpoint sources were not included in the modeling due to one of the following reasons: 1) the sources are only reported by a few states or agencies, 2) the sources are 'atypical' and small, and/or 3) there are other data available that appears to be more accurate. Additional details on nonpoint source processing can be found in the Version 5, 2007 Emissions Modeling Platform documentation discussed earlier.

In the rest of this section, each of the platform sectors into which the 2008 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 PM emission estimates for 2008 NEI nonpoint SCCs identified by EPA staff as fugitive dust sources. Categories included in this sector are paved roads, unpaved roads and airstrips, construction (residential, industrial, road and total), agriculture production and all of the mining 10-digit SCCs beginning with the digits

"2325." It does not include fugitive dust from grain elevators because these are elevated point sources.

This sector is separated from other nonpoint sectors to allow for the application of "transport fraction," and meteorology/precipitation-based reductions. These adjustments are applied via sector-specific scripts and make use of land use-based gridded transport fractions. 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 then applied to remove all emissions for days on which measureable rain occurs or there is snow on the ground. Both the transport fraction and meteorological adjustments are based on the gridded meteorological data; therefore, different emissions could result from different grid resolutions. Application of the transport fraction and meteorological adjustments reduces the overestimation of fugitive dust impacts in the grid modeling as compared to ambient samples.

3.2.2.2 Agricultural Ammonia Sector (ag)

The agricultural NH_3 "ag" sector is comprised of livestock and agricultural fertilizer application emissions from the nonpoint sector of the 2008 NEI. The livestock and fertilizer emissions were extracted based on SCC. The "ag" sector includes all of the NH_3 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. To prevent double-counting, emissions were not included in the nonpoint ag inventory for counties in which they were in the point source inventory. A significant error in the 2008 NEI was corrected in the modeling platform ag sector. A fertilizer application source "N-P-K (multi-grade nutrient fertilizers)" (SCC=2801700010) in Luna county New Mexico (FIPS=35025), was 6,953 tons of NH3 in the 2008 NEI. This source was corrected by a factor of 1,000 to be 6.953 tons in the modeling platform.

Monthly NH3 emissions provided by the Lake Michigan Air Directors Consortium were used to replace NEI ag sector emissions in that region due to the improved temporal resolution. 2008 NEI (annual) ag sector emissions were used in all other states. A new temporal allocation methodology for animal NH3 was implemented for this modeling platform that allocates monthly emissions down to the hourly level by taking into account temperature and wind speed. This method is discussed in more detail in the emission modeling portion of this chapter.

3.2.2.3 Other Nonpoint Sources (nonpt)

Stationary nonpoint sources that were not included in the afdust, ag or nonpt sectors were assigned to the "nonpt" sector. In preparing the nonpt sector, catastrophic releases were excluded since these emissions were dominated by tire burning, which is an episodic, location-specific emissions category. Tire burning accounts for significant emissions of particulate matter in some parts of the country. Because such sources are reported by a very small number of states, and are inventoried as county/annual totals without the information needed to temporally and spatially allocate the emissions to the time and location where the event occurred, catastrophic releases were excluded. All fire emissions, including agricultural, wildfire, and prescribed burning, were

removed and substituted with SMARTFIRE emissions (see the "ptfire" sector). Locomotives and CMV mobile sources from the 2008 NEI nonpoint inventory are described in the mobile sources section.

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.

Some adjustments to the 2008 NEI nonpoint data were made using data from regional planning organizations (RPOs) as follows:

- Replaced 2008 NEI oil and gas emissions (SCCs beginning with "23100") with year 2006 Phase III oil and gas emissions for several basins in the WRAP RPO states. These WRAP Phase III emissions contain point and nonpoint formatted data are discussed in greater detail at: http://www.wrapair2.org/PhaseIII.aspx. These changes were made only in counties for which there was WRAP data.
- Replaced 2008 NEI nonpoint agriculture burning emissions with year 2008 SMARTFIRE day-specific county-based emissions aggregated to monthly totals.
- Replaced open burning "land clearing" (SCC=2610000500) emissions in Florida and Georgia with SESARM-provided daily point data, but aggregated to county and monthly resolution.
- Replaced open burning data (SCCs beginning with 261000x) in MARAMA states with RPO-proved data.
- Removed industrial coal combustion emissions (SCC=2102002000) in Tennessee.
- Replaced, removed and modified much of the residential wood combustion (RWC) emissions in the MARAMA, MWRPO and SESARM states with RPO data and non-RPO corrections, modified the outdoor hydronic heater (OHH) emissions in all states and indoor furnaces in MWRPO states.
- Removed EPA-estimated commercial cooking (SCCs 2302002100 and 2302002200) duplicate PM emissions in California.
- Removed duplicate "Industrial Processes; Food and Kindred Products; Total" source (SCC=23020000000) in Maricopa county Arizona (FIPS=04013).

The oil and gas changes were already discussed in the ptnonipm section. Other significant changes are discussed below.

Ag burning
2008 NEI agricultural burning estimates were replaced with more specific data from the Fire Characteristic Classification System (FCCS) module fuel loadings map in the BlueSky

23

Framework (http://blueskyframework.org/modules/fuel-loading/fccs). Year 2008-specific fire locations from SMARTFIRE version 1 (Sullivan, et al., 2008) were read into the FCCS module and intersected with the FCCS fuel-loading dataset. The module assigned an FCCS code to each fire record that reflects the ecosystem geography and potential natural vegetation based on remote sensing data. Prescribed or unclassified fires having an FCCS code equal to zero (0) were assumed to be agricultural fires. ArcGIS was used to categorize the fires as occurring on rangeland, cropland or other land use via USGS 2006 National Land Cover Database (NLCD). Activity data were analyzed to restrict to cropland fires and assign state and crop-specific emission factors. Emissions were then appropriately weighted based on known statistics about each state's crop mix.

These SMARTFIRE-based ag burning emissions were provided in at 1km point source and day-specific resolution. State-county FIPS codes were assigned using GIS. The emissions were aggregated to county and monthly resolution and converted to SMOKE nonpoint FF10 format. This SMARTFIRE-based ag burning dataset includes emissions for all but these 7 of the lower 48 states: CT, DC, MA, ME, NH, RI and VT. These 7 states did not contain any cropland burning estimates for year 2008 based on this SMARTFIRE approach.

Open burning RPO data
All 2008 NEI open burning emissions (CAPs only) were replaced in the MARAMA states with the 2007 MARAMA open burning inventory. These MARAMA open burning emissions include estimates for household waste (SCC=2610030000), land clearing (2610000500) and yard waste leaf and brush (2610000100 and 2610000400 respectively).
The 2008 NEI land clearing emissions in Georgia and Florida were replaced with SESARM-based year-2007 data. The SESARM land clearing emissions are based on daily point emissions from the CONSUME v3.0 model (SESARM, 2012a). These daily point-format emissions were aggregated to county and monthly resolution as a separate FF10 nonpoint monthly inventory.

TN coal combustion
Tennessee nonpoint industrial coal combustion (SCC=2102002000) emissions are significantly overestimated in the 2008 NEI because of incorrect reconciliation with the point source inventory. Nonpoint industrial coal combustion emissions were estimated by subtracting point source emissions rather than activity. By not accounting for controlled sources, the remaining activity for nonpoint coal combustion is significantly overestimated. EPA NEI experts determined that it would be more appropriate to completely remove the nonpoint component of this sector than to leave the values as they were. The reality for TN industrial coal combustion nonpoint sector emissions is likely much closer to zero than the value in the 2008 NEI because these emissions are accounted for in the point source inventory.

Residential Wood Combustion
There were many modifications to the RWC emissions data. First, all RWC outdoor wood burning devices such as "fire pits and chimneys" (SCC=2104008700) were removed because they were only reported in a couple of states, RPO inventories did not include them for most states and emissions were generally insignificant. A market research report (Frost and Sullivan, 2010) developed in support of the potential RWC New Source Performance Standard (NSPS) indicated slower sales of outdoor hydronic heaters compared to what was assumed for growth

estimates in the 2008 NEI. Therefore, outdoor hydronic heater appliance counts and emissions estimates (SCC=2104008610) were recomputed for all states, resulting in a 51% reduction to outdoor hydronic heater emissions for all states.

In addition, all emissions in the SESARM states (i.e., AL, FL, GA, KY, MS, NC, SC, TN, VA, WV), including Virginia, were replaced with the SESARM year-2007 inventory (SESARM, 2012b). Urban area RWC were lower than the NEI estimates partially because of the assumptions about greater penetration of natural gas fireplaces, less access to inexpensive wood supplies and a lower proportion of housing units with wood burning appliances as primary heating units than rural areas. Overall, the SESARM RWC estimates are considerably lower than the 2008 NEI estimates for several states, particularly for "uncertified" and "general" wood stoves and insert categories: FL, KY, NC, TN, VA and WV. However, emissions in Mississippi are only slightly reduced and emissions in AL, GA and SC are very similar to those in the 2008NEIv2.

The Midwest RPO (LADCO) states (i.e., IL, IL, MI, OH, WI, MN) year-2007 RWC inventory was similar to the 2008 NEI for most source types. However, the pellet stoves (SCC=2104008400), indoor furnaces (2104008510), and outdoor hydronic heater (OHH, SCC=2104008610) estimates were updated to reallocate the indoor furnaces and OHHs to non-MSA counties (LADCO, 2012) for several urban areas. Some double counting of appliances was also fixed in Wisconsin and Michigan. Overall, the MWRPO states totals are very similar to the 2008 NEI; however, emissions are spatially redistributed from urban to rural areas. Therefore, for the MWRPO states, the 2008 NEI emissions were used for all RWC sources except the three aforementioned SCCs that use the 2007 MWRPO data.

Emissions from indoor wood fired furnaces (SCC=2104008510) in several MWRPO states based were also recomputed based on newer, improved survey data from Minnesota. The 2008 NEI for these sources started with an assumption of year 2002 Minnesota wood burning survey data of 38 indoor furnaces per 100 woodstoves for Illinois, Indiana, Michigan, Ohio, and Wisconsin. More recent year 2007 MN survey data resulted in the much lower ratio of 7.3 indoor furnaces per 100 wood stove units. Thus, for the other five MWRPO states previously listed, the indoor furnace emissions are normalized by setting the indoor furnace count ratio to wood stoves to match the 7.6% reported value in Minnesota. The resulting adjustment factors reduce the indoor furnace emissions in these states by 67% (Wisconsin) to as much as 83% in Ohio.

The MARAMA states (i.e., CT, DE, DC, ME, MD, MA, NH, NJ, NY, PA, RI, VT) year 2007 RWC inventory was either unchanged from the 2008 NEI, or was missing for most states. The exceptions were New York and Pennsylvania which includes significantly revised RWC estimates compared to the 2008 NEI. For New York, the MARAMA estimates were not split out into the refined set of 10 RWC appliance types/SCCs in the NEI. New York only reported "general" fireplaces (SCC=2104008100) and "EPA certified, non-catalytic" woodstoves (SCC=2104008320). However, similar to the SESARM and MWRPO improvements, the MARAMA NY RWC estimates were spatially reallocated from urban to more rural areas and were also lower state-wide than the NEI. For Pennsylvania, MARAMA RWC estimates were not much different state-wide on the aggregate, but were refined by SCC and spatially compared to the 2008 NEI. Therefore, the MARAMA 2007 RWC data is used for New York and

Pennsylvania and the 2008 NEI emissions are used for all RWC sources in the rest of the MARAMA states.

The uniform temporalization from month to day was modified to be day-of-year specific as discussed in more detail in the emissions modeling section. In short, the SMOKE program (GenTPRO) is used to distribute annual RWC emissions to the coldest days of the year, using maximum temperature thresholds by-state and/or by-county. On days where the low temperature does not drop below this threshold, RWC emissions are zero. Conversely, the program temporally allocates the most relative emissions to the coldest days. This meteorological-based temporal allocation can have a substantial impact on the amount of RWC emissions in an area on any given day, particularly in the winter.

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 estimates of the emissions from each fires value. The ptfire sector for the 2010 Platform excludes agricultural burning and other open burning sources, which are included in the nonpt sector. The agricultural burning and other open burning sources are in the nonpt sector because these categories were not factored into the development of the ptfire sector. Additionally, their year-to-year impacts are not as variable as wildfires and non-agricultural prescribed/managed burns.

The ptfire sector includes a 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. Note that agricultural burning is not included in the ptfire sector but is included in the nonpt sector. The point source day-specific emission estimates for 2010 fires rely on the Satellite Mapping Automated Reanalysis Tool for Fire Incident Reconciliation Version 2 (SMARTFIRE2) system (Raffuse, et al., 2012). Activity data was used from the Monitoring Trends in Burn Severity (MTBS) project, Incident Command Summary Reports (ICS-209), and the National Oceanic and Atmospheric Administration's (NOAA's) Hazard Mapping System (HMS).

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. The ICS-209 reports for each large wildfire are created daily to enable fire incident commanders to track the status and resources assigned to each large fire (100 acre timber fire or 300 acre rangeland fire). The SMARTFIRE system of reconciliation with ICS-209 reports is described in an Air and Waste Management Association report (Raffuse, et al., 2007). Once the fire reconciliation process is completed, the emissions are calculated using the U.S. Forest Service's CONSUMEv3.0 fuel consumption model and the FCCS fuel-loading database in the BlueSky Framework (Ottmar, et. al., 2007). The detection of fires with this method is satellite-based. Additional sources of information used in the fire classification process included MODIS satellite and fuel moistures derived from fire weather observational data.

The ICS-209 reports for each large wildfire are created daily to enable fire incident commanders to track the status and resources assigned to each large fire (100 acre timber fire or 300 acre rangeland fire). Note that the distinction between wildfire and prescribed burn is not as precise as

with ground-based methods. The fire size was based on the number of satellite pixels and a nominal fire size of 100 acres/pixel was assumed for a significant number of fire detections when the first detections were not matched to ICS 209 reports, so the fire size information is not as precise as ground-based methods.

