Technical Support Document
EPA's 2014 National Air Toxics Assessment

2014 NATA TSD

August 2018

Office of Air Quality Planning and Standards
Research Triangle Park
North Carolina 27711

Contents

Tables

Figures

Common Acronyms and Abbreviations

µg/m^3	microgram(s)/cubic meter
AERMOD	atmospheric dispersion model developed by the American Meteorological Society and the U.S. Environmental Protection Agency's Regulatory Model Improvement Committee
ASPEN	Assessment System for Population Exposure Nationwide
ATSDR	Agency for Toxic Substances and Disease Registry
CAA	Clean Air Act
CHAD	Consolidated Human Activity Database
CAP	criteria air pollutant
CMAQ	Community Multiscale Air Quality model
CONUS	Contiguous United States (modeling domain for CMAQ)
EC	exposure concentration
EPA	Environmental Protection Agency
EGU	electricity generating unit
HAP	hazardous air pollutant
HAPEM	Hazardous Air Pollutant Exposure Model
HEM	Human Exposure Model
HI	hazard index
HQ	hazard quotient
IRIS	Integrated Risk Information System
ISC	Industrial Source Complex
MOVES	Motor Vehicle Emissions Simulator
NATA	National-scale Air Toxics Assessment
NEI	National Emissions Inventory
NMIM	National Mobile Inventory Model
OAQPS	Office of Air Quality Planning and Standards
PAH	polycyclic aromatic hydrocarbon
PM	particulate matter
PM$_{2.5}$	particulate matter less than 2.5 microns in diameter
PM$_{10}$	particulate matter less than 10 microns in diameter
POM	polycyclic organic matter
RfC	reference concentration
RTR	Risk and Technology Review
SCC	Source Classification Code
S/L/T	state, local or tribal agency
URE	unit risk estimate
WRF	Weather Research Forecasting model

1. Introduction

1.1. Overview

The National Air Toxics Assessment (NATA) is the U.S. Environmental Protection Agency's (EPA; referred to throughout this document as "we") ongoing thorough evaluation of air toxics across the United States. EPA developed NATA as a state-of-the-science tool to inform both national and localized efforts to collect air toxics information, characterize emissions and help prioritize pollutants and areas of interest for further study to gain a better understanding of risks. The goal of NATA is to identify those air toxics which are of greatest potential concern in terms of contribution to population risk. Ambient and exposure concentrations and estimates of risk and hazard for air toxics in each state are typically generated at the census tract level. The 2014 NATA is the sixth version of this national assessment.

This NATA Technical Support Document (TSD) describes the data and approaches EPA used to conduct the 2014 NATA, including descriptions of how we:

■ compiled emissions data and prepared them for use as model inputs (Section 2);

■ estimated ambient concentrations of air toxics (Section 3);

■ estimated exposures to air toxics for populations (Section 4);

■ selected toxicity values (Section 5);

■ characterized human-health risks and hazards (Section 6); and

■ addressed variability and uncertainty (Section 7).

References to additional documents are included (Section 8) to facilitate access to more detailed technical information on the emissions inventories, dispersion modeling, photochemical modeling, exposure modeling and toxicity values.

The TSD also includes the following appendixes:

■ Appendix A – a glossary of the key terms and their definitions

■ Appendix B – a list of air toxics included in NATA

■ Appendix C – a list of source classification codes (SCCs) for diesel particulate matter (diesel PM)

■ Appendix D – procedures used to estimate NATA background concentrations

■ Appendix E – additional model evaluation summaries

■ Appendix F – documentation on HAPEM7 and its use in NATA

We also provide a "Supplemental Data" folder with this document that contains the Microsoft® Access™ and Microsoft® Excel™ files referenced throughout this TSD.

This TSD satisfies basic documentation protocol expected of EPA products and provides a resource for the technically oriented user community by summarizing the data sources, methods, models and assumptions used in the 2014 NATA. This document does not provide quantitative results for NATA and thus presents no exposure or risk estimates. You can find results and other specific information for NATA, including for the 2014 NATA and previous assessments, on the NATA website (https://www.epa.gov/national-air-toxics-assessment).

1.2. Purpose and Overview of NATA Steps

Each NATA involves the same basic steps needed to produce the final assessment. These NATA steps are depicted in Figure 1-1.
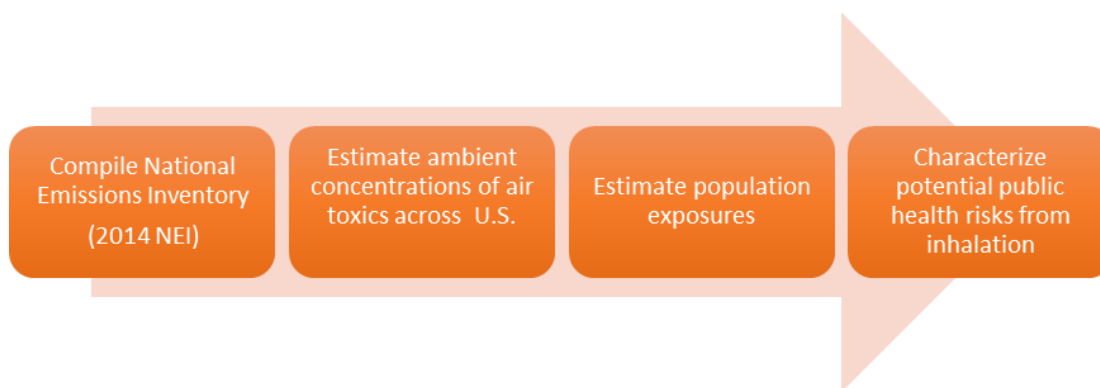
Figure 1-1. 2014 NATA – basic steps

The first and most time-consuming step in NATA is assembling the National Emissions Inventory (NEI), a detailed, nationwide inventory of air toxics emissions. This NATA's emissions data are from the 2014 NEI version 2. The NEI includes emissions from point, nonpoint and mobile sources, as well as emissions from biogenic sources and fires. These source data are the foundation of NATA's air quality modeling. Section 2 of this document details the steps EPA used to construct this extensive database.

After preparing the NEI emissions and other needed data (e.g., meteorological data), we use these data as inputs to two air quality models used to estimate ambient air concentrations of air toxics: the American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD) atmospheric dispersion model and the Community Multiscale Air Quality (CMAQ) photochemical model. AERMOD is used for all NATA air toxics modeled, and CMAQ is used for a list of 52 air toxics that are incorporated into CMAQ multipollutant version 5.2. CMAQ provides the overall mass, chemistry and formation for hazardous air pollutants (HAPs) formed secondarily in the atmosphere (e.g., formaldehyde, acetaldehyde and acrolein), whereas AERMOD provides spatial granularity and more detailed source attribution. CMAQ also provides the biogenic and fire concentrations, as these sources are not run in AERMOD. Only AERMOD is run in areas outside the lower 48, or contiguous, U.S. states (referred to as CONUS) that are included in NATA: Alaska, Hawaii, Puerto Rico and the Virgin Islands. Special steps are taken to estimate secondary HAPs, fires and biogenics in these areas. Section 3 of this document details the steps EPA used to model for NATA.

For the HAPs modeled in both CMAQ and AERMOD, we combine the model-calculated annual average concentrations using a hybrid approach. We next use these concentrations, along with the other

concentrations calculated by AERMOD, to prepare census tract-level concentrations of all modeled air toxics. Then, using the HAPEM7 exposure model, we account for human activity patterns and develop exposure concentrations, or ECs, for each census tract. Finally, we estimate census tract-level risks by applying health benchmark data to the ECs. Sections 4 through 6 of this TSD detail these steps.

In every NATA assessment, EPA's state, local and tribal (S/L/T) air agency partners play an integral role. First, S/L/T specialists review early versions of NEI's source data for their area, working with local industry and other emissions sources to develop and forward corrections to their area's emissions data. They also review preliminary risk results during early stages of the NATA modeling process, which often reveals other inaccuracies in the data. This review and feedback process helps ensure that NATA's input data are as accurate as possible in the final version.

For the 2014 NATA, we first modeled using version 1 of the 2014 NEI (v1), performing the initial modeling runs in late 2016. Our S/L/T partners reviewed these v1 source data as well as two rounds of first-pass modeling and risk results. We incorporated their corrections into what became the 2014 NEI version 2 and fully remodeled using these data in the summer of 2018. This produced the final NATA results. We then previewed these final results with S/L/T agencies and released the 2014 NATA in August 2018.

Figure 1-2 provides a more detailed flowchart showing the emissions sources used in the air quality models and how the hybrid approach fits into the overall approach.
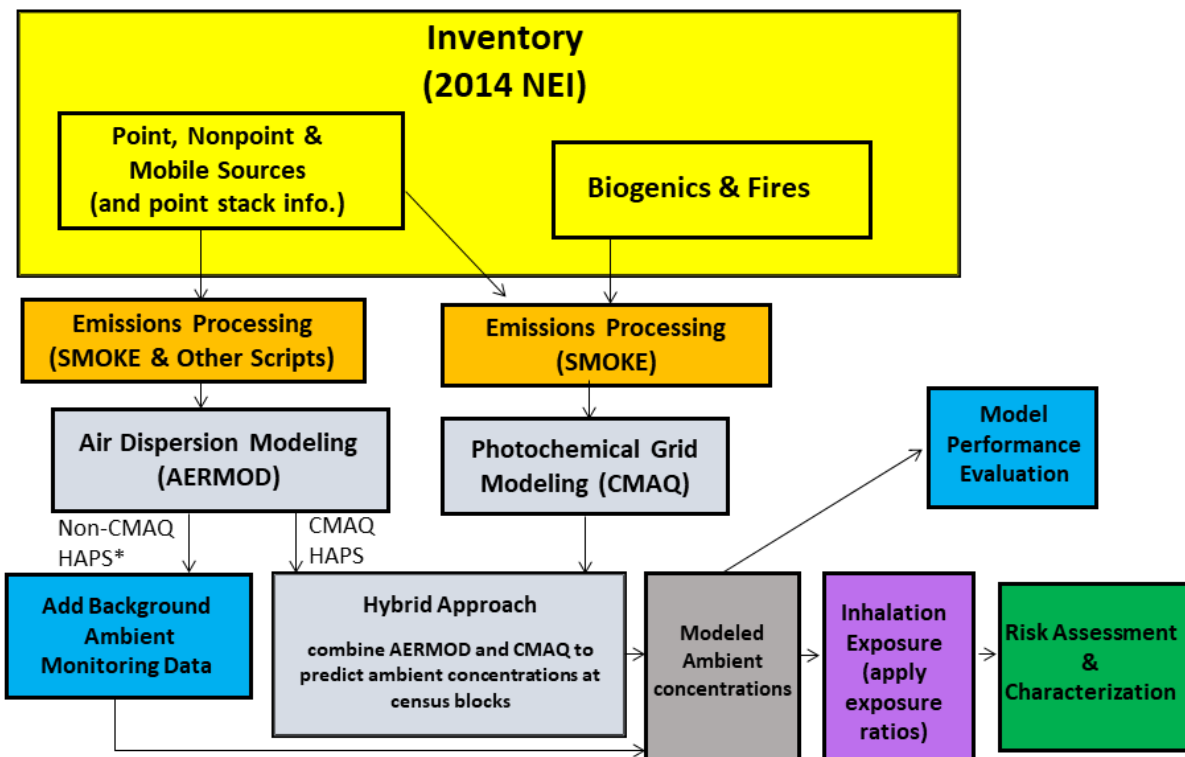
Figure 1-2. Detailed steps and approach used for 2014 NATA

1.3. What NATA Is

NATA is a first-pass, screening tool intended to evaluate the human-health risks posed by air toxics across the United States. NATA provides screening-level estimates of the risk of cancer and other potentially serious health effects from inhaling air toxics.

NATA uses emissions data compiled for a single year as inputs to air quality models. The models use these source data along with meteorological data for the same year to estimate ambient air concentrations of certain air toxics. EPA then combines these modeled concentrations with census data and other information to calculate exposure concentrations of the air toxics. We also estimate cancer risks and potential noncancer health effects associated with chronic inhalation exposure to the toxics.

EPA generates NATA results for each U.S. state at county and census tract levels. We also generate results for Puerto Rico, the Virgin Islands and the District of Columbia. These results help state, local and tribal agencies prioritize air toxics, emission sources and locations of interest for further study. They also help air agencies plan and implement national, regional and local efforts to reduce toxic air pollution.

NATA provides a "snapshot" of outdoor air quality as it relates to air toxics. It also suggests the risks to human health if air toxic emission levels were to remain the same as those estimated for the assessment year. The estimates only reflect risks associated with chronic (assumed to be lifetime) inhalation exposures to air toxics at the population level. Assumptions and methods we use to complete the assessment limit the types of questions that NATA can answer reliably. You should consider these limitations, described throughout this document and summarized in Section 7, when interpreting the NATA results or when using them to address questions posed outside of NATA.

NATA results can provide general answers to questions about emissions, ambient air concentrations and exposures and risks across broad geographic areas (such as counties, states and the nation) for the year modeled in the assessment.

NATA can answer questions such as the following:

- Which air toxics pose the greatest potential risk of cancer or adverse noncancer effects across the entire United States?

- Which air toxics pose the greatest potential risk of cancer or adverse noncancer effects in certain areas of the United States?

- Which air toxics pose less, but still significant, potential risk of cancer or adverse noncancer effects across the entire United States?

- When risks from long-term inhalation exposures to all outdoor air toxics are considered together, how many people could experience a lifetime cancer risk greater than levels of concern (e.g., 1-in-1 million or 100-in-1 million)?

- When considering potential adverse noncancer effects from long-term exposures to all outdoor air toxics together for a given target organ or system, how many people could experience exposures that exceed the reference levels intended to protect against those effects (i.e., a hazard quotient greater than 1)?

1.4. The History of NATA

As discussed on the NATA website (https://www.epa.gov/national-air-toxics-assessment), EPA's first national-scale air toxics study was the Cumulative Exposure Project (Caldwell et al. 1998). EPA developed this project based on estimates of air toxics emissions present before the Clean Air Act (CAA) was amended in 1990. The Cumulative Exposure Project estimated outdoor air toxics concentrations in each contiguous-U.S. census tract.

For the first NATA, EPA enhanced the Cumulative Exposure Project framework to include estimates of population exposure and health risk. The first NATA used a more refined inventory of air toxics emissions developed for 1996, known at that time as the National Toxics Inventory. EPA submitted this assessment for a technical peer review in January 2001 to a panel of EPA's Science Advisory Board (EPA 2001b). The panel provided detailed comments later that year on the validity of the overall approach, the elements of the assessment (including the data, models and methods used) and the manner in which these components were integrated into a national-scale assessment (EPA 2001a).

EPA incorporated many of the Science Advisory Board's suggestions into NATA and published the results of that assessment in 2002. Since then, four assessments have been completed – representative of air toxic emissions in 1999, 2002, 2005 and 2011, respectively – based on significant triennial updates of the national emission inventories. In general, the scope of NATA has progressively expanded with subsequent versions, and some methods have been refined and improved. Table 1-1 summarizes the six NATAs EPA has conducted to date.

Table 1-1. Summary of the six completed NATAs

Inventory Year	Year Completed/ Published	Air Toxics Modeled [a,b]	Key Attributes
1996	2002	33 – Includes 32 HAPs, focusing on those of concern in urban areas, plus diesel PM	ASPEN used to model ambient concentrations HAPEM4 used to model inhalation exposures
1999	2006	177 – Includes 176 HAPs, including all those with chronic-health toxicity values at the time, plus diesel PM	ASPEN used to model ambient concentrations HAPEM5 used to model inhalation exposures Doubled the number of emission sources covered compared to 1996 NATA
2002	2009	181 – Includes 180 HAPs, including four4 with additional health information, plus diesel PM	ASPEN and HEM (with ISCST3) used to model ambient concentrations HAPEM5 used to model inhalation exposures

Inventory Year	Year Completed/ Published	Air Toxics Modeled [a,b]	Key Attributes
2005	2010	179 – Includes178 HAPs for which emissions data and chronic-health toxicity values were available, plus diesel PM	Emissions inventory updated to include recent information on industrial sources, residual-risk assessments, lead emissions from airports and other sources ASPEN and HEM-3 (with AERMOD, a more refined dispersion model) used to model ambient concentrations; HEM used for more source types than in 2002 Exposure factors derived from 2002 NATA used to estimate inhalation exposures CMAQ model (EPA 2015f) used to estimate secondary formation of acetaldehyde, acrolein and formaldehyde and decay of 1,3-butadiene to acrolein
2011	2015	180 – Includes 179 HAPs for which emissions data and chronic-health toxicity values were available, plus diesel PM	CMAQ and HEM-3 more fully integrated as a hybrid modeling system for about 40 HAPS and diesel PM to improve mass conservation. HEM-3 used with background for remaining HAPs not covered by the hybrid approach (also for areas outside the contiguous U.S. CMAQ modeling domain) HAPEM7 modeled inhalation exposures for a subset of air toxics and used to provide exposure factors for the remaining air toxics
2014	2018	181 – Includes 180 HAPs for which emissions data and chronic-health toxicity values were available, plus diesel PM	AERMOD used as main dispersion model. CMAQ and AERMOD fully integrated as a hybrid modeling system for 51 HAPS and diesel PM. AERMOD used with background for remaining HAPs not covered by the hybrid approach (also for areas outside the contiguous U.S. CMAQ modeling domain) HAPEM7 modeled inhalation exposures for a subset of air toxics and used to provide exposure factors for the remaining air toxics

[a] Note that "air toxics" and "HAPs" are sometimes used interchangeably. In this document, however, "air toxics" refers to HAPs plus diesel PM. HAPs are those air toxics which we are required to control under Section 112 of the 1990 CAA Amendments (EPA 2015h). Diesel PM is not a HAP but is likely carcinogenic to humans, although we have not yet developed a unit risk estimate for it. Given these concerns, the adverse noncancer effects of diesel PM are estimated in NATA (using an Integrated Risk Information System reference concentration) but its cancer risks are not estimated.

[b] The number of air toxics included in a NATA emission inventory can be slightly larger than the number of air toxics actually modeled. Some air toxics are not modeled because of uncertainty in the emissions numbers or in the ability to model air concentrations or health risk accurately. For example, asbestos is included in the inventory but not modeled and not included in the counts presented in this table.

Notes: HAPs = hazardous air pollutants; diesel PM = diesel particulate matter; ASPEN = Assessment System for Population Exposure Nationwide; HAPEM4, HAPEM5, HAPEM7 = Hazardous Air Pollutant Exposure Model, version 4, 5 and 7; HEM = Human Exposure Model; CMAQ = Community Multiscale Air Quality model. ISC and AERMOD are Gaussian dispersion models.

1.5. How EPA and State, Local and Tribal Air Agencies Use NATA Results

We designed NATA to help guide efforts to reduce toxic air pollution and to provide information that can be used to further the already significant emissions reductions achieved in the United States since 1990. EPA and S/L/T air agencies use NATA to identify those air toxics and source sectors (e.g., point or mobile sources) having the highest exposures and health risks. The assessment results also help us

identify geographic patterns and ranges of risks across the country and across individual states and county areas within states.

Specifically, we use NATA results to:

- identify pollutants and industrial source categories of greatest concern;

- improve understanding of health risks posed by air toxics;

- help set priorities for the collection of additional information;

- set priorities for improving emission inventories;

- expand and prioritize EPA's network of air-toxics monitors;

- support communities in designing their own local assessments;

- enhance targeted risk-reduction activities; and

- provide a multiple-pollutant modeling framework linking air toxics to the Criteria Pollutant Program (EPA 2015c).

Similarly, S/L/T air agencies use NATA to:

- prioritize pollutants and emission source types;

- identify places of interest for further study;

- get a starting point for local assessments;

- focus community efforts; and

- inform monitoring programs.

1.6. How NATA Results Should *Not* Be Used

As described above, NATA is a screening-level assessment, designed to answer specific types of questions. The underlying assumptions of NATA and its methods limit the range of questions that can be answered reliably.

NATA results should *not* be used:

- as a definitive means to pinpoint specific risk values within a census tract;

- to characterize or compare risks at local levels (such as between neighborhoods);

- to characterize or compare risks between states;

- to examine trends between or otherwise compare NATAs;

- as the sole basis for developing risk reduction plans or regulations;

- as the sole basis for determining appropriate controls on specific sources or air toxics; or

- as the sole basis to quantify benefits of reduced air toxic emissions.

The limitations of the assessment methods prevent NATA from serving as a stand-alone tool. Furthermore, although EPA reports results at the census tract level in NATA, average risk estimates are far more uncertain at this level of spatial resolution than at the county or state level. To analyze air toxics in smaller areas, such as census blocks or in a suspected "hotspot," other tools such as site-specific monitoring and local-scale assessments coupled with refined and localized data should be used.

These caveats are integral to the proper interpretation of NATA results. You should use NATA results only to address those questions for which the assessment methods are suited.

Moreover, as noted above, NATA results from different assessment years generally should not be compared to each other. From one assessment to the next, EPA has improved its methodology and incorporated additional data. These improvements make comparing assessments inappropriate. Differences in emissions, ambient or exposure concentrations, or risks between two assessments may be due to improvement in the assessment methodology or to actual changes in emissions, populations or other "real-life" characteristics – or to some combination of all these.

EPA does not use NATA as the sole source of information leading to regulations or guiding the enforcement of existing rules. Some of the methods used to conduct NATA are like those used in air-related risk assessments conducted under the CAA mandate (such as residual risk assessments of HAP emissions from point sources, or assessments of exposures to criteria air pollutants (CAPs) for evaluations of National Ambient Air Quality Standards), NATA fundamentally differs from such assessments in that it is not a regulatory program.

1.7. The Risk Assessment Framework NATA Uses

In NATA we use methods consistent with the general risk assessment framework used throughout EPA. This section overviews EPA's risk assessment framework and summarizes the NATA process. Later sections detail the analytical components of this process.

EPA has published a series of guidelines (EPA 2015d) that establishes and explains the methods recommended for assessing human-health risks from environmental pollution. This series makes recommendations for carcinogen risk assessment, exposure assessment, chemical mixtures risk assessment and other major EPA-wide risk assessments. EPA has also developed the three-volume Air Toxics Risk Assessment (ATRA) Reference Library (EPA 2004a,b; EPA 2006a) as a reference for those conducting air toxics risk assessments. This library details the fundamental principles of risk-based assessment for air toxics, how to apply those principles in various settings, and strategies for reducing risk at the local level. EPA's guidelines and methods are consistent with the National Research Council's recommendations on conducting risk assessments (NRC 1983, 1994).

As described in more detail in these guidelines and documents, EPA's risk assessment process has three phases (Figure 1-3), the second of which has two parts.

- The first phase (problem formulation) comprises the initial planning and scoping activities and definition of the problem, which results in the development of a conceptual model.

- The second phase (analysis) includes two components:
  - Exposure assessment; and
  - Toxicity assessment.

- The third phase is risk characterization, a synthesis of the outputs of the exposure and toxicity assessments to characterize health risks for the scenario described in the initial phase.
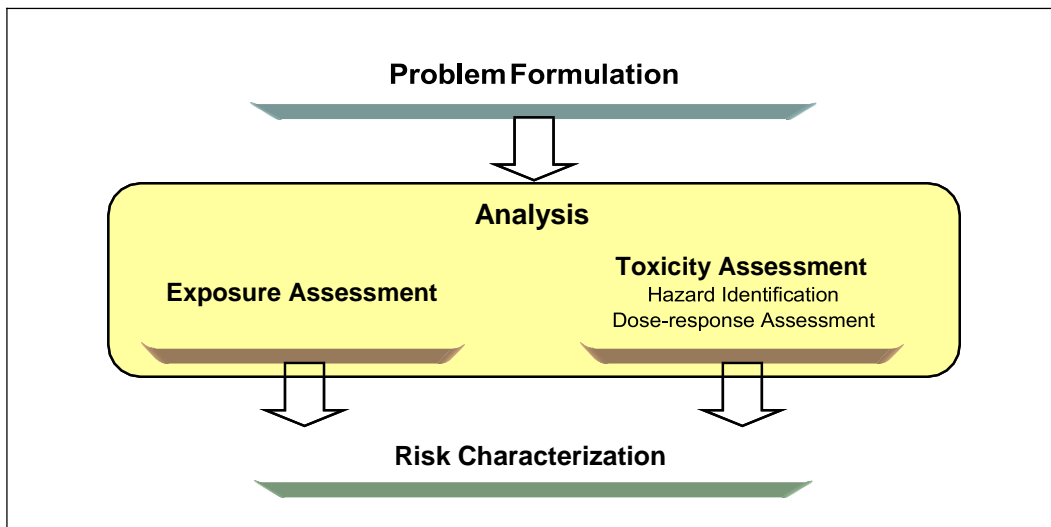
Figure 1-3. The general air toxics risk assessment process

Source: Adapted from EPA (2004a).

An air toxics risk assessment starts with problem formulation. This first step begins with the systematic planning and scoping needed before any analyses are begun. This planning process helps ensure that the objectives of the assessment are met, resources are used efficiently and the overall effort is successful.

One important product of the problem formulation step is a conceptual model that describes how releases of air toxics might pose risks to people. The conceptual model serves as a guide or "road map" to the assessment. It defines the physical boundaries, potential sources, emitted air toxics, potentially exposed populations, chemical fate and transport processes, expected routes of exposure and potential health effects. The planning and scoping activities and problem formulation we conduct before carrying out the analyses, are critical – they set the course for the assessment and inform EPA's decisions regarding specific methods, models and data sources to use. The following section (1.8) describes the conceptual model developed for NATA – the product of the first phase.

Meanwhile, the rest of this document is concerned primarily with describing the analysis phase of the general air toxics risk assessment process (and specifically with describing the analyses conducted for NATA). The analysis phase is the stage at which we use the risk assessment processes to evaluate the problem at hand. Section 1.9 outlines the analytical steps, with detailed descriptions of each step presented in later sections of this document.

1.8. The Scope of NATA

The national-scale assessment described in this document is consistent with EPA's definition of a cumulative risk assessment, as stated in EPA's Framework for Cumulative Risk Assessment (EPA 2003, p. 6), as "an analysis, characterization, and possible quantification of the combined risks to health or the environment from multiple agents or stressors." The Framework emphasizes that a conceptual model is an important output of the problem formulation phase of a cumulative risk assessment. The conceptual model defines the actual or predicted relationships among exposed individuals, populations or ecosystems and the chemicals or stressors to which they might be exposed. Specifically, the conceptual model lays out the sources, stressors, environmental media, routes of exposure, receptors and endpoints (i.e., measures of effects) relevant to the problem or situation that is being evaluated. This model takes the form of a written description and a visual representation of the relationships among these components.

The conceptual model can sometimes include components that are not addressed specifically or quantitatively by an assessment, but that are nevertheless important to consider.

Section 2.4 of the report for the 1996 NATA presented to EPA's Science Advisory Board for review (EPA 2001b) included a conceptual model. Some of the specifics included in that conceptual model have since evolved as newer assessments have been completed (for example, the number of air toxics evaluated has increased substantially since the 1996 NATA). The fundamental components included in NATA and the relationships among them, however, have been generally consistent for all six NATAs completed to date. Moreover, the conceptual model described in this document is very similar to the one presented in the documentation for the 1996 NATA.

NATA is national in scope, covering the United States, Puerto Rico and the U.S. Virgin Islands. It focuses on long-term inhalation exposures to air toxics. In general, NATA is intended to provide EPA with the best possible national-scale population-level estimates of exposure to and risks associated with air toxics, considering data availability, technical capabilities and other potentially limiting factors. The conceptual model for the 2014 NATA is presented in Figure 1-4. Each component included in the model is described briefly in the sections that follow.

1.8.1. Sources of Air Toxic Emissions That NATA Addresses

Sources of primary air toxic emissions included in NATA (i.e., the NATA categories) are point, nonpoint, mobile, biogenics and fires in the contiguous United States. Point, nonpoint and mobile sources are included from Alaska, Hawaii, Puerto Rico and the U.S. Virgin Island (with impacts from biogenics and fires estimated). Examples of point sources are large waste incinerators and factories. Nonpoint sources include residential wood combustion (RWC), commercial cooking, and consumer and commercial solvents. Mobile sources include vehicles found on roads and highways, such as cars and trucks, and nonroad equipment, including lawn mowers and construction equipment. Nonroad sources also include marine vessels, trains and aircraft.
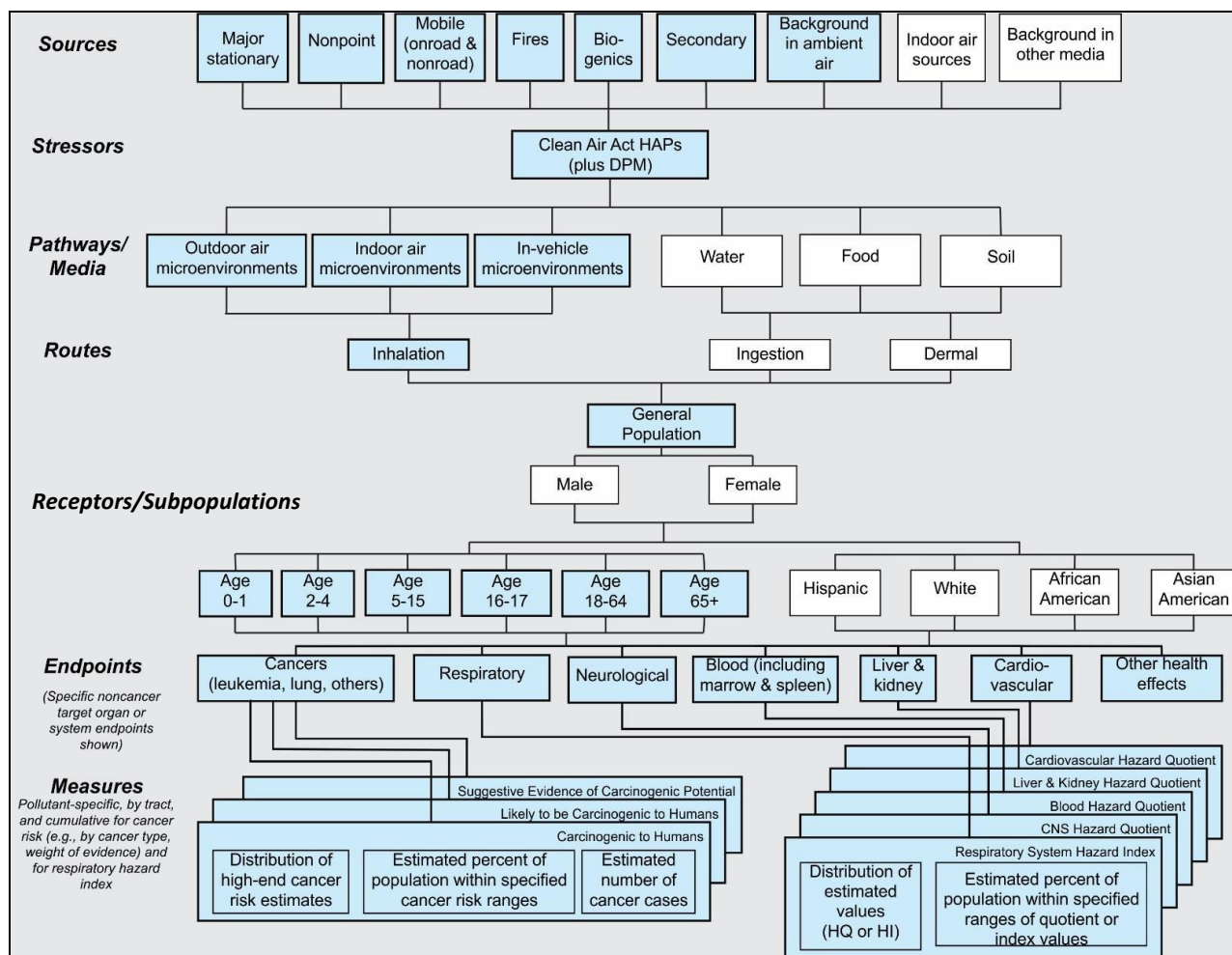
Figure 1-4. Conceptual model for NATA

Blue boxes indicate elements included in the 2014 NATA; clear boxes indicate elements that could be included in future assessments. In the "Sources" included here, "Major stationary" includes both major and area sources as defined for regulatory purposes in the CAA. "Nonpoint" refers to smaller (and sometimes less discrete) sources that are typically estimated on a top-down basis (e.g., by county). Additional explanation of source types included in NATA is presented in Section 2. DPM refers to diesel particulate matter. HQ and HI refer to hazard quotient and hazard index, respectively.

NATA only considers outdoor sources of air toxics. It does not address indoor sources of toxics, for example, those emitted from household chemicals. In addition, NATA background estimates do not consider background air toxics from other media, such as water.

NATA presents results by both these broad source categories and by more detailed NATA source groups. Details on this and other aspects of emission sources are presented in Section 2; details on air quality modeling and characterization are presented in Section 3.

1.8.2. Stressors that NATA Evaluates

The stressors evaluated through NATA can include any of the 187 HAPs defined in the 1990 CAA (190 HAPs were included originally but 3 have since been removed from the list). The set of air toxics included in NATA is determined by the emission and toxicity data available at the time of the

assessment. Diesel PM, an indicator of diesel exhaust, is included in the set of stressors for NATA. The spreadsheet file "NATA_Pollutants.xlsx" within the Supplemental Data folder accompanying this TSD lists the pollutants that the 2014 NATA assesses and provides more detailed information on the NEI and NATA pollutants. In Appendix B of this document, Table B-I lists the CAA pollutants that are not included in the 2014 NATA and the reason for their omission.

1.8.3.   Exposure Pathways, Routes and Time Frames for NATA

Exposure to air toxics from all sources is determined by multiple interactions among complex factors, including the locations and nature of the emissions, the emission-release conditions, local meteorology, locations of receptor populations, and the specific behaviors and physiology of individuals in those populations. The combination of air toxics that people inhale, and the chemical interactions among those air toxics, influence the risks associated with these exposures. This high level of complexity makes aggregating risk across both substances and sources useful for depicting the magnitude of risks associated with inhalation of air toxics.

The air quality modeling step of NATA includes evaluating the transport of emitted particles and gases through the air to receptors. NATA modeling accounts for transformation of substances in the atmosphere (also referred to as secondary formation) and losses of substances from the air by deposition, where data are available and the modeling approach supports it. For air toxics with sufficient ambient monitoring data, or with emissions data primarily due to point sources, we estimate background concentrations. With fate and transport of emissions considered, and the presence of some background concentrations, NATA estimates outdoor ambient concentrations across the nation.

NATA focuses on exposures due to inhalation of ambient air. Human receptors are modeled to account for an individual's movement among microenvironments, such as residences, offices, schools, exterior work sites and automobiles, where concentration levels can be quite different from general outdoor concentrations. The exposure assessment estimates air concentrations for each substance within each modeled microenvironment. The exposure assessment also accounts for human activities that can affect the magnitude of exposure (e.g., exercising, sleeping). This component of NATA accounts for the difference between ambient outdoor concentrations and the exposure concentrations (ECs; i.e., long-term-average concentrations to which people are exposed after accounting for human activities).

To date, NATA has not estimated air toxic concentrations in water, soil or food associated with deposition from air, or the bioaccumulation of air toxics in tissues. Similarly, NATA has not estimated human exposures to chemicals via ingestion or dermal contact. EPA considers these pathways important, but refined tools and data required to model multipathway concentrations and human exposures on the national scale are not yet readily available for use for many air toxics.

NATA estimates average annual outdoor concentrations, which are used to develop long-term inhalation exposures for each of the air toxics. For cancer and chronic (long-term) health effects, the exposure is assumed to be continuous over a lifetime (i.e., 70 years for the purposes of this analysis). Subchronic and acute (lasting less than 24 hours) exposures are not estimated in NATA because the emissions database contains only annual-total emissions. If the emission inventories are later expanded to cover short-term (e.g., hourly, daily) emission rates, we would consider incorporating shorter exposure times into NATA.

1.8.4. Receptors that NATA Characterizes

NATA characterizes average risks to people belonging to distinct human subpopulations. The overall population is divided into cohorts based on residential location, life stage (age) and daily activity pattern. A cohort is generally defined as a group of people within a population assumed to have identical exposures during a specified exposure period. Residential locations are specified according to U.S. Census tracts, which are geographic subdivisions of counties that vary in size but typically contain about 4,000 residents each. Life stages are stratified into six age groups: 0–1, 2–4, 5–15, 16–17, 18–64, and 65 and older. Daily-activity patterns specify time spent in various microenvironments (e.g., indoors at home, in vehicles, outdoors) at various times of day. For each combination of residential census tract and age, 30 sets of age-appropriate daily activity patterns are selected to represent the range of exposure conditions for residents of the tract. A population-weighted typical exposure estimate is calculated for each cohort, and this value is used to estimate representative risks, as well as the range of risks, for a "typical" individual residing in that tract. Risk results for individual cohorts are not included in NATA results.

To date, NATA evaluations have not included non-human receptors (e.g., wildlife and native plants). The complexity of the varied ecosystems across the vast area NATA covers precludes considering potential adverse ecological impacts at this time. Local- and urban-scale assessments can be developed to include non-human receptors, contingent on the availability of necessary resources, data and methodologies. We currently, however, have no plans to include non-human receptors in NATA.

1.8.5. Endpoints and Measures: Results of NATA

NATA reports estimated cancer risks and noncancer hazards attributed to modeled sources. Key measures of cancer risk developed for the 2014 NATA include:

- upper-bound estimated lifetime individual cancer risk; and

- estimated numbers of people within specified risk ranges (e.g., number of individuals with estimated long-term cancer risk of 1-in-1 million or greater or less than 100-in-1 million).

For noncancer effects, the key measures presented in the 2014 NATA are hazard indexes summed across all air toxics modeled for the respiratory system. Other target organs and systems are also shown.

NATA characterizes cancer risk and potential noncancer effects based on estimates of inhalation exposure concentrations determined at the census tract level. This approach is used only to determine geographic patterns of risks within counties, and not to pinpoint specific risk values for each census tract. We are reasonably confident that the patterns (i.e., relatively higher levels of risk within a county) represent actual differences in overall average population risks within the county. We are less confident that the assessment pinpoints the exact locations where higher risks exist, or that the assessment captures the highest risks in a county. EPA provides the risk information at the census tract level rather than just the county level, however, because the county results are less informative (in that they show a single risk number to represent each county). Information on variability of risk within each county would be lost if tract-level estimates were not provided. This approach is consistent with the purpose of NATA, which is to provide a means to inform both national and more localized efforts to collect air toxics information and to characterize emissions (e.g., to help prioritize air toxics and areas of interest for more refined data collection such as emissions testing or monitoring). Nevertheless, the assumptions

made in allocating mobile and nonpoint source emissions within counties can result in significant uncertainty in estimating risk levels, even though general spatial patterns are reasonably accurate.

1.9. Model Design

Consistent with the general approach for air toxics risk assessment described in Section 1.7 and illustrated in Figure 1-3, the analysis phase of NATA includes two main components: estimating exposure and estimating toxicity. The outputs of these analyses are used in the third phase, risk characterization, which produces health-risk estimates that can be used to inform research or risk management. These two phases (analysis and risk characterization) represent the "core" of EPA's assessment activities associated with NATA. This set of activities is referred to here as the "NATA risk assessment process."

The NATA risk assessment process can be characterized by four main steps:

1. compiling the nationwide inventory of emissions from outdoor sources;

2. estimating nationwide ambient outdoor concentrations of the emitted air toxics;

3. estimating population exposures to these air toxics via inhalation; and

4. characterizing potential health risks associated with these inhalation exposures.

The fourth step (risk characterization) also requires that quantitative dose-response or other toxicity values be identified for each air toxic included in the assessment. These values are taken from those developed by other EPA and non-EPA programs. Although this step does not require a "new" quantitative dose-response assessment to be conducted as part of NATA, it does require that we make important scientific and policy decisions regarding the appropriate values to use in NATA. Because these decisions are critical to the risk results, the identification of appropriate dose-response values is also described in this TSD in Section 5. The NATA risk assessment process is illustrated in Figure 1-5. The development of the emission inventory, air quality modeling, inhalation exposure modeling and risk characterization must be conducted sequentially – completing each step requires outputs from the previous step, and toxicity values are required to carry out the risk-characterization calculations. Cancer risks and the potential for noncancer health effects are estimated using available information on health effects of air toxics, risk-assessment and risk-characterization guidelines, and estimated population exposures.

Each of these five components is described briefly here and explained in detail in the remainder of this document:

■   Section 2 explains the source types and air toxics included in the NATA emissions inventory. It also describes the processes we carried out to prepare the emissions for the air quality models.

■   Section 3 discusses the models and procedures used to estimate ambient concentrations of air toxics, with links and references to technical manuals and other detailed documentation for the models used for NATA.
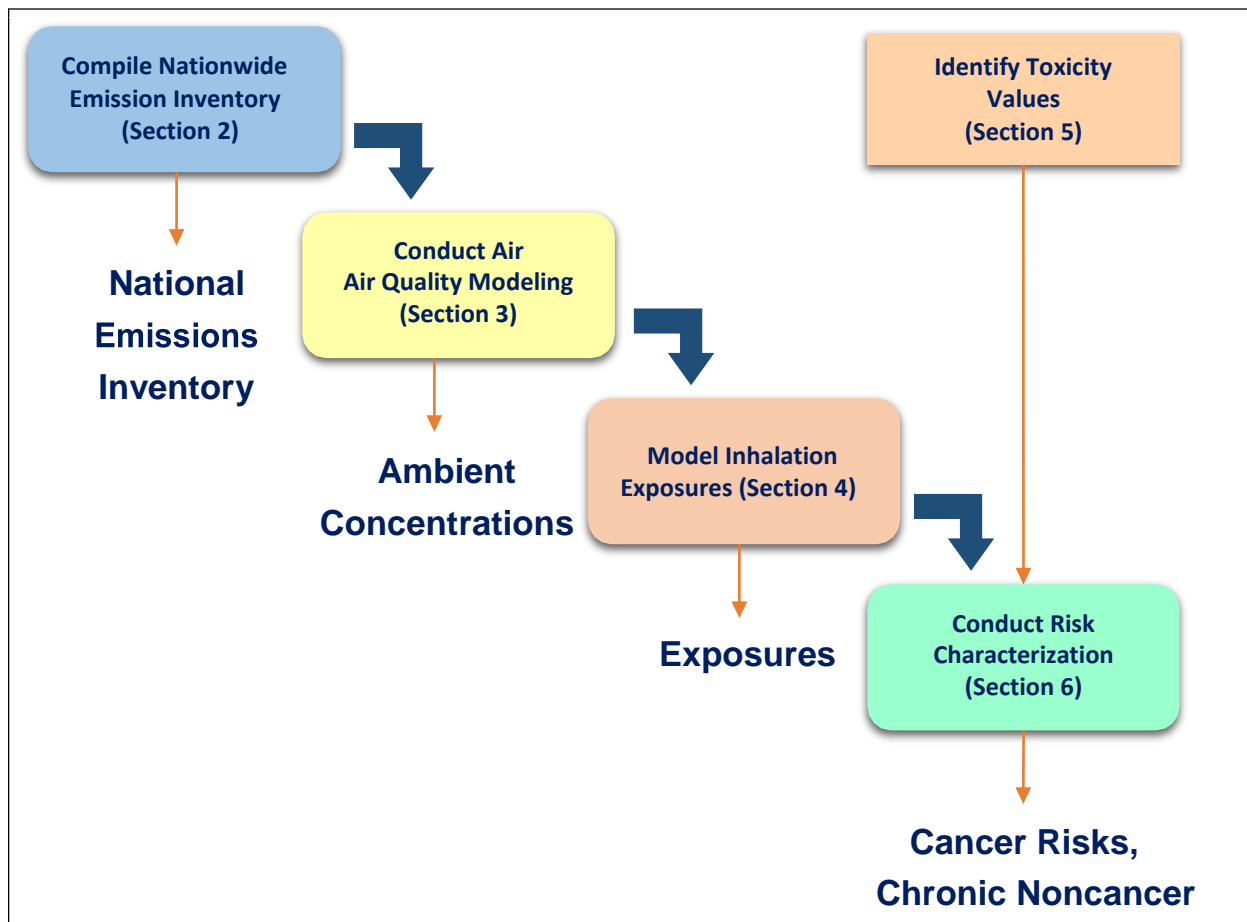
Figure 1-5. The NATA risk assessment process and corresponding sections of this TSD

- Section 4 explains the processes used to estimate population-level exposure to outdoor ambient levels of air toxics, accounting for information on activities and other characteristics that can affect inhalation exposures.

- Section 5 discusses the dose-response values used for NATA, the sources from which these values are obtained and assumptions made specific to NATA.

- Section 6 provides the calculations used to estimate cancer risk and potential noncancer hazard.

- Section 7 describes the uncertainties and limitations associated with the NATA process that must be considered when interpreting NATA results.

As noted at the beginning of this section, this document is intended to serve as a resource accompanying the most recent national-scale assessment – the 2014 NATA. Accordingly, although the following sections contain information on the NATA process that are generally applicable to all previous NATAs, references to specific technical processes and supporting details typically emphasize what we did for the 2014 NATA.

1.9.1. The Strengths and Limitations of the Model Design

EPA developed NATA to inform both national and localized efforts to characterize air toxics emissions and health risks (e.g., prioritize air toxics or areas of interest for monitoring and community assessments). Because of this targeted objective, tools other than NATA may be more appropriate for assessing health risks outside the specific purpose of NATA (e.g., for evaluating risks from either a broader or more specific perspective).

To further define and clarify what NATA should not be used for, this section contains descriptions of some of the important data and results that are not included in NATA:

- NATA does not include information that applies to specific locations. The assessment focuses on variations in air concentration, exposure and risk among geographic areas such as census tracts, counties and states. All questions asked, therefore, must focus on the variations among these geographic areas (census tracts, counties, etc.). Moreover, as previously mentioned, results are far more uncertain at the census tract level than for larger geographic areas such as states or regions. (Section 7 contains discussions on the higher uncertainty at small geographic scales such as census tracts.)

- NATA does not include data appropriate for addressing epidemiological questions such as the relationship between cancer risks or noncancer health effects and proximity of residences to point sources, roadways and other sources of air toxics emissions.

- The results do not include impacts from sources in Canada or Mexico other than as general background sources. Thus, the results for states bordering these countries do not comprehensively reflect sources of transported emissions that could be significant.

- NATA does not include results for individuals. Within a census tract, all individuals are assigned the same ambient air concentration, chosen to represent a typical ambient air concentration. Similarly, the exposure assessment uses activity patterns that do not fully reflect the actual variations among individuals.

- The results do not include exposures and risk from all compounds. For example, of the 181 air toxics included in the 2014 NATA, only 138 air toxics have been assigned dose-response values. In EPA's judgment, the remaining air toxics do not have adequate data to quantitatively assess their impacts on health. Therefore, they do not contribute to the aggregate cancer risk or target-organ-specific hazard indexes estimated in NATA. Of note, the assessment does not quantify cancer risk from diesel PM, although EPA has concluded that the general population is exposed to levels close to or overlapping with levels that have been linked to increased cancer risk in epidemiology studies. NATA, however, does quantify noncancer effects of diesel PM.

- Other than lead, which is both a CAP and a HAP, the results do not include the air pollutants, known as CAPs (particulate matter, ground-level ozone, carbon monoxide, sulfur oxides, nitrogen oxides), for which the CAA requires EPA to set National Ambient Air Quality Standards (other than CAP impacts on secondary formation of formaldehyde, acetaldehyde and acrolein).

- The results do not reflect all pathways of potential exposure. The assessment includes risks only from direct inhalation of the emitted air toxics compounds. It does not consider air toxics

compounds that may deposit onto soil, water and food and subsequently enter the body through ingestion or skin contact.

- The assessment results reflect exposure at outdoor, indoor and in-vehicle locations, but only to compounds released into the outdoor air. The assessment does not include exposure to air toxics emitted indoors, such as those from stoves, those that out-gas from building materials or those from evaporative benzene emissions from cars in attached garages. The assessment also does not consider toxics released directly to water and soil.

- The assessment does not fully reflect variation in background ambient air concentrations. Background ambient air concentrations are average values over broad geographic regions.

- The assessment may not capture all sources that have episodic emissions (e.g., facilities with short-term deviations in emissions resulting from startups, shutdowns, malfunctions and upsets). Where available, episodic emission information is used (e.g., for electricity generating units). In the absence of additional data, we assume emission rates are uniform throughout the year.

- Short-term (acute) exposures and risks are not included in NATA.

- Atmospheric transformation and losses from the air by deposition are not accounted for in NATA air toxics that are not modeled in CMAQ.

- The evaluations to date have not assessed ecological effects, given the complexity of the varied ecosystems across the vast area covered by NATA.

2. Emissions

The systematic compilation of a detailed, nationwide inventory of air toxics emissions is the first major step in the NATA risk assessment process. This section contains descriptions of the emissions used for the 2014 NATA. Section 2.1 describes the emissions data sources and preparation of the emissions used in NATA. Section 2.2 discusses the processing of emissions for input into the CMAQ model, and Section 2.3 discusses the processing for input into the AERMOD model.

2.1. Sources of Emissions

NATA is intended to address outdoor emissions of all hazardous air pollutants (HAPs) and diesel particulate matter (PM), together called "air toxics" in this document. To model air toxics, emissions of both air toxics and criteria air pollutants (CAPs, including CAP precursors such as ammonia and volatile organic compounds) are used to address the chemical interactions that occur across all pollutants.

> Sometimes "air toxics" and "HAPs" are used interchangeably. In this document, however, "air toxics" refers to the HAPs that EPA is required to control under Section 112 of the 1990 Clean Air Act (EPA 2015h) plus diesel PM. The 1990 Clean Air Act Amendments required EPA to control 189 HAPs (EPA 2018) and provided for revisions to be made to that list. Currently, the list includes 187 HAPs. Diesel PM is not a HAP; however, it has been included in this and previous NATAs. Some evidence indicates that localized high lifetime cancer risks are associated with exposure to diesel PM. However, EPA currently does not have sufficient evidence to develop a unit risk estimate for it. Therefore, the potential adverse noncancer effects associated with diesel PM are estimated in NATA (using an Integrated Risk Information System reference concentration), but its cancer risks are not.

The 2014 NATA combines modeling from CMAQ and AERMOD for the contiguous United States. CMAQ multipollutant modeling addresses all sources in the 2014 National Emissions Inventory (NEI) for CAPs and about 50 HAPs. Emissions from outside the United States are represented by CMAQ boundary conditions as discussed in Section 3.3.2. For the remaining "non-CMAQ" HAPs and non-CMAQ parts of the modeling domain (i.e., Alaska, Hawaii, Puerto Rico and the U.S. Virgin Islands), only AERMOD is used. For these pollutants and geographic regions, spatially uniform background concentrations based on remote concentrations are added to the AERMOD-modeled data to represent influences from transport and emissions outside the modeling domain (Section 3.6.1). Like previous NATAs, AERMOD modeling addresses all pollutants covered by NATA – and all anthropogenic sources except prescribed and agricultural burning.

The main source of the emissions data for the CAPs and HAPs modeled for NATA is the 2014 NEI version 2. The NEI is a comprehensive and detailed estimate of air emissions of CAPs and HAPs from all air emissions sources in the United States, including the territories of Puerto Rico and the U.S. Virgin Islands, and offshore sources and commercial marine vessels (CMVs) in federal waters. A complete NEI, consisting of point stationary sources, nonpoint sources, mobile sources and fires, is prepared every 3 years by EPA. It is based primarily upon emission estimates and emission model inputs provided by S/L/T air agencies for sources in their jurisdictions, supplemented by data developed by EPA. These data are submitted electronically to the Emissions Inventory System (EIS). CAPs are required under EPA's Air Emissions Reporting Requirements (AERR). HAPs are submitted voluntarily. Lead is both a HAP and a CAP, so it must be submitted under the AERR. For the 2014 NEI, states are required to report facilities with lead emissions greater than or equal to 0.5 tons per year (TPY). In addition to CAPs and HAPs, the 2014 NEI includes speciated particulate matter (PM) and diesel PM. In previous NATAs, diesel PM was

generated in an emissions modeling step after the NEI was produced. But in the 2014 NEI, diesel PM is available along with NEI CAPs and HAPs. As in previous NATAs, the NEI diesel PM is computed as the PM10 emissions for on-road and nonroad engines burning diesel or residual oil fuels. Although stationary engines also can burn diesel fuel, only mobile source sectors are used for estimating diesel PM emissions.

To build as complete an NEI as possible, EPA augments the S/L/T-submitted data using various sources of information, including the Toxics Release Inventory (TRI), and applies HAP-to-CAP emission-factor ratios to CAP emissions reported by S/L/T.

Table 2-1 contains a summary of the sources of emissions data in the NEI. More detailed information on all data sources can be found in the 2014 NEI documentation.

Table 2-1. Summary of emissions sources in the 2014 NEI v2

Source	Description
Stationary point	Most stationary point-source HAP data were submitted voluntarily by S/L/T. For some point sources, EPA gap-filled HAPs. Sources of data included: TRI data for 2014, augmentation of HAPs using emission-factor ratios (of HAP to CAP) applied to S/L/T-reported CAP emissions and rule-based emission factors (e.g., emissions factors using the 2010 test program conducted in support of Mercury and Air Toxics rule), and methane emissions reported by landfill operators in compliance with Subpart HH of the Greenhouse Gas Reporting Program (GHGRP) as a "surrogate" activity indicator. In some cases, 2011 data were pulled forward for some facilities.
Point airports	EPA estimates used the Federal Aviation Administration (FAA) Emission Dispersion Modeling System using landing and takeoff (LTO) information from FAA databases and updated where S/L/T-provided LTO. For smaller airports (general aviation) without detailed aircraft-specific activity data, straight emission factors were used. Lead emissions were estimated based on per-LTO emissions factors, assumptions about lead content in the fuel and lead-retention rates in the piston engines and oil. For some airports, estimates were provided by S/L/T. NEI has over 19,000 airports (including heliports and seaplanes); all are inventoried as point sources. In addition to LTOs, EPA's emissions estimates for airports included emissions of ground support equipment.
Point rail yards	The 2014v2 NEI includes non-zero emissions estimates for nearly 1000 rail yards. EPA emission estimates are associated with the operation of switcher engines at each Class I rail yard. EPA estimates were developed by the Eastern Regional Technical Advisory Committee's (ERTAC) rail group using a "top-down" approach that apportions 2014 national fuel use data to rail yards and used national fleet-wide information to create weighted average emission factors. For some rail yards, emission estimates were provided by S/L/T. HAP emissions were estimated by applying fractions to the VOC or PM estimates. There are also railyard emissions in the nonpoint inventory (see locomotives).
Stationary nonpoint	Includes many different source stationary source categories that are generally too ubiquitous to be inventoried as point sources and are therefore estimated at the county level. Examples of these sources include residential heating, consumer and commercial product usage, commercial cooking, oil and gas production, and industrial, commercial and institutional fuel combustion (where not in the point inventory). Emission estimates for these are developed by EPA and/or submitted by S/L/T. Where S/L/T submitted CAPs but not HAPs, missing HAP emissions were augmented.

Source	Description
Biogenics	Based on Biogenic Emission Inventory System (BEIS3.61) using 2014 meteorology from the Weather Research Forecasting Model (WRF). Gridded emissions were used in NATA and summed to annual county-level estimates for the NEI. Includes VOC, NO_X and three HAPs: formaldehyde, acetaldehyde and methanol.
Locomotives	Emissions at county-level resolution for Class I line haul, Class II/III line haul, passenger, commuter and rail yards. ERTAC's rail group developed EPA estimates for Class I line haul and Class II/III line haul. S/L/T also submitted data for locomotives. All passenger, commuter and county-level rail yards estimates were from S/L/T. EPA rail yard emissions were included as point-source rail yards as described above. HAPs were estimated by applying toxic fractions to the VOC or PM estimates.
CMVs	Emissions from category 1 and category 2 (C1/C2) and category 3 (C3) marine vessels at ports or underway. C1/C2 includes fishing boats, ferries and tugboats and is assumed to use diesel fuel; C3 includes oceangoing vessels and large ships and are assumed to use residual fuel. Emission estimates were developed using 1) activity data (kilowatt hours or kW), 2) engine operating load factors and 3) emission factors and HAP speciation profiles. This "bottom up" approach was used for the first time for the 2014 NEI. For some areas, estimates were provided by S/L/T. For 2014v2, the Lake Michigan Air Directors Consortium (LADCO) submitted emissions estimate for several states. HAPs were added based on the toxic fractions used in the EPA estimates.
On-road	Except for California, on-road emissions were generated using the SMOKE-MOVES emissions modeling framework, which leverages emission factors generated by the latest released version of the Motor Vehicle Emissions Simulator ([MOVES](#)) (MOVES2014a, code version: 20151201, database version: movesdb20151028); county and SCC-specific activity data and hourly meteorological data. These models used state- or EPA-provided input details specific to each county. California's emissions were developed via their EMFAC on-road model, but VOC HAPs were speciated from California-reported VOCs consistent with the MOVES2014a speciation. EPA added DIESEL-PM10 and DIESEL-PM25 for all diesel fuel SCCs, and they were set equal to the PM_{10} and $PM_{2.5}$ emissions from these diesel SCCs.
Nonroad, excluding airports, locomotives and CMVs	Except for California, EPA estimated these emissions using the MOVES2014a model, which incorporates the NONROAD2008 model for criteria pollutants. Several S/L provided inputs to the model. MOVES2014a also replaces toxic emission estimates for nonroad previously generated using the National Mobile Inventory Model (NMIM), which was used for 2011 and earlier NEIs. MOVES2014a incorporates updated nonroad emission factors for HAPs. MOVES2014a was used for all states other than California, which uses their own model. EPA added DIESEL-PM10 and DIESEL-PM25 for all diesel fuel SCCs, and they were set equal to the PM_{10} and $PM_{2.5}$ emissions from these diesel SCCs.
Fires	For purposes of NATA, fires include agricultural burning, prescribed burning and wildfires. EPA estimated agricultural burning (included in stationary nonpoint) using remote-sensing data, crop-usage maps and emission factors as daily point estimates. Many states submitted their own data as county estimates; these were used ahead of EPA estimates. EPA estimates were modeled as daily point estimates, and the state-submitted data were converted to point estimates for modeling. EPA developed day- and location-specific prescribed burning and wildfire emissions via the SMARTFIRE2 system (which includes the BlueSky modeling framework) with inputs from state agencies where available. Georgia and Washington state submitted emission estimates (day and location specific).

2.1.1. Pollutants and Pollutant Groups

NATA air quality modeling requires emissions of criteria air pollutants (CAPs) and their precursors in addition to HAP and diesel PM. We also need to aggregate the emissions of certain NEI pollutants to match the NATA pollutants. This section discusses the pollutants and pollutant groups in the NEI used in NATA and the aggregation needed to generate the NATA pollutants.

In CMAQ, we model CAPs and precursors and about 50 air toxics, including diesel particulate matter. Table 2-2 shows the specific air toxics used in CMAQ. For AERMOD, we model nearly all HAPs covered by the NEI; those that are not modeled are due to lack of emissions or risk considerations. Table B-I in Appendix B provides more detail about each HAP excluded. For NATA, we aggregated NEI pollutants that can be reported as either a group or as specific individual pollutants belonging to the group into pollutant groups for the NATA modeling and results. This aggregation was done in the emissions modeling process, prior to the air quality modeling. Table 2-2 lists the groups. For example, individual glycol ethers are grouped into the single NATA HAP "glycol ethers." The spreadsheet file "NATA_Pollutants.xlsx" within the Supplemental Data folder accompanying this TSD shows the individual HAPs and groups used in AERMOD and CMAQ. The following subsections give more details about some specific NATA pollutant groups.

Table 2-2. Pollutant groups

Group
Chromium Vi (Hexavalent)
Cresol cresylic acid (mixed isomers)
Cyanide compounds
Glycol ethers
Nickel compounds
PAHPOM
Polychlorinated biphenyls (aroclors)
Xylenes (mixed isomers)

2.1.1.1. Assignment of PAHs into PAH modeling groups

Polycylic aromatic hydrocarbons and polycylic organic matter (PAHPOM) are lumped together as a single group in NATA. However, the individual compounds in these groups have widely varying risks. As a result, we modeled the PAHPOM in separate groups based on their unit risk estimate (URE). For NATA, we summed the concentrations and risks across all PAHPOM risk groups. For the tabular emission summaries on the NATA website (https://www.epa.gov/national-air-toxics-assessment), the individual PAHPOM compounds (i.e., in Table 2-3) are provided.

The PAH groups are based on the groups established for the 2014 NATA and are listed below along with the individual PAHs assigned to each. Note that two pollutants representing polycylic organic matter in the NEI are unspeciated: pollutant code 250 (PAH/POM–Unspecified) and pollutant code 130498292 (PAH, total). These are assigned to PAH_880E5.

Table 2-3. PAH/POM pollutant groups

PAH Group	NEI Pollutant Code	NEI Pollutant Description	URE 1/(µg/m³)
PAH_000E0	120127	Anthracene	0
PAH_000E0	85018	Phenanthrene	0
PAH_000E0	129000	Pyrene	0
PAH_101E2	56495	3-Methylcholanthrene	0.01
PAH_114E1	57976	7,12-Dimethylbenz[a]Anthracene	0.114
PAH_176E2	189640	Dibenzo[a,h]Pyrene	9.6E-03
PAH_176E2	189559	Dibenzo[a,i]Pyrene	9.6E-03
PAH_176E2	191300	Dibenzo[a,l]Pyrene	9.6E-03
PAH_176E3	50328	Benzo[a]Pyrene	9.6E-04
PAH_176E3	192654	Dibenzo[a,e]Pyrene	9.6E-04
PAH_176E3	53703	Dibenzo[a,h]Anthracene	9.6E-04
PAH_176E3	194592	7H-Dibenzo[c,g]carbazole	9.6E-04
PAH_176E3	3697243	5-Methylchrysene	9.6E-04
PAH_176E3	41637905	Methylchrysene	9.6E-04
PAH_176E4	56553	Benz[a]Anthracene	9.6E-05
PAH_176E4	205992	Benzo[b]Fluoranthene	9.6E-05
PAH_176E4	205823	Benzo[j]fluoranthene	9.6E-05
PAH_176E4	226368	Dibenz[a,h]acridine	9.6E-05
PAH_176E4	224420	Dibenzo[a,j]Acridine	9.6E-05
PAH_176E4	193395	Indeno[1,2,3-c,d]Pyrene	9.6E-05
PAH_176E4	5522430	1-Nitropyrene	9.6E-05
PAH_176E5	207089	Benzo[k]Fluoranthene	9.6E-06
PAH_176E5	86748	Carbazole	9.6E-06
PAH_176E5	218019	Chrysene	9.6E-06
PAH_192E3	8007452	Coal Tar	9.9E-04
PAH_880E5	83329	Acenaphthene	4.8E-05
PAH_880E5	208968	Acenaphthylene	4.8E-05
PAH_880E5	203338	Benzo(a)Fluoranthene	4.8E-05
PAH_880E5	195197	Benzo(c)phenanthrene	4.8E-05
PAH_880E5	192972	Benzo[e]Pyrene	4.8E-05
PAH_880E5	203123	Benzo(g,h,i)Fluoranthene	4.8E-05
PAH_880E5	191242	Benzo[g,h,i,]Perylene	4.8E-05
PAH_880E5	56832736	Benzofluoranthenes	4.8E-05
PAH_880E5	91587	2-Chloronaphthalene	4.8E-05
PAH_880E5	284	Extractable Organic Matter (EOM)	4.8E-05
PAH_880E5	206440	Fluoranthene	4.8E-05
PAH_880E5	86737	Fluorene	4.8E-05
PAH_880E5	779022	9-Methyl Anthracene	4.8E-05
PAH_880E5	26914181	Methylanthracene	4.8E-05

PAH Group	NEI Pollutant Code	NEI Pollutant Description	URE 1/(µg/m³)
PAH_880E5	2422799	12-Methylbenz(a)Anthracene	4.8E-05
PAH_880E5	65357699	Methylbenzopyrene	4.8E-05
PAH_880E5	90120	1-Methylnaphthalene	4.8E-05
PAH_880E5	91576	2-Methylnaphthalene	4.8E-05
PAH_880E5	832699	1-Methylphenanthrene	4.8E-05
PAH_880E5	2531842	2-Methylphenanthrene	4.8E-05
PAH_880E5	2381217	1-Methylpyrene	4.8E-05
PAH_880E5	130498292	PAH, total	4.8E-05
PAH_880E5	198550	Perylene	4.8E-05
PAH_880E5	250	PAH/POM – Unspecified	4.8E-05

2.1.1.2. Metal groups

NATA includes metal compound groups consistent with metal emissions in the NEI. Metal emissions in the 2014 NEI represent only the mass of the metal with a few exceptions for specific compounds of **hexavalent chromium** (chromium VI) and **nickel** of known composition. Prior to modeling, we

> **Example: Adjusting Emissions for Chromium VI Compounds**
>
> Chromic acid (VI) (H_2CrO_4) has a molecular weight of about 118.01. Chromium, with an atomic mass of 52, is the toxic element of interest in this metal compound. Emissions reported in NEI are therefore multiplied by 0.4406 (i.e., 52 / 118.01), and the resulting emission rate is used in NATA modeling.

applied factors that convert the emissions of specific metal compounds to the portion of the compound that is metal. Table 2-4 shows the HAPs that have metal speciation factors other than 1.

The three nickel compounds and three chromium VI compounds in the 2014 NEI are shown in the table below with the corresponding adjustment factors to compute the emissions that account for just the metal portion of the compound. Note that after applying the adjustments, the chromium VI compounds are grouped into chromium VI and the nickel compounds are grouped into nickel. These are generally small in mass compared to the metal-only pollutants (nickel and chromium VI) and are only present for stationary sources.

Table 2-4. Metal speciation factors for NEI metal compounds

Description	NATA Pollutant Group	pollutant_cd (CAS)	Metal Speciation Factor
Nickel oxide	NICKEL COMPOUNDS	1313991	0.7858
Chromium trioxide	CHROMIUM VI (HEXAVALENT)	1333820	0.52
Chromic acid (VI)	CHROMIUM VI (HEXAVALENT)	7738945	0.4406

We made two changes to the 2014 metal CN factors from 2011:

■ We removed the 0.5 factor from calcium cyanamide because the unit risk is for that particular HAP and not cyanide compounds.

■ We corrected the nickel oxide factor from 0.7412 to 0.7858, which is the ratio of the molecular weight of nickel (58.6934) to the molecular weight of nickel oxide (74.6928).

2.1.1.3. Diesel PM

Diesel PM is neither a CAP nor HAP as defined by Section 112 of the CAA, however it was identified as a mobile source air toxic in EPA's 2007 rule, "Control of Hazardous Air Pollutants From Mobile Sources final rule" (EPA 2007a). Prior to the 2014 NEI, it was generated separately for NATA modeling from the NEI and was not included as a separate NEI pollutant. However, starting with the 2014 NEI, diesel PM emissions are included in the NEI, as discussed above. The NEI-generated diesel PM emissions from the mobile-source, engine-exhaust PM_{10} emissions were used for engines burning diesel or residual-oil fuels. These sources include on-road, nonroad, point-airport-ground support equipment, point-locomotives, nonpoint locomotives, and all PM from diesel or residual-oil-fueled nonpoint CMVs. Diesel PM emissions were set equal to PM_{10} emissions for these engines. Although stationary engines also can burn diesel fuel, only mobile-related diesel engine SCCs were used.

2.1.1.4. Pollutant information file

The spreadsheet file "NATA_Pollutants.xlsx" within the Supplemental Data folder accompanying this TSD includes a crosswalk that contains NEI pollutant codes/descriptions, NATA group information, CMAQ names, metal adjustment factors, URE, RfC and target organ information.

2.1.2. Emissions Categorization: NEI and NATA

As explained on the NEI website, the 2014 NEI includes five data categories: point, nonpoint (formerly called "stationary area"), nonroad mobile, on-road mobile, and events consisting of wild and prescribed fires. NEI summaries are generally provided by sectors and tiers, which describe the type of emission source (e.g., industrial processes – oil and gas production). Some sectors and tiers cut across data categories since stationary sources are inventoried as both point and nonpoint. For example, the NEI sector "Fuel Combustion – Commercial/Institutional – Oil" results from large institutions inventoried as point sources (e.g., large universities with onsite steam plants) as well as commercial/institutional entities that are small and ubiquitous in nature, so their emissions are inventoried as county sums.

