Decision Points for Configuring APEX for Air Toxics Exposure Assessments

Prepared by:

ICF Consulting
Research Triangle Park, NC

Under contract to:

U.S. EPA
Office of Air Quality Planning and Standards
Research Triangle Park, NC

December 12, 2005

FINAL

TABLE OF CONTENTS

[This page intentionally left blank.]

1. INTRODUCTION

This document describes the decisions involved in configuring the Air Pollutants Exposure (APEX) model for air toxics applications and provides some helpful information for making these decisions. The types of applications for which this document might be useful include those with the following characteristics:

- Spatial scale of the study area is local (e.g., single facility) to regional (e.g., multiple facilities geographically dispersed);

- Indoor sources do not contribute to exposures; and

- Exposure durations of interest are chronic and/or acute.

Note that these are the types of applications this document was designed to address and they are not necessarily limitations of APEX.

This document is not intended to replace the APEX User's Guide (EPA 2003); instead, it provides supplementary material that users may find useful in designing and setting up an APEX application for air toxics. Before using this document, users should familiarize themselves with the basic concepts and terminology presented in Chapter 2 of the APEX User's Guide. Note that all discussion of specific input file formats is based on the file formats for APEX Version 3.3.0; however, the general approach presented here should apply to all subsequent versions of APEX.

The second section of this document describes the steps involved in determining the scope of an analysis using APEX. The third section provides some general guidelines and other useful information for configuring APEX input files for air toxics exposure assessments. The final section provides the references for the document.

[This page intentionally left blank.]

2. DETERMINING SCOPE OF ANALYSIS

The first phase of setting up and performing an APEX analysis is to determine the scope of the analysis. For analyses involving air toxics, there are usually four steps involved in this phase:

- Selecting the modeling mode;
- Selecting the spatial resolution;
- Defining the study area(s); and
- Selecting commuting option.

The decisions made during these steps will dictate how the APEX input files will be configured (described in Chapter 3). The following sections provide information on what each of these steps involves and how users should approach them for air toxics analyses.

2.1 SELECT THE MODELING MODE

As described in the APEX User's Guide, APEX can be run in two different modes: DOS mode or TRIM (i.e., MIMS) mode. Users must decide which one of these modes best fits their needs. Note that regardless of which mode is selected, all of the APEX input files (with the exception of the *params* file) must be prepared in the same way for both modes. The advantages and disadvantages of both modes are summarized below.

2.1.1 DOS mode

Advantages

- Less software to install. This mode does not require the installation of the TRIM modeling system.

- No new software to learn. In DOS mode, users can set up APEX using only a simple text editor.

Disadvantages

- Cannot track exposures from multiple chemicals, sources, facilities, and study areas in a single simulation.

- If TRIM.Risk will be used to calculate risk metrics, additional processing of APEX outputs will be required if this mode is selected.

- No GUI. All inputs and model options are specified in text files.

2.1.2 TRIM mode

Advantages

- Provides a GUI for specifying inputs and model options.

- Can track exposures from multiple chemicals, sources, facilities, and study areas in a single simulation. The primary advantage of this feature is that the user would be able to create tables and figures comparing the exposures across chemicals, sources, facilities, and/or study areas using the Data Analysis and Visualization Engine (DAVE), which can be installed with the TRIM modeling system. Note that this feature can only be used when each unique source, facility, and study area combination is being modeled for the same year.

- Exposure estimates can be automatically used as inputs to TRIM.Risk to calculate various, user-specified risk metrics (using DOS mode, the exposure outputs would require additional processing before risk metrics could be calculated).

Disadvantages

- Requires installation of additional software (e.g., MIMS, MySQL). Note, however, that all of the required software is available at no charge from the Internet.

- For users unfamiliar with TRIM or MIMS, using the TRIM mode will require additional time getting familiar with navigating the MIMS environment and understanding the different options for running APEX in TRIM.

2.2   DEFINING STUDY AREAS

After selecting the appropriate modeling mode, users should next define their study area(s). There are two steps involved in defining a study area in APEX: (1) defining the center and radius of the study area and (2) defining the county list (optional) Both of these steps is described below. For analyses with more than one study area, these steps would need to be repeated for each one.

2.2.1 Study area centroid and radius

The first step in defining the study area is to specify the geographic center (centroid) and radius of the study area. APEX uses these two inputs, in combination with other information, to determine which sectors will be included in the study area. If the distance from a sector to the centroid is greater than the specified radius, the sector is not included in the study area. When specifying a value for the radius, users should consider the sector locations of interest, making sure that all locations of interest are captured within this radius.

The APEX study area has traditionally been on the scale of a city or slightly larger metropolitan area, although larger areas such as consolidated metropolitan statistical areas (CMSAs) have also been modeled. Even larger study areas are possible, depending primarily on computing capabilities, available data, and the desired precision of the run.

2.2.2 County list

Users have the option of providing a list of counties (using county FIPS codes) that is used to further restrict the size of the study area. If a county list is provided, any sectors within the study area radius but not within one of the listed counties will not be included the study area. Likewise, any sectors outside of the study area radius but within one of the listed counties will not be included in the study area.

2.2.3 Air districts and temperature zones

The study area can be further restricted based on the definition of the air districts and temperature zones. Sectors are excluded from the study area in they do not fall within at least one air district and one temperature zone. This is explained in more detail in Sections 2.3.2 and 2.3.3.

2.3 SELECTING SPATIAL RESOLUTION

After selecting a modeling mode and defining the study area(s), the next step is to select the spatial resolution for the application. Specifying the spatial resolution for an APEX application consists of selecting the resolution of the sectors, districts, and temperature zones. The **sector** is the fundamental spatial unit in APEX. For each sector the model requires demographic information representing the distribution of age, gender, race, and work status within the study population. Each sector must have a location specified by latitude and longitude for some representative point (e.g., centroid).

The spatial units for ambient air quality data are called **districts**. The districts are used to assign pollutant concentrations to the sectors and microenvironments being modeled. The ambient air quality data must be provided by the user as an hourly time series for one location (e.g., centroid) within each district. The locations could be monitoring sites, centers of census units such as blocks, or even receptors modeled by an air quality model. Each district is defined by its center location (using latitude and

longitude coordinates) and a radius that the air concentrations associated with that location could be applied.