The activity data and other information were used within the BlueSky Framework to model vegetation distribution, fuel consumption, and emission rates, respectively. Latitude and longitude locations were incorporated as a post processing step. The method to classify fires as WF, WFU, RX (FCCS > 0), and unclassified (FCCS > 0) 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.

Because the HMS satellite product from NOAA is based on daily detections, the emission inventory represents a time-integrated emission estimate. For example, a large smoldering fire will show up on satellite for many days and would count as acres burned on a daily basis; whereas a ground-based method would count the area burned only once even it burns over many days.

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). The BEIS3.14 model creates gridded, hourly, model-species emissions from vegetation and soils. It estimates CO, VOC, and NO_X 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).

3.2.5 *Biogenic Sources (beis)*

For CMAQ, biogenic emissions were computed with the BEIS3.14 model within SMOKE using 2010 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.

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)*

The 2010 onroad emissions are broken out into two sectors: "onroad" and "onroad_rfl". Aircraft emissions are in the nonEGU point inventory. The locomotive and commercial marine emissions are divided into two sectors: "c1c2rail" and "c3marine", and the "nonroad" sector contains the

remaining nonroad emissions. Note that the 2008 NEI includes state-submitted emissions data for nonroad, but the modeling performed for this platform does not incorporate state-submitted emissions for the onroad or nonroad sectors, except for California. All tribal data from the mobile sectors have been dropped because we do not have spatial surrogate data, and the emissions are small.

The onroad and onroad_rfl sectors are processed separately to allow for different spatial allocation to be applied to onroad refueling via a gas station surrogate, versus onroad vehicles that are spatially allocated based on roads and population. Except for California, all onroad and onroad refueling emissions are generated using the SMOKE-MOVES emissions modeling framework that leverages MOVES2010b-generated outputs (http://www.epa.gov/otaq/models/moves/index.htm) and gridded hourly meteorological data. Emissions for onroad (including refueling), nonroad and c1c2rail sources in California were provided by the California Air Resources Board (CARB).

The nonroad sector is based on NMIM except for California which uses data provided by the California Air Resources Board (CARB). 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 2010 NMIM nonroad emissions were generated using activity (e.g., fuels, vehicle population, etc) data that represent 2010. All nonroad emissions are compiled at the county/SCC level. Detailed inventory documentation for the 2008 NEI nonroad sectors is available at http://www.epa.gov/ttn/chief/net/2008inventory.html#inventorydoc. Neither NMIM nor MOVES generates tribal data.

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 2008 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)*

For the Version 5 modeling platform, EPA estimated emissions for every county in the continental U.S. except for California using similar methods as for the 2008 NEI Versions 2 and 3. The modeling framework took into account the temperature sensitivity of the onroad emissions. Specifically, county-specific inputs and tools were used that integrated the MOVES model with the SMOKE emission inventory model to take advantage of the gridded hourly temperature information available from meteorology modeling used for air quality modeling. This integrated "SMOKE-MOVES" tool was developed by EPA in 2010 and is in use by states and regional planning organizations for regional air quality modeling. SMOKE-MOVES requires emission rate "lookup" tables generated by MOVES that differentiate emissions by process (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 146 "representative counties," to which every other county could be mapped as detailed below. Using the MOVES emission rates, SMOKE selected appropriate emissions rates for each county, hourly temperature, SCC, and speed bin and multiplied the emission rate by 2010-specific activity (i.e., VMT (vehicle miles travelled) or vehicle population) to produce emissions. These calculations were done for every county, grid cell, and hour in the continental United States. SMOKE-MOVES can be used with different versions of the MOVES model. For the Version 5 modeling platform, EPA used the latest publically released version: MOVES2010b (http://www.epa.gov/otaq/models/moves/index.htm). Fuels representative of 2010 were used, with temperature and humidity values from 2010.

The steps to apply SMOKE-MOVES to create emissions for modeling requires numerous steps, as follows:

- Determine which counties will be used to represent other counties in the MOVES runs.
- Determine which months will be used to represent other month's fuel characteristics.
- Create MOVES inputs needed only for MOVES runs. MOVES requires county-specific information on vehicle populations, age distributions, and inspection-maintenance programs for each of the representative counties.
- Create inputs needed both by MOVES and by SMOKE, including a list of year-specific temperatures and activity data.
- Run MOVES to create emission factor tables using year-specific fuel information.
- Run SMOKE to apply the emission factors to activities to calculate emissions.
- Aggregate the results at the county-SCC level for summaries and quality assurance.

Some data used in the SMOKE-MOVES process is year-specific. When MOVES was run to generate the emission factors, gasoline and diesel properties for representing counties were based on 2010 fuel information (i.e., RegionalFuels_2010_20120802. The temperature and humidity inputs were also based on 2010 values. The VMT used by SMOKE-MOVES was generated by taking 2010 VMT by state and freeway/non-freeway from FHWA VM-2 tables and allocating to county and month and roadtype using the 2008 NEI VMT. The VMT was allocated to vehicle type using FHWA's VM-4 table and to MOVES sourcetype using ratios from MOVES. Vehicle populations were then generated by applying VMT/vehicle default ratios from MOVES to the VMT. The same speed data used for the 2008 NEI were also used for this study.

The California emissions were post-processed to incorporate both CARB supplied inventories and the shape of the meteorologically-based SMOKE-MOVES results by scaling the SMOKE-MOVES generated totals to match CARB-provided totals. Because CARB provide 2007 and 2011 emissions data, the data for 2010 were linearly interpolated between 2007 and 2011 levels. For more details on this process, see the Version 5 platform documentation.

3.2.8 Onroad Refueling (onroad_rfl)

Onroad refueling was modeled very similarly to the other onroad emissions. MOVES2010b was used produce emission factors (EFs) for refueling. These EFs are at the resolution of the onroad SCC and were run separately from the other onroad mobile sources to allow for different spatial allocation. To facilitate this, the EFs were separated into refueling and non-refueling tables. SMOKE-MOVES was then run using these EF tables as inputs and the results spatially allocated based on a gas stations spatial surrogate. For California, the SMOKE-MOVES generated emissions were used for onroad refueling without any adjustments because there were no CARB-supplied refueling emissions.

3.2.9 Nonroad Mobile Sources — NMIM-Based (nonroad)

The nonroad sector includes monthly exhaust, evaporative and refueling emissions from nonroad engines (not including commercial marine, aircraft, and locomotives) that are derived from NMIM for all states except California. NMIM was run using 2010 meteorological and fuel data to create county-SCC emissions by month for the 2010 nonroad mobile CAP and HAP sources. This version of NMIM ran the NR08a version of NONROAD. The run incorporated Bond rule revisions to some of the base case inputs and the Bond Rule controls did not take effect until future years. NMIM provides nonroad emissions for VOC by three emission modes: exhaust, evaporative and refueling. Unlike the onroad sector, refueling emissions from nonroad sources are not separated into a different sector.

EPA default inputs were replaced by state inputs where such data were provided via the 2008 NEI process. The 2008 NEI documentation describes this and other details of the NMIM nonroad emissions development. CAPs and only the necessary HAPs for the nonroad sector (i.e., BAFM, butadiene, and naphthalene) were included. For this study, NMIM was run separately for each county. To aid with the processing by SMOKE, the mode was appended to the pollutant name and the California NMIM data was replaced with state-supplied data.

For California, year 2010 nonroad emissions values were interpolated between the 2007 and 2011 emissions provided by CARB.   The CARB-supplied nonroad annual inventory to monthly emissions values by using the aforementioned EPA NMIM monthly inventories to compute monthly ratios by pollutant and SCC. Some adjustments to the CARB inventory were needed to convert the provided total organic gas (TOG) to the VOC that was needed by SMOKE.

3.2.10 Nonroad Mobile Sources: Commercial Marine C1, C2, and Locomotive (c1c2rail)

The c1c2rail sector contains CAP and HAP emissions from locomotive and commercial marine sources, except for the category 3/residual-fuel (C3) commercial marine vessels (CMV) found in the c3marine sector. The "c1c2" portion of this sector name refers to the Class I/II CMV emissions, not the railway emissions. Railway maintenance emissions are included in the nonroad sector because these are included in the nonroad NMIM monthly inventories. The C3 CMV emissions are in the c3marine sector. Except for California, the emissions in the c1c2rail sector are year 2008 and are composed of the following SCCs: 2280002100 (CMV diesel, ports), 2280002200 (CMV diesel, underway), 2285002006 (locomotives diesel line haul Class I), 2285002007 (locomotives diesel line haul Class II/III), 2285002008 (locomotives diesel line haul

passenger trains), 2285002009 (locomotives diesel line haul commuter lines), and 2285002010 (locomotives diesel, yard).

The 2008 NEI Version 2 was the starting point for this sector, but several adjustments were made. First, the 2008 NEI point inventory contains rail yard emissions for several states and counties. The NEI point and nonpoint inventories were reviewed for counties with significant rail yard emissions in both inventories. It was assumed that the point inventory contained more accurate information when both inventories contained rail yard emissions. Therefore, nonpoint rail yards were removed from the c1c2rail sector for certain counties in California, Maryland, Oregon and Arizona. For more information, see the Version 5 2007 platform documentation.

Analysis of the total rail emissions in the 2008 NEI showed what appeared to be missing rail line emissions in Texas. It was determined that line haul emissions from Texas were essentially zero in the 2008 NEI. Therefore, all line haul emissions from the 2008 NEI were removed and information from an EPA default dataset of Texas line haul emissions was added. These EPA line haul emissions are restricted to the Class I and Class II/III operations and add approximately 52,000 tons of NOX to Texas that would otherwise be missing.

For several Texas counties, the C1/C2 CMV emissions in the 2008 NEI included EPA gap filled values where shape IDs were not populated in the state submittal. The intended Texas submittal was often much smaller than the EPA-estimated default value for several counties. An example of this is Harris county (FIPS=48201) where the Texas submittal was approximately 1,200 tons of NOX for port and underway emissions but not all shape IDs were included. The NEI methodology used EPA emissions where Texas did not provide estimates and the resulting double count and overestimate of this top-down method resulted in over 49,000 tons of NOX in the 2008 NEI in Harris County, Texas. Therefore, the modeling platform used the original Texas submittal, did not append any EPA emissions, and summed up port and underway for the modeling files to the county level. Similar corrections to these may have been included in Version 3 of the 2008 NEI. Other states were impacted by a similar error in the 2008 NEI Version 2, but for many of these states alternative data were used as discussed below.

For California, the California Air Resources Board (CARB) provided year 2007 and 2011 emissions for all mobile sources, including C1/C2 CMV and rail. These emissions are documented in a staff report available at:
http://www.arb.ca.gov/regact/2010/offroadlsi10/offroadisor.pdf. The modeling platform uses 2010 emissions interpolated between the 2007 and 2011 emissions. The C1/C2 CMV emissions were obtained from the CARB nonroad mobile dataset and include the regulations to reduce emissions from diesel engines on commercial harbor craft operated within California waters and 24 nautical miles of the California baseline. These emissions were developed using Version 1 of the 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 here: http://www.arb.ca.gov/msei/categories.htm#offroad_motor_vehicles. The CARB inventory TOG emissions were converted to VOC by dividing the inventory TOG by the available VOC-to-TOG speciation factor.

Year-2007 inventories provided by MARAMA, SESARM and the MWRPO were used for the c1c2rail sector emissions in their respective states. Emissions data from MARAMA rather than SESARM was used for Virginia because the SESARM data included some rather large emissions for Commuter Lines (SCC=2285002009) that were not reflected in the 2008 NEI nor the MARAMA dataset. The MWRPO year-2007 c1c2rail data were obtained from a subset of their version 7 emissions modeling file "nrinv.mwrpo_alm.baseCv7.annual.orl.txt", where MWRPO NEI Inventory Format (NIF)-formatted data were converted to SMOKE ORL format. The MARAMA dataset was obtained from a subset of their version 3.3 January 27, 2012 vintage file "ARINV_2007_MAR_Jan2012.txt". The SESARM dataset was obtained from a subset of the file "nrinv.alm.semap.base07.v093010.orl.txt" developed for the Southeastern Modeling, Analysis, and Planning (SEMAP) project. All RPO datasets were edited to remove non-c1c2rail sources.

3.2.11 Nonroad mobile sources: C3 commercial marine (c3marine)

The c3marine sector emissions data were 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). These emissions consist of large marine diesel engines (at or above 30 liters/cylinder) that until very recently, were allowed to meet relatively modest emission requirements, often burning residual fuel. The emissions in this sector are comprised of primarily foreign-flagged ocean-going vessels, referred to as Category 3 (C3) CMV ships.

The c3marine inventory includes these ships in several intra-port modes (cruising, hoteling, reduced speed zone, maneuvering, and idling) and underway mode and includes near-port auxiliary engines. 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 NOX, SO2, and PM C3 emissions can be found at: http://www.epa.gov/oms/regs/nonroad/marine/ci/420r09019.pdf. The resulting ECA-IMO coordinated strategy, including emission standards under the Clean Air Act for new marine diesel engines with per-cylinder displacement at or above 30 liters, and the establishment of Emission Control Areas is 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, thereby allowing for the emissions to be allocated to modeling layers above the surface layer. As described in the paper, the ASCII raster dataset was converted to latitude-longitude, mapped to state/county FIPS codes that extended up to 200 nautical miles (nm) from the coast, assigned stack parameters, and monthly ASCII raster dataset emissions were used to create monthly temporal profiles. 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. All non-US emissions (i.e., in waters considered outside of the 200nm EEZ, and hence out of the U.S. territory) are assigned a dummy state/county FIPS code=98001. The SMOKE-ready data were cropped from the original ECA-IMO data to cover only the 36-km CMAQ domain, which is the largest domain used for this effort, and larger than the 12km domain used in this project.

