**Online Resource**

**Article title**:A sensitivity analysis of a human exposure model using the Sobol method

**Journal**:Stochastic Environmental Research and Risk Assessment

**Authors**: John Langstaff a, Graham Glen b, Chris Holder b, Stephen Graham a, Kristin Isaacs c

**Author affiliations**:

a Office of Air Quality Planning and Standards (OAQPS), U.S. EPA, Research Triangle Park, NC 27711, USA

b ICF, Inc., Durham, NC 27713, USA

c Office of Research and Development (ORD), U.S. EPA, Research Triangle Park, NC 27711, USA

**Corresponding author**: John Langstaff, [langstaff.john@epa.gov](mailto:langstaff.john@epa.gov)

This Online Resource supplements the main article by providing additional helpful information on

* how the APEX model generally functions (Section A);
* the formulas calculating the Sobol sensitivity indices (Section B);
* an illustration of the Chicago study area used in the main article’s case study (Section C);
* the distributions assigned to the model microenvironmental parameters (Section D); and
* additional descriptions of the meaning of the stochastic model-input variables of the sensitivity analyses conducted in the main article (Section E).

# A. APEX Workflow and Exposure and Dose Equations

A high-level overview of the APEX modeling workflow is available in Section 2.2. of U.S. EPA (2017) and we reproduce it below. Further below, we provide a brief discussion on how APEX calculates exposures and doses.

## Workflow

APEX estimates human inhalation exposure to criteria pollutants and air toxics at the local, urban, or regional level. There are seven broad steps in the APEX model, as described briefly in Figure A-1 and illustrated and described with more detail in Figure A-3–Figure A-9. These figures are overviews; for a more comprehensive understanding of these steps, refer to U.S. EPA (2019a,b).

Characterize study area

APEX models sectors within a study area corresponding to Census data

Generate simulated profiles

APEX stochastically generates a set of profiles (simulated people) based on the demographics of the study area

Construct activity events

For each profile, APEX constructs an activity diary for each day in the simulation

Estimate energy expenditures

APEX constructs energy expenditures for each profile based on activity sequence

Calculate concentrations in ME

Based on air quality input data, APEX calculates the pollutant air concentration in each microenvironment (ME) in each air quality sector, for each timestep, for the period of the simulation

Calculate exposures for each pollutant

APEX assigns a concentration to each exposure event based on the location of the activity

Calculate doses

APEX optionally calculates dose values for each of the profiles, for each timestep

1

2

3

4

5

6

7

Figure A-1. Overview of Process Steps in APEX

A picture containing text

Description automatically generated

Figure A-2. Legend for  
APEX Overview Figures (Figures A-3–A-9)

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated with low confidence | In the first step, the user defines the study area, the time period of interest, and the number of profiles to be used in the simulation. The number of profiles should be large enough to provide a representative sample of people in the area. The user must also provide inputs on air concentrations and meteorology data for the study area. |

Figure A-3. APEX General Step 1

|  |  |
| --- | --- |
| A picture containing map  Description automatically generated | The model then generates a representative, random sample of profiles with sex, age, and racial characteristics in proportion to those within the study area, usually based on U.S. Census Data. APEX randomly selects each profile’s physiological attributes (height, weight, etc.) based on age- and sex-specific probability distributions. Employment status of each profile is assigned based on employment probabilities for the study area. APEX also assigns each profile housing parameters such as whether they have air conditioning and gas stoves based on probability distributions. Finally, APEX assigns each profile a home location (and a work location if employed.) |

Figure A-4. APEX General Step 2

|  |  |
| --- | --- |
|  | Next, the model generates an activity diary for each profile for every day of the simulation. The activity patterns of the simulated individuals are assumed by the model to be comparable to those of individuals with similar demographic characteristics, according to activity data such as diaries compiled in EPA’s Consolidated Human Activities Database (CHAD) (McCurdy 2000; U.S. EPA 2019c). Simulated demographic information along with the time period of the simulation and input meteorology data are used by the model to sample activity diaries for each profile for each study day. |

Figure A-5. APEX General Step 3

|  |  |
| --- | --- |
| Diagram  Description automatically generated | The model then estimates the energy expenditure and resulting ventilation rate of each activity for each profile based on his or her demographic and physiological variables (age, height, weight, resting metabolic rate, etc.). |