The spatial unit for which temperature data are provided is called a temperature zone. The temperature zones are used to assign temperature data to the sectors and microenvironments being modeled. At a minimum, the daily maximum one-hour temperature is required; however, other daily temperature data (e.g., average) may also be required depending on how the user defines the microenvironments. Each temperature zone is defined by its center location (using latitude and longitude coordinates) and a radius that the temperature data associated with that location could be applied.

The selection of the appropriate spatial resolution is very important because it determines how well the model captures the spatial variability of the air quality data, temperature data, and population demographics. It also affects how some of the other APEX input files will be configured. The following sections describe how the resolutions for sectors and districts are specified.

2.3.1 Sectors

For APEX simulations, sectors are typically selected to be at the census block or census tract level. When selecting sector resolution, users should consider the goals of the analysis. If interested in the maximum exposures or "hotspots," users should consider using census block resolution. Census blocks are typically used for assessments where concentration gradients are likely to be steep and concentration peaks may be limited in spatial extent (e.g., single facility assessment). If interested in population exposures over larger study areas, users should consider using census tract resolution. Census tracts are typically used for assessments with smaller spatial differences in concentrations (e.g., secondary pollutant assessments).

Files containing sectors at the census tract level are available from the EPA website and can be used for any analysis in the continental U.S. Files containing sectors at the census block level can be created for a particular study area using the APEX Block File Builder, a utility available from the EPA website. Other sector sizes can be used, but significant setup time would be required to create the sectors files as well as a number of other files.

When deciding whether to use census tract or block data for sectors, users should consider whether they are interested in analyzing racial distinctions in their results. The census tract data include population counts for each race by age, whereas census block data include population counts by age and population counts by race but not population counts for each race by age.

2.3.2 Districts

Unlike the sector locations, there are no prepackaged district locations accompanying APEX. When deciding the appropriate location and radius of the air districts, the user should attempt to define the air districts so that the concentration at the

center of each district is approximately the average of the concentrations across the district. Users are provided significant flexibility for specifying district center locations. Typically, however, they are either modeled receptor locations or the locations of monitors. For example, if air quality model outputs were generated at census block centroids, the district locations would be set to these centroid locations.

The air district radius defines the maximum distance (in km) from an air district center location (e.g., monitor, modeled receptor location) that the air concentrations associated with that location could be applied. APEX matches each sector in the study area to the closest air district center location, provided at least one air district location is within the air district radius from the sector center location. **If no air district locations are within this radius from the sector location, the sector will not be included in the study area.** Note that air districts can reside outside of the study area and still be matched with sectors.

For applications where the sector and air district center locations are the same (e.g., census block centroids), the air district radius is not important because all districts will be automatically matched with sectors at the same location (i.e., census block centroid). In this case, the user could set the radius to some small value (e.g., 0.1 km). Otherwise, users should consider the chemical(s) they are modeling and the desired spatial resolution when selecting a value for this parameter.

2.3.3   Temperature zones

In APEX, temperature data can be used to assign human activity patterns or to assign conditional probability distributions to microenvironmental parameters. As was the case with air districts, there are no prepackaged temperature zone locations accompanying APEX. Users are provided significant flexibility in defining temperature zone locations. These locations typically correspond to the location of a meteorological station. The temperature zone radius defines the maximum distance (in km) from a temperature zone location (e.g., meteorological station) that the temperature data associated with that location could be applied. APEX matches each sector in the study area to the closest temperature zone location, provided at least one temperature zone location is within the temperature zone radius from the sector location**. If no temperature zone locations are within this radius from the sector location, the sector will not be included in the study area.** Note that, as with air districts, a temperature zone location can reside outside of the study area and still be matched with sectors.

When selecting temperature zones for an application, users should consider the variability in temperatures across their study area and the temperature cut points defined for the application (e.g., used for selecting activity patterns or air exchange rates). If temperatures do not typically vary significantly across the study area, a single temperature zone may be sufficient. In this case, the temperature zone radius must be set large enough to encompass the entire study area. For study areas where the variability in temperatures is larger (e.g., larger study areas or coastal areas), users should consider using multiple temperature zones. In this case, the temperature zone radius must be

specified such that each sector in the study area is within this radius of a temperature zone location.

2.4   SELECTING COMMUTING OPTION

Users have the option of modeling home-to-work commuting in their APEX analysis. When selected, this option allows the modeled individuals to move from one sector (home sector) to another (work sector) and back again based on employment probabilities and commuting pattern data. If users select this option, they must provide two additional inputs: the *employment probabilities* file and the *commuting flow* file. If this option is not selected, modeled individuals are assumed to remain in their home sector for the duration of the simulation.

It is recommended that users select this option for their simulations. It provides a more accurate representation of exposures of the modeled population and very little additional effort is required to configure the required input files (as described in Chapter 3).

3. CONFIGURING INPUT FILES

After determining the scope of the assessment, the next step in performing an APEX analysis is to configure the required input files. The APEX input files fall into the following general categories:

- Human activity and exposure factors data files;
- Population data files;
- Microenvironment files;
- Air quality data files;
- Temperature data files; and
- Simulation setting file.

Figure 1 provides an overview of how these files are developed and how the files are related to each other. Files in ***red italics*** are input files available from the EPA website; the user must create all of the remaining input files. The following sections briefly describe each of these file types and provide information that may be helpful to a user in configuring these files for an air toxics exposure assessment.

3.1 HUMAN ACTIVITY AND CHARACTERISTICS DATA

There are five APEX input files that provide human activity and characteristics data:

- *Personal info* file;
- *Diary events* file;
- *Activity-specific METS* file;
- *Physiological parameters* file; and
- *Ventilation input* file.

The *personal info*, *diary events*, and *activity-specific METS* files are derived from the Consolidated Human Activity Database (CHAD) and are available from the EPA website in "ready-to-run" format for any application in the continental U.S. Users should not need to perform any additional processing of these input files.

The *physiological parameters* and *ventilation input* files are used by APEX for criteria pollutants. However, although they are not used internally by APEX to perform any calculations for air toxics, these files are required inputs and APEX will produce an error if they are not provided. The files provided on the EPA website are appropriate for any air toxics applications in the continental U.S.