The base year ECA inventory is 2002 and consists of these CAPs: PM10, PM2.5, CO, CO_2, NH3, NOX, SOX (assumed to be SO2), and Hydrocarbons (assumed to be VOC). The EPA developed regional growth (activity-based) factors that we applied to create the 2007v5 inventory from the 2002 data. These growth factors are provided in Table 3-4. The East Coast and Gulf Coast regions were divided along a line roughly through Key Largo (longitude 80° 26' West).

Table 3-4. Growth factors to project the 2002 ECA inventory to 2010

Region	EEZ FIPS	NO$_X$	PM$_{10}$	PM$_{2.5}$	VOC	CO	SO$_2$
East Coast (EC)	85004	1.258	0.478	0.475	1.436	1.436	0.513
Gulf Coast (GC)	85003	1.096	0.415	0.411	1.251	1.252	0.448
North Pacific (NP)	85001	1.158	0.452	0.444	1.310	1.309	0.508
South Pacific (SP)	85002	1.314	0.499	0.495	1.489	1.486	0.580
Great Lakes (GL)	n/a	1.061	0.387	0.384	1.157	1.156	0.408
Outside ECA	98001	1.300	1.396	1.396	1.396	1.396	1.396

A modification to the original ECA-IMO c3marine dataset include updating the state of Delaware county total emissions to reflect comments received during the Cross-State Air Pollution Rule (CSAPR) emissions modeling platform development: http://www.epa.gov/ttn/chief/emch/index.html#final. The original ECA-IMO inventory also did not delineate between ports and underway (or other C3 modes such as hoteling, maneuvering, reduced-speed zone, and idling) emissions; however, we used a U.S. ports spatial surrogate dataset to assign the ECA-IMO emissions to ports and underway SCCs - 2280003100 and 2280003200, respectively. This has no effect on temporal allocation or speciation because all C3 emissions, unclassified/total, port and underway, share the same temporal and speciation profiles.

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). The assignment of U.S. FIPS was also restricted to state-federal water boundaries data from the Mineral Management Service (MMS) that extended only (approximately) 3 to 10 miles off shore. Emissions outside the 3 to 10 mile MMS boundary but within the approximately 200 nm EEZ boundary in Figure 2 8 were projected to year 2010 using the same regional adjustment factors as the U.S. emissions; however, the FIPS codes were assigned as "EEZ" FIPS. 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. The classification of emissions to U.S. and Canadian FIPS codes is primarily needed only for inventory summaries and is irrelevant for air quality modeling except potentially for source apportionment of states contributions to transport.

Factors were applied to compute HAP emissions (based on emissions ratios) to VOC to obtain HAP emissions values. Table 3-5 below shows these factors. Because HAPs were computed directly from the CAP inventory and the calculations are therefore consistent, the entire c3marine sector utilizes CAP-HAP VOC integration to use the VOC HAP species directly, rather than VOC speciation profiles.

Table 3-5. HAP emission ratios for generation of HAP emissions from criteria emissions for C3 commercial marine vessels

Pollutant	Apply to	Pollutant Code	Factor
Acetaldehyde	VOC	75070	0.0002286
Benzene	VOC	71432	9.80E-06
Formaldehyde	VOC	50000	0.0015672

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

The emissions from Canada, Mexico, and offshore 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 code, 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. All "oth" emissions are CAP-only inventories.

For Canada, year-2006 Canadian emissions were used but several modifications were applied to the inventories:

1. Wildfires or prescribed burning were not included because Canada does not include these inventory data in their modeling.
2. In-flight aircraft emissions were not included because we do not include these for the U.S. and we do not have a finalized approach to include in our modeling.
3. A 75% reduction ("transport fraction") was applied to PM for the road dust, agricultural, and construction emissions in the Canadian "afdust" inventory. This approach is more simplistic than the county-specific approach used for the U.S., but a comparable approach was not available for Canada.
4. Speciated VOC emissions from the ADOM chemical mechanism were not included because we use speciated emissions from the CB5 chemical mechanism that Canada also provided.
5. Residual fuel CMV (C3) SCCs (22800030X0) were removed because these emissions are included in the c3marine sector, which covers not only emissions close to Canada but also emissions far at sea. Canada was involved in the inventory development of the c3marine sector emissions.
6. Wind erosion (SCC=2730100000) and cigarette smoke (SCC=2810060000) emissions were removed from the nonpoint (nonpt) inventory; these emissions are also absent from our U.S. inventory.
7. Quebec PM2.5 emissions (2,000 tons/yr) were removed for one SCC (2305070000) for Industrial Processes, Mineral Processes, Gypsum, and Plaster Products due to corrupt

fields after conversion to SMOKE input format. This error should be corrected in a future inventory.

8. Excessively high CO emissions were removed from Babine Forest Products Ltd (British Columbia SMOKE plantid='5188') in the point inventory.

9. The county part of the state/county FIPS code field in the SMOKE inputs were modified in the point inventory from "000" to "001" to enable matching to existing temporal profiles.

For Mexico, year 2008 emissions were used that are projections of their 1999 inventory originally developed by Eastern Research Group Inc., (ERG, 2006) 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-2008 Mexico emissions from the 1999 inventory is available at:
http://www.wrapair.org/forums/ef/inventories/MNEI/index.html.

The offshore emissions include point source offshore oil and gas drilling platforms. We used emissions from the 2008 NEI point source inventory. The offshore sources were provided by the Mineral Management Services (MMS).

3.2.13 *SMOKE-ready non-anthropogenic chlorine inventory*

The ocean chlorine gas emission estimates are based on the build-up of molecular chlorine (Cl_2) concentrations in oceanic air masses (Bullock and Brehme, 2002). Data at 36 km and 12 km resolution were available and were not modified other than the name "CHLORINE" was changed to "CL2" because that is the name required by the CMAQ model. 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 emissions data to be input as hourly rates of specific gas and particle species for the horizontal and vertical grid cells contained within the modeled region (i.e., modeling domain). To provide emissions in the form and format required by the model, it is necessary to "pre-process" the "raw" emissions (i.e., emissions input to SMOKE) for the sectors described above. In brief, the process of emissions modeling transforms the emissions inventories from their original temporal resolution, pollutant resolution, and spatial resolution into the hourly, speciated, gridded resolution required by the air quality model. The pre-processing steps involving temporal allocation, spatial allocation, pollutant speciation, and vertical allocation of point sources are referred to as emissions modeling.

The temporal resolution of the emissions inventories input to SMOKE for the modeling platform varies across sectors, and may be hourly, monthly, or annual total emissions. The spatial resolution, which also can be different for different sectors, may be at the level of individual

point sources, county totals, province totals for Canada, or municipio totals for Mexico. This section provides some basic information about the tools and data files used for emissions modeling as part of the Version 5 platform. The emissions inventories were discussed in detail earlier. Therefore, the descriptions of data in this section are limited to the ancillary data used by SMOKE to perform the emissions modeling steps.

3.3.1 The SMOKE Modeling System

For this study, emission inventories were processed into CMAQ-ready inputs using SMOKE version 3.1. 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 CMAQ was used, and therefore source-based emissions files were created rather than the much larger three-dimensional files. For quality assurance purposes, emissions totals by specie for the entire model domain are output as reports that are then compared to inventory level reports generated by SMOKE to ensure mass is not lost or gained during this conversion 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. 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 for the individual SMOKE programs for each sector. Table 3-6 summarizes the major processing steps of each platform sector. The "Spatial" column shows the spatial approach: "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 BEIS3 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 which will have 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 and ptfire sectors are the only sectors with only "in-line" emissions, meaning that all of the emissions are placed in aloft layers and thus there are no emissions for those sectors in the two-dimensional, layer-1 files created by SMOKE. In addition to the other settings, no grouping of stacks was performed using the PELVCONFIG file because grouping done for "in-line" processing will not give identical results as "offline" (i.e., processing whereby SMOKE creates 3-dimensional files). The only way to get the same results

between in-line and offline is to choose to have no grouping.

Table 3-6. Key emissions modeling steps by sector

Platform sector	Spatial	Speciation	Inventory resolution	Plume rise
Ptipm	Point	Yes	daily & hourly	in-line
Ptnonipm	Point	Yes	annual	in-line
Ptfire	Point	Yes	Daily	in-line
Othpt	Point	Yes	annual	in-line
c3marine	Point	Yes	annual	in-line
Ag	Surrogates	Yes	annual & monthly	
Afdust	Surrogates	Yes	annual	
Beis	pre-gridded landuse	in BEIS	computed hourly	
c1c2rail	Surrogates	Yes	annual	
Nonpt	surrogates & area-to-point	Yes	annual & monthly for ag burning and SESARM open burning	
Nonroad	surrogates & area-to-point	Yes	monthly	
Onroad	Surrogates	Yes	computed hourly	
onroad_rfl	Surrogates	Yes	computed hourly	
Othar	Surrogates	Yes	annual	
Othon	Surrogates	Yes	annual	

3.3.3   *Spatial Configuration*

For this study, SMOKE and CMAQ were run for a 12-km modeling domain shown in Figure 3-1 (12US1). 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 3-1. CMAQ Modeling Domain

3.3.4 Chemical Speciation Configuration

The emissions modeling step for chemical speciation creates "model species" needed by the air quality model for a specific chemical mechanism. These model species are either individual chemical compounds or groups of species, called "model species." The chemical mechanism used for this study is the Carbon Bond 05 (CB05) mechanism (Yarwood, 2005) with secondary organic aerosol (SOA) and HONO enhancements as described in http://www.cmascenter.org/help/model_docs/cmaq/4.7/ RELEASE_NOTES.txt. The mapping of inventory pollutants to model species is shown in Table 3-7. From the perspective of emissions preparation, the CB05 with SOA mechanism is the same as was used in the 2005 platform. 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 in CMAQ.

Table 3-7. Model Species Produced by SMOKE for CB05

Inventory Pollutant	Model Species	Model Species Description
CO	CO	Carbon monoxide
NO_X	NO	Nitrogen oxide
	NO_2	Nitrogen dioxide
SO_2	SO_2	Sulfur dioxide
	SULF	Sulfuric acid vapor
NH_3	NH_3	Ammonia
VOC	ALD2	Acetaldehyde
	ALDX	Propionaldehyde and higher aldehydes
	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
Various additional VOC species from the biogenics model which do not map to the above model species	TERP	Terpenes
PM_{10}	PMC	Coarse PM > 2.5 microns and ≤ 10 microns
$PM_{2.5}$	PEC	Particulate elemental carbon ≤ 2.5 microns
	PNO_3	Particulate nitrate ≤ 2.5 microns
	POC	Particulate organic carbon (carbon only) ≤ 2.5 microns
	PSO_4	Particulate sulfate ≤ 2.5 microns
	PMFINE	Other particulate matter ≤ 2.5 microns

The approach for speciating PM2.5 emissions supports both CMAQ 4.7.1 with five species (i.e., AE5) and CMAQ 5.0 and includes speciation of PM2.5 into 17 PM model species (i.e., AE6). The TOG and $PM_{2.5}$ speciation factors that are the basis of the chemical speciation approach were developed from the SPECIATE4.3 database (http://www.epa.gov/ttn/chief/software/speciate) and is the EPA's repository of TOG and PM speciation profiles of air pollution sources. A few of the profiles used in the v5 platform will be published in later versions of the SPECIATE database. The SPECIATE database development and maintenance is a collaboration involving the EPA's ORD, OTAQ, and the Office of Air Quality Planning and Standards (OAQPS), and 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 PM2.5. The database also contains the PM2.5, 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. These simplified components for AE5 are:

- PSO$_4$: primary particulate sulfate
- PNO$_3$: primary particulate nitrate
- PEC: primary particulate elemental carbon
- POC: primary particulate organic carbon
- PMFINE: other primary particulate, less than 2.5 micrograms in diameter

NOX can be speciated into NO, NO2, and/or HONO. For the non-mobile sources, a single profile is used "NHONO" to split NOX into NO and NO2 with 10% NO2 and 90% NO. 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 NOX into NO, NO2, and HONO with 90% NO, 9.2% NO2 and 0.8% HONO. The onroad sector does not use the "HONO" profile to speciate NOX. Instead, MOVES2010b produces speciated NO, NO2, and HONO by source, including emission factors for these species in the emission factor tables used by SMOKE-MOVES. Within MOVES, the HONO fraction is a constant 0.008 of NOX. The NO fraction varies by heavy duty versus light duty, fuel type, and model year. The NO2 fraction = 1 – NO – HONO. For more details on the NOX fractions within MOVES, see http://www.epa.gov/otaq/models/moves/documents/420r12022.pdf. The SMOKE-MOVES system is configured to model these species directly without further speciation.

The approach for speciating VOC emissions from non-biogenic sources has the following characteristics: 1) for some sources, HAP emissions are used in the speciation process to allow integration of VOC and HAP emissions in the NEI; and, 2) for some mobile sources, "combination" profiles are specified by county and month and emission mode (e.g., exhaust, evaporative). SMOKE computes the resultant profile on-the-fly given the fraction of each specific profile specified for the particular county, month and emission mode. The SMOKE feature called the GSPRO_COMBO file supports this approach.

The VOC speciation approach for the 2010 Platform includes HAP emissions from the NEI in the speciation process for some sectors. That is instead of speciating VOC to generate all of the species needed by the model, emissions of the 4 HAPs, benzene, acetaldehyde, formaldehyde and methanol (BAFM) from the NEI were integrated with the NEI VOC. The integration process combines the BAFM HAPs with the VOC in a way that does not double-count emissions and uses the BAFM directly in the speciation process. Generally, the HAP emissions from the NEI are believed to be more representative of emissions of these compounds than their generation via VOC speciation.

