MEMORANDUM

SUBJECT: **Analysis Of Uncertainty In Ozone Population Exposure Modeling**

FROM: John Langstaff

TO: Ozone NAAQS Review Docket (OAR-2005-0172)

DATE: January 31, 2007

The attached technical memorandum describes a methodology for an analysis of the uncertainties associated with the population exposure modeling conducted in support of the review of the ozone NAAQS, and the results of this analysis.

Analysis Of Uncertainty In Ozone Population Exposure Modeling

Technical Memorandum

January 2007

John E. Langstaff

Ambient Standards Group
Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

DISCLAIMER

This draft document has been prepared by staff from the Ambient Standards Group, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency. Any opinions, findings, conclusions, or recommendations are those of the author and do not necessarily reflect the views of the EPA. A previous draft of this document was circulated to obtain review and comment from the Clean Air Scientific Advisory Committee (CASAC) and the general public. Comments on this document should be addressed to John E. Langstaff, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, C504-06, Research Triangle Park, North Carolina 27711 (email: langstaff.john@epa.gov).

ACKNOWLEDGEMENT

The execution of the APEX uncertainty analysis was greatly aided by the computational support received from the Center for Environmental Computing (CEC), a branch of EPA's Office of Environmental Information (OEI)/Office of Technology Operations and Planning (OTOP)/National Computer Center (NCC). CEC's innovative use of compute grid technology enabled the efficient harvesting of unused cycles on existing low-end systems to run thousands of simulations. This work was completed with support from Rob McCauley of Lockheed-Martin under EPA's Information Technology Solutions-Enterprise Systems Engineering (ITS-ESE) Contract (contract number 68-W-04-005).

Contents

Tables

Figures

Acronyms and Abbreviations

Acronym	Meaning of Acronym/Abbreviation	page first used
AER	air exchange rate	23
APEX	Air Pollution Exposure Model (US EPA, 2006b,c)	1
AQI	air quality index	48
BM	body mass	32
BMI	body mass index	34
BSA	body surface area	51
CD	Ozone Air Quality Criteria Document (US EPA, 2006a)	1
CDC	The Centers for Disease Control and Prevention	33
CDS	Child Development Supplement (CDS, 2005)	58
CFR	Code of Federal Regulations	11
CHAD	Consolidated Human Activity Database (US EPA,2002)	29
CMAQ	Community Multiscale Air Quality Model (www.epa.gov/asmdnerl/cmaq/)	19
CSA	combined statistical area (OMB, 2005)	9
CV	coefficient of variation	11
ECF	energy conversion factor	51
EPOC	excess post-exercise oxygen consumption	50
EVR	effective ventilation rate	32
FEM	Federal equivalent method	11
FRM	Federal reference method	11
GM	geometric mean	5
GSD	geometric standard deviation	5
HVAC	heating, ventilation, and air conditioning	26
MET	metabolic equivalent	32
MSA	metropolitan statistical area	49
NAAQS	national ambient air quality standard	1
NHANES	National Health and Nutrition Examination Survey	33
NHAPS	National Human Activity Pattern Study	30
NO	nitric oxide	18
NO_x	nitrogen oxides	9
OES	Occupational Employment Statistics Survey	49
PAI	physical activity index	32
PEM	personal exposure monitor	8
ppm	parts per million	11
RMR	resting metabolic rate	32
RMSE	root mean square error	12
TSD	technical support document	1
UI	uncertainty interval	54
UV	ultraviolet	11
V_E	expiration ventilation rate	51
VO_2	oxygen consumption rate	50
VOC	volatile organic compounds	9

INTRODUCTION

When models are applied in decision-making processes an evaluation of the uncertainty of the model predictions is of importance. The decision maker needs to know whether or not the level of uncertainty in modeled results are acceptable in the context of the decisions to be made. Without some knowledge of the uncertainty, the model essentially lacks useful predictive power. In practice it is difficult, if not impossible, to gain a complete understanding of all of the sources of model uncertainty and take them into account. However, it is incumbent on modelers to assess and report model uncertainty to the extent that it is feasible.

At the time the 1996 Ozone Criteria Document was published, available information indicated that only 40 percent of the variability in personal exposures was explained by exposure models (US EPA, 2006a, Appendix AX3, page 162). Since that time there have been considerable improvements in population exposure models and data for these models. However, a comprehensive evaluation of population exposure models for ambient air pollutants has never been performed, and significant uncertainties in the predictions of these models remain.

The importance of specific limitations of exposure models is application- and pollutant-specific. For example, the distribution of air exchange rates is one of the more important model input data for exposure modeling of particulate matter. For some air toxics, uncertainties in the emissions and air concentrations of the pollutant will be the overriding limitation. For pollutants where time spent outdoors is an important parameter (for example, ozone), activity diary construction may be a significant source of uncertainty.

The analysis of model uncertainty described in this report has been performed as part of the exposure modeling conducted in support of EPA's ozone NAAQS review, described in Chapter 4 of the Ozone Staff Paper (US EPA, 2007a) and the Exposure Analysis Technical Support Document (US EPA, 2007b). The exposure model, APEX, is documented in a user's guide and technical document (US EPA, 2006b,c). We will refer to these four documents in the remainder of this report as the Staff Paper, the Exposure Analysis TSD, the APEX User's Guide, and the APEX TSD. We refer to the Ozone Criteria Document (US EPA, 2006a) as the CD.

In the remainder of this section, we cover some of the basic concepts of model variability and uncertainty. The next section gives an overview of our approach for quantitatively characterizing the uncertainty of the exposure modeling performed as part of the ozone NAAQS review, followed by sections on estimation of the uncertainty of the inputs to the exposure model APEX and treatment of the uncertainty of the formulation of the APEX model. Results of the uncertainty analysis for the 2002 Boston base case scenario are presented in the final section.

Concepts

Uncertainties arise from errors in the values of data and parameters input to the model and the necessarily simplified representation by the model of complex physical and human behavioral processes. The model inputs are intended to be representative of the area being modeled, and many of them are (e.g., population demographics, air quality and meteorological data). However, some of the inputs are derived from data collected at locations and/or time periods that differ from those being modeled, and these can contribute to the uncertainty of the model results. It is difficult to judge the significance of these different sources of uncertainty without conducting a thorough assessment of the uncertainties and also of the variability of the model inputs and results. The distinctions between uncertainty and variability and between sensitivity and uncertainty analyses are fundamental to this discussion. These are defined as follows.

Uncertainty refers to the lack of knowledge of the actual values of physical variables (parameter uncertainty) and of physical systems (model uncertainty). For example, parameter uncertainty can result when non-representative sampling (to measure the distribution of parameter values) gives sampling errors. Model uncertainty results from simplification of complex physical systems. Uncertainty can be reduced through improved measurements and improved model formulation.

Variability represents the diversity or heterogeneity in a population or property, and is an inherent property of a physical property or population characteristic. This is sometimes referred to as *natural variability*. Examples are the variation in the heights of people and the variation of temperature over time. Variability cannot be reduced by using more measurements or measurements with increased precision (taking more precise measurements of people's heights does not reduce the natural variation in heights). *Inter-individual (between-individual) variability* refers to the differences in a property between individuals in a population. The variation of a property for one individual over time is *intra-individual (within-individual) variability*.

Sensitivity Analysis assesses the effect of changes in individual model input parameters on model predictions. This is often done by systematically varying one or more parameters at a time and recording the associated changes in model response. One primary objective of a sensitivity analysis is to rank the input parameters on the basis of their influence on model output.

Uncertainty Analysis involves the propagation of uncertainties and natural variability in a model's inputs to calculate the uncertainty and variability in the model outputs. It can also involve an analysis of the uncertainties resulting from model formulation. The contributions of the uncertainty and variability of specific model inputs to the uncertainty and variability of the model predictions can in some cases be explicitly quantified.

Data Uncertainty and Model Uncertainty

In general, limitations and uncertainties result from erroneous or uncertain inputs, errors in coding, simplifications of physical, chemical, and biological processes to form the conceptual model, and flaws in the conceptual model. One important class of deficiencies in a conceptual model is due to variability not modeled or modeled incorrectly. Sources of uncertainty in exposure modeling can be classified into two primary areas: errors in the model input data and parameters, and errors in the formulation of the model itself (structural uncertainty).

Parameter or Input Data Uncertainty. When parameters or input data are estimated from measurements or samples from within a larger population, uncertainties can arise from:

- small sample sizes
- imprecise measurements (systematic and random errors)
- non-representative samples, extrapolation errors
- temporal period and/or spatial extent too limited to detect trends
- flawed study design (systematic errors in the data collection process)
- flawed statistical estimation method
- the use of surrogate measures

Model Formulation or Structural Uncertainty. Model uncertainty can result from:

- simplifying assumptions
- incorrect assumptions
- incomplete knowledge of the physicochemical processes
- not accounting for important variables
- variability not modeled
- temporal and spatial aggregation errors
- mis-specification of the problem
- applying a model in a situation for which it was not designed

A simple example which illustrates the difference between model input uncertainty and structural uncertainty is modeling the distribution of heights in a population by a normal distribution, parameterized by the mean and variance. Estimates of the mean and variance are the "model input data." The uncertainty which results from the difference between the shape of the true distribution and the normal distribution leads to structural uncertainty. The parameters of the distribution are estimated by measuring a sample of the population, and thus are subject to sampling errors, which result in the model inputs uncertainty. Increasing the sample size will reduce these errors and the associated uncertainty of the modeled distribution. However, if the form of the distribution is incorrect, increasing the sample size will help only up to a point, and then model uncertainty will dominate. The only way to reduce the uncertainty further would be to improve the model by finding a distribution whose shape more accurately reflects the true distribution of heights.

Note that an input value can be very uncertain and yet have little contribution to the uncertainty of the model results. This depends on the degree of leverage or influence the particular model input has. Thus the most uncertain inputs do not necessarily contribute the most to the uncertainty of the model results. To further complicate matters, in some cases the amount of influence that a parameter has can depend on values of other parameters. For example, the number of houses with air conditioning may be an influential model input when temperatures are high, but not when temperatures are low.

The primary difficulty in performing an uncertainty analysis is the quantitative characterization of the uncertainties of the model inputs and model formulation. We often have information about the variability of model inputs, and sometimes the variability and uncertainty combined, but it is usually difficult to estimate the uncertainty separately from the variability. We seldom know the quantitative uncertainty resulting from model formulation, except in cases where a model evaluation has been performed.

The Uncertainty of Uncertainty Analysis

If all of the important sources of uncertainty are not taken into account, an analysis of uncertainty will give a misleading picture. Unfortunately, the major sources of uncertainty tend to be the most difficult to characterize, since if we have data for good quantitative characterization of uncertainty, these data can often then be used to reduce the uncertainty. Thus, estimates of uncertainty are themselves uncertain. Model evaluation, where model predictions are compared with measured values for a specific model application, can give useful insights in this context.

APPROACH FOR ASSESSMENT OF EXPOSURE MODELING UNCERTAINTY

The goal of this uncertainty analysis is to quantify the overall uncertainty in the APEX model output, resulting from uncertainty in the model inputs and uncertainty due to the model itself.

There are two general methods used here to assess the uncertainty due to uncertain model inputs. The primary method involves first quantifying the uncertainties of each of the model inputs, and then propagating those uncertainties through the model to estimate the resulting uncertainty of the model results. We do this using the Monte Carlo method, which has the advantage of being very flexible and comprehensive (Morgan and Henrion, 1990; Vose, 1996).

The second method involves sensitivity analyses. Certain model inputs are very complex and difficult to treat with a Monte Carlo approach, and we conduct sensitivity analyses to quantify their effect on uncertainty.

APEX is a probabilistic model which uses Monte Carlo simulation to explicitly incorporate the inherent variability of the modeled population and physical processes leading to exposures. The majority of the inputs to APEX are distributions characterizing the natural variability of model inputs. For example, instead of using a single decay rate for the decay of ozone indoors, a distribution of hourly decay rates is input to APEX, specified by its form (lognormal) and parameters (a geometric mean (GM) of 2.5 and geometric standard deviation (GSD) of 1.5), as shown in Figure 3. The development of the distributions representing variability which are input to APEX is described in the Exposure Analysis TSD (US EPA, 2007b).

The Monte Carlo approach to quantification of uncertainty entails performing many model runs with model inputs randomly sampled from distributions reflecting the uncertainty of the model inputs. For example, for ozone decay rates, we are assuming that the form of the distribution is approximately correct, but realize that the GM and GSD are not known precisely. Suppose we estimate that the errors of the GM are normally distributed with mean 0 and standard deviation 0.18 (Figure 1), and that the errors of the GSD are normal with mean 0 and standard deviation 0.05 (Figure 2). Then we run APEX numerous times, and for each run we randomly select values from these error distributions, add them to the GM (2.5) and GSD (1.5) of the decay rates, and use these for model inputs. Figure 4 illustrates six decay rate distributions that result from adding these uncertainty terms (randomly selected from the distributions depicted in Figure 1 and Figure 2) to the GM and GSD of the base distribution. These six distributions would be used as input to six separate APEX runs. If APEX is run 500 times in this way, we then have 500 values of any measure calculated from the APEX model results. This collection of values would quantitatively indicate the extent of the uncertainty in the APEX results due to the uncertainty in the decay rates input to the model.

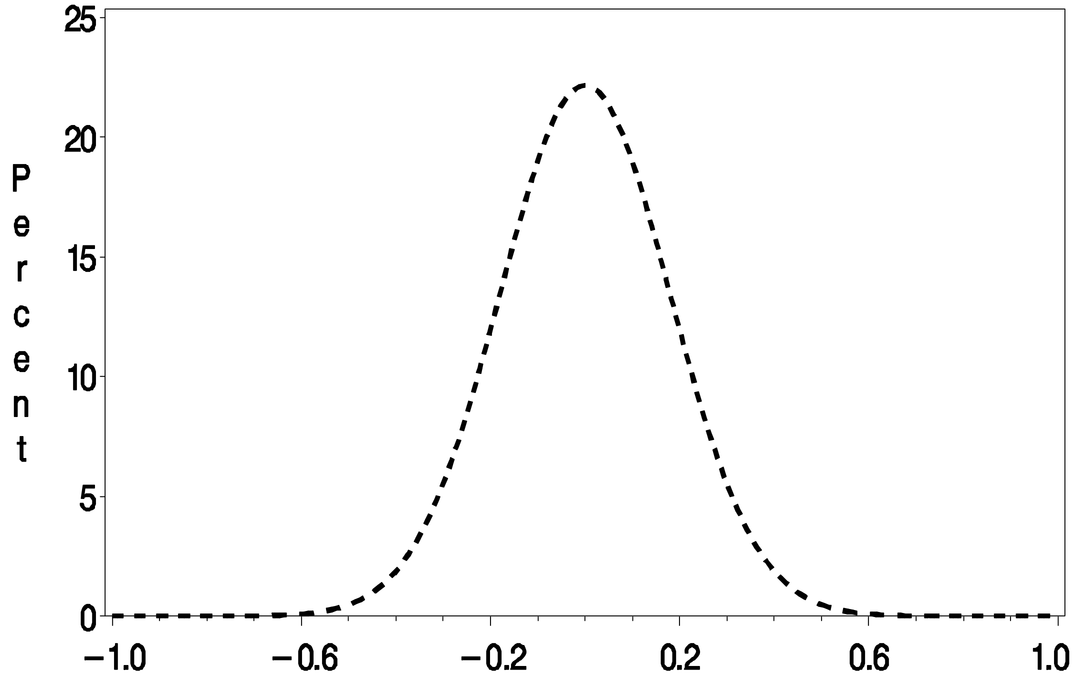

Figure 1. The distribution of the uncertainty of the GM (normal, mean=0, stdev=0.18)

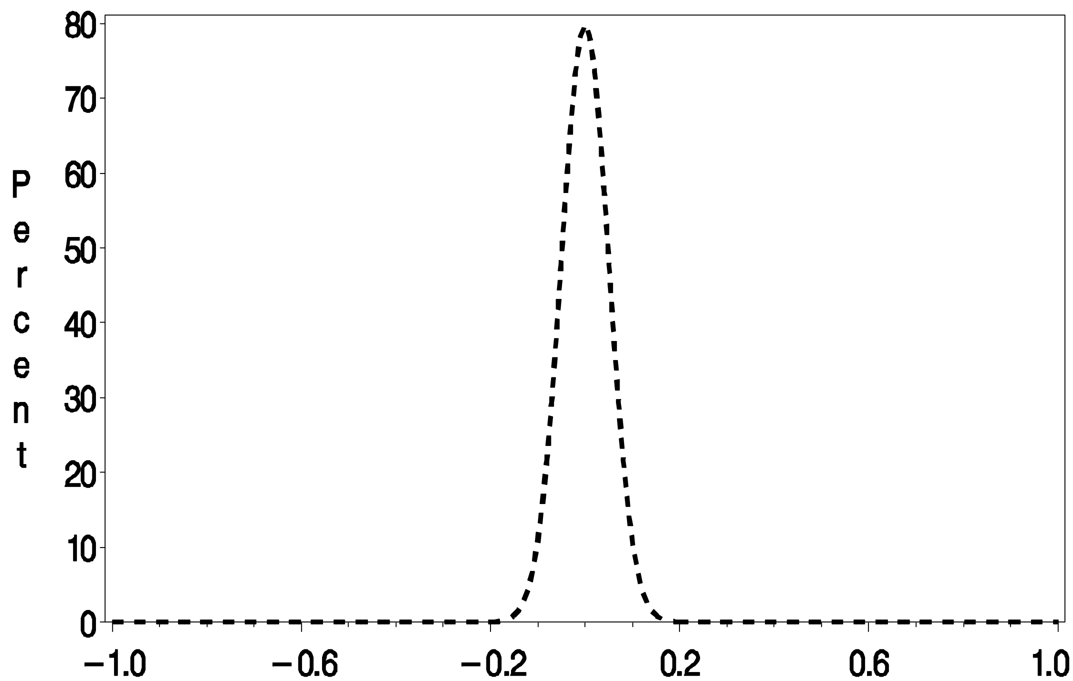

Figure 2. The distribution of the uncertainty of the GSD (normal, mean=0, stdev=0.05)

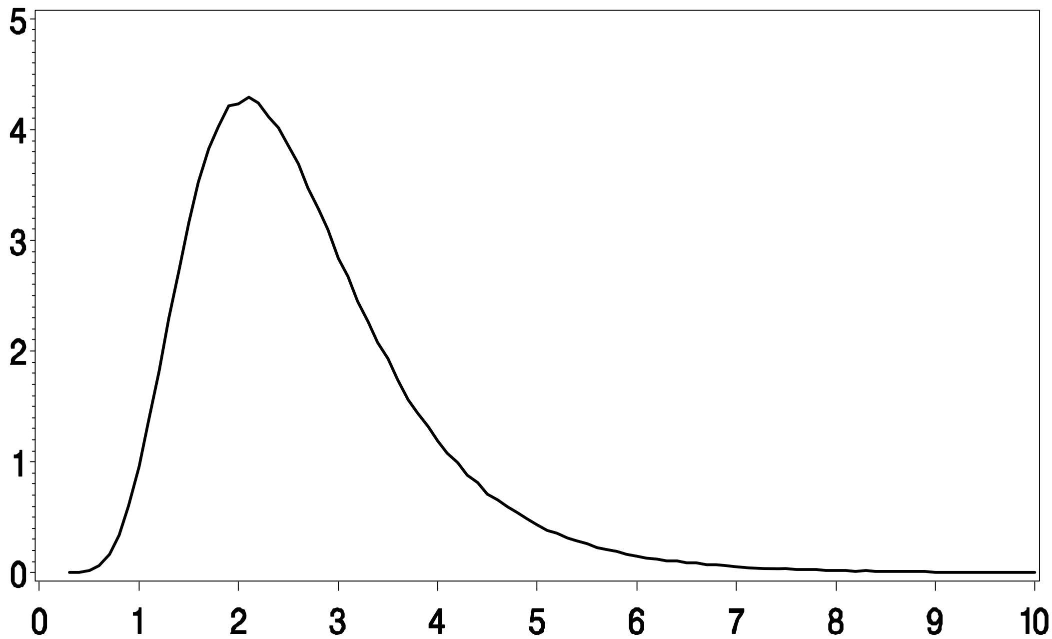

Figure 3. The base decay rates variability distribution (lognormal, GM=2.5, GSD=1.5)

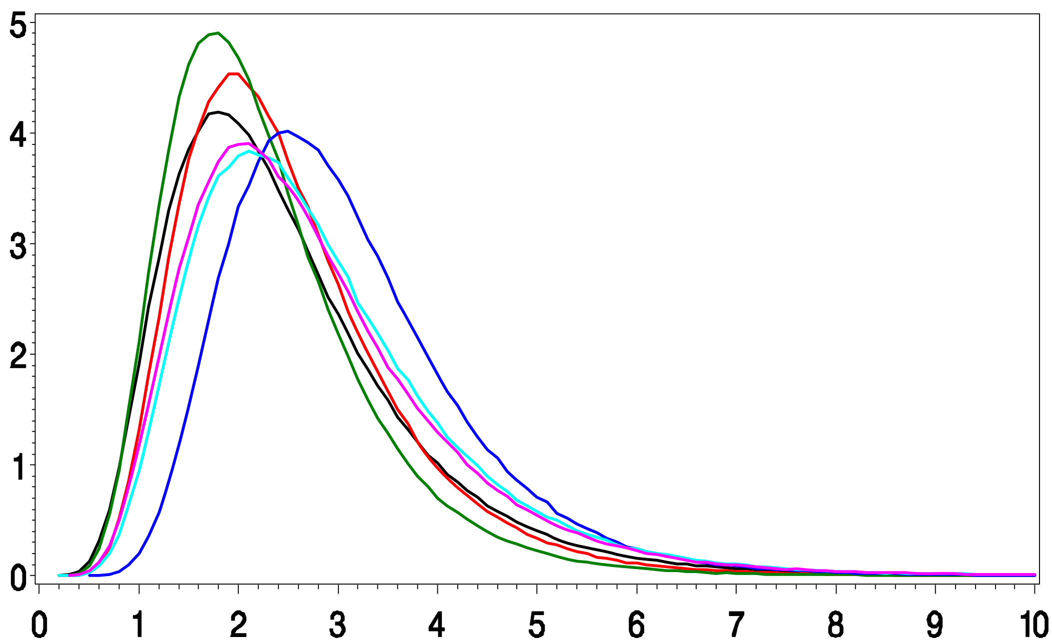

Figure 4. Six realizations of the combined variability and uncertainty distribution

Our approach to the assessment of the uncertainty resulting from model formulation and structure primarily involves a careful review of the scientific basis of the algorithms that make up APEX. We have also conducted a limited evaluation of APEX by comparing its predictions to 6-day average personal exposure measurements of ozone (see the Exposure Analysis TSD). A diagnostic evaluation with personal exposure monitors (PEMs), indoor, and outdoor measurements of ozone with shorter averaging times (1 hour or less) would be very informative, if the data were available.

The primary obstacle to performing an acceptable uncertainty analysis for this type of modeling is the quantitative characterization of the uncertainties of the model inputs. Developing appropriate distributions representing variability and uncertainty in various model inputs (e.g., air exchange rates, ozone decay rates, physiological parameters) is a key part of this modeling effort.

A Note About the Lognormal Distribution

Most of the inputs to APEX which have population variability are best fit with a lognormal distribution, and in some cases only the parameters of lognormal fits to data are reported in the literature. Typically there are not much data or information available for estimating the uncertainty of the distributions representing variability which are input to APEX, and a decision must be made about the distributional form of the uncertainty. Given an estimate of the uncertainty of an unbiased estimate of the GM, the question arises whether the uncertainty interval for the GM should be symmetric about the GM (e.g., [GM–Δ, GM+Δ]) or symmetric in the data space (symmetric about the GM multiplicatively, e.g., [GM/Δ, GM$\cdot\Delta$]), and whether or not the GSD should be varied concurrently with the GM. Changing the GM (or the GSD) changes both the mean and the standard deviation of a distribution, so care must be exercised when varying one or the other of these to ensure that the GM,GSD pair is valid. For example, if the estimate of the mean of a lognormal distribution is unbiased, then the GM and GSD must be varied concurrently (not independently, as in the decay rate example above) in the Monte Carlo simulations in such a way that the average of the means of the Monte Carlo distributions is approximately equal to the original estimate of the mean.

