

**Technical Support Document
EPA's Air Toxics Screening Assessment**

2020 AirToxScreen TSD

February 2024

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

Contents

Contents.....	ii
Tables.....	vi
Figures.....	vi
Common Acronyms and Abbreviations	viii
1. Introduction	1
1.1. Overview	1
1.2. Purpose and Overview of AirToxScreen Steps.....	2
1.3. What AirToxScreen Is.....	4
1.4. Air Toxics Screening Assessment History.....	5
1.5. How EPA and State, Local and Tribal Air Agencies Use AirToxScreen Results.....	6
1.6. How AirToxScreen Results Should <i>Not</i> Be Used.....	7
1.7. The Risk Assessment Framework AirToxScreen Uses	8
1.8. The Scope of AirToxScreen	9
1.8.1. Sources of Air Toxic Emissions That AirToxScreen Addresses.....	10
1.8.2. Stressors that AirToxScreen Evaluates	11
1.8.3. Exposure Pathways, Routes and Time Frames for AirToxScreen.....	11
1.8.4. Receptors that AirToxScreen Characterizes.....	12
1.8.5. Endpoints and Measures: Results of AirToxScreen.....	12
1.9. Model Design	13
1.9.1. The Strengths and Limitations of the Model Design.....	15
2. Emissions.....	17
2.1. Sources of Emissions.....	17
2.1.1. Pollutants and Pollutant Groups.....	20
2.1.2. Emissions Categorization: NEI and AirToxScreen	24
2.1.3. Differences Between the current NEI and Emissions Used for AirToxScreen	26
2.1.4. Overview of Differences in Emissions for CMAQ and AERMOD	28
2.2. Preparation of Emissions Inputs for CMAQ.....	30
2.2.1. Sectors in the CMAQ AirToxScreen Platform.....	30
2.2.2. Fires and Biogenics.....	33
2.2.3. Speciation.....	34
2.2.4. Temporalization	34
2.2.5. Spatial Allocation	34
2.3. Preparation of Emissions Inputs for AERMOD.....	35
2.3.1. Source and Run Groups – Overview	35
2.3.2. Point Sources Excluding Airports	43
2.3.3. Airport Point Sources	47
2.3.4. Nonroad, On-road and Nonpoint – County-level Sources.....	50
2.3.5. Residential Wood Combustion	53
2.3.6. Oil and Gas	53
2.3.7. Agricultural Livestock.....	54

2.3.8. Commercial Marine Vessels – County/Shape-level Sources.....	54
2.3.9. Urban/Rural Determination for All Emission Sources	56
3. Air Quality Modeling and Characterization	61
3.1. Modeling Overview.....	61
3.1.1. Photochemical Model Selection	61
3.1.2. Dispersion Model Selection	63
3.2. Meteorological Data	63
3.2.1. Meteorological Data Inside the Contiguous States	63
3.2.2. Meteorological Data Outside the Contiguous States	65
3.3. CMAQ Setup.....	66
3.3.1. Sources Modeled in CMAQ.....	66
3.3.2. Boundary and Initial Conditions.....	72
3.4. AERMOD Setup	74
3.4.1. Sources Modeled in AERMOD.....	74
3.4.2. Receptor Placement.....	74
3.4.3. Model Options	79
3.4.4. AERMOD Simulations.....	79
3.4.5. Post-processing of AERMOD Results	80
3.5. Hybrid Modeling	82
3.5.1. Overview	82
3.5.2. Treatment of Species	84
3.5.3. Ambient Monitoring Data Preparation.....	86
3.5.4. Model Performance Statistics.....	87
3.5.5. Hybrid Evaluation.....	89
3.5.6. Non-hybrid Evaluation	91
4. Estimating Exposures for Populations.....	94
4.1. Estimating Exposure Concentrations	94
4.2. About HAPEM	94
4.3. HAPEM Inputs and Application	96
4.3.1. Data on Ambient Air Concentrations	96
4.3.2. Population Demographic Data	96
4.3.3. Data on Population Activity.....	96
4.3.4. Microenvironmental Data	97
4.4. Exposure Factors.....	99
4.5. Evaluating Exposure Modeling.....	100
4.6. Summary	100
5. Characterizing Effects of Air Toxics	101
5.1. Toxicity Values and Their Use in AirToxScreen	101
5.2. Types of Toxicity Values	102
5.2.1. Cancer URE.....	102
5.2.2. Noncancer Chronic RfC	105
5.3. Data Sources for Toxicity Values.....	105
5.3.1. U.S. EPA Integrated Risk Information System	106

5.3.2. U.S. Department of Health and Human Services, Agency for Toxic Substances and Disease Registry	106
5.3.3. California Environmental Protection Agency Office of Environmental Health Hazard Assessment	106
5.3.4. U.S. EPA Health Effects Assessment Summary Tables	106
5.3.5. World Health Organization International Agency for Research on Cancer.....	107
5.4. Additional Toxicity Decisions for Some Chemicals.....	107
5.4.1. Polycyclic Organic Matter.....	107
5.4.2. Glycol Ethers	108
5.4.3. Acrolein	108
5.4.4. Metals	108
5.4.5. Adjustment of Mutagen UREs to Account for Exposure During Childhood	109
5.4.6. Diesel Particulate Matter	110
5.5. Summary	111
6. Characterizing Risks and Hazards in AirToxScreen.....	112
6.1. The Risk-characterization Questions AirToxScreen Addresses	112
6.2. How Cancer Risk Is Estimated	113
6.2.1. Individual Pollutant Risk.....	113
6.2.2. Multiple-pollutant Risks	113
6.3. How Noncancer Hazard is Estimated	114
6.3.1. Individual Pollutant Hazard.....	114
6.3.2. Multiple-pollutant Hazard.....	115
6.4. How Risk Estimates and Hazard Quotients Are Calculated for AirToxScreen at Tract, County and State Levels	116
6.4.1. Model Results for Point Sources: Aggregation to Tract-level Results	Error! Bookmark not defined.
6.4.2. Background Concentrations and Secondary Pollutants: Interpolation to Tract-level Results.....	Error! Bookmark not defined.
6.4.3. Aggregation of Tract-level Results to Larger Spatial Units.....	116
6.5. The Risk Characterization Results That AirToxScreen Reports	116
6.6. Summary	118
7. Variability and Uncertainty Associated with AirToxScreen.....	120
7.1. Introduction	120
7.2. How AirToxScreen Addresses Variability	120
7.2.1. Components of Variability.....	121
7.2.2. Quantifying Variability	122
7.2.3. How Variability Affects Interpretation of AirToxScreen Results	124
7.3. How AirToxScreen Addresses Uncertainty	124
7.3.1. Components of Uncertainty.....	125
7.3.2. Components of Uncertainty Included in AirToxScreen.....	126
7.4. Summary of Limitations in AirToxScreen.....	131

8. References	134
Appendix A. Glossary.....	A-1
Appendix B. Air Toxics Modeled in AirToxScreen	B-1
Appendix C. Estimating Background Concentrations for AirToxScreen	C-3
Appendix D. Model Evaluation Summaries	D-1
Appendix E. Exposure Factors for AirToxScreen	E-1
Introduction to HAPEM and its Use in AirToxScreen	Error! Bookmark not defined.

Tables

Table 2-1. Summary of emissions sources in the 2019 NEI	18
Table 2-2. Pollutant groups.....	21
Table 2-3. PAH/POM pollutants group	21
Table 2-4. Metal speciation factors for NEI metal compounds	23
Table 2-5. Map of NEI data categories to AirToxScreen categories	25
Table 2-6. Release parameter defaults/changes to the FF10 inventory files for point sources.....	27
Table 2-7. Differences in spatial characterization of sources between CMAQ and AERMOD.....	29
Table 2-8. Platform sectors for the 2019 emissions modeling platform	31
Table 2-9. Run groups for AERMOD.....	36
Table 2-10. Resolution of the run groups modeled as gridded sources.....	37
Table 2-11. Non-gridded AERMOD run groups.....	37
Table 2-12. AERMOD gridded run groups.....	38
Table 2-13. AirToxScreen source groups	41
Table 2-14. Assignment of AERMOD source type for point sources	45
Table 2-15. Options for temporal variation specification in helper files	47
Table 2-16. Lead emissions (kg/yr) at SMO in 2019 by aircraft operation mode	49
Table 2-17. Airport emissions file format	49
Table 2-18. NEI SCCs covered in the CMV run group.....	54
Table 3-1. CMAQ HAPs.....	61
Table 3-2. Vertical layer structure for WRF and CMAQ (heights are layer top)	64
Table 3-3. WRF Options Used for the Alaska, Hawaii, and Puerto Rico/Virgin Islands domains.....	66
Table 3-4. Boundary conditions from 2019 remote concentration estimates	73
Table 3-5. CMAQ HAP boundary conditions applied as zero value	73
Table 3-9. List of hybrid HAPs evaluated	89
Table 4-1. Key differences between recent versions of HAPEM	95
Table 4-2. Microenvironments used in HAPEM modeling for AirToxScreen	99
Table 6-1. Criteria establishing AirToxScreen drivers and contributors of health effects for risk characterization.....	117

Figures

Figure 1-1. AirToxScreen – basic steps.....	2
Figure 1-2. Detailed steps and approach used in AirToxScreen	4
Figure 1-3. The general air toxics risk assessment process	8
Figure 1-4. Conceptual model for AirToxScreen.....	10
Figure 1-5. The AirToxScreen risk assessment process and corresponding sections of this TSD	14

Figure 2-1. Example fugitive source characterization: NEI length = 1897 feet, width = 680 feet and angle = 22	46
Figure 2-2. Port shapes for Los Angeles and Long Beach, California	55
Figure 3-1(a). Map of the CMAQ modeling domain; the blue box denotes the 12 km CONUS modeling domain	68
Figure 3-1(b). Map of the CMAQ modeling domain; the green box denotes the 9 km Alaska modeling domain	69
Figure 3-1(c). Map of the CMAQ modeling domain; the red box denotes the 3 km Hawai'i modeling domain	70
Figure 3-1(d). Map of the CMAQ modeling domain; the red box denotes the 3 km Puerto Rico/Virgin Islands modeling domain	71
Figure 3-2. CBSAs exceeding 1 million people	Error! Bookmark not defined.
Figure 3-3. Dense (left) and coarse (right) receptor grid layout in CMAQ Lambert Projection.....	76
Figure 3-4. Example grid cell with subgrid cells and census blocks.....	77
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.....	83

Common Acronyms and Abbreviations

$\mu\text{g}/\text{m}^3$	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 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 Air Toxics Screening Assessment (AirToxScreen) 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 AirToxScreen 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. AirToxScreen is the successor to the National Air Toxics Assessment, or NATA.

The goal of AirToxScreen is to identify those air toxics which are of greatest potential concern in terms of contribution to population risk. Screening-level ambient and exposure concentrations and estimates of risk and hazard for air toxics in each state have previously been generated at the census tract level, but for the 2020 version of AirToxScreen, have been generated at the census block level.

This AirToxScreen Technical Support Document (TSD) describes the data and approaches EPA used to conduct AirToxScreen, 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 AirToxScreen
- Appendix C – procedures used to estimate AirToxScreen background concentrations
- Appendix D – additional model evaluation summaries
- Appendix E – documentation on HAPEM8 and its use in AirToxScreen

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 AirToxScreen. This document does not provide quantitative results for AirToxScreen and thus presents no exposure or risk estimates. You can find results and other specific information for AirToxScreen on the AirToxScreen website (<https://www.epa.gov/AirToxScreen>). Links to previous assessments are also available on this site.

1.2. Purpose and Overview of AirToxScreen Steps

AirToxScreen follows several basic steps to produce the final assessment. These AirToxScreen steps are depicted in Figure 1-1.

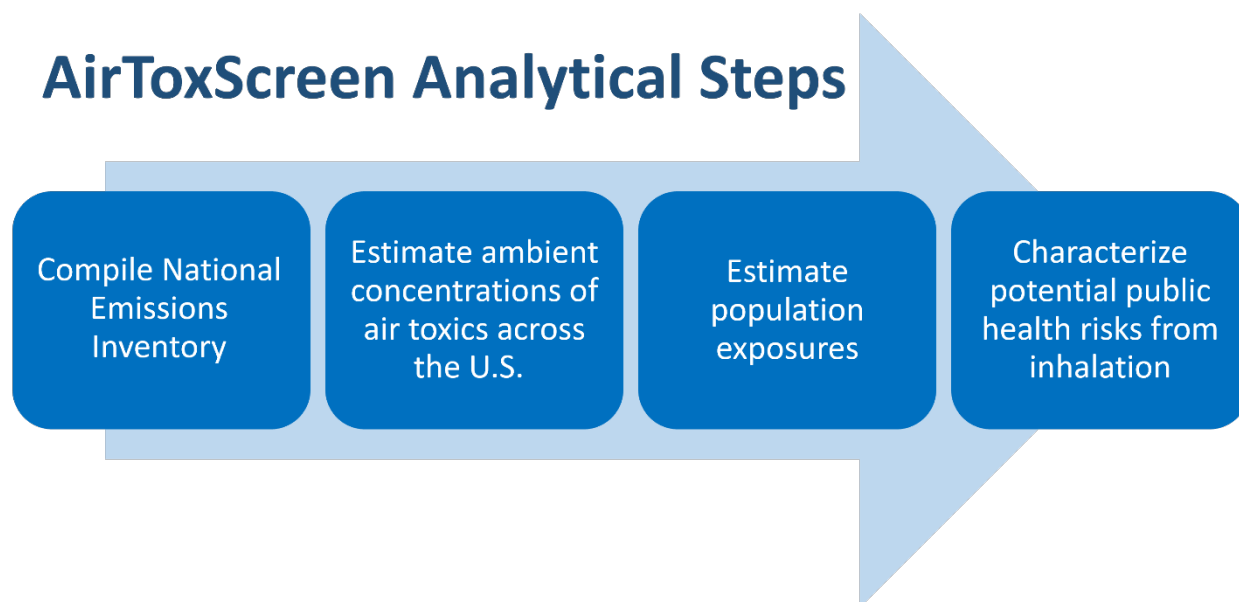


Figure 1-1. AirToxScreen – basic steps

The first and most time-consuming step in AirToxScreen is assembling the National Emissions Inventory (NEI), a detailed, nationwide inventory of air toxics emissions. The NEI includes emissions from point, nonpoint and mobile sources, as well as emissions from biogenic sources and fires. These source data form the EPA’s air emissions modeling platform and are the foundation of AirToxScreen’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 AirToxScreen air toxics modeled, and CMAQ is used for a list of 52 air toxics that are incorporated into CMAQ multipollutant version 5.4. 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

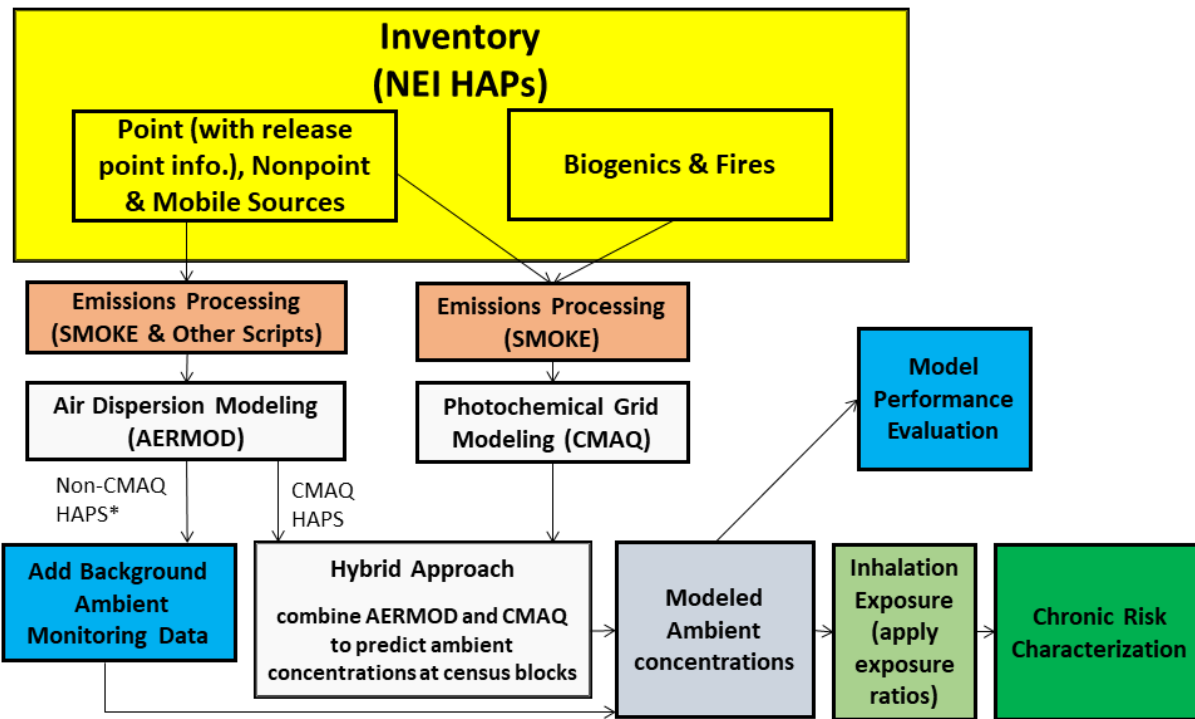
detailed source attribution. CMAQ also provides the biogenic and fire concentrations, as these sources are not run in AERMOD. Section 3 of this document details the steps EPA used to model for AirToxScreen.

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 block-level concentrations of all modeled air toxics. Then, using the HAPEM8 exposure model, we account for human activity patterns and develop exposure concentrations, or ECs. Finally, we estimate census block-level risks by applying health benchmark data to the ECs. Sections 4 through 6 of this TSD detail these steps.

EPA's state, local and tribal (S/L/T) air agency partners play an integral role in AirToxScreen. 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 AirToxScreen modeling process, which often reveals other inaccuracies in the data. This review and feedback process helps ensure that AirToxScreen's input data are as accurate as possible in the final version.

For the 2020 AirToxScreen, we modeled the 2020 NEI. For point sources, a draft version of the 2020 NEI was modeled using AERMOD in the Summer of 2023, the results of which were used to prioritize review of emissions that might contribute to high risk in the final results. Our S/L/T partners then reviewed the point source emissions in the draft 2020 NEI (without airports and railyards) along with draft risk estimates. We reviewed and incorporated S/L/T point source changes to emissions and release point locations and parameters into the final 2020 NEI. The full AirToxScreen modeling process for all emissions sectors using the hybrid approach (AERMOD and CMAQ) was performed. We then previewed these final results with S/L/T agencies before releasing the 2020 AirToxScreen in XX 2024.

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.



*includes all HAPs in AK/HI/PR/VI since not part of CMAQ modeling domain

Figure 1-2. Detailed steps and approach used in AirToxScreen

1.3. What AirToxScreen Is

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

AirToxScreen 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 AirToxScreen results for each U.S. state at county and census block 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.

AirToxScreen 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 a 70 year

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 AirToxScreen can answer reliably. You should consider these limitations, described throughout this document and summarized in Section 7, when interpreting the AirToxScreen results or when using them to address questions posed outside of AirToxScreen.

AirToxScreen 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.

AirToxScreen 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. Air Toxics Screening Assessment History

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 National Air Toxics Assessment (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, five versions of NATA have been completed – representative of

air toxic emissions in 1999, 2002, 2005, 2011 and 2014, respectively – based on significant triennial updates of the national emission inventories. In general, the scope of NATA progressively expanded with subsequent versions, and some methods were refined and improved.

Beginning with emissions year 2017, NATA was succeeded by the AirToxScreen assessment, which use the same basic methods as those used in the 2014 NATA. The goal of AirToxScreen was to build upon the methodological foundation of NATA and to provide more frequent (*i.e.*, annual) updates. AirToxScreen for emission years 2017, 2018, and 2019 have been released with results at the census tract level, consistent with NATA. As we strive to continue updating and improving our national scale screening-level approach, the 2020 version of AirToxScreen incorporates the most significant update to the air toxics screening assessment approach to date -- a new receptor approach to provide results at the census block level. This shift from census tracts to census blocks provides results at a more granular geographic scale and helps to better answer the questions outlined in Section 1.3 and the goals outlined below in Section 1.5.

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

We designed AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 2024a).

Similarly, S/L/T air agencies use AirToxScreen 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 AirToxScreen Results Should *Not* Be Used

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

AirToxScreen results should *not* be used:

- as a definitive means to pinpoint specific risk values within a census tract or census block;
- to characterize or compare risks at local levels (such as between neighborhoods);
- to characterize or compare risks between states;
- 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 AirToxScreen from serving as a stand-alone tool. Furthermore, although EPA reports results currently at the census block or previously at the census tract level in AirToxScreen, the average risk estimates are far more uncertain at these levels of spatial resolution than at the county or state level. To analyze air toxics in smaller geographic areas, such as within 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 AirToxScreen results. You should use AirToxScreen results only to address those questions for which the assessment methods are suited. EPA does not use AirToxScreen as the sole source of information leading to regulations or guiding the enforcement of existing rules. Some of the methods used to conduct AirToxScreen 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), AirToxScreen fundamentally differs from such assessments in that it is not a regulatory program.

1.7. The Risk Assessment Framework AirToxScreen Uses

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

EPA has published a series of guidelines (EPA 2024b) 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.

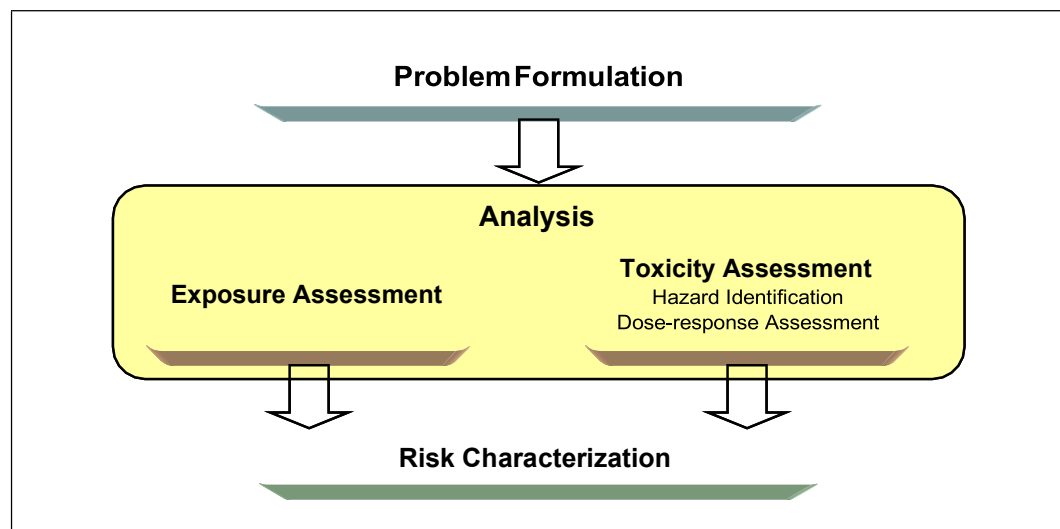


Figure 1-3. The general air toxics risk assessment process

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 AirToxScreen – 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 AirToxScreen). 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 AirToxScreen

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 and for AirToxScreen. Moreover, the conceptual model described in this document is very similar to the one presented in the documentation for the 1996 NATA.

AirToxScreen 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, AirToxScreen 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 AirToxScreen 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 AirToxScreen Addresses

Sources of primary air toxic emissions included in AirToxScreen (i.e., the AirToxScreen categories) are point, nonpoint, mobile, biogenics and fires in the contiguous United States and Alaska, Hawaii, Puerto Rico, and the U.S. Virgin Islands. 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.

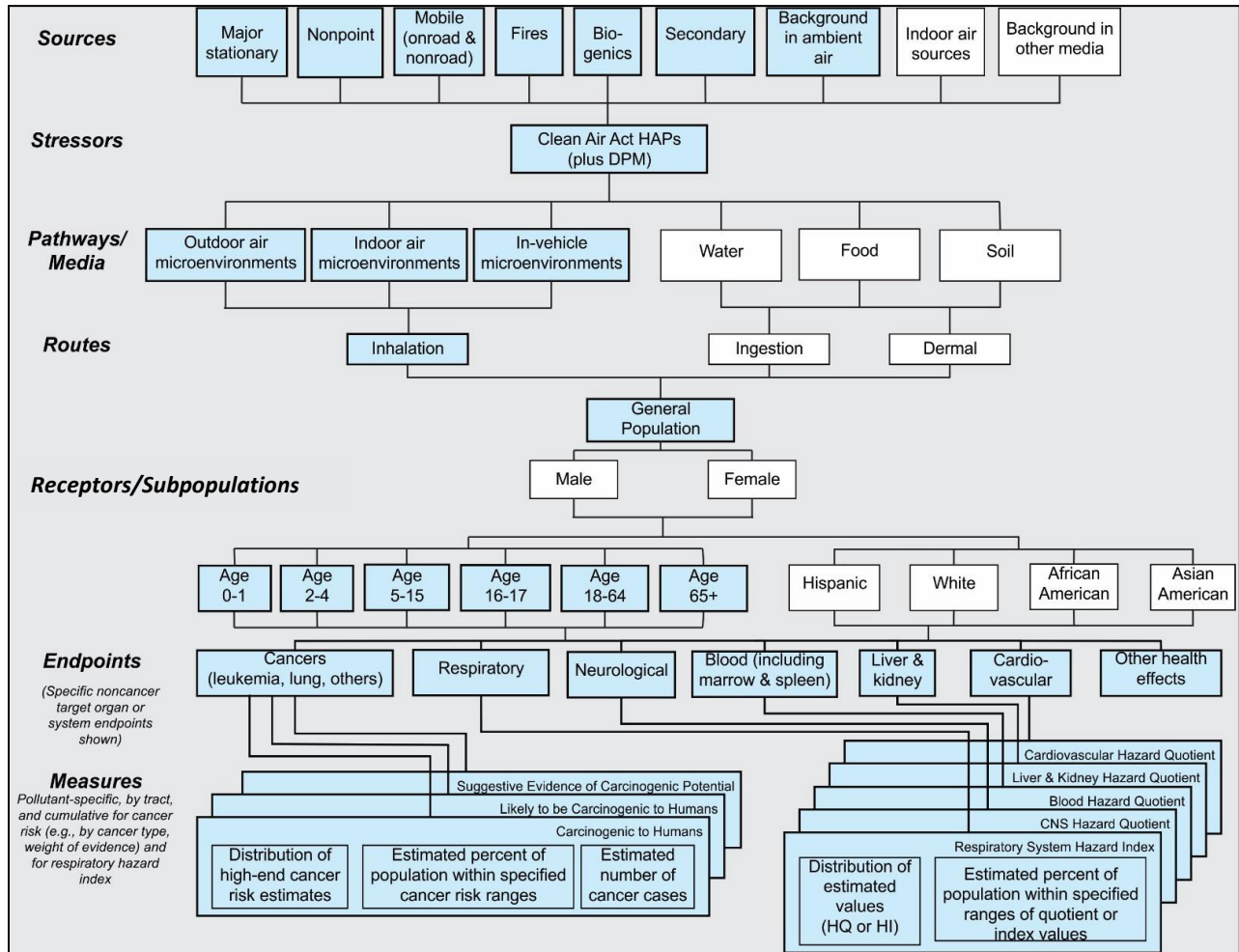


Figure 1-4. Conceptual model for AirToxScreen

Blue boxes indicate elements included in AirToxScreen; 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 AirToxScreen is presented in Section 2. DPM refers to diesel particulate matter. HQ and HI refer to hazard quotient and hazard index, respectively.

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

AirToxScreen presents results by both these broad source categories and by more detailed AirToxScreen 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 AirToxScreen Evaluates

The stressors evaluated through AirToxScreen can include any of the 188 current HAPs defined in the CAA and diesel PM. The set of air toxics included in AirToxScreen 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 AirToxScreen. The spreadsheet file “AirToxScreen_Pollutants.xlsx” within the [Supplemental Data](#) folder accompanying this TSD lists the pollutants that AirToxScreen assesses and provides more detailed information on the NEI and AirToxScreen pollutants. In Appendix B of this document, Table B-1 lists the CAA pollutants that are not included in AirToxScreen and the reason for their omission.

1.8.3. Exposure Pathways, Routes and Time Frames for AirToxScreen

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 AirToxScreen includes evaluating the transport of emitted particles and gases through the air to receptors. AirToxScreen 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, AirToxScreen estimates outdoor ambient concentrations across the nation.

AirToxScreen 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 AirToxScreen

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, AirToxScreen 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, AirToxScreen 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.

AirToxScreen 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 AirToxScreen 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 AirToxScreen.

1.8.4. Receptors that AirToxScreen Characterizes

AirToxScreen characterizes average risks to people belonging to distinct human subpopulations using exposure factors estimated at census tract scales. 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 AirToxScreen results.

AirToxScreen does not include non-human receptors (e.g., wildlife and native plants). The complexity of the varied ecosystems across the vast area AirToxScreen 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 AirToxScreen.

1.8.5. Endpoints and Measures: Results of AirToxScreen

AirToxScreen reports estimated cancer risks and noncancer hazards attributed to modeled sources. Key measures of cancer risk developed for AirToxScreen 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 AirToxScreen are hazard indexes summed across all air toxics modeled for the respiratory system. Other target organs and systems are also shown.

AirToxScreen characterizes cancer risk and potential noncancer effects based on estimates of inhalation exposure concentrations determined at the census block level. This approach is used only to determine geographic patterns of risks within counties, and not to pinpoint specific risk values for each census block. 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 block 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 block-level estimates were not provided. This approach is consistent with the purpose of AirToxScreen, 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 AirToxScreen 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 AirToxScreen. This set of activities is referred to here as the “AirToxScreen risk assessment process.”

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

compiling the nationwide inventory of emissions from outdoor sources;

estimating nationwide ambient outdoor concentrations of the emitted air toxics;

estimating population exposures to these air toxics via inhalation; and

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 AirToxScreen, it does require that we make important scientific and policy decisions regarding the appropriate values to use in AirToxScreen.

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 AirToxScreen 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 AirToxScreen 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 AirToxScreen.