The BAFM HAPs were chosen for this special treatment because, with the exception of BENZENE, they are the only explicit VOC HAPs in the base version of CMAQ 4.7 model. By "explicit VOC HAPs," we mean 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 was added as a model species in the base version of CMAQ 4.7, 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, we included ethanol in our integration. To differentiate when a source was integrating BAFM versus EBAFM (ethanol in addition to BAFM), the speciation profiles which do not include ethanol are referred to as an "E-profile", for example E10 headspace gasoline evaporative speciation profile 8763 where ethanol is speciated from VOC, versus 8763E where ethanol is obtained directly from the inventory. The specific profiles used in 2010 are the same as used for the 2007 platform (see 2007 speciation in Table 3-6 in the 2007v5 TSD). The only differences between 2010 and 2007 are the GSPRO_COMBOs, which represent a different mixture of E0 and E10 by county between the two modeling years.

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 the particular HAPs to integrate and the particular sources to integrate. The HAPs to integrate are specified in the INVTABLE file, and the sources to integrate are based on the NHAPEXCLUDE file (which lists the sources that are *excluded* from integration). For the "integrate" sources, SMOKE subtracts the "integrate " 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. SMOKE computes NONHAPTOG and then applies the speciation profiles to allocate the NONHAPTOG to the other CMAQ VOC species not including the integrated HAPs.

CAP-HAP integration was considered for all sectors and "integration criteria" were developed for some of those. Table 3-8 summarizes the integration approach for each platform sector. For the c1c2rail sector, the integration criteria were (1) that the source had to have at least one of the 4 HAPs and (2) that the sum of BAFM could not exceed the VOC emissions. For the nonpt sector, the following integration criteria were used to determine the sources to integrate:

1. Any source for which the sum of B, A, F, or M is greater than the VOC was not integrated, since this clearly identifies sources for which there is an inconsistency between VOC and VOC HAPs.

2. For some source categories (those that comprised 80% of the VOC emissions), sources were selected for integration in the category per specific criteria. For most of these source categories, sources may be integrated if they had the minimum combination of B, A, F, and M. For some source categories, all sources were designated as "no-integrate".

3. For source categories that do not comprise the top 80% of VOC emissions, as long as the source has emissions of one of the B, F, A or M pollutants, then it can be integrated.

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

Platform Sector	Approach for Integrating NEI emissions of Benzene (B), Acetaldehyde (A), Formaldehyde (F) and Methanol (M)
Ptipm	No integration because emissions of BAFM are relatively small for this sector
Ptnonipm	No integration because emissions of BAFM are relatively small for this sector and it is not expected that criteria for integration would be met by a significant number of sources
Ptfire	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
C1c2rail	Partial integration
C3marine	Full integration
Nonpt	Partial integration
Nonroad	Partial integration—did not integrate California emissions, CNG or LPG sources (SCCs beginning with 2268 or 2267) because NMIM computed only VOC and not any HAPs for these SCCs.
Onroad	Full integration
Othar	No integration—not the NEI
Othon	No integration—not the NEI
Othpt	No integration—not the NEI

The SMOKE feature to compute speciation profiles from mixtures of other profiles in user-specified proportions was used in this project. 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 for onroad and nonroad mobile and gasoline-related related stationary sources. Since the ethanol content varies spatially (e.g., by state or sources use fuels with varying ethanol content, and therefore the speciation profiles require different combinations of gasoline and E10 profiles by county), temporally (e.g., by month) and by modeling year (i.e., future years have more thanol) the combo feature allows combinations to be specified at various levels for different years.

The INVTABLE and NHAPEXCLUDE SMOKE input files have a critical function in the VOC speciation process for emissions modeling cases utilizing HAP-CAP integration, as is done for the 2010 Platform. Two different types of INVTABLE files were developed to use with different sectors of the platform. For sectors in which we chose no integration across the entire sector a "no HAP use" INVTABLE was developed in which the "KEEP" flag is set to "N" for BAFM pollutants. Thus, any BAFM pollutants in the inventory input into SMOKE are dropped. This both avoids double-counting of these species and assumes that the VOC speciation is the best available approach for these species for the sectors using the approach. The second INVTABLE is used for sectors in which one or more sources are integrated and causes SMOKE to keep the BAFM pollutants and indicates that they are to be integrated with VOC (by setting the "VOC or TOG component" field to "V" for all four HAP pollutants. This integrate INVTABLE is further differentiated into sectors that integrate BAFM versus those that integrate EBAFM (e.g., the onroad and onroad_rfl sectors).

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 MOVES2010b speciated PM does not map 1-to-1 to either the AE5 or AE6 species. Table 3-9 shows the relationship between MOVES2010b exhaust $PM_{2.5}$ related species and CMAQ AE5 PM species.

Table 3-9. MOVES exhaust PM species versus AE5 species

MOVES2010b Pollutant Name	Variable name for Equations	Relation to AE5 model species
Primary Exhaust PM2.5 – Total	PM25_TOTAL	
Primary $PM_{2.5}$ - Organic Carbon	PM25OM	Sum of POC, PNO3 and PMFINE
Primary $PM_{2.5}$ - Elemental Carbon	PM25EC	PEC
Primary $PM_{2.5}$ - Sulfate Particulate	PM25SO4	PSO4

MOVES species are related as follows:
PM25_TOTAL = PM25EC + PM25OM + PSO4

The five CMAQ AE5 species also sum to total $PM_{2.5}$:
PM2.5 = POC+PEC+PNO3+PSO4+PMFINE

The basic problem is to differentiate MOVES species "PM25OM" into the component AE5 species (POC, PNO3 and PMFINE). The Moves2smkEF post-processor script takes the MOVES2010b species (EF tables) and calculates the appropriate AE5 $PM_{2.5}$ species and converts them into a format that is appropriate for SMOKE (see http://www.smoke-model.org/version3.1/html/ch05s02s04.html for details on the Moves2smkEF script).

For brake wear and tire wear PM, total $PM_{2.5}$ (not speciated) comes directly from MOVES2010b. These PM modes are speciated by SMOKE. PMFINE from onroad exhaust is further speciated by SMOKE into the component AE6 species.

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 species SESQT. The profile code associated with BEIS3.14 profiles for use with CB05 was "B10C5."

3.3.4 Temporal Processing Configuration

Temporal allocation or temporalization is the process of distributing aggregated emissions to a finer temporal resolution, such as 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.

Typically, emissions inventories are annual or monthly in nature. Temporalization takes these annual emissions and distributes them to the month, the monthly emissions to the day, and the daily emissions to the hour. This process is typically done by applying temporal profiles—monthly, day of the week, and diurnal—to the inventories.

The monthly, weekly, and diurnal temporal profiles and associated cross references used to create the hourly emissions inputs for the air quality model were similar to those used for the 2005v4.3 platform. Some new methodologies are introduced in this platform and updated profiles are discussed. Temporal factors are typically applied to the inventory by some combination of country, state, county, SCC, and pollutant.

Table 3-10 summarizes the temporal aspect of the emissions processing configuration. It compares the key approaches used for temporal processing across the sectors. The temporal aspects of SMOKE processing are controlled through (a) the scripts T_TYPE (Temporal type) and M_TYPE (Mergetype) settings and (b) ancillary data files. In the table, "Daily temporal approach" refers to the temporal approach for getting daily emissions from the inventory using the Temporal program. The "Merge processing approach" refers to the days used to represent other days in the month for the merge step. If not "all", then the SMOKE merge step runs only for representative days, which could include holidays as indicated by the right-most column. In addition to the resolution, temporal processing includes a ramp-up period for several days prior to January 1, 2010, intended to mitigate the effects of initial condition concentrations. The ramp up period for the national 12km grid was 10 days. For most sectors, the emissions from late December of 2009 were used to provide emissions for the end of December, 2010.

The Flat File 2010 format (FF10) is a new inventory format for SMOKE that provides a more consolidated format for monthly, daily, and hourly emissions inventories. Previously, 12 separate inventory files would be required to process monthly inventory data. 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, individual records contain data for all days in a month and all hours in a day in the daily and hourly FF10 inventories, respectively.

SMOKE 3.1 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; rather, it should only have month to day and diurnal temporalization. This is particularly important when 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.

Table 3-10. Temporal Settings Used for the Platform Sectors in SMOKE

Platform sector	Inventory resolution	Monthly profiles used?	Daily temporal approach 1,2	Merge processing approach 1,3	Process Holidays as separate days?
Ptipm	daily & hourly		All	all	yes
Ptnonipm	annual	yes	Mwdss	all	yes
Ptfire	Daily		All	all	yes
Ag	annual & monthly	yes	all	all	yes
Afdust	annual	yes	Week	all	yes
Beis	hourly		n/a	all	yes
c3marine	annual	yes	Aveday	aveday	
c1c2rail	annual	yes	Mwdss	mwdss	
Nonpt	annual & monthly	yes	All	all	yes
Nonroad	monthly		Mwdss	mwdss	yes
Onroad	annual & monthly[a]		all	all	yes
onroad_rfl	annual & monthly[a]		All	all	yes
Othar	annual	yes	Week	week	
Othon	annual	yes	Week	week	
Othpt	annual	yes	Mwdss	mwdss	

[1] **Definitions for processing resolution:**
all   = hourly emissions computed for every day of the year
week   = hourly emissions computed for all days in one "representative" week, representing all weeks for each month, which means emissions have day-of-week variation, but not week-to-week variation within the month
mwdss   = hourly emissions for one representative Monday, representative weekday, representative Saturday and representative Sunday for each month, which means emissions have variation between Mondays, other weekdays, Saturdays and Sundays within the month, but not week-to-week variation within the month. Also Tuesdays, Wednesdays and Thursdays are treated the same.
aveday   = hourly emissions computed for one representative day of each month, which means emissions for all days of each month are the same.
[2] Daily temporal approach refers to the temporal approach for getting daily emissions from the inventory using the Temporal program. The values given are the values of the L_TYPE setting.
[3] Merge processing approach refers to the days used to represent other days in the month for the merge step. If not "all", then the SMOKE merge step just run for representative days, which could include holidays as indicated by the rightmost column. The values given are the values of the M_TYPE setting.
[a] For onroad and onroad_rfl, the annual and monthly refers to activity data (VMT and VPOP). Emissions are computed on an hourly basis.

For the EGU emissions in the ptipm sector, hourly CEM NO_X and SO_2 data were used directly for sources that match CEMs. For other pollutants, hourly CEM heat input data were used to allocate the NEI annual values. For sources not matching CEM data ("non-CEM" sources), daily emissions were computed from the NEI annual emissions using a structured query language (SQL) program and state-average CEM data. To allocate annual emissions to each month, state-specific three-year averages of 2008-2010 CEM data were created. These average annual-to-month factors were assigned to non-CEM sources within each state. To allocate the monthly emissions to each day, the 2010 CEM data to compute state-specific month-to-day factors, averaged across all units in each state. These daily emissions were calculated outside of SMOKE and the resulting daily inventory is used as an input into SMOKE.

The daily-to-hourly allocation was performed in SMOKE using diurnal profiles. The state-specific and pollutant-specific diurnal profiles for use in allocating the day-specific emissions for non-CEM sources in the ptipm sector were updated. The 2010 CEM data was used to create state-specific, day-to-hour factors, averaged over the whole year and all units in each state. Diurnal factors were calculated using CEM SO_2 and NO_X emissions and heat input. SO_2 and NO_X-specific factors were computed from the CEM data for these pollutants. All other pollutants used factors created from the hourly heat input data. The resulting profiles were assigned by state and pollutant.

Two updated diurnal temporal profiles were incorporated into the 2010 modeling platform. For all agricultural burning, we used a diurnal temporal profile (McCarty et al., 2009) that puts more of the emissions during the actual work day and suppresses the emissions during the middle of the night was used. Note that all states used a uniform day of week profile for all agricultural burning emissions, except for the following states that for which state-specific day of week profiles were used: Arkansas, Kansas, Louisiana, Minnesota, Missouri, Nebraska, Oklahoma, and Texas. For residential wood combustion, a profile was used that placed more of the emissions in the morning and the evening when people are typically using these sources. This profile is based on an average of 2004 MANE-VU survey based temporal profiles (see http://www.marama.org/publications_folder/ResWoodCombustion/ Final_report.pdf). When this 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 it was found that the updated RWC profile generally tracked the concentration based temporal patterns.

The temporal profile assignments for the Canadian 2006 inventory were provided by Environment Canada along with the inventory. They provided profile assignments that rely on the existing set of temporal profiles in the 2002 Platform. For point sources, they provided profile assignments by PLANTID.

3.3.5 Meteorological-based Temporal Profiles

A significant improvement over previous platforms is the introduction of meteorologically-based temporalization. We recognize that there are many factors that impact the timing of when emissions occur. The benefits of utilizing meteorology as method of temporalizing are: (1) a consistent meteorological dataset as is used by the AQ model (e.g. WRF) is available; (2) the meteorological model data is highly resolved in terms of spatial resolution; and (3) the meteorological variables vary at hourly resolution which can translate to hour-specific temporalization.

The SMOKE program GenTPRO provides a method for developing meteorologically-based temporalization. Currently, the program can utilize three types of temporal algorithms: RWC, agricultural livestock ammonia, and a generic meteorology based algorithm. For the 2007 platform, we used the RWC and ag NH3 GenTPRO generated profiles. GenTPRO reads in gridded meteorology data (MCIP) and 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 algorithm and the run parameters. For more details on the development of these algorithms and running GenTPRO, see the GenTPRO documentation http://www.smoke-

model.org/version3.1/GenTPRO_TechnicalSummary_Aug2012_Final.pdf and the SMOKE manual section http://www.smoke-model.org/version3.1/html/ch05s03s07.html.