NATA summaries are provided by NATA broad summary categories and by more detailed source groups. The broad NATA summary categories are point, nonpoint, on-road, nonroad, fires, biogenics and secondary. Some of these categories are named the same as the NEI data categories, but they are not identical. For example, the NEI nonpoint category includes CMVs and locomotives, while the NATA category does not. Similarly, the NATA nonroad category includes airports, CMVs and locomotives, while the NEI category does not. Table 2-5 contains comparisons between the NEI data categories and the NATA categories. "Secondary" is not included in Table 2-5 since it not a primary emissions category covered in the NEI, but rather a result of atmospheric chemistry from the modeled emissions of CAPS and HAPs.

Table 2-5. Map of NEI data categories to NATA categories

NEI Data Category	NATA Category (Reflecting NATA Summary Results)
Point	*Point[1]*
Emissions estimates for sources that are individually inventoried and usually located at a fixed, stationary location (although portable sources such as some asphalt- or rock-crushing operations are also included). Point sources include large industrial facilities and electric power plants but also increasingly include many smaller industrial and commercial facilities, such as dry cleaners and gas stations, that had traditionally been included as nonpoint sources. The choice of whether these smaller sources are estimated individually and included as point sources or inventoried as a nonpoint source aggregated to county or tribal areas is determined by the separate S/L/T air agency.	Same as NEI point except: Excludes portable sources, which are not modeled in either CMAQ or AERMOD because no geographic information other than the state code is included. Excludes airports and railyards, which are nonroad mobile.
Nonpoint	*Nonpoint*
Sources that individually are too small in magnitude or too numerous to inventory as individual point sources and that can often be estimated more accurately as a single aggregate source for a county or tribal area. Examples are residential heating and consumer solvent use. Agricultural, CMVs and locomotive emissions are included. Biogenic emissions that come from vegetation are also included.	Same as NEI nonpoint except excludes locomotive, CMV, biogenic emissions and agricultural fires.
On-road	*On-road*
Emissions estimates for mobile sources, such as cars, trucks and buses. EPA's MOVES2014a model currently generates these estimates (except in California, which uses different models).	Same as NEI on-road.
Nonroad	*Nonroad*
Emissions estimates for nonroad equipment such as lawn and garden equipment, agricultural, construction, industrial and commercial equipment and recreational equipment. EPA's MOVES2014a model generates these estimates (except in California, which uses different models).	Same as NEI nonroad, but the NATA nonroad also includes CMVs, locomotives, aircraft engine emissions occurring during LTOs, and the ground support equipment and auxiliary power units associated with the aircraft.

NEI Data Category	NATA Category (Reflecting NATA Summary Results)
Event	*Fires*
Prescribed and wildfire emissions computed as day- and location-specific events	Wildfires, prescribed burning and agricultural burning. These are modeled in CMAQ but not AERMOD.
	Wildfires and prescribed burning are generated via the SMARTFIRE2 model at specific geographic coordinates for each day, and are assigned to 12-km grid cells for input into CMAQ.
	Agricultural burning estimates are in the NEI as county-level emissions, but EPA-derived data are developed as day- and location-specific emissions, and S/L/T-submitted data are county level but prepared for CMAQ as point sources with day-specific emissions as explained in the 2014 modeling platform documentation.
	Biogenic Emissions
	Emissions of formaldehyde, acetaldehyde and methanol from vegetation (trees, plants and soils) computed from the Biogenic Emission Inventory System within CMAQ. These are gridded to 12-km cells and modeled in CMAQ, but are not modeled in AERMOD.

[1]In results presented online for assessments for the 2002 and earlier NATA inventories, point sources were divided into major sources and area sources; these were sometimes referred to as stationary sources. Major sources are defined in the CAA as stationary sources that have the potential to emit either at least 10 TPY of a HAP or at least 25 TPY of any combination of HAPs. Area sources are stationary sources for which the locations are known but that emit at levels below the major source emissions thresholds. This terminology is not used in the 2014 NATA, and stationary-source emissions are referred to only as point-source or nonpoint-source emissions. Point sources in the NATA results refer to those sources, including smaller sources, for which a specific location for their emissions is identified by latitude and longitude descriptions, and nonpoint sources are those stationary sources that are not point sources.

2.1.3. Differences Between the 2014 NEI and Emissions Used for NATA

Two modeling platforms were developed for the two air quality models run for NATA, CMAQ and AERMOD. The starting-point emission files for both were based primarily on "flat file" formats (FF10) of the 2014 NEI produced by the EIS for the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. On-road emissions were generated by SMOKE-MOVES (Motor Vehicle Emissions Simulator; EPA 2015g), which produced the FF10 used as the starting point for the AERMOD emissions processing.

Although 2014 NEI v2 is the main basis of the emissions fed into the air quality models for NATA, there were minor differences between the 2014 NEI v2 and emissions data used for the NATA modeling.

These differences are:

■ FIPS code corrections were made to the 2014 NEI v2 Alaska CMV emissions for counties 02201, 02232 and 02280 to put them into nonretired counties. See Section 2.3.8 for more details.

■ Small changes in the data occurred between the time the flat file was exported from EIS and the NEI was posted on the public website. For the modeling platforms, the data were exported on November 1, 2017. A few emissions changes were made to the 2014 NEI prior to its release on EPA's

website in February 2018. Most emissions changes were incorporated into the FF10; those that were not were insignificant or were for pollutants not used in the platform:

♦ Change for primary $PM_{2.5}$ and PM_{10} from 0.754 tons to 0.846 tons affecting facility 16691911 (SOL 95-18 Wastewater Management Facility) in Wyoming

♦ Changes to the following pollutants not used in the platform: condensable and filterable portion of particulate matter and the speciated components (EC, OC, SO4, NO3 and PMFINE) affecting facility 16691911 (SOL 95-18 Wastewater Management Facility) in Wyoming and facility 4195111 (Covanta Alexandria/Arlington Inc) in Virginia.

In addition, facility inventory changes (e.g., geographic coordinates, FIPS and release parameters) were made in the EIS after the modeling file was generated. These changes are reflected in the summaries of 2014 NEI v2 on EPA's website. Facility ID and FIPs changes will cause differences in facility or county summaries between the modeling file and 2014 NEI v2. Table 2-6 summarizes the differences in point inventory between the 2014 NEI v2 and NATA's modeling files.

Table 2-6. Point inventory differences between the modeling file and 2014 NEI v2

FACILITY_ID CHANGES	
Alabama	Facility 17057711 (Integrity Cabinets) has two HAPs split between this ID and 17133511 in the FF10, but all emissions are under 17057711 in the NEI. Same total emissions.
Minnesota	Geringhoff Manufacturing has ID 17066711 in the FF10, and 17903811 in the NEI.
Ohio	TRINITY HIGHWAY PRODUCTS LLC has two HAPs under ID 8102011 in the FF10; entire facility is ID 16535611 in the NEI; ID 8102011 is an alternate ID for this facility in EIS. WHEMCO - OHIO FOUNDRY INC has HAPs under alternate ID 8103111 in EMF; entire facility is ID 16806011 in the NEI; ID 8103111 is an alternate ID for this facility in the NEI. Facility 14756011 (Gerken Materials Inc) has FIPS 39069 in the FF10, 39043 in the NEI.
FIPS CHANGES	
Louisiana	Facility 8024611 (Noranda Alumina LLC) has FIPS 22095 in the FF10, 22093 in the NEI.
Minnesota	Facility 6384211 (Viking Gas Transmission - Cushing) has FIPS 27097 in the FF10, 27153 in the NEI. Facility 9496511 (University of MN - Landscape Arboretum) has FIPS 27053 in the FF10, 27019 in the NEI. Facility 6455811 (Hardrives Inc - Plant 701) has FIPS 27053 in the FF10, 27145 in the NEI. Facility 6951111 (Brown-Wilbert Vault Co - Lakeville) has FIPS 27039 in the FF10, 27139 in the NEI. Facility 7038911 (Northern Improvement Co - Nonmetallic) has FIPS 27027 in the FF10, 27021 in the NEI. Facility 13596111 (Cannon Falls Energy Center) has FIPS 27037 in the FF10, 27049 in the NEI. Facility 17095711 (Two Harbors Birch Bark Processing & Pellet Plant) has FIPS 27137 in the FF10, 27075 in the NEI.
Ohio	Facility 14756011 (Gerken Materials Inc) has FIPS 39069 in the FF10, 39043 in the NEI.
Utah	Facility 6432511 (Northwest Pipeline GP- Moab Compressor Station) has FIPS 49037 in the FF10, 49019 in the NEI. Facility 8178511 (Chevron Products Co - Salt Lake Refinery) has FIPS 49035 in EMF, 49011 in the NEI.
Virginia	Facility 6683711 (CHURCH & DWIGHT CO INC) has FIPS 51570 in EMF, 51041 in the NEI.
RELEASE POINT COORDINATE AND PARAMETER CHANGES	
	S/L/T made facility inventory changes after the SMOKE FF10 was produced due to the submittal of the year 2016 emissions inventory. These differences were not quantified.

Differences that result from differences in emissions processing (which reflect the specific role and function of the resulting inventory within the context of the NATA risk assessment process) are more accurately described as post-processing procedures rather than substantive changes.

Release parameter changes were made to the SMOKE flat file to fill in missing data or change out-of-range stack parameters. Data reporters of point sources must provide height, diameter, temperature and either velocity or flowrate, but they do not have to provide fugitive parameters.

Prior to emissions processing, we default missing or out-of-range stack or fugitive parameters and compute the velocity from the flowrate (if velocity is not provided). We chose to do the defaulting prior to the emissions processing and include the defaulted parameters directly into the FF10 prior to input into SMOKE. This is done for two reasons: 1) to provide better transparency in the FF10 files with respect to the data used in the model, and 2) to ensure that emission inputs are consistent across CMAQ and AERMOD models, since both use the FF10 as the starting point. The out-of-range parameters were chosen to be consistent with the range checks used in the Emissions Inventory System (EIS). The fugitive defaults were consistent with what has been used in previous NATAs.

Table 2-7 shows the changes made and why. The "Records changed" column indicates how many records were changed and provide the keywords used in the FF10 that indicate that a release parameter was changed and the situation. Even though SMOKE does not use the fugitive release point parameters, they are included in the table to make it complete.

Table 2-7. Release parameter defaults/changes to the FF10 inventory files for point sources

Field	Existing Value	New Value	Conditions/Notes	Number of Records Defaulted or Changed and Comment[1]
For point *sources* with stack releases (ERPtype NOT equal to "1")				
stkhgt	missing	use pstk[2] or global defaults[2]	None	None
stkdiam	missing	use pstk[2] or global defaults[3]	None	None
stkvel	missing	calculate from stkflow and stkdiam if not missing; otherwise reference by SCC from pstk[2] or global defaults[3]	vel = 4*stkflow/(pi*stkdiam^2) If the flow and diam are missing such that you cannot compute, use new value based pstk or global defaults	1,473,185 No pstk values used *ERPVelCompute*
stktemp	missing	use pstk[2] or global defaults[3]	None	None
stkhgt	Outside SMOKE range	use minimum value or maximum value in feet	Less than 1 ft (0.3048 m) or greater than 1300 ft (396 m)	None

Field	Existing	New Value	Conditions/Notes	Number of Records
stkdiam	Outside SMOKE range	use minimum value or maximum value in ft	Less than 0.001 ft (0.0003048 m) or greater than 300 ft (91.4 m)	None
stkvel	Outside SMOKE range	use minimum value or maximum value in ft/s	Less than 0.001 ft/s (0.0003048 m/s) or greater than 1000 ft/s (304.8 m/s)	Below min[4]: 18,817 Above max[4]: 11,742 *ERPVelRange*
stktemp	Outside SMOKE range	use minimum value or maximum value in F	Less than -30 F (-34.4 C or 248.15 K) or greater than 4000 F (2204.4 C or 2477.6 K)	None
For fugitive release points (not used in CMAQ)				
fug_width_ydim	missing	32.808 ft	None	3,856,867; *ERPFugMissing*
fug_length_xdim	missing	32.808 ft	None	3,888,847; *ERPFugMissing*
fug_angle	missing	0	None	3,932,478; *ERPFugMissing*
fug_height	missing	10 ft	fug_width_ydim and /or fug_length_xdim are missing	3,556,330; *ERPFugMissing*
fug_height	missing	0	WHEN fug_width_ydim and fug_length_xdim are not missing and > 0	12,742; *ERPFugHeight0*
For coke ovens: any release point that emits coke oven emissions (pollutant code 140)–all pollutants at that				
stkhgt	<126 ft	126 ft	erptype NOT = "1"	2159; *ERPCokeoven126*
fug_height	<126 ft	126 ft	erptype = "1"	2829; *ERPCokeoven126*
fug_length_xdim	<50 ft	50 ft	erptype = "1"	2767; *ERPCokeovenFug50*
fug_width_ydim	<50 ft	50 ft	erptype = "1"	2767; *ERPCokeovenFug50*

Comments were put into the modeling file to indicate why a record was changed:

ERPVelCompute – velocity computed from the flowrate provided in the inventory

ERPHtRange – height in the inventory was out of range

ERPDiamRange – diameter in the inventory was out of range

ERPVelRange – velocity in the inventory or velocity calculated from the flowrate in the inventory was out of range

ERPTempRange – temperature in the inventory was out of range

ERPFugHeight0 – fugitive height in the inventory was set to 0 because the width and length were not missing

ERPFugMissing – fugitive height, length and width were missing or fugitive length and/or width were missing.

ERPCokeoven126 – fugitive or stack height of release point emitting coke oven emissions was less than 126 ft

ERPCokeovenFug50 – fugitive length or width was less than 50 ft.

Pstk provides default stack parameters and is provided with other SMOKE ancillary files (ge_dat directory) on the website. The pstk file is formatted: region_cd, scc, stkhgt (m), stkdiam (m), stktemp (K) and stkvel (m/s)

Global defaults (converted to English): stkvel = 13.1234 ft; stktemp=72.05 F, stkdiam=0.6562 ft, stkhgt=9.8425 ft

Out-of-range values exist because the flow range checks in EIS allow some velocities to be above or below the range, and we run the velocity check after computing the missing flowrates.

Other emissions processing changes include:

- For on-road emissions, there were differences in the HAP VOCs in California. These were due to changes in the post-processing approach to adjust California-submitted pollutants consistent with the MOVES2014a speciation. The NATA inventory also includes a more refined set of SCCs that includes road type to support spatial allocation of county-level emissions to finer scales.

- Sources with FIPS state-county codes ending in 777 (in-flight lead and asphalt plants that have no geographic coordinates) were removed from the inventory.

- Nonpoint tribal data (FIPS beginning in 88) were not used in the modeling because spatial surrogates were not available and possible double counting would introduce uncertainty.

- To minimize over-concentration of fire emissions into individual grid cells, we spatially allocated fires (from the event category, modeled only in CMAQ) that were larger than 20,000 acres on a single day. Those fires were allocated into overlapping CMAQ grid cells based on fire polygons, where available, or circles otherwise.

- Other miscellaneous changes included air toxic name conversions, placing individual air toxics into groups, and similar transcription and phraseology conversions (e.g., to crosswalk the identity of an emitted air toxic to a substance with a quantitative dose-response value).

2.1.4. Overview of Differences in Emissions for CMAQ and AERMOD

By design, there were differences in the sources of emissions used by CMAQ and AERMOD. Differences in the emissions inputs to the two models were due to design differences in how the models were run. The emissions input into AERMOD excluded NATA categories more appropriately addressed by CMAQ, namely biogenics and three types of fires: wildfires, prescribed burning and agricultural-field burning. Biogenic emissions were generated within CMAQ using the Biogenic Emission Inventory System (BEIS) model with hourly meteorological inputs to generate hourly gridded (12 km x 12 km) emissions of several photochemical-model species, including three HAPs: formaldehyde, acetaldehyde and methanol. This category of emissions is routinely part of CMAQ runs and is more appropriately modeled in CMAQ due to its broad spatial and refined temporal resolution and meteorological dependence. Wildfires and prescribed and agricultural burning were also included in the CMAQ run but not in AERMOD because CMAQ provides for in-line plume rise of fires to higher vertical layers based on the acres burned. These algorithms are also used for agricultural burning, which is grouped with the other fires to allow us to retain source attribution from the fire and biogenic CMAQ zero-out runs (although not between the different fire types) discussed in Section 3.3.1.

In addition to differences in the sources of emissions used for the two models, there were also differences in the way we processed or modeled the emissions. Emissions modeling transforms the emissions inventory into the format needed by the air quality model and provides the source characterization. The emission inventory primarily contains annual emissions represented as either point sources with locations specified by latitude and longitude or county-level sources specified by FIPS code (county code). In the inventory, large facilities are inventoried as point sources; while more ubiquitous sources, such as wood stoves, solvent use, and cars and trucks, are inventoried at the county level.

Emissions modeling is performed in three main steps: spatial allocation, temporal allocation and speciation. Spatial allocation provides the models with the horizontal characterization of the emissions source. For example, emissions modeling provides AERMOD with locations for point sources and locations and spatial extent of sources emitted over large or small areas, and it provides CMAQ with gridded emissions. Vertical allocation of emissions is performed within the air quality models using stack parameter information and other model inputs.

Temporal allocation produces hourly variation in emissions based on the monthly, day-of-week and/or diurnal variation associated with the specific type of source.

Speciation takes the inventory pollutant and converts it to the pollutant used by the air quality model (or in subsequent processes). For example, the compound nickel oxide is converted to nickel because the risk information is only for the nickel portion of the compound. For AERMOD, sources also must be characterized using parameters that may not be in the inventory, such as release heights (which are not inventoried for mobile and nonpoint sources) and initial vertical dispersion (which is not in the inventory for any source).

Emissions processing for CMAQ and AERMOD is described in Sections 2.2 and 2.3, respectively. However, Section 2.2 contains only a summary since a separate technical support document is available that describes the CMAQ emissions modeling in detail.

Table 2-8 summarizes the spatial allocation differences between the models. As seen in Table 2-8, CMAQ uses 12-km horizontal resolution along with vertical resolution for point sources and fires, whereas AERMOD spatial resolution depends on the source category. Both models use hourly emissions, however CMAQ uses pollutant-specific hourly emissions whereas AERMOD uses the same hourly variation for all sources in a "run group," as described in Section 2.3.

Table 2-8. Differences in spatial characterization of sources between CMAQ and AERMOD

Category	NEI Resolution	Spatial Approach for AERMOD	Spatial Approach for CMAQ
Point (excluding airports)	Point	Point – vertical stack and fugitive based on NEI information on emission-release point	12-km grid cells
Airports	Point	Point – runways & 10 m^2 areas consistent with NEI geographic coordinates	12-km grid cells
Locomotives	Point (railyards) and County/Shape	Nonpoint – 12-km grid cells in the CONUS domain, tract (non-CONUS) Point – point fugitives	12-km grid cells
CMVs, ports and underway	County/Shape	Shapes from the NEI; separate shapes used for CMV at ports versus underway	12-km grid cells
On-road, nonroad equipment and other nonpoint	County	12-km or 4-km grid cells, depending on the category, in the CONUS domain; 9-km grid cells for AK; 3-km grid cells for HI, PR and VI	12-km grid cells
Agricultural burning and biogenic emissions	County	Not modeled	12-km grid cells
Fires (prescribed and wild)	Point	Not modeled	12-km grid cells

2.2. Preparation of Emissions Inputs for CMAQ

EPA routinely prepares emissions for photochemical grid models such as CMAQ by developing an emissions modeling platform, and the SMOKE modeling system is used as the primary emissions modeling tool. An emissions modeling platform includes the emission inventories, the ancillary data files (e.g., for speciation, temporal allocation and spatial allocation) and the approaches used to transform inventories for use in air quality modeling.

The platform used for this study has been developed based on the 2014 NEI version 2 and is the multipollutant 2014v7.1 platform (modeling case name: "2014fd_nata_cb6_14j"). We will denote this as the CMAQ NATA platform. This platform generated the emission inputs for the version of CMAQ used for NATA (version 5.2) using all CMAQ HAPs from the NEI emissions as opposed to generating them from speciation of VOC.

This version of CMAQ includes more air toxics than previous multipollutant versions – 51 HAPs and diesel PM. Specifically, chloroprene, ethyl benzene, acrylic acid, styrene, acetonitrile, methyl chloride, carbonyl sulfide and hexane were added in CMAQ for this 2014 NATA application.

A TSD describing this platform is found on the emissions modeling platforms website. See in particular: https://www.epa.gov/air-emissions-modeling/2014-version-71-platform. Emissions inputs and ancillary data for speciation and for temporal and spatial allocation are available at the emissions modeling platform ftp site.

A summary of the platform's key features is below.

2.2.1. Modeling Domain

For CMAQ, emissions are prepared for the CMAQ 12-km contiguous United States (or "CONUS") modeling domain (12US2) shown in Figure 2-1.

This grid uses a Lambert-Conformal projection, with Alpha = 33º, Beta = 45º and Gamma = -97º, with a center of X = -97º and Y = 40º. Table 2-9 describes the grid.

Figure 2-1. Air quality modeling domain

Table 2-9. Description of the platform grid

Common Name	Grid Cell Size	Description (see Figure 2-1)	Grid Name	Parameters Listed in SMOKE Grid Description (GRIDDESC) File: projection name, xorig, yorig, xcell, ycell, ncols, nrows, nthik
US 12 km or "smaller" CONUS-12	12 km	Smaller 12 km CONUS plus some of Mexico/Canada	12US2	'LAM_40N97W', -2412000, -1620000, 12.D3, 12.D3, 396, 246, 1

2.2.2. Sectors in the CMAQ NATA Platform

For the purposes of preparing the CMAQ model-ready emissions, the 2014 NEI v2 was split into finer-grained sectors used for emissions modeling. The significance of an emissions modeling or "platform sector" is that the data are run through all the SMOKE programs independently from the other sectors except the final merge program (Mrggrid). The sectors used for the NATA CMAQ platform are listed in Table 2-10.

Table 2-10. Platform sectors for the 2014v7.1 emissions modeling platform

Platform Sector: *abbreviation*	NEI Data Category	Description and Resolution of the Data Input to SMOKE
EGU units: *ptegu*	Point	2014 NEI v2 point source EGUs. The 2014 emissions are replaced with hourly 2014 continuous emissions monitoring system (CEMS) values for NO_X and SO_2 for any units that are matched to the NEI, and other pollutants for matched units are scaled from the 2014 NEI v2 using CEMS heat input. Emissions for all sources not matched to CEMS data come from the 2014 NEI v2. CEMS data come from EPA's Clean Air Markets data (http://ampd.epa.gov/ampd/). Annual resolution for sources not matched to CEMS data, hourly for CEMS sources.
Point source oil and gas: *pt_oilgas*	Point	2014 NEI v2 point sources that include oil and gas production and related processes based on facilities with the following NAICS: 2111, 21111, 211111, 211112 (Oil and Gas Extraction); 213111 (Drilling Oil and Gas Wells); 213112 (Support Activities for Oil and Gas Operations); 2212, 22121, 221210 (Natural Gas Distribution); 48611, 486110 (Pipeline Transportation of Crude Oil); 4862, 48621, 486210 (Pipeline Transportation of Natural Gas). Includes offshore oil and gas platforms in the Gulf of Mexico. Annual resolution.
Remaining non-EGU point: *ptnonipm*	Point	All 2014 NEI v2 point source records not matched to the ptegu or pt_oilgas sectors. Includes all aircraft and airport ground support emissions and some rail yard emissions. Annual resolution.
Agricultural: *ag*	Nonpoint	Nonpoint livestock and fertilizer application emissions. Livestock includes ammonia and other pollutants (except $PM_{2.5}$). Fertilizer includes only ammonia. County and daily resolution for livestock; county and annual resolution for fertilizer.
Agricultural fires with point resolution: *ptagfire*	Nonpoint	2014 NEI v2 agricultural fire sources that were developed by EPA as point sources with day-specific emissions or reported by S/L/T as county-level emissions and were apportioned to point sources. They are in the nonpoint NEI data category, but in the platform, they are treated as point sources.
Area fugitive dust: *afdust*	Nonpoint	PM_{10} and $PM_{2.5}$ fugitive dust sources from the 2014 NEI v2 nonpoint inventory; including building construction, road construction, agricultural dust and road dust. The NEI emissions are reduced during modeling according to a transport fraction and a meteorology-based (precipitation and snow/ice cover) zero-out. County and annual resolution.
Biogenic: *beis*	Nonpoint	Year 2014, hour-specific, grid cell-specific emissions generated from the BEIS3.61 model within SMOKE, including emissions in Canada and Mexico using BELD v4.1 land-use data (slightly updated from the BELDv4.1 used in 2014v7.0).
Category 1, 2 CMV: *cmv_c1c2*	Nonpoint	Category 1 (C1) and category 2 (C2) commercial marine vessel (CMV) emissions sources from the 2014 NEI v2 nonpoint inventory, except that it does not use C1/C2 from the 2014 NEI in federal waters. County and annual resolution; see othpt sector for all non-U.S. C3 emissions.
Category 3 CMV: *cmv_c3*	Nonpoint	Category 3 (C3) cmv emissions converted to point sources based on the center of the grid cells.
locomotives: *rail*	Nonpoint	Rail locomotives emissions from the 2014 NEI v2. County and annual resolution.

Platform Sector: *abbreviation*	NEI Data Category	Description and Resolution of the Data Input to SMOKE
Remaining nonpoint: *Nonpt*	Nonpoint	2014 NEI v2 nonpoint sources not included in other platform sectors, with adjustments to remove chromium from fugitive dust categories (paved and unpaved roads, construction and crops and livestock). County and annual resolution.
Nonpoint source oil and gas: *np_oilgas*	Nonpoint	2014 NEI v2 nonpoint sources from oil and gas-related processes. County and annual resolution.
Residential Wood Combustion: *Rwc*	Nonpoint	2014 NEI v2 nonpoint sources with residential wood combustion (RWC) processes. County and annual resolution.
Nonroad: *Nonroad*	Nonroad	2014 NEI v2 nonroad equipment emissions developed with the MOVES2014a using NONROAD2008 version NR08a with newer HAP emission factors than had been used in the 2011NEI. MOVES used for all states except California, which submitted their own emissions. County and monthly resolution.
On-road: *Onroad*	On-road	2014 on-road mobile source gasoline and diesel vehicles from moving and non-moving vehicles and vehicle refueling. Includes the following modes: exhaust, extended idle, auxiliary power units, evaporative, permeation, refueling and brake and tire wear. For all states except California, developed using winter and summer MOVES emission factor tables produced by MOVES2014a.
On-road California: *onroad_ca_adj*	On-road	2014 California-provided CAP and metal HAP on-road mobile source gasoline and diesel vehicles submitted to the NEI, gridded and temporalized using MOVES2014a. Volatile organic compound (VOC) HAP emissions derived from California-provided VOC emissions and MOVES-based speciation.
Point source fires: *ptfire*	Events	Point source day-specific wildfires and prescribed fires for 2014 computed using SMARTFIRE2 for both flaming and smoldering processes.
Non-US. fires: *ptfire_othna*	N/A	Point source day-specific wildfires and prescribed fires for 2014 for Canada and Mexico.
Other dust sources not from the 2014 NEI: *othafdust*	N/A	Fugitive dust sources from Canada's 2013 inventory. Transportable emissions are reduced via a transport fraction and a meteorology-based (precipitation and snow/ice cover) zero-out. County and annual resolution.
Other point sources not from the 2014 NEI: *othpt*	N/A	Point sources from Canada's 2013 inventory and Mexico's 2014 inventory, annual resolution. Also includes all non-U.S. C3 CMV.
Other non-NEI nonpoint and nonroad: *othar*	N/A	Year 2013 Canada emissions: monthly for agricultural ammonia and nonroad sources; annual for rail, CMV and other nonpoint Canada sectors. Year 2014 Mexico (municipio resolution): annual nonpoint and nonroad mobile inventories.
Other non-NEI on-road sources: *onroad_can*	N/A	Monthly year 2013 Canada on-road mobile inventory.
Other non-NEI on-road sources: *onroad_mex*	N/A	Monthly year 2014 Mexico (municipio resolution) on-road mobile inventory.

2.2.3. Fires and Biogenics

To approximate how much fires and biogenic primary emissions contribute to overall air toxics concentrations, we use a concept known as "zero-out" runs in CMAQ. This technique lets us split out the individual impacts from these source categories in the NATA results.

CMAQ can compute biogenic emissions during a run, and it has options to take multiple sets of point-source fire emission files as input. These features were used in this study to quantify the impacts of biogenic and fire emissions by running CMAQ three times:

1. The base-case run used all fire and anthropogenic emissions from the 2014fd_nata case with the option to generate biogenic emissions turned on.

2. The biogenic zero-out run used all fire and anthropogenic emissions from the 2014fd_nata case but with the option to generate biogenic emissions turned off.

3. The fire zero-out run used all anthropogenic emissions from the 2014fd_nata case and had the option to generate biogenic emissions turned on, but it excluded the input files for wild, prescribed and anthropogenic fires.

2.2.4. Speciation

The emissions modeling step for chemical speciation creates the "model species" needed by the air quality model for a specific chemical mechanism. These model species are either individual chemical compounds (i.e., "explicit species") or groups of species (i.e., "lumped species"). Model species are created in the emissions modeling process by directly mapping emissions from the emission inventory to the appropriate model species or by speciation of inventory species where a one-to-one match does not exist. For example, VOCs are speciated into numerous VOC-related model species defined by the chemical mechanism. The chemical mechanism used for the 2014 platform is the CB6 mechanism (Yarwood et al. 2010). We used a version of CB6 that we refer to as "CB6-CMAQ" that breaks out explicit naphthalene (NAPH) from the lumped CB6 model species XYL and PAR (resulting in model species NAPH, XYLMN and a revised PAR), and uses an additional VOC secondary organic aerosol tracer called SOAALK (Pye and Pouliot 2012). This platform generates the $PM_{2.5}$ model species associated with the CMAQ Aerosol Module version 6 (AE6).

In the NATA modeling platform, all CMAQ species that are explicit NATA HAPs were generated by directly mapping the NEI emissions for these HAPs; no HAPs in the United States were generated through speciation of VOC or $PM_{2.5}$ in the emissions modeling step.

Table 2-11 through Table 2-13 list the model species produced by SMOKE in the 2014 platform. Eight species (chloroprene, ethyl benzene, acrylic acid, styrene, acetonitrile, methyl chloride, carbonyl sulfide and hexane) were added to CMAQ for the 2014 NATA (these are shown with an asterisk in the below tables). The HAP metals, also from the inventory, are speciated within SMOKE into coarse and fine components used in CMAQ.

More information on speciation approach is available in the technical support document for the 2014 v2 modeling platform.

Table 2-11. Emission model species produced for CB6-CMAQ

Inventory Pollutant	Model Species[1]	Model Species Description
Cl_2	CL2	Atomic gas-phase chlorine
HCl	HCL	Hydrogen chloride (hydrochloric acid) gas
CO	CO	Carbon monoxide
NO_X	NO	Nitrogen oxide
	NO2	Nitrogen dioxide
	HONO	Nitrous acid
SO_2	SO2	Sulfur dioxide
	SULF	Sulfuric acid vapor
NH_3	NH3	Ammonia
VOC	ACET	Acetone
	ALD2	Acetaldehyde[2]
	ALDX	Propionaldehyde and higher aldehydes
	BENZ	Benzene[2]
	CH4	Methane[3]
	ETH	Ethene
	ETHA	Ethane
	ETHY	Ethyne
	ETOH	Ethanol
	FORM	Formaldehyde[2]
	KET	Ketone groups
	IOLE	Internal olefin carbon bond (R-C=C-R)
	ISOP	Isoprene
	MEOH	Methanol[2]
	NAPH	Naphthalene[2]
	OLE	Terminal olefin carbon bond (R-C=C)
	PAR	Paraffin carbon bond
	PRPA	Propane
	TOL	Toluene and other monoalkyl aromatics
	XYLMN	Xylene and other polyalkyl aromatics, minus naphthalene
	SOAALK	Lumped SOA tracer
Naphthalene	NAPH	Naphthalene from inventory
Benzene	BENZ	Benzene from inventory
Acetaldehyde	ALD2	Acetaldehyde from inventory
Formaldehyde	FORM	Formaldehyde from inventory
Methanol	MEOH	Methanol from inventory
VOC species from the biogenics model that do not map to model species above	SESQ	Sesquiterpenes
	TERP	Terpenes
PM_{10}	PMC	Coarse PM > 2.5 microns and ≤ 10 microns

Inventory Pollutant	Model Species[1]	Model Species Description
PM$_{2.5}$	PAL	Aluminum
	PCA	Calcium
	PCL	Chloride
	PEC	Particulate elemental carbon $\leq$ 2.5 microns
	PFE	Iron
	PK	Potassium
	PH2O	Water
	PMG	Magnesium
	PMN	Manganese
	PMOTHR	PM$_{2.5}$ not in other AE6 species
	PNA	Sodium
	PNCOM	Non-carbon organic matter
	PNO3	Particulate nitrate $\leq$ 2.5 microns
	PNH4	Ammonium
	POC	Particulate organic carbon (carbon only) $\leq$ 2.5 microns
	PSI	Silica
	PSO4	Particulate sulfate $\leq$ 2.5 microns
	PTI	Titanium
Sea-salt species (non-anthropogenic)[4]	PCL	Particulate chloride
	PNA	Particulate sodium

[1]Additional HAPs used outside of CB6-CMAQ for NATA are provided in Table 2-12.
[2]Naphthalene, benzene, acetaldehyde, formaldehyde and methanol (NBAFM) is produced via VOC speciation for Canada and Mexico, for other than on-road mobile sources in Mexico, or in very small quantities due to mixtures in the speciation profiles.
[3]Technically, CH$_4$ is not a VOC but part of TOG. Although emissions of CH$_4$ are derived, the AQ models do not use these emissions because the anthropogenic emissions are dwarfed by the CH$_4$ already in the atmosphere.
[4]These emissions are created outside of SMOKE.

Table 2-12. Additional HAP gaseous model species produced for CMAQ multipollutant specifically for NATA (not used within CB6)

Inventory Pollutant	Model Species
Acetaldehyde	ALD2_PRIMARY
Acetonitrile[1]	ACETONITRILE
Acrolein	ACROLEIN
Acrylic acid[1]	ACRYLICACID
Acrylonitrile	ACRYLONITRILE
1,3-Butadiene	BUTADIENE13
Carbon tetrachloride	CARBONTET
Carbonyl sulfide[1]	CARBSULFIDE
Chloroform	CHCL3
Chloroprene[1]	CHLOROPRENE
1,4-Dichlorobenzene(p)	DICHLOROBENZENE
1,3-Dichloropropene	DICHLOROPROPENE

Inventory Pollutant	Model Species
Ethylbenzene[1]	ETHYLBENZ
Ethylene dibromide (Dibromoethane)	BR2_C2_12
Ethylene dichloride (1,2‑Dichloroethane)	CL2_C2_12
Ethylene oxide	ETOX
Formaldehyde	FORM_PRIMARY
Hexamethylene-1,6‑diisocyanate	HEXAMETH_DIIS
Hexane[1]	HEXANE
Hydrazine	HYDRAZINE
Maleic anyhydride	MAL_ANYHYDRIDE
Methyl chloride[1]	METHCLORIDE
Methylene chloride (Dichloromethane)	CL2_ME
Specific PAHs assigned with URE =0	PAH_000E0
Specific PAHs assigned with URE =9.6E-06 (previously 1.76E-5)	PAH_176E5
Specific PAHs assigned with URE =4.8E-05 (previously 8.8E-5)	PAH_880E5
Specific PAHs assigned with URE =9.6E-05 (previously 1.76E-4)	PAH_176E4
Specific PAHs assigned with URE =9.6E-04 (previously 1.76E-3)	PAH_176E3
Specific PAHs assigned with URE =9.6E-03 (previously 1.76E-2)	PAH_176E2
Specific PAHs assigned with URE =0.01 (previously 1.01E-2)	PAH_101E2
Specific PAHs assigned with URE =1.14E-1	PAH_114E1
Specific PAHs assigned with URE =9.9E-04 (previously 1.92E-3)	PAH_192E3
Propylene dichloride (1,2‑Dichloropropane)	PROPDICHLORIDE
Quinoline	QUINOLINE
Styrene[1]	STYRENE
1,1,2,2-Tetrachloroethane	CL4_ETHANE1122
Tetrachloroethylene (Perchloroethylene)	CL4_ETHE
Toluene	TOLU
2,4-Toluene diisocyanate	TOL_DIIS
Trichloroethylene	CL3_ETHE
Triethylamine	TRIETHYLAMINE
m-xylene, o-xylene, p-xylene, xylenes (mixed isomers)[2]	XYLENES
Vinyl chloride	CL_ETHE

[1]New to CMAQ5.2 – version of CMAQ used for 2011 NATA did not include these HAPs.

[2]In 2011 NATA, these were separated into three model species: MXYL, OXYL and PXYL; in 2014 they are combined into XYLENES.

Table 2-13. Additional HAP particulate model species produced for CMAQ multipollutant specifically for NATA

Inventory Pollutant	Model Species
Arsenic	ARSENIC_C, ARSENIC_F
Beryllium	BERYLLIUM_C, BERYLLIUM_F
Cadmium	CADMIUM_C, CADMIUM_F
Chromium VI, chromic acid (VI), chromium trioxide	CHROMHEX_C, CHROMHEX_F

Inventory Pollutant	Model Species
Chromium III	CHROMTRI_C, CHROMTRI_F
Lead	LEAD_C, LEAD_F
Manganese	MANGANESE_C, MANGANESE_F
Mercury[1]	HGIIGAS, HGNRVA, PHGI
Nickel, nickel oxide, nickel refinery dust	NICKEL_C, NICKEL_F
Diesel-PM10, diesel-PM25	DIESEL_PMC , DIESEL_PMFINE, DIESEL_PMEC, DIESEL_PMOC, DIESEL_PMNO3, DIESEL_PMSO4

[1]Mercury is multiphase.

2.2.5. Temporalization

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. Temporal allocation (i.e., temporalization) is the process of distributing aggregated emissions to a finer temporal resolution, thereby converting annual emissions to hourly emissions as is required by CMAQ.

Temporalization takes these aggregated emissions and distributes them to the month, and then distributes the monthly emissions to the day and the daily emissions to the hours of each day. This process is typically done by applying temporal profiles to the inventories in this order: monthly, day of the week, and diurnal, with monthly and day-of-week profiles applied only if the inventory is not already at that level of detail.

The temporal factors applied to the inventory are selected using some combination of country, state, county, SCC and pollutant. Table 2-14 summarizes the temporal aspects of emissions modeling by comparing the key approaches used for temporal processing across the sectors. In the table, "Daily temporal approach" refers to the temporal approach for getting daily emissions from the inventory using the SMOKE Temporal program. The "Merge processing approach" refers to the days used to represent other days in the month for the merge step. If this is not "all," then the SMOKE merge step runs only for representative days, which could include holidays as indicated by the rightmost column.

Table 2-14. Temporal settings used for the platform sectors in SMOKE

Platform Sector Short Name	Inventory Resolutions	Monthly Profiles Used?	Daily Temporal Approach	Merge Processing Approach	Process Holidays as Separate Days
afdust_adj	Annual	Yes	week	All	Yes
ag	Monthly	No	all	All	Yes
beis	Hourly	No	n/a	all	Yes
cmv_c1c2	Annual	Yes	aveday	aveday	No
cmv_c3	Annual	Yes	aveday	aveday	No
nonpt	Annual	Yes	week	week	Yes
nonroad	Monthly	No	mwdss	mwdss	Yes
np_oilgas	Annual	Yes	week	week	Yes
onroad	Annual & monthly[1]	No	all	all	Yes

Platform Sector Short Name	Inventory Resolutions	Monthly Profiles Used?	Daily Temporal Approach	Merge Processing Approach	Process Holidays as Separate Days
onroad_ca_adj	Annual & monthly[1]	No	all	all	Yes
othafdust_adj	Annual	Yes	week	all	No
othar	Annual & monthly	Yes	week	week	No
onroad_can	Monthly	No	week	week	No
onroad_mex	Monthly	No	week	week	No
othpt	Annual & monthly	Yes	mwdss	mwdss	No
pt_oilgas	Annual	Yes	mwdss	mwdss	Yes
ptegu	Annual & hourly	Yes[3]	all	all	Yes
ptnonipm	Annual	Yes	mwdss	mwdss	Yes
ptagfire	Daily	No	all	all	Yes
ptfire	Daily	No	all	all	Yes
ptfire_othna	Daily	No	all	all	Yes
rail	Annual	Yes	aveday	aveday	No
rwc	Annual	No[2]	met-based	all	Yes

[1]Note the annual and monthly "inventory" refers to the activity data (vehicle miles travelled and vehicle population) for onroad. The actual emissions are computed on an hourly basis.
[2]Except for 2 SCCs that do not use met-based speciation.
[3]Only units that do not have matching hourly CEMS data use monthly temporal profiles.

In the table, the value "all" means that hourly emissions are computed for every day of the year and that emissions potentially have day-of-year variation. The value "week" means that hourly emissions are computed for all days in one "representative" week, representing all weeks for each month. This means emissions have day-of-week variation, but not week-to-week variation within the month. The value "mwdss" means hourly emissions are computed for a representative Monday, representative weekday (Tuesday through Friday), representative Saturday, and representative Sunday for each month. This means emissions have variation between Mondays, other weekdays, Saturdays and Sundays within the month, but not week-to-week variation within the month. The value "aveday" means hourly emissions are computed for one representative day of each month, meaning emissions for all days within a month are the same. Sector-specific details, including meteorological-based temporal allocation, are provided in the 2014v7.1 platform TSD. Ancillary data are available at the emissions modeling platform ftp site.

2.2.6. Spatial Allocation

For CMAQ, all emissions were allocated to 12-km grid cells. Sources with geographic coordinates such as point sources and fires are mapped to the appropriate grid cell based on those coordinates. Sectors with county-level resolution were allocated to 12-km grid cells using spatial surrogates, which are developed based on shapefiles of data with spatial patterns expected for the emissions category.

Table 2-15 through Table 2-17 list the spatial surrogates available for the CONUS domain used by CMAQ. In Table 2-15, surrogate names and codes listed in *italics* are not directly assigned to any sources for the platform, but they are sometimes used to gap-fill other surrogates or as an input for merging two surrogates to create a new surrogate that is used.

Table 2-15. U.S. surrogates available for the 2014v7.1 modeling platform

Code	Surrogate Description	Code	Surrogate Description
N/A	Area-to-point approach (for jet refueling)	505	Industrial Land
100	Population	506	Education
110	*Housing*	*507*	*Heavy Light Construction Industrial Land*
131	*Urban Housing*	*510*	*Commercial plus Industrial*
132	*Suburban Housing*	*515*	*Commercial plus Institutional Land*
134	*Rural Housing*	*520*	*Commercial plus Industrial plus Institutional*
137	*Housing Change*	*525*	*Golf Courses plus Institutional plus Industrial plus Commercial*
140	*Housing Change and Population*	*526*	*Residential - Non-Institutional*
150	Residential Heating – Natural Gas	*527*	*Single Family Residential*
160	*Residential Heating – Wood*	535	Residential + Commercial + Industrial + Institutional + Government
170	Residential Heating – Distillate Oil	*540*	*Retail Trade (COM1)*
180	Residential Heating – Coal	*545*	*Personal Repair (COM3)*
190	Residential Heating – LP Gas	*555*	*Professional/Technical (COM4) plus General Government (GOV1)*
201	*Urban Restricted Road Miles*	560	Hospital (COM6)
202	Urban Restricted AADT	*575*	*Light and High Tech Industrial (IND2 + IND5)*
205	Extended Idle Locations	*580*	*Food Drug Chemical Industrial (IND3)*
211	*Rural Restricted Road Miles*	*585*	*Metals and Minerals Industrial (IND4)*
212	*Rural Restricted AADT*	*590*	*Heavy Industrial (IND1)*
221	*Urban Unrestricted Road Miles*	*595*	*Light Industrial (IND2)*
222	*Urban Unrestricted AADT*	*596*	*Industrial plus Institutional plus Hospitals*
231	*Rural Unrestricted Road Miles*	650	Refineries and Tank Farms
232	*Rural Unrestricted AADT*	670	Spud Count - CBM Wells
239	Total Road AADT	671	Spud Count - Gas Wells
240	Total Road Miles	672	Gas Production at Oil Wells
241	*Total Restricted Road Miles*	673	Oil Production at CBM Wells
242	All Restricted AADT	674	Unconventional Well Completion Counts
243	*Total Unrestricted Road Miles*	*676*	*Well Count - All Producing*
244	All Unrestricted AADT	*677*	*Well Count - All Exploratory*
258	Intercity Bus Terminals	678	Completions at Gas Wells
259	Transit Bus Terminals	679	Completions at CBM Wells
260	*Total Railroad Miles*	681	Spud Count - Oil Wells
261	NTAD Total Railroad Density	683	Produced Water at All Wells
271	NTAD Class 1 2 3 Railroad Density	685	Completions at Oil Wells
272	*NTAD Amtrak Railroad Density*	*686*	*Completions at All Wells*

Code	Surrogate Description	Code	Surrogate Description
273	NTAD Commuter Railroad Density	687	Feet Drilled at All Wells
275	ERTAC Rail Yards	691	Well Counts - CBM Wells
280	Class 2 and 3 Railroad Miles	692	Spud Count - All Wells
300	NLCD Low Intensity Development	693	Well Count - All Wells
301	NLCD Med Intensity Development	694	Oil Production at Oil Wells
302	NLCD High Intensity Development	695	Well Count - Oil Wells
303	NLCD Open Space	696	Gas Production at Gas Wells
304	NLCD Open + Low	697	Oil Production at Gas Wells
305	NLCD Low + Med	698	Well Count - Gas Wells
306	NLCD Med + High	699	Gas Production at CBM Wells
307	NLCD All Development	710	Airport Points
308	NLCD Low + Med + High	711	Airport Areas
309	NLCD Open + Low + Med	801	Port Areas
310	NLCD Total Agriculture	805	Offshore Shipping Area
318	NLCD Pasture Land	806	Offshore Shipping NEI2014 Activity
319	NLCD Crop Land	807	Navigable Waterway Miles
320	NLCD Forest Land	808	2013 Shipping Density
321	NLCD Recreational Land	820	Ports NEI2014 Activity
340	NLCD Land	850	Golf Courses
350	NLCD Water	860	Mines
500	Commercial Land	890	Commercial Timber

Land cover data from the 2011 National Land Cover Database (NLCD) were used to create several revised and new surrogates for the 2014 NATA modeling. We used the attributes in the NLCD that classify land as "Developed, Low Intensity," "Developed, Medium Intensity," "Developed, High Intensity," and "Developed, Open Space," along with combinations of these classifications.

In addition, we processed annual average daily traffic (AADT) data from the highway performance monitoring system (HPMS) to develop an activity-based on-network surrogate, and we updated Urban/Rural Restricted/Unrestricted road miles to use urban and rural distinctions consistent with the roadway classifications in MOVES.

Spatial surrogates were assigned to emissions sources based on SCC. Ancillary files providing the surrogate assignments and that describe the data used for the surrogates are available in the Supplemental Data provided with this TSD. Summaries of emissions by surrogate code are provided in the 2014v7.1 TSD.

For the onroad sector, the on-network (RPD) emissions were allocated differently from the off-network (RPP and RPV). On-network emissions were allocated based on AADT data and off network used land use surrogates as shown in Table 2-16. Emissions from the extended (i.e., overnight) idling of trucks were assigned to surrogate 205 that is based on locations of overnight truck parking spaces. This surrogate's underlying data were updated for use in the 2014 NATA modeling to include additional data sources and corrections based on comments received on the 2011 NATA.

Table 2-16. Off-network mobile source surrogates

Source type	Source Type Name	Surrogate ID	Description
11	Motorcycle	307	NLCD All Development
21	Passenger Car	307	NLCD All Development
31	Passenger Truck	307	NLCD All Development
32	Light Commercial Truck	308	NLCD Low + Med + High
41	Intercity Bus	258	Intercity Bus Terminals
42	Transit Bus	259	Transit Bus Terminals
43	School Bus	506	Education
51	Refuse Truck	306	NLCD Med + High
52	Single Unit Short-haul Truck	306	NLCD Med + High
53	Single Unit Long-haul Truck	306	NLCD Med + High
54	Motor Home	304	NLCD Open + Low
61	Combination Short-haul Truck	306	NLCD Med + High
62	Combination Long-haul Truck	306	NLCD Med + High

For the oil and gas sources in the np_oilgas sector, the spatial surrogates were updated to those shown in Table 2-17 using 2014 data consistent with that used to develop the 2014 NEI nonpoint oil and gas emissions.

Table 2-17. Spatial surrogates for oil and gas sources

Surrogate Code	Surrogate Description
670	Spud Count - CBM Wells
671	Spud Count - Gas Wells
672	Gas Production at Oil Wells
673	Oil Production at CBM Wells
674	Unconventional Well Completion Counts
676	Well Count - All Producing
677	Well Count - All Exploratory
678	Completions at Gas Wells
679	Completions at CBM Wells
681	Spud Count - Oil Wells
683	Produced Water at All Wells
685	Completions at Oil Wells
686	Completions at All Wells
687	Feet Drilled at All Wells
691	Well Counts - CBM Wells
692	Spud Count - All Wells
693	Well Count - All Wells
694	Oil Production at Oil Wells
695	Well Count - Oil Wells
696	Gas Production at Gas Wells

Surrogate Code	Surrogate Description
697	Oil Production at Gas Wells
698	Well Count - Gas Wells
699	Gas Production at CBM Wells

2.3. Preparation of Emissions Inputs for AERMOD

Unlike emissions modeling for CMAQ, there were no tools[1] to perform emissions modeling for AERMOD until the 2014 NATA. For the 2014 NATA, we developed the capability in SMOKE to produce helper files that could be further post-processed to generate the source (SO) Pathway for AERMOD. The SO Pathway contains the source location and parameter information used by AERMOD. Helper files provide information about source location, release characteristics and temporal variability. They also provide source emissions. For sources in the CONUS, helper files also provide the column and row of the 12-km meteorological grid cell associated with the appropriate meteorological data (see also Section 3) to use for the emission source. Helper file formats and information vary by the type of source. For example, some emissions sources are point sources, some are gridded and some (like ports) are defined by polygon shapes. Gridded sources use different grid resolutions depending on their location (e.g., all Alaska gridded sources are 9 km[2]) and/or the type of source.

The SMOKE interface for AERMOD models each source using a unit emissions rate rather than actual pollutant-specific emissions. The location, source parameter and temporal helper files join to produce the SO Pathway for each source's AERMOD run. AERMOD provides source-specific concentrations per unit emissions rate (also known as χ/Q, abbreviated here as X/Q) that are not pollutant specific (but are source specific). These concentrations are then combined with pollutant-specific emissions for each source, using the emissions helper file from the emissions modeling process, to create the pollutant- and source-specific concentrations as shown below.

$$C_{pollutant\ i} = \frac{X}{Q} \times Emissions_{pollutant\ i}$$

Because we are using a unit emission rate in AERMOD, we do not account for chemistry or deposition that would be specific to any individual pollutant.

The format and content of the helper files vary by type of source, and we have chosen to organize the creation of separate sets of helper files by "run group." A run group is a set of sources that use similar source characterization methods or parameters. Run groups allow us to combine sources with similar characteristics and run them in AERMOD together, even though they may have very different emissions. The emissions from the specific sources in the run group can then be applied to the source group X/Qs to create source-specific concentrations. We also combine the sources within the run group into source groups, which is the source-level resolution of concentration and risk results for NATA.

A difference from previous NATAs is the allocation of county-level emissions to grid cells, as opposed to census tract sources. This was done for both CONUS and non-CONUS (i.e., Alaska, Hawaii, Puerto Rico and the Virgin Islands) sources; however, different grid cell resolutions were used, as discussed below.

---

[1]EMS-HAP (EPA 454/B-03-006, 8/2004) is written in SAS and requires a SAS license. In addition, it was designed based on a previous inventory schema and would not be able to be used with the current NEI.

2.3.1. Source and Run Groups – Overview

Source and run groups are similar, but serve different functions in NATA. Source groups provide source attribution for the NATA census tract risk results.

Run groups organize sources for modeling in AERMOD to minimize the number of separate model runs. It would be resource prohibitive to run the thousands of different source classification codes separately in AERMOD, so we grouped them into a manageable number of run groups that all have the same spatial resolution, approach for temporal variation and release characteristics (height and sigma-z). Run groups typically have multiple source groups, though two of them (OILGAS and RWC) have the same source group as run group.

The inventory run groups are shown in Table 2-18. Those modeled in AERMOD as AREA sources using grid cells (i.e., "AREA (gridcell)") were gridded from county level to grid cells using the same underlying surrogate data (shape files) used in CMAQ.

Table 2-18. Run groups for AERMOD

Run Group	NEI Data Category	NATA Category	Resolution of Inventory Prior to Emissions Modeling	Modeled in AERMOD as:
Point	Point	Point, Nonroad[1]	Point	POINT, AREA (fugitive)
Airport	Point	Nonroad	Point	POINT, AREA (runway)
Onroad On-network LD (LDON)	Onroad	Onroad	County	AREA (gridcell)
Onroad Off-network LD (LDOFF)	Onroad	Onroad	County	AREA (gridcell)
Onroad On-network HD (HDON)	Onroad	Onroad	County	AREA (gridcell)
Onroad Off-network HD (HDOFF)	Onroad	Onroad	County	AREA (gridcell)
Onroad Off-network-hoteling (extended idling and auxiliary power units) (HOTEL)	Onroad	Onroad	County	AREA (gridcell)
Onroad On-network LD (LDON)	Onroad	Onroad	County	AREA (gridcell)
Nonroad (NONRD)	Nonroad	Onroad	County	AREA (gridcell)
Nonpoint 10-meter release height (NPHI)	Nonpoint	Nonpoint	County	AREA (gridcell)
Nonpoint low-level release height (NPLO)	Nonpoint	Nonpoint, Nonroad[2]	County	AREA (gridcell)
Nonpoint Oil and Gas (OILGAS)	Nonpoint	Nonpoint	County	AREA (gridcell)
Nonpoint Residential Wood Combustion (RWC)	Nonpoint	Nonpoint	County	AREA (gridcell)
Nonpoint Agricultural Livestock (AG)	Nonpoint	Nonpoint	County	AREA (gridcell)
Nonpoint Oil and Gas (OILGAS)	Nonpoint	Nonpoint	County	AREA (gridcell)
Commercial Marine Vessels (CMV)	Nonpoint	Nonroad	Shapes	POLYGON

[1]Rail yards.
[2]Locomotives.

The run groups modeled as grid cells have different resolutions: 3 kilometers for all grid cells in Hawaii, Puerto Rico and the Virgin Islands; 9 kilometers for all grid cells in Alaska, and either 12 or 4 kilometers for grid cells in the CONUS, depending on the run group. These are summarized in Table 2-19.

Table 2-19. Resolution of the run groups modeled as gridded sources

Run Group (resolution is the value of the number in the run group abbreviation)	CONUS GRID: 12US1	Alaska GRID: 9AK1	Hawaii GRID = 3HI1	Puerto Rico & Virgin Islands GRID = 3PR1
Onroad On-network LD (LDON)	LDON4	LDON9AK	LDON3HI	LDON3PR
Onroad Off-network LD (LDOFF)	LDOFF12	LDOFF9AK	LDOFF3HI	LDOFF3PR
Onroad On-network HD (HDON)	HDON4	HDON9AK	HDON3HI	HDON3PR
Onroad Off-network HD (HDOFF)	HDOFF12	HDOFF9AK	HDOFF3HI	HDOFF3PR
Onroad Off-network-hoteling (extended idling and auxiliary power units) (HOTEL)	HOTEL4	HOTEL9AK	N/A	N/A
Nonroad (NONRD)	NONRD12	NONRD9AK	NONRD3HI	NONRD3PR
Nonpoint 10-meter release height (NPHI)	NPHI12	NPHI9AK	NPHI3HI	NHI3PR
Nonpoint low-level release height (NPLO)	NPLO12	NPLO9AK	NPLO3HI	NPLO3PR
Nonpoint Oil and Gas (OILGAS)	OILGAS4	OILGAS9AK	OILGAS3HI	N/A
Nonpoint Residential Wood Combustion (RWC)	RWC12	RWC9AK	RWC3HI	N/A
Nonpoint Agricultural Livestock (AG)	AG12	N/A	N/A	N/A

N/A means there are no emissions for this run group in this grid.

Table 2-20 and Table 2-21 describe the source characteristics of these run groups and how they relate to the platform sectors used for CMAQ emissions modeling. Note that the run groups developed for the county-level emissions were treated differently for CONUS versus non-CONUS domains. For the CONUS, we allocated to grid cells using the spatial resolution in the table. For the non-CONUS areas (AK, HI, PR, VI), we used these same run groups but allocated to tracts rather than grid cells. Design documents for the generation of the helper files for on-road, nonroad and nonpoint sources in CONUS and non-CONUS areas are provided in the Supplemental Data files.

Table 2-20. Non-gridded AERMOD run groups

Run Group	NEI Category and NATA CMAQ Sector	AERMOD Modeling Features: Release Height (RH; meters), Initial Vertical Dispersion (σ_z; meters) and Temporal Approach	Description of Sources
POINT	**NEI:** point **Platform:** ptegu, pt_oilgas, ptnonipm	Release parameters for individual release points at each facility are taken from the 2014 NEI, with defaulting done for missing or out-of-range parameters. **Spatial**: Use specific geographic coordinates for stacks ("POINT") and geographic coordinates along with fugitive length and width for fugitives ("AREA"). **Temporal**: Temporal profiles based on SCC codes for ptnonipm; based on continuous emissions monitoring data for ptegu.	All NEI point sources containing NATA HAP emissions except for facility source type 100 (airports).

Run Group	NEI Category and NATA CMAQ Sector	AERMOD Modeling Features: Release Height (RH; meters), Initial Vertical Dispersion (σ_z; meters) and Temporal Approach	Description of Sources
AIRPORTS	**NEI:** point Platform: ptnonipm	RH=3 σ_z =0 **Spatial:** For runway-area (line) sources: length based on a database containing runway endpoint coordinates; 50-m width for the major airports, 25-m width for the OTAQ-provided (smaller) airports. All facility emissions (NEI) spread equally over the runway(s). For nonrunway sources: 10-m-square area centered on NEI coordinates. **Temporal:** monthly/day-of-week/hourly profiles. Different profiles for Alaska seaplanes.	Airport Facilities in the point inventory – all emissions where facility source type code = 100 (airports). Includes seaplane ports and heliports, Emissions used at the facility unit's process release point. For these facilities all sub-facility emissions (airplane landing and takeoffs, ground support equipment) are summed to the facility level, and are treated the same way with regards to spatial and temporal allocation.
CMV	**NEI:** Nonpoint Platform: cmv	C1/C2 uses **RH** = 8.4 and σ_z = RH/2.15. C3 uses RH = 20 and σ_z = 40.7 based on CMAQ vertical emissions. **Spatial**– polygon shapes updated from the 2014 NEI v1-inventory shapes (there could be multiple polygons per port). Underway shapes use the same simplified shapes developed for the 2011 NATA. **Temporal:** monthly profiles in the SMOKE ancillary file. C1/C2 is flat except for Great Lakes state in which the monthly profile varies by county. C3 uses a monthly profile that puts slightly more emissions in the summer months.	C1/C2 and C3 commercial marine vessels at ports or underway.

Table 2-21. AERMOD gridded run groups

Run Group	NEI Category and NATA CMAQ Sector	AERMOD Modeling Features: Release Height (RH; meters), Initial Vertical Dispersion (σ_z; meters) and Temporal Approach	Description of Sources
LDON4 LDON9AK LDON3HI LDON3PR	**NEI:** on-road **Platform:** on-road	**RH** = 1.3, σ_z = 1.2, Resolution is 4 km for CONUS, 9 km for AK and 3 km for HI, PR, VI. **Temporal:** monthly temporal variation is pollutant-specific and county-specific. County-specific hourly profiles are the same for all pollutants based on benzene hourly emissions from SMOKE-MOVES (aggregate only SCCs in this run group).	On-network light duty mobile emissions such as passenger car exhaust and light duty passenger truck brake and tire wear. Emissions derived from SMOKE-MOVES. Includes refueling since temporal profile and spatial surrogate for on-network is a better match for refueling than off-network.

Run Group	NEI Category and NATA CMAQ Sector	AERMOD Modeling Features: Release Height (RH; meters), Initial Vertical Dispersion (σ_z; meters) and Temporal Approach	Description of Sources
LDOFF12 LDOFF9AK LDOFF3HI LDOFF3PR	**NEI:** on-road **Platform:** on-road	**RH** = 0.5, σ_z = 0.5, Resolution is 12 km for CONUS, 9 km for AK and 3 km for HI, PR, VI. **Temporal:** same approach as LDON4, but aggregate only SCCs in the LDOFF run group to compute hourly profiles.	Off-network light duty mobile emissions such as passenger car and passenger light truck start emissions. Derived from SMOKE-MOVES. Tailpipe height (no turbulence).
HDON12 HDON9AK HDON3HI HDON3PR	**NEI:** on-road **Platform:** on-road	**RH** = 3.4, σ_z = 3.2, Resolution is 4 km for CONUS, 9 km for AK and 3 km for HI, PR, VI. **Temporal:** monthly temporal variation is pollutant-specific and county-specific. County-specific hourly profiles based on **PM$_{2.5}$** hourly emissions from SMOKE-MOVES (aggregate only SCCs in this run group).	On-network heavy duty mobile emissions such as running exhaust from combination long haul and short haul trucks and buses. Includes brake and tire wear. Derived from SMOKE-MOVES.
HDOFF12 HDOFF9AK HDOFF3HI HDOFF3PR	**NEI:** on-road **Platform:** on-road	**RH** = 3.4, σ_z = 0.5, Resolution is 12 km for CONUS, 9 km for AK and 3 km for HI, PR, VI. **Temporal:** same approach as HDON run groups, but use only SCCs in the HDOFF run groups to compute hourly profiles.	Off-network heavy duty mobile emissions such as start emissions from combination long haul and short haul trucks and buses: Derived from SMOKE-MOVES. Minimal dispersion for start emissions. Tailpipe height (no turbulence.
HOTEL4 HOTEL9AK	**NEI:** on-road **Platform:** on-road	**RH** = 3.4, σ_z = 0.5, Resolution is 4 km for CONUS, 9 km for AK and 3 km for HI, PR, VI. **Temporal:** same approach as HDON run groups, but use only SCCs in the HOTEL run groups to compute hourly profiles.	Extended idling and auxiliary power units (APU) that occur at truck stops. Minimal dispersion for hoteling (e.g., extended idling) emissions. Tailpipe height (no turbulence).
NONRD12 NONRD9AK NONRD3HI NONRD3PR	**NEI:** nonroad **Platform:** nonroad	**RH** = 2, σ_z = 1, Resolution is 12 km for CONUS, 9 km for AK and 3 km for HI, PR, VI. **Temporal:** monthly temporal variation is pollutant-specific and county-specific. County-specific diurnal scalars computed based on **benzene** hourly emissions for the aggregate of all SCCs in this run group summed across all days of the year.	Nonroad equipment such as lawn mowers, turf equipment, agriculture and construction equipment, commercial generators, power-washing equipment, pleasure craft, recreational off-road.
NPHI12 NPHI9AK NPHI3HI NPHI3PR	**NEI:** nonpoint **Platform:** some of nonpt	**RH**=10, σ_z = 4.7, Resolution is 12 km for CONUS, 9 km for AK and 3 km for HI, PR, VI **Temporal:** uniform monthly/day-of-week. diurnal: Use SMOKE hourly profile 26 – mostly daytime emissions-(use qflag HROFDY).	Industrial processes (e.g., chemical plants, refineries, mines, metals); fuel combustion – industrial, commercial, institutional and residential (except wood); waste disposal.

Run Group	NEI Category and NATA CMAQ Sector	AERMOD Modeling Features: Release Height (RH; meters), Initial Vertical Dispersion (σ_z; meters) and Temporal Approach	Description of Sources
NPLO12 NPLO9AK NPLO3LO NPLO3PR	**NEI:** nonpoint **Platform:** some of nonpt	RH = 3.9, σ_z = 3.6, Resolution is 12 km for CONUS, 9 km for AK and 3 km for HI, PR, VI. **Temporal:** same as NPHI run groups.	Solvents (consumer, commercial); surface coating; commercial cooking; locomotives, bulk terminals, gas stations (stage 1); miscellaneous non-industrial (portable gas cans, auto repair shops, structure fires, and nonpoint mercury categories such as human cremation, dental amalgam).
OILGAS4 OILGAS9AK OILGAS3HI	**NEI:** nonpoint **Platform:** np_oilgas	RH=10, σ_z =4.7, Resolution is 4 km for CONUS and 9 km for AK. There are no emissions in HI, PR or VI. **Temporal:** monthly profiles that align with the county/scc temporal profiles used in CMAQ v7.1 platform. Use qflag= MONTH and generate county-specific run group monthly profiles based on benzene emissions aggregated over all SCCs within each county.	Oil and gas sources reported in the nonpoint NEI data category (i.e., county-level emissions).
RWC12 RWC9AK RWC3HI	**NEI:** nonpoint **Platform:** rwc	RH = 6.4, σ_z = 3.2, Resolution is 12 km for CONUS, 9 km for AK and 3 km for HI. There are no emissions in PR or VI. **Temporal:** hourly by grid cell based on county-specific hourly emissions (created by SMOKE using year-to-day factors derived from meteorological data that are used for many SCCs in this sector)-for all 8760 hours in the year-based on benzene emissions summed by hour and county across all SCCs in the run group.	Fireplaces, woodstoves, hydronic heaters used for residential heating.
AG12	**NEI:** nonpoint **Platform:** ag	RH = 1, σ_z = RH/2.15 = 0.465 m. Resolution is 12 km. There are no emissions in AK, HI, PR or VI. **Temporal:** hourly profile based on ammonia hourly emissions for the Ag emissions sector (which includes both livestock and fertilizer).	Miscellaneous area sources; agriculture production – livestock, beef cattle – finishing operations, dairy cattle – drylot/pasture dairy-confinement; swine production – operations; poultry production.

Each run group contains one or more source groups for purposes of presenting the NATA results. There are no situations where a source group fits into multiple run groups. The 2014 NATA uses groups similar to the 2011 NATA. The key differences are that mobile source groups were expanded and some nonpoint source groups from the 2011 NATA were combined. Table 2-22 lists these source groups.