Figure 1. Overview of APEX Input Files
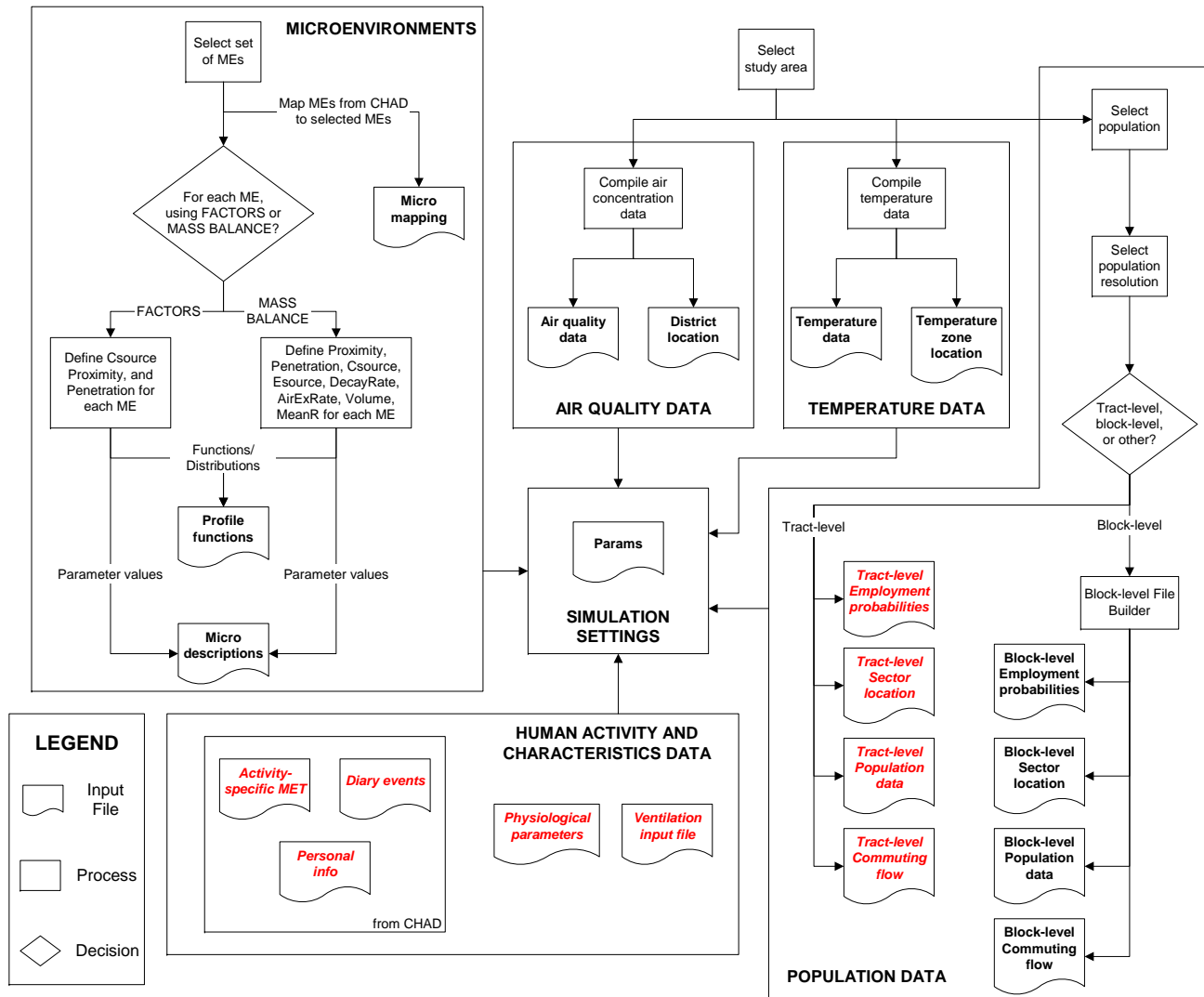
3.2   POPULATION DATA

There are four types of APEX input files that provide information on the population of interest:

- *Sector location* file;
- *Population counts* files;
- *Employment by age group* file; and
- *Commuting flow* file.

The content of these files varies based on the scoping decisions made by the user. APEX is very flexible in the resolution of population data that can be used. As long as the data are in the appropriate format, and the s*ector location*, *population counts, employment by age group*, and c*ommuting flow* files use the same set of sectors, any resolution can be used (e.g., county, census tract level, etc.). However, typically either census tract- or block-level resolution data are used. Users may choose to use population data at a resolution other than tract or block, but this would involve a significant amount of additional work. The remainder of this section briefly describes these files and the steps involved in creating them at tract- and block-level resolution.

Note about the APEX Block File Builder

When users run the APEX Block File Builder for a particular study area, all of the required population data files (*sector location*, *population estimates*, *employment by age group*, and *commuting flow*) are created for the study area in a single run of the processor. There is no need to run this processor more than once for a particular study area.

3.2.1   Sector location

This file contains a list of the locations of all of the potential sectors in the study area. As described in Chapter 2, the list of sectors included in the study area begins with the sectors in the sector location file. Users can limit the sectors included in the study area using the specified locations and radii of the study area, air districts, and temperature zones, as well as (optionally) the list of counties.

Block-level resolution: If the user wishes to use block-level population resolution, a *sector location* file can be created using the APEX Block File Builder. Users specify the county or counties they wish to include in their study area and the processor creates a "ready-to-run" *sector location* file containing data for these counties (from the 2000 U.S. Census). If the user is only interested in select census blocks within a county, they should select this county and then limit the sectors included in the study area using the study area radius, air districts, and/or temperature zones.

Tract-level resolution: If the user wishes to use tract-level population resolution, the *sector location* file available from the EPA website can be used. This file contains tract-level data for the entire continental U.S. (from the 2000 U.S. Census) and is provided in "ready-to-run" format – it should not require any additional processing.

3.2.2 Population counts

The *population counts* files contain the number of people (by age group) in each sector in the sector location file. Each population file contains these population counts for a specific gender-race combination. Depending on the application, users must specify one or more *population counts* files. Which files are specified depends on the selected resolution and the gender(s) and race(s) of interest. If block-level data are selected, two *population counts* files (one for males, one for females) are available; these files do not distinguish between races. If tract-level data are selected, 10 *population counts* files (one for each unique combination of five races and two genders) are available.