For the RWC algorithm, GenTPRO uses the daily minimum temperature to determine the temporal allocation of emissions to days. GenTPRO was run to create an annual-to-day temporal profile for the RWC sources within the nonpt sector. 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 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, just the distribution of the emissions within the year. Initially, the RWC algorithm used a default temperature threshold of 50 ˚F. For most of the country, this produced a reasonable distribution of emissions, but for a few Southern counties all of the emissions were compressed into a few days creating excessively high daily emissions. GenTPRO was then modified to accept an optional input that defines a county/state specific alternative temperature threshold. In addition, an alternative RWC algorithm was created to avoid negative RWC emissions when the daily minimum temperature was greater than 53.3 ˚F. For the v5 platform, the alternative RWC algorithm was used for the whole country, with the default 50 ˚F threshold for the majority of the states, and a 60 ˚F threshold for the following states: Alabama, Arizona, California, Florida, Georgia, Louisiana, Mississippi, South Carolina, and Texas.

For the agricultural livestock NH3 algorithm, GenTPRO algorithm is based on the Russel and Cass (1986) equation. This algorithm uses county-average hourly temperature and wind speed to calculate the temporal profile. GenTPRO was run to create month-to-hour temporal profiles for these sources. Because these profiles distribute to the hour based on monthly emissions, the emissions will either come from a monthly inventory or from an annual inventory that has been temporalized already to the month.

For the onroad and onroad_rfl sectors, meteorology is not used in the development of the temporal profiles; rather, but meteorology impacts the calculation of the hourly emissions through the program Movesmrg. The result is that the emissions will vary at the hourly level by grid cell. More specifically, the on-network (RPD) and the off-network (RPV) exhaust, evaporative, and evaporative permeation modes use the gridded meteorology (MCIP) directly. Movesmrg determines the temperature for each hour and grid cell and uses it to select the appropriate EF for that SCC/pollutant/mode. For the off-network rate per profile (RPP) emissions, Movesmrg uses the Met4moves output for SMOKE (daily minimum and maximum temperatures by county) to determine the appropriate EF for that hour and SCC/pollutant. The result is that the emissions will vary hourly by county. The combination of these three processes (RPD, RPV, and RPP) is the total onroad emissions, while the combination of the two processes (RPD, RPV) for the refueling mode only is the total onroad_rfl emissions. Both sectors will show a strong meteorological influence on their temporal patterns.

3.3.6   *Vertical Allocation of Emissions*

Table 3-6 specifies the sectors for which plume rise is calculated. If there is no plume rise for a sector, the emissions are placed into layer 1 of the air quality model. Vertical plume rise was performed in-line within CMAQ for all of the SMOKE point-source sectors (i.e., 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. See http://www.smoke-model.org/version2.7/html/ch06s07.html for full documentation of Laypoint. The selection of point sources for plume rise is pre-determined in SMOKE using the Elevpoint program (http://www.smoke-model.org/ version2.7/html/ch06s03.html). 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 Ancillary Files

The methods used to perform spatial allocation for the 2007 platform are summarized in this section. For the 2007 platform, spatial factors are typically applied by country and SCC. As described earlier, spatial allocation was performed for a national 12-km domain. To accomplish this, SMOKE used national 12-km spatial surrogates and a SMOKE area-to-point data file. For the U.S., the spatial surrogates used 2010-based data (e.g., population) wherever possible. For Mexico, the same spatial surrogates were used in the 2005 platform. For Canada we used a set of Canadian surrogates provided by Environment Canada, also unchanged from the 2005v4.3 platform. The U.S., Mexican, and Canadian 12-km surrogates cover the entire CONUS domain 12US1 shown in Figure 3-1. The remainder of this subsection provides further details on the origin of the data used for the spatial surrogates and the area-to-point data.

The SMOKE ancillary data files, particularly the cross-reference files, provide the specific inventory resolution at which spatial, speciation, and temporal factors are applied. For the 2010 Platform, spatial factors were generally applied by country/SCC, speciation factors by pollutant/SCC or (for combination profiles) state/ county FIPS code and month, and temporal factors by some combination of country, state, county, SCC, and pollutant.

3.3.7.1 Surrogates for U.S. Emissions

More than sixty spatial surrogates were used to spatially allocate U.S. county-level emissions to the CMAQ 12-km grid cells. The Surrogate Tool was used to generate all of the surrogates. The shapefiles input to the Surrogate Tool are provided and documented at http://www.epa.gov/ttn/chief/emch/spatial/spatialsurrogate.html. The tool and updated 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. The detailed steps in developing the county boundaries for the surrogates are documented at ftp://ftp.epa.gov/EmisInventory/ emiss_shp2006/us/metadata_for_2002_ county_boundary_shapefiles_rev.pdf.

Table 3-11 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, and 2010 National Transportation Atlas Data for ports and navigable waterways. For this project "Version 3" of the 2010-baed spatial surrogates was used. Not all of the available surrogates are used to spatially allocate sources in the 2007 platform; that is, some surrogates shown in Table 3-11 were not assigned to any SCCs. An area-to-point approach overrides the use of surrogates for some airport-related sources.

Alternative surrogates for ports (801) and shipping lanes (802) were developed from the 2008 NEI shapefiles: Ports_032310_wrf and ShippingLanes_111309FINAL_wrf. These surrogates were used for c1 and c2 commercial marine emissions instead of the standard 800 and 810 surrogates, respectively. For the onroad sector, the on-network (RPD) emissions were spatially allocated to roadways, which the off-network (RPP and RPV) emissions were allocated to parking areas. For the onroad_rfl sector, the emissions were spatially allocated to gas station locations.

For the oil and gas sources in the nonpt sector, the WRAP Phase III sources have detailed basin-specific spatial surrogates shown in Table 3-12. The remaining oil and gas sources used the 2005-based surrogate "Oil & Gas Wells, IHS Energy, Inc. and USGS" (680) developed for oil and gas SCCs. The surrogates in Table 3-12 were applied for the counties listed in Table 3-13.

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

There are numerous airport-related emission sources in the 2005 NEI, such as aircraft, airport ground support equipment, and jet refueling. In the 2002 platform most of these emissions were contained in sectors with county-level resolution — alm (aircraft), nonroad (airport ground support) and nonpt (jet refueling), but in the 2005 and 2008 platforms aircraft emissions are included as point sources as part of the ptnonipm sector.

For the 2010 platform, the SMOKE "area-to-point" approach was used for 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.

Nearly the same ARTOPNT file was used to implement the area-to-point approach as was done

for the CAP and HAP-2002-based Platform. This was slightly updated from the CAP-only 2002 Platform by further allocating the Detroit-area airports into multiple sets of geographic coordinates to support finer scale modeling. The updated file was retained for the 2010 Platform.

3.3.7.4 Surrogates for Canada and Mexico Emission Inventories

The Mexican single surrogate (population) was the same as was used in the 2002 and 2005 Platforms. For Canada, surrogates provided by Environment Canada with the 2006 emissions were used to spatially allocate the 2006 Canadian emissions for the 2005 and 2010 Platforms.

The Canadian surrogate data described in Table 3-14 came from Environment Canada. They provided both the surrogates and cross references; the surrogates were outputs from the Surrogate Tool (previously referenced). Per Environment Canada, the surrogates are based on 2001 Canadian census data. The cross-references that Canada originally provided were updated as follows: all assignments to surrogate '978' (manufacturing industries) were changed to '906' (manufacturing services), and all assignments to '985' (construction and mining) and '984' (construction industries) were changed to '907' (construction services) because the surrogate fractions in 984, 978 and 985 did not sum to 1. Codes for surrogates other than population that did not begin with the digit "9" were also changed.

Table 3-11. U.S. Surrogates Available for the 2010 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
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

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 and Gas
280	**Class 2 and 3 Railroad Miles**	700	Airport Areas
300	Low Intensity Residential	**710**	**Airport Points**
310	Total Agriculture	720	Military Airports
312	Orchards/Vineyards	**800**	**Marine Ports**
320	Forest Land	**801**	**NEI Ports**
330	Strip Mines/Quarries	**802**	**NEI Shipping Lanes**
340	**Land**	**807**	**Navigable Waterway Miles**
350	Water	810	Navigable Waterway Activity
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

Table 3-12. Spatial Surrogates for WRAP Oil and Gas Data

Country	Code	Surrogate Description
USA	699	Gas production at CBM wells
USA	698	Well count - gas wells
USA	697	Oil production at gas wells
USA	696	Gas production at gas wells
USA	695	Well count - oil wells
USA	694	Oil production at Oil wells
USA	693	Well count - all wells
USA	692	Spud count
USA	691	Well count - CBM wells
USA	690	Oil production at all wells
USA	689	Gas production at all wells

Table 3-13. Counties included in the WRAP Dataset

FIPS	State	County	FIPS	State	County
8001	Colorado	Adams	30075	Montana	Powder River
8005	Colorado	Arapahoe	35031	New Mexico	Mc Kinley
8007	Colorado	Archuleta	35039	New Mexico	Rio Arriba
8013	Colorado	Boulder	35043	New Mexico	Sandoval
8014	Colorado	Broomfield	35045	New Mexico	San Juan
8029	Colorado	Delta	49007	Utah	Carbon
8031	Colorado	Denver	49009	Utah	Daggett

8039	Colorado	Elbert		49013	Utah	Duchesne
8043	Colorado	Fremont		49015	Utah	Emery
8045	Colorado	Garfield		49019	Utah	Grand
8051	Colorado	Gunnison		49043	Utah	Summit
8063	Colorado	Kit Carson		49047	Utah	Uintah
8067	Colorado	La Plata		56001	Wyoming	Albany
8069	Colorado	Larimer		56005	Wyoming	Campbell
8073	Colorado	Lincoln		56007	Wyoming	Carbon
8075	Colorado	Logan		56009	Wyoming	Converse
8077	Colorado	Mesa		56011	Wyoming	Crook
8081	Colorado	Moffat		56013	Wyoming	Fremont
8087	Colorado	Morgan		56019	Wyoming	Johnson
8095	Colorado	Phillips		56023	Wyoming	Lincoln
8103	Colorado	Rio Blanco		56025	Wyoming	Natrona
8107	Colorado	Routt		56027	Wyoming	Niobrara
8115	Colorado	Sedgwick		56033	Wyoming	Sheridan
8121	Colorado	Washington		56035	Wyoming	Sublette
8123	Colorado	Weld		56037	Wyoming	Sweetwater
8125	Colorado	Yuma		56041	Wyoming	Uinta
30003	Montana	Big Horn		56045	Wyoming	Weston

Table 3-14. Canadian Spatial Surrogates for Canadian Emissions

Code	Description	Code	Description
9100	Population	9493	Warehousing and storage
9101	Total dwelling	9494	Total Transport and warehouse
9102	Urban dwelling	9511	Publishing and information services
9103	Rural dwelling	9512	Motion picture and sound recording industries
9104	Total Employment	9513	Broadcasting and telecommunications
9106	ALL_INDUST	9514	Data processing services
9111	Farms	9516	Total Info and culture
9113	Forestry and logging	9521	Monetary authorities - central bank
9114	Fishing hunting and trapping	9522	Credit intermediation activities
9115	Agriculture and forestry activities	9523	Securities commodity contracts and other financial investment activities
9116	Total Resources	9524	Insurance carriers and related activities
9211	Oil and Gas Extraction	9526	Funds and other financial vehicles
9212	Mining except oil and gas	9528	Total Banks
9213	Mining and Oil and Gas Extract activities	9531	Real estate

Code	Description	Code	Description
9219	Mining-unspecified	9532	Rental and leasing services
9221	Total Mining	9533	Lessors of non-financial intangible assets (except copyrighted works)
9222	Utilities	9534	Total Real estate
9231	Construction except land subdivision and land development	9541	Professional scientific and technical services
9232	Land subdivision and land development	9551	Management of companies and enterprises
9233	Total Land Development	9561	Administrative and support services
9308	Food manufacturing	9562	Waste management and remediation services
9309	Beverage and tobacco product manufacturing	9611	Education Services
9313	Textile mills	9621	Ambulatory health care services
9314	Textile product mills	9622	Hospitals
9315	Clothing manufacturing	9623	Nursing and residential care facilities
9316	Leather and allied product manufacturing	9624	Social assistance
9321	Wood product manufacturing	9625	Total Service
9322	Paper manufacturing	9711	Performing arts spectator sports and related industries
9323	Printing and related support activities	9712	Heritage institutions
9324	Petroleum and coal products manufacturing	9713	Amusement gambling and recreation industries
9325	Chemical manufacturing	9721	Accommodation services
9326	Plastics and rubber products manufacturing	9722	Food services and drinking places
9327	Non-metallic mineral product manufacturing	9723	Total Tourism
9331	Primary Metal Manufacturing	9811	Repair and maintenance
9332	Fabricated metal product manufacturing	9812	Personal and laundry services
9333	Machinery manufacturing	9813	Religious grant-making civic and professional and similar organizations
9334	Computer and Electronic manufacturing	9814	Private households
9335	Electrical equipment appliance and component manufacturing	9815	Total other services
9336	Transportation equipment manufacturing	9911	Federal government public administration

Code	Description	Code	Description
9337	Furniture and related product manufacturing	9912	Provincial and territorial public administration (9121 to 9129)
9338	Miscellaneous manufacturing	9913	Local municipal and regional public administration (9131 to 9139)
9339	Total Manufacturing	9914	Aboriginal public administration
9411	Farm product wholesaler-distributors	9919	International and other extra-territorial public administration
9412	Petroleum product wholesaler-distributors	9920	Total Government
9413	Food beverage and tobacco wholesaler-distributors	9921	Commercial Fuel Combustion
9414	Personal and household goods wholesaler-distributors	9922	TOTAL DISTRIBUTION AND RETAIL
9415	Motor vehicle and parts wholesaler-distributors	9923	TOTAL INSTITUTIONAL AND GOVERNEMNT
9416	Building material and supplies wholesaler-distributors	9924	Primary Industry
9417	Machinery equipment and supplies wholesaler-distributors	9925	Manufacturing and Assembly
9418	Miscellaneous wholesaler-distributors	9926	Distribution and Retail (no petroleum)
9419	Wholesale agents and brokers	9927	Commercial Services
9420	Total Wholesale	9928	Commercial Meat cooking
9441	Motor vehicle and parts dealers	9929	HIGHJET
9442	Furniture and home furnishings stores	9930	LOWMEDJET
9443	Electronics and appliance stores	9931	OTHERJET
9444	Building material and garden equipment and supplies dealers	9932	CANRAIL
9445	Food and beverage stores	9933	Forest fires
9446	Health and personal care stores	9941	PAVED ROADS
9447	Gasoline stations	9942	UNPAVED ROADS
9448	clothing and clothing accessories stores	9943	HIGHWAY
9451	Sporting goods hobby book and music stores	9944	ROAD
9452	General Merchandise stores	9945	Commercial Marine Vessels
9453	Miscellaneous store retailers	9946	Construction and mining
9454	Non-store retailers	9947	Agriculture Construction and mining
9455	Total Retail	9950	Intersection of Forest and Housing
9481	Air transportation	9960	TOTBEEF
9482	Rail transportation	9970	TOTPOUL

Code	Description	Code	Description
9483	Water Transportation	9980	TOTSWIN
9484	Truck transportation	9990	TOTFERT
9485	Transit and ground passenger transportation	9993	Trail
9486	Pipeline transportation	9994	ALLROADS
9487	Scenic and sightseeing transportation	9995	30UNPAVED_70trail
9488	Support activities for transportation	9996	Urban area
9491	Postal service	9997	CHBOISQC
9492	Couriers and messengers	9991	Traffic

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.