Table 2-22. NATA source groups

2014 NATA Source Group Name	NATA Category	Source Group Description	Source Group Definition Compared to 2011 NATA	2014 Run Group
PT-StationaryPoint	POINT	All facilities in the POINT data category that have geographic coordinates other than those with facility type = 100 (which are in the "Airport" source group) or facility type = 151 (which are in the "Railyard" source group)	Same as 2011 NATA	POINT
OR-LightDuty-OffNetwork-Gas	Onroad	On-road, light duty, nondiesel (i.e., gasoline & ethanol blends) vehicles – off network processes (e.g., "starts")	Different grouping	LDOFF
OR-LightDuty-OffNetwork-Diesel	Onroad	On-road, light duty, diesel vehicles – off network processes (e.g., "starts")	Different grouping	LDOFF
OR-HeavyDuty-OffNetwork-Gas	Onroad	On-road, heavy duty, **nondiesel** vehicles – off network processes (e.g., "starts")	Different grouping	HDOFF
OR-HeavyDuty-OffNetwork-Diesel	Onroad	On-road, heavy duty, diesel vehicles – off network processes (e.g., "starts")	Different grouping	HDOFF
OR-LightDuty-OnNetwork-Gas	Onroad	On-road, light duty, **nondiesel** vehicles – on network processes (e.g., running emissions)	Different grouping	LDON
OR-LightDuty-OnNetwork-Diesel	Onroad	On-road, light duty, diesel vehicles – on network processes (e.g., running emissions)	Different grouping	LDON
OR-HeavyDuty-OnNetwork-Gas	Onroad	On-road, heavy duty, **nondiesel** vehicles – on-network processes (e.g., "running")	Different grouping	HDON
OR-HeavyDuty-OnNetwork-Diesel	Onroad	On-road, heavy duty, diesel vehicles – on-network processes (e.g., "running")	Different grouping	HDON
OR-Refueling	Onroad	On-road refueling	Same as 2011 NATA	LDON
OR-HeavyDuty-Hoteling	Onroad	On-road, heavy duty diesel vehicle extended idling and auxiliary power units	Different grouping	HOTEL
NR-Recreational-inc-PleasureCraft	Nonroad	Off-road motorcycles, snow mobiles, golf carts, outboard pleasure craft, personal watercraft, inboard/sterndrive pleasure craft, etc.	New category – nonroad are broken out into equipment categories	NONRD
NR-Construction	Nonroad	Paving equipment, plate compactors, trenchers, tampers/rammers, tractors/loaders/backhoes, cranes, signal boards/light plants, etc.	Same as above	NONRD
NR-CommercialLawnGarden	Nonroad	Mowers, leaf blowers, turf equipment, chippers/ stump grinders, tillers, chainsaws, snow blowers, etc.	Same as above	NONRD

2014 NATA Source Group Name	NATA Category	Source Group Description	Source Group Definition Compared to 2011 NATA	2014 Run Group
NR-ResidentialLawnGarden	Nonroad	Mowers, leaf blowers, tillers, chainsaws, snow blowers, shredders, etc.	Same as above	NONRD
NR-Agriculture	Nonroad	Agricultural tractors, combines, sprayers, balers, tillers, irrigation sets, swathers, etc.	Same as above	NONRD
NR-CommercialEquipment	Nonroad	Generator sets, pressure washers, pumps, hydropower units, etc.	Same as above	NONRD
NR-AllOther	Nonroad	Industrial Equipment (Forklifts, sweepers/scrubbers, other oil field equipment), logging equipment, railroad maintenance, underground mining equipment	Same as above	NONRD
NR-CMV_C1C2_ports	Nonroad	Commercial marine vessel (CMV) emissions – C1/C2 while at ports	C1/C2 are separate from C3; in 2011 they were combined	NR-CMV
NR-CMV_C3_ports	Nonroad	CMV emissions – C3 while at ports	Same as above	NR-CMV
NR-CMV_C1C2C3_underway	Nonroad	CMV emissions – both C1/C2 and C3 marine – while underway	Same as 2011 NATA	NR-CMV
NR-Locomotives	Nonroad	Locomotive emissions	Same as 2011 NATA	NPLO12
NR-Point-Airports	Nonroad	Point source inventory where facility type = 100 "Airport"	Same as 2011 NATA	AIRPORT
NR- Point-Railyards	Nonroad	Point source inventory where facility type = 151 "Rail Yard"	Same as 2011 NATA	POINT
NP-industrial	Nonpoint	Nonpoint mining and quarrying, paved and unpaved roads, food and kindred products, mineral products, chemical manufacturing, non-ferrous metals and nonpoint industrial processes not elsewhere classified	Combines: NP refinery, NP chemical mfg, NP-nonferrous metals, NP- mining, NP NEC	NPHI
NP-CommercialCooking	Nonpoint	Commercial cooking (charbroiling, frying)	Same as 2011 NATA	NPLO
NP-OilGas	Nonpoint	Oil and gas operations (pumps, dehydrators, tanks, engines	Same as 2011 NATA	OILGAS
NP-SolventsCoatings	Nonpoint	Degreasing, dry cleaning, consumer and commercial solvents, industrial surface coating, non-industrial surface coating	Adds degreasing, dry cleaning, consumer and commercial solvents, industrial surface coating, non-industrial surface coating	NPLO

2014 NATA Source Group Name	NATA Category	Source Group Description	Source Group Definition Compared to 2011 NATA	2014 Run Group
NP-StorageTransfer_BulkTerminals_GasStage1	Nonpoint	Bulk terminals, petroleum, organic and inorganic chemical storage and transport – has gas station emissions of only stage 1 (tank refueling) not stage 2 (vehicle refueling)	Bulk terminals, storage and transport – has gas station emissions of only stage 1 (tank refueling) not stage 2 (vehicle refueling)	NPLO
NP-MiscellaneousNonindustrial	Nonpoint	Poultry/livestock, laboratories, dental alloy, motor vehicle fires, portable fuel containers, residential charcoal grilling	Same as 2011 NATA except that charcoal grilling is a new NEI category for 2014	NPLO
NP-FuelCombustion_not_RWC	Nonpoint	Residential, industrial, commercial and institutional (ICI) fuel combustion, but excludes residential wood combustion	Combines residential and ICI fuel combustion	NPHI
NP-ResidentialWoodCombustionRWC	Nonpoint	Residential wood combustion – woodstoves, fireplaces	Same as 2011 NATA	RWC
NP-WasteDisposal	Nonpoint	Open burning, yard and household waste, managed burning, slash (logging debris), onsite incineration, waste water treatment, composting, landfills, scrap materials	Combines landfills and other waste disposal	NPHI
NP- AgriculturalLivestock	Nonpoint	Emissions from livestock waste (HAPs only)	New for 2014	AG
FIRE	FIRE	CMAQ category only: wildfires, prescribed fires and agricultural burning; while the modeled results are not separated by type of fire, the tabular emissions summary breaks out the three fire types separately		
BIOGENICS	BIOGENICS	CMAQ category only: VOCs and particular HAPs (formaldehyde, acetaldehyde and methanol) emitted from vegetation (and NOX from soils)		
SECONDARY	SECONDARY	Not an emissions group: formed in the atmosphere due to photochemical reactions (CMAQ)		
BACKGROUND	BACKGROUND	Not an emissions group: concentrations due to ubiquitous nature of some HAPs with long residence time or coming into the modeling domain from outside the domain such as carbon tetrachloride		

The next section describes in more detail the development of emissions and characterization of the emissions for the run groups by NEI data category.

2.3.2. Point Sources Excluding Airports

Point sources used in AERMOD are all sources in the point data category in the NEI that are in the United States, Puerto Rico or Virgin Islands;[2] have geographic coordinates present in the NEI and emit at least one HAP. This section discusses the nonairport point sources; airports are discussed in Section 2.3.3.

The starting point for the emissions modeling for point sources is the SMOKE modeling file, or FF10 discussed in Section 2.1.3. The format of this point FF10 file is in the SMOKE User's Manual https://www.cmascenter.org/smoke/documentation/4.0/html/ch08s02s08.html#sect_input_ptinv_ff10 and the file is posted on the emissions modeling platform ftp site.

The SMOKE FF10 was split into the CMAQ platform sectors (see Table 2-10) ptnonipm, ptegu and pt_oilgas, and the ptnonipm sector was further split out into airports and nonairports using the FACILITY_SOURCE_TYPE field. The FACILITY_SOURCE_TYPE indicates the type of facility (where available) and a value of "100" is used for airports (including large and small airports, heliports and seaplanes).

For purposes of AERMOD, the ptnonipm (after removal of airports) and pt_oilgas were run in the same way and were not distinguished from one another. The ptegu sector was treated differently due to different temporalization. The ptegu sector allowed for hourly variation at the unit level, whereas the ptnonipm (including pt_oilgas) used the monthly, day-of-week and diurnal profiles used by SMOKE. Within the run, each ptegu unit was temporalized using hourly emission values as discussed in Section 2.3. Non-EGU units were modeled in AERMOD using temporal allocation factors derived from the temporal profiles used in SMOKE for CMAQ. Many facilities included a mixture of EGU and non-EGU processes. In such cases, all sources at a given facility were modeled in the same AERMOD run.[3] This ensured that ambient impacts were calculated for a consistent set of receptor locations for all sources at the facility.

2.3.2.1. Point source characterization for AERMOD

In the SMOKE FF10 modeling file, a source is a unique combination of EIS process id and EIS release point id. In AERMOD, we chose to group together sources with the same release point characteristics (geographic coordinates, release point type and release point parameters) and temporal profile since they will have the same X/Qs, and grouping them reduces the number of AERMOD runs needed. This is because many processes use the same temporal profiles, so there are fewer unique combinations of these parameters than process id/release point id combinations. In the 2014 NEI, there are 75 percent fewer unique HAP sources (based on temporal and release parameters) than process-id/release point id combinations. The FF10 fields that must be unique (when combined) are: FACILITY_ID, MONTHLY temporal profile code, WEEKLY temporal profile code, ALLDAY temporal profile, ERPTYPE (except that ERPTYPEs 3, 4 and 6 are treated as equivalent), STKHT, STKVEL, STKTEMP, STKDIAM, FUG_HEIGHT, FUG_WIDTH_YDIM, FUG_LENGTH_XDIM, FUG_ANGLE, LATITUDE and LONGITUDE. Sources within the

---

[2]Offshore platforms (i.e., FIPS = 85; meaning federal waters) are excluded.
[3]AERMOD temporalization is performed at the level of source IDs, so using different temporalization schemes at one facility is possible.

same facility were run together in AERMOD. To assign the proper meteorological data, each facility was assigned to a grid cell based on the geographic coordinates of the release points.

Helper files were developed for point locations, point source parameters, area source parameters, temporal parameters and emissions. Point locations were based on the latitude and longitudes in the FF10 (the release point coordinates) and were converted to UTM coordinates. Point source parameters used the release point type code (field name is ERPTYPE in the FF10 file) and associated stack or fugitive parameter information from the FF10.

Table 2-23 lists the release point types and how each is assigned an AERMOD type.

Table 2-23. Assignment of AERMOD source type for point sources

Emission Release Point Type Code (NEI)	Emission Release Point Type Description	AERMOD Source Type Code or Special Adjustment
1	Fugitive	AREA
2	Vertical	POINT
3	Horizontal	POINTHOR
4	Goose neck	POINTHOR
5	Vertical with rain cap	POINTCAP
6	Downward-facing vent	POINTHOR

For AERMOD area sources, the following fields are used to characterize a fugitive release (units in parentheses):

- Latitude and longitude (decimal degrees)
- FUG_HEIGHT – fugitive height (ft)
- FUG_WIDTH_YDIM[4] – fugitive width (east/west), (ft)
- FUG_LENGTH_XDIM[4] – fugitive length (north/south), (ft)
- FUG_ANGLE– fugitive angle (degrees)

The NEI allows fugitive release angles of 0 to 89 degrees. While the fugitive release point latitude and longitude are not specified to be any particular location in the area source in the emission inventory system, due to the limits on the angle and the general conventions for specifying the source in AERMOD, we interpret the release point as the follows: 1) the lat/lon is treated as the most western corner; 2) the angle is measured clockwise from true (geodetic) north (not magnetic north); 3) length is the measure along the side that would run in the north-south direction if the angle were 0 degrees and 4) width is the measure along the side that would run in the east-west direction if the angle were 0 degrees.

As shown in Figure 2-2, the release point is at the push pin, the width is 680 feet, the length is 1897 feet and the angle is 22 degrees.

Geographic coordinates were converted to UTM, with all release points within the same facility assigned to the same UTM zone. Release parameters are converted to meters in SMOKE as required by AERMOD.

---

[4]These SMOKEFF10 variables will be renamed to FUG_WIDTH_XDIM and FUG_LENGTH_YDIM to keep with the convention of X as the east-west direction and Y as the north-south direction.

The initial vertical dispersion (σ_z) was set to FUG_HEIGHT/4.3 if the FUG_HEIGHT was greater than 10 meters, and 0 m if otherwise.

For AERMOD point sources, the following fields are used to characterize a stack release. The SMOKE FF10 units are listed below; they are converted to metric (m, m/s, Kelvin) as required by AERMOD.

- Latitude and longitude (decimal degrees)
- STKHGT– height of stack (ft)
- STKDIAM – diameter of stack (ft)
- STKTEMP – temperature of stream exiting stack (°F)
- STKVEL – velocity of stream exiting stack (ft/s)

Separate helper files are derived for AREA versus POINT AERMOD source types because the source parameters required for these two types differ.

Figure 2-2. Example fugitive source characterization:
NEI length = 1897 feet, width = 680 feet and angle = 22

2.3.2.2.   Point source temporalization

As in CMAQ, point sources in the ptnonipm sector of the NATA CMAQ platform were modeled differently in AERMOD from those in the ptegu sector with respect to the temporalization of the emissions. The ptegu sectors were temporalized allowing for hourly variation at the unit level, whereas the ptnonipm and pt_oilgas sectors used the monthly, day-of-week and diurnal profiles used by SMOKE for the CMAQ platform. Within the run, each ptegu unit was temporalized using hourly emission values

as discussed below. Non-EGU units were modeled in AERMOD using temporal allocation factors derived from the temporal profiles used in SMOKE for CMAQ. Many facilities included a mixture of EGU and non-EGU processes. In such cases, all sources at a given facility were modeled in the same AERMOD run.[5] This ensured that ambient impacts were calculated for a consistent set of receptor locations for all sources at the facility.

For non-EGUs, temporal helper files were prepared based on the temporal cross reference file for point sources (PTREF) and the temporal profile files (TPRO) used by SMOKE (same data as are used for CMAQ). The AERMOD scalars reflect diurnal, day-of-week and/or monthly variability and are determined from the SMOKE diurnal, weekly and monthly temporal profile values. Table 2-24 shows the temporal variability options for non-EGUs.

Table 2-24. Options for temporal variation specification in helper files

Temporal Variability	TPRO	Qflag (for AERMOD)	Number of Scalars	Notes and Calculation of Scalar Using SMOKE TPRO Fractions1
Uniform – every hour of every day emits the same amount	Monthly profile code 262, daily profile code 7, hourly profile code 24	MONTH	12	AERMOD Scalar value is equal to 1/12 for each of the 12 months (identical to profile code 262)
Monthly Variation only	Day-of-week profile code 7, hourly profile code 24	MONTH	12	AERMOD Scalar value month is $monthfrac_m$ Where $monthfrac_m$ is monthly fraction from the TPRO monthly profile assigned to the source
Diurnal Variation Only	Monthly profile code 262, day-of-week profile code 7	HROFDAY	24	AERMOD Scalar value month is $dayfrac_d$ Where $dayfrac_d$ is the diurnal fraction from the TPRO diurnal profile assigned to the source
Month and hour of day type variation in which M-F is the same, but Sat and Sunday can be different	Diurnal profile code same for all weekdays	MHRDOW	864 (=24hrs*3day types*12 months)	AERMOD Scalar value = $$monthfrac_m * dayfrac_d * hourfrac_h$$ Where $monthfrac_m$ is the monthly fraction, $dayfrac_d$ is the day-of-week fraction and $hourfrac_h$ is the diurnal fraction for each hour for weekdays, Saturday and Sunday from the TPRO 5 × sum of weekday scalars + sum of Saturday scalars + sum of Sunday scalars = 1

---

[5] AERMOD temporalization is performed at the level of source IDs, so using different temporalization schemes at one facility is possible.

Temporal Variability	TPRO	Qflag (for AERMOD)	Number of Scalars	Notes and Calculation of Scalar Using SMOKE TPRO Fractions1
Month and hour of day variation and every day of the week could be different	Diurnal profile code varies by day-of-week	MHRDOW7	2016 (=24 hrs*7 day types * 12 months)	AERMOD Scalar value = $$monthfrac_m * dayfrac_d * hourfrac_h$$ Where $monthfrac_m$ is the monthly fraction, $dayfrac_d$ is the day-of-week fraction and $hourfrac_h$ is the diurnal fraction for each hour every day of the week from the TPRO Sum of scalars is 1

For EGUs, hourly temporalization was used, based on continuous emission monitoring data for heat input. The approach is the same as used in CMAQ and is described in more detail in the 2014v7.1 modeling platform TSD. For AERMOD, separate facility-specific helper files were developed with the hourly scalars for each hour of the year.

2.3.3. Airport Point Sources

Airports are in the SMOKE point FF10 file and can be identified by the FACILITY_SOURCE_TYPE field "100" (for an airport facility). We modeled airports as one of two types: runway line sources or small (10- by 10-meter) area sources. The runway airports are those for which we have geographic coordinates for the runway endpoints from either the 2014 National Transportation Atlas Database or the OTAQ-supplied database used for the 2011 NATA. The nonrunway airports have no runway information and were modeled as small area sources. These include smaller airports, seaplanes and heliports. Although there may be multiple processes at an airport (e.g., commercial aircraft emissions and ground support equipment emissions), we did not treat them differently with respect to source characterization. Emissions at an airport were summed and apportioned equally among the runways or to the small area source. Airports have multiple sources in the AERMOD helper files only if the airport had more than one runway.

Runway information is not in the NEI. The NEI has a site latitude and longitude, and all release points (fugitives) use the site coordinates. To get runway endpoints, we matched the NTAD/OTAQ databases with the NEI emissions. Non-matches (NEI facilities with no match to NTAD/OTAQ based on airport identifiers, or NEI facilities with coordinates inconsistent with the runway locations) were modeled as area sources based on the NEI coordinates. Matches were done using the airport location id (typically a 3- or 4-character field), which is an alternate facility id in EIS and is the id used for the NTAD/OTAQ airports. The starting-point airport facilities and final runway airports resulting from the matching effort, and detailed set of process steps taken to do the matching, are available in the Supplemental Data folder. This process resulted in 7,384 nonrunway airports and 11,991 runway airports.

2.3.3.1. Airport characterization for AERMOD

For runway airports, an airport source is a runway at an airport facility; for nonrunway airports, an airport source is a 10-by-10-meter area source at an airport facility. Runways are modeled as AERMOD "LINE" sources, using the coordinates of the runway endpoints, which are assumed to be in the center of the runway width. All runways are assumed to be 50 meters wide (NTAD-based runways) or 25 meters wide (OTAQ-based runways) and have a release height and initial vertical dispersion of 3 meters. All pollutant emissions are divided equally across all runways at the airport.

Airports without runway characterization are modeled as 10m by 10 m area sources with the same release height and vertical dispersion as runway airports. The NEI airport coordinates are used as the southwest corner of the area source and the angle is 0.

2.3.3.2. Airport temporalization

Airports used a single set of temporal profiles for all processes within the airport, since the emissions from all processes are combined on the airport runways or as the small area sources.

All airports use the same set of monthly, day-of-week and diurnal profiles except for Alaska seaplanes.

For airports other than Alaska seaplanes, weekly and monthly temporal profiles are based on 2014 data from the FAA Operations Network Air Traffic Activity System (Figure 2-3) (http://aspm.faa.gov/opsnet/sys/Terminal.asp). A report of all airport operations (takeoffs and landings) by day for 2014 was generated. These data were then summed to month and day-of-week to derive the monthly and weekly temporal profiles. An overview of the Operations Network data system is at http://aspmhelp.faa.gov/index.php/Operations_Network_%28OPSNET%29. These are profile codes 727 (monthly) and 747 (weekly). For Alaska seaplanes, we used the flat weekday profile and spring/summer weighted profile as used in the 2011 NATA. The Alaska sea plane profile codes are 8888 (monthly) and 7 (weekly).

For airports other than Alaska seaplanes, hourly airport operations data were obtained from the Aviation System Performance Metrics (ASPM) Airport Analysis website (https://aspm.faa.gov/opsnet/sys/airport.asp). A report of 2014 hourly Departures and Arrivals for Metric Computation was generated. An overview of the ASPM metrics is at http://aspmhelp.faa.gov/index.php/Aviation_Performance_Metrics_%28APM%29. For Alaska seaplanes, we used the same spring/summer-weighted monthly profile that was developed in the 2011 NATA. The profile code is 777. For Alaska seaplanes, we used the same diurnal profile as 2011 NATA, which zeroes out emissions prior to 6 am or after 10 pm (see Figure 2-4).
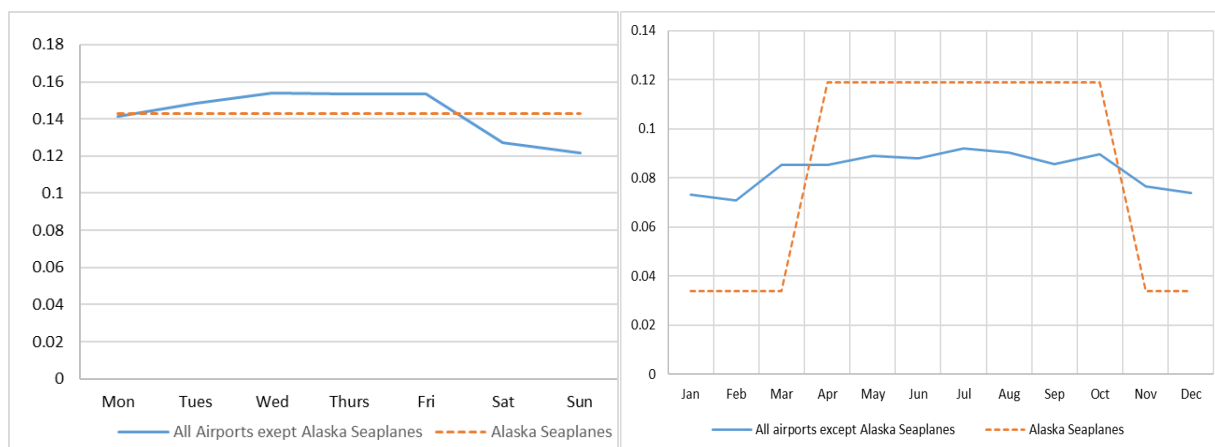
Figure 2-3. Airport weekly and monthly temporal profiles
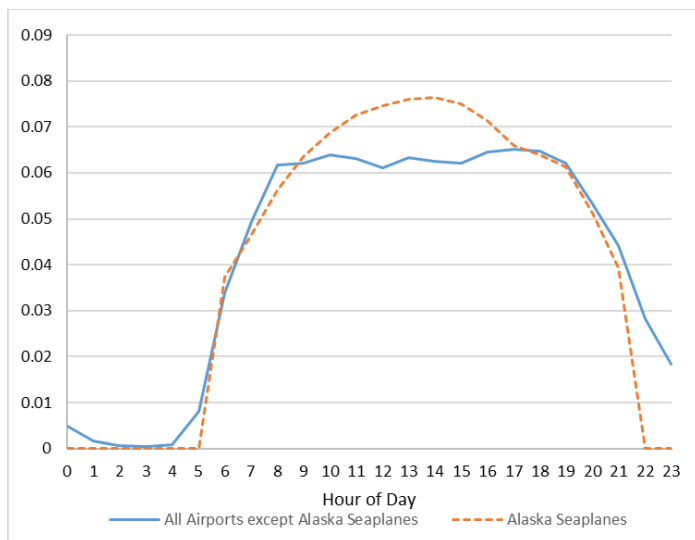
Figure 2-4. Airport diurnal temporal profiles

2.3.3.3. Airport emissions

Emissions are applied to X/Qs for airports differently than other point sources. Instead of multiplying the source-based emissions to the X/Qs, we multiply the *total facility* emissions to a facility-aggregated X/Q, which for runway airports is the aggregate across all runways. In creating the emissions inputs to AERMOD, the unit emission rate, 1000 g/s, is the unit emission rate for the entire airport, not each runway. To apportion the 1000 g/s to each runway, the 1000 g/s is divided by the number of runways. AERMOD then outputs the annual average concentration for the entire airport, not each runway. The annual average concentrations reflect the apportionment between the different runways. Because it can be assumed that each runway will get the same share of actual HAP emissions as the unit emission rate based on the fraction variable, the specific HAP emissions for the entire airport can be used instead of runway-specific emissions. This approach reduces the number of records in both the emissions helper file used for X/Q and the AERMOD annual average concentration output file based on the unit emission rate.

The NEI estimated that lead emissions from piston aircraft totaled 456 TPY, nationally. Of these emissions, the NEI estimated that lead emitted during several modes of operation at and near airports totaled 228 TPY. These estimates were used in NATA. The NEI also estimated in-flight lead emissions at the state level, which happened to equal the lead emitted from the piston aircraft (228 TPY). In-flight lead emissions were not included in NATA. Lead emissions near airports included lead emitted during the climb-out and approach modes, which occur at altitude and are not included in NATA. To account for this, we adjusted the NEI-specific lead[6] emissions estimates used in AERMOD down by 50 percent, based on previous modeling conducted at the Santa Monica (SMO) airport indicating that nearly 50 percent of emissions occurred in these higher-altitude modes (see Table 2-25).

---

[6]We designed the programs to reduce all metals from the following 2 SCCs: 2275050011=General Aviation /Piston and 2275060011=Air Taxi /Piston by 50%; however, the only metal emissions from these 2 SCCs are lead, so only lead was reduced.

Table 2-25. Lead emissions (kg/yr) at SMO in 2008 by aircraft operation mode

Mode	Emissions (% of Total)
Taxi to runway	20.4 (17.6%)
Run-up	13.5 (11.4%)
Takeoff roll	10.0 (8.4%)
Climb-out	37.9 (32.7%)
Approach	17.9 (15.8%)
Landing	9.4 (7.9%)
Taxi to apron	9.5 (8.4%)

The emissions helper file format for runway and nonrunway airports is the same and is shown in Table 2-26.

Table 2-26. Airport emissions file format

Field Name	Description
State abbrev	2-character abbrev. Use TB for tribe
Facility id	Identifer for facility (i.e., EIS ID)
Facility name	Facility name
FAC_SOURCE_TYPE	Code indicating facility type (from SMOKEFF)
pollutant name	Use SMOKE shortname
Emissions (tons)	Emissions from the SrcID multiplied by the Metal/CN speciation factor (column AI) from the NATA_Pollutants.xslx) For airports, for metals only, multiply metal pollutant emissions from (2275050011=General Aviation /Piston; 2275060011=Air Taxi /Piston) by 0.5

2.3.4. Nonroad, On-road and Nonpoint – County-level Sources

Nonroad, on-road and nonpoint (other than CMVs and locomotives) sources in the NEI are estimated at the county level and use various levels of temporal variation. For AERMOD, we characterize these as "AREAPOLY" sources, which provide an emission flux over a defined polygon shape. We also characterize the temporal variability based on the various options available in AERMOD.

The main steps needed to prepare county-level sources for AERMOD are spatial allocation, assignment of temporal scalar factors and assignment of release parameters. County-level sources must undergo spatial allocation prior to being input into AERMOD. All county-level sources are allocated to finer resolution using spatial surrogates. For CONUS domain sources, they are allocated to 4-km or 12-km resolution, and the size is primarily based on our confidence in the surrogate data at fine resolution and its representativeness of the source category for which it is being used. For non-CONUS county-level sources, they are allocated to 9-km resolution (Alaska) or 3-km resolution (Hawaii, Puerto Rico and the Virgin Islands). This is a major difference from the 2011 NATA approach, which allocated county-level emissions to census tracts. Mostly, the same underlying surrogate data are used for AERMOD as CMAQ and are discussed in detail in the TSD developed for CMAQ emissions (2014v7.1 Emissions TSD).

Table 2-15 through Table 2-17 provide the available surrogates. One difference is that the National Land Cover (NLCD) data are not available for Hawaii, Puerto Rico and the Virgin Islands (i.e., grids 3HI and 3PR), so an alternative land cover database, the Coastal Change Analysis Program (C-CAP) database was used for the land cover categories. C-CAP does not have as many categories as the NLCD; low and medium intensity land was not available and other land categories were substituted. For example, if the surrogate called for open+low, then only open was used in the 3HI and 3PR grids. Another difference is that CMAQ uses 12 km for all gridded emissions whereas AERMOD uses 4 km for some of the source groups in the CONUS; also, CMAQ does not cover non-CONUS areas. As in CMAQ, the spatial surrogates are assigned according to SCC codes. The SCC-to-surrogate cross reference and surrogate documentation workbook is provided in the Supplemental Data folder.

In addition to spatial allocation, the emissions are temporally allocated to capture the temporal variability of emissions throughout the year, day and/or hour of day. Temporal allocation is done because not all emissions are emitted uniformly throughout the year. Ideally each specific SCC (and in some cases SCC and county) would be given a different temporal profile as is done when preparing emissions for CMAQ. But because all sources in the same run group must have the same scalars, we need temporal scalars that best represent the aggregation of the individual SCCs. In addition, these scalars, apart from onroad and nonroad mobile run groups, do not vary by pollutant because AERMOD is run at a unit emissions rate and the pollutant-specific emissions are applied after the model is run. The scalars can vary by county. For the onroad and nonroad mobile categories, AERMOD preserves the pollutant-specific monthly variation.

For each run group, SMOKE is used to produce AERMOD helper files that provide the locations and coordinates of the gridded sources, the area source parameters, the temporal scalars and the emissions by source and pollutant, broken out by run group.

The subsections below provide more information on how the emissions are developed and modeled in AERMOD. Detailed information on the development of the emissions at county level is provided in the 2014 NEI v2 TSD.

2.3.4.1. On-road

On-road uses 5 run groups: LDON4, LDOFF12, HDON4, HDOFF12 and HOTEL4, as described in Table 2-21. These run groups distinguish on-network (ON) versus off-network (OFF) sources, heavy-duty versus light-duty sources, and hoteling (extended idling and auxiliary power units).

For characterization, the release height of LD ON vs OFF is different (1.3 m vs. 0.5 m) to account for added dispersion from the vehicle wake. Sigma-z is also different (1.2 vs. 0.5) to account for more dispersion on the roads. Off-network sources use minimal dispersion associated with start emissions (no turbulence). HD on and off and hoteling (which is pertinent only to HD vehicles) use the same release height based on average tailpipe height (3.4 m) but different dispersion. Sigma-z for on-network is 3.2 m, but for off-network and hoteling it is 0.5 m to account for less dispersion during idling.

On-network sources and hoteling use finer resolution (4 km vs 12 km) in the CONUS domain than for off-network. This is because we have more confidence in the spatial surrogates used to allocate the county-level emissions. The on-network sources use AADT data, whereas the off-network emissions are allocated to broad classes of land use data. Off-network exhaust and evaporative emissions from passenger cars are allocated to the "all development" land category, which consists of the sum of the following land categories: developed, open space; developed, low intensity; developed, medium

intensity; and developed, high intensity. This is land with greater than zero percent impervious surface. For HI, PR and VI, which use C-CAP, there is no low or medium intensity, so it is the sum of open space and high intensity. Hoteling uses a set of truck stops that have been updated since the 2011 NATA.

The temporal resolution for onroad run groups is hourly across the year. Different profiles were developed for off-network than on-network due to the types of mobile processes (running vs. starts, for example) that occur. They are also different for heavy duty versus light duty because of the different vehicle mixes (trucks vs. cars). For the v2 modeling platform, more refined temporal profiles were developed for the diurnal variation in vehicle miles travelled based on telematics data analyzed under the CRC Project A-100 (ERG 2017).

In developing the hourly scalars for CONUS areas, we used the SMOKE-MOVES hourly emissions reports by county and SCC for a key pollutant for each run group. We chose $PM_{2.5}$ for HD run groups and benzene for LD run groups. We then aggregated emissions across all SCCs in that run group. Hourly emissions were converted to local time (considering daylight saving time where appropriate). SMOKE-MOVES was run only for the CONUS. National averages were used for non-CONUS temporal profiles.

Because we run grid cells (and not counties) in AERMOD, we assigned each grid cell to a county for purposes of assigning county-specific hourly emissions. Due to the unit emissions rate, we could not account for differences across pollutants in temporal variation, which can occur due to different temperature impacts on EFs that vary by pollutant, except at the monthly level. We imparted monthly variation to on-road run groups by pollutant by running unit emissions rates by SEASON to get season-specific X/Q values, and then by post-processing the X/Q by at the seasonal level (i.e., applying seasonal emissions). This is further discussed in Section 3.4.

The temporal variability of on-road mobile sources varies by emissions process (e.g., running vs. idling), vehicle type, road type and pollutant. Different pollutants can have different temporal variation because some pollutants' emission factors depend on meteorology, which has diurnal variation.

Figure 2-5 and Figure 2-6 show example Wednesday and Sunday diurnal profiles in April for King County, WA (55033), Cook County, IL (17031) and Wake County, NC (37183). The single-unit short-haul starts appear to peak in the morning, and that peak is particularly pronounced for $PM_{2.5}$. Other pollutants do not show such a large gradient.

For non-CONUS areas, we used hourly scalars by run group that were developed by taking a national average of the hourly scalars in the CONUS area.

2.3.4.2. Nonroad

Nonroad equipment, i.e., emissions other than planes, trains and ships that are generated by MOVES, are put into a single run group. While there are differences in types of equipment and therefore emissions characterization, for this national study we chose to use a more simplified approach by characterizing equipment with the same spatial resolution, temporal profile and release characteristics.

The release characteristics are a release height of 2 meters and an initial vertical dispersion of 1 meter, the same as used for the 2011 NATA. We chose a spatial resolution of 12 km for the CONUS domain; however, we used a variety of spatial surrogates for different SCCs. For example, for agricultural nonroad SCCs such as combines, tractors and balers, agricultural land from NLCD was used, whereas for

pleasure craft, water (also from NLCD) was used. For non-CONUS, we also used land use surrogates and allocated to grid cells.

For the 2014 NATA, we revised the temporal allocation approach used in the 2011 NATA to be more consistent with changes made for the 2014 modeling platform for the diurnal variation. We changed the diurnal profile from profile 26, which puts a considerable portion of emissions at night (as shown in Figure 2-7), to county-specific nonroad average profiles based on the equipment-specific profiles shown in Figure 2-7. These profiles remove the nighttime emissions from lawn & garden and agriculture (profile code 25a), remove night time and early morning emissions from residential lawn and garden (profile code 27), and reduce the nighttime emissions from construction (profile code 26a). County-specific profiles enable us to account for the different mix of nonroad equipment in different counties. We used benzene to create the county-specific averages. The county-specific profiles are represented by the gray band in Figure 2-8; selected counties are provided for reference. Similar to the 2011 NATA, we did not use the weekday variation used in the modeling platform, which is less important than diurnal variation in terms of impact of emissions/meteorological iterations.

We revised the seasonal variation to run AERMOD monthly by applying pollutant specific monthly emissions to the X/Q. Note that the month-specific emissions applied were at source group resolution, not SCC-level resolution.
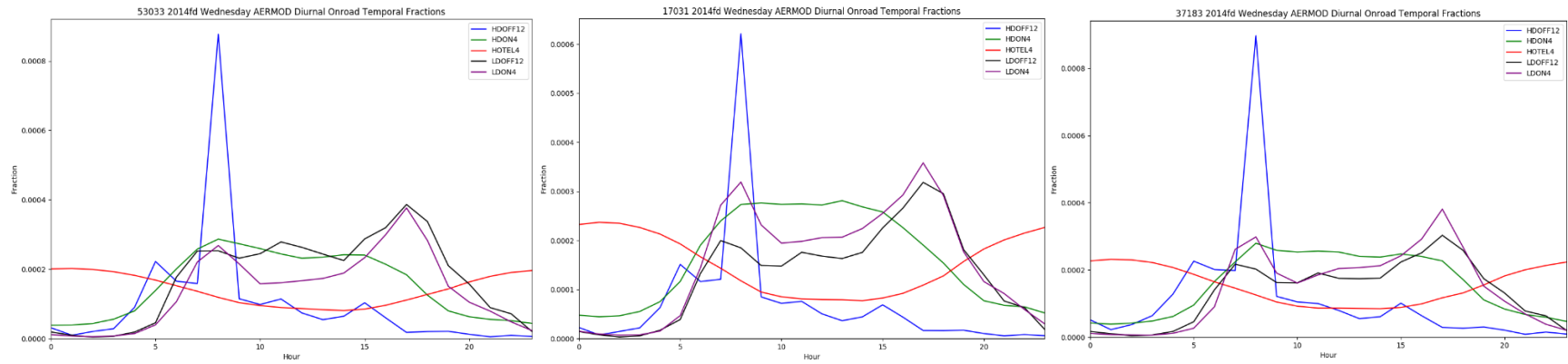
Figure 2-5. On-road weekday diurnal temporal profiles for King County, WA (55033); Cook County, IL (17031) and Wake County, NC (37183)
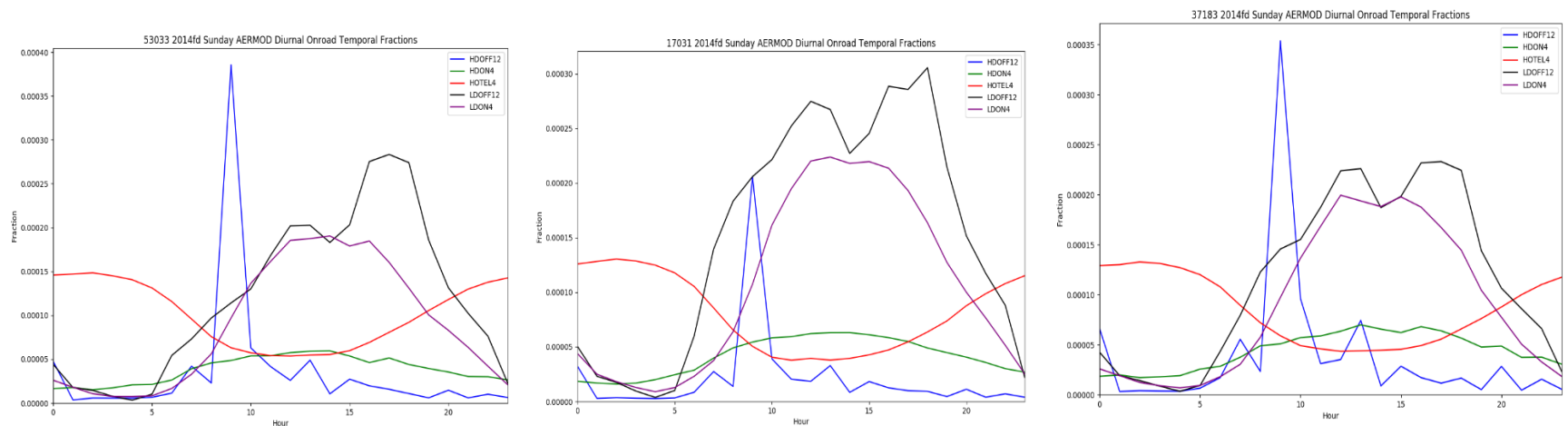
Figure 2-6. On-road Sunday diurnal temporal profiles for King County, WA (55033); Cook County, IL (17031); and Wake County, NC (37183)
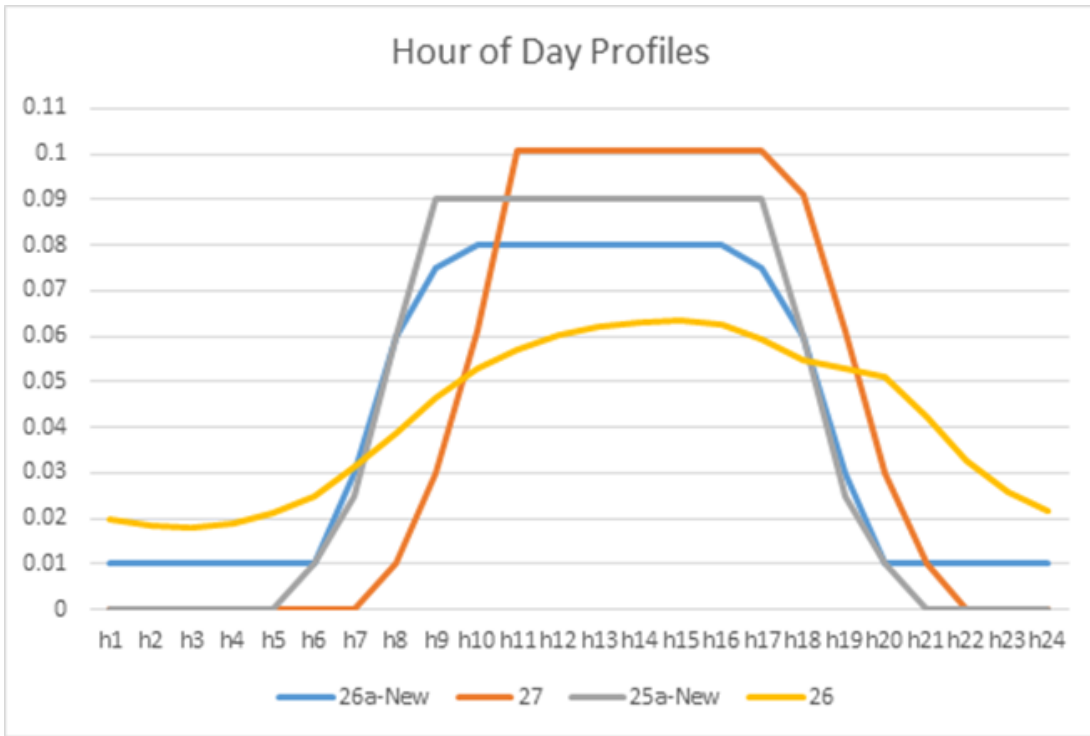
Figure 2-7. Hourly pattern of activity for SMOKE Profile 26
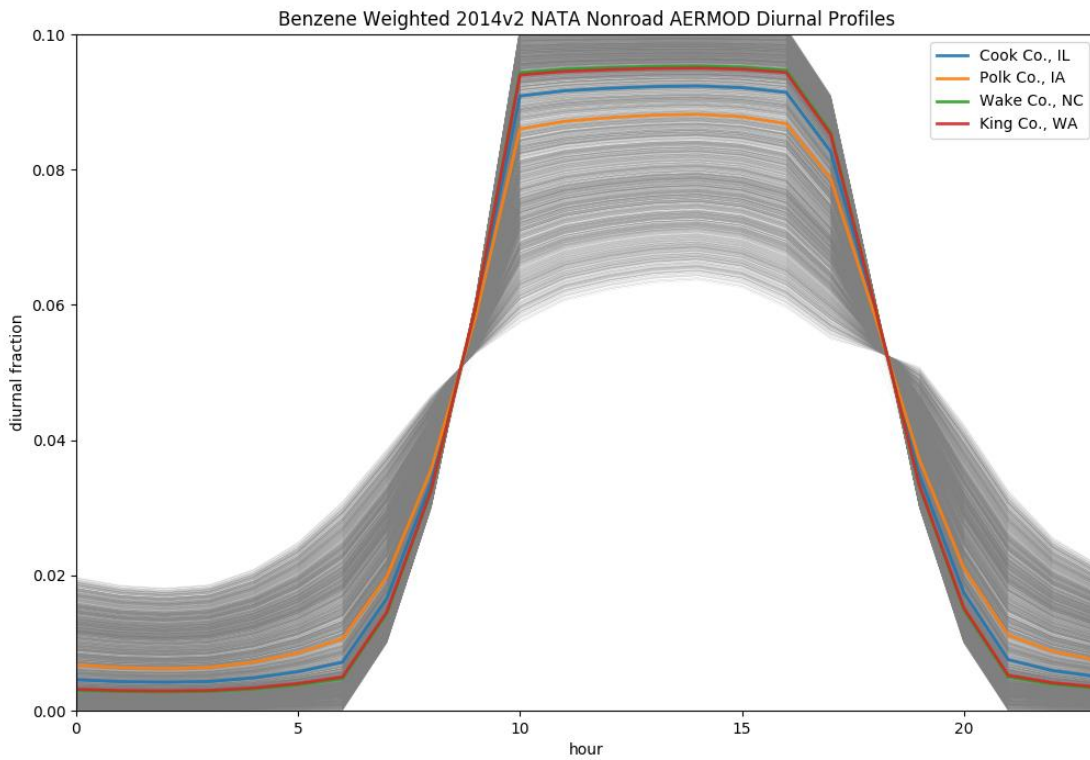
Figure 2-8. Hourly pattern of activity for county-average profiles used for the nonroad run group in AERMOD modeling

2.3.4.3. Nonpoint – NPHI12 and NPLO12

Nonpoint stationary sources are broken out into several run groups: NPHI, NPLO, RWC and OILGAS. NPHI and NPLO are very similar; they use the same temporal profile (diurnal profile 26, which is shown in Figure 2-7 and uniform weekday and monthly profiles. They are allocated to a 12-km spatial resolution (12 km for CONUS sources and a 9- or 3-km resolution for non-CONUS). NPHI uses a release height of 10 meters and an initial vertical dispersion of 4.7 meters, which is based on release height/2.15. NPLO uses a release height of 3.9 meters and an initial vertical dispersion of 3.6 meters. Because we combined all fuel combustion other than Residential Wood Combustion (RWC) into a single run group, we had to move residential nonwood combustion, which had been characterized with the lower release height/vertical dispersion in the 2011 NATA, to the NPHI12, which uses the same release height as industrial, commercial and institutional (ICI) fuel combustion.

Other than locomotive emissions, the two nonpoint run groups are the same as the CMAQ "nonpt" sector. The CMAQ "nonpt" sector excludes locomotive emissions, which are in the NPLO run group.

2.3.5. Residential Wood Combustion

As in the 2011 NATA, this is a source group that has its own run group; i.e., the run group and source group are identical. It is separate from other run groups because its temporalization, which depends on daily temperature, is unique. This run group comprises wood stoves (indoor and outdoor) and fireplaces. The NEI SCC codes are 2104008* (includes wood stoves and fireplaces, fire pits/chimneas and hydronic heaters) and 2104009000 (fire logs).

We used the same release height and initial vertical dispersion as in the 2011 NATA – 6.4 m height and 3.2 m vertical dispersion. All SCCs in the source group used the same spatial surrogate: NLCD Low Intensity Development.

The temporal approach is the same as used in 2011, which used temperature data to assign emissions to days of the year. This was done using the "Gentpro" feature of SMOKE that was developed specifically for this sector, as the level of residential wood combustion activity depends on the daily minimum temperature. For 2014, we took the same approach using 2014 meteorological data. Diurnal profiles based on device type were also applied. For AERMOD, we took the SMOKE-generated hourly emissions by SCC and created the temporal scalars using benzene emissions for all SCCs by hour of the year. Hourly scalars were computed by county as:

$$Scalar_{hour\text{-}i} = 8760 * (Benzene)_{county, hour\text{-}i} / \Sigma_{all\ hours, county}(Benzene)$$

To assign grid cells to counties, we used the county having the most RWC HAP emissions.

Puerto Rico and the Virgin Islands do not have any emissions for this source group. For Alaska and Hawaii, the hourly scalars were computed in the same way as the CONUS scalars. However, the underlying data SMOKE used for generating the hourly benzene emissions were not based on day-specific meteorological data. Instead, monthly temporal profiles for Alaska and Hawaii were calculated from the national average of the 2014 meteorology-based profiles, and SMOKE SCC-to-profile mappings were used for daily and diurnal variation.

2.3.6. Oil and Gas

This run group covers all nonpoint oil and gas sector emissions in the NEI. It contains only one source group – the oil and gas source group. The main purpose for breaking nonpoint oil and gas sources out of the NPHI run group (as they were characterized in the 2011 NATA modeling) was to allow a finer spatial allocation resolution – 4 km as opposed to 12 km in the CONUS. A finer spatial resolution was chosen because the surrogates are based on well locations from a commercially available database called HPDI that compiles oil and gas data (such as individual well locations, production information, drilling information and well completion data) from state databases – the same underlying data used by the Oil and Gas tool during the creation of the NEI. Because of the different types of activity data underlying the tool, there are numerous spatial surrogates applied to the various SCCs that make up the nonpoint oil and gas category. The same source of data (HPDI) was used to develop temporal profiles by county and SCC. This run group uses only monthly variation. County-specific monthly scalars were computed from benzene monthly emissions across all SCCs from the run group in the county (as shown below). If for a certain county there are no benzene emissions, then the sum of all HAPs was used.

$$\text{Scalar}_i = \frac{\text{Benzene}_i}{\text{\# of days}_i} \times \frac{1}{\sum \frac{\text{Benzene}_i}{\text{\# of days}_i}}$$

Where:

    Benzene$_i$ = the tons of benzene in month i for the county associated with the met grid cell across all SCCs in OILGAS

    # of days$_i$ = the number of days in month i

The release height chosen is the same as NPHI – 10 m height and 4.7 m for the initial vertical dispersion.

2.3.7. Agricultural Livestock

Like oil and gas, this run group contains just the one source group – agricultural livestock. These emissions are new to the 2014 NEI v2, as previously neither VOC nor the HAP components were estimated. We kept this run group separate due to the different temporal nature of livestock emissions from other nonpoint run groups and to be consistent with the temporal allocation used in CMAQ. In CMAQ, these emissions were modeled in the "ag" sector and use a meteorologically based temporalization. For AERMOD, a separate run group was created for consistent temporalization across the two models. For AERMOD, hourly scalars were computed from hourly ammonia emissions. The release height chosen was 1 m, and 0.465 m for the initial vertical dispersion.

2.3.8. Commercial Marine Vessels – County/Shape-level Sources

This run group has the same sources as the cmv_c1c2 and cmv_c3 CMAQ modeling sectors except that for AERMOD, we include only sources in U.S. state or territory waters, whereas in CMAQ, ships in federal waters were also included. The NEI inventory uses different SCCs to distinguish between ships that are at ports (hoteling or maneuvering) versus ships that are moving along waterways, or "underway" (cruising or in low speed zones), and uses different SCCs for diesel versus residual oil ships. The NEI provides emissions by port or underway "shape," which are drawn to provide sub-county geographic resolution of the cmv emissions. In all, there are four SCCs in this run group. Table 2-27 lists the SCCs in the CMV run group.

Table 2-27. 2014NEI SCCs covered in the CMV run group

SCC	Sector	Description: Mobile Sources Prefix for All
2280002100	cmv	Marine Vessels; Commercial; Diesel; Port
2280002200	cmv	Marine Vessels; Commercial; Diesel; Underway
2280003100	cmv	Marine Vessels, Commercial; Residual; Port emissions
2280003200	cmv	Marine Vessels, Commercial; Residual; Underway emissions

For AERMOD, each source is characterized by a polygon shape with multiple vertices and SCC. The polygon shape is based on the NEI shape. It is assumed that the emissions are homogeneous in each polygon. The port polygon shapes are identical to the NEI port shapes; these were updated for the 2014 NEI v2. Note that in the NEI, a single port may be characterized by multiple NEI polygon port shapes. The polygon emissions for all polygons within a single port are proportional to the total port emissions based on the shape's area compared to the total area of all polygons for that port. The underway shapes were the same as used in the 2011 NATA except for a few new shapes in Puerto Rico and a few other areas. The NATA underway polygons were simplified from the NEI underway shapes such that multiple simplified polygons were constructed to represent complex NEI underway shapes. For underway shapes, the emissions of the simplified shape are assumed proportional to the NEI shape based on the simplified shape area. An example of a port shape is shown in Figure 2-9.

Figure 2-9. Port shapes for Los Angeles and Long Beach, California

One change from the NEI was to correct the county FIPS that were inadvertently used in the NEI but were no longer valid for year 2014 due to census changes. For NATA, these emissions were re-allocated to the valid counties as follows:

shape_id	old_fips	new_fips
20171	02280	02198
20190	02232	02105
20191	02232	02105
20192	02232	02105
20336	02232	02230
20539	02280	02195
20598	02201	02198
20599	02280	02275
20601	02232	02105
20602	02201	02198
20603	02201	02198
20604	02201	02275
20605	02201	02198
20619	02201	02198
20837	02232	02105

Each source's temporal profile was based on the SCC code and/or county; CMV uses only monthly variation; daily and diurnal profiles are uniform.

For the 2014 NATA, release characteristics are different for C1/C2 vessels than for C3. For C1/C2, we used a release height of 8.4 meters and an initial vertical dispersion of 3.907 meters. Category 3 vessels are larger and different release parameters were used for consistency with CMAQ. In CMAQ, C3 vessels were characterized as point sources so the model would compute plume rise. We developed the C3 AERMOD release height and initial vertical dispersion to be consistent with the characterization in CMAQ. We used the same release height, 20 m. For the initial vertical dispersion, we examined summaries of plume rise of C3 from CMAQ; in particular, we looked at a SMOKE report of the annual NO_X vertical distribution by county for the CMAQ cmv_c3 sector. For every county, we found that 50 percent of the emissions were above layer 3 and 50 percent were below. The midpoint of layer 3 is 60.7 meters. Therefore, we computed sigma-z as 60.7-20 = 40.7 m.

2.3.9. Urban/Rural Determination for All Emission Sources

The urban/rural determination for all emissions sources was based on the urban/rural classification of nearby census blocks. We classified each populated census block in the United States as urban if its centroid falls within an "urbanized area." For the 2010 Census, an urbanized area comprises a densely settled core of census tracts or census blocks, or both, that meet minimum population density requirements (50,000 or more people), along with adjacent territory. The adjacent territory generally is also densely settled, but it may also contain some lower-density areas as well as nonresidential urban land uses. About 500 such areas are included in the 2010 Census.

For point sources, each facility was designated as urban or rural based on the urban/rural classification of the nearest census-block centroid receptor to the average facility latitude/longitude (averaged across

all release points emitting HAPs). Each gridded source was designated based on the urban/rural classification of the nearest block receptor to the center of the 12-km grid cell (so all 4-km sources within the same 12-km grid cell are treated the same, based on the center of the 12-km grid cell). There were 353 grid cells on the borders of the United States that were inadvertently excluded from the approach; these were assigned a classification of rural, which would be expected based on their locations.

For CMV ports, each individual port shape was designated based on the urban/rural classification of the census block nearest to the center of the port shape. All CMV underway sources were designated as rural.

3. Air Quality Modeling and Characterization

The NATA emission estimates described in Section 2 are used as inputs to EPA air quality models to estimate ambient concentrations of emitted air toxics. An air quality model is a set of mathematical equations that uses emissions, meteorological and other information to simulate the behavior and movement of air toxics in the atmosphere. Air quality models estimate outdoor concentrations of air toxics at specified locations. The modeling approach for the HAPs in the NATA includes development and application of a hybrid approach blending a chemical transport model (CMAQ) with a dispersion model (AERMOD) to estimate ambient concentrations of about 50 of the more prevalent and higher risk HAPs as described in Section 3.5. The HAPs modeled in the hybrid approach capture a vast majority of the total risk nationally. Treatment for the remaining "non-hybrid" HAPs are described in Section 3.1.2.

3.1. Modeling Overview

3.1.1. Photochemical Model Selection

For 2014 NATA photochemical modeling, we used CMAQv5.2 (Appel et al., 2018) with the Carbon-Bond 6 (CB6-CMAQ) chemical mechanism. CMAQ is a comprehensive, three-dimensional grid-based Eulerian air quality model designed to simulate the formation and fate of gaseous and particulate species, including ozone, oxidant precursors, primary and secondary PM concentrations, and sulfur and nitrogen deposition over urban and regional spatial scales. 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. CMAQv5.2 includes many updates from the previous version 5.1. The model enhancements in version 5.2 include updates to the organic aerosol treatment (via new source pathways and updated properties of secondary organic aerosol (SOA)), updates to the gas-phase photochemistry (improved treatment of rural and remote chemistry and oceanic halogen chemistry), as well as various updates to the cloud treatment. While most compounds are grouped when model chemistry is applied, the CB6-CMAQ chemical mechanism treats formaldehyde, acetaldehyde, benzene, methanol and naphthalene explicitly within the chemistry. In this version, xylene concentrations are aggregated across the xylene isomers rather than separately generating concentrations for each isomer. For more information on CMAQ, see https://www.epa.gov/air-research/community-multi-scale-air-quality-cmaq-modeling-system-air-quality-management or http://www.cmascenter.org.

Table 3-1 lists HAPs included in the multipollutant version of CMAQv5.2 used for the 2014 NATA. The table also identifies the HAPs not run in CMAQ for the 2011 NATA that are included in this assessment.

Table 3-1. CMAQ HAPs

Air Toxic	Added for 2014	2014 CMAQ Species Name(s)
1,1,2,2-Tetrachloroethane	-	CL4_ETHANE1122
1,3-Butadiene	-	BUTADIENE13
1,3-Dichloropropene	-	DICHLOROPROPENE
1,4-Dichlorobenzene(p)	-	DICHLOROBENZENE
2,4-Toluene diisocyanate	-	TOL_DIIS
Acetaldehyde	-	ALD2, ALD2_PRIMARY
Acetonitrile	YES	ACETONITRILE

Air Toxic	Added for 2014	2014 CMAQ Species Name(s)
Acrolein	-	ACROLEIN, ACROLEIN_PRIMARY
Acrylic acid	YES	ACRYLICACID
Acrylonitrile	-	ACRYLONITRILE
Arsenic	-	AASI, AASJ, ASSK
Benzene (including benzene from gasoline)	-	BENZENE
Beryllium	-	ABEK, ABEI, ABEJ
Cadmium	-	ACDI, ACDJ, ACDK
Carbon tetrachloride	-	CARBONTET
Carbon tetrachloride (without background) *	YES	CARBONTET_NOBC
Carbonyl sulfide	YES	CARBONYLSULFIDE
Chlorine	-	CL2
Chloroform	-	CHCL3
Chloroprene	YES	CHLOROPRENE
Chromium Compounds	-	ACR_VIK, ACR_VIJ, ACR_VII
Chromium Compounds	-	ACR_IIIK, ACR_IIII, ACR_IIIJ
Diesel PM *	-	ADE_ECI, ADE_ECJ, ADE_OCI, ADE_OCJ, ADE_SO4J, ADE_NO3J, ADE_OTHRI, ADE_OTHRK, ADE_K
Ethyl benzene	YES	ETHYLBENZ
Ethylene dibromide (Dibromoethane)	-	BR2_C2_12
Ethylene dichloride (1,2-Dichloroethane)	-	CL2_C2_12
Ethylene oxide	-	ETOX
Formaldehyde	-	FORM, FORM_PRIMARY
Hexamethylene-1,6-diisocyanate	-	HEXAMETHY_DIIS
Hexane	YES	HEXANE
Hydrazine	-	HYDRAZINE
Hydrochloric acid	-	HCL
Lead Compounds	-	APBK, APBJ, APBI
Maleic anhydride	-	MAL_ANHYDRIDE
Manganese Compounds	-	AMN_HAPSK, AMN_HAPSJ, AMN_HAPSI
Mercury Compounds	-	HG,HGIIGAS, APHGI, APHGJ (there is no APHGK)
Methanol	-	MEOH
Methyl chloride (Chloromethane)	YES	METHYLCHLORIDE
Methylene chloride (Dichloromethane)	-	CL2_ME
m-xylene, p-xylene, o-xylene and xylenes (isomers and mixture)	-	XYLENES
Naphthalene	-	NAPHTHALENE
Nickel Compounds	-	ANIK, ANII, ANIJ
Polycyclic Organic Matter	-	PAH_000E0
Polycyclic Organic Matter	-	PAH_176E5
Polycyclic Organic Matter	-	PAH_880E5

Air Toxic	Added for 2014	2014 CMAQ Species Name(s)
Polycyclic Organic Matter	-	PAH_176E4
Polycyclic Organic Matter	-	PAH_176E3
Polycyclic Organic Matter	-	PAH_192E3
Polycyclic Organic Matter	-	PAH_101E2
Polycyclic Organic Matter	-	PAH_176E2
Polycyclic Organic Matter	-	PAH_114E1
Propylene dichloride (1,2-Dichloropropane)	-	PROPDICHLORIDE
Quinoline	-	QUINOLINE
Styrene	YES	STYRENE
Tetrachloroethylene (Perchloroethylene)	-	CL4_ETHE
Toluene	-	TOLU
Trichloroethylene	-	CL3_ETHE
Triethylamine	-	TRIETHYLAMINE
Vinyl chloride	-	CL_ETHE

3.1.2. Dispersion Model Selection

For 2014 NATA air dispersion modeling, we used AERMOD (Cimorelli et al. 2005; EPA 2015e), a steady-state plume model that incorporates air dispersion based on planetary boundary layer turbulence structure and scaling concepts. AERMOD is EPA's preferred near-field modeling system of emissions for distances up to 50 km (EPA 2017). AERMOD version 15181 was used for all NATA rungroups except CMV due to a bug related to polygon sources in version 15181. For CMV sources, AERMOD version 16216 was used to take advantage of the bug fix for polygon sources.

3.2. Meteorological Data

3.2.1. Input WRF Data

For use in all NATA modeling, we derived gridded meteorological data for all of year 2014 for the contiguous United States from version 3.8 of the Weather Research and Forecasting Model (WRF), Advanced Research WRF (ARW) core (Skamarock et al. 2008). The WRF Model is a state-of-the-science mesoscale numerical weather prediction system developed for both operational forecasting and atmospheric research applications (https://www.mmm.ucar.edu/weather-research-and-forecasting-model). The WRF simulation used the same CMAQ map projection, a Lambert Conformal projection centered at coordinates (-97, 40) with true latitudes at 33 and 45 degrees north. The 12-km WRF domain consisted of 396 by 246 grid cells and 35 vertical layers up to 50 millibars.

3.2.2. MCIP Processing for CMAQ

The 2014 WRF meteorological outputs were processed using the Meteorology-Chemistry Interface Processor (MCIP) package (Otte and Pleim 2010), version 4.2.5, to derive the specific inputs to CMAQ: horizontal wind components (i.e., speed and direction), temperature, moisture, vertical diffusion rates and rainfall rates for each grid cell in each vertical layer. Table 3-2 shows the vertical layer structure

used in WRF and the layer-collapsing approach used to generate the CMAQ meteorological inputs. CMAQ resolved the vertical atmosphere with 25 layers, preserving greater resolution in the planetary boundary layer.

Table 3-2. Vertical layer structure for WRF and CMAQ (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
Surface		1.0000	0

3.2.3. MMIF Processing for AERMOD

WRF output was processed through the Mesoscale Model Interface (MMIF) program to create AERMOD-ready meteorological input data. The MMIF produces the surface and profile files used by AERMOD. MMIF was processed for AERMOD in accordance with EPA guidance, Guidance on the Use of the Mesoscale Model Interface Program (MMIF) for AERMOD Applications (U.S. EPA 2016). As options, we used the default FLM layers (see MMIF User's Guide for details), TOP for vertical interpolation and MMIF-calculated mixing heights.

3.2.4. Meteorological Data Outside the Contiguous States

For meteorological data covering areas outside the contiguous states (Alaska, Hawaii, Puerto Rico and the U.S. Virgin Islands), we used National Weather Service observations processed through the AERMET program (version 15181). AERMET is a general purpose meteorological preprocessor that organizes meteorological data into a format compatible with AERMOD. In AERMET, we included the adjusted surface friction velocity option.

3.3. CMAQ Setup

3.3.1. Sources Modeled in CMAQ

The 2014 NATA CMAQ modeling included a 2014 base year run (primary and secondary annual average concentrations) and "zero-out" runs for biogenics and fires (primary annual average concentrations). The 2014 annual simulations included a "ramp-up" period of 10 days to mitigate the effects of initial concentrations. All 365 model days were used in the annual average levels of air toxics modeled.

The 2014 CMAQ model runs were performed for a domain covering the contiguous United States (CONUS), as shown in Figure 3-1. This single domain covers the entire CONUS and large portions of Canada and Mexico using 12-km by 12-km horizontal grid spacing. The model extends vertically from the surface to 50 millibars (approximately 17,600 meters) using a sigma-pressure coordinate system. Air quality conditions at the outer boundary of the 12-km domain were taken from a global model. Table 3-3 provides some basic geographic information regarding the 12-km CMAQ domain.
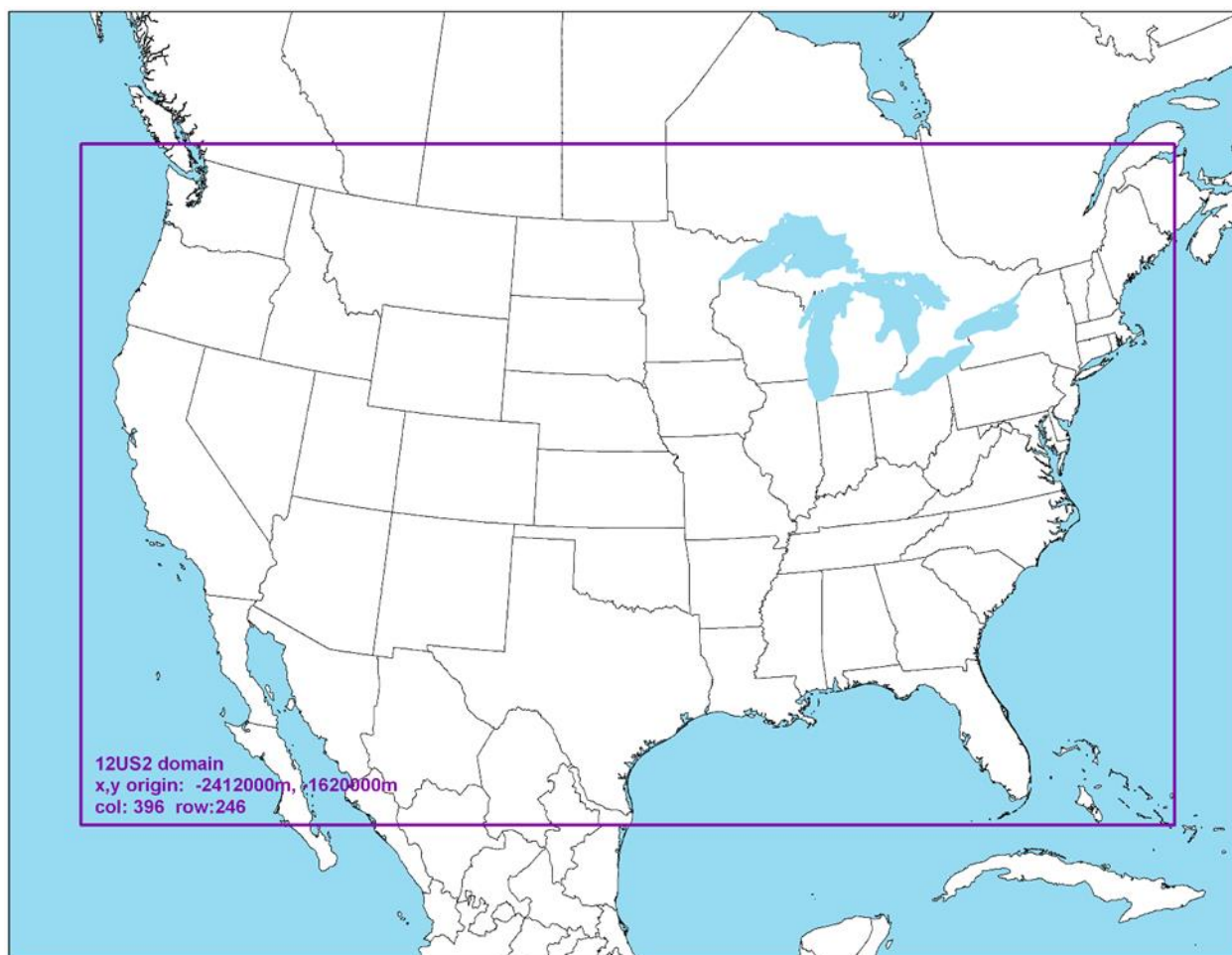
Figure 3-1. Map of the CMAQ modeling domain; the purple box denotes the 12-km national modeling domain

Table 3-3. Geographic information for the 12-km CMAQ modeling domain

12-km CMAQ Modeling Configuration	
Map Projection	Lambert Conformal Projection
Grid Resolution	12 km
Coordinate Center	97 W, 40 N
True Latitudes	33 and 45 N
Dimensions	396 x 246 x 25
Vertical Extent	25 Layers: Surface to 50 mb level (see Table 3-2)

3.3.2. Boundary and Initial Conditions

The lateral boundary and initial species concentrations are provided by a three-dimensional global atmospheric chemistry model, the GEOS-Chem (Bey et al. 2001) model version 11-01, with the SOA chemistry option and updates for lightning NOx inputs from Murray et al. (personal communication 2017-09-28). The global GEOS-Chem model simulates atmospheric chemical and physical processes

driven by assimilated meteorological observations from NASA's Goddard Earth Observing System (GEOS-5). This model was run for 2014 with a grid resolution of 4.0 degrees x 45 degrees (latitude-longitude) and 47 vertical layers. The predictions were processed using pseudonetcdf (Akhtar et al. 2012) using chemical translations comparable to Henderson et al. (2014) at three-hour intervals. Boundary conditions of formaldehyde, acetaldehyde and benzene were explicitly provided from the 2014 GEOS-Chem run. More information is available about the GEOS-Chem model and other applications using this tool at: http://gmao.gsfc.nasa.gov/GEOS/ and http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-5.