Figure A-6. APEX General Step 4

|  |  |
| --- | --- |
| Graphical user interface, application  Description automatically generated | For each simulation, the user defines microenvironments, which are different spaces with relatively homogeneous air pollution concentrations in which activities takes place. Typical microenvironments may include indoor spaces such as residences or offices and outdoor areas like parks. APEX uses the input ambient concentrations of pollutants to calculate the concentration for each pollutant in every microenvironment for every timestep. Every activity included in the activity diaries is mapped to a microenvironment, however the calculations of pollutant concentration in microenvironments occur independently of profiles and their activities (that is, APEX will calculate timestep concentrations in microenvironments regardless of if a profile occupies that microenvironment in that timestep). |

Figure A-7. APEX General Step 5

|  |  |
| --- | --- |
| Diagram  Description automatically generated | Next, APEX calculates the inhalation exposure for profiles for each timestep based on the activity location, activity duration within the microenvironment, and microenvironment pollutant concentration. |

Figure A-8. APEX General Step 6

|  |  |
| --- | --- |
| Text  Description automatically generated with medium confidence | Finally, the model uses the ventilation rate associated with a profile and the exposure metrics to calculate the inhaled dose for each profile as a function of timestep. |

Figure A-9. APEX General Step 7

## Exposure and Dose Equations

Here is the ozone concentration (µg/m3) in the microenvironment indicated by the assigned activity diary at time *t*, and is the ventilation rate (m3/hr) for that profile performing the corresponding activity as indicted by the diary. The concentration is related to the simultaneous ambient air quality in one of two ways, depending on the microenvironment. In the “factors” method, it is just a linear regression (using microenvironment-specific parameters) on the concurrent ambient air of the nearest ambient site. In the “mass-balance” method, depends on the ambient air, microenvironmental properties such as volume, air-exchange rate, and other loss rates, as well as on the concentration at the previous hour.

The activity diaries assigned to each profile affect which microenvironments (and the duration in each) are visited. They also determine the activities on which the ventilation rate is based. Personal variables such as age and sex affect the diary selection, as well as affecting directly.

Each activity diary is 24 hours long, and both hourly and daily totals of exposure and dose are produced. When there are multiple activities in the same hour, the exposure and dose are time-weighted over them to obtain hourly averages:

Here the summation is over the *N* activity diary events within hour *h*. There is not a fixed number in each hour, *N* =1 is typical while sleeping while *N* >1 is typical while awake. Duration is the time (in hours) that that diary event lasts, expressed as (duration in minutes/60). Within each hour the durations always sum to one hour. Thus, is the average exposure for that hour, with units of µg/m3, and is the average pollutant inhalation rate in µg/hr for that hour. Since this applies for one hour, it is also the mass inhaled in µg during that hour.

Each day, the 24 one-hour averages may be averaged to obtain the daily average, or else the largest of them may be used as a measure of maximum hourly exposure (or dose) on that day.

Each profile is followed for the duration of the simulation, which was 7 months in the scenario examined in this paper. Over this time, the daily averages may either be averaged to obtain the overall average exposure or dose, or else the largest single daily value may be retained as a measure of extreme exposure (or dose) for that profile. A total of 25,000 profiles were simulated in this way in each of these runs in this paper.

Interactions of Variables, Particularly in Relation to Sobol Analysis

The underlying equations in APEX are simple. The complexity in APEX comes from the interaction of many variables. Each profile is independent of all others. Profiles are assigned a sequence of activity diaries (one per simulation day) that match age, sex, and propensity to spend time outdoors, all of which are assigned to each profile using population weighed and study-area-specific data. Nevertheless, there are enough activity diaries in the input database so that every profile is assigned a substantially different set, even if the demographic variables match those of another profile. In addition, the properties of each microenvironment, while following the same rules for each profile, are profile-specific because they are resampled for each profile. For example, each house samples its air-exchange rate independently. With different activity dairies, each profile visits a different set of microenvironments for differing durations, and the air quality in each differs from that experienced by other profiles. The results for any one profile are the result of thousands of random samples, but as usual in Monte Carlo modeling, the overall statistical properties of the set of profiles should provide meaningful results, as stochastic variation converges to small values when averaged over enough cases. This is true for exposure and dose directly, and in Sobol analysis it is also true for the estimates of the indices. The complicated interplay of the large number of variables in APEX means that the Sobol indices cannot be evaluated analytically, as is possible in models with (say) fewer than 10 randomly-sampled inputs which are combined in one (or a few) equations. Therefore, in APEX the Sobol indices must be evaluated empirically by running enough profiles.