Block-level resolution: If the user wishes to use block-level population resolution, two *population counts* files (one for males, one for females) can be created using the APEX Block File Builder. Population counts for each sector are provided in one-year increments (age groups), from one to 100. Users specify the county or counties they wish to include in their study area and the processor creates two "ready-to-run" *population counts* files containing data for these counties (from the 2000 U.S. Census). If the user is only interested in select census blocks within a county, they should select this county and then limit the sectors included in the study area using the study area radius, air districts, and/or temperature zones. Note that these files will not identify population counts by race. If analyzing racial distinctions among exposure estimates is of interest, block-level data should not be used.

Tract-level resolution: If the user wishes to use tract-level population resolution, the ten *population counts* files available from the EPA website can be used. Population counts for each sector are provided in one-year increments (age groups), from one to 100. These files contain tract-level data for the entire continental U.S. (from the 2000 U.S. Census) and are provided in "ready-to-run" format – they should not require any additional processing. These files provide data for males and females of the following five races:

- White;
- Black or African American;
- American Indian or Alaska Native;
- Asian and Pacific Islander; and
- Other.

3.2.3   Employment by age group

The *employment by age group* file contains employment probabilities by sector. The file is broken down by gender and age group, so that each gender/age group combination is given an employment probability fraction (ranging from 0 to 1) within each sector.

Block-level resolution: If the user wishes to use block-level population resolution, an *employment by age group* file can be created using the APEX Block File Builder. Users specify the county or counties they wish to include in their study area and the processor creates a "ready-to-run" *employment by age group* file containing data for these counties (from the 2000 U.S. Census). If the user is only interested in select census blocks within a county, they should select this county and then limit the sectors included in the study area using the study area radius, air districts, and/or temperature zones. The age groupings used in the created file are: 16-19, 20-21, 22-24, 25-29, 30-34, 35-44, 45-54, 55-59, 60-61, 62-64, 65-69, 70-74, and >75. It is assumed that all individuals younger than 16 have zero probability of being employed.

Tract-level resolution: If the user wishes to use tract-level population resolution, the *employment by age group* file available from the EPA website can be used. This file contains tract-level data for the entire continental U.S. (from the 2000 U.S. Census) and is provided in "ready-to-run" format – it should not require any additional processing. The age groupings used in this file are: 16-19, 20-21, 22-24, 25-29, 30-34, 35-44, 45-54, 55-59, 60-61, 62-64, 65-69, 70-74, and >75. It is assumed that all individuals younger than 16 have zero probability of being employed.

3.2.4   Commuting flow

The *commuting flow* file contains the fraction of the employed population in each sector that commutes within their home sector or to another sector in the study area. Each row of the file contains a home sector, a work sector, and the fraction of workers in the home sector that commute to the work sector.

Block-level resolution: If the user wishes to use block-level population resolution, a *commuting flow* file can be created using the APEX Block File Builder. Users specify the county or counties they wish to include in their study area and the processor creates a "ready-to-run" *commuting flow* file containing data for these counties (from the 2000 U.S. Census). If the user is only interested in select census blocks within a county, they should select this county and then limit the sectors included in the study area using the study area radius, air districts, and/or temperature zones. This file is created using the tract-level commuting file and available land use data, so the assumptions made in creating the tract-level file apply to these data as well.

Tract-level resolution: If the user wishes to use tract-level population resolution, the *commuting flow* file available from the EPA website can be used. This file contains tract-level data for the entire continental U.S. (from the 2000 U.S. Census) and is provided in "ready-to-run" format – it should not require any additional processing.

3.3   MICROENVIRONMENTS

APEX estimates inhalation exposures by simulating the movement of individuals through time and space. At each time step, each modeled individual is assigned to a single sector (either home or work). This sector is associated with a single district, which provides the ambient air concentrations to which the individual may be exposed. However, because individuals do not spend all of their time outdoors exposed to ambient air, APEX accounts for the differences in exposure concentrations associated with different locations (known as **microenvironments**) within each sector. Users must define these microenvironments based on: (1) the 115 locations tracked in the Consolidated Human Activity Database (CHAD) and (2) any application-specific considerations. The following three input files are required by APEX to define the microenvironments for an application:

- *Micro mapping* file;
- *Micro description* file; and
- *Profile functions* file.

The remainder of this section provides basic information on how to select microenvironments; brief descriptions of the three microenvironment files and information that will help users configure these files for an air toxics application; and potential data sources that can be used to configure these files.

3.3.1   How to select microenvironments

The selection and definition of microenvironments is one of the most important (and time-consuming) parts of configuring an APEX application. Users must consider how pollutant concentrations will vary between different environments based on chemical properties (degradation rates, particle size, etc.) and properties of the different environments of interest for the analysis (e.g., for indoor environments, is ambient air filtered before being brought inside?).

The first step in defining microenvironments is to select the microenvironments of interest for the analysis. A good place to start is the list of 115 locations tracked in the CHAD database. Each of these locations must be assigned to a user-selected microenvironment. Users should examine these 115 locations and determine how they can be combined into groups. This determination should be based on whether the user believes that pollutant concentrations in each of a set of locations would be approximately the same if they were all located in the same air district. For example, the CHAD defines 10 different locations within an individual's home. For most air toxics applications with no indoor sources, the differences in concentrations between these

locations will not be significant, allowing the user to combine these locations into a single microenvironment. APEX applications to date have generally defined between 10 and 40 microenvironments; however, users can certainly define more or fewer microenvironments depending on the needs of their assessment. The *micro mapping* file is created based on the selected microenvironments and how they match with the CHAD locations (as described in Section 3.3.2).

The next step in defining microenvironments is specifying how exposure concentrations will be calculated for each microenvironment. For each microenvironment, users must first decide whether to use the MASS BALANCE or FACTORS method for calculating microenvironment concentrations. The differences between these methods are described in the APEX User's Guide. Users should keep in mind that there are no restrictions on which method can be used for a particular type of microenvironment. After selecting the calculation method, users must provide values for the inputs required by the selected method. Each method has different data requirements, so users should consider the availability of these data when selecting a method. The *micro descriptions* file is used to define how concentrations in each microenvironment will be calculated (including which method should be used) and the *profile functions* file is used to define functions that are used in the *micro descriptions* file. These files are discussed in Section 3.3.3 and 3.3.4, respectively.

3.3.2   Microenvironment mapping

The *micro mapping* file is used to match each of the 115 locations in the CHAD to one of the user-defined microenvironments. The easiest way to create this file is to start with an existing *micro mapping* file (e.g., from one of the APEX applications available from http://www.epa.gov/ttn/fera) and edit it to correspond to the microenvironments defined for the analysis. This prevents the user from needing to enter each CHAD location and reduces the possibility of typographical errors. However, before using an existing file as a template, users should confirm that this file was created based on the most recent version of the CHAD (i.e., contains the same 115 locations).