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

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

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. Available at: http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&ved=0CDAQFjAC&url=ftp%3A%2F%2Fwsip-70-164-45-196.dc.dc.cox.net%2Fpublic%2FSESARM%2FRevised%2520Final%2FSESARM%2520Base%2520Year%2520Revised%2520Final%2520Report_Jan2012.docx&ei=xU-AUPu1F4WA0AHC5YHYCg&usg=AFQjCNFhjgx3Ej-hbfYmMUP4zGI_HBiqZA&sig2=hWWN0m3WYPSO28QSzn5BIA.

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

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

4.0 CMAQ Air Quality Model Estimates

4.1 Introduction to the CMAQ Modeling Platform

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

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

- Improve the environmental management community's ability to evaluate the impact of air quality management practices for multiple pollutants at multiple scales.

- Improve the scientist's ability to better probe, understand, and simulate chemical and physical interactions in the atmosphere.

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

CMAQ was also designed to have multi-scale capabilities so that separate models were not needed for urban and regional scale air quality modeling. The grid spatial resolutions in past annual CMAQ runs have been 36 km x 36 km per grid for the "parent" domain, and nested within that domain are 12 km x 12 km grid resolution domains. The parent domain typically covered the

continental United States, and the nested 12 km x 12 km domain covered the Eastern or Western United States. The CMAQ simulation performed for this 2010 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 recommended for most applications as the highest resolution. 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/CMAQ or http://www.cmascenter.org.

4.1.1 Advantages and Limitations of the CMAQ Air Quality Model

An advantage of using the CMAQ model output for characterizing air quality for use in comparing with health outcomes is that it provides a complete spatial and temporal coverage across the U.S. CMAQ is a three-dimensional Eulerian photochemical air quality model that simulates the numerous physical and chemical processes involved in the formation, transport, and destruction of ozone, particulate matter and air toxics for given input sets of initial and boundary conditions, meteorological conditions and emissions. The CMAQ model includes state-of-the-science capabilities for conducting urban to regional scale simulations of multiple air quality issues, including tropospheric ozone, fine particles, toxics, acid deposition and visibility 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 summarized in Simon et al., 2012[6].

---

6 Heather Simon, Kirk R. Baker, Sharon Phillips. (2012) Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012. *Atmospheric Environment* **61**, 124-139. Online publication date: 1-Dec-2012.

4.2   CMAQ Model Version, Inputs and Configuration

This section describes the air quality modeling platform used for the 2010 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[7] model as part of the 2010 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 2010 CMAQ simulation along with the results of a model performance evaluation in which the 2010 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 2010 analysis employed CMAQ version 4.7.1[8] which reflects updates to version 4.7 to improve the underlying science which include aqueous chemistry mass conservation improvements and improved vertical convective mixing. . The CMAQ model version 4.7 was most recently peer-reviewed in February of 2009 for the U.S. EPA.[9] The model enhancements in version 4.7.1 also include:

1. Aqueous chemistry
   - Mass conservation improvements
     - Imposed one second minimum timestep for remainder of the cloud lifetime after 100 'iterations' in the solver
     - Force mass balance for the last timestep in the cloud by limiting oxidized amount to mass available

---

7 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.

8 CMAQ version 4.7.1 model code is available from the Community Modeling and Analysis System (CMAS) at: http://www.cmascenter.org.

9 Allen, D., Burns, D., Chock, D., Kumar, N., Lamb, B., Moran, M. (February 2009 Draft Version). Report on the Peer Review of the Atmospheric Modeling and Analysis Division, NERL/ORD/EPA. U.S. EPA, Research Triangle Park, NC. CMAQ version 4.7 was released on December, 2008. It is available from the Community Modeling and Analysis System (CMAS) as well as previous peer-review reports at: http://www.cmascenter.org.

- Implemented steady state assumption for OH
- Only allow sulfur oxidation to control the aqueous chemistry solver timestep (previously, reactions of OH, GLY, MGLY, and Hg for multipollutant model also controlled the timestep)

2. Advection
   - Added additional divergence-based constraint on advection timestep
   - Vertical advection in the Yamo module is now represented with the PPM scheme to limit numerical diffusion

3. Model time step determination
   - Fixed a potential advection time step error
     - The sum of the advection steps for a given layer time step might not equal the output time step duration in some extreme cases
     - Ensured that the advection steps sum up to the synchronization step

4. Horizontal diffusion
   - Fixed a potential error
     - Concentration data may not be correctly initialized if multiple sub-cycle time steps are required
     - Fix to initialize concentrations with values calculated in the previous sub-time step

5. Emissions
   - Bug fix in EMIS_DEFN.F to include point source layer 1 NH3 emissions
   - Bug fix to calculate soil NO "pulse" emissions in BEIS
   - Remove excessive logging of cases where ambient air temperature exceeds 315.0 Kelvin. When this occurs, the values are just slightly over 315
   - Bug fix for parallel decomposition errors in plume rise emissions

6. Photolysis
   - JPROC/phot_table and phot_sat options
     - Expanded lookup tables to facilitate applications across the globe and vertical extent to 20km
     - Updated temperature adjustments for absorption cross sections and quantum yields
     - Revised algorithm that processes TOMS datasets for OMI data format
   - In-line option
     - Asymmetry factor calculation updated using values from Mie theory integrated over log normal particle distribution; added special treatment for large particles in asymmetry factor algorithm to avoid numerical instabilities

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 19 km) 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	459 x 299 x 25
Vertical Extent	25 Layers: Surface to 50 mb level (see Table 4-2)

12km CONUS nationwide domain
x,y: -2556000,-1728000
col: 459 row: 299

Figure 4-1. Map of the CMAQ Modeling Domain. The blue 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 2010. The 2010 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

2010 Emissions: The emissions inventories used in the 2010 air quality modeling are described in Section 3, above.

Meteorological Input Data: The gridded meteorological data for the entire year of 2010 at the 12 km continental United States scale domain was derived from version 3.4 of the Weather Research and Forecasting Model (WRF), Advanced Research WRF (ARW) core.[1] 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 2010 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 and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme.[2]

The WRF meteorological outputs were processed using the Meteorology-Chemistry Interface Processor (MCIP) package[3], version 4.1.2, to derive the specific inputs to CMAQ: horizontal wind components (i.e., speed and direction), temperature, moisture, and its related speciated components was conducted for vertical diffusion rates, and rainfall rates for each grid cell in each vertical layer. The WRF simulation used the same CMAQ map projection, a Lambert Conformal projection centered at (-97, 40) with true latitudes at 33 and 45 degrees north. The 12 km WRF domain consisted of 459 by 299 grid cells and 35 vertical layers with a surface layer of approximately 38 meters. Table 4-2 shows the vertical layer structure used in WRF and the layer

---

[1] 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.

[2] 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.

[3] 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.

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 2010 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 2010 WRF and CMAQ simulations (heights are layer top).

CMAQ Layers	WRF Layers	Sigma P	Approximate Height (m)
25	35	0.0000	17,556
	34	0.0500	14,780
24	33	0.1000	12,822
	32	0.1500	11,282
23	31	0.2000	10,002
	30	0.2500	8,901
22	29	0.3000	7,932
	28	0.3500	7,064
21	27	0.4000	6,275
	26	0.4500	5,553
20	25	0.5000	4,885
	24	0.5500	4,264
19	23	0.6000	3,683
18	22	0.6500	3,136
17	21	0.7000	2,619
16	20	0.7400	2,226
15	19	0.7700	1,941
14	18	0.8000	1,665
13	17	0.8200	1,485
12	16	0.8400	1,308
11	15	0.8600	1,134
10	14	0.8800	964
9	13	0.9000	797
	12	0.9100	714
8	11	0.9200	632
	10	0.9300	551

7	9	0.9400	470
	8	0.9500	390
6	7	0.9600	311
5	6	0.9700	232
4	5	0.9800	154
	4	0.9850	115
3	3	0.9900	77
2	2	0.9950	38
1	1	0.9975	19
0	0	1.0000	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[10] 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). This model was run for 2010 with a grid resolution of 2.0 degrees x 2.5 degrees (latitude-longitude) and 46 vertical layers up to 0.01 hPa. The predictions were processed using the GEOS-2-CMAQ tool and used to provide one-way dynamic boundary conditions at one-hour intervals.[11] Ozone was evaluated from these GEOS-Chem runs by comparing to satellite vertical profiles and ground-based measurements and found acceptable model performance. More information is available about the GEOS-CHEM model and other applications using this tool at: http://www-as.harvard.edu/chemistry/trop/geos.

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 20010 simulation using state/local monitoring sites data in order to estimate the ability of the CMAQ modeling system to replicate the 2010 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, normalized mean bias and fractional bias; and two error metrics, normalized mean error and fractional error. Normalized mean bias (NMB) is used as a normalization to facilitate a range of concentration magnitudes. This statistic

---

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

11 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.

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:

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

Fractional bias is defined as:

$$FB = \frac{1}{n}\left(\frac{\sum_{1}^{n}(P-O)}{\sum_{1}^{n}\left(\frac{(P+O)}{2}\right)}\right)*100$$

FB is a useful model performance indicator because it has the advantage of equally weighting positive and negative bias estimates. The single largest disadvantage in this estimate of model performance is that the estimated concentration (i.e., prediction, P) is found in both the numerator and denominator.

Fractional error (FE) is similar to fractional bias except the absolute value of the difference is used so that the error is always positive. Fractional error is defined as:

$$FE = \frac{1}{n}\left(\frac{\sum_{1}^{n}|P-O|}{\sum_{1}^{n}\left(\frac{(P+O)}{2}\right)}\right)*100$$

In addition to the performance statistics, regional maps which show the normalized mean bias and error 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 2010 in the continental U.S. were included in the evaluation and were taken from the 2010 State/local monitoring site data in the Air Quality System (AQS) Aerometric Information Retrieval System (AIRS). 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. domain12: 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 normalized mean bias and error for individual monitors are shown in Figures 4-2 through 4-3. 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 2010 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. Specifically, 8-hour ozone in the summer is slightly over predicted with the greatest over prediction in Central U.S. (NMB is 20.6 percent). Ozone performance in spring shows better performance with slight over predictions in most of the subregions except in the Northeast (slight under prediction of 2.2). In the winter, when concentrations are generally low, the model slightly under predicts 8-hour ozone with the exception of the West (NMB is 5.6). In the fall, when concentrations are also relatively low, ozone is also slightly over predicted; with NMBs less than 7 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. The information in Figure 4-2 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 and in Seattle, WA. 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 4-2 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 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 Figure 4-3, is 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 and Maine, Baton Rouge, LA, Cleveland, OH, and Seattle, WA.

---

12 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.

Table 4-4. Summary of CMAQ 20010 8-Hour Daily Maximum Ozone Model Performance Statistics by Subregion, by Season.

