The HAPEM User's Guide Hazardous Air Pollutant Exposure Model, Version 8

December 2023

This page intentionally left blank.

Contents

Figures

Tables

This page intentionally left blank.

New Features in HAPEM8

The Hazardous Air Pollutant Exposure Model, version 8 (HAPEM8) includes a number of updated features. These updated features better reflect the residential locations, work locations, commuting habits, and activity patterns of the current (2020) U.S. population. They also are designed to provide exposure estimates that better characterize the variability across the population. These updated features are summarized in the list below and detailed in other portions of this User's Guide.

- Data on population, commuting patterns, and residential proximity to major roads have been updated based on information from the 2020 census where possible.

- Activity-pattern data have been updated based on the April 2020 version of CHAD-Master.

This page intentionally left blank.

1.   Introduction

The Hazardous Air Pollutant Exposure Model, version 8 (HAPEM8) User's Guide is designed to assist exposure analysts with running and interpreting results from HAPEM8. Throughout the User's Guide, for easier identification, the input filenames and file types are in italics (usually lowercase), model program names are uppercase underlined, and model variables are in bold italics. When presented, input and output data and program source codes will be presented in a single lined box, indicating that the text inside the box is shown exactly as it exists in its electronic form. In addition, shaded text boxes appear throughout the document providing useful information and tips to users.

Most of the material in this HAPEM8 User's Guide was taken from the HAPEM7 User's Guide.

1.1. Organization of the User's Guide

The User's Guide is organized into six chapters and an appendix. Chapters 1 and 2 provide a general overview of the background functionality of HAPEM, as well as basic instructions for running the model. The remaining chapters are designed to provide the user with more detailed information on the components of HAPEM. These chapters are designed to be easily referenced without requiring the entire document to be read. We suggest, however, that the novice user read all of the chapters at least once to gain a better understanding of HAPEM.

Chapter 1	Introduction. Provides a brief introduction to HAPEM modeling fundamentals, including a brief history of the development of HAPEM.
Chapter 2	Getting Started—An Overview of HAPEM. Provides an overview of the various components of HAPEM and basic information needed to run the model.
Chapter 3	HAPEM Input Files. Provides a description of the format, data, and options for each HAPEM input file.
Chapter 4	HAPEM Output Files. Provides a description of the format and data associated with each HAPEM output file.
Chapter 5	HAPEM Programs. Provides a description of the purpose, operations, inputs, and outputs, including a brief description of the computer code, for each HAPEM computer program.
Chapter 6	References.
Appendix A	A 2023 technical memorandum from ICF to EPA, providing a thorough description of the process of updating the default input files and model source code for HAPEM8.

1.2. Background

The Hazardous Air Pollutant Exposure Model, version 8 (HAPEM8) is a screening-level exposure model appropriate for assessing average long-term inhalation exposures of the general population, or a specific sub-population, over spatial scales ranging from urban[1] to national. HAPEM provides a relatively transparent set of exposure assumptions and approximations, as is appropriate for a screening-level model.

HAPEM uses the general approach of tracking representatives (termed "replicates") of specified demographic groups as they move among indoor and outdoor microenvironments (MEs) and among geographic locations. The estimated HAP concentrations in each ME visited are combined into a time-weighted average concentration, which is assigned to members of the demographic group.

> A microenvironment (**ME**) is a three-dimensional space in which human contact with an environmental pollutant takes place and which can be treated as a well-characterized, relatively homogeneous location with respect to pollutant concentrations for a specified time period.

HAPEM uses four primary sources of information: population data from the U.S. Census Bureau (census), population activity data from the U.S. Environmental Protection Agency (EPA) Consolidated Human Activity Database (CHAD), air quality data, and ME data. These data will be discussed briefly below, and in greater detail later in this User's Guide.

1.2.1. Population Data

The census is the primary source of most population demographic data. The census collects, among other things, information on where people live, their demographic makeup (e.g., age, gender, ethnic group; note that only age currently is used in HAPEM), employment (which is not explicitly used in HAPEM), and commuting behavior. The default population data for HAPEM currently are derived from the 2020 census reported at the spatial resolution of census tracts. Census tracts are small, relatively permanent statistical subdivisions of a county and usually contain between 2,500 and 8,000 residents. The six HAPEM age groups are: 0–1, 2–4, 5–15, 16–17, 18–64, and 65 years and older.

A second type of population data used in HAPEM is an estimate of the fraction of the population of each census tract that lives within certain distances of major roadways. These estimates were derived using geospatial software to perform proximity analyses on roadway location data from the 2022 census TIGER/Line database (see Appendix A for more details on the HAPEM default input files). They are used, in conjunction with the *PROX* factors described below, to account for the enhanced outdoor concentrations of HAPs emitted from onroad vehicles at locations near major roadways, and the associated enhanced indoor concentrations.

1.2.2. Activity Data

HAPEM uses four types of population activity data: activity-pattern data, commuting-flow data, commuting-time data, and commuting-fraction data. Human activity-pattern data are used to determine the frequency and duration of exposure for specific groups within various MEs. Activity-pattern data are taken from demographic surveys compiled in CHAD (Graham et al.

---

[1] Urban refers to a scale that encompasses the size of a large city, generally on the order of tens of kilometers.

2019) of individuals' daily activities, the amount of time spent engaged in those activities, and the locations where the activities occur. The version of CHAD current in April 2020 was used in the current HAPEM. CHAD contains the sequential patterns of activities for each individual diary-day, and each activity event has a corresponding location code so that the ME of each activity event is known. It is composed of over 178,000 person-days of activity-pattern data, including 315 specific activities and 110 specific ME locations, collected and organized from 23 human activity-pattern surveys.

In addition to recording the duration and location of a person's activities, these surveys also collect important demographic information about the person. The demographic information usually includes the person's age, gender, and race/ethnicity group. Most activity-pattern studies also try to collect information on other attributes of a respondent, such as highest level of education completed, number of people in their household, whether the person or anyone in their household is a smoker, employment status, and the number of hours spent outdoors. For the purposes of the HAPEM default files, age is the only CHAD demographic attribute used, although the current HAPEM activity input file includes gender and race/ethnicity. A commuting-status indicator also is included in the current HAPEM activity file and used in the modeling, as is an indication of the day type of the diary-day (HAPEM currently has three day types: summer weekdays, other [non-summer] weekdays, and weekends).

The ME categories currently incorporated into the default population activity file for HAPEM are presented in Table 1-1 (see Appendix A for more details on the HAPEM default input files).

Table 1-1.
HAPEM MEs

ME Number	ME Description	Broader ME (for clustering)	Commuting?
1	Residential	1 Indoors Residence	No
2	School	2 Indoors Other	No
3	Hospital	2 Indoors Other	No
4	Office	2 Indoors Other	No
5	Public Access	2 Indoors Other	No
6	Bar/Restaurant	2 Indoors Other	No
7	Car/Truck	5 In-vehicle	Yes - Private Transit
8	Public Transit	5 In-vehicle	Yes - Public Transit
9	Air Travel	2 Indoors Other	No
10	Waiting Indoors for Public Transit	2 Indoors Other	Yes - Public Transit
11	Waiting Outdoors for Public Transit	3 Outdoors Near-roadway	Yes - Public Transit
12	Motorcycle/Bicycle	3 Outdoors Near-roadway	Yes - Private Transit
13	Ferryboat	4 Outdoors Other	Yes - Public Transit
14	Residential Garage	3 Outdoors Near-roadway	No
15	Outdoors, Near Roadway	3 Outdoors Near-roadway	No
16	Outdoors, Service Station	3 Outdoors Near-roadway	No
17	Outdoors, Parking Garage	3 Outdoors Near-roadway	No
18	Outdoors, Other	4 Outdoors Other	No

Because available activity data are not adequate to estimate the exposure of each individual in a population, HAPEM groups activity-pattern data together for people with similar demographic characteristics that are expected to influence exposure to HAPs (e.g., age and commuting status), and it makes exposure estimates for these groups. The activity profiles for each replicate in an demographic group have an equal chance of being selected from the activity

database (see Section 1.2.5 [Stochastic Elements]). The result is that HAPEM provides a distribution of exposure concentrations for each demographic group in each census tract.

The commuting-flow data contained in the current HAPEM default file were derived by the U.S. Department of Transportation Federal Highway Administration (FHWA) from the 2012–2016 five-year data from the U.S. Census Bureau's American Community Survey (ACS), as part of the Census Transportation Planning Package (CTPP) and commissioned by the American Association of State Highway and Transportation Officials (see the CTPP web site https://ctpp.transportation.org/2012-2016-5-year-ctpp/). The data files specify the number of residents of each census tract that work in that tract and every other tract (i.e., the population associated with each home-tract/work-tract pair). The geographies of the data were for 2010 tracts, so a 2010-to-2020 tract-relationship file from the Census Bureau was used to approximate the data for 2020 tract geographies. For the current HAPEM, the distance between the centroids of the home and work tracts were calculated outside of the CTPP, using the 2020 census gazetteer spatial files and geostpatial algorithms (see Appendix A for more details on the current HAPEM default input files). HAPEM uses these data in coordination with the activity-pattern data to place a replicate who commutes to work either in the home tract or the work tract at each time step.

For each census tract, the current HAPEM default commuting-time file contains the proportion of commuting workers using public transit and the proportion using private transit, and it also contains the average commute time stratified by public or private transit, as derived from the 2016–2020 five-year ACS (see Appendix A for more details on the current HAPEM default input files). These data are combined with data on the centroid-to-centroid distances between tracts (see Section 5.2.5 [HAPEM]) to estimate the commuting time for each commuting replicate.

Data specifying the fraction of each demographic group in each census tract that commutes to work, as contained in the current HAPEM default commuting-fraction file, were derived from the 2016–2020 five-year ACS (see Appendix A for more details on the current HAPEM default input files).

1.2.3. Air-quality Data

Some previous versions of HAPEM relied on measured outdoor HAP concentration data for the exposure calculations. This limited both the extent of the modeling domain and HAPs, because exposures could only be calculated for locations and HAPs with large monitoring networks. Typically, sufficient data were only available for large metropolitan areas and for the criteria pollutants.[2]

HAPEM is able to estimate exposures over the entire US at spatial scales as small as a census tract. In order to preserve any characteristic diurnal patterns in ambient concentrations that might be important in the estimation of population exposure, HAPEM can treat annual-average concentration estimates that are stratified by time of day in the air-quality input file. The time steps in the air-quality data must be an integral factor of the number of time steps in the activity input file (see Section 2.1.2 [The DURAV Program and the *Activity* and *Cluster* Files]). For example, the current HAPEM default activity file contains data in (24) 1-hour time blocks, so an air-quality file used with the default activity file must contain data in (24) 1-hour time blocks, or (12) 2-hour time blocks, or (8) 3-hour time blocks, and so on. The air-quality data are combined

---

[2] Criteria pollutants are those for which a National Ambient Air Quality Standard (NAAQS) has been set. They are ground-level ozone, carbon monoxide, sulfur dioxide, nitrogen dioxide, lead, and particulate matter.

in HAPEM with activity data to estimate exposure concentrations. The air-quality data also can be decomposed to reflect the contributions from various emission sources. The number of sources is a user-specified variable.

HAPEM also is able to incorporate spatial variability of air quality within each census tract. That is, the air quality within a tract is not limited to a single point estimate (diurnally- and source-stratified). Spatial variability may be incorporated in two different ways. One method is to characterize the air quality in a census tract by a set of up to 500 diurnally- and source-stratified values. How HAPEM handles these datasets is explained below in Section 1.2.5 (Stochastic Elements).

When air quality is characterized by a single point estimate (diurnally- and source-stratified), a second method allows the user to specify a scalar factor to be applied to the census-tract air-quality values, with the scalar dependent on the distance of the replicate's residence from a major roadway. This approach also is discussed in Section 1.2.5 (Stochastic Elements).

1.2.4. ME Data

To calculate the exposure concentration for each demographic group, an estimate is required of the concentration in each ME specified by the activity pattern. In HAPEM, these ME concentration estimates are derived from the outdoor-concentration estimate for the census tract and a set of three ME factors: *PEN*, *PROX*, and *ADD*. These respectively account for penetration of outdoor air into the ME, concentration enhancement due to proximity of the ME to the emission source, and emission sources within the ME (note that *ADD* factors are currently set to zero, as discussed in Section 2.1.6 [The <u>HAPEM</u> Program, the ME *Factors* and *Mobiles* Files, and the Activity *Cluster-transition* File]).

The ME factors are entered into the model as data from input files that contain estimates of distributions for *PEN*, *PROX, and ADD* for three phases of HAPs: gases, particles, and HAPs that might be either phase depending on various conditions. The current HAPEM default *PEN* distributions were obtained from an extensive review of literature and databases on indoor/outdoor ratios of HAPs. The current HAPEM default *PROX* distributions for onroad-mobile sources were derived from modeling studies of the concentration gradients of HAPs near major roadways.[3] How the distributions are utilized in HAPEM is discussed below in Section 1.2.5 (Stochastic Elements).

As is the case with all other HAPEM input files, these data can be modified by the user. The ME factors should be updated as needed to reflect current knowledge, as available.

1.2.5. Stochastic Elements

Although it would be difficult to accurately represent the activities of an individual due to day-to-day variation, the general behavior of population groups can be well represented using stochastic processes. This makes it possible for estimates of population exposure to be characterized as distributions rather than point estimates. HAPEM incorporates six stochastic elements, as described below.

---

[3] The default PROX values for other emission source categories are point values of 1.0 (i.e., no concentration enhancement due to proximity), and the default ADD values are point values of 0.0 (i.e., no indoor emission sources).

1.2.5.1.   Commuting Status

The first stochastic element in the construction of a replicate is the determination of the commuting status (yes or no), according to the probabilities specific to census tracts and demographic groups.

1.2.5.2.   Activity Patterns

The second stochastic element is the selection of daily activity patterns to represent the demographic group and commuting status of the replicate. HAPEM estimates long-term-average concentrations, but the available sequences of population activity data are specified for 24-hour periods only. The general approach used by HAPEM for constructing long-term-average activity sequences from short-term records is composed of several steps (see Appendix A for a detailed discussion, which is briefly summarized here). The first is to select three sets of 24-hour activity patterns, where each set is used to construct an average pattern for an individual for one of the three specified HAPEM day types. A set of patterns, rather than a single pattern, is selected for each day type to reflect the day-to-day variability of activity patterns for an individual. How the set of patterns is combined into an average pattern for the day-type is explained later in this section.

Next, the corresponding exposure concentration is calculated for each of the three day-type-average activity patterns. Then, a weighted average of the three exposure concentrations is calculated to represent the annual-average concentration, where the weightings represent the number of days per year for each day type (i.e., 65 for summer weekdays, 196 for other weekdays, and 104 for weekends). This process is repeated for several replicates[4] for each combination of census tract and demographic group, to create a set of annual exposure-concentration estimates for each group in each tract.

To implement this approach, first all the activity-pattern data are grouped according to demographic group, day type, and commuting status. Then, for each group/commuting /day combination, the activity patterns are stratified into one to three categories, based on similarity of time spent in the various MEs, as determined by cluster analysis (see Appendix A for a detailed discussion on clustering).

Transition probabilities between categories are derived from empirical data of sequenced diary records. Given that the first day of a 2-day sequence falls into category X, the transition probabilities specify the relative frequency of the second day falling into each possible category. For example, if half of the 2-day sequences with the first day in category X also have the second day in category X, the X-to-X transition probability would be 0.5.

The HAPEM algorithms construct an average activity pattern for each replicate by randomly selecting one activity pattern from each category and combining them with weighted averaging. The weights represent the relative frequency of days from each category for the replicate represented. To determine the averaging weights to use, the algorithms perform a Markov process based on the category-to-category transition probabilities. For example, suppose the day type is summer weekday. Because there are 65 summer weekdays in a year, 65 random selections are made of categories. The category for the first day is selected randomly from the set of categories using the relative frequency of each category as the probability of selection. The category for the second day is selected according to the transition probabilities from the first day's category. The category for the third day is selected according to the transition probabilities

---

[4] The number of replicates is a user-specified variable.

from the second day's category. This is repeated until 65 category selections are made. The weight given each activity pattern in the averaging process is the number of times its category was selected in the Markov process.

1.2.5.3. Work Tract

Another stochastic process is applied in HAPEM for replicates that commute to work. For those groups, a work census tract is selected at random from the set of work tracts specified for that home tract, using the proportion of workers commuting to each work tract for its selection probability.

1.2.5.4. ME Factors

Another stochastic feature of HAPEM is the ability to characterize ME factors as variable, instead of uniform over the population. That is, three of the four ME factors (*PEN*, *PROX*, and *ADD*) can be represented by probability distributions rather than point estimates.[5] Several distribution types may be used, as discussed in Section 3.10 (ME *Factors* and *Mobiles* Files). For each replicate, a different set of ME factors is randomly selected.

1.2.5.5. Air Quality—General

HAPEM has the ability to characterize outdoor air concentrations as spatially variable within a census tract. It can do this in two different ways. One approach is to characterize the air quality for each tract as a dataset with up to 500 sets of values (i.e., diurnally- and source-stratified). Then, for each replicate, a different set of ambient air concentrations is selected for the home (and work) tract to reflect the spatial variability in air quality within the tract.

1.2.5.6. Air Quality—Onroad Vehicles

When air quality is characterized by a single point estimate (diurnally- and source-stratified), another approach is used to account for enhanced onroad-vehicle-related HAP concentrations in the vicinity of major roadways. To implement this approach, the distance of the replicate's home (and workplace) from a major roadway is randomly selected based on probabilities specific to census tracts and demographic groups. A *PROX* factor is then selected from a distribution and applied to the census-tract air-quality values for onroad-mobile sources, with the distribution dependent on the selected distance.

1.3. Strengths and Limitations of HAPEM

All models have strengths and limitations. Therefore, for each application, it is important to carefully select the model that has the desired attributes. The following sections provide a summary of the strengths and potential limitations of HAPEM. However, this is not an exhaustive list and may not address features important for specific applications of an exposure model.

---

[5] As noted above, in practice the default PROX values for emission source categories other than onroad vehicles are point values of 1.0 (i.e., no concentration enhancement due to proximity), and the default ADD values are point values of 0.0 (i.e., no indoor emission sources). However, HAPEM8 contains the structure to characterize these as distributions if appropriate data are available.

1.3.1. Strengths

One strength of HAPEM is the ability to use air-concentration estimates from modeling, allowing exposure to population groups to be simulated at the census-tract level rather than relying solely data from the limited (in both areal extent and HAPs measured) nationwide network of fixed-site monitors.

Another important feature of HAPEM is its versatility. The model is designed so that input data specific to different applications can be used without having to rewrite the computer source code. This flexibility is possible because most specifications are not "hard wired" into the model's code. Instead, the necessary input data are entered through external databases and the modeling parameters are specified through an external file. This feature allows easier use of new data, or other information (e.g., ME factors) used by the model, as they become available.

Another strength of HAPEM is its ability to estimate the exposures of workers in the geographic area where they work, in addition to the geographic area where they live, since the HAP concentrations in these locations may be very different.

Another important feature of HAPEM is the incorporation of stochastic processes for the selection activity patterns, work census tracts, ambient air quality among locations within a tract, and ME factors, so that more of the variability in the exposure estimates can be captured than simply the variability associated with residential tract.

Exposure assessment with HAPEM has also been facilitated by development of default input files derived from the databases discussed above: national census population and commuting information, CHAD activity data, and variable ME factors for gases, particles and those HAPs that might be either gaseous or particulate depending on conditions.

1.3.2. Limitations

HAPEM calculates long-term average exposure concentrations in order to address exposures to HAPs with carcinogenic and other long-term effects. Thus, HAPEM does not preserve the time-sequence of exposure events when sampling from the time-activity databases. The result is that information used to evaluate possible correlations in exposures to different HAPs due to activities that are related in time is not preserved.

HAPEM only estimates exposures experienced through inhalation. For certain HAPs, inhalation might not be the major route of exposure, and, therefore, HAPEM may underestimate exposures in these instances. Also, although HAPEM is an inhalation-exposure model, it does not include any measures of the ventilation rate associated with an activity, so there is no ability to calculate the potential dose received when engaging in various activities.

Uncertainty in the prediction distributions is not addressed. Some of the uncertainties are as follows.

- The activity-pattern data are limited. Only three of the 23 studies in the version of CHAD used for the current HAPEM were national in scope (with several other studies covering multiple metropolitan areas or state-scale); therefore, the combined dataset does not constitute a representative sample, at least with respect to geographic region.

- Commuting-pattern data address only home-to-work travel. The population not employed outside the home is assumed to always remain in the residential census tract.

Further, although several of the HAPEM MEs account for time spent in travel, the travel is assumed to always occur either in the home or work tract. No provision is made for the possibility of passing through other tracts during travel.

- The ME *PEN* factor distributions incorporated into the current HAPEM were derived from reported measurement studies. The data available were quite limited. As a result, most factors were not derived from a representative sample of measurements, and many were inferred from measurements of different HAPs and/or MEs that would be expected to be similar. In addition, the derivation of the *PEN* factors assumed that measured indoor:outdoor ratios of 1.0 or less indicate the absence of indoor emission sources. Because this assumption is unlikely to be uniformly valid, *PEN* factors are likely to overestimate penetration by some unknown amount.

- The ME *PROX* factor distributions incorporated into the current HAPEM for the onroad-vehicle source category were derived from modeling studies for Portland, Oregon. They are subject to the standard uncertainties of air-dispersion modeling. They also are subject to the uncertainties of extrapolating from the traffic patterns of Portland to other locations.