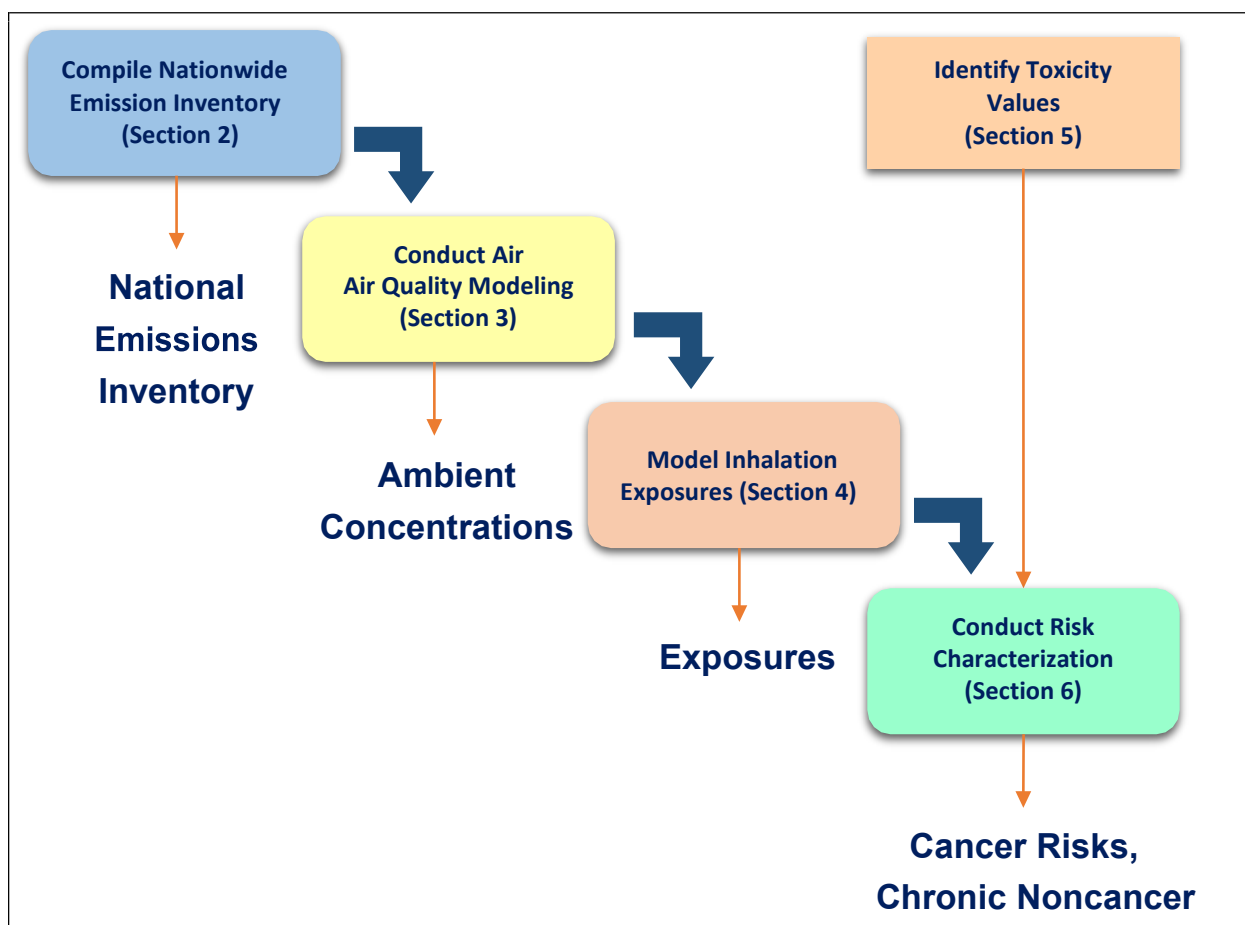


Figure 1-5. The AirToxScreen risk assessment process and corresponding sections of this TSD

-
- 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 AirToxScreen, the sources from which these values are obtained and assumptions made specific to AirToxScreen.
 - Section 6 provides the calculations used to estimate cancer risk and potential noncancer hazard.
 - Section 7 describes the uncertainties and limitations associated with the AirToxScreen process that must be considered when interpreting AirToxScreen 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. Accordingly, although the following sections contain information on the AirToxScreen process that are generally applicable to all previous assessments, references to specific technical processes and supporting details typically emphasize what we did for the current version of AirToxScreen.

1.9.1. The Strengths and Limitations of the Model Design

EPA developed AirToxScreen 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 AirToxScreen may be more appropriate for assessing health risks outside the specific purpose of AirToxScreen (e.g., for evaluating risks from either a broader or more specific perspective).

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

- AirToxScreen 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 blocks, counties and states. All questions asked, therefore, must focus on the variations among these geographic areas (census blocks, counties, etc.). Moreover, as previously mentioned, results are far more uncertain at the census block 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 blocks.)
- AirToxScreen 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.
- AirToxScreen does not include results for individuals. Within a census block, 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 AirToxScreen, only 127 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 AirToxScreen. 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. AirToxScreen, 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 AirToxScreen.
 - Atmospheric transformation and losses from the air by deposition are not accounted for in AirToxScreen 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 AirToxScreen.

2. Emissions

The systematic compilation of a detailed, nationwide inventory of air toxics emissions is the first major step in the AirToxScreen risk assessment process. This section contains descriptions of the emissions used for AirToxScreen. Section 2.1 describes the emissions data sources and preparation of the emissions used in AirToxScreen. 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

AirToxScreen 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.

AirToxScreen combines modeling from CMAQ and AERMOD for the contiguous United States. CMAQ multipollutant

modeling addresses all sources in the National Emissions Inventory (NEI) for CAPs and about 52 air toxics including diesel particulate matter. 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). AERMOD modeling addresses all pollutants covered by AirToxScreen and all anthropogenic sources except prescribed and agricultural burning.

The main source of the emissions data for the CAPs and HAPs modeled for AirToxScreen is the [National Emissions Inventory](#), or NEI. 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. Currently, 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 NEI includes speciated particulate matter (PM) and diesel PM. The NEI diesel PM is computed

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 2016a) plus diesel PM. The 1990 Clean Air Act Amendments required EPA to control [HAPs](#) (EPA 2018) and provided for revisions to be made to that list. Currently, the list includes 188 HAPs. Diesel PM is not a HAP; however, it has been included in AirToxScreen. 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 AirToxScreen (using an Integrated Risk Information System reference concentration), but its cancer risks are not.

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 [2020 NEI documentation](#) and the [2020 Emissions Modeling Platform documentation](#).

Table 2-1. Summary of emissions sources in the 2020 NEI

Source	Description
Stationary point	<p>Most stationary point-source HAP data were submitted voluntarily by S/L/T. For some point sources, EPA gap-filled HAPs. Sources of gap-filled HAPs include: TRI data and augmentation using emission-factor ratios (of HAP to CAP) applied to S/L/T-reported CAP emissions. This source group includes the following modeling platform sectors: Electric Generating Units (ptegu), Point source oil and gas (pt_oilgas), and remaining non-EGU point (Ptnonipm), with results of point airports and point railyards published separately.</p> <p>Ptegu: 2020 NEI point source EGUs, replaced with hourly Continuous Emissions Monitoring System (CEMS) values for NOX and SO2, and the remaining pollutants temporally allocated according to CEMS heat input where the units are matched to the NEI. Emissions for all sources not matched to CEMS data come from 2020 NEI point inventory. Annual resolution for sources not matched to CEMS data, hourly for CEMS sources. EGUs closed in 2020 are not part of the inventory.</p> <p>Pt_oilgas: 2020 NEI point sources that include oil and gas production emissions processes for facilities with North American Industry Classification System (NAICS) codes related to Oil and Gas Extraction, Natural Gas Distribution, Drilling Oil and Gas Wells, Support Activities for Oil and Gas Operations, Pipeline Transportation of Crude Oil, and Pipeline Transportation of Natural Gas. Includes U.S. offshore oil production.</p> <p>Ptnonipm: All 2020 NEI point source records not matched to the airports, ptegu, or pt_oilgas sectors. 2020 NEI rail yards, while in the ptnonipm modeling platform sector, are published separately for AirToxScreen (see below).</p>
Point airports	<p>2020 NEI CAP and HAP emissions for aircraft operations including commercial, general aviation, air taxis and military aircraft, auxiliary power units and ground support equipment computed by the EPA for approximately 20,000 airports. Methods include the use of the Federal Aviation Administration's (FAA's) Aviation Environmental Design Tool (AEDT) For further information, see Section 3.2 of 2020 National Emissions Inventory Technical Support Document: Point Sources and 2020 National Emissions Inventory: Aviation Component.</p>

Source	Description
Point rail yards	The 2020 NEI includes estimates compiled by the EPA for most rail yards in the US. Yard emissions are associated with the operation of switcher engines at each yard. S/L/T data also provided by California, the District of Columbia, Minnesota, Texas, and Washoe County Nevada. Switch yards are reported as point sources to SCC 28500201. Some states report switch yards to nonpoint (2285002010); however, EPA prefers that these emissions be reported as point sources and may be retiring this SCC in the next NEI cycle. Details for rail yards are documented in a report, " 2020 National Emissions Inventory Locomotive Methodology ", on the 2020 Supplemental data FTP site under the nonpoint rail folder. S/L/Ts submitted point rail yard emissions were compared to EPA-computed emissions to avoid double counting between the S/L/T and EPA emissions.
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. For further information, see 2020NEI TSD: Section 6 nonpoint Overview
Biogenics	The biogenic emissions for the 2020 National Emissions Inventory (NEI) were computed based on 2020 meteorology data from the Weather Research and Forecasting (WRF) model version 3.8 (WRFv3.8) and using the Biogenic Emission Inventory System, version 4 (BEIS4) model. Includes VOC, NO _x and three HAPs: formaldehyde, acetaldehyde and methanol. For further information, see 2020 NEI TSD Section 8: Biogenics .
Locomotives	The 2020 locomotive sector includes railroad locomotives powered by diesel-electric engines. A diesel-electric locomotive uses 2-stroke or 4-stroke diesel engines and an alternator or a generator to produce the electricity required to power its traction motors. The locomotive source category is further divided up into categories: Class I line haul, Class II/III line haul, Passenger, and Commuter, and Yard. Railyard emissions are included as point sources as described above. The locomotives sector includes data from SLT agency-provided emissions data, and an EPA dataset of locomotive emissions. HAP emissions were estimated by applying speciation profiles to the VOC or PM estimates. These "HAP fractions" were updated for 2017 NEI. These profiles are posted in the workbook "2017Rail_HAP_AugmentationProfileAssignmentFactors_20200128.xlsx" on the 2017 Supplemental data FTP site. 12-3 HAP estimates were calculated at the yard and link level, after the criteria emissions had been allocated. Where submitting agencies did not supply HAPs, those estimates were also derived via this VOC/PM speciation method. For further information, see 2020 NEI TSD Section 12. Locomotives .
CMVs	EPA's Commercial Marine Vessels estimates are computed using detailed satellite-based automatic identification system (AIS) activity data from the US Coast Guard. The details of these calculation are available in the documents " 2020NEI_C1C2_Documentation " and " 2020 C3 Marine Emissions Tool Documentation " on the 2020 Supplemental data FTP site under folder nonpoint, CMV.

Source	Description
On-road	<p>Onroad mobile sources include emissions from motorized vehicles that normally operate on public roadways. This includes passenger cars, motorcycles, minivans, sport-utility vehicles, light-duty trucks, heavy-duty trucks, and buses. The sector includes emissions generated from parking areas, emissions from short-duration idle during pickups/deliveries, emissions from vehicles when they start, and emissions while the vehicles are moving. The sector also includes “hoteling” emissions, which refers to the time spent idling in a diesel long-haul combination truck during federally mandated rest periods of long-haul trips. Onroad emissions in the 2020 NEI are comprised of emission estimates calculated based on version 3 of the MOVES model run with State, Local, and Tribal (S/L/T)-submitted activity data and other MOVES inputs when provided, except for California and tribes, for which the NEI includes submitted emissions. In cases where S/L/T submitted data are not provided, EPA-developed default activity based on data from the Federal Highway Administration (FHWA) and other data sources. EPA also developed default data for all other inputs required by MOVES which are used where S/L/T data of sufficient quality are not available. EPA added DIESEL-PM10 and DIESEL-PM25 for all diesel fuel SCCs, and they were set equal to the PM₁₀ and PM_{2.5} emissions from these diesel SCCs. For more information, see 2020 NEI TSD Section 5 Onroad Mobile.</p>
Nonroad, excluding airports, locomotives and CMVs	<p>The mobile nonroad equipment data category includes all mobile source emissions that do not operate on roads, excluding commercial marine vehicles, railways, and aircraft. The emissions included in the NEI for this category are modeled using MOVES and cover nonroad equipment in 10 broad economic sectors: construction, agriculture, industrial, lawn & garden (commercial and residential), commercial, logging, railroad support (excluding locomotives), recreational vehicles, recreational marine (pleasure craft; excluding commercial marine vessels), and underground mining. Nonroad equipment emissions were computed by running the MOVES3,1 which incorporates the NONROAD model. MOVES3 and its predecessor MOVES2014b incorporated updated nonroad engine population growth rates, nonroad Tier 4 engine emission rates, and sulfur levels of nonroad diesel fuels. MOVES3 was used for all states other than California, which developed their own emissions using their own tools. For further information, see 2020NEI TSD: Section 4 Nonroad Equipment.</p>
Fires	<p>Wildfires and prescribed burns that occur during the inventory year are included as “non-point” sources beginning with the 2020 NEI. Previous NEIs had wildland fires labeled as “events” sources. Emissions from these fires, as well as agricultural fires, make up the National Fire Emissions Inventory (NFEI). Estimated emissions from all these fire types in the 2020 NEI are calculated from burned area data. Input data sets are collected from State/Local/Tribal (S/L/T) agencies and from national agencies and organizations. Raw burned area data compiled from S/L/T agencies and national data sources are cleaned and combined to produce a comprehensive burned area data set. Emissions are then calculated using fire emission tools/models that rely on burned area as well as fuel and climatological weather information. For more information, including HAP factors used, see 2020NEI TSD, Section 7: Wild and Prescribed Fires and Field Burning. The resulting emissions are compiled by date and location as day-specific emission estimates.</p>

2.1.1. Pollutants and Pollutant Groups

AirToxScreen 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 AirToxScreen pollutants. This section discusses the pollutants and pollutant groups in the NEI used in AirToxScreen and the aggregation needed to generate the AirToxScreen pollutants.

In CMAQ, we model CAPs and precursors and about 52 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-1 in Appendix B provides more detail about each HAP excluded. For AirToxScreen, 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 AirToxScreen 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 AirToxScreen HAP “glycol ethers.” The spreadsheet file “AirToxScreen_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 AirToxScreen 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 (aroclor)
Xylenes (mixed isomers)

2.1.1.1. Assignment of PAHs into PAH modeling groups

In AirToxScreen, Polycyclic aromatic hydrocarbons and polycyclic organic matter are represented by a lumped group, “PAHPOM”.

The individual compounds in the PAHPOM group have widely varying risks. As a result, we modeled the PAHPOM in separate risk-based groups based on their unit risk estimate (URE). For AirToxScreen, we summed the concentrations and risks across all PAHPOM risk groups. For the tabular emission summaries on the AirToxScreen website (<https://www.epa.gov/AirToxScreen>), the individual PAHPOM compounds (i.e., in Table 2-3) are provided.

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

Table 2-3. PAH/POM pollutants group

PAH Group	NEI Pollutant Code	NEI Pollutant Description	URE 1/($\mu\text{g}/\text{m}^3$)
PAH_000E0	120127	Anthracene	
PAH_000E0	85018	Phenanthrene	
PAH_000E0	129000	Pyrene	
PAH_101E2	56495	3-Methylcholanthrene	0.01

PAH Group	NEI Pollutant Code	NEI Pollutant Description	URE 1/($\mu\text{g}/\text{m}^3$)
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
BENZOAPYRNE	50328 (see Note 1)	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 (see Note 2)	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 (see Note 3)	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	Methylantracene	4.8E-05
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 Group	NEI Pollutant Code	NEI Pollutant Description	URE 1/($\mu\text{g}/\text{m}^3$)
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
PAH_880E5	N590	Polycyclic aromatic compounds (includes 25 specific compounds)	4.8E-05

Note 1: Benzo[a]pyrene is the only PAHPOM with an RfC ($2\text{E}-6 \mu\text{g}/\text{m}^3$). Therefore, chronic noncancer HI values for the PAHPOM group are based solely on benzo[a]pyrene concentrations.

Note 2: Even though chrysene URE is $9.6\text{E}-7$, put into $9.6\text{E}-6$ group (there is no lower risk group other than 0 so this is conservative).

Note 3: pollutant code retired in 2016

2.1.1.2. Metal groups

AirToxScreen includes metal compound groups consistent with metal emissions in the NEI. Metal emissions in the 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 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.

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 AirToxScreen modeling.

The three nickel compounds and three chromium VI compounds in the 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	AirToxScreen 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

* Metal speciation factor is the ratio of the molecular weight of the metallic element to the molecular weight of the compound.

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 AirToxScreen 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 PM₁₀ 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 PM₁₀ 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 "AirToxScreen_Pollutants.xlsx" within the [Supplemental Data](#) folder accompanying this TSD includes a crosswalk that contains NEI pollutant codes/descriptions, AirToxScreen group information, CMAQ names, metal adjustment factors, URE, RfC and target organ information.

2.1.2. Emissions Categorization: NEI and AirToxScreen

As explained on the [NEI website](#), the 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.

AirToxScreen summaries are provided by AirToxScreen broad summary categories and by more detailed source groups. The broad AirToxScreen 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 AirToxScreen category does not. Similarly, the AirToxScreen 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 AirToxScreen 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 AirToxScreen categories

NEI Data Category	AirToxScreen Category (Reflecting AirToxScreen Summary Results)
Point	Point¹
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. AirToxScreen includes airports and railyards in the nonroad sector.
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. Wild fires, prescribed fires, and agricultural fires are included., CMVs and locomotive emissions are included. Agricultural activities (crops & livestock dust, fertilizer application, and livestock waste) are included. Biogenic emissions that come from vegetation are also included.	Same as NEI nonpoint except: <ul style="list-style-type: none"> • AirToxScreen includes wild fires, prescribed fires and agricultural fires as FIRES. • AirToxScreen includes CMV and line-haul locomotives in Nonroad • AirToxScreen has a separate category for biogenics
Onroad	Onroad
Emissions estimates for mobile sources, such as cars, trucks and buses.	Same as NEI onroad.
Nonroad	Nonroad
Emissions estimates for nonroad equipment such as lawn and garden equipment, agricultural, construction, industrial and commercial equipment and recreational equipment	Same as NEI nonroad, but the AirToxScreen 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	AirToxScreen Category (Reflecting AirToxScreen 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.
	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.

¹In results presented online for assessments for the 2002 and early 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 AirToxScreen, and stationary-source emissions are referred to only as point-source or nonpoint-source emissions. Point sources in the AirToxScreen 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 current NEI and Emissions Used for AirToxScreen

Although 2020 NEI is the main basis of the emissions fed into the air quality models for AirToxScreen, there were several differences between the 2020 NEI and emissions data used for the AirToxScreen modeling.

State, local, and tribal agencies reviewed emissions information along with scaled preliminary risk information, which were estimated from pollutant-specific preliminary draft AERMOD 2020 results. EPA reviewed the comments and changes and incorporated the accepted changes into the modeling platform.

Differences that result from differences in emissions processing (which reflect the specific role and function of the resulting inventory within the context of the AirToxScreen 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 2-6 shows the changes made and why. Even though SMOKE does not use the fugitive release point parameters, they are included in the table to make it complete.

Table 2-6. Release parameter defaults/changes to the FF10 inventory files for point sources

Field	Existing Value	New Value	Conditions/Notes ¹
For point sources with stack releases (ERPtype NOT equal to "1")			
stkhgt	missing	use pstk ² or global defaults ²	None
stkdiam	missing	use pstk ² or global defaults ³	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
stktemp	missing	use pstk ² or global defaults ³	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)
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)
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)
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)
For fugitive release points (not used in CMAQ)			
fug_width_ydim	missing	32.808 ft	None
fug_length_xdim	missing	32.808 ft	None
fug_angle	missing	0	None
fug_height	missing	10 ft	fug_width_ydim and /or fug_length_xdim are missing
fug_height	missing	0	WHEN fug_width_ydim and fug_length_xdim are not missing and > 0
For coke ovens: any release point that emits coke oven emissions (pollutant code 140)—all pollutants at that release point are changed to the below			
stkhgt	<126 ft	126 ft	erptype NOT = "1"
fug_height	<126 ft	126 ft	erptype = "1"
fug_length_xdim	<50 ft	50 ft	erptype = "1"
fug_width_ydim	<50 ft	50 ft	erptype = "1"
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			

Field	Existing Value	New Value	Conditions/Notes ¹
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, California provided emissions are adjusted in the modeling platform, and adjusted model-ready files sum to annual totals from California, but have the temporal and spatial patterns reflecting the highly resolved meteorology and SMOKE MOVES. The AirToxScreen 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 (for example, 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.
- 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 are differences in the sources of emissions used by CMAQ and AERMOD. Differences in the emissions inputs to the two models are due to design differences in how the models are run. The emissions input into AERMOD exclude AirToxScreen categories more appropriately addressed by CMAQ, namely biogenics and three types of fires: wildfires, prescribed burning and agricultural-field burning. Biogenic emissions are 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 are 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 are also differences in the way we process or model 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 2-7 summarizes the spatial allocation differences between the models. As seen in Table 2-7, 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 2-7. 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 (CONUS), 9-km (AK), and 3-km (HI, PR, and VI) grid cells
Airports	Point	Point – runways & 10 m ² areas consistent with NEI geographic coordinates	12-km (CONUS), 9-km (AK), and 3-km (HI, PR, and VI) grid cells

Category	NEI Resolution	Spatial Approach for AERMOD	Spatial Approach for CMAQ
Locomotives	Point (railyards) and County/Shape	Nonpoint – 12-km grid cells in the CONUS domain, 9-km grid cells for AK; 3-km grid cells for HI, PR and VII Point – point fugitives	12-km (CONUS), 9-km (AK), and 3-km (HI, PR, and VI) grid cells
CMVs, ports and underway	County/Shape	Shapes from the NEI; separate shapes used for CMV at ports versus underway	12-km (CONUS), 9-km (AK), and 3-km (HI, PR, and VI) 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 (CONUS), 9-km (AK), and 3-km (HI, PR, and VI) grid cells
Agricultural burning and biogenic emissions	County	Not modeled	12-km (CONUS), 9-km (AK), and 3-km (HI, PR, and VI) grid cells
Fires (prescribed and wild)	Point	Not modeled	12-km (CONUS), 9-km (AK), and 3-km (HI, PR, and VI) 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 is described in depth in the technical support document found on the [emissions modeling platforms website](#). For 2020 information, please see in particular: <https://www.epa.gov/air-emissions-modeling/2020-emissions-modeling-platform-technical-support-document>. 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. Sectors in the CMAQ AirToxScreen Platform

For the purposes of preparing the CMAQ model-ready emissions, the NEI is 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 AirToxScreen CMAQ platform are listed in Table 2-8.

Table 2-8. Platform sectors for the 2020 emissions modeling platform

Platform Sector: <i>abbreviation</i>	NEI Data Category	Description and Resolution of the Data Input to SMOKE
EGU units: <i>ptegu</i>	Point	2020 NEI point source EGUs, replaced with hourly Continuous Emissions Monitoring System (CEMS) values for NOX and SO2, and the remaining pollutants temporally allocated according to CEMS heat input where the units are matched to the NEI. Emissions for all sources not matched to CEMS data come from 2020 NEI point inventory. Annual resolution for sources not matched to CEMS data, hourly for CEMS sources. EGUs closed in 2020 are not part of the inventory.
Point source oil and gas: <i>pt_oilgas</i>	Point	2020 NEI point sources that include oil and gas production emissions processes for facilities with North American Industry Classification System (NAICS) codes related to Oil and Gas Extraction, Natural Gas Distribution, Drilling Oil and Gas Wells, Support Activities for Oil and Gas Operations, Pipeline Transportation of Crude Oil, and Pipeline Transportation of Natural Gas. Includes U.S. offshore oil production.
Aircraft and ground support equipment: <i>airports</i>	Point	2020 NEI point source emissions from airports, including aircraft and airport ground support emissions. Annual resolution.
Remaining non-EGU point: <i>ptnonipm</i>	Point	All 2020 NEI point source records not matched to the airports, <i>ptegu</i> , or <i>pt_oilgas</i> sectors. Includes 2020 NEI rail yard emissions. Annual resolution.
Agricultural: <i>livestock</i>	Nonpoint	2020 NEI nonpoint livestock emissions. Livestock includes ammonia and other pollutants (except PM2.5). County and annual resolution.
Agricultural Fertilizer	Nonpoint	2020 agricultural fertilizer ammonia emissions computed inline within CMAQ.
Area fugitive dust: <i>afdust</i>	Nonpoint	PM10 and PM2.5 fugitive dust sources from the 2020 NEI nonpoint inventory; including building construction, road construction, agricultural dust, and paved and unpaved road dust. The emissions modeling system applies a transport fraction reduction and a zero-out based on 2020 gridded hourly meteorology (precipitation and snow/ice cover). Emissions are county and annual resolution.
Biogenic: <i>beis</i>	Nonpoint	Year 2020 emissions from biogenic sources. These were left out of the CMAQ-ready merged emissions, in favor of inline biogenic emissions produced during the CMAQ model run itself. Version 4 of the Biogenic Emissions Inventory System (BEIS) was used with Version 6 of the Biogenic Emissions Landuse Database (BELD6). Therefore, the biogenic emissions used here are similar to the 2020 NEI biogenic emissions, but not exactly the same.
Category 1, 2 CMV: <i>cmv_c1c2</i>	Nonpoint	2020 NEI Category 1 (C1) and Category 2 (C2), commercial marine vessel (CMV) emissions based on Automatic Identification System (AIS) data. Point and hourly resolution.
Category 3 CMV: <i>cmv_c3</i>	Nonpoint	2020 NEI Category 3 (C3) commercial marine vessel (CMV) emissions based on AIS data. Point and hourly resolution.
locomotives: <i>rail</i>	Nonpoint	Line haul rail locomotives emissions from 2020 NEI. County and annual resolution.
Nonpoint source oil and gas: <i>np_oilgas</i>	Nonpoint	Nonpoint 2020 NEI sources from oil and gas-related processes. County and annual resolution.

Platform Sector: abbreviation	NEI Data Category	Description and Resolution of the Data Input to SMOKE
Residential Wood Combustion: <i>Rwc</i>	Nonpoint	2020 NEI nonpoint sources with residential wood combustion (RWC) processes. County and annual resolution.
Solvents: <i>np_solvents</i>	Nonpoint	Emissions of solvents from the 2020 NEI (Seltzer, 2021). Includes household cleaners, personal care products, adhesives, architectural and aerosol coatings, printing inks, and pesticides. Annual and county resolution.
Remaining nonpoint: <i>Nonpt</i>	Nonpoint	2020 NEI nonpoint sources not included in other platform sectors. County and annual resolution.
Nonroad: <i>Nonroad</i>	Nonroad	2020 NEI nonroad equipment emissions developed with MOVES3, including the updates made to spatial apportionment that were developed with the 2016v1 platform. MOVES3 was used for all states except California, which submitted their own emissions for the 2020 NEI. County and monthly resolution.
On-road: <i>Onroad</i>	On-road	Onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles from 2020 NEI. Includes the following emission processes: exhaust, extended idle, auxiliary power units, evaporative, permeation, refueling, vehicle starts, off network idling, long-haul truck hoteling, and brake and tire wear. MOVES3 was run for 2020 to generate emission factors.
On-road California: <i>onroad_ca_adj</i>	On-road	California-provided 2020 CAP and HAP (VOCs and metals) onroad mobile source gasoline and diesel vehicles from parking lots and moving vehicles based on Emission Factor (EMFAC), gridded and temporalized based on outputs from MOVES3. Polycyclic aromatic hydrocarbon (PAH) emissions are based on MOVES3.
Point source agricultural fires: <i>ptagfire</i>	Nonpoint	Agricultural fire sources for 2020 developed by EPA as point and day-specific emissions. Only EPA-developed ag. fire data are used in this study, thus 2020 NEI state submissions are not included. Agricultural fires are in the nonpoint data category of the NEI, but in the modeling platform, they are treated as day-specific point sources. Updated HAP-augmentation factors were applied.
Point source prescribed fires: <i>ptfire-rx</i>	Events	Point source day-specific prescribed fires for 2020 NEI computed using SMARTFIRE 2 and BlueSky Pipeline. The ptfire emissions were run as two separate sectors: ptfire-rx (prescribed, including Flint Hills / grasslands) and ptfire-wild.
Point source wildfires: <i>ptfire-wild</i>		Point source day-specific wildfires for 2020 NEI computed using SMARTFIRE 2 and BlueSky Pipeline.
Non-US. fires: <i>ptfire_othna</i>	N/A	Point source day-specific wildfires and agricultural fires outside of the U.S. for 2020. Canadian fires for May through December are provided by ECCC. All other fire emissions, including Canadian emissions from January through April, as well as Mexico, Caribbean, Central American, and other international fires, are from v2.5 of the Fire INventory (FINN) from National Center for Atmospheric Research (Wiedinmyer, C., 2023).
Canada Area Fugitive dust sources not from the NEI: <i>canada_afdust</i>	N/A	Area fugitive dust sources from ECCC for 2020 with transport fraction and snow/ice adjustments based on 2020 meteorological data. Annual and province resolution.
Canada Point Fugitive dust sources: <i>canada_ptdust</i>	N/A	2020 point source fugitive dust sources from ECCC with transport fraction and snow/ice adjustments based on 2020 meteorological data. Monthly and province resolution.

Platform Sector: <i>abbreviation</i>	NEI Data Category	Description and Resolution of the Data Input to SMOKE
Canada and Mexico stationary point sources: <i>canmex_point</i>	N/A	Canada and Mexico point source emissions not included in other sectors. Canada point sources for 2020 were provided by ECCC and Mexico point source emissions for 2016 were provided by SEMARNAT. Mexico sources were projected from 2019ge (EPA, 2023a) with COVID adjustments applied. Canada monthly temporalization adjusted for COVID. Annual and monthly resolution.
Canada and Mexico agricultural sources: <i>canmex_ag</i>		Canada and Mexico agricultural emissions. Canada point sources for 2020 were provided by ECCC and Mexico emissions for 2016 were provided by SEMARNAT and adjusted to 2019. COVID adjustments were not applied to the ag sector. Annual resolution.
Canada low-level oil and gas sources: <i>canada_og2D</i>		2020 Canada emissions from upstream oil and gas. This sector contains the portion of oil and gas emissions which are not subject to plume rise. The rest of the 2020 Canada oil and gas emissions are in the <i>canmex_point</i> sector. Provided by ECCC with COVID-adjusted monthly temporalization. Monthly resolution.
Canada and Mexico nonpoint and nonroad sources: <i>canmex_area</i>	N/A	2020 Canada and Mexico nonpoint source emissions not included in other sectors. Canada: ECCC provided a 2020 inventory and surrogates. Mexico: applied COVID adjustments to 2019ge. Monthly temporalization adjusted for COVID.
Canada on-road sources: <i>Canada_onroad</i>	N/A	Canada onroad emissions. 2020 Canada inventory provided by ECCC and processed using updated surrogates. COVID impacts applied to monthly profiles (not to annual totals). Province and monthly resolution.
Mexico on-road sources: <i>Mexico_onroad</i>	N/A	Mexico onroad emissions. 2020 MOVES-Mexico with COVID adjustments applied. Municipio and monthly resolution.

2.2.2. 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 AirToxScreen 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 with the option to generate biogenic emissions turned on.

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

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

2.2.3. 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.

In the AirToxScreen modeling platform, all CMAQ species that are explicit AirToxScreen 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.

More information on speciation approach and detailed tables of modeling species are available in section 3.2 of the [technical support document](#) for the 2020 emissions modeling platform.

2.2.4. 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.

Sector-specific information, including meteorological-based temporal allocation and detailed charts of the temporal settings used for each sector in the emissions modeling platform can be found in section 3.3 of the [technical support document](#). Ancillary data are available at the [emissions modeling platform ftp site](#).

2.2.5. Spatial Allocation

For CMAQ, all emissions were allocated to 12-km grid cells for the CONUS domain, 9-km grid cells for the AK domain, and 3-km grid cells for HI domain and VI/PR domain. 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 (or 9-km/3-km) grid cells using spatial surrogates, which are developed based on shapefiles of data with spatial patterns expected for the emissions category.

Spatial surrogates were assigned to emissions sources based on SCC. A detailed description of these assignments can be found in section 3.4 of the emissions modeling platform [technical support document](#).

2.3. Preparation of Emissions Inputs for AERMOD

For AirToxScreen, we use the capability in SMOKE to produce helper files that can 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 km²) 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.

$$C_{\text{pollutant } i} = \frac{X}{Q} \times \text{Emissions}_{\text{pollutant } i}$$

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/Q s 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 AirToxScreen.