Subregion	Season	No. of Obs	NMB (%)	NME (%)	FB (%)	FE (%)
Northeast	Winter	6,015	-20.3	22.9	-23.0	26.7
	Spring	13,825	-2.2	12.2	-2.0	13.1
	Summer	16,359	3.3	14.1	3.9	14.6
	Fall	11,373	4.1	16.2	6.1	17.5
Midwest	Winter	2,720	-23.6	27.7	-27.0	32.7
	Spring	12,066	3.9	12.0	4.6	12.5
	Summer	16,389	8.3	16.3	7.8	16.1
	Fall	9,666	4.0	14.2	5.5	15.2
Central States	Winter	11,505	-4.6	15.8	-5.0	17.5
	Spring	16,330	4.1	13.2	4.9	13.7
	Summer	17,956	20.6	26.7	20.6	26.2
	Fall	14,781	3.4	16.8	4.9	17.3
Southeast	Winter	7,112	-4.7	12.5	-4.6	13.5
	Spring	20,567	4.9	12.8	6.1	13.5
	Summer	22,162	17.4	22.4	18.1	22.4
	Fall	16,908	5.3	15.3	6.7	15.9
West	Winter	22,696	5.6	20.9	8.2	23.0
	Spring	27,849	1.4	11.6	2.1	12.1
	Summer	30,279	7.2	16.0	7.6	16.7
	Fall	22,984	6.9	15.6	7.7	16.2

O3_8hrmax NMB (%) for run 2010ef_v5_10f_12US1 for 20100501 to 20100930
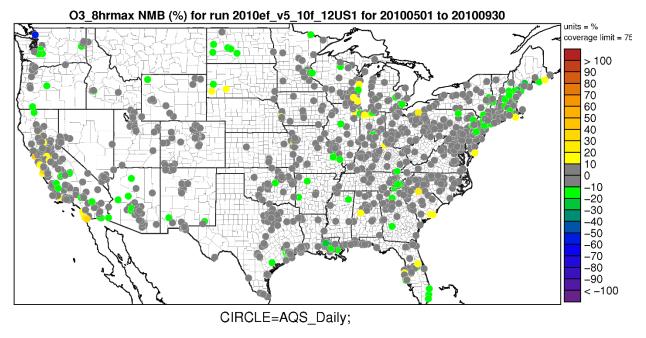
CIRCLE=AQS_Daily;

Figure 4-2. Normalized Mean Bias (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2010 at monitoring sites in the continental U.S. modeling domain.
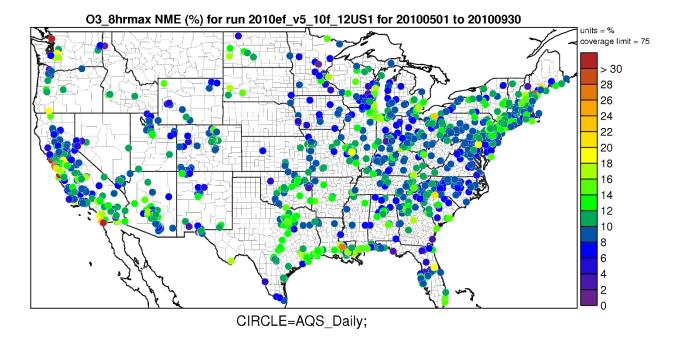

O3_8hrmax NME (%) for run 2010ef_v5_10f_12US1 for 20100501 to 20100930

CIRCLE=AQS_Daily;

Figure 4-3. Normalized Mean Error (%) of 8-hour daily maximum ozone greater than 60 ppb over the period May-September 2010 at monitoring sites in the continental U.S. modeling domain.

Evaluation for Annual PM$_{2.5}$: The PM evaluation focuses on PM$_{2.5}$ total mass and its 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 PM$_{2.5}$ bias and error performance statistics were calculated on an annual basis for each subregion (Table 4-5). PM$_{2.5}$ measurements for 2010 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 PM$_{2.5}$ and its 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 normalized mean bias and error by site for each PM$_{2.5}$ species are provided in Figures 4-4 through 4-17.

As indicated by the statistics in Table 4-5, annual CMAQ PM$_{2.5}$ for 2009 shows over predictions at rural IMPROVE monitoring sites and urban CSN monitoring sites in each subregion except in the Southeast at CSN sites (which shows an under prediction in NMB of 17 percent) and in the Southeast, Central and West at IMPROVE sites (which shows a slight under prediction in NMB of 2 to 11 percent). Although not shown here, the mean observed concentrations of PM$_{2.5}$ are more than twice as high at the CSN sites (~11µg m^{-3}) as the IMPROVE sites (~5 µg m^{-3}), thus illustrating the statistical differences between the urban CSN and rural IMPROVE networks.

Annual average sulfate is consistently under predicted at CSN, IMPROVE, and CASTNet monitoring sites across the modeling domain, with NMB values ranging from near negligible to -29 percent. Sulfate performance shows moderate error, ranging from 24 to 45 percent. Figures 4-6 and 4-7, 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 70 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 (NMB in the range of 3 to 40 percent), except in the West where nitrate is under predicted on average by 26 percent. The bias statistics indicate that the model performance for nitrate is generally best at the urban CSN monitoring sites. Model performance of total nitrate at sub-urban CASTNet monitoring sites shows an over prediction across all subregions (NMB in the range of 2 to 27 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 4-8. 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 Figure 4-9, 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 -7 percent) and over predict across the urban CSN sites (ranging from 6 to 13 percent). 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 shows bias within ±30 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 under predicted across most subregions in rural IMPROVE areas (NMB ranging from -12 to -17 percent), except in the Northeast where the bias on average is 30.5 percent. However, the model over predicted annual average organic carbon in most subregions at urban CSN sites (NMB ranging from 1 to 34 percent), except in the Southeast where the bias on average is -5.7 percent. 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 (42 to 81 percent).

Table 4-5. Summary of CMAQ 2010 Annual PM Species Model Performance Statistic

Pollutant	Monitor Network	Subregion	No. of Obs	NMB (%)	NME (%)	FB (%)	FE (%)
PM₂.₅	CSN	Northeast	2,882	0.7	39.9	1.1	39.4
		Midwest	2,210	9.5	36.6	6.2	35.6
		Southeast	2,441	-17.0	35.6	-23.5	39.7
		Central	1,671	3.9	39.0	-0.1	38.4
		West	1,297	0.6	58.7	-3.0	54.0
	IMPROVE	Northeast	2,101	9.0	46.2	6.0	44.2
		Midwest	562	0.4	38.2	-1.4	40.0
		Southeast	1,940	-10.5	38.4	-17.9	43.3
		Central	2,429	-2.4	40.9	-7.2	42.3
		West	10,060	-8.7	49.4	-7.7	48.3
Sulfate	CSN	Northeast	2,885	-10.9	30.3	-0.1	30.7
		Midwest	2,226	-6.6	32.1	-0.5	32.4

Pollutant	Monitor Network	Subregion	No. of Obs	NMB (%)	NME (%)	FB (%)	FE (%)
		Southeast	2,453	-20.5	30.9	-22.8	35.4
		Central	1,789	-11.9	37.4	-7.9	39.3
		West	1,308	1.0	43.1	5.6	42.2
	IMPROVE	Northeast	2,035	0.1	32.4	13.6	36.5
		Midwest	583	-12.2	32.8	1.9	37.3
		Southeast	1,919	-17.5	30.3	-17.5	35.3
		Central	2,261	-14.0	35.2	-6.7	37.6
		West	9,832	-0.1	45.1	15.6	48.7
	CASTNet	Northeast	758	-15.7	22.3	-9.3	22.4
		Midwest	618	-20.4	23.7	-18.8	24.7
		Southeast	1,095	-24.5	26.7	-28.3	31.5
		Central	378	-28.9	31.4	-29.1	34.3
		West	1,017	-17.5	32.9	-9.8	36.3
	CSN	Northeast	2,885	-0.5	57.2	-37.4	81.8
		Midwest	2,226	-0.3	49.3	-21.6	67.8
		Southeast	2,453	17.5	81.1	-50.0	101.0
		Central	1,690	-4.1	53.4	-45.0	88.7
		West	1,308	-40.6	71.6	-66.8	94.5
Nitrate	IMPROVE	Northeast	2,031	40.3	89.2	-29.8	97.5
		Midwest	580	2.6	64.4	-40.8	92.6
		Southeast	1,919	22.2	95.2	-59.7	116.0
		Central	2,261	15.4	62.4	-32.0	92.9
		West	9,765	-26.4	80.0	-82.3	121.0
	CASTNet	Northeast	758	27.2	37.9	20.3	38.8
		Midwest	618	7.6	33.3	13.8	30.0
Total Nitrate (NO$_3$ + HNO$_3$)		Southeast	1,095	18.9	42.3	13.8	42.4
		Central	378	4.7	31.3	2.0	30.7
		West	1,016	1.6	34.8	12.5	39.3
Ammonium	CSN	Northeast	2,885	7.9	40.8	34.3	44.2

Pollutant	Monitor Network	Subregion	No. of Obs	NMB (%)	NME (%)	FB (%)	FE (%)
		Midwest	2,226	10.0	38.1	21.6	41.5
		Southeast	2,453	9.6	37.2	12.9	38.0
		Central	1,789	13.0	44.6	20.7	47.1
		West	1,308	5.8	77.6	30.3	65.2
	CASTNet	Northeast	758	0.0	27.7	2.0	28.0
		Midwest	618	-2.2	27.3	3.7	27.0
		Southeast	1,095	-6.7	25.9	-7.5	28.8
		Central	378	-0.3	30.7	-0.5	32.9
		West	1,017	-6.8	38.3	-4.4	40.9
Elemental Carbon	CSN	Northeast	2,638	32.9	59.2	29.4	48.5
		Midwest	2,191	49.6	68.4	35.1	52.5
		Southeast	2,403	9.6	47.6	12.1	45.6
		Central	1,629	74.8	91.8	46.6	61.2
		West	1,227	50.9	87.3	32.8	65.7
	IMPROVE	Northeast	2,117	24.6	53.4	11.5	46.8
		Midwest	583	13.0	46.0	2.7	48.1
		Southeast	1,924	4.0	46.9	-1.3	45.6
		Central	2,506	17.1	52.5	7.7	43.4
		West	10,118	22.3	72.7	8.0	56.8
Organic Carbon	CSN	Northeast	2,112	1.0	61.1	-6.4	56.3
		Midwest	2,180	16.7	54.7	12.3	49.1
		Southeast	2,390	-5.7	42.6	-10.2	44.0
		Central	1,624	23.0	59.2	13.9	48.7
		West	1,218	34.0	80.6	27.9	63.9
	IMPROVE	Northeast	2,624	30.5	65.9	22.7	55.6
		Midwest	575	-11.6	48.6	-17.5	54.8
		Southeast	1,931	-16.3	47.7	-33.3	54.6
		Central	2,502	-17.2	51.5	-30.5	55.5
		West	9,919	-15.1	59.9	-12.9	58.1

PM_TOT NMB (%) for run 2010ef_v5_10f_12US1 for 20100101 to 20101231

CIRCLE=IMPROVE; TRIANGLE=CSN;

Figure 4-4. Normalized Mean Bias (%) of annual PM$_{2.5}$ mass at monitoring sites in the continental U.S. modeling domain.

Figure 4-5. Normalized Mean Error (%) of annual PM$_{2.5}$ mass at monitoring sites in the continental U.S. modeling domain.
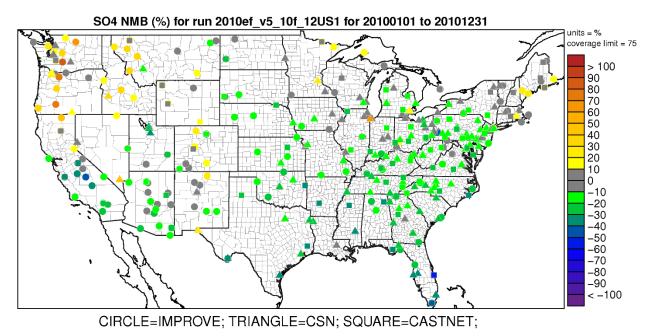

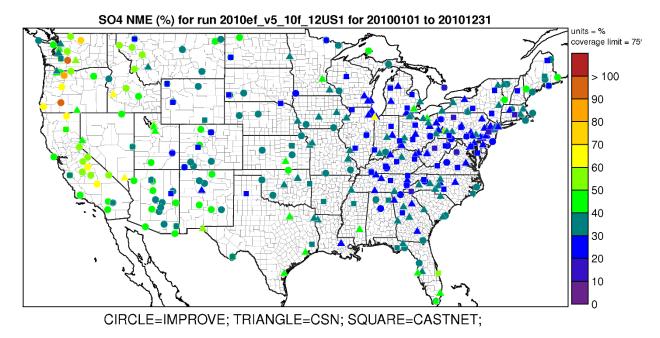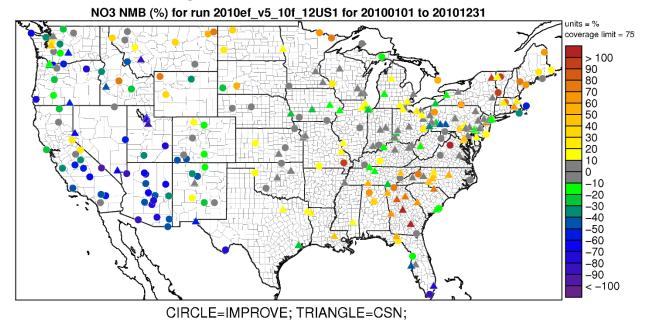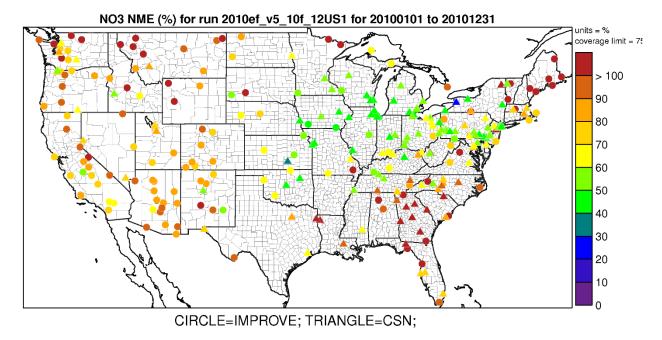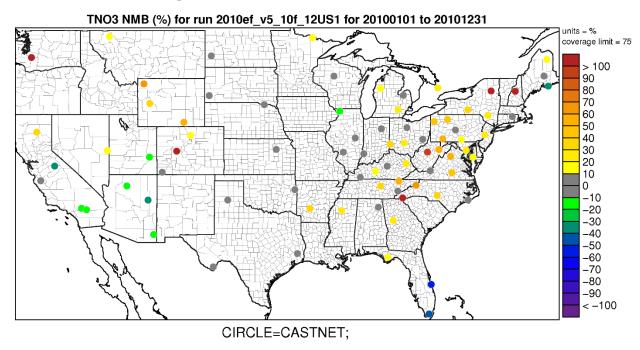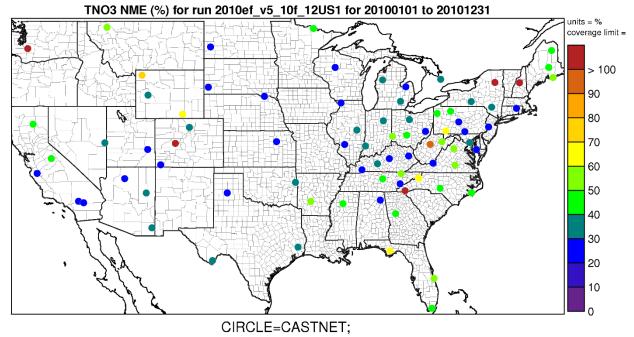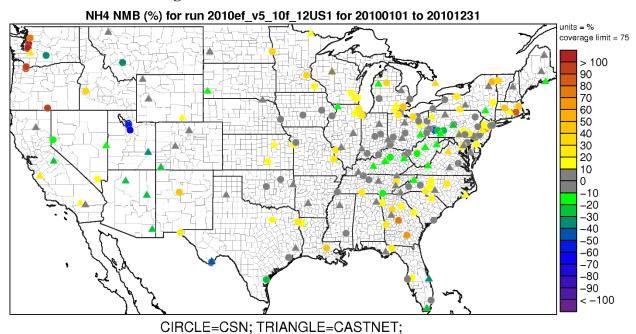**Figure 4-6. Normalized Mean Bias (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.**
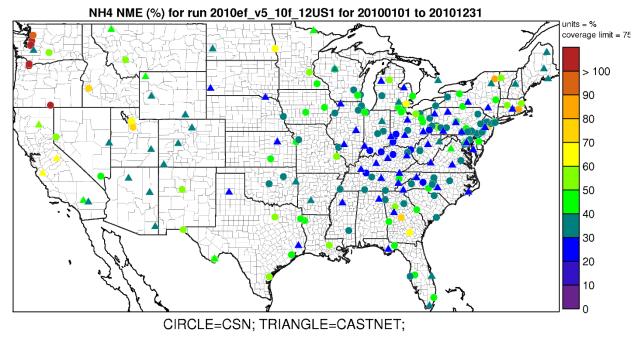
SO4 NME (%) for run 2010ef_v5_10f_12US1 for 20100101 to 20101231