Because GEOS-Chem does not include all modeled HAPs, we also used remote concentration estimates as nonvarying background (in space and time). These were computed based on data from the five NOAA GMD sites: Cape Kumukahi, Hawaii (KUM); Mauna Loa, Hawaii (MLO); Niwot Ridge, Colo. (NWR); Barrow, Alaska (BRW); and Alert, Canada (ALT) and the Trinidad Head Site (AGAGE). More information on how these were derived is in Appendix D.

For the remaining CMAQ HAP BCs not provided by GEOS-Chem nor estimated using remote concentrations listed in Table 3-4, a value of zero was applied (shown in Table 3-5) due to a lack of data.

Table 3-4. Boundary conditions from 2014 remote concentration estimates

Pollutant	2014 Remote Concentration at STP (µg/m3)	RCE (pptv) 2014	Remote Network	Location(s)	CMAQ HAP
Chloroform	0.070	0.072	AGAGE	Trinidad Head	x
Methyl chloride (chloromethane)	1.16	0.99	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x
Carbon tetrachloride	0.530	84.4	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x
Methyl bromide (bromomethane)	0.028	7.2	NOAA GMD	KUM, MLO, NWR, BRW, ALT	
Methyl chloroform (1,1,1-trichloroethane)	0.020	3.7	NOAA GMD	KUM, MLO, NWR, BRW, ALT	
Dichloromethane (methylene chloride)	0.20	58.3	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x
Tetrachloroethene (perchloroethylene, tetrachloroethylene)	0.013	1.9	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x

Table 3-5. CMAQ HAP boundary conditions applied as zero value

Air Toxic	2014 CMAQ Species Name(s)
1,3-Butadiene	BUTADIENE13
1,3-Dichloropropene	DICHLOROPROPENE
1,4-Dichlorobenzene(p)	DICHLOROBENZENE
2,4-Toluene diisocyanate	TOL_DIIS
Acetonitrile	ACETONITRILE
Acrolein	ACROLEIN, ACROLEIN_PRIMARY
Acrylic acid	ACRYLICACID

Air Toxic	2014 CMAQ Species Name(s)
Acrylonitrile	ACRYLONITRILE
Arsenic	AASI, AASJ, ASSK
Beryllium	ABEK, ABEI, ABEJ
Cadmium	ACDI, ACDJ, ACDK
Carbon tetrachloride	CARBONTET_NOBC
Carbonyl sulfide	CARBONYLSULFIDE
Chlorine	CL2
Chloroprene	CHLOROPRENE
Chromium Compounds	ACR_VIK, ACR_VIJ, ACR_VII
Chromium Compounds	ACR_IIIK, ACR_IIII, ACR_IIIJ
Diesel PM*	ADE_ECI, ADE_ECJ, ADE_OCI, ADE_OCJ, ADE_SO4J, ADE_NO3J, ADE_OTHRI, ADE_OTHRK, ADE_K
Ethyl benzene	ETHYLBENZ
Ethylene dibromide (Dibromoethane)	BR2_C2_12
Ethylene dichloride (1,2-Dichloroethane)	CL2_C2_12
Ethylene oxide	ETOX
Hexamethylene-1,6-diisocyanate	HEXAMETHY_DIIS
Hexane	HEXANE
Hydrazine	HYDRAZINE
Hydrochloric acid	HCL
Lead Compounds	APBK, APBJ, APBI
Maleic anhydride	MAL_ANHYDRIDE
Manganese Compounds	AMN_HAPSK, AMN_HAPSJ, AMN_HAPSI
Mercury Compounds	HG,HGIIGAS, APHGI, APHGJ (there is no APHGK)
Methanol	MEOH
m-xylene, p-xylene, o-xylene and xylenes (isomers and mixture)	XYLENES
Naphthalene	NAPHTHALENE
Nickel Compounds	ANIK, ANII, ANIJ
Polycyclic Organic Matter	PAH_000E0
Polycyclic Organic Matter	PAH_176E5
Polycyclic Organic Matter	PAH_880E5
Polycyclic Organic Matter	PAH_176E4
Polycyclic Organic Matter	PAH_176E3
Polycyclic Organic Matter	PAH_192E3
Polycyclic Organic Matter	PAH_101E2
Polycyclic Organic Matter	PAH_176E2
Polycyclic Organic Matter	PAH_114E1
Propylene dichloride (1,2-Dichloropropane)	PROPDICHLORIDE
Quinoline	QUINOLINE
Styrene	STYRENE
Toluene	TOLU

Air Toxic	2014 CMAQ Species Name(s)
Trichloroethylene	CL3_ETHE
Triethylamine	TRIETHYLAMINE
Vinyl chloride	CL_ETHE

3.4. AERMOD Setup

3.4.1. Sources Modeled in AERMOD

AERMOD modeling comprised point, nonpoint, on-road and nonroad sources. We excluded fires (agricultural burning, wildfires and prescribed fires) and biogenic emissions.

3.4.2. Receptor Placement

For the CONUS domain, we used the following receptors:

1. Equally spaced "gridded" receptors (1 km in highly populated areas, 4 km otherwise)

2. 2010 populated census-block centroid receptors (discussed in Section 3.4.2.2)

3. 2014 monitoring site receptors (discussed in Section 3.4.2.3)

To facilitate the CMAQ/AERMOD hybrid modeling, we used these receptors to compute an AERMOD average concentration corresponding to each CMAQ grid cell. These concentrations could then be used in the hybrid equation. Gridded receptors were also used for interpolating to block receptors and monitor receptors. Receptors at the monitoring locations were used in the model evaluation.

For non-CONUS areas (AK, HI, PR, VI), where we do not run CMAQ, we used:

1. Census-block centroid receptors (both non-populated and populated)

2. Monitoring site receptors

3.4.2.1. Gridded receptors

We used gridded receptors throughout the CONUS area. The purposes of the gridded receptors were to provide a uniform grid in the CMAQ grid cells to adequately capture near-field concentration gradients of sources in and surrounding the grid cells and to provide a grid used to interpolate census block and monitor receptors post-modeling. We based spacing of the gridded receptors on the 2013 populations of Core Base Statistical Areas (CBSA). For CMAQ grid cells that intersected a CBSA with a population of 1 million people or more (Figure 3-2), the receptors were placed 1 km apart, resulting in 144 receptors per 12-km CMAQ grid cell. Otherwise, the receptors were placed 4 km apart, resulting in nine receptors per 12-km CMAQ grid cell. This resulted in 1.4 million receptors nationwide, with 1-km spacing used in 6,935 12-km cells and 4-km spacing used in 49,490 12-km cells. Correct receptor placement was verified by overlaying receptors with CBSA's in ArcGIS.
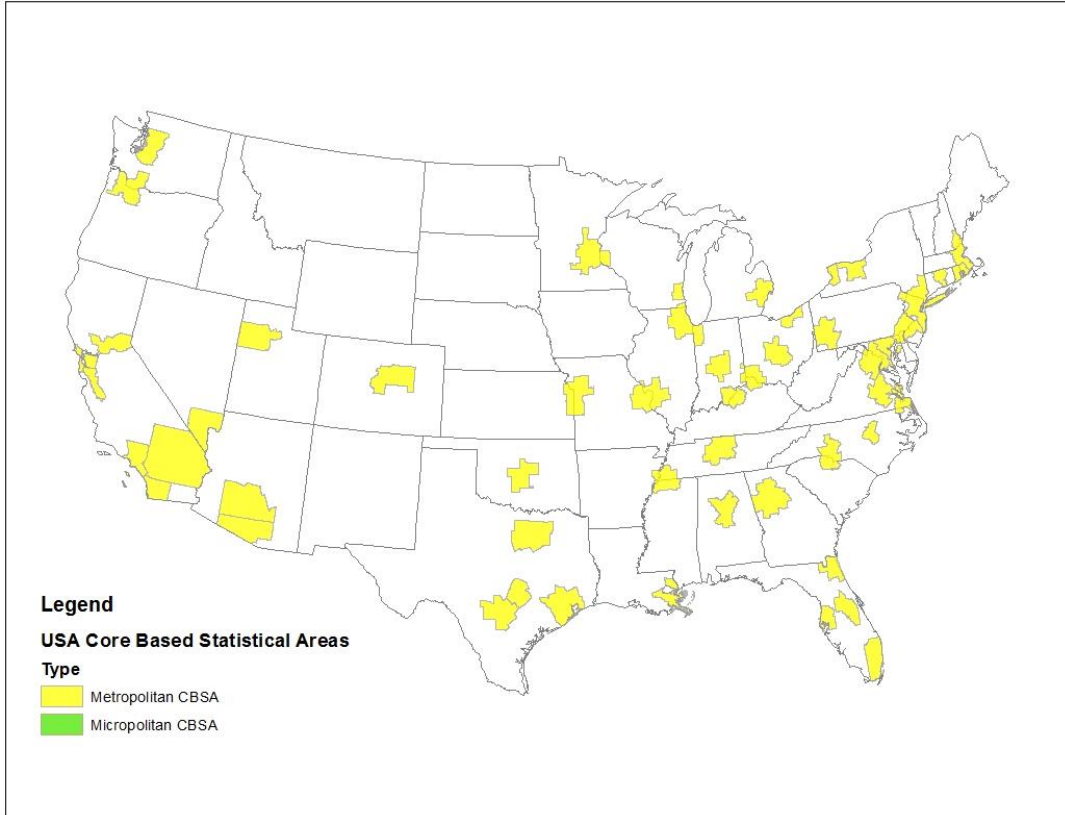
Figure 3-2. CBSAs exceeding 1 million people

Each gridded receptor, with either 1- or 4-km spacing, represented the center of a subgrid cell within the 12-km CMAQ grid cells (Figure 3-3). These gridded receptors, plus populated block and monitor receptors when available within a subgrid cell, were averaged (Figure 3-4). These subgrid-cell averages were then used to calculate the overall AERMOD average of the 12-km grid cell, which was then used in the hybrid equation that combines CMAQ and AERMOD results. Receptor elevations and hill heights were determined using AERMAP (version 11103).
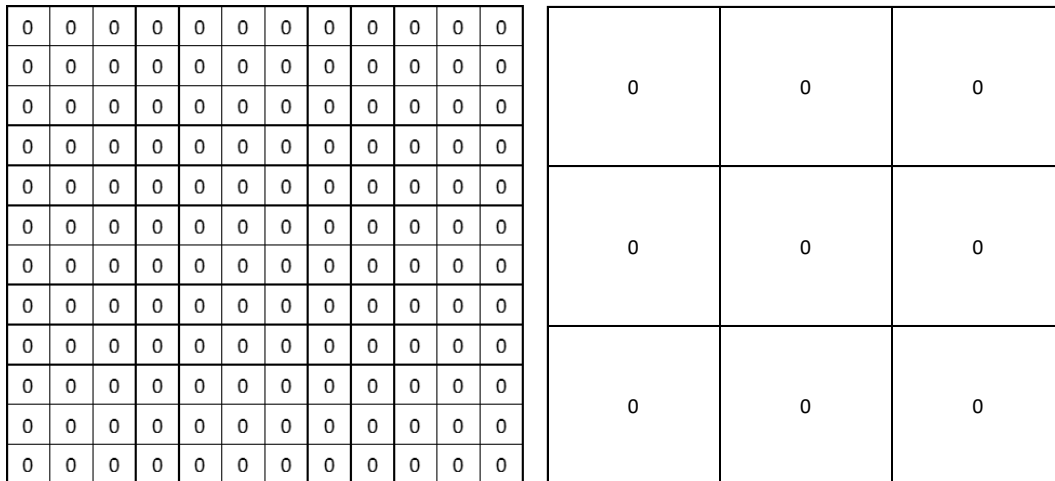
Figure 3-3. Dense (left) and coarse (right) receptor grid layout in CMAQ Lambert Projection
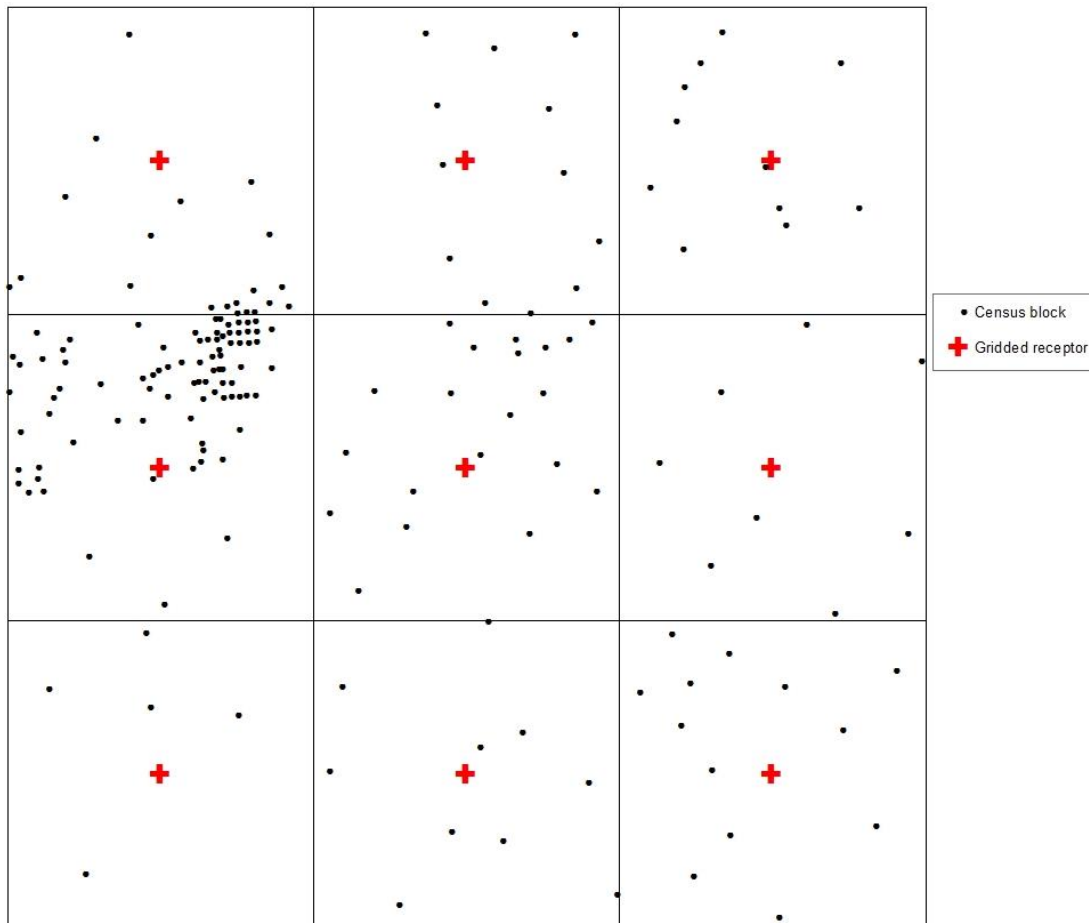
Figure 3-4. Example grid cell with subgrid cells and census blocks

When performing the dispersion modeling for point sources, gridded receptors within 50 km of any emission point at the facility were explicitly modeled in AERMOD. For airports, any gridded receptor within 50 km of any point along a runway or within 50 km of any part of the 100-by-100 m area sources was explicitly modeled. For CONUS gridded sources, any gridded receptor within 50 km of the center of the 12-km or 4-km gridded source was explicitly modeled. For ports and underway sources, any gridded receptor that was within 50 km of a side of the polygon or within 50 km of the center of the source was explicitly modeled. Gridded receptors were not modeled for the non-CONUS grid sources.

3.4.2.2. Census-block centroid receptors

The locations of census-block centroids were based on the 2010 U.S. Census (2010). When performing the dispersion modeling for point sources in the CONUS area, populated block receptors within 10 km of any emission point at the facility were explicitly modeled in AERMOD. For non-CONUS point sources, the distance was 50 km. For airports in the CONUS area, any populated block receptor within 10 km of any point along a runway or 10 km from any point of the 100-by-100 m area sources was modeled. For airports in the non-CONUS areas, the distance was 50 km. For gridded CONUS sources, census blocks were not modeled; these were later interpolated from gridded receptors during model post-processing. For ports and underway sources in the CONUS area, any block receptor within 10 km of a side of the polygon or within 10 km of the center of the source was explicitly modeled. For ports and underway emissions in the non-CONUS areas, the distance was 50 km. For the non-CONUS gridded emissions, we

used the same methods as used for CONUS gridded emissions, except the distance was 50 km instead of 10 km. For all non-CONUS sources, point, airports, etc., non-populated blocks were modeled as receptors as well using the same methodology and distance criteria as used for populated blocks. Non-populated blocks were not modeled in the CONUS area.

3.4.2.3. AERMOD receptors at monitoring sites

The Air Toxics Archive of monitoring data sites was used for the model evaluation (Section 3.7.1). Therefore, we obtained a unique set of geographic coordinates for all monitors to include as receptors in the AERMOD modeling. AERMAP (version 11103) was used to generate receptor elevations and hill heights for input into AERMOD. The modeling distance criteria used for census blocks was used for the monitor receptors.

3.4.3. Model Options

For all AERMOD runs, the FASTALL option was used to decrease model runtimes, especially for gridded, airport, CMV shapes and tract sources. For all AERMOD runs excluding point and airport sources, the FLAT option was used (terrain ignored).

For sources determined to be urban, the AERMOD urban option was used.

3.4.4. AERMOD Simulations

We ran AERMOD for each of the run groups described in Table 2-20. For the run groups that included the gridded source types in the CONUS area, gridded sources were also used as the sources in the non-CONUS areas. Receptor placement was as described in Section 3.4.2, and each source was assigned the meteorological files corresponding to the WRF grid cells in which they were located for the CONUS gridded sources. For all non-CONUS sources, the closest meteorological station to the source was used.

For the point sources and airports, each facility or airport was run in its own AERMOD run. For the point sources, each AERMOD source in the facility was its own AERMOD source group with the source group name corresponding to the AERMOD source ID. For each airport, all emissions sources were assigned to a total group (group ALL). Similarly, for the CMV runs and 12 gridded sources, each CMV shape or 12-km grid cell was run in its own AERMOD run. The CMV emissions sources in each AERMOD run were assigned to groups based on the source. For underway emissions, a single group representing all three ship types was used. For the port emissions, groups were assigned: 1) a group for C1 and C2 and 2) a C3 group. . For the 12-km gridded sources and non-CONUS gridded sources, the source was assigned to source group with the same name as the AERMOD source ID. For the 4-km gridded sources, all sources that shared the same parent 12-km grid cell and thus the same meteorological data were run together in one AERMOD run. Each 4-km source was given its own source group in the AERMOD output.

As discussed in Section 2.3 for the various run groups, different temporalization resolutions were used (hourly, monthly, etc.). The scalars from the temporal helper files were used in conjunction with the unit emission rate of 10,000 tons/year to develop hourly unit emissions for the modeled sources. For hourly emissions, this was simply the product of the hourly scalar and 10,000 tons/year, then applying a

conversion factor of 251.9957[7] to yield an emission rate in g/s. If a source was an hourly source, the resulting emission rate was divided by the area of the source to yield the correct units for AERMOD.

For sources using monthly emission factors, a weighting factor was calculated by multiplying each monthly scalar by the number of days for the month. These products were summed across all 12 months. The hourly emission rate E for each month i was then calculated as:

$$E_i = \frac{10000 \times \left(\frac{EF_i}{Weight}\right)}{24} \times 251.9957$$

Where EF_i is the monthly scalar factor from the temporal helper file and 10,000 is the unit emission rate in tons/year. If the source was an area source, E_i was divided by the area of the source. Since each hour for a particular month has the same emission rate, the AERMOD EMISFACT keyword was used to represent the hourly emissions. The emission rate calculated above was used for the factor, and the base emission rate on the SRCPARAM line was set to 1.0 g/s or 1.0 g/s/m^2 for area sources. For emissions that varied by hour of day only, the AERMOD EMISFACT HROFDY was used and the hourly emission rate E_i for hour i (1–24) was calculated as:

$$E_i = \frac{10000 \times EF_i}{365} \times 251.9957$$

Where 365 represents the number of days per year.

Area sources' emission rates were divided by the area of the source.

For emissions that varied by month, day of the week, MHRDOW or MHRDOW7, and hour of day, the month weighting factor was applied in the temporal helper file. The hourly emission rate applied in AERMOD was calculated as the product of 10,000 tons/year, the emission factor, and the conversion factor of 215.9957. As with the other emission factors, area sources' emission rates were divided by the area of the source.

3.4.5. Post-processing of AERMOD Results

Post-processing of the AERMOD runs consisted of two steps: 1) interpolation of populated census blocks and monitors from gridded receptors in the CONUS domain and 2) calculation of HAP-specific concentrations. For all NATA run groups except the gridded sources, interpolation of concentrations to census blocks and monitors occurred at blocks and monitors from 10 to 50 km from the source in the CONUS domain. For the CONUS gridded source types, census blocks and monitors were interpolated 0 to 50 km from the sources. Interpolation was done for each individual facility (point source, airport, CMV shape) or gridded source. Note that interpolation was not done for the non-CONUS AERMOD runs, as all census blocks and monitors within 50 km of a source were explicitly modeled.

For a receptor located between four gridded receptors, the concentration at the receptor was based on linear interpolation between the four gridded receptors. For receptors near the 50-km edge of the

---

[7]215.9957 is the conversion factor from tons/hr to g/s.

modeling domain and thus without four receptors around it, the interpolated receptor's concentration was the average of the nearest gridded receptors.

For nonmobile NATA run groups (point, airports, CMV, nonpoint, RWC, oil and gas, non-CONUS nonroad) the HAP/source group-specific concentrations at each receptor were calculated by dividing the HAP emissions (tons/year) for each source group at each AERMOD source by 10,000 tons/year.

For the mobile run groups (excluding nonroad in non-CONUS areas), the HAP/source group-specific concentrations were calculated based on monthly emissions. The AERMOD output for these rungroups was average concentrations by month. AERMOD also output the hours that had missing or calm meteorological data. Only hours that were noncalm and not missing in the meteorological data were included in the averaging, consistent with the calculation of long-term concentrations in AERMOD. The following methodology was used to calculate the annual average HAP/source group specific concentrations at each receptor:

1. Multiply the concentration from the monthly output file by the number of valid hours for the particular month based on review of the AERMOD errors file. Add the resulting product to a running total of concentration for each month and modeled source group.

2. Interpolate the results from step 1 to unmodeled census/blocks and monitors within 50 km for each facility or gridded source.

3. For each season, multiply the total calculated in step 1 and 2 by the ratio of the source group's monthly HAP emissions (tons/month) to the modeled monthly emissions (based on 10,000 tons/year) listed in the AERMOD output file.

4. Add the result from step 3 to a running total across all months. Also for each month, loop through the months to determine the total number of valid hours for the year. If all meteorological data hours were noncalm and not missing, this results in a total of 8,760 hours. The number may be less if calms and missing data are present in the meteorological data.

5. Divide running total concentration from step 4 by the running total of hours from step 4 to calculate an annual average HAP/source group-specific concentration. This division of concentration by hours is equivalent to how AERMOD calculates annual average concentrations in a simulation, a sum of hourly concentrations divided by the number of noncalm and nonmissing hours.

Once HAP/source group-specific concentrations are calculated for each gridded receptor, block receptor and monitor receptor, the resulting concentrations were output for input into the hybrid program.

3.5. Hybrid Modeling

The hybrid approach combines the annual concentration results from the AERMOD and CMAQ models to compute ambient concentrations at census block receptors. The 2014 NATA used largely the same approach used for the 2011 NATA. The main differences are the computation of the average AERMOD concentration for each grid cell, the inclusion of more HAPs in CMAQ (and thus the hybrid and the treatment of carbon tetrachloride, a HAP present mainly in background concentrations). The subsections below contain discussions on the hybrid air modeling approach used for the 2014 NATA.

3.5.1. Overview

For 52 of the most prevalent and highest risk air toxics (see Table 3-1), we used a hybrid air quality modeling method combining the fine spatial scale and source attributions of AERMOD (Cimorelli et al. 2005; EPA 2015e) with the full treatment of chemistry and transport afforded by CMAQ. In this application, AERMOD treated all species as chemically nonreactive. The emissions and meteorological data sets used in CMAQ were processed further to generate AERMOD inputs consistent with CMAQ. AERMOD receptor locations were based on the centroids of populated census blocks, monitoring-site positions, and evenly distributed points within each 12-km horizontal CMAQ grid cell in the CONUS (see Figure 3-1 and Figure 3-5), resulting in at least nine, and sometimes more than 10,000, receptors per cell and 6.5 million receptors nationwide.
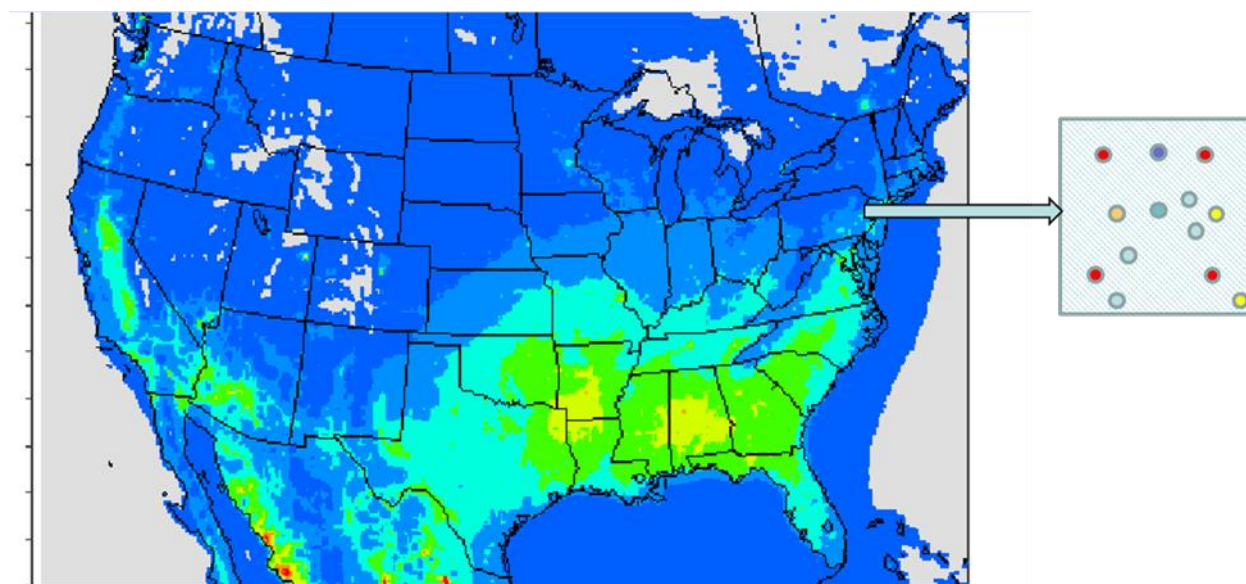
Figure 3-5. CMAQ domain with expanded cell showing hybrid receptors; colors indicate modeled concentrations; dots in inset show locations of receptors within a grid cell

The equation below was used to calculate the 2014 annual average estimates of air concentrations at receptor locations, which were constrained to CMAQ-grid-average values, with AERMOD providing subgrid-scale spatial texture.

$$C = AERMOD_{REC} \times \left(\frac{CMAQ_{PNFB}}{AERMOD_{GRIDAVG}} \right) + CMAQ_{SEC} + CMAQ_{PFIRES} + CMAQ_{PBIOGENICS} + CMAQ_{BACKGROUND}$$

Where:

C	=	concentration at a receptor
CMAQ_PNFB	=	concentration in CMAQ grid cell, contributed by primary emissions, excluding fires and biogenics
AERMOD_REC	=	concentration at AERMOD receptor

$AERMOD_{GRIDAVG}$ = average of all AERMOD results within a CMAQ grid, calculated through averaging the receptors in each of the nine 4-km cells in a CMAQ grid cell, and then taking the average of the nine grid cells

$CMAQ_{SEC}$ = contribution from atmospheric reactions in CMAQ grid cell

$CMAQ_{PFIRES}$ = contribution from primary emissions of fires in CMAQ grid cell

$CMAQ_{PBIOGENICS}$ = contribution from primary emissions of biogenics in CMAQ grid cell

$CMAQ_{BACKGROUND}$ = contribution from background in CMAQ grid cell for carbon tetrachloride

This hybrid approach, which builds on earlier area-specific applications to Philadelphia, Pa. (Isakov et al. 2007) and Detroit, Mich. (Wesson et al. 2010), attempts to characterize nonreactive and reactive species across multiple spatial scales. However, the blending of two different modeling platforms challenges adherence to basic mass-conservation principles. To address this issue, CMAQ tracks primary and secondary contributions by source type, enabling the AERMOD estimate at each receptor location to be normalized to the CMAQ primary contribution. Anchoring concentration averages to CMAQ largely retains mass conservation. The constraint to CMAQ average grid values imposed by the above equation minimized possible redundancies and allows us to combine results from these two very different models.

3.5.2. Treatment of Species

As noted above, we applied the hybrid model to 52 of the highest risk air toxics (shown in Table 3-1) among the 181 air toxics included in the 2014 NATA. CMAQ bases its treatment of atmospheric chemistry on gas-phase reaction processes optimized to characterize ozone, linked with heterogeneous and thermodynamic processes for PM formation. This structure allows us to model explicit chemical species. For example, formaldehyde and acetaldehyde generate significant amounts of peroxy radicals, which lead to enhanced ozone production and secondary PM formation. This demonstrates the multipollutant linkages driven by atmospheric processes that CMAQ can simulate.

Chemical species not incorporated as explicit species in chemical mechanisms are added as nonreactive tracers (e.g., several halogenates) or included in simple reaction schemes, such as 1,3-butadiene decay and subsequent acrolein generation, decoupled from the chemical mechanism. The emissions mass of several less reactive VOCs, such as the prevalent benzene, toluene and xylene species, are tracked as nonreactive tracers. CMAQ treats these as lumped carbon bond species in its reaction calculations, assuming that atmospheric chemistry minimally influences air concentrations. AERMOD, which treats all pollutants as nonreactive, was applied to the remaining air toxics not incorporated within CMAQ.

The calculation of the AERMOD grid cell average, $AERMOD_{GRIDAVG}$ in the hybrid equation is a three-step process:

1. Calculate a total AERMOD concentration at each gridded receptor, block receptor and monitor receptor by reading the individual post-processed run group concentrations and adding to a running total.

2. Calculate the average concentration at each subgrid cell centered on each gridded receptor in each CMAQ grid cell (see Section 3.4.2.1 for details on gridded receptors and example subgrid cells). This results in either 9 or 144 subgrid-cell averages in each CMAQ cell.

3. Average the 9 or 144 averages in each grid cell to calculate an overall average for the grid cell, $AERMOD_{GRIDAVG}$.

After calculating $AERMOD_{GRIDAVG}$ the following steps are taken:

1. For secondary HAPs (acetaldehyde, formaldehyde and acrolein), calculate the secondary concentration by subtracting the CMAQ primary concentration from the total CMAQ concentration at each grid cell. For nonsecondary HAPs, secondary concentrations are zero.

2. For fire and biogenic HAPs, calculate the primary concentration at each grid cell by subtracting the no-fire/no-biogenic CMAQ concentration from the base CMAQ concentration, where base concentration includes all sources. For nonfire and nonbiogenic HAPs, secondary concentrations are zero.

3. At each grid cell, calculate the primary anthropogenic concentrations by subtracting primary fire and primary secondary concentrations from the total primary concentration from the base CMAQ results.

4. Divide the CMAQ primary concentration from step 3 by the $AERMOD_{GRIDAVG}$ and multiply by the AERMOD total concentration at each receptor in the CMAQ grid cell. This is the primary hybrid concentration.

5. To yield the individual source groups' hybrid concentrations (e.g., nonroad construction equipment), multiply each receptor's hybrid primary concentration by the ratio of the receptor's source group concentration to its total AERMOD concentration.

6. For carbon tetrachloride, two CMAQ species were used, one for which boundary conditions and initial conditions were set to zero (CARBONTET_NOBC) and one with non-zero initial and boundary conditions (CARBONTET). Then calculate background concentrations for carbon tetrachloride for each CMAQ grid cell by subtracting CARBONTET_NOBC from CARBONTET. Set total primary concentrations equal to the CARBONTET_NOBC concentrations.

7. Calculate total hybrid concentrations at each receptor by adding the total primary hybrid from step 4 with the secondary, fire and biogenic concentrations from steps 1 and 2.

For CMAQ HAPs in non-CONUS areas, the hybrid program calculated the background, fire, secondary and biogenic concentrations as discussed in Section 3.6 and added them to the total AERMOD concentrations. For non-CMAQ HAPs in all areas, relevant background was added to the total AERMOD concentrations. The hybrid program then output hybrid concentrations for each run group for each HAP. Separate files were created for gridded receptors, block receptors and monitors. The block and monitor files contained all areas, including non-CONUS areas comprising the AERMOD concentrations with relevant background, fire, secondary and biogenic concentrations. The program output the AERMOD results for the non-CMAQ HAPs in a similar format.

Hybrid concentrations for the non-populated blocks in the CONUS area were interpolated from the hybrid concentrations for the gridded receptors using methods similar to those used for the post-AERMOD interpolation of populated blocks (described in Section 3.4.5). For NATA, the hybrid ambient concentrations at the block level were used to estimate exposures. However, they were also summarized at the tract level. After calculating concentrations for the non-populated blocks, tract-average concentrations for the entire United States, Puerto Rico and the Virgin Islands were calculated by area-weighting the block (populated and non-populated) hybrid concentrations in each tract.

3.6. Other Source Characterizations

3.6.1. Background Concentrations Used for non-CMAQ Areas and Pollutants

For non-CMAQ HAPs and non-CMAQ areas, background concentrations were included with the hybrid results. Table 3-6 lists the pollutants for which 2014 background was included in non-CONUS areas. For methyl bromide and methyl chloroform, the background was also included in the CONUS area, as these two HAPs were not CMAQ HAPs.

For background, we used the remote concentration estimates from five NOAA GMD sites: Cape Kumukahi, Hawaii (KUM); Mauna Loa, Hawaii (MLO); Niwot Ridge, Colo. (NWR); Barrow, Alaska (BRW); Alert, Canada (ALT); and the Trinidad Head Site (AGAGE). These are the same values used for CMAQ boundary conditions for pollutants not estimated by GEOS-Chem. Appendix D details how these were developed.

Table 3-6. Background Concentrations used for non-CMAQ areas and pollutants

Pollutant	2014 Remote Concentration at STP (µg/m³)	RCE (pptv) 2014	Remote Network	Location(s)
Chloroform	0.070	0.072	AGAGE	Trinidad Head
Methyl chloride (chloromethane)	1.16	0.99	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Benzene	0.109	0.096	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Carbon tetrachloride	0.530	84.4	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Methyl bromide (bromomethane)	0.028	7.2	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Methyl chloroform (1,1,1-trichloroethane)	0.020	3.7	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Dichloromethane (methylene chloride)	0.20	58.3	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Tetrachloroethene (perchloroethylene, tetrachloroethylene)	0.013	1.9	NOAA GMD	KUM, MLO, NWR, BRW, ALT

For background, we used the remote concentration estimates from five NOAA GMD sites: Cape Kumukahi, Hawaii (KUM); Mauna Loa, Hawaii (MLO); Niwot Ridge, Colo. (NWR); Barrow, Alaska (BRW); Alert, Canada (ALT); and the Trinidad Head Site (AGAGE). These are the same values used for CMAQ boundary conditions for pollutants not estimated by Geos-Chem. Appendix D details how these were developed.

3.6.2. Fires, Biogenics and Secondary Concentrations Used for Non-CMAQ Situations

In the CONUS, fires, biogenics and secondary concentrations are determined by CMAQ. For non-CONUS areas, we developed an approach to estimate these components based on available information, including hemispheric CMAQ concentrations for formaldehyde and acetaldehyde for secondary concentrations.

3.6.3. Non-CONUS Fires

For the HAPs that include fire contributions, an approach for the non-CONUS areas was developed using average CMAQ fire concentrations of states with similar fire emission densities in the CONUS area. For example, North Dakota had a similar emissions density for formaldehyde and acrolein as Alaska. North Dakota's formaldehyde density was 75 and Alaska's 89. For acrolein, North Dakota's formaldehyde density was 11 and Alaska's was 13. For Hawaii (formaldehyde density = 685, acrolein density = 118), the state with the closest densities was Louisiana (formaldehyde density = 557, acrolein density = 99). For Puerto Rico (formaldehyde density = 74, acrolein density = 13), Maryland was chosen as the representative state (formaldehyde density = 79, acrolein density = 13). For the Virgin Islands, fire emissions were assumed zero.

To calculate the fire concentrations for fire HAPs in Alaska, Hawaii and Puerto Rico, the average primary fire concentration was calculated for each of the representative states based on the CMAQ gridded results that overlapped the states. The resulting averages were then applied to the appropriate non-CONUS areas. Table 3-7 lists the fire HAPs with the Alaska, Hawaii and Puerto Rico fire concentrations.

Table 3-7. Fire concentrations (μg/m³) for selected HAPs in Alaska, Hawaii and Puerto Rico

HAP	Alaska	Hawaii	Puerto Rico
Acetaldehyde (primary)	9.3×10^{-3}	1.1×10^{-2}	6.0×10^{-2}
Acetonitrile	2.9×10^{-3}	4.1×10^{-3}	1.1×10^{-2}
Acrolein (primary)	3.8×10^{-3}	5.6×10^{-3}	3.4×10^{-2}
Acrylic acid	2.2×10^{-4}	3.5×10^{-4}	2.1×10^{-3}
PAH_00E0	2.0×10^{-5}	2.9×10^{-5}	1.1×10^{-4}
PAH_176E4	1.5×10^{-5}	2.2×10^{-5}	8.1×10^{-5}
Benzene	9.9×10^{-3}	1.6×10^{-2}	5.3×10^{-2}
PAH_880E5	9.6×10^{-5}	1.4×10^{-4}	5.1×10^{-4}
PAH_176E3	1.5×10^{-5}	2.2×10^{-5}	8.1×10^{-5}
1,3-butadiene	1.1×10^{-3}	1.4×10^{-3}	1.6×10^{-2}
Carbonyl sulfide	4.3×10^{-7}	8.9×10^{-7}	6.7×10^{-6}
PAH_176E5	1.5×10^{-5}	2.2×10^{-5}	8.1×10^{-5}
Formaldehyde (primary)	1.3×10^{-2}	1.8×10^{-2}	1.1×10^{-1}
Hexane	4.6×10^{-4}	5.2×10^{-4}	2.3×10^{-3}
Methanol	3.9×10^{-2}	3.3×10^{-2}	1.3×10^{-1}
Methyl chloride	2.3×10^{-3}	3.4×10^{-3}	1.2×10^{-2}
Naphthalene	2.9×10^{-3}	4.2×10^{-3}	2.4×10^{-2}
Styrene	6.8×10^{-11}	2.3×10^{-6}	4.8×10^{-5}
Toluene	4.3×10^{-3}	6.3×10^{-3}	2.8×10^{-2}
Xylene	1.3×10^{-3}	9.2×10^{-4}	8.0×10^{-3}

3.6.4. Secondary Concentrations

For formaldehyde and acetaldehyde in non-CONUS areas, secondary concentrations were based on results from a 2011 hemispheric CMAQ simulation. This simulation did not include HAPs other than formaldehyde and acetaldehyde, which are explicit model species in the CB05 chemical mechanism. The simulation used a 108-km resolution hemispheric grid that included these areas, and emission inputs for these areas used the 2011 NEI v2 (Eyth et al. 2011). Hemispheric CMAQ results were used to determine a representative secondary concentration for formaldehyde and acetaldehyde for non-CONUS areas. For each area, the secondary concentration was assumed to be 75 percent of the total concentration from the hemispheric CMAQ simulations. The total concentration used for the calculations was based on visual inspections of the modeled concentrations over each area.

Acrolein was processed differently than formaldehyde and acetaldehyde, as hemispheric CMAQ results were not available for acrolein. To determine the secondary concentration for acrolein, the statewide primary acrolein concentrations averages for Alaska and Hawaii (territory averages for Puerto Rico and Virgin Islands) were calculated based on the AERMOD results. Averages were based on modeled census blocks and monitor receptors. The acrolein secondary concentration was assumed to be 25 percent of the primary concentration, as 25 percent is the average ratio of secondary to primary concentrations in the CONUS.

Table 3-8 lists the secondary concentrations for acetaldehyde, acrolein and formaldehyde, and Figure 3-6 and Figure 3-7 show the spatial distribution of the total formaldehyde and acetaldehyde hemispheric CMAQ concentrations. For Alaska and Puerto Rico, the chosen values were based on an average of the range of concentrations across each area. For Alaska, this was the mainland area. The value chosen for Hawaii was based on the areas with populated areas, and the value for the Virgin Islands was the CMAQ grid cell value.

Table 3-8. Concentrations (µg/m³) used to derive secondary concentrations for formaldehyde, acetaldehyde and acrolein; acrolein is primary concentrations only

HAP	Alaska	Hawaii	Puerto Rico	Virgin Islands
Acetaldehyde (hemispheric average)	0.2	0.225	0.45	0.37
Formaldehyde (hemispheric average)	0.1	0.225	0.60	0.30
Acrolein (non-CONUS average)	0.007	0.007	0.01	0.005
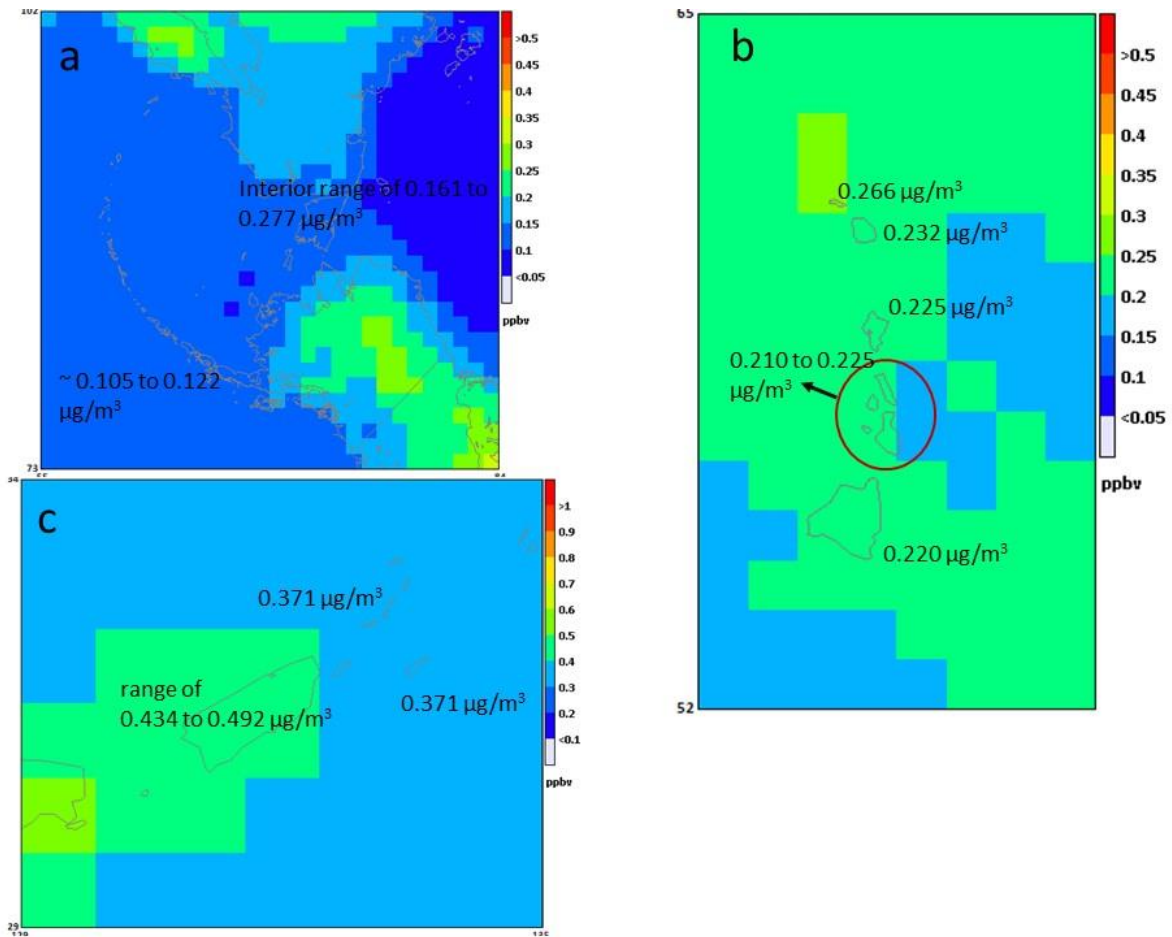
Figure 3-6. Acetaldehyde total concentrations for a) Alaska, b) Hawaii and c) Puerto Rico/Virgin Islands
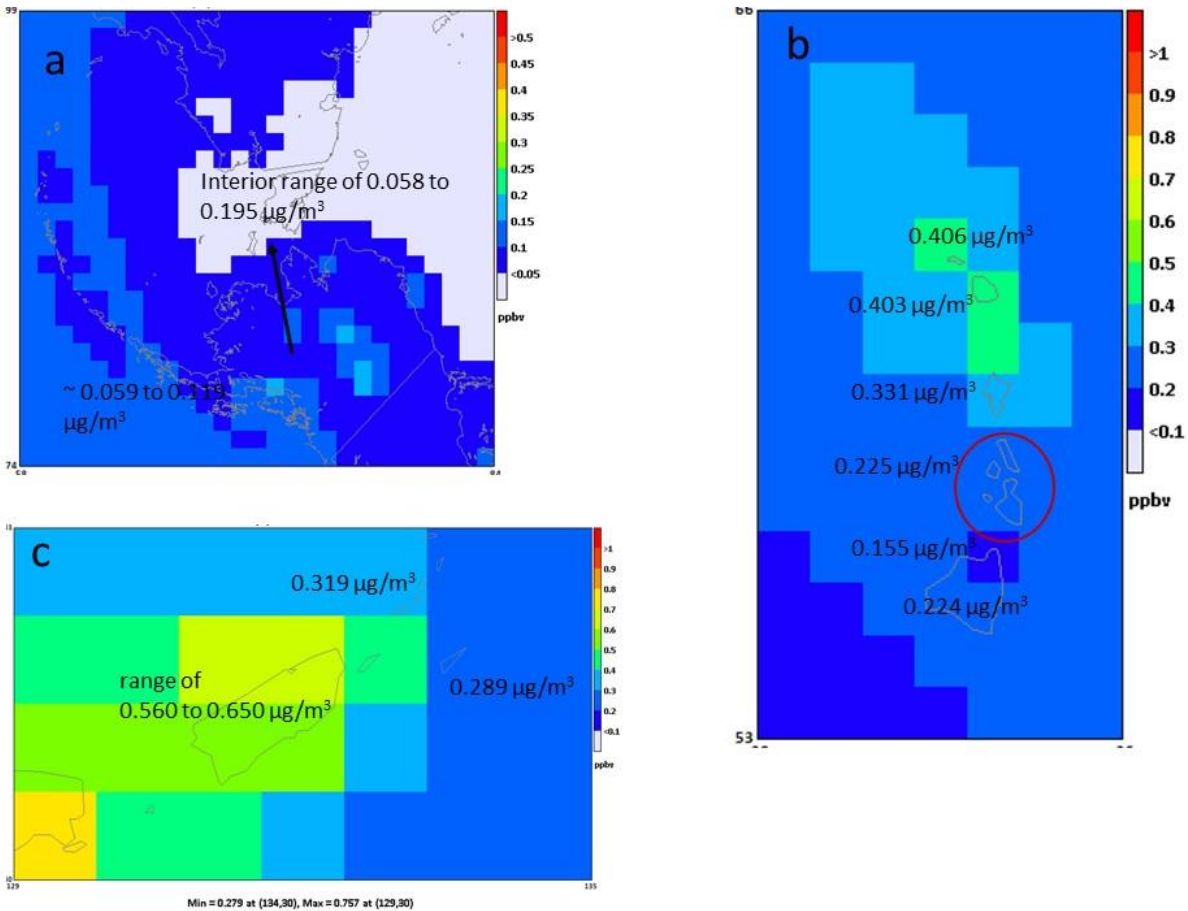
Figure 3-7. Formaldehyde total concentrations for a) Alaska, b) Hawaii and c) Puerto Rico/Virgin Islands

3.6.5. Biogenics

To calculate formaldehyde and acetaldehyde biogenic concentrations for non-CONUS areas, we used the total hemispheric CMAQ concentrations shown in Table 3-6. Biogenic concentrations were assumed to be 20 percent of the total concentrations. For methanol, biogenic concentrations in the non-CONUS areas were assumed to be zero.

3.7. Model Evaluation

An operational model performance evaluation of the HAPs simulated for this 2014 NATA was conducted using the Air Toxics Archive Phase12 for the year 2014 (https://www3.epa.gov/ttnamti1/toxdat.html#data); more details found in Section 3.7.1 below). The model evaluation included both the hybrid air toxics and non-hybrid air toxics. The hybrid evaluation compared the HAPs for which there are valid ambient data (i.e., completeness criteria protocol) to compare against CMAQ, AERMOD and the hybrid model predictions. Likewise, the HAP non-hybrid evaluation used similar observational completeness criteria constraints to compare against HAPs estimated by adding AERMOD to remote ambient concentrations (where available) that are assumed to reflect background conditions.

Note that when pairing observed to model data, there are spatial-scale differences between CMAQ, AERMOD and the hybrid model predictions. A CMAQ concentration represents a 12-km grid-cell volume-averaged value. The AERMOD model concentration represents a specific point within the modeled domain. The hybrid model concentration combines the AERMOD point concentration gradients with the 12-km CMAQ grid-cell volume average. The ambient observed measurements are made at specific spatial locations (latitude/longitude). Several annual graphical presentations and statistics of model performance were calculated and prepared. Graphical presentations include:

1. Box and whisker plots that show the distribution and the bias of the predicted and observed data, and

2. Regional maps that show the mean bias and error calculated at individual monitoring sites.

3.7.1. Ambient Monitoring Data Preparation

EPA has created annual average concentrations for year 2014 using data in the Air Toxics Archive Phase12 (https://www3.epa.gov/ttnamti1/toxdat.html#data). These data primarily come from AQS; however, they may also come directly from special studies that may not have been included in AQS. In addition, these data have been converted to micrograms per cubic meter ($\mu g/m^3$) from the unit they were originally reported in (typically parts per billion) using local meteorological data where available or standard conditions otherwise. To create annual averages, EPA used the following procedures:

1. Separated out (did not mix) data with different duration description at a monitoring site (e.g., hourly, 3-hour, 5-hour, 24-hour are not averaged together)

2. Averaged subdaily data (hourly) to create daily averages

3. Required 70 percent of expected days to be present (a nondetect ["ND"] is counted as present) for each quarter

4. Required 3 of 4 quarters to be 70 percent complete

5. Used the regression on order statistics (ROS) for creating annual averages where less than 80 percent of the data were ND; removed site/pollutants where greater than 80 percent of the data were ND

6. Removed site/pollutant data where the difference between the ROS-computed mean and the zero-substituted mean was more than 30 percent

7. The value we used to compare to the modeled value is the median for the year 2014. We computed site/pollutant medians using ROS so data that were ND would be treated as censored values rather than zero. The ROS requires the detection limit to be used as the censor value. Because the instrument detection limit was not available, we chose the minimum measured value for the year or the method detection limit (MDL) as the censor value. The MDL is not the instrument detection limit, but rather is defined as the minimum concentration of a substance that can be measured and reported with 99 percent confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte. MDLs are determined through statistical procedures. For this the model evaluation, we used data below the MDL in the computation of the annual statistics if the data were not ND (i.e., equal to zero).

8. For some monitors, we summed annual medians because NATA models the sum of the monitored components as the NATA modeled pollutant and not the individual components. This occurred, for example, for PAH groups (summed individual PAHs belonging to the PAH groups), xylenes (summed m/p with o-xylene) and 1,3-dichloropropylene (summed cis and trans).

9. For some monitors, the same NATA pollutant had multiple 2014 median values to use for the evaluation (for example, if there were 1-hour and 24-hour sampling durations and both met the data completeness criteria). These multiple 2014 median observation data are noted in Tables 3-6 and 3-8 as compared to the model HAPs (e.g., modeled 1,3-butadiene is evaluated against median observations for BUTADIENE13_1_HOUR and BUTADIENE13_24_HOUR).

The data used in the model evaluation are provided in the [Supplemental Data](#) folder.

3.7.2. Model Performance Statistics

The Atmospheric Model Evaluation Tool (AMET) was used to conduct the 2014 NATA HAP evaluation (Appel et al., 2011). There are various statistical metrics used by the science community for model performance evaluation. For a robust evaluation, the principal evaluation statistics used to evaluate model performance are based on the following metrics: two-bias metrics (mean bias and normalized mean bias); three-error metrics (mean error, normalized mean error and root mean square error) and correlation coefficient.

Common Variables:

M = predicted concentration

O = observed concentration

X = predicted or observed concentration

σ = standard deviation

I. Mean Bias, Mean Error and Root Mean Square Error (µg/m³)

$$Mean\ Bias = \frac{1}{n}\sum_{1}^{n}(M - O)$$

$$Mean\ Error = \frac{1}{n}\sum_{1}^{n}|M - O|$$

$$Root\ Mean\ Square\ Error = \sqrt{\frac{\sum_{1}^{n}(M - O)^2}{n}}$$

Mean Bias (MB) quantifies the tendency of the model to overestimate or underestimate values, while Mean Error (ME) and Root Mean Square Error (RMSE) measure the magnitude of the difference

between modeled and observe values regardless of whether the modeled values are higher or lower than observations.

II. Normalized Mean Bias and Error (unitless)

$$Normalized\ Mean\ Bias = \frac{\sum_{1}^{n}(M-O)}{\sum_{1}^{n}(O)}$$

Normalized mean bias (NMB) is used as a normalization to facilitate a range of concentration magnitudes. This statistic averages the difference (modeled minus 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.

$$Normalized\ Mean\ Error = \frac{\sum_{1}^{n}|M-O|}{\sum_{1}^{n}(O)}$$

Normalized mean error (NME) is 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 minus observed) over the sum of observed values.

III. Correlation Coefficient (unitless)

$$Correlation = \frac{1}{(n-1)}\sum_{1}^{n}\left(\left(\frac{O-\overline{O}}{\sigma_o}\right)*\left(\frac{M-\overline{M}}{\sigma_m}\right)\right)$$

Correlation coefficient (r) provides an indication of the strength of linear relationship and is signed positive or negative based on the slope of the linear regression.

3.7.3. Hybrid Evaluation

An annual operational model performance evaluation for HAPs used in the hybrid model calculation was conducted to estimate the ability of the hybrid model as well as to compare to the predictions from the CMAQ and AERMOD modeling systems to replicate the 2014 HAP observed ambient concentrations. Inclusion of all three model results is intended to demonstrate the merged attributes of the hybrid model used for this 2014 NATA. Statistical assessments of each model versus observed pairs were paired in time and space and aggregated on an annual basis. Table 3-9 provides a list of HAPs evaluated in the hybrid model performance evaluation and the number of observed monitoring sites (based on completeness criteria of observations, Section 3.7.2). Figure 3-8 shows the 2014 HAP monitoring locations.

Table 3-9. List of hybrid HAPs evaluated

Model Air Toxic	Measured Air Toxic	No. of Sites
Acetonitrile	ACETONITRILE_24_HOURS	51
Acrolein	ACROLEIN_24_HOURS	30
Acrylonitrile	ACRYLONITRILE_24_HOURS	18
Acetaldehyde	ALD2_24_HOURS	111
Arsenic	ARSENIC_PM10_24_HOURS	39
	ARSENIC_PM25_24_HOURS	213
	ARSENIC_TSP_24_HOURS	44
Benzene	BENZENE_1_HOUR	15
	BENZENE_24_HOURS	211
	BENZENE_5_MINUTES	12
Beryllium	BERYLLIUM_PM10_24_HOURS	32
	BERYLLIUM_TSP_24_HOURS	43
Ethylene dibromide	BR2_C2_12_24_HOURS	9
1,3-Butadiene	BUTADIENE13_1_HOUR	14
	BUTADIENE13_24_HOURS	154
Cadmium	CADMIUM_PM10_24_HOURS	38
	CADMIUM_PM25_24_HOURS	99
	CADMIUM_TSP_24_HOURS	53
Carbon tetrachloride	CARBONTET_24_HOURS	184
	CARBONTET_5_MINUTES	11
Chloroform	CHCL3_24_HOURS	170
Vinyl chloride	CL_ETHE_24_HOURS	48
Chlorine	CL2_24_HOURS	324
Ethylene dichloride	CL2_C2_12_24_HOURS	109
Methylene chloride	CL2_ME_24_HOURS	184
	CL2_ME_5_MINUTES	12
Trichloroethylene	CL3_ETHE_24_HOURS	53
1,1,2,2-Tetrachloroethane	CL4_ETHANE1122_24_HOURS	10
Tetrachloroethylene	CL4_ETHE_24_HOURS	134
	CL4_ETHE_5_MINUTES	12
Chromium Compounds	CR_VI_PM10_24_HOURS	4
1,4-Dichlorobenzene(p)	DICHLOROBENZENE_24_HOURS	49
1,3-Dichloropropene	DICHLOROPROPENE_24_HOURS	26
Ethyl benzene	ETHYLBENZENE_1_HOUR	15
	ETHYLBENZENE_24_HOURS	184
Formaldehyde	FORM_24_HOURS	111
Hexane	HEXANE_1_HOUR	15
	HEXANE_24_HOURS	130
Lead Compounds	LEAD_NON_FEM_TSP_24_HOURS	161
	LEAD_PM10_24_HOURS	56

Model Air Toxic	Measured Air Toxic	No. of Sites
	LEAD_PM25_24_HOURS	320
	LEAD_TSP_24_HOURS	29
Manganese Compounds	MANGANESE_PM10_24_HOURS	43
	MANGANESE_PM25_24_HOURS	322
	MANGANESE_TSP_24_HOURS	73
Methyl chloride	METHCHLORIDE_24_HOURS	161
	METHCHLORIDE_5_MINUTES	12
Naphthalene	NAPHTHALENE_24_HOURS	35
Nickel Compounds	NICKEL_PM10_24_HOURS	39
	NICKEL_PM25_24_HOURS	266
	NICKEL_TSP_24_HOURS	59
Polycyclic Organic Matter	PAH_000E0_NA	31
Polycyclic Organic Matter	PAH_176E3_NA	20
Polycyclic Organic Matter	PAH_176E4_NA	23
Polycyclic Organic Matter	PAH_176E5_NA	27
Polycyclic Organic Matter	PAH_880E5_NA	28
Propylene dichloride	PROPDICHLORIDE_24_HOURS	22
Styrene	STYRENE_1_HOUR	15
	STYRENE_24_HOURS	132
Toluene	TOLU_1_HOUR	15
	TOLU_24_HOURS	215
Xylene	XYLENE_24_HOURS	194
	XYLENE_1_HOUR	7
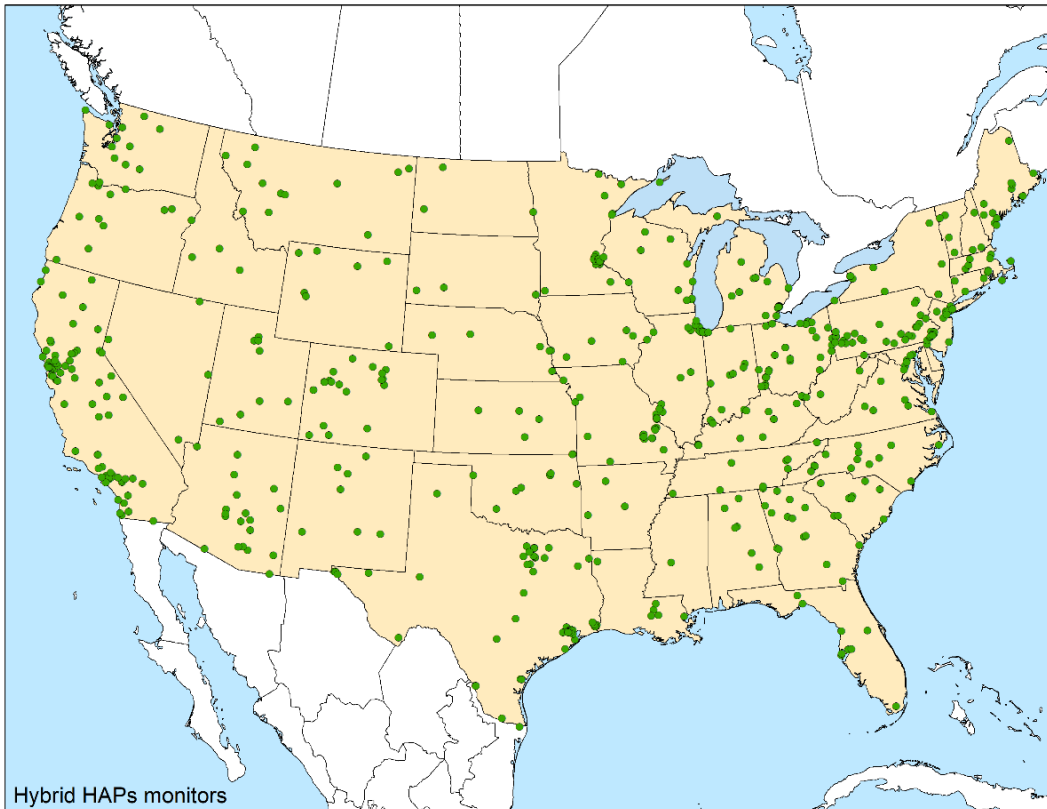
Figure 3-8. 2014 monitoring locations for the hybrid HAPs evaluation

In this section, we present paired annual model-to-monitor site comparisons for the hybrid model, along with CMAQ and AERMOD, for three key HAPs: acetaldehyde, formaldehyde and benzene. The annual model performance statistical results for these three HAPs are presented below in Table 3-10. Boxplots showing model distribution (units of μg/m^3) and bias differences (units of μg/m^3) as compared to ambient observations are presented in this statistical analysis. These boxplots display boxed interquartile ranges of 25th to 75th, along with whiskers from the 5th to 95th, quartiles. Also plotted on these boxplots are summary statistics of correlation (r), RMSE, NMB, NME, MB and ME. Regional spatial maps that show the mean bias and error calculated at individual monitoring sites are also provided for acetaldehyde, formaldehyde and benzene. Appendix E presents more details of the hybrid evaluation.

Both CMAQ and hybrid model predictions of annual formaldehyde, acetaldehyde and benzene showed relatively small to moderate bias and error percentages when compared to observations. AERMOD showed larger biases and errors; these underestimates are expected for secondarily formed HAPs (e.g., -84.8% for acetaldehyde and -83.9% for formaldehyde) given the exclusion of atmospheric chemistry in AERMOD. Differences in bias and error statistics between the hybrid and CMAQ models were negligible for formaldehyde and acetaldehyde. Technical issues in the HAPs data consist of (1) uncertainties in monitoring methods; (2) limited measurements in time/space to characterize ambient concentrations ("local in nature"); (3) commensurability issues between measurements and model predictions; (4) emissions and science uncertainty issues may also affect model performance and (5) limited data for estimating intercontinental transport that effects the estimation of boundary conditions (i.e., boundary estimates for some species are much higher than predicted values inside the domain).