In AirToxScreen, we allocate county-level emissions to grid cells for both CONUS and non-CONUS (i.e., Alaska, Hawaii, Puerto Rico and the Virgin Islands) sources; however, we use different grid cell resolutions, as discussed below.

2.3.1. Source and Run Groups – Overview

Source and run groups are similar, but serve different functions in AirToxScreen. Source groups provide source attribution for the AirToxScreen census block 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 2-9. 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 2-9. Run groups for AERMOD

Run Group	NEI Data Category	AirToxScreen 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) Ports	Nonpoint	Nonroad	Shapes	POLYGON
Commercial Marine Vessels (CMV) Underway	Nonpoint	Nonroad	Gridcell	AREA (gridcell)

¹Rail yards.

²Locomotives.

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 2-10.

Table 2-10. 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 2-11 and Table 2-12 describe the source characteristics of these run groups and how they relate to the platform sectors used for CMAQ emissions modeling.

Table 2-11. Non-gridded AERMOD run groups

Run Group	NEI Category and AirToxScreen 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 2017 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 AirToxScreen HAP emissions except for facility source type 100 (airports).

Run Group	NEI Category and AirToxScreen 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.
CMVP	NEI: Nonpoint Platform: cmv_c12 and cmv_c3	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 NEI inventory shapes (there could be multiple polygons per port). Temporal: hourly data based on AIS activity data	C1/C2 and C3 commercial marine vessels at ports.

Table 2-12. AERMOD gridded run groups

Run Group	NEI Category and AirToxScreen 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.
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).

Run Group	NEI Category and AirToxScreen CMAQ Sector	AERMOD Modeling Features: Release Height (RH; meters), Initial Vertical Dispersion (σ_z ; meters) and Temporal Approach	Description of Sources
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 AirToxScreen 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.

Run Group	NEI Category and AirToxScreen CMAQ Sector	AERMOD Modeling Features: Release Height (RH; meters), Initial Vertical Dispersion (σ_z ; meters) and Temporal Approach	Description of Sources
CMVU12 CMVU9AK CMVU3HI CMVU3PR	NEI: nonpoint Platform: cmv_c12 and cmv_c3	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. Resolution is 12 km for CONUS, 9 km for AK and 3 km for HI and PR/VI Temporal: hourly by grid cell based on AIS activity data	C1/C2 and C3 commercial marine vessels underway

Each run group contains one or more source groups for purposes of presenting the AirToxScreen results. There are no situations where a source group fits into multiple run groups. Table 2-13 lists these source groups.

Table 2-13. AirToxScreen source groups

Source Group Name	AirToxScreen Category	Source Group Description	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)	POINT
OR-LightDuty-OffNetwork-Gas	Onroad	On-road, light duty, nondiesel (i.e., gasoline & ethanol blends) vehicles – off network processes (e.g., “starts”)	LDOFF
OR-LightDuty-OffNetwork-Diesel	Onroad	On-road, light duty, diesel vehicles – off network processes (e.g., “starts”)	LDOFF
OR-HeavyDuty-OffNetwork-Gas	Onroad	On-road, heavy duty, nondiesel vehicles – off network processes (e.g., “starts”)	HDOFF
OR-HeavyDuty-OffNetwork-Diesel	Onroad	On-road, heavy duty, diesel vehicles – off network processes (e.g., “starts”)	HDOFF
OR-LightDuty-OnNetwork-Gas	Onroad	On-road, light duty, nondiesel vehicles – on network processes (e.g., running emissions)	LDON
OR-LightDuty-OnNetwork-Diesel	Onroad	On-road, light duty, diesel vehicles – on network processes (e.g., running emissions)	LDON
OR-HeavyDuty-OnNetwork-Gas	Onroad	On-road, heavy duty, nondiesel vehicles – on-network processes (e.g., “running”)	HDON
OR-HeavyDuty-OnNetwork-Diesel	Onroad	On-road, heavy duty, diesel vehicles – on-network processes (e.g., “running”)	HDON
OR-Refueling	Onroad	On-road refueling	LDON
OR-HeavyDuty-Hoteling	Onroad	On-road, heavy duty diesel vehicle extended idling and auxiliary power units	HOTEL

Source Group Name	AirToxScreen Category	Source Group Description	Run Group
NR-Recreational-inc-PleasureCraft	Nonroad	Off-road motorcycles, snow mobiles, golf carts, outboard pleasure craft, personal watercraft, inboard/sterndrive pleasure craft, etc.	NONRD
NR-Construction	Nonroad	Paving equipment, plate compactors, trenchers, tampers/rammers, tractors/loaders/backhoes, cranes, signal boards/light plants, etc.	NONRD
NR-CommercialLawnGarden	Nonroad	Mowers, leaf blowers, turf equipment, chippers/ stump grinders, tillers, chainsaws, snow blowers, etc.	NONRD
NR-ResidentialLawnGarden	Nonroad	Mowers, leaf blowers, tillers, chainsaws, snow blowers, shredders, etc.	NONRD
NR-Agriculture	Nonroad	Agricultural tractors, combines, sprayers, balers, tillers, irrigation sets, swathers, etc.	NONRD
NR-CommercialEquipment	Nonroad	Generator sets, pressure washers, pumps, hydropower units, etc.	NONRD
NR-AllOther	Nonroad	Industrial Equipment (Forklifts, sweepers/scrubbers, other oil field equipment), logging equipment, railroad maintenance, underground mining equipment	NONRD
NR-CMV_C1C2_ports	Nonroad	Commercial marine vessel (CMV) emissions – C1/C2 while at ports	NR-CMV Ports
NR-CMV_C3_ports	Nonroad	CMV emissions – C3 while at ports	NR-CMV Ports
NR-CMV_C1C2C3_underway	Nonroad	CMV emissions – both C1/C2 and C3 marine – while underway	NR-CMV Underway
NR-Locomotives	Nonroad	Locomotive emissions	NPLO
NR-Point-Airports	Nonroad	Point source inventory where facility type = 100 "Airport"	AIRPORT
NR- Point-Railyards	Nonroad	Point source inventory where facility type = 151 "Rail Yard"	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	NPFI
NP-CommercialCooking	Nonpoint	Commercial cooking (charbroiling, frying)	NPLO
NP-OilGas	Nonpoint	Oil and gas operations (pumps, dehydrators, tanks, engines)	OILGAS
NP-SolventsCoatings	Nonpoint	Degreasing, dry cleaning, consumer and commercial solvents, industrial surface coating, non-industrial surface coating	NPLO

Source Group Name	AirToxScreen Category	Source Group Description	Run Group
NP-StorageTransfer_BulkTerminals_GasStage 1	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)	NPLO
NP-MiscellaneousNonindustrial	Nonpoint	Poultry/livestock, laboratories, dental alloy, motor vehicle fires, portable fuel containers, residential charcoal grilling	NPLO
NP-FuelCombustion_not_RWC	Nonpoint	Residential, industrial, commercial and institutional (ICI) fuel combustion, but excludes residential wood combustion	NPHI
NP-ResidentialWoodCombustionRWC	Nonpoint	Residential wood combustion – woodstoves, fireplaces	RWC
NP-WasteDisposal	Nonpoint	Open burning, yard and household waste, managed burning, slash (logging debris), onsite incineration, waste water treatment, composting, landfills, scrap materials	NPHI
NP- AgriculturalLivestock	Nonpoint	Emissions from livestock waste (HAPs only)	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;¹ 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.

¹Offshore platforms (i.e., FIPS = 85; meaning federal waters) are excluded.

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.

The SMOKE FF10 was split into the CMAQ platform sectors (see Table 2-8) 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.² 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. 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 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 2-14 lists the release point types and how each is assigned an AERMOD type.

²AERMOD temporalization is performed at the level of source IDs, so using different temporalization schemes at one facility is possible.

Table 2-14. 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_YDIM³ – fugitive width (east/west), (ft)
- FUG_LENGTH_XDIM⁴ – 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 2-1, 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. 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)

³These 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.

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



Figure 2-1. 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 AirToxScreen 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.⁴ 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

⁴AERMOD temporalization is performed at the level of source IDs, so using different temporalization schemes at one facility is possible.

the SMOKE diurnal, weekly and monthly temporal profile values. Table 2-15 shows the temporal variability options for non-EGUs.

Table 2-15. 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 Fractions ¹
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
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 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 National Transportation Atlas Database or an OTAQ-supplied database. 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. 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.

See the [Emissions Modeling Platform](#) documentation for more information on temporalization of airport emissions.

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.

NEI Lead emissions near airports included lead emitted during the climb-out and approach modes, which occur at altitude and are not included in AirToxScreen. To account for this, we adjusted the NEI-specific lead⁵ 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 2-16).

Table 2-16. 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 2-17.

Table 2-17. Airport emissions file format

Field Name	Description
State abbrev	2-character abbrev. Use TB for tribe
Facility id	Identifier 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 AirToxScreen_Pollutants.xlsx) For airports, for metals only, multiply metal pollutant emissions from (2275050011=General Aviation /Piston; 2275060011=Air Taxi /Piston) by 0.5

⁵We 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.

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 method was first used in the 2014 NATA. Mostly, the same underlying surrogate data are used for AERMOD as CMAQ and are discussed in detail in the technical support document of the emissions modeling platform developed for CMAQ emissions. 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. As in CMAQ, the spatial surrogates are assigned according to SCC codes.

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 national emissions inventory technical support document.

2.3.4.1. On-road

On-road uses 5 run groups: LDON4, LDOFF12, HDON4, HDOFF12 and HOTEL4, as described in Table 2-12. 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 was updated for the 2020 Modeling Platform.

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 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.

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. 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.

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 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. Residential non-wood combustion is in 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

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/chimineas and hydronic heaters) and 2104009000 (fire logs).

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

The temporal approach uses 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. We took the same approach using 2020 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:

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

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 2020 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. This allows a fine spatial allocation resolution of 4 km. This 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
- $\# \text{ of days}_i$ = 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. 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 2-18 lists the SCCs in the CMV run group.

Table 2-18. NEI 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

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

For AERMOD, each port 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. 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 12 km CMAQ grid cells for CONUS and 9 km grid cells for Alaska and 3 km grid cells for Hawaii and Puerto Rico. The AirToxScreen 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 2-2.



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

For AirToxScreen, 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 2020 AirToxScreen differed from that of previous AirToxScreen assessments (EPA, 2022). For previous assessments, the urban/rural determination was based on population and on a source-by-source basis. For 2020, a more holistic approach was used in that the urban/rural status of the CMAQ grid cells was determined and that any source within the grid cell (including the gridded sources) would use the same urban/rural classification of the source's parent grid cell. The urban/rural classification is based on methodology outlined in Section 7.2.1 of the *Guideline on Air Quality Models* (EPA, 2023b). A source is considered urban if 50% or more of the area within 3 km of the source is commercial, industrial, or high intensity residential land use areas or if the population density in the same 3 km radius circle is 750 people/km². To determine the land use criterion, the AERSURFACE tool (EPA, 2020) was used with 2016 National Land Cover Data. AERSURFACE, normally used to calculate albedo, Bowen ratio, and surface roughness for AERMET, gives a count of land use types in a circle of user defined radius. The 2016 NLCD categories of interest were categories 23 and 24 (developed medium and developed high intensity, respectively). For the population density criterion, the 2020 Census block populations were used with the census block centroid used to determine if the block was within 3-km of the grid cell. The following methodology was used for the CONUS domain:

1. For nine points on or in a 12 km CMAQ grid cell (four corners, cell center, and locations between each corner), AERSURFACE was run for a 3-km radius circle for each point (Figure 2-3).

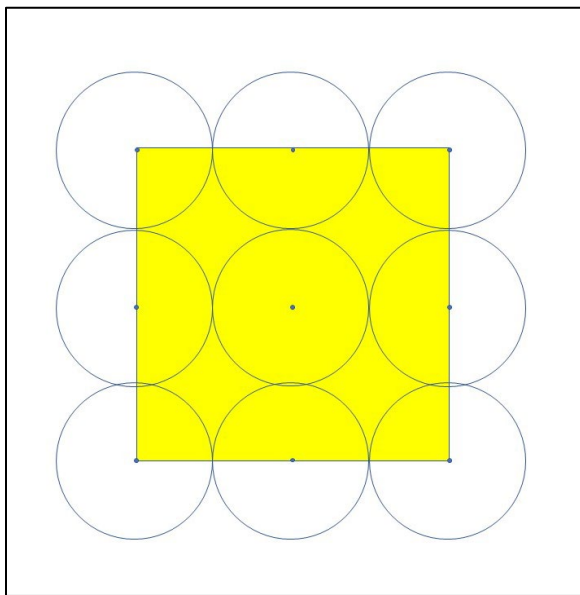


Figure 2-3. Points used for urban/rural determination for 12 km CMAQ grid cells.

2. For each of the nine points, the count of the number of categories 23 and 24 30x30 m cells within the 3-km radius circle was computed. Additionally, the number of cells where the impervious land cover is < 50% of the cell is calculated and subtracted from the number of cells with categories 23 and 24. For example, if there were 100 cells with categories 23 and 24 and 50 of those cells had 50% or less impervious coverage, then the final count is 50 cells. Given the definitions of categories 23 and 24, there should not be any cells with impervious coverage < 50% but this was done as a double check.
3. For each of the nine points, calculate the total population of 2020 census blocks within the 3 km radius circle for each point. The census block centroids were used as the locations to determine if a block was within 3 km.
4. Calculate the population density of the 3-km area for each point.
5. Determine if each point is urban based on land use (50% of the area in the 3-km circle is category 23 and 24) as well as if the point is urban based on population density. Each point has an urban designation based on land use and an urban designation based on population.
6. For each grid cell, if three or more of the points is considered urban either by land use or population density, then the grid cell is considered urban. Otherwise, it is rural.

The result was 954 grid cells were considered urban (Figure 2-4).



Figure 2-4. Urban grid cells for the CONUS domain.

Because recent land use data was not available for Alaska, Hawaii, Puerto Rico, and Virgin Islands, only population density was considered when determining the urban/rural status of the CMAQ grid cells for those domains. The use of population density alone is adequate since it was found for the contiguous US, the urban designation was mostly driven by the number of urban points based on population density. Because population is currently not available for the Virgin Islands, the 3-km CMAQ cells over the Virgin Islands are considered rural. When population data becomes available, the urban/rural classification for the Virgin Islands CMAQ cells will be assessed. For Alaska, nine points are used as described for the 12 km CMAQ cells over the contiguous US and an example layout is shown in Figure 2-5. For Hawaii and Puerto Rico, five points were used, the four corners and the center (Figure 2-6). For Alaska, if more than three points are considered urban based on population density, the cell is considered urban. For Hawaii and Puerto Rico if at least one of the points is considered urban based on population density, then the cell is considered urban. For Alaska, no cell is designated urban and for Hawaii, 56 cells are designated as urban (Figure 2-7) and 84 cells are designated urban for Puerto Rico (Figure 2-8).

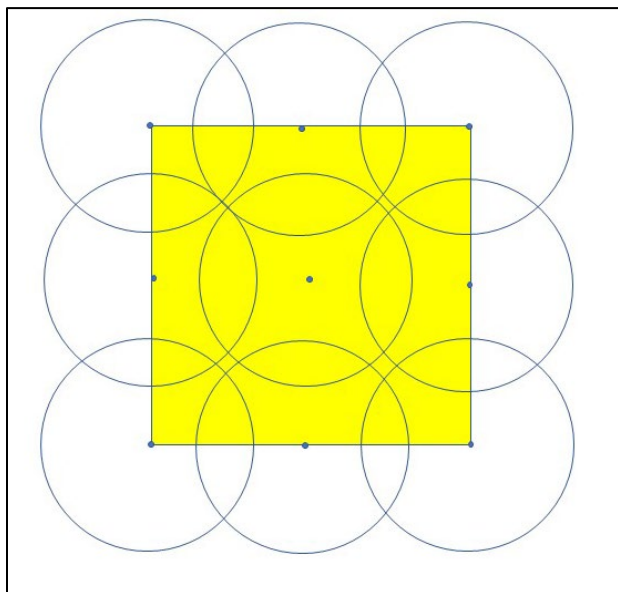


Figure 2-5. Points used for urban/rural determination for 9 km Alaska CMAQ grid cells.

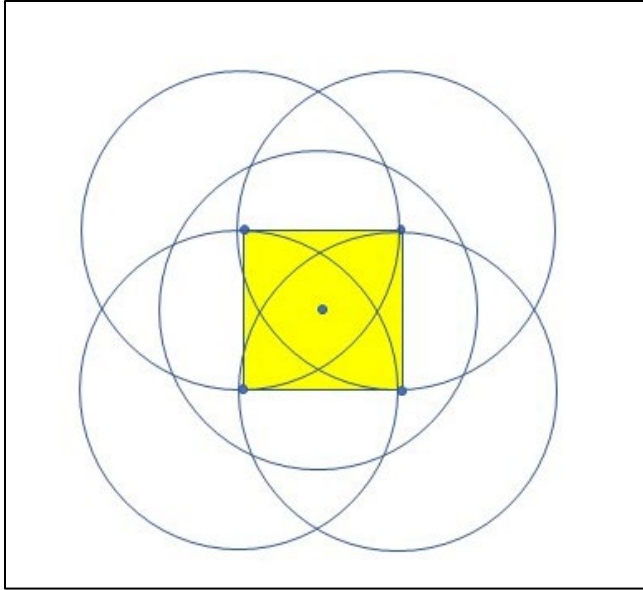


Figure 2-6. Points used for urban/rural determination for 3 km Hawaiian and Puerto Rico/Virgin Islands CMAQ grid cells.

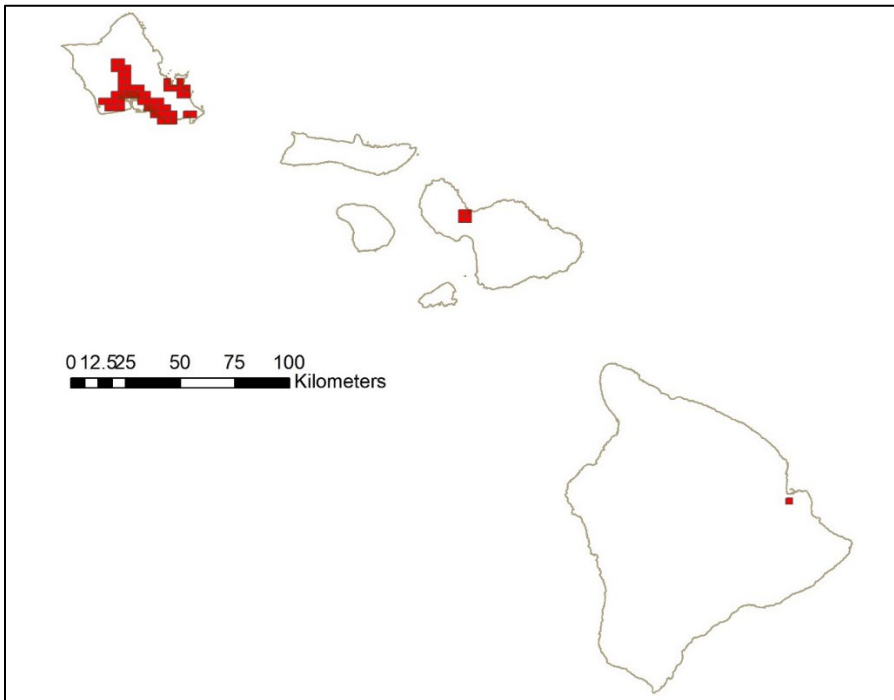


Figure 2-7. Urban grid cells for the Hawaiian domain.

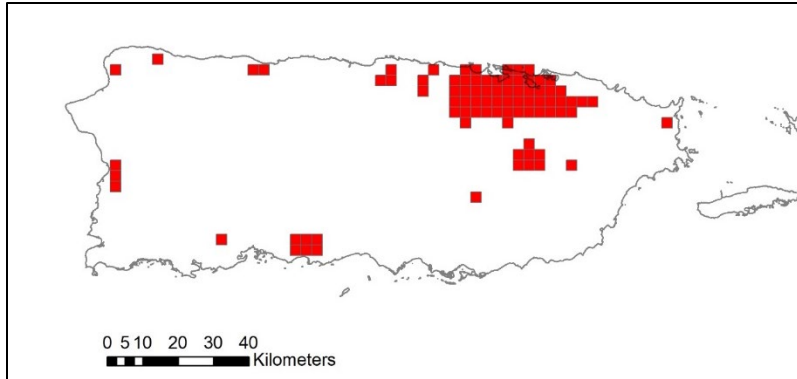


Figure 2-8. Urban grid cells for the Puerto Rico/Virgin Islands domain.

The input urban population needed by AERMOD is calculated for each of the urban cells in each domain (US, Hawaii, and Puerto Rico). To calculate the populations, each census block centroid is assigned to a CMAQ grid cell and each cell's total population is calculated. This is done for all CMAQ cells that contain a census block centroid. To calculate the urban population for each urban cell, the population of that cell and the surrounding eight cells are totaled, and that number represents the urban population for the urban cell.

3. Air Quality Modeling and Characterization

The AirToxScreen 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 AirToxScreen 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 AirToxScreen photochemical modeling, we used CMAQv5.4 (<https://doi.org/10.5281/zenodo.7218076>; <https://www.epa.gov/cmaq>; <http://www.cmascenter.org>.) with the Carbon-Bond 6r5 (CB6r5-CMAQ) chemical mechanism, AERO7 aerosol module with non-volatile Primary Organic Aerosol (POA). 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 (Appel et al., 2018). While most compounds are grouped when model chemistry is applied, the CB6r5-CMAQ chemical mechanism treats formaldehyde, acetaldehyde, benzene, methanol, and naphthalene explicitly. 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 CMAQv5.4 used for AirToxScreen.

Table 3-1. CMAQ HAPs

Air Toxic	CMAQ Species Name(s)
1,1,2,2-Tetrachloroethane	CL4_ETHANE
1,3-Butadiene	BUTADIENE13
1,3-Dichloropropene	DICL_PROPENE
1,4-Dichlorobenzene(p)	DICL_BENZENE
2,4-Toluene diisocyanate	TOL_DIIS
Acetaldehyde	ALD2, ALD2_PRIMARY
Acetonitrile	ACET_NITRILE
Acrolein	ACROLEIN, ACROLEIN_PRIMARY
Acrylic acid	ACRYACID
Acrylonitrile	ACRY_NITRILE

Air Toxic	CMAQ Species Name(s)
Arsenic	AASI, AASJ, ASSK
Benzene (including benzene from gasoline)	BENZENE
Benzo-A-Pyrene	ABENAPYI, ABENAPYJ, BENAPY
Beryllium	ABEK, ABEI, ABEJ
Cadmium	ACDI, ACDJ, ACDK
Carbon tetrachloride	CARBONTET
Carbon tetrachloride (without background)	CARB_TET_NBC
Carbonyl sulfide	CARBSULFIDE
Chlorine	CL2
Chloroform	CHCL3
Chloroprene	CHLOROPRENE
Hexavalent Chromium Compounds	ACR_VIK, ACR_VIJ, ACR_VII
Trivalent Chromium Compounds	ACR_IIIK, ACR_IIII, ACR_IIIJ
Diesel PM	ADE_ECI, ADE_ECJ, ADE_OCI, ADE_OCI, ADE_SO4J, ADE_NO3J, ADE_OTHRI, ADE_OTHRK, ADE_K
Ethyl benzene	ETHYLBENZENE
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	HEXMETH_DIS
Hexane	HEXANE
Hydrazine	HYDRAZINE
Hydrochloric acid	HCL
Lead Compounds	APBK, APBJ, APBI
Maleic anhydride	MAL_ANHYDRID
Manganese Compounds	AMN_HAPSK, AMN_HAPSJ, AMN_HAPSI
Mercury Compounds	HG, HGIIGAS, APHGI, APHGJ, APHGK
Methanol	MEOH
Methyl chloride (Chloromethane)	METHCHLORIDE
Methylene chloride (Dichloromethane)	CL2_ME
m-xylene, p-xylene, o-xylene and xylenes (isomers and mixture)	XYLENE
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

Air Toxic	CMAQ Species Name(s)
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)	PROPYL_DICL
Quinoline	QUINOLINE
Styrene	STYRENE
Tetrachloroethylene (Perchloroethylene)	CL4_ETHE
Toluene	TOLU
Trichloroethylene	CL3_ETHE
Triethylamine	TRIETHYLAMIN
Vinyl chloride	CL_ETHE

3.1.2. Dispersion Model Selection

For AirToxScreen air dispersion modeling, we used AERMOD (Cimorelli et al. 2005), 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 22112 was used for all AirToxScreen run groups.

3.2. Meteorological Data

3.2.1. Meteorological Data Inside the Contiguous States

3.2.1.1. Input WRF Data

For use in all AirToxScreen modeling, we derived gridded meteorological data for the contiguous United States (CONUS) from version 4.1.1 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 (<http://wrf-model.org/>). The CONUS WRF simulation used the same 12-km 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 459 by 299 grid cells and 35 vertical layers up to 50 millibars. The 12-km CONUS WRF model was initialized using the 12-km North American Model (12NAM; Only, <http://nomads.ncdc.noaa.gov/data.php>; download from ftp://nomads.ncdc.noaa.gov/NAM/analysis_only/) analysis product provided by National Climatic Data Center (NCDC). Where 12NAM data was unavailable, the 40-km Eta Data Assimilation System (EDAS) analysis (ds609.2) from the National Center for Atmospheric Research (NCAR) was used. Analysis nudging for temperature, wind, and moisture was applied above the boundary layer only. The model simulations were conducted continuously. The 'ipxwrf' program was used to initialize deep soil moisture at the start of the run using a 10-day spin-up period. The 2020 WRF meteorology simulated was based on 2011 National Land Cover Database (NLCD; National Land Cover Database 2011, <http://www.mrlc.gov/nlcd2011.php>). The WRF simulation included the physics options of the Pleim-Xiu land surface model (LSM), Asymmetric Convective Model version 2 planetary boundary

layer (PBL) scheme, Morrison double moment microphysics, Kain- Fritsch cumulus parameterization scheme utilizing the moisture-advection trigger (Ma and Tan 2009) and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme (Gilliam and Pleim, 2010). In addition, the Group for High Resolution Sea Surface Temperatures (GHRSSST; Stammer et al. 2003, analysis, <https://www.ghrsst.org/>) 1-km SST data was used for SST information to provide more resolved information compared to the more coarse data in the NAM analysis.

3.2.1.2. MCIP Processing for CMAQ

The 2020 WRF meteorological outputs were processed using the Meteorology-Chemistry Interface Processor (MCIP) package (Otte and Pleim 2010), version 5.3.3, 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 CMAQ meteorological inputs.

Table 3-2. Vertical layer structure for WRF and CMAQ (heights are layer top)

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

WRF & CMAQ Layers	Sigma P	Approximate Height (m)
9	0.9400	470
8	0.9500	390
7	0.9600	311
6	0.9700	232
5	0.9800	154
4	0.9850	115
3	0.9900	77
2	0.9950	38
1	0.9975	19
0	1.0000	0

3.2.1.3. MMIF Processing for AERMOD

WRF output was processed through the Mesoscale Model Interface (MMIF) program (version 3.4.2) to create AERMET-ready meteorological input data and processed in AERMET (version 22112). 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 2023c). As options, we used the default FLM layers (see MMIF User’s Guide for details), TOP for vertical interpolation and MMIF-calculated mixing heights. AERMET (version 22112) was run with the adjusted u* option to better represent concentrations in AERMOD under low-wind stable conditions.

3.2.2. Meteorological Data Outside the Contiguous States

3.2.2.1. Input WRF Data

The 2020 gridded meteorological data covering areas outside the contiguous states, the 3-km Hawaii and Puerto Rico/Virgin Islands and the 9-km Alaska domains, was derived from the publicly available WRF version 4.1.1, ARW core (<https://www2.mmm.ucar.edu/wrf/users/>).

Grid Domain Specifications:

- Alaska with 9 km horizontal grid spacing (312 cells in the x-direction, 252 cells in the y-direction) with 35 layers centered on 63° N and 155° W.
- Hawaii with 3 km horizontal grid spacing (225 cells in the x-direction, 201 cells in the y-direction) with 35 layers centered on 21° N and 157° W
- Puerto Rico and Virgin Islands with 3 km horizontal grid spacing (150 cells in the x-direction, 150 cells in the y-direction) with 35 layers centered on 18° N and 66° W.

The WRF model was initialized using the 0.25-degree Global Forecast System (GFS) model analysis (NCEP/NWS/NOAA/USDC, 2015, <https://doi.org/10.5065/D65D8PWK>). Analysis nudging for temperature, wind, and moisture was applied above the boundary layer only. The model simulations were conducted continuously. The ‘ipxwrf’ program was used to initialize deep soil moisture at the start of the run using a 10-day spin-up period. The 2020 WRF meteorology simulated was based on IGBP-

Modified MODIS 20-category Land Use data. The WRF simulation included the physics options indicated in Table 3-3. In addition, the Group for High Resolution Sea Surface Temperatures (GHR SST; Stammer et al. 2003, analysis, <https://www.ghrsst.org/>) 1-km SST data was used for SST information to provide more resolved information compared to the coarser data in the NAM analysis.

Table 3-3. WRF Options Used for the Alaska, Hawaii, and Puerto Rico/Virgin Islands domains.

Option	Settings for ALASKA	Settings for HI & PR/VI
Cumulus	Kain-Fritsch (w/ Trigger 1) w/ cu_rad_feedback = .true.	Kain-Fritsch (w/ Trigger 1) w/ cu_rad_feedback = .true.
Precipitation	Hourly Incremental	Hourly Incremental
PBL	MYNN2	YSU
Sea Surface Temperatures	GHR SST	GHR SST
LSM	MYNN	Noah
Microphysics	New Eta (Ferrier, mp_physics = 5)	New Eta (Ferrier, mp_physics = 5)
Longwave and Shortwave radiation	RRTMG	RRTMG
Model Top	50 mb	50 mb

3.2.2.2. MCIP Processing for CMAQ

The 2020 WRF meteorological outputs for the 3-km Hawaii and Puerto Rico/Virgin Islands and the 9-km Alaska domains were processed using the Meteorology-Chemistry Interface Processor (MCIP) package (Otte and Pleim 2010), version 5.3.3, 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 CMAQ meteorological inputs.

3.2.2.3. MMIF Processing for AERMOD

For meteorological data covering areas outside the contiguous states (Alaska, Hawaii, Puerto Rico and the U.S. Virgin Islands), the same methodology used for the contiguous U.S. described in Section 3.2.1.3 was used.

3.3. CMAQ Setup

3.3.1. Sources Modeled in CMAQ

AirToxScreen CMAQ modeling included a base year run (primary and secondary annual average concentrations) and “zero-out” runs for biogenics and fires (primary annual average concentrations).

The annual simulations included a “ramp-up” period of 10 days to mitigate the effects of initial concentrations. All 366 model days were used in the annual average levels of air toxics modeled. The model extends vertically from the surface to 50 millibars (approximately 17,600 meters) using a sigma-pressure coordinate system.

The CMAQ model runs were performed for four domains covering:

- (1) Contiguous United States (CONUS), as shown in Figure 3-1(a). This single domain covers the entire CONUS and large portions of Canada and Mexico using 12 km by 12 km horizontal grid spacing.
- (2) Alaska (Figure 3-1(b)) with 9 km by 9 km horizontal grid spacing.
- (3) Hawaii (Figure 3-1(c)) with 3 km by 3 km horizontal grid spacing.
- (4) Puerto Rico and Virgin Islands (Figure 3-1(d)) with 3 km by 3 km horizontal grid spacing.

Table 3-4 provides some basic geographic information regarding the four CMAQ domains. Air quality conditions at the outer boundary of the four CMAQ domains were taken from a global model (described in Section 3.3.2).