QUANTIFYING THE UNCERTAINTY OF APEX MODEL INPUTS

In this section we describe how the distributions of uncertainty were developed for this assessment of uncertainty of our application of APEX to model population exposures to ozone pollution for the 2002 Boston base case scenario. The APEX inputs for this base case are described in the Staff Paper and in the Exposure Analysis TSD.

Ambient Air Quality Concentrations

Hourly ambient concentrations are input to APEX, accounting for temporal variability. If concentrations from only one monitor are used, then spatial variability is not accounted for and cannot be properly modeled. If multiple monitors are used, then spatial variability is accounted for, but some uncertainty remains for concentrations at locations distant from monitors. The uncertainties associated with these concentrations in relation to spatial representativeness can be significant. For this modeling analysis, there is reasonable spatial coverage of the areas modeled. Table 1 lists the numbers of monitoring sites in the study areas for the years modeled. Using Boston as an example, the placement of monitors for the Boston greater metropolitan area is shown in Figure 5 (the monitoring sites are indicated by squares and the modeled region, the combined statistical area (CSA), by the heavy black lines). However, spatial variations in ozone concentrations can be considerable, resulting in uncertainty if these are not accounted for by the model (CD, Section 3.3).

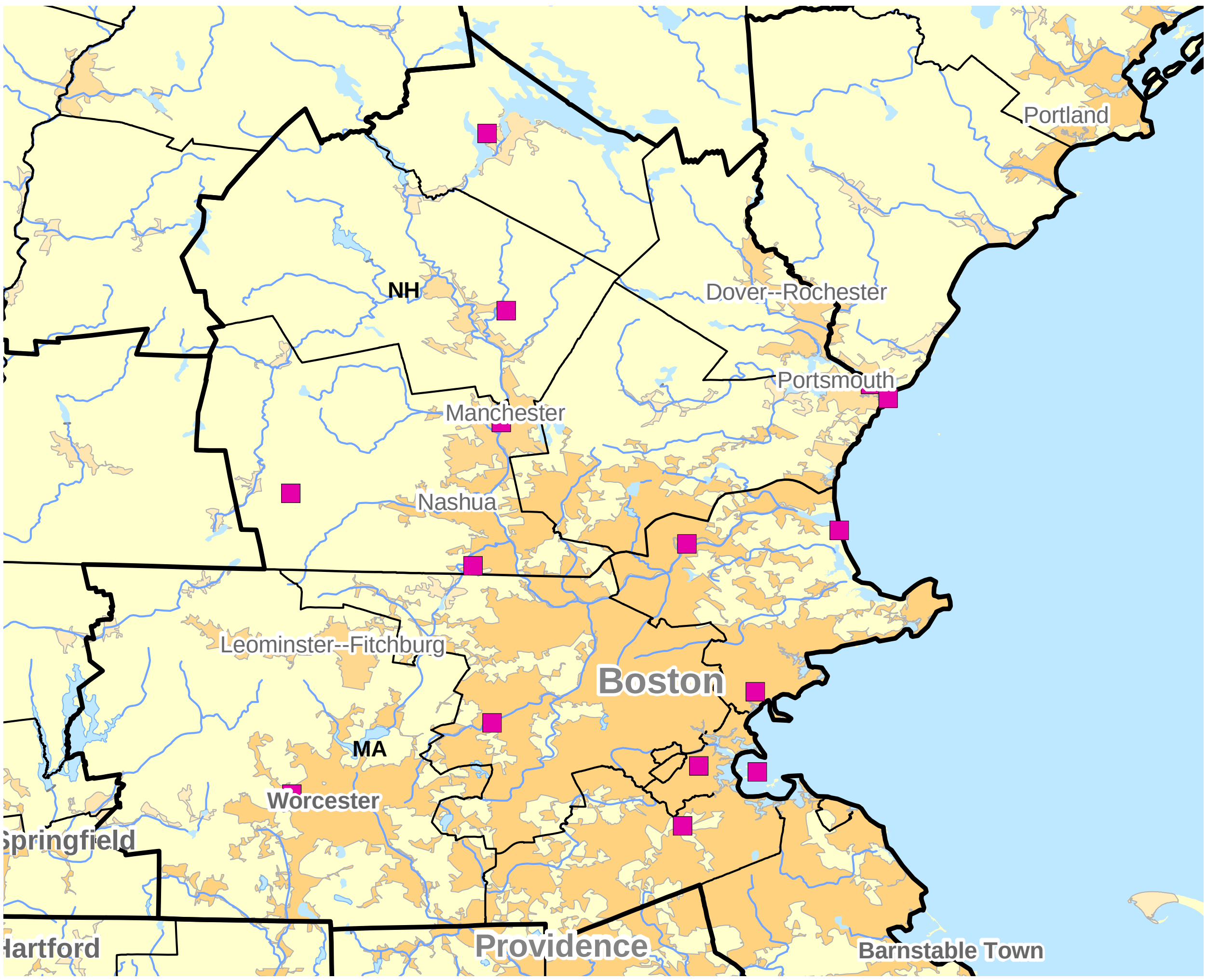

Figure 5. Boston CSA with ozone monitoring sites

If a single ozone season is modeled, another source of uncertainty results from the year-to-year variability of ozone concentrations, meteorology and NO_x and VOC emissions. We have modeled the 2002, 2003, and 2004 ozone seasons, which have different ozone concentrations due to a combination of different weather patterns and emissions of ozone precursors. In this way we account for the sensitivity of the exposure modeling results to year-to-year variability of air quality and meteorology. Modeling additional years might give a more complete picture of year-to-year variability, but we want the model results to represent recent air quality, and these three years provide a good range of weather patterns, which we feel is sufficient for this analysis.

Table 1. The number of ozone monitors in each of the study areas

Study Area (CSA)	**Number of monitors**		
	2002	**2003**	**2004**
Atlanta-Sandy Springs-Gainesville, GA-AL	13	12	12
Boston-Worcester-Manchester, MA-NH	17	19	15
Chicago-Naperville-Michigan City, IL-IN-WI	32	30	27
Cleveland-Akron-Elyria, OH	11	11	11
Detroit-Warren-Flint, MI	10	10	10
Houston-Baytown-Huntsville, TX	21	23	21
Los Angeles-Long Beach-Riverside, CA	45	43	44
New York-Newark-Bridgeport, NY-NJ-CT-PA	30	30	29
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	18	17	16
Sacramento--Arden-Arcade--Truckee, CA-NV	21	22	22
St. Louis-St. Charles-Farmington, MO-IL	18	18	17
Washington-Baltimore-N. Virginia, DC-MD-VA-WV	28	28	26

In addition to modeling exposures for three years, we are modeling exposures for scenarios of attainment of the current ozone standard and a number of potential alternative standards. For areas which do not meet these standards for these modeled years, attainment of these hypothetical scenarios would occur in the future. Modeling exposures for future years under different emission control strategies has, in addition to the uncertainties involved with modeling historical scenarios, the uncertainties of the complex process of projecting to future years air quality, population demographics, activity patterns, and other parameters which change over time. We employ a quadratic rollback technique to estimate ozone concentrations for these scenarios. This technique and the rationale for using it are described in the draft Staff Paper.

The primary uncertainties in the air quality data input to the model, discussed in the remainder of this section, result from:

- Instrument measurement error
- Estimation of missing data (temporal interpolation)
- Estimation of neighborhood-scale concentrations at locations which are not close to monitoring sites (spatial interpolation)
- Estimation of micro-scale outdoor concentrations (e.g., near-roadway)

- Adjustment of concentrations to reflect attainment of alternative standards (rollback)

Uncertainty Due To Measurement Error

The Federal reference method (FRM) for measuring concentrations of ambient ozone is based on chemiluminescence. However, chemiluminescence has not been widely used in instrumentation since the mid-1980s. Most instruments in use today employ ultra-violet (UV) absorption, a Federal equivalent method (FEM) (CD, Section 2.6). Federal reference and equivalent ozone monitoring methods are required to have a lower detectable limit of 0.01 ppm and precision of 0.01 ppm for 1-hour average concentrations (40 CFR Ch. 1, §53.21). Interference with other pollutants and humidity can lead to errors in measurements. This does not seem to be well quantified (CD, pages 2-25 to 2-26).

For this uncertainty analysis we estimate distributions of errors of hourly average measurements from the site-specific single point precision and bias estimates from the *2003 Criteria Pollutant Quality Indicator Summary Report* (Battelle, 2004) and the *2004 Single Point Precision and Bias Graphics for Criteria Pollutants* (US EPA, 2005a,b). Figure 6, taken from the Battelle report, illustrates these errors for some of the Massachusetts monitors. The “Bias” is the 95% confidence upper limit on the mean of the absolute values of relative percent differences for the monitoring season, and the “CV” is the 90% confidence upper limit of the coefficient of variation (CV) of relative percent difference values for the monitoring season. A positive bias means that the monitor readings are too high. Table 2 lists the available 2003 and 2004 values for monitors in the Boston CSA. Note that the 2003 and 2004 values do not correlate well, which indicates that the bias may be random for a given monitor. We estimate the measurement error uncertainty as normally distributed, with mean and CV taken to be the overall average of the bias and precision values. The values in Table 2 give an average bias of 1.2% and an average CV of 4.4%. Note that this will tend to overestimate the mean and CV of the measurement errors, since these are upper confidence limits.

Table 2. 2003 and 2004 Single Point Precision and Bias for Boston Monitors

	2003		2004	
AQS ID	**Bias (%)**	**CV (%)**	**Bias (%)**	**CV (%)**
250090005-1	-2.71	0.90	+2.66	2.55
250092006-1	+3.73	3.74	-5.70	5.71
250094004-1	+3.89	4.96	+2.49	2.28
250095005-1			+0.72	0.81
250171102-1	-5.86	0.91	-5.73	6.23
250213003-1	-2.41	3.21	-2.82	2.89
250250041-1	-7.35	7.87	-4.01	1.66
250250042-1	-1.07	0.94	3.54	4.51
250270015-1	-2.63	3.37	+4.09	2.40
330012004-1	+5.23	7.32		
330110020-1	+3.4	4.26		
330111010-1	3.05	4.47		
330115001-1	3.83	4.18		
330130007-1	-3.38	3.98		

AQS ID	2003 Bias (%)	2003 CV (%)	2004 Bias (%)	2004 CV (%)
330150012-1	-2.38	2.93		
330150013-1	+11.28	14.29		
330150015-1	9.28	11.90		
330173002-1	-5.36	6.79		
Average				

Uncertainty in Estimation of Missing Data

Missing air quality data were estimated by the following procedure. If there were consecutive strings of missing values (data gaps) of less than 6 hours, missing values were estimated by linear interpolation between the valid values at the ends of the gap. Remaining missing values at a monitor were estimated by fitting linear regression models for each hour of the day, with each of the other monitors, and choosing the model which maximizes R^2 for each hour of the day, subject to the constraints that R^2 be greater than 0.5 and the number of regression data values is at least 50. If there were any remaining missing values at this point, for gaps of less than 9 hours, missing values were estimated by linear interpolation between the valid values at the ends of the gap. Any remaining missing values were replaced with the regionwide mean for that hour.

The uncertainty of this method for filling in missing data was estimated by a jackknife-type approach where subsets of the data are randomly designated as "missing," then these missing values are filled in using the above procedure, and the filled in values are compared with the original values to see how well they are estimated. Since longer gap lengths generally engender more uncertainty, we calculate the frequencies of different gap lengths in the original data and set data to missing in such a way that these frequencies are reproduced. These errors turn out to be generally less than 0.004 ppm. Table 3 shows that replacement of missing data for the Boston CSA had little effect on the mean and standard deviation of the hourly ozone concentrations at each monitor. The root mean square error (RMSE) was generally less than 0.004 ppm, with insignificant bias, and the distribution of errors can be reasonably approximated by a normal distribution with mean zero and standard deviation 0.004 ppm.

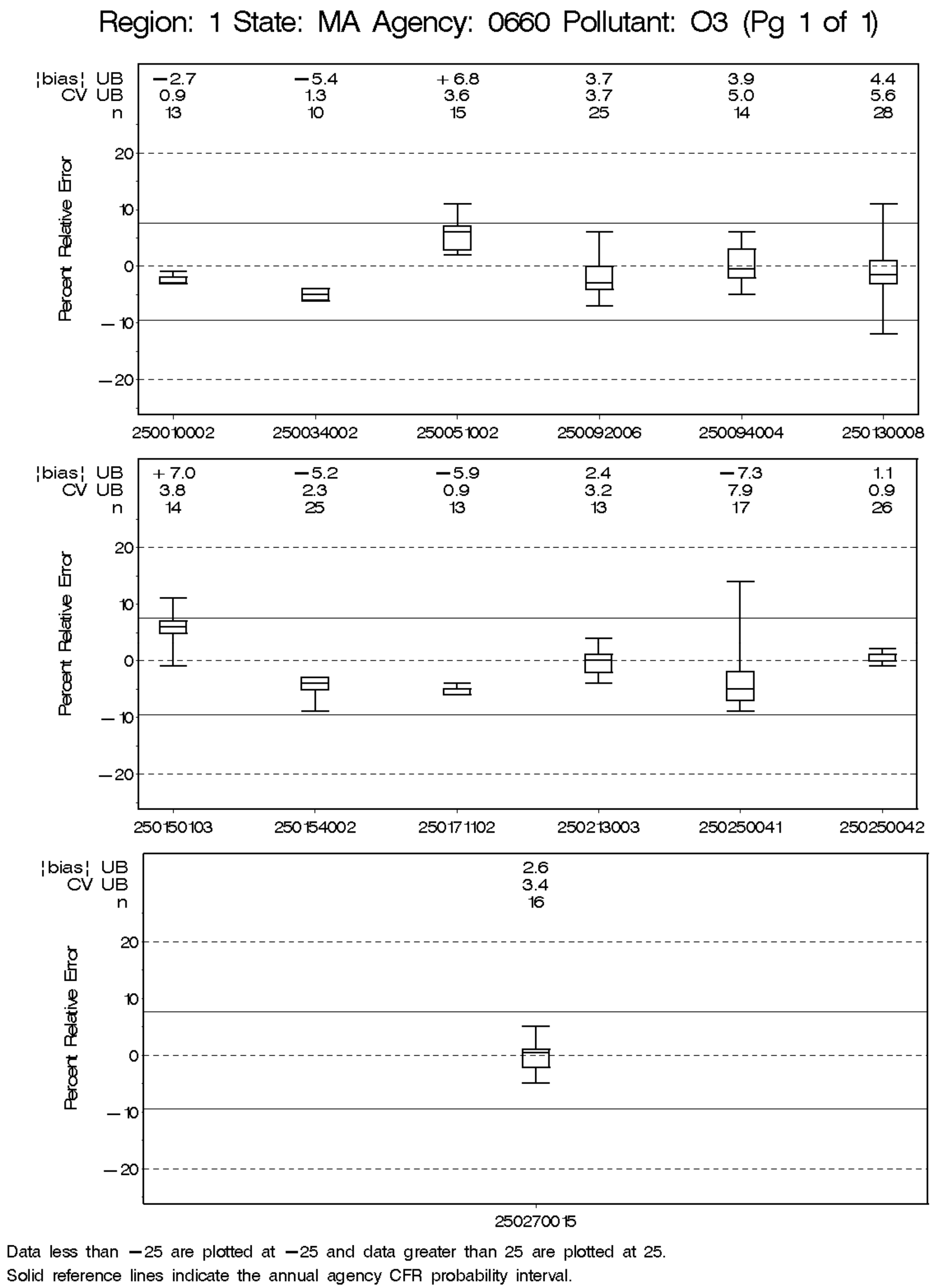

Figure 6. 2003 Precision and Accuracy for Massachusetts Monitors

Table 3. Effect of missing data replacement on the distribution of 2002 hourly ozone for monitors in the Boston CSA (ppb)

Monitor	# of hours missing	Mean of original data	Mean of filled data	Difference	St. dev. of original data	St. dev. of filled data	Difference
2500900051	290	0.0296	0.0292	0.00042	0.0188	0.0187	0.00009
2500920061	171	0.0369	0.0368	0.00013	0.0197	0.0195	0.00014
2500940041	141	0.0375	0.0374	0.00010	0.0178	0.0178	-0.00001
2501711021	483	0.0360	0.0362	-0.00016	0.0199	0.0199	-0.00004
2502130031	143	0.0427	0.0425	0.00015	0.0204	0.0203	0.00012
2502500411	157	0.0366	0.0365	0.00019	0.0187	0.0186	0.00009
2502500421	353	0.0258	0.0263	-0.00053	0.0165	0.0174	-0.00091
2502700151	1723	0.0431	0.0446	-0.00151	0.0179	0.0178	0.00014
3300120041	49	0.0373	0.0373	0.00001	0.0149	0.0149	0.00004
3301100201	47	0.0311	0.0311	-0.00001	0.0177	0.0176	0.00007
3301110101	47	0.0334	0.0334	0.00002	0.0195	0.0195	0.00007
3301150011	2634	0.0464	0.0479	-0.00145	0.0186	0.0158	0.00288
3301300071	50	0.0282	0.0283	-0.00008	0.0188	0.0188	0.00002
3301500121	39	0.0343	0.0344	-0.00003	0.0171	0.0170	0.00002
3301500131	141	0.0332	0.0333	-0.00011	0.0193	0.0192	0.00012
3301500151	162	0.0315	0.0316	-0.00010	0.0184	0.0181	0.00027
3301730021	40	0.0339	0.0339	-0.00005	0.0167	0.0167	0.00001

Uncertainty in Spatial Interpolation

The ambient ozone concentrations were interpolated from the monitoring sites to each Census tract by assigning the concentration at the monitor closest to each tract ("nearest neighbor" interpolation). We employed two approaches to estimate the impacts of the errors of spatial interpolation of ozone concentrations, jackknife estimation and a sensitivity analysis varying the radius of influence of the monitors.

Jackknife Estimates of Uncertainty

We estimate the errors of spatial interpolation using the jackknife method (Efron, 1980; Stone, 1974), in which we drop out one monitor, use the spatial interpolation method to estimate concentrations at the location of that monitor (not using nearby monitors), and compare the predicted to the observed values for each hour, giving a distribution of errors for that monitor. We do this for every monitor in the study area, thereby characterizing the errors of spatial interpolation in that area by a single distribution. This method tends to overestimate the size of the errors, because all monitors are used in the actual interpolation, reducing the interpolation errors to zero at the locations where the errors are estimated.

For each site, we calculate the observed/predicted ratio for each hour, and the 25th and 75th percentiles of these ratios. For the Boston 2002 measurements, the means (over all sites) of these site-specific quartiles are 0.94 and 1.2, with a central value of 1.06, and we approximate the uncertainty of the spatial interpolation by a normal distribution of observed/predicted ratios with quartiles 0.94 and 1.2, giving a standard deviation of 0.2. Although it may appear that the interpolation generally underpredicts in this case (ratio > 1), we cannot conclude this with confidence since all sites are used in the interpolation, which acts to correct this bias. If a bias remains, we do not know whether it is positive or negative; therefore, we assume that the interpolation is unbiased.

Decreasing Radius of Representativeness of Monitors

In general, the closer a location is to a given monitoring site, the better the measurements represent concentrations at that location. APEX allows the user to specify a radius of representativeness for the air quality monitors, and only models exposures to the population residing in Census tracts located within this radius of a monitoring site (the center of the tract is required to be within this distance of a monitor). Conversely, the further away that locations are from monitoring sites, the more uncertain the spatially interpolated concentrations tend to be at these locations. In choosing the radius of representativeness there is a tradeoff between more accurate concentrations (smaller radius) and better characterization of the population (larger radius).

We conducted a series of APEX simulations varying the radius of representativeness of air quality monitors from 10 km to unlimited (within the modeled area) for the Boston CSA using the nearest-neighbor spatial interpolation method. The APEX inputs for the Boston 2002 and 2004 base case and current standard scenarios were used, except for varying the concentration fields. Note that the "unlimited" radius simulations are the same as the simulations used in the exposure analysis described in the Staff Paper. The results of these simulations are depicted in Figure 7 (2004 base case), Figure 8 (2002 base case), and Figure 9 (2002 current standard). The vertical axes are the fractions of the population, and the values for the unlimited radius are plotted at 60 km in these figures. Table 4 shows the 2002 population coverage for the different radii analyzed. This analysis indicates that the nearest-neighbor method of spatial interpolation may be introducing a small positive bias into the exposure modeling results for these scenarios. However, using a different radius results in a different population being modeled, and this also contributes to the differences seen.