Table 3-10. 2014 annual air toxics performance statistics for the Hybrid, CMAQ and AERMOD models

Air Toxic Species	Model	MB (µg/m3)	ME (µg/m3)	NMB (%)	NME (%)
Acetaldehyde (ALD2_24_HOURS)	Hybrid	-0.3	0.5	-21.0	35.2
	CMAQ	-0.3	0.5	-21.8	35.5
	AERMOD	-1.2	1.2	-84.8	84.8
Formaldehyde (FORM_24_HOURS)	Hybrid	-0.9	1.0	-40.6	44.7
	CMAQ	-0.9	1.0	-41.5	45.2
	AERMOD	-1.9	1.9	-83.9	85.2
Benzene (BENZENE_24_HOURS)	Hybrid	-0.1	0.2	-14.5	38.8
	CMAQ	-0.2	0.2	-26.9	39.1
	AERMOD	-0.2	0.3	-38.3	48.6

(a)

(b)
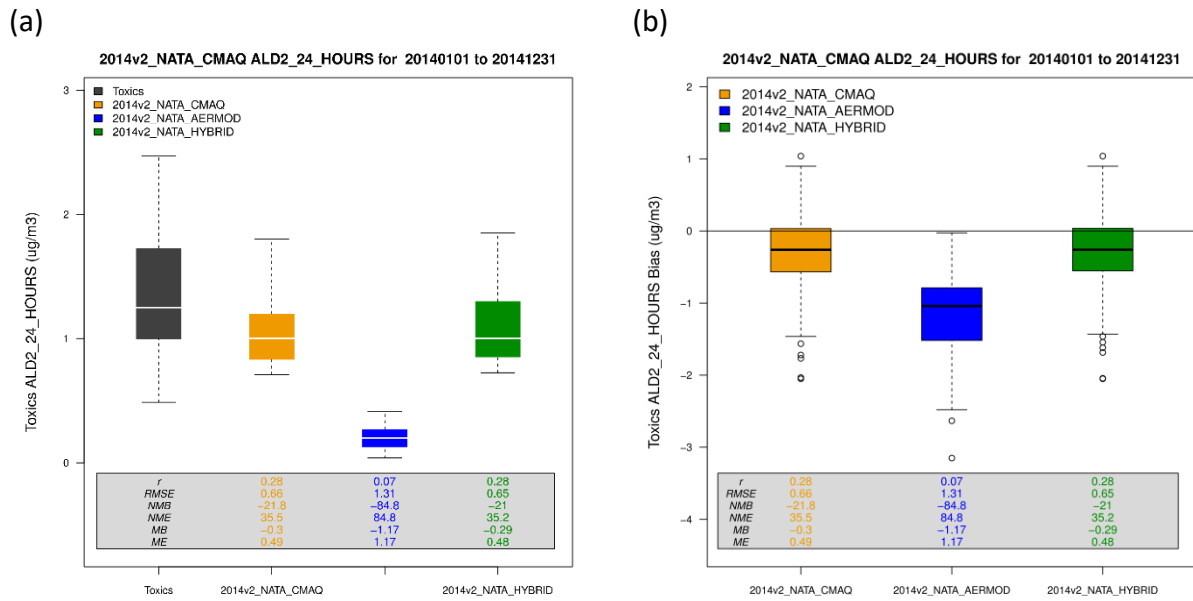
Figure 3-9. Acetaldehyde boxplots of (a) distribution (µg/m³) and (b) bias difference (µg/m³) for CMAQ, AERMOD and Hybrid models compared to ambient observations
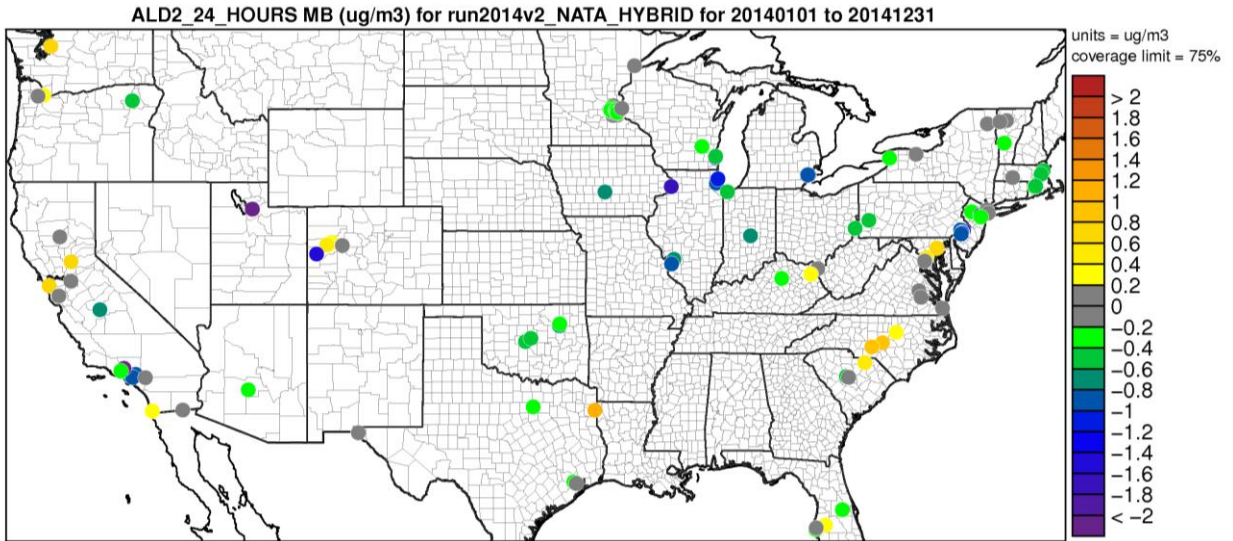
Figure 3-10. Mean bias (%) for acetaldehyde at 2014 monitoring sites in the Hybrid modeling domain
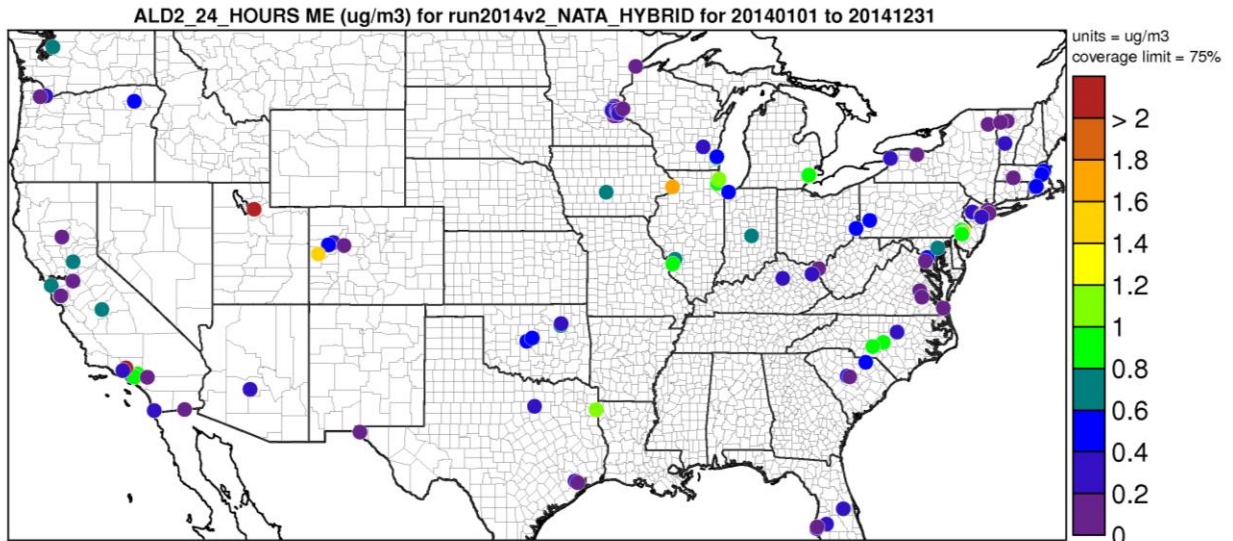
Figure 3-11. Mean error (%) for acetaldehyde at 2014 monitoring sites in the Hybrid modeling domain
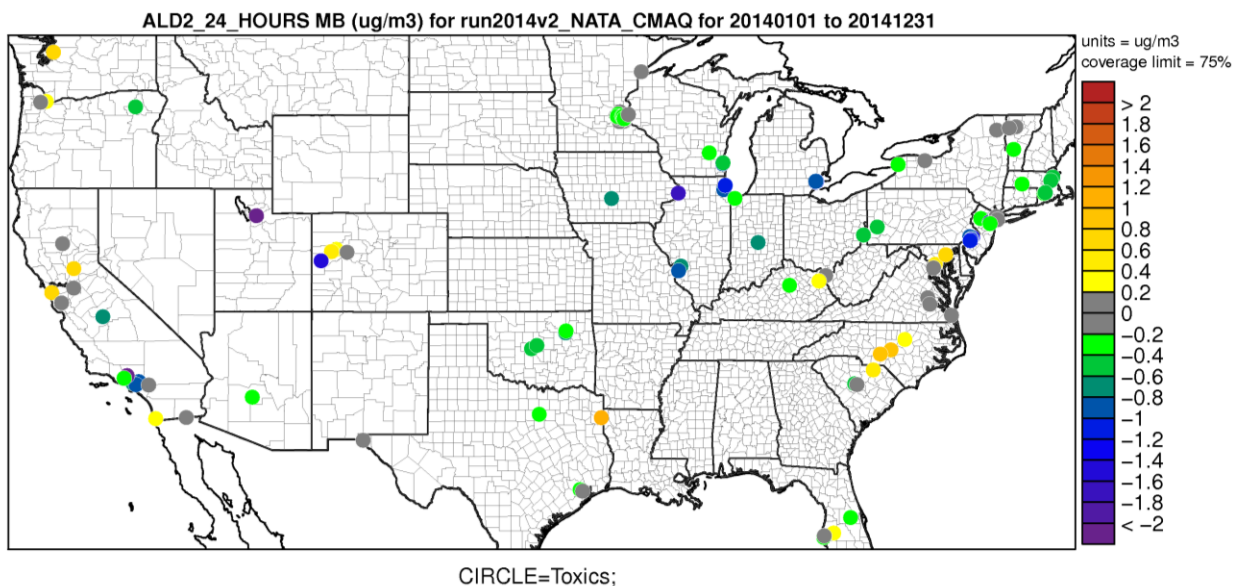
Figure 3-12. Mean bias (%) for acetaldehyde at 2014 monitoring sites in the CMAQ modeling domain
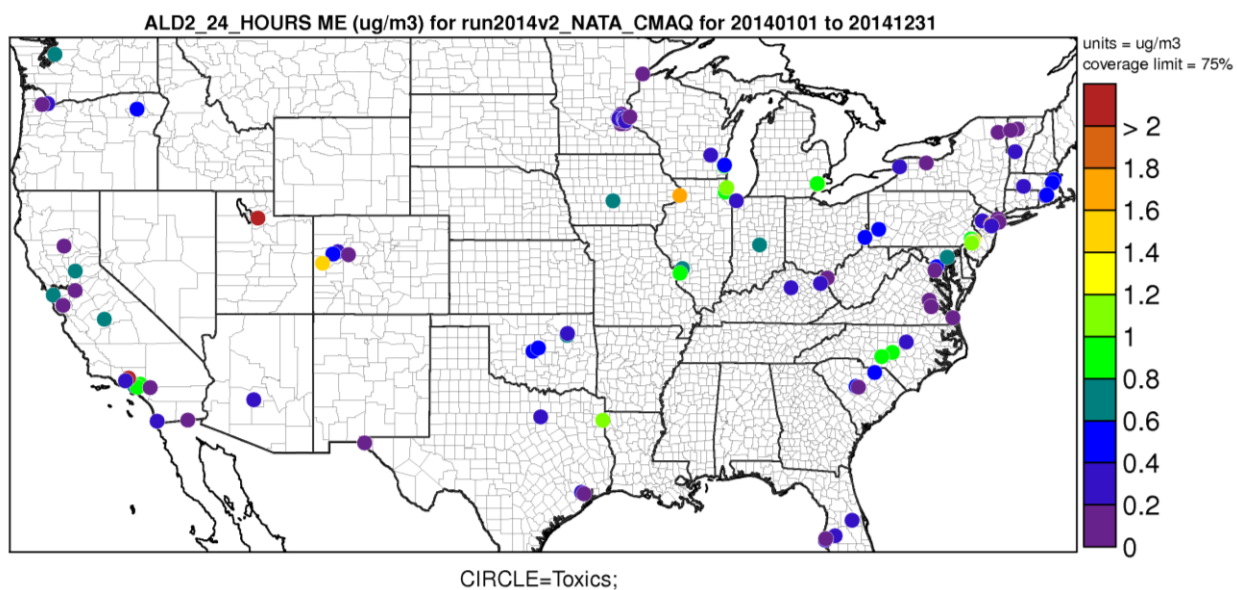
Figure 3-13. Mean error (%) for acetaldehyde at 2014 monitoring sites in the CMAQ modeling domain
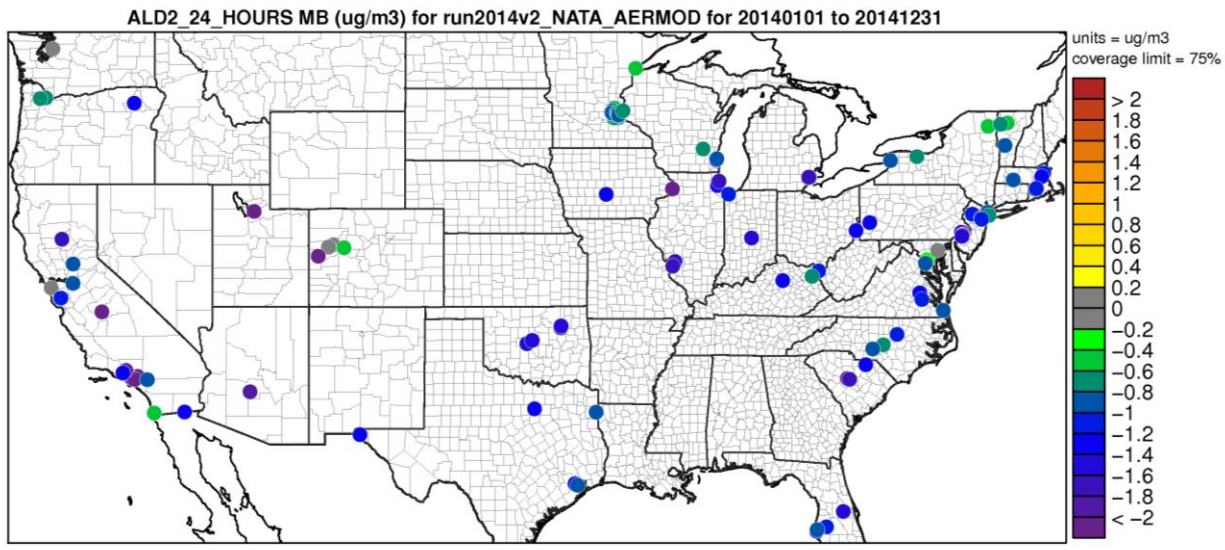
Figure 3-14. Mean bias (%) for acetaldehyde at 2014 monitoring sites in the AERMOD modeling domain
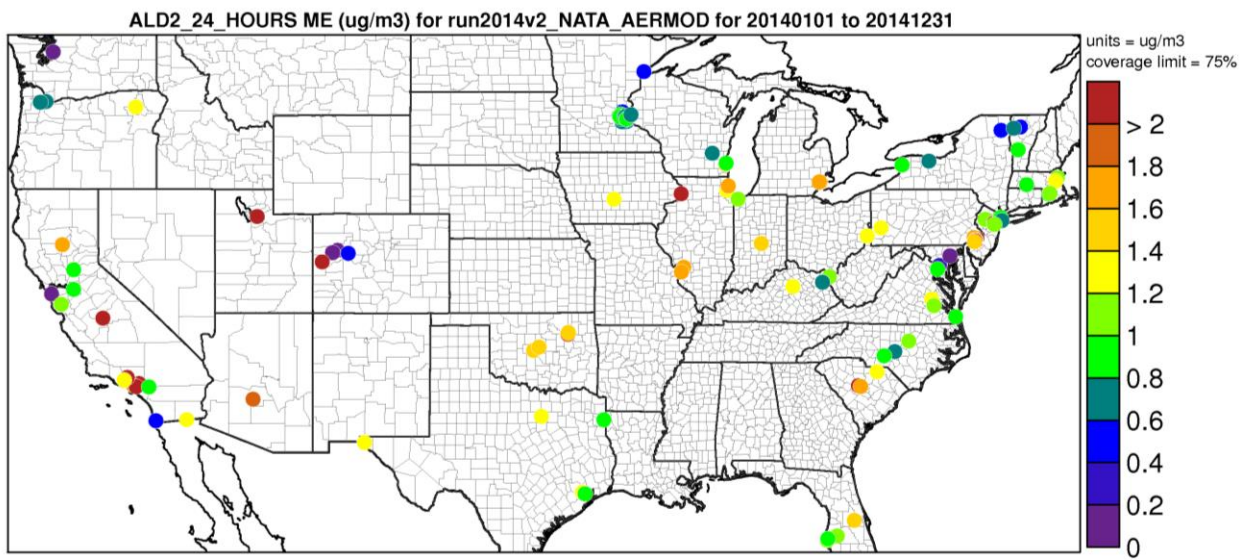
Figure 3-15. Mean error (%) for acetaldehyde at 2014 monitoring sites in the AERMOD modeling domain
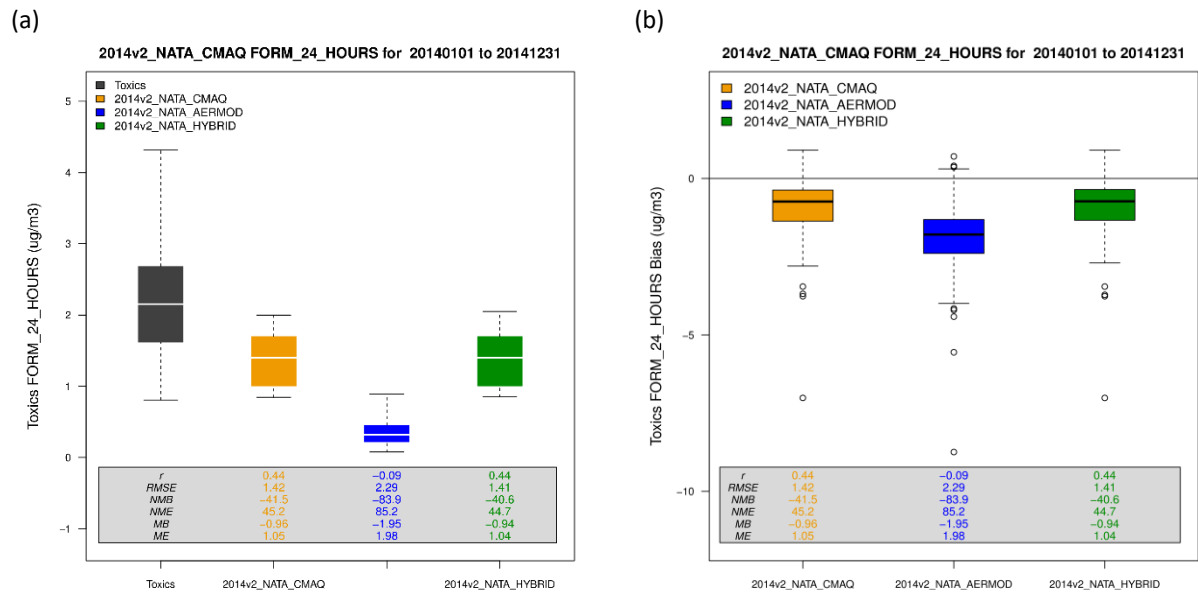
Figure 3-16. Formaldehyde boxplots of (a) distribution (μg/m³) and (b) bias difference (μg/m³) for CMAQ, AERMOD and Hybrid models compared to ambient observations
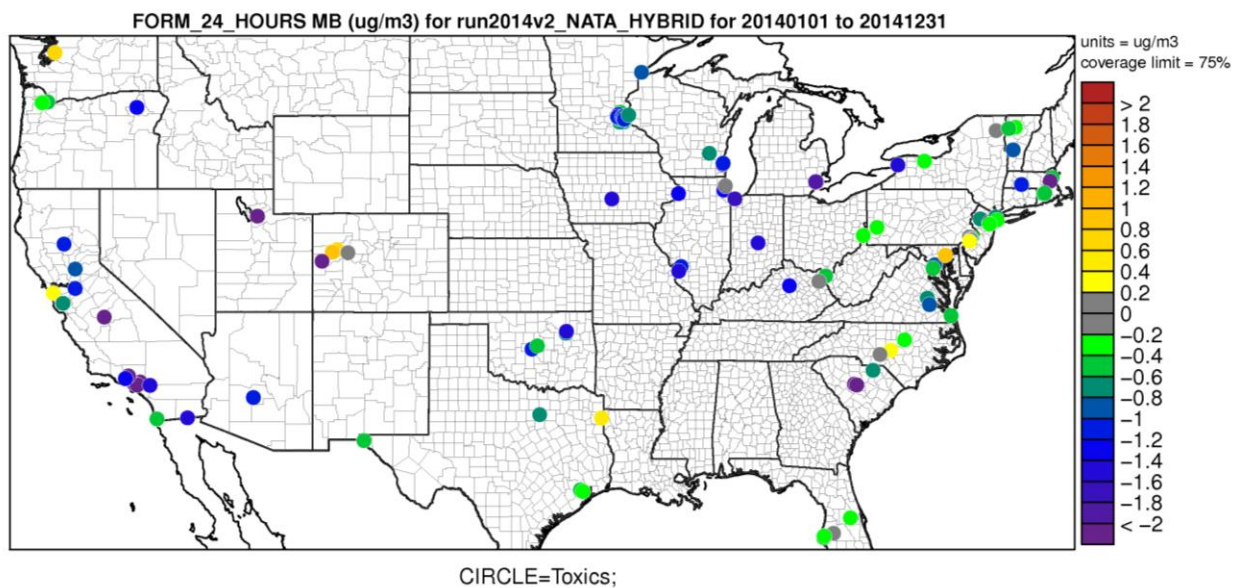
Figure 3-17. Mean bias (%) for formaldehyde at 2014 monitoring sites in the Hybrid modeling domain
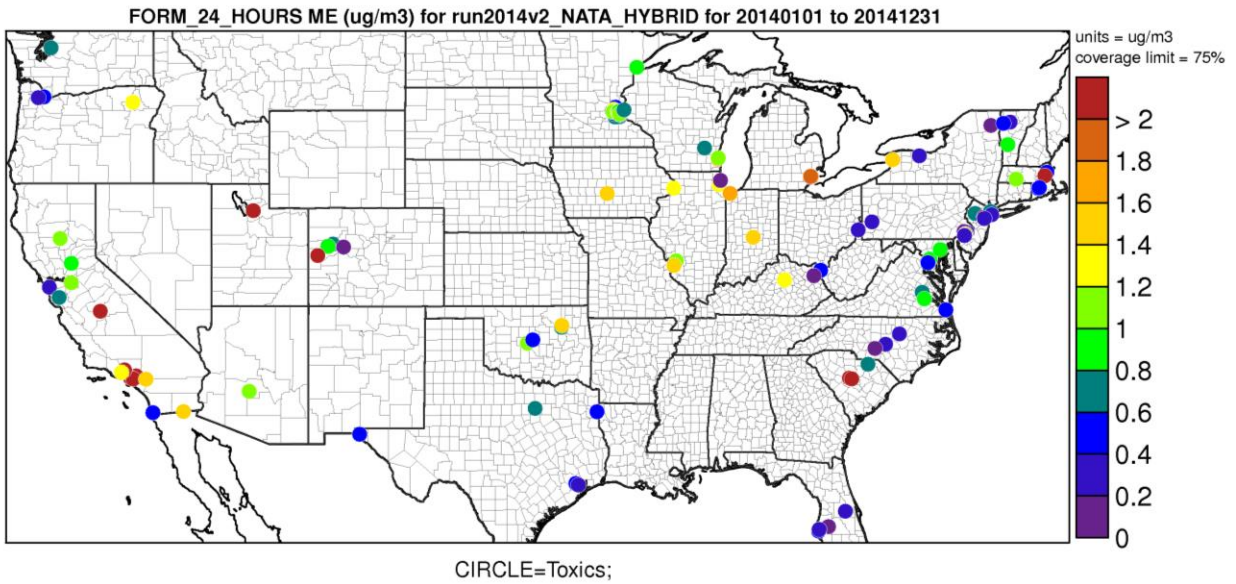
Figure 3-18. Mean error (%) for formaldehyde at 2014 monitoring sites in the Hybrid modeling domain
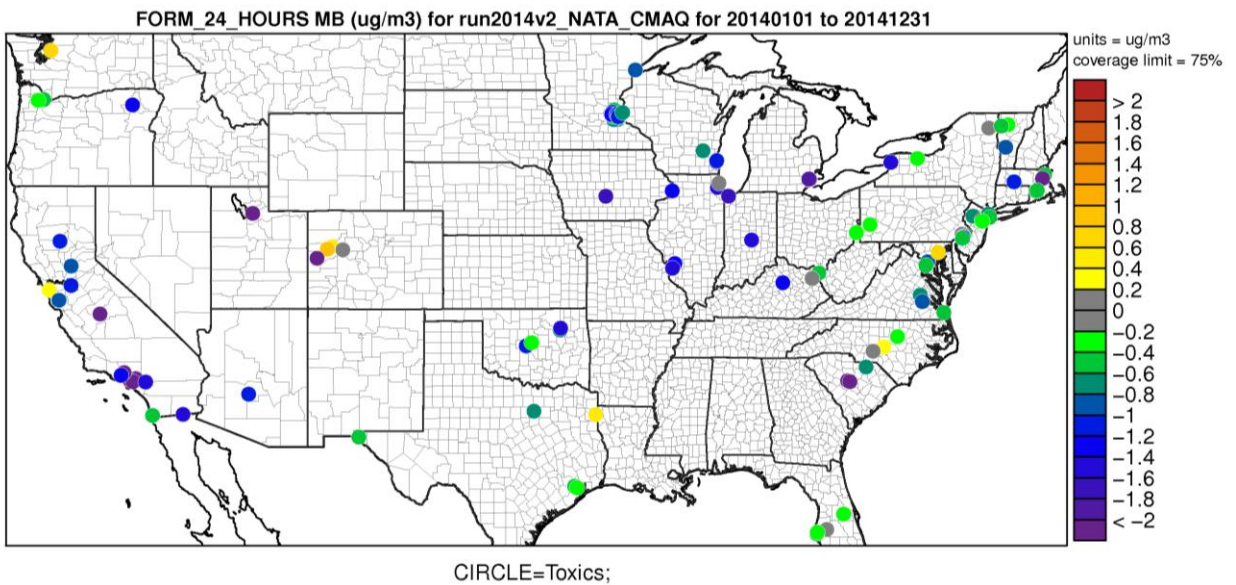
Figure 3-19. Mean bias (%) for formaldehyde at 2014 monitoring sites in the CMAQ modeling domain

FORM_24_HOURS ME (ug/m3) for run2014v2_NATA_CMAQ for 20140101 to 20141231
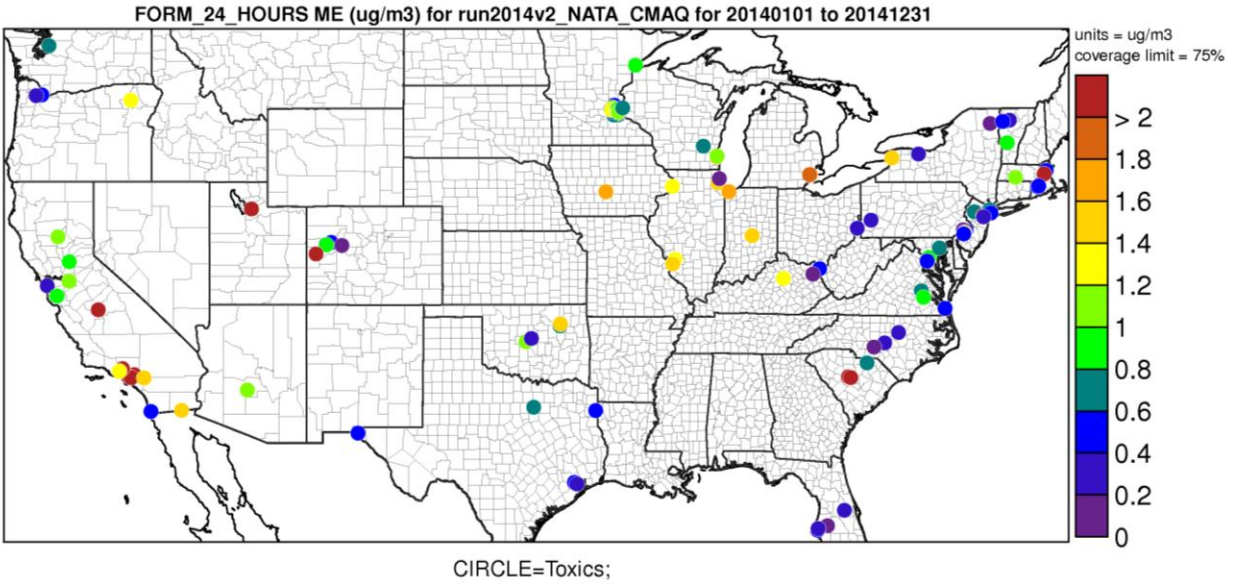
Figure 3-20. Mean error (%) for formaldehyde at 2014 monitoring sites in the CMAQ modeling domain

FORM_24_HOURS MB (ug/m3) for run2014v2_NATA_AERMOD for 20140101 to 20141231
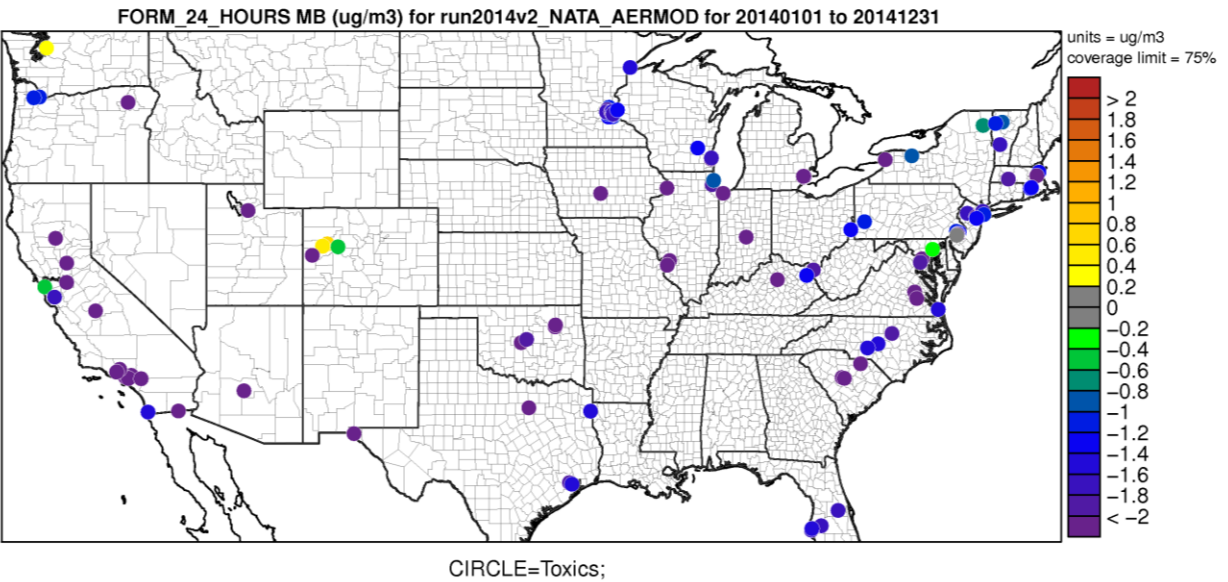
Figure 3-21. Mean bias (%) for formaldehyde at 2014 monitoring sites in the AERMOD modeling domain
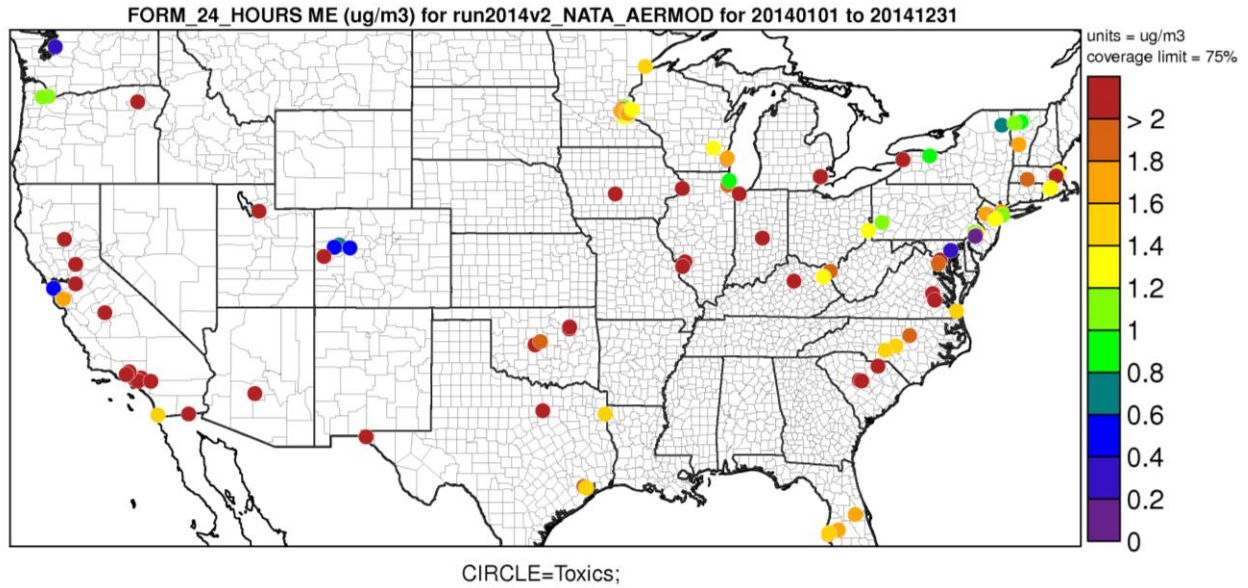
Figure 3-22. Mean error (%) for formaldehyde at 2014 monitoring sites in the AERMOD modeling domain
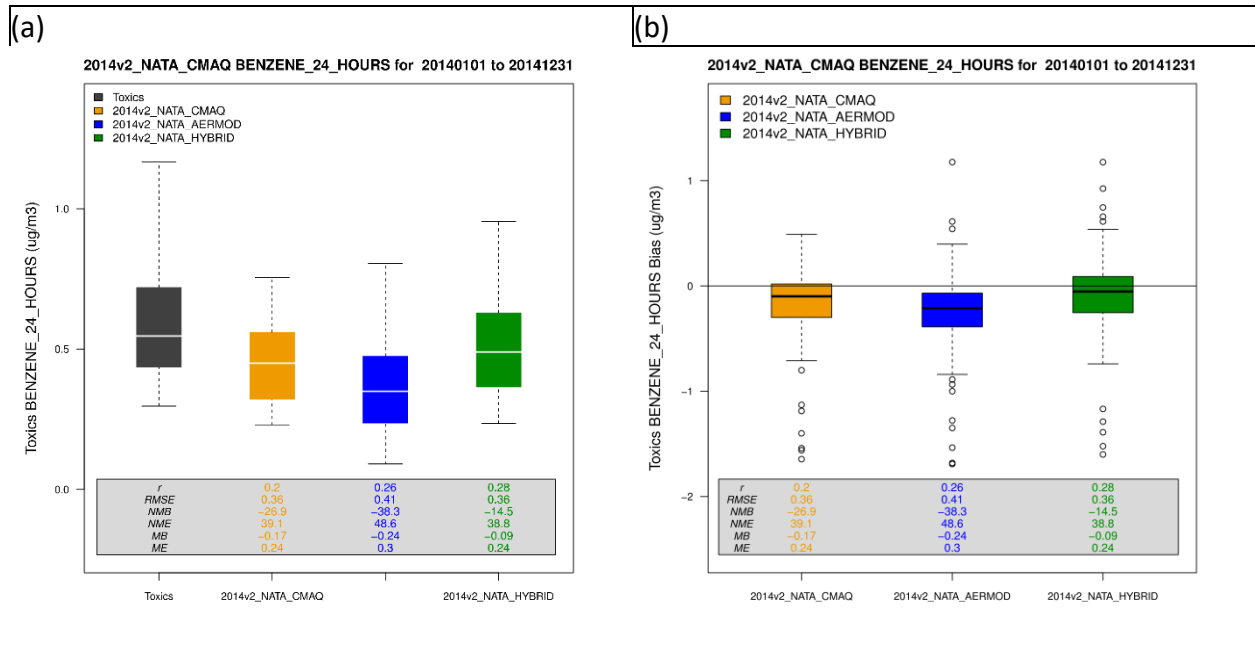
Figure 3-23. Benzene boxplots of (a) distribution (μg/m³) and (b) bias difference (μg/m³) for CMAQ, AERMOD and Hybrid models compared to ambient observations (BENZENE_24_HOURS)
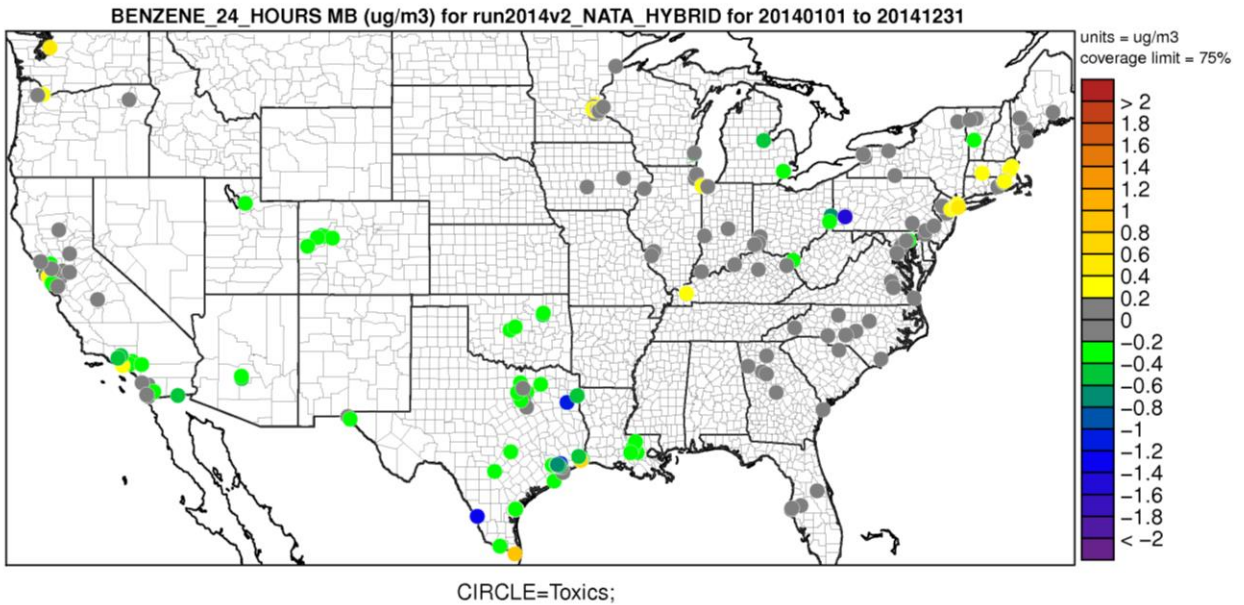
Figure 3-24. Mean bias (%) for benzene at 2014 monitoring sites (BENZENE_24_HOURS) in the Hybrid modeling domain
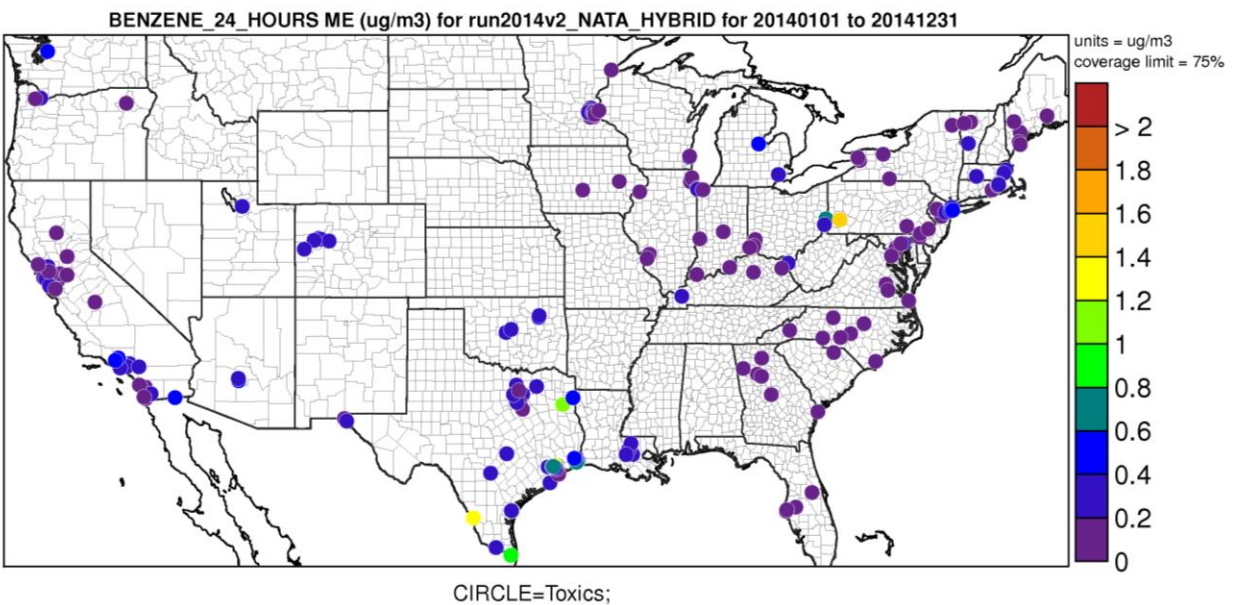
Figure 3-25. Mean error (%) for benzene at 2014 monitoring sites (BENZENE_24_HOURS) in the Hybrid modeling domain
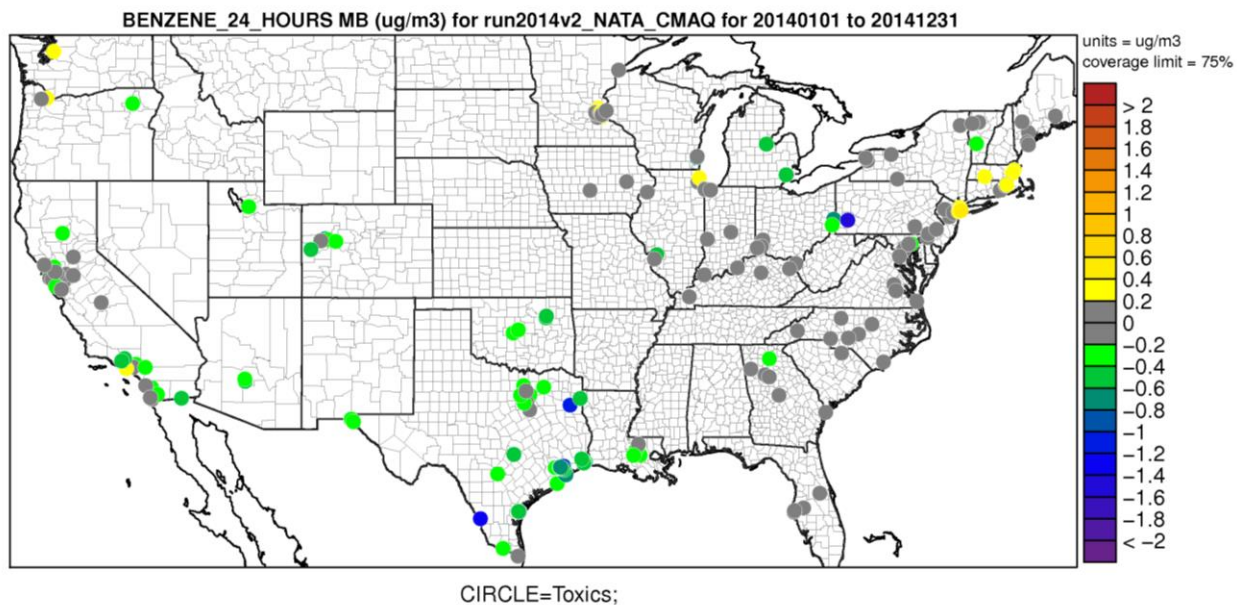
Figure 3-26. Mean bias (%) for benzene at 2014 monitoring sites (BENZENE_24_HOURS) in the CMAQ modeling domain
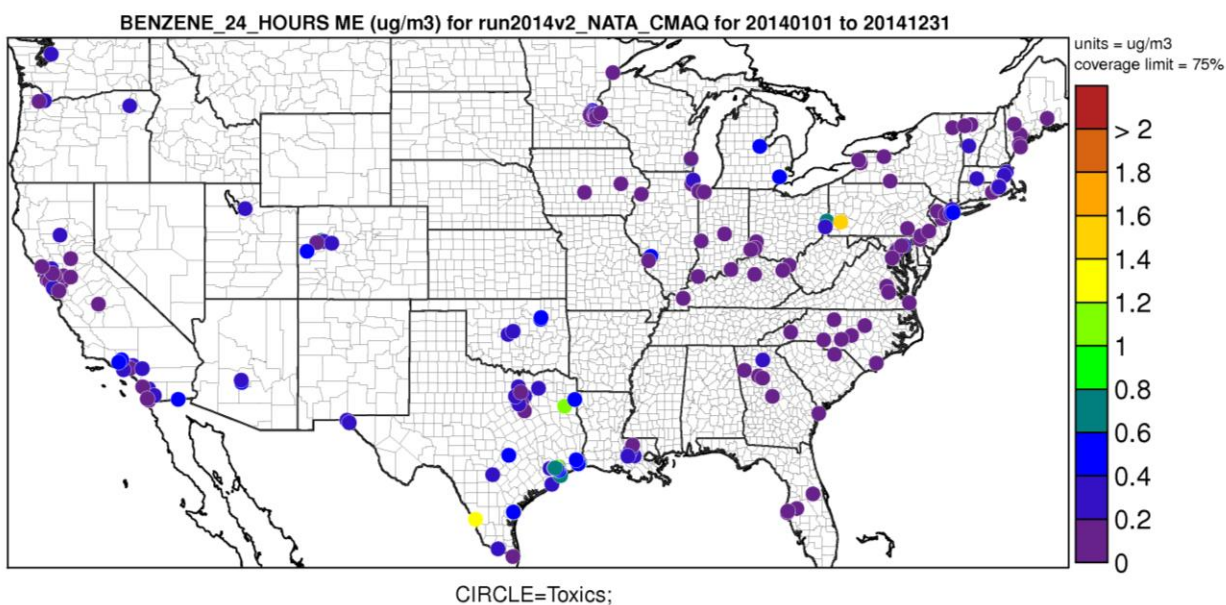
Figure 3-27. Mean error (%) for benzene at 2014 monitoring sites (BENZENE_24_HOURS) in the CMAQ modeling domain
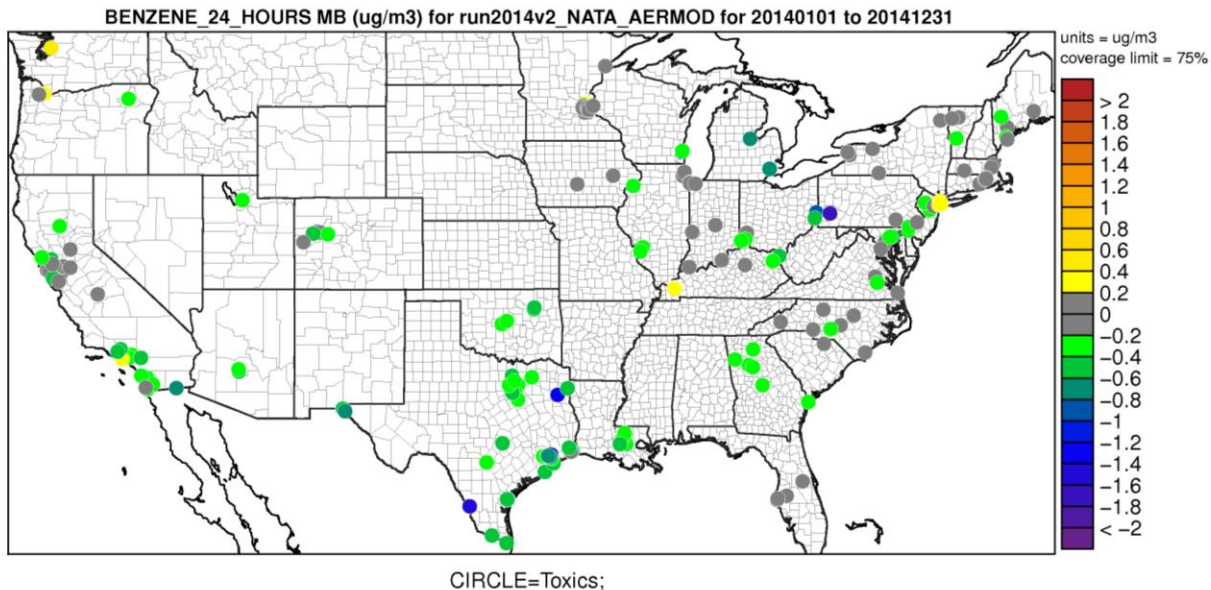
Figure 3-28. Mean bias (%) for benzene at 2014 monitoring sites (BENZENE_24_HOURS) in the AERMOD modeling domain
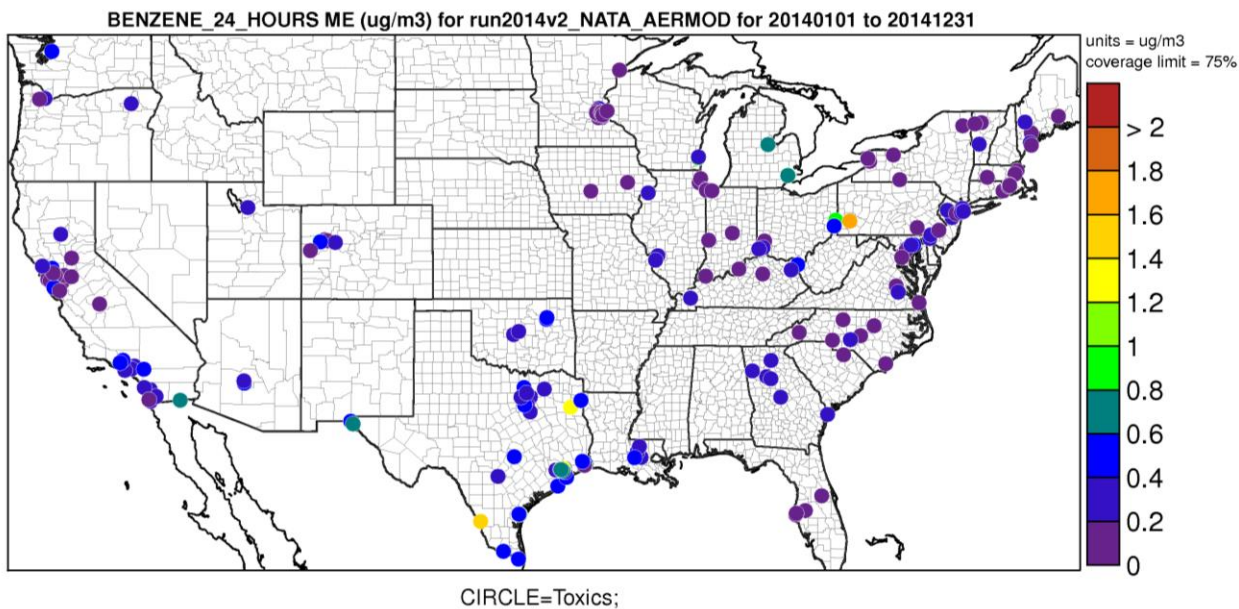
Figure 3-29. Mean error (%) for benzene at 2014 monitoring sites (BENZENE_24_HOURS) in the AERMOD modeling domain

3.7.4. Non-hybrid Evaluation

An annual operational model performance evaluation for HAPs used in the non-hybrid model calculation was conducted to estimate the ability of the AERMOD model to replicate the 2014 HAP observed ambient concentrations. Statistical assessments of modeled results versus observed pairs were paired in time and space and aggregated on an annual basis. Table 3-11 provides a list of HAPs evaluated in the

non-hybrid model performance evaluation and the number of pairs (based on completeness criteria of observations, Section 3.7.1) used in the annual median. Figure 3-30 shows the 2014 non-hybrid HAP monitoring locations. Results from the non-hybrid evaluation are presented in Appendix E.

Table 3-11. List of non-hybrid HAPs evaluated

Model Air Toxic	Measured Air Toxic	No. of Sites
Antimony	ANTIMONY_PM25_24_HOURS	156
	ANTIMONY_PM10_24_HOURS	24
	ANTIMONY_TSP_24_HOURS	33
Cobalt	COBALT_PM25_24_HOURS	169
	COBALT_PM10_24_HOURS	29
	COBALT_TSP_24_HOURS	33
Selenium	SELENIUM_PM25_24_HOURS	279
	SELENIUM_PM10_24_HOURS	26
	SELENIUM_TSP_24_HOURS	32
Methyl Bromide (Bromomethane)	METHYLBROM_5_MINUTES	12
	METHYLBROM_24_HOURS	110
Methyl Chloroform (1,1,1-Trichloroethane)	MTHYLCHLRF_5_MINUTES	12
	MTHYLCHLRF_24_HOURS	79
Carbon disulfide	CARBNDISULF_24_HOURS	72
Methyl Isobutyl Ketone (4-Methyl-2-pentanone)	MIBK_24_HOURS	50
Propanal (Propionaldehyde)	PROPIONAL_24_HOURS	79
Cumene (Isopropylbenzene)	CUMENE_24_HOURS	29
	CUMENE_1_HOUR	15
2,2,4-Trimethylpentane	TRMEPN224_24_HOURS	77
	TRMEPN224_1_HOUR	15
Bromoform (Tribromomethane)	BROMOFORM_24_HOURS	2
Chlorobenzene	CHLROBZNE_24_HOURS	35
1,2,4-Trichlorobenzene	TRICBZ124_24_HOURS	19
Benzyl Chloride (alpha-Chlorotoluene)	BENZYLCHLO_24_HOURS	13
Hexachloro-1,3-butadiene	HEXCHLRBT_24_HOURS	8
Methyl tert-butyl ether	MTBE_24_HOURS	19
p-Dioxane (1,4-Dioxane)	P_DIOXANE_24_HOURS	9
Vinyl Acetate	VINYLACET_24_HOURS	13
Ethyl Chloride (Chloroethane)	ETHYLCHLRD_24_HOURS	16
Ethylidene Dichloride (1,1-Dichloroethane)	ETHIDDICHLD_24_HOURS	16
Methyl Methacrylate	MMETACRYLAT_24_HOURS	4
Vinylidene Chloride (1,1-Dichloroethylene)	VINYLIDCLOR_24_HOURS	5
1,1,2-Trichloroethane	TRICLA112_24_HOURS	3

Non-Hybrid HAPs monitors
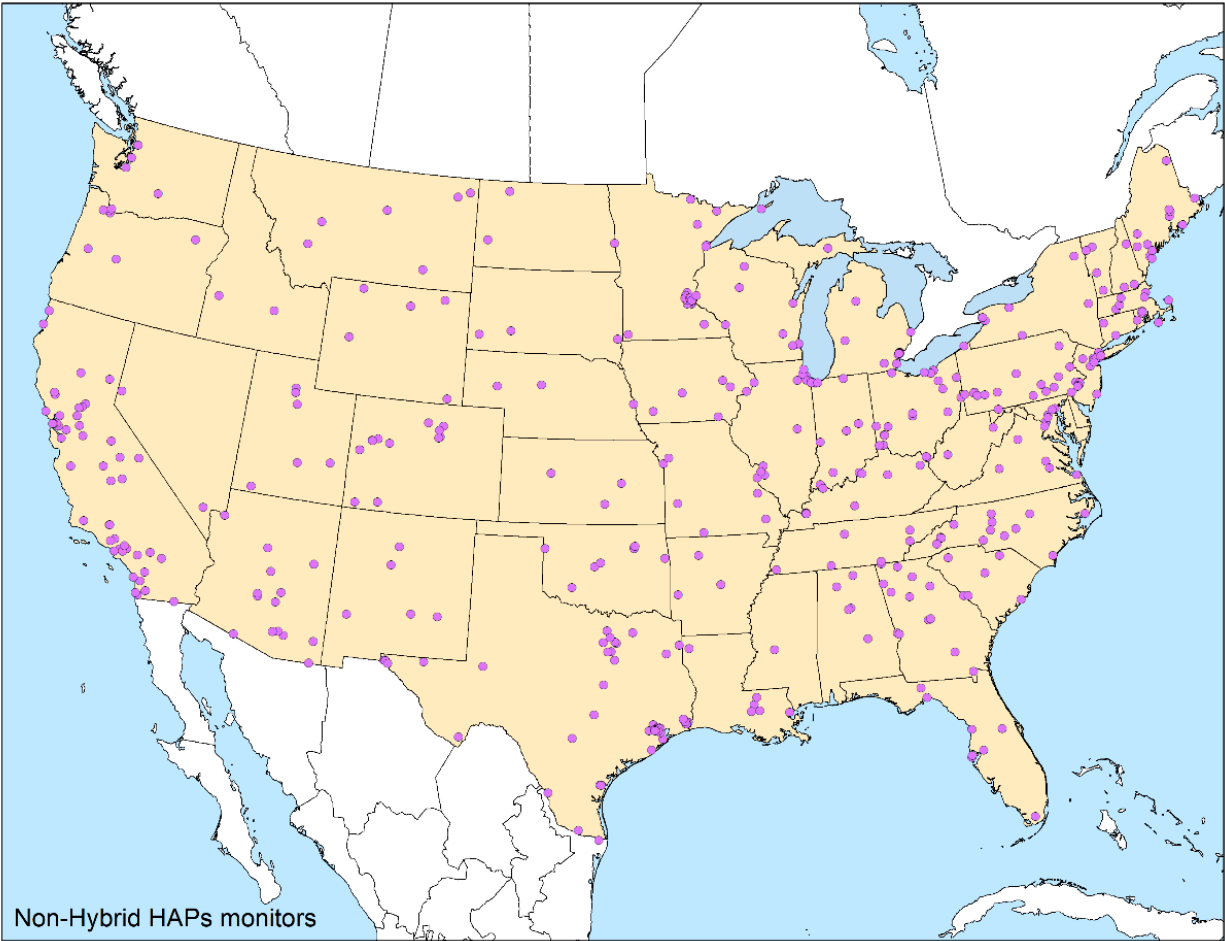
Figure 3-30. 2014 monitoring locations for the non-hybrid HAPs evaluation

4. Estimating Exposures for Populations

Estimating inhalation exposure concentrations (ECs) is a critical step in determining potential health risks. Ambient concentrations do not consider that people move through locations and microenvironments where pollutant concentrations can differ. Different people have different daily activities, spend different amounts of time engaged in those activities, and engage in those activities in different locations. Most activities occur indoors (e.g., in homes, workplaces, schools and vehicles) where pollutant concentrations can differ from those outdoors. Therefore, the average concentration of a pollutant that people breathe can differ significantly from the ambient concentration at a fixed outdoor location.

This section contains a discussion of how EPA estimated ECs for the 2014 NATA. It begins with an overview of the surrogate approach used that included new exposure modeling for some NATA pollutants and applications of exposure-to-ambient concentration ratios for the remaining pollutants. This is followed by a more detailed description of this approach, a summary of the user inputs and other data required, and an overview of the quality-assurance measures included in estimating exposures. Further details on the exposure calculations for the 2014 NATA can be found in Appendix F.

4.1. Estimating Exposure Concentrations

For the 2014 NATA, EPA used a combination of direct modeling and exposure factors to estimate inhalation ECs for NATA. This approach used census tract-level ambient concentrations estimated with air quality models, as described in Section 3, and yielded census tract-level exposure concentration estimates that we used to determine potential health risks for NATA.

EPA used version 7 of the EPA Hazardous Air Pollutant Exposure Model (HAPEM7) to conduct direct exposure modeling for the 2014 NATA. HAPEM, described in detail in Section 4.2, is a screening-level exposure model that estimates inhalation ECs corresponding to estimated ambient-pollutant concentrations. We used HAPEM7 for a selected group of surrogate pollutants and source categories. For each surrogate pollutant and NATA category (i.e., point, nonpoint, on-road mobile and nonroad mobile), EPA calculated the ratio of EC to ambient concentration (i.e., an exposure factor) for each census tract. Using each pollutant's chemical properties, we then mapped each pollutant/category combination to the surrogate pollutants and source categories. Per census tract, we multiplied the ambient concentration of the pollutant by the surrogate's exposure factor, resulting in estimated ECs. Section 4.4 further describes this exposure-factor approach.

4.2. About HAPEM

Nearly two decades ago, EPA developed HAPEM for Mobile Sources (HAPEM-MS) to assess inhalation exposure to air toxics from highway mobile sources. This version of HAPEM used carbon monoxide as a tracer for highway mobile-source air toxic emissions. Today, HAPEM7 predicts inhalation ECs for a wide range of air toxics using either modeled ambient concentrations or measured data (without regard to source category), and the model no longer uses carbon monoxide as a tracer.

Following other improvements, HAPEM version 4 and later (including HAPEM7) can estimate annual average human-exposure levels nationwide at a spatial resolution as fine as the census tract level (EPA 2002b, EPA 2005c, EPA 2007b, EPA 2015b). These changes make HAPEM7 suitable for regional and

national inhalation risk assessments such as NATA. The 1996 and 1999 NATAs estimated Inhalation ECs using HAPEM4 and HAPEM5 respectively; the 2011 and 2014 NATAs both used HAPEM7 (EPA did not use HAPEM6 for NATA). Table 4-1 outlines some key differences between these three HAPEM versions. A complete history of the model can be found in the User's Guide for HAPEM7 (EPA 2015b).

Table 4-1. Key differences between recent versions of HAPEM

Characteristic	HAPEM4	HAPEM5	HAPEM7
Data source for population demographics	1990 U.S. Census	2000 U.S. Census	2010 U.S. Census
Characterization of microenvironmental factors	Point estimates	Probability distributions	Same as HAPEM5
Method for creation of annual average activity patterns from daily activity-pattern data	Resampling of daily diaries for each of 365 days without accounting for autocorrelation	Sampling a limited number of daily diaries to represent an individual's range of activities, accounting for	Same as HAPEM5, except now includes commuter-status criterion
Interpretation of exposure-concentration range for a given cohort/tract combination	Uncertainty for the average annual EC for the cohort/tract combination	Variability of annual ECs across cohort/tract members	Same as HAPEM5, except now includes adjustments based on proximity to roadway

HAPEM uses a general approach of tracking representative individuals of specified demographic groups as they move among indoor and outdoor microenvironments and between locations. As described in the following section, personal-activity and commuting data, specific to a hypothetical person's demographic groups, are used to determine the census tracts containing residential and work locations and the microenvironments within each tract. Using stochastic sampling, the model estimates ECs by selecting empirically based factors reflecting the relationship between ECs within each microenvironment and the outdoor (ambient) air concentrations at that location.

To estimate long-term ECs for a hypothetical person, the pollutant concentrations in each microenvironment visited are first combined into a daily-average concentration. The daily averages are then combined with proper weighting for season and day type to calculate a long-term average. Finally, the long-term averages are stratified by demographic group and census tract to create a distribution of ECs for each stratum. The median of each distribution represents the best estimate of exposure for a "typical" person of that demographic group in that census tract. In this case, "typical" does not refer to a specific individual in the population or even the average over a group of individuals. Rather, this is a hypothetical person living at the centroid of a census tract who engages in the usual activities (both indoor and outdoor) for someone in that demographic group and census tract.

Additional technical information on HAPEM can be found in the User's Guide for HAPEM7 (EPA 2015b).

4.3. HAPEM Inputs and Application

HAPEM requires four main types of information to estimate ECs: (1) ambient concentrations of air toxics, (2) population data from the U.S. Census Bureau, (3) population-activity data and (4) microenvironmental data. The subsections below discuss these inputs, along with descriptions of the

data used for NATA and related information on how EPA configured the model and applied it to conduct direct exposure modeling.

4.3.1. Data on Ambient Air Concentrations

HAPEM typically uses annual average, diurnally distributed ambient air concentrations as input data. These concentrations can be monitoring data or concentrations estimated using a dispersion model or other air quality model.

For the 2014 NATA, EPA estimated annual average ambient concentrations for each census tract using a hybrid CMAQ-AERMOD approach discussed in Section 3.5. To ensure that exposure concentrations were completed in time for NATA's release, the ambient data used in the EC calculations came from a preliminary assessment based on NEI version 1 emissions data; this concession does not significantly affect the ECs calculated. EPA stratified the air quality outputs for a selected group of pollutants by one or more of the four principal NATA categories (i.e., point, nonpoint, on-road mobile and nonroad mobile), using those results as surrogates for the remaining pollutants not modeled in CMAQ-AERMOD. Thus, exposure-model results generated for NATA can be summarized for each principal NATA category or any combination of those categories.

4.3.2. Population Demographic Data

HAPEM divides the exposed population into cohorts such that each person in the population is assigned to one and only one cohort, and all the cohorts combined make up the entire population. A cohort is defined as a group of people whose exposure is expected to differ from exposures of other cohorts due to certain characteristics shared by the people within that cohort. For the 2014 NATA, we specified cohorts by residential census tract and age, with the population in each census tract divided into six age groups: 0–1, 2–4, 5–15, 16–17, 18–64 and ≥65 years of age. These groups were developed using demographic data derived from the 2010 U.S. Census. EPA aggregated the predicted inhalation ECs across cohorts to estimate ECs for the general population.

4.3.3. Data on Population Activity

HAPEM uses two types of data to define activities for the modeled population: activity-pattern data (specifying the frequency, location and duration of daily activities) and commuting-pattern data (specifying the work tracts for people living in each home tract). HAPEM uses these data together to place a hypothetical commuter in either the home or work tract and in a specific microenvironment at each 3-hour time step (the time step used for NATA). The microenvironment assignments and locations derived from these data are then used to calculate ECs (as explained in the next section).

Data on human activity patterns are used to determine the frequency and duration of exposure within various microenvironments (such as indoors at home, in-vehicle and outdoors). Activity-pattern data are taken from demographic surveys of individuals' daily activities that specify the sequence, duration and locations of those activities. The default source of activity-pattern data used by HAPEM and for NATA is EPA's Consolidated Human Activity Database (CHAD; EPA 2015a). To develop the version of CHAD used in NATA (version June 2014), data from 21 individual U.S. studies of human activities were combined into one complete data system containing over 45,000 person-days of activity-pattern records (NATAs before 2011 used previous versions of CHAD containing fewer studies and person-days of data). Because of design limitations in the studies from which it is derived, CHAD may not well represent all

demographic groups, particularly ethnic minorities and low-income populations. Also, the activity-pattern data in CHAD is mostly limited to one- or two-day periods. Extrapolation of these short-term records to the annual activity patterns needed for assessments of air toxics exposure introduces some uncertainty into the analysis.

To address this extrapolation uncertainty, HAPEM uses a stochastic process to create simulated long-term (multi-day) activity patterns from daily activity-pattern data that account for day-to-day autocorrelation. These algorithms create annual average activity patterns from daily activity-pattern data to better represent the variability among individuals within a cohort-tract combination. For each day type and demographic group, daily-activity diaries were divided into three groups based on similarity using a cluster analysis. To simulate the activities of an individual, one diary was selected from each group for each day type, resulting in nine diaries in total. Then, for each day type, the sequence of the selected diaries was determined according to the probability of transition from one cluster group to another, as determined by analyses of the CHAD data. The simulation was repeated 30 times, resulting in a set of 30 estimates of annual ECs for each demographic group in each census tract. Use of a limited number of diaries and the transition probabilities is a way to account for day-to-day autocorrelation of activities for an individual, so each exposure-concentration estimate represents an estimate for an individual rather than an average for the group. Therefore, with this approach, the range represents the variability of ECs across the group. These algorithms were used beginning with HAPEM5 (i.e., beginning with the 1999 NATA).

Commuting-pattern data, the second type of population activity data used in HAPEM, are derived for each cohort from a U.S. Census database containing information on tract-to-tract commuting patterns. These data specify the number of residents in each tract that work in that tract and every other census tract (i.e., the population associated with each home-tract/work-tract pair) and the distance between the centroids of the two tracts. An important limitation is that the commuting-pattern data included in HAPEM do not account for the movement of school-age children who travel (or commute) to a school located outside of their home tract.

4.3.4. Microenvironmental Data

A microenvironment is a three-dimensional space in which human contact with an environmental pollutant occurs. In HAPEM, this space is treated as a well-characterized, relatively homogenous location with respect to pollutant concentrations for a specified period. The inhalation exposure estimate is determined by the sequence of microenvironments visited by the individual. The concentration in each microenvironment is estimated by using the three microenvironmental factors listed below to adjust the ambient-concentration estimate for the census tract where it is located:

- a penetration factor that is an estimate of the ratio of the microenvironmental concentration to the concurrent outdoor concentration in the immediate vicinity of the microenvironment; penetration factors are pollutant-specific estimates that are derived from reported measurement studies;

- a proximity factor that is an estimate of the ratio of the outdoor concentration in the immediate vicinity of the microenvironment to the outdoor concentration represented by the ambient air concentration input to the model; and

- an additive factor that accounts for emission sources within or near a particular microenvironment, such as indoor emission sources. As noted below, the additive factor is not used for NATA.

The relationship between the estimated ECs, the input ambient concentrations and these three factors is demonstrated by the equation below.

$$C_{(i,k,t)} = CONC_{(i,t)} \times PEN_k \times PROX_k + ADD_k$$

Where:

$C_{(i,k,t)}$	=	EC predicted within census tract *i* and microenvironment *k* for time step *t*, in units of μg/m^3
$CONC_{(i,t)}$	=	ambient concentration for census tract *i* for time step *t*, in units of μg/m^3
PEN_k	=	penetration factor for microenvironment *k*
$PROX_k$	=	proximity factor for microenvironment *k*
ADD_k	=	additive factor accounting for sources within microenvironment *k*, in units of μg/m^3

Stochastic processes can be used to select work tracts, ambient air concentrations and microenvironmental factors. This important feature allows exposures to be characterized with probability distributions rather than point estimates, which more accurately reflect the variability of these components and simulate some of the variability found in measurement studies.

In HAPEM, the characteristics of each microenvironment are used to assign each microenvironment to one of three groups: indoors, outdoors and in-vehicle. The 2014 NATA used the 18 microenvironments shown in Table 4-2. The microenvironments in the indoor group were further classified as associated with either residence or other buildings, while those in the outdoor group were categorized as either near-road or away-from-road. Each group consists of microenvironments expected to have similar penetration factors, thus allowing microenvironmental factors developed for one microenvironment to be applied to other microenvironments in the same group. Within each census tract, HAPEM uses estimates of the number of people living within each of three distance-from-road bins to stochastically vary the proximity factor based on distance-from-road (i.e., proximity factors are higher for microenvironments near major roadways, lower for microenvironments relatively far from major roadways). The additive factor (ADD_k) in the expression for EC, above, was set to zero for NATA because indoor-source data are currently incomplete (recall that NATA covers only pollutants derived from outdoor sources).

An important consideration is that data to support quantitative microenvironmental factors are not well developed for many of the air toxic compounds and for most of the microenvironments, which introduces uncertainty into the analysis of exposures. Section 7 contains a discussion on uncertainty and variability regarding this and other issues for NATA.

Table 4-2. Microenvironments used in HAPEM modeling for the 2014 NATA

Indoors	Outdoors	In Vehicle
Residence Residential *Other Building* Air Travel Bar/Restaurant Hospital Office Public Access School Waiting Inside for Public Transit	*Near-road* Motorcycle/Bicycle Outdoors, Near Roadway Outdoors, Parking Garage Outdoors, Service Station Residential Garage Waiting Outdoors for Public Transit *Away-from-road* Ferryboat Outdoors Other	Car/Truck Public Transit

4.4. Exposure Factors

HAPEM exposure modeling for NATA requires substantial time and resources for data collection and processing, computing and model processing. Due to these requirements, EPA conducted HAPEM modeling for the 2014 NATA only for selected pollutants, which we present below along with how we used them to estimate ECs for the remaining NATA pollutants.

- **Coke oven emissions** (emitted by point sources and present in ambient air as either particulates or gases) and **diesel PM** (modeled as particulates from nonpoint and mobile sources) were special cases that EPA modeled as themselves in HAPEM and not used as surrogates for any other pollutants not modeled in HAPEM.