- Air-quality data from modeling studies are uncertain, due to simplifications incorporated into modeling algorithms and limitations of input data (e.g., emissions, meteorology). Air-quality measurements also are uncertain due to limitations of measurement technology (e.g., minimum detection limits) and unknown representativeness of monitoring locations.

1.4. Applicability

HAPEM is a screening-level exposure model appropriate for assessing average long-term inhalation exposures of the general population, or a specific sub-population, over spatial scales ranging from urban to national. Due to its design features, HAPEM is not appropriate for modeling short-term (e.g., hourly or daily) exposure events, nor should the model be used to assess the exposure of individuals.

The model is designed to look at the "typical" inhalation exposures of different groups, including their variance across the population. However, it should not be used to quantify episodic "high-end" inhalation exposure that results from highly localized HAP concentrations and/or activities that, by their nature, could result in potentially high exposures (e.g., occupational exposures). Furthermore, HAPEM cannot address cumulative exposure from multiple HAPs or HAP mixtures.

1.5. Brief History of the Hazardous Air Pollutant Exposure Model

In 1985, the EPA's Office of Mobile Sources (OMS)[6] developed a model for estimating human exposure to nonreactive pollutants emitted by mobile sources. This model was similar to the probabilistic National Ambient Air Quality Standards Exposure Model (pNEM) in that both simulated the movements of population groups between home and work locations and through

---

[6] The EPA changed this name to the Office of Transportation and Air Quality in 1999.

various MEs. They differed, however, in several respects. The pNEM provided minute-by-minute exposure estimates, which could be averaged over longer time periods, whereas the model now known as HAPEM provided annual-average exposure estimates. The pNEM included stochastic processes for estimating uncertainty and variability, while HAPEM provided only point estimates. HAPEM also included the ability to estimate cancer incidence using risk factors developed by EPA—a capability not available to pNEM.

OMS extended the modeling methodology in 1991 to estimate annual-average carbon monoxide (CO) exposures in urban and rural areas under specified control scenarios. The model was renamed the Hazardous Air Pollutant Exposure Model for Mobile Sources (HAPEM-MS). HAPEM-MS used the estimated annual-average CO exposures to estimate annual-average exposures to various HAPs associated with mobile sources. This was achieved by assuming the annual-average exposure to each HAP was linearly proportional to the annual-average CO exposure. The model was limited by the fact that it could be run only for specified urban areas with ambient fixed-site CO monitors.

Shortly thereafter, EPA's Office of Research and Development (ORD) developed an enhanced version of HAPEM-MS, called HAPEM-MS2. HAPEM-MS2 sub-divided the annual exposures by calendar quarter (i.e., 3-month periods) to more accurately estimate exposures to mobile sources as a function of outdoor air temperature. HAPEM-MS2 also increased the number of MEs from 5 to 37, increased the number of demographic groups from 11 to 23, and increased the size of the activity-pattern database.

In 1996, ORD further enhanced HAPEM by creating another generation of the model called HAPEM-MS3. These enhancements included adding the ability to customize the demographic groups, updating the census data using the 1990 census, and developing an algorithm for estimating ambient impacts in residences with attached garages.

Until the spring of 1998, HAPEM-MS3 could only be run on an EPA mainframe computer. During early model development, use of the mainframe was necessary because the model required the storage of large data files and the calculation of large internal arrays. After 1998, with advances in computing technology, it became possible for HAPEM-MS3 to be executed on a "workstation." To this end, in the spring of 1998, HAPEM-MS3 was migrated (i.e., transferred) to the UNIX operating system on a workstation. During the migration, further enhancements to the model were made, including a new time-activity database derived from CHAD, a new air-quality program that automatically selects air-pollutant monitoring sites, and a more efficient implementation of the commuting algorithm.

Immediately after the release of the UNIX-version of HAPEM-MS3, ORD, in association with the EPA's Office of Air Quality Planning and Standards (OAQPS), again made substantial improvements to the model. The newer model had two distinct improvements over the 1998 UNIX-version. First, the flexibility of the model was expanded to allow the use of modeled air-quality data as well as measured data. This added functionality allowed the second improvement: expanding the areal extent of the model to include the entire contiguous US at the census-tract level. With these improvements, the model was able to directly estimate exposures to HAPs, and hence the model was again renamed by dropping the mobile source (-MS) acronym.

An earlier version of the model, HAPEM4, had other enhancements as well. These included broader flexibility in defining the study area (this can range from a single census tract up to the entire contiguous US), population and commuting data for all census tracts in the country, a

database of (non-variable) ME factors for more than 30 HAPs, stochastic selection of activity data, and the ability to allow the user to change internal modeling parameters such as the number of MEs. EPA used HAPEM4 in its National Air Toxics Assessment (NATA) for 1996—a periodic assessment designed to help assess the prevalence of air toxics in the US and an important part of EPA's Integrated Urban Air Toxics Strategy.

HAPEM5 incorporated additional enhancements. These included the use of variable ME factors and air-quality data that are spatially variable within census tracts. It also contained a more refined approach for extrapolating short-term (24-hour) activity patterns into annual activity patterns, to better reflect the day-to-day variability in an individual's activities. HAPEM5 was applied as part of the NATA for 1999.

HAPEM6 included the ability to account for enhanced onroad-vehicle-related HAP concentrations in the vicinity of major roadways, a more accurate characterization of the fraction of the population of each census tract that commutes to work, and a more accurate estimate of the duration of commuting to work.

HAPEM7 and HAPEM8 are not fundamentally different from HAPEM6. Both include updates to all census- and CHAD-related data in the default input files. The HAPEM7 update included 18 default microenvironments (up from 14 in HAPEM6). HAPEM7 was applied as part of the NATA for 2011, while HAPEM8 was applied as part of the AirToxScreen for 2020 (AirToxScreen—Air Toxics Screening Assessment—is the successor to NATA).

NOTE: HAPEM currently contains enhanced algorithms for estimating exposure concentrations from indoor emission sources. However, the algorithms have undergone only limited testing, and the development is not complete of the databases required to implement these algorithms. Therefore, we do not recommend the use of these algorithms at the present time.

This page intentionally left blank.

2. Getting Started—An Overview of HAPEM

This chapter provides the user the basic information needed to run HAPEM. The topics addressed in this chapter include the functions of the programs that are contained in HAPEM, the contents of the various input and output files, and the meanings of parameter values. The chapter has been separated into the following sections.

Section 2.1   Model Structure. Describes the general structure of HAPEM, the input and output files, and the parameter settings.

Section 2.2   Changing the Parameter Settings. Discusses considerations for changing parameter settings.

Section 2.3   Setting Up a HAPEM Run. Provides instructions for setting up and running HAPEM.

Figure 2-1 presents a graphical overview of HAPEM, including the types of data needed and the types of output produced by the model. The user should refer back to the figure while reading this chapter to understand how all the pieces of the model fit together.

Figure 2-1.
Overview of HAPEM
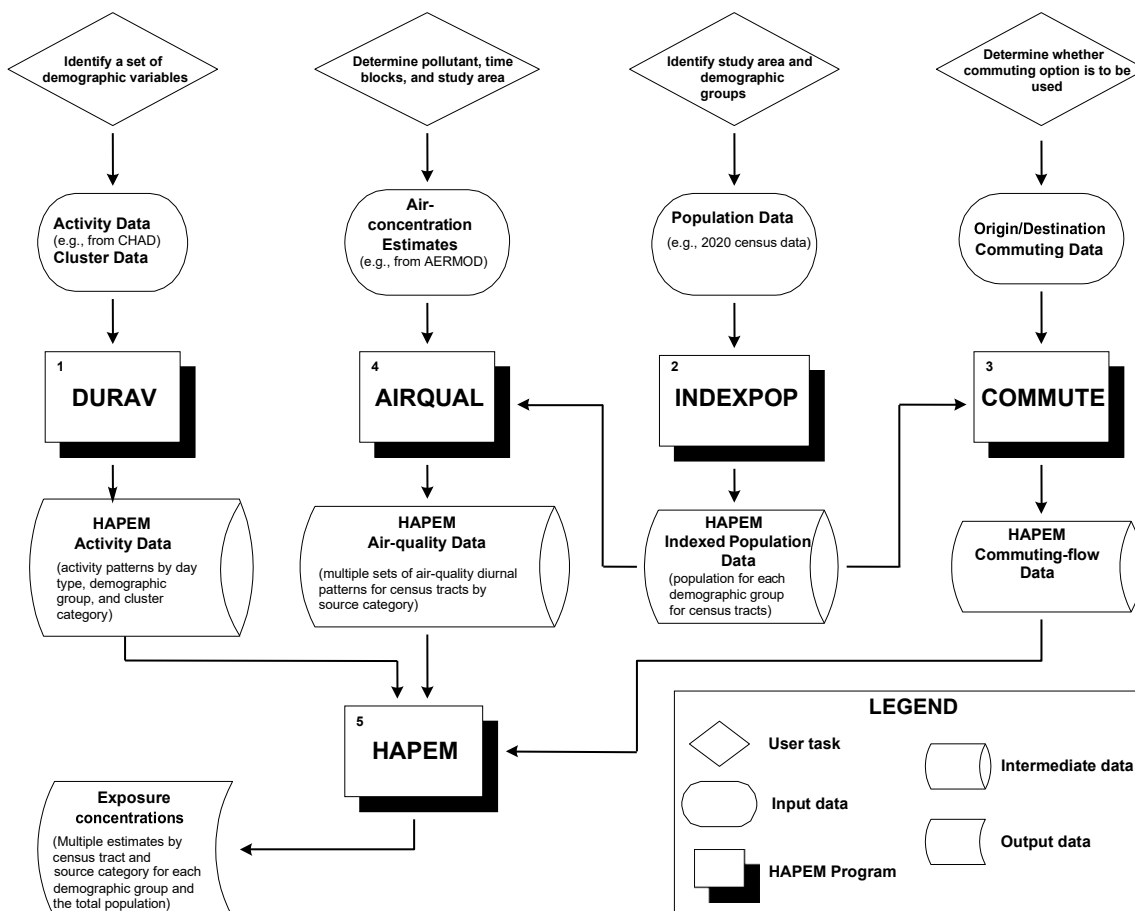
2.1. Model Structure

HAPEM contains five programs. These are listed below.

1. <u>DURAV</u>

2. <u>INDEXPOP</u>

3. <u>COMMUTE</u>

4. <u>AIRQUAL</u>

5. <u>HAPEM</u>

Because several output files of these programs are used as inputs to other programs of the set, it is important to execute them in the order presented. The <u>COMMUTE</u> program is omitted if commuting is not included in the exposure assessment.

For a given modeling domain (e.g., a state, a set of states, the entire US), programs 1–3 need to be executed only once, even if several different air-quality scenarios/HAPs are evaluated. Programs 4–5 need to be executed one time each for each air-quality scenario/HAP. The modeling domain for running programs 4–5 must be included in the modeling domain used for running Programs 1–3, but it may be smaller. For example, if programs 1–3 are run for the entire US, the output files from these runs may then be used by programs 4–5 for evaluating a single state or set of states.

The model programs use 12 groups of user-supplied input data files, and two or more *parameter* files. All are in American Standard Code for Information Exchange (ASCII) format. A *parameter* file identifies the user-supplied input files, the output files available to the user, and specifies the parameter settings for a model run.

2.1.1. *Parameter* Files

The information required in the *parameter* files is presented in Table 2-1 in a way that shows what information is supplied by user-defined files, what is supplied by user-defined parameters, and which model program requires the information. With one exception, noted below, any information in the *parameter* files that is not required will be ignored by the program. This allows wide flexibility in the use of *parameter* files. For example, one approach would be to construct and use a separate *parameter* file for each model program, with each *parameter* file including only the information required by its corresponding program. An alternative approach is to use the same *parameter* file for running more than one program by aggregating all the information needed for each program into the file. We recommend using one *parameter* file for running programs 1–3, and a separate *parameter* file for each set of program 4–5 runs (i.e., for each air-quality scenario). This configuration provides a balance between avoiding errors in duplicating information used by more than one program, and keeping track of the input files used for each air-quality scenario. To avoid using the wrong *parameter* file, a checking feature has been included in programs 1–3 so that they will stop if the keyword **nreplic** (required by

> We recommend that the user prepare a separate *parameter* file for each air-quality scenario/pollutant evaluation. Using distinct files, rather than re-using the same file repeatedly (i.e., by editing it between runs), will assist the user in keeping track of the differences between various model runs, because the *parameter* file serves as a record of the job settings.

programs 4–5) is encountered in the *parameter* file. The name of the *parameter* file is specified on the command line just after the name of the executable file to be run.

Table 2-1.
Keywords for *parameter* files and example filenames

User/Model Defined	Inputs	Outputs
DURAV_HAPEM8.f90		
User-defined files	*activity* file (e.g., *durhw_HAPEM8.txt*) *cluster* file (e.g., *cluster_HAPEM8.txt*)	*log_file.txt* *counter.dat*
User-defined parameters	**nmicro** **hblock** **ngroup** **nblock** **ntype**	
Model-defined files		*durhw_HAPEM8.da* *durhw_HAPEM8.nonzero*
INDEXPOP_HAPEM8.f90		
User-defined files	*population* file (e.g., *population_HAPEM8.txt*) *distance-to-road* file (e.g., *proximity_road_HAPEM8.txt*) *commuting-time* file (e.g., *commute_time_HAPEM8.txt*) *commuting-fraction* file (e.g., *commute_fraction_HAPEM8.txt*) *statefip* file (e.g., *FIPS_StateCrosswalk_HAPEM8.DAT*)	*log_file.txt* *counter.dat*
User-defined parameters	**region1** **region2** **ngroup**	
Model -defined files		*population_HAPEM8.da* *population_HAPEM8_direct.ind* *population_HAPEM8.county_tract_pop_range* *population_HAPEM8.state_county_pop_range* *proximity_road_HAPEM8.STIDX* *proximity_road_HAPEM8.dat* *commute_time_HAPEM8.STIDX* *commute_time_HAPEM8.dat* *commute_fraction_HAPEM8.STIDX* *commute_fraction_HAPEM8.dat*
COMMUTE_HAPEM8.f90		
User-defined files	*commuting* file (e.g., *commute_flow_HAPEM8.txt*) *population* file (e.g., *population_HAPEM8.txt*) *distance-to-road* file (e.g., *proximity_road_HAPEM8.txt*) *commuting-time* file (e.g., *commute_time_HAPEM8.txt*) *commuting-fraction* file (e.g., *commute_fraction_HAPEM8.txt*) *statefip* file (e.g., *FIPS_StateCrosswalk_HAPEM8.DAT*)	*log_file.txt* *counter.dat* *mistract.dat*
User-defined parameters	**region1** **region2** **keep**	
Model-defined files	*population_HAPEM8_direct.ind* *population_HAPEM8.county_tract_pop_range* *population_HAPEM8.state_county_pop_range* *proximity_road_HAPEM8.STIDX* *proximity_road_HAPEM8.dat* *commute_time_HAPEM8.STIDX* *commute_time_HAPEM8.dat* *commute_fraction_HAPEM8.STIDX* *commute_fraction_HAPEM8.dat*	*commute_flow_HAPEM8st_comm1_ fip_ range* *commute_flow_HAPEM8.da* *commute_flow_HAPEM8.ind*

User/Model Defined	Inputs	Outputs
	AIRQUAL_HAPEM8.f90	
User-defined files	*air quality* file *population* file (e.g., *population_HAPEM8.txt*) *distance-to-road* file (e.g., *proximity_road_HAPEM8.txt*) *statefip* file (e.g., *FIPS_StateCrosswalk_HAPEM8.DAT*)	*log_file.txt* *counter.dat* *mistract.dat*
User-defined parameters	**hblock** **ngroup** **region2** **nsource** **region1** **nreplic**	
Model-generated files	*population_HAPEM8.da* *proximity_road_HAPEM8.STIDX* *population_HAPEM8_direct.ind* *proximity_road_HAPEM8.dat*	*HAP.state_air_fip_range* *HAP.state_air1_fip_range* *HAP.state_air2_fip_range* *HAP.da* *HAP.pop_air_da* *HAP.air_da*
	HAPEM_HAPEM8.f90	
User-defined files	*factors* file (e.g., *factors_*_HAPEM8.txt*) *mobiles* file (e.g., *factors_OnroadMobile_*_HAPEM8.txt*) *population* file (e.g., *population_HAPEM8.txt*) *air quality* file *commuting* file (e.g., *commute_flow_HAPEM8.txt*) *activity* file (e.g., *durhw_HAPEM8.txt*) *cluster-transition* file (e.g., *clustrans_HAPEM8.txt*) *Product* files[1] (specify path only) *AutoPduct* file[1]	*log_file.txt* *mistract.dat* *counter.dat* *afile* file (path only)
User-defined parameters	**pollutant** **backg** **Rseed1** **CAS**[1] **sarod** **Rseed2** **unit** **nmicro** **Rseed3** **EPA** **hblock** **B_00** **nmobiles** **ntype** **B_02** **nemicro** **ngroup** **B_05** **nbmicro** **nsource** **B_16** **nvehicles** **nreplic** **B_18** **npublict** **region1** **B_65** **year** **region2**	
Model-generated files	*commute_flow_HAPEM8st_comm1_fip_ range* *commute_flow_HAPEM8.da* *HAP.air_da* *commute_flow_HAPEM8.ind* *HAP.pop_air_da* *activity_CHAD_HAPEM8.da* *HAP.state_air_fip_range* *activity_CHAD_HAPEM8.nonzero* *HAP.state_air1_fip_range* *HAP.da*	

Note: some entries in the above table are presented side-by-side instead of down the page, to save space

[1] A path to one or more indoor-emission-source inputs for the indoor-source algorithms is specified in these statements (with the *AutoPduct* statement including a filename). These algorithms are included in the HAPEM program but have not yet been tested and reviewed. Therefore, they are currently not recommended for use, and instructions for their use are omitted from this document. The Chemical Abstract Service (CAS) registry number is used to identify files for inputs to the HAPEM indoor-source algorithms. To disable the indoor source algorithms, set keyword **CAS** to 99999, and specify any existing path (and file for *AutoPduct*; other than those otherwise specified for input or output for the HAPEM program) since no indoor source files will then actually be utilized by the HAPEM program.

For a record in the *parameter* file to be processed by the program, it must contain an equal sign ("="). Other records in the file are ignored by the program. The left side of the equal sign contains a user-supplied key word or phrase for each user defined file and parameter, as indicated in Table 2-1. Note that the word "file" is part of the file key phrase (e.g., "activity file"). On the right side of the equal sign is specified a full file pathname (all files except the *final exposure* output files and the indoor-source files), a pathname (the *final exposure* output files

and the indoor source files[7]), or a parameter value. As currently configured, HAPEM creates an exposure output file for each state/HAP combination. The names of these files are constructed by the program based on the HAP's SAROAD code and the state's Federal Information Processing Standard (FIPS) code, so that the user need not supply names for these files in the *parameter* file. However, the user must supply the SAROAD code for the HAP in the *parameter* file of the AIRQUAL and HAPEM programs as the value for the parameter **sarod**.

The names of the other user-defined input and output files should consist of two parts, separated by a dot ("."). The part of the name preceding the dot, including the path, is the root and the part following the dot is the extension. Note that the program will not process a record in the *parameter* file that is longer than 120 characters, including the key word/phrase, the equal sign, and the filename/path or parameter value. The number of spaces between the keywords and the "=" signs and between the "=" signs and the filenames are not fixed and therefore can be any reasonable number. Figure 2-2a and Figure 2-2b present example *parameter* files that can be used to run programs 1–3 and 4–5, respectively. Note that the input and output filenames must be listed before the parameter settings.

Figure 2-2a.
Example *parameter* file for running model programs 1–3 (DURAV, INDEXPOP, and COMMUTE)

```
INPUT FILES:
  activity file    = input/activity pattern/durhw_HAPEM8.txt
  cluster file     = input/activity pattern/cluster_HAPEM8.txt
  population file  = input/population/population_HAPEM8.txt
  commuting file   = input/commute/commute_flow_HAPEM8.txt
  CommutTime file  = input/others/commute_time_HAPEM8.txt
  CommutFrac file  = input/others/Commute_fraction_HAPEM8.txt
  DistToRoad file  = input/others/proximity_road_HAPEM8.txt
  statefip file    = input/FIPS_StateCrosswalk_HAPEM8.dat

OUTPUT FILES:
  log file         = output/log_1-3.txt
  counter file     = output/counter.dat
  mistract file    = output/mistract_1-3.dat


PARAMETER SETTINGS:
  keep       = YES
  region1    = 1
  region2    = 53
    nmicro   = 18          ! Number of microenvironments
    nblock   = 24          ! Number of time blocks/day in CHAD file
    hblock   = 8           ! Number of time blocks/day in
    ntype    = 3           ! Number of day types
    ngroup   = 6           ! Number of demographic groups
```

---

[7] Indoor-source algorithms are included in the HAPEM program but have not yet been tested and reviewed. Therefore, they are currently not recommended for use, and instructions for their use are omitted from this document. To disable the indoor-source algorithms, set keyword **CAS** to 99999.