## References

McCurdy T, Glen G, Smith L, Lakkadi Y (2000) The National Exposure Research Laboratory’s Consolidated Human Activity Database. J Expo Sci Environ Epidemiol 10:566–578. <https://doi.org/10.1038/sj.jea.7500114>

U.S. EPA (2017) An Introduction to APEX. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711. Available at: <https://www.epa.gov/fera/apex-user-guides>.

U.S. EPA (2019a) Air pollutants exposure model documentation (APEX, Version 5.2) Volume I: User’s Guide. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711. EPA-452/R-19-005a. Available at: <https://www.epa.gov/fera/apex-user-guides>.

U.S. EPA (2019b) Air pollutants exposure model documentation (APEX, Version 5.2) Volume II: Technical Support Document. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711. EPA-452/R-17-001b. Available at: <https://www.epa.gov/fera/apex-user-guides>.

U.S. EPA (2019c) The consolidated human activity database (CHAD) Documentation and Users’ Guide. U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711. EPA-452/B-19-001. Available at: <https://www.epa.gov/fera/consolidated-human-activity-database-chad-documentation-chad-master>.

# B. Sobol Equations as Implemented in this Paper

From Theorems 1 and 2 in Glen and Isaacs (2012), the Sobol sensitivity indices for group *j* are:

Here is the expected value of the correlation in model output variable between model runs that share the stochastic input values for the variables in group and no others. On the other hand, refers to the expected value of the correlation in model output variable between model runs that share the stochastic input values for all variables except those in group

Simplified Explanation:

If the variables in have no effect on model output, then because all the stochastic variation comes from the variable set , which is sampled independently in those runs. Similarly, because the set is common to both runs and is responsible for all variation in model output. In this case the two runs have identical output, and any function has a correlation of one with itself, so 1.

At the other extreme, suppose alone fully determines the model output. Then because the outputs are identical. Similarly, because the variables with common values in both runs (the set ) contribute nothing to model output, so the expected correlation in the model outputs is zero; that is, .

The second case is not interesting since the output depends on just one input, making sensitivity analysis unnecessary. The usual case is somewhere in between, with and . The first case is relevant because model inputs not affecting a given output variable have indices of zero. For example, physiological variables have indices of zero for exposure (but not for dose) because their effect in the model occurs only after exposure has been determined.

## Reference

Glen G, Isaacs K (2012) Estimating Sobol indices using correlations. Environ Model Softw 37:157–166. <https://doi.org/10.1016/j.envsoft.2012.03.014>