Table 4. Population coverage of 2002 ozone monitors in Boston CSA

Radius about monitors (km)	Population coverage within the radius
10	47%
15	65%
20	81%
25	89%
50	99%
unlimited	100%

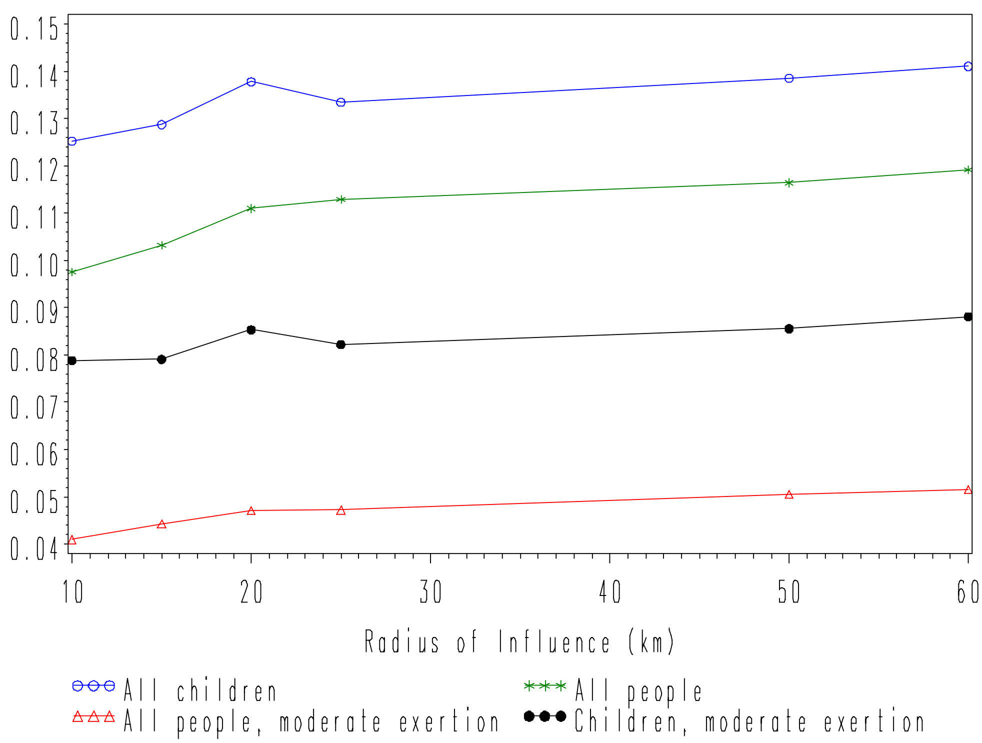

Figure 7. Sensitivity to monitor radius of influence of the fractions of four population groups with 8-hour exposures > 0.08 ppm-8hr, Boston, 2004 base case

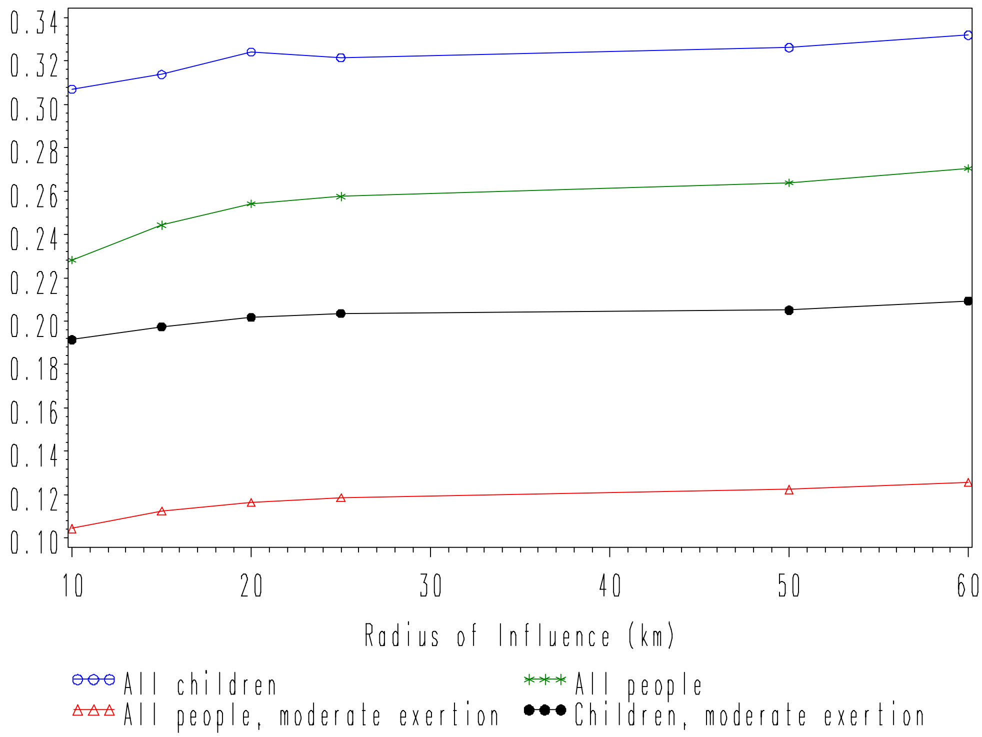

Figure 8. Sensitivity to monitor radius of influence of the fractions of four population groups with 8-hour exposures > 0.08 ppm-8hr, Boston, 2002 base case

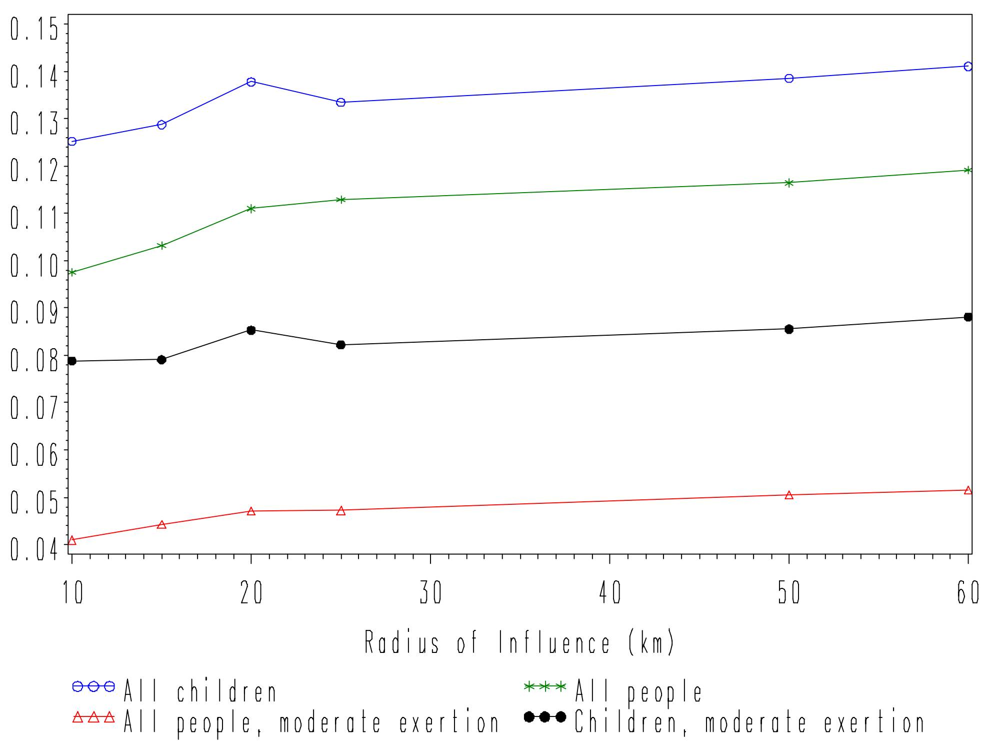

Figure 9. Sensitivity to monitor radius of influence of the fractions of four population groups with 8-hour exposures > 0.08 ppm-8hr, Boston, 2002 current standard

Summary of Uncertainty in Neighborhood-Scale Concentrations

Table 5 summarizes the uncertainties of neighborhood-scale hourly concentrations for the Boston CSA. As discussed above, we are assuming that these uncertainties can be adequately characterized by normal distributions. It seems reasonable to assume that these three components of uncertainty are independent. The uncertainties of measurement error and missing data replacement are additive, while the spatial interpolation uncertainties are multiplicative. The spatial interpolation uncertainties are at least an order of magnitude greater than the other uncertainties, and we approximate the combined uncertainties by the spatial uncertainties (Table 5).

Table 5. Uncertainty distribution parameters for neighborhood-scale concentrations

Component of uncertainty	Mean (bias)	Standard deviation
Measurement error (additive)	small	0.00135 (ppm)
Missing data replacement (additive)	insignificant	0.004 (ppm)
Spatial interpolation (ratios)	none	0.2 (dimensionless)
Combined uncertainties (ratios)	none	0.2 (dimensionless)

Uncertainty of Outdoor Near-Roadway Concentrations

Concentrations of ozone near roadways are particularly difficult to estimate due to the rapid reaction of ozone with nitric oxide (NO) emitted from motor vehicles (forming NO_2 and O_2), which reduces ozone concentrations in the vicinity of the roadway.

APEX adjusts ambient ozone concentrations for NO titration near roadways through the use of proximity factors. Proximity factors which adjust concentrations according to the locations of people's activities can be input as single values or distributions to APEX. They are intended to scale the concentrations measured at fixed-site monitors to better represent the concentrations at other locations. In APEX they can serve the dual purpose of incorporating random concentration variability into the model.

We developed distributions for near-roadway proximity factors based on data from the 1994 Cincinnati Ozone Study (American Petroleum Institute, 1997, Appendix B; Johnson et al. 1995). Table 6 lists these distributions. Vehicle miles traveled in 2003 by city and road type obtained from the Federal Highway Administration were used to estimate the distribution of road types (local, urban, interstates) for each modeled city. The development of these proximity factor distributions is described in Appendix A of the Exposure Analysis TSD.

Table 6. Near-roadway proximity factor distributions

Location	Mean	Standard Deviation	Lower Bound	Upper Bound
outdoors near road and parking lots	0.755	0.203	0.422	1.0
in-vehicle, local roads	0.755	0.203	0.422	1.0
in-vehicle, urban roads	0.754	0.243	0.355	1.0
in-vehicle, interstates	0.364	0.165	0.093	1.0

We conducted a review of literature on near-roadway titration of ozone by NO to obtain information which could be used to estimate the uncertainty of the near-roadway proximity factor distributions. Rodes and Holland (1981) found reductions in ozone downwind of a Los Angeles freeway ranging from more than 90% at 8 meters to small reductions at 500 meters from the roadway. Lin et al. (2001) report a 30-40% reduction in ozone in a high traffic-density neighborhood. Suppan and Schadler (2004) in a modeling study using CMAQ predict ozone reductions from 3 to 20 ppb downwind of a major highway, with small changes in ozone concentrations as far as 40 km from the highway. Beckerman et al. (2006) measured ozone and other pollutants at various distances from a heavily traveled highway in Toronto and find significant variation even in 7-day average ozone concentrations, as shown in Figure 10.

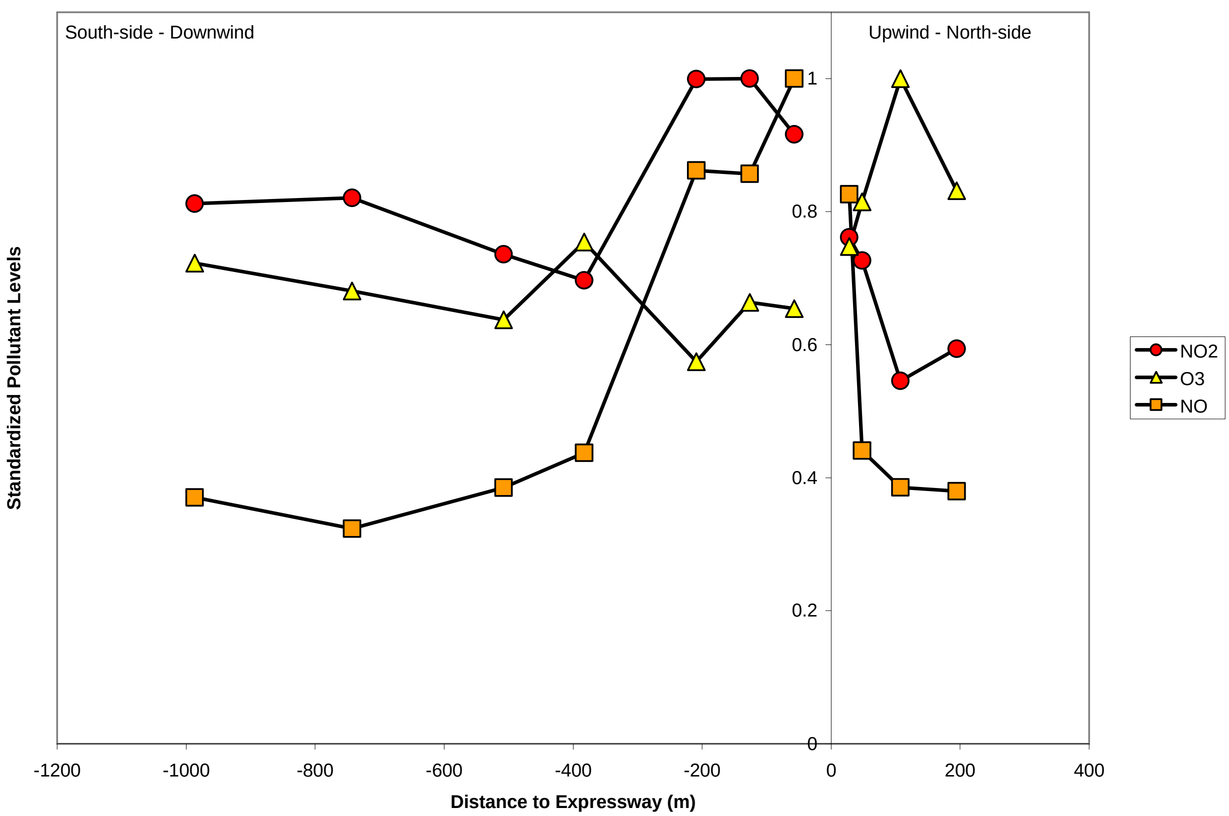

Figure 10. Pollutant concentrations around Highway 401, Toronto (Beckerman et al., 2006)

Based on this limited information we estimate the uncertainty of the means of the near-roadway proximity factor distributions to be uniformly distributed as summarized in Table 7. Uncertainties of the standard deviations of the near-roadway proximity factor distributions have a lesser effect than uncertainties of the means, and we are not assigning uncertainties to them. The vehicle miles traveled are much less uncertain than the titration adjustments, and therefore we do not need to take into account their uncertainty.

Table 7. Uncertainty of the means of near-roadway proximity factor distributions

Location	Uncertainty of the mean of the distribution	Distribution of uncertainty (normal)		
		Uncertainty	Mean	Standard deviation
outdoors near road and parking lots	90% within [0.605, 0.905]	90% within [−0.15, 0.15]	0	0.09
in-vehicle, local roads	90% within [0.605, 0.905]	90% within [−0.15, 0.15]	0	0.09
in-vehicle, urban roads	90% within [0.604, 0.904]	90% within [−0.15, 0.15]	0	0.09
in-vehicle, interstates	90% within [0.214, 0.514]	90% within [−0.15, 0.15]	0	0.09

Uncertainty of Indoor Near-Roadway Concentrations

APEX considers a person to be near a roadway when their activity diary puts them in the near-road microenvironment. There is no consideration of the effects of roadways on the concentrations in residences near roadways, and this is an additional source of uncertainty, since a significant portion of the population live near roadways (the 2001 American Housing Survey [U.S. Census Bureau, 2002] estimated that an eighth of the housing units in the U.S. are within 300 feet of a four or more lane highway, railroad, or airport). We quantify the effects of this uncertainty by performing the exposure modeling to account for this in a simplistic way, and comparing those modeling results with the standard APEX results. We base this analysis on a data base that specifies the fraction of the population in each Census tract that live within 75 m from a major roadway. This data base is described in Appendix I in the Exposure Analysis TSD.

We performed APEX simulations using the outdoor near-road proximity factors to decrease the ambient concentrations outside residences within 75 m from a major roadway. A comparison of these model results with the standard simulations shows only slight decreases in population exposures to high 8-hour average ozone levels. For example, the number of children (ages 5 to 18) predicted by APEX to experience one or more exposures above 0.07 ppm-8hr concomitant with moderate exertion decreased by three percent.

Note that the phenomenon of titration by NO near roadways also has the potential to influence the exposure results in the other direction, in cases where the ozone monitors are located in areas of high traffic. Then the measurements could be low in comparison with other locations not affected by traffic emissions. However, there are criteria for siting monitors, which specify how far to site monitors from roads to avoid interference that would make the monitor unrepresentative of the surrounding area (US EPA, 1998).

Uncertainty of the Vertical Profile of Concentrations

Ozone concentrations vary with height within the lower boundary layer of the atmosphere, which can lead to exposure error for people living in high-rise apartment buildings (when significantly higher than the ozone monitors) and in cases where an ozone monitor is significantly higher than the surrounding population. The CD (page AX3-202) states that:

> A study of the effect of elevation on O_3 concentrations found that concentrations increased with increasing elevation. The ratio of O_3 concentrations at street level (3 m) compared to the rooftop (25 m) was between 0.12 and 0.16, though the actual concentrations were highly correlated (r = 0.63) (Väkevä et al., 1999). Differential O_3 exposures may, therefore, exist in apartments that are on different floors. Differences in elevation between the monitoring sites in Los Angeles and street level samples may have contributed to the lower levels measured by Johnson (1997). Furthermore, since O_3 monitors are frequently located on rooftops in urban settings, the concentrations measured there may overestimate the exposure to individuals outdoors in streets and parks, locations where people exercise and maximum O_3 exposure is likely to occur.

We do not intend to address this source of uncertainty at this time, due to a lack of data on the vertical distribution of concentrations near the surface in urban areas.

Uncertainty in Concentration Rollback to Reflect Alternative Standards

One method for assessing the uncertainty of the rollback adjustments used in our modeling analyses is to apply the rollback procedure to historical air quality data and compare the observed air concentrations with the rolled-back concentrations (Rizzo, 2005). There are difficulties in translating this uncertainty into uncertainties of the APEX model inputs, and so we employ a different approach.[1]

This approach entails using the rollback method to adjust ozone concentrations for each of the 3-year periods 2000-2002, 2001-2003, and 2002-2004, to reflect air quality representative of just meeting the current 8-hr ozone standard of 0.08 ppm. For each of these 3-year periods, design values for the current standard were calculated and hourly ozone concentrations rolled-back to meet the standard. Since each of these 3-year sets of concentrations represents just attaining the current standard, differences between them are due to uncertainty of the rollback

1 Staff notes that this evaluation is ongoing. Staff anticipates completing this analysis and presenting the results in a staff memo to be made available in the docket for this rulemaking.

method. Each of these 3-year periods are being modeled using APEX for all 12 cities. The variability in the model results for each city provide estimates of a lower bound of the uncertainty of rollback in the modeled exposures. Comparisons are based on the distributions of modeled exposures over the 3-year periods, since it is the 3-year period which is being brought to just attaining the standard, and not each individual year.

Meteorological Data

Uncertainty of Ambient Temperatures

Temperatures are the only meteorological inputs to APEX for this application. Temperatures input to APEX are specified not as distributions but as hourly and daily values measured from one or more monitors. Thus, temporal and spatial variability are accounted for. Due to the smooth nature of the temporal and spatial variability of temperatures, the uncertainty of the temperature inputs is typically small and therefore we can ignore this source of uncertainty.

Most of the temperature sites have no missing data; a few have 1 or 2 days missing during the year. Thus, the uncertainty from the estimation of missing temperature data is insignificant.

Modeling Concentrations in Microenvironments

The importance of estimation of concentrations in indoor microenvironments (homes, offices, schools, restaurants, vehicles, etc.) is underscored by the finding that personal exposure measurements of ozone are often not well-correlated with ambient measurements (CD, pages 3-59 to 3-61).

The microenvironmental characteristics used to model the concentrations in microenvironments tend to be highly variable, both in different microenvironments (e.g., different houses have varying characteristics) and within a single microenvironment (e.g., the characteristics of a specific house can vary over time). Since APEX is a probabilistic model, if data accurately characterizing this variability can be provided to the model, this will not result in uncertainties. However, even if we can appropriately characterize the distributions of each microenvironmental parameter, there will be significant uncertainties unless we appropriately model the relationships (e.g., correlations) between the different microenvironmental parameters, as well as the relationships between the microenvironmental parameters and other components of the exposure model (e.g., people's activities). There are 12 microenvironments modeled in APEX for this application, listed in Table 8. The mass balance and factors models used to calculate ozone concentrations in these microenvironments are described in the APEX TSD.

Table 8. Microenvironments Modeled For Ozone Exposure Assessment

Microenvironment	Model	Parameters[1]
Indoors – Residence	Mass balance	AER and DE
Indoors – Bars and restaurants	Mass balance	AER and DE
Indoors – Schools	Mass balance	AER and DE
Indoors – Day-care centers	Mass balance	AER and DE
Indoors – Office	Mass balance	AER and DE
Indoors – Shopping	Mass balance	AER and DE
Indoors – Other	Mass balance	AER and DE
In-vehicle – Cars and Trucks	Factors	PE and PR
In-vehicle - Mass Transit	Factors	PE and PR
Outdoors – Near road	Factors	PR
Outdoors – Public garage - parking lot	Factors	PR
Outdoors – Other	Factors	PR

[1] AER: Air Exchange Rate, DE: Decay rate; PE: Penetration factor; PR: Proximity factor

Uncertainty of Air Exchange Processes

The air exchange rate (AER) is one of the most important factors in determining the ratio of outdoor to indoor concentrations of ozone and therefore in determining exposures while indoors. AERs are highly variable at hourly and daily time scales, both within a microenvironment over time and between microenvironments of the same type in different buildings. AERs depend strongly on the physical characteristics of a microenvironment and also on the behavior of the occupants of the microenvironment. For example, the concentration in a house when a person enters the house will depend on the AER of the preceding hour, which could depend on whether or not there was someone else already in the house. There is also some dependence on the atmospheric conditions (temperature, humidity, and wind speed), both directly (higher wind speeds result in higher AERs in most circumstances) and indirectly (occupants can open and close windows in response to the outdoor temperature).

AER measurements (which are used to derive the APEX input distributions for a city) typically involve fitting tracer concentrations to simple mass balance models. This analysis of AER uncertainty currently does not take into account the uncertainty in this measurement process.

Air Exchange Rates

City-specific lognormal distributions of AERs for use with the APEX ozone model were developed based on an analysis of AER data from several studies, described in the Exposure Analysis TSD, Appendix A. The parameters of these distributions depend on the outside temperature and whether or not the residence has air conditioning.

We assess the within-city uncertainty by using a bootstrap distribution to estimate the effects of sampling variation on the fitted geometric means (GMs) and standard deviations

(GSDs) for each city. This analysis is described in the Exposure Analysis TSD. The bootstrap is a nonparametric method for estimating uncertainty which accounts for the correlation between the GMs and GSDs (e.g., see Figure 12), so that there are not unrealistic combinations of GMs and GSDs. The bootstrap distributions assess the uncertainty due to random sampling variation but do not address uncertainties due to the lack of representativeness of the available study data. This could be assessed, to some extent, by comparing AER distributions from different studies in the same city. However, data are not available to do this, and we assign a ten percent uncertainty to the potential non-representativeness of the measured AER distributions.

Several bootstrap distributions were developed for residential air exchange rates, one for each city-temperature-A/C combination. Examples of two of the bootstrap uncertainty distributions, which illustrate that the GMs and GSDs are not independent, are provided in Figure 11 and Figure 12. Figure 11 shows the uncertainty distribution around the model input values GM=0.916 and GSD=2.451, which specify the distribution of AERs of residences in Houston without A/C when ambient temperatures are above 20 degrees C (24-hour average). Similarly, Figure 12 shows the uncertainty distribution for the AER distribution parameters for Los Angeles for residences without A/C when the ambient temperature is above 25 degrees C. Note that in each of these figures there is only one “original data” point (this is the APEX input value), indicated by the intersection of the cross-hairs in the figure. The clouds of points are bootstrapped values.

In the Monte Carlo uncertainty simulations, a GM, GSD pair is selected at random from the bootstrap uncertainty distribution for each temperature-A/C combination, and used for input to APEX. (APEX then selects AER values randomly from the log-normal distribution with the bootstrap GM and GSD). The uncertainty of non-residential air exchange rates was modeled in the same way, using bootstrap distributions of GM, GSD pairs.

Uncertainty of Residential Air Conditioning Prevalence and Use

The AER distributions input to APEX are conditioned on the presence or absence of air conditioning, and estimates of residential air conditioning prevalence rates for each modeled area were obtained from the American Housing Survey of 2003. Appendix F of the Exposure Analysis TSD gives confidence intervals for the air conditioning prevalence rates, reproduced here in Table 9. We model the uncertainty of the prevalence rates with zero-mean normal distributions with standard deviations equal to the standard errors given in Table 9.

In addition to the uncertainty of prevalence rates, there is uncertainty about the amount of use of A/C given that a house or office has A/C. However, most of the studies of AERs that we used to develop AER distributions report presence or absence of air conditioning, and not whether the A/C was being used (Appendix A, Exposure Analysis TSD). Thus, the variability resulting from the use or non-use of A/C is built into the AER distributions, and is being taken into account. If, in the future, we have sufficient data to allow us to characterize AERs separately for conditions of use and non-use, then we can supply APEX with these distributions, as well as distributions for use vs. non-use.