Figure 2-2b.
Example *parameter* file for running model programs 4–5 (<u>AIRQUAL</u> and <u>HAPEM</u>)

```
INPUT FILES:
  activity file    = input/activity pattern/durhw_HAPEM8.txt
  ClusTrans file   = input/Activity Pattern/clustrans_HAPEM8.txt
  population file  = input/population/population_HAPEM8.txt
  commuting file   = input/commute/commute_flow_HAPEM8.txt
  air quality file = input/airqual/2020benzene_AirToxScreen2023.txt
  factors file     = input/factor/factors_gas_HAPEM8.txt
  mobiles file           = input/factor/factors_OnroadMobile_Benzene_HAPEM8.txt
  CommutTime file  = input/others/commute_time_HAPEM8.txt
  DistToRoad file  = input/others/proximity_road_HAPEM8.txt
  statefip file    = input/FIPS_StateCrosswalk_HAPEM8.dat
  product file Pathname    = input/Add/
  AutoPduct file     = input/Add/AutoGarage.txt

Demographic Groups:
  B_00  =  Ages 0-1
  B_02  =  Ages 2-4
  B_05  =  Ages 5-15
  B_16  =  Ages 16-17
  B_18  =  Ages 18-64
  B_65  =  Ages >= 65

OUTPUT FILES:
  log file         = output/benzene/log_4-5_benzene.txt
  counter file     = output/counter_AirToxScreen2023.dat
  mistract file    = output/benzene/mistract_4-5_benzene.dat
  afile            = output/benzene/
  bfile            = output/benzene/
  Keep intermediate files = YES

PARAMETER SETTINGS:
  pollutant  = Benzene
  CAS        = 99999
  units      = ug/m3
  year       = 2020
  region1    = 1
  region2    = 53
  EPA Region = 1
  sarod      = 45201
  Rseed1   = -10    ! Random Seed (Negative) for Selecting Activity Pattern Data
  Rseed2   = -1     ! Random Seed (Negative) for Selecting Micro Factors
  Rseed3   = -1     ! Random Seed (Negative) for Selecting Air Quality Dataset
    backg    = 0.0
    nmicro   = 18    ! Number of microenvironments
    nblock   = 24    ! Number of time blocks/day in CHAD file
    hblock   = 8     ! Number of time blocks/day in airqual file
    ntype    = 3     ! Number of day types
    ngroup   = 6     ! Number of demographic groups
    nsource  = 4     ! Number of source categories
    nmobiles = 3     ! Sequence # of source categories which are on-road mobile
    nbmicro  = 1     ! beginning # of micros which are indoor environments
    nemicro  = 10    ! ending # of micros which are indoor environments
    nvehicles= 7 12  ! Sequence # of micros which are cars/trucks used in commuting
    npublict = 8 10 11 13  ! Sequence # of micros which are public transit
    nreplic  = 30    ! Number of replicates/demo group in output file
```

The model programs also create several intermediate output files that are used as input to other programs in the model set but are not directly useful for the user. The model programs generate the names of the intermediate output files by changing the filename extensions (i.e., the text after the dot) of the input filenames. An example set of filenames, including the intermediate files generated by the programs, is shown in Table 2-1, with example user-defined filenames in parentheses. In the COMMUTE program, two of these intermediate files (*population_HAPEM8.county_tract_pop_range* and *population_HAPEM8.state_county_pop_range)* will be deleted at the end of the program unless the keyword variable *keep* is set to "yes".

Besides the input and output files, the model programs create a set of user-defined diagnostic output files. The main one is a *log* file, which records information about the execution of the programs, including some error messages. Another is a *counter* file that keeps track of the numbers of elements in various processed files, some of which are used by subsequent programs. A third diagnostic file is the *mistract* file which keeps track of census tracts in the *population* file that are not matched by tracts in the *commuting* file, tracts in the *population* file that are not matched by tracts in the *air quality* file, and of tracts in the *commuting* file that are not matched by tracts in the *air quality* file. Only tracts included in both the *population* and *air quality* files are processed by the model since both these pieces of information about a tract (*population* and *air quality*) are needed to make an exposure estimate. If commuting is included in the simulation and the tract is missing from the *commuting* file, it is assumed that all workers residing in that tract stay in the home tract for work.

2.1.2. The DURAV Program and the *Activity* and *Cluster* Files

The DURAV program performs the three main functions listed below.

- It categorizes and groups human activity data extracted from CHAD into demographic groups, day types, commuting status, and cluster categories.

- If a different number of daily time blocks is specified for the analysis than in the activity file, it processes the activity records so that the number of time blocks matches the number specified for the analysis.

- It creates a sequential ASCII file of the activity-pattern records for use by the HAPEM program.

The *activity* file is the primary input file for the DURAV program. The default file, currently *durhw_HAPEM8.txt*, contains data extracted from CHAD describing the amount of time spent in various MEs by individuals. Each record in the *activity* file consists of one person-day (i.e., 1,440 minutes for an individual) of activity data. This information is not an activity sequence; rather, it is the total number of minutes spent in each ME during each block of time throughout the day (i.e., the time increments used per 24-hour period).

For example, in the current HAPEM default *activity* file, *durhw_HAPEM8.txt*, there are 18 MEs, (24) 1-hour time blocks, and 2 exposure districts (home and work), resulting in a total of 864 duration values. The duration in each of the 18 MEs for the first hour comes first in the *activity* file, followed by the 18 durations for the second hour, etc. This pattern is repeated for all 24 hours for the home exposure district, and then for the 24 hours and 18 MEs of the work district (see Appendix A for more details on the current HAPEM default input files).

The number of time blocks in the *activity* file is specified by the user in the *parameter* file of the DURAV, INDEXPOP, and COMMUTE programs as **nblock**. The number of MEs in both the *activity* file and the *factors* and *mobiles* files (discussed below) must be the same and is specified in all the *parameter* files as **nmicro**.[8] The number of duration values in the *activity* file must equal twice the product of the values of the **nmicro** and **nblock** settings in the *parameter* file. The sum of the duration values for each individual profile should always equal 1,440 minutes (i.e., there should be no unaccounted time); otherwise, the program will stop. Each duration must be specified as a whole number (i.e., no decimals; this number can be zero) of minutes in each ME.

The number of time blocks for the analysis is specified in all the *parameter* files as **hblock**. The number may be less than or equal to **nblock**; however, it must be an integral factor of **nblock**, so that the activity time blocks can be combined if necessary to match to match **hblock.** For example, if **nblock** is 24 and **hblock** is set to 8, the DURAV program will combine the (24) 1-hour activity time blocks into (8) 3-hour activity time blocks.

Each record in the *activity* file also contains information about the individual from whose activities the data were derived, so that the records can be classified into demographic groups. The definitions of these groups are part of the DURAV program source code, so that to change the definitions of the groups, the source code must be modified and recompiled. Similarly, the definitions of day types, pertaining to season and day-of-week for categorizing activity patterns, are part of the DURAV program source code. The number of groups is specified as **ngroup** in all the *parameter* files. The number of day types, **ntype**, is specified on the *parameter* files of the DURAV and HAPEM programs.

The cluster category for each CHAD record, identified by CHAD identification code, is specified in the *cluster* file. The current version of the DURAV program divides the activity data into 12 person groups, based on demographic (six categories) and commuting status (yes or no). Activity-pattern data also are separated into three day types: summer weekdays, other weekdays, and weekends. The number of clusters, derived from a statistical cluster-analysis procedure, ranges from one to three, depending on the group and day type (see Appendix A for a detailed discussion). The current HAPEM requires that the CHAD diaries are in the same order in the *activity* and *cluster* files.

2.1.3. The INDEXPOP Program and the *Population*, *Distance-to-road*, *Commuting-time*, and *Commuting-fraction* Files

The INDEXPOP program performs the three main functions listed below.

- It creates a direct-access file of population data to be used in the AIRQUAL program.

- It creates sequential ASCII index files for the population data census tracts, to facilitate file searching in the COMMUTE and AIRQUAL programs.

- It creates direct-access files and associated index files of the data in the *distance-to-road*, *commuting-time*, and *commuting-fraction* files, to be used in the COMMUTE and AIRQUAL programs.

---

[8] As explained in Section 2.1.6 (The HAPEM Program, the ME *Factors* and *Mobiles* Files, and the Activity *Cluster-transition* File), there must be **nmicro** records for each onroad-mobile source category in the mobiles file.

The main input file to the <u>INDEXPOP</u> program is the *population* file, which provides the number of people in each demographic group (defined in the <u>DURAV</u> program source code) for each census tract in the study area. The data must be sorted according to the state, county, and tract FIPS codes. These data are typically obtained from the census surveys (see Appendix A for more details on the current HAPEM default input files).

Other input files with census-tract-specific information about the population, such as the *distance-to-road*, *commuting-time*, and *commuting-fraction* files, also are first processed in this program. The *distance-to-road* file provides information on the fraction of each demographic group in each tract that resides within three different distance categories of major roadways, as well as the fraction of the tract area that is within each distance category. The *commuting-time* file provides information on the average commuting time for commuters residing in each tract. The *commuting-fraction* file provides information on the fraction of workers in each group that resides in each tract and that commutes to work (see Appendix A for more details on the current HAPEM default input files).

2.1.4. The <u>COMMUTE</u> Program and the *Commuting*, *Distance-to-road*, *Commuting-time*, and *Commuting-fraction* Files

The <u>COMMUTE</u> program performs the three main functions listed below.

- It creates a file identifying for each census tract (i.e., home tract), the associated set of work tracts (i.e., tracts in which the residents of the home tract work), the fraction of workers residing in that home tract and working in each work tract, and the normalized centroid-to-centroid distance between home tract and each work tract. The normalized distance is the distance/(average distance). The normalized distance is combined with the average commuting time for the tract to estimate the commuting time for the home-tract/work-tract pair in the <u>HAPEM</u> program.

- It creates a sequential index file to facilitate file searching in the <u>HAPEM</u> program.

- It adds the census-tract-specific information from the distance-to-road, commuting-time, and commuting-fraction direct-access files (created in the <u>INDEXPOP</u> program) to the commuting index file.

The *commuting* file is the main input file to the <u>COMMUTE</u> program. It specifies the fraction of residents of each home census tract that work in that tract and every other tract (i.e., the population associated with each home-tract/work-tract pair), which is typically derived from census data (see Appendix A for more details on the current HAPEM default input files). While there are hundreds of million pairs of tracts nationwide within a reasonable commuting distance of each other, only about 6 million of these pairs have a non-zero flow of commuters. Only those pairs with non-zero flows are included in the *commuting* file.

An important issue pertaining to this commuting data is that workers do not always travel daily between their home and work locations. The larger the distance between home and work, the greater the likelihood that daily commuting does not occur. For example, places of residence in the lower 48 states appear with Alaskan places of work. These workers are almost surely not commuting daily between the continental US and Alaska. To address this issue, the commuting flows were examined as a function of distance. To examine how the decline in commuting flow is affected by distance, researchers plotted the natural log of the natural log of the total flow versus distance. This plot revealed that the ln(ln(total flow)) is nearly linear for distances ranging

from 0 to about 100 km. For distances greater than 100 km, the graph exhibits a decreasingly negative slope with distance (i.e., the curve "flattens out"). These findings suggest that people's "commuting behavior" is fairly consistent, on an aggregate basis, to a distance of approximately 100 km. Then, at greater distances, factors other than daily commuting may become increasingly important. Therefore, in constructing the commuting distance distributions for each census tract, commuting distances greater than 120 km are assumed to be atypical for a daily commuter and the COMMUTE program ignores these longer commutes.

2.1.5. The AIRQUAL Program and the *Air Quality* and *Distance-to-road* Files

The AIRQUAL program performs the four main functions listed below.

- It creates a sequential file of air-quality data to be used in the HAPEM program.

- It determines the number of data records for each census tract in the *air quality* file.

- It creates index files to facilitate file searching in the HAPEM program.

- It adds the tract-specific information from the distance-to-road direct-access file (created in the INDEXPOP program) to the air-quality index files.

The *air quality* file contains the ambient air concentrations that are used by the AIRQUAL program. The file records can have concentration contributions from multiple emission source categories for multiple time blocks for a census tract, as well as a time-invariant location-specific background concentration. There may be multiple such records for each tract, representing spatial variability throughout the tract. The AIRQUAL program requires a separate *air quality* file for each HAP being evaluated. Details about the format of the *air quality* file can be found in Section 3.9 (*Air Quality* File).

The number of outdoor emission-source categories is specified in the *parameter* files of the AIRQUAL and HAPEM programs as **nsource**, and it must match the number in the *factors* file (see Section 2.1.6 [The HAPEM Program, the ME *Factors* and *Mobiles* Files, and the Activity *Cluster-transition* File]). The user specifies the number of time blocks for the analysis in all the *parameter* files as **hblock**. As discussed above, this value must be an integral factor of **nblock**, the number of time blocks in the *activity* file, so that the activity time blocks can be combined if necessary to match to match **hblock.** Similarly, **hblock** may also be greater than or equal to the number of time blocks in the *air quality* file, but it must be an integral multiple of the number of air-quality time blocks, so that the air-quality values can be replicated if necessary to create **hblock** air-quality values. For example, suppose the *air quality* input file has 8 3-hour time blocks per day; if **hblock** is set to 24, the AIRQUAL program will create 24 air-quality time blocks with three replicates of each of the 8 air-quality values.

2.1.6. The HAPEM Program, the ME *Factors* and *Mobiles* Files, and the Activity *Cluster-transition* File

The HAPEM program performs the six main functions listed below.

- For each demographic group in each census tract, it randomly selects **nreplic** sets of ME factors based on the distribution data provided in the *factors* and *mobiles* files. Each set contains a subset of ME factors randomly selected for each of the time blocks (for

the *PEN* and *ADD* factors) or each of the sources (for the *PROX* and *LAG* factors). Each subset contains randomly selected ME factors for each of **nmicro** MEs.

- For each demographic group in each census tract, it randomly selects **nreplic** sets of air-quality data from the datasets available for a tract.

- For each demographic group in each census tract, it creates **nreplic** sets of average activity patterns, where a set contains one average pattern for each day type. An average activity pattern for each day type is calculated as a weighted average of activity patterns randomly selected from each cluster in a group/day-type/commuting-status combination. The weights are determined by the relative frequencies of cluster types randomly selected in a one-stage Markov process,[9] based on the cluster-transition probabilities provided in the *cluster-transition* file.

- For each activity pattern for a commuting demographic group, it randomly selects a work census tract with probability weighting based on the fraction of home-tract residents that work in that tract.

- For each census tract, it estimates the concentration in each ME based on ME factors and outdoor concentrations.

- It combines activity patterns, commuting status, and estimates of ME concentration to calculate **nreplic** annual-average exposure concentrations for each demographic group in each census tract.

The ME *factors* and *mobiles* files provide the factors used to calculate an estimated ME concentration from an outdoor concentration. This methodology allows the user to specify values (distributions or point estimates) for three types of ME factors: penetration factors, proximity factors, and additive factors. These factors are combined with the outdoor concentration estimates according to the following algorithm.

$$\text{ME concentration} = PROX \times PEN \times \text{outdoor concentration} + ADD$$

The outdoor concentration is the sum of the concentration contributions from each outdoor emission-source category and background.

The penetration factor, *PEN*, is an estimate of the ratio of the ME concentration contribution (from a given emission-source category) to the concurrent outdoor-concentration contribution in the immediate vicinity of the ME.

The proximity factor, *PROX*, is an estimate of the ratio of the outdoor concentration in the immediate vicinity of the ME to the outdoor concentration represented by the air-quality data. The air-quality data represent an average over some geographic area (i.e., some subset of a census tract, or an average across the whole tract). For most situations, the current default *factors* file specifies a *PROX* value of 1.0 (i.e., an outdoor-concentration contribution in the immediate vicinity of the tract equal to the tract-average concentration contribution). However, when assessing exposure to motor vehicle emissions, for MEs near roadways (e.g., in-vehicle, residences near major roadways) the HAP concentration contribution in the immediate vicinity of the ME is expected to be higher than the average HAP concentration contribution over the

---

[9] A one-stage Markov process is a sequence of events, such that at every step in the Markov chain the probability distribution for the next event depends on what the current event is.

census tract (i.e., *PROX* is expected to be greater than 1.0), and this is reflected in the current default *factors* and *mobiles* files.

ADD is an additive factor that accounts for emission sources within or near to a ME (i.e., indoor emission sources). Unlike the other two factors, the *ADD* factor is itself a concentration and therefore has units of mass/volume. The actual units used must be the same as those in the *air quality* file.[7]

A fourth factor, *LAG*, is used to account for the possibility of very slow HAP diffusion and penetration, so that the relevant air-concentration value may be from the previous time block. A value of zero for *LAG* indicates no time lag (i.e., use the concurrent air-concentration value; otherwise, the previous time-block value is used).

The *factors* file includes distributions for each of these factors for each ME/emission-source-category combination, with the exception of *PROX* and *LAG* factors for onroad-mobile-source emissions, which are contained in the *mobiles* file with separate distributions specified for three distance-from-roadway categories. As noted above, the number of MEs in the *factors* and *mobiles* files must match the number in the *activity* file (i.e., **nmicro**). Similarly, the number of outdoor-emission source categories (i.e., **nsource**) must match the number in the *air quality* file. The *mobiles* file must contain **nmicro** records for each onroad-mobile source category specified with **nmobiles**.

There are three default *factors* files: one each for gaseous HAPs, particulate HAPs, and HAPs which could be either phase depending on conditions. There are four default *mobiles* files for onroad-mobile sources: one each for benzene, 1,3-butadiene, diesel particulate matter (PM), as well as one for non-specific HAPs.

The default *factors* and *mobiles* files contain ME factors applicable to all the MEs included in the default *activity* file, for **nsource** emission-source categories (e.g., point, non-point, onroad mobile, and nonroad mobile). These category-specific estimates were derived from reported measurement and modeling studies. Because, as noted above, a new approach to evaluating indoor sources is in development, the *ADD* factors are uniformly set to zero. And, due to lack of data, *LAG* is uniformly set to zero. For onroad-mobile sources, the *PROX* and *LAG* values in the *mobiles* files will override those in the *factors* files.

The *cluster-transition* file specifies, for each combination of demographic group and day type, the number of activity patterns in each of two to three clusters (derived from cluster analysis on the activity-pattern data from CHAD), along with the cluster-to-cluster transition probabilities (derived from the transition frequencies for multiple-day activity-pattern records from CHAD; see Appendix A for more details on the current HAPEM default input files). These values are used to create weights for averaging selected activity patterns, one from each cluster, to represent an individual within the demographic group for that day type.

2.1.7. The *Statefip* File

The *statefip* file cross-references the two-character state FIPS codes for each U.S. state (plus Puerto Rico, the U.S. Virgin Islands, and Washington, DC) to its numerical ranking on the list. The numerical rankings range from 1 to 53 in the default file, although the FIPS codes range from "01" to "78", since several possible codes in the sequence are skipped (i.e., not assigned to a state, district, or territory).

The *statefip* file is used in conjunction with the parameters **region1** and **region2** (used in all the *parameter* files to specify the group of states to be included in the analysis according to numerical ranking). For example, setting **region1** to 1 and **region2** to 53 results in assessment of all the states, districts, and territories in the default *statefip* file (assuming the input files contain all the necessary data). Alternatively, setting both **region1** and **region2** to 5 results in assessment of the fifth state only: California with FIPS code "06".

The region range need not be the same for each of the five model programs; the range for each program may be the same as or smaller than the range for the preceding program, where the order of the programs is as specified above. For example, the INDEXPOP and COMMUTE programs could be run for region range 1 to 53, while the AIRQUAL and HAPEM programs are run for a single state.

Note that the **region1** and **region2** parameters specify the states for which the program will look for data in the input files. However, the input files need not contain data for every tract within the specified states. For example, if the *air quality* file contains data for only a subset of census tracts within a state, the AIRQUAL and HAPEM programs will simply make estimates for those tracts, as long as the state or states are specified within the **region1** and **region2** range.

2.1.8. Background Concentration

In addition to estimating exposure-concentration contributions for each emission-source category for which data are provided in the *air quality* file, the HAPEM program also estimates the exposure-concentration contribution from the background outdoor concentration. The background concentration is an estimate of the outdoor concentration that would occur in the absence of any anthropogenic emissions within the modeling domain. It includes concentration contributions from natural sources, re-entrainment, global transport, and other anthropogenic sources outside the modeling domain. This background exposure contribution is added together with the emission-source-category contributions; the total exposure concentration is reported in the exposure output files.

The background concentration is composed of two parts, either or both of which may be used. The first is a uniform background concentration throughout the study area, with the single value is specified as **backg** in the *parameter* file of the AIRQUAL and HAPEM programs. The units of measurement must be the same as those used in the *air quality* file. The second background-concentration specification is a single value for each location specified in the *air quality* file, representing a spatially variable component of the background concentration.

2.1.9. Exposure Output Files

As currently configured, the model creates an exposure output file for each state/HAP combination. The names of these files are constructed by the model based on the HAP SAROAD code (specified by **sarod** in the *parameter* file) and the state FIPS code (as SAROAD.FIPS.dat).

These output files contain **nreplic** records for each combination of census tract and demographic group. Each record identifies the census tract, the group, the number of people to which the exposure estimates apply (i.e., 1/**nreplic** of the population of the group in the tract), and exposure-concentration contribution estimates: one each for the **nsource** outdoor-emission-source categories, one for background, one for each of four indoor-source categories,

and a total of the contributions from all outdoor-emission-source categories, background, and indoor sources.

2.2. Changing the Parameter Settings

HAPEM was designed to be as easy to use as possible. With this in mind, the model's structure is such that, for routine applications, no changes need be made to the model's computer code. For most applications, the user need only supply the model with the appropriately formatted input files and parameter specifications declared in the *parameter* files.

However, there are several changes that users can make to HAPEM to "tailor" the model to their needs. Changes or modifications to the model are most easily accomplished by altering the parameter settings. The following discussion describes those parameters that can be altered.