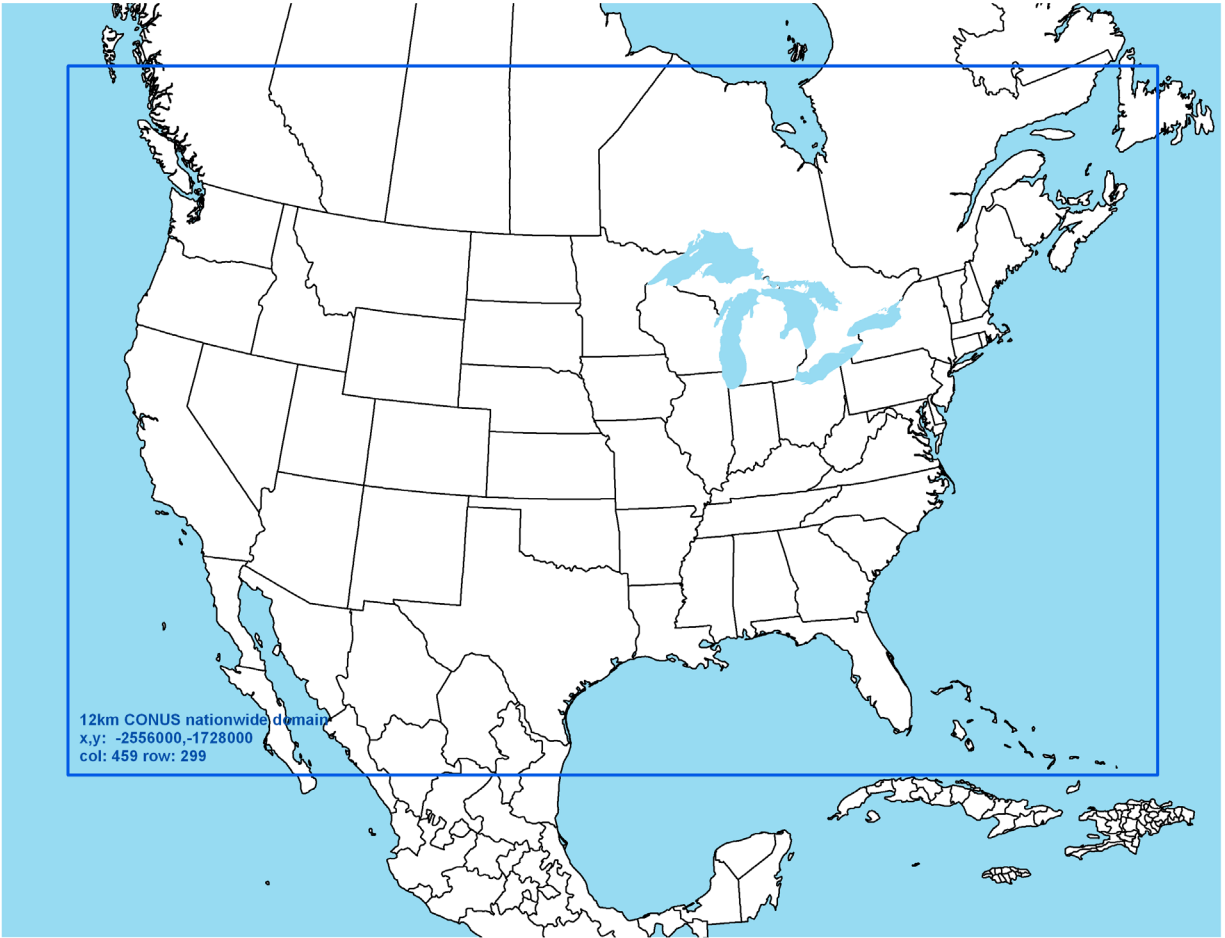


Figure 3-1(a). Map of the CMAQ modeling domain; the blue box denotes the 12 km CONUS modeling domain.



Figure 3-2(b). Map of the CMAQ modeling domain; the green box denotes the 9 km Alaska modeling domain.

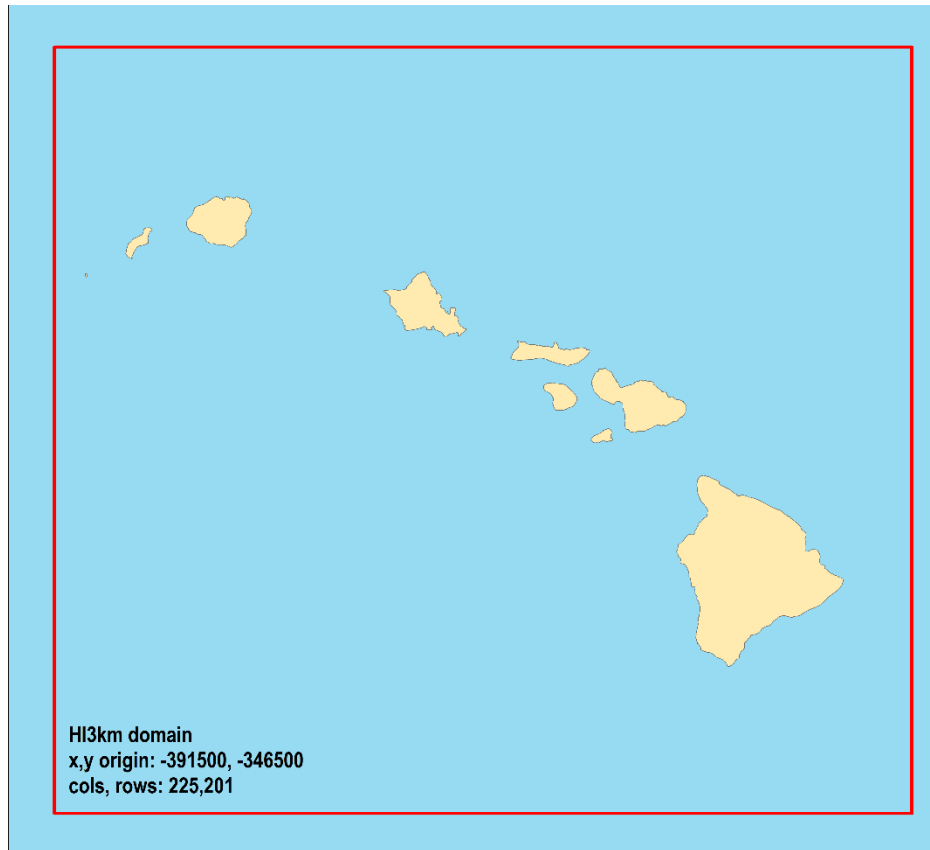


Figure 3-3(c). Map of the CMAQ modeling domain; the red box denotes the 3 km Hawaii modeling domain.

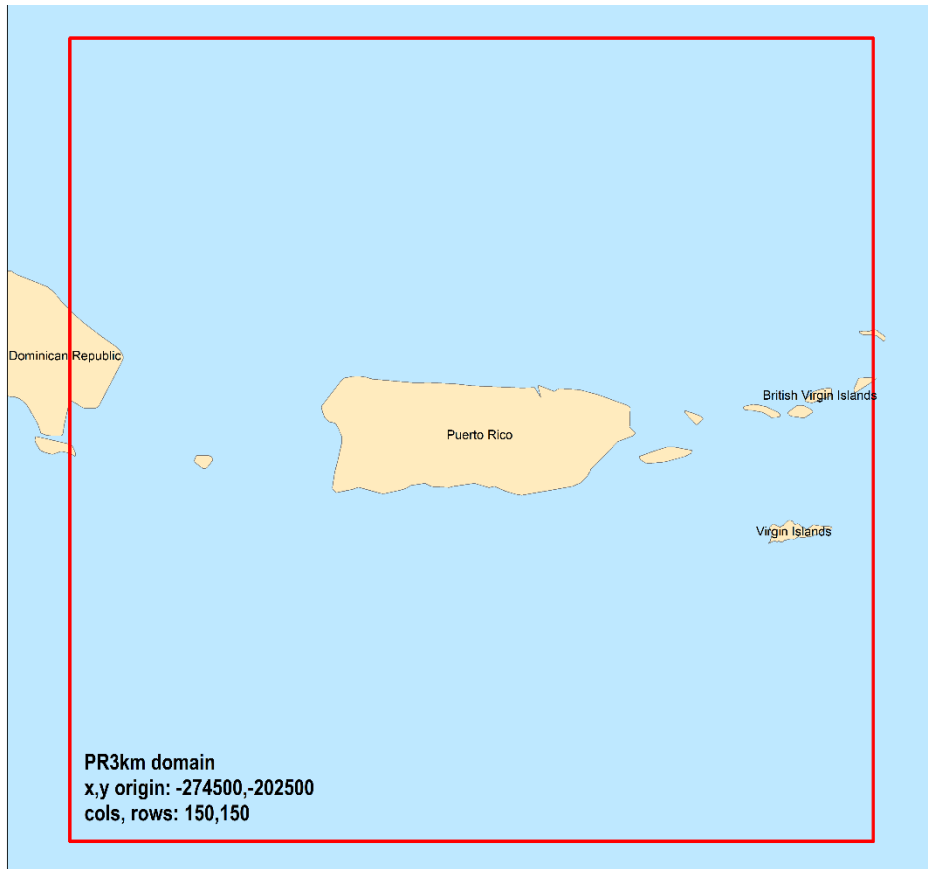


Figure 3-4(d). Map of the CMAQ modeling domain; the red box denotes the 3 km Puerto Rico/Virgin Islands modeling domain.

Table 3-4. Geographic information for the CMAQ modeling domains

CMAQ Modeling Domain Configurations					
Domain	Map Projection	Grid Resolution	Coordinate Center	True Latitudes	Dimensions & Vertical Extent
Continental US	Lambert Conformal Projection	12 km	97 W, 40 N	33 and 45 N	459 x 299 x 35
Alaska	Lambert Conformal Projection	9 km	63 N, 155 W	60 and 70 N	312 x 252 x 35
Hawaii	Lambert Conformal Projection	3 km	21 N, 157 W	19 and 22 N	225 x 201 x 35
Puerto Rico/Virgin Islands	Lambert Conformal Projection	3 km	18 N, 66 W	17 and 19 N	150 x 150 x 35

3.3.2. Boundary and Initial Conditions

The 2020 annual lateral boundary and initial species concentrations were provided using a global 3-D GEOS-Chem v14.0.1. GEOS-Chem is a 3-D model of atmospheric chemistry driven by meteorological inputs from the Goddard Earth Observing System of the National Aeronautics and Space Administration (NASA) Global Modeling Assimilation Office. GEOS-Chem was run using the standard (or default) options and full atmospheric chemistry (<https://geoschem.github.io/index.html>). The GEOS-Chem simulation was performed at 2 x 2.5-degree horizontal resolution with a 72-layer vertical structure (36 layers in troposphere, hybrid terrain following coordinate). Simulation used full chemistry including an online stratosphere, non-local planetary boundary layer, and simple secondary organic aerosols. The 2020 simulation required extending the methane inputs to the year 2020, updating lightning inputs, and other parameters for 2020. Emissions included online Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 2.1 (Guenther et al. 2012), online DUST module, and online sea salt module. Global Fire Emissions Database (GFED; <https://www.globalfiredata.org/>) were monthly mean. Anthropogenic emissions included fugitive, combustion, and industrial dust (Philip et al. 2017). Marine emissions were based on Community Emissions Data System (CEDS) version 2 including shipping vessels (<https://www.pnnl.gov/projects/ceds>). Aircraft Emissions Inventory Code (AEIC; Simone et al. 2013) monthly aircraft input data. In addition, CEDS and AEIC was scaled by Covid-19 adjustment Factors for Emissions (CONFORM) dataset (Doubria et al. 2021). Meteorology used in this 2020 GEOS-Chem run was from Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA2; GMAO 2015) meteorology at 2 x 2.5-degree. With the exception of input updates for 2020, these were the default options and inputs distributed with v14.0.1.

In addition to the standard chemistry, we also included Persistent Organic Pollutants (POPs) and mercury from specialized applications of GEOS-Chem. The POPs model configuration (Friedman et al. 2014) in v12.0.1 for 2018 applied to PHE, and PYR using 4x5 degree resolution. For mercury, we applied the mercury model configuration (Selin et al., 2007) from a 2013 v9 simulation with 47 layers and 2x2.5-degree resolution. The mercury concentrations were averaged to months and used as a climatology. Benzo-a-pyrene was applied as a monthly value from previous 2018 boundary conditions data files.

Where CMAQ species were not available from either the standard chemistry, the POPs or mercury simulation, then they were treated as climatology using values from previous multipollutant boundary condition applications.

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, Colorado (NWR); Barrow, Alaska (BRW); and Alert, Canada (ALT) and the Trinidad Head Site (AGAGE). More information on how these were derived is in Appendix C.

For the remaining CMAQ HAP BCs not provided by H-CMAQ 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 2020 remote concentration estimates

Pollutant	RCE at 298K and 1 atm (µg/m ³) 2020	RCE (pptv) 2020	Remote Network	Location(s)	CMAQ HAP
Chloroform	0.075	15.3	AGAGE	Trinidad Head	x
Methyl chloride (chloromethane)	1.155	559.7	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x
Carbon tetrachloride	0.487	77.5	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x
Methyl bromide (bromomethane)	0.028	7.1	NOAA GMD	KUM, MLO, NWR, BRW, ALT	
Methyl chloroform (1,1,1-trichloroethane)	0.007	1.2	NOAA GMD	KUM, MLO, NWR, BRW, ALT	
Dichloromethane (methylene chloride)	0.238	68.5	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x
Tetrachloroethene (perchloroethylene, tetrachloroethylene)	0.011	1.7	NOAA GMD	KUM, MLO, NWR, BRW, ALT	x

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

Air Toxic	2020 CMAQ Species Name(s)
1,1,2,2-Tetrachloroethane	CL4_ETHANE
1,3-Dichloropropene	DICL_PROPENE
1,4-Dichlorobenzene(p)	DICL_BENZENE
2,4-Toluene diisocyanate	TOL_DIIS
Acetonitrile	ACET_NITRILE
Acrylic acid	ACRYACID
Acrylonitrile	ACRY_NITRILE
Arsenic	AASI, AASJ, ASSK
Beryllium	ABEK, ABEI, ABEJ
Benzo-A-Pyrene	ABENAPYI, ABENAPYJ, BENAPY
Cadmium	ACDI, ACDJ, ACDK
Carbon tetrachloride	CARB_TET_NBC
Carbonyl sulfide	CARBSULFIDE
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	ETHYLBENZENE
Ethylene dibromide (Dibromoethane)	BR2_C2_12
Ethylene dichloride (1,2-Dichloroethane)	CL2_C2_12
Ethylene oxide	ETOX

Air Toxic	2020 CMAQ Species Name(s)
Hexamethylene-1,6-diisocyanate	HEXMETH_DIS
Hexane	HEXANE
Hydrazine	HYDRAZINE
Lead Compounds	APBK, APBJ, APBI
Maleic anhydride	MAL_ANHYDRID
Manganese Compounds	AMN_HAPSK, AMN_HAPSJ, AMN_HAPSI
Mercury Compounds	APHGI, APHGK
m-xylene, p-xylene, o-xylene and xylenes (isomers and mixture)	XYLENE
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)	PROPYL_DICL
Quinoline	QUINOLINE
Styrene	STYRENE
Trichloroethylene	CL3_ETHE
Triethylamine	TRIETHYLAMIN
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 at 1 km resolution. This differs from previous editions of AirToxScreen which used 1 km spacing in highly populated areas and 4 km in all other areas.
2. Populated 2020 census-block c receptors (discussed in Section 3.4.2.2)
3. Monitoring site receptors (discussed in Section 3.4.2.3)

For non-CONUS areas (AK, HI, PR, VI), we used:

1. Equally spaced “gridded” receptors at 1 km resolution. In previous editions of AirToxScreen, Alaska was modeled at 3 km resolution.
2. Populated 2020 census-block receptors
3. Monitoring site receptors

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 averaging concentrations for the gridded sources (nonpoint and mobile) and interpolating to block receptors and monitor receptors. Receptors at the monitoring locations were used in the model evaluation.

3.4.2.1. Gridded receptors

We used gridded receptors throughout the CONUS area and the three non-CONUS areas. 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. New to 2020, the gridded receptors were also used to calculate a gridded average concentration for the gridded source types (nonpoint and mobile) during AERMOD post-processing. For all domains, a uniform resolution of 1 km was used for the gridded receptors. This differed from previous editions of AirToxScreen where gridded receptor resolution was based on the population of the 2013 Core Base Statistical Areas (CBSA). See EPA (2022) for details on the previous receptor strategy. The updated strategy resulted in 144 receptors per 12-km CMAQ grid cell in the CONUS domain, nine receptors in each 3-km CMAQ grid cell in Hawaii and Puerto Rico/Virgin Islands, and 81 receptors in each 9-km CMAQ grid cell in Alaska.

Each gridded receptor represented the center of a subgrid cell within the CONUS 12-km CMAQ grid cells, Hawaii and Puerto Rico/Virgin Islands 3-km CMAQ grid cells, and Alaska 9-km CMAQ grid cells (Figure 3-2). These gridded receptors, plus populated block and monitor receptors when available within a subgrid cell, were averaged (Figure 3-3). These subgrid-cell averages were then used to calculate the overall AERMOD average within each 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 18081).

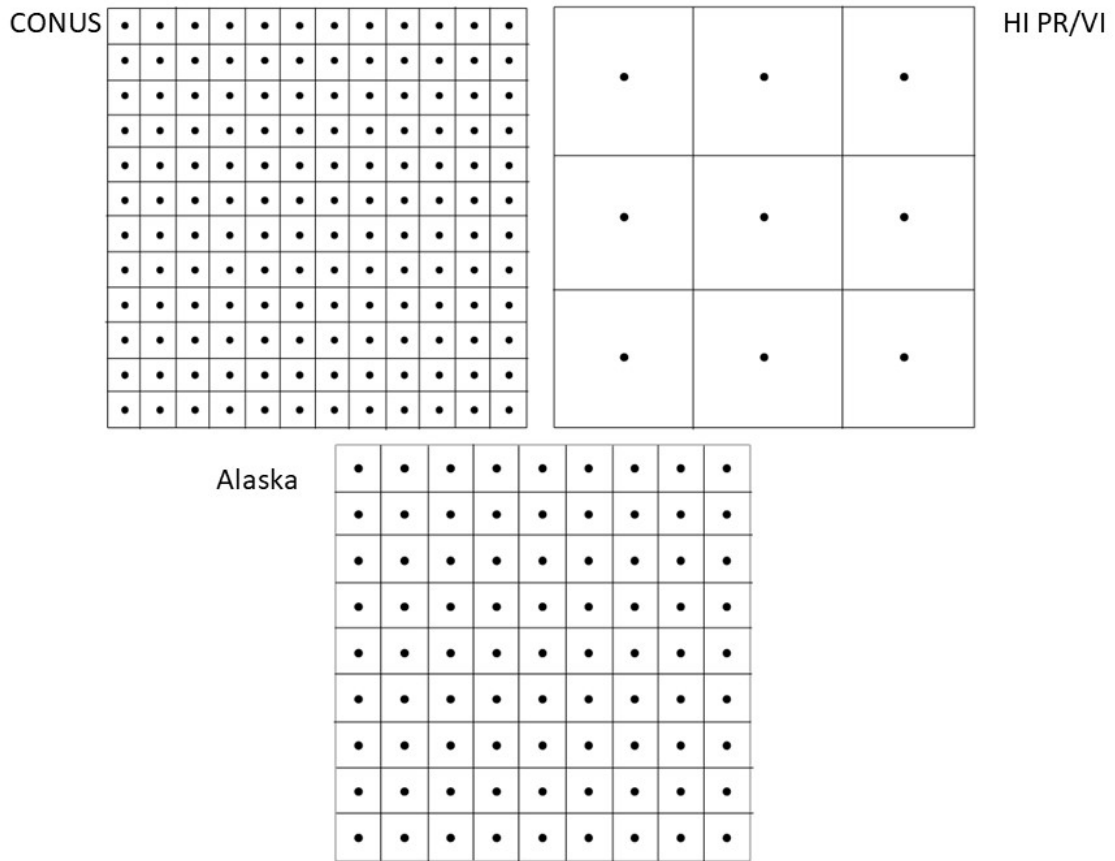


Figure 3-5. CONUS, Hawaii, Puerto Rico/Virgin Islands, and Alaskan receptor grid layouts in CMAQ Lambert Projection

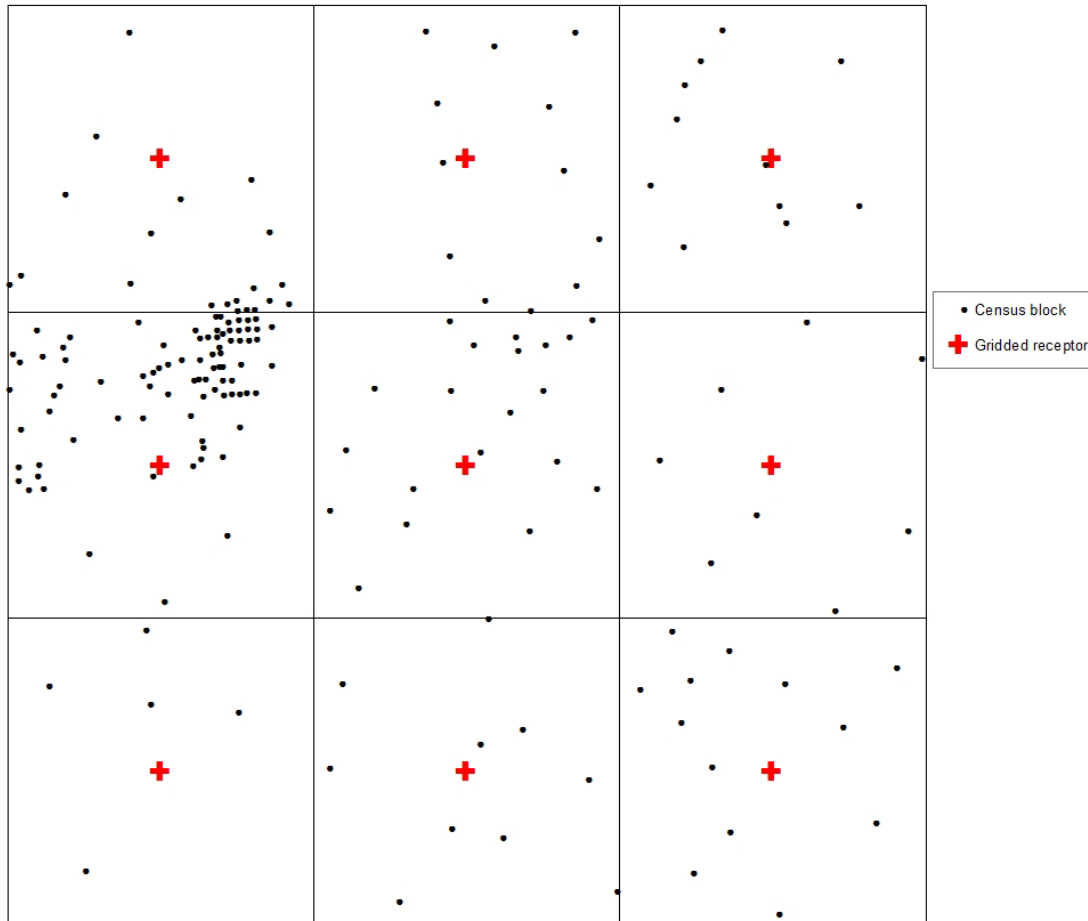


Figure 3-6. Example grid cell with subgrid cells and census blocks

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 gridded sources (all domains), any gridded receptor within a grid cell that was within 50 km of the center of the 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.

3.4.2.2. Census-block receptors

The locations of census-block receptors were based on the 2020 U.S. Census. Prior to the 2020 AirToxScreen, the census block centroids were used to represent the census blocks in modeling. Beginning with the 2020 AirToxScreen, census block concentrations, risks, and hazard quotients are reported. With this new emphasis on the census block results, EPA developed a new receptor strategy to represent the census blocks. Within each census block, a grid of receptors was created to calculate concentrations throughout the census block, not just at the census block centroid. This new approach would allow us to account for non-ambient air (i.e., receptors on facility property) and allow for more spatially representative concentrations for each census block. An example would be large rural blocks previously represented by the census block centroid but now the block has a grid to calculate more representative concentrations for the block (Figure 3-4).

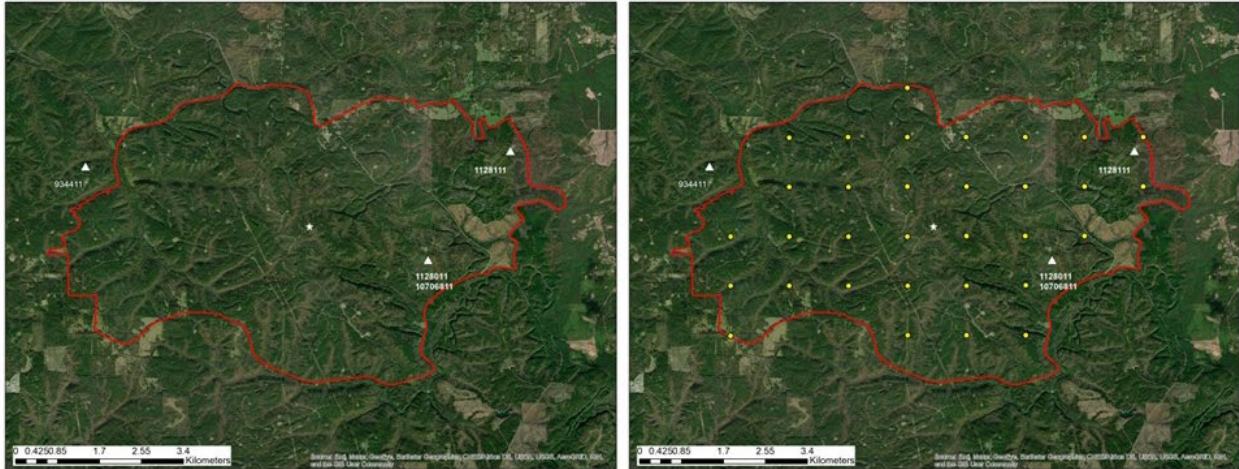


Figure 3-7. Rural census block represented by census block centroid on left and grid of receptors on right.

The grid resolution within each census block varied from 25 meters to 10 km at resolutions of 25 m, 50 m, 100 m, 250 m, 500 m, 1 km, 2 km, 5 km and 10 km, depending on the size of the census block. When developing the resolution of a census block, the initial resolution was 100 m. If there were fewer than five receptors in the block at 100 m resolution, the resolution would be increased to 50 m and subsequently to 25 m if there were still fewer than five receptors in the block at 50 m resolution. Alternatively, if there were more than 50 receptors at 100 m resolution, then the resolution was decreased to a coarser resolution (250 m, 500 m, etc.) until the number of gridded receptors was between 4 and 50 receptors. If the number of receptors was less than four and the resolution was 25 m or the census block was less than 500 x 500 m, then census block centroid was used to represent the census block in modeling. When performing the dispersion modeling for point sources in all domains, populated block receptors within 5 km of any emission point at the facility were explicitly modeled in AERMOD. For airports in all domains, any populated block receptor within 5 km of any point along a runway or 5 km from any point of the 100-by-100 m area sources was modeled. For gridded sources in all domains, census blocks were not modeled; concentrations were assigned to the census block receptors during the hybrid calculations. . For ports and underway sources in all domains, any block receptor within 5 km of a side of the polygon or within 10 km of the center of the source was explicitly modeled. Non-populated blocks were not modeled in AirToxScreen.

With the new census block receptor strategy, the probability that a receptor was located in non-ambient air (inside the fence line of a facility) was more likely than when using the older census block centroid receptor strategy. To check for ambient air and modeled receptors for a facility, port, or airport, receptor locations were checked to see if they were within a specified distance of sources for a facility or, when available, within a pre-determined property boundary of the facility. The use of property boundaries or an ambient air distance were used to flag receptors that do not represent ambient conditions for a particular census block and thus lead to a less representative block average concentration. For the 2020 AirToxScreen, 137 facilities used a property boundary to determine if a

receptor was located in ambient air relative to the facility⁶. If a receptor was located within the property boundary it was considered non-ambient and flagged in the AERMOD input file. Facilities without a property boundary used a specified distance to check receptor-source distances. The default distance used to check was 30 m. Ten facilities used distances other than 30 m, with those distances ranging from 35 to 150 m. If a receptor was within the specified distance from any source at the facility, it was considered non-ambient and flagged in the AERMOD input file. For airports and ports, property boundaries were not used and the default 30 m distance was used to check for ambient air. For airports, this distance was used to make sure no receptors on the runways for airports with runways. Treatment of non-ambient receptors in post-processing is discussed in Section 3.4.5 below.

3.4.2.3. AERMOD receptors at monitoring sites

The Ambient Monitoring Archive for HAPs monitoring data sites was used for the model evaluation (found in the [Supplemental Data](#) folder on the AirToxScreen website). Therefore, we obtained a unique set of geographic coordinates for all monitors to include as receptors in the AERMOD modeling. AERMAP (version 18081) 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. Unlike the census blocks, monitors were not checked for ambient air status.

3.4.3. Model Options

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

For sources determined to be urban based on the parent CMAQ grid cell, the AERMOD urban option was used with the population of the grid cell.

3.4.4. AERMOD Simulations

We ran AERMOD for each of the run groups described in Table 2-11. 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 gridded sources.

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 gridded sources, each CMV shape or 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

⁶ As new editions of AirToxScreen are released, the number of facilities with property boundaries will increase with each edition.

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 251.9957⁷ 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:

$$E_i = \frac{10000 \times (EF_i / \text{Weight})}{24} \times 251.9957$$

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/m² 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:

$$E_i = \frac{10000 \times EF_i}{366} \times 251.9957$$

Where 366 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 for point sources, airports and ports or 1) averaging of gridded receptors within a CMAQ grid cell for the gridded sources, and 2) calculation of HAP-specific concentrations. For all AirToxScreen run groups except the gridded sources, interpolation of concentrations to census blocks and monitors occurred at blocks and monitors from 5 to 50 km from the

⁷215.9957 is the conversion factor from tons/hr to g/s.

source. Interpolation was done for each individual facility (point source, airport, CMV shape) or gridded source. For the gridded source types, the gridded 1 km receptors were averaged within the CMAQ grid cell and the averaging matched the resolution of the source type.

As noted in Section 3.4.2.2, we flagged census block receptors that were within facility property boundaries or within the facility specific ambient air distance. During post-processing for a particular facility, the concentration for any receptor deemed non-ambient is ignored for that facility unless one of the following conditions are found:

- If census block is represented by the census block centroid, then the concentration is included for the facility so that the census block has some representation for the facility,
- If a census block has multiple receptors and all receptors for the census block are non-ambient one of them is selected to represent the census block for that facility.

Interpolation of census block receptors and monitors was done for the point sources, airports, and ports. 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 modeling domain and thus without four receptors around it, the interpolated receptor's concentration was the average of the nearest gridded receptors.

For nonmobile AirToxScreen 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, the HAP/source group-specific concentrations were calculated based on monthly emissions. The AERMOD output for these run groups 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. 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.
3. 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 8,784 hours (2020 was a leap year). The number may be less if calms and missing data are present in the meteorological data.
4. 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.