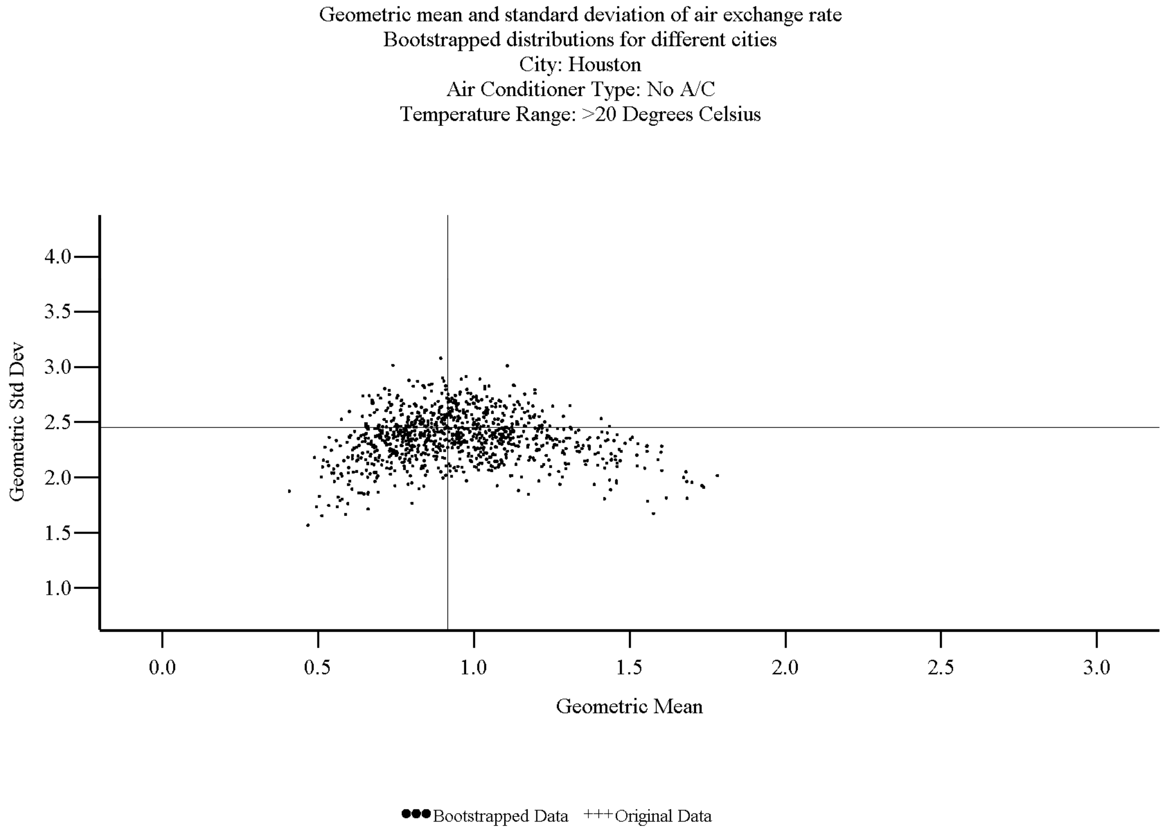

Figure 11. Bootstrap distribution of AER uncertainty for Houston, no A/C, >20 C

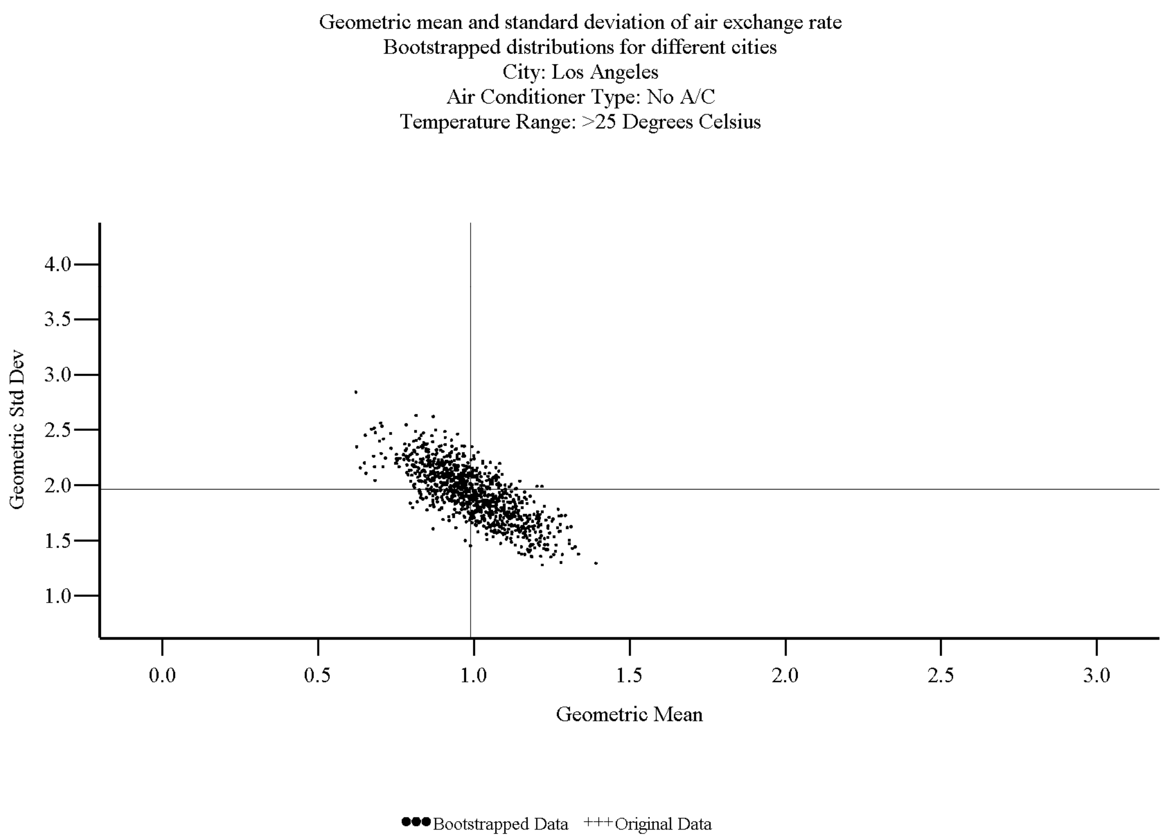

Figure 12. Bootstrap distribution of AER uncertainty for Los Angeles, no A/C, >25 C

In each of these figures there is only one "original data" point (the APEX input value), indicated by the intersection of the cross-hairs in the figure. The clouds of points are bootstrapped values.

Table 9. Uncertainty of air conditioning prevalence rates

City	Prevalence rate (%)	Standard error	Lower 95% confidence bound	Upper 95% confidence bound
Atlanta, 2003	97.0	1.18	94.7	99.3
Boston, 2003	85.2	2.14	81.0	89.4
Chicago, 2003	87.1	1.39	84.4	89.8
Cleveland, 2003	74.6	3.38	68.0	81.3
Detroit, 2003	81.4	1.76	78.0	84.9
Houston, 2003	98.7	0.67	97.4	100.0
Los Angeles, 2003	55.1	1.70	51.7	58.4
New York, 2003	81.6	1.27	79.1	84.1
Philadelphia, 2003	90.6	1.30	88.1	93.2
Sacramento, 2003	94.6	1.93	90.8	98.4
St. Louis, 2003	95.5	1.67	92.3	98.8
Washington DC, 2003	96.5	1.00	94.5	98.4

Uncertainty of Deposition, Filtration, and Chemical Reaction Processes

The removal of ozone from a microenvironment due to deposition, filtration, and chemical reaction processes is modeled in APEX by a distribution of ozone decay rates. The rate of deposition of ozone to a surface depends on the material the surface is made of, the humidity, and the concentration of ozone. The rate of removal of ozone due to deposition in a specific microenvironment also depends on the dimensions, surface coverings, furnishings, and the ratio of surface area to volume in the microenvironment. The degree of ozone loss through filtration is a function of the HVAC system in the microenvironment. Other chemical processes that contribute to reduction in ozone concentrations indoors include reaction with NO emitted from gas stoves and reaction with VOCs from cleaning products.

The distribution of ozone decay rates used in the present study represents the decay rates measured in a study of 17 residences in Southern California (Lee et al., 1999). A lognormal distribution was fit to the measurements from this study, yielding a geometric mean (GM) of 2.5 and a geometric standard deviation (GSD) of 1.5. These values are constrained to lie between 0.95 and 8.05 hour^{-1}. We estimate the uncertainty of this distribution using a bootstrap method described by Cullen and Frey (1999). This method quantifies sampling uncertainty nonparametrically. The standard deviations of the bootstrap samples for the GM and GSD are given in Table 10; however, bear in mind that the GMs and GSDs in these samples are not independent.

The bootstrap approach quantifies sampling uncertainty, but does not account for uncertainty resulting from nonrepresentativeness of the study in relation to the residences we are modeling. Weschler (2000) summarizes the results of several studies which measured rates of ozone removal in indoor environments, and concludes that "earlier measurements in homes and offices are in good agreement with the values reported by Lee and coworkers," referring to the Lee et al. (1999) study, which he characterizes as "the most extensive set of measurements in the literature." Based on this review, we estimate the nonrepresentativeness uncertainty by assuming that the GM of 2.5 is unbiased but is correct to within 10 percent with 90 percent confidence, and represent this uncertainty with a normal distribution with a standard deviation of 0.15.

It is reasonable to assume that this uncertainty is independent of the sampling uncertainty, and therefore we combine these uncertainties independently. Table 10 summarizes our estimates of the uncertainty of ozone decay rates.

Table 10. Uncertainty of lognormal distributions of ozone decay rates input to APEX

Source of Uncertainty	Uncertainty of Geometric Mean (hr^{-1})	Uncertainty of Geometric Standard Deviation (hr^{-1})
Finite sample	bootstrap (mean = 0, st. dev. = 0.1)	bootstrap (mean = 0, st. dev. = 0.05)
Nonrepresentativeness of the study	normal distribution, mean = 0, st. dev. = 0.15	none

Uncertainty of Vehicle Penetration Factors

A vehicle penetration factor distribution (normal, mean 0.3, standard deviation 0.232, lower bound 0.1, upper bound 1.0) was developed with data from the Cincinnati Ozone Study (Johnson et al, 1995). This was a scripted study using three cars in one city in 1994, designed for the purpose of developing distributions for exposure modeling. This distribution is consistent with other studies of concentration ratios inside and outside of vehicles (Chan et al., 1991; Chan and Chung, 2003; Riediker et al., 2003), but we have not found data that would support quantitative estimates of the uncertainty of this vehicle penetration factor distribution. For this analysis we assume that the mean of the vehicle penetration factor distribution is likely to lie within ±50 percent of the APEX input value. We feel that this estimate is more realistic than the implied uncertainty of zero if we do not include this source of uncertainty due to a lack of data. We represent this uncertainty with normal distributions such that the mean values input to APEX are between 0.15 and 0.45 with 90 percent probability, with an average value of 0.3.

Characterization of Population Demographics

Uncertainty of Demographic Model Inputs

Data from the 2000 Census provide the demographics of the modeled populations. When modeling a year close to the year of the Census, the uncertainty of the demographic mix of the population is relatively small, compared with the other uncertainties of APEX, and therefore we are not treating this as an explicit source of uncertainty in this analysis. The Census data input to APEX at a tract level are:

- age
- gender
- race (not used in this modeling analysis)
- home location (Census tract)
- work location (Census tract)
- employment probabilities (by age, gender, tract)
- between-tract commuting probabilities

However, we can quantify changes in the size of the total populations between the year of the Census (2000) and the year being modeled. Table 11 lists the percent increase in population from 2000 to 2002 and 2004 for the 12 modeled CSAs (calculated from the *Subcounty Population Estimates, April 1, 2000 to July 1, 2004*, Population Estimates Program, U.S. Bureau of the Census Release dated June 30, 2005).

Table 11. Change in populations from 2000 to 2002 and 2004

Urban Area (CSA)	2000 to 2002 % change	2000 to 2004 % change
Atlanta, GA	5	10
Boston, MA	1	1
Chicago, IL	2	3
Cleveland, OH	0	0
Detroit, MI	0	1
Houston, TX	5	9
Los Angeles, CA	3	7
New York, NY	1	2
Philadelphia, PA	1	2
Sacramento, CA	6	11
St. Louis, MO	1	2
Washington, DC	3	5

The biases resulting from population changes likely cancel to a large degree when assessing relative differences between exposure scenarios.

Modeling People's Activity Patterns

APEX models the variability of activities of individuals by random sampling of daily activity patterns in the Consolidated Human Activity Database (CHAD). CHAD consists of a collection of 24-hour "diaries" compiled from several studies. Each diary specifies the activities of an individual during the day, the locations of the individual during the activities, and the time period of each activity. The durations of the events in the diaries range from a few minutes to several hours.

Uncertainty of the Activity Pattern Data

The activity pattern database (CHAD) input to APEX is a very complex multivariate database which, due to its complexity, is less amenable than other model inputs to the Monte Carlo approach to uncertainty analysis. In particular, it would be very difficult to vary a set of characteristics of CHAD and generate different diary databases reflecting the varied characteristics. In addition, we don't know a priori what the important characteristics of CHAD are with respect to uncertainties of exposure modeling. A further complication is that we must consider the uncertainties of CHAD in the context of the formation of a year-long activity sequence made up of diary days sampled from CHAD, for each individual simulated by APEX. The uncertainty that results from the method for assembling diary-days for each individual could also be important. The following are limitations of CHAD that result in uncertainties in modeling exposures.

- Diary errors, particularly the recall studies (72% of CHAD diaries are recall). There is extensive literature on diary errors; Takarangi et al. (2006) provide an instructive commentary on factors which conspire to produce inaccurate diary data.
- Incompatibility of the CHAD categories/codes with the coding schemes in the different studies in CHAD (each study's codes are mapped to the CHAD codes)
- Nonrepresentativeness of non-random studies
- Nonrepresentativeness of older studies (42% are pre-1990, 98% are pre-1995)
- Geographic (city-specific) nonrepresentativeness
- Sample size limitation. This is particularly important because of the stratification required for appropriate use of the data in exposure modeling.
- Longitudinal autocorrelation of activities is not characterized.
- Geographical locations of activities away from the home are unknown.

Since is difficult to characterize the uncertainties in CHAD and then to propagate these uncertainties to the model results using the Monte Carlo approach, we performed a sensitivity analysis, comparing the APEX results based on all of the CHAD diaries with results from running APEX using only the CHAD diaries from the National Human Activity Pattern Study (NHAPS), the most comprehensive and nationally representative study in CHAD. NHAPS is national in scope, with a random design, and comprises more than half of the CHAD diaries for all ages, and 43 percent of the diary days in CHAD for children ages 5 to 18.

APEX simulations were performed using only NHAPS diaries for all 12 urban areas, for the 2002 base case and the scenario of meeting the current NAAQS. The results of this comparison for the 2002 base case simulations are presented in Table 12 for the percent of children at moderate exertion with 8-hour exposures above exposure levels of 0.06, 0.07, and 0.08 ppm-8hr. The comparison of estimated reductions in exposures to children at moderate exertion in going from the base case to the current standard is presented in Table 13.

There is very good agreement between the APEX results, whether all of CHAD or only the NHAPS component of CHAD is used, indicating that the model results are not being unduly influenced by any single study in CHAD. This also indicates that the method used in APEX for stratifying diaries when assigning diaries to simulated individuals is appropriate.

Table 12. Comparison of APEX 2002 base case simulations: All CHAD vs. the NHAPS part of CHAD. Percent of children at moderate exertion with 8-hour exposures above levels of 0.06, 0.07, 0.08 ppm-8hr.

	Above 0.06 ppm-8hr			Above 0.07 ppm-8hr			Above 0.08 ppm-8hr		
CSA	All CHAD	NHAPS only	Absolute difference	All CHAD	NHAPS only	Absolute difference	All CHAD	NHAPS only	Absolute difference
Atlanta	64%	65%	1%	34%	38%	3%	11%	14%	4%
Boston	62%	60%	(2%)	41%	41%	0%	20%	21%	1%
Chicago	67%	66%	(0%)	40%	42%	2%	15%	16%	1%
Cleveland	74%	73%	(1%)	57%	58%	0%	31%	34%	3%
Detroit	70%	69%	(0%)	46%	49%	3%	18%	20%	2%
Houston	55%	55%	(0%)	26%	28%	2%	11%	12%	2%
Los Angeles	61%	60%	(1%)	35%	35%	(0%)	16%	17%	1%
New York	71%	70%	(1%)	49%	49%	0%	25%	25%	1%
Philadelphia	74%	73%	(1%)	57%	55%	(1%)	34%	34%	0%
Sacramento	64%	65%	1%	36%	39%	2%	13%	16%	3%
St. Louis	70%	69%	(1%)	50%	50%	(0%)	21%	22%	1%
Washington	72%	72%	(0%)	50%	51%	1%	25%	27%	2%

Table 13. Comparison of APEX simulations: All CHAD vs. the NHAPS part of CHAD. Percent reduction[1] from the 2002 base case to the current standard of the number of children at moderate exertion with 8-hour exposures above levels of 0.06, 0.07, 0.08 ppm-8hr.

	Above 0.06 ppm-8hr			Above 0.07 ppm-8hr			Above 0.08 ppm-8hr		
CSA	All CHAD	NHAPS only	Absolute difference	All CHAD	NHAPS only	Absolute difference	All CHAD	NHAPS only	Absolute difference
Atlanta	26%	23%	(3%)	57%	50%	(7%)	76%	71%	(4%)
Boston	21%	19%	(2%)	41%	39%	(2%)	58%	55%	(3%)
Chicago	27%	24%	(3%)	54%	55%	1%	84%	82%	(2%)
Cleveland	17%	16%	(2%)	46%	41%	(5%)	79%	79%	(1%)
Detroit	17%	15%	(2%)	44%	42%	(3%)	84%	79%	(5%)
Houston	58%	54%	(4%)	77%	73%	(3%)	91%	89%	(2%)
Los Angeles	88%	85%	(3%)	98%	96%	(1%)	100%	99%	(0%)
New York	37%	34%	(3%)	71%	67%	(4%)	91%	88%	(3%)
Philadelphia	18%	18%	(0%)	42%	39%	(2%)	74%	69%	(5%)
Sacramento	51%	48%	(3%)	81%	75%	(6%)	93%	91%	(2%)
St. Louis	9%	9%	(0%)	27%	26%	(1%)	53%	50%	(3%)
Washington	22%	20%	(2%)	49%	47%	(2%)	73%	73%	(0%)

[1] The percent reductions are calculated as 100(base case results – current standard results)/(base case results).

Uncertainty of Longitudinal Diary Assembly

The method in APEX for assembling longitudinal diaries is intended to capture the tendency of individuals to repeat activities (this method is described in detail in the Exposure Analysis TSD). There are two model input parameters that control the strength of this tendency in the simulated individuals, a population diversity statistic (***D***) and a within-person autocorrelation statistic (***A***). For the current application, these statistics are based on the time a person spends outdoors each day, which is one of the most important determinants of exposure to ozone. The ***D*** statistic reflects the relative importance of within-person variance and between-person variance in the outdoor time. The ***A*** statistic specifies the day-to-day autocorrelation of outdoor time. The values used for this analysis (0.2 for ***D*** and 0.2 for ***A***) are based on one study of school age children, and may be considerably uncertain. To reflect this uncertainty in the Monte Carlo analysis, we allow ***D*** and ***A*** to vary independently, uniformly within a factor of two of their base values (varying from 0.1 to 0.4) Table 14 gives the distributions of (additive) uncertainty about the base values. The uncertainty of longitudinal diary assembly is discussed further in the Model Uncertainty section below.

Table 14. Uncertainty of longitudinal diary parameters

Parameter	Distribution of uncertainty
population diversity statistic (***D***)	Uniform on [−0.1, 0.2]
within-person autocorrelation statistic (***A***)	Uniform on [−0.1, 0.2]

Modeling Physiological Processes

The uncertainties of the inputs to the physiological model are discussed in this section. Also see the discussion of the physiological model in APEX in the Model Uncertainty section below.

Uncertainty of Physiological Model Inputs

The physiological model inputs to APEX are provided as parameters for distributions reflecting population variability. These have been recently updated by Isaacs and Smith (2005). The following distributions and parameters are input to APEX:

- Body mass (BM) (kg) distributions by age and gender
- Normalized maximal oxygen uptake (NVO_2max) distributions by age and gender
- Resting metabolic rate (RMR) (kcal/min) age- and weight-specific regression equations
- Metabolic equivalent (MET) distribution for each activity type (dimensionless). Distributions for a few activities are occupation- and age-dependent.
- Effective ventilation rate (EVR) cutpoints for specifying levels of exertion (liters of oxygen per minute per m^2) (e.g., 1-hour average EVR > 16 indicates moderate or greater exertion)
- Active PAI cutpoint (a person is characterized as “active” if the median of all of their daily PAI values is > 1.75) (dimensionless)

Body Mass Distributions

The distributions of body mass come from the most recent data from the National Health and Nutrition Examination Survey (NHANES), compiled for the years 1999-2004 (CDC, 2005). The NHANES body mass data are sampled and weighted to provide unbiased national estimates of body mass. There will be some uncertainty due to regional/city differences. However, the uncertainty in the body mass distributions is small compared to the other uncertainties in the APEX input data, and we are treating it as insignificant.

NVO_2max Distributions

NVO_2max is used in the calculation of the maximum metabolic activity level that can be sustained for about five minutes, which ensures that modeled values do not exceed realistic limits. Since this rarely occurs, NVO_2max is not an influential model input. These distributions were recently updated based on an extensive review of the literature and acquisition of data (Isaacs and Smith, 2005). The uncertainty of this input is significantly less influential than other uncertainties, and we are treating it as insignificant.

Resting Metabolic Rates

The age- and weight-specific RMR regression equation coefficients input to APEX are integral to the RMR model in APEX and are not intended to be modified by the user. These are model inputs only for the purpose of facilitating sensitivity analyses. The uncertainty of the model for predicting resting metabolic rates is discussed below in the section on model uncertainty.

MET Distributions

The distributions of activity-specific MET values are of fundamental importance to the physiological model in APEX. Johnson (2003, section 9.6) states:

> Perhaps the weakest link in the algorithm is the step which requires the analyst to provide a distribution of possible MET values for each activity code. These distributions are currently based on distributions provided by the developers of CHAD (McCurdy et al., 2000). Because available data were often insufficient to accurately define a distribution for each activity code, the developers tended to follow a conservative approach and over-estimate the variability of each distribution. Consequently, the Ve values produced by the ventilation rate algorithm may exhibit an excessive degree of variability.

McCurdy et al. (2000), in a paper describing the development of the MET distributions in CHAD, state:

> At this stage of development, the METs distribution assignment effort should be viewed as being preliminary in nature. More work is needed to better relate activity codes used in human activity pattern surveys to those long used by exercise physiologists and clinical nutritionists.

Most of the MET distributions in CHAD were developed based on Ainsworth et al. (1993), which has been updated and revised in 2000 (Ainsworth et al., 2000; Ainsworth, 2003). CHAD has not yet been updated with this newer information.

There is uncertainty in the MET distributions related to the question of how well the MET distributions for defined activities represent the actual exertion during the discrete event duration. For example, a diary event for an hour may be coded "play basketball" (which has a relatively high MET value), but in reality the MET value may be much lower for the hour, since it is likely that the hour-long event contains periods of rest. Also, there is uncertainty in the use of the CHAD MET distributions for children, the elderly, and persons with compromised health, since they were derived from healthy adults. Puyau et al. (2002) show that adult-derived MET cutpoints are not applicable to children.

Although the uncertainty of the assignment of MET to specific individuals and activities in APEX may be high, it is the MET distributions over the populations modeled that is crucial to producing realistic model results. In order to assess this we compared the distributions of MET generated in APEX simulations to values in the literature. Brochu, et al. (2006) provide summaries of a recent study which collected over 20,000 days of physiological measurements from over 1,000 individuals. Table 15 compares the mean daily MET by gender and age group reported by Brochu et al. with the corresponding values from the APEX Boston simulations for normal-weight individuals (age 3 to 19 years with BMI less than the 85th percentile by age; ages 20 to 96 years with BMI from 18.5 to 25 kg/m^2). The agreement is very good, with the APEX values generally higher by 0.1 to 0.2. For the Monte Carlo uncertainty simulations, we shift the means of the MET distributions randomly (uniform distribution) within ±5 percent.