2.2.1. Changing the Number of MEs

In principle, the model will work with any number of MEs. The number, specified as **nmicro** in all the *parameter* files, must match the number actually used in the *activity* file and the *factors* and *mobiles* files. Definitions of the MEs do not appear anywhere in model code.

The model programs should be able to accommodate anywhere from one up to at least 100 MEs. However, large numbers of MEs could result in input-file line lengths beyond a system's limits (particularly in the case of the *activity* file) if other parameters (such as the number of time blocks) also are set to large values.

2.2.2. Changing the Number and/or Definitions of the Demographic Groups

The number of demographic groups, specified as **ngroup** in all the *parameter* files, must be consistently represented in

- the number of columns in the *population* file,

- the number of columns in the *commuting-fraction* file,

- the number of columns in the *distance-to-road* file, and

- the number of demographic groups specified in the *cluster* and *cluster-transition* files.

The definitions of the groups are listed in the *parameter* file for the AIRQUAL and HAPEM programs so that they can be repeated at the start of the final output file for tracking. This listing in the *parameter* file has no impact on the exposure results.

The six current age groups are as follows, in years.

- 0–1

- 2–4

- 5–15

- 16–17

- 18–64

- 65+

The number of demographic groups is unlimited. However, the user is cautioned that for narrowly defined groups, there might not be enough activity-pattern data to calculate a reliable group average or create meaningful activity-pattern clusters. An extreme example of this is where no activity patterns fit a group's definition, resulting in incorrect exposure calculations (i.e., exposure concentrations equal to zero) for that group.

2.2.3. Changing the Number and/or Definitions of Day Types

Day types are used to guide the selection of the activity patterns. Demographic studies indicate that typical weekday (Monday–Friday) and weekend (Saturday–Sunday) activities differ significantly for most working people and school children. Furthermore, in certain respects, activities in summer (or warm weather) might differ from those in winter (or cold weather), especially for children or other non-workers. Currently, season and day of week are used to determine three day types as

- weekdays in summer (June–August),

- other weekdays, or

- weekends.

In principle, year, month, day, season, temperature, rainfall, other meteorological variables, or even geographical variables could be used to assign day type. However, if there are too many day types, or if they are too narrowly defined, then there may not be enough activity-pattern data fitting the day type definition to allow the determination of a reliable average or to create meaningful activity-pattern clusters. If additional variables are used to define day types, then the programmer is advised to check that there are an adequate number of activity-pattern profiles for each new day type.

2.2.4. Changing the Number and/or Definitions of Time Blocks

The traditional method for running HAPEM has been to use one-hour time increments (referred to as time blocks). However, the current model was designed to allow more flexibility in the selection of time blocks. Time blocks can range between one minute (the finest resolution available for the activity data) and one day, so in principle, there can be any number from one to 1,440 time blocks. In most practical applications, the number of time blocks will be 24 or less. To accommodate the possible adjustment of time blocks from *nblock* to *hblock* as discussed above, the time blocks must each be of equal size.

2.3. Setting Up a HAPEM Run

This section shows how to set up and conduct a simple HAPEM model run. Subsequent sections and chapters provide more detailed explanation about HAPEM's input and output files and the model's programs. The example shown in this section is for a hypothetical HAPEM8 analysis of benzene.

The most important consideration for making a HAPEM run is ensuring that the input files are accurate and correctly formatted. This is the responsibility of the user. To run the model, the user must provide 11–12 data input files (depending on the HAP and source category), the *parameter* files defining the run, and the five executable files for the five programs that make up

the model. The programs can be run consecutively by using a "batch" file, or they can be run independently.

Parameter Files

The *parameter* files for this example, presented above in Figure 2-2a and Figure 2-2b, can be used for running the five executables. The name of the *parameter* file must be specified in the command line immediately after the executable name. As the first three programs in the model sequence (<u>DURAV</u>, <u>INDEXPOP</u>, and <u>COMMUTE</u>) require different inputs from the final two programs (<u>AIRQUAL</u> and <u>HAPEM</u>), it is suggested that two separate *parameter* files be generated for the model sequence of a given simulation or set of simulations. The first *parameter* file (Figure 2-2a) should be used for the first three programs and the second *parameter* file (Figure 2-2b) should be used for the final two programs.

Input/output Files

As seen in Figure 2-2a and Figure 2-2b, the input files (including full pathnames) are identified in the *parameter* files. The input files reside in a subdirectory named "input/". The main exposure output files (*afile*) are sent to a subdirectory named "/output/", along with the diagnostic output files (the *log* file, the *mistract* file, and the *counter* file). When the full pathname is identified for an input or output file, it is not required that it reside in the same subdirectory as the executables.

The names of the input and output files must be identified in the *parameter* files before the parameter settings.

As noted above, an existing pathname should be specified for the *product* files, and the full pathname of any existing file (except other model input or output files) must be specified as the *AutoPduct* file in the *parameter* file used with the model. In the future, these files will be part of the input for evaluating indoor sources, but for now the file will not actually be utilized by the <u>HAPEM</u> program. To disable the indoor-source algorithms, set keyword **CAS** to "99999".

Parameter Settings

The "PARAMETER SETTINGS" in the *parameter* files shown in Figure 2-2a and Figure 2-2b indicate that the region to be modeled is 1 through 53 (all states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands), and the HAP SAROAD code is 45201 (benzene).

The last group of information in the *parameter* file shows that there are 18 MEs to be modeled (**nmicro**). This number of MEs must be consistent with the number of ME factors specified in the *factors* and *mobiles* files (i.e., *factors_*_HAPEM8.txt* and *factors_OnroadMobile_*_HAPEM8.txt*) and the number of duration values specified in the *activity* file (i.e., *durhw_HAPEM8.txt*). The number of time blocks per day in the *activity* file is 24 (**nblock**), but the number of time blocks per day for the analysis is 8 (**hblock**), which is an integral factor of the **nblock** value, as explained above. The number of outdoor emission-source categories is 4 (**nsource**). The data in the *air quality* file (for this example the file is *2020benzene_AirToxScreen2023.txt*) must be consistent with **nsource**, and the number of time blocks must be an integral factor of **hblock**, as explained above. The number of demographic groups (**ngroup**) must be consistent with the groups specified in the <u>DURAV</u> source code and in the *population, commuting-fraction, distance-to-road, cluster, and cluster-transition* files (i.e., *population_HAPEM8.txt, commute_fraction_HAPEM8.txt, proximity_road_HAPEM8.txt,*

---

cluster_HAPEM8.txt, *clustrans_HAPEM8.txt*). The number of replicates to be simulated for each group in each tract is 30 (**nreplic**).

In addition, there are five parameter settings that specify the sequence numbers of particular emission-source categories and ME types that are subject to special treatment in HAPEM. In the example, the sequence number for the onroad-mobile source category in the *air quality* file is 3 (**nmobiles**). The sequence numbers of the indoor MEs (including in-vehicle) in the *factors* and *mobiles* files are 1–10 (**nbmicro** through **nemicro**). There are two MEs for private commuting, with sequence numbers 7 and 12 (**nvehicles**), and there are four MEs for public-transit commuting, with sequence numbers 8, 10, 11, and 13 (**npublict**). (There may be up to 10 values each for **nmobiles**, **nvehicles**, and **npublict**.)

The HAP name (**pollutant**), measurement units (**units**), target year for the analysis (**year**), and the definitions of demographic groups are listed here by the user so that they can be repeated at the beginning of the final output file for tracking. They have no effect on the exposure results.

2.3.1. Running HAPEM as a "Batch" Job

When running HAPEM by submitting batch jobs, each job should be allowed to finish before submitting the next job.

For this example, a simple batch file was written to run the five model programs sequentially, with all five programs residing in the same directory as the batch file. The batch file is shown in Figure 2-3.

**Figure 2-3.
Example "batch" file for
running the five model
programs**

```
durav_HAPEM8.exe p1.txt
indexpop_HAPEM8.exe p1.txt
commute_HAPEM8.exe p1.txt
airqual_HAPEM8.exe p2.txt
HAPEM_HAPEM8.exe p2.txt
```

Because the *parameter* files specify the names or paths of all the input and output files as well as the parameter settings, the batch file simply specifies the order in which the HAPEM executable programs will be run.

2.3.2. Running HAPEM Programs Individually

Any of the model programs can be run individually. The user must ensure that the required input files exist and are in the same location specified in the *parameter* file.

If a user is interested in running the <u>DURAV</u> program (this is typically the first program that is run when doing an exposure analysis), they would go to the subdirectory containing the executable program and type the following command on the command line:

```
durav_HAPEM8.exe p1.txt
```

The other model programs are run similarly.

As indicated in Table 2-1, <u>COMMUTE</u>, <u>AIRQUAL</u> and <u>HAPEM</u> all require input files that are generated from running other model programs. Therefore, if any of these programs is run alone, the user must ensure that the required model-generated input files exist and are in the same subdirectory as the original input file from which their filenames were derived (see Table 2-1). For example, running <u>AIRQUAL</u> requires two files with filenames derived from the *population* file and two files with names derived from the *distance-to-road* file. For this example, these files are *population_HAPEM8.da*, *population_HAPEM8_direct.ind*, *proximity_road_HAPEM8.STIDX*, and *proximity_road_HAPEM8.dat*, with filenames derived from *population_HAPEM8.txt* and *proximity_road_HAPEM8.txt*. Therefore, the *parameter* file for running <u>AIRQUAL</u> and <u>HAPEM</u> must specify the full pathname of the *population* and *distance-to-road* files, and the four intermediate files must exist and reside at those paths.

If the user wishes to run the model for multiple pollutants using the same regions and settings, it should be noted that <u>DURAV</u>, <u>INDEXPOP</u>, and <u>COMMUTE</u> need only be run in sequence one time. <u>AIRQUAL</u> and <u>HAPEM</u> may then be run for multiple pollutants without rerunning <u>DURAV</u>, <u>INDEXPOP</u>, and <u>COMMUTE</u>. This may be accomplished by either running the programs individually, as directed above, or by creating one batch file for the execution of <u>DURAV</u>, <u>INDEXPOP</u>, and <u>COMMUTE</u> and then a batch file for each successive run of <u>AIRQUAL</u> and <u>HAPEM</u>. If the user chooses to do this, it is suggested that upon completing the execution of <u>DURAV</u>, <u>INDEXPOP</u>, and <u>COMMUTE</u> that the user save the *log* and *mistract* files, if needed, both before and after the execution of <u>AIRQUAL</u> and <u>HAPEM</u>, as each successive run will overwrite these files. The *log* and *mistract* files saved before the execution of <u>AIRQUAL</u> and <u>HAPEM</u> apply to each successive run, as they include the information from the first three programs in the sequence.

3. HAPEM Input Files

The model programs use 11–12 user-supplied data input files (depending on the HAP and source category), and two or more *parameter* files. All are in ASCII format. The function of each of the files and their relationship to the structure of HAPEM are discussed in Chapter 2. The reader is referred to that chapter for an overview of HAPEM input files. This chapter summarizes that information and presents the format of each of the user-supplied input files.

The *parameter* files are the central input files for HAPEM simulations, and customized *parameter* files should be prepared for every simulation (or set of simulations). It is best to save the *parameter* file used for each simulation under a unique name, so that the files from earlier simulations are not overwritten. A consistent naming system should be developed to pair each *parameter* file with the output files generated by the simulation or set of simulations. This pairing serves as one form of documentation for the model simulations, so the user can later determine which settings produced which results. Another form of documentation is the repetition of the parameter settings at the start of the final output file.

The remaining filenames used by the model programs are input from the *parameter* files. Thus, the user must check that the *parameter* files refer to the correct filenames before conducting a simulation. Which of the user-supplied files and model-generated files are required for each of the five programs that HAPEM contains is discussed in Chapter 2 and presented in Table 2-1.

As explained in Chapter 2, there are default files available for 11 of the 12 user-supplied input files. These 11 default files are listed below.

Default input files available for HAPEM

population	(e.g., *population_HAPEM8.txt*; national scope)
activity	(e.g., *durhw_HAPEM8.txt*)
cluster	(e.g., *cluster_HAPEM8.txt*)
cluster-transition	(e.g., *clustrans_HAPEM8.txt*)
commuting	(e.g., *commute_flow_HAPEM8.txt*; national scope)
commuting-time	(e.g., *commute_time_HAPEM8.txt*; national scope)
commuting-fraction	(e.g., *commute_fraction_HAPEM8.txt*; national scope)
distance-to-road	(e.g., *proximity_road_HAPEM8.txt*; national scope)
factors	(one each for gaseous HAPs, particulate HAPs, and HAPs which could be either phase depending on conditions; e.g., *factors_gas_HAPEM8.txt*, *factors_particulate_HAPEM8.txt*, and *factors_mixed_HAPEM8.txt*, respectively)
mobiles	(one each for benzene, 1,3-butadiene, diesel PM, and non-specific HAPs; e.g., *factors_OnroadMobile_Benzene_HAPEM8.txt*, *factors_OnroadMobile_Butadiene_HAPEM8.txt*, *factors_OnroadMobile_Diesel_HAPEM8.txt*, and *factors_OnroadMobile_Other_HAPEM8.txt*, respectively)
statefip	(e.g., *FIPS_StateCrosswalk_HAPEM8.DAT*; national scope)

See Appendix A for more details on the current HAPEM default input files. The user may provide his or her own files as replacements for any or all of these files, using the file formats described in this chapter.

The twelfth user-supplied file, the *air quality* file, must be provided by the user with the format described in this chapter.

3.1. *Parameter* Files

The *parameter* files contain the seven types of information listed below for use in HAPEM runs.

- Paths and filenames for the input data files (except the indoor-source files, which currently are not used) and the diagnostic-type output files.

- Paths for the *final exposure*-output files.

- Identification of the set of states (optionally including the District of Columbia, Puerto Rico, and the U.S. Virgin Islands) for the simulation.

- Identification of the HAP, the units of measurement, the target year of the analysis, and the definitions of demographic groups.

- A spatially uniform background concentration.

- Internal parameter settings.

- Seed values for three random-number generators.

All of this information is identified using keywords. The required *parameter*-file information for running each of the five model programs is presented in Table 2-1 of Chapter 2 as user-defined files and user-defined parameters. The contents and format of each of the user-defined files is described below. As explained in Chapter 2, with one exception any information in the *parameter* file in addition to that required by a program will be ignored by the program. (The exception is that programs 1–3 will stop if the keyword **nreplic**—required by the AIRQUAL and HAPEM programs—is encountered in the *parameter* file.) Therefore, although a separate *parameter* file may be used for each program in the model set, it is possible to use the same *parameter* file for running programs 1–3 and another for running programs 4–5 by aggregating all the information needed for each program in the file. The format (including keywords) of a *parameter* file for running the model programs is presented in Figure 2-2a and Figure 2-2b in Chapter 2.

The model programs only scan lines containing an equal ("=") sign. The word or words to the left of the equal sign identify which variable is being set and thus should not be changed. The data to the right of the equal sign are the values or settings that the user selects for the model run. The pathnames should precede the parameter settings in the file. The user can add additional lines (e.g., comments) anywhere to the *parameter* file. It is safest if these lines do not contain an equal sign, which could cause them to be parsed accidentally by the model. To ensure that all the necessary information is specified, it is safest to edit an existing *parameter* file, changing only the comments and the right-hand sides of the equal signs.

3.1.1. Specifying the Location and Names of Input and Output Files

In editing the *parameter* file, the user should typically provide the full pathnames for input and output files (except the indoor-source files [not currently used] and the *final exposure* output files). The names can be up to 100 characters in length and should not use quotation marks to enclose the filenames. If the full pathnames exceed 100 characters, the user may use abbreviated paths (locations of the files relative to the *parameter* file's directory) but must always update these abbreviated paths if the *parameter* file is moved. Some PC systems might require backslashes ("\") in pathnames, rather than the forward slashes ("/") used in UNIX and other systems.

In addition to the input files discussed above, there are three diagnostic output files and a set of final output files (i.e., one file for each state, district, or territory included in the simulation) for which full pathnames must be specified. The diagnostic output files are *log*, *counter*, and *mistract*.

As explained in Chapter 2, HAPEM creates an exposure output file for each state/HAP combination. The names of these files are constructed by the program based on the HAP SAROAD code (specified as the value of the **sarod** parameter) and the state FIPS code. Thus, the pathnames, but not the filenames, for these files must be specified in the *parameter* file.

3.1.2. Identifying the Uniform Component of the Background Concentration

In addition to estimating exposure-concentration contributions for each emission-source category for which data are provided in the *air quality* file, HAPEM also estimates the exposure-concentration contribution from the background outdoor concentration. This background-exposure contribution, of which there are two possible components, is added together with the contributions from the source categories to calculate the total exposure concentration. One component of the background concentration is assumed uniform throughout the study area (i.e., a single value is specified as the **backg** parameter, in the same units as those used in the *air quality* file). The uniform component of the background concentration is an estimate of the outdoor concentration that would occur in the absence of any local anthropogenic emissions. It includes concentration contributions from natural sources, re-entrainment, and/or global transport. The second component of the background concentration is provided in the *air quality* file as a single time-invariant value for each location specified in the *air quality* file. This component typically represents either the impact of anthropogenic emissions released outside of the modeling domain or a combination of those emissions and the outdoor concentration that would occur in the absence of all anthropogenic emissions. In the latter case, the value of **backg** would be set to 0.0, since its constituents would be included in the location-specific background value.

3.1.3. Setting the Internal Parameters

The 12 internal parameter settings (***nmicro***, ***nblock***, ***hblock, ntype***, ***ngroup***, ***nsource***, ***nreplic***, ***nmobiles, nbmicro***, ***nemicro***, ***nvehicles***, and ***npublict***) are specified by the user in one or more of the *parameter* files and must be consistent with the structure of the other input data files. Each of these parameters is defined in the adjacent text box. Thus, if the user wishes to change the number of MEs, for example, the input files that specify MEs must also be altered in a consistent manner.

As explained in Chapter 2, the value of the ***hblock*** parameter (the number of time blocks per day for the analysis) must be selected to meet the criteria listed below.

- The value of ***hblock*** must be an integral factor of ***nblock*** (the number of time blocks per day in the *activity* file) so that the activity time blocks can be combined if necessary to match to match ***hblock.***

- The value of ***hblock*** must be an integral multiple of the number of time blocks per day in the *air quality* file, so that the air-quality values can be replicated if necessary to create ***hblock*** air-quality values.

Internal Parameters	
nmicro	number of MEs in the *activity*, *factors*, and *mobiles* files
nblock	number of time blocks per day in the *activity* file
hblock	number of time blocks per day for the analysis
ntype	number of day types in the <u>DURAV</u> source code
ngroup	number of demographics in the <u>DURAV</u> source code and the *population* file
nsource	number of emission-source categories in the *air quality* file
nreplic	number of replicates to be simulated for each demographic group in each census tract
nmobiles	sequence numbers of up to 10 onroad-mobile emission-source categories in the *air quality* file
nbmicro	sequence number of the first indoor ME (including in-vehicle) in the *activity*, *factors*, and *mobiles* files
nemicro	sequence number of the last indoor ME (including in-vehicle) in the *activity*, *factors*, and *mobiles* files
nvehicles	sequence numbers of up to 10 MEs for private commuting in the *activity*, *factors*, and *mobiles* files (e.g., cars, trucks)
npublict	sequence numbers of up to 10 MEs for public-transit commuting in the *activity*, *factors*, and *mobiles* files (e.g., buses, trains)

3.2. *Activity* File

The *activity* file, the primary input to the <u>DURAV</u> program, contains information on the time individuals spent in various MEs. This information is not presented as an activity sequence; rather, it is presented in the *activity* file as the total time spent in each ME during each block of time and at each location throughout the day.

3.2.1. Variables and Format of the Default File

The first line of the *activity* file is a text header that indicates the order of the variables in each record, although it does not explicitly name the contents that make up the bulk of the file—the minutes spent in each ME for each combination of commute status and day type. The header in the current default *activity* file is as follows.

Header of default *activity* file
(in "wrapped" view)

```
CHADID    ZIP   ST COU SEX RACE WORK YEAR MN DY   AGE   G DT CT
```

Although most of the header record of the *activity* file is not used by the model programs, it provides documentation to inform the user of the meaning of the data fields. The exception is the specification of the number of time blocks per day, ***nblock***, which the DURAV program checks against the value of the ***nblock*** parameter specified in the *parameter* file for consistency. If inconsistent, an error message is sent to the *log* file and the program stops.

Each fixed-width, space-delimited record following the header record consists of one person-day (1,440 minutes) of activity data. The variables in the default *activity* file, extracted from CHAD, are defined in Table 3-1. See Appendix A for more details on the current HAPEM default input files.

Following the commuting indicator is a series of duration values. The values specify the integral number of minutes (possibly zero) spent in each ME/time-block/location combination, where locations are at home or at work. The current default *activity* file, with 18 MEs (listed in Table 1-1), 24 time blocks per day, and two locations, has a total of 864 duration values. These values are sequenced so that the 18 ME durations for the first time block in the home location come first, followed by the 18 ME durations for the second time block 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. An example of a record from the current *activity* file is presented below.