For 2020 AirToxScreen, we used a new approach for the gridded sources. In previous editions of AirToxScreen, census block receptors and monitors were interpolated from the gridded receptors modeled for these source types. This interpolation implies a certain level of precision in the spatial allocation of emissions. Within the AERMOD modeling for these gridded sources, gridded receptors near the center of the source have the highest concentrations within the grid cell and receptors near the edge of the grid cell, especially the upwind directions, would have the lowest concentrations of the grid cell. Interpolation to block receptors or monitors would then lead to the census blocks near the center of the grid cell having the highest concentrations and the census blocks near the edges would have lower concentrations. This concentration distribution would not necessarily be reflective of emissions but more of the meteorology. Depending on the source type, the emissions are spread throughout the CMAQ grid or in the case of certain CONUS onroad and oil & gas emissions, spread over 4 km cells. For example, consider the commercial cooking nonpoint source emissions. The emissions are spread over the 12 km grid cell and uniform throughout. Using the older interpolation method would lead to the highest block concentrations near the center of the 12 km cell. However, there may be no actual commercial cooking facilities near those blocks. To have a more representative gridded source concentrations, we decided to average the gridded receptors' concentrations within each CMAQ grid cell and that averaging matched the emissions resolution. Using the commercial cooking example once more for a 12 km CMAQ cell, the 144 gridded receptors' commercial cooking concentrations were averaged together for each HAP, resulting in a single commercial cooking value for each HAP. Similarly, for Alaska, the 81 gridded receptors were averaged for the 9 km gridded sources, and so on for the 3 km cells in Hawaii and Puerto Rico/Virgin Islands. For the 4 km gridded sources in the CONUS domain, the receptors were averaged within 4 km mini-cells within the parent 12 km cell. So, for a particular 12 km grid cell, there would be nine 4 km cell averages.

These gridded averages were then assigned to the block receptors and monitors during the hybrid calculations.

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 subsections below contain discussions on the hybrid air modeling approach used for AirToxScreen for the CONUS.

3.5.1. Overview

For 58 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 2023d) 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 horizontal CMAQ grid cell in the four domains (see Figure 3-1 and Figure 3-5 for an example in the CONUS domain), resulting in at least 144, and sometimes more than 10,000, receptors per cell and 6.5 million receptors nationwide in the CONUS domain.

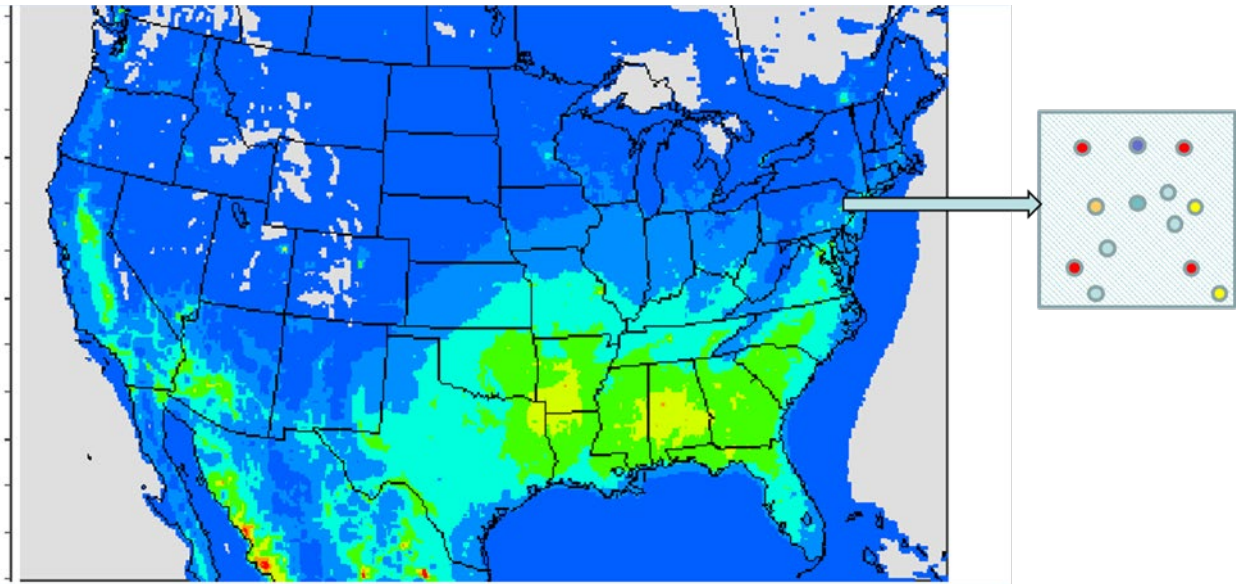


Figure 3-8. CMAQ domain with expanded cell showing hybrid receptors; colors indicate modeled concentrations; dots in inset show locations of receptors within a grid cell

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

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

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 |
| $CMAQ_{PFIRE}$ | = | 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, MI (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 58 of the highest risk air toxics (shown in Table 3-1) among the 181 air toxics included in AirToxScreen. 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 144 subgrid-cell averages in each CMAQ cell in the CONUS domain, 81 subgrid cell averages in Alaska, or 9 averages in the other non-CONUS domains.
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.
 8. For each census block, average across all the receptors for the block for each source group and total concentration.

In the non-CONUS domains, the hybrid calculations shown above were not performed. Instead, the hybrid concentration, C , is the total AERMOD concentration at each census block receptor and the secondary, biogenics, fires, and appropriate background from CMAQ were added to the AERMOD results.

For receptors outside the CMAQ domain in Alaska, the CMAQ domain average of fire, biogenic, secondary, and carbon tetrachloride background were used.

For non-CMAQ HAPs in all areas, relevant background was added to the total AERMOD concentrations for methyl bromide ($0.027 \mu\text{g}/\text{m}^3$) and methyl chloroform ($0.006 \mu\text{g}/\text{m}^3$). Background for the two HAPs were based on 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). The hybrid program then output hybrid concentrations for each run group, including fires, biogenics, secondary and total concentrations for each HAP. Separate files were created for gridded receptors, block receptors and monitors by CMAQ domain. The program output the AERMOD results for the non-CMAQ HAPs in a similar format.

For AirToxScreen, the hybrid ambient concentrations at the block level were used to estimate exposures.

An operational model performance evaluation of the HAPs simulated for the 2020 AirToxScreen was conducted using the Ambient Monitoring Archive for HAPs for the year 2020 (<https://www.epa.gov/amtic/amtic-air-toxics-data-ambient-monitoring-archive>); more details can be found in the [Supplemental Data](#) folder on the AirToxScreen website). 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 and
3. Box and whisker plots which show the model-to-monitor ratios.

3.5.3. Ambient Monitoring Data Preparation

EPA has created annual average concentrations of the ambient monitoring data for the year 2020 using data in the [Ambient Monitoring Archive for HAPs](#). These data primarily come from AQS; however, they also come from non-AQS sources (e.g., air agencies that do not submit some of their data to AQS, other federal agencies that collect relevant data, special studies, etc.). All annual averages are in units of micrograms per cubic meter ($\mu\text{g}/\text{m}^3$) using local meteorological data when available or standard conditions otherwise. An annual average is created for each unique pollutant/monitoring site/sampling duration using the following procedures:

1. There must be at least 3 valid quarters of monitoring data for the year. A quarter is considered valid if it contains at least 7 daily averages.
2. Hourly monitoring data are averaged to daily (and sub-hourly data are averaged to hourly) using the following criteria translating to the ceiling of 75% completeness:

Sampling Duration	Averaging To	Minimum Count
5 MINUTES	HOURLY	9
10 MINUTES	HOURLY	5
15 MINUTES	HOURLY	3
30 MINUTES	HOURLY	2

150 MINUTES	DAILY	8
90 MINUTES	DAILY	12
1 HOUR	DAILY	18
2 HOUR	DAILY	9
3 HOURS	DAILY	6
4 HOUR	DAILY	5
5 HOUR	DAILY	4
6 HOUR	DAILY	3
8 HOUR	DAILY	3
12 HOUR	DAILY	2
24 HOURS	DAILY	1

3. The median Regression on Order Statistic (ROS) is the annual average used, where the percentage of data below the method detection limit (MDL) cannot exceed 80%.
4. Some pollutants were summed to better reflect the AirToxScreen modeled pollutant. This occurred for PAH groups (individual PAHs belonging to the same PAH groups were summed), xylenes (m/p- with o-xylene were summed), and 1,3-dichloropropylene (cis- and trans- were summed).

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

3.5.4. Model Performance Statistics

The Atmospheric Model Evaluation Tool (AMET) was used to conduct AirToxScreen 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 ($\mu\text{g}/\text{m}^3$)

$$\text{Mean Bias} = \frac{1}{n} \sum_1^n (M - O)$$

$$\text{Mean Error} = \frac{1}{n} \sum_1^n |M - O|$$

$$\text{Root Mean Square Error} = \sqrt{\frac{\sum_1^n (M - O)^2}{n}}$$

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)

$$\text{Normalized Mean Bias} = \frac{\sum_1^n (M - O)}{\sum_1^n (O)}$$

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.

$$\text{Normalized Mean Error} = \frac{\sum_1^n |M - O|}{\sum_1^n (O)}$$

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)

$$\text{Correlation} = \frac{1}{(n-1)} \sum_1^n \left(\left(\frac{O - \bar{O}}{\sigma_o} \right) * \left(\frac{M - \bar{M}}{\sigma_m} \right) \right)$$

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.

3.5.5. 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 2020 HAP observed ambient concentrations. Inclusion of all three model results is intended to demonstrate the merged attributes of the hybrid model used for AirToxScreen. Statistical assessments of each model versus observed pairs were paired in time and space and aggregated on an annual basis. Results from the hybrid evaluation are presented in Appendix E. 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, found in the model evaluations in the [Supplemental Data](#) folder on the AirToxScreen website). Figure 3-8 shows the 2020 HAP monitoring locations.

Table 3-6. List of hybrid HAPs evaluated

Model Air Toxic	Measured Air Toxic	No. of Sites
Acetonitrile	ACET_NITRILE_24_HOURS	31
Acrolein	ACROLEIN_24_HOURS	56
Acrylonitrile	ACRY_NITRILE_24_HOURS	9
Acetaldehyde	ALD2_24_HOURS	73
Arsenic	ARSENIC_PM10_24_HOURS	40
	ARSENIC_PM25_24_HOURS	173
Benzene	BENZENE_1_HOUR	15
	BENZENE_24_HOURS	181
	BENZENE_5_MINUTES	7
	BENZENE_3_HOURS	1
Beryllium	BERYLLIUM_PM10_24_HOURS	32
Ethylene dibromide	BR2_C2_12_24_HOURS	12
1,3-Butadiene	BUTADIENE13_1_HOUR	14
	BUTADIENE13_24_HOURS	108
Benzo-a-pyrene	BaP_PM10_24_HOURS	27
Cadmium	CADMIUM_PM10_24_HOURS	38
	CADMIUM_PM25_24_HOURS	128
Carbon tetrachloride	CARBONTET_24_HOURS	173
	CARBONTET_1_HOUR	3
Carbonyl sulfide	CARBSULFIDE_5_MINUTES	7
Chloroform	CHCL3_24_HOURS	153
Chloroprene	CHLOROPRENE_24_HOURS	6
Vinyl chloride	CL_ETHE_24_HOURS	24
Chlorine	CL2_PM2_5_24_HOURS	279
Ethylene dichloride	CL2_C2_12_24_HOURS	63
Methylene chloride	CL2_ME_24_HOURS	63

Model Air Toxic	Measured Air Toxic	No. of Sites
	CL2_ME_5_MINUTES	7
Trichloroethylene	CL3_ETHE_24_HOURS	66
1,1,2,2-Tetrachloroethane	CL4_ETHANE_24_HOURS	14
Tetrachloroethylene	CL4_ETHE_24_HOURS	137
	CL4_ETHE_5_MINUTES	7
Chromium Compounds (only hexavalent chromium was modeled)	CR_VI_PM10_24_HOURS	7
1,4-Dichlorobenzene(p)	DICL_BENZENE_24_HOURS	70
1,3-Dichloropropene	DICL_PROPENE_24_HOURS	5
Ethyl benzene	ETHYLBENZENE_1_HOUR	15
	ETHYLBENZENE_24_HOURS	175
Ethylene Oxide	ETOX_24_HOURS	33
Formaldehyde	FORM_24_HOURS	73
Hexane	HEXANE_1_HOUR	15
	HEXANE_24_HOURS	107
Lead Compounds	LEAD_PM10_24_HOURS	43
	LEAD_PM25_24_HOURS	277
Manganese Compounds	MANGANESE_PM10_24_HOURS	42
	MANGANESE_PM25_24_HOURS	277
Methyl chloride	METHCHLORIDE_24_HOURS	150
	METHCHLORIDE_5_MINUTES	7
Naphthalene	NAPHTHALENE_24_HOURS	52
Nickel Compounds	NICKEL_PM10_24_HOURS	35
	NICKEL_PM25_24_HOURS	277
Polycyclic Organic Matter	PAH_000E0_24_HOURS	29
Polycyclic Organic Matter	PAH_176E4_24_HOURS	28
Polycyclic Organic Matter	PAH_176E5_24_HOURS	11
Polycyclic Organic Matter	PAH_880E5_24_HOURS	29
Propylene dichloride	PROPYL_DICL_24_HOURS	22
Styrene	STYRENE_1_HOUR	14
	STYRENE_24_HOURS	131
Toluene	TOLUENE_1_HOUR	4
	TOLUENE_24_HOURS	131
Xylene	XYLENE_24_HOURS	128
	XYLENE_1_HOUR	4

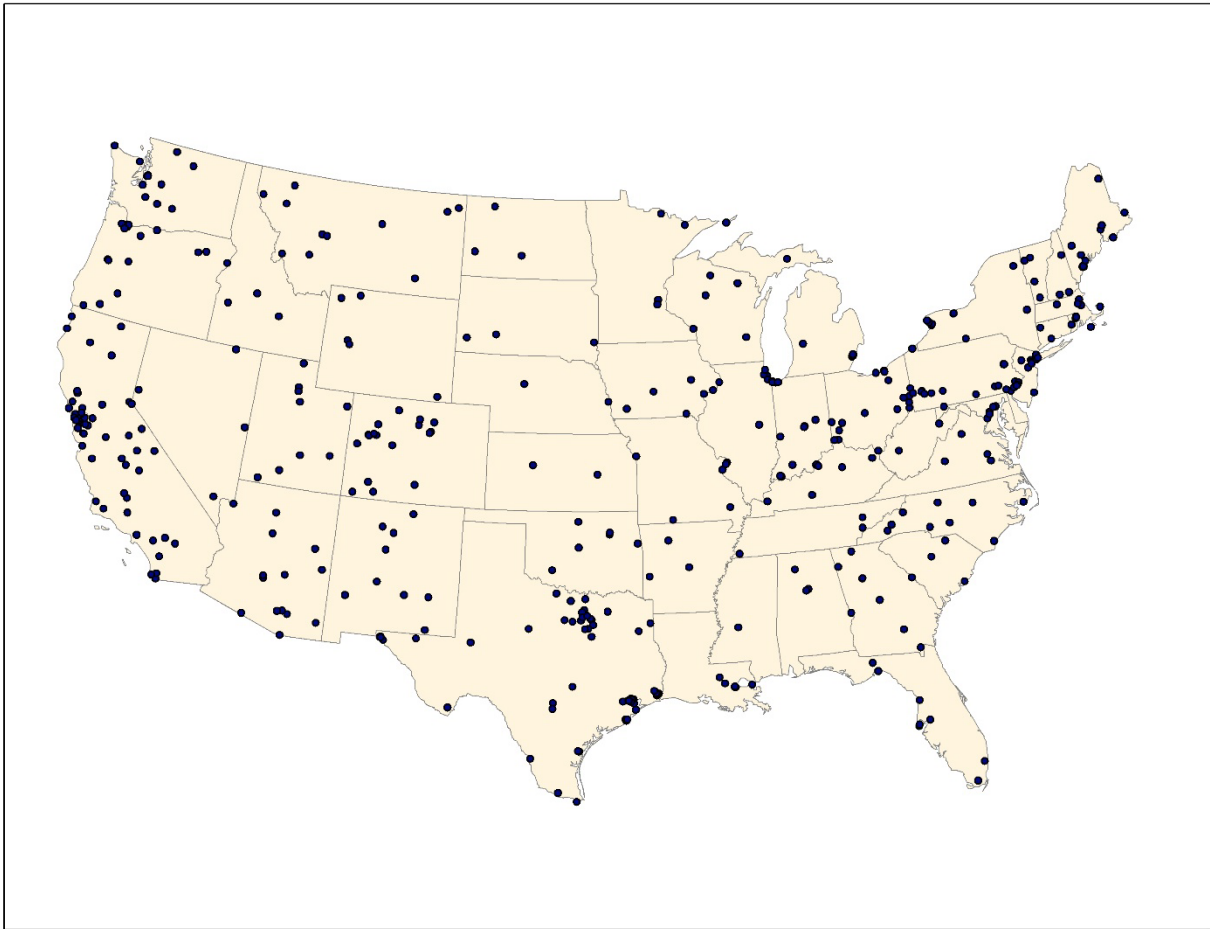


Figure 3-8. 2020 monitoring locations for the hybrid HAPs evaluation

3.5.6. 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 2020 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, found in the [Supplemental Data](#) folder on the AirToxScreen website) used in the annual median. Figure 3-30 shows the 2020 non-hybrid HAP monitoring locations. Results from the non-hybrid evaluation are presented in Appendix E.

Table 3-10. List of non-hybrid HAPs evaluated

Model Air Toxic	Measured Air Toxic	No. of Sites
Antimony	ANTIMONY_PM25_24_HOURS	130
	ANTIMONY_PM10_24_HOURS	26
Benzyl Chloride (alpha-Chlorotoluene)	BENZYLCHLO_24_HOURS	3
Bromoform (Tribromomethane)	BROMOFORM_24_HOURS	34
Carbon disulfide	CARBNDISULF_24_HOURS	37
Chlorobenzene	CHLROBZNE_24_HOURS	52
Cobalt	COBALT_PM25_24_HOURS	127
	COBALT_PM10_24_HOURS	29
Cumene (Isopropylbenzene)	CUMENE_24_HOURS	28
	CUMENE_1_HOUR	14
	CUMENE_3_HOURS	1
Ethyl Chloride (Chloroethane)	ETHYLCHLRD_24_HOURS	47
Ethylidene Dichloride (1,1-Dichloroethane)	ETHIDDICHLD_24_HOURS	20
Hexachloro-1,3-butadiene	HEXCHLRBT_24_HOURS	11
Methyl Bromide (Bromomethane)	METHYLBROM_5_MINUTES	7
	METHYLBROM_24_HOURS	104
Methyl Chloroform (1,1,1-Trichloroethane)	MTHYLCHLRF_5_MINUTES	6
	MTHYLCHLRF_24_HOURS	47
	MTHYLCHLRF_1_HOUR	3
Methyl Iodide (Iodomethane)	MTHYLIODID_24_HOURS	5
Methyl Isobutyl Ketone (4-Methyl-2-pentanone)	MIBK_24_HOURS	59
Methyl Methacrylate	MMETACRYLAT_24_HOURS	13
Methyl tert-butyl ether	MTBE_24_HOURS	11
Propanal (Propionaldehyde)	PROPIONAL_24_HOURS	63
p-Dioxane (1,4-Dioxane)	P_DIOXANE_24_HOURS	6
Selenium	SELENIUM_PM25_24_HOURS	277
	SELENIUM_PM10_24_HOURS	28
	SELENIUM_TSP_24_HOURS	9
1,2,4-Trichlorobenzene	TRICBZ124_24_HOURS	24
2,2,4-Trimethylpentane	TRMEPN224_24_HOURS	68
	TRMEPN224_1_HOUR	15
	TRMEPN224_3_HOURS	1
Vinyl Acetate	VINYLACET_24_HOURS	21
Vinylidene Chloride (1,1-Dichloroethylene)	VINYLDICLOR_24_HOURS	18
1,1,2-Trichloroethane	TRICLA112_24_HOURS	22

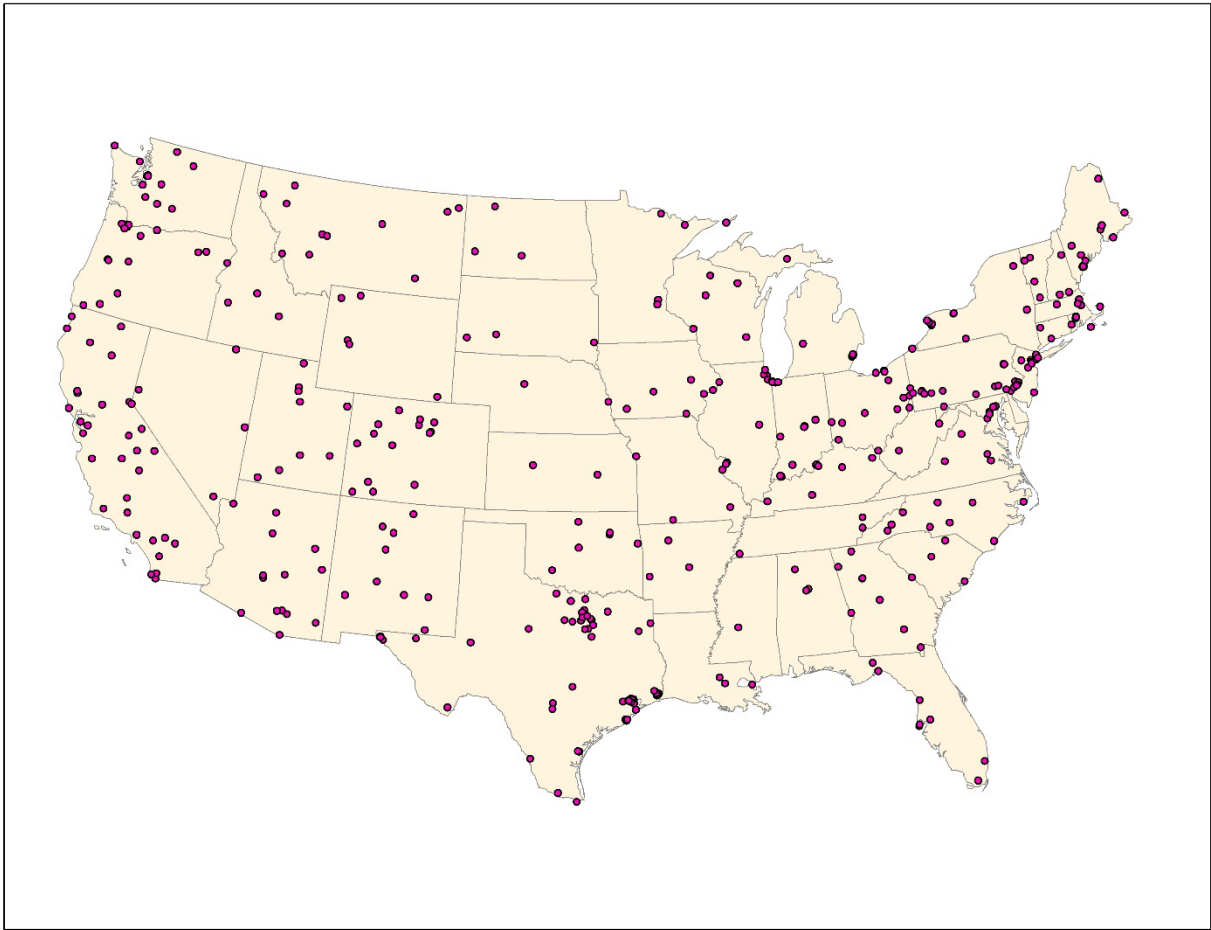


Figure 3-9. 2020 monitoring locations for the non-hybrid HAPs evaluation

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 AirToxScreen. It begins with an overview of the surrogate approach used that included new exposure modeling for some AirToxScreen 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 AirToxScreen can be found in Appendix E.

4.1. Estimating Exposure Concentrations

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

EPA used version 8 of the EPA Hazardous Air Pollutant Exposure Model (HAPEM8) to conduct direct exposure modeling for AirToxScreen. 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 HAPEM8 for a selected group of surrogate pollutants and source categories to estimate census tract level exposure factors. Census tracts were used because most HAPEM inputs are available at the census tract level. For each surrogate pollutant and AirToxScreen 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 block, we multiplied the ambient concentration of the pollutant by the surrogate's exposure factor for the census tract containing the census block, 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, HAPEM8 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 HAPEM8) 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 20124). These changes make HAPEM8 suitable for regional and national inhalation risk assessments such as AirToxScreen. The 1996 and 1999 NATAs estimated Inhalation ECs using HAPEM4 and HAPEM5 respectively; the 2011 and 2014 NATAs as well as 2017-2019 AirToxScreens used HAPEM7 (EPA did not use HAPEM6 for NATA). 2020 AirToxScreen uses HAPEM8. 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 HAPEM8 (EPA 2024c).

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

Characteristic	HAPEM5	HAPEM7	HAPEM8
Data source for population demographics	1990 U.S. Census	2000 U.S. Census	2020 U.S. Census
Characterization of microenvironmental factors	Probability distributions	Same as HAPEM5	Same as HAPEM7
Method for creation of annual average activity patterns from daily activity-pattern data	Sampling a limited number of daily diaries to represent an individual’s range of activities, accounting for autocorrelation	Same as HAPEM5, except now includes commuter-status criterion	Same as HAPEM7
Interpretation of exposure-concentration range for a given cohort/tract combination	Variability of annual ECs across cohort/tract members	Same as HAPEM5, except now includes adjustments based on proximity to roadway	Same as HAPEM7

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 HAPEM8 (EPA 2024c).

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 AirToxScreen 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 AirToxScreen, EPA estimated annual average ambient concentrations for each census tract using a hybrid CMAQ-AERMOD approach discussed in Section 3.5 to calculate block-level averages, and then used populating weighting to estimate tract-level averages. EPA stratified the air quality outputs for a selected group of pollutants by one or more of the four principal AirToxScreen 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 AirToxScreen can be summarized for each principal AirToxScreen 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 AirToxScreen, 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 AirToxScreen). 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 AirToxScreen is EPA's Consolidated Human Activity Database (CHAD; EPA 2024d). To develop the version of CHAD (version April 2020) used in AirToxScreen, 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. 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.

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 AirToxScreen.

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 $\mu\text{g}/\text{m}^3$
$CONC_{(i,t)}$	=	ambient concentration for census tract i for time step t , in units of $\mu\text{g}/\text{m}^3$
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 $\mu\text{g}/\text{m}^3$

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. AirToxScreen uses 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 AirToxScreen because indoor-source data are currently incomplete (recall that AirToxScreen 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 AirToxScreen.

Table 4-2. Microenvironments used in HAPEM modeling for AirToxScreen

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

4.4. Exposure Factors

HAPEM exposure modeling for AirToxScreen requires substantial time and resources for data collection and processing, computing and model processing. Due to these requirements, EPA conducted HAPEM modeling for AirToxScreen only for selected pollutants, which we present below along with how we used them to estimate ECs for the remaining AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 E to this document contains further details on the application of HAPEM8 in the AirToxScreen analysis. A spreadsheet file 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 AirToxScreen website at <https://www.epa.gov/AirToxScreen>.)

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 are not included in AirToxScreen).

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 AirToxScreen 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 use in air toxics screening assessments. 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.

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 exposure factors for each pollutant/source group/census tract for AirToxScreen using the HAPEM8 model, which were then used to estimate exposure concentrations at the census block level.
- These ECs can be used to determine census block-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 AirToxScreen risk assessment process. Specifically, the sections that follow provide an overview of the cancer and noncancer toxicity values used in AirToxScreen and the primary sources of these values. They also describe several adjustments and assumptions to toxicity values specific to the AirToxScreen risk assessment process.

5.1. Toxicity Values and Their Use in AirToxScreen

The toxicity values used in AirToxScreen 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 AirToxScreen is focused on long-term exposures, the toxicity values used in AirToxScreen 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 AirToxScreen 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 AirToxScreen, 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. 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 AirToxScreen are consistent with those EPA's Office of Air Quality Planning and Standards (OAQPS) has compiled for chronic inhalation exposures to air toxics. The full set of toxicity values used for AirToxScreen (and their sources) can be accessed via the AirToxScreen [Supplemental Data](#) file; see Appendix B for details. Sources of chronic dose-response assessments used for AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 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).

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 ($\mu\text{g}/\text{m}^3$) 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.

For an inhalation risk assessment (and for AirToxScreen), 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 toxic 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 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.

**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

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 AirToxScreen 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 AirToxScreen in Section 6 of this document.

5.3. Data Sources for Toxicity Values

Information on dose-response assessments for evaluating chronic exposures for AirToxScreen 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 "AirToxScreen_Pollutants.xlsx" within the [Supplemental](#)

[Data](#) folder accompanying this TSD contains the toxicity values for both cancer and noncancer chronic effects used in AirToxScreen. 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 AirToxScreen. 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 AirToxScreen 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 AirToxScreen, 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 AirToxScreen, 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. AirToxScreen 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 AirToxScreen 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 AirToxScreen.

5.4.1. Polycyclic Organic Matter

A substantial proportion of polycyclic organic matter (POM) reported in the NEI was not speciated into individual compounds. For example, some emissions of POM were reported in NEI as “PAH, total” or “PAH/POM-Unspecified.” In other cases, individual POM compounds were reported for which no quantitative cancer dose-response value has been published in the sources used for AirToxScreen. 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.

For AirToxScreen, we divided unspciated POM emissions into ten POM groups. We discuss the POM groups used for the AirToxScreen 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 AirToxScreen. 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 AirToxScreen.

5.4.4. Metals

Several decisions made for AirToxScreen 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 do not use the lead RfC to calculate noncancer adverse effects. Instead, we estimate ambient concentrations which can be compared to the primary National Ambient Air Quality Standard (NAAQS) for lead. The primary NAAQS for lead 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).

Nickel compounds. The cancer inhalation URE for most of the emissions of nickel compounds included in AirToxScreen (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 AirToxScreen 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 AirToxScreen. 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 AirToxScreen.

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 AirToxScreen, 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 2011).

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 AirToxScreen 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 $\mu\text{g DPM}/\text{m}^3$) 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 $5 \mu\text{g}/\text{m}^3$ 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 m³ 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 AirToxScreen

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

6.1. The Risk-characterization Questions AirToxScreen Addresses

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

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:

$$Risk_{tot} = Risk_1 + Risk_2 + \dots + Risk_i$$

Where:

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

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

In AirToxScreen, the estimated ECs are not considered upper bounds. Rather, they represent central-tendency estimates of ECs at the census block 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 AirToxScreen, 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 AirToxScreen.

Information on non-additive interactions is not readily available in a form that can be used for AirToxScreen. 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 AirToxScreen. 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 for 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 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 2016b). EPA has also published a [framework for using studies of these and other effects in inhalation risk assessment](#) (EPA 1994).

EPA’s Chronic Noncancer Guidelines

- [Mutagenicity](#) (EPA 1986)
- [Developmental Toxicity](#) (EPA 1991)
- [Neurotoxicity](#) (EPA 1998)
- [Reproductive Toxicity](#) (EPA 1996)

6.3.1. Individual Pollutant Hazard

EPA estimated chronic noncancer hazards for AirToxScreen 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/m³

RfC = the corresponding reference concentration for that air toxic, in units of mg/m³

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 *i*th 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 AirToxScreen, 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 AirToxScreen at Block, 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 AirToxScreen 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. Aggregation of Block-level Results to Larger Spatial Units

Block-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 block-level concentrations by the population of each block, summing these population-weighted concentrations, and dividing by the total county population encompassing all blocks to obtain a final population-weighted, county-level concentration. The process for aggregating from the block to the county level can be expressed using the following equation:

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

Where:

- $Conc_{county_k}$ = population-weighted concentration for county k
- $Conc_{block_i}$ = ambient concentration in block i (contained within county k)
- Pop_{block_i} = population in block i (contained within county k)
- Pop_{county_k} = 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. AirToxScreen results include ambient concentrations, ECs, cancer risks, and noncancer HQs at the block, 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 AirToxScreen 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.