Table 15. Comparison of measured daily MET values with APEX modeled values

Age group (years)	Measured DMET (males)	Measured DMET (females)	APEX DMET (males)	APEX DMET (females)
3 to < 10	1.5 ±0.2	1.5 ±0.2	1.7 ±0.2	1.6 ±0.2
10 to < 18	1.7 ±0.2	1.7 ±0.3	1.8 ±0.3	1.8 ±0.3
18 to < 30	1.8 ±0.2	1.8 ±0.3	2.0 ±0.5	1.9 ±0.4
30 to < 60	1.8 ±0.2	1.8 ±0.3	2.0 ±0.4	1.9 ±0.4
60+	1.6 ±0.3	1.6 ±0.3	1.8 ±0.4	1.7 ±0.4

[1] Measured values from Table Web-4, Brochu, et al. (2006). (dimensionless)

EVR Exertion Level Cutpoints

The EVR cutpoints input to APEX define moderate and heavy exertion levels, used to stratify exposures by the level of exertion during exposures for the risk calculations (discussed in the Staff Paper) and are not considered as uncertain here. (As opposed to the values of EVR calculated by APEX, which are uncertain.)

Active PAI cutpoint

The Active PAI cutpoint is used to classify a simulated individual as active or not. In order to address the uncertainty of the PAI cutpoint used in the exposure modeling analysis, one must have a clear definition of what it means for a person to be characterized as active. Then one could assess the extent to which the PAI cutpoint classification is accurate. We do not have such a definition, and have essentially been using the PAI cutpoint as defining an active person. We discuss this further in the section on model uncertainty below.

Simulation Convergence

APEX is a probabilistic model with numerous inputs and parameters defined in terms of probability distributions which reflect the natural variability of the physiology and activities of individuals and of physical processes. In order to realistically estimate distributions of population exposures, a sufficient number of individuals must be simulated by APEX to allow the input distributions to be sampled enough times so they are adequately represented.

For this discussion we denote the number of simulated individuals in an APEX run by N_S. As N_S for a model run increases, the predicted exposure distributions converge to a limiting distribution. If too few individuals are simulated, then the results of simulations with identical inputs will differ because too few values from the input distributions are being sampled to properly characterize them.

To illustrate this phenomenon, we ran thousands of APEX simulations with identical inputs, but with varying N_S. From each APEX run we calculated statistics based on the predicted distributions of exposures, for example, the fraction of the population who experience one or more hourly exposures greater than 0.12 ppm-hr. For runs with very few people simulated, these statistics are not stable and can vary widely; but for runs with many people simulated, the statistics have values that are closer together for the different model runs. This is illustrated in Figure 13, where we have plotted the spread of one statistic against N_S. The horizontal axis gives N_S, for 1000, 2000, up to 15,000, simulated in each APEX run. The vertical axis is the fraction of the population who experience one or more hourly exposures greater than 0.12 ppm-hr, and the collection of those values for all runs with a given N_S is presented as a box plot (each model run provides one value). The bottom and top edges of a box indicate the 25th and 75th percentiles; whiskers are at the 5th and 95th percentiles; and squares indicate values outside this range. We see that for N_S = 1000, this statistic ranges from 0.172 to 0.242 (±17% from the median), while for N_S = 15,000 the range is only from 0.198 to 0.213 (±4% from the median).

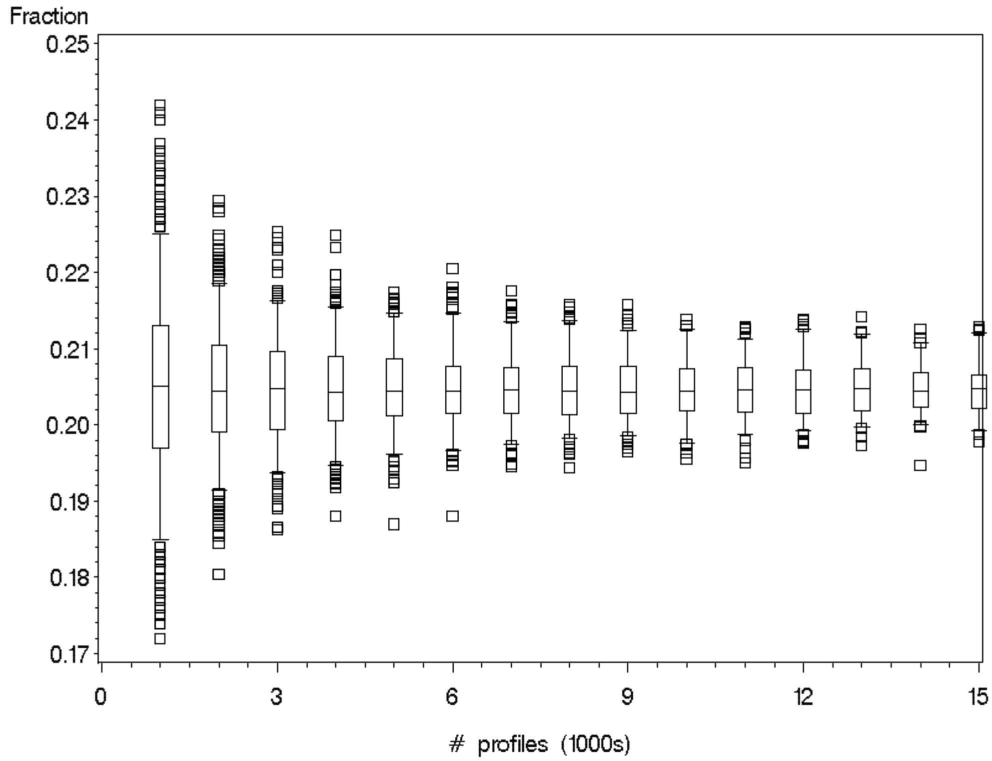

Figure 13. Distribution of the predicted fractions of population who experience any hourly exposures > 0.12 ppm-hr as a function of the number of profiles simulated

In practice, we model the distribution of exposures with a single simulation, and the deviation of this distribution from the limiting distribution (obtained with very large N_S) is an error, or uncertainty, due to lack of convergence. Since model run time is proportional to N_S, the N_S that one can simulate depends on the computing capacity and the time requirements. For the hundreds of APEX runs performed in support of the ozone NAAQS review, we simulated 60,000 individuals in each APEX run, to balance the desire for convergence with time limitations.

We have assessed the extent of "non-convergence uncertainty" for N_S = 60,000 for one city, Atlanta, for the 2002 base case scenario, by conducting several APEX simulations identical to the single simulation whose results are used in the exposure assessment. Figure 14 (children) and Figure 15 (all people) illustrate this uncertainty with the distributions of the number of people predicted by APEX who experience one or more 8-hour average exposures above 0.08 ppm-8hr, concomitant with moderate or greater exertion. This distribution is made up of the predicted values for 1,268 APEX runs (one value from each run).

In the next four tables, we describe 12 such distributions; Table 16 and Table 17, respectively for children and all people, under moderate exertion, and Table 18 and Table 19 respectively for children and all people, under any exertion level. The last row in Table 16 corresponds to Figure 14 and the last row in Table 17 corresponds to Figure 15. For example, in the distribution in Figure 14, 90 percent of the values are within 4.5 percent of the median (the median should be close to the limiting value as N_S become large).

As expected, convergence is poorer for statistics that are in the tails of the distribution of population exposures. As the exposure cutoff level increases (e.g., going down any column in these tables) or as the population group looked at becomes smaller (e.g., children vs. adults), a larger N_S is required to achieve the same level of convergence. This is illustrated in the summary provided by Table 20.

In these simulations conducted to assess convergence, we allow the starting seed of the sequence of random numbers generated by APEX to be picked randomly based on the date and time of the start of the run, so each simulation has a different starting seed. In the exposure simulations for the 12 cities described in the draft Staff Paper, we used different starting seeds for each city and year simulated, but used the same seed for all runs for a given city and year. For example, the same seed was used for the nine 2002 New York simulations (base case, current standard, 7 alternative standards). In this way the non-convergence uncertainty largely cancels out from the comparisons of the runs for a given city, although we have yet to assess the extent of this cancellation.

Since random seeds are used in the Monte Carlo uncertainty simulations, this aspect of uncertainty is accounted for in the Monte Carlo results.

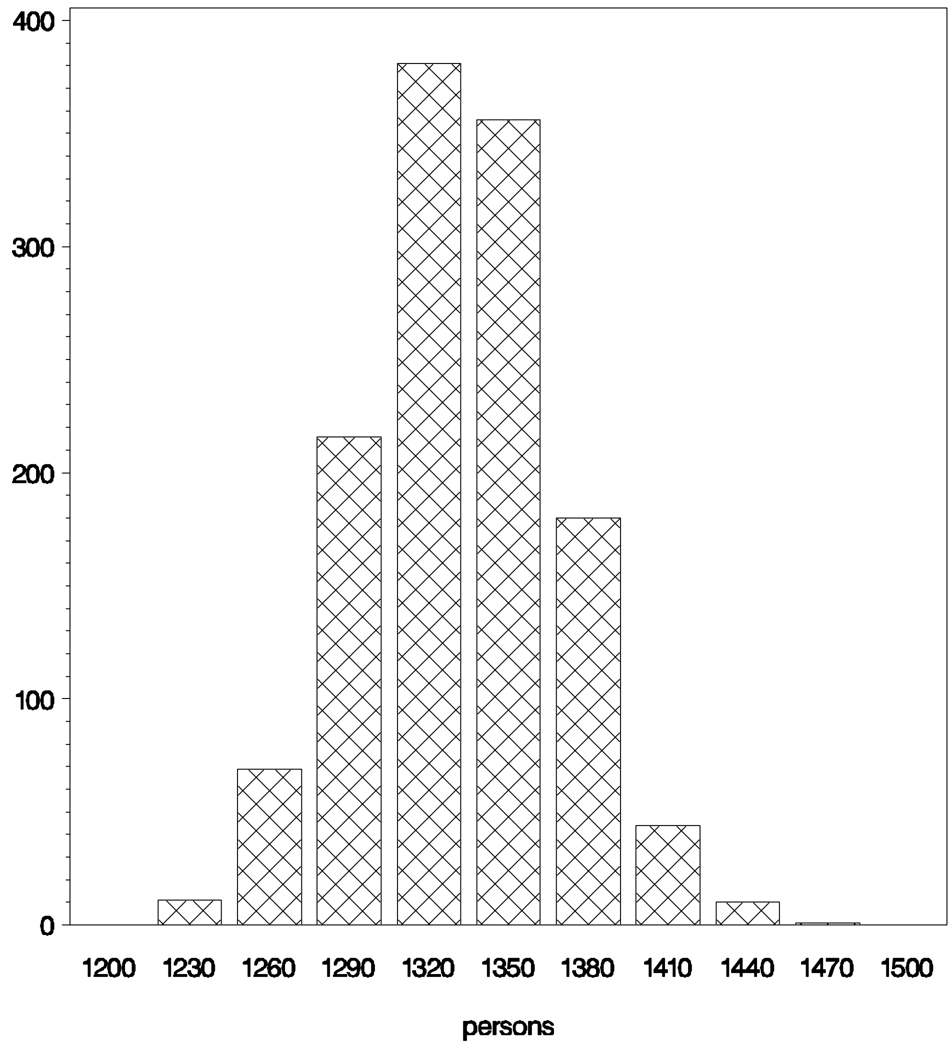

Figure 14. The distribution of the predicted number of children with at least one 8-hour exposure above 0.08 ppm-8hr at moderate or greater exertion for 1,268 repeated simulations of the Atlanta 2002 base case with 60,000 profiles

Table 16. Variability of replicate APEX simulations of 60,000 persons: Medians, 5th, 10th, 25th, 75th, 90th, 95th percentiles, and the percent differences of these from the medians of the number of persons with exposures above different daily maximum 8-hour exposure levels (ppm-8hr) – All children, moderate exertion

Exposure level	median	5th percentile		10th percentile		25th percentile		75th percentile		90th percentile		95th percentile	
0.06	7,873	7,727	(1.9%)	7,761	(1.4%)	7,815	(0.7%)	7,928	0.7%	7,977	1.3%	8,008	1.7%
0.07	4,277	4,169	(2.5%)	4,194	(1.9%)	4,234	(1.0%)	4,323	1.1%	4,361	2.0%	4,384	2.5%
0.08	1,332	1,272	(4.5%)	1,282	(3.8%)	1,306	(2.0%)	1,356	1.8%	1,378	3.5%	1,392	4.5%

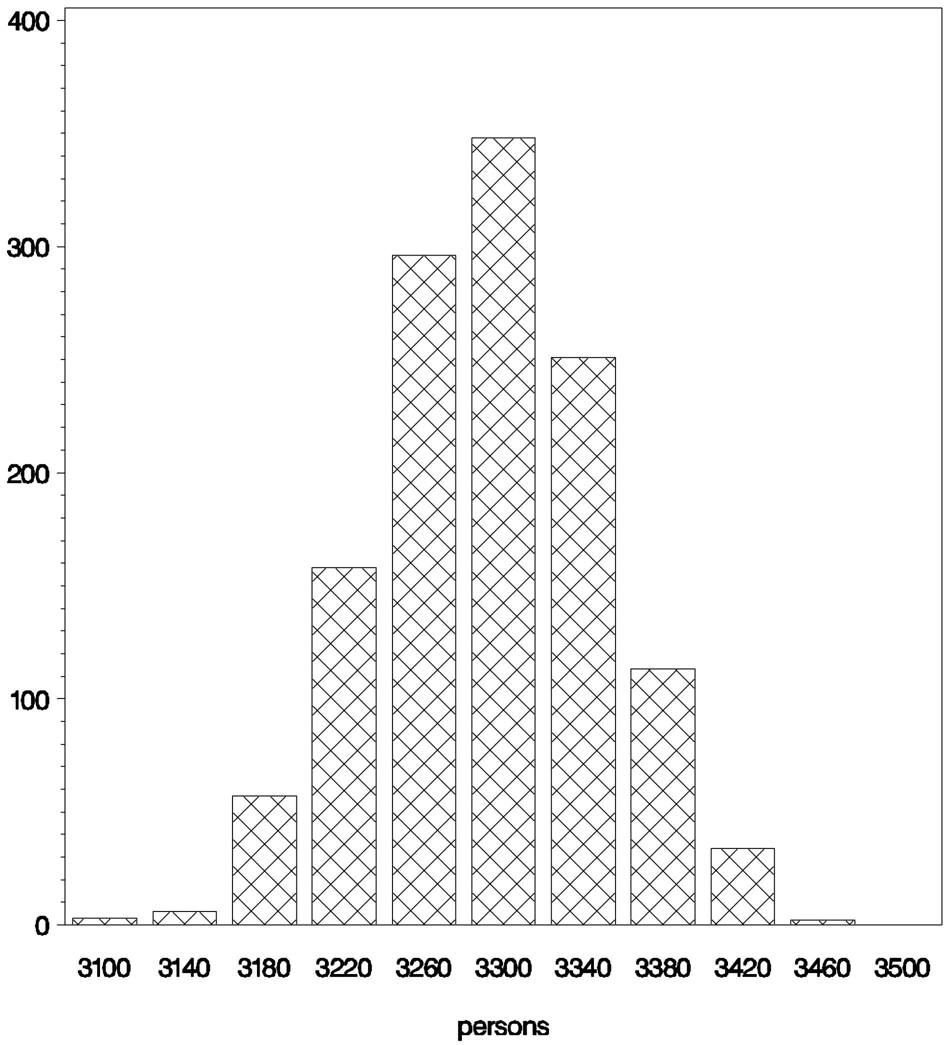

Figure 15. The distribution of the predicted number of people with at least one 8-hour exposure above 0.08 ppm-8hr at moderate or greater exertion for 1,268 repeated simulations of the Atlanta 2002 base case with 60,000 profiles

Table 17. Variability of replicate APEX simulations of 60,000 persons: Medians, 5th, 10th, 25th, 75th, 90th, 95th percentiles, and the percent differences of these from the medians of the number of persons with exposures above different daily maximum 8-hour exposure levels (ppm-8hr) – All people, moderate exertion

Exposure level	median	5th percentile		10th percentile		25th percentile		75th percentile		90th percentile		95th percentile	
0.06	21,816	21,622	(0.9%)	21,664	(0.7%)	21,734	(0.4%)	21,892	0.3%	21,968	0.7%	22,004	0.9%
0.07	10,804	10,642	(1.5%)	10,679	(1.2%)	10,740	(0.6%)	10,865	0.6%	10,931	1.2%	10,960	1.4%
0.08	3,294	3,199	(2.9%)	3,218	(2.3%)	3,254	(1.2%)	3,331	1.1%	3,364	2.1%	3,383	2.7%

Table 18. Variability of replicate APEX simulations of 60,000 persons: Medians, 5th, 10th, 25th, 75th, 90th, 95th percentiles, and the percent differences of these from the medians of the number of persons with exposures above different daily maximum 8-hour exposure levels (ppm-8hr) – All children

Exposure level	median	5th percentile		10th percentile		25th percentile		75th percentile		90th percentile		95th percentile	
0.06	10,405	10,253	(1.5%)	10,281	(1.2%)	10,344	(0.6%)	10,470	0.6%	10,521	1.1%	10,558	1.5%
0.07	6,373	6,237	(2.1%)	6,266	(1.7%)	6,317	(0.9%)	6,419	0.7%	6,465	1.5%	6,499	2.0%
0.08	2,082	2,011	(3.4%)	2,025	(2.7%)	2,052	(1.4%)	2,114	1.5%	2,142	2.9%	2,157	3.6%

Table 19. Variability of replicate APEX simulations of 60,000 persons: Medians, 5th, 10th, 25th, 75th, 90th, 95th percentiles, and the percent differences of these from the medians of the number of persons with exposures above different daily maximum 8-hour exposure levels (ppm-8hr) – All people

Exposure level	median	5th percentile		10th percentile		25th percentile		75th percentile		90th percentile		95th percentile	
0.06	44,107	43,912	(0.4%)	43,962	(0.3%)	44,030	(0.2%)	44,175	0.2%	44,242	0.3%	44,278	0.4%
0.07	25,819	25,613	(0.8%)	25,659	(0.6%)	25,736	(0.3%)	25,904	0.3%	25,974	0.6%	26,017	0.8%
0.08	9,463	9,322	(1.5%)	9,353	(1.2%)	9,403	(0.6%)	9,520	0.6%	9,570	1.1%	9,605	1.5%

Table 20. Summary of convergence statistics for the number of people predicted by APEX who experience one or more 8-hour average exposures above exposure levels of 0.06, 0.07, and 0.08 ppm-8hr: 90 percent confidence intervals around the medians

Exposure level (ppm-8hr)	Children, moderate exertion	Children, any exertion	All people, moderate exertion	All people, any exertion
0.06	± 1.8%	± 1.5%	± 0.9%	± 0.4%
0.07	± 2.5%	± 2.1%	± 1.5%	± 0.8%
0.08	± 4.5%	± 3.5%	± 2.8%	± 1.5%

APEX MODEL FORMULATION UNCERTAINTY

Uncertainties are inherent in modeled representations of physical reality due to simplifying assumptions and other aspects of model formulation. The methods for assessing input parameter uncertainty and model formulation or structure uncertainty are different. It is difficult to incorporate the uncertainties due to the model formulation into a quantitative assessment of uncertainty in a straightforward manner. The preferred way to assess model formulation uncertainty is by comparing model predictions with measured values, while having fairly complete knowledge of the uncertainty due to input parameters. Whence the importance of model evaluation and the availability of data suitable to model evaluation. In the absence of measurements that can be used to estimate model uncertainty, one must rely on informed judgment.

Our approach to assessing model formulation uncertainty is to partition this uncertainty into that of the components, or algorithms, of the model. For each of the algorithms within the model, we will discuss the simplifying assumptions and those uncertainties associated with the algorithms which are distinct from the input data uncertainties. Where possible, we will evaluate these algorithms by comparing their predictions with measured data. Otherwise, we will formulate an informed judgment as to a range of plausible uncertainties for the algorithms. We will assemble the different types of uncertainties to present an integrated assessment of model uncertainty.

It should be noted that data collection efforts in the near future would best serve to reduce uncertainties by improving the inputs to the current algorithms and not to derive better algorithms. Uncertainty would be reduced significantly just by the use of better inputs. For example, APEX can model the dependence of AER distributions on hourly temperature, humidity, and wind speed, which are known to influence AERs, but data are not available to characterize these relationships. APEX has the flexibility to take advantage of much more data than are currently available.

There are several algorithms in APEX containing simplifying assumptions that have the potential to introduce uncertainty into the model, including the following:

- demographic profiles model
- longitudinal diary construction model
- collapsing the numerous microenvironments in the diaries to 12 modeled microenvironments
- modeling movements of individuals (commuting, school, shopping, etc.)
- microenvironment concentration model – factors approach
- microenvironment concentration model – mass balance approach
- modeling near-roadway titration of ozone by NO
- model for assigning physiological characteristics to individuals
- MET model
- ventilation model

The Treatment of Variability and Covariability in Apex

Assessment of the extent to which APEX correctly models variability and covariability is central to an understanding of the model uncertainty, and is summarized here.

APEX simulates individuals and then computes exposures to ozone concentrations for each of these simulated individuals. The individuals are selected to represent a random sample from a defined population. The collection of individuals represents the variability of the target population, and accounts for several types of variability, including demographic, physiological, and human behavior. Typically more than 50,000 individuals are modeled in order to capture a wide range of variability.

APEX incorporates stochastic processes representing the natural variability of personal profile characteristics, activity patterns, and microenvironment parameters. In this way, APEX is able to represent much of the variability in the exposure estimates resulting from the variability of the factors effecting human exposure. APEX is also designed to account for covariability, or linear and nonlinear correlation, among the model inputs.

APEX models variability and covariability in two ways:

- **Stochastic**. The user provides APEX with probability distributions characterizing the variability of input parameters. These are treated stochastically in the model and the computed distributions of exposures reflect this variability. For example, the rate of decay of ozone in houses depends in a complex way on several factors which we are not able to explicitly model at this time. However, we can specify a distribution of decay rates which reflects observed variations in ozone decay rates. APEX randomly samples from this distribution to obtain values which are used in the mass balance model. Covariability is modeled through the use of conditional distributions. If two or more parameters are related, conditional distributions which depend on the values of the related parameters are input to APEX. For example, the distribution of air exchange rates (AERs) in a house depends on the outdoor temperature and whether or not air conditioning (A/C) is in use. In this case, a set of AER distributions is provided to APEX for different ranges of temperatures and A/C use, and the selection of the distribution in APEX is driven by the temperature and A/C status at that time.
- **Explicit**. For some variables used in modeling exposure, APEX models variability and covariability explicitly and not stochastically. For example, hourly-average ambient ozone concentrations and temperatures are used in model calculations. These are input to the model for every hour in the time period modeled, and in this way the variability and covariability of concentrations and hourly temperatures are modeled explicitly.

Each of these methods allows for linear and nonlinear relationships between variables to be modeled. Table 21 lists the components of exposure variability which are modeled by APEX.