# C. Chicago Study Area

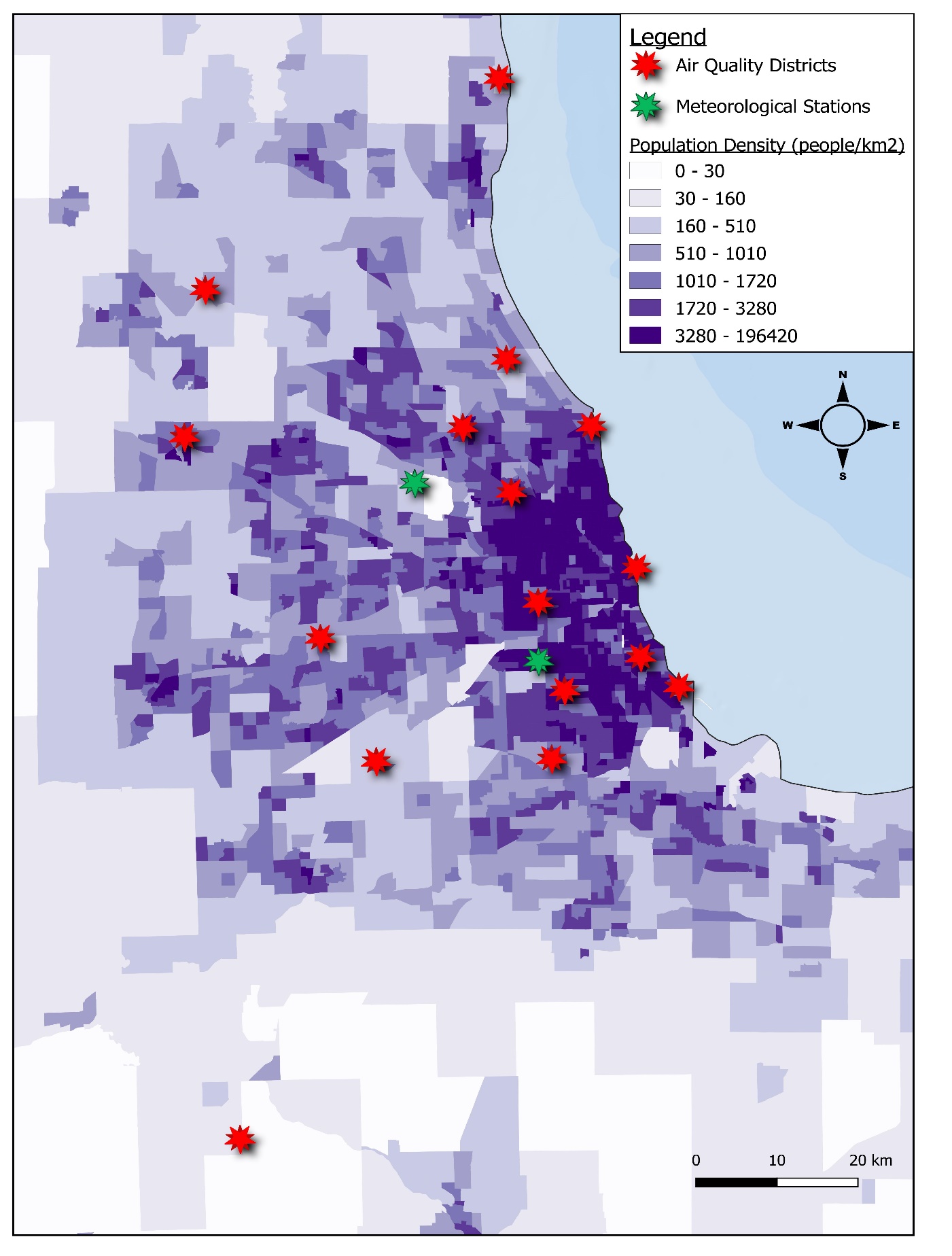


Figure C-1. Spatial Information on the Chicago Case Study

# D. APEX Microenvironmental Parameters

Table D-1. Distributions Used for Microenvironmental Parameters in the Chicago Case Study

|  |  |  |  |
| --- | --- | --- | --- |
| **Microenvironment** | **Parameter** | **Conditions** | **Distribution** |
| Indoors - residence | AER | Temp < 50; A/C | LogN(0.744, 1.982, 0.1, 10) |
| Temp 50–67; A/C | LogN(0.811, 2.653, 0.1, 10) |
| Temp 68–77; A/C | LogN(0.785, 2.817, 0.1, 10) |
| Temp > 77; A/C | LogN(0.916, 2.671, 0.1, 10) |
| Temp < 50; no A/C | LogN(0.760, 1.747, 0.1, 10) |
| Temp 50–67; no A/C | LogN(1.447, 2.950, 0.1, 10) |
| Temp 68–77; no A/C | LogN(1.531, 2.472, 0.1, 10) |
| Temp > 77; no A/C | LogN(1.901, 2.524, 0.1, 10) |
| Indoors - residence | A/C prevalence | year 2009 data | 95% |
| Indoors - school | AER | All | Weibull(0.5, 1.5, 0.07, 4) |
| Indoors - restaurant | AER | All | LogN(3.712, 1.855, 1.46, 9.07) |
| Indoors - office buildings, hospitals, stores, other | AER | All | LogN(1.109, 3.015, 0.07, 13.8) |
| Indoors - All | O3 decay rate | All | LogN(2.51, 1.53, 0.95, 8.05) |
| O3 proximity | All | LogN(0.988, 1.689, 0.5, 1.5) |
| Outdoors - near road | O3 proximity | All | Normal(0.755, 0.203, 0.422, 1.0) |
| Outdoors - other | O3 proximity | All | LogN(0.988, 1.689, 0.5, 1.5) |
| In-vehicle | O3 proximity | local roads (13%) | Normal(0.755, 0.203, 0.422, 1.0) |
| urban roads (56%) | Normal(0.754, 0.243, 0.355, 1.0) |
| interstates (31%) | Normal(0.364, 0.165, 0.093, 1.0) |
| O3 penetration | All | Normal(0.300, 0.232, 0.1, 1.0) |