Table 3-1.
Variables in the default *activity* file

Variable	Description
CHADID	9-character string identifying the data record; e.g., the corresponding person-day in the CHAD activity database. This information is used by the DURAV program only to identify faulty records in the diagnostic output files.
ZIP	5-character string identifying the zip code of respondent's residence. If a ZIP code is missing, it is reported as "00000". This information is not used by the current version of the DURAV program.
ST	2-character string identifying the FIPS code of the state where the activities took place. This information is not used by the current version of the DURAV program.
COU	3-character string identifying the FIPS code of the county where the activities took place. This information is not used by the current version of the DURAV program.
SEX	1-character string, indicating the sex of the respondent, with values as follows: "1" = female "2" = male "9" = unknown This information is not used by the current version of the DURAV program.
RACE	1-character string, indicating race/ethnic group of the respondent, with values as follows: "1" = White (non-Hispanic) "2" = Black (non-Hispanic) "3" = Hispanic (any race) "4" = Asian or Other (non-Hispanic) "9" = unknown This information is not used by the current version of the DURAV program.

Variable	Description
WORK	1-character string, indicating employment status of respondent, with values as follows: "Y" = Yes "N" = No "X" = missing This information is not used by the current version of the <u>DURAV</u> program.
YEAR, MN, DY	Numeric variables (four-digit year) that identify the date when the activities actually took place. This information is not used by the current version of the <u>DURAV</u> program.
AGE	Integer indicator of the age of the respondent (missing = -999.00)
G	Integer indicator of the HAPEM age group of the respondent, with values as follows: 1 = 0–1 years old 2 = 2–4 3 = 5–15 4 = 16–17 5 = 18–64 6 = 65+
DT	Integer indicator of day type for classification, with values as follows: 1 = summer weekday 2 = non-summer weekday 3 = weekend
CT	Integer indicator of whether the respondent is a commuter, with values as follows: 0 = no commuting 1 = commuting
DURATION (MICRO,BLOCK,HW)	Duration of event (minutes). There are 864 of these fields, cycling through each of the 18 MEs; cycling through the MEs for the first time block for home locations, and so on through the last time block, then repeating the cycle for work locations..

Example data record from default *activity* file
(in "wrapped" view)

```
CHADID    ZIP   ST COU SEX RACE WORK YEAR MN DY   AGE    G DT CT
CAC01166A 93277 06 000 1   1    N    1989 6  16   1.67   1 1 1   60  0  0  0  0
0   0   0   0   0   0   0   0   0   0   0   0   0   60  0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   60  0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   60  0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   60  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   60  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   60  0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   60  0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   60  0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   60  0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   60  0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   60  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   60  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   60  0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   30  0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   30  30  0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   30  60  0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   45  0   0   0   0   0   15  0   0   0   0   0   0   0
0   0   0   0   60  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   35  0   25  0   0   0   0   0   0   0   0   0   0   0   60  0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   60  0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   60  0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   60  0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

3.2.2. Replacing or Modifying the Default File

If the user wishes to replace or modify the default *activity* file, they must ensure that the following two conditions are met.

- The number of duration values in each record must equal twice the product of the values of ***nmicro*** and ***nblock*** as specified in the *parameter* file.

- The sum of the duration values in each record must total 1,440 minutes (i.e., no time is unaccounted for); otherwise, the <u>DURAV</u> program will stop.

In addition, the user must ensure that the *activity* file is consistent with a feature of the <u>DURAV</u> source code: the record length of the *activity* file (unit 11) is specified in the <u>DURAV</u> program. If the user constructs a replacement activity file with a record length different from that of the default *activity* file, corresponding changes need to be made in the <u>DURAV</u> program.

The variables used by the <u>DURAV</u> program for classifying activity records (i.e., day type, demographic, and commute status), as well as the activity-duration values, are identified by the program by their position in the data record. If the user constructs a replacement activity file with these variables positioned differently, corresponding changes need to be made in <u>DURAV</u> for unit 11 as well as in <u>HAPEM</u> for unit 21 which is produced by <u>DURAV</u>. The number of demogrpahic groups and day types is unlimited. However, the user is cautioned that for narrowly-defined groups and day types, there might not be enough activity-pattern data to calculate a reliable group average or create meaningful activity-pattern clusters. An extreme example of this is where no activity patterns fit a group's definition, resulting in incorrect exposure calculations (i.e., exposure concentrations equal to zero) for that group.

The number, definition, and order of MEs must be the same in both the *activity* file and the *factors* and *mobiles* files (see Section 3.10 [ME *Factors* and *Mobiles* Files]). The number is specified in the *parameter* files as **nmicro**.

The *activity* file is read by the <u>DURAV</u> program, which creates two intermediate output files with the same path and root filename, but with different filename extensions. Thus, the user should NOT name an *activity* file with any of the following filename extensions: *.da* and *.nonzero*.

As with other model input files, the user can add comments or other information after the last data record in the file. To prevent the program from reading these comments as data, a blank line must be inserted after the last data record and before any comments.

3.3. *Cluster* File

This file provides information on demographic group, day type, and cluster type of each complete (i.e., with 1,440 minutes per day) CHAD record in *activity* file. The file is used in <u>DURAV</u> to group CHAD records according to cluster. See Appendix A for more details on the current HAPEM input files.

3.3.1. Variables and Format of the Default File

The first line of the fixed-width, space-delimited *cluster* file is a text header that indicates the order of the variables in each record. The header in the current default *cluster* file is as follows.

Header of default *cluster* file

```
g   dt   ct    chadid    clus nclus
```

Although the header record of the *cluster* file is not used by the model programs, it provides documentation to inform the user of the meaning of the data fields. The first four variabiles have the same meaning as in the *activity* file (see Table 3-1), while "clus" refers to the cluster category number for that record, and "nclus" refers to the total number of cluster categories for that demographic-group/day-type/commuting-status combination.

An extract from the current default *cluster* file is shown below. These cluster categories were determined using cluster analysis, as explained in Appendix A.

Extract from default *cluster* file

```
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
```

3.3.2. Replacing or Modifying the Default File

If the user wishes to replace or modify the default *cluster* file, they must ensure that the file is properly formatted and the following two conditions are met.

- There should be one record for every valid record in the corresponding *activity* file (i.e., one with 1,440 minutes, a demographic specification, and a day-type designation, and a commuting-status specification). Any record in the *activity* file without a corresponding record in the *cluster* file will not be used.

- The records should be sorted in the same way as the *activity* file.

3.4. *Population* File

The *population* file, the primary input to the INDEXPOP program, provides the number of people in each demographic group residing in each census tract of the study area. The data must be sorted according to state FIPS, county FIPS, and tract FIPS. The data are typically derived from the census data. The group definitions are presented in Section 2.2.2 (Changing the Number and/or Definitions of the Demographic Groups). See Appendix A for more details on the current HAPEM default input files.

3.4.1. Variables and Format of the Default File

The *population* file begins with two text header records, followed by one data record for each census tract. The first header record indicates the order of the variables in each of the data records. The first and second header records of the current default *population* file are as follows.

Header of default *population* file

TRACT	B_00	B_02	B_05	B_16	B_18	B_65
	COM	COM	COM	COM	COM	COM

Although the header of the *population* file is not used by the model programs, it provides documentation to inform the user of the meaning of the data fields. Each fixed-width, space-delimited data record following the header consists of a census-tract identifier and a population

value for each of the indicated demographic groups in that tract. The definitions of the data fields in the current default *population* file are presented in Table 3-2.

Table 3-2.
Variables in the default *population* file

Variable	Description
TRACT	11-character string uniquely identifying a U.S. census tract. The first two characters identify the state FIPS code, the next three characters the county FIPS code. The remaining six characters consist of the four-character tract code followed by its two-character extension. If there is no extension for the tract, "00" is used.
B_YY	Integer specifying the number of tract residents with age in category YY. The age category definitions are: 00 = 0–1 years old 02 = 2–4 years old 05 = 5–15 years old 16 = 16–17 years old 18 = 18–64 years old 65 = 65 years or older

An extract from the current default *population* file is presented below.

Extract from default *population* file

```
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
```

3.4.2. Replacing or Modifying the Default File

If the user wishes to replace or modify the default *population* file, they must ensure that the definitions and ordering of the demographic groups in the *population* file corresponds to the ordering in the output file from <u>DURAV</u> that is subsequently used in the <u>HAPEM</u> program. In addition, the user must ensure that the record length is consistent with its specification in the <u>INDEXPOP</u> program (unit 14).

The *population* file is read by the <u>INDEXPOP</u> program, which creates several intermediate output files with the same path and root filename, but with different filename extensions. Thus, the user should NOT name a *population* file with any of the following filename extensions: *.da*, *.county_tract_pop_range*, and *.state_county_pop_range*. There also is an intermediate file with the characters *_direct.ind* attached to the *population* file root name.

As with other model input files, the user can add comments or other information after the last data record in the file. To prevent the program from reading these comments as data, a blank line must be inserted after the last data record and before any comments.

3.5. *Commuting-time* File

The *commuting-time* file provides data for each tract on the proportion of commuting workers who take public transit and private transit, and their respective round-trip average commuting times (minutes). This information is combined with data on the centroid-to-centroid commuting distances for workers in the tract, provided in the *commuting* file described below, to estimate a commuting time for each replicate that is probabilistically selected to commute to work, according to the data provided in the *commuting-fraction* file described below. The HAPEM program then adjusts the selected activity patterns for that replicate to reflect the estimated commuting time (see Section 5.2.5 [HAPEM] for more details on the algorithm).

The *commuting-time* file has no header records, only data records. Each tab-delimited data record contains the following five variables, derived from U.S. Census data for the default file.

Variables in the *commuting-time* file

Home tract ID	(11-character string: state FIPS, county FIPS, and tract FIPS)
Proportion of commuters who travel by public transit	(decimal number)
Proportion of commuters who travel by private vehicle	(decimal number)
Average round-trip commuting time for public-transit commuters	(minutes)
Average round-trip commuting time for private-transit commuters	(minutes)

The default *commuting-time* file is sorted by tract ID, smallest to largest in numerical order. Several example data records from the current default *commuting-time* file are presented below. See Appendix A for more details on the current HAPEM default input files.

Extract from the default *commuting-time* file

```
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
```

3.6. *Commuting-fraction* File

The *commuting-fraction* file provides data for each tract on the proportion of workers in each demographic group that commutes to work. This information is used by the HAPEM program to determine for each replicate in each group whether they commute to work, and therefore, which set of activity patterns should be sampled to represent that replicate. The data in the default *commuting-fraction* file are derived from census data.

The *commuting-fraction* file has no header records, only data records. Each tab-delimited data record contains 13 variables, as follows.

Variables in the *commuting-fraction* file

Home tract ID	(11-character string: state FIPS, county FIPS, and tract FIPS)
Proportion of workers in demographic-group 1 does not commute to work	(decimal number)
Proportion of workers in demographic-group 1 that commutes to work	(decimal number)
Repeat the latter two above for groups 2–6	

The default *commuting-fraction* file is sorted by tract ID, smallest to largest in numerical order. Several example data records from the current default *commuting-fraction* file are presented below. See Appendix A for more details on the current HAPEM default input files.

Extract from the default *commuting-fraction* file
(in "wrapped" view)

```
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
```

3.7. *Distance-to-road* File

The *distance-to-road* file provides data for each tract on the proportion of tract area and the proportion of each demographic group that resides within three distance categories from a major roadway: 0–75 meters, 75–200 meters, and greater than 200 meters. This information is used by the HAPEM program to determine, for each replicate for each ME, the distance from a major roadway and, therefore, which *PROX* factor distributions in the *mobiles* file (described below) to sample from for the onroad-mobile source categories.

The *distance-to-road* file has no header records, only data records. Each tab-delimited data record contains the 22 variables listed below, derived in the current default file using census data as well as census data processed by a third party.

Variables in the *distance-to-road* file

Tract ID	(11-character string: state FIPS, county FIPS, and tract FIPS)
Fractions of tract area within each of three distance categories from a major roadway: 0–75 meters, 75–200 meters, greater than 200 meters	(decimal number)
Fractions of demographic-group 1 that reside within each of three distance categories from a major roadway: 0–75 meters, 75–200 meters, greater than 200 meters	(decimal number)
Repeat the latter one for groups 2–6	

The default *distance-to-road* file is sorted by tract ID, smallest to largest in numerical order. Several example data records from the current default *distance-to-road* file are presented below. See Appendix A for more details on the current HAPEM default input files.

Extract from the default *distance-to-road* file
(in "wrapped" view)

```
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 0.7819 0.0412 0.0743 0.8845 0.0607 0.0998 0.8395
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 0.8243 0.0525 0.0752 0.8722
01001020300 0.0740 0.1116 0.8143 0.0598 0.0978 0.8423 0.0598 0.0978 0.8423 0.0403
0.0752 0.8844 0.0572 0.0898 0.8530 0.0501 0.0951 0.8548 0.0937 0.1450 0.7614
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 0.6892
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 0.9126 0.0235 0.0460 0.9306
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 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 0.8607
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 0.6583
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 0.8149
01001020801 0.0152 0.0248 0.9601 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 0.8638
```

3.8. *Commuting* File

The *commuting* file, the main input file to the <u>COMMUTE</u> program, provides data on the commuting flows (i.e., the proportion of commuters) between pairs of census tracts. The default *commuting* file was derived from census data identifying the tract of work and tract of residence for individuals in all 50 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. Only those home-tract/work-tract pairs with non-zero flows are included in the default *commuting* file.

The *commuting* file has data records with no header records. Each fixed-width, space-delimited data record contains five variables (the first being empty), as follows.

Variables in the *commuting* file

Leading space in file	
Home tract ID	(11-character string: state FIPS, county FIPS, and tract FIPS)
Work tract ID	(11-character string: state FIPS, county FIPS, and tract FIPS)
Distance apart in kilometers	(decimal number)
Fraction of workers in the commuting flow	(decimal number; sums to 1 across all instances of a home tract)

An extract from the current default *commuting* file is presented below. See Appendix A for more details on the current HAPEM default input files.

Extract from the default *commuting* file

```
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
```

3.8.1. Replacing or Modifying the Default File

The *commuting* file is read by the COMMUTE program, which creates several intermediate output files with the same path and root filename, but with different filename extensions. Thus, the user should NOT name a *commuting* file with any of the following filename extensions: *.da*, *.ind*, and *.st_comm1_fip_range*.

As with other model input files, the user can add comments or other information after the last data record in the file. To prevent the program reading these comments as data, a blank line must be inserted after the last data record and before any comments.

3.9. *Air Quality* File

The *air quality* file contains the ambient-air concentrations that are used by the AIRQUAL program. AIRQUAL requires a separate *air quality* file for each HAP being evaluated.

The *air quality* file must begin with at least one text header record, followed by one or more data records for each census tract to be evaluated. The required text header is used by the AIRQUAL program to determine the number of time blocks per day (of equal size) in the air-quality data. This value should be indicated immediately following the last instance of the

character string "block". For example, the sixth header record of an AERMOD-derived *air quality* file used for the recent AirToxScreen analysis, which indicates the order of the variables in each of the data records, is as follows.

Example header record from an *air quality* file
(in "wrapped" view)

```
FIPS    Tract   Backgrd_Conc    Conc_block1     Conc_block2     Conc_block3
Conc_block4     Conc_block5     Conc_block6     Conc_block7     Conc_block8
Conc_block1     Conc_block2     Conc_block3     Conc_block4     Conc_block5
Conc_block6     Conc_block7     Conc_block8     Conc_block1     Conc_block2
Conc_block3     Conc_block4     Conc_block5     Conc_block6     Conc_block7
Conc_block8     Conc_block1     Conc_block2     Conc_block3     Conc_block4
Conc_block5     Conc_block6     Conc_block7     Conc_block8
```

For this example, AIRQUAL will interpret the number of time blocks per day as 8. As noted elsewhere, the number of time blocks per day in the *air quality* file must be an integral factor of **hblock**, the number of time blocks per day for the analysis as specified in the *parameter* file; otherwise, the program will stop. If the number of time blocks per day in the *air quality* file is less than **hblock**, AIRQUAL will replicate the values to create **hblock** concentration values. The other information in this header record and all other header records is ignored by AIRQUAL.

After the required header information is found, AIRQUAL identifies data records by finding a numerical digit in the fourth data field. To avoid a mistaken identification, the user should ensure that header records do not contain a numerical digit in the fourth data field.

The fields in the data records are defined as follows.

Variables in the *air quality* file

Leading spaces in file	
Home tract ID	(11-character string: state FIPS, county FIPS, and tract FIPS)
Space-delimited air concentrations for spatially-variable (but temporally constant) background concentration	(decimal number; optionally in exponential format, e.g., X.XXE-XX)
Space-delimited air concentrations for each combination of emission-source category and time block	(decimal numbers; optionally in exponential format, e.g., X.XXE-XX)

The number of non-background concentration values in each data record must equal the product of the number of outdoor-emission-source categories (i.e., the value of **nsource** in the *parameter* file) and the number of time blocks per day, as indicated in the text header record discussed above. The values are ordered beginning with the first time block of the first emission source, followed by the second time block of the first emission source, and so on. An example data record is presented below for **nsource** = 4.

Example data record from an *air quality* file
(in "wrapped" view)