Table 21. Components of exposure variability modeled by APEX

Parameter	Dimensions of Variation in APEX	Treatment in APEX
Population demographics (age, gender, race, employment, residence location, work location)	Individuals, by Census tract	Random samples from Census tracts
Commuting	Individuals, by Census tract	Random samples from Census tracts
Physiology (weight)	Individuals	Distributions by age and gender
Physiology (resting metabolic rate, maximum level of sustained metabolic activity, oxygen uptake per unit of energy expended)	Individuals	See section 4.3 in the APEX TSD.
Physiology (blood volume, lung diffusivity, endogenous CO production rate, amount of hemoglobin in the blood)	Individuals	See section 4.3 in the APEX TSD. These parameters are not used for modeling ozone.
Ambient pollutant concentrations	Space and time (hourly)	Hourly values at a set of locations are input; values from the closest location are used.
Ambient meteorological data	Space and time (hourly and daily)	Hourly values at a set of locations are input; values from the closest location are used; daily values are calculated in APEX.
Spatial concentration variability within microenvironments	Microenvironment type and geographical region	This variability can be incorporated into the variability of mass balance or factors model parameters.
Spatial concentration variability within air quality districts	Microenvironment type and geographical region	This variability can be incorporated into the variability of mass balance or factors model parameters.
Within-hour concentration variability	Microenvironment type and geographical region	This variability can be incorporated into the variability of mass balance or factors model parameters.
Microenvironment	Microenvironment type	APEX can model any number of user-defined microenvironments

There are also model inputs which are not tied to the individual which contribute to the variability of the modeling results. These include spatially and temporally varying air quality concentrations and meteorological variables, as well as a number of factors involved in the calculation of indoor and in-vehicle microenvironmental concentrations. The variability of air quality and meteorological data is modeled by providing hourly average, spatially varying inputs to APEX. Variability for these inputs for time scales less than one hour can be modeled with parameters of the microenvironment model. The variability of other parameters is treated by specifying distributions for these parameters, from which APEX randomly samples values.

Correlations and non-linear relationships between variables input to the model can result in the model producing incorrect results if the inherent relationships between these variables are not preserved.

APEX has a sophisticated method for modeling linearly and non-linearly correlated input data. This is accomplished by providing inputs that enable the correlation to be modeled explicitly within APEX. For example, there are non-linear relationships between the outdoor temperature and rates of air exchange in homes (or automobiles). One factor that contributes to this is that windows tend to be closed more often when temperatures are low or high than when temperatures are moderate. This relationship is explicitly modeled in APEX by specifying different probability distributions of air exchange rates for different ambient temperatures.

Thus, the APEX formulation allows for relationships between input data to be modeled, provided that enough is known about these relationships to specify them. The degree to which these relationships are unknown contributes to the uncertainty of the results. For those relationships which APEX explicitly models the correlation, uncertainty arises from misspecification of the correlation in the model inputs.

Table 22 lists different types of covariability and how they are modeled in APEX. The center column of this table indicates whether or not APEX explicitly models this type of covariability.

Table 22. Important components of covariability

Type of Covariability	Modeled in APEX?	Treatment in APEX / Comments
Within-profile correlations [1]	Yes	Activities, physiology, microenvironments
Between-profile correlations	No	Not important
Correlations between profile variables and microenvironment parameters	Yes	Profiles are assigned microenvironment parameters
Correlations between profile variables (age, gender) and activities	Yes	Age and gender are used in activity diary selection
Correlations between activities and microenvironment parameters	No	E.g., opening windows when cooking or smoking. Might be important, but do not have data.
Correlations among microenvironment parameters in the same microenvironment	Yes	Modeled with joint conditional variables

Type of Covariability	Modeled in APEX?	Treatment in APEX / Comments
Correlations between demographic variables and air quality.	Yes	This is modeled with the spatially varying demographic variables and air quality input to APEX.
Correlations between meteorological variables and activities	Yes	Temperature is used in activity diary selection
Correlations between meteorological variables and microenvironment parameters	Yes	The distributions of microenvironment parameters can be functions of temperature
The consistency of the occupation (and time spent commuting) for an individual from one working day to the next.	No	Simulated individuals who are employed are assigned activity diaries without regard to occupation. This would be important for modeling outdoor workers.

[1] We use the term "correlation" to encompass linear and nonlinear relationships.

Errors in Coding

APEX has undergone fairly extensive testing, but has not been subjected to a rigorous, exhaustive test regime. Incorrect implementation of algorithms as documented falls into the realm of coding errors. We will not attempt to quantify the uncertainties in the model predictions that might be the result of coding errors.

Errors in Algorithms

The likelihood of errors in algorithms can be reduced by a scientific peer review of the documentation of the model algorithms. We will not attempt to quantify a likely range of uncertainties due to possible errors in algorithms. However, we present an example of such an error which resulted in increased uncertainty of our exposure modeling results.

In our review of the APEX modeling results, we uncovered an error in the algorithm for estimating ventilation rates. This algorithm (section 2.5.1, Exposure Analysis TSD) included terms for uncertainty as well as variability. Since only variability should be reflected by the algorithm, this error erroneously inflates the variability of ventilation rates, most noticeably for older adults. This error primarily affects the highest percentiles of the distributions of ventilation rates. For adults 70 years of age and older, the 99.9th percentile of the ventilation rates distribution is a factor of two too high; for children, the difference is less than 1.5% at the 99.9th percentile. Therefore, while the estimates of exertion levels are acceptable for children, they are overestimated for the general population.

Ambient Air Quality Concentrations

Ambient concentrations are not explicitly modeled by APEX; they are provided as input data. APEX is capable of using input data with highly resolved spatial and temporal resolution. Model uncertainty associated with ambient concentrations results from erroneous characterization of the levels and/or variability of concentrations in very localized areas, e.g., close to sources or sinks.

For ozone modeling, one important process that may not be adequately modeled is the effect on exposures of the decrease in ozone concentrations downwind of roadways due to titration by NO emitted by cars and trucks. APEX does simulate the decrease in ozone levels downwind of roadways, and the effect of this on exposures of people engaged in activities near roadways, but does not differentially model the affects on people in homes close to roadways (vs. homes not close to roadways). A sensitivity analysis (described above) indicates that this has a small effect on the estimated frequencies of 8-hour average exposures at levels above 0.06 ppm-8hr.

Meteorological Data

Meteorological variables are not explicitly modeled by APEX; they are provided as input data. APEX is capable of using input data with highly resolved spatial and temporal resolution, and we do not consider model structure uncertainty associated with meteorological data to be an issue.

Modeling Concentrations in Microenvironments

There are two models in APEX for calculating concentrations in microenvironments, the mass balance and the factors models (see the APEX TSD for details):

$$\frac{dC}{dt} = C_{ambient}\ x\ f_{proximity}\ x\ f_{penetration}\ x\ f_{air\ exchange}\ x\ f_{decay} \qquad \text{(mass balance model)}$$

$$C = C_{ambient}\ x\ f_{proximity}\ x\ f_{penetration} \qquad \text{(factors model)}$$

One can raise questions as to the appropriateness of the assumptions of the mass balance model in APEX for estimating concentrations in microenvironments, such as linearity assumptions and assumptions that parameters (e.g., air exchange rates, source strengths, infiltration factors, and deposition rates) can be treated as constant in time over an hour. However, of much greater importance for model uncertainty is how the inputs to the mass balance model (air exchange rates, decay rates, etc.) are modeled, so our discussion will focus on these. The factors model formulation has no model uncertainty, by definition of that model's parameters.

Air Exchange Rate

APEX models the dependence of AERs on the microenvironment characteristics and temperature, but not the behavior of the occupants, which is known to influence AERs. The analysis of the uncertainties of the AER distributions input to APEX encompasses this aspect of model uncertainty.

Deposition Processes

The rate of deposition of ozone to some materials diminishes with cumulative exposure to ozone. This is not necessarily a small effect. In one study, 60 to 90 percent more ozone was scavenged by fiberglass insulation that had not been previously exposed to ozone, than by insulation with no previous exposure (Liu and Nazaroff, 2001; CD, Appendix AX3, page 179). Data are not available to allow this to be explicitly modeled by APEX or to assess the uncertainty that may result from this process.

Chemical Reaction Processes

Ozone reacts with a number of indoor pollutants, such as NO from gas stoves and VOCs from consumer products. Titration of ozone by NO from gas stoves reduces the concentration of ozone indoors. Lee et al. (1999) find ozone concentrations dropping by a factor of five within seven minutes of a gas stove being turned on. If this process were modeled, it would have the effect of slightly reducing some people's exposures.

Ozone reacts slowly with most other indoor pollutants, and in general this is a minor removal process compared to air exchange and surface removal (Weschler, 2000). Aside from the gas stove effect, the lack of a more refined treatment of indoor air chemistry is not considered to be a limitation of APEX for modeling ozone exposures.

Characterization of Population Demographics

The population demographics are taken directly from the 2000 Census and not modeled by APEX. Therefore, although there is uncertainty in the values input to APEX, there is no model structure uncertainty associated with this characterization.

Modeling Activity Patterns

The following are population characteristics that contribute to the variability of exposures but which are not fully modeled in APEX. Of course, some of these are more important than others. Additional data collection will be required to assess the extent to which these are limitations.

- Occupational category
- Life cycle (see, e.g., Zuzanek and Smale, 1992)
- Socio-economic status and educational level
- Longitudinal stability in occupation, exercise levels, and leisure activities
- Geographical locations of activities away from the home
- The specific microenvironments visited away from home

Even though some of these may influence exposures, they will not necessarily have much effect on the population distribution of exposures. In this case, there would be no reason to model them explicitly, unless one wanted to break out results by that variable. The uncertainty of the activity data input to APEX is likely larger than the model structure uncertainty associated with these limitations.

Behavior changes in response to ozone pollution or in response to air quality index (AQI) notification ("averting behavior") is not being taken into account in our exposure modeling. Eiswerth et al. (2005) find that increased ozone levels appear to influence the amount of time that asthmatic adults spend in different activities. In a national survey, Mansfield and Corey (2003) find a significant fraction of the people surveyed modifying their activities in response to ozone alerts. We do not feel that this is a relatively influential uncertainty at this time, however, this aspect of people's activities presumably will become more important in the future.

The algorithm in APEX which sequentially (longitudinally) assigns activity diaries to simulated individuals introduces a degree of realism reflecting the ways that people tend to repeat certain activity patterns. We have performed sensitivity analyses to assess the impact of this treatment (described in the Exposure Analysis TSD), and its parameters are included in the Monte Carlo uncertainty analysis. Additional data on longitudinal activity patterns are needed to be able to evaluate this model.

The assignment of activity diaries to individuals is the primary determinant of the frequency of repeated exposures for individuals. This is an important consideration, since multiple exposures pose a greater health concern than single exposures. The new longitudinal methodology does increase the similarity of daily activities for a given simulated individual in terms of the time spent outdoors, and some simulated individuals tend to spend more time outdoors than others, compared to a more random assignment of diaries from CHAD to modeled individuals. However, repeated routine behavior from one weekday to the next is not simulated. For example, there are no simulated individuals representing children in summer camps who spend a large portion of their time outdoors, or adults with well-correlated weekday schedules.

There are not sufficient multiple diaries from single individuals in CHAD to be able to directly evaluate the implications of this shortcoming; and we performed an assessment of APEX's predictions of outdoor workers' exposures to evaluate the repeated exposure results generated by APEX. Individuals who work outdoors tend to experience the highest 8-hour exposures to ozone compared to other groups of adults, and a significant number of outdoor workers work at an elevated activity level. Thus, this is an important segment of the population to consider in an ozone exposure and risk assessment. APEX does not adequately model exposures for this group since there is not have sufficient information to properly model the populations of outdoor workers in urban areas, due to the lack of data on activity patterns and exertion levels representative of this group.

In order to investigate the uncertainty of the exposure analysis with respect to repeated exposures we calculate for two urban areas (Atlanta and Sacramento) crude estimates of the exposures to the population that work outdoors during the day and discuss the results in the context of the APEX results for the general population of working adults.

Our estimate of the number of outdoor workers in an area is based on the May 2005 estimates of employment by occupation from the Occupational Employment Statistics (OES) Survey (US Bureau of Labor Statistics, 2005), coupled with estimated ranges for the proportion of all-day outdoor workers in each employment category. Specifically, for each employment category (7-digit Standard Occupational Classification code) we assigned high and low estimates of the fraction that work outdoors 8 or more daytime hours five days per week and the fraction that work outdoors 8 or more daytime hours only three days per week. For example, the estimates for the category "Landscaping and groundskeeping workers" are:

Low estimates	High estimates
50% 5 days/week	80% 5 days/week
10% 3 days/week	10% 3 days/week

Appendix A lists the OES employment data, our low and high estimates for the proportions of all-day outdoor workers, and the resulting low and high estimates of the numbers of outdoor workers for 65 out of approximately 700 OES employment categories, for the Atlanta-Sandy Springs-Marietta, Georgia and the Sacramento--Arden-Arcade--Roseville, CA MSAs. This results in the outdoor worker estimates in Table 23.

Table 23. Estimates of 8-hour outdoor workers in the Atlanta and Sacramento MSAs

	Low Estimates		High Estimates	
	3-day workers	**5-day workers**	**3-day workers**	**5-day workers**
Atlanta MSA	22,500	40,400	57,800	90,500
Sacramento MSA	11,300	22,000	23,900	43,800

We estimate exposures for the outdoor workers by assuming that the 5-day workers are exposed to the 9:00 AM to 5:00 PM 8-hour average outdoor ozone concentrations on all weekdays in the ozone season and that the 3-day workers are exposed to random subsets of 3/5 of these days. This is not necessarily the period of the daily maximum 8-hour average outdoor concentrations of ozone, although there is usually a large overlap of these periods. Also, it is not necessarily the same period of time that an individual is working outdoors.

Table 24 summarizes the comparisons performed for Atlanta and Sacramento for repeated exposures to levels above 0.06, 0.07, and 0.08 ppm-8hr for the 2002 base cases. The APEX numbers are for all workers, of which outdoor workers is a subset.

Table 24. Comparison of estimated outdoor workers' repeated exposures with APEX results for all workers, in Atlanta and Sacramento, 2002. Numbers of people with at least six repeated 8-hour exposures above 0.06, 0.07, and 0.08 ppm-8hr.[1]

	# above 0.06 ppm-8hr		# above 0.07 ppm-8hr		# above 0.08 ppm-8hr	
	Est. outdoor workers	APEX all workers	Est. outdoor workers	APEX all workers	Est. outdoor workers	APEX all workers
Atlanta	63,000 – 150,000	74,000	62,000 – 140,000	220	41,000 – 94,000	0
Sacramento	30,000 – 61,000	30,000	27,000 – 55,000	95	21,000 – 42,000	0

[1] The numbers in this table have been rounded to two significant digits.

Comparison of estimates of repeated exposures to outdoor workers with the corresponding APEX estimates for all workers reveals that APEX significantly underestimates the number of multiple exposures for a large subgroup of working adults. As discussed above, this underestimation results primarily from the way that people's activities are modeled, which does not properly account for repeated behavior of individuals.

Modeling Physiological Processes

Overview of the Physiological Model

The model in APEX of physiological processes that are relevant to inhalation exposure and dose is significantly improved over earlier (pre-2005) versions of APEX. APEX currently has a physiological model for ventilation rates (the primary driver of dose of ozone) which accounts for prior energy expenditure patterns (also known as oxygen debt [fatigue] and excess post-exercise oxygen consumption [EPOC]), described in Appendix B of the Exposure Analysis TSD. The physiological calculations do not directly affect APEX's estimation of exposures; rather, they are used to characterize the population according to exertion levels and "active" or not. These are important for exposure-based estimates of risk.

The physiological model produces two quantities which are used in this exposure assessment, an effective ventilation rate (EVR), which is used to characterize levels of exertion in compiling summary exposure tables, and a physical activity index (PAI_{med}), used to characterize simulated individuals as "active."

One of the key variables in this model is the "MET" (metabolic equivalent, Ainsworth, 2002), defined as the ratio of the metabolic rate of energy consumption for an activity to the resting metabolic rate (RMR). For each simulated individual, APEX generates an event-by-event time-series of MET values based on activity-specific MET distributions (some MET distributions are occupation/age dependent as well). Events are specified in the diaries, and last from 1 to 60 minutes. This time series of MET values is then adjusted for fatigue and excess post-exercise oxygen consumption (MET_{adj}). Then the oxygen consumption rate, VO_2, is calculated for each event as $VO_2 = MET_{adj} \cdot RMR \cdot ECF$, where ECF is a person-specific energy conversion factor and RMR is the person-specific resting (basal) metabolic rate. The expired ventilation rate V_E, is calculated by a stochastic model using VO_2, body mass, age, and gender. The effective ventilation rate EVR = V_E / BSA is averaged over one and eight hours, and used to characterize average levels of exertion. Body surface area (BSA) is currently modeled as a simple deterministic function of body mass (BM), and there is some uncertainty in the regression equation parameters.

Thus, EVR is a complex function of the activity-specific MET and person-specific RMR, ECF, BSA, and BM, which vary with age and gender. The person-specific parameters are modeled in such a way as to realistically reflect variability in the populations. For example, different 36-year old males will have different physiological parameters reflective of the variation observed across the population of 36-year old males.

Once the final MET time series is for a person calculated, a daily average physical activity index (PAI) for the simulated individual is calculated as the time-weighted average of MET values for each day. The median of the daily PAI values is calculated for each profile. This median daily PAI value (PAI_{med}) is used in the characterization of persons as "active."

Uncertainty of the Physiological Model

<u>The Resting Metabolic Rate Model</u>

The RMRs for individuals are estimated by a regression equation with coefficients specific to age and gender, which were developed by Schofield (1985). (See Johnson, 2003, Table 9-11.) Since then, studies have improved on this model. For example, Huang et al. (2004) find that the best predictive equation for RMR for obese adults include terms for age, gender, weight, height, and diabetes. RMRs are used in APEX to derive ventilation rates, which we are able to evaluate against values in the literature. Therefore, we are not investigating the uncertainties of RMRs in APEX at this time.

<u>The Ventilation Rate Model</u>

We evaluated the APEX algorithm for calculating ventilation rates by comparing them with values in the literature. Detailed distributions of measured ventilation rates are reported by Brochu. et al. (2006) for normal-weight individuals (age 3 to 19 years with BMI less than the 85th percentile by age; ages 20 to 96 years with BMI from 18.5 to 25 kg/m^2). Table 25 compares the mean daily ventilation rates by gender and age group reported by Brochu et al. with the corresponding values from the APEX Boston simulations. The APEX ventilation rates are significantly higher (by 2 or more) for ages less than 7 and greater than 39, and in fairly good agreement with the measured values for ages 7 to 39.

Table 25. Comparison of measured ventilation rate distributions with APEX modeled values (m^3/day)

Age group (years)	Measured V_E (males)	APEX V_E (males)	Measured V_E (females)	APEX V_E (females)
2 to < 5	7.6 ±1.3	10.7 ±2.2	7.1 ±1.2	10.4 ±2.1
5 to < 7	8.6 ±1.2	11.2 ±2.1	8.2 ±1.3	10.4 ±2.0
7 to < 11	10.6 ±2.0	12.1 ±2.6	9.8 ±1.7	11.7 ±2.5
11 to < 23	17.2 ±3.7	15.1 ±4.5	13.3 ±2.6	13.2 ±3.6
23 to < 30	17.5 ±2.8	16.5 ±5.2	13.7 ±2.3	13.3 ±3.9
30 to < 40	16.9 ±2.5	16.9 ±5.2	13.7 ±1.8	13.9 ±3.7
40 to < 65	16.2 ±2.7	18.2 ±5.2	12.3 ±2.1	14.4 ±3.6
65 to ≤ 96	13.0 ±2.5	16.3 ±4.6	9.8 ±2.2	12.2 ±3.1

[1] Measured values from Table 2, Brochu, et al. (2006), rounded to 1 decimal place.

<u>Classification of Individuals as Active</u>

This is an area where a great deal of research is being done for both adults and children. Duke et al. (2003) summarize nationally representative information about levels and types of physical activity among children aged 9–13 years. Puyau et al. (2002) show that adult-derived MET cutpoints are not applicable to children and can lead to erroneous conclusions regarding physical activity levels in children.

APEX uses the PAI cutpoint to classify individuals as active or not active. There are two shortcomings of this method. First, it is not clear what the relevant classification of specific activities is in terms of levels of physical exertion. Second, given such a classification of activities, it is not clear how to best characterize a given individual as “active” or not.

There are various “definitions” or interpretations of how to classify levels of exertion in the literature. The Centers for Disease Control and Prevention (CDC) and the American College of Sports Medicine categorize physical activity levels in adults as *light*: < 3 MET, *moderate*: 3 to 6 MET, and *vigorous*: >6 MET. Reland et al. (2004) use low activity: < 4,185 kJ/week (1,000 kcal/week) high: > 8,370 (2,000) moderate: in between, where 1 MET = 4.185 kJ kg^{-1} h^{-1}. Marty et al. (2002) categorize activity levels according to ventilation rates (l/min per kg body weight), with different classifications for ages > 12 and ≤ 12 years. McCurdy and Graham (2004) present a survey of the exercise physiology literature of different measures used to define moderate and vigorous physical activity, and find many different ways that researchers are categorizing activity levels.

There is less research in the area of characterizing individuals (as opposed to activities) in terms of how active they are, particularly in the context of how it influences their health risk from exposure to ozone.

Sapkota et al. in their 2005 study on adult participation in recommended levels of physical activity, use a definition of "regular physical activity" given by CDC[2]. Moderate-intensity activity is described to respondents as any activity "that causes small increases in breathing and heart rate," and vigorous-intensity activity is described as any activity "that causes large increases in breathing or heart rate." Respondents are classified as active at the minimum recommended level if they report moderate-intensity activity at least 30 minutes per day, 5 or more days per week, or vigorous-intensity activity at least 20 minutes per day, 3 or more days per week. The Behavioral Risk Factor Surveillance System (BRFSS) survey for 2003 reports 46% of U.S. adults to be active by this definition (Sapkota et al., 2005).

APEX characterizes a person as active if their median daily average MET is greater than 1.75. One factor that contributes to the uncertainty of this model is the fact that the daily average MET is largely influenced by the number of hours spent sleeping, which is not correlated with most definitions of active people. A better characterization might be the daily maximum 12-hour average MET, which would reflect levels of activity while not sleeping. In order to characterize the uncertainty associated with the estimation of exposures to an "active" population, we can use the CDC definition of "regular physical activity" for the definition of an active person.[3]

Unknown Model Uncertainty

There are structural uncertainties of APEX of which we are currently unaware. We will attempt to characterize their uncertainties as they come to light. We are proceeding on two fronts to uncover additional uncertainties: peer review and diagnostic model evaluation.

Perhaps there is a correlation between the tendency of people to open windows and their tendency to engage in outdoor activities. If so, people who spend more time outdoors would tend to have higher AERs at home and work, and ignoring this could bias the modeled distribution of exposures. This is conjecture, but is an example of potential unknown model uncertainty.

---

2 http://www.cdc.gov/nccdphp/dnpa/physical/terms/index.htm

3 Analysis of uncertainty of the exposure modeling results uncovered an error in how children are characterized as active, which resulted in an overestimate of the number of active children in the population. Thus, an evaluation of the uncertainty associated with the characterization of children as active is not included at this time.

UNCERTAINTY ANALYSIS RESULTS

As discussed above, a Monte Carlo approach was selected for a detailed uncertainty analyses. Monte Carlo methods for analysis of model uncertainty use statistical sampling techniques to estimate statistics which characterize uncertainty. Essentially, a Monte Carlo approach involves performing many model runs with model inputs randomly sampled from distributions reflecting the uncertainty of the inputs. This propagates the uncertainty of the model inputs through to the model results, taking into account input parameter dependencies and the interaction of uncertainties within the model. These simulations provide uncertainties of model results in terms of uncertainty distributions of the model outputs. From these we are able to calculate 95 percent uncertainty intervals (UI) for a particular model result as the interval from the 2.5th to the 97.5th percentile of the uncertainty distribution for that result.

The Monte Carlo uncertainty analysis performed accounts for the following sources of uncertainty (described above in the section Quantifying the Uncertainty of Apex Model Inputs):

- Ambient air concentrations measurement error
- Spatial interpolation of ambient concentrations
- Air exchange rates
- Air conditioning prevalence rates
- Ozone deposition and decay rates
- Vehicle penetration factors
- Longitudinal diary assembly parameters
- Metabolic equivalents (MET)
- Model convergence

The Monte Carlo uncertainty analysis was performed for Boston 2002, based on approximately 2000 APEX simulations for each of the base case and the current standard scenarios. Each pair of simulations uses the same uncertain inputs and differs only by the air quality concentrations input to the model, so that we can assess the uncertainty of estimates of reductions in exposures in going from the base case to the current standard as well as the uncertainty of the estimates of exposures. Uncertainties of model results for other areas and years are expected to be similar.