- **Benzene** and **1,3-butadiene** are gas-phase pollutants emitted by many processes (and all four principal NATA categories) in nearly all U.S. locations. EPA selected benzene as the surrogate for all other gas-phase pollutants not modeled in HAPEM (EPA considers benzene modeling in NATA to be more reliable than 1,3-butadiene modeling).

- Unspeciated, generic PAHs ("**PAH, total**"), which are pollutants that can be present in either gas phase or particulate phase in ambient air, are emitted by all four principal NATA categories and from a wide variety of processes. EPA selected "PAH, total" as the surrogate for all other mixed-phase pollutants not modeled by HAPEM.

- **Chromium (VI)** is a highly toxic particulate-phase pollutant emitted by all four principal NATA categories. EPA selected it as the surrogate for all other particulate pollutants not modeled in HAPEM and emitted by point or nonpoint sources.

- EPA selected **nickel**, a particulate-phase pollutant emitted by a variety of processes spread across the United States, as the surrogate for all other particulate pollutants not modeled in HAPEM and emitted by mobile sources.

Appendix F to this document contains further details on the application of HAPEM7 in the 2014 NATA analysis. The spreadsheet file "2014_NATA_Exposure_Factors.xlsx" within the Supplemental Data folder accompanying this TSD spreadsheet contains the tract-level exposure-to-ambient concentration ratios (i.e., exposure factors) for each surrogate pollutant. (Access the NATA website at https://www.epa.gov/national-air-toxics-assessment.)

Overall, the HAPEM exposure predictions are lower than the corresponding predicted air quality values. This reduction likely results from the inability of many pollutants to penetrate efficiently into an indoor environment. (Recall that indoor sources of air toxics have not been included in any versions of NATA completed to date).

4.5. Evaluating Exposure Modeling

A model-performance evaluation can provide valuable information regarding model uncertainty when using computer-simulation models of human exposures to pollutants. Also, a well-conducted evaluation can greatly increase confidence in model results for a given application or use. One type of performance evaluation uses measurements and environmental data as a benchmark to compare modeling estimates. EPA has worked with the Mickey Leland Center (NUATRC 2011) on past assessments to help identify new and independent sources of personal-monitoring data for use in comparison with the NATA results.

Extensive peer review involving independent scientific and technical advice from scientists, engineers and economists can be another valuable component of a model evaluation. In July 2000, HAPEM4 underwent external peer review by technical experts for both the microenvironmental factors used in the model and the overall application of the model for NATA. A discussion of several of the issues addressed by these reviews is included in Appendix A of the report for the 1996 NATA presented to EPA's Science Advisory Board for review (EPA 2001b). In 2001, EPA's Scientific Advisory Board reviewed the application of HAPEM4 as part of the 1996 NATA review (EPA 2001a). Although several limitations were identified in the current methodology, HAPEM4 was acknowledged as an appropriate tool to help better understand the relationship of human exposures to ambient-concentration levels. Since then, numerous studies have used and cited subsequent HAPEM versions, including HAPEM7.

4.6. Summary

- Estimating inhalation ECs is a critical step in determining potential health risks because ambient concentrations do not account for movements of individuals among locations and microenvironments where pollutant concentrations can differ.

- We estimated inhalation ECs for each pollutant/source group/census tract for the 2014 NATA using the HAPEM7 model.

- These ECs can be used to determine census tract-level potential health risks.

5. Characterizing Effects of Air Toxics

Exposure to air toxics is associated with increased incidence of cancer and a variety of adverse noncancer health effects. The type and severity of effects depends on several factors, including the identity and nature of the chemical to which an individual is exposed, the magnitude and duration of exposure, and the unique behaviors and sensitivities of exposed individuals.

EPA uses a toxicity assessment to identify and quantify the adverse health effects associated with exposure to a chemical, following EPA risk assessment methods. As indicated in Figure 1-3 of this document and described in more detail in Volume 1 of EPA's ATRA Reference Library (EPA 2004a), two processes constitute toxicity assessment: hazard identification (during which the specific adverse effects are identified that can be causally linked with exposure to a given chemical) and dose-response assessment (which characterizes the quantitative relationship between chemical dose or concentration and adverse effects, that is, the hazard(s) identified in the first step).

> The phrase "dose-response" is used generally throughout this document to refer to the relationship between a level of a chemical and a physical response. The values EPA uses for inhalation, however, are derived for exposure concentration, although with consideration of dose. Consideration of the relationship between exposure concentration, dose and dosimetry (how the body handles a chemical once it is inhaled) is inherent in the derivation of values. The term "toxicity values" is used here to refer to the RfCs and UREs used in inhalation risk assessment.

Ultimately, the results of the toxicity assessment, referred to in this document as "toxicity values," are used along with exposure estimates to characterize health risks for exposed populations (as described in Section 6). Although the toxicity assessment is integral to the overall air toxics risk assessment, it is usually done prior to the risk assessment. EPA has completed this toxicity assessment for many air toxics and has made available the resulting toxicity information and dose-response values, which have undergone extensive peer review.

This section explains how toxicity assessments are used in the NATA risk assessment process. Specifically, the sections that follow provide an overview of the cancer and noncancer toxicity values used in NATA and the primary sources of these values. They also describe several adjustments and assumptions to toxicity values specific to the NATA risk assessment process.

5.1. Toxicity Values and Their Use in NATA

The toxicity values used in NATA are quantitative expressions used to estimate the likelihood of adverse health effects given an estimated level and duration of exposure. These toxicity values are based on the results of dose-response assessments, which estimate the relationship between the dose and the frequency or prevalence of a response in a population or the probability of a response in any individual. Because NATA is focused on long-term exposures, the toxicity values used in NATA are based on the results of chronic dose-response studies when such data are available. Chronic dose-response assessments can be used to help evaluate the specific 70-year-average (i.e., "lifetime") ECs associated with cancer prevalence rates, or, for noncancer effects, the concentrations at which noncancer adverse health effects might occur given exposure over an extended time (possibly a lifetime, but the time frame also can be shorter).

The toxicity values that are combined with ECs to conduct the risk characterization in NATA are based on the results of quantitative dose-response assessments. The actual values used, however, are not strictly considered dose-response or concentration-response values. To estimate cancer risks in NATA, EPA converted the results of cancer dose-response assessments for a given chemical to a URE that incorporates certain exposure assumptions. This value can be multiplied by the 70-year-average EC to obtain a lifetime cancer risk estimate for each individual. To evaluate the potential for noncancer adverse health effects, we used chronic dose-response data to estimate a threshold that is the EC in air at which adverse health effects are assumed to be unlikely (i.e., the RfC). These two types of values are described in more detail in the following section.

The toxicity values used in NATA are consistent with those OAQPS has compiled for chronic inhalation exposures to air toxics. The full set of toxicity values used for the 2014 NATA (and their sources) can be accessed via the NATA Supplemental Data file; see Appendix B for details. Sources of chronic dose-response assessments used for the 2014 NATA were prioritized according to OAQPS risk assessment guidelines and level of peer review, as discussed below.

5.2. Types of Toxicity Values

Each toxicity value used in NATA is best described as an estimate within a range of possible values appropriate for screening-level risk assessments. Note that the uncertainty in the dose-response assessments and toxicity values that NATA relies on is to some extent one-sided, providing a conservative (health-protective) estimate of risk. The "true" cancer risk and potential for adverse noncancer impacts are believed to be lower than those estimated in this assessment, although the possibility remains that they could be greater. Uncertainty in the derivation of the dose-response values and in other aspects of the NATA process is discussed in Section 7.

5.2.1. Cancer URE

A cancer dose-response curve is used to demonstrate the quantitative relationship between dose and the likelihood of contracting cancer. If the dose-response relationship is linear, the cancer response is assumed to increase proportionally with the dose (which might be expressed as an EC, an absorbed internal dose, a dose to a specific organ or tissue, or other measure). We have proposed that linear extrapolation of carcinogenic risk in the low-dose region of the curve is a reasonable approach for estimating risk at relatively low exposures, such as

> The **URE** is the upper-bound excess lifetime cancer risk estimated to result from continuous exposure to an agent at a concentration of 1 microgram per cubic meter ($\mu g/m^3$) in air. UREs are considered upper-bound estimates, meaning they represent a plausible upper limit to the true value. The true risk is likely to be less, but could be greater.

those typically experienced by the general population for air toxics (i.e., the true value of the risk is unknown, and could be as low as zero). An upper-bound lifetime cancer risk represents a plausible upper limit to the true probability that an individual will contract cancer due to exposure over a 70-year lifetime to a given hazard (e.g., exposure to an air toxic).

For an inhalation risk assessment (and for NATA), a URE can be used to calculate the estimated cancer risk from inhalation ECs. A URE is calculated by using dose-response information for a chemical and developing a factor in the appropriate units that can be combined directly with ECs in air to estimate individual cancer risks, given certain assumptions regarding the exposure conditions. Specifically, the

URE represents the upper-bound of the excess cancer risk estimated to result from continuous exposure to a concentration of 1 µg of a substance per m3 of air, over a 70-year lifetime and assuming a daily inhalation rate of about 20 m3/day. The risk value is derived from the slope of the dose-response curve as estimated using a linearized multistage statistical model in the low-dose portion of the curve. The interpretation of the URE is as follows: If the URE is 3×10-6 µg/m3, no more than three excess tumors would develop per 1 million people if they were exposed daily for a lifetime to a concentration of 1 µg/m3. To the extent that true dose-response relationships for some air toxics compounds are not strictly linear, this assumption could result in overestimates of cancer risk. The upper bound is not a true statistical confidence limit because the URE reflects unquantifiable assumptions about effects at low doses. Thus, although the actual carcinogenic risk is likely to be lower than what is reflected in the URE, it also might be higher.

The URE estimates the toxic potency of a chemical. EPA's weight-of-evidence (WOE) descriptors provide estimates of the level of certainty regarding a chemical's carcinogenic potential. We evaluate three broad categories of toxicological data to make a WOE determination: (1) human data (primarily epidemiological); (2) animal data (results of long-term experimental animal bioassays) and (3) supporting data, including a variety of short-term tests for genotoxicity and other relevant properties, pharmacokinetic and metabolic studies and structure-activity relationships. We evaluate these data in combination to characterize the extent to which they

> **EPA's Weight of Evidence (WOE) Descriptors**
> (EPA 2005a)
>
> - Carcinogenic to humans
> - Likely to be carcinogenic to humans
> - Suggestive evidence of carcinogenic potential
> - Inadequate information to assess carcinogenic potential
> - Not likely to be carcinogenic to humans

support the hypothesis that an agent or chemical causes cancer in humans. The approach outlined in EPA's Guidelines for Carcinogen Risk Assessment (EPA 2005a) considers available scientific information regarding carcinogenicity. It provides a narrative approach to characterizing carcinogenicity rather than assigning chemicals to specific categories (as was done previously by EPA according to the 1986 guidelines). To provide some measure of clarity and consistency in an otherwise free-form, narrative characterization, standard descriptors are used as part of the hazard narrative to express the conclusion regarding the WOE for carcinogenic-hazard potential. The five recommended standard hazard descriptors are described below.

Carcinogenic to Humans: This descriptor indicates strong evidence of human carcinogenicity. It is appropriate when the epidemiologic evidence of a causal association between human exposure and cancer is convincing. Alternatively, this descriptor might be equally appropriate with a lesser weight of epidemiologic evidence that is strengthened by other lines of evidence. It can be used when all the following conditions are met: (1) evidence of an association between human exposure and either cancer or the key precursor events of the agent's mode of action is strong but insufficient for a causal association; (2) evidence of carcinogenicity in animals is extensive; (3) the mode(s) of carcinogenic action and associated key precursor events have been identified in animals and (4) evidence is strong that the key precursor events that precede the cancer response in animals are anticipated to occur in humans and progress to tumors, based on available biological information.

Likely to Be Carcinogenic to Humans: This descriptor is appropriate when the WOE is adequate to demonstrate carcinogenic potential to humans but does not reach the WOE for the descriptor "Carcinogenic to Humans." Adequate evidence consistent with this descriptor covers a broad spectrum.

At one end of the spectrum is evidence for an association between human exposure to the agent and cancer and strong experimental evidence of carcinogenicity in animals. At the other end, with no human data, the weight of experimental evidence shows animal carcinogenicity by a mode or modes of action that are relevant or assumed to be relevant to humans. The use of the term "likely" as a WOE descriptor does not correspond to a quantifiable probability. Moreover, additional data, such as information on the mode of action, might change the choice of descriptor for the illustrated examples.

Suggestive Evidence of Carcinogenic Potential: This descriptor is appropriate when the WOE is suggestive of carcinogenicity; that is, a concern for potential carcinogenic effects in humans is raised, but the data are judged insufficient for a stronger conclusion. This descriptor covers a spectrum of evidence associated with varying levels of concern for carcinogenicity, ranging from a positive cancer result in the only study on an agent to a single positive cancer result in an extensive database that includes negative studies in other species. Depending on the extent of the database, additional studies may or may not provide further insights.

Inadequate Information to Assess Carcinogenic Potential: This descriptor is appropriate when available data are judged inadequate for applying one of the other descriptors. Additional studies generally would be expected to provide further insights.

Not Likely to Be Carcinogenic to Humans: This descriptor is appropriate when the available data are considered robust for deciding no basis for human hazard concern exists. In some instances, positive results in experimental animals can occur when the evidence is strong and consistent that each mode of action in experimental animals does not operate in humans. In other cases, there can be convincing evidence in both humans and animals that the agent is not carcinogenic. A descriptor of "not likely" applies only to the circumstances supported by the data. For example, an agent might be "Not Likely to Be Carcinogenic" by one route but not necessarily by another. In those cases that have one or more positive animal experiments but the results are judged to be not relevant to humans, the narrative discusses why the results are not relevant. As with the "likely" descriptor, the term "not likely" here does not correspond to a quantifiable probability.

Important to note is that these WOE categories express only a relative level of certainty that these substances might cause cancer in humans. The categories do not specifically connote relative levels of hazard or the degree of conservatism applied in developing a dose-response assessment. For example, a substance with suggestive evidence of carcinogenic potential might impart a greater cancer risk to more people than another substance that is carcinogenic to humans.

The process of developing UREs includes several important sources of uncertainty. Many of the air toxics in NATA are classified as "likely" carcinogens. The term likely, as used in this instance, means that data are insufficient to prove these substances definitively cause cancer in humans. That some are not human carcinogens at environmentally relevant ECs is possible, and the true cancer risk associated with these air toxics might be zero. UREs for most air toxics were developed from animal data using health-protective methods to extrapolate to humans. Actual human responses may differ from those predicted. For more information, see EPA's [Guidelines for Carcinogen Risk Assessment](EPA 2005a).

5.2.2. Noncancer Chronic RfC

The RfC is an estimate of a continuous inhalation exposure that is thought to be without an appreciable risk of adverse health effects over a lifetime. The population considered when deriving RfCs includes sensitive subgroups (i.e., children, asthmatics and the elderly). The RfC is derived by reviewing a health-effects database for a chemical and identifying the most sensitive and relevant endpoint, along with the principal study or studies demonstrating that endpoint. The value is calculated by dividing the no-observed-adverse-effect level (or an analogous exposure level obtained with an alternate approach, e.g., a lowest-observed-adverse-effect level or a benchmark dose) by uncertainty factors reflecting the limitations of the data used.

> The **RfC** is an estimate (with uncertainty spanning perhaps an order of magnitude) of a continuous inhalation exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime.

As with UREs for cancer risk assessment, the process of developing RfCs includes several important sources of uncertainty, which span perhaps an order of magnitude. Uncertainty factors are intended to account for (1) variation in sensitivity among the individuals in the population, (2) uncertainty in extrapolating laboratory animal data to humans, (3) uncertainty in extrapolating from data obtained in a study involving a less-than-lifetime exposure, (4) uncertainty in using lowest-observed-adverse-effect-level or other data rather than no-observed-adverse-effect-level data and (5) inability of any single study to address all possible adverse outcomes in humans adequately. Additionally, an adjustment factor is sometimes applied to account for scientific uncertainties in the data or study design not explicitly captured in the uncertainty factors (e.g., a statistically inadequate sample size or poor exposure characterization). For more information, refer to EPA's Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry (EPA 1994).

Unlike linear dose-response assessments for cancer, noncancer risks generally are not expressed as a probability that an individual will experience an adverse effect. Instead, in an air toxics risk assessment, the potential for noncancer effects in humans is typically quantified by calculating the ratio of the inhalation EC to the RfC. This ratio is referred to as the hazard quotient (HQ). For a given air toxic, exposures at or below the RfC (i.e., HQs are 1 or less) are not likely to be associated with adverse health effects. As exposures increase above the RfC (i.e., HQs are greater than 1), the potential for adverse effects also increases. The HQ, however, should not be interpreted as a probability of adverse effects.

Additional information is provided in the description of risk characterization for NATA in Section 6 of this document.

5.3. Data Sources for Toxicity Values

Information on dose-response assessments for evaluating chronic exposures for NATA was obtained from multiple sources and prioritized according to OAQPS risk assessment guidelines and level of peer review. Our approach for selecting appropriate toxicity values generally places greater weight on the EPA-derived toxicity values than those from other agencies (listed below).

Additionally, the approach of favoring EPA values (when they exist) has been endorsed by EPA's Science Advisory Board. This ensures the use of values most consistent with well-established and scientifically based EPA science policy. A spreadsheet file "NATA_Pollutants.xlsx" within the Supplemental Data folder

accompanying this TSD contains the toxicity values for both cancer and noncancer chronic effects used in the 2014 NATA. Cancer effects are characterized according to the extent to which available data support the hypothesis that a pollutant causes cancer in humans.

5.3.1. U.S. EPA Integrated Risk Information System

We disseminate dose-response assessment information in several forms, depending on the level of internal review. The Integrated Risk Information System (**IRIS**) is an electronic database prepared and maintained by EPA that contains information on human-health effects that could result from exposure to various substances in the environment. These assessments have undergone external peer review and subsequent revision, compliant with requirements EPA instituted in 1996 for the IRIS review process.

Externally peer-reviewed assessments under development for IRIS were given first consideration for NATA. These assessments, which reflect the most recent available toxicity information and data analysis, were used in some cases to supersede existing values on IRIS. Current IRIS values were used for NATA when peer-reviewed IRIS values under development were not available.

5.3.2. U.S. Department of Health and Human Services, Agency for Toxic Substances and Disease Registry

The Agency for Toxic Substances and Disease Registry (ATSDR) publishes minimal risk levels (MRLs) for many substances based on health effects other than cancer. The MRL is defined as an estimate of human exposure to a substance that is likely to be without an appreciable risk of adverse effects (other than cancer) over a specified duration of exposure. For noncancer values in the 2014 NATA, inhalation MRLs were used when IRIS RfC values were not available or when the ATSDR value was based on more recent, peer-reviewed data and analysis methods than the IRIS value (because the ATSDR concept, definition and derivation are analogous to IRIS). ATSDR does not develop assessments based on carcinogenicity. After internal and external review, MRLs are published in pollutant-specific toxicological-profile documents. ATSDR regularly updates these toxicological-profile documents; they are available at Toxic Substances Portal MRLs (ATSDR 2015).

5.3.3. California Environmental Protection Agency Office of Environmental Health Hazard Assessment

California's Office of Environmental Health Hazard Assessment (OEHHA) develops UREs based on carcinogenicity and reference exposure levels (RELs) based on health effects other than cancer. The REL is defined as a concentration level at or below which no adverse health effects are anticipated. For cancer and noncancer values in the 2014 NATA, OEHHA UREs and inhalation RELs were used when their derivation was determined to be consistent with the concepts and definitions of IRIS or ATSDR. OEHHA dose-response information is available at Air Toxicology and Epidemiology (OEHHA 2016). Technical support documents for assessing hot spots are available on the OEHHA website at Hot Spots Guidelines (OEHHA 2015).

5.3.4. U.S. EPA Health Effects Assessment Summary Tables

The Health Effects Assessment Summary Tables (EPA 1997) are a comprehensive listing consisting almost entirely of provisional UREs, RfCs and other risk assessment information of interest that various EPA offices have developed. The assessments, which have never been submitted for EPA consensus,

were last updated in 2001. NATA uses information from these tables only when no values from the sources discussed in Sections 5.3.1 through 5.3.3 are available.

5.3.5. World Health Organization International Agency for Research on Cancer

The International Agency for Research on Cancer of the World Health Organization (WHO) coordinates and conducts research on cancer and provides information on related cancer research and epidemiology. Although the agency does not develop quantitative dose-response values, it has published a series of monographs (WHO 2018) on the carcinogenicity of a wide range of substances. The following "degrees of evidence" published by the International Agency for Research on Cancer were used when EPA WOE determinations were not available for a substance or are out of date (see the NATA Glossary of Terms in Appendix A for definitions of each):

- Group 1:   Carcinogenic to humans;

- Group 2A: Probably carcinogenic to humans;

- Group 2B: Possibly carcinogenic to humans;

- Group 3:   Not classifiable as to human carcinogenicity; and

- Group 4:   Probably not carcinogenic to humans.

5.4. Additional Toxicity Decisions for Some Chemicals

After the dose-response information was prioritized, we made additional changes to some of the chronic inhalation exposure values to address data gaps, increase accuracy and avoid underestimating risk for NATA.

5.4.1. Polycyclic Organic Matter

A substantial proportion of polycyclic organic matter (POM) reported in the 2014 NEI was not speciated into individual compounds. For example, some emissions of POM were reported in NEI as "7-PAH" or "16-PAH," representing subsets of certain POM, or simply as "total PAH" or "polycyclic organic matter." In other cases, individual POM compounds were reported for which no quantitative cancer dose-response value has been published in the sources used for NATA. As a result, we made simplifying assumptions that characterize emissions reported as POM. This allows us to quantitatively evaluate cancer risk for these species without substantially under- or overestimating risk (which can occur if all reported emissions of POM are assigned the same URE). To accomplish this, POM emissions as reported in NEI were grouped into categories. EPA assigns dose-response values based on the known or estimated toxicity for POM within each group and on information for the POM speciation of emission sources, such as wood fires and industrial processes involving combustion.

For the 2014 NATA, we divided unspeciated POM emissions into eight POM groups. The first two groups included unspeciated POM (including "total PAH") and individual POM species with no URE assigned. Both groups were assigned a URE equal to 5 percent of that for pure benzo[a]pyrene, accounting for toxicity and the estimated emission profile of POM compounds. Groups 3 through 7 comprised POM

compounds for which emissions were reported as individual compounds and for which UREs have been estimated. Compounds in these groups were categorized based on toxicity, and an appropriate URE was assigned to each category based on toxicity of the compounds included in the group. Category 8 was composed of unspeciated polynuclear aromatic hydrocarbons reported as 7-PAH and was assigned a URE equal to 18 percent of that for pure benzo[a]pyrene. We discuss the POM groups used for the 2014 NEI in Section 2.1.1.1.

We concluded that three PAHs – anthracene, phenanthrene and pyrene – are not carcinogenic, and therefore no URE was assigned to these for the 2014 NATA. Details of the analysis that led to this conclusion can be found in the document entitled [Development of a Relative Potency Factor (RPF) Approach for Polycyclic Aromatic Hydrocarbon (PAH) Mixtures: In Support of Summary Information of the Integrated Risk Information System (IRIS)](EPA 2010a).

5.4.2. Glycol Ethers

Much of the emission-inventory information for the glycol ether category reported only the total mass for the entire group without distinguishing among individual glycol ether compounds. In other cases, emissions of individual glycol ether compounds that had not been assigned dose-response values were reported. Individual glycol ether compounds vary substantially in toxicity. To avoid underestimating the health hazard associated with glycol ethers, we protectively applied the RfC for **ethylene glycol methyl ether** (the most toxic glycol ether for which an assessment exists) to glycol ether emissions of unspecified composition.

5.4.3. Acrolein

EPA first derived an IRIS RfC for acrolein in 2003 (EPA 2003a). This value was based on a 1978 subchronic rodent study that identified a lowest-observed-adverse-effect level (LOAEL) for nasal lesions (Feron et al. 1978). In 2008, the OEHHA derived a chronic reference exposure level for acrolein that was based on a more recent subchronic rodent study. The newer study identified a no-observed-adverse-effect level (NOAEL) for nasal lesions (OEHHA 2008; Dorman et al. 2008). Because both studies identified nasal lesions as the critical effect and because the Dorman et al. (2008) study identified a NOAEL, we used the more recent OEHHA REL for acrolein in the 2014 NATA.

5.4.4. Metals

Several decisions made for the 2014 NATA regarding the toxicity values used for metal compounds are discussed in this section.

Chromium (VI) compounds. We used the IRIS RfC for particulate chromium (VI) instead of the RfC for chromic acid mists and dissolved aerosols to avoid underestimating the health hazard associated with these compounds. The RfC for particulate chromium (VI) is less than those RfCs for chromic acid mists and dissolved aerosols.

Lead. We consider the primary National Ambient Air Quality Standard (NAAQS) for lead, which incorporates an ample margin of safety, to be protective of all potential health effects for the most susceptible populations. The NAAQS, developed using the EPA Integrated Exposure, Uptake, Biokinetic Model, was preferred over the RfC for noncancer adverse effects because the NAAQS for lead was developed using more recent toxicity and dose-response information on the noncancer adverse impacts

of lead. The NAAQS for lead was set to protect the health of the most susceptible children and other potentially at-risk populations against an array of adverse health effects, most notably including neurological effects, particularly neurobehavioral and neurocognitive effects (which are the effects to which children are most sensitive). The lead NAAQS, a rolling 3-month average level of lead in total suspended particles, was used as a long-term value in NATA.

Nickel compounds. The cancer inhalation URE for most of the emissions of nickel compounds included in NATA (including unspecified nickel emissions reported as "nickel compounds") was derived from the IRIS URE for insoluble nickel compounds in crystalline form. Soluble nickel species, and insoluble species in amorphous form, do not appear to produce genotoxic effects by the same toxic mode of action as insoluble crystalline nickel. Nickel speciation information for some of the largest nickel-emitting sources, including oil and coal combustion, suggests that at least 35 percent of total nickel emissions could be soluble compounds. The remaining insoluble nickel emissions, however, are not well characterized. Consistent with this limited information, we conservatively assumed for NATA that 65 percent of emitted nickel is insoluble and that all insoluble nickel is crystalline. Because the nickel URE listed in IRIS is based on nickel subsulfide and represents pure insoluble crystalline nickel, it was adjusted to reflect an assumption that 65 percent of the total mass of emitted nickel might be carcinogenic. In cases where a chemical-specific URE was identified for a reported nickel compound, it was used without adjustment. Furthermore, the MRL in Table 2 of the ATSDR is not adjusted because the noncancer effects of nickel are not thought to be limited to the insoluble crystalline form.

Manganese. We used the ATSDR MRL for manganese (Mn) as the chronic inhalation reference value in the 2014 NATA. There is an existing IRIS RfC for Mn (EPA 1993), and ATSDR published an assessment of Mn toxicity which includes a chronic inhalation reference value (i.e., an ATSDR Minimal Risk Level, MRL). (ATSDR 2012). Both the 1993 IRIS RfC and the 2012 ATSDR MRL were based on the same study (Roels et al. 1992); however, ATSDR used updated dose-response modeling methodology (benchmark dose approach) and considered recent pharmacokinetic findings to support their MRL derivation. Because of the updated methods, we determined that the ATSDR MRL is the appropriate noncancer health reference value for manganese in NATA.

5.4.5. Adjustment of Mutagen UREs to Account for Exposure During Childhood

For carcinogenic chemicals acting via a mutagenic mode of action (i.e., chemicals that cause cancer by damaging genes), we recommend that estimated risks reflect the increased carcinogenicity of such chemicals during childhood. This approach is explained in detail in the Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens (EPA 2005b). Where available data do not support a chemical-specific evaluation of differences between adults and children, the Supplemental Guidance recommends adjusting for early-life exposures by increasing the carcinogenic potency by 10-fold for children up to 2 years old and by 3-fold for children 2 to 15 years old. These adjustments have the aggregate effects of increasing the estimated risk by about 60 percent (a 1.6-fold increase) for a lifetime of constant inhalation exposure. EPA recommends making these default adjustments only for carcinogens known to be mutagenic and for which data to evaluate adult and juvenile differences in toxicity are not available.

For NATA 2014, we adjusted the UREs for acrylamide, benzidine, chloroprene, coke oven emissions, ethyl carbamate, ethylene oxide, methylene chloride, nitrosodimethylamine and PAHs upward by a factor of 1.6 to account for the increased risk during childhood exposures. Although trichloroethylene is

carcinogenic by a mutagenic mode of action, the age-dependent adjustment factor for the URE only applies to the portion of the slope factor reflecting risk of kidney cancer. For full lifetime exposure to a constant level of trichloroethylene exposure, we adjusted the URE upward by a factor of 1.12 (rather than 1.6 as discussed above). For more information on applying age-dependent adjustment factors in cases where exposure varies over the lifetime, see Toxicological Review of Trichloroethylene (EPA 2014).

These nine air toxics were the only ones that met the adjustment criteria described above at the time of this assessment. The overall lifetime adjustment was applied because a single, lifetime-average EC was estimated for NATA rather than age-group-specific exposures. Note that the IRIS assessment contains vinyl chloride UREs for both exposure from birth and exposure during adulthood. However, the overall vinyl chloride URE already includes exposure from birth. Thus, we used that value with no additional factor applied.

5.4.6. Diesel Particulate Matter

Diesel PM (DPM) mass (expressed as μg DPM/m3) has historically been used as a surrogate measure of exposure for whole diesel exhaust. Although uncertainty exists as to whether DPM is the most appropriate parameter to correlate with human health effects, it is considered a reasonable choice until more definitive information about the mechanisms of toxicity or mode(s) of action of diesel exhaust becomes available.

In EPA's 2002 Diesel Health Assessment Document (Diesel HAD), exposure to diesel exhaust was classified as likely to be carcinogenic to humans by inhalation from environmental exposures, in accordance with the revised draft 1996/1999 EPA cancer guidelines. Several other agencies (National Institute for Occupational Safety and Health, the International Agency for Research on Cancer, the World Health Organization, California EPA and the U.S. Department of Health and Human Services) had made similar hazard classifications prior to 2002. EPA also concluded in the 2002 Diesel HAD that it was impossible to calculate a cancer unit risk for diesel exhaust due to limitations in the exposure data for the occupational groups or the absence of a dose-response relationship.

In the absence of a cancer unit risk, the Diesel HAD sought to provide additional insight into the significance of the diesel exhaust cancer hazard by estimating possible ranges of risk that might be present in the population. An exploratory analysis was used to characterize a possible risk range, and found that environmental risks from diesel exhaust exposure could plausibly range from a low of 10^{-5} to as high of 10^{-3} for long-term exposures. Because of uncertainties, the analysis acknowledged that the risks could be lower than 10^{-5}, and a zero risk from diesel exhaust exposure was not ruled out.

Noncancer health effects of acute and chronic exposure to diesel exhaust emissions are also of concern to EPA. The agency derived a diesel exhaust reference concentration (RfC) after considering four well-conducted chronic rat inhalation studies showing adverse pulmonary effects. The RfC is 5 μg/m^3 for diesel exhaust measured as diesel particulate matter. This RfC does not consider allergenic effects such as those associated with asthma, immunologic effects or the potential for cardiac effects. Emerging evidence in 2002, discussed in the Diesel HAD, suggested that exposure to diesel exhaust could exacerbate these effects, but the exposure-response data were lacking at that time to derive an RfC.

The Diesel HAD noted that the cancer and noncancer hazard conclusions applied to the general use of diesel engines then on the market, and as cleaner engines replace a substantial number of existing ones, the applicability of the conclusions would need to be reevaluated.

Several studies published since 2002 continue to report increased lung cancer with occupational exposure to older engine diesel exhaust. Of note since 2011 are three new epidemiology studies that examined lung cancer in occupational populations (for example, in truck drivers, underground nonmetal miners and other diesel motor related occupations). These studies (Garshick et al. 2012; Silverman et al. 2012; Olsson et al. 2011) reported increasing risks with exposure to diesel exhaust, and positive exposure-response relationships were also evident to varying degrees. These newer studies (along with others that have appeared in the scientific literature) add to the evidence EPA evaluated in the 2002 Health Assessment Document and further reinforce the lung cancer hazard concern. The findings from these newer studies do not necessarily apply to newer technology diesel engines since the newer engines have large reductions in the emission constituents.

In June 2012, the World Health Organization's International Agency for Research on Cancer (IARC) evaluated the full range of cancer-related health-effects data for diesel engine exhaust (IARC 2014). IARC concluded that diesel exhaust should be regarded as "carcinogenic to humans." This designation was an update from its 1988 evaluation that considered the evidence to be indicative of a "probable human carcinogen." IARC is a recognized international authority on the carcinogenic potential of chemicals and other agents.

Also in 2012, EPA and industry asked the Health Effects Institute to convene a panel to review recently published epidemiology studies of occupational exposures to diesel engine exhaust. The request was in part to determine whether new studies could be used in a quantitative risk assessment (QRA) to calculate a cancer URE. In a final report published at the end of 2015 (HEI 2015), the panel concluded that newer studies made considerable progress toward addressing a number of major limitations in previous epidemiologic studies of diesel engine exhaust. It further stated that, although uncertainties still remain, these newer studies provide a useful basis for potentially conducting a QRA of diesel engine exhaust exposures, but specifically to diesel engine exhaust from older diesel engines. Currently, there are no ongoing activities at EPA related to conducting a QRA for diesel engine exhaust.

5.5. Summary

- To evaluate the potential of a given air toxic to cause cancer and other adverse health effects, we identified potential adverse effects that the substance causes and evaluated the specific ECs at which these effects might occur.

- The URE represents the upper-bound excess cancer risk estimated to result from continuous exposure to a concentration of 1 μg of a substance per m^3 of air over a 70-year lifetime.

- The RfC is an estimate of a continuous inhalation EC over a 70-year lifetime that is thought to be without an appreciable risk of adverse effects. The population considered in the derivation of RfCs includes sensitive subgroups (i.e., children, asthmatics and the elderly).

- Dose-response-assessment information for chronic exposure was obtained from multiple sources and prioritized according to conceptual consistency with OAQPS risk assessment guidelines and level of peer review.

- After considering dose-response information, EPA adjusted some chronic-toxicity values to increase accuracy and to avoid underestimating risk.

6. Characterizing Risks and Hazards in NATA

Risk characterization, the final step in our risk assessment process for air toxics, combines the information from modeled exposure estimates with the dose-response assessment. The result is a quantitative estimate of potential cancer risk and noncancer hazard associated with real-world exposure to air toxics. The term "risk" implies a statistical probability of developing cancer over a lifetime. Noncancer "risks," however, are not expressed as a statistical probability of developing a disease. Rather, noncancer "hazards" are expressed as a ratio of an exposure concentration (EC) to a reference concentration (RfC) associated with observable adverse health effects (i.e., a hazard quotient).

This section contains information on the risk characterization conducted for NATA. After a brief overview of the risk-related questions that NATA is intended to address, the methods used to characterize cancer risk and noncancer hazards for NATA are described. A discussion of the quantitative results included in NATA follows this description.

6.1. The Risk-characterization Questions NATA Addresses

NATA risk characterization considers both cancer risk and the potential for noncancer effects from inhalation of air toxics nationwide, in both urban and rural areas. The purpose of NATA is to understand cancer risks and noncancer hazards to help EPA and others identify air toxics and source categories of greatest potential concern and to set priorities for collecting additional data. The assessment represents a "snapshot" in time for characterizing risks from exposure to air pollutants; it is not designed to characterize risks sufficiently for regulatory action. The risk characterization for NATA, which was limited to inhalation risk from outdoor sources, was designed to answer the following questions:

- Which air toxics pose the greatest potential risk of cancer or adverse noncancer effects across the entire United States?

- Which air toxics pose the greatest potential risk of cancer or adverse noncancer effects in specific areas of the United States?

- Which air toxics pose less, but still significant, potential risk of cancer or adverse noncancer effects across the entire United States?

- When risks from inhalation exposures to all outdoor air toxics are considered in combination, how many people could experience a lifetime cancer risk greater than levels of concern (such as 1-in-1 million or 100-in-1 million)?

- When potential adverse noncancer effects from long-term exposures to all outdoor air toxics are considered in combination for a given target organ or system, how many people could experience exposures that exceed the reference levels intended to protect against those effects (i.e., a hazard quotient greater than 1)?

6.2. How Cancer Risk Is Estimated

To estimate cancer risks in NATA, the results of cancer dose-response assessments for a given chemical were converted to a unit risk estimate (URE). That URE was then multiplied by the estimated inhalation

exposure concentration to obtain an estimate of individual lifetime cancer risk. The approach used in NATA for characterizing cancer risk is consistent with EPA's 2005 final [Guidelines for Carcinogen Risk Assessment](#) (EPA 2005a). When used with the cancer UREs described in Section 5, the approach is also consistent with EPA's [Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens](#) (EPA 2005b).

6.2.1. Individual Pollutant Risk

In NATA, individual lifetime cancer risk associated with exposure to a single air pollutant was estimated by multiplying an average estimated long-term exposure concentration by the corresponding URE for that pollutant. Thus, the equation below estimates the probability of an individual developing cancer over a lifetime from the exposure being analyzed due to a given inhalation exposure, over and above that due to any other factors.

$$Risk = EC \times URE$$

Where:

> *Risk* = estimated incremental lifetime cancer risk for an individual due to exposure to a specific air toxic, unitless (expressed as a probability)

> *EC* = estimate of long-term inhalation exposure concentration for a specific air toxic, in units of $\mu g/m^3$

> *URE* = the corresponding inhalation unit risk estimate for that air toxic, in units of $1/(\mu g/m^3)$

Note that UREs are typically upper-bound estimates, so actual risks may be lower than predicted. Also, the true value of the risk is unknown.

6.2.2. Multiple-pollutant Risks

EPA estimates individual lifetime cancer risks resulting from exposure to multiple air toxics by summing the chronic cancer risk for each air toxic that can be quantified. This estimate of risk focuses on the additional lifetime risk of cancer predicted from the exposure being analyzed, over and above that due to any other factors. The following equation estimates the predicted cumulative individual cancer risk from inhalation of multiple substances:

$$Risk_{tot} = Risk_1 + Risk2 + ...+ Risk_i$$

Where:

> $Risk_{tot}$ = total cumulative individual lifetime cancer risk, across *i* substances

> $Risk_i$ = individual risk estimate for the *i*th substance

In NATA, the estimated ECs are not considered upper bounds. Rather, they represent central-tendency estimates of ECs at the census tract level. Because cancer slope factors are 95-percent upper-confidence

intervals (not "most probable estimates"), summing traditional risk levels can cause the resulting sum to overestimate a 95-percent upper-confidence-level risk for a mixture.

In NATA, we assume that exposures to multiple carcinogens can be added together to estimate risks. This approach has drawbacks: Effects from multiple chemicals may be greater or less than additive, and statistical limitations exist. But this straightforward calculation is widely used to estimate cumulative risks, especially in screening assessments like NATA.

Information on non-additive interactions is not readily available in a form that can be used for NATA. Without this specific information, cancer risk from various chemicals is conservatively assumed to be additive. Thus, the cancer risks from all air toxic compounds listed as carcinogenic or likely carcinogenic to humans were summed to determine cumulative cancer risks for NATA. More information on EPA's methods for conducting risk assessment of mixtures can be found in the Framework for Cumulative Risk Assessment (EPA 2003b).

6.3. How Noncancer Hazard is Estimated

To evaluate the potential for noncancer adverse health effects, EPA uses chronic dose-response data to estimate a reference concentration (RfC) – the EC at which adverse health effects are assumed to be unlikely. (See Section 5.2.2 5.2.2for more information on noncancer RfCs.) Due to the wide variety of endpoints, hazard-identification procedures for noncancer effects have not been described as completely in EPA guidance as those for carcinogens. But EPA has published guidelines for assessing several

EPA's Chronic Noncancer Guidelines
• Mutagenicity (EPA 1986)
• Developmental Toxicity (EPA 1991)
• Neurotoxicity (EPA 1998)
• Reproductive Toxicity (EPA 1996)

specific types of chronic noncancer effects (mutagenicity, developmental toxicity, neurotoxicity and reproductive toxicity). These can be found at Products and Publications Relating to Risk Assessment Produced by the Office of the Science Advisor (EPA 2016). EPA has also published a framework for using studies of these and other effects in inhalation risk assessment (EPA 1994).

6.3.1. Individual Pollutant Hazard

EPA estimated chronic noncancer hazards for NATA by dividing a chemical's estimated long-term EC by the RfC for that chemical to yield a hazard quotient (HQ). The following equation estimates the noncancer hazard due to a given inhalation exposure:

$$HQ = \frac{EC}{RfC}$$

Where:

HQ　　=　　the hazard quotient for an individual air toxic, unitless

EC　　=　　estimate of long-term inhalation exposure concentration for a specific air toxic, in units of mg/m^3

RfC　　=　　the corresponding reference concentration for that air toxic, in units of mg/m^3

An HQ value less than or equal to 1 indicates that the exposure is not likely to result in adverse noncancer effects. An HQ value greater than 1, however, does not necessarily suggest a likelihood of adverse health effects and cannot be interpreted to mean that adverse health effects are statistically likely to occur. The statement is simply whether, and by how much, an EC exceeds the RfC, indicating that a potential exists for adverse health effects.

6.3.2. Multiple-pollutant Hazard

We estimated chronic noncancer hazards for multiple air toxics by summing chronic noncancer HQs for individual air toxics that cause similar adverse health effects. The result is a hazard index (HI). Aggregation in this way produces a target-organ-specific HI, defined as a sum of HQs for individual air toxics that affect the same organ or organ system. More information on chemical mixtures risk assessment methods can be found in the EPA supplementary guidance for risk assessment of mixtures (EPA 2000).

The following equation estimates the HI from inhalation of multiple substances:

$$HI = HQ_1 + HQ_2 + \dots + HQ_i$$

Where:

> HI = the hazard index for chronic exposure to air toxics 1 through i, unitless
>
> HQ_i = the hazard quotient for the ith air toxic, where all i air toxics are assumed to affect the same target organ or organ system, unitless

As with the HQ, an HI value less than or equal to 1 indicates that the exposure is not likely to result in adverse noncancer effects. An HI value greater than 1, however, does not necessarily suggest a likelihood of adverse health effects and cannot be interpreted as a statistical probability of adverse effects occurring.

This equation assumes an additive effect from simultaneous exposures to several chemicals. Summing HQs is inappropriate when effects from multiple chemicals are synergistic (greater than additive) or antagonistic (less than additive). As is the case with cancer risk, quantitative information on non-additive interactions resulting in noncancer hazards is not readily available; consequently, the noncancer HQs are assumed to be additive for chemicals with the same target organ or organ system.

For NATA, we report HIs for 14 target organs or systems. Results indicate that respiratory hazards are the primary noncancer hazards for inhalation exposures to the modeled chemicals.

6.4. How Risk Estimates and Hazard Quotients Are Calculated for NATA at Tract, County and State Levels

The cancer risk and HQs for each modeled air toxic are estimated from ECs (not ambient concentrations) by combining them with UREs and inhalation RfCs (or their equivalents). As described previously, the modeling conducted for NATA resulted in ambient concentrations for each air toxic emitted by modeled sources, with the level of spatial resolution varying by source type and the corresponding modeling approach (see Section 3).

6.4.1. Model Results for Point Sources: Aggregation to Tract-level Results

The NATA modeling step generates ambient concentrations at the block level. For risk and exposure calculations, we aggregated these results from the block level to the tract level by taking a population-weighted average of all block-level concentrations within a given tract, as follows:

$$Conc_{tract_i} = \frac{\sum Conc_{block_j} \times Pop_{block_j}}{\sum Pop_{block_j}}$$

Where:

$Conc_{tracti}$ = ambient concentration for census tract *i*

$Conc_{blockj}$ = ambient concentration for census block *j* (contained within tract *i*), estimated by AERMOD

Pop_{blockj} = population of blocks contained in tract *i*

6.4.2. Background Concentrations and Secondary Pollutants: Interpolation to Tract-level Results

Background concentrations, as well as estimated concentrations of secondary pollutants generated by the CMAQ model, were estimated for levels other than census tract and thus required interpolation "down" to the tract level. Background concentrations were estimated at the county level. To obtain tract-level concentrations, the county-level estimate was assigned to all census tracts within that county. For secondary pollutants, concentrations were estimated using CMAQ. The results for each grid were then applied evenly to all tracts located within the grid.

6.4.3. Aggregation of Tract-level Results to Larger Spatial Units

Tract-level ambient concentrations were aggregated up to the county, state and national levels using a method that weights concentration according to the population within a region. For a county, for example, a population-weighted ambient concentration was estimated by multiplying the tract-level concentrations by the population of each tract, summing these population-weighted concentrations, and dividing by the total county population encompassing all tracts to obtain a final population-weighted, county-level concentration. The process for aggregating from the tract to the county level can be expressed using the following equation:

$$Conc_{county_k} = \frac{\sum Conc_{tract_i} \times Pop_{tract_i}}{Pop_{county_k}}$$

Where:

$Conc_{countyk}$ = population-weighted concentration for county *k*

$Conc_{tracti}$ = ambient concentration in tract *i* (contained within county *k*)

Pop_{tracti} = population in tract *i* (contained within county *k*)

$Pop_{countyk}$ = population in county *k*

This same method was applied when aggregating up to the state or national level, using the appropriate concentration and population values. NATA results include ambient concentrations, ECs, cancer risks, and noncancer HQs at the tract, county, state and national levels.

The ambient concentrations derived at the block level also were used to estimate ECs using either direct exposure modeling with HAPEM or with the exposure factors derived from HAPEM modeling (i.e., ratios of EC to estimated ambient concentration). (See Section 4 for a more thorough discussion of NATA exposure modeling and estimates.) Because the exposure factors were applied at the tract level, each census block was assigned the tract-level EC or exposure factor and then the census block-level ECs are estimated. As was done with the ambient-level concentrations, the block-level ECs were used to estimate cancer and noncancer effects and to aggregate these concentrations up to larger spatial scales. To aggregate tract-level concentrations up to the county-, state- or national-level concentrations, the tract-level concentrations were population-weighted.

6.5. The Risk Characterization Results That NATA Reports

NATA provides a snapshot of the outdoor air quality and the risks to human health that would result if air toxic emission levels remain unchanged. The assessment was based on an inventory of air toxics emissions from 2014. Individuals were assumed to spend their entire lifetimes exposed to these air toxics. Therefore, we did not account for the reductions in emissions that have occurred since the year of the assessment, or those that might happen in the future due to regulations for mobile and industrial sources.

Each NATA represents an update and enhancement to the previous NATA. Because improvements in methodology are made with each successive assessment, comparing assessment results from year to year is not meaningful. Any change in emissions, ambient concentrations or risks might be due to improvement of methodology, to real changes in emissions or source characterization, or both.

The evaluation of national-scale results and comparison of risks among chemicals make it possible to estimate which air toxics pose the greatest potential risk to human health in the United States. NATA reports a summary of these findings. Cancer risks are presented as lifetime risks, meaning the risk of developing cancer due to inhalation exposure to each air toxic compound over a normal lifetime of 70 years. Noncancer hazards are presented in terms of the ratio between the exposure and an RfC for inhalation exposures (i.e., the HQ). As described previously in this section, HQs are combined across chemicals where a common target organ or system is expected to estimate HI (i.e., for respiratory).

Using these quantitative results, NATA classifies certain pollutants as drivers or contributors at the national or regional scale based on certain criteria. Table 6-1 contains the criteria for classifying the air toxics included in NATA at the regional and national level. In general, drivers and contributors were defined as air toxics showing a particular level of risk or hazard for some number of people exposed.

For example, for a pollutant to be categorized in NATA as a cancer contributor at the national level, the individual lifetime cancer risk for that pollutant must have been shown by the assessment to be at least 1-in-1 million and the number of people exposed to that pollutant must have been shown to be at least 25 million. For a pollutant to be categorized in NATA as a regional driver of noncancer health effects, the chronic hazard index for that pollutant must have been shown to exceed 1 and the number of people exposed to that pollutant must have been shown to be at least 10,000.

The NATA results for 2014 indicated that most individuals' estimated risk was between 1-in-1 million and 100-in-1 million, although a small number of localized areas showed risks of higher than 100-in-1 million. Although individuals and communities may be concerned about these results, recall that NATA is not designed to assess specific risk values at local levels. The results are best used as a tool to prioritize pollutants, emissions sources and locations of interest for further investigation.

Table 6-1. Criteria establishing NATA drivers and contributors of health effects for risk characterization

Risk-characterization Category	Criterion (Criteria in both columns must be met)	
	Individual Health Risk or Hazard Index Exceeds…	Minimum Number of People Exposed (in millions) is…
Cancer Risk *(value in first column represents individual lifetime cancer risk, in 1 million)[a]*		
National cancer driver	10	25
Regional cancer driver	10	1
(either set of criteria can be used)	100	0.01
National cancer contributor	1	25
Regional cancer contributor	1	1
Hazard Index *(value in first column represents chronic hazard index for any organ/organ system)[b]*		
National noncancer driver	1	25
Regional noncancer driver	1	0.01

[a] Cancer risks are upper-bound lifetime cancer risks; that is, a plausible upper limit to the true probability that an individual will contract cancer over a 70-year lifetime as a result of a given hazard (such as exposure to a toxic chemical). This risk can be measured or estimated in numerical terms (e.g., one chance in a hundred).
[b] Hazard index is the sum of the HQs for substances that affect the same target organ or organ system. Because different pollutants can cause similar adverse health effects, combining HQs associated with different substances is often appropriate to understand the potential health risks associated with aggregate exposures to multiple pollutants.

Furthermore, the risks estimated by the assessment do not consider indoor sources of air toxics or ingestion exposure to any pollutants. Also, although NATA estimates cancer and noncancer risks for numerous pollutants, additional chemicals might exist that are not identified or for which toxicity information is unavailable. Therefore, these risk estimates represent only a subset of the total potential cancer and noncancer risk associated with air toxics.

Analytical results (including modeled ambient concentrations, exposure and risks) for each NATA are also provided at the census tract, county and state levels for those who wish to do their own technical analyses using the most refined output available. In performing such analyses, users must be extremely mindful of the purposes for which NATA was developed. NATA was developed as a screening tool to inform both national and more localized efforts to collect air toxics information and characterize emissions (e.g., prioritize pollutants or areas of interest for more refined data collection such as monitoring). The results are most meaningful when viewed at the state or national level. Nevertheless, reported spatial patterns within a county likely represent actual variations in *overall average* population risks. Less likely, however, is that the assessment pinpoints the exact locations where higher risks exist or that the assessment captures the highest risks in a county.

Using these results alone to draw conclusions about local concentrations and risk is inappropriate. This assessment does not attempt to identify areas or populations that have significantly higher risks than others. Rather, it focuses on characterizing geographic patterns and ranges of risk across the country. In general, however, spending time in larger urban areas tends to pose greater risks than spending time in

smaller urban and rural areas because the emissions of air toxics tend to be higher and more concentrated in areas with more people. This trend is not universal, however, and can vary from pollutant to pollutant according to its sources. The trend also can be affected by exposures and risk from non-inhalation and indoor sources of exposure.

Based on the NATA results, millions of people live in areas where air toxics pose potential health concerns. Although air quality continues to improve, more needs to be done to meet the CAA's requirements to reduce the potential exposure and risk from these chemicals. EPA will continue to develop air toxic regulations and cost-effective pollution prevention and other control options to address indoor and urban pollutant sources that significantly contribute to risk.

You can access all NATA results via the NATA website (https://www.epa.gov/national-air-toxics-assessment).

6.6. Summary

- The purpose of NATA is to understand cancer risk and noncancer health effects to help EPA and others identify pollutants and source categories of greatest potential concern and to set priorities for collecting additional information to improve future assessments.

- Cancer risk is expressed as a statistical probability that an individual will develop cancer. Cancer risks were assumed to be additive across chemicals for NATA.

- Noncancer hazard is expressed as an HQ, which is the ratio of the EC to an RfC associated with observable adverse effects. Noncancer hazards were assumed to be additive across chemicals that effect the same target organ or organ system.

- NATA estimates most individuals' cancer risk to be between 1-in-1 million and 100-in-1 million, although a small number of localized areas showed risk higher than 100-in-1 million.

- Air toxics data for NATA are presented at the national, state, county and census tract levels. The results are most meaningful when used to identify patterns of risk over larger areas. Using these results in the absence of additional information to draw conclusions about local concentrations and risk is inappropriate.

7. Variability and Uncertainty Associated with NATA

7.1. Introduction

This section discusses the variability and uncertainty associated with NATA. Clearly understanding these two concepts – found in all broad-scale assessments that rely on models and data – will help you understand which questions NATA can answer appropriately, and which it cannot.

As stated in Section 1, NATA results should not be used for limited-scale or site-focused applications. NATA results are intended to characterize broad-scale risk to help identify those air toxics and source types associated with the highest exposures and posing the greatest potential health risks. The results are intended to identify geographic patterns and ranges of risks across the country. To avoid over-interpretation and misapplication of the results, users must first understand the concepts of variability and uncertainty and then must recognize the role that these elements play in NATA results.

> **Key Definitions for this Section**
>
> **Variability** represents the *diversity or heterogeneity in a population or parameter* (e.g., variation in heights of people). Variability cannot be reduced by taking more (or better) measurements; however, it can be accounted for by a more detailed modeling approach (e.g., modeling peoples' heights in terms of age will reduce the unexplained variability due to variation in heights).
>
> **Uncertainty** refers to the *lack of knowledge* regarding the actual values of model input variables (parameter uncertainty) and of physical systems (model uncertainty). Uncertainty can be reduced through improved measurements and improved model formulation.

Air toxic emissions, air concentrations and exposures are not the same throughout the United States, and the risks associated with air toxics are not the same for all people. Some areas have higher concentrations than others. At certain times, the concentration is higher at a given location than at other times. The risks for some individuals are below the national average, while for others the risks are above the national average. For these reasons, understanding how the ambient (outdoor) air concentration, exposure and risks from air toxics vary throughout the United States is essential for understanding NATA. This information comes from a process called variability analysis.

EPA seeks to protect health with reasonable confidence based on the best data available. Estimates of air concentrations, exposures and risks, however, must always involve assumptions. Assumptions are necessary to simplify the problem at hand. They make NATA possible given available information and resources. That said, assumptions introduce uncertainties into NATA results because we can never be fully confidence that the assumptions are entirely correct. Understanding the extent of these uncertainties, the level of confidence that can be placed in statements related to the assessment, and how this confidence affects the ability to make reasoned decisions is essential. This information comes from a process called uncertainty analysis.

7.2. How NATA Addresses Variability

The NATA process focuses on the variation in ambient air concentrations, exposures and risks across the United States, Puerto Rico and the U.S. Virgin Islands. Included, for example, are variations in source locations and the amounts of pollutants these sources emit, variations in meteorological conditions across the country, and variations in the daily activities of people. This section presents information on the key components that drive variability in risks associated with air toxics and the variability

components that NATA addresses. A brief explanation is also provided on how NATA results should be interpreted given variability.

7.2.1. Components of Variability

NATA results show how air concentrations, exposures and risks vary across broad areas of the country. They do not fully characterize how concentration, exposure and risk vary among individuals, except to the extent these individuals live in different geographic regions and are affected by the values typical of a census tract in that region. NATA results also do not fully characterize how ambient air concentrations might vary temporally and they do not show how concentrations vary spatially within a census tract. The following list explains some of the components of variability that determine differences in ambient air concentrations and individual risks.

Temporal. Sources do not emit pollutants at constant rates. Similarly, the meteorological conditions that affect dispersion in the atmosphere vary over time. Thus, the ambient air concentration at a given location can vary over time.

Geographic. The influence of pollutant emissions on a location's ambient concentration depends on the degree of atmospheric dispersion of the emissions as they travel from the source to the location. Dispersion depends on both meteorological conditions, which vary from place to place, and the travel distance from the source. As a result, the ambient air concentration can vary greatly among different locations. The NATA analysis accounts for some geographic variation by using available meteorology data representative of the location and by modeling ambient concentrations for census areas, but the spatial resolution of model predictions is limited.

Individual location. Two individuals might live at different locations within the same census tract. The ambient concentration estimated for the tract is only an approximation of conditions at all locations in the tract. Different locations within that tract may have different average ambient concentrations. Therefore, exposures and risks may also vary.

Individual activity patterns. Two people may live at the same location but engage in different activities (called an "activity pattern") during each day. Concentrations of substances indoors often differ from concentrations outdoors. If one person spends more time indoors than the other person, the average air concentration to which the two are exposed will differ, even though the ambient air concentration is the same. Similarly, one person may spend more time in a car and be exposed to an air concentration that is typical near roads. The net effect would be that the concentration of each pollutant in the air inhaled by these two people would differ. In other words, their exposure differs.

In addition, the amount of outdoor pollution that penetrates into buildings and vehicles varies due to differences in ventilation and structural integrity. Thus, two people who live in the same location and spend the same amount of time indoors or in vehicles can still be exposed to different pollutant concentrations.

Susceptibility. Two individuals may live at the same location and engage in the same activities, but one person may be more susceptible than another. Susceptibility refers to the extent to which an individual takes a pollutant into the body, transports it into an organ or tissue that might be adversely affected by it, or develops an adverse effect.

A more susceptible person may develop a higher concentration of a pollutant in his or her organs or tissues, or have a higher chance of developing an adverse health effect, than another individual, even when exposures for both individuals are the same. For example, people breathe at different rates; two people breathing the same air may bring different amounts of a pollutant into their bodies. The amount of a pollutant reaching an organ or tissue can also vary between individuals, even if both bring the same amount into their lungs. The amount of time the pollutant remains in the body may also differ. Finally, the innate sensitivity to the effect may vary, even at equal doses in the tissues. The net effect of these factors is that either the dose of the pollutant delivered to the organs or tissues of the body or the level of response (or both) can differ substantially between two people even if they are exposed to the same pollutant concentrations.

The extent to which each factor described above influences variation in individual risk can depend on the age, gender or ethnic group to which an individual belongs, as well as on that individual's lifestyle.

These groups comprise different receptor populations, or cohorts, and the exposures and risks can differ among them.

7.2.2. Quantifying Variability

EPA conducts NATA to understand how ambient air concentrations, exposures and risks vary geographically – not among specific individuals. EPA calculates the ambient air concentrations for each specific, discrete location (i.e., census-block centroid or census-tract centroid; see discussion below) based on the emission sources and meteorological conditions affecting those specific tracts. Some temporal variation is accounted for in NATA calculations. For example, meteorological data used for air quality modeling is temporally dynamic. The air quality modeling therefore captures important variations in ambient conditions on an hourly basis before the resulting modeled ambient air concentrations are time averaged. The ambient concentration inputs to HAPEM are stratified into eight 3-hour time blocks; HAPEM then calculates ECs for each 3-hour time block before calculating an overall, long-term average EC. Although this approach to air quality and exposure modeling considers some important temporal variations, these time-stratified model outputs are averaged prior to the risk characterization step and are not included in the NATA results reported by EPA.

The NATA concentrations and risks, however, do reflect a degree of geographic variation. The smallest geographic area for which NATA results are reported is the census tract. Although results are reported at the census tract level, average risk estimates are far more uncertain at this level of spatial resolution than at the county or state level. Census tracts are small, relatively permanent statistical subdivisions of a county, typically having between 1,200 and 8,000 residents, with most having close to 4,000. Census tracts do not cross county boundaries. Their areas vary widely depending on the density of settlement. Census tracts tend to be small in densely populated areas but can be very large in sparsely populated areas.

Within census tracts are census blocks, the smallest geographic areas that the Census Bureau uses. Blocks are bounded by visible or virtual features such as streets, streams, and city or town boundaries. Census blocks are typically small in area and population; for example, in an urban area, a census block might correspond to a block bounded by city streets. In remote areas, however, census blocks might be large and irregular, comprising many square miles. Blocks typically have approximately 50 residents.

Air concentrations are estimated in NATA at various levels of resolution depending upon the source type modeled. Secondary formation, fires and biogenics (modeled in the CONUS) are at 12-km grid cell resolution. Other sources use census block resolution, though the emissions for some sources are at the grid cell level – these grid cell-level emissions originate from even broader geographic scales (county and national level) and are less certain at these finer scales, as discussed below.

For a given source type and modeling approach, variation in ambient air concentrations within a grid cell or census block is not explicitly modeled. For estimates at the block level, a representative ambient air concentration is estimated for a single location near the center of the block (i.e., the centroid, which is typically, but not always, the geographic center of the block chosen by the U.S. Census Bureau as a reference point). EPA then averages ambient concentrations estimated at the block level for the encompassing census tract, with concentration and risk results reported at the tract level. Assessment results do not reflect variations in the susceptibility of people within a census tract because the focus is to compare typical exposures and risks in different tracts. As a result, individual exposures or risks might differ by as much as a factor of 10 in either direction. You should consider exposures or risks determined in NATA as representative of the geographic area where an individual lives, but not as that individual's personal risk.

Thus, the results of the NATA analysis do not allow for a comparison of ambient air concentrations, exposures or risks between two individuals. They do, however, enable you to understand the variation in typical values for these quantities among counties or states and to a lesser degree among census tracts. For an individual, however, the values may differ from the typical value for the county or state if that individual lives in a part of the area that has a higher or lower than typical value, has an activity pattern that causes a higher or lower exposure than is typical, or is more (or less) susceptible than a "typical" person used in this assessment.

For the purposes of estimating and reporting risk, EPA assumes that individuals within a census tract have the same exposure and risk. This assumption allows NATA users to examine variation in individual exposure among census tracts, but not the variation within a census tract. Activity patterns are included for each of six cohorts defined by age. Even within a receptor population, some variability in activity patterns among individuals is considered. Differences in susceptibility, however, are not included in NATA. EPA took this approach for NATA for two primary reasons:

- **An overall purpose of NATA is to examine broad differences driven by geography.** NATA considers only geographic differences in pollutant concentration, exposure and risk. The goal is to understand how these three factors differ among people living in different geographic areas. EPA assesses these differences, as mentioned above, by tracking differences in air concentration in different census tracts, producing differences in the typical pollutant concentrations, exposures and risks in different tracts. Differences in susceptibility, however, can produce differences in risk between two individuals in the same census tract, and reporting on these differences is not a purpose of NATA.

- **The variability in susceptibility is difficult to model at the national scale.** Very limited information is available on differences in susceptibility among individuals. Even if EPA were to choose to calculate and report differences among individuals in a census tract, scientifically reliable information necessary to produce these calculations is not available for many of the pollutants. Given current information, it may be possible to estimate variability in the rates at which people breathe, but this is only a small part of the overall variation in susceptibility. EPA therefore has chosen not to incorporate this source of variation between individuals.

Considering these limitations, EPA elected to incorporate differences in emissions and meteorology (resulting in differences in ambient air concentration) and differences in location of typical individuals (resulting in differences in exposure) among census tracts. Variation in activity patterns for different age groups is reflected in the assessments to the degree that the age of residents varies by location. Variability in susceptibility is not included for the reasons given above. We address temporal variation in inputs by developing time-weighted averages of emissions characteristics, meteorological conditions and ECs. Temporal variation in the estimated ambient air concentrations, however, is not reflected in the results (only time-weighted annual averages are presented).

7.2.3. How Variability Affects Interpretation of NATA Results

The NATA analysis illustrates how ambient air concentration, exposure and risk vary throughout the United States. The assessment does not focus on the variation in exposure and risk among individuals. It focuses on variation among well-defined geographic areas, such as counties or states, based on calculations of ambient air concentration, exposure and risk in various census tracts. To a lesser degree, variation among demographic groups is also addressed by NATA in that differences in activity patterns are considered in modeling ECs using HAPEM. Risk results, however, are not presented separately for individual demographic groups.

The information contained in the maps, charts and tables produced in NATA display predictions of cancer risk and noncancer hazard. Cancer risk results include statements such as:

> "X percent of the census tracts in a given area are characterized by a typical lifetime excess cancer risk of less than R."

For this statement, if X is 25 percent and R is 1-in-1 million, the result would be:

> "25 percent of the census tracts are characterized by a typical risk of less than 1-in-1 million."

This statement does not necessarily mean that 25 percent of individuals in the specified area have a cancer risk of less than 1-in-1 million. Some people in these census tracts would be expected to have a risk above 1-in-1 million. Although a person may live in a census tract where the typical or average risk is less than 1-in-1 million, that person may live nearer the source than the average person in the census tract, may have an activity pattern that leads to greater exposure, or may be more susceptible. All these factors could cause that individual to experience a risk above the typical value for that census tract. Conversely, the individual could have a lower risk by living farther from the source, having an activity pattern that produces lower exposures, or being less susceptible.

The important point to remember when interpreting the maps and charts of the NATA analysis is that they show variation among values of ambient air concentration, exposure or risk in census tracts or larger areas such as counties. This presentation allows the user to identify geographic regions (counties or states) where these values are higher or lower than the aggregated national average for all census tracts. It does not allow users to identify individuals who have higher or lower values of ambient air concentration, exposure or risk. Nevertheless, individuals with a high risk are more likely to be in geographic regions characterized by a high risk than in those geographic regions characterized by a low risk. The same can be said for exposure (i.e., individuals with a high exposure are more likely to be found in geographic regions characterized by high exposure than in those regions characterized by low exposure).

7.3. How NATA Addresses Uncertainty

No scientific statement (in risk assessment or other areas of science) can be made with complete confidence. Risk estimates are always uncertain to some degree due to issues such as those discussed below. To maintain transparency and openness in the presentation of risk results, the party conducting a risk assessment must explain these uncertainties and how these uncertainties increase or decrease confidence.

The NATA analysis produces statements about variability in ambient air concentrations, exposures and risks across geographic regions for typical individuals, as described in Section 7.2. In this section, the discussion of uncertainty is intended to address the confidence with which these statements regarding variability can be made. Note that uncertainty does not prevent EPA from making a statement of risk, nor does it prevent EPA from taking reasonable actions. Uncertainty does require, however, that the nature of the uncertainty, and the implications for decisions, be understood so the degree of support for the statement can be correctly and properly interpreted.

7.3.1. Components of Uncertainty

Uncertainty arises from a variety of sources. To understand the sources of uncertainty affecting a risk assessment, it is instructive to consider the process by which a study such as NATA is performed, as described in the following sections.

Problem formulation. We must first define the problem to be addressed. For example, a question that might help define the problem could include, "Is the occurrence of adverse human health effects correlated with emissions from industrial facilities?" What the study is intended to address and how the results will be used should be clear at the outset. This initial step in the analysis introduces problem-formulation uncertainty. The purpose of NATA is described in Section 1 of this document, where the question addressed in the assessment is defined as precisely as possible (e.g., that the study is limited to estimates of health effects in human populations) and the reader is informed of the limitations of the assessment. The issue of problem-formulation uncertainty is not considered further in this document.

Defining the analysis components. This step describes what can influence the answer to the problem. In NATA, the multiple influences include emissions from a variety of sources (e.g., mobile, stationary, biogenic); atmospheric dispersion and chemistry; activity patterns for different cohorts; UREs and RfCs and other considerations. Where the science is poorly developed, the factors that must be included might not be clear. Resources may be limited, making the inclusion of all factors in the study infeasible. This step in the analysis, which results in the conceptual model for the assessment, introduces conceptual uncertainty. This issue is also addressed in the discussion of the limitations of NATA in Section 1, where the aspects of the problem that are (and are not) included in the study are addressed (e.g., that the study addresses inhalation of air toxics only). The issue of conceptual uncertainty is not considered further here.

Selecting models. All risk assessments use models. The NATA analysis uses a series of mathematical models. EPA uses models in NATA to produce the emissions inventory, to calculate ambient air concentrations, to calculate exposures and to calculate risks (for cancer and noncancer effects). All scientific models involve uncertainties because a model reduces a (potentially very complex) set of chemical, biological, physical, social or other processes to manageable algorithms that can be used to

perform calculations and make forecasts. The simplifications inherent in the development of a model introduce uncertainties.

Typically, more than one model is available to apply to a problem, and those models can produce different results. Thus, uncertainty is introduced as to which model, and which model results, should be used. As a simple example, NATA uses a linear statistical model to relate EC and cancer risk: Cancer risk equals the exposure (air concentration) multiplied by a URE. Uncertainty analysis involves asking a series of questions: Are we certain this linear relationship is correct? Could the relationship be quadratic (i.e., risk equals exposure multiplied by the square of the dose)? Could the relationship have a threshold (i.e., no risk is apparent until the exposure becomes sufficiently large)? What are the implications for estimates of risk if these different models are used? What are the implications for decisions if a clear choice among the models cannot be made?