3.3.3   Microenvironment descriptions

The *micro descriptions* file is used to define how exposure concentrations will be calculated for each user-defined microenvironment. Users must specify the name of the microenvironment, which calculation method should be used, and the parameter values that will be used to calculate the exposure concentrations in the microenvironment. These parameter values are defined as distributions from which APEX will randomly sample for each time step. For microenvironments using the MASS BALANCE method, distributions must be defined for eight parameters, and for those using the FACTORS method, distributions must be defined for three parameters. Different distributions can be used for the same parameter based on the time of day, day of the week, and/or season. In addition, the parameter values can be specified as dependent on other user-defined functions contained in the *profile functions* file (described in Section 3.3.4). If the default value is used for any of these parameters (provided in the APEX User's Guide), they do not need to be included in the *micro descriptions* file; APEX assumes the default if no

distribution is specified. Some basic information about specifying values for air toxics assessments for the parameters required by each calculation method is provided in Table 1. Refer to the APEX User's Guide for more information on configuring these files.

Table 1. Setting Parameters in the *Micro Descriptions* File for Air Toxics

Parameter	How to Set Values for Air Toxics Assessments
	Mass Balance Method
Proximity	Accounts for the proximity of the microenvironment to an emission source. When setting this value, users should consider the sources of interest and the spatial resolution being used. For applications using census block resolution and where on-road mobile sources are not of interest, the default value of 1 is generally appropriate. If census tract resolution is used or if on-road mobile sources are important for the application, a Proximity value other than 1 might be appropriate.
Penetration	Can be accounted for using air exchange rate (AirExRate). If this is done, the default value of 1 should be used.
CSource	If the user is only interested in the impacts of sources captured by the ambient concentrations, CSource should be set to zero.
DecayRate	This parameter should be set based on the degradation of the pollutant in the microenvironment. If the pollutant degrades very slowly or if the user wishes to perform a more conservative (i.e., risk-protective) assessment, the DecayRate can be assumed to be zero.
AirExRate	Must be specified by the user.
Volume	Used in calculating contribution of indoor sources. If the user is only interested in the impacts of sources captured by the ambient concentrations, Volume can be set to zero.
MeanR	By default MeanR is defined as AirExRate + DecayRate. For most applications, the default value is appropriate.
ESource	If the user is only interested in the impacts of sources captured by the ambient concentrations, ESource can be set to zero.
	Factors Method
Proximity	Accounts for the proximity of the microenvironment to an emission source. When setting this value, users should consider the sources of interest and the spatial resolution being used. For applications using census block resolution and where on-road mobile sources are not of interest, the default value of 1 is generally appropriate. If census tract resolution is used or if on-road mobile sources are important for the application, a Proximity value other than 1 might be appropriate.
Penetration	Accounts for the presence of barriers that might change the concentration of pollutant within the microenvironment compared to the ambient concentration at the district location. Generally for outdoor microenvironments, the Penetration can be assumed to be 1.
CSource	If the user is only interested in the impacts of sources captured by the ambient concentrations, CSource should be set to zero.

3.3.4 Profile functions

The *profile functions* file contains the user-defined functions that can be used in the *micro descriptions* file. These functions are used to specify different distributions for one of the required parameters for a microenvironment based on factors that may affect exposures (e.g., presence of air conditioning). There are currently 10 built-in functions that can be defined in this file:

- MaxTempCat;
- AvgTempCat;
- Diary Pools;
- IDGRP;
- HasGasStove;
- HasPilot;
- AC_Home;
- AC_Car;
- WindowPos; and
- SpeedCat.

Users must define each of these functions regardless of whether they are used in the *micro descriptions* file. However, if they are not used, the specified values do not impact the results. Users should keep in mind that several of the required functions were originally designed for criteria pollutant applications (e.g., HasGasStove, HasPilot) and may not apply to air toxics assessments. These functions must be parameterized but might not be used.

Users can define up to three additional functions to account for other factors that may affect exposures. Any of the 13 available functions (10 required, 3 user-specified) can be defined using values from another function. For example, the WindowPos function could be a function of MaxTemp, AvgTemp, and AC_Home (as in the example provided in the APEX User's Guide). Thus, functions included in the *profile functions* file may be used without being referenced directly in the *micro descriptions* file. Refer to the APEX User's Guide for more information on configuring this file.

3.3.5 Potential data sources

Several excellent data sources exist that can be used to develop inputs for the microenvironments. The U.S. EPA is planning to use APEX to estimate ozone exposure in 12 cities or metropolitan areas in the United States (EPA 2005). As part of this effort, distributions of residential and non-residential air exchange rates (AER) for indoor microenvironments, as well as proximity and penetration factors for outdoor and in-vehicle microenvironments, were developed as APEX inputs for the cities to be modeled. These AER distributions and proximity and penetration factors are independent of the chemical being modeled and thus may be appropriate for use in a wide variety of analyses using APEX. The analyses used to develop these distributions are briefly described below.

Indoor AERs

EPA's air exchange analysis demonstrated that the AER distributions vary greatly across cities and with differences in A/C usage and ambient temperatures; thus, the AER distributions developed were dependent upon these three variables. The studies used to develop these AER distributions included the following:

Avol: Avol et al, 1998. In this study, ozone concentrations and AERs were measured at 126 residences in the greater Los Angeles metropolitan area between February and December, 1994. Measurements were taken in four communities: Lancaster, Lake Gregory, Riverside, and San Dimas.

RTP Panel: Williams et al, 2003a, 2003b. In this study particulate matter concentrations and daily average AERs were measured at 37 residences in central North Carolina during 2000 and 2001 (averaging about 23 AER measurements per residence).

RIOPA: Meng et al, 2004, Weisel et al, 2004. The Relationship of Indoor, Outdoor, and Personal Air (RIOPA) study was undertaken to estimate the impact of outdoor sources of air toxics to indoor concentrations and personal exposures.

TEACH: Chillrud at al, 2004, Kinney et al, 2002, Sax et al, 2004. The Toxic Exposure Assessment, a Columbia/Harvard (TEACH) study was designed to characterize levels of and factors influencing exposures to air toxics among high school students living in inner-city neighborhoods of New York City and Los Angeles, CA.