Figure 16 illustrates the uncertainty distributions for one model result, the percent of children with exposures above 0.06 ppm-8hr while at moderate exertion. This distribution results from approximately 2000 Monte Carlo APEX simulations of the Boston 2002 base case with model inputs varied randomly according to their uncertainty. The "point estimate" of 62 percent is the result from the APEX simulation using our best estimates of the model inputs, as described in Section 4.5 of the Staff Paper. The corresponding result from the Monte Carlo simulations ranges from 56 to 67 percent, with a 95 percent UI of 58 to 65 percent. Figure 17 and Figure 18 illustrate the uncertainty distributions for two other model results, the percents of children with exposures above 0.07 and 0.08 ppm-8hr while at moderate exertion.

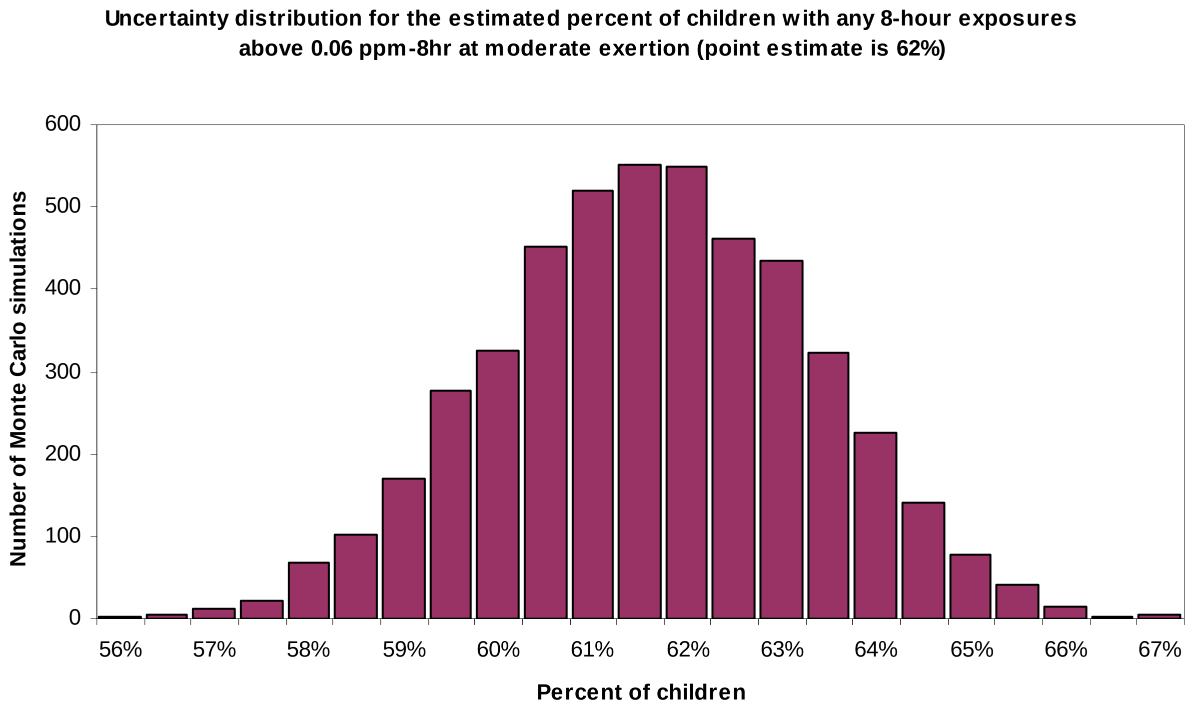

Figure 16. Uncertainty of percent of children with exposures above 0.06 ppm-8hr (Boston 2002 base case)

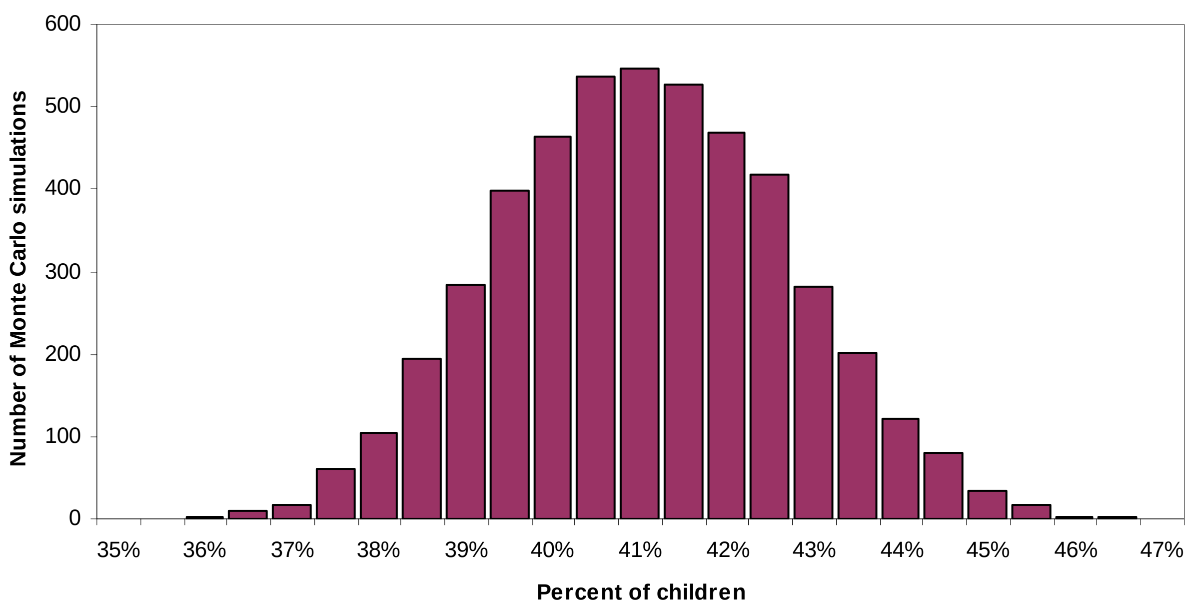

Figure 17. Uncertainty of percent of children with exposures above 0.07 ppm-8hr (Boston 2002 base case)

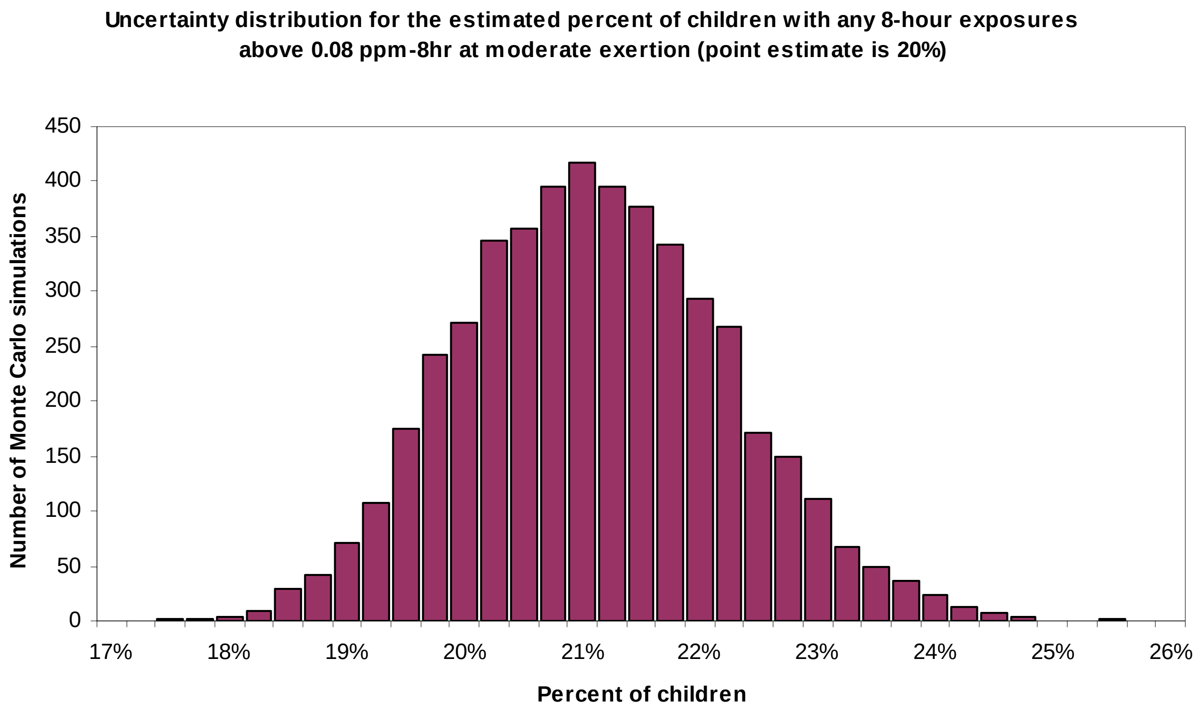

Figure 18. Uncertainty of percent of children with exposures above 0.08 ppm-8hr (Boston 2002 base case)

Uncertainty intervals are presented in Table 26 and Table 27 for the estimated percentages of all children and asthmatic children with exposures above different 8-hour exposure levels under moderate exertion. The UIs for the estimated reductions in exposures, going from the 2002 base case to the current standard, for these two groups are given in Table 28. Across these three tables, the spans of the 95 percent UIs range from 2 to 10 percentage points, and the point estimates are generally within 5 percentage points of the UI endpoints. The uncertainties of the exposures to asthmatic children are slightly higher than for all children. These results are very positive, and the modeling uncertainty is small enough to lend confidence to the use of the model results.

Table 26. Uncertainty of the estimated percent of children exposed at moderate exertion, Boston, 2002

Exposure level (ppm-8hr)	Air quality scenario	Point estimate	95% UI
0.06	base case	62%	58-65%
0.07	base case	41%	38-44%
0.08	base case	20%	19-24%
0.06	current standard	49%	46-52%
0.07	current standard	24%	23-27%
0.08	current standard	8.5%	8-10%

Table 27. Uncertainty of the estimated percent of asthmatic children exposed at moderate exertion, Boston, 2002

Exposure level (ppm-8hr)	Air quality scenario	Point estimate	95% UI
0.06	base case	65%	60-67%
0.07	base case	43%	39-46%
0.08	base case	21%	19-25%
0.06	current standard	52%	48-56%
0.07	current standard	24%	23-30%
0.08	current standard	9%	8-11%

Table 28. Uncertainty of the estimated percent reduction, from the base case to the current standard, of all children and asthmatic children exposed at moderate exertion, Boston, 2002

	All children		Asthmatic children	
Exposure level (ppm-8hr)	Point estimate	95% UI	Point estimate	95% UI
0.06	21%	18-22%	19%	16-22%
0.07	41%	38-42%	43%	37-45%
0.08	58%	55-59%	58%	53-63%

Key Findings

Uncertainty of the APEX model predictions results from uncertainties in the spatial interpolation of measured concentrations, the microenvironment models and parameters, people's activity patterns, and, to a lesser extent, model structure. The predominant sources of uncertainty appear to be the activity pattern information and the spatial interpolation of ambient concentrations from monitoring sites to other locations. The primary findings of these analyses are the following:

- The Monte Carlo analysis of the uncertainties of the APEX model estimates of exposure distributions indicates that the uncertainty is relatively small. The APEX estimates of the percent of children or asthmatic children with exposures above 0.06, 0.07, or 0.08 ppm-8hr under moderate exertion have 95% uncertainty intervals of at most ±6 percentage points.
- The non-representativeness of the CHAD activity diaries with respect to the specific urban areas and time periods modeled indicates uncertainties of only a few percent in the APEX estimates of the numbers of children with exposures above 0.06, 0.07, or 0.08 ppm-8hr under moderate exertion.
- The effect on exposures in residences of the titration of ozone by mobile source NO is small, on the order of 1 to 3 percent.
- APEX significantly underestimates the frequency of occurrence of individuals experiencing repeated 8-hour average exposures greater than 0.06 ppm-8hr. The reasons for this are understood, and further research will be required to address this.

In the future, we expect to have better data for characterizing personal exposure and dose to ozone and other pollutants and integrating these with controlled human exposure health studies and with epidemiological analyses. Important research needs to reduce uncertainties associated with the current ozone exposure analysis include conducting studies to provide better information for refining methods for assessing exposure to ozone as well as other pollutants. E.g., activity diaries for sensitive groups; distributions of short-term ozone concentrations near roadways and inside homes as functions of influential covariates. There is also a need for personal exposure monitors with shorter averaging times and lower detection limits.

The activity diary data base CHAD could be updated to include recent studies, such as the second phase of the Panel Study of Income Dynamics Child Development Supplement (CDS, 2005). This is a longitudinal study of a representative sample of U.S. individuals and families which collected time diary data for almost 3,000 children and adolescents aged 5-18 years. Characterization of repetitive activity patterns is particularly important.

The most pressing need at this time is for evaluation of existing exposure models and evaluation of the specific algorithms which make up these models. This would greatly improve our understanding of how well current models perform and aid in prioritizing future data collection and model development efforts.

REFERENCES

Ainsworth, B. (2003). The Compendium of Physical Activities. President's Council on Physical Fitness and Sports Research Digest. 4(2), 1-8.

Ainsworth, B., Haskell, W., Whitt, M., Irwin, M., Swartz, A., Strath, S., O'Brien, W., Bassett, D. Jr, Schmitz, K., Emplaincourt, P., Jacobs, D. Jr, Leon, A. (2000). Compendium of Physical Activities: An update of activity codes and MET intensities. Medicine and Science in Sports and Exercise, 2000;32 (Suppl):S498-S516.

Ainsworth, B., Haskell, W., Leon, A., Jacobs, D. Jr., Montoye, H., Sallis, J., Paffenbarger, R., Jr. (1993). Compendium of physical activities: classification of energy costs of human physical activities. Med. Sci. Sports Exer. 25: 71-80.

Battelle (2004). 2003 Criteria Pollutant Quality Indicator Summary Report For July 14, 2004, AQS Data. Final Technical Report. September 9, 2004. Prepared For U.S.Environmental Protection Agency, Research Triangle Park, North Carolina.

Beckerman, B., Brook, J., Verma, D., Arain, A., Finkelstein, M., Jerrett, M. (2006). Correlation of nitrogen dioxide to other traffic pollutants near a major expressway. Journal of Exposure Science and Environmental Epidemiology, to appear.

Brochu, P., J-F. Ducré-Robitaille and J. Brodeur (2006). Physiological daily inhalation rates for free-living individuals aged 2.6 months to 96 years based on doubly labeled water measurements: comparison with time-activity-ventilation and metabolic energy conversion estimates. International Journal of Human and Ecological Risk Assessment 12(4):736-761.

CDS (2005). The Panel Study of Income Dynamics Child Development Supplement: User Guide for CDS-II. The Institute for Social Research, Ann Arbor, MI. Available at: http://psidonline.isr.umich.edu/CDS

Centers for Disease Control and Prevention (CDC) 2005. National Health and Nutrition Examination Survey Data 1999-2004 (NHANES). National Center for Health Statistics , U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. June, 2005. http://www.cdc.gov/nchs/nhanes.htm

Chan, A.T. and Chung, M.W., 2003. Indoor–outdoor air quality relationships in vehicle: effect of driving environment and ventilation modes. Atmospheric Environment 37: 3795–3808.

Chan, C.C., Ozkaynak, H., Spengler, J.D., Sheldon, L., 1991. Driver exposure to volatile organic compounds, CO, ozone, and NO_2 under different driving conditions. Environmental Science and Technology 25 (5), 964–972.

Cullen, A.C. and Frey, H.C. (1999). Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Springer.

Duke, J., Huhman, M., Heitzler, C. (2003). Physical Activity Levels Among Children Aged 9–13 Years - United States, 2002. Morbidity and Mortality Weekly Report (MMWR) Weekly August 22, 2003, 52(33). Centers for Disease Control and Prevention (CDC).

Efron, Bradley (1980). The Jackknife, The Bootstrap, and Other Resampling Plans. Technical Report No. 163, Department of Statistics, Stanford University, Stanford, California.

Eiswerth, M., Shaw, D., and Yen, S. (2005). Impacts of ozone on the activities of asthmatics: Revisiting the data. J. Env. Mgt. 77, 1:56-63.

Graham, S. and McCurdy, T. (2005). "Revised Ventilation Rate (V_E) Equations for Use in Inhalation-Oriented Exposure Models, A NERL Internal Research Report," NERL, EPA. EPA/600/X-05/008, February 2005.

Graham, S. and McCurdy, T. (2004). Developing meaningful cohorts for human exposure models. J. Exp. Anal. and Env. Epi., 14, 23-43.

Huang, K., Kormas, N., Steinbeck, K., Loughnan, G., Caterson, I. (2004). Resting metabolic rate in severely obese diabetic and nondiabetic subjects. Obes Res. May 2004; 12(5):840-5.

Isaacs, K. and Smith, L. (2005). New Values for Physiological Parameters for the Exposure Model Input File 'Physiology.txt.' Technical memorandum to Tom McCurdy, EPA, December 20, 2005.

Johnson, T. (2003). A Guide to Selected Algorithms, Distributions, and Databases Used in Exposure Models Developed by the Office of Air Quality Planning and Standards. Revised Draft. Prepared for U.S. Environmental Protection Agency under EPA Grant No. CR827033. June 30, 2003.

Johnson, T. (1997). A pilot study in Los Angeles to measure personal ozone exposures during scripted activities. Washington, DC: American Petroleum Institute, Health and Environmental Sciences Department; API publication no. DR 218.

Kroenke et al., 2005. Depressive Symptoms and Prospective Incidence of Colorectal Cancer in Women. Am J Epidemiol. 162: 839-848.

Lee, Vallarino, Dumyahn, Ozkaynak, and Spengler (1999). Ozone decay rates in residences. JAWMA. 49: 1238-1244.

Lin, T-Y., Young, L-H., and Wang, C-S. (2001). Spatial variations of ground-level ozone concentrations in areas of different scales. Atmos. Environ. 35: 5799-5807.

Mansfield, C. and Corey, C. (2003). Analysis of Survey Data on Ozone Alert Days, Final Report, September 2003. Prepared for the U.S. Environmental Protection Agency by RTI International, Research Triangle Park, NC.

Marty, M.A., Blaisdell, R.J., Broadwin, R., Hill, M., Shimer, D., Jenkins, M. (2002). Distribution of Daily Breathing Rates For Use in California's Air Toxics Hot Spots Program Risk Assessments. Human and Ecological Risk Assessment 8(7):1723-1737.

McCurdy, T. and Graham, S. (2004). Analyses to Understand Relationships Among Physiological Parameters in Children and Adolescents Aged 6-16. A NERL Internal Research Report, Office of Research and Development, U.S. Environmental Protection Agency, July 2004. Available on request from Tom McCurdy.

OMB (2005). U.S. OMB Bulletin No. 05-02. Subject: Update of Statistical Area Definitions and Guidance on Their Uses. February 22, 2005.

Puyau, M., Adolph, A., Vohra, F., Butte, N. (2002). Validation and calibration of physical activity monitors in children. Obes Res. Mar 2002; 10(3):150-7.

Reland, S., Ville, N., Wong, S., Senhadji, L. and Carre, F. (2004). Does the level of chronic physical activity alter heart rate variability in healthy older women? Clinical Science (2004) 107, 29–35.

Riediker, M., Williams, R., Devlin, R., Griggs, T., Bromberg, P. (2003). Exposure to Particulate Matter, Volatile Organic Compounds, and Other Air Pollutants Inside Patrol Cars. Environ. Sci. Technol. 37:2084-2093.

Rodes, C., Holland, D. (1981). Variations of NO, NO_x, and O_3 concentrations downwind of a Los Angeles freeway. Atmos Environ 1981;15:242-50.

Sapkota, S., Bowles, H., Ham, S., Kohl, H. (2005). Adult Participation in Recommended Levels of Physical Activity - United States, 2001 and 2003. Morbidity and Mortality Weekly Report (MMWR) Weekly, December 2, 2005, 54(47):1208-1212. Centers for Disease Control and Prevention (CDC).

Schofield, W. N. (1985). Predicting basal metabolic rate, new standards and review of previous work. Clin Nutr. 1985; 39C(suppl): 5–41.

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. J. Royal Statistical Society, Series B, 36, pp. 111-147.

Suppan, P., Schadler, G. (2004). The impact of highway emissions on ozone and nitrogen oxide levels during specific meteorological conditions. Science of the Total Environment 334-335, 215-222.

Takarangi, Melanie, Maryanne Garry and Elizabeth F. Loftus (2006) Dear Diary, Is Plastic Better Than Paper? I Can't Remember: Comment on Green, Rafaeli, Bolger, Shrout, and Reis (2006), Psychological Methods, Volume 11, Issue 1, March 2006, Pages 119-122.

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.

U.S. Bureau of Labor Statistics, 2005. The May 2005 Occupational Employment and Wage Estimates, Metropolitan Area Cross-Industry Estimates. U.S. Department of Labor, Bureau of Labor Statistics. Available at: http://stat.bls.gov/oes/oes_dl.htm#2005_m .

U.S. Census Bureau (2002). Current Housing Reports, Series H150/01, American Housing Survey for the United States: 2001. U.S. Government Printing Office, Washington, DC, October, 2002. Available at: http://www.census.gov/hhes/www/housing/ahs/ahs.html.

U.S. Census Bureau (2004). Current Housing Reports, Series H150/03, American Housing Survey for the United States: 2003. U.S. Government Printing Office, Washington, DC, September, 2004. Available at: http://www.census.gov/hhes/www/housing/ahs/ahs.html.

U.S. Census Bureau (2005). Subcounty Population Estimates, April 1, 2000 to July 1, 2004, Population Estimates Program, U.S. Bureau of the Census Release dated June 30, 2005.

U.S. Environmental Protection Agency (1998). Guideline on Ozone Monitoring Site Selection. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC.

U.S. Environmental Protection Agency (2002). Consolidated Human Activities Database (CHAD) Users Guide. The database and documentation are available at: http://www.epa.gov/chadnet1/.

U.S. Environmental Protection Agency (2005a). Summary Report of the Completeness, Precision and Bias Assessments for Criteria Pollutants, Calendar Year 2004. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. Available at: http://www.epa.gov/ttn/amtic/parslist.html

U.S. Environmental Protection Agency (2005b). 2004 Single Point Precision and Bias Graphics For Criteria Pollutants Companion Document. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. Available at: http://www.epa.gov/ttn/amtic/parslist.html

U.S. Environmental Protection Agency (2006a). Air Quality Criteria for Ozone and Related Photochemical Oxidants (Final). National Center for Environmental Assessment, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA/600/R-05/004aF-cF. Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=149923

U.S. Environmental Protection Agency (2006b). Total Risk Integrated Methodology (TRIM) - Air Pollutants Exposure Model Documentation (TRIM.Expo / APEX, Version 4) Volume I: User's Guide. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. June 2006. Available at: http://www.epa.gov/ttn/fera/human_apex.html

U.S. Environmental Protection Agency (2006c). Total Risk Integrated Methodology (TRIM) - Air Pollutants Exposure Model Documentation (TRIM.Expo / APEX, Version 4) Volume II: Technical Support Document. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. June 2006. Available at: http://www.epa.gov/ttn/fera/human_apex.html

U.S. Environmental Protection Agency (2007a). Review of National Ambient Air Quality Standards for Ozone: Assessment of Scientific and Technical Information (OAQPS Staff Paper). EPA-452/D-05-001. Office of Air Quality Planning and Standards, Research Triangle Park, NC. Available at: http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_cr_sp.html

U.S. Environmental Protection Agency (2007b). Ozone Population Exposure Analysis for Selected Urban Areas. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. Available at: http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_cr_td.html

Väkevä, M.; Hämeri, K.; Kulmala, M.; Lahdes, R.; Ruuskanen, J.; Laitinen, T. (1999). Street level versus rooftop concentrations of submicron aerosol particles and gaseous pollutants in an urban street canyon. Atmos. Environ. 33: 1385-1397.