This step in the analysis introduces model uncertainty. Judging model uncertainty can be both quantitative and qualitative. Qualitative issues involve the scientific plausibility of the model. Does the model include all important processes? Does it explain the phenomenon (e.g., atmospheric dispersion) well? Is the model well accepted in the scientific community – has it passed critical tests and been subject to rigorous peer review?

Quantitative issues involve comparing model results against sets of data (although this also involves issues of parameter uncertainty discussed in the next bullet). Does the model generally predict these data accurately? Are the predictions accurate to within a factor of 2; a factor of 4? What is the effect of any approximation methods used in the model?

Applying models. The models used in the NATA analysis require parameter inputs such as emission rates, stack heights, fractions of time spent indoors and UREs. Although models describe general relationships among properties of the real world (e.g., the linear relationship between exposure and cancer risk), parameters quantify these properties for specific cases (e.g., the numerical value of the URE for benzene). Parameters provide the numbers needed in the models. Various databases are available from which we can estimate these parameters, and the methods used to collect the data and to compile the databases introduce uncertainties. These factors all introduce parameter uncertainty.

Although parameter uncertainty has both quantitative and qualitative aspects, common practice is to characterize this source of uncertainty quantitatively, with some qualitative caveats. For example, parameter uncertainty might be characterized by a confidence interval, which states that the true value of the parameter (such as the stack height for a facility) probably lies somewhere between 40 and 60 meters or that the stack height is "known to be within" a factor of 1.2, or that the stack height is "accurate to within" 20 percent. Attached to this quantitative characterization of uncertainty will be a qualitative caveat such as "the estimate of this uncertainty is based on measurements made in 1990 at facilities similar to the one considered in this study, but a change in the design of stacks might have been made since 1990." This qualitative statement provides some idea of the confidence with which the quantitative assessment of uncertainty can be applied.

7.3.2. Components of Uncertainty Included in NATA

For this discussion, we have divided the uncertainties in NATA into three sources, based on the three steps leading from the estimate of emissions to the calculations of risk. Uncertainty in ambient air concentrations is due to uncertainty in the emissions estimates and in the air quality models. Uncertainty in exposure is due to uncertainty in activity patterns, locations of individuals

NATA Components that Include Uncertainty
• Ambient concentrations
• Exposure estimates
• Risk estimates

within a census tract, and microenvironmental concentrations as reflected in the exposure model. Finally, uncertainty in risk is due to uncertainty in the shape of the relationship between exposure and effects, the URE and the RfC. These three sources of uncertainty are discussed below.

Ambient air concentrations. Considering first the predictions of ambient air concentrations, the specific sources of uncertainty derive from the parameters for the following: emissions, stack data, particle sizes and reactivity, chemical speciation, terrain, boundary conditions, background concentration, meteorology and model equations. These sources of uncertainty are discussed briefly in this section.

Emissions parameters, including emission rates and locations of sources, are taken from the NEI database. The NEI is a composite of estimates produced by state and local regulatory agencies, industry and EPA. Some of these data were further modified during the NATA review. We have not fully assessed the quality of specific emission rates and locations in the NEI and resultant NATA emissions (e.g., industrial emissions from a specific census tract), although we have conducted reviews. Some of the parameter values may be out of date, errors might have been introduced in transcribing raw data to a computer file, and other data-quality issues may be present. Emission estimates use a variety of methods such as emission factors, material balances, engineering judgment and source testing. Some release point locations use an average facility location instead of the location of each specific unit within the facility. Release point parameters may be defaulted for some situations. Fugitive release parameters are not required and are defaulted where missing. In addition, TRI data does not provide release-point parameters other than identifying sources as "stack" or "fugitive"; the release parameters used historical defaults from previous inventories or new defaults.

Uncertainty also is inherent in the emission models used to develop inventory estimates. For example, we estimate county-level air toxic emissions from nonroad equipment by applying fractions of toxic total hydrocarbons to estimates of county-level hydrocarbons for gaseous air toxics and gas-phase PAHs and fractions of toxic particulate matter to estimates of county-level particulate matter for particle-phase PAHs. We use emission factors based on milligrams per mile for metals. The toxic fractions are derived from speciation data, based on limited testing of a few equipment types. The estimates of county-level total organic gases and particulates are derived from the EPA MOVES model. In MOVES, uncertainties are associated with emission factors, activity and spatial-allocation surrogates. National-level emissions for nonroad equipment in MOVES are allocated to the county level using surrogates, such as construction costs adjusted for geographic construction material cost (to allocate emissions of construction equipment) and employees in manufacturing (to allocate industrial equipment). Availability of more specific local data on equipment populations and usage will result in more accurate inventory estimates. For onroad sources, activity data either come from states or are allocated from states to counties using surrogates. Different surrogates are used for different emission types and geographic locations. These surrogates vary in quality from location to location.

For mobile and nonpoint sources, we typically allocate emissions rates from the county level to grid cells through a surrogate such as traffic data, land cover or other land use types (such as truck stops for heavy duty diesel idling emissions). This allocation introduces additional uncertainty because the data on the surrogates also have uncertainty, and the correlations between the surrogates and the emissions are imperfect.

The health effects of a pollutant depend on its chemical form when inhaled. For many sources, the emissions reported to the NEI or NEI database itself do not include information on chemical speciation of the pollutants of interest, but instead contains the total rate of pollutant emitted in all its forms. For example, chromium obtained from TRI or from some states is reported as chromium and are speciated into hexavalent and trivalent forms in an NEI augmentation step. We make assumptions about chemical speciation based on values estimated to be representative at such sources, considering information on source type, typical feedstock materials, knowledge of the process involved, or other relevant factors. Any one source, however, may have different values than the ones assumed.

The dispersion, or movement, of pollutants in the atmosphere is influenced by the topography of the area surrounding a source, which is characterized by terrain parameters. Although CMAQ model estimates consider topography, the AERMOD model estimates as implemented for NATA do not in all cases. AERMOD estimates for point sources consider topography, but the estimates for the emissions sources modeled as census-tract area sources do not because considering topography in the model requires a single source elevation, which is not always possible for large grid cell sources. Not accounting for terrain introduces uncertainty into predictions of ambient air concentrations, particularly in areas with hills or mountains.

Other sources of uncertainty in the modeling of ambient air concentrations are the values used for the boundary conditions in CMAQ and the background concentration estimates added to AERMOD concentrations for the non-CMAQ HAPs. These sources may include, for example, contributions from long-range transport of compounds from other counties and states. For more details on background concentrations, refer to the discussion in Section 3.

The model equations used in the air quality models represent another source of uncertainty. The AERMOD dispersion model uses steady-state Gaussian equations, which make several assumptions that simplify plume dispersion. The CMAQ model is more complex in its treatment of pollutant dispersion and atmospheric dynamics than AERMOD; nevertheless, many assumptions underlie its Eulerian approach to dispersion, which are outlined further in the science documentation for the CMAQ model.

While the hybrid approach of combining the CMAQ and AERMOD model output improves NATA's treatment of chemistry and transport, there are uncertainties in the implementation. The approach requires emissions and meteorological inputs to be consistent between the models. While we treated emissions as consistently as possible, some simplifications were necessary. The main difference was in the temporal treatment of the emissions. The temporal allocation used in AERMOD was not exactly the same as in CMAQ for the county-level sources, though average profiles based on the CMAQ temporal approach were developed for use in AERMOD. There were also differences in the spatial treatment of CMVs, though for the large vessels, AERMOD source characterization parameters were developed based on summaries of the CMAQ vertical distribution of emissions. In addition to inconsistencies in model inputs, the hybrid approach uses an AERMOD grid cell average to normalize the individual AERMOD concentrations within a grid cell. The AERMOD values are less representative of the true AERMOD

average in grid cells where there are fewer gridded receptors (i.e., nine gridded receptors were used in less-populated areas).

To help characterize the aggregate uncertainty of the predictions of the air quality models, EPA compared modeled concentrations to available monitoring data on ambient air quality. For each monitor-pollutant combination, we compared the predicted annual average concentrations at the monitor location to the sampled annual average concentrations. We present these comparisons in Section 3.7 of this document.

Measured concentrations were taken from EPA's Ambient Monitoring Archive, which includes National Air Toxics Trends Stations and state and local monitors reported to the Air Quality System. For the 2014 NATA, the exact locations of the monitors were used for the model-to-monitor comparison, an approach that increases accuracy over previous assessments. For more details about the model-to-monitor analyses for previous assessments, see Comparison of 1996 ASPEN Modeling System Results to Monitored Concentrations (EPA 2002c), Comparison of 1999 Model-Predicted Concentrations to Monitored Data (EPA 2006b), Comparison of 2002 Model-Predicted Concentrations to Monitored Data (EPA 2009) and Comparison of 2005 Model-Predicted Concentrations to Monitored Data (EPA 2010b).

Discrepancies between model predictions and concentration measurements can be attributed to five sources of uncertainty:

- emission characterization (e.g., specification of source location, emission rates and release characterization);

- meteorological characterization (e.g., representativeness);

- model formulation and methodology (e.g., characterization of dispersion, plume rise, deposition, chemical reactivity);

- monitoring; and

- boundary conditions/background concentrations.

Underestimates for some pollutants could be a result of the following:

- The NEI may be missing specific emission sources.

- Emission rates may be underestimated or overestimated due to emission-estimation techniques and/or spatial allocation of national estimates to county, and county estimates to grid cells.

- The accuracy of the monitor averages is uncertain; the monitors, in turn, have their own sources of uncertainty. Sampling and analytical uncertainty, measurement bias and temporal variation all can cause the ambient concentrations to be inaccurate or imprecise representations of the true atmospheric averages.

Exposure. Sources of uncertainty in the relationship between ambient air concentrations and ECs include those associated with microenvironmental factors and activity patterns. HAPEM calculates the EC in various microenvironments (e.g., indoors at home, in a car) based on inputs of predicted ambient air concentrations and microenvironmental factors. The factors are characterized as probability

distributions to reflect the variability found in air toxics measurements more fully. For many air toxics, the measurement studies needed to estimate microenvironmental factors are not available, so the values used are based on measurement studies of similar compounds in similar situations. This practice introduces uncertainty into the estimation of ECs for such compounds. In addition, even for air toxics with measurement studies, the estimated microenvironmental factors have some uncertainty because the number of such studies is limited. Furthermore, the uniform application of the microenvironmental factors to all census tracts introduces uncertainty by not accounting for possible geographic differences among tracts (e.g., different window-opening behavior, different levels of building integrity).

The activity-pattern sequences for individuals used in HAPEM are based on the Consolidated Human Activity Database (CHAD). As explained in Section 4.3.3 HAPEM algorithms consider the variability in activity patterns among individuals within a cohort-tract combination. They do this largely by addressing correlation between subsequent activity patterns assumed to occur for each cohort-tract combination. The representativeness of the daily diaries in CHAD is uncertain because they are a compilation of many studies, including some older studies and some for which the data are based on non-random sampling. It is also uncertain how well the model algorithms represent actual daily autocorrelation between types of activity. This latter issue, however, pertains only to the variability of the ECs across the demographic group and not the median EC, which is the concentration reported by NATA.

The commuting data used in HAPEM are based on an EPA analysis of information from a special study by the U.S. Census Bureau (2010). HAPEM uses this information, reflecting 2010 data, in coordination with the activity-pattern data to place an individual in either the home tract or work tract at each time step. These data introduce some uncertainty because they simplify commuting patterns to a pair of home and work census tracts and may not reflect certain details of some commutes (e.g., the additional census tracts encountered by commuters who travel to non-adjacent tracts or more complex commuting patterns that are not point to point). An additional important consideration is that the commuting-pattern data included in HAPEM do not account for the movement of school-age children who travel (or commute) to a school located outside the tracts in which they reside.

Risk. Concerning the predictions of risk, the specific sources of uncertainty in dose-response relationships (in addition to those considered for ambient air concentration and exposure) are hazard identification, dose-response models for carcinogens, UREs and RfCs.

One component of predicting risk is hazard identification. NATA's cancer-risk estimates assume that a compound either is a carcinogen or produces a noncancer effect. We base this on the results of a hazard-identification stage that assesses the evidence that an air toxic produces either cancer or a noncancer effect. Because the evidence for either judgment is never unequivocal, a compound labeled as a carcinogen or one deemed to produce noncancer effects may produce neither effect in humans. This possibility introduces uncertainty into the calculation of risk – the risk could be zero. As the evidence for the original conclusion (i.e., that the compound produces the effect) increases, this uncertainty decreases.

NATA's cancer-risk estimates assume that the relationship between exposure and probability of cancer is linear. In other words, the probability of developing cancer is assumed proportional to the exposure (equal to the exposure multiplied by a URE). This type of dose-response model is used routinely in regulatory risk assessment because it is believed to be conservative; that is, if the model is incorrect, it is more likely to lead to an overestimate of risks than to an underestimate. Other scientifically valid, biologically based models are available. These produce estimates of cancer risk that differ from those

obtained from the linear model. Uncertainty in risk estimates is therefore introduced by the inability to justify completely the use of one model or the other (because each model has some scientific support). An essential consideration is that this uncertainty is, to some extent, one-sided. In other words, conservatism when uncertainty exists allows more confidence in the conclusion that true risks are less than predicted than in the conclusion that risks are greater than predicted.

URE parameters have associated uncertainty. In some cases, the UREs are based on maximum-likelihood estimates of the slope of the dose-response relationship derived from reliable data. In other cases, the UREs are based on "upper-bound" estimates (i.e., the slope is not the best estimate, but is a conservative value that is likely to lead to overestimates of risk) derived from less reliable data. For some compounds, the UREs are derived from human-exposure studies, but for others they are from animal exposures. These considerations introduce uncertainty into the URE values, and the amount of uncertainty varies among pollutants.

Another source of uncertainty in estimating risk derives from the values chosen for the RfC parameters used to calculate an HQ for noncancer health risk. The RfC, which (like the URE) is based on limited information, is uncertain; as a result, the value of HQ is uncertain. As is the case for UREs, the uncertainty in the RfC is generally one-sided, and the risk is unlikely to be greater than predicted.

7.4. Summary of Limitations in NATA

EPA developed this assessment to inform both national and more localized efforts to collect information and characterize or reduce air-toxics emissions (e.g., to prioritize pollutants or areas of interest for monitoring and community assessments). As described above, uncertainty and variability characterize many of the elements in the assessment process for NATA, as in other assessments that derive results from environmental data and modeling of environmental data. Because of this, EPA suggests exercising caution when using the results of these assessments, as the overall quality and uncertainty of each assessment vary from location to location and from pollutant to pollutant. More localized assessments, using local-scale monitoring and modeling, are often needed to better characterize local-level risks.

Recognizing the specific limitations in NATA results is critical to interpreting and using them properly, including that the results:

- apply to geographic areas, not specific locations;

- do not include comprehensive impacts from sources in Canada or Mexico;

- are restricted to the year to which the assessment pertains (because the assessment uses emissions data only from that year);

- do not reflect exposures and risk from all compounds;

- do not reflect all pathways of exposure;

- reflect only compounds released into the outdoor air;

- do not fully capture variations in background ambient air concentrations;

- may underestimate or overestimate ambient air concentrations for some compounds due to spatial uncertainties;

- are based on default, or simplifying, assumptions where data are missing or of poor quality; and

- may not accurately capture sources with episodic emissions or other uncertainties.

The results apply to geographic areas, not specific locations. The assessment focuses on variations in air toxics concentrations, exposures and risks among areas such as census tracts, counties and states. All questions asked, therefore, must focus on the variations among different areas. They cannot be used to identify "hot spots" where concentrations, exposures or risks may be significantly higher than other locations. Furthermore, this type of modeling assessment cannot address the kinds of questions an epidemiology study might, such as the relationship between cancer risks and proximity of residences to point sources, roadways and other sources of pollutant emissions.

The results do not include comprehensive impacts from sources in Canada or Mexico. The NATA results for states that border these countries do not thoroughly reflect these potentially significant sources of transported emissions.

The results apply to groups, not to specific individuals. Within a census tract, all individuals are assigned the same ambient concentration, chosen to represent a typical concentration.

Similarly, the exposure assessment uses activity patterns that do not fully reflect variations among individuals. As a result, the exposures and risks in a census tract should be interpreted as typical values rather than as means, medians or some other statistical average. The values are likely to be in the midrange of values for all individuals in the census tract.

The results for the 2014 NATA are restricted to 2014 because the assessment used emissions data from that year. Also, the assumption regarding emissions in the assessment is that the levels remain constant throughout one's lifetime (the emissions are not today's levels nor are they projected levels). Emissions continue to decrease, however, as (1) mobile-source regulations are phased in over time, (2) EPA-issued air-toxics regulations for major industrial sources reach compliance due dates, (3) state and industry initiatives to reduce air pollutants continue and (4) some facilities are closed or have made process changes or other changes that have significantly reduced their emissions since 2014.

The results do not reflect exposures and risk from all compounds. Only 138 of the 181 air toxics (180 CAA HAPs plus diesel PM) modeled in NATA have dose-response values. The remaining air toxics are not considered in the aggregate cancer risk or target-organ-specific hazard indexes. Of significance is that the assessment does not quantify cancer risk from diesel PM, although EPA has classified diesel PM as likely to be carcinogenic to humans by inhalation from environmental exposures. Currently, a URE for diesel PM has not yet been derived; therefore, a quantitative estimate of the cancer risks was included in the 2014 NATA. An IRIS RfC for diesel PM allows NATA to include a quantitative estimate of its noncancer effects.

The results do not reflect all pathways of exposure. The assessment includes only risks from direct inhalation of the emitted pollutants. It does not consider pollutants that might then deposit onto soil and into water and food, and therefore enter the body through ingestion or skin contact. Consideration of these routes of exposure could increase estimates of exposure and risk.

The assessment results reflect only compounds released into the outdoor air. The assessment does not include exposure to pollutants produced indoors, such as from stoves or out-gassing from building materials, or evaporative benzene emissions from cars in attached garages. For some compounds such as formaldehyde, these indoor sources can contribute significantly to the total exposure for an individual, even if only inhalation exposures are considered. In addition, the assessment does not consider pollutants released directly to water and soil. It does include secondary formation, the transformation of one pollutant into another in the atmosphere.

The assessment does not utilize CMAQ in all areas (i.e., not in Alaska, Hawaii, Puerto Rico and the U.S. Virgin Islands) and therefore does not estimate fires, biogenics and secondary formation based on location-specific data in these areas. It also does not use CMAQ for all pollutants and hence may not appropriately estimate long-range transport for these non-CMAQ pollutants. For pollutants not estimated in CMAQ, the assessment uses background ambient air concentrations based on remote concentration estimates, but these would not account for variations due to regional transport of these pollutants.

NATA may underestimate or overestimate ambient air concentrations for some compounds in some locations due to spatial uncertainty in mobile and nonpoint emissions, which are more uncertain at finer geographic scales.

The assessment uses default, or simplifying, assumptions where data are missing or of poor quality. Data for some variables used in the modeling for emissions and dispersion of pollutants (such as stack height and facility location) may be unavailable or flawed. In such instances, these values are replaced by default assumptions. For example, a stack height for a facility might be set equal to stack heights at comparable facilities or the location of the release points within a facility might be placed at the center of the facility. These substitutions introduce uncertainty into the final predictions of ambient concentrations, exposures and risks.

NATA may not accurately capture sources with episodic emissions except for those with continuous emissions monitoring (CEMS) data, which use hourly emissions. NATA also does not include any short-term (a few days or weeks) deviations from a facility's typical emissions pattern, such as during startups, shutdowns, malfunctions and upsets. NATA modeling uses temporal profiles for sources without CEMS that would not capture non-routine emissions spikes.

8. References

Akhtar, F., Henderson, B., Appel, W., Napelenok, S., Hutzell, B., Pye, H. and 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.

Appel, K.W., Gilliam, R.C., Davis, N., Zubrow, A. and Howard, S.C. 2011. Overview of the Atmospheric Model Evaluation Tool (AMET) v1.1 for Evaluating Meteorological and Air Quality Models. *Environ. Modell. Softw.*, 26(4): 434–443. Available online at http://www.cmascenter.org/.

Appel, K.W., Napelenok, S., Hogrefe, C., Pouliot, G., Foley, K.M., Roselle, S.J., Pleim, J.E., Bash, J., Pye, H.O.T., Heath, N., Murphy, B., Mathur, R. 2018. Overview and evaluation of the Community Multiscale Air Quality Model (CMAQ) modeling system version 5.2. In Mensink C., Kallos G. (eds), Air Pollution Modeling and its Application XXV. ITM 2016. Springer Proceedings in Complexity. Springer, Cham. Available online at https://doi.org/10.1007/978-3-319-57645-9_11.

ATSDR. 2012. (Agency for Toxic Substances & Disease Registry). Toxicological Profile for Manganese. Available online at http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=102&tid=23.

ATSDR. 2015. Toxic Substances Portal Minimal Risk Levels (MRLs) for Hazardous Substances. Available online at http://www.atsdr.cdc.gov/mrls/index.asp. Last updated 28 October 2015. Last accessed 10 December 2015.

Bey, I., Jacob, D.J., Yantosca, R.M., Logan, J.A., Field, B.D., Fiore, A.M., Li, Q., Liu, H.Y., Mickley, L.J., Schultz, M.G. 2001. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. Journal of Geophysical Research 106, 23073. https://doi.org/10.1029/2001JD000807.

Caldwell, J.C., Woodruff, T.J., Morello-Frosch, R. and Axelrad, D.A. 1998. Application of Health Information to Hazardous Air Pollutants Modeled in EPA's Cumulative Exposure Project. *Toxicology and Industrial Health,* 14(3): 429–454.

Cimorelli, A.J., Perry, S.G., Venkatram, A., Weil, J.C., Paine, R.J., Wilson, R.B., Lee, R.F., Peters, W.D. and Brode, R.W. 2005. AERMOD: A Dispersion Model for Industrial Source Applications. Part I: General Model Formulation and Boundary Layer Characterization. *Journal of Applied Meteorology,* 44: 682–693.

Cook, R., Phillips, S., Houyoux, M., Dolwick, P., Mason, R., Yanca, C., Zawacki, M., Davidson, K., Michaels, H., Harvey, C., Somers, J. and Luecken, D. 2011. Air Quality Impacts of Increased Use of Ethanol under the United States' Energy Independence and Security Act. *Atmospheric Environment,* 45: 7714–7724.

Dorman DC, Struve MF, Wong BA, Marshall MW, Gross EA and Willson GA, 2008. Respiratory tract responses in male rats following subchronic acrolein inhalation. Inhal Toxicol 20(3): 205-16.

EPA (U.S. Environmental Protection Agency). 1986. Guidelines for Mutagenicity Risk Assessment. EPA/630/R-98/003. EPA, Washington, DC. Available online at

http://www2.epa.gov/risk/guidelines- mutagenicity-risk assessment. Last accessed 10 December 2015.

EPA. 1991. Guidelines for Developmental Toxicity Risk Assessment. EPA/600/R-91/001. EPA, Washington, DC. Available online at http://www2.epa.gov/risk/guidelines-developmental-toxicity-risk assessment. Last accessed 26 October 2015.

EPA. 1993. Integrated Risk Information System Review of Manganese. Available online at https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=373.

EPA. 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry. EPA/600/8-90/066F. EPA Office of Research and Development (ORD), Washington, DC. Available online at http://www2.epa.gov/risk/methods-derivation-inhalation-reference-concentrations- and-application-inhalation-dosimetry. Last accessed 27 October 2015.

EPA. 1996. Guidelines for Reproductive Toxicity Risk Assessment. EPA/630/R-96/009. EPA, Washington, DC. Available online at http://www2.epa.gov/risk/guidelines-reproductive-toxicity-risk assessment. Last accessed 10 December 2015.

EPA. 1997. Health Effects Assessment Summary Tables (HEAST). EPA NCEA, Washington, DC. Available online at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=2877.

EPA. 1998. Guidelines for Neurotoxicity Risk Assessment. EPA/630/R-97/0. EPA, Washington, DC. Available online at http://www2.epa.gov/risk/guidelines-neurotoxicity-risk assessment. Last accessed 10 December 2015.

EPA. 2000. Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures. EPA/630/R-00/002. Risk Assessment Forum, Washington, DC. Available online at http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=4486. Last accessed 8 December 2015.

EPA. 2001a. NATA – Evaluating the National-scale Air Toxics Assessment 1996 Data – An SAB Advisory. EPA/SAB/EC/ADV-02/001. Science Advisory Board, Washington, DC. Available online at http://archive.epa.gov/airtoxics/nata/web/pdf/sabrept1201.pdf. Last accessed 29 November 2015.

EPA. 2001b. National-scale Air Toxics Assessment for 1996. Draft for EPA Science Advisory Board Review: January 18, 2001. EPA-453/R-01-003. EPA Office of Air Quality Planning and Standards (OAQPS), Research Triangle Park, NC. Available online at http://archive.epa.gov/airtoxics/nata/web/html/sabrev.html. Last accessed 29 November 2015.

EPA. 2002a. Health Assessment Document for Diesel Engine Exhaust. EPA/600/8-90/057F. EPA ORD/ National Center for Environmental Assessment (NCEA), Washington, DC. Available online at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=29060. Last accessed 4 December 2015.

EPA. 2002b. The HAPEM User's Guide Hazardous Air Pollutant Exposure Model, Version 4. EPA OAQPS, Research Triangle Park, NC. Available online at http://archive.epa.gov/airtoxics/nata/web/zip/hapem4guide4.zip. Last accessed 29 November 2015.

EPA. 2002c. Comparison of ASPEN Modeling System Results to Monitored Concentrations. EPA. Available online at http://archive.epa.gov/airtoxics/nata/web/html/mtom_pre.html. Last updated 22 October 2015. Last accessed 30 November 2015.

EPA. 2003a. Integrated Risk Information System Review of Acrolein. Available online at https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=364.

EPA. 2003b. Framework for Cumulative Risk Assessment. EPA/630/P-02/001F. EPA ORD/NCEA, Washington, DC. Available online at http://www2.epa.gov/risk/framework-cumulative-risk assessment. Last accessed 26 October 2015.

EPA. 2004a. Air Toxics Risk Assessment Reference Library. Volume 1: Technical Resource Manual. EPA-453/K-04-001A. EPA OAQPS, Research Triangle Park, NC. Available online at http://www2.epa.gov/sites/production/files/2013-08/documents/volume_1_reflibrary.pdf. Last accessed 2 December 2015.

EPA. 2004b. Air Toxics Risk Assessment Reference Library. Volume 2: Facility-Specific Assessment. EPA-453/K-04-001B. EPA OAQPS, Research Triangle Park, NC. Available online http://www2.epa.gov/sites/production/files/2013-08/documents/volume_2_facilityassess.pdf. Last accessed 2 December 2015.

EPA. 2005a. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. EPA, Washington, DC. Available online at http://www2.epa.gov/risk/guidelines-carcinogen-risk assessment. Last accessed 26 October 2015.

EPA. 2005b. Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. EPA/630/R-03/003F. EPA, Washington, DC. Available online at http://www2.epa.gov/osa/memoranda-about-implementation-cancer-guidelines-and-accompanying- supplemental-guidance-science. Last accessed 4 December 2015.

EPA. 2005c. The HAPEM User's Guide Hazardous Air Pollutant Exposure Model, Version 5. EPA OAQPS, Research Triangle Park, NC. Available online at http://www2.epa.gov/sites/production/files/2013-08/documents/hapem5_guide.pdf. Last accessed 29 November 2015.

EPA. 2006a. Air Toxics Risk Assessment Reference Library. Volume 3: Community-Scale Assessment. EPA-453/K-06-001C. EPA OAQPS, Research Triangle Park, NC. Available online at http://www2.epa.gov/sites/production/files/2013-08/documents/volume_3_communityassess.pdf. Last accessed 2 December 2015.

EPA. 2006b. Comparison of 1999 Model-Predicted Concentrations to Monitored Data. EPA OAQPS. Available online at http://archive.epa.gov/airtoxics/nata1999/web/html/99compare.html. Last updated 13 September 2015. Last accessed 30 November 2015.

EPA. 2007a. Control of Hazardous Air Pollutants from Mobile Sources Regulatory Impact Analysis. EPA420-R-07-002. EPA Office of Transportation and Air Quality (OTAQ) Assessment and Standards Division, Ann Arbor, MI. Available online at https://nepis.epa.gov/Exe/ZyPdf.cgi?Dockey=P1004LNN.PDF. Last accessed 16 July 2018.

EPA. 2007b. The HAPEM User's Guide Hazardous Air Pollutant Exposure Model, Version 6. EPA OAQPS, Research Triangle Park, NC. Available online at http://www2.epa.gov/sites/production/files/2013-08/documents/hapem6_guide.pdf. Last accessed 29 November 2015.

EPA. 2009. Comparison of 2002 Model-Predicted Concentrations to Monitored Data. EPA OAQPS. Available online at http://archive.epa.gov/nata2002/web/html/compare.html. Last accessed 30 November 2015.

EPA. 2010a. Development of a Relative Potency Factor (RPF) Approach for Polycyclic Aromatic Hydrocarbon (PAH) Mixtures: In Support of Summary Information of the Integrated Risk Information System (IRIS) (External Review Draft). EPA/635/R-08/012A. EPA, Washington, DC. Available online at http://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=194584. Last accessed 19 October 2015.

EPA. 2010b. Results of the 2005 NATA Model-to-Monitor Comparison, Final Report. Prepared by Eastern Research Group for EPA OAQPS. Available online at http://www3.epa.gov/ttn/atw/nata2005/05pdf/nata2005_model2monitor.pdf. Last accessed 30 November 2015.

EPA. 2014. Toxicological Review of Trichloroethylene. EPA NCEA, Washington, DC. Available online at http://cfpub.epa.gov/ncea/iris/search/index.cfm?keyword=trichloroethylene. Last updated 12 September 2014. Last accessed 10 December 2015.

EPA. 2015a. Consolidated Human Activity Database (CHAD). EPA, Washington, DC. Available online at http://www2.epa.gov/healthresearch/consolidated-human-activity-database-chad-use-human-exposure- and-health-studies-and. Last updated 30 September 2015. Last accessed 17 November 2015.

EPA. 2015b. The HAPEM User's Guide Hazardous Air Pollutant Exposure Model, Version 7. EPA OAQPS, Research Triangle Park, NC. Available online at http://www.epa.gov/fera/hazardous-air-pollutant-exposure-model-hapem-users-guides. Last accessed 29 November 2015.

EPA. 2015c. What Are the Six Common Air Pollutants? EPA, Washington, DC. Available online at http://www3.epa.gov/airquality/urbanair/. Last updated 18 September 2015. Last accessed 2 December 2015.

EPA. 2015d. Risk Assessment Guidance and Tools. EPA, Washington, DC. Available online at http://www.epa.gov/risk/guidance.htm. Last updated 1 December 2015. Last accessed 2 December 2015.

EPA. 2015e. User's Guide for the AMS/EPA Regulatory Model – AERMOD. EPA-454/B-03-001. Addendum June 2015. EPA, Research Triangle Park, NC.

EPA. 2015f. Community Multiscale Air Quality (CMAQ). EPA, Washington, DC. Available online at http://www.epa.gov/air-research/community-multi-scale-air-quality-cmaq-modeling-system-air-quality- management. Last updated 8 December 2015. Last accessed 10 December 2015.

EPA. 2015g. MOVES 2014. EPA OTAQ, Washington, DC. Available online at
http://www3.epa.gov/otaq/models/moves/moves-docum.htm. Last updated 5 November 2015.
Last accessed 10 December 2015.

EPA. 2015h. Overview by Section of CAA. EPA OAR, Washington, DC. Available online at
http://www.epa.gov/ttn/atw/overview.html. Last updated 10 September 2015. Last accessed 10
December 2015.

EPA. 2016. Products and Publications Relating to Risk Assessment Produced by the Office of the Science
Advisor (OSA). EPA OSA, Washington, DC. Available online at http://www.epa.gov/osa/products-
and-publications-relating-risk-assessment-produced-office-science- advisor. Last updated 23
August 2016. Last accessed 16 August 2018.

EPA. 2017. Revisions to the Guideline on Air Quality Models: Enhancements to the AERMOD Dispersion
Modeling System and Incorporation of Approaches To Address Ozone and Fine Particulate
Matter. 82 Federal Register 10 (17 January 2017), pp. 5182-5231.

EPA. 2018. Initial List of Hazardous Air Pollutants with Modifications. EPA Office of Air and Radiation
(OAR), Washington, DC. Available online at h https://www.epa.gov/haps/initial-list-hazardous-
air-pollutants-modifications. Last updated 16 March 2017. Last accessed 18 June 2018.

ERG. 2017. Improvement of Default Inputs for MOVES and SMOKE-MOVES: CRC Project A-100. February
28, 2017. Prepared for: The Coordinating Research Council. ERG No. 4020.00.001.001.

Eyth, Alison, J. Vukovitch, M. Strum, P. Dolwick, G. Pouliot, C. Allen, J. Beidler, B. Baek and Z. Adeleman.
2016. "Development of 2011 Hemispheric Emissions for CMAQ," Presented at the 2016 CMAS
Conference, October 24, 2016, University of North Carolina, Chapel Hill.

Feron, V.J., Kryusse, A., Til H.P., et al. 1978. Repeated exposure to acrolein vapor: subacute studies in
hamsters, rats and rabbits. *Toxicology* 9:47-57.

Garshick, E., Laden, F., Hard, J.E., Davis, M.E., Eisen, E.A. and Smith, T.J. 2012. Lung Cancer and
Elemental Carbon Exposure in Trucking Industry Workers. *Environmental Health Perspectives,*
120:1301–1306. Available online at http://dx.doi.org/10.1289/ehp.1204989. Last accessed 10
December 2015.

HEI (Health Effects Institute). 2015. Diesel Emissions and Lung Cancer: An Evaluation of Recent
Epidemiological Evidence for Quantitative Risk Assessment. HEI, Boston, MA. Available online at
http://pubs.healtheffects.org/view.php?id=446. Last updated 24 November 2015. Last accessed
10 December 2015.

Henderson, B.H., Akhtar, F., Pye, H.O.T., Napelenok, S.L. and Hutzell, W.T. 2014. A Database and Tool for
Boundary Conditions for Regional Air Quality Modeling: Description and Evaluation. *Geosci.
Model Dev.,* 7:339–360.

IARC (International Agency for Research on Cancer). 2014. IARC Monographs on the Evaluation of
Carcinogenic Risks to Humans, Volume 105 (2014). Available online at
http://monographs.iarc.fr/ENG/Monographs/vol105/index.php. Last accessed 10 December
2015.

Isakov, V., Irwin., J. and Ching, J.K. 2007. Using CMAQ for Exposure Modeling and Characterizing the Sub-grid Variability for Exposure Estimates. *Journal of Applied Meteorology and Climatology,* 46:1354– 1371.

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Committee on the Institutional Means for Assessments of Risk to Public Health, Commission on Life Sciences. NRC. National Academy Press, Washington, DC.

NRC. 1994. Science and Judgment in Risk Assessment. Committee on Risk Assessment of Hazardous Air Pollutants, Board on Environmental Sciences and Technology, Commission on Life Sciences. NRC. National Academy Press, Washington, DC.

NUATRC (Mickey Leland National Urban Air Toxics Research Center). 2011. Available online at http://www.sph.uth.tmc.edu/mleland/. Last updated 30 December 2011. Last accessed 10 December 2015.

OEHHA (Office of Environmental Health Hazard Assessment, California). 2008. California Environmental Protection Agency. Technical Support Document for the Derivation of Noncancer Reference Exposure Levels. Available online at https://oehha.ca.gov/media/downloads/crnr/appendixd1final.pdf.

OEHHA. 2015. Hot Spots Guidelines. OEHHA, Sacramento, CA. Available online at http://www.oehha.ca.gov/air/hot_spots/index.html. Contents last updated 6 March 2015.

OEHHA. 2016. Air Toxicology and Epidemiology. OEHHA, Sacramento, CA. Available online at http://www.oehha.ca.gov/air/allrels.html. Values last updated 28 June 2016.

Otte T.L. and 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.

Pye, H.O. and Pouliot, G.A. 2012. Modeling the Role of Alkanes, Polycyclic Aromatic Hydrocarbons, and Their Oligomers in Secondary Organic Aerosol Formation. *Environ. Sci. Technol*, 46, 11, 6041-6047.

Roels HA, Ghyselen P, Buchet JP, et al. 1992. Assessment of the permissible exposure level to manganese in workers exposed to manganese dioxide dust. Br J Ind Med 49:25-34.

Silverman, D.T., Samanic, C.M., Lubin, J.H., Blair, A.E., Stewart, P.A., Vermeulen, R., Coble, J.B., Rothman, N., Schleiff, P.L., Travis, W.D., Ziegler, R.G., Wacholder, S. and Attfield, M.D. 2012. The Diesel Exhaust in Miners Study: A Nested Case-Control Study of Lung Cancer and Diesel Exhaust. Available online at http://dx.doi.org/10.1093/jnci/djs034. Last accessed 10 December 2015.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X., Wang, W. and Powers, J.G. 2008. A Description of the Advanced Research WRF Version 3. Available online at http://www2.mmm.ucar.edu/wrf/users/pub-doc.html. Last updated 5 December 2014. Last accessed 16 December 2015.

U.S. Census Bureau. 2010. Decennial Census of Population and Housing. Available online at https://www.census.gov/programs-surveys/decennial-census/decade.2010.html. Last accessed 20 August 2018.

Wesson, K., Fann, N., Morris, M., Fox, T. and Hubbell, B. 2010. A Multi-pollutant, Risk-based Approach to Air Quality Management: Case Study for Detroit. *Air Pollut. Res*., 1:296–304. doi:10.5094/APR.2010.037.

Yarwood, G., J. Jung, G. Whitten, G. Heo, J. Mellberg and M. Estes. 2010: Updates to the Carbon Bond Chemical Mechanismfor Version 6 (CB6). Presented at the 9th Annual CMAS Conference, Chapel Hill, NC. Available at https://www.cmascenter.org/conference/2010/abstracts/emery_updates_carbon_2010.pdf.

This page intentionally left blank.

Appendix A. Glossary

"N"-in-1 million cancer risk:
A risk level of "N"-in-1 million implies that up to "N" people out of one million equally exposed people would contract cancer if exposed continuously (24 hours per day) to the specific concentration over 70 years (an assumed lifetime). This would be in addition to cancer cases that would normally occur in one million unexposed people. Note that NATA looks at *lifetime* cancer risks. This shouldn't be confused with or compared to *annual* cancer risk estimates. To compare an annual cancer risk estimate with NATA results, multiply the annual estimate by 70 (or divide the lifetime risk by 70).

Activity-pattern data:
Data that depict actual human physical activity, the location of the activity and the time of day it takes place. The Hazardous Air Pollution Model (HAPEM) uses activity-pattern data from EPA's Comprehensive Human Activity Database (CHAD).

Adverse health effect:
A change in body chemistry, body function or cell structure that could lead to disease or health problems.

AMS/EPA Regulatory Model (AERMOD):
EPA's preferred model to simulate near-field (i.e., within 50 km) dispersion of emissions. AERMOD models near-surface (boundary-layer) air turbulence in simple and complex terrain. This allows AERMOD to simulate how pollutants move and disperse in the air. It calculates pollutant concentrations from surface and elevated point, area, line and volume sources at many discrete points (receptors).

Air toxics:
Pollutants known to cause or suspected of causing cancer or other serious health effects. Air toxics are also known as toxic air pollutants or hazardous air pollutants.* Health concerns are linked to both short- and long-term exposures to these pollutants. Many air toxics cause respiratory, neurological, immune or reproductive effects, particularly for more susceptible or sensitive groups such as children. Five important air pollutants are not included in the list of air toxics because the Clean Air Act addresses them separately as "criteria pollutants." These are particulate matter (PM), nitrogen oxides (NO_x), sulfur oxides (SO_x), ozone and carbon monoxide. Lead is both a criteria pollutant and an air toxic. Criteria pollutants are not addressed in NATA.

*Diesel particulate matter is not a hazardous air pollutant but is included in the NATA air toxics.

Ambient:
Surrounding, as in the surrounding environment. In NATA, ambient air refers to the outdoor air surrounding a person through which pollutants can be carried. Therefore, the ambient concentrations estimated by NATA are concentrations estimated in the outdoor environment. NATA also estimates exposure concentrations that result when a person moves through various microenvironments, including the indoor environment.

Ambient air monitoring:
Process of collecting outdoor air samples to determine how much of an air pollutant is present at a location. Monitoring is used to:
- assess the extent of pollution;
- provide air pollution data to the public in a timely manner;
- help implement air quality goals or standards;
- evaluate whether emissions control strategies are effective;
- provide data on air quality trends;
- help evaluate air quality models or modeling results; and
- support research (for example, long-term studies of the health effects of air pollution).

State, local and tribal agencies often monitor near places where screening tools (such as NATA) suggest the chance of high concentrations or risks.

Area and other sources:
Sources of air pollution that, by themselves, generally have lower emissions than "major sources" of air pollution (like factories). Area sources are often too small or too widespread to be inventoried as individual sources. They include facilities with air toxics emissions below the major source threshold as defined in the Clean Air Act (less than 10 tons of a single toxic air pollutant or less than 25 tons of multiple toxic air pollutants emitted in any one year). Area sources include smaller facilities, such as dry cleaners.

As a separate definition, area sources in air quality modeling refer to those modeled in two dimensions (with length and width), as compared to point sources modeled at a single location.

Assessment System for Population Exposure Nationwide (ASPEN):
A computer model used to estimate toxic air pollutant concentrations. The ASPEN model includes:
- rate of pollutant release;
- location of and height from which the pollutants are released;
- wind speeds and directions from the meteorological stations nearest to release;
- breakdown of the pollutants in the atmosphere after release (i.e., reactive decay);
- settling of pollutants out of the atmosphere (i.e., deposition);
- transformation of one pollutant into another (i.e., secondary formation or decay).

Atmospheric transformation (secondary formation):
The process by which chemicals are transformed into other chemicals in the air (atmosphere). When a chemical is transformed, the original pollutant no longer exists; it is replaced by one or more new chemicals. Compared to the original chemical, the transformed chemical can have more, less or the same toxicity. Transformations and removal processes affect both the fate of the chemical and how long it stays in the air, called its persistence. Persistence is important because human exposure to a chemical depends on the length of time the chemical remains in the air. In NATA, we use both "atmospheric transformation" and "secondary formation"; they mean the same thing.

Background concentrations:
The amount of a pollutant that exists in the air that does not come from a specific source. These pollutants may come from a natural source or from distance sources. Background concentrations can explain pollutant concentrations found even without recent human-caused emissions. In NATA, we add background concentrations to AERMOD concentrations but not to CMAQ concentrations, which include background already. Most risk from NATA background concentrations is from carbon tetrachloride, a common pollutant that has few emission sources but is persistent due to its long half-life.

Biogenic emissions:
Biogenic emissions are emissions from natural sources, such as plants and trees. These sources emit formaldehyde, acetaldehyde and methanol; formaldehyde and acetaldehyde are key risk drivers in NATA. Biogenic sources also emit large amounts of other nonhazardous VOCs. We estimate biogenic emissions with a model that uses vegetation and land use data with temperature and solar radiation data. In addition to being a primary source of air toxics, compounds emitted by biogenic sources sometimes react with human-caused pollutants to form secondary pollutants. The NATA biogenics source group includes only the primary emissions.

Cancer Risk:
The probability of contracting cancer over the course of a lifetime, assuming continuous exposure (assumed in NATA to be 70 years).

Carcinogen:
A chemical, physical or biological agent that can cause cancer.

Carcinogenicity:
Ability to produce cancer cells from normal cells.

Chemical Abstracts Service (CAS) number:
A unique number assigned to a chemical by the Chemical Abstracts Service, a service of the American Chemical Society. The purpose is to make database searches easier, as chemicals often have many names.

Census tracts:
Land areas defined by the U.S. Census Bureau. Tracts usually contain from 1,200 to 8,000 people, with most having close to 4,000 people. Census tracts are usually smaller than 2 square miles in cities, but are much larger in rural areas.

Cohort:
A group of people assumed to have identical exposures during a certain period. Using cohorts makes modeling exposures of a large population easier to manage. In NATA, we divide the entire population into a set of cohorts. Each person is assigned to one and only one cohort, and all the cohorts combined equal the entire population.

Community Multiscale Air Quality (CMAQ) modeling system:
An air quality model used in NATA. CMAQ estimates how pollutants move and disperse in the air. It includes the effect of atmospheric chemistry – how pollutants react in the air – a unique feature of the model. CMAQ's structure allows it to calculate concentrations over a very large area, including many emission sources.

Concentration:
A way to describe how much of a pollutant is in the air. Concentration is usually shown as an amount, or mass, of pollutant per certain volume of air. In NATA, most concentrations are in micrograms (μg) of air pollutant per cubic meter (m^3) of air (a "box" of air one meter on each side).

Consolidated Human Activity Database (CHAD):
An in-depth EPA database of human activity. CHAD includes data from over 20 activity studies dating to 1982. It also includes data from other assessments of human exposure, intake dose and risk.

Diesel particulate matter:
A mixture of particles that is part of diesel exhaust. EPA lists diesel exhaust as a mobile-source air toxic due to the cancer and noncancer health effects linked to exposure to whole diesel exhaust. Diesel PM (expressed as grams diesel PM/m^3) has been used as a surrogate exposure measure for whole diesel exhaust.

Dispersion model:
A computerized set of equations that uses emissions and meteorological data to simulate how air pollutants behave and move in the air. A dispersion model estimates outdoor concentrations of individual air pollutants at chosen locations (called receptors).

Emission Inventory System (EIS):
An EPA information system for collecting emission inventory data and generating emission inventories.

Emissions:
Pollutants released into the air.

Emission inventory:
A listing, by source, of the location and amount of air pollutants released into the air during some period (in NATA, a single year).

Exposure assessment:
An exposure assessment is part of an air toxics risk assessment such as NATA. The assessment determines (or estimates):
- how a person may be exposed to chemicals (for example, by breathing);
- how much of a chemical to which a person is likely to be exposed;
- how long and/or how often they will be exposed; and
- how many people are likely to be exposed.

HAP:
Hazardous air pollutant; another name for air toxics.

Hazard index (HI):
The sum of hazard quotients for toxics that affect the same target organ or organ system. Because different air toxics can cause similar adverse health effects, combining hazard quotients from different toxics is often appropriate. As with the hazard quotient, exposures below an HI of 1.0 likely will not result in adverse noncancer health effects over a lifetime of exposure. An HI equal to or greater than 1.0, however, doesn't necessarily suggest a likelihood of adverse effects.

Hazard quotient (HQ):
The ratio of the potential exposure to a substance and the level at which no adverse effects are expected (calculated as the exposure divided by the appropriate chronic or acute value). A hazard quotient less than or equal to 1.0 indicates that adverse noncancer effects are not likely to occur, and thus can be considered to have negligible hazard. For HQs greater than 1.0, the potential for adverse effects increases, but we do not know by how much.

Hazardous Air Pollutant Exposure Model (HAPEM):
A computer model designed to estimate inhalation exposure for specified population groups and air toxics. The model uses census data, human-activity patterns, ambient air quality levels, and indoor/outdoor concentration relationships to estimate an expected range of inhalation exposure concentrations for groups of people.

Human Exposure Model (HEM):
A computer model used primarily for conducting inhalation risk assessments for sources emitting air toxics to ambient air.

Inhalation exposure:
Introducing air toxics (or other pollutants) into the body via breathing. Once inhaled, air toxics can be deposited in the lungs, taken into the blood, or both.

Integrated Risk Information System (IRIS):
An EPA program that identifies and characterizes the health hazards of chemicals found in the environment. IRIS is EPA's preferred source of toxicity information.

Lifetime cancer risk:
The probability of contracting cancer over the course of a lifetime (assumed to be 70 years for the purposes of NATA).

Major sources:
Defined by the Clean Air Act as those stationary facilities that emit or have the potential to emit 10 tons of any one toxic air pollutant or 25 tons of more than one toxic air pollutant per year.

Median:
The middle value of a set of ordered values (i.e., half the numbers are less than or equal to the median value). A median is the 50th percentile of the data.

Metropolitan statistical area (MSA):
A region with a relatively high population density at its core and close economic ties throughout the area. As defined by the U.S. Census Bureau, an MSA must have at least one urban area of 50,000 or more inhabitants.

Microenvironment:
A small space in which human contact with a pollutant takes place. NATA models cohort activities in indoor, outdoor and in-vehicle microenvironments:
1. Indoor locations:
   - Residence
   - Office
   - Store

- School
- Restaurant
- Church
- Manufacturing facility
- Auditorium
- Healthcare facility
- Service station
- Other public building
- Garage

2. Outdoor locations:
   - Parking lot/garage
   - Near road
   - Motorcycle
   - Service station
   - Construction site
   - Residential grounds
   - School
   - Sports arena
   - Park/golf course

3. In-vehicle locations:
   - Car
   - Bus
   - Truck
   - Train/subway
   - Airplane
   - Other

Microgram:
One-millionth of a gram. One gram is about one twenty-eighth of an ounce, or about the weight of a raisin or paper clip.

Mobile source:
Air pollution sources that can move from place to place, like cars or trucks. Mobile sources are divided into two categories: on-road and nonroad vehicles/engines.

Monitoring:
See Ambient air monitoring.

Motor Vehicle Emission Simulator (MOVES):
An emissions modeling system that estimates emissions for mobile sources at the national, county, and project level for criteria air pollutants, air toxics and greenhouse gases.

National Air Toxics Assessments (NATA):
EPA's ongoing thorough review of air toxics in the United States. NATA results help scientists focus on pollutants, emission sources and places that may need further study to better understand risks. NATA also spurs improvements in what we know about U.S. air toxics. This includes expanding air toxics monitoring, improving and updating emission inventories, improving air quality modeling, driving research on health effects and exposures to both ambient and indoor air, and improving assessment tools.

National Emissions Inventory (NEI):
A national database of air emissions data. EPA prepares NEI with input from many state and local air agencies, from tribes and from industry. This database contains information on stationary and mobile sources that emit criteria air pollutants and their precursors, as well as hazardous air pollutants. NEI includes estimates of annual emissions, by

source, of air pollutants in each area of the country. NEI includes emission estimates for all 50 states, the District of Columbia, Puerto Rico and the U.S. Virgin Islands.

National Mobile Inventory Model (NMIM):
Computer tool containing EPA's NONROAD model for estimating county-level inventories of nonroad mobile emissions.

Noncancer risks:
Risks associated with health effects other than cancer.

Nonroad mobile sources:
Mobile sources not used on roads and highways for transportation of passengers or freight. Nonroad sources include:
- aircraft;
- heavy equipment;
- locomotives;
- marine vessels;
- recreation vehicles (snowmobiles, all-terrain vehicles, etc.); and
- small engines and tools (lawnmowers, etc.).

On-road mobile sources:
Mobile sources used on roads and highways for transportation of passengers or freight. On-road sources include:

- passenger cars and trucks
- commercial trucks and buses; and
- motorcycles.

Percentile:
Any one of the points dividing a set of values into parts that each contain 1/100 of the values. For example, the 75th percentile is a value such that 75 percent of the values are less than or equal to it.

Polycyclic organic matter (POM):
A broad class of compounds that includes polycyclic aromatic hydrocarbons. Polycyclic organic matter (POM) compounds form mainly from combustion and are present in the air as particles. Sources of POM emissions include:
- vehicle exhaust;
- forest fires and wildfires;
- asphalt roads;
- coal;
- coal tar;
- coke ovens;
- agricultural burning;
- residential wood burning; and
- hazardous waste sites.

Not all POM reported to EPA's National Emission Inventory is broken down by pollutant. So, we make some simplifying assumptions to model and assess the risk from the different pollutants that make up polycyclic organic matter.

Reference concentration (RfC):
An estimate of a continuous inhalation exposure unlikely to cause adverse health effects during a lifetime. This estimate includes sensitive groups such as children, asthmatics and the elderly.

Risk:
The probability that adverse effects to human health or the environment will occur due to a given hazard (such as exposure to a toxic chemical or mixture of toxic chemicals). We can measure or estimate some risks in numerical terms (for example, one chance in a hundred).

Rural:
A county is considered "rural" if it does not contain a metropolitan statistical area with a population greater than 250,000 and the U.S. Census Bureau designates 50 percent or less of the population as "urban." Note that this definition does not necessarily apply for any regulatory or implementation purpose. It is consistent with the definition EPA used in the analyses to support the Integrated Urban Air Toxics Strategy.

Science Advisory Board (SAB):
A panel of scientists, engineers and economists who provide EPA with independent scientific and technical advice.

Secondary formation:
See "Atmospheric transformation (Secondary Formation)"

Secondary sources:
See "Atmospheric transformation (Secondary Formation)"

Sparse Matrix Operator Kernel Emissions (SMOKE):
A modeling system that processes emissions data for use in air quality models. It uses the Biogenic Emission Inventory System (BEIS) to model biogenic emissions. It also has a feature to use MOVES emission factors, activity data and meteorological data to compute hourly gridded on-road mobile emissions.

Stationary sources:
Sources of air emissions that do not move. Stationary sources include large industrial sources such as power plants and refineries, smaller industrial and commercial sources such as dry cleaners, and residential sources such as residential wood combustion and consumer products usage. Stationary sources may be "major" or "area" sources based on definitions in the Clean Air Act. In NATA, we present sources as "point" and "nonpoint" rather than "major" and "area" sources. "Point" and "nonpoint" reflect how we modeled each emission source. Some smaller sources that are area sources in the inventory (based on the amount of their emissions) are modeled as point sources because their location was identified with latitude and longitude coordinates.

Susceptibility:
The increased likelihood of an adverse effect. Susceptibility is often discussed in terms of relationship to a factor describing a human population (for example, life stage, demographic feature or genetic trait).

Toxicity weighting:
A way to prioritize pollutant emissions based on risk. To calculate toxicity-weighted emissions, we multiply emissions from a facility or source being assessed by a toxicity factor for each pollutant. Pollutants that are more harmful (for the same emission rate) have a higher toxicity factor. By weighting the amount of a pollutant released to its toxicity, we can compare relative risk from different pollutants in emission inventories. Toxicity weighting is very useful if the number of pollutants is large, helping risk assessors focus on pollutants that contribute the most to risk.

Typical:
Describes a hypothetical person living at a census-tract centroid (a reference point usually but not always located at the geographic center of a census tract) and doing the types of things (indoors and outdoors) that most people living in that tract would do. To describe that person's risk, NATA divides the population into cohorts (groups assumed to have the same exposures each day) based on where they live, how old they are and their daily activity patterns. Fora census tract, we select age-appropriate activity patterns to model the range of exposure conditions for residents of the tract. We can then calculate a population-weighted typical exposure estimate for each cohort. We use this value to estimate risks for a "typical" individual residing in that tract.

Unit risk estimate (URE):
The upper-bound excess lifetime cancer risk estimated to result from continuous exposure to an air toxic at a concentration of 1 microgram per cubic meter ($\mu g/m^3$) in air. You can interpret the URE as follows: If the URE = 3 x 10^{-6} per $\mu g/m^3$, as many as three more people might be expected to develop cancer per one million people exposed daily for a lifetime to 1 microgram (μg) of the chemical in 1 cubic meter (m^3) of air. UREs are considered upper-bound estimates designed to keep us from underestimating risks. The true risk may be lower and is considered unlikely to

be higher. In NATA, we multiply the model-output concentrations for an air toxic by that pollutant's URE to calculate exposure risks from that air toxic.

Upper-bound:
A likely upper limit to the true value of a quantity. This is usually not a true statistical confidence limit.

Upper-bound lifetime cancer risk:
A likely upper limit to the true probability that a person will contract cancer over a 70-year lifetime due to a given hazard (such as exposure to a toxic chemical). This risk can be measured or estimated in numerical terms (for example, one chance in a hundred).

Urban:
A county is considered "urban" if it either includes a metropolitan statistical area with a population greater than 250,000 or the U.S. Census Bureau designates more than 50 percent of the population as "urban." Note that this definition does not necessarily apply for any regulatory or implementation purpose. It is consistent with the definition EPA used in the analyses to support the Integrated Urban Air Toxics Strategy.

Volatile organic compounds (VOCs):
Chemicals emitted as gases from certain solids or liquids. VOCs are known for being common indoor air pollutants. EPA regulates VOCs in the ambient air because some cause adverse health effects and because they can react with other pollutants to form ozone and secondary air toxics. Cars and trucks, some industries, and even plants and trees emit VOCs.

Weight-of-evidence for carcinogenicity (WOE):
A system used by the EPA for characterizing the extent to which available data support the hypothesis that an agent causes cancer in humans. The approach, outlined in EPA's *Guidelines for Carcinogen Risk Assessment* (2005), considers all scientific information in determining the WOE. Five standard descriptors are used as part of the WOE narrative:
1. Carcinogenic to humans.
2. Likely to be carcinogenic to humans.
3. Suggestive evidence of carcinogenic potential.
4. Inadequate information to assess carcinogenic potential.
5. Not likely to be carcinogenic to humans.

Each of these descriptors is explained in its own glossary entry. You can read more details about WOE narratives in the NATA Technical Support Document.

Carcinogenic to humans:
This descriptor indicates strong evidence of human carcinogenicity. It covers different combinations of evidence. This descriptor is appropriate when there is convincing epidemiologic evidence of a link between human exposure and cancer. It might also be appropriate when a lesser weight of epidemiologic evidence is strengthened by other lines of evidence.

This descriptor can be used when *all* of the following conditions are met:
- there is strong evidence of an association between human exposure and either cancer or the key precursor events of the agent's mode of action, but not enough for a causal association;
- there is extensive evidence of carcinogenicity in animals;
- the mode(s) of carcinogenic action and associated key precursor events have been identified in animals; and
- there is strong evidence that the key precursor events that precede the cancer response in animals are anticipated to occur in humans and progress to tumors, based on available biological information.

Likely to be carcinogenic to humans:
This descriptor is appropriate when the weight of evidence is enough to show the potential to cause cancer in humans but does not meet all conditions necessary to be called "carcinogenic to humans." Adequate evidence consistent with this descriptor covers a broad spectrum. At one end of the spectrum is a plausible association

between human exposure to the agent and cancer and strong experimental evidence of carcinogenicity in animals. At the other, with no human data, the weight of experimental evidence shows animal carcinogenicity by a mode or modes of action that are relevant or assumed to be relevant to humans. The use of the term "likely" as a WOE descriptor does not correspond to a quantifiable probability. Moreover, additional information, for example, on mode of action, might change the choice of descriptor for the illustrated examples.

Suggestive evidence of carcinogenic potential:
This descriptor is appropriate when the weight of evidence suggests carcinogenicity, raising concern for potential carcinogenic effects in human, but the data are judged insufficient for a stronger conclusion. This descriptor covers a spectrum of evidence ranging from a positive cancer result in the only study on an agent to a single positive cancer result in an extensive database that includes negative studies in other species. Depending on the extent of the database, additional studies might or might not provide further insights.

Inadequate information to assess carcinogenic potential:
This descriptor is appropriate when available data are judged inadequate for applying one of the other descriptors. Additional studies generally would be expected to provide further insights.

Not likely to be carcinogenic to humans:
This descriptor is appropriate when the available data are considered strong enough for deciding that there is no basis for cancer concerns for humans. In some cases, there can be positive results in experimental animals, but the evidence is strong and consistent that each mode of action does not operate in humans. In other cases, the evidence can be convincing that the agent is not carcinogenic in humans or animals. "Not likely" applies only to the circumstances supported by the data. For example, an agent might be "not likely to be carcinogenic" by one route but not necessarily by another. In cases having positive animal experiment(s), but the results are judged not to be relevant to humans, the narrative discusses why the results are not relevant.

Weather Research and Forecasting (WRF) model:
A mesoscale numerical weather-prediction system for atmospheric research and weather forecasting. It can generate atmospheric conditions using real input data or idealized conditions.

Appendix B. Air Toxics Modeled in the 2014 NATA

A master pollutant list for NATA in spreadsheet format, "NATA_Pollutants.xlsx," is provided in the Supplemental Data folder on the NATA website (https://www.epa.gov/national-air-toxics-assessment). This file includes all air toxics modeled in NATA and indicates the inventory types(s) reporting them. The file also includes the toxicity values used in NATA. The names shown in this spreadsheet match the terminology used in the 1990 Clean Air Act (CAA) Amendments; for example, the file lists "chromium compounds" but does not indicate which individual compounds containing chromium were modeled, and it lists four forms of xylenes (o-, m-, p- and mixed isomers), but these were grouped and modeled as a single entity. The file also contains indications about whether cancer risks and chronic noncancer hazard quotients were estimated for each air toxic.

The spreadsheet also contains an air toxic names crosswalk and metal speciation factors used to conduct the modeling of emissions. This crosswalk contains a link between lists of air toxic names in two data bases used for NATA:

- the names used in the National Emissions Inventory (NEI); and

- the names used for NATA.

In addition, the file contains the speciation of metal chemicals based on their metal mass fractions. The metal speciation factor was used to adjust modeled mass emissions prior to modeling and conducting risk calculations because metal toxicity is usually evaluated relative to the amount of metal ion present rather than the total mass of the metal compound. Most metal and cyanide compounds are reported in the 2014 NEI as just the metal or cyanide parts; consequently, most fractions are 1, including the two cyanide compounds. If the NEI data reporters did not adjust the emissions downward to account for just the metal part, a more health-protective (higher risk) result would be obtained.

Table B-I contains the air toxics that were *not* modeled for the 2014 NATA and why. Note that although diesel PM was modeled for NATA and is included in the list of air toxics modeled, it is not categorized as a HAP in the CAA. Diesel PM emissions were computed based on PM_{10} emissions from on-road and nonroad mobile sources burning diesel or residual fuels (see Appendix C).

Note that NEI = National Emissions Inventory.

Table B-I. Pollutants excluded from NATA

Pollutant	NEI Pollutant Code (CAS Number) a	Reason for Exclusion	In Previous NATAs?
Chromium III		– In building the NEI, chromium VI is emphasized, and chromium III may be missing data.	n[b]
2,3,7,8-Tetrachlorodibenzo-p-dioxin	1746016	Dioxins and furans are not in the 2011 NEI due to uncertainty in the completeness or accuracy of the S/L/T agency data for this group of pollutants. In addition, the most significant exposure route for dioxin is ingestion, not inhalation, so dioxin's relative contribution to NATA's inhalation risk estimates likely would not be large.	n[b]
Other dioxins/furans	multiple		n[b]
Radionuclides		Radionuclides are not in the 2011 NEI due to uncertainty in the completeness or accuracy of the S/L/T agency data for this group of pollutants. In addition, the NEI currently is not compatible with emissions reported in units other than mass, and therefore suitable emissions data have not been compiled for these substances on a national scale.	n[b]
DDE	72559 incorrectly referred to in the Section 112(b) list as 3547-04-4	This pollutant was not reported to the 2014 NEI.	n[b]
Fine mineral fibers (including rockwool, slagwool and fine mineral fibers)	Fine mineral fibers: 383 Rockwool: 617 Slagwool: 616	Rockwool and slagwool have zero emissions and fine mineral fibers are excluded from previous assessments.	n[b]
Asbestos	1332214	Air concentrations of asbestos are often measured in terms of numbers of fibers per unit volume, but the NEI provides tons, which cannot be converted.	n[b]
Diazomethane	334883	This pollutant has 0 emissions in the 2014NEI.	n[b]
Beta-propiolactone	57578	This pollutant has 0 emissions in the 2014NEI.	y

[a]In most cases, the NEI pollutant code is the same as the CAS number. In a few cases (e.g., coke oven emissions) a CAS number has not been assigned, and NEI uses a unique pollutant code.
[b]Also excluded from the 2011 NATA

Appendix C. Source Classification Codes that Define Diesel Particulate Matter

This appendix lists the source classification codes (SCC) for which the PM_{10} emissions were considered to be diesel particulate matter (PM). Diesel PM emissions were computed based on PM_{10} emissions from on-road and nonroad mobile sources burning diesel or residual fuels.

Note that NEI = National Emissions Inventory.