CIRCLE=IMPROVE; TRIANGLE=CSN; SQUARE=CASTNET;

Figure 4-7. Normalized Mean Error (%) of annual sulfate at monitoring sites in the continental U.S. modeling domain.

NO3 NMB (%) for run 2010ef_v5_10f_12US1 for 20100101 to 20101231
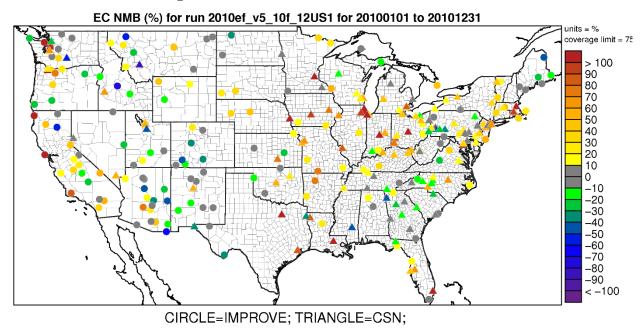
CIRCLE=IMPROVE; TRIANGLE=CSN;

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

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

Figure 4-10. Normalized Mean Bias (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.
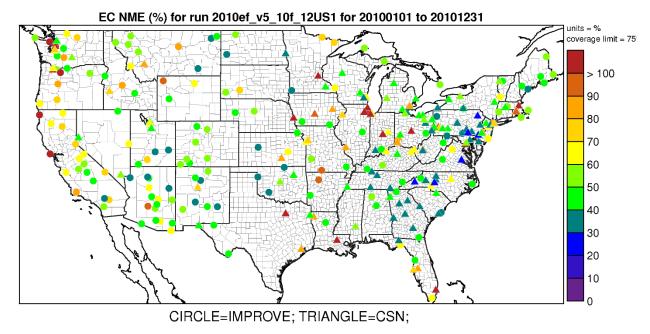
Figure 4-11. Normalized Mean Error (%) of annual total nitrate at monitoring sites in the continental U.S. modeling domain.
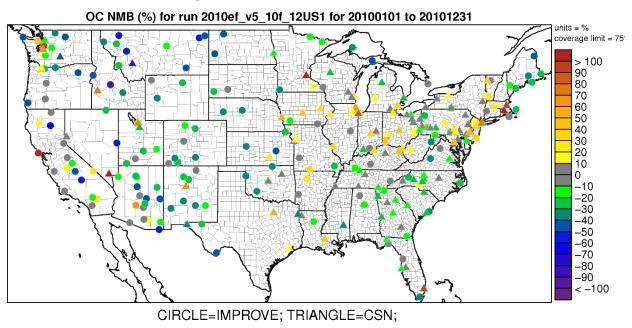

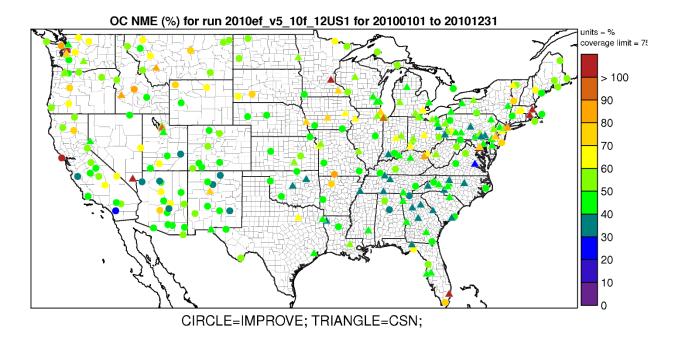**Figure 4-12. Normalized Mean Error (%) of annual ammonium at monitoring sites in the continental U.S. modeling domain.**

CIRCLE=CSN; TRIANGLE=CASTNET;

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

CIRCLE=IMPROVE; TRIANGLE=CSN;

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

Figure 4-15. Normalized Mean Error (%) of annual elemental carbon at monitoring sites in the continental U.S. modeling domain.

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

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

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

5.1 Introduction

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

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*.
~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.
~β₀(s,t) is the intercept, and is composed of both a global and a local component.
β₁(t) is the global slope; local components of the slope are contained in the *~x(s,t)* term.
ε(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 *~β₀* and *β₁* 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 2010 US census tract centroids across the continental U.S. using 2010 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 2010. 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 2010 about half of the US Census tracts (36206 out of 72283) experienced at least one day with an ozone value above the NAAQS of 75 ppb.
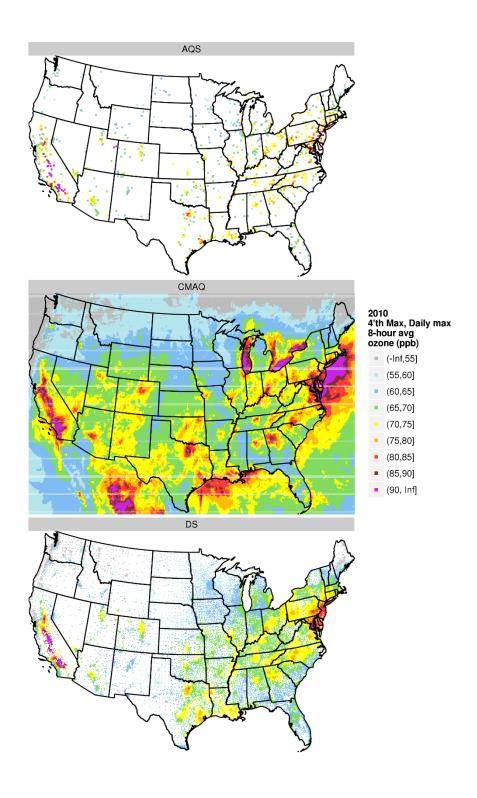
Figure 5-1. Annual 4th max (daily max 8-hour ozone concentrations) derived from AQS, CMAQ and DS data.

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 2010. Figure 5-2 shows annual means and Figure 5-3 shows 98'th percentiles of 24-hour PM2.5 concentrations for AQS observations, CMAQ model predictions and DS model results. The DS model estimated that for 2010 about 33% of the US Census tracts (23547 out of 72283) experienced at least one day with a PM2.5 value above the 24-hour NAAQS of 35 ug/m3.
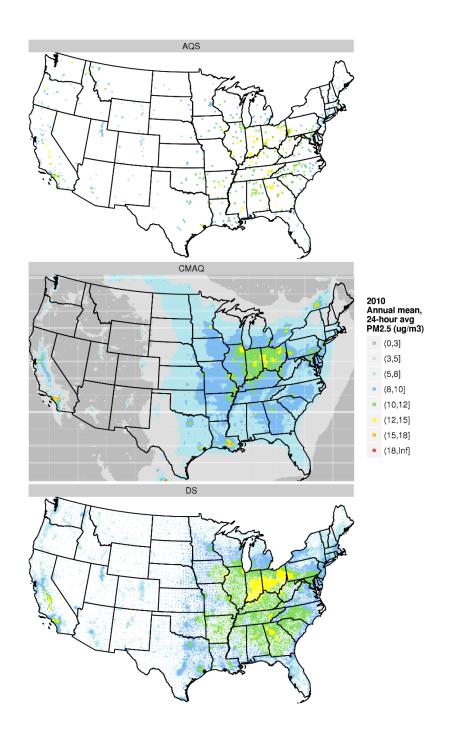
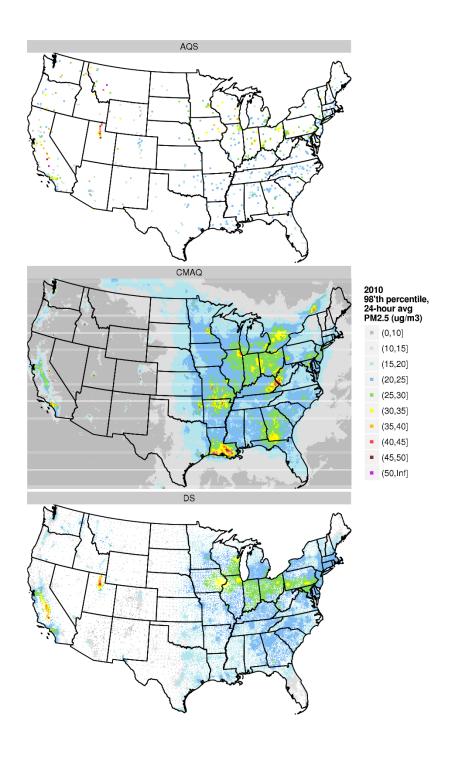**Figure 5-2. Annual mean PM₂.₅ concentrations derived from AQS, CMAQ and DS data.**

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

5.4 Downscaler Uncertainties

5.4.1 Standard Errors

As mentioned above, the DS model works by drawing random samples from built-in distributions during its parameter estimation. The standard errors associated with each of these populations provide a measure of uncertainty associated with each concentration prediction. Figure 5-4 shows the percent errors resulting from dividing the DS standard errors by the associated DS prediction. The black dots on the maps show the location of EPA sampling network monitors whose data was input to DS via the AQS datasets (Chapter 2). The maps show that, in general, errors are relatively smaller in regions with more densely situation monitors (ie the eastern US), and larger in regions with more sparse monitoring networks (ie western states). These standard errors could potentially be used to estimate the probability of an exceedance for a given point estimate of a pollutant concentration.
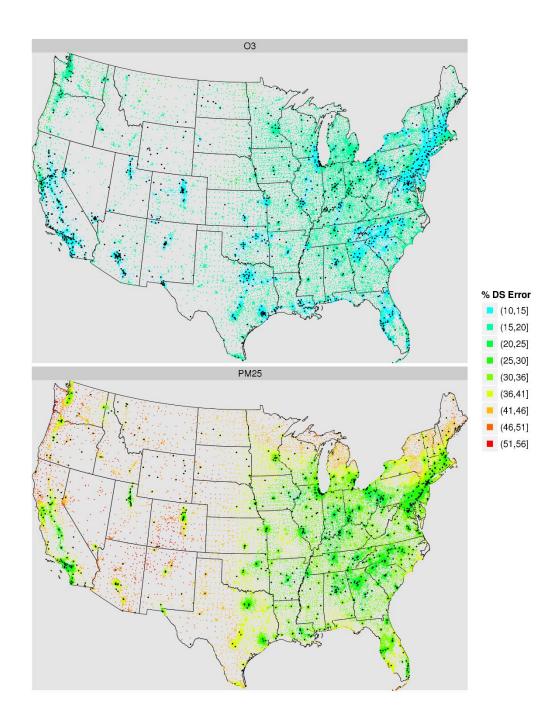
Figure 5-4. Annual mean relative errors (standard errors divided by predictions) from the DS 2010 runs. The black dots show the locations of monitors that generated the AQS data used as input to the DS model.

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	901	6.20e-3	2.88	0.96
O3	1239	1.35e-2	5.12	0.96

Table 5-1. Cross-validation statistics associated with the 2010 DS runs.

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

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

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

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

5.5   Summary and Conclusions

The results presented in this report are from an application of the DS fusion model for characterizing national air quality for Ozone and $PM_{2.5}$. DS provided spatial predictions of daily ozone and $PM_{2.5}$ at 2010 U.S. census tract centroids by utilizing monitoring data and CMAQ output for 2010. 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