6.5. The Risk Characterization Results That AirToxScreen Reports

AirToxScreen 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 2020. 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.

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. AirToxScreen 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, AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 AirToxScreen 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.

AirToxScreen results indicate that most individuals' estimated risk is between 1-in-1 million and 100-in-1 million, although a small number of localized areas show risks of higher than 100-in-1 million. Although individuals and communities may be concerned about these results, recall that AirToxScreen 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 AirToxScreen 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 (either set of criteria can be used)	10	1
	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).

^bHazard 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 AirToxScreen 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 AirToxScreen are also provided at the census block, 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 AirToxScreen was developed. AirToxScreen 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 AirToxScreen 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 AirToxScreen results via the following website: <https://www.epa.gov/AirToxScreen>.

6.6. Summary

- The purpose of AirToxScreen 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 AirToxScreen.

-
- 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.
 - AirToxScreen estimates most individuals' cancer risk to be between 1-in-1 million and 100-in-1 million, although a small number of localized areas show risk higher than 100-in-1 million.
 - Air toxics data for AirToxScreen are presented at the national, state, county and census block 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 AirToxScreen

7.1. Introduction

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

As stated in Section 1, AirToxScreen results should not be used for limited-scale or site-focused applications. AirToxScreen 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 AirToxScreen 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 AirToxScreen. 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 AirToxScreen possible given available information and resources. That said, assumptions introduce uncertainties into AirToxScreen results because we can never be fully confident 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 AirToxScreen Addresses Variability

The AirToxScreen 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 AirToxScreen addresses. A brief explanation is also provided on how AirToxScreen results should be interpreted given variability.

7.2.1. Components of Variability

AirToxScreen 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 block in that region. AirToxScreen 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 block. 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 AirToxScreen 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 block. The ambient concentration estimated for the block is only an approximation of conditions at all locations in the block. Different locations within that block 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 AirToxScreen 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 based on the emission sources and meteorological conditions affecting those specific blocks. Some temporal variation is accounted for in AirToxScreen 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 AirToxScreen results reported by EPA.

The AirToxScreen concentrations and risks, however, do reflect a degree of geographic variation. The smallest geographic area for which AirToxScreen results are reported is the census block. Although results are reported at the census block level, average risk estimates are far more uncertain at this level of spatial resolution than at the county or state level. Census blocks are 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 AirToxScreen 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.

Assessment results do not reflect variations in the susceptibility of people within a census block because the focus is to compare typical exposures and risks in different blocks. As a result, individual exposures or risks might differ in either direction. You should consider exposures or risks determined in AirToxScreen as representative of the geographic area where an individual lives, but not as that individual's personal risk.

Thus, the results of the AirToxScreen 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 blocks. 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 block have the same exposure and risk. This assumption allows AirToxScreen users to examine variation in individual exposure among census blocks, but not the variation within a census block. 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 AirToxScreen. EPA took this approach for AirToxScreen for two primary reasons:

- **An overall purpose of AirToxScreen is to examine broad differences driven by geography.** AirToxScreen 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 blocks, producing differences in the typical pollutant concentrations, exposures and risks in different blocks. Differences in susceptibility, however, can produce differences in risk between two individuals in the same census block, and reporting on these differences is not a purpose of AirToxScreen.
- **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 block, 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 blocks. 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 AirToxScreen Results

The AirToxScreen 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 blocks. To a lesser degree, variation among demographic groups is also addressed by AirToxScreen 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 AirToxScreen display predictions of cancer risk and noncancer hazard. Cancer risk results include statements such as:

“X percent of the census blocks 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 blocks 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 blocks would be expected to have a risk above 1-in-1 million. Although a person may live in a census block 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 block, 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 block. 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 AirToxScreen analysis is that they show variation among values of ambient air concentration, exposure or risk in census blocks 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 blocks. 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 AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 AirToxScreen 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 AirToxScreen, 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 AirToxScreen 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 AirToxScreen analysis uses a series of mathematical models. EPA uses models in AirToxScreen 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, AirToxScreen 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 AirToxScreen 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 AirToxScreen

For this discussion, we have divided the uncertainties in AirToxScreen 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 within a census block, 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.

AirToxScreen Components that Include Uncertainty
<ul style="list-style-type: none">• Ambient concentrations• Exposure estimates• Risk estimates

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 AirToxScreen review. We have not fully assessed the quality of specific emission rates and locations in the NEI and resultant AirToxScreen emissions (e.g., industrial emissions from a specific census block), 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 AirToxScreen do not in all cases. AERMOD estimates for point sources consider topography, but the estimates for the emissions sources modeled as gridded 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 AirToxScreen'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 the model evaluations found in the [Supplemental Data](#) folder on the AirToxScreen website.

Measured concentrations were taken from EPA's [Ambient Monitoring Archive for HAPs](#). These data primarily come from AQS; however, they also come from non-AQS sources (e.g., air agencies that do not submit some of their data to AQS, other federal agencies that collect relevant data, special studies, etc.). For AirToxScreen, 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 AirToxScreen.

The commuting data used in HAPEM are based on an EPA analysis of information from a special study by the U.S. Census Bureau (2020). HAPEM uses this information, reflecting 2020 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. AirToxScreen'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.

AirToxScreen'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 AirToxScreen

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 AirToxScreen, 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 AirToxScreen 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 AirToxScreen 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 block, 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 block 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 block.

The results for the current AirToxScreen are restricted to 2020 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 2020.

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 AirToxScreen 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 AirToxScreen. An IRIS RfC for diesel PM allows AirToxScreen 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 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.

AirToxScreen 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.

AirToxScreen may not accurately capture sources with episodic emissions except for those with continuous emissions monitoring (CEMS) data, which use hourly emissions. AirToxScreen 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. AirToxScreen 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 <https://www.atsdr.cdc.gov/mrls/index.html>. 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 D.C., Struve M.F., Wong B.A., Marshall M.W., Gross E.A., and Willson G.A., 2008. Respiratory tract responses in male rats following subchronic acrolein inhalation. *Inhal Toxicol* 20(3): 205-16.
- Doumbia, T., Granier, C., Elguindi, N., Bouarar, I., Darras, S., Brasseur, G., Gaubert, B., Liu, Y., Shi, X., Stavrou, T., Tilmes, S., Lacey, F., Deroubaix, A., and Wang, T., 2021. Changes in global air

-
- pollutant emissions during the COVID-19 pandemic: a dataset for atmospheric modeling, *Earth Syst. Sci. Data*, 13, 4191–4206, <https://doi.org/10.5194/essd-13-4191-2021>.
- East, J.D., Henderson, B. H., Napelenok, S. L., Koplitz, S. N., Sarwar, G., Gilliam, R., Lenzen, A., Tong, D., Pierce, R. B., Garcia-Menendez, F., 2022. Inferring and evaluating satellite-based constraints on NO_x emissions estimates in air quality simulations. *Atmospheric Chemistry and Physics*, Vol 22, issue 24, 15981-16001, <https://doi.org/10.5194/acp-22-15981-2022>.
- EPA (U.S. Environmental Protection Agency). 1986. Guidelines for Mutagenicity Risk Assessment. EPA/630/R-98/003. EPA, Washington, DC. Available online at <https://www.epa.gov/sites/default/files/2013-09/documents/mutagen2.pdf>. Last accessed 10 December 2015.
- EPA. 1991. Guidelines for Developmental Toxicity Risk Assessment. EPA/600/R-91/001. EPA, Washington, DC. Available online at https://www.epa.gov/sites/default/files/2014-11/documents/dev_tox.pdf. 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 https://www.epa.gov/sites/default/files/2014-11/documents/rfc_methodology.pdf. Last accessed 27 October 2015.
- EPA. 1996. Guidelines for Reproductive Toxicity Risk Assessment. EPA/630/R-96/009. EPA, Washington, DC. Available online at https://www.epa.gov/sites/default/files/2014-11/documents/guidelines_repro_toxicity.pdf. 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 https://www.epa.gov/sites/default/files/2014-11/documents/neuro_tox.pdf. 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 https://www.epa.gov/sites/default/files/2014-11/documents/frmwrk_cum_risk_assmnt.pdf. 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 https://www.epa.gov/sites/default/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 https://www.epa.gov/sites/default/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 https://www.epa.gov/sites/default/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf. 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 https://www.epa.gov/sites/default/files/2013-09/documents/childrens_supplement_final.pdf. 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://www.epa.gov/fera/hazardous-air-pollutant-exposure-model-hapem-users-guides>. 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

https://www.epa.gov/sites/default/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 <https://www.epa.gov/fera/hazardous-air-pollutant-exposure-model-hapem-users-guides>. 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 <https://www.epa.gov/sites/default/files/2021-03/documents/day3techregioommenmodeltomonitorcomparison.pdf>. Last accessed 30 November 2015.

EPA. 2011. Toxicological Review of Trichloroethylene. EPA NCEA, Washington, DC. Available online at <https://cfpub.epa.gov/ncea/iris/search/index.cfm?keyword=trichloroethylene>. Last updated 11 September 2011.

EPA. 2016a. Overview by Section of CAA. EPA OAR, Washington, DC. Available online at <https://www3.epa.gov/airtoxics/overview.html>. Last updated 23 February 2016. Last accessed 23 May 2024.

EPA. 2016b. 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 <https://www.epa.gov/haps/initial-list-hazardous-air-pollutants-modifications>. Last updated 16 March 2017. Last accessed 18 June 2018.
- EPA. 2020. AERSURFACE. Available online at: <https://www.epa.gov/scram/air-quality-dispersion-modeling-related-model-support-programs#aersurface>. Last updated 29 February 2020. Last accessed 23 May 2024.
- EPA. 2022. Technical Support Document: EPA’s Air Toxics Screening Assessment: 2018 AirToxScreen TSD. EPA-454/B-22-002. August 2022. EPA, Research Triangle Park, NC.
- EPA. 2023a. Technical Support Document (TSD): Preparation of Emissions Inventories for the 2020 North American Emissions Modeling Platform. EPA-454/B-23-004. December 2023. Available at: https://www.epa.gov/system/files/documents/2023-12/2020_emismod_tsd_dec2023_4.pdf.
- EPA, 2023b. *Guideline on Air Quality Models*. 40 CFR Part 51 Appendix W.
- EPA. 2023c. Guidance on the Use of the Mesoscale Model Interface Program (MMIF) for AERMOD applications. EPA-454/B-24-006. October 2023. EPA, Research Triangle Park, NC.
- EPA. 2023d. User’s Guide for the AMS/EPA Regulatory Model – AERMOD. EPA-454/B-23-008. October 2023. EPA, Research Triangle Park, NC.
- EPA. 2024a. Criteria Air Pollutants. EPA, Washington, DC. Available online at <https://www.epa.gov/criteria-air-pollutants>. Last updated 2 May 2024. Last accessed 17 May 2024.
- EPA. 2024b. Risk Assessment Guidance and Tools. EPA, Washington, DC. Available online at <https://www.epa.gov/risk/risk-assessment-guidance>. Last updated 19 January 2024. Last accessed 17 May 2024.
- EPA. 2024c. The HAPEM User’s Guide Hazardous Air Pollutant Exposure Model, Version 8. EPA OAQPS, Research Triangle Park, NC. Available online at <http://www.epa.gov/fera/hazardous-air-pollutant-exposure-model-hapem-users-guides>. Last accessed 17 May 2024.
- EPA. 2024d. Consolidated Human Activity Database (CHAD). EPA, Washington, DC. Available online at <https://www.epa.gov/healthresearch/consolidated-human-activity-database-chad-use-human-exposure-and-health-studies-and>. Last updated 17 September 2023. Last accessed 17 May 2024.
- EPA. 2024a. The HAPEM User’s Guide Hazardous Air Pollutant Exposure Model, Version 8. EPA OAQPS, Research Triangle Park, NC. Available online at <http://www.epa.gov/fera/hazardous-air-pollutant-exposure-model-hapem-users-guides>. Last accessed 17 May 2024.
- EPA. 2024b. Consolidated Human Activity Database (CHAD). EPA, Washington, DC. Available online at <https://www.epa.gov/healthresearch/consolidated-human-activity-database-chad-use-human-exposure-and-health-studies-and>. Last updated 17 September 2023. Last accessed 17 May 2024.
- 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.
- Friedman, C.L., Y. Zhang, and N.E. Selin, 2014. Climate change and emissions impacts on atmospheric PAH transport to the Arctic, *Environ. Sci. Technol.*, 48 (1), 429-437, doi:10.1021/es403098w.
- 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.
- Gilliam, R.C., Pleim, J.E., 2010. Performance Assessment of New Land Surface and Planetary Boundary Layer Physics in the WRF-ARW. *Journal of Applied Meteorology and Climatology* 49, 760-774.
- Guenther, A.B., Jiang, X., Heald, C.L., Sakulyanontvittaya, T., Duhl, T., Emmons, L.K., and Wang, X, 2012. The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *GMD*, Volume 5, Issue 6, 1471-1492.
- Global Modeling and Assimilation Office (GMAO), 2015. Inst3_3d_asm_Cp; MERRA-2 IAU State Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4, Greenbelt, MD, USA: Goddard Space Flight Center (GSFC DAAC). Doi: 10.5067/VJAFPL1CSIV.
- 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 <https://www.healtheffects.org/publication/diesel-emissions-and-lung-cancer-evaluation-recent-epidemiological-evidence-quantitative>. 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.
- Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q., 2018: Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), *Geosci. Model Dev.*, 11, 369-408, doi: 10.5194/gmd-11-369-2018.
- 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.
- Janssens-Maenhout, G., Dentener, F., Van Aardenne, J., Monni, S., Pagliari, V., Orlandini, L., Klimont, Z., Kurokawa, J., Akimoto, H., Ohara, T., others, 2012. EDGAR-HTAP: a harmonized gridded air pollution emission dataset based on national inventories. European Commission Publications Office, Ispra (Italy). JRC68434, EUR report No EUR 25, 299–2012.
- Kang, D.; Willison, J.; Sarwar, G.; Madden, M.; Hogrefe, C.; Mathur, R.; Gantt, B. and Saiz-Lopez, A., 2021 (October). Improving the Characterization of Natural Emissions in CMAQ, Environmental Manager, A&WMA.
- Keller, C.A., M.S. Long, R.M. Yantosca, A.M. Da Silva, S. Pawson, and D.J. Jacob, 2014. HEMCO v1.0: A versatile, ESMF-compliant component for calculating emissions in atmospheric models, *Geosci. Model Devel.*, 7, 1409-1417, doi:10.5194/gmd-7-1409-2014.
- Luecken, D. J., Yarwood, G., and Hutzell, W. T, 2019. Multipollutant modeling of ozone, reactive nitrogen and HAPs across the continental US with CMAQ-CB6, *Atmos Environ*, 201, 62-72, 10.1016/j.atmosenv.2018.11.060.
- Ma, L-M. and Tan, Z-M, 2009. Improving the behavior of the Cumulus Parameterization for Tropical Cyclone Prediction: Convection Trigger. *Atmospheric Research* 92 Issue 2, 190-211. <http://www.sciencedirect.com/science/article/pii/S0169809508002585>.
- Mathur, R., Xing, J., Gilliam, R., Sarwar, G., Hogrefe, C., Pleim, J., Pouliot, G., Roselle, S., Spero, T. L., Wong, D. C., and Young, J, 2017. Extending the Community Multiscale Air Quality (CMAQ) modeling system to hemispheric scales: overview of process considerations and initial applications, *Atmos Chem Phys*, 17, 12449-12474, 10.5194/acp-17-12449-2017.
- National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. 2015, updated daily. NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. <https://doi.org/10.5065/D65D8PWK>.
- 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 https://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/outlinks.centers/center/136. 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/cnr/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.
- Philip, S., Martin, R.V., Snider, G., Weagle, C.L., van Donkelaar, A., Brauer, M., Henze, D.K., Klimont, Z., Venkataraman, C., Guttikunda, S.K., and Zhang, Q., April 2017. “Anthropogenic fugitive, combustion and industrial dust is a significant, underrepresented fine particulate matter source in global atmospheric models.” *Environmental Research Letters*; Bristol, Vol. 12, Iss. 4. Doi:10.1088/1748-9326/aa65a4.
- Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., Liu, Z. Q., Snyder, C., Chen, F., Barlage, M. J., Yu, W., and Duda, M. G., 2017. THE WEATHER RESEARCH AND FORECASTING MODEL Overview, System Efforts, and Future Directions, *B Am Meteorol Soc*, 98, 1717-1737, 10.1175/Bams-D-15-00308.1.
- 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.
- Selin, N.E., D.J. Jacob, R.J. Park, R.M. Yantosca, S. Strode, L. Jaeglé, and D. Jaffe, 2007. Chemical cycling and deposition of atmospheric mercury: Global constraints from observations, *J. Geophys. Res.*, 112, D02308, doi:10.1029/2006JD007450.
- 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.
- Simone, N.W., Stettler, M.E.J., Barrett, S.R.H., 2013. Rapid estimation of global civil aviation emissions with uncertainty quantification, *Transportation Research Part D: Transport and Environment*, Volume 25, 33-41, ISSN 1361-9209, <https://doi.org/10.1016/j.trd.2013.07.001>.
- Simpson, D., 2018. Soil N emissions for 2000-present. (D81.3.6.1.) [dataset].
- 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

<https://opensky.ucar.edu/islandora/object/technotes:500>. Last updated 5 December 2014. Last accessed 16 December 2015.

Stammer, D., F.J. Wentz, and C.L. Gentemann, 2003, Validation of Microwave Sea Surface Temperature Measurements for Climate Purposes, *J. Climate*, 16, 73-87.

U.S. Census Bureau. 2020. Decennial Census of Population and Housing. Available online at <https://www.census.gov/programs-surveys/decennial-census/decade/2020/2020-census-main.html>. Last accessed 23 May 2024.

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.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J., 2011. The Fire INventory from NCAR (FINN): a high-resolution global model to estimate the emissions from open burning, *Geosci. Model Dev.*, 4, 625-641, 10.5194/gmd-4-625-2011.

Xu, L., Pye, H. O. T., He, J., Chen, Y. L., Murphy, B. N., and Ng, N. L., 2018. Experimental and model estimates of the contributions from biogenic monoterpenes and sesquiterpenes to secondary organic aerosol in the southeastern United States, *Atmos Chem Phys*, 18, 12613-12637, 10.5194/acp-18-12613-2018.

Yarwood, G., J. Jung, G. Whitten, G. Heo, J. Mellberg and M. Estes. 2010: Updates to the Carbon Bond Chemical Mechanism for 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 may contract cancer if exposed continuously (24 hours per day) to the specific concentration over 70 years (an assumed lifetime). For example, a 1-in-1 million risk means that one person may develop cancer for every 1 million people exposed. A 10-in-1 million risk means 10 people may develop cancer per 1 million exposed, etc. These risks would be in addition to cancer cases that would normally occur in one million unexposed people. Note that AirToxScreen 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 AirToxScreen 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.

Air Toxics Screening Assessment (AirToxScreen):

EPA's ongoing thorough review of air toxics in the United States. AirToxScreen results help scientists focus on pollutants, emission sources and places that may need further study to better understand risks. AirToxScreen 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.

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 AirToxScreen.

*Diesel particulate matter is not a hazardous air pollutant but is included in the AirToxScreen air toxics.

Ambient:

Surrounding, as in the surrounding environment. In AirToxScreen, ambient air refers to the outdoor air surrounding a person through which pollutants can be carried. Therefore, the ambient concentrations estimated by AirToxScreen are concentrations estimated in the outdoor environment. AirToxScreen 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 AirToxScreen) 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 AirToxScreen, 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 AirToxScreen, we add background concentrations to AERMOD concentrations but not to CMAQ concentrations, which include background already. Most risk from AirToxScreen 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 AirToxScreen. 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 AirToxScreen 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 AirToxScreen 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 blocks:

Statistical areas bounded by visible features such as roads, streams, and railroad tracks. Generally small in area. In a city, a census block looks like a city block bounded on all sides by streets. Census blocks in suburban and rural areas may be large, irregular, and bounded by a variety of features, such as roads, streams, and transmission lines. In remote areas, census blocks may encompass hundreds of square miles. Census blocks are not delineated based on population, and many census blocks do not have any population..

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 AirToxScreen, 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 AirToxScreen. 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 AirToxScreen, 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 AirToxScreen, a single year).

Exposure assessment:

An exposure assessment is part of an air toxics risk assessment such as AirToxScreen. 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 AirToxScreen).

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. AirToxScreen 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 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 AirToxScreen, 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, AirToxScreen 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. For a 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 ($\mu\text{g}/\text{m}^3$) in air. You can interpret the URE as follows: If the URE = 3×10^{-6} per $\mu\text{g}/\text{m}^3$, 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 AirToxScreen, 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 AirToxScreen 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 AirToxScreen

A master pollutant list for AirToxScreen in spreadsheet format, “2018AirToxScreen_Pollutants.xlsx,” is provided in the [Supplemental Data](#) folder on the AirToxScreen website (<https://www.epa.gov/AirToxScreen>). This file includes all air toxics modeled in AirToxScreen and indicates the inventory types(s) reporting them. The file also includes the toxicity values used in AirToxScreen. 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 AirToxScreen:

- the names used in the National Emissions Inventory (NEI); and
- the names used for AirToxScreen.

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 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-1 contains the air toxics that were *not* modeled for AirToxScreen and why. Note that although diesel PM was modeled for AirToxScreen 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 PM₁₀ emissions from on-road and nonroad mobile sources burning diesel or residual fuels.

Note that NEI = National Emissions Inventory.

Table B-1. Pollutants excluded from AirToxScreen

Pollutant	NEI Pollutant Code (CAS Number) a	Reason for Exclusion	In Previous Assessments?
Chromium III		– In building the NEI, chromium VI is emphasized, and chromium III may be missing data.	n
2,3,7,8-Tetrachlorodibenzo-p-dioxin	1746016	Dioxins and furans are not in the 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 AirToxScreen’s inhalation risk estimates likely would not be large.	n
Other dioxins/furans	multiple		n
Radionuclides		Radionuclides are not in the 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
DDE	72559 incorrectly referred to in the Section 112(b) list as 3547-04-4	This pollutant was not reported to the 2018 NEI.	n
Fine mineral fibers (including rockwool, slagwool and fine mineral fibers)	Fine mineral fibers: 383 Rockwool: 617 Slagwool: 616	Rockwool has zero emissions in the 2017 NEI. Slagwool and fine mineral fibers are excluded from previous assessments.	n
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
Diazomethane	334883	This pollutant has 0 emissions in the 2017 NEI.	n
Beta-propiolactone	57578	This pollutant has 0 emissions in the 2017 NEI.	y

^aIn 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.

Appendix C. Estimating Background Concentrations for AirToxScreen

This appendix contains the methods we used to estimate background concentrations for AirToxScreen.

Data

Overview

Boundary conditions are calculated for the following 14 pollutants: chloroform, methyl chloride (chloromethane), benzene, carbon tetrachloride, methyl bromide (bromomethane), methyl chloroform (1,1,1-trichloroethane), dichloromethane (methylene chloride), tetrachloroethene (perchloroethylene, tetrachloroethylene), chromium VI, arsenic, chromium, lead, manganese, and nickel.

Air Quality Data

Air quality data from remote locations come from three sources. The Advanced Global Atmospheric Gases Experiment (AGAGE) is a global network supported by several international institutions and organizations that have monitored trace gases since 1978 (<https://agage.mit.edu/>). The National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Laboratory (GML) “conducts research that addresses three major challenges: greenhouse gas and carbon cycle feedbacks, changes in clouds, aerosols, and surface radiation, and recovery of stratospheric ozone” (<https://gml.noaa.gov/>). The Halocarbons And other Trace Species (HATS) is a program within GML that quantifies “the distributions and magnitudes of the sources and sinks for important ozone-depleting and greenhouse gases” (<https://gml.noaa.gov/>). Lastly, The Interagency Monitoring of Protected Visual Environments (IMPROVE) network is a monitoring network cooperative represented by several federal agencies and state organizations, along with some international representation that focuses on visibility in Class I federal areas (<http://vista.cira.colostate.edu/Improve/improve-program/>). Below are the data sources and sites used for each pollutant.

Pollutant	Data Source	Sites
Chloroform	AGAGE	THD
Methyl chloride (Chloromethane)	NOAA	KUM, MLO, NWR, BRW, ALT
Benzene	NOAA and IMPROVE	KUM, MLO, NWR, BRW, ALT
Carbon tetrachloride	NOAA	MLO, NWR, BRW, SMO, SPO
Methyl bromide (Bromomethane)	NOAA	KUM, MLO, NWR, BRW, ALT
Methyl chloroform (1,1,1-Trichloroethane)	NOAA	ALT, SUM, BRW, MHD, THD, NWR, KUM, MLO, SMO, CGO, PSA, SPO
Dichloromethane (Methylene chloride)	NOAA	KUM, MLO, NWR, BRW, ALT
Tetrachloroethene (Perchloroethylene, Tetrachloroethylene)	NOAA	KUM, MLO, NWR, BRW, ALT
Chromium VI (PM2.5)	IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Arsenic (PM2.5)	IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Chromium (PM2.5)	IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Lead (PM2.5)	IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR

Manganese (PM2.5)	IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR
Nickel (PM2.5)	IMPROVE	DENA, KALM, PORE, REDW, TRCR, TUXE, HACR

Below is location information for each AGAGE and NOAA site.

Code	Name	Country	Lat.	Lon.	Elev. (m)
ALT	Alert, Nunavut	Canada	82.451	-62.507	185
BRW	Barrow Atmospheric Baseline Observatory	USA	71.323	-156.611	11
CGO	Cape Grim, Tasmania	Australia	-40.683	144.69	94
KUM	Cape Kumukahi, Hawaii	USA	19.561	-154.888	8
MHD	Mace Head, County Galway	Ireland	53.326	-9.899	5
MLO	Mauna Loa, Hawaii	USA	19.536	-155.576	3397
NWR	Niwot Ridge, Colorado	USA	40.053	-105.586	3523
PSA	Palmer Station, Antarctica	USA	-64.774	-64.053	10
SMO	Tutuila	USA	-14.247	-170.564	42
SPO	South Pole, Antarctica	USA	-89.98	-24.8	2810
SUM	Summit	Greenland	72.596	-38.422	3209.5
THD	Trinidad Head, California	USA	41.054	-124.151	107

Air Quality Data – AGAGE

The AGAGE air quality data came from the following URL:

Chloroform –

https://agage2.eas.gatech.edu/data_archive/agage/gc-md/complete/california/

Air Quality Data – NOAA

The NOAA air quality data came from the following URLs:

Methyl chloride –

https://gml.noaa.gov/aftp/data/hats/solvents/CH3CCl3/flasks/GCMS/CH3CCL3_GCMS_flask.txt

Benzene – Because benzene data are not posted publicly (<https://gml.noaa.gov/hats/flask/flasks.html>), US benzene data are obtained from the latest version of the Ambient Monitoring Archive (AMA;

<https://www.epa.gov/amtic/amtic-ambient-monitoring-archive-haps>) under the link “All .Rda data files by year”. All US benzene data (AQS parameter code 45201) are obtained from the following NOAA sites:

NOAA Site Code	AMA ID	City	State
KUM	15-001-NKUM	Pahoa	HI
MLO	15-001-NMLO	Kaulapuu	HI
NWR	08-013-NNWR	Boulder	CO
BRW	02-185-NBRW	Barrow	AK

Benzene data from the NOAA site ALT are obtained from Steven Montzka (stephen.a.montzka@noaa.gov) via the ERG contractor Regi Oommen (Regi.Oommen@erg.com).

Carbon tetrachloride – <https://gml.noaa.gov/aftp/hats/solvents/CCl4/insituGCs/CATS/hourly/>
Methyl bromide – https://gml.noaa.gov/aftp/data/hats/methylhalides/ch3br/flasks/CH3BR_GCMS_flask.txt
Methyl chloroform –
https://gml.noaa.gov/aftp/data/hats/solvents/CH3CCl3/flasks/GCMS/CH3CCl3_GCMS_flask.txt
Dichloromethane – https://gml.noaa.gov/aftp/data/hats/solvents/CH2Cl2/flasks/ch2cl2_GCMS_flask.txt
Tetrachloroethene – https://gml.noaa.gov/aftp/data/hats/solvents/C2Cl4/flasks/pce_GCMS_flask.txt
Air Quality Data – IMPROVE
Login information is needed to access IMPROVE data (<http://vista.cira.colostate.edu/Improve/improve-data/>). However, IMPROVE data are obtained from the AMA (<https://www.epa.gov/amtic/amtic-ambient-monitoring-archive-haps>) under the link “All .Rda data files by year”. The tables below indicate the AQS site codes and parameter codes pulled from the IMPROVE sites.

IMPROVE Site Code	AQS ID	City/Location	State
DENA	02-068-0003	Denali	AK
KALM	41-033-0010	Kalmiopsis	OR
PORE	06-041-0002	Sausalito	CA
REDW	06-015-0002	Redwood National Park	CA
TRCR	02-170-0011	Trapper Creek	AK
TUXE	02-122-9000	Tuxedni	AK
HACR	15-009-9001	Haleakala National Park	HI

Pollutant	AQS Parameter Code
Chromium VI	14115
Arsenic	88103
Chromium	88112
Lead	88128
Manganese	88132
Nickel	88136

Meteorology Data

Pressure (determined from elevation) and temperature data are needed to convert the VOC air quality data to local conditions. Some sites have collocated meteorology data, and some sites do not. For those that do not, a nearby meteorology site is used. The speciated PM data collected from the IMPROVE sites are already in local conditions and therefore, have no unit conversion. Meteorology data are collected for all NOAA and AGAGE sites. The temperature data come from three different sources. The HATS sites BRW, MLO, SMO, SPO, and SUM have collocated meteorology data. The remainder of the sites (except for THD) obtained meteorology data from NOAA’s Global Historical Climatology Network daily (GHCNd). This network is comprised of over 100,000 stations located across the globe monitoring a variety of different meteorological parameters (<https://www.ncei.noaa.gov/products/land-based-station/global-historical->

[climatology-network-daily](#)). The AGAGE site THD had collocated meteorology data in the past; however, data were obtained via email through NOAA and ERG contacts (see Appendix 2 and Appendix 3). Below outlines which sites use which meteorology data sources.

Meteorology Data Source	Site(s)
Collocated NOAA	BRW, MLO, SMO, SPO, SUM
GHCNd	KUM, NWR, ALT, MHD, CGO, PSA
Contact	THD

Meteorology Data – Collocated Meteorology Data

The collocated meteorology data (temperature at 2 meters) come from the following URLs (YYYY is the four-digit year):

Site	URL
BRW	https://gml.noaa.gov/aftp/data/meteorology/in-situ/brw/met_brw_insitu_1_obop_hour_YYYY.txt
MLO	https://gml.noaa.gov/aftp/data/meteorology/in-situ/mlo/met_mlo_insitu_1_obop_hour_YYYY.txt
SMO	https://gml.noaa.gov/aftp/data/meteorology/in-situ/smo/met_smo_insitu_1_obop_hour_YYYY.txt
SPO	https://gml.noaa.gov/aftp/data/meteorology/in-situ/spo/met_spo_insitu_1_obop_hour_YYYY.txt
SUM	https://gml.noaa.gov/aftp/data/meteorology/in-situ/sum/met_sum_insitu_1_obop_hour_YYYY.txt

Meteorology Data – GHCNd

Meteorology data (average daily temperature; field name “TAVG”) come from the closest GHCNd station (https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/) from the available inventory of stations (<https://www.ncei.noaa.gov/pub/data/ghcn/daily/ghcnd-inventory.txt>). As of this writing, the follow URLs are used.

Site	URL	Start Yr	End Yr	Dist. (m)
ALT*	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/CA002400305.csv.gz	2004	2023	6,033
BRW	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/USW00027502.csv.gz	1945	2023	6,111
CGO*	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/ASN00091245.csv.gz	1988	2023	22
HFM	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/USW00094746.csv.gz	1998	2023	38,519
KUM*	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/USW00021504.csv.gz	1998	2023	24,320
LEF	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/USR0000WGLI.csv.gz	1997	2021	30,271
MHD*	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/EIM00003962.csv.gz	1973	2023	95,369
MLO	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/USR0000HPTA.csv.gz	1999	2021	16,567
NWR*	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/US00005J08S.csv.gz	1989	2023	2,605
PSA*	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/AYM00089061.csv.gz	1976	2023	1,683
SMO	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/AQW00061705.csv.gz	1945	2023	18,604
SPO	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/AYW00090001.csv.gz	1994	2023	2,234
SUM	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/GLE00146908.csv.gz	2004	2023	586,671
THD	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/USR0000CSHH.csv.gz	2001	2021	22,643
USH	https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/ARM00087938.csv.gz	1957	2023	1,113

*Sites where GHCNd meteorological data are used.

Other Data Files

The elevation for each site is determined from

<https://gml.noaa.gov/dv/site/gmdsites.php?program=all&projtable=1> under the table “HATS Flask Sampling”. Data from this table are saved as NOAA_SiteMetaInfo.csv.

Methodology

Meteorology Data Averages

Collocated meteorology data. The collocated meteorology data are measured hourly. All hourly values are averaged up to daily after removing null values for each site if there are a minimum of 18 non-null hours for a given day. The temperature (°C) at 2 meters is selected for the temperature variable. Daily temperature is then averaged up to monthly for each site.

GHCNd meteorology data. GHCNd meteorology data are measured daily, and average temperature data are selected (tenths °C). Therefore, the variable “TAVG” is divided by 10 and is used as the temperature variable (°C). Daily temperature is averaged up to monthly temperature for each site.

THD meteorology data. The variable “TemperatureC” is pulled for the sub-hourly temperature data from THD. After subsetting only to values with the proper QC flag (“TemperatureC.QcDataDescriptor” equal to “S”), an hourly average is calculated if there are at least nine sub-hourly values, and a daily value is calculated if there are at least 18 hourly values. From there, daily values are averaged to monthly values. See Appendix 2 and Appendix 3 for email correspondence regarding the THD data.

Air Quality Data Averages

NOAA and AGAGE monthly average. NOAA and AGAGE data have a sub-daily sampling duration with an approximately weekly sampling frequency (with the exception of the in situ data for carbon tetrachloride). The average of each pollutant/site is calculated for each month. A minimum number of two values is needed to calculate a monthly average. All zeros are removed for chloroform before taking a monthly average. The following equation is used to calculate the monthly average for site i , pollutant j , year k , month l for M concentrations within a given month:

$$d_{i,j,k,l} = \frac{1}{M} \sum_{m=1}^M c_{i,j,k,l}$$

Converting units. Elevation and monthly temperature data are paired for every pollutant site/month average. Each site/month average is converted from its original units (i.e., ppt) to $\mu\text{g}/\text{m}^3$ standard conditions and $\mu\text{g}/\text{m}^3$ local conditions using the following equations for site i , pollutant j , year k , and month l :

$$d_{i,j,k,l} (\mu\text{g}/\text{m}^3 \text{ STD}) = \frac{d_{i,j,k,l} (\text{ppt})}{1000} \times \frac{MW_j}{24.46}$$

Where MW_j is the molecular weight of pollutant j (g/mol); 24.46 is the volume (L/mol) of gas at normal temperature and pressure. Molecular weight is pulled from the Archive (from AMA_POLLUTANT_CODES_DICTIONARY).

$$d_{i,j,k,l} (\mu\text{g}/\text{m}^3 \text{ LC}) = \frac{d_{i,j,k,l} (\text{ppt})}{1000} \times \frac{MW_j}{(24.46 * 101325 * (t_{i,k,l} + 273)/(298 * p_i))}$$

Where $t_{i,k,l}$ is the temperature (°C) for site i , year k , and month l ; normal temperature and pressure is defined as 101325 Pa (1 atm) and 298 K; p_i is the pressure (Pa) for site i defined as

$$p_i = 101325 * (1 - 0.0000225577 * Elev_i)^{5.25588}$$

Where $Elev_i$ is the elevation (m) at site i .

The equation for pressure can be found here: https://www.engineeringtoolbox.com/air-altitude-pressure-d_462.html. All three units (i.e., ppt, $\mu g/m^3$ standard conditions, and $\mu g/m^3$ local conditions) are retained for annual site and overall averages.

Annual average. The annual average is taken for all three units for site i , pollutant j , year k for L monthly averages using the following equation:

$$\mu_{i,j,k} = \frac{1}{L} \sum_{l=1}^L d_{i,j,k,l}$$

Latitude-weighted averages. The final boundary condition for pollutant j and year k is calculated by summing a latitude-weighted annual site-level averages for N sites using the following equation:

$$G_{j,k} = \frac{\sum_{i=1}^N \cos\left(\frac{\pi lat_i}{180}\right) * \mu_{i,j,k}}{\sum_{i=1}^N \cos\left(\frac{\pi lat_i}{180}\right)}$$

Where lat_i is the latitude for site i . The methodology for latitude-weightings have been described in Montzka et al. (2011; 1999).

IMPROVE data. IMPROVE data are averaged differently than NOAA and AGAGE. Because of the large percentage of non-detects in the IMPROVE data, the Method Detection Limit (MDL) is an upper bound and used as a surrogate for ambient concentrations. Unlike NOAA and AGAGE, the boundary condition is calculated more directly by averaging all the MDLs for all collected samples across the entire year for each pollutant (with the exception of chromium VI, explained below). All IMPROVE data are collected in units of $\mu g/m^3$ local conditions. No unit conversion or latitude-weighting is done.

Chromium VI. Chromium VI is calculated differently than the other IMPROVE parameters. The boundary condition for chromium VI is calculated by weighting the boundary condition for chromium by a ratio using the following equation:

$$G_{Chromium\ VI,k} = \frac{b_{Chromium\ VI,k}}{b_{Chromium\ TSP,k}} * G_{Chromium,k}$$

Where $G_{Chromium\ VI,k}$ is the final boundary condition for chromium VI for year k , $G_{Chromium,k}$ is the final boundary condition for chromium for year k , $b_{Chromium\ VI,k}$ is the annual average for chromium VI (TSP) LC (AQS parameter code 14115) only from NATTS across all available sites for year k , and $b_{Chromium\ TSP,k}$ is the annual average for chromium (TSP) LC (AQS parameter code 14112) across all networks and across all available sites for year k . With $b_{Chromium\ VI,k}$ and $b_{Chromium\ TSP,k}$, non-detects are treated as zeros and are included in the calculation. All data are pulled from the AMA.

Results

Name	ppt	$\mu g/m^3$ (STD)	$\mu g/m^3$ (LC)
Chloroform	15.27	0.0745	0.0774
Benzene	30.51	0.0974	0.0878

Bromomethane	7.12	0.0276	0.0236
Chloromethane	559.74	1.1554	0.9824
Dichloromethane	68.51	0.2379	0.2043
Tetrachloroethylene	1.68	0.0114	0.0099
Carbon tetrachloride	77.46	0.4871	0.4074
Methyl chloroform	1.23	0.0067	0.0063
Arsenic (PM2.5)	---	---	2.152E-04
Chromium (PM2.5)	---	---	1.334E-04
Lead (PM2.5)	---	---	5.070E-04
Manganese (PM2.5)	---	---	2.492E-04
Nickel (PM2.5)	---	---	1.108E-04
Chromium VI	---	---	2.034E-07

Name	% Below MDL
Arsenic (PM2.5)	98.9
Chromium (PM2.5)	86.8
Lead (PM2.5)	81.4
Manganese (PM2.5)	61.0
Nickel (PM2.5)	90.1

Citations

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-CO2 greenhouse gases and climate change. *Nature*, 476(7358), 43-50, doi: 10.1038/nature10322, August 3.

Appendix D. Model Evaluation Summaries

EPA performed model evaluations for other AirToxScreen pollutants. These evaluations, including graphics, can be found in the [Supplemental Data](#) folder on the AirToxScreen website.

Appendix E. Exposure Factors for AirToxScreen

The memorandum contained within this Appendix (below) describes in detail how EPA developed exposure factors for each chemical assessed in AirToxScreen. These calculated exposure factors can be found in the [Supplemental Data](#) folder accompanying this TSD.

MEMORANDUM

To: Matt Woody, Rod Truesdell, and Michael Moeller

From: ICF: Minti Patel, Chris Holder, Aishwarya Javali, Jared Wang, Graham Glen, and Melissa Polansky
Innovate! Inc.: David Yarnell, Ben Holloway, and Michael Blair

Date: December 4, 2023

Re: Updating the Hazardous Air Pollutant Exposure Model (HAPEM) for Use in the 2020 Air Toxics Screening Assessment (AirToxScreen)

ICF (“we”) updated the default input files accompanying the Hazardous Air Pollution Exposure Model (HAPEM), and we updated some of the HAPEM source code to accommodate the new default files. The resulting new version of HAPEM (i.e., HAPEM8), with its default files, simulates exposure concentrations for all populated census tracts using 2020 census data, commuting data from the 2012–2016 and 2015–2020 American Community Survey (ACS), and time-activity data from the April 2020 version of the U.S. Environmental Protection Agency (EPA) Consolidated Human Activity Database (CHAD). In this technical memorandum, we describe how we updated the default files and model source code, including the quality-assurance (QA) steps we used and the format of the final default files. HAPEM8 and its updated default files will be available for download as EPA’s latest, default version of HAPEM.¹ We modeled exposure concentrations using HAPEM8 for the 2020 Air Toxics Screening Assessment (AirToxScreen), as described in a separate memorandum.²

1. Introduction to HAPEM and its Use in AirToxScreen

HAPEM is a model used by EPA to perform screening-level assessments of long-term inhalation exposures to hazardous air pollutants (HAPs). Exposure concentrations output

¹ We anticipate HAPEM8 and its User’s Guide will be made available by EPA online in Winter 2023–2024. As of April 26, 2021, HAPEM7 is available for download at <https://www.epa.gov/fera/human-exposure-modeling-hazardous-air-pollutant-exposure-model-hapem>.

² We describe the use of HAPEM8 in the 2020 AirToxScreen in the ICF Memorandum “HAPEM8 Modeling for the 2020 Air Toxics Screening Assessment (AirToxScreen)” dated December 4, 2023, to Matt Woody, Rod Truesdell, and Michael Moeller of EPA’s Office of Air Quality Planning and Standards.

by HAPEM are stratified by location (i.e., U.S. census tract), age group, and the individual source categories and HAPs being modeled. The model's default files cover all 50 states in the US, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands (USVI).

AirToxScreen uses HAPEM with these default files. Therefore, exposure concentrations produced for the AirToxScreen have the same stratifications discussed above, though AirToxScreen-specific post-processing includes accumulating exposure concentrations into a lifetime period of exposure (defined as 70 years). AirToxScreen (the successor to the National Air Toxics Assessment, or NATA) is a nationwide modeling assessment of air concentrations, exposure concentrations, and potential human health cancer risks and chronic hazards associated with exposure to HAP emissions from man-made and naturally occurring sources. These results are spatially partitioned by various census geographies. EPA models air concentrations using two air-concentration models: AERMOD (the atmospheric dispersion model developed by the American Meteorological Society and the EPA Regulatory Model Improvement Committee) and CMAQ (EPA's Community Multiscale Air Quality model). Those modeled air concentrations are the "air quality" inputs for HAPEM. AirToxScreen is not an enforcement tool to determine compliance with various standards of emissions, air quality, or health impacts; rather, it is a screening-level tool used to rank HAPs based on potential health impacts (nationally and locally), estimate the numbers of people and demographics potentially subject to health risks above levels of concern, identify gaps in data, and prioritize locations, source categories, and HAPs to inform additional data collection and assessment.

Data on where people live and work, and otherwise how they spend their time, are critical to the completeness of the exposure modeling conducted with HAPEM. The version of HAPEM currently available for download (HAPEM7) uses census data from the year 2010 and activity patterns gleaned from the 2014 version of CHAD.³ We have updated the default files used by HAPEM to reflect or approximate 2020 census data and the version of CHAD available in April 2020. We also have updated HAPEM source code as necessary, mostly to accommodate the sizes of the updated inputs.

2. Updating Census-based Data

a. Population File – "population_HAPEM8.txt"

The HAPEM default population input file ("population_HAPEM8.txt" in HAPEM8) provides the number of people in each HAPEM age group residing in each tract in the 50 states

³ The content, functionality, and implementation of HAPEM7 are discussed in the HAPEM7 User's Guide, available as of April 26, 2021 at <https://www.epa.gov/fera/hazardous-air-pollutant-exposure-model-hapem-users-guides>.

plus the District of Columbia, Puerto Rico, and the USVI. The HAPEM default ages are binned into six groups: 0–1, 2–4, 5–15, 16–17, 18–64, and 65 years and older.

HAPEM7: For the previous HAPEM model (HAPEM7), the population data were derived from the 2010 census Summary File 1: Table PCT12 (“Sex by Age”), available separately for males and females, and provided by each year of age. Population data for the USVI were not available from Table PCT12, but they were available by querying the census American FactFinder web page. For the purposes of HAPEM, the male and female data from Table PCT12 were aggregated male+female and into the HAPEM age groups. The American FactFinder USVI data were available by groups of ages which did not match the HAPEM age groups. For the purposes of fitting the USVI age groups to the HAPEM age groups, it was assumed that population counts were evenly distributed among the incremental years represented in the USVI 0–4-year group (i.e., two fifths being 0–1 and three fifths being 2–4 years old) and in the 15–17 group (i.e., one third being 15 and two thirds being 16–17 years old); all other USVI age groups (e.g., 5–9, 10–14, 18–19, ..., 62–64, 65–66, ..., 85 and over) required no subdivision to fit into the HAPEM age groups.

HAPEM8: For HAPEM8, we used the 2020 census’ Table PCT12 (“Sex by Single-year Age”) to update the HAPEM population file for all areas except the USVI. We obtained population data for the USVI from the 2020 census’ Table PCT1 (“Sex by Single Years of Age”, U.S. Virgin Islands). We summed the population information across the two sexes and aggregated the single-age data into the six default HAPEM age groups.

i. Quality Assurance

We checked that the HAPEM8 default population file contained all the expected census geographies (i.e., all the 2020 tracts) by comparing against the 2020 census gazetteer tract file⁴ (and tigerweb.geo.census for the USVI). We created the file using Microsoft® Excel, where we cross-checked our processing formulas to ensure individual ages were accurately summed into the HAPEM age groups. We also compared the grand total of those binned population numbers to the grand total of the raw census data of individual ages. Lastly, we compared the HAPEM8 population file against the HAPEM7 file to ensure proper formatting.

ii. Content and Format

The HAPEM8 population data are contained in a fixed-width, space-delimited text file with characteristics shown in Table 1. The file contains seven columns and a total of 85,427 rows of data (after two header rows). Each data row corresponds to a tract, where

⁴ As of August 2023, the census gazetteer files are available at <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html>.

the first field identifies the tract using Census Federal Information Processing Series (FIPS) coding⁵, and fields 2–7 contain population counts per age group. Population counts are whole numbers (no commas separating thousands). The first header row labels the fields, where the age-group columns are identified by the youngest age within the group (i.e., B_00 for age group 0–1 years old, B_02 for age group 2–4, and so on). The second header row serves an unknown purpose, but we retained it from the HAPEM7 population file. In Figure 1 we show the first ten data rows of the population file. On the whole, the HAPEM8 total tract populations range from 0 (for 617 tracts across 42 states and territories, which is less than 1 percent of all tracts) to 37,892, with an average of 3,919. The total population in this file is 334,822,301.

Table 2. Characteristics of the HAPEM8 Population File

Variable	Description	Character Start Position on Data Row	Character Length on Data Row ^a
TRACT	Full census FIPS code for home tract	1	11
B_00	Total population ages 0–1 years	17	8
B_02	Total population ages 2–4	25	8
B_05	Total population ages 5–15	33	8
B_16	Total population ages 16–17	41	8
B_18	Total population ages 18–64	49	8
B_65	Total population ages 65 and older	57	8

Note: FIPS = Census Federal Information Processing Series

^a Any unused character space after a number and/or between fields consists of blank spaces.

⁵ The full tract identifier used by census consists of a 2-digit state code, a 3-digit county code, and a 6-digit tract code, concatenated together to form an 11-digit code.

TRACT	B_00	B_02	B_05	B_16	B_18	B_65
	COM	COM	COM	COM	COM	COM
01001020100	30	67	252	56	1086	284
01001020200	33	78	283	77	1268	316
01001020300	60	109	471	91	1887	598
01001020400	81	137	596	88	2383	961
01001020501	75	115	625	138	2599	770
01001020502	84	144	569	105	2090	292
01001020503	81	120	542	104	2188	581
01001020600	77	141	642	101	2198	570
01001020700	91	161	491	86	2105	475
01001020801	44	104	487	131	1861	516

Figure 9. Excerpt from the HAPEM8 Population File

b. Commuting-flow File – “commute_flow_HAPEM8.txt”

In HAPEM, the tract where a person resides is their home tract, and the tract where a person works is their work tract. Some people work within their home tract (i.e., the work tract is the home tract); the remaining employed people work outside their home tract. For the employed people in each home tract, the HAPEM default commuting-flow input file (“commute_flow_HAPEM8.txt” in HAPEM8) provides the fraction of those people who work within their home tract and the fraction that commute to work in each other tract. For each home tract, the file contains only the tract(s) where residents of the home tract work (i.e., there are no fractions of 0). These commuting data are provided for nearly all the (home) tracts contained in the HAPEM population file, with exceptions noted in the discussion below.

HAPEM7: For the previous HAPEM model (HAPEM7), the commuting-flow data were derived from data provided by the U.S. Department of Transportation (DOT) Federal Highway Administration (FHWA)—specifically, their Microsoft® Access—based Census Transportation Planning Products (CTPP) 2006–2010 file, based on 2006–2010 five-year summary data from the ACS and commissioned by the American Association of State Highway and Transportation Officials (AASHTO). This Access database contains estimates of the total number of workers commuting within or between tracts.

HAPEM8: For the HAPEM8 commuting-flow file, we used the FHWA CTPP data based on the 2012–2016 five-year summary data from the ACS (the most current available).⁶ The

⁶ As of August 2023, the 2012–2016 CTPP data are available at <https://ctpp.transportation.org/2012-2016-5-year-ctpp/>.

data are available by state, and we downloaded the state files and concatenated them into the overall commuting-flow file.

Because the 2012–2016 CTPP data uses geography (census tracts) for the 2010 census, we mapped the data to 2020 census tracts using a 2020 relationship file made available by the Census Bureau.⁷ The relationship file provides a one-to-many crosswalk from 2010 tracts to 2020 tracts, indicating the surface area of overlap between the two vintages of tracts. We used proportion of total overlapping tract area (sum of land area and water area) to redistribute the CTPP commuter data to the 2020 tracts.

To produce the commuter fractions, we divided the number of workers in each home-tract/work-tract pair by the total number of workers residing in the home tract. We calculated the distance between each home-tract/work-tract pair using the 2020 census coordinates of tract internal points (i.e., centroids), available from the 2020 census gazetteer.⁴ More specifically, we used the `distGeo` function in the `geosphere` package of R⁸ to calculate the distance between the internal-point tract coordinates.

A small number of tracts were absent from the CTPP data as home tracts (totaling 920 tracts, 1 percent of all tracts; 469 of these were unpopulated, while 451 were populated). HAPEM will model each missing tract as if all its employed residents work within the home tract (i.e., for the purposes of HAPEM modeling, they essentially do not commute), so we did not insert any data for these missing tracts. Additionally, the CTPP contained no data on all 32 tracts in the USVI. To prevent the USVI from being conspicuously missing from the commuting file, we inserted one record for each USVI tract, where work tract equals home tract and the commute distance is 0 kilometer (km), which is how HAPEM would model them if they remained missing from the file.

i. Quality Assurance

We ensured that the data downloaded from the CTPP website matched that obtained from the CTPP online queries, by randomly checking four tracts from five different states. We confirmed the numbers of home and work tracts at various stages of the analysis. We also ensured the accuracy of the commuting fractions including the usage of the relationship file to estimate flows between the 2020 census tracts, through a thorough check of the calculations for Alaska. We ensured that the cumulative commuting fraction equaled 1 for each home tract (with an allowance for very small rounding errors). We used

⁷ As of August 2023, the census relationship files are available at <https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.2020.html#tract>.

⁸ As of August 2023, the R `geosphere` package files are available at <https://www.rdocumentation.org/packages/geosphere/versions/1.5-18>.

mapping software to check a small number of the commuting distances calculated by the distGeo function.

ii. **Content and Format**

The HAPEM8 commuting-flow data are contained in a fixed-width, space-delimited text file with characteristics shown in Table 2. The file contains five columns (the first being empty) and a total of 6,004,343 rows of data with no header rows. Each data row corresponds to a unique home-tract/work-tract pair, where the second and third fields respectively contain the home and work tract identifiers using FIPS coding, and the fourth and fifth fields respectively contain the commuting distance (in km) and the fraction of workers commuting between the associated home and work tracts. Distance values are presented to no more than two decimal places (i.e., hundredths of km, which is tens of meters), while commuting fractions are presented to no more than eight decimal places. In Figure 2 we show the first ten data rows of the commuting-flow file. On the whole, the data show on average there are 71 work tracts per home tract, up to a maximum of 313 work tracts. In 351 home tracts (which is less than 1 percent of home tracts), all workers worked within their home tract.

Table 3. Characteristics of the HAPEM8 Commuting-flow File

Field Number	Description	Character Start Position on Data Row	Character Length on Data Row ^a
1	Leading space in file	1	1
2	Full census FIPS code for home tract	2	11
3	Full census FIPS code for work tract	14	11
4	Distance in kilometers between home and work tract	26	8
5	Fraction of workers in the home tract commuting to the work tract	34	10

Note: FIPS = Census Federal Information Processing Series

^a Any unused character space after a number and/or between fields consists of blank spaces.

01001020100	01001020100	0.00	0.03045067
01001020100	01001020803	6.98	0.00365408
01001020100	01001020200	1.92	0.04263094
01001020100	01001020300	3.12	0.04872107
01001020100	01001020400	4.56	0.01827040
01001020100	01001020501	7.51	0.07308161
01001020100	01001020502	7.07	0.02436054
01001020100	01001020503	6.25	0.03654080
01001020100	01001020600	4.08	0.02436054
01001020100	01001020700	7.52	0.06090134

Figure 10. Excerpt from the HAPEM8 Commuting-flow File

Commuting distances greater than 120 km are assumed in HAPEM to be very atypical for a daily commuter. As noted in the HAPEM User’s Guide,¹ during an earlier development stage of HAPEM, commuting flows were examined as a function of distance. The analysis revealed that commute flows generally decreased linearly in log space with increasing distance, but at commute distances greater than about 100 km that trend flattened. This suggested that those longer commutes likely did not occur daily. Since HAPEM is designed to construct daily commutes for simulated workers, it would not be appropriate for HAPEM to model daily commutes longer than about 120 km, and thus HAPEM ignores these longer commutes in constructing the commute distance distributions for each tract. Most home tracts have at least one work tract that is more than 120 km away; that is, in approximately 64 percent of home tracts there is at least one person residing there who commutes farther than 120 km. However, this affects only 3 percent of home-tract/work-tract pairs. Ignoring these records with commuting distances greater than 120 km, the average tract-to-tract distance is 22.5 km (weighting all tract pairs equally, not by numbers of people performing those commutes; that average is 42.1 km when commuting distances greater than 120 km are included).

c. Commuting-time File – “commute_time_HAPEM8.txt”

While the HAPEM commuting-flow file (see Section 2.2) contains information on the frequency distribution of commuting distances for workers in a given home tract, the HAPEM commuting-time file (“commute_time_HAPEM8.txt” in HAPEM8) contains information on the method of commuting (public versus private transit) and the average commuting time per person. These commuting-time data are provided for all the tracts contained in the HAPEM population file, though no commuting data were available for the USVI, as discussed below.

HAPEM7: For the previous HAPEM model (HAPEM7), the commuting-time file data were derived for 2010 from the 2006–2010 five-year summary data from the ACS Tables B08301 (“Means of Transportation to Work for Workers 16+ Years”), C08134 (“Means of Transportation to Work by Travel Time to Work for Workers 16+ Years who Did Not Work at Home”), and C08136 (“Aggregate Travel Time to Work (in Minutes) by Means of Transportation to Work for Workers 16+ Years who Did Not Work at Home”).

HAPEM8: For the HAPEM8 commuting-time file, relative to the HAPEM7 file, we identified equivalent data for the year 2020 from other tables from the ACS 2016–2020 five-year summary data, as detailed in the following paragraphs.

Table B08134 (“Means of Transportation to Work by Travel Time to Work for Workers 16+ Years who Did Not Work at Home”) contains the numbers of people commuting to work, irrespective of commuting time, for specific means of transit in broader groups than in Table B08301 that was used for HAPEM7. We used Table B08134 to derive the proportion of workers traveling by public transit (i.e., bus, trolley bus, streetcar, trolley car, subway, elevated train, railroad, and ferryboat) and the proportion of commuters traveling by private transit (i.e., car, truck, van, taxicab, motorcycle, bicycle, any other non-public means except walking). People working from home (i.e., workers not commuting) were not included in this dataset. We excluded people walking to work, which are cases where we assume people work within their home tract and thus are not considered commuters for the purposes of HAPEM exposure modeling. As such, the fractions of workers commuting by public and private transit sum to 1, except a relatively small number of tracts (approximately 1,064, or 1 percent of all tracts) where the survey recorded no commuting activity.

ACS Table B08136 (“Aggregate Travel Time to Work (in Minutes) by Means of Transportation to Work for Workers 16+ Years who Did Not Work at Home”) contains travel times to work by the same transit means as in Table B08134, summed across all people who use those means. We divided these aggregate travel times by the corresponding population counts from Table B08134, resulting in average per-person travel times to work, by public transit and by private transit. We then multiplied the average per-person travel times by two to derive the round-trip time used in HAPEM8 commuting-time file. Commuting times related to public transit include time spent waiting at a bus or train stop, and commuting times (and population counts from Table B08134) related to private transit include walking commuters; these times are included in our calculations because they cannot be disaggregated from the total commuting time. If the data derived from Table B08134 (used for the proportions of workers commuting by public and private means) indicated that a tract had no commuters using public means, then we set commuting times to 0 for public means; similarly, we set private commuting

times to 0 if there were no private commuters, and we set both public and private commuting times to 0 if there were no commuters at all.

Commuting data for the USVI were not available from the ACS, so we set all their workers to work in their home tract (i.e., commute neither by public nor private transit, with commuting times equal to 0). This is consistent with how we approached USVI data in the commuting-flow file.

Aggregate commuting-time data also were unavailable from Table B08136 (either missing entirely from the table, or present in the table but with flags [or “-“ value entries] indicating a lack of reliable data) for 87 percent⁹ of tracts in areas outside the USVI. We used county-average aggregate times for 68 percent of these missing tracts (i.e., for 32 percent of all tracts outside of the USVI) and state times for the remaining 32 percent of missing tracts (i.e., for 5 percent of all tracts outside of the USVI). We divided those county and state aggregate times by the county and state counts of commuters to produce average, per-person, one-way commuting times, and we multiplied by two to obtain round-trip times. We stratified these county and state averages by public- and private-transit means.

For the State of Wyoming, although state time aggregates had a null value, values were available for some counties. To derive a state-level aggregate, we summed values across all counties. We used this to substitute as a state-level value in cases of missing tract- and county-level aggregates in Wyoming.

i. Quality Assurance

We checked that the HAPEM8 default commuting-time file contained all the expected census geographies (i.e., all the 2020 tracts) by comparing against the default population file (see Section 2.1). We spot-checked several very different tracts (e.g., rural Alaska, city in Alaska, Queens County in New York City) to ensure that the ACS data pulled into our Excel processing file matched the raw data displayed on the ACS website. We checked each of our Excel processing formulas, including aggregations across census transit types, the calculations of county- and state-average data, and the compilation of those data into a complete set of tract data. We ensured that the public and private commuting proportions summed to 1 for every record except the tracts with 0 commuters. For consistency, we confirmed that tracts with commuting workers (from the HAPEM8 default commuting-fraction file, discussed later in Section 2.4) had non-0 commuting-time values in the final file.

⁹ It was unclear why a large percentage of these data were missing or marked as insufficient.

ii. Content and Format

The HAPEM8 commuting-time data are contained in a tab-delimited text file with characteristics shown in Table 3. The file contains five columns and a total of 85,427 rows of data with no header rows. Each row corresponds to a tract, where the first field contains the tract identifier using census FIPS coding, the second and third fields respectively contain the proportion of commuters who travel by public transit (excluding taxicabs) and by private transit (including taxicabs), and the fourth and fifth fields respectively contain the average round-trip times (in minutes) commuting to work by public transit and by private transit. All values in fields 2–5 are displayed to four decimal places. In Figure 3 we show the first ten data rows of the commuting-time file. On the whole (except the USVI), the data show that 86 percent of commuters used private transit, and all commuters in 40 percent of census tracts used private transit. The conditional-average round-trip private-transit commute was 53 minutes (100 minutes for public transit) (conditional averaging considers only non-zero values). This statistic treats every tract equally, rather than weighting by commuting population, and it includes county and state averages where we used them. The longest round-trip commuting times in the data set are 163 minutes for private transit and 336 minutes for public transit.

Table 4. Characteristics of the HAPEM8 Commuting-time File

Field Number	Description
1	Full census FIPS code for home tract
2	Proportion of workers commuting outside of the home by public transit
3	Proportion of workers commuting outside of the home by private transit
4	Average round-trip commuting time for workers commuting outside of the home by public transit
5	Average round-trip commuting time for workers commuting outside of the home by private transit

Note: The position where table values begin and the number of characters per value are not relevant in a tab-delimited format.

01001020100	0.0000	1.0000	0.0000	50.8320
01001020200	0.0000	1.0000	0.0000	50.8320
01001020300	0.0000	1.0000	0.0000	49.5615
01001020400	0.0377	0.9623	89.9083	50.8320
01001020501	0.0166	0.9834	89.9083	50.8320
01001020502	0.0000	1.0000	0.0000	50.8320
01001020503	0.0000	1.0000	0.0000	50.8320
01001020600	0.0000	1.0000	0.0000	50.8320
01001020700	0.0000	1.0000	0.0000	50.8320
01001020801	0.0000	1.0000	0.0000	50.8320

Figure 11. Excerpt from the HAPEM8 Commuting-time File

d. Commuting-fraction File – “commute_fraction_HAPEM8.txt”

The HAPEM commuting-fraction file (“commute_fraction_HAPEM8.txt” in HAPEM8) contains the fraction of workers in each tract who commute to work and the fraction who do not commute, stratified by age group. Workers who walk to work are not included as commuters for HAPEM8.

HAPEM7: The HAPEM7 commuting-fraction data were derived for 2010 from the 2006–2010 five-year summary data from the ACS—specifically, ACS Table B23001 (“Sex by Age by Employment Status for the Population 16 Years and Over”) and ACS Table B08101 (“Means of Transportation to Work by Age for Workers 16+ Years”). HAPEM7 included Armed Forces members but did not include those walking to work.

HAPEM8: For the HAPEM8 commuting-fraction file, relative to the HAPEM7 file, we identified equivalent data for 2020 from Table B08101 of the ACS 2016–2020 five-year

summary data, not including those walking to work. Detailed calculation methods are discussed in the following paragraphs.

ACS Table B08101 (“Means of Transportation to Work by Age for Workers 16+ Years”) contains the numbers of people per age group commuting to work by various means of transit (e.g., “Total”, “Car, truck, or van: Drove alone”, “Car, truck, or van: Carpooled”, “Public Transportation (excluding taxicab)”). We used this table to derive 1) the numbers of workers who commuted by means other than walking and 2) the number of people per HAPEM age group who are workers. As we did in calculating the proportion of workers commuting by public and private transit (see Section 2.3), we excluded people walking to work because they likely work within their home tract, and for simplicity we consider them not to be commuters in HAPEM.

For each tract and HAPEM age group, we calculated the fraction of workers commuting as $(\text{number of people aged 16+ years who commute to work other than by walking}) \div (\text{number of workers aged 16+ years})$. The fraction of workers not commuting is 1 minus the above fraction.

Commuting data for the USVI were not available from the ACS, so we set data in the commuting-fraction file such that all workers in the USVI work in their home tract (i.e., did not commute). This is consistent with how we treated USVI data in the commuting-flow and commuting-time files (see Sections 2.2 and 2.3, respectively).

i. Quality Assurance

We performed systematic data processing using R. As a thorough check, we also repeated the processing in Excel (by a separate person than who authored the R code), finding that both methods of processing resulted in the same values. We checked that all commuting-fraction numbers were between 0 and 1. We ensured that the fractions of workers in each age group commuting and not commuting summed to 1 for every record. We compared the HAPEM8 and HAPEM7 files to ensure proper layout.

ii. Content and Format

The HAPEM8 commuting-fraction data are contained in a tab-delimited text file with characteristics shown in Table 4. The file contains five columns and a total of 85,427 rows of data with no header rows. Each row corresponds to a tract, where the first field contains the tract identifier using census FIPS coding, the second and third fields respectively contain the fraction of workers aged 0–1 years who do not commute and who do commute, and the remaining fields show the same data for each of the other five HAPEM age groups. All values in fields 2–13 are displayed to four decimal places. Nobody younger than 16 years is considered employed and a commuter, so all values for “does

not commute to work” are 1 and all values for “commutes to work” are 0 for the first three HAPEM age groups. In Figure 4 we show the first ten data rows of the commuting-fraction file. On the whole (except the USVI), the data show that the average tract commuting fraction is 0.80 (80 percent of workers commuting) for ages 16–17 years, 0.89 for ages 18–64 years, and 0.83 for 65+ years. This statistic treats every tract equally, rather than weighting by commuting population.

Table 5. Characteristics of the HAPEM8 Commuting- fraction File

Field Number	Description
1	Full census FIPS code for home tract
2	Proportion of age group 1 (ages 0–1 years) that does not commute to work
3	Proportion of age group 1 (ages 0–1) that commutes to work
4	Proportion of age group 2 (ages 2–4) that does not commute to work
5	Proportion of age group 2 (ages 2–4) that commutes to work
6	Proportion of age group 3 (ages 3–15) that does not commute to work
7	Proportion of age group 3 (ages 3–15) that commutes to work
8	Proportion of age group 4 (ages 16–17) that does not commute to work
9	Proportion of age group 4 (ages 16–17) that commutes to work
10	Proportion of age group 5 (ages 18–64) that does not commute to work
11	Proportion of age group 5 (ages 18–64) that commutes to work
12	Proportion of age group 6 (ages 65 and older) that does not commute to work
13	Proportion of age group 6 (ages 65 and older) that commutes to work

Note: The position where table values begin and the number of characters per value are not relevant in a tab-delimited format.

01001020100	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0101	0.9899
	0.0000	1.0000								
01001020200	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0149	0.9851
	0.0000	1.0000								
01001020300	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0909	0.9091	0.0334	0.9666
	0.0000	1.0000								
01001020400	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0379	0.9621
	0.1374	0.8626								
01001020501	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0589	0.9411
	0.0000	1.0000								
01001020502	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0051	0.9949
	1.0000	0.0000								
01001020503	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0896	0.9104
	1.0000	0.0000								
01001020600	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0717	0.9283
	0.1000	0.9000								
01001020700	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0178	0.9822
	0.0000	1.0000								
01001020801	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.1304	0.8696	0.0917	0.9083
	0.0000	1.0000								

Note: Contents wrap around due to space constrictions in this figure.

Figure 12. Excerpt from the HAPEM8 Commuting- fraction File

e. Distance-to-road File –
“proximity_road_HAPEM8.txt”

The HAPEM distance-to-road file (“proximity_road_HAPEM8.txt” in HAPEM8) contains information on the fraction of a tract’s residents that live within each of three categories of distance from a major roadway, by age group. These distances are 0–75 m, greater than 75 m up to 200 m, and greater than 200 m. The file contains these data for all tracts in the HAPEM8 population file. We conducted the proximity assessment at the level of census blocks and stratified by age and sex, and then we aggregated the block-level results up to the tract level and stratified only by age group.

We used block-level geographies from the 2020 Census TIGER/Line Shapefiles¹⁰ for all areas except the USVI. We used block-level population data from the 2020 census’ Table P12 (“Sex by Age for Selected Age Categories”). For the USVI, we used 2020 tract-level geographies and population data from the 2020 census’ Table PCT1 (“Sex by Single Years of Age”, U.S. Virgin Islands). We downloaded the geometries and demographic data separately before joining them into a single table in a PostGIS server.

We compiled roadway location data from the 2022 Census TIGER/Line “All Roads” U.S. roadway layer. We considered the three roadway types shown in Table 5 to be major roads for the purposes of evaluating enhanced pollutant exposure to people living near heavy-use roads, assuming that other features such as traffic circles, cul-de-sacs, local or neighborhood roads, rural roads, and city streets do not meet the definition.

Table 6. Types of “Major” Roads Included in the Roadway-proximity Assessment

Roadway Type	Definition
Primary Road	Generally divided, limited-access highways within the interstate highway system or under state management, and distinguished by the presence of interchanges. Accessible by ramps and may include some toll highways.
Ramp	Allows controlled access from adjacent roads onto a limited-access highway, often in the form of a cloverleaf interchange.
Secondary Road	Main arteries, usually in the U.S., state, or county highway systems. Have one or more lanes of traffic in each direction, may or may not be divided, and usually have at-grade intersections with many other roads and driveways.

We used Postgres software utilizing PostGIS to perform the steps noted below for the roadway-proximity geospatial analyses.

¹⁰ As of August 2023, the U.S. Census TIGER/Line data available at <https://www.census.gov/geo/maps-data/data/tiger-line.html>.

-
1. We created 75- and 200-m buffers around all major roadways. We clipped these buffers at the boundaries of census blocks, such that no buffer crossed a block boundary.
 2. We assumed uniform population across blocks and used area analysis to calculate the ratio of each block area within each buffer. For each block, we calculated the fraction of the area that was within the 75-m buffer, the fraction that was within the 200-m buffer (subtracting the 75-m portion to create results for the 75-to-200-m distance), and the fraction that was outside the 200-m buffer. We calculated the ratios of (block area that fell within each of the major-roadway buffers) divided by (total block area).
 3. For each block and buffer, we multiplied the ratio from Step 2 above by the block population count per gender and age group. These are the numbers of people residing 0–75 m, greater than 75 m up to 200 m, and greater than 200 m of a major roadway, at the block level and stratified by sex and age.
 4. We aggregated the data from Step 3 above to the tract level and summed together the male and female data. We then divided the population counts within the major-roadway buffers by the total tract population, stratified by each of the six HAPEM age groups. The result for each age group is the fraction of residents who live within each of the three distance buffers of a major roadway.

i. Quality Assurance

We implemented several layers of QA with multiple staff members at different stages of the processing. A major focus was on calculations performed in Step 4 above (i.e., the final steps of processing population data and aggregating to the tract level). We reviewed the block-level population data to ensure they were complete, and we reviewed our processed block-level results to ensure they included all blocks nationwide.

We checked that the major-roadway buffer ratios from Step 3 summed to 1 for every block (and in Step 4 summed to 1 for every tract). In this process, we implemented post-processing algorithms to remove rounding errors so that fractions summed to 1 where appropriate (when processed at 4 decimal places).

We spot-checked that the processed tract population data summed to the correct state-total populations and summed correctly across age groups. We also noted that we should not always expect the fraction of tract area within the individual major-roadway buffers to equal the fraction of tract population within the buffers. This is because we performed the assessment at the block level and then aggregated to the tract level, where each block has a unique population density that makes aggregated populations unequal to aggregated areas.

We also discovered that HAPEM8 throws an error if any age group in a tract has all its population living within 75 m of a major roadway. This happened with a single tract, and

we worked around the model error by setting 99.98% (0.9998) living in that buffer, with 0.01% (0.0001) living in the second buffer and again 0.01% living in the third buffer.

ii. **Content and Format**

The HAPEM8 distance-to-road data are contained in a tab-delimited text file with characteristics shown in Table 6. The file contains 22 columns and a total of 85,427 rows of data with no header rows. Each row corresponds to a tract, where the first field contains the tract identifier using census FIPS coding, fields 2–4 contain the fractions of tract area within each of the three roadway buffers, and the remaining fields show similar data for the fractions of people in each HAPEM age group who reside within those buffers. All values in fields 2–22 are displayed to four decimal places. The population fractions in tracts with 0 residents are shown as 0 values. In Figure 5 we show the first ten data rows of the distance-to-road file.

Table 7. Characteristics of the HAPEM8 Distance-to-road File

Field Number	Description
1	Full census FIPS code for home tract
2	Proportion of tract area located 0–75 m from major roadway
3	Proportion of tract area located beyond 75 m of major roadway, up to 200 m
4	Proportion of tract area located beyond 200 m of major roadway
5	Proportion of age group 1 (ages 0–1 years) residing 0–75 m from major roadway
6	Proportion of age group 1 (ages 0–1) residing >75 m of major roadway, up to 200 m
7	Proportion of age group 1 (ages 0–1) residing >200 m of major roadway
8	Proportion of age group 2 (ages 2–4) residing 0–75 m from major roadway
9	Proportion of age group 2 (ages 2–4) residing >75 m of major roadway, up to 200 m
10	Proportion of age group 2 (ages 2–4) residing >200 m of major roadway
11	Proportion of age group 3 (ages 5–15) residing 0–75 m from major roadway
12	Proportion of age group 3 (ages 5–15) residing >75 m of major roadway, up to 200 m
13	Proportion of age group 3 (ages 5–15) residing >200 m of major roadway
14	Proportion of age group 4 (ages 16–17) residing 0–75 m from major roadway
15	Proportion of age group 4 (ages 16–17) residing >75 m of major roadway, up to 200 m
16	Proportion of age group 4 (ages 16–17) residing >200 m of major roadway
17	Proportion of age group 5 (ages 18–64) residing 0–75 m from major roadway
18	Proportion of age group 5 (ages 18–64) residing >75 m of major roadway, up to 200 m
19	Proportion of age group 5 (ages 18–64) residing >200 m of major roadway
20	Proportion of age group 6 (ages 65 and older) residing 0–75 m from major roadway
21	Proportion of age group 6 (ages 65 and older) residing >75 m of major roadway, up to 200 m

Field Number	Description
22	Proportion of age group 6 (ages 65 and older) residing >200 m of major roadway

Note: The position where table values begin and the number of characters per value are not relevant in a tab-delimited format.

01001020100	0.0601	0.0865	0.8534	0.0778	0.1077	0.8145	0.0778	0.1077	0.8145	0.0512
	0.0889	0.8599	0.0903	0.1279	0.7818	0.0412	0.0743	0.8845	0.0607	0.0998
01001020200	0.0526	0.0644	0.8830	0.0562	0.0716	0.8722	0.0562	0.0716	0.8722	0.0421
	0.0552	0.9027	0.0543	0.0742	0.8715	0.0586	0.1170	0.8244	0.0525	0.0752
01001020300	0.0740	0.1116	0.8144	0.0598	0.0978	0.8424	0.0598	0.0978	0.8424	0.0403
	0.0752	0.8845	0.0572	0.0898	0.8530	0.0501	0.0951	0.8548	0.0937	0.1450
01001020400	0.1151	0.1740	0.7109	0.1024	0.1859	0.7117	0.1024	0.1859	0.7117	0.1000
	0.2404	0.6596	0.1250	0.2122	0.6628	0.1107	0.1902	0.6991	0.1082	0.2026
01001020501	0.0800	0.1126	0.8074	0.0322	0.0527	0.9151	0.0322	0.0527	0.9151	0.0240
	0.0453	0.9307	0.0378	0.0791	0.8831	0.0305	0.0568	0.9127	0.0235	0.0460
01001020502	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	1.0000
01001020503	0.0563	0.0932	0.8505	0.0185	0.0312	0.9503	0.0185	0.0312	0.9503	0.0294
	0.0496	0.9210	0.0181	0.0305	0.9514	0.0303	0.0512	0.9185	0.0519	0.0874
01001020600	0.1428	0.2165	0.6407	0.1410	0.2209	0.6381	0.1410	0.2209	0.6381	0.1288
	0.2146	0.6566	0.0941	0.1996	0.7063	0.1301	0.2333	0.6366	0.1211	0.2206
01001020700	0.0524	0.0783	0.8693	0.0544	0.0994	0.8462	0.0544	0.0994	0.8462	0.0490
	0.1127	0.8383	0.0523	0.1090	0.8387	0.0596	0.1172	0.8232	0.0648	0.1203
01001020801	0.0152	0.0248	0.9600	0.0733	0.0347	0.8920	0.0733	0.0347	0.8920	0.0298
	0.0320	0.9382	0.0403	0.0548	0.9049	0.0422	0.0493	0.9085	0.0693	0.0669

Note: Contents wrap around due to space constrictions in this figure.

Figure 13. Excerpt from the HAPEM8 Distance-to-road File

3. Updating Activity Files – “durhw_HAPEM8.txt”, “cluster_HAPEM8.txt”, and “clustertrans_HAPEM8.txt”

We updated the HAPEM activity file (“durhw_HAPEM8.txt” in HAPEM8) to reflect the most recent version of CHAD as of April 20 20. This version of CHAD has nearly four times the number of activity diaries as the version used for HAPEM7. Accordingly, we also updated the HAPEM cluster file (“cluster_HAPEM8.txt” in HAPEM8) and the HAPEM cluster-transition file (“clustertrans_HAPEM8.txt” in HAPEM8).

Starting with HAPEM5, we analyzed CHAD data to create longitudinal activity patterns using Markov chains. In HAPEM8, we refit the Markov chain model to the most recent CHAD to include more activity-pattern studies and, thus, more daily activity patterns. The data analysis groups the daily patterns into one, two, or three activity categories (or

“clusters”) of similar activity patterns for each of 36 combinations of type of day (the three day types of HAPEM: summer weekday, non-summer weekday, and weekend), age (the six age groups discussed in this memo), and commuter type (two types: commutes or does not commute). Whether one, two, or three activity clusters are assigned to a day-age-commuter combination depends on the availability of CHAD data. For HAPEM8, 17 day-age-commuter combinations were assigned three clusters, 2 were assigned two clusters, and 17 were assigned one cluster. We defined clusters based on similar times spent in five broad microenvironments (i.e., indoors residence, indoors other, outdoors near-roadway, outdoor other, and in-vehicle).

In HAPEM, for each day-age-commuter combination, one daily activity pattern is randomly selected from all the CHAD data that correspond to that combination. The starting activity category (i.e., for the first day) is selected according to the relative frequencies of each category. The activity category for the second day is selected according to the transition probabilities from the starting category. Transition probabilities are the relative frequencies of each activity category when the same subject was in the starting category on the first day and the given activity category on the next day. The activity category for the third day is selected according to the transition probabilities from the second day’s category. This is repeated for all days in the day type, producing a sequence of daily activity categories. For a given simulated person, each day is assigned an activity pattern representative of the day’s activity category. Once a particular activity pattern is selected as representative of an activity category, that pattern is always used for that category for that simulated person. Further details on the cluster and cluster-transition approach can be found in Appendix A of the HAPEM7 User’s Guide (a 2015 memorandum from ICF to EPA’s Ted Palma and Terri Hollingsworth).

For HAPEM8, we also forced our analysis of CHAD to consider children in the first three age groups (through age 15 years) to never be commuters (even if CHAD has them “working”). This was to better comply with the census-based commuting data (discussed in this memo) where workers start at age 16 years. We had to create “dummy” records in the cluster-transition file for commuting children, since HAPEM8 expects these records to be present in the file even though they are never used by the model. The result is that the “clustrans_HAPEM8.txt” output file has 36 data records, one for each combination of the 6 demographic groups, 3 day types, and 2 commuting categories, and the 9 categories of commuting children under 16 years old are dummy records.

f. Quality Assurance

We ensured that each CHAD record was represented in the activity file and formatted appropriately, including the proper sets of columns for each day-age-commuter

combination. We ensured that the same CHAD records were represented in the cluster file and formatted appropriately. We ensured that the cluster-transition file contained the correct combinations of day-age-commuter and was formatted appropriately.

The HAPEM8 activity and cluster files both contain 178,621 records with the same CHADID on the same record in both files. This was a change for HAPEM8 that allowed us to simplify HAPEM8 algorithms (counterbalanced by more complex code to develop the activity files). By having matching CHADIDs, all diaries are available for use in HAPEM8 and will remain available for future CHAD updates.

g. Content and Format

The HAPEM8 activity data are contained in a fixed-width, space-delimited text file with characteristics shown in Table 7. The file contains 878 columns and a total of 178,621 rows of data with one header row. Each row corresponds to a person-day of activity in CHAD, where the first field contains an identifier for the record, the next 12 fields can be used together to describe the study respondent, and the remaining fields contain duration values for how long the subject spends in each microenvironment, for each hour of a day, and at work versus at home. All values in fields 15–878 are displayed as whole numbers (i.e., whole minutes). In Figure 6 we show the header and first data row of the HAPEM8 activity file. This record is from a white non-Hispanic female from an unspecified county in California. She was a child between 1 and 2 years old (unemployed and non-commuting). This record was from 16 June 1989, which was a summer weekday. She spent most of her day indoors at home, except in the afternoon when she was outdoors for 1 hour total, in a vehicle for 40 minutes total, and in some other indoor location for 35 minutes total.

Table 8. Characteristics of the HAPEM8 Activity File

Variable Number	Variable	Description	Character Start Position on Data Row	Character Length on Data Row ^a
1	CHADID	ID of event in CHAD	1	10
2	ZIP	ZIP code of subject's residence	11	6
3	ST	2-character FIPS code of state where event took place	17	3
4	COU	3-character FIPS code of county where event took place	20	4
5	SEX	Gender of subject (1=female, 2=male, 9=unknown)	24	4
6	RACE	Race of subject (1=white non-Hispanic, 2=black non-Hispanic, 3=Hispanic any race, 4=Asian or other non-Hispanic, 9=unknown)	28	5
7	WORK	Employment status of subject ("Y"=employed, "N"=unemployed, "X"=missing)	33	5
8	YEAR	Year when the event took place	38	5 or 6, depending on the next field
9	MN	Month when the event took place	Field length varies such that the last digit of each month entry lines up	
10	DY	Day of month when event took place	Field length varies such that the last digit of each day entry lines up	
11	AGE	Age of subject (presented to two decimal places)	51	6
12	G	HAPEM8 age group (1-6)	57	3
13	DT	Type of day when the event took place (1=summer weekday, 2=non-summer weekday, 3=weekend)	60	3
14	CT	Commuter status of subject (1=does not commute, 2=commutes)	63	4
15-878	No header text	Duration of event (minutes). There are 864 of these fields, cycling through each of the 18 microenvironments, 24 hours of the day, and 2 commute types. The values are sequenced so that the 18 microenvironment durations for the first hour in the home location come first, followed by the 18 microenvironment durations for the second hour in the home location, and so on, until all the 432 values for the home location are specified. These are followed by the 432 values for the work location.	Field lengths vary such that the last digit of each duration entry lines up down the file	

1. ^a Any unused character space before a number or character and/or between fields consists of blank spaces.

corresponding to each data row in the activity file, plus one header row. The first field contains an identifier for the record, the next three fields together identify the age-day-commuter combination of the event, and the final two fields respectively identify the cluster number of the event and the number of clusters that exist for all records corresponding to the age-day-commuter combination. In Figure 7 we show the first ten data rows of the HAPEM8 cluster file, indicating all are in the first age group, on summer weekdays, for non-commuters, and all are in the first cluster and all belong to only one cluster.

Table 9. Characteristics of the HAPEM8 Cluster File

Variable Number	Variable	Description	Character Start Position on Data Row	Character Length on Data Row ^a
1	g	HAPEM8 age group (1-6)	1	5
2	dt	Type of day when the event took place (1=summer weekday, 2=non-summer weekday, 3=weekend)	6	5
3	ct	Commuting status of subject (1=does not commute, 2=commutes)	11	5
4	chadid	ID of event in CHAD	16	12
5	clus	Cluster category of event	28	5
6	nclus	Number of clusters for the corresponding combination of g, dt, and ct	33	1

2. ^a Any unused character space before a number or character and/or between fields consists of blank spaces.

g	dt	ct	chadid	clus	nclus
1	1	1	CAC01166A	1	1
1	1	1	CAC01251A	1	1
1	1	1	CAC01489A	1	1
1	1	1	CAC01562A	1	1
1	1	1	CAC01568A	1	1
1	1	1	CAC01809A	1	1
1	1	1	CAC01830A	1	1
1	1	1	CAC01982A	1	1
1	1	1	CAC02036A	1	1
1	1	1	CAC02132A	1	1

Figure 15. Excerpt from the HAPEM8 Cluster File

The HAPEM8 activity-cluster-transition data are contained in a fixed-width, space-delimited text file with characteristics shown in Table 9. The file contains 16 columns, with a data row corresponding to each age-day-commuter combination, plus a header row. The first three fields identify the age group, day type, and commuter status, while the fourth field identifies the number of clusters that exist for that age-day-commuter

combination, fields 5–7 contain the cumulative fractions of the combination within each cluster, and the remaining fields identify the cumulative transition probabilities of all possible combinations of the subject’s cluster number on day X and the subject’s cluster number on day X+1. In Figure 8 we show the age group # 1 data rows of the HAPEM8 cluster-transition file. Several of the day-commuter combinations shown in the excerpt (day-type 1 with both commuting status, day-type 2 with commuting, and day-type 3 with commuting) are all in cluster # 1. For day-type 2 non-commuting, about 49% of diaries are in cluster # 1, 89% are in cluster # 2, and all are in cluster # 3. Logically there is a 100% probability of a cluster # 1, 2, or 3 diary transitioning to a cluster # 1, 2, or 3 diary (i.e., values of 1.00000). There are relatively high probabilities of a cluster # 1 diary transitioning to a cluster # 1 diary (0.60714) or a cluster # 2 diary transitioning to a cluster # 1 or 2 diary (0.86207). There are relatively low probabilities of a cluster # 2 or # 3 diary transitioning to a cluster # 1 diary (0.13793 and 0.11111, respectively) or a cluster # 3 diary transitioning to a cluster # 1 or 2 diary (0.33333). For day-type 3 non-commuting, about 55% (0.54806) of diaries are in cluster # 1, 89% (0.88666) are in cluster # 2, and all are in cluster # 3 (1.00000). Logically there is a 100% probability of a cluster # 1, 2, or 3 diary transitioning to a cluster # 1, 2, or 3 diary (i.e., values of 1.00000). There are relatively high probabilities of a cluster # 1 diary transitioning to a cluster # 1 diary (0.90000), a cluster # 2 diary transitioning to a cluster # 1 or 2 diary (0.62500), or a cluster # 3 diary transitioning to a cluster # 1 or 2 diary (0.75000), with a 50% (0.50000) probability of a cluster # 3 diary transitioning to a cluster # 1 diary. There is a relatively low probability of a cluster # 2 diary transitioning to a cluster # 1 diary (0.37500).

Table 10. Characteristics of the HAPEM8 Cluster-transition File

Variable Number	Variable	Description ^a	Character Start Position on Data Row	Character Length on Data Row ^b
1	g	HAPEM8 age group (1-6)	1	4
2	dt	Type of day when the event took place (1=summer weekday, 2=non-summer weekday, 3=weekend)	5	4
3	ct	Commuting status of subject (1=does not commute, 2=commutes)	9	4
4	nclus	Number of clusters for the corresponding combination of g, dt, and ct	13	3
5	clust1	Cumulative fraction of g/dt in cluster # 1	16	8
6	clust2	Cumulative fraction of g/dt in clusters # 1-2	24	8
7	clust3	Cumulative fraction of g/dt in clusters # 1-3	32	8
8	prob11	Cumulative transition probability from cluster # 1 to # 1	40	8

Variable Number	Variable	Description ^a	Character Start Position on Data Row	Character Length on Data Row ^b
9	prob12	Cumulative transition probability from cluster # 1 to clusters # 1-2	48	8
10	prob13	Cumulative transition probability from cluster # 1 to clusters # 1-3	56	8
11	prob21	Cumulative transition probability from cluster # 2 to # 1	64	8
12	prob22	Cumulative transition probability from cluster # 2 to clusters # 1-2	72	8
13	prob23	Cumulative transition probability from cluster # 2 to clusters # 1-3	80	8
14	prob31	Cumulative transition probability from cluster # 3 to # 1	88	8
15	prob32	Cumulative transition probability from cluster # 3 to clusters # 1-2	96	8
16	prob33	Cumulative transition probability from cluster # 3 to clusters # 1-3	104	7

^a For the cluster* fields, if nclus = 1 then clust2 and clust3 = 0 in the file; similarly, if nclus = 2 then clust3 = 0. The same is true for the prob* fields (if nclus = 1 then prob12, prob13, prob21, prob22, prob23, prob31, prob32, and prob33 = 0, and if nclus = 2 then prob13, prob23, prob31, prob32, and prob33 = 0).

^b Any unused character space before a number or character and/or between fields consists of blank spaces.

```

g dt ct nclus clust1 clust2 clust3 prob11 prob12 prob13 prob21 prob22
prob23 prob31 prob32 prob33
1 1 1 1 1.00000 0.00000 0.00000 1.00000 0.00000 0.00000 0.00000 0.00000
0.00000 0.00000 0.00000 0.00000
1 1 2 1 1.00000 0.00000 0.00000 1.00000 0.00000 0.00000 0.00000 0.00000
0.00000 0.00000 0.00000 0.00000
1 2 1 3 0.48625 0.88567 1.00000 0.60714 1.00000 1.00000 0.13793 0.86207
1.00000 0.11111 0.33333 1.00000
1 2 2 1 1.00000 0.00000 0.00000 1.00000 0.00000 0.00000 0.00000 0.00000
0.00000 0.00000 0.00000 0.00000
1 3 1 3 0.54806 0.88666 1.00000 0.90000 1.00000 1.00000 0.37500 0.62500
1.00000 0.50000 0.75000 1.00000
1 3 2 1 1.00000 0.00000 0.00000 1.00000 0.00000 0.00000 0.00000 0.00000
0.00000 0.00000 0.00000 0.00000

```

Note: Contents wrap around due to space constrictions in this figure.

Figure 16. Excerpt from the HAPEM8 Cluster-transition File

3. Updating Source Code

We made several modifications to various source-code modules for HAPEM8. Most modifications were minor and functioned either to ensure proper execution from the command line or to ensure that data-array dimensions were large enough to accommodate the revised default model input data discussed in this memorandum. The changes to the “durav” module were more significant. We describe below the specific changes we made to the specific modules.

“durav_HAPEM8.f90” (compiled into an executable named “durav_HAPEM8.exe”):

- Simplified code since updated activity input files already were sorted consistently.
- The number of code lines now is reduced by more than half.

“indexpop_HAPEM8.f90” (compiled into an executable named “indexpop_HAPEM8.exe”):

- No changes.

“commute_HAPEM8.f90” (compiled into an executable named “commute_HAPEM8.exe”):

- Increased seven array bounds from 80000 to 99000.
- Updated the status for two files to eliminate a compiler warning.

“airqual_HAPEM8.f90” (compiled into an executable named “airqual_HAPEM8.exe”):

- No changes.

“hapem_HAPEM8.f90” (compiled into an executable named “hapem_HAPEM8.exe”):

- Broke down one large seven-dimensional array into six six-dimensional arrays, one for each demographic group.
- Revised the reading of the commuting database file to six times (for six demographic groups) rather than one time overall.

Table E-1. HAPs Assessed in AirToxScreen, with their HAPEM8 HAP Phases and Surrogate Chemical Assignments

Chem.				AirToxScreen					
Name	Boiling Point		Phase ^c	Name	Phase ^c	Matching Chemical Modeled in HAPEM ^d			
	VValue (°C) ^a	Source ^b				PP	NNP	MM-OR	MM-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	BS2TYLHXYP	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.				AirToxScreen					
Name	Boiling Point		Phase ^c	Name	Phase ^c	Matching Chemical Modeled in HAPEM ^d			
	VValue (°C) ^a	Source ^b				PP	NNP	MM-OR	MM-NR
Chromium III	2,672	http://books.google.com/books?id=SFD30BvPBh0C&pg=PA123&lpq=PA123&dq=chromium+III+meltin+g+point&source=bl&ots=upHljDrKMy&sig=dISMKFL5z0sVI0z8Z4NhsFHggE&hl=en&sa=X&ei=4nklVP LvJ4LS8AGbiYD4DA&ved=0CFkQ6AEwCQ#v=onepage&q=chromium%20III%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	DINITOCRES	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.				AirToxScreen					
Name	Boiling Point		Phase ^c	Name	Phase ^c	Matching Chemical Modeled in HAPEM ^d			
	VValue (°C) ^a	Source ^b				PP	NNP	MM-OR	MM-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/noreq/110-00977.pdf&fromPage=GetDoc	G						
Methoxytriglycol	249	http://msdssearch.dow.com/PublishedLiteratureDOWCOM/dh_012d/0901b8038012d976.pdf?filepath=oxysolvents/pdfs/noreq/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/noreq/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

Chem.				AirToxScreen					
Name	Boiling Point		Phase ^c	Name	Phase ^c	Matching Chemical Modeled in HAPEM ^d			
	VValue (°C) ^a	Source ^b				PP	NNP	MM-OR	MM-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.				AirToxScreen					
Name	Boiling Point		Phase ^c	Name	Phase ^c	Matching Chemical Modeled in HAPEM ^d			
	VValue (°C) ^a	Source ^b				PP	NNP	MM-OR	MM-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	PAH_176E4	P	Cr6	Cr6	Ni	Ni
Methylchrysene	449 P	CS	P						
1-Nitropyrene	445 P	CS	P						
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						
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						

Chem.				AirToxScreen					
Name	Boiling Point		Phase ^c	Name	Phase ^c	Matching Chemical Modeled in HAPEM ^d			
	VValue (°C) ^a	Source ^b				PP	NNP	MM-OR	MM-NR
Benzo[g,h,i]Perylene	550	NIH	P						
Benzo[fluoranthenes	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/reg3hwmd/bf-r/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	PNTCHLNZ	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.				AirToxScreen					
Name	Boiling Point		Phase ^c	Name	Phase ^c	Matching Chemical Modeled in HAPEM ^d			
	VValue (°C) ^a	Source ^b				PP	NNP	MM-OR	MM-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	VINYLDICLOR	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 “AirToxScreen” header (note the occasional many-to-one relationship between the “Name” column under “Chem” and the “Name” column under “AirToxScreen”).

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 E-2. Boiling-point Definitions Used to Classify HAPs for HAPEM8 Modeling for the 2020 AirToxScreen

HAPEM8 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 2 February 2023 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.

Table E-3. HAPs Modeled in HAPEM8 for AirToxScreen

AirToxScreen HAP	HAPEM8 HAP Phase ^a	Modeled in HAPEM8 ^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 E-4. Truncations Applied to Exposure Factors for AirToxScreen

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