Table C-I. Source classification codes for which PM₁₀ emissions were assigned to diesel particulate matter

NEI Category	SCC	Description
Point	28500201	Internal Combustion Engines;Railroad Equipment;Diesel;Yard Locomotives
	2270008005	*M;Off-highway Vehicle Diesel;Airport Ground Support Equipment;Airport Ground Support Equipment
Nonpoint	2280002100	*M;Marine Vessels, Commercial;Diesel;Port emissions
	2280002200	*M;Marine Vessels, Commercial;Diesel;Underway emissions
	2285002006	*M;Railroad Equipment;Diesel;Line Haul Locomotives: Class I Operations
	2285002007	*M;Railroad Equipment;Diesel;Line Haul Locomotives: Class II / III Operations
	2285002008	*M;Railroad Equipment;Diesel;Line Haul Locomotives: Passenger Trains (Amtrak)
	2285002009	*M;Railroad Equipment;Diesel;Line Haul Locomotives: Commuter Lines
	2285002010	*M;Railroad Equipment;Diesel;Yard Locomotives
	2280003100	*M;Marine Vessels, Commercial;Residual;Port emissions
	2280003200	*M;Marine Vessels, Commercial;Residual;Underway emissions
	2270001060	*M;Off-highway Vehicle Diesel;Recreational Equipment;Specialty Vehicles/Carts
	2270002003	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Pavers
	2270002006	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Tampers/Rammers
	2270002009	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Plate Compactors
	2270002015	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Rollers
	2270002018	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Scrapers
	2270002021	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Paving Equipment
	2270002024	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Surfacing Equipment
	2270002027	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Signal Boards/Light Plants
	2270002030	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Trenchers
	2270002033	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Bore/Drill Rigs
	2270002036	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Excavators
	2270002039	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Concrete/Industrial Saws
	2270002042	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Cement and Mortar Mixers
	2270002045	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Cranes
	2270002048	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Graders
	2270002051	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Off-highway Trucks
	2270002054	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Crushing/Processing Equipment
	2270002057	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Rough Terrain Forklifts

NEI Category	SCC	Description
	2270002060	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Rubber Tire Loaders
	2270002066	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Tractors/Loaders/Backhoes
	2270002069	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Crawler Tractor/Dozers
	2270002072	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Skid Steer Loaders
	2270002075	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Off-highway Tractors
	2270002078	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Dumpers/Tenders
	2270002081	*M;Off-highway Vehicle Diesel;Construction and Mining Equipment;Other Construction Equipment
	2270003010	*M;Off-highway Vehicle Diesel;Industrial Equipment;Aerial Lifts
	2270003020	*M;Off-highway Vehicle Diesel;Industrial Equipment;Forklifts
	2270003030	*M;Off-highway Vehicle Diesel;Industrial Equipment;Sweepers/Scrubbers
	2270003040	*M;Off-highway Vehicle Diesel;Industrial Equipment;Other General Industrial Equipment
	2270003050	*M;Off-highway Vehicle Diesel;Industrial Equipment;Other Material Handling Equipment
	2270003060	*M;Off-highway Vehicle Diesel;Industrial Equipment;ACRefrigeration
	2270003070	*M;Off-highway Vehicle Diesel;Industrial Equipment;Terminal Tractors
	2270004031	*M;Off-highway Vehicle Diesel;Lawn and Garden Equipment;Leafblowers/Vacuums (Commercial)
	2270004036	*M;Off-highway Vehicle Diesel;Lawn and Garden Equipment;Snowblowers (Commercial)
	2270004046	*M;Off-highway Vehicle Diesel;Lawn and Garden Equipment;Front Mowers (Commercial)
	2270004056	*M;Off-highway Vehicle Diesel;Lawn and Garden Equipment;Lawn and Garden Tractors (Commercial)
	2270004066	*M;Off-highway Vehicle Diesel;Lawn and Garden Equipment;Chippers/Stump Grinders (Commercial)
	2270004071	*M;Off-highway Vehicle Diesel;Lawn and Garden Equipment;Turf Equipment (Commercial)
	2270004076	*M;Off-highway Vehicle Diesel;Lawn and Garden Equipment;Other Lawn and Garden Equipment (Commercial)
	2270005010	*M;Off-highway Vehicle Diesel;Agricultural Equipment;2-Wheel Tractors
	2270005015	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Agricultural Tractors
	2270005020	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Combines
	2270005025	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Balers
	2270005030	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Agricultural Mowers
	2270005035	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Sprayers
	2270005040	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Tillers >6 HP
	2270005045	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Swathers
	2270005055	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Other Agricultural Equipment
	2270005060	*M;Off-highway Vehicle Diesel;Agricultural Equipment;Irrigation Sets
	2270006005	*M;Off-highway Vehicle Diesel;Commercial Equipment;Generator Sets

NEI Category	SCC	Description
	2270006010	*M;Off-highway Vehicle Diesel;Commercial Equipment;Pumps
	2270006015	*M;Off-highway Vehicle Diesel;Commercial Equipment;Air Compressors
	2270006020	*M;Off-highway Vehicle Diesel;Commercial Equipment;Gas Compressors
	2270006025	*M;Off-highway Vehicle Diesel;Commercial Equipment;Welders
	2270006030	*M;Off-highway Vehicle Diesel;Commercial Equipment;Pressure Washers
	2270006035	*M;Off-highway Vehicle Diesel;Commercial Equipment;Hydro-power Units! new SCC in 2002v2
	2270007010	*M;Off-highway Vehicle Diesel;Logging Equipment;Shredders >6 HP
	2270007015	*M;Off-highway Vehicle Diesel;Logging Equipment;Forest Eqp - Feller/Bunch/Skidder
	2270009010	*M;Off-highway Vehicle Diesel;Underground Mining Equipment;Other Underground Mining Equipment
	2270010010	*M;Off-highway Vehicle Diesel;Industrial Equipment;Other Oil Field Equipment
	2282020005	*M;Pleasure Craft;Diesel;Inboard/Sterndrive
	2282020010	*M;Pleasure Craft;Diesel;Outboard
	2285002015	*M;Railroad Equipment;Diesel;Railway Maintenance
On-road	2202210181	*D;Passenger Cars;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202210281	*D;Passenger Cars;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202210381	*D;Passenger Cars;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202210481	*D;Passenger Cars;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202210581	*D;Passenger Cars;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202310181	*D;Passenger Trucks;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202310281	*D;Passenger Trucks;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202310381	*D;Passenger Trucks;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202310481	*D;Passenger Trucks;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202310581	*D;Passenger Trucks;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202320181	*D;Light Commercial Trucks;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202320281	*D;Light Commercial Trucks;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202320381	*D;Light Commercial Trucks;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202320481	*D;Light Commercial Trucks;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202320581	*D;Light Commercial Trucks;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202410181	*D;Intercity Buses;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202410281	*D;Intercity Buses;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202410381	*D;Intercity Buses;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202410481	*D;Intercity Buses;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling

NEI Category	SCC	Description
	2202410581	*D;Intercity Buses;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202420181	*D;Transit Buses;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202420281	*D;Transit Buses;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202420381	*D;Transit Buses;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202420481	*D;Transit Buses;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202420581	*D;Transit Buses;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202430181	*D;School Buses;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202430281	*D;School Buses;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202430381	*D;School Buses;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202430481	*D;School Buses;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202430581	*D;School Buses;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202510181	*D;Refuse Trucks;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202510281	*D;Refuse Trucks;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202510381	*D;Refuse Trucks;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202510481	*D;Refuse Trucks;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202510581	*D;Refuse Trucks;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202520181	*D;Single Unit Short-haul Trucks;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202520281	*D;Single Unit Short-haul Trucks;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202520381	*D;Single Unit Short-haul Trucks;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202520481	*D;Single Unit Short-haul Trucks;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202520581	*D;Single Unit Short-haul Trucks;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202530181	*D;Single Unit Long-haul Trucks;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202530281	*D;Single Unit Long-haul Trucks;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202530381	*D;Single Unit Long-haul Trucks;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202530481	*D;Single Unit Long-haul Trucks;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202530581	*D;Single Unit Long-haul Trucks;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202540181	*D;Motor Homes;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202540281	*D;Motor Homes;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202540381	*D;Motor Homes;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202540481	*D;Motor Homes;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202540581	*D;Motor Homes;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202610181	*D;Combination Short-haul Trucks;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling

NEI Category	SCC	Description
	2202610281	*D;Combination Short-haul Trucks;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202610381	*D;Combination Short-haul Trucks;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202610481	*D;Combination Short-haul Trucks;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202610581	*D;Combination Short-haul Trucks;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202620153	*D;Combination Long-haul Trucks;Off-network: Extended Idle Exhaust
	2202620181	*D;Combination Long-haul Trucks;Off-network: All Exhaust, Evaporative, Brake, and Tire Except Refueling and Hoteling
	2202620191	*D;Combination Long-haul Trucks;Off-network: Auxiliary Power Exhaust
	2202620281	*D;Combination Long-haul Trucks;Rural Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling and
	2202620381	*D;Combination Long-haul Trucks;Rural Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202620481	*D;Combination Long-haul Trucks;Urban Restricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling
	2202620581	*D;Combination Long-haul Trucks;Urban Unrestricted Access: All Exhaust, Evaporative, Brake, and Tire Except Refueling

*M= Mobile Sources. *D=Highway Vehicles–Diesel

Appendix D. Estimating Background Concentrations for the 2014 NATA

The memo in this appendix contains the methods we used to estimate background concentrations for NATA 2014.

Technical Memorandum

August 17, 2016   STI-915217-6569-TM

To: Madeleine Strum, U.S. Environmental Protection Agency

From: Michael McCarthy

Re: NATA Remote Background Estimates Updated for 2014

Background

EPA contracted with STI to update the National Air Toxics Assessment (NATA) remote background estimates developed for the 2011–2014 values. In previous work, STI developed methods for estimating background concentrations for 2011 (McCarthy et al., 2015).

For the 2011 NATA, we investigated multiple types of "background" concentration estimates; for this project, we updated only the Remote Northern Hemisphere background estimates with 2014 values.

Remote Northern Hemisphere (NH) background is the annual mean concentration at remote receptor areas not impacted by local-scale (50 km) or regional-scale emissions of the pollutant. Examples of remote locations include the summit of Mauna Loa, the Aleutian Islands in Alaska, portions of the Pacific Coast between Oregon and California, and some parts of the Mountain West. These estimates represent the lowest concentrations that would be observed in the United States.

Methods

STI used concentrations from networks with monitoring sites in remote locations, where little influence from regional emissions is expected, to estimate background concentrations for some of the most important (from a risk perspective) air toxics. Remote National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Division (GMD) sites[1] and the Trinidad Head Advanced Global Atmospheric Gases Experiment (AGAGE) site[2] primarily measure gases, and remote IMPROVE sites[3] measure metals.

---

[1] www.esrl.noaa.gov/gmd/
[2] http://agage.mit.edu/
[3] http://vista.cira.colostate.edu/improve/

Annual mean concentrations for 2014 were generated at five NOAA GMD sites: Cape Kumukahi, HI (KUM); Mauna Loa, HI (MLO); Niwot Ridge, CO (NWR); Barrow, AK (BRW); and Alert, Canada (ALT). Other remote sites are available but were not selected, to be consistent with previous NATA background estimates. The annual mean concentrations were cosine-weighted by latitude to generate annual mean background estimates for carbon tetrachloride, benzene, dichloromethane, tetrachloroethene, and bromomethane. Measurement methods from this network have been reported in peer-reviewed articles such as Montzka et al. (2011; 1999). Cosine-weighted latitude averaging accounts for differences in the amount of NH air (i.e., there is more air at the equator then at the poles).

Annual mean concentrations of chloroform for 2014 were measured at AGAGE network sites. While multiple AGAGE sites are available in remote locations such as Mace Head, Ireland, and Cape Grim, Tasmania, the only measurements representative of the eastern Pacific Ocean are made at Trinidad Head, California; data from this site were used to generate remote concentration estimates. AGAGE data can be accessed at http://cdiac.ornl.gov/ndps/alegage.html (DOI: 10.3334/CDIAC/atg.db1001) and the original reference for the network is available from Prinn et al. (2000).

Metals are measured at the IMPROVE sites across the United States. Some of these sites are representative of clean air coming off the Pacific Ocean and were used to generate remote background concentration estimates for lead, manganese, and nickel. Annual mean concentrations from 2014 were generated for data from the Denali, Alaska (DENA); Kalmiopsis, Oregon (KALM); Point Reyes, California (PORE); Redwoods, California (REDW); Trapper Creek, Alaska (TRCR); Tuxedni, Alaska (TUXE); and Haleakala, Hawaii (HACR) sites. These were cosine-latitude averaged to generate mean concentrations for the toxics measurements made at those sites. The IMPROVE network data were used as available from the air toxics monitoring archive. Recent work on IMPROVE detection limits was described by Hyslop and White (2011).

Results

Estimates of NH concentrations from the remote network are presented in Table 1. In contrast to the 2011 table, we now present concentrations for the gas-phase species in units of pptv, $\mu g/m^3$ at standard temperature and pressure, and $\mu g/m^3$ at local conditions. Average concentrations of metals measured at IMPROVE sites were almost all below method detection limits (MDL) when using MDL/2 substitution for non-detect values. Thus, the remote concentration estimates should be considered upper-limit thresholds for concentrations.

For chromium VI, the total chromium estimate was multiplied by the average ratio of chromium VI:chromium TSP (0.0125) seen in air toxics archive measurements for previous years. This is still an upper-limit threshold, as measurements at IMPROVE sites were below the average MDL at those sites in 2014.

Input files and R code used to generate these estimates are provided as separate electronic file attachments in the delivery email.

Table 1. Remote concentration estimates (RCE) for 2014.

Pollutant	RCE STP[a] (μg/m³) 2014	RCE LC[b] (μg/m³) 2014	RCE (pptv) 2014	Remote Network	Location(s)
Chloroform	0.070	0.072	14.3	AGAGE	Trinidad Head
Methyl chloride (chloromethane)	1.16	0.99	563	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Benzene	0.109	0.096	34.0	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Carbon tetrachloride	0.530	0.453	84.4	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Methyl bromide (bromomethane)	0.028	0.024	7.2	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Methyl chloroform (1,1,1-trichloroethane)	0.020	0.017	3.7	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Dichloromethane (methylene chloride)	0.20	0.17	58.3	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Tetrachloroethene (perchloroethylene, tetrachloroethylene)	0.013	0.011	1.9	NOAA GMD	KUM, MLO, NWR, BRW, ALT
Chromium VI	<1.4E-06			IMPROVE Cr and NATTS Cr VI:Cr ratio	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Arsenic	<2.1E-4			IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Chromium	<1.1E-4			IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Lead	<6.9E-04			IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Manganese	3.8E-04			IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Nickel	<1.0E-4			IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR

[a] Standard temperature and pressure, 298 K and 101.325 kPa
[b] Local conditions – pressure was calculated using elevation and temperature based on average annual temperature measured at the site.

References

Hyslop N.P. and White W.H. (2011) Identifying sources of uncertainty from the inter-species covariance of measurement errors. *Environ. Sci. Technol.*, 45(9), 4030-4037, doi: 10.1021/es102605x. Available at http://pubs.acs.org/doi/abs/10.1021/es102605x.

McCarthy M., O'Brien T., Du Y., and Russell A. (2015) Methods for estimating background concentrations for the National Air Toxics Assessment (NATA) 2011. Final report prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-915110-6315, August 13.

Montzka S.A., Butler J.H., Elkins J.W., Thompson T.M., Clarke A.D., and Lock L.T. (1999) Present and future trends in the atmospheric burden of ozone-depleting halogens. *Nature*, 398, 690-694.

Montzka S.A., Dlugokencky E.J., and Butler J.H. (2011) Non-CO_2 greenhouse gases and climate change. *Nature*, 476(7358), 43-50, doi: 10.1038/nature10322, August 3.

Prinn R.G., Weiss R.F., Fraser P.J., Simmonds P.G., Cunnold D.M., Alyea F.N., O'Doherty S., Salameh P., Miller B.R., Huang J., Wang R.H.J., Hartley D.E., Harth C., Steele L.P., Sturrock G., Midgley P.M., and McCulloch A. (2000) A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *J. Geophys. Res*, 105, 17,751-717,792.

Appendix E. Model Evaluation Summaries

In addition to the evaluations show in Section 3.7, EPA performed model evaluations for other NATA pollutants. These evaluations, including graphics, can be found in the Supplemental Data folder on the NATA website.

Appendix F.   Exposure Factors for the 2014 NATA

The memorandum contained within this Appendix (below) describes in detail how EPA developed exposure factors for each chemical assessed in NATA. These calculated exposure factors can be found in the Supplemental Data folder accompanying this TSD.

MEMORANDUM

To: Ted Palma, U.S. EPA Office of Air Quality Planning and Standards

From: John Hader and Chris Holder, ICF

Date: June 25, 2018

Re: HAPEM7 Modeling for the 2014 National Air Toxics Assessment (NATA)

The National Air Toxics Assessment (NATA) is a nationwide EPA modeling assessment of the air concentrations, exposure concentrations,[1] and potential human health cancer and noncancer risks and hazards associated with inhalation exposure to Hazardous Air Pollutants (HAPs) emitted by manmade and natural sources of outdoor origin. As part of NATA, the Hazardous Air Pollutant Exposure Model (HAPEM), which is a U.S. Environmental Protection Agency (EPA) model, is used to perform screening-level assessments of long-term inhalation exposures to HAPs.

For the 2014 NATA, ICF ("we") used **Version 7 of HAPEM (HAPEM7),** with modeled air-concentration data provided by EPA, to **model exposure concentrations for seven selected HAPs, stratified by census tract and emission source type**. These HAPs also are used as surrogates for the many other HAPs included in NATA, as described in detail below. We used the annual-average exposure concentrations by age group, provided by the HAPEM modeling, to estimate lifetime-average exposure concentrations. **We then divided the lifetime exposure concentrations by the corresponding annual-average air concentrations creating "exposure factors".** EPA can then estimate exposure to each NATA HAP in each census tract by multiplying the air concentration for the HAP by the exposure factor derived for the chemical's surrogate HAP.

In this memorandum, we discuss HAPEM7, how we assigned the gas or particulate phase of the NATA HAPs, how EPA selected the seven HAPs modeled in HAPEM, how we set up the HAPEM runs, and how we developed the exposure factors.

---

[1] Exposure concentrations are time-averaged air concentrations to which a simulated individual is exposed. The time averaging considers the air concentrations in each place where the individual is simulated to spend time, and how long he or she spends there.

1. Introduction to HAPEM and its Use in NATA

HAPEM is a model used by EPA to perform screening-level assessments of long-term inhalation exposures to HAPs. We updated HAPEM and its default, ancillary files in 2015, as discussed in the 2011 NATA documentation.[2] This latest version of HAPEM (HAPEM7)[3] estimates exposure concentrations using

- demographic and behavior data from the 2010 U.S. Census (covering all 50 U.S states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands),[4]

- compiled activity data from a recent version of EPA's Consolidated Human Activity Database (CHAD),[5] and

- updated data on population proximities to major roadways.

HAPEM7 stratifies exposure concentrations by

- location (U.S. Census tract),

- time of day,

- age group, and

- the individual emission source types and HAPs being modeled.

HAPEM is a probabilistic model that selects some model inputs from distributions of data to characterize exposure. These elements include commuting patterns, activity patterns, employment work location, microenvironmental factors, and (if desired) air concentrations. Including more simulated individuals (termed "replicates" in HAPEM) can increase the range of values selected for these probabilistic elements, across the simulated population.

---

[2] Appendix G of the Technical Support Document of EPA's 2011 National Air Toxics Assessment. December 2015. EPA Office of Air Quality Planning and Standards. Available at https://www.epa.gov/sites/production/files/2015-12/documents/2011-nata-tsd.pdf.

[3] As of October 23, 2017, HAPEM7 is available for download from https://www.epa.gov/fera/human-exposure-modeling-hazardous-air-pollutant-exposure-model-hapem. The HAPEM7 User's Guide also is available from that link, and it contains detailed discussions of the model's functionality and input files.

[4] For additional information on the census data parameterized in HAPEM, see Appendix B of The HAPEM User's Guide, Version 7 (July, 2015), available at https://www.epa.gov/sites/production/files/2015-12/documents/hapem7usersguide.pdf.

[5] For additional information on CHAD and how its data were parameterized in HAPEM, see https://www.epa.gov/healthresearch/consolidated-human-activity-database-chad-use-human-exposure-and-health-studies-and and Appendix B of The HAPEM User's Guide, Version 7 (July, 2015), available at https://www.epa.gov/sites/production/files/2015-12/documents/hapem7usersguide.pdf.

ICF

2. Air Quality Inputs to HAPEM7

2.1. HAPs Assessed for the 2014 NATA

For the 2014 NATA, EPA compiled an inventory of the annual mass of HAPs emitted from four NATA source types, namely

- point,

- non-point,

- on-road mobile, and

- non-road mobile.

These emissions were nationwide, typically at spatial resolutions of 12-km grids down to specific points. The assessed HAPs are shown in the far left column of Table 1 (under the "Chem" heading). In some cases, emissions of several chemicals were aggregated and modeled as a group, as indicated on the right side of Table 1 (under the "NATA" heading; the columns under the "NATA" heading provide the chemicals and chemical groups as they were modeled for NATA). Using these estimated emissions, temporal emission profiles based on source type, and hourly meteorology data selected by proximity, EPA used AERMOD[6] and CMAQ[7] to estimate ambient air concentrations. **The EPA air-modeling results provided for HAPEM exposure modeling were annual-average air concentrations at the level of census tracts, stratified by HAP, source type, and at three-hour increments across the day.**

2.2. Specifying Chemical Phases for HAPEM Modeling

For all source types, **HAPEM7 models concentrations for three chemical phases**. These phases are

- gaseous ("G"),

- particulate ("P"), and

- variable for typical atmospheric conditions ("G/P").

HAPEM estimates concentrations for user-specified microenvironments (ME). An ME is a generic location, such as indoors at home or outdoors near roadway, where a person spends time and breathes air, leading to chemical exposure. In HAPEM, MEs are treated as having well-mixed air concentrations at a given time, and the model estimates those ME concentrations relative to the outdoor air concentrations provided as input to the model. HAPEM7 is parameterized for 18 MEs, including eight indoor MEs (including residential, school, office,

---

[6] AERMOD: the atmospheric dispersion model developed by the American Meteorological Society and the EPA Regulatory Model Improvement Committee. See https://www.epa.gov/scram/air-quality-dispersion-modeling-preferred-and-recommended-models#aermod

[7] CMAQ: EPA's Community Multiscale Air Quality model. See https://www.epa.gov/cmaq

ICF

bar/restaurant, and similar), two in-vehicle MEs (personal vehicle and public transit), six outdoor MEs near roadways, and two outdoor MEs not near roadways.

Starting with outdoor ambient chemical concentrations, the model uses phase-specific penetration and proximity factors to estimate the ME chemical concentrations. A penetration factor describes how much of the chemical in the outdoor air penetrates the indoor or in-vehicle air. The penetration factor is a ratio in the form of [chemical air concentration in an indoor/in-vehicle ME] to [chemical air concentration directly outdoors of the ME]. A proximity factor parameterizes the heterogeneity of air concentrations across a Census tract. It describes how much higher or lower we expect the outdoor air concentration to be in the immediate vicinity of the ME relative to the Census tract outdoor concentration supplied in the input file, and it is a ratio in the form of [outdoor chemical air concentration in the immediate vicinity of the ME] to [tract-level outdoor chemical air concentration input to the model]. HAPEM7 has one set of penetration and proximity factors for mobile on-road sources, and a second set for all other source types. For mobile on-road sources, there are four categories of factors: three are for specific HAPs (benzene, 1,3-butadiene, and diesel particulate matter [DPM][8]) and a fourth category representing all other HAPs.

To prepare for NATA exposure modeling, **we categorized each of the NATA HAPs as G, P, or G/P based on available boiling-point data**, as defined in Table 2. We provide each HAP's boiling point and assigned HAPEM7 phase in Table 1 (left side under the "Chem" heading). If chemicals were modeled as a group, we assigned the group a G, P or G/P designation, as indicated in Table 1 (right side under the "NATA" heading) and as discussed later in this section.

We obtained the vast majority of boiling-point values from

- the Centers for Disease Control (CDC),[9]

- the National Institutes of Health (NIH),[10] or

- the Royal Society of Chemistry using their ChemSpider web site.[11]

These were judged to be the most reputable, comprehensive, user-friendly, and readily available sources of chemical boiling-point data. Each source allows the user to search by chemical name or Chemical Abstract Service (CAS) Registry Number. For HAPs whose boiling points we could not identify using these three sources, we searched a variety of additional data sources. We provide the source of each chemical's boiling-point value in the "Source" column of Table 1.

We could not identify empirical boiling-point data for some of the HAPs. For 25 HAPs (9 percent of all evaluated HAPs), we used predictive boiling points in order to categorize their chemical phases. These values came from ChemSpider, which generates estimated boiling points using three software modules: EPA's EPIsuite, the ACD/Labs Percepta Platform - PhysChem Module, and ChemAxon's predictive software platform. We typically selected the ACD/Labs values when available because these values were presented with confidence intervals and the conditions

---

[8] DPM is not a HAP as defined under Section 112(b) of the Clean Air Act; however, for ease of discussion we refer to it as a HAP in this memorandum.
[9] CDC: http://www.cdc.gov/niosh/npg.
[10] NIH: http://pubchem.ncbi.nlm.nih.gov/.
[11] CS: http://www.chemspider.com.

ICF

under which the boiling-point values were predicted (typically standard temperature and pressure); the other two platforms did not provide such information. If an ACD prediction was not available, we used the EPA EPIsuite value. Those boiling points that are predictive rather than empirical are flagged with a "P" in Table 1.

Note that the boiling-point ranges in Table 2 have imprecise endpoints (e.g., the high end of boiling points for G HAPs covers a range of 240 to 260°C). Forty of NATA HAPs have boiling points within these imprecise endpoints, meaning there was some uncertainty associated with assigning the phases for these HAPs. In order to make appropriate phase designations, we conducted a literature review for each of these HAPs to identify relevant information regarding its typical physical state. For example, 1-nitropyrene has a boiling point of 445°C, within the overlap of G/P and P boiling points. A review of the literature yielded several studies and reports identifying 1-nitropyrene as a particulate at typical atmospheric conditions, leading us to assign a designation of "P" to this HAP with a high degree of confidence. Where literature searches were uninformative, we assigned HAP phase based on the categorizations used for HAPEM5 to support the 1999 NATA. The combination of the additional literature review and consultation of the HAPEM5 designations allowed us to make phase designations for these HAPs. Seventeen HAPs have boiling points within the 240–260°C range; based on the literature review and HAPEM5 designations, we categorized 16 as G and 1 as G/P. Twenty HAPs have boiling points within the 400–480°C range; we categorized 16 as P and 4 as G/P. Three HAPs have wide-ranging boiling-point values; we categorized the two with higher ranges of boiling points (extending into the 400–480°C range) as G/P, and we categorized the one with a lower range of boiling points (extending only into the 240–260°C range) as G.

In addition to the above 40 HAPs, boiling-point values were specified by the CDC or EPA as "variable" for three HAPs (see the "V" designations in the boiling-point-value column in Table 1). We categorized two of these HAPs (coke oven emissions and cyanide) as G/P. We categorized coke oven emissions based on an EPA characterization of this pollutant as consisting of a mixture of particulates, volatiles, and semi-volatiles.[12] We characterized cyanide based on the fact that cyanide is not typically found in isolation in nature, but rather in a variety of compounds, some of which are typically solid (e.g., calcium cyanide, sodium cyanide) and some of which are typically gaseous (e.g., hydrogen cyanide).[13] The third HAP with "variable" boiling points was DPM and we assumed that it is typically a particulate.

For five HAPs, boiling-point data were either unavailable or were ill-defined (see the "NA", "D", and "S" designations in the boiling-point-value column in Table 1); one was Fine Mineral Fibers, which we categorized as P, and we left the remaining four (Extractable Organic Matter (EOM), propoxur, quinone, and toxaphene) uncategorized. For these four HAPs, EPA was health protective and set exposure concentrations equal to ambient outdoor concentrations per census tract, as indicated by "EF=1" designations in the "Matching Chemical Modeled in HAPEM" columns in Table 1. A setting of EF=1 is health-protective because it assumes that people always breathe outdoor ambient air, receiving no protection that may otherwise be afforded by the barriers and filtration systems of a building or vehicle (that is, a penetration factor of 1 rather than below 1). Without indoor sources of emissions (which are not included in NATA), outdoor air concentrations necessarily are the same or higher than those of indoor or in-vehicle MEs.

---

[12] https://www.epa.gov/sites/production/files/2016-01/documents/coke_rra.pdf
[13] http://www.atsdr.cdc.gov/

ICF

Where EPA modeled several HAPs as one group, we assigned the group's chemical phase based on the phase most frequently assigned to the group's component chemicals. Two NATA chemical groups did not easily accommodate this criterion, as discussed in the bullets below.

- For NATA chemical group CHROMHEX, its two components are Chromium (VI) and Chromium Trioxide, which are P and G/P phases, respectively. We assumed that Chromium (VI) was the primary chemical in the group (the one emitted in higher amounts) and assigned the group to P accordingly.

- For NATA chemical group CYANIDE, its two components are cyanide and hydrogen cyanide (HCN). Cyanide has a variable boiling point, while HCN's boiling point is 26ºC (making it a G chemical). Because HCN has the higher level of emissions, we assigned the group to G accordingly.

2.3. HAPs Modeled in HAPEM7 for the 2014 NATA

As noted previously, except for air-concentration data, HAPEM7 does not use any settings or inputs that are specific to each individual NATA HAP. Penetration and proximity factors vary only by the phase of the HAP, not by each HAP individually (except for mobile on-road sources, where benzene, 1,3-butadiene, and DPM have specific penetration and proximity factors). Therefore, except for those three HAPs for mobile on-road sources, the model inputs for HAPs of a given phase vary only by air concentration, and so their exposure concentrations output by HAPEM (i.e., the time-averaged air concentrations across the MEs where individuals are simulated to spend time) vary only by their air concentrations. This means that we can estimate the exposure concentration of one HAP of a given phase in HAPEM, and then in post-processing we can calculate the relationship between its air and exposure concentrations and apply this relationship to other HAPs of the same phase, significantly reducing the level of effort needed to conduct the exposure modeling.

EPA identified a small subset of NATA HAPs for which to conduct HAPEM exposure modeling. NATA uses the modeling results of this subset as-is, and uses the results as surrogate values for the remaining NATA HAPs not modeled in HAPEM. EPA used the following decision criteria in identifying this subset of HAPs to model.

1. The subset must include at least one HAP per phase (at least one G HAP, one P HAP, and one G/P HAP).

2. Each emission source type must be represented for each phase (for example, G HAPs must collectively cover the point, non-point, on-road mobile, and non-road mobile source types, and so on for P and G/P HAPs).

3. It is preferred that the selected HAPs have the potential to pose higher risks to human health nationwide, relative to other HAPs.

4. It is preferred that the selected HAPs be emitted in many spatially-diverse locations across the United States.

Using these criteria, **EPA selected seven HAPs for exposure modeling**. We list these HAPs below and in Table 3. The right four columns of Table 1 indicate how these modeled HAPs were mapped to the other NATA HAPs not modeled in HAPEM.

- **Benzene** and **1,3-butadiene** are emitted by many processes (and all four modeled source types) in nearly all U.S. locations. They each also have specific penetration and proximity factors for emissions from mobile on-road sources. **Benzene was selected to be the surrogate for all other G HAPs**.

- Unspeciated, generic PAHs ("**PAH, total**") are emitted by all four source types and from a wide variety of processes, so it was **selected to be the surrogate for all G/P HAPs except coke oven emissions,** which is a special case that was modeled by itself for point sources but not used as a surrogate for any other HAPs.

- **Chromium (VI)** is a highly toxic HAP that was **selected as the surrogate for all P HAPs emitted by point and non-point sources, except for DPM,** which was modeled as itself for non-point and mobile sources. DPM has specific penetration and proximity factors for emissions from mobile on-road sources. Note that the NATA air-concentration modeling included chromium (VI) emissions from all four source types, but its use as an exposure surrogate included only point and non-point sources because those are its major emitters.

- **Nickel was selected as the surrogate for P HAPs emitted by mobile sources**, except for DPM, which was modeled as itself for non-point and mobile sources. Nickel is emitted by a variety of processes across the United States. Note that NATA air-concentration modeling included nickel emissions from all four source types, but its use as an exposure surrogate included only mobile sources because chromium (VI) was designated as the more appropriate surrogate for point and non-point sources.

We used the air-concentration modeling outputs provided by EPA for these seven HAPs, stratified by source type, three-hour increment of the day, and census tract, as the air-quality inputs for **seven HAPEM7 runs**.

3. HAPEM7 Runs

3.1. Design

For each of the seven HAPEM7 runs (one run for each modeled chemical), we used the HAPEM7 **default census- and CHAD-based files**. Each run assessed the **18 HAPEM7 MEs** and **all populated census tracts in the United States, Puerto Rico, and the U.S. Virgin Islands**. The air-quality inputs were in three-hour periods for each HAP, tract, and source type (that is, for each day, average air concentrations for the first three hours of the day, the second three hours, etc., totaling eight values per day for each HAP, tract, and source type). We used the **six default HAPEM7 age groups** (discussed below in Section 3.3) and **three day-types** (summer weekdays, non-summer weekdays, and weekends). We linked each HAP to its appropriate HAPEM7 penetration- and proximity-factors files, and we used **30 replicates (simulated individuals) per age group** evaluated per tract. ICF has previously shown that the tract-mean exposure concentrations of an age group are reasonably stable when using 30 replicates, striking a good balance between model runtime and the stability of model outputs (that is, at a 95-percent probability level, 30 replicates are needed to estimate the tract-mean exposures to within 20 percent).[14]

---

[14] Source: "Benzene Case Study for PO3-NTA006-ICF", a technical memorandum from Arlene Rosenbaum, Michael Huang, and Jonathan Cohen (ICF) to Ted Palma (U.S. EPA), September 30, 2002.

ICF

3.2. Quality Assurance and Quality Control

We reviewed the HAPEM7 control files ("parameters" files) for accuracy, and then we reviewed the log, "counter," and "mistract" HAPEM7 output files to identify any potential errors in the modeling. One tract identifier was modified by the U.S. Census after 2010 (after the HAPEM7 population files were developed using the 2010 census)—the 2014 NATA's air-concentration data used the tract's latest identifier of 51019050100, which we renamed to its old identifier (51515050100) for HAPEM7 modeling (its geography did not change). We switched it back to the current identifier in the results delivered to EPA. We detected no other errors in the inputs or outputs. We present below additional information gleaned from the QA/QC activities.

- EPA modeled all U.S. census tracts to develop air concentration estimates. However, 584 tracts (less than 1 percent of U.S. tracts) were not modeled in HAPEM because the 2010 Census indicated zero population.

- A total of 73,450 tracts were modeled for exposure concentrations for the 2014 NATA.

3.3. Post-processing for Exposure Factors

We utilized the HAPEM7 outputs for the seven modeled HAPs to estimate exposure factors that EPA then applies to all HAPs assessed in NATA, based on HAP phase and source type. For each tract and source type in a HAP-specific HAPEM7 run, we calculated the estimated **70-year lifetime-average exposure concentration for each modeled replicate**. We calculated these average concentrations as the time-weighted average of exposures for the six HAPEM7 age groups, as shown below.

```
Lifetime average exposure conc. =
   [ages 00-01 exposure conc. × (2/70)]
 + [ages 02-04 exposure conc. × (3/70)]
 + [ages 05-15 exposure conc. × (11/70)]
 + [ages 16-17 exposure conc. × (2/70)]
 + [ages 18-64 exposure conc. × (47/70)]
 + [ages 65+   exposure conc. × (5/70)]
```

We then calculated the **population-median lifetime-average exposure concentration in each tract** (median calculated from the 30 replicates in a tract), stratified by HAP and source type. For a given HAP and tract, summing these median concentrations across source types yields an estimated "typical" lifetime-average HAP exposure concentration from all contributing sources.

For each assessed HAP in a tract, **we calculated exposure factors by dividing these median lifetime-exposure concentrations by the corresponding annual-average outdoor air concentrations** (stratified by source type as well as the all-source total). **EPA can then multiply these exposure factors by the air concentrations of any HAP of the same phase-type, resulting in estimated lifetime-exposure concentrations for that HAP**. For example, for a given tract, to estimate the exposure concentrations of fluorene (a G/P HAP) emitted by non-point sources, EPA will multiply its non-point air concentrations by the non-point exposure factors for "PAH, total" (the surrogate for G/P HAPs). The exposure factors typically were between roughly 0.4 and 1.0 (larger factors generally for on-road mobile sources and gases; smaller factors typically for the other source types and particulates).

ICF

The primary limitation of this exposure-factor method is related to commuting. HAPEM7 simulates the movement of replicates to and from work. Some replicates will work in the same tract where they live, so that the outdoor air concentrations of their "home tract" equal those of their "work tract". For these replicates, the exposure-factor method is accurate, because the exposure factors account only for home-tract exposure to the modeled HAP, and the air-concentration denominator in the factor calculation also corresponds only to home-tract air concentrations of the unmodeled HAP. For replicates who commute outside their home tract, they work in a tract with different air concentrations than in their home tract, affecting their exposure during work hours (usually less than half the day). When the work air tract concentrations are similar to the home tract air concentrations, then the exposure factor method provides a good approximation of exposure for unmodeled HAPs. However, when work tract and home tract air concentrations are highly dissimilar, the exposure factors method estimates are less accurate because the exposure factors account for some home-tract exposures and some very different work-tract exposures to the modeled HAP, while the air-concentration denominator in the factor calculation corresponds only to home-tract air concentrations of the unmodeled HAP.

Instances of highly dissimilar work tract and home tract concentrations are most easily detected when exposure factors are above 1, indicating the work-tract air concentrations are higher than those of the home tract. For a relatively small number of tracts, exposure factors were larger than 10 and, for seven tracts, exposure factors were larger than 100. In an effort to improve the representativeness of the modeled exposure factors and their application to unmodeled HAPs, **we limited exposure factors to be no larger than the values shown in** Table 4 (exposure factors that were above these values were set to the maximum values). To define these upper limits on the exposure factors, we calculated the median exposure factor and standard deviation of exposure factors across all tracts for each combination of HAP and source type; the upper limit was then defined as *median+standard deviation*. The calculations for medians and standard deviations did not include exposure factors of 100 or larger, which we considered to be outliers. They also did not consider instances where air concentrations or exposure concentrations were zero, which occurred at 584 tracts as discussed in Section 3.2. In these instances, we set the exposure factors equal to 1.0, indicating that the exposure concentration equals the air concentration. All applications of the exposure factors for the 2014 NATA use these truncated values, including exposure factors initially set equal to 1.0 and including the HAPs explicitly modeled in HAPEM7.

4. Outputs

We performed quality assurance checks of the post-processing by thoroughly reviewing and testing the post-processing R code in order to ensure calculations were being performed correctly, and we also performed other broad checks to ensure all records were properly read in and that flagged records were properly processed.

![ICF logo]

Table 1. HAPs Assessed in the 2014 NATA, with their HAPEM7 HAP Phases and Surrogate Chemical Assignments

Chem.				NATA					
	Boiling Point					Matching Chemical Modeled in HAPEM[d]			
Name	Value (°C)[a]	Source[b]	Phase[c]	Name	Phase[c]	P	NP	M-OR	M-NR
Acetaldehyde	20	CDC	G	ACETALD	G	Benz	Benz	Benz	Benz
Acetamide	165	CDC	G	ACETAMIDE	G	Benz	Benz	Benz	Benz
Acetonitrile	82	CDC	G	ACETONIT	G	Benz	Benz	Benz	Benz
Acetophenone	202	NIH	G	ACETOPHEN	G	Benz	Benz	Benz	Benz
Acrolein	53	CDC	G	ACROLEI	G	Benz	Benz	Benz	Benz
Acrylic acid	141	CDC	G	ACRYLCACID	G	Benz	Benz	Benz	Benz
Acrylamide	175	CDC	G	ACRYLMID	G	Benz	Benz	Benz	Benz
Acrylonitrile	77	CDC	G	ACRYLONITRL	G	Benz	Benz	Benz	Benz
2-Acetylaminofluorene	400 P	CS	G/P	ACTYALFLUR2	G/P	PAH	PAH	PAH	PAH
Allyl chloride	45	CDC	G	ALLYLCHLORD	G	Benz	Benz	Benz	Benz
4-Aminobiphenyl	302	CDC	G/P	AMNOBIPNYL4	G/P	PAH	PAH	PAH	PAH
Aniline	184	CDC	G	ANILINE	G	Benz	Benz	Benz	Benz
Antimony	1,587	CDC	P	ANTIMONY	P	Cr6	Cr6	Ni	Ni
Arsenic	612	CDC	P	ARSENIC	P	Cr6	Cr6	Ni	Ni
Benzene	80	CDC	G	BENZENE	G	Benz	Benz	Benz	Benz
Benzidine	400	CDC	G/P	BENZIDINE	G/P	PAH	PAH	PAH	PAH
Benzotrichloride	221	NIH	G	BENZOTRICHL	G	Benz	Benz	Benz	Benz
Benzyl chloride	179	NIH	G	BENZYLCHLO	G	Benz	Benz	Benz	Benz
BERYLLIUM	2,500	NIH	P	BERYLLIUM	P	Cr6	Cr6	Ni	Ni
Biphenyl	256	NIH	G	BIPHENYL	G	Benz	Benz	Benz	Benz
Bis(chloromethyl)ether	106	CDC	G	BISCHLROME	G	Benz	Benz	Benz	Benz
Bromoform	149	CDC	G	BROMOFORM	G	Benz	Benz	Benz	Benz
Bis(2-ethylhexyl)phthalate	386	CDC	G/P	BS2TYLHXYLP	G/P	PAH	PAH	PAH	PAH
beta-Propiolactone	162	http://www.cdc.gov/niosh/docs/81-123/pdfs/0528.pdf	G	BT-PRPLACTN	G	Benz	Benz	Benz	Benz
1,3-Butadiene	138	NIH	G	BUTADIE	G	Buta	Buta	Buta	Buta
Cadmium	765	CDC	P	CADMIUM	P	Cr6	Cr6	Ni	Ni
Calcium cyanamide	>2,444	CDC	P	CALCIUMCYA	P	Cr6	Cr6	Ni	Ni
Captan	314	CS	G/P	CAPTAN	G/P	PAH	PAH	PAH	PAH
Carbaryl	315	CS	G/P	CARBARYL	G/P	PAH	PAH	PAH	PAH
Carbon disulfide	47	CDC	G	CARBNDISULF	G	Benz	Benz	Benz	Benz
Carbonyl sulfide	-50	NIH	G	CARBNYLSUL	G	Benz	Benz	Benz	Benz
Carbon tetrachloride	77	CDC	G	CARBONTET	G	Benz	Benz	Benz	Benz
Catechol	245	CDC	G	CATECHOL	G	Benz	Benz	Benz	Benz
Chloroform	62	CDC	G	CHCL3	G	Benz	Benz	Benz	Benz
Chloroacetic acid	106	CDC	G	CHLACETACD	G	Benz	Benz	Benz	Benz
2-Chloroacetophenone	244	CDC	G	CHLACETPH2	G	Benz	Benz	Benz	Benz
Chloramben	312	CS	G/P	CHLORAMBEN	G/P	PAH	PAH	PAH	PAH
Chlordane	175	NIH	G	CHLORDANE	G	Benz	Benz	Benz	Benz
Chlorine	-33	CDC	G	CHLORINE	G	Benz	Benz	Benz	Benz
Chloroprene	59	CDC	G	CHLOROPRENE	G	Benz	Benz	Benz	Benz
Chlorobenzilate	146	NIH	G	CHLROBZLAT	G	Benz	Benz	Benz	Benz
Chlorobenzene	132	CDC	G	CHLROBZNE	G	Benz	Benz	Benz	Benz
Chloromethyl methyl ether	59	CDC	G	CHLROMME	G	Benz	Benz	Benz	Benz
Chromium (VI)	2,642	CDC	P	CHROMHEX	P	Cr6	Cr6	Ni	Ni
Chromium Trioxide	250	CDC	G/P						

Chem.				NATA					
	Boiling Point					Matching Chemical Modeled in HAPEM[d]			
Name	Value (°C)[a]	Source[b]	Phase[c]	Name	Phase[c]	P	NP	M-OR	M-NR
Chromium III	2,672	http://books.google.com/books?id=SFD30BvPBhoC&pg=PA123&lpg=PA123&dg=chromium+III+melting+point&source=bl&ots=upHljDrKMy&sig=dISMKFL5z0sVl0z8Z4NhlsFHqgE&hl=en&sa=X&ei=4nkIVPLvJ4LS8AGbiYD4DA&ved=0CFkQ6AEwCQ#v=onepage&q=chromium%20IlI%20melting%20point&f=false	P	CHROMTRI	P	Cr6	Cr6	Ni	Ni
Ethylene dichloride	83	CDC	G	CL2_C2_12	G	Benz	Benz	Benz	Benz
Trichloroethylene	87	CDC	G	CL3_ETHE	G	Benz	Benz	Benz	Benz
Cobalt	3,100	CDC	P	COBALT	P	Cr6	Cr6	Ni	Ni
Coke oven emissions	V	CDC	G/P	COKEOVEN	G/P	Coke	Coke	Coke	Coke
Cresol/Cresylic Acid (Mixed Isomers)	202	CDC	G	CRESOLS	G	Benz	Benz	Benz	Benz
m-Cresol	202	CDC	G						
o-Cresol	191	CDC	G						
p-Cresol	202	CDC	G						
Cumene	152	CDC	G	CUMENE	G	Benz	Benz	Benz	Benz
Cyanide	V	CDC	G/P	CYANIDE	G	Benz	Benz	Benz	Benz
Hydrogen cyanide	26	CDC	G						
2,4-Dichlorophenoxy Acetic Acid	345 P	CS	G/P	D24SALTS	G/P	PAH	PAH	PAH	PAH
Diethyl sulfate	210	NIH	G	DETHLSULFAT	G	Benz	Benz	Benz	Benz
Diazomethane	-23	CDC	G	DIAZOMETHAN	G	Benz	Benz	Benz	Benz
Dibenzofuran	287	NIH	G/P	DIBENZOFUR	G/P	PAH	PAH	PAH	PAH
1,2-Dibromo-3-chloropropane	196	CDC	G	DIBRM3CHLPR	G	Benz	Benz	Benz	Benz
Dibutyl Phthalate	340	CDC	G/P	DIBUTYLPHTH	G/P	PAH	PAH	PAH	PAH
3,3'-Dichlorobenzidine	400	NIH	G/P	DICHLBZD33P	G/P	PAH	PAH	PAH	PAH
Dichlorvos	140 at 40 mmHG	NIH	G	DICHLORVOS	G	Benz	Benz	Benz	Benz
1,4-Dichlorobenzene	173	CDC	G	DICHLRBNZN	G	Benz	Benz	Benz	Benz
Dichloroethyl ether	177	CDC	G	DICHLROETET	G	Benz	Benz	Benz	Benz
1,3-Dichloropropene	108	NIH	G	DICLPRO13	G	Benz	Benz	Benz	Benz
Diesel PM	V	http://www.epa.gov/region1/eco/airtox/diesel.html	P	DIESEL_PM10	P	DPM	DPM	DPM	DPM
Diethanolamine	268	NIH	G/P	DIETHNLAMIN	G/P	PAH	PAH	PAH	PAH
Dimethylcarbamoyl chloride	167	NIH	G	DIMTHYCARB	G	Benz	Benz	Benz	Benz
N,N-Dimethylformamide	153	CDC	G	DIMTHYFORM	G	Benz	Benz	Benz	Benz
4-Dimethylaminoazobenzene	371	CS	G/P	DIMTHYLAMAZ	G/P	PAH	PAH	PAH	PAH
Dimethyl phthalate	284	NIH	G/P	DIMTHYLPHTH	G/P	PAH	PAH	PAH	PAH
Dimethyl sulfate	188	NIH	G	DIMTHYLSULF	G	Benz	Benz	Benz	Benz
3,3'-Dimethylbenzidine	300	CDC	G/P	DIMTYLBZ33P	G/P	PAH	PAH	PAH	PAH
1,1-Dimethyl Hydrazine	64	CDC	G	DIMTYLHYD11	G	Benz	Benz	Benz	Benz
3,3'-Dimethoxybenzidine	391	CS	G/P	DIMTYOXB33P	G/P	PAH	PAH	PAH	PAH
2,4-Dinitrotoluene	300	NIH	G/P	DINITOTOL24	G/P	PAH	PAH	PAH	PAH
4,6-Dinitro-o-cresol	312	CDC	G/P	DINITROCRES	G/P	PAH	PAH	PAH	PAH
2,4-Dinitrophenol	312	CS	G/P	DINITROPH24	G/P	PAH	PAH	PAH	PAH
1,2-Diphenylhydrazine	293	NIH	G/P	DIPYLHYZ12	G/P	PAH	PAH	PAH	PAH
N,N-dimethylaniline	192	CDC	G	DMTYLANILNN	G	Benz	Benz	Benz	Benz
Epichlorohydrin	118	NIH	G	EPICHLORO	G	Benz	Benz	Benz	Benz
1,2-Epoxybutane	63	NIH	G	EPOXYBUTA12	G	Benz	Benz	Benz	Benz
Ethylene dibromide	131	CDC	G	ETHDIBROM	G	Benz	Benz	Benz	Benz
Ethyleneimine	56	CDC	G	ETHENIMINE	G	Benz	Benz	Benz	Benz
Ethylene glycol	197	CDC	G	ETHGLYCOL	G	Benz	Benz	Benz	Benz

Chem.				NATA					
	Boiling Point					Matching Chemical Modeled in HAPEM[d]			
Name	Value (°C)[a]	Source[b]	Phase[c]	Name	Phase[c]	P	NP	M-OR	M-NR
Ethylidene dichloride	-17	CDC	G	ETHIDDICHLD	G	Benz	Benz	Benz	Benz
Ethylene thiourea	230	CDC	G	ETHTHUREA	G	Benz	Benz	Benz	Benz
Ethyl acrylate	99	CDC	G	ETHYLACRYL	G	Benz	Benz	Benz	Benz
Ethyl benzene	136	CDC	G	ETHYLBENZ	G	Benz	Benz	Benz	Benz
Ethyl carbamate	185	NIH	G	ETHYLCARBA	G	Benz	Benz	Benz	Benz
Ethyl chloride	-139	NIH	G	ETHYLCHLRD	G	Benz	Benz	Benz	Benz
Ethylene oxide	11	CDC	G	ETOX	G	Benz	Benz	Benz	Benz
Formaldehyde	-21	CDC	G	FORMALD	G	Benz	Benz	Benz	Benz
1,2-Dimethoxyethane	82	NIH	G	GLYCOLETHR	G	Benz	Benz	Benz	Benz
2-(Hexyloxy)ethanol	258	NIH	G						
2-Butoxyethyl acetate	192	CDC	G						
2-Propoxyethyl Acetate	184	CS	G						
Butyl carbitol acetate	245	NIH	G						
Carbitol acetate	219	NIH	G						
Cellosolve Acetate	145	CDC	G						
Cellosolve Solvent	124	CDC	G						
Diethylene glycol diethyl ether	189 E	CS	G						
Diethylene glycol dimethyl ether	161 E	CS	G						
Diethylene glycol ethyl methyl ether	168 P	CS	G						
Diethylene glycol monobutyl ether	230	NIH	G						
Diethylene glycol monoethyl ether	196	NIH	G						
Diethylene glycol monomethyl ether	194 E	CS	G						
Ethoxytriglycol	256	NIH	G						
Ethylene glycol methyl ether	124	CDC	G						
Ethylene Glycol Monomethyl Ether Acetate	145	CDC	G						
Ethylene glycol mono-sec-butyl ether	192	CS	G						
Glycol ethers	120–240	http://msdssearch.dow.com/PublishedLiteratureDOWCOM/dh_012d/0901b8038012d976.pdf?filepath=oxysolvents/pdfs/noreg/110-00977.pdf&fromPage=GetDoc	G						
Methoxytriglycol	249	http://msdssearch.dow.com/PublishedLiteratureDOWCOM/dh_012d/0901b8038012d976.pdf?filepath=oxysolvents/pdfs/noreg/110-00977.pdf&fromPage=GetDoc	G						
N-Hexyl carbitol	260 E	CS	G						
Phenyl cellosolve	245 E	CS	G						
Propyl cellosolve	150	http://msdssearch.dow.com/PublishedLiteratureDOWCOM/dh_012d/0901b8038012d976.pdf?filepath=oxysolvents/pdfs/noreg/110-00977.pdf&fromPage=GetDoc	G						
Triethylene Glycol Dimethyl Ether	215 E	CS	G						
Triglycol monobutyl ether	278	NIH	G/P						
Hydrochloric acid	-85	CDC	G	HCL	G	Benz	Benz	Benz	Benz
Heptachlor	392	CS	G/P	HEPTACHLOR	G/P	PAH	PAH	PAH	PAH
Hexamethylene Diisocyanate	212	NIH	G	HEXAMTHLE	G	Benz	Benz	Benz	Benz
Hexane	69	CDC	G	HEXANE	G	Benz	Benz	Benz	Benz

ICF

Chem.				NATA					
	Boiling Point					Matching Chemical Modeled in HAPEM[d]			
Name	Value (°C)[a]	Source[b]	Phase[c]	Name	Phase[c]	P	NP	M-OR	M-NR
Hexachloroethane	187	CDC	G	HEXCHLETHN	G	Benz	Benz	Benz	Benz
Hexachlorocyclopentadiene	238	CDC	G	HEXCHLPNTD	G	Benz	Benz	Benz	Benz
Hexachlorobutadiene	215	CDC	G	HEXCHLRBT	G	Benz	Benz	Benz	Benz
Hexachlorobenzene	325	NIH	G/P	HEXCHLROBZ	G/P	PAH	PAH	PAH	PAH
Hexamethylphosphoramide	233	NIH	G	HEXMTHPHO	G	Benz	Benz	Benz	Benz
Hydrogen Fluoride	19	CDC	G	HF	G	Benz	Benz	Benz	Benz
Mercury	356	CDC	G/P	HGSUM	G/P	PAH	PAH	PAH	PAH
1,2,3,4,5,6-Hexachlorocyclohexane	323	CDC	G/P	HXCCL123456	G/P	PAH	PAH	PAH	PAH
Hydrazine	113	CDC	G	HYDRAZINE	G	Benz	Benz	Benz	Benz
Hydroquinone	285	CDC	G/P	HYDROQUIN	G/P	PAH	PAH	PAH	PAH
Isophorone	215	CDC	G	ISOPHORONE	G	Benz	Benz	Benz	Benz
LEAD	1,740	CDC	P	LEAD	P	Cr6	Cr6	Ni	Ni
Maleic anhydride	202	CDC	G	MALANHYD	G	Benz	Benz	Benz	Benz
Manganese	1,962	CDC	P	MANGANESE	P	Cr6	Cr6	Ni	Ni
4,4'-Methylenebis(2-Chloraniline)	209	NIH	G	MB2CLRAN44P	G	Benz	Benz	Benz	Benz
Methylene chloride	39	CDC	G	MECL	G	Benz	Benz	Benz	Benz
Methanol	64	CDC	G	METHANOL	G	Benz	Benz	Benz	Benz
Methoxychlor	89	NIH	G	METHOXYCHL	G	Benz	Benz	Benz	Benz
Methyl bromide	3	CDC	G	METHYLBROM	G	Benz	Benz	Benz	Benz
Methyl isobutyl ketone	116	CDC	G	MIBK	G	Benz	Benz	Benz	Benz
Fine Mineral Fibers	NA	http://www.usg.com/content/dam/USG_Marketing_Communications/united_states/product_promotional_materials/finished_assets/usg-mineral-wool-300a-msds-en-75850002.pdf	P	MINFIB	P	Cr6	Cr6	Ni	Ni
Methyl methacrylate	101	CDC	G	MMETACRYLAT	G	Benz	Benz	Benz	Benz
Methyl tert-butyl ether	55	NIH	G	MTBE	G	Benz	Benz	Benz	Benz
4,4'-Methylenedianiline	397	CDC	G/P	MTHYDIAN44P	G/P	PAH	PAH	PAH	PAH
Methyl chloride	-98	CDC	G	MTHYLCHLRD	G	Benz	Benz	Benz	Benz
Methyl Chloroform	74	CDC	G	MTHYLCHLRF	G	Benz	Benz	Benz	Benz
4,4'-Methylenediphenyl Diisocyanate	313	CS	G/P	MTHYLDPNLDS	G/P	PAH	PAH	PAH	PAH
Methylhydrazine	88	CDC	G	MTHYLHYZIN	G	Benz	Benz	Benz	Benz
Methyl iodide	43	CDC	G	MTHYLIODIDE	G	Benz	Benz	Benz	Benz
Methyl isocyanate	39	CDC	G	MTHYLISOCY	G	Benz	Benz	Benz	Benz
Naphthalene	260	CDC	G	NAPHTH	G	Benz	Benz	Benz	Benz
Nickel	2,913	CDC	P	NICKEL	P	Cr6	Cr6	Ni	Ni
Nickel oxide	1,955	NIH	P						
Nickel refinery dust	2,730	http://www.cdc.gov/niosh/docs/81-123/pdfs/0445.pdf	P						
4-Nitrobiphenyl	340	CDC	G/P	NITROBIPHL4	G/P	PAH	PAH	PAH	PAH
Nitrobenzene	211	CDC	G	NITROBNZNE	G	Benz	Benz	Benz	Benz
4-Nitrophenol	279	NIH	G/P	NITROPHENL4	G/P	PAH	PAH	PAH	PAH
2-Nitropropane	121	CDC	G	NITROPROPA2	G	Benz	Benz	Benz	Benz
N-Nitrosodimethylamine	152	CDC	G	NNITROSDIM	G	Benz	Benz	Benz	Benz
N-Nitrosomorpholine	224	NIH	G	NNITROSMPH	G	Benz	Benz	Benz	Benz
N-Nitroso-n-methylurea	164 P	CS	G	NNITROSURE	G	Benz	Benz	Benz	Benz
o-Anisidine	225	CDC	G	o-Anisidine	G	Benz	Benz	Benz	Benz
o-Toluidine	200	CDC	G	O-TOLUIDINE	G	Benz	Benz	Benz	Benz
Anthracene	342	NIH	G/P	PAH_000E0	G/P	PAH	PAH	PAH	PAH
Phenanthrene	340	NIH	G/P						
Pyrene	404	NIH	P						
3-Methylcholanthrene	178	http://www.speclab.com/compound/c50328.htm	G	PAH_101E2	G	Benz	Benz	Benz	Benz

Chem.				NATA					
	Boiling Point					Matching Chemical Modeled in HAPEM[d]			
Name	Value (°C)[a]	Source[b]	Phase[c]	Name	Phase[c]	P	NP	M-OR	M-NR
7,12-Dimethylbenz[a]anthracene	122	http://www.sigmaaldrich.com/catalog/product/supelco/442425?lang=en&region=US	G	PAH_114E1	G	Benz	Benz	Benz	Benz
Dibenzo[a,h]pyrene	308 E	CS	G/P	PAH_176E2	P	Cr6	Cr6	Ni	Ni
Dibenzo[a,i]pyrene	552 P	CS	P						
Dibenzo[a,l]pyrene	552 P	CS	P						
5-Methylchrysene	449 P	CS	P	PAH_176E3	P	Cr6	Cr6	Ni	Ni
7H-Dibenzo[c,g]carbazole	544 P	CS	P						
Benzo[a]pyrene	360	http://www.speclab.com/compound/c50328.htm	G/P						
Coal tar	>250	http://www.inchem.org/documents/icsc/icsc/eics1415.htm	G/P						
Dibenzo[a,e]pyrene	552 P	CS	P						
Methylchrysene	449 P	CS	P						
1-Nitropyrene	445 P	CS	P	PAH_176E4	P	Cr6	Cr6	Ni	Ni
Benz[a]anthracene	438	NIH	P						
Benzo[b]fluoranthene	4,665 P	CS	P						
Benzo[j]fluoranthene	480 E	CS	P						
Dibenz[a,h]acridine	534 P	CS	P						
Dibenzo[a,j]Acridine	534 P	CS	P						
Indeno[1,2,3-c,d]pyrene	530	http://www.speclab.com/compound/c193395.htm	P						
Benzo[k]fluoranthene	480	http://www.speclab.com/compound/c207089.htm	P	PAH_176E5	P	Cr6	Cr6	Ni	Ni
Carbazole	355	http://www.sigmaaldrich.com/catalog/product/sigma/c5132?lang=en&region=US	G/P						
Chrysene	448	http://www.speclab.com/compound/c218019.htm	P						
Dibenzo[a,h]anthracene	262	http://www.sigmaaldrich.com/catalog/product/supelco/48574?lang=en&region=US	G/P	PAH_192E3	G/P	PAH	PAH	PAH	PAH
12-Methylbenz(a)anthracene	410 P	CS	P	PAH_880E5	G/P	PAH	PAH	PAH	PAH
1-Methylnaphthalene	240	NIH	G						
1-Methylphenanthrene	359	http://www.nature.nps.gov/hazardssafety/toxic/phen1met.pdf	G/P						
1-Methylpyrene	372	http://www.chemicalbook.com/ChemicalProductProperty_EN_CB7421679.htm	G/P						
2-Chloronaphthalene	256	http://www.chemicalbook.com/ChemicalProductProperty_EN_CB8854627.htm	G						
2-Methylnaphthalene	241	http://www.speclab.com/compound/c91576.htm	G						
2-Methylphenanthrene	339	CS	G/P						
9-Methyl Anthracene	196	CS	G						
Acenaphthene	279	NIH	G/P						
Acenaphthylene	265	NIH	G/P						
Benzo(a)fluoranthene	295	NIH	G/P						
Benzo(c)phenanthrene	430 P	CS	P						
Benzo(g,h,i)fluoranthene	406 P	CS	P						
Benzo[e]pyrene	465 P	CS	P						

ICF

Chem.				NATA					
	Boiling Point					Matching Chemical Modeled in HAPEM[d]			
Name	Value (°C)[a]	Source[b]	Phase[c]	Name	Phase[c]	P	NP	M-OR	M-NR
Benzo[g,h,i,]Perylene	550	NIH	P						
Benzofluoranthenes	406 P	CS	P						
Extractable Organic Matter (EOM)	NA	NA	NA						
Fluoranthene	384	NIH	G/P						
Fluorene	295	NIH	G/P						
Methylanthracene	360 P	CS	G/P						
Methylbenzopyrene	479 P	CS	G/P						
PAH, total	240–400	http://www.epa.gov/reg3h wmd/bf-lr/regional/analytical/semi-volatile.htm	G/P						
Perylene	276	http://www.sigmaaldrich.com/catalog/product/aldrich/394475?lang=en&region=US	G/P						
Parathion	375	CDC	G/P	PARATHION	G/P	PAH	PAH	PAH	PAH
2,4,4'-Trichlorobiphenyl (PCB-28)	164	CS	G	PCB	G/P	PAH	PAH	PAH	PAH
4,4'-Dichlorobiphenyl (PCB-15)	144	CS	G						
Decachlorobiphenyl (PCB-209)	460 P	CS	P						
Heptachlorobiphenyl	415 P	CS	P						
Hexachlorobiphenyl	396 P	CS	G/P						
Pentachlorobiphenyl	365 E	CS	G/P						
Polychlorinated biphenyls	365 E	CS	G/P						
Tetrachlorobiphenyl	360 P	CS	G/P						
p-Dioxane	101	CDC	G	P-DIOXANE	G	Benz	Benz	Benz	Benz
Tetrachloroethylene	121	CDC	G	PERC	G	Benz	Benz	Benz	Benz
Phenol	182	CDC	G	PHENOL	G	Benz	Benz	Benz	Benz
p-Phenylenediamine	267	CDC	G/P	PHNYLNDIAMI	G/P	PAH	PAH	PAH	PAH
Phosgene	8	CDC	G	PHOSGENE	G	Benz	Benz	Benz	Benz
Phosphine	88	CDC	G	PHOSPHINE	G	Benz	Benz	Benz	Benz
Phosphorus	280	CDC	G/P	PHOSPHORS	G/P	PAH	PAH	PAH	PAH
Phthalic anhydride	295	CDC	G/P	PHTHANHYDR	G/P	PAH	PAH	PAH	PAH
Pentachloronitrobenzene	328	NIH	G/P	PNTCHLNBZ	G/P	PAH	PAH	PAH	PAH
Pentachlorophenol	309	CDC	G/P	PNTCLPHENOL	G/P	PAH	PAH	PAH	PAH
Propylene dichloride	97	CDC	G	PROPDICLR	G	Benz	Benz	Benz	Benz
Propionaldehyde	48	NIH	G	PROPIONAL	G	Benz	Benz	Benz	Benz
1,3-Propanesultone	180	NIH	G	PROPNESLT13	G	Benz	Benz	Benz	Benz
Propoxur	D	CDC	NA	PROPOXUR	NA	EF=1	EF=1	EF=1	EF=1
Propylene oxide	34	CDC	G	PROPYLENEOX	G	Benz	Benz	Benz	Benz
1,2-Propylenimine	66	NIH	G	PROPYLENIMI	G	Benz	Benz	Benz	Benz
Quinoline	238	NIH	G	QUINOLINE	G	Benz	Benz	Benz	Benz
Quinone	S	CDC	NA	QUINONE	NA	EF=1	EF=1	EF=1	EF=1
Selenium	685	CDC	P	SELENIUM	P	Cr6	Cr6	Ni	Ni
Styrene	145	CDC	G	STYRENE	G	Benz	Benz	Benz	Benz
Styrene oxide	194	http://www.sigmaaldrich.com/catalog/product/aldrich/s5006?lang=en&region=US	G	STYROXIDE	G	Benz	Benz	Benz	Benz
Titanium tetrachloride	136	http://www.sigmaaldrich.com/catalog/product/aldrich/697079?lang=en&region=US	G	TITATETRA	G	Benz	Benz	Benz	Benz
2,4-Toluene diisocyanate	251	NIH	G	TOL_DIIS	G	Benz	Benz	Benz	Benz
Toluene-2,4-Diamine	292	CDC	G/P	TOLDIAM24	G/P	PAH	PAH	PAH	PAH
Toluene	111	CDC	G	TOLUENE	G	Benz	Benz	Benz	Benz
Toxaphene	D	CDC	NA	TOXAPHENE	NA	EF=1	EF=1	EF=1	EF=1
2,4,5-Trichlorophenol	247	NIH	G	TRCLPHNL245	G	Benz	Benz	Benz	Benz

Chem.				NATA					
						Matching Chemical Modeled in HAPEM[d]			
	Boiling Point								
Name	Value (°C)[a]	Source[b]	Phase[c]	Name	Phase[c]	P	NP	M-OR	M-NR
2,4,6-Trichlorophenol	246	NIH	G	TRCLPHNL246	G	Benz	Benz	Benz	Benz
1,2,4-Trichlorobenzene	213	CDC	G	TRICBZ124	G	Benz	Benz	Benz	Benz
1,1,2-Trichloroethane	114	CDC	G	TRICLA112	G	Benz	Benz	Benz	Benz
Triethylamine	89	CDC	G	TRIETHLAMN	G	Benz	Benz	Benz	Benz
Trifluralin	140	http://www.speclab.com/compound/c1582098.htm	G	TRIFLURALIN	G	Benz	Benz	Benz	Benz
2,2,4-Trimethylpentane	99	NIH	G	TRMEPN224	G	Benz	Benz	Benz	Benz
1,1,2,2-Tetrachloroethane	147	CDC	G	TTCLE1122	G	Benz	Benz	Benz	Benz
Vinyl chloride	-14	CDC	G	VINYCHLRI	G	Benz	Benz	Benz	Benz
Vinyl acetate	72	CDC	G	VINYLACET	G	Benz	Benz	Benz	Benz
Vinyl bromide	16	CDC	G	VINYLBROM	G	Benz	Benz	Benz	Benz
Vinylidene chloride	32	CDC	G	VINYLIDCLOR	G	Benz	Benz	Benz	Benz
m-Xylene	139	CDC	G	XYLENES	G	Benz	Benz	Benz	Benz
o-Xylene	144	CDC	G						
p-Xylene	138	CDC	G						
Xylenes (Mixed Isomers)	139	NIH	G						

Note: Information under the "Chem" header indicates chemicals for which EPA developed emissions data; we identified boiling points and HAPEM phase designations for these chemicals. When EPA ultimately conducted air-concentration modeling, for use then in exposure and risk modeling, they grouped some chemicals together as indicated in the information under the "NATA" header (note the occasional many-to-one relationship between the "Name" column under "Chem" and the "Name" column under "NATA"). Assigning phase designations to groups of chemicals required professional judgment, as discussed in Section 2.2.

[a] D=decomposes; E=experimental; NA=not available; P=predicted; S=sublimes; V=varies depending on compound.

[b] CDC=http://www.cdc.gov/niosh/npg; CS=http://www.chemspider.com; NIH=http://pubchem.ncbi.nlm.nih.gov/.

[c] G=gaseous; G/P=gaseous or particulate depending on conditions; P=particulate; NA=unknown.

[d] Source Types: P =point; NP=non-point; M-OR=mobile on-road; M-NR=mobile non-road.
Matching Chemicals: Benz=benzene; Buta=1,3-butadiene; Coke=Coke Oven Emissions; PAH=PAH, total; Ni=Nickel; Cr6=Chromium (VI); DPM = diesel particulate matter; "EF=1" means no surrogate assignment, so the exposure factor is set to 1 (before any truncations are then applied), meaning all exposure concentrations are set equal to ambient outdoor concentrations.

Table 2. Boiling-point Definitions Used to Classify HAPs for HAPEM7 Modeling for the 2014 NATA

HAPEM7 HAP Phase	Boiling-point Range (°C)
G (Gaseous)	< 240-260
G/P (Either gaseous or particulate depending on conditions)	240-260 to 400-480
P (Particulate)	> 400-480

Source: Adapted from the "Classification of Inorganic Pollutants" table at EPA's Volatile Organic Compound web page (available as of 23 October 2017 at https://www.epa.gov/indoor-air-quality-iaq/technical-overview-volatile-organic-compounds), as adapted from: World Health Organization, 1989. "Indoor air quality: organic pollutants." Report on a WHO Meeting, Berlin, 23-27 August 1987. Euro Reports and Studies 111. Copenhagen, World Health Organization Regional Office for Europe.

ICF

Table 3. HAPs Modeled in HAPEM7 for the 2014 NATA

NATA HAP	HAPEM7 HAP Phase[a]	Modeled in HAPEM7[b]			
		P	NP	M-OR	M-NR
Benzene	G	✓	✓	✓	✓
1,3-butadiene	G	✓	✓	✓	✓
Coke Oven Emissions	G/P	✓			
PAH, Total (polycyclic aromatic hydrocarbons; aggregate mass of unspecified congeners)	G/P	✓	✓	✓	✓
Chromium (VI) (compounds of hexavalent chromium)	P	✓	✓		
DPM (diesel particulate matter)	P		✓	✓	✓
Nickel	P			✓	✓

Note: For chromium (VI), DPM, and nickel, EPA modeled air concentrations corresponding to emissions from all four source types, but per EPA direction we only generated exposure concentrations for the subset of source types shown in the table. See discussion at the bottom of Section 2.3.

[a] G=gaseous; G/P=gaseous or particulate depending on conditions; P=particulate.
[b] P =point; NP=non-point; M-OR=mobile on-road; M-NR=mobile non-road.

Table 4. Truncations Applied to Exposure Factors for the 2014 NATA

HAP	Source Type				Total (Aggregate of all Source Types)
	Point	Non-point	Mobile On-road	Mobile Non-road	
Benzene	1.00	0.89	1.09	0.89	0.97
1,3-butadiene	0.99	0.92	1.18	0.92	1.03
Coke oven emissions	0.87	NA	NA	NA	0.87
PAH, total	0.70	0.66	0.83	0.66	0.72
Chromium (VI)	0.60	0.94	NA	NA	0.45
Diesel PM	NA	1.27	0.60	0.48	0.53
Nickel	NA	NA	0.56	0.48	0.52