```
        01001          020100 0.00E+00       3.18E-03       3.17E-03       2.13E-03
    1.02E-03      8.42E-04      1.85E-03      3.00E-03      3.07E-03      2.98E-02
    3.44E-02      3.63E-02      2.40E-02      1.83E-02      5.47E-02      8.24E-02
    4.59E-02      2.13E-02      3.40E-02      8.71E-02      4.99E-02      3.04E-02
    7.28E-02      7.77E-02      4.01E-02      1.58E-02      1.79E-02      1.89E-02
    1.16E-02      9.37E-03      2.20E-02      2.74E-02      2.01E-02
```

The *air quality* file is read by the AIRQUAL program, which creates several intermediate output files with the same path and root filename, but with different filename extensions. Thus, the user should NOT name an *air quality* file with any of the following filename extensions: *.da*, *.air_da*, *.pop_air_da*, *.state_air_fip_range*, *.state_air1_fip_range*, and *.state_air2_fip_range*.

As with other model input files, the user can add comments or other information after the last data record in the file. To prevent the program reading these comments as data, a blank line must be inserted after the last data record and before any comments.

3.10. ME *Factors* and *Mobiles* Files

The ME *factors* and *mobiles* files provide probability distributions for the factors used to calculate an estimated ME concentration from an outdoor concentration. The files contain probability distributions for three of the four factors for each ME, and a single value for the fourth factor. These factors are used in the HAPEM algorithm, as follows.

$$ME(m, c, t, s, d) = PROX(m, s, d) \times PEN(m, t) \times AMB(c, t_{LAG(m)}, s)$$

$$ME(m, t, i) = ADD(m, t)$$

$$ME(m, c, t, b) = PROX(m., s) \times PEN(m, t) \times [bckgd\_u + bckgd\_v(c)]$$

$$ME(m, c, t. d) = \sum_s ME(m, c, t, s, d) + ME(m, t, i) + ME(m, c, t, b)$$

where:

ME(m,c,t,s,d):	concentration in ME *m* located in census tract *c* at time *t* due to source category *s* and at distance from source *d*,
PROX(m,s,d):	proximity factor for ME *m*, source category *s*, and distance from source *d* (defined below),
PEN(m,t):	penetration factor for ME *m* at time *t* (defined below),
AMB(c,t,s):	ambient concentration for census tract *c* at time *t* for source category *s* from the *air quality* file,
$t_{LAG(m)}$:	time *t* if *LAG(m)* = 0; time *t-1*, otherwise,
ME(m,t,i):	concentration in ME *m* at time *t* due to indoor sources,
ADD(m,t):	additive factor for ME *m* at time *t* (defined below),
ME(m,c,t,b):	concentration in ME *m* located in census tract *c* at time *t*, due to the background concentration,

      bckgd_u: uniform component of ambient background concentration,

      bckgd_v(c):   spatially-variable component of background concentration, and

      ME(m,c,t,d): total concentration in ME *m* located in census tract *c* at time *t* at distance from source *d*.

The penetration factor, *PEN*, is an estimate of the ratio of the ME-concentration contribution (from a given emission-source category) to the concurrent outdoor-concentration contribution in the immediate vicinity of the ME. That is,

$$PEN = \frac{\text{indoor or in-vehicle ME concentration}}{\text{outdoor concentration in immediate vicinity of indoor or in-vehicle ME}}$$

The proximity factor, *PROX*, is an estimate of the ratio of the outdoor concentration in the immediate vicinity of the ME (or in the ME for outdoor MEs) to the outdoor concentration represented by the air-quality data. That is,

$$PROX_I = \frac{\text{outdoor concentration in immediate vicinity of indoor or in-vehicle ME}}{\text{concentration from } air\,quality \text{ file}}$$

$$PROX_O = \frac{\text{outdoor ME concentration}}{\text{concentration from } air\,quality \text{ file}}$$

Air-quality data used in the model typically represent a spatial average over the census tract. For most MEs, the default *factors* file specifies a *PROX* value of 1.0 (i.e., an outdoor-concentration contribution in the immediate vicinity of the ME equal to the spatial-average contribution over the census tract). However, when assessing exposure to motor-vehicle emissions for MEs near roadways (e.g., in-vehicle, indoor MEs situated near roadways), the HAP-concentration contribution in the immediate vicinity of the ME is expected to be higher than the spatial-average HAP-concentration contribution of the census tract (i.e., *PROX* is expected to be greater than 1.0). This is because the concentration gradient near roadways tends to be relatively steep. This condition for onroad-mobile emissions is reflected in the default *mobiles* file, which contains *PROX* distributions and *LAG* factors for onroad-mobile emissions.

ADD is an additive factor that accounts for emission sources within or near to an ME (i.e., indoor-emission sources). Unlike the other two factors, the *ADD* factor is itself a concentration and therefore has units of mass/volume. The actual units used must be the same as those in the *air quality* file.[10]

LAG is used to account for the possibility of very slow HAP diffusion and penetration, so that the relevant air-quality concentration value may be from the previous time block. A value of zero for *LAG* indicates no time lag (i.e., use the concurrent air-quality value); otherwise, the previous time-block value is used. Due to lack of sufficient data to make estimates for *LAG*, the default file contains a uniform value of zero for all MEs.

---

[10] A database of distributions of indoor-source-concentration contributions for several indoor-source categories and subcategories is currently under development. The current version of the HAPEM program contains untested algorithms to utilize the developing database. Therefore, it is currently recommended that indoor sources be omitted from HAPEM applications until the database and algorithms have been tested and reviewed. To disable the indoor-source algorithms, set keyword **CAS** to 99999.

The *factors* and *mobiles* files have no header records. The *factors* file contains a set of records (one for each ME) for each outdoor-source category being modeled (the number identified as **nsource**), in the same order as the source categories are specified in the *air quality* file. The *mobiles* file contains a set of records (one for each ME) for the onroad-mobile source category identified with **nmobiles** and for each distance-from-road category.[11] The MEs must be the same number, definition, and order as the MEs in the *activity* file. The files are read in free format, once for each ME, with fields as specified in Table 3-3a and Table 3-3b. All values are decimal numbers.

Table 3-3a.
Format for the *factors* file

ME Factor	Field Num.	Parameter		
(N/A)	1	Number of ME (1–18)		
PEN	2	Indicate Distribution Type 1 - Normal 2 - Lognormal 3 - Uniform 4 - Triangular 5 - Dataset		
	3	Distribution Type Normal Lognormal Uniform Triangular Dataset	→	Indicate Parameter Mean Mean Minimum Minimum Number of data points
	4	Distribution Type Normal Lognormal Uniform Triangular Dataset	→	Indicate Parameter Standard deviation Standard deviation Maximum Maximum First data point in the set
	5	Distribution Type Normal Lognormal Triangular Dataset	→	Indicate Parameter 0 (always) 0 (always) Mode Second data point in the set
	6	Distribution Type Normal Lognormal Dataset	→	Indicate Parameter Lower bound (optional) Lower bound (optional) Third data point in the set
	7	Distribution Type Normal Lognormal Dataset	→	Indicate Parameter Upper bound (optional) Upper bound (optional) Fourth data point in the set
	8	Distribution Type Dataset	→	Indicate Parameter Fifth data point in the set
	9	Distribution Type Dataset	→	Indicate Parameter Sixth data point in the set
	10	Distribution Type Dataset	→	Indicate Parameter Seventh data point in the set

---

[11] Note that a *PROX*-factor distribution is specified in the *factors* file for the onroad-mobile source category as a place-holder and such values should be set to 1. The *PROX*-factor distributions in the *mobiles* file are then multiplied by the distributions from the *factors* file.

ME Factor	Field Num.	Parameter		
	11	Distribution Type Dataset	→	Indicate Parameter Eighth data point in the set
	12	Distribution Type Dataset	→	Indicate Parameter Ninth data point in the set
	13	Distribution Type Dataset	→	Indicate Parameter Tenth data point in the set
ADD	14–25	Repeat fields 2–13 for additive factor		
PROX **Source 1** **Source 2** **Source 3** **Source 4**	 26–37 39–50 52–63 65–76	 Repeat fields 2–13 for proximity factor Repeat fields 2–13 for proximity factor Repeat fields 2–13 for proximity factor Repeat fields 2–13 for proximity factor		
LAG **Source 1** **Source 2** **Source 3** **Source 4**	 38 51 64 77	 0 (no lag) or 1 (lag of 1 time block) 0 (no lag) or 1 (lag of 1 time block) 0 (no lag) or 1 (lag of 1 time block) 0 (no lag) or 1 (lag of 1 time block)		

Table 3-3b.
Format for the *mobiles* file (one onroad-mobile source category)

ME Factor	Field Num.	Parameter		
(N/A)	1	Number of ME (1–18)		
PROX for Onroad-mobile Source Category: **Distance-from-source Category 1**	2	Indicate Distribution Type 1 - Normal 2 - Lognormal 3 - Uniform 4 - Triangular 5 - Dataset		
	3	Distribution Type Normal Lognormal Uniform Triangular Dataset	→	Indicate Parameter Mean Mean Minimum Minimum Number of data points
	4	Distribution Type Normal Lognormal Uniform Triangular Dataset	→	Indicate Parameter Standard deviation Standard deviation Maximum Maximum First data point in the set
	5	Distribution Type Normal Lognormal Triangular Dataset	→	Indicate Parameter 0 (always) 0 (always) Mode Second data point in the set
	6	Distribution Type Normal Lognormal Dataset	→	Indicate Parameter Lower bound (optional) Lower bound (optional) Third data point in the set
	7	Distribution Type Normal Lognormal Dataset	→	Indicate Parameter Upper bound (optional) Upper bound (optional) Fourth data point in the set
	8	Distribution Type Dataset	→	Indicate Parameter Fifth data point in the set

ME Factor	Field Num.	Parameter		
	9	Distribution Type Dataset	→	Indicate Parameter Sixth data point in the set
	10	Distribution Type Dataset	→	Indicate Parameter Seventh data point in the set
	11	Distribution Type Dataset	→	Indicate Parameter Eighth data point in the set
	12	Distribution Type Dataset	→	Indicate Parameter Ninth data point in the set
	13	Distribution Type Dataset	→	Indicate Parameter Tenth data point in the set
LAG for Onroad-mobile Source Category: **Distance-from-source Category 1**	14	0 (no lag) or 1 (lag of 1 time block)		
Distance-from-source Category 2	15–27	Repeat fields 2–14		
Distance-from-source Category 3	28–40	Repeat fields 2–14		

The fields in the *factors* file include *PROX* distributions (one per ME and source category), *PEN* distributions (one per ME), *ADD* distributions (one per ME), and *LAG* factors (one per ME; *LAG* factors are either 0 or 1).

The fields in the *mobiles* file include distributions of *PROX* and *LAG* factors for onroad-mobile source category identified with **nmobiles**. Distributions of *PROX* factors in the *mobiles* files are stratified for each of three distance-from-source categories: 0–75 meters, 75–200 meters, and beyond 200 meters, and this information is combined with the data in the *distance-to-road* file in the HAPEM program to determine from which probability distribution the *PROX* factor should be selected for a given tract/ME combination (see Section 5.2.5 [HAPEM] for more details). The distributions in the *mobiles* file override those in the *factors* file for the onroad-mobile source category identified with **nmobiles**.

Distributions can take any of five different forms: normal, lognormal, uniform, triangular, or dataset. The dataset is composed of up to 10 values, each of which is selected with equal probability. The parameters that need to be specified for each type of distribution are listed below.

Distribution types used in the *factors* and *mobiles* files

Normal	arithmetic mean, arithmetic standard deviation, lower bound (optional), upper bound (optional) [Note: If both the lower and upper bounds are set to 0.0, then the distribution is sampled as if unbounded]
Lognormal	geometric mean, geometric standard deviation, lower bound (optional), upper bound (optional) [Note: If both the lower and upper bounds are set to 0.0, then the distribution is sampled as if unbounded]
Uniform	minimum, maximum
Triangular	minimum, maximum, mode
Dataset	number of data values in the set (1–10), each value

For HAPEM, default *factors* files are provided for each of three phases of HAPs: gaseous, particulate, and HAPs that might be either phase depending on various conditions. Default *mobiles* files are provided for benzene 1,3-butadiene, diesel PM, and non-specific HAPs (formatted for a single onroad-mobile source category). As noted above, because a new

approach to evaluating indoor sources is in development, the default *ADD* factors are uniformly set to zero. Due to lack of data, default *LAG* factors are uniformly set to zero. Excerpts from the default *factors* and *mobiles* files for gaseous HAPs and non-specific HAPs, respectively, are presented below. See Appendix A for more details on the current HAPEM default input files.

As with other model input files, the user can add comments or other information after the last data record in the file. In this case a blank line need not be inserted after the last data record before the comments.

Extract from a default *factors* file
(in "wrapped" view)

```
1       1       5       3       0.8     0.8     1       0       0       0       0       0       0
        0       5       1       0       0       0       0       0       0       0       0       0
        0       5       1       1       0       0       0       0       0       0       0       0
        0       0       5       1       1       0       0       0       0       0       0       0
        0       0       0       5       1       1       0       0       0       0       0       0
        0       0       0       0       5       1       1       0       0       0       0       0
        0       0       0       0       0
2       5       5       0.33    0.67    0.71    1       1       0       0       0       0       0
        5       1       0       0       0       0       0       0       0       0       0       0
        5       1       1       0       0       0       0       0       0       0       0       0
        0       5       1       1       0       0       0       0       0       0       0       0
        0       0       5       1       1       0       0       0       0       0       0       0
        0       0       0       5       1       1       0       0       0       0       0       0
        0       0       0       0
3       5       5       0.33    0.67    0.71    1       1       0       0       0       0       0
        5       1       0       0       0       0       0       0       0       0       0       0
        5       1       1       0       0       0       0       0       0       0       0       0
        0       5       1       1       0       0       0       0       0       0       0       0
        0       0       5       1       1       0       0       0       0       0       0       0
        0       0       0       5       1       1       0       0       0       0       0       0
        0       0       0       0
```

Extract from a default *mobiles* file
(in "wrapped" view)

```
1   2   2.477 2.0477 0      1       8.0532 0   0       0       0       0       0       0
    2   1.6113 1.9292 0     1       4.7492 0   0       0       0       0       0       0
    5   1      1      0      0       0      0   0       0       0       0       0       0
2   2   2.477 2.0477 0      1       8.0532 0   0       0       0       0       0       0
    2   1.6113 1.9292 0     1       4.7492 0   0       0       0       0       0       0
    5   1      1      0      0       0      0   0       0       0       0       0       0
3   2   2.477 2.0477 0      1       8.0532 0   0       0       0       0       0       0
    2   1.6113 1.9292 0     1       4.7492 0   0       0       0       0       0       0
    5   1      1      0      0       0      0   0       0       0       0       0       0
```

3.11.   *Cluster-transition* File

The *cluster-transition* file specifies, for each combination of demographic group, day type and commuting status, the number of activity patterns in each of 1–3 clusters (derived from cluster analysis on the activity-pattern data from CHAD) and the cluster-to-cluster transition probabilities (derived from the transition frequencies for multiple-day activity-pattern records

from CHAD). These values are used to create weights for averaging selected activity patterns, one from each cluster, to represent an individual within the group for that day type.

The *cluster-transition* file begins with a text header record, followed by one data record for each combination of day type and demographic group. The header record indicates the order of the variables in each of the data records. Although the header record of the *cluster-transition* file is not used by the model programs, it provides documentation to inform the user of the meaning of the data fields. The header record of the current default *cluster-transition* file is as follows.

Header record from the default *cluster-transition* file

```
g  dt  ct nclus clust1  clust2  clust3  prob11  prob12  prob13  prob21  prob22  prob23
prob31  prob32  prob33
```

The *cluster-transition* file is read in free format with the variables defined in Table 3-4 for each combination of day type and demographic group.

Table 3-4.
Variables in the *cluster-transition* file

Variable	Description
g	demographic group
dt	day type
ct	commuting status of subject
nclus	number of clusters for the group/day type (1–3)
clust1	cumulative fraction of group/day type in cluster #1
clust2	cumulative fraction of group/day type in clusters #1–2
clust3	cumulative fraction of group/day type in clusters #1–3
prob11	cumulative transition probability from cluster #1 to #1
prob12	cumulative transition probability from cluster #1 to clusters #1–2
prob13	cumulative transition probability from cluster #1 to clusters #1–3
prob21	cumulative transition probability from cluster #2 to #1
prob22	cumulative transition probability from cluster #2 to clusters #1–2
prob23	cumulative transition probability from cluster #2 to clusters #1–3
prob31	cumulative transition probability from cluster #3 to #1
prob32	cumulative transition probability from cluster #3 to clusters #1–2
prob33	cumulative transition probability from cluster #3 to clusters #1–3

An extract from the current default *cluster-transition* file is presented below. See Appendix A for more details on the current HAPEM default input files.

Extract from a default *cluster-transition* file
(in "wrapped" view)

```
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
```

3.12.   *Statefip* File

The *statefip* file cross-references the two-character FIPS code for each U.S. state (and District of Columbia, Puerto Rico, and the U.S. Virgin Islands, totaling 53 areas) to its numerical ranking on the list. The format of each record is as follows.

Variables in the statefip file

Numerical Rank	(integer)
State (or district or territory) FIPS code	(2-character string)

As discussed in Section 2.1.7 (The *Statefip* File), the *statefip* file is used in conjunction with **region1** and **region2** specified in all the *parameter* files to specify the areas to be included in the analysis, according to numerical ranking.

Default *statefip* file and corresponding state names

1	01	Alabama
2	02	Alaska
3	04	Arizona
4	05	Arkansas
5	06	California
6	08	Colorado
7	09	Connecticut
8	10	Delaware
9	11	District Of Columbia
10	12	Florida
11	13	Georgia
12	15	Hawaii
13	16	Idaho
14	17	Illinois
15	18	Indiana
16	19	Iowa
17	20	Kansas
18	21	Kentucky
19	22	Louisiana
20	23	Maine
21	24	Maryland
22	25	Massachusetts
23	26	Michigan
24	27	Minnesota
25	28	Mississippi
26	29	Missouri
27	30	Montana
28	31	Nebraska
29	32	Nevada
30	33	New Hampshire
31	34	New Jersey
32	35	New Mexico

33	36	New York
34	37	North Carolina
35	38	North Dakota
36	39	Ohio
37	40	Oklahoma
38	41	Oregon
39	42	Pennsylvania
40	44	Rhode Island
41	45	South Carolina
42	46	South Dakota
43	47	Tennessee
44	48	Texas
45	49	Utah
46	50	Vermont
47	51	Virginia
48	53	Washington
49	54	West Virginia
50	55	Wisconsin
51	56	Wyoming
52	72	Puerto Rico
53	78	U.S. Virgin Islands

4. HAPEM Output Files

HAPEM creates three diagnostic output files and a set of *final exposure* output files. The diagnostic files record error messages and information about the parameters of the simulations. The names for these three files are specified by the user in the *parameter* files of each model program. The *final exposure* output files contain all the exposure estimates from a model run. The pathnames for these files are specified by the user in the *parameter* file for the <u>AIRQUAL</u> and <u>HAPEM</u> programs.

4.1. *Log* File

The *log* file contains a record of a model analysis. Three of the model programs (<u>INDEXPOP</u>, <u>COMMUTE</u>, and <u>HAPEM</u>) will append records onto an existing *log* file, as specified their *parameter* files, without overwriting previous records. The <u>DURAV</u> and <u>AIRQUAL</u> programs will overwrite any records on an existing *log* file. Therefore, if a single *log* filename is used to run all the model programs, a running record will be written for the <u>DURAV</u>, <u>INDEXPOP</u>, and <u>COMMUTE</u> programs, but then the <u>AIRQUAL</u> program will erase those records and begin a new record, and the <u>HAPEM</u> program will add to it. To maintain a complete *log* file record of a HAPEM simulation, the two alternatives below can be used.

- If a single *parameter* file is for a complete simulation, so that the *log* filename is the same for all five programs, manually rename the *log* file created by the first three programs before <u>AIRQUAL</u> is run.

- Use a different *parameter* file for running <u>AIRQUAL</u> and <u>HAPEM</u> than for the other programs, with a different name specified for the *log* file.

If the model programs experience no fatal errors during a simulation, there are several items written to the *log* file by each of the programs. The first record written to the file by each program identifies the program and its start time. The start time consists of three numbers—the current time, the size of the time increment equivalent to one second, and the maximum value allowed for the current time before it is reset to zero. All three of these quantities are system-dependent. An example record of this type is presented below.

Example *log* record: program and start time

```
DURAV Start time=    34862630 1000 86399999
```

The last two records written to the *log* by each model program report the ending time and the total job time for the particular program. For the total-job-time record, the job time is converted into seconds. Note that the total job time will not be correct if the clock maximum is exceeded during the job. An example of these types of records is presented below.

Example *log* records: program, stop time, and run time

```
DURAV End time =   34880980
DURAV Job time =   18.3500004
```

If an error occurs that HAPEM considers to be fatal, a diagnostic message will be written to the *log* file and the program will stop. For example, if <u>DURAV</u> finds that the number of time blocks per day specified in the *activity*-file header does not match the value of **nblock** specified in the *parameter* file, it will write a message to the *log* file and stop. An example of this type of record is presented below.

Example *log* record: error message

```
number of time blocks in activity file does not equal nblock 999
```

4.1.1. <u>DURAV</u> Output to the *Log* File

Apart from the text produced by all model programs, each program writes some specialized information to the *log* file. The <u>DURAV</u> program writes the names of the input *activity* and *cluster* files and the output file (the averaged activity database). An example of these types of records is presented below.

Example *log* records: input and intermediate files

```
CHAD data from file=input/activity pattern/durhw_HAPEM8.txt
Clustering from file=input/activity pattern/cluster_HAPEM8.txt
Output data on file=input/activity pattern/durhw_HAPEM8.da
```

The <u>DURAV</u> program also records the number of records (person-days) extracted from the *activity* file, and it produces a table of frequency counts for each combination of demographic group, commute status, and day type (a matrix whose elements should sum to the total number of records extracted). If any elements of this matrix are zero then there are groups that have no activity patterns and thus are undefined. If the numbers are positive but small (e.g., less than ten), then there is a chance that the exposure results might not be representative for the group. An example of a part of this type of matrix is presented below.

Example *log* records: file matrix

```
Frequency table for diaries (non-commute)
By demographic group (rows) & day type (cols)
  198   691   697
 1275  2126  1968
 2682  7664  7171
  469  2095  2465
 7969 16967 41348
 4404  9678 11555
Frequency table for diaries (commute)
By demographic group (rows) & day type (cols)
    0     0     0
    0     0     0
    0     0     0
  146   382   534
11360 28136 13653
  747  1560   681
```

4.1.2. <u>INDEXPOP</u> Output to the *Log* File

In addition to the program name and the start-, stop-, and job-time information provided to the *log* file by all the model programs, the <u>INDEXPOP</u> program writes two other records to the *log* file. The first confirms that all the input files were successfully opened, and the second records the total number of tract records in the *population* file. An example of these two records is presented below.

Example *log* records: opened files and tract counts

```
Finished opening files
 total number of tracts is     85427
```

4.1.3. <u>COMMUTE</u> Output to the *Log* File

The <u>COMMUTE</u> program writes no information to the *log* file other than the program name and the start, stop, and job time.

4.1.4. <u>AIRQUAL</u> Output to the *Log* File

In addition to the program name and the start-, stop-, and job-time information provided to the *log* file by all the model programs, the <u>AIRQUAL</u> program writes several other records to the *log* file. First, a summary of the *air quality* file is written to document the number of census tracts and distinct counties found in the file. These tracts are then paired with the tracts found in the *population* file. The number of tracts found in the *air quality* file but not in the *population* file is recorded in the line containing the phrase "unpaired air tracts". This is followed by the list (if any) of unpaired tracts. Then, the tracts in the *population* file are compared to the tracts in the *air quality* file—the number of tracts in the *population* file but not in the *air quality* file is reported, along with the number of matching tracts as well as the number of *population* tracts with multiple *air quality* tracts. Next, similar statistics are given comparing tracts within the counties in the *air quality* file to those in the *population* file. Any tract in the *population* file but not the *air quality* file will not be modeled; similarly, any tract in the *air quality* file but not in the *population* file will not

be modeled. An example of the log output produced by the <u>AIRQUAL</u> program is presented below.

Example *log* records: <u>AIRQUAL</u> statistics

```
# air tracts =        84810       # of air records =       84810
 # counties on air file =       3224
 There were            0  unpaired air tracts.
 Overall, there were:
        617  unpaired census tracts.
      84810  census tracts with a matching air tract.
          0  census tracts with 2 or more air tracts.
 Within the counties on the air file, ther were:
        617  unpaired census tracts.
      84810  census tracts with a matching air tract.
          0  census tracts with 2 or more air tracts.
```

4.1.5. <u>HAPEM</u> Output to the *Log* File

In addition to the program name and the start-, stop-, and job-time information provided to the *log* file by all the model programs, the <u>HAPEM</u> program writes two other records to the *log* file. It reports the time when dynamic array allocation is complete and the number of tracts used in the analysis (i.e., that had data in the *air quality*, *population*, and *commuting* files). An example of the *log* output produced by the <u>HAPEM</u> program is presented below.

Example *log* record: <u>HAPEM</u> array allocation and tracts