Wilson 1984: Wilson et al, 1986, 1996. In this 1984 study, AER and other data were collected at about 600 southern California homes with three seven-day tests (in March and July 1984, and January, 1985) for each home.

Wilson 1991: Wilson et al, 1996. Colome et al, 1993, 1994. In this 1991 study, AER and other data were collected at about 300 California homes with one two-day test in the winter for each home.

Murray and Burmaster: Murray and Burmaster (1995). For this article, Murray and Burmaster corrected and compiled nationwide residential AER data from several studies conducted between 1982 and 1987. These data were originally compiled by the Lawrence Berkeley National Laboratory.

To estimate AER distributions for non-residential, indoor environments (e.g., offices and schools), two AER data sets were analyzed: "Turk" (Turk et al, 1989) and "Persily" (Persily and Gorfain 2004; Persily et al. 2005). The more recent "Persily" data were derived from the U.S. EPA Building Assessment Survey and Evaluation (BASE) study, which was conducted to assess indoor air quality, including ventilation, in a large number of randomly selected office buildings throughout the U.S.

Outdoor and In-vehicle Factors

For the outdoor and in-vehicle microenvironmental proximity and penetration factors, the Cincinnati Ozone Study (American Petroleum Institute, 1997, Appendix B; Johnson et al. 1995) is an excellent source of data. The study is particularly useful for in-vehicle data, as the study specifies the vehicle type, road type, vehicle speed, and ventilation mode.

3.4   AIR QUALITY DATA

There are two APEX input files that provide air quality data: the *district location* file and the *air quality data* file. These files provide the locations of air concentration data (districts) and hourly air concentrations at each of these locations. APEX allows users a lot of flexibility in defining these districts. They can correspond to monitoring stations, locations modeled using an air quality model, or any other location with available air concentration estimates. Note that not all districts in these files need to be paired with a sector. Any districts that are not paired with at least one sector are simply not used in the simulation. This simplifies the process of preparing these files because users do not need to determine which districts will match with which sectors prior to starting to simulation. The remainder of this section briefly describes these files and the steps involved in creating them for an air toxics exposure assessment.

3.4.1   District location

The *district location* file provides the locations (latitude and longitude) of each air district and the dates during which air quality data are provided for each district in the *air quality data* file. If the ISCST3 model is being used to generate air concentration data, users may find it convenient to model concentrations at census block centroids and then use these locations as air districts in APEX. This would allow them to use the ISCAPEX utility (available for download from http://www.epa.gov/ttn/fera) to create the *district location* and *air quality data* files. For instructions on using ISCAPEX, refer to the README file accompanying the software.

If air concentrations are generated using a model other than ISCST3 or measurement data will be used, users will need to create the *district location* file in the format specified in the APEX User's Guide.

3.4.2   Air quality data

The *air quality data* file provides hourly air concentrations for each of the specified district locations. For simulations with multiple chemical/source/facility/study area combinations, a separate air quality file must be provided for each combination. Note, however, that all air quality files for a particular simulation must be for the same year. Available utilities for converting air quality model output into APEX *air quality data* file format are described in Section 3.4.3. If users do not use one of these utilities, they must create the *air quality data* file manually in the format specified in the APEX User's Guide.

3.4.3 Utility programs to convert air quality data

Utility programs can be used to convert air concentration data generated by air quality models into the *air quality data* file format required by APEX. Currently, a utility program is available for converting outputs from ISCST3 into the format required by APEX. This program can be used to convert ambient air concentrations estimated by ISCST3 at census block centroids into the *air quality data* file format, provided the appropriate ISCST3 output options are specified when the model is run (see ISCAPEX README file). ISCAPEX also creates the *district location* file in the appropriate format. No other utilities are currently available to assist in converting *air quality data* files; however, additional utilities may be made available in the future to convert outputs from other air quality models (e.g., AERMOD, CMAQ) into the format required by APEX.

3.5 TEMPERATURE DATA

There are two files that are used to input temperature data to APEX: the *temperature zone location* file and the *temperature data* file. The *temperature zone location* file defines the locations of the temperature zones and the *temperature data* file contains the required temperature data for each of these zones. For typical air toxics analyses, temperature data from a single meteorological data station are sufficient for an APEX simulation. However, if temperatures vary significantly over the study area users may choose to include more than one temperature zone.

3.5.1 Temperature zone location

The *temperature zone location* file contains the locations of one or more meteorological data stations, expressed in latitude and longitude coordinates. The file also contains the dates during which temperature data are provided for each zone in the *temperature data* file. For most air toxics analyses, this file will contain the location of one temperature zone and the start and end dates will be specified as the start and end dates of the simulation (indicating that the data for this zone cover the entire modeling period).

3.5.2 Temperature data

The *temperature data* file contains the daily maximum and daily average temperature readings for the period being modeled from the location(s) described in the temperature zone location file. For most air toxics analyses, this file will contain temperature data covering the entire modeling duration for one temperature zone. If the required air quality data were generated using an air quality model, the meteorological input file used by the model may contain the necessary data for compiling the *temperature data* file. If not, this type of meteorological data is available for locations around the United States from the National Climatic Data Center (http://www.ncdc.noaa.gov).

3.6   SIMULATION SETTINGS

The input and output file locations, simulation settings, and output table settings for an APEX run are specified in the *params* file. This section discusses each of the simulation settings and provides information that may be useful in specifying values for these parameters for air toxics exposure assessments. The variable name for each parameter is provided in ***bold italics*** (with the keyword used in the *params* file in parentheses), followed by information on that parameter. More complete descriptions of these parameters can be found in the APEX User's Guide.

Pollutant (POLL). This parameter specifies the name of the pollutant being modeled. It is only used by APEX for labeling outputs and can be set to any name (less than 40 characters long) selected by the user.

InUnits (INPUTUNIT). This parameter specifies the units of the concentration data in the *air quality data* file. It must be either "ppm" or "ug/m3". For most air toxics applications, it should be set to ug/m3 as these are the units typically output by air quality models.

OutUnits (OUTPUTUNIT). This parameter specifies the units of the exposure concentrations that will be generated by APEX. It must be either "ppm" or "ug/m3". For most air toxics applications, it should be set to ug/m3 as these are the units typically used for calculating risks and hazards.

Location (LOCATION). This parameter specifies the name of the location being modeled. It is only used by APEX for labeling outputs and can be set to any name (less than 40 characters long) selected by the user.