Notes: AER = air-exchange rate; Temp = daily-average temperature in degrees Fahrenheit; A/C = usage of air conditioning; LogN(W,X,Y,Z) = lognormal(geometric mean, geometric standard deviation, minimum, maximum); Normal(W,X,Y,Z) = Gaussian(mean, standard deviation, minimum, maximum); Weibull(W,X,Y,Z) = Weibull(scale, shape, minimum, maximum); O3 = ozone.

Values below the minimum or above the maximum are discarded until an acceptable value occurs.

# E. The Stochastic Model-input Variables Used in this Analysis

Table E-1. Descriptions for Each Main Type of Stochastic Model-input Variable

| **Input Random Variable** | **Description** |
| --- | --- |
|
| Sex | Male or female |
| Age | Age in full years since birth (rounded down). Held constant over the simulation. |
| HomeSec | Home tract. The census tract where the profile is when at home, as indicated by their activity diary. |
| WorkSec | Work tract. The census tract where the profile is when at work, as indicated by their activity diary. |
| Race | Race. Part of the census data but not relevant in these runs. Therefore, placed in group zero in these runs. |
| Employ | Employment status. Binary switch indicating whether the profile has a job or not, based on employment data by age, sex, and home tract. If they work, a suitable “WorkSec” is randomly chosen using the commuting flow database. |
| ProfFactor | Location-dependent profile factor. Not used in these APEX runs. |
| GasStove | Has gas stove. Not relevant in these runs since they do not produce ozone. |
| GasPilot | Has gas pilot on stove. Not relevant in these runs. |
| AC\_Home | Type of air conditioning system at home. Air-exchange rate (AER) distributions for the home depend on this variable. |
| AC\_Car | Has air conditioning in car. Air quality in cars may depend on this variable. |
| ProfCond1-5 | Profile conditional variables #1 to #5. Not used in these APEX runs. |
| RegCond1-5 | Regional conditional variables #1 to #5. Not used in these APEX runs. |
| OtherD | Air Quality (AQ) district other than home or work. Determines which AQ site drives the microenvironmental air quality when the profile is neither at home nor at work, as indicated by their activity diary. |
| NearHome | AQ district when in Near Home location. Similar to OtherD, is the location for diary events coded as “near home”, which includes activities such as walking, banking, grocery shopping, visiting the dry cleaners, etc. |
| NearWork | AQ district when in Near Work location. Similar to NearHome, applies to activities such as lunch on work days or other errands performed during work but not specifically at the work location. |
| WindowRes | Daily window status at residence (open, closed). Not used in these runs. Can be reset daily for each profile and would affect the AER, if used. AER distributions used in these runs depended on air conditioning type and outdoor temperature but did not explicitly indicate window status. |
| WindowCar | Daily window status in car. Similar to WindowRes and not used in these runs. |
| SpeedCat | Daily speed category for travel. Average driving speed can affect the air quality in cars, but this variable was not used in these runs. |
| DayCond1-3 | Daily conditional variables #1 to #3. DailyConditional#1 represents the predominant roadway type (changed daily). Affects the penetration factor for vehicles. Local roads allow more ozone to penetrate (on average) than do large interstate highways. |
| BodyMass | Body mass of the profile in kilograms. Assumed constant over the simulation. |
| Height | Height of the profile in centimeters. Assumed constant over the simulation. |
| BSA | Body surface area. Could affect dose for some chemicals, but has no effect in these runs. |
| RMR | Resting metabolic rate. Determines the breathing ventilation rate while resting. Held constant over the simulation. |
| VEAge | Regression terms for breathing versus age. Affects the ventilation rate while resting, and together with VO2max affects the ventilation rate for other activities. |
| Disease | Presence of disease. Not used in these runs. |