```
 HAPEM Allocation =  1317787200
  There were           85427   tracts in the study area.
```

4.2. *Counter* File

A second diagnostic file created by HAPEM is the *counter* file. The *counter* file records the number of records in various data-input and -output files, which can also be a useful tool for troubleshooting and keeping track of which files were used in the simulation.

It is important to use same *counter* file for all the model programs in a simulation—the programs use some of the information recorded by previous programs for dynamic memory allocation of arrays. If the expected records from previous programs are not in the *counter* file, an error will occur.

The model programs add records to the *counter* file by appending to the end of the records generated by the previous programs, where programs are run in the expected order as described in Section 2.1 (Model Structure; though running the <u>COMMUTE</u> program is optional). For example, the <u>INDEXPOP</u> program reads records generated by the <u>DURAV</u> program, and then it begins its own recording. If the <u>INDEXPOP</u> program is run a second time using the same *counter* file, the second run will overwrite the previous records generated by the <u>INDEXPOP</u> program.

The specific information recorded in the *counter* file is provided in Table 4-1. An example counter file is also shown below.

Table 4-1.
Variables in the *counter* file

HAPEM Program	Record Number	Description
<u>DURAV</u>	1	-number of data records (person-days) in the *activity* file -number of *activity*-file data records (person-days) with 1,440 total minutes
<u>INDEXPOP</u>	2	-number of data records (tracts) in the *population* file -number of counties in the *population* file (
<u>COMMUTE</u>	3	-number of data records in the population index file (e.g., *population_HAPEM8_direct.ind*) -number of data records (home-tract/work-tract pairs) in the *commuting* file
	4	-number of data records in the work-tract file (e.g., *commute_flow_HAPEM8.da*) -number of records in the commuting index file *(e.g., commute_flow_HAPEM8.ind*)

HAPEM Program	Record Number	Description
AIRQUAL	5	-number of matching tracts in the *air quality* (e.g., *benzene.txt*) and population index (e.g., *population_HAPEM8.ind*) files -number of counties with matching tracts in the *air quality* (e.g., *benzene.txt*) and population index (e.g., *population_HAPEM8.ind*) files
	6	-number of tracts in the *air quality* file (e.g., *benzene.txt*) -number of data records in the *air quality* file (e.g., *benzene*.txt)

Example *counter* file

```
178621        178621
 85427          3224
 85427       6004343
5831073         85427
 84810          3224
 84810         84810
```

The relationships listed below are expected among the numbers in the *counter* file.

- The number of records in the *population* file, the population index file (e.g., *population_HAPEM8.ind*), and the commuting index file *(e.g., commute_flow_HAPEM8.ind*) should all be the same.

- The number of records in the work-tract file (e.g., *commute_flow_HAPEM8.da*) may be larger or smaller than the number of records in the *commuting* file. It may be larger because the COMMUTE program will create a "commuting" flow for a tract that is in the *population* file but is not a home tract in the *commuting* file (using the *population* tract as both the home and work tract). It may also be smaller if the study area in the *population* file is smaller than the study area in the *commuting* file (the default *commuting* file is all U.S. states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands).

4.3. *Mistract* File

A third diagnostic file created by the COMMUTE, AIRQUAL, and HAPEM programs is the *mistract* file. If the same *mistract* filename is used for the COMMUTE and AIRQUAL programs, the COMMUTE program's information will be overwritten by that of the AIRQUAL program. The HAPEM program will then append records onto an existing *mistract* file. To maintain a complete record of this information for a HAPEM simulation, either different *mistract* filenames should be used for the COMMUTE and AIRQUAL programs (requiring different *parameter* files), or the *mistract* file should be manually renamed after the COMMUTE program is run.

Each of the three programs records a different set of information about the consistency of census tracts included in the input files, as detailed in the list below. Below the list are example excerpts from each program's *mistract* file.

- The COMMUTE program's *mistract* file records the state, county, and tract FIPS codes of each tract in the *population* file that is not matched by a home tract in the *commuting* file. These unmatched tracts are still processed by the COMMUTE program, as explained in the previous section, by creating a "commuting" flow using the *population* tract as both the home and work tract.

- The <u>AIRQUAL</u> program's *mistract* file records the record number and the state, county, and tract FIPS codes of each tract in the *population* file that is not matched by a tract in the *air quality* file. Only tracts that are included in both the files are processed by HAPEM, since both these pieces of information about a tract (population and air quality) are needed to make an exposure estimate.

- The <u>HAPEM</u> program's *mistract* file records the state, county, and tract FIPS codes of each home tract in the *commuting* file that is not matched by a tract in the air-quality index files. These air-quality index files contain information on tracts that were included in both the *population* and *air quality* files. The unmatched home tracts are not processed further. The <u>HAPEM</u> program's *mistract* file also records each instance of a work tract that is not matched by a tract in the *air quality* file; for these cases, the work tract is assigned the air-quality values of the home tract.

Example excerpt from the <u>COMMUTE</u> program's *mistract* file

```
MISSING TRACTS OF COMMUTE & AIRQUAL in COMMUTE
      44        2        1 01003990000  0
     109        8        1 01015981903  0
     139       13        1 01025957601  0
[etc.]
```

Example excerpt from the <u>AIRQUAL</u> program's *mistract* file

```
MISSING TRACTS for AIRQUAL & POPULATION DATA in AIRQUAL
    2375 04013980500
    5785 06037320000
    7049 06037980001
[etc.]
```

Example excerpt from the <u>HAPEM</u> program's *mistract* file

```
MISSING TRACTS of AIRQUAL & COMMUTE IN HAPEM
 airtract match with worktract not found
home     3   2375 04013980500  0
 airtract match with worktract not found
[etc.]
```

4.4. *Final Exposure* File

As explained in Section 2.1.9 (Exposure Output Files), HAPEM creates an exposure output file for each combination of state and HAP. The names of these files are constructed by the model based on the HAP SAROAD code (specified by **sarod** in the *parameter* file) and the state FIPS code (as SAROAD.FIPS.dat).

The *final exposure* output files each begin with a repetition of some of the information specified in the *parameter* file for the <u>AIRQUAL</u> and <u>HAPEM</u> programs, as listed below.

Information at the top of the *final exposure* output file

State FIPs code	
HAP SAROAD code	(sarod)
HAP name	(pollutant)
HAP CAS number	(CAS)
Air quality data units	(units)
Year of air quality data	(year)
Number of outdoor air emission source categories	(nsource)
Random number seed for activity pattern selection	(Rseed1)
Random number seed for ME factors selection	(Rseed2)
Random number seed for air quality data selection	(Rseed3)
Number of indoor product emission source types	([a])
Number of indoor material emission source types	([a])
Number of indoor combustion emission source types	([a])
Number of vehicle in residential garage emission source types	([a])
EPA Region of indoor emission source data	([a])
Number of demographic groups	(ngroup)
Number of replicates for each demographic group	(nreplic)
Definition of each demographic group, ordered as in the *population* file	(under "Demographic Groups:" heading in the parameter file for the <u>AIRQUAL</u> and <u>HAPEM</u> programs)

[a] Indoor-source algorithms are included in the HAPEM program but have not yet been tested and reviewed. Therefore, they are currently not recommended for use, and instructions for their use are omitted from this document. To disable the indoor-source algorithms, set keyword CAS to 99999.

This information is followed by a header record defining the fields in the data records. An example header record is presented below.

Example header record for *final exposure* output file
(in "wrapped" view)