Scenario (SCENARIO). This parameter is used to describe the scenario being modeled. It is only used by APEX for labeling outputs and can be set to any name (less than 40 characters long) selected by the user.

#Profiles (PROFILE). This parameter specifies the number of modeled individuals that will be used in the simulation. It should be set to a value large enough to generate a statistically valid representation of the population being modeled. Test simulations performed by EPA indicate that values between 20,000 and 35,000 may be sufficient for most applications. Note, however, that more profiles may be needed if exposure outputs will be used in TRIM.Risk to calculate some population-based risk metrics.

#Micros (MICRO). This parameter indicates the number of microenvironments to be modeled in this simulation. This value should correspond to the number of microenvironments in the *micro mapping* and *micro descriptions* files.

#Sources (SOURCE). This parameter indicates the maximum number of concentration sources in any microenvironment. This parameter should be set to zero unless the user specified any concentration sources in the *micro descriptions* file.

Start_Date (START_). This parameter indicates the start date for the simulation.

End_Date (END_). This parameter indicates the end date for the simulation.

Latitude (LATIT). This parameter indicates the latitude (in decimal degrees) of the center of the study area.

Longitude (LATIT). This parameter indicates the longitude (in decimal degrees) of the center of the study area.

CityRadius (CITYRAD). This parameter indicates the radius (km) of the study area. Any sectors outside of this distance from the center of the study area are not included in the simulation.

AirRadius (AIRRAD). This parameter indicates the maximum application radius (km) of air quality data from an air quality station location (i.e., district centroid). Any sectors that are not within this distance of at least one air district are not included in the simulation. If census blocks are used for sector **and** district locations, this parameter can be set to some small number (e.g., 0.1) because all sectors will be co-located with a district.

ZoneRadius (ZONERAD).   This parameter indicates the maximum application radius (km) of temperature data from a temperature station location (i.e., zone centroid). Any sectors that are not within this distance of at least one temperature zone are not included in the simulation. Users must make sure that all sectors they wish to include in the simulation are within this radius from at least one temperature zone location.

CountyList (COUNTYL). This parameter indicates whether the user will provide a list of counties to which the study area will be restricted. If it is set to YES, the user must provide a list of counties (using the ***County*** parameter) and then only sectors included in one of the listed counties will be included in the simulation. The default value for this parameter is NO.

County (COUNTY). This parameter is used to specify a list of counties (using FIPS codes) to which the study area will be limited. It can be repeated multiple times to add multiple counties (i.e., a separate line in the *params* file should be used for each county in the list).

Commuting (COMMUT). This parameter is used to specify whether home-to-work commuting should be simulated. If this is set to YES, the user must provide a *commuting flow* file using the same resolution as selected for the sectors. If it is set to NO, all workers are assumed to work in their home sector.

AgeMin (AGEMIN). This parameter specifies the minimum age of the modeled individuals. For most air toxics applications, it is set to zero. However, if users wish to

limit the modeled individuals persons older than a particular age, this parameter could be set to a higher number.

AgeMax (AGEMAX). This parameter specifies the maximum age of the modeled individuals. For most air toxics applications, it is set to 99 (to correspond with the maximum age in the population data files). However, if users wish to limit the modeled individuals persons younger than a particular age, this parameter could be set to a lower number.

DSTAdjust (DST). This parameter is used to specify whether Daylight Savings Time (DST) should be accounted for in processing the ambient air concentrations provided as input to APEX. If the application being performed is for an area that uses DST, users should consider setting this parameter to YES.

HourlyOut (HOURLY). This parameter is used to specify whether the model should output hourly exposure and (if modeled) dose files. This parameter is typically set to NO because these output files can be quite large and are not necessary for assessing chronic exposures and risks.

DoDose (DODOSE). This parameter is used to specify whether dose calculations should be performed. APEX can currently only calculate doses for CO; thus, for air toxics assessments this parameter should be set to NO.

RollBack (ROLLBACK). This parameter is used to specify whether to use air quality rollback adjustments. Using this option, users can adjust the concentrations in the *air quality data* file before exposure concentrations are calculated. It allows users to estimate the impact of various control options that would be used to reduce ambient air concentrations. Because most air toxics assessments are based on facility emissions (as opposed to ambient air concentrations), this option should be set to NO for most air toxics assessments.

RbTarget (RBTARGET). This parameter is used to specify the target concentration for rollback adjustments. If the rollback option is not selected, it should be set to zero.

RbBackgnd (RBBACK). This parameter is used to specify the background concentration for rollback adjustments. If the rollback option is not selected, it should be set to zero.

RbMax (RBMAX). This parameter is used to specify the maximum concentration for rollback adjustments. If the rollback option is not selected, it should be set to zero.

PPMFactor (PPM). This parameter is used to specify the conversion factor between PPM and ug/m3. It is used when source strengths are expressed in ug/m3 but air concentrations are in ppm and when InUnits and OutUnits are in different units. For most air toxics assessments, all concentrations will be specified as ug/m3 and this parameter can be set to one.

MissGender (MISSGEND). This parameter specifies the probability that a CHAD profile with no specified gender will be selected. If it is set to zero, diaries with missing

genders will never be selected. If it is set to one, diaries with missing genders are as likely to be selected as diaries with specified genders. It can be set to a small non-zero value (e.g., 0.1) to increase the size of the pool of diaries without permitting much chance of selecting a diary with missing gender (which would cause APEX to stop prematurely).

MissEmployment (MISSEMP). This parameter specifies the probability that a CHAD profile with no specified employment status will be selected. If it is set to zero, diaries with missing employment status will never be selected. If it is set to one, diaries with missing employment status are as likely to be selected as diaries with specified employment status. It can be set to a small non-zero value (e.g., 0.1) to increase the size of the pool of diaries without permitting much chance of selecting a diary with missing employment status (which would cause APEX to stop prematurely).

MissAge (MISSAGE). This parameter specifies the probability that a CHAD profile with no specified age will be selected. If it is set to zero, diaries with missing ages will never be selected. If it is set to one, diaries with missing ages are as likely to be selected as diaries with specified ages. It can be set to a small non-zero value (e.g., 0.1) to increase the size of the pool of diaries without permitting much chance of selecting a diary with missing age (which would cause APEX to stop prematurely).