| VO2max | Maximum oxygen (O2) consumption. Maximum rate of O2 uptake which corresponds to METmax (maximum metabolic equivalent of task). Specific to each profile. |
| ECF | Energy-conversion factor. Volume of O2 needed to support one kilocalorie of metabolic expenditure. |
| RecoveryT | Recovery time for O2 debt. Time needed to be able to exercise at full rate after becoming exhausted. |
| Hemog | Hemoglobin density in blood. Needed when the pollutant is carbon monoxide (CO), but has no effect in these runs. |
| MaxOxD | Maximum possible O2 debt. Heavy exertion can create an oxygen debt. Then the limit on the metabolic rate for subsequent activities becomes more severe, until that debt is cleared. |
| EndogCO | Endogenous CO production. Affects CO but not ozone, so not used in these runs. |
| BloodVol | Blood volume regression terms. The terms in the ratio of blood volume to body weight. |
| SFast | Slope of fast O2 debt recovery. O2 debt recovery has two phases: fast and slow. This term describes the fast rate, while the recovery time determine the slow rate. |
| FEVB1-9 | dFEV terms #1 to #9. Coefficients for a regression of lung function decrement versus intake. Response variable that does not affect either exposure or dose. |
| FEVreg | FEV regression parameters. See above, has no effect on exposure or dose. |
| FEVU | dFEV parameter U. See above, has no effect on exposure or dose. |
| FEVE1 | dFEV parameter E1. See above, has no effect on exposure or dose. |
| FEVE2 | dFEV variation parameter E2. See above, has no effect on exposure or dose. |
| VEBM | Regression terms for breathing ventilation. Affect breathing ventilation rate needed to support a given activity level. |
| MET | Activity-specific MET values. Represents the ratio of activity-specific metabolic energy expenditure to the baseline level (while resting). |
| AQData | AQ data drawn from distributions. Not used in these APEX runs. |
| DiarySel | Daily activity-diary selection. Vector of random samples (one per simulation day) which determines which activity diary is selected from the appropriate diary pool. The pool not random but is determined by the age, sex, day type (weekend or weekday) and temperature, evaluated on each simulation day. Within each pool the diaries are assigned selection weights (also not random) based on their similarity to the profile’s characteristics. “DiarySel” randomly chooses one 24-hour long activity diary on each simulation day, using these selection weights. |
| DAutoCor | Autocorrelation of diaries. Quantifies the autocorrelation in outdoor time from one diary day to the next, which affects the diary selection weights. |
| Clus1 | First diary-clustering parameter. Not used in these runs. |
| Clus2 | Second diary-clustering parameter. Not used in these runs. |
| MP1 | AER inside residence. AER between inside the home and the outdoor air. |
| MP2 | AER inside office buildings or hospitals. AER between inside large office-like work buildings and the outdoor air. |
| MP3 | AER inside schools. AER between inside school buildings and the outdoor air. |
| MP4 | AER inside stores. AER between inside stores and the outdoor air. |
| MP5 | AER inside restaurants. AER between inside restaurants and the outdoor air. |
| MP6 | Decay rate (DE) inside residence. Ozone loss rate to chemical reactions inside the home. |
| MP7 | DE inside office buildings or hospitals. Ozone loss rate inside offices. |
| MP8 | DE inside schools. Ozone loss rate inside schools. |
| MP9 | DE inside stores. Ozone loss rate inside stores. |
| MP10 | DE inside restaurants. Ozone loss rate inside restaurants. |
| MP11 | Proximity factor (PR) factor for residence. PR accounts for differences in ambient concentrations between the AQ district location (e.g., the monitor) and the residence. |
| MP12 | PR for office buildings or hospitals. |
| MP13 | PR for schools. |
| MP14 | PR for stores. |
| MP15 | PR for restaurants. |
| MP16 | PR for outdoors - general. |
| MP17 | PR for outdoors - near major road. |
| MP18 | PR for vehicles. |
| MP19 | Penetration (PE) for vehicles. |