Vose, David (1996). *Quantitative Risk Analysis: A Guide to Monte Carlo Simulation*. John Wiley & Sons Ltd., West Sussex, England.

Weschler, C. J. (2000) Ozone in indoor environments: concentration and chemistry. Indoor Air 10: 269-288.

Zuzanek, J. and Smale, B. (1992). Life-cycle variations in across-the-week allocation of time to selected daily activities. Society Leisure 15:559-586.

APPENDIX A

Occupational Employment Statistics Survey employment data for the Atlanta and Sacramento MSAs and low and high estimates for the fractions of all-day outdoor workers in each employment category

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				11-0000	**Management occupations**	156,630	42,220
				13-0000	**Business and financial operations occupations**	112,510	55,060
				15-0000	**Computer and mathematical occupations**	75,960	24,970
				17-0000	**Architecture and engineering occupations**	37,350	18,960
				19-0000	**Life, physical, and social science occupations**	15,110	9,740
				21-0000	**Community and social services occupations**	19,550	13,500
				23-0000	**Legal occupations**	18,260	7,720
				25-0000	**Education, training, and library occupations**	130,850	58,250
				27-0000	**Arts, design, entertainment, sports, and media occupations**	23,490	10,870
				29-0000	**Healthcare practitioners and technical occupations**	90,910	36,120
				31-0000	**Healthcare support occupations**	37,370	17,980
				33-0000	**Protective service occupations**	49,040	22,030
				35-0000	**Food preparation and serving related occupations**	183,860	67,980
				37-0000	**Building and grounds cleaning and maintenance occupations**		
				37-1011	First-line supervisors/managers of housekeeping and janitorial workers	3,470	900
0.00	0.00	0.30	0.50	37-1012	First-line supervisors/managers of landscaping, lawn service, and groundskeeping workers	3,090	750
				37-2011	Janitors and cleaners, except maids and housekeeping cleaners	28,170	10,210
				37-2012	Maids and housekeeping cleaners	12,160	4,530
0.05	0.05	0.10	0.20	37-2021	Pest control workers	1,620	690
0.10	0.50	0.10	0.80	37-3011	Landscaping and groundskeeping workers	15,020	8,070
0.10	0.50	0.10	0.80	37-3012	Pesticide handlers, sprayers, and applicators, vegetation	270	70
0.10	0.70	0.10	0.80	37-3013	Tree trimmers and pruners	340	70
0.10	0.50	0.10	0.80	37-3019	Grounds maintenance workers, all other	0	0
				39-0000	Personal care and service occupations	52,760	18,690
				41-0000	**Sales and related occupations**		
				41-1011	First-line supervisors/managers of retail sales workers	20,450	6,710
				41-1012	First-line supervisors/managers of non-retail sales workers	5,780	1,530
				41-2011	Cashiers	54,960	21,130
				41-2021	Counter and rental clerks	6,430	4,370
				41-2022	Parts salespersons	3,330	1,550
				41-2031	Retail salespersons	75,950	27,370
				41-3011	Advertising sales agents	3,130	850
				41-3021	Insurance sales agents	4,600	1,820
				41-3031	Securities, commodities, and financial services sales agents	3,170	2,490
				41-3041	Travel agents	1,950	340
				41-3099	Sales representatives, services, all other	6,780	5,280
				41-4011	Sales representatives, wholesale and manufacturing, technical and scientific products	12,110	2,060
				41-4012	Sales representatives, wholesale and manufacturing, except technical and scientific products	33,520	6,960

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				41-9011	Demonstrators and product promoters	2,420	450
				41-9021	Real estate brokers	1,050	30
				41-9022	Real estate sales agents	4,420	330
				41-9031	Sales engineers	1,720	860
				41-9041	Telemarketers	6,730	800
				41-9091	Door-to-door sales workers, news and street vendors, and related workers		
				41-9091A	Door-to-door sales workers (est 50% of 41-9091)	210	1,255
0.20	0.60	0.00	1.00	41-9091B	news and street vendors (est 50% of 41-9091)	210	1,255
				41-9099	Sales and related workers, all other	3,380	1,310
				43-0000	**Office and administrative support occupations**		
				43-1011	First-line supervisors/managers of office and administrative support workers	32,380	15,290
				43-2011	Switchboard operators, including answering service	3,980	1,110
				43-2021	Telephone operators	0	0
				43-2099	Communications equipment operators, all other	50	30
				43-3011	Bill and account collectors	13,220	2,470
				43-3021	Billing and posting clerks and machine operators	8,710	2,710
				43-3031	Bookkeeping, accounting, and auditing clerks	27,690	12,530
				43-3051	Payroll and timekeeping clerks	3,410	1,640
				43-3061	Procurement clerks	1,110	340
				43-3071	Tellers	9,190	3,530
				43-4011	Brokerage clerks	1,380	490
				43-4021	Correspondence clerks	540	50
				43-4031	Court, municipal, and license clerks	1,770	0
				43-4041	Credit authorizers, checkers, and clerks	0	330
				43-4051	Customer service representatives	53,340	12,650
				43-4061	Eligibility interviewers, government programs	40	0
				43-4071	File clerks	3,000	1,660
				43-4081	Hotel, motel, and resort desk clerks	3,770	920
				43-4111	Interviewers, except eligibility and loan	2,130	2,040
				43-4121	Library assistants, clerical	1,230	960
				43-4131	Loan interviewers and clerks	0	2,930
				43-4141	New accounts clerks	3,050	680
				43-4151	Order clerks	4,260	0
				43-4161	Human resources assistants, except payroll and timekeeping	0	2,470
				43-4171	Receptionists and information clerks	17,590	7,900
				43-4199	All other information and record clerks	6,530	1,420
				43-5011	Cargo and freight agents	1,560	0
				43-5021	Couriers and messengers	1,410	710
				43-5031	Police, fire, and ambulance dispatchers	1,600	0
				43-5032	Dispatchers, except police, fire, and ambulance	4,110	920

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
0.10	0.50	0.20	0.80	43-5041	Meter readers, utilities	930	410
				43-5051	Postal service clerks	1,480	670
0.10	0.40	0.20	0.80	43-5052	Postal service mail carriers	5,660	2,320
				43-5053	Postal service mail sorters, processors, and processing machine operators	4,200	1,290
				43-5061	Production, planning, and expediting clerks	7,090	1,830
				43-5071	Shipping, receiving, and traffic clerks	12,150	5,470
				43-5081	Stock clerks and order fillers	34,340	10,710
				43-5111	Weighers, measurers, checkers, and samplers, recordkeeping	1,340	640
				43-6011	Executive secretaries and administrative assistants	30,520	11,480
				43-6012	Legal secretaries	0	2,460
				43-6013	Medical secretaries	0	2,060
				43-6014	Secretaries, except legal, medical, and executive	27,980	6,690
				43-9011	Computer operators	2,660	1,590
				43-9021	Data entry keyers	6,760	2,450
				43-9022	Word processors and typists	2,660	1,980
				43-9031	Desktop publishers	420	110
				43-9041	Insurance claims and policy processing clerks	4,120	4,080
				43-9051	Mail clerks and mail machine operators, except postal service	3,080	1,080
				43-9061	Office clerks, general	39,740	33,990
				43-9071	Office machine operators, except computer	1,050	520
				43-9081	Proofreaders and copy markers	0	30
				43-9111	Statistical assistants	1,800	110
				43-9199	Office and administrative support workers, all other	4,160	0
				45-0000	**Farming, fishing, and forestry occupations**		
0.10	0.10	0.20	0.20	45-1011	First-line supervisors/managers of farming, fishing, and forestry workers	100	3,620
0.10	0.10	0.20	0.20	45-2011	Agricultural inspectors	130	190
0.10	0.10	0.20	0.20	45-2041	Graders and sorters, agricultural products	320	100
0.10	0.50	0.20	0.80	45-2091	Agricultural equipment operators	0	320
0.10	0.50	0.20	0.80	45-2092	Farmworkers and laborers, crop, nursery, and greenhouse	0	2,090
0.10	0.50	0.20	0.80	45-2093	Farmworkers, farm and ranch animals	350	160
0.10	0.40	0.20	0.80	45-2099	Agricultural workers, all other	0	240
0.10	0.40	0.20	0.80	45-4011	Forest and conservation workers	0	190
				45-4021	Fallers	0	40
0.10	0.40	0.20	0.80	45-4022	Logging equipment operators	260	60
				47-0000	**Construction and extraction occupations**		
0.10	0.10	0.20	0.20	47-1011	First-line supervisors/managers of construction trades and extraction workers	11,230	4,970
				47-2011	Boilermakers	680	
0.40	0.10	0.30	0.60	47-2021	Brickmasons and blockmasons	810	390
0.40	0.10	0.30	0.60	47-2022	Stonemasons	0	100
0.20	0.25	0.30	0.40	47-2031	Carpenters	10,120	11,390

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				47-2041	Carpet installers	0	1,050
				47-2044	Tile and marble setters	0	1,320
0.00	0.30	0.30	0.60	47-2051	Cement masons and concrete finishers	3,420	2,840
0.20	0.10	0.30	0.60	47-2053	Terrazzo workers and finishers	0	0
0.10	0.60	0.10	0.90	47-2061	Construction laborers	18,270	8,950
0.10	0.60	0.10	0.90	47-2071	Paving, surfacing, and tamping equipment operators	1,230	240
0.10	0.60	0.10	0.90	47-2072	Pile-driver operators	0	40
0.10	0.60	0.10	0.90	47-2073	Operating engineers and other construction equipment operators	9,190	2,060
0.05	0.05	0.30	0.20	47-2081	Drywall and ceiling tile installers	2,540	2,700
0.05	0.05	0.30	0.20	47-2082	Tapers	130	1,240
0.05	0.05	0.20	0.10	47-2111	Electricians	11,650	5,150
				47-2121	Glaziers	1,400	0
				47-2131	Insulation workers, floor, ceiling, and wall	240	0
				47-2132	Insulation workers, mechanical	0	0
0.20	0.10	0.30	0.30	47-2141	Painters, construction and maintenance	2,740	2,500
				47-2142	Paperhangers	0	0
				47-2151	Pipelayers	2,410	380
				47-2152	Plumbers, pipefitters, and steamfitters	6,230	4,500
0.20	0.10	0.30	0.30	47-2161	Plasterers and stucco masons	0	1,810
0.20	0.30	0.10	0.90	47-2171	Reinforcing iron and rebar workers	170	0
0.25	0.50	0.10	0.90	47-2181	Roofers	1,850	1,300
				47-2211	Sheet metal workers	3,360	1,300
0.40	0.30	0.10	0.90	47-2221	Structural iron and steel workers	850	0
0.40	0.10	0.30	0.60	47-3011	Helpers--brickmasons, blockmasons, stonemasons, and tile and marble setters	460	620
0.20	0.25	0.30	0.40	47-3012	Helpers--carpenters	0	0
0.05	0.05	0.20	0.10	47-3013	Helpers--electricians	2,140	400
0.20	0.10	0.30	0.30	47-3014	Helpers--painters, paperhangers, plasterers, and stucco masons	130	230
				47-3015	Helpers--pipelayers, plumbers, pipefitters, and steamfitters	2,230	0
0.25	0.50	0.10	0.90	47-3016	Helpers--roofers	590	240
0.40	0.10	0.30	0.60	47-3019	Helpers, construction trades, all other	680	140
0.05	0.05	0.20	0.10	47-4011	Construction and building inspectors	1,650	1,010
				47-4021	Elevator installers and repairers	340	0
0.30	0.30	0.10	0.90	47-4031	Fence erectors	0	290
				47-4041	Hazardous materials removal workers	510	150
0.30	0.30	0.10	0.90	47-4051	Highway maintenance workers	810	450
0.20	0.10	0.30	0.60	47-4071	Septic tank servicers and sewer pipe cleaners	150	0
0.20	0.10	0.30	0.60	47-4099	Construction and related workers, all other	370	270
				47-5021	Earth drillers, except oil and gas	130	0
				47-5051	Rock splitters, quarry	0	0
				47-5081	Helpers--extraction workers	30	0

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				47-5099	Extraction workers, all other	0	0
				49-0000	**Installation, maintenance, and repair occupations**		
				49-1011	First-line supervisors/managers of mechanics, installers, and repairers	10,740	2,580
				49-2011	Computer, automated teller, and office machine repairers	3,420	1,190
				49-2021	Radio mechanics	370	0
				49-2022	Telecommunications equipment installers and repairers, except line installers	5,170	0
				49-2091	Avionics technicians	190	40
				49-2092	Electric motor, power tool, and related repairers	310	80
				49-2094	Electrical and electronics repairers, commercial and industrial equipment	2,310	0
				49-2096	Electronic equipment installers and repairers, motor vehicles	120	60
				49-2097	Electronic home entertainment equipment installers and repairers	1,520	280
				49-2098	Security and fire alarm systems installers	1,390	300
				49-3011	Aircraft mechanics and service technicians	0	220
0.10	0.00	0.50	0.10	49-3021	Automotive body and related repairers	2,640	950
				49-3022	Automotive glass installers and repairers	0	0
				49-3023	Automotive service technicians and mechanics	10,510	4,970
				49-3031	Bus and truck mechanics and diesel engine specialists	3,660	1,500
0.10	0.00	0.50	0.00	49-3041	Farm equipment mechanics	130	120
				49-3042	Mobile heavy equipment mechanics, except engines	2,110	710
				49-3051	Motorboat mechanics	170	160
				49-3052	Motorcycle mechanics	130	130
				49-3053	Outdoor power equipment and other small engine mechanics	370	130
				49-3091	Bicycle repairers	30	70
				49-3092	Recreational vehicle service technicians	0	0
0.10	0.00	0.50	0.00	49-3093	Tire repairers and changers	1,230	1,050
				49-9012	Control and valve installers and repairers, except mechanical door	530	270
0.00	0.00	0.20	0.00	49-9021	Heating, air conditioning, and refrigeration mechanics and installers	1,900	680
				49-9031	Home appliance repairers	1,140	170
				49-9041	Industrial machinery mechanics	3,220	460
				49-9042	Maintenance and repair workers, general	19,670	6,880
				49-9043	Maintenance workers, machinery	1,080	180
				49-9044	Millwrights	610	160
0.20	0.30	0.10	0.90	49-9051	Electrical power-line installers and repairers	2,670	0
0.20	0.10	0.10	0.70	49-9052	Telecommunications line installers and repairers	2,440	850
				49-9062	Medical equipment repairers	520	200
				49-9069	Precision instrument and equipment repairers, all other	480	60
				49-9091	Coin, vending, and amusement machine servicers and repairers	410	120
				49-9093	Fabric menders, except garment	0	0
				49-9094	Locksmiths and safe repairers	410	200
				49-9095	Manufactured building and mobile home installers	0	0

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				49-9096	Riggers	0	0
0.10	0.00	0.30	0.10	49-9098	Helpers--installation, maintenance, and repair workers	3,580	620
0.10	0.00	0.30	0.10	49-9099	Installation, maintenance, and repair workers, all other	2,050	1,080
				51-0000	**Production occupations**		
				51-1011	First-line supervisors/managers of production and operating workers	9,860	2,150
				51-2021	Coil winders, tapers, and finishers	500	70
				51-2022	Electrical and electronic equipment assemblers	2,100	510
				51-2023	Electromechanical equipment assemblers	460	130
				51-2031	Engine and other machine assemblers	450	350
				51-2041	Structural metal fabricators and fitters	1,160	0
				51-2091	Fiberglass laminators and fabricators	310	3,650
				51-2092	Team assemblers	18,110	690
				51-2099	Assemblers and fabricators, all other	4,150	890
				51-3011	Bakers	2,540	950
				51-3021	Butchers and meat cutters	3,140	300
				51-3022	Meat, poultry, and fish cutters and trimmers	1,940	40
				51-3023	Slaughterers and meat packers	1,080	690
				51-3092	Food batchmakers	1,100	130
				51-4011	Computer-controlled machine tool operators, metal and plastic	600	310
				51-4012	Numerical tool and process control programmers	120	30
				51-4021	Extruding and drawing machine setters, operators, and tenders, metal and plastic	1,580	0
				51-4022	Forging machine setters, operators, and tenders, metal and plastic	220	0
				51-4023	Rolling machine setters, operators, and tenders, metal and plastic	510	130
				51-4031	Cutting, punching, and press machine setters, operators, and tenders, metal and plastic	2,840	870
				51-4032	Drilling and boring machine tool setters, operators, and tenders, metal and plastic	310	30
				51-4033	Grinding, lapping, polishing, and buffing machine tool setters, operators, and tenders, metal and plastic	740	200
				51-4034	Lathe and turning machine tool setters, operators, and tenders, metal and plastic	330	30
				51-4035	Milling and planing machine setters, operators, and tenders, metal and plastic	180	40
				51-4041	Machinists	3,600	1,120
				51-4051	Metal-refining furnace operators and tenders	80	0
				51-4052	Pourers and casters, metal	50	0
				51-4062	Patternmakers, metal and plastic	30	0
				51-4072	Molding, coremaking, and casting machine setters, operators, and tenders, metal and plastic	1,790	0
				51-4081	Multiple machine tool setters, operators, and tenders, metal and plastic	1,190	80

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				51-4111	Tool and die makers	790	0
0.10	0.30	0.10	0.90	51-4121	Welders, cutters, solderers, and brazers	4,850	1,130
				51-4122	Welding, soldering, and brazing machine setters, operators, and tenders	610	0
				51-4191	Heat treating equipment setters, operators, and tenders, metal and plastic	110	0
				51-4192	Lay-out workers, metal and plastic	40	40
				51-4193	Plating and coating machine setters, operators, and tenders, metal and plastic	270	80
				51-4194	Tool grinders, filers, and sharpeners	60	60
				51-4199	Metal workers and plastic workers, all other	80	80
				51-5011	Bindery workers	970	310
				51-5012	Bookbinders	140	0
				51-5021	Job printers	480	0
				51-5022	Prepress technicians and workers	870	420
				51-5023	Printing machine operators	4,900	660
				51-6011	Laundry and dry-cleaning workers	4,420	940
				51-6021	Pressers, textile, garment, and related materials	1,060	450
				51-6031	Sewing machine operators	3,300	370
				51-6041	Shoe and leather workers and repairers	0	0
				51-6051	Sewers, hand	0	0
				51-6052	Tailors, dressmakers, and custom sewers	210	0
				51-6061	Textile bleaching and dyeing machine operators and tenders	370	0
				51-6062	Textile cutting machine setters, operators, and tenders	310	0
				51-6063	Textile knitting and weaving machine setters, operators, and tenders	820	0
				51-6064	Textile winding, twisting, and drawing out machine setters, operators, and tenders	1,470	0
				51-6091	Extruding and forming machine setters, operators, and tenders, synthetic and glass fibers	340	0
				51-6092	Fabric and apparel patternmakers	110	0
				51-6093	Upholsterers	240	0
				51-6099	Textile, apparel, and furnishings workers, all other	250	0
				51-7011	Cabinetmakers and bench carpenters	1,510	920
				51-7031	Model makers, wood	0	160
				51-7041	Sawing machine setters, operators, and tenders, wood	460	550
				51-7042	Woodworking machine setters, operators, and tenders, except sawing	1,270	580
				51-7099	Woodworkers, all other	0	0
				51-8012	Power distributors and dispatchers	0	50
				51-8021	Stationary engineers and boiler operators	0	310
				51-8031	Water and liquid waste treatment plant and system operators	1,760	330
				51-8091	Chemical plant and system operators	320	0
				51-8099	Plant and system operators, all other	70	0
				51-9011	Chemical equipment operators and tenders	340	0

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				51-9012	Separating, filtering, clarifying, precipitating, and still machine setters, operators, and tenders	320	0
				51-9021	Crushing, grinding, and polishing machine setters, operators, and tenders	320	120
				51-9022	Grinding and polishing workers, hand	300	180
				51-9023	Mixing and blending machine setters, operators, and tenders	2,520	350
				51-9031	Cutters and trimmers, hand	610	90
				51-9032	Cutting and slicing machine setters, operators, and tenders	1,540	100
				51-9041	Extruding, forming, pressing, and compacting machine setters, operators, and tenders	1,090	90
				51-9051	Furnace, kiln, oven, drier, and kettle operators and tenders	280	100
				51-9061	Inspectors, testers, sorters, samplers, and weighers	6,170	1,190
				51-9071	Jewelers and precious stone and metal workers	510	70
				51-9081	Dental laboratory technicians	1,720	290
				51-9082	Medical appliance technicians	130	90
				51-9111	Packaging and filling machine operators and tenders	7,450	1,000
				51-9121	Coating, painting, and spraying machine setters, operators, and tenders	860	250
0.00	0.00	0.40	0.20	51-9122	Painters, transportation equipment	1,110	340
0.00	0.00	0.30	0.10	51-9123	Painting, coating, and decorating workers	510	0
				51-9131	Photographic process workers	360	80
				51-9132	Photographic processing machine operators	800	320
				51-9192	Cleaning, washing, and metal pickling equipment operators and tenders	180	1,290
				51-9193	Cooling and freezing equipment operators and tenders	110	0
				51-9194	Etchers and engravers	60	0
				51-9195	Molders, shapers, and casters, except metal and plastic	0	100
				51-9196	Paper goods machine setters, operators, and tenders	2,070	180
				51-9198	Helpers--production workers	9,450	1,150
				51-9199	Production workers, all other	4,770	2,350
				53-0000	**Transportation and material moving occupations**		
				53-1021	First-line supervisors/managers of helpers, laborers, and material movers, hand	3,670	1,010
				53-1031	First-line supervisors/managers of transportation and material-moving machine and vehicle operators	4,780	1,210
				53-2012	Commercial pilots	380	70
				53-2021	Air traffic controllers	670	230
				53-2022	Airfield operations specialists	130	0
				53-3011	Ambulance drivers and attendants, except emergency medical technicians	0	0
				53-3022	Bus drivers, school	4,890	1,190
				53-3031	Driver/sales workers	8,450	2,260
				53-3032	Truck drivers, heavy and tractor-trailer	29,200	7,570
				53-3033	Truck drivers, light or delivery services	16,770	5,930
				53-3041	Taxi drivers and chauffeurs	1,130	690

3 days low	5 days low	3 days high	5 days high	SOC Code	Occupation Title	Atlanta employment	Sacramento employment
				53-3099	Motor vehicle operators, all other	1,350	270
				53-4011	Locomotive engineers	40	0
				53-4013	Rail yard engineers, dinkey operators, and hostlers	30	0
				53-4099	Rail transportation workers, all other	50	0
				53-5021	Captains, mates, and pilots of water vessels	0	30
				53-5022	Motorboat operators	0	30
0.30	0.30	0.10	0.90	53-6021	Parking lot attendants	2,210	1,060
0.10	0.10	0.10	0.50	53-6031	Service station attendants	1,290	330
				53-6041	Traffic technicians	0	220
				53-6051	Transportation inspectors	730	90
				53-6099	Transportation workers, all other	0	390
				53-7011	Conveyor operators and tenders	580	220
0.20	0.30	0.10	0.90	53-7021	Crane and tower operators	860	90
0.20	0.30	0.10	0.90	53-7032	Excavating and loading machine and dragline operators	360	250
0.10	0.00	0.50	0.20	53-7051	Industrial truck and tractor operators	15,930	2,940
				53-7061	Cleaners of vehicles and equipment	6,400	2,270
0.10	0.00	0.50	0.20	53-7062	Laborers and freight, stock, and material movers, hand	48,480	12,290
				53-7063	Machine feeders and offbearers	1,540