AgeCutPCT (AGECUT). This parameter is used to specify the range of years around the modeled individuals age from which diary data can be selected. It is specified as the percentage of the modeled individual's age and a single value is applied to all modeled individuals. The resulting range of years is referred to as the "main age window" and is calculated by multiplying the value of this parameter by the age of the modeled individual. For example, for the age of a modeled individual is 40 and AgeCutPCT is 25, the main age window would be 10 years (40 x 25%) and APEX would select diaries for persons 30 to 50 for this modeled individual. Larger AgeCutPCT values result in a larger pool of diaries from which activity data are selected for each individual.

Age2Probability (AGE2PROB). This parameter is used to specify the probability that activity data will be selected from the optional "shoulder" ages. The shoulder windows of ages are immediately outside of the main age window and have the same width as the main age windows. Using the example used for AgeCutPCT, the shoulder ages would be 20-29 and 51-60. If it is set to zero, diaries with from the shoulder ages will never be selected. If it is set to one, diaries from the shoulder ages are as likely to be selected as diaries in the main age window. It can be set to a small non-zero value (e.g., 0.1) to increase the size of the pool of diaries without allowing APEX to select too many diaries in the shoulder ages.

Altitude (ALTITUDE). This parameter is only used for dose calculations for CO and is therefore not relevant to air toxics assessments. It can be set to zero for all pollutants other than CO.

COHbFactor (COHBFACT). This parameter is only used for dose calculations for CO and is therefore not relevant to air toxics assessments. It can be set to zero for all pollutants other than CO.

COThreshold (COTHRESH). This parameter is only used for dose calculations for CO and is therefore not relevant to air toxics assessments. It can be set to zero for all pollutants other than CO.

DebugLevel (DEBUG). This parameter is used to specify whether diagnostic information will be output to the log file. It does not impact any of the APEX calculations.

Randomseed (RANDOM). This parameter is used drive the selection of all of the probabilistically sampled data (e.g., profiles, activity diaries). If it is set to zero, the random seed to generated dynamically by APEX using the computer's internal clock. Otherwise, APEX uses the user-supplied random seed. If two runs use the same population data files and the same random seed, they will use the same modeled individuals. If two runs use all of the same inputs but different random seeds, the combined results from these runs is equivalent to a single run with the total number of modeled profiles equal to the sum of the profiles used in these two runs.

4.   REFERENCES

American Petroleum Institute (API) 1997. *Sensitivity testing of pNEM/O3 exposure to changes in the model algorithms.* Health and Environmental Sciences Department.

Avol, E. L., W. C. Navidi, and S. D. Colome 1998. Modeling ozone levels in and around southern California homes. *Environ. Sci. Technol.* 32, pp. 463-468.

Chilrud, S. N., D. Epstein, J. M. Ross, S. N. Sax, D. Pederson, J. D. Spengler, P. L. Kinney 2004. Elevated airborne exposures of teenagers to manganese, chromium, and iron from steel dust and New York City's subway system. *Environ. Sci. Technol.* 38, pp. 732-737.

EPA 2005. Ozone Population Exposure Analysis for Selected Urban Areas (draft). Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. November. Available electronically on the internet at: http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_cr_td.html.

EPA 2003. Total Risk Integrated Methodology, TRIM.Expo$_{Inhalation}$ User's Document. Volume I: Air Pollutants Exposure Model (APEX, version 3) User's Guide. Office of Air Quality Planning and Standards. Research Triangle Park, NC. April.

Johnson, T., A. Pakrasi, A. Wisbeth, G. Meiners, W. M. Ollison 1995. Ozone exposures within motor vehicles – results of a field study in Cincinnati, Ohio. *Proceedings 88th annual meeting and exposition of the Air & Waste Management Association, June 18-23, 1995.* San Antonio, TX. Preprint paper 95-WA84A.02.

Meng, Q. Y., B. J. Turpin, L. Korn, C. P. Weisel, M. Morandi, S. Colome, J. J. Zhang, T. Stock, D. Spektor, A. Winer, L. Zhang, J. H. Lee, R. Giovanetti, W. Cui, J. Kwon, S. Alimokhtari, D. Shendell, J. Jones, C. Farrar, S. Maberti 2004. Influence of ambient (outdoor) sources on residential indoor and personal PM$_{2.5}$ concentrations: Analyses of RIOPA data. *Journal of Exposure Analysis and Environ Epidemiology*. Preprint.

Murray, D. M. and D. E. Burmaster 1995. Residential Air Exchange Rates in the United States: Empirical and Estimated Parametric Distributions by Season and Climatic Region. *Risk Analysis*, Vol. 15, No. 4, pp. 459-465.

Persily, A., J. Gorfain, G. Brunner.2005. Ventilation design and performance in U.S. office buildings. *ASHRAE Journal.* pp. 30-35. April.

Persily, A. and J. Gorfain.2004. *Analysis of ventilation data from the U.S. Environmental Protection Agency Building Assessment Survey and Evaluation (BASE) Study.* National Institute of Standards and Technology, NISTIR 7145. December.

Turk, B. H., D. T. Grimsrud, J. T. Brown, K. L. Geisling-Sobotka, J. Harrison, R. J. Prill 1989. *Commercial building ventilation rates and particle concentrations*. ASHRAE, No. 3248.

Williams, R., J. Suggs, A. Rea, K. Leovic, A. Vette, C. Croghan, L. Sheldon, C. Rodes, J. Thornburg, A. Ejire, M. Herbst, W. Sanders Jr. 2003a. The Research Triangle Park particulate matter panel study: PM mass concentration relationships. *Atmos Env* 37, pp. 5349-5363.

Williams, R., J. Suggs, A. Rea, L. Sheldon, C. Rodes, J. Thornburg 2003b. The Research Triangle Park particulate patter panel study: modeling ambient source contribution to personal and residential PM mass concentrations. *Atmos Env* 37, pp. 5365-5378.

Wilson, A. L., S. D. Colome, Y. Tian, P. E. Baker, E. W. Becker, D. W. Behrens, I. H. Billick, C. A. Garrison 1996. California residential air exchange rates and residence volumes. *Journal of Exposure Analysis and Environ Epidemiology*. Vol. 6, No. 3.

Wilson, A. L., S. D. Colome, P. E. Baker, E. W. Becker 1986. *Residential Indoor Air Quality Characterization Study of Nitrogen Dioxide, Phase I, Final Report.* Prepared for Southern California Gas Company, Los Angeles.