```
ST CTY CENSUS GRUP POPUL   SOURCE01   SOURCE02   SOURCE03   SOURCE04  BackgConc
IndCon_Pro IndCon_Mat IndCon_Com IndCon_Veh Total Conc
```

The header record is then followed by **nreplic** data records for each combination of group and tract combination. The format of each data record, assuming **nsource** = 4, is provided in Table 4-2.

Table 4-2.
Variables in the *final exposure* output file (assuming *nsource* = 4)

Field Numbers	Description	Format
1	leading space	
2–3	state FIPS code	2-character string
5–7	county FIPS code	3-character string
9–14	tract FIPS code	6-character string
16–17	demographic-group indicator	integer 1–10, ordered as in the *population* input file
19–25	number of people to which the exposure estimates in the data record apply	decimal number equal to the population of the group/tract combination divided by ***nreplic***
27–36	estimated exposure-concentration contribution from emission-source-category 1	decimal number in scientific notation; units of measurement as in the *air quality* file
38–47	estimated exposure-concentration contribution from emission-source-category 2	decimal number in scientific notation; units of measurement as in the *air quality* file
49–58	estimated exposure-concentration contribution from emission-source-category 3	decimal number in scientific notation; units of measurement as in the *air quality* file
60–69	estimated exposure-concentration contribution from emission-source-category 4	decimal number in scientific notation; units of measurement as in the *air quality* file
71–80	estimated exposure-concentration contribution from background	decimal number in scientific notation; units of measurement as in the *air quality* file; derived from the sum of the uniform background—***backg***—and the variable background concentrations
82–91	estimated exposure-concentration contribution from indoor-product emission sources	decimal number in scientific notation; units of measurement as in the *air quality* file
93–102	estimated exposure-concentration contribution from building-materials indoor emissions	decimal number in scientific notation; units of measurement as in the *air quality* file
104–113	estimated exposure-concentration contribution from indoor-combustion emission sources	decimal number in scientific notation; units of measurement as in the *air quality* file
115–124	estimated exposure-concentration contribution from vehicles in attached garages	decimal number in scientific notation; units of measurement as in the *air quality* file
126–135	estimated total-exposure concentration	decimal number in scientific notation; units of measurement as in the *air quality* file; the sum of the preceding contribution values

An example of a set HAPEM exposure-output records (for 30 replicates of one demographic group in one tract) is presented below. The total population for group 1 in this tract is 36 and ***nreplic*** = 30, so that the number of people to which the exposure estimates in each record apply is 36/30 = 1.200.

Example set of exposure-output records (for 30 replicates of one demographic group in one tract)

ST	CTY	CENSUS	GRUP	POPUL	SOURCE01	SOURCE02	BackgConc	IndCon_Pro	IndCon_Mat	IndCon_Com	IndCon_Veh	Total Conc
78	010	970100	1	1.200	0.4119E+00	0.1527E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4121E+00
78	010	970100	1	1.200	0.3939E+00	0.1699E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3941E+00
78	010	970100	1	1.200	0.3360E+00	0.1102E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3361E+00
78	010	970100	1	1.200	0.4791E+00	0.1654E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4793E+00
78	010	970100	1	1.200	0.4057E+00	0.1228E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4058E+00
78	010	970100	1	1.200	0.4652E+00	0.1805E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4654E+00
78	010	970100	1	1.200	0.4535E+00	0.1430E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4536E+00
78	010	970100	1	1.200	0.4412E+00	0.1815E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4414E+00
78	010	970100	1	1.200	0.3891E+00	0.1564E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3893E+00
78	010	970100	1	1.200	0.3771E+00	0.1278E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3772E+00
78	010	970100	1	1.200	0.3785E+00	0.1366E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3786E+00
78	010	970100	1	1.200	0.4700E+00	0.1662E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4702E+00
78	010	970100	1	1.200	0.3918E+00	0.1333E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3919E+00
78	010	970100	1	1.200	0.3923E+00	0.1632E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3925E+00
78	010	970100	1	1.200	0.3228E+00	0.8932E-04	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3229E+00
78	010	970100	1	1.200	0.2741E+00	0.8380E-04	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.2742E+00
78	010	970100	1	1.200	0.5163E+00	0.2219E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.5165E+00
78	010	970100	1	1.200	0.3965E+00	0.1391E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3966E+00
78	010	970100	1	1.200	0.3831E+00	0.1792E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3833E+00
78	010	970100	1	1.200	0.3829E+00	0.1528E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3831E+00
78	010	970100	1	1.200	0.3717E+00	0.1006E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3718E+00
78	010	970100	1	1.200	0.4407E+00	0.1510E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4409E+00
78	010	970100	1	1.200	0.3463E+00	0.1926E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3465E+00
78	010	970100	1	1.200	0.5035E+00	0.1562E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.5037E+00
78	010	970100	1	1.200	0.4628E+00	0.1433E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4629E+00
78	010	970100	1	1.200	0.3847E+00	0.1175E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3848E+00
78	010	970100	1	1.200	0.3232E+00	0.3503E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3236E+00
78	010	970100	1	1.200	0.3479E+00	0.1098E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.3480E+00
78	010	970100	1	1.200	0.4049E+00	0.3598E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4053E+00
78	010	970100	1	1.200	0.4263E+00	0.1639E-03	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.4265E+00

5. HAPEM Programs

This section contains detailed descriptions of the five programs that are contained in the model: <u>DURAV</u>, <u>INDEXPOP</u>, <u>COMMUTE</u>, <u>AIRQUAL</u>, and <u>HAPEM</u>. The first four programs (<u>DURAV</u>, <u>INDEXPOP</u>, <u>COMMUTE</u>, and <u>AIRQUAL</u>) are pre-processers that convert input-data files into the form required for efficient exposure calculations. The final program (<u>HAPEM</u>) performs the exposure calculations and summarizes the results.

It is important to note that some knowledge of Fortran programming is necessary to understand all the programming details discussed in this section. However, all the general concepts related to the programs should be clear to all users.

5.1. Programming Guidelines Used to Develop HAPEM

The source code for each of the five model programs is written in Fortran 90 and designed so that it can be compiled and executed on various platforms (e.g., UNIX, DOS, Windows) with little or no programming changes required.

The model programs incorporate a structured programming style as summarized by the attributes listed below.

- No "GO TO" statements or line numbers are in any of the programs. Program flow is direct from the beginning to the end within each program, thus making the code easy to follow. The only looping is within "DO" blocks.

- No filenames appear in source code. Instead, this information is specified in the *parameter* file, which is read in from the command line.

- Most parameter values are input from the *parameter* file so that the programs themselves only allocate space for carrying out as many calculations as are necessary.

- Most arrays depending on variable parameters are dynamically allocated.

- All variables are declared (no implicit typing), with comments at the end of most declarations to assist in interpretation. Comment lines are inserted between the logical blocks of code for clarity.

5.1.1. Common Structural Elements

All the model programs consist of a declarations section, a parameters section, a setup section, a primary section that processes the data, and a wrap-up section.

In the declarations section, all variables are explicitly typed. Most lines include a trailing comment to indicate the general purpose of the variable(s). Arrays that are to be dynamically allocated are fixed in rank (number of dimensions), with a colon used to defer the size specification.

The second program section, referred to as the params section, reads the *parameter* file to determine the specific input filenames and the parameter settings. This section is similar in all the model programs, except that only the names of files needed by each job are retained as variables. Each line of the *parameter* file is read in as a character string (maximum length of 120 characters) and inspected for an equals sign ("="). If there is no equals sign, then the line is ignored. This allows the programmer to add comments and other lines directly to the *parameter* file without altering its performance. Lines containing an equals sign are divided into two parts at the equals sign. The part to the left of the sign is scanned for keywords. All keywords are in lower case. If the string 'file' is found, then the line is assumed to specify one of the input or output files. For these lines, a second keyword is searched for. Possible keywords are provided in Table 5-1. Which filenames and paths are required by each model program are shown in Table 2-1 as user-defined files.

Table 5-1.
The filename keywords in the *parameter* files
recognized by the model programs

Keyword	Definition
activity	name of the *activity* file (input)
cluster	name of the *cluster* file (input)
ClusTrans	name of the *cluster-transition* probability file (input)
populat	name of the *population* file (input)
CommutTime	name of the *commuting-time* file (input)
CommutFrac	name of the *commuting-fraction* file (input)
DistToRoad	name of the *distance-to-road* file (input)
commut	name of the *commuting* file (input)
quality	name of the *air quality* file (input)
factors	name of the *factors* file (input)
mobiles	name of the *mobiles* file (input)
statefip	name of the *statefip* file (input)
log	name of the *log* file (output)
counter	name of the *counter* file (output)
mistract	name of the *mistract* file (output)
afile	path of *final exposure* file (output)
Product[1]	path of indoor source files (input)
AutoPduct[1]	Name of file for automobile-related consumer products (input)

[1] A path to one or more indoor-emission-source inputs for the indoor-source algorithms is specified in these statements (with the *AutoPduct* statement including a filename). These algorithms are included in the HAPEM program, but they have not yet been tested and reviewed. Therefore, they are currently not recommended for use, and instructions for their use are omitted from this document. To disable the indoor-source algorithms, set keyword **CAS** to 99999, and specify any existing path (and file for *AutoPduct*; other than those otherwise specified for input or output for the HAPEM program) since no indoor-source files will then actually be utilized by the HAPEM program.

The model user can use the above keywords in lines that do not contain an equals sign, or in comments containing an equals sign as long as the word "file" does not also appear left of the equals sign. The strings containing the directory and filenames should not exceed 100 characters. If they do, then use an alias or a logical drive specification to identify most of the path, and thereby reduce the length to less than 100 characters. As described earlier in this guide, each of the input files requires a certain format for the data. It is the responsibility of the user to ensure that this format specification is met.

The setup section allocates and initializes the dynamic arrays that can be sized from the parameter settings specified in the *parameter* file. Other arrays that are dependent on the number of records in an input file are allocated elsewhere. The dynamic allocation saves space and time by only using as much space as is necessary, allows for the parameters to be increased or decreased without recompiling the program, and allows vector and array operations to be programmed more simply since they can be applied to the entire array rather than only to certain elements.

5.2. Program Descriptions

This section describes the purpose and structure of the processing section of each of the five model programs.

5.2.1. DURAV

As explained in Section 2.1.2 (The DURAV Program and the *Activity* and *Cluster* Files), the DURAV program performs the two main functions listed below.

- If a different number of daily time blocks is specified for the analysis than in the *activity* file, it processes the activity records so that the number of time blocks matches the number specified for the analysis.

- It creates a sequential ASCII file of the activity pattern records for use by the HAPEM program.

The six age groups in HAPEM are as follows, in years.

- 0–1
- 2–4
- 5–15
- 16–17
- 18–64
- 65+

Currently, season and day of week are used to determine three day types as

- weekdays in summer (June–August),
- other weekdays, or
- weekends.

Cluster types are used to represent variations in activity pattern within each combination of demographic group, day type, and commuting status. There are 1–3 cluster types for each combination of group, day type, and commuting status. Each CHAD record in the *activity* file has been assigned a cluster type based on the cluster analyses. While DURAV makes use of the age groups, day types, and commuting-status categories, those are already on the *activity* and *cluster* input files.

DURAV Processing Operations

In addition to the operations discussed above, the params section of DURAV conducts the operations described below.

- The *parameter* file, which stores input model variables and input file names, is read in from the command line.

- The values of **nblock** (the number of time blocks per day in the *activity* file) and **hblock** (the number of time blocks per day for the analysis), specified in the *parameter* file, are checked for compatibility. As explained elsewhere in this guide, **hblock** must be an integral factor of **nblock**, so that the activity time blocks can be combined if necessary to match **hblock.** If the check fails, then an error message is written to the *log* file and the program stops.

Further, the setup section of DURAV conducts the operation described below.

- The number of time blocks per day in the *activity* file is determined from the header record, as explained elsewhere in this guide. This number is checked against the value of **nblock** specified in the *parameter* file. If the values are different, an error message is written to the *log* file and the program stops.

Finally, the main processing section of DURAV conducts the several operations described below.

- The number of data records in the *activity* file is determined so that memory can be allocated for various arrays used to hold the input data records and other data derived from them.

- Each activity record is checked to ensure that the total activity time is 1,440 minutes. If the check fails, then a message is written to the screen. This should never occur if it is checked for when developing the *activity* file.

- The **nblock** time blocks in each activity record are aggregated, if necessary, to create **hblock** time blocks.

- The total number of data records in the *activity* file, and the total number with activity durations of 1,440 minutes, are recorded in the *counter* file.

- The number of aggregated records in each combination of demographic group, day type, and commuting status is determined.

- The number of aggregated activity records in each combination of demographic group, day type, commuting status, and cluster, and the number of clusters in each combination of group, day type, and commuting status, are recorded in an intermediate file with filename extension *.nonzero*.[12] This information is used in the HAPEM program, as described in Section 5.2.5 (HAPEM).

---

[12] The *.nonzero file also records a flag for each combination of demographic group and day type, indicating whether 10 percent of the activity patterns include commuting. This flag was used by an earlier version of the HAPEM program, but it is not used in this version.

- The total number of aggregated activity records processed and their allocation among demographic group, day types, and commuting status is written into the *log* file.

- The activity patterns are written into a sequential file with filename extension *.da* sorted by demographic group, day type, commuting status, and cluster type, and the filename is recorded in the *log* file.

5.2.2. INDEXPOP

As explained in Section 2.1.3 (The INDEXPOP Program and the *Population*, *Distance-to-road*, *Commuting-time*, and *Commuting-fraction* Files), the INDEXPOP program performs the two main functions listed below.

- It creates a direct-access file of population data to be used in the AIRQUAL program.

- It creates sequential ASCII index files for the population data census tracts, to facilitate file searching in the COMMUTE and AIRQUAL programs.

- It creates direct-access files and associated index files of the data in the *distance-to-road*, *commuting-time*, and *commuting-fraction* files, to be used in the COMMUTE and AIRQUAL programs.

INDEXPOP Processing Operations

The specific operations performed in the main processing section of INDEXPOP are described below.

- The *parameter* file, which stores input model variables and input file names, is read in from the command line.

- Each record in the *distance-to-road* file is read and written into a direct-access file with filename extension *.dat*, and an associated index file is created with filename extension *.STIDX*.

- Each record in the *commuting-time* file is read and written into a direct-access file with filename extension *.dat*, and an associated index file is created with filename extension *.STIDX*.

- Each record in the *commuting-fraction* file is read and written into a direct-access file with filename extension*.dat*, and an associated index file is created with filename extension *.STIDX*.

- The number of data records in the *population* file is determined so that memory can be allocated for various arrays used to hold the input data records and other data derived from them.

- Each data record in the *population* file is read. The population array is recorded in a direct-access file with the filename extension *.da*. The state FIPS, county FIPS, tract FIPS, and serial record number are recorded in a direct-access file with the filename extension *_direct.ind*.

- The total number of tract records in each county is determined.

- The total number of counties included in the *population* file that are in each state is determined.

- A sequential index file is created with filename extension *.county_tract_pop_range*. For each county in the *population* file, there is a record in this file indicating the serial record numbers of the first and last data record for tracts in that county in the **.da* and **_direct.ind* files.

- A sequential index file is created with filename extension *.state_county_pop_range*. For each county, there is a record in this file indicating the serial record numbers of the first and last data record for counties in that state in the **.county_tract_pop_range* file.

- The total number of records (tracts) and counties in the *population* file is added to the *counter* file.

5.2.3. COMMUTE

As explained in Section 2.1.4 (The COMMUTE Program and the *Commuting*, *Distance-to-road*, *Commuting-time*, and *Commuting-fraction* Files), the COMMUTE program performs the three main functions described below.

- It creates a file identifying the set of work tracts (i.e., tracts in which the residents of the home tract work) associated with each census tract (i.e., home tract), the fraction of workers residing in that home tract and working in each work tract, and the normalized centroid-to-centroid distance between home tract and each work tract. The normalized distance is the distance/(average distance). The normalized distance is combined with the average commuting time for the tract to estimate the commuting time for the home-tract/work-tract pair in the HAPEM program.

- It creates a sequential index file to facilitate file searching in the HAPEM program.

- It adds the census-tract-specific information from the distance-to-road, commuting-time, and commuting-fraction direct-access files (created in the INDEXPOP program) to the commuting index file.

COMMUTE Processing Operations

The specific operations performed in the main processing section of COMMUTE are as follows.

- The *parameter* file, which stores input model variables and input file names, is read in from the command line.

- The distance-to-road index file (filename extension *.STIDX*, created in INDEXPOP) is read twice: first to determine the number of records for array allocation, and then to populate the arrays with the data in the file.

- The commuting-time index file (filename extension *.STIDX*, created in INDEXPOP) is read twice: first to determine the number of records for array allocation, and then to populate the arrays with the data in the file.

- The commuting-fraction index file (filename extension *.STIDX*, created in INDEXPOP) is read twice: first to determine the number of records for array allocation, and then to populate the arrays with the data in the file.

- The number of data records in the *commuting* file is determined so that memory can be allocated for various arrays used to hold the input-data records and other data derived from them.

- The number of *commuting* records with home tracts in each state is determined.

- For each state, the sequence numbers of the first and last data record indicating a home tract in that state are determined.

- The number of records in the *population* file is read from the *counter* file, so that memory can be allocated for various arrays used to hold the input-data records and other data derived from them.

- All the tract FIPS are read from the *_direct.ind* file created by <u>INDEXPOP</u>, using the indices from the *.state_county_pop_range* and *.county_tract_pop_range* files created by <u>INDEXPOP</u>.

- For each tract in the *_direct.ind* file created by <u>INDEXPOP</u>, all matching home tracts in the *commuting* file are found. (There is one home-tract record for every commuting flow originating in that tract). For each matched home tract, the FIPS and number of work tracts within 120 km are determined. For each home tract, the fractions of total commuting flow to work tracts, which are specified in the *commuting* file, are adjusted to the fractions of the total commuting flow within 120 km.

- For each home-tract/work-tract pair, the centroid-to-centroid distance from the *commuting* file is determined and a normalized distance is calculated as distance/(average distance).

- Each work-tract FIPS, its adjusted flow fraction, and its normalized distance are recorded in a sequential file with filename extension *.da* (one record for each work tract).

- If no matching home tracts are found in the *commuting* file for a *population* tract, an entry is recorded in the *mistract* file, indicating the tract FIPS and the indices of the tract in the *.state_county_pop_range*, *.county_tract_pop_range*, and *_direct.ind* files.

- For *population* tracts with no matching *commuting* home tracts, a record is recorded in the *.da* file indicating the *population* tract as the work tract, with fractional commuting flow of 1.0 (i.e., all work takes place in the home tract).

- For each *population* tract, a record is written into a temporary index file. The fields in the record are the *population* tract FIPS, the sequence numbers of the first and last work tract record in the *.da* file, and a flag indicating whether the *population* tract was matched by a home tract in the *commuting* file (0=no; 1=yes).

- Two records are added to the *counter* file. The first record indicates the number of records found in the *_direct.ind* file (created by <u>INDEXPOP</u>) and the number of data records found in the *commuting* file. The second record records the number of records in the *.da* file and the number of records in the *.ind* file.

- A sequential index file is created with filename extension *.st_comm1_fip_range*. For each state, there is a record in this file indicating the sequence numbers of the first and last data record for tracts for that state in the temporary index file.

- The temporary index file is read from the beginning. Each record is matched by tract with a record in the distance-to-road, commuting-time, and commuting-fraction direct-access files (filename extensions of *.dat*, created in <u>INDEXPOP</u>). The combined data for each tract are written into a direct-access file with the root filename of the *commuting* file and the filename extension *.ind*.

5.2.4. <u>AIRQUAL</u>

As explained in Section 2.1.5 (The <u>AIRQUAL</u> Program and the *Air Quality* and *Distance-to-road* Files), the <u>AIRQUAL</u> program performs the four main functions listed below.

- It creates a sequential file of air-quality data to be used in the <u>HAPEM</u> program.

- It determines the number of data records for each census tract in the *air quality* file.

- It creates index files to facilitate file searching in the <u>HAPEM</u> program.

- It adds the tract-specific information from the distance-to-road direct-access file (created in the <u>INDEXPOP</u> program) to the air-quality index files.

<u>AIRQUAL</u> Processing Operations

The specific operations performed in the main processing section of <u>AIRQUAL</u> are described below.

- The *parameter* file, which stores input model variables and input file names, is read in from the command line.

- The number of data records in the *air quality* file is determined so that memory can be allocated for various arrays used to hold the input-data records and other data derived from them.

- The number of time blocks in the *air quality* file is determined from the header record. It is checked for compatibility with the value of **hblock** (the number of time blocks for the analysis, as specified in the *parameter* file). As explained in Section 2.1.5 (The <u>AIRQUAL</u> Program and the *Air Quality* and *Distance-to-road* Files), **hblock** must be an integral multiple of the number of air-quality time blocks, so that the air-quality values can be replicated if necessary to create **hblock** air-quality values. If this check fails, an error message is written to the *log* file and the program stops.

- Each data record in the *air quality* file is read and, if necessary, the concentration values for each time block are replicated to create **hblock** values.

- The concentrations in each record are recorded in a sequential file with the root name of the *air quality* file and the filename extension .da, (e.g., *HAP.da*) to be used in <u>HAPEM</u>.

- The index ranges for the multiple data records in each tract are determined and stored in an index array.

- All the unique county FIPS in the *air quality* file are counted and the values saved into an array.

- The number of records in the *population* file is read from the *counter* file.

- An attempt is made to match each *population* tract specified in the **_direct.ind* file (created by <u>INDEXPOP</u>) with a tract in the *air quality* file. If a match is found, the population array from the **.da* file (created by <u>INDEXPOP</u>) is recorded in a sequential file with the root name of the *population* file and the filename extension *.pop_air_da* (e.g., *population_HAPEM8.pop_air_da)*. The tract code (state FIPS, county FIPS, and tract FIPS) and the indices range for data records in a tract (from the index array) are recorded in a sequential file with the root name of the *air quality* file and the filename extension *.air_da*, (e.g., *HAP.air_da*). If no match is found, the serial record number of the tract in the **_direct.ind* file (created by <u>INDEXPOP</u>) and the tract code are recorded in the *mistract* file.

- For each state, the number of tracts in the **.air_da* file is determined.

- For each county in the **.air_da* file, the number of tracts is determined.

- A sequential index file is created with filename extension *.state_air_fip_range*. For each county, there is a record in this file indicating the serial record numbers of the first and last data records in the **.pop_air_da* and **.air_da* files.

- A sequential index file is created with filename extension *.state_air1_fip_range*. For each state, there is a record in this file indicating the serial record numbers of the first and last data records in the **.state_air_fip_range* file.

- A sequential index file is created with filename extension *.state_air2_fip_range*. For each state, there is a record in this file indicating the serial record numbers of the first and last data records in the **.pop_air_da* and **.air_da* files.

- Two records are added to the *counter* file. The first record indicates the number of tracts in the **.pop_air_da* and **.air_da* files, and the number of counties in the **.state_air_fip_range* file. The second record indicates the number of census tracts in the *air quality* file and the number of data records in the *air quality* file.

5.2.5. <u>HAPEM</u>

As explained in Section 2.1.6 (The <u>HAPEM</u> Program, the ME *Factors* and *Mobiles* Files, and the Activity *Cluster-transition* File), the <u>HAPEM</u> program performs the six main functions described below.

- For each demographic group in each census tract, it randomly selects **nreplic** sets of ME factors based on the distribution data provided in the *factors* and *mobiles* files. Each set contains a subset of ME factors randomly selected for each of the time blocks (for the *PEN* and *ADD* factors) or each of the sources (for the *PROX* and *LAG* factors). Each subset contains randomly selected ME factors for each of **nmicro** MEs.

- For each demographic group in each census tract, it randomly selects **nreplic** sets of air-quality data from the datasets available for a tract.

- For each demographic group in each census tract, it creates **nreplic** sets of average activity patterns, where a set contains one average pattern for each day type. An average activity pattern for each day type is calculated as a weighted average of activity patterns randomly selected from each cluster in a group/day-type/commuting-status combination. The weights are determined by the relative frequencies of cluster types

randomly selected in a one-stage Markov process,[9] based on the cluster-transition probabilities provided in the *cluster-transition* file.

- For each activity pattern for a commuting demographic group, it randomly selects a work census tract with probability weighting based on the fraction of residents that work in that tract.

- For each census tract, it estimates the concentration in each ME based on ME factors and outdoor concentrations.

- It combines activity patterns, commuting status, and estimates of ME concentration to calculate **nreplic** annual-average exposure concentrations for each demographic group in each census tract.

HAPEM Processing Operations

The specific operations performed in the main processing section of HAPEM are described below.

- The *parameter* file, which stores input model variables and input file names, is read in from the command line.

- The distribution data of ME factors for each of **nmicro** MEs is read from the *factors* and *mobiles* files (as appropriate) and saved into arrays.

- For the *PROX* distributions in the *mobiles* file for onroad-mobile sources, the average *PROX* factor for the second distance category (75–200 meters) over all the indoor MEs is calculated. (This value will be used later to calculate the ambient concentration for the third distance category [beyond 200 meters], as described below.)

- For each combination of demographic group, day type, and commuting status, the number of activity patterns for each cluster is read from the **.nonzero* file created in DURAV.

- For each combination of demographic group, day type, and commuting status, the frequency of each cluster, and the cluster-to-cluster transition probabilities, are read from the *cluster-transition* file.

- For each combination of demographic group, day type, commuting status, and cluster with a positive number of activity records, the activity-pattern records are read from the **.da* file (created in DURAV) and the values saved into an array.

- Each activity pattern is checked to ensure a total activity time of 1,440 minutes. If this check fails, an error message is written to the *log* file and the program stops.

- Several values are read from the *counter* file to allocate memory for various arrays.

- Indices are read from the **.state_air_fip_range* and **.state_air1_fip_range* files (created by AIRQUAL).

- Data are read from the **.pop_air_da* file and the index ranges for air records from the **.air_da* file (created by AIRQUAL).

- Air-data records are read from **.da* files created by AIRQUAL.

- Indices are read from the **.st_comm1_fip_range* and **.ind* files created by COMMUTE, and data are read from the **.da* file created by COMMUTE.

- For each tract in the **.ind* file created by COMMUTE, an attempt is made to find a matching tract in the **.state_air_fip_range* file created by AIRQUAL. If a match is not found, the *commuting* tract is recorded in the *mistract* file.

- For each demographic group in each census tract, **nreplic** sets of ME factors are randomly selected based on the distribution data provided in the *factors* and *mobiles* files, using subroutines "DISTRIBUTION" and "DATASET" (i.e., distribution_HAPEM8.FOR and dataset_HAPEM8.f90). Each set contains a subset of ME factors randomly selected for each time block (for the *PEN* and *ADD* factors) or each source (for the *PROX* factor). For onroad-mobile source categories, first a distance-from-source category is selected for each indoor ME based on the population fractions in each distance category that were taken from the *distance-to-road* file and added to the commuting index file in COMMUTE.[13] Then, a *PROX* factor for each indoor ME is selected from the appropriate distribution. Each subset contains randomly selected ME factors for each of **nmicro** MEs.

- For each demographic group in each census tract, **nreplic** sets of air-quality data are randomly selected from the datasets available for the census tract in the **.da*" file created by AIRQUAL.

- When a single set of ambient concentrations are provided for each tract in the *air quality* file (as is typically the case), they represent spatial averages over the tract, excluding locations very close to an emission source. For onroad-mobile source categories, it is assumed that the ambient concentrations in the *air quality* file represent spatial averages over the second and third distance categories (the distances 75–200 meters and beyond 200 meters) for the *distance-to-road* and *mobiles* files. Because HAPEM estimates the ambient concentration for the second distance category by applying a *PROX* factor to the "tract-average" ambient concentration, the ambient concentration for the third distance category also is adjusted to make the area-weighted average over these two distance categories equal to the "tract average". This is done as shown below.

$$CONC_{AQ} = AREA_{D3} \times CONC_{D3} + AREA_{D2} \times CONC_{D2} \qquad \text{or}$$

$$CONC_{AQ} = AREA_{D3} \times CONC_{D3} + AREA_{D2} \times PROX_{D2} \times CONC_{D3} \qquad \text{or}$$

$$CONC_{D3} = \frac{CONC_{AQ}}{(AREA_{D3} + AREA_{D2} \times PROX_{D2})}$$

where:

$CONC_{AQ}$:   the "tract-average" concentration from the *air quality* file,

$CONC_{D2}$:   average ambient concentration in second distance category (75–200 meters),

---

[13] It is assumed that the spatial distribution of all indoor MEs in a tract with respect to distance from major roadways is the same as for residences.

$CONC_{D3}$: average ambient concentration in third distance category (beyond 200 meters),

$AREA_{D2}$: fraction of the tract area in the second distance category (from the *distance-to-road* file),

$AREA_{D3}$: fraction of the tract area in the third distance category (from the *distance-to-road* file), and

$PROX_{D2}$: average *PROX* factor for the second distance category over all the indoor MEs (calculated above).[14]

- The randomly selected air-quality data from the *.da* file created by <u>AIRQUAL</u> for each matched tract is combined with the randomly selected ME factors to estimate the concentrations for each ME/time-block combination for that tract.

- For each demographic group in each census tract, the background-exposure-concentration contributions are calculated for each ME/time-block combination based on the uniform value of the **backg** parameter (specified in the *parameter* file), the variable background-concentration values for each data record in *.da* file created by <u>AIRQUAL</u>, and the randomly selected ME factors.

- For each census-tract, demographic-group, and day-type replicate, a commuting status is selected based on the data from the *commuting-fraction* file (that were added to the commuting index file in <u>COMMUTE</u>). If the replicate is a commuter, then a commuting mode (public or private transit) is randomly selected based on the data from the *commuting-time* file (that were added to the commuting index file in <u>COMMUTE</u>). This selection also determines an associated average commuting time for the tract.

- For each replicate that commutes, a work tract is randomly selected for each selected activity pattern, using the subroutine "RANDOMR" within <u>HAPEM</u>. The work tract is selected from the set of work tracts corresponding to that home tract, as specified in the *.da* file created by <u>COMMUTE</u>. The air-quality data for that work tract are randomly selected from the datasets available for the work tract in the *.air_da* file created by <u>AIRQUAL</u>. If the work tract cannot be found in the *.air_da* file, the air-quality data for the home tract are used. The air-quality data are adjusted and combined with the ME factors randomly selected in the same way as the home tract, to estimate the concentrations for each ME/time-block combination for that work tract.

- For each replicate/day-type combination, an average activity pattern is calculated as the weighted average of activity patterns randomly selected from each cluster in a combination of demographic group, day type, and commuting status in the *.da* file created in <u>DURAV</u>. The weights are determined by the relative frequencies of cluster types randomly selected in a Markov process, based on the cluster-transition probabilities provided in the *cluster-transition* file.

- The average activity pattern for the day-type is adjusted so that the commuting time for the replicate is equal to the product of the tract-average commuting time for the

---

[14] As implied by the equations above, the onroad-mobile-source PROX distributions are estimated as the ratios between the near-roadway concentration and the concentration distant from the roadway, rather than the ratios between the near-roadway concentration and the "tract-average" concentration.

commuting mode selected above, and the normalized home-tract/work-tract distance calculated in <u>COMMUTE</u> and recorded in the commuting direct-access file (created in <u>COMMUTE</u>). The adjustments are made by uniform scaling of the time in each time block for commuting MEs (so that the sum matches the total calculated commuting time) and corresponding uniform scaling of the time in each time block for non-commuting MEs.

- The ME/time-block time durations of the weighted-average activity patterns are combined with the estimated ME/time-block concentrations for the home tract and the work tracts to estimate **nreplic** exposure concentrations for each combination of demographic group and day type. A separate set of estimates is made for each emission-source category. The algorithm for each combination of group and day type in the tract is as follows.

$$ExpConc = \frac{\sum_{TimeBlocks} \sum_{Microenvironments} Conc_{t,m} \times Duration_{t,m}}{\sum_{TimeBlocks} \sum_{Microenvironments} Duration_{t,m}}$$

where:

$Conc_{t,m}$:   the emission-source-category concentration during time-block t in ME m, and

$Duration_{t,m}$:   the duration of activity during time-block t in ME m.

- The exposure concentrations for each day type are combined with weighted averaging to create an annual-average exposure concentration. The weights are the relative frequencies of the day types: 0.178 for summer weekday, 0.537 for other weekdays, and 0.285 for weekends.

- A total annual-average exposure concentration is calculated by summing the annual-average values for each emission-source category, the background contribution, and from the indoor-source *ADD* factor.

- The results are written into the *final exposure* output files, with **nreplic** records for each demographic group in each tract. The format of the files is described in Section 4.4 (*Final Exposure* File).

This page intentionally left blank.

6.   References

Graham, S., K. Isaacs, T. McCurdy, J. Langstaff, P. Hartman, C. Stevens, H. Hubbard, S. Hartley, J. Cohen, A. Bordner, C. Holder, N. Vetter, A.J Overton, I. Warren, C. Cavanagh, B. Luukinen, and W. Mitchell, 2019: The Consolidated Human Activity Database (CHAD) Documentation and Users' Guide. EPA-452/B-19-001. U.S. Environmental Protection Agency, Research Triangle Park, NC. https://www.epa.gov/sites/default/files/2019-11/documents/chadreport_october2019.pdf.

This page intentionally left blank.

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

This page intentionally left blank.

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] 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.

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 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.

---

[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.

2. Updating Census-based Data

2.1. 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 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.

2.1.1. 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 file[4] (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

---

[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.

ages. Lastly, we compared the HAPEM8 population file against the HAPEM7 file to ensure proper formatting.

2.1.2. 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 the first field identifies the tract using Census Federal Information Processing Series (FIPS) coding[5], 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 off all tracts) to 37,892, with an average of 3,919. The total population in this file is 334,822,301.

Table 1. 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 COM	B_02 COM	B_05 COM	B_16 COM	B_18 COM	B_65 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 1. Excerpt from the HAPEM8 Population File

2.2. 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. More specifically, we used the distGeo function in the geosphere package of R[8] 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.

2.2.1. 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.

2.2.2. 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 2. 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 2. 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).

2.3. 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 percent[9] 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.

2.3.1. 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

---

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

HAPEM8 default commuting-fraction file, discussed later in Section 2.4) had non-0 commuting-time values in the final file.

2.3.2. 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 3. 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 3. Excerpt from the HAPEM8 Commuting-time File

2.4. 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 (number of people aged 16+ years who commute to work other than by walking) ÷ (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).

2.4.1. 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.

2.4.2. 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 4. 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 4. Excerpt from the HAPEM8 Commuting-fraction File

2.5. 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 Shapefiles[10] 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 5. 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.

2.5.1. 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 <u>area</u> within the individual major-roadway buffers to equal the fraction of tract <u>population</u> 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.

2.5.2. 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 6. 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

Field Number	Description
21	Proportion of age group 6 (ages 65 and older) residing > 75 m of major roadway, up to 200 m
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 0.8395
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 0.8723
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 0.7613
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 0.6892
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 0.9305
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 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 0.8607
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 0.6583
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 0.8149
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 0.8638
```

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

Figure 5. 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 2020. 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.

3.1. 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.

3.2. 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 7. 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	

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

```
CHADID      ZIP    ST COU SEX RACE WORK YEAR MN DY   AGE   G DT CT
CAC01166A 93277 06 000 1    1    N    1989 6  16   1.67  1 1  1   60  0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    60   0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    60   0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    60   0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    60   0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    60   0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   60  0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    60  0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    60   0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    60   0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    60   0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    60   0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    60   0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   60  0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    30  0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    30   30   0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    30   60   0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    45   0    0    0    0    0    15   0    0    0   0   0   0   0
0    0    0    0    60   0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    35   0    25   0    0    0    0    0    0    0    0    0   0   0   60  0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    60  0   0   0   0
0    0    0    0    0    0    0    0    0    0    60   0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    60   0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0   0
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   0   0   0
```

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

Figure 6. Excerpt from the HAPEM8 Activity File

The HAPEM8 activity–cluster data are contained in a fixed–width, space–delimited text file with characteristics shown in Table 8. The file contains six columns and a data row 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 8. 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

[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 7. 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 9. 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

Variable Number	Variable	Description[a]	Character Start Position on Data Row	Character Length on Data Row[b]
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
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 8. Excerpt from the HAPEM8 Cluster–transition File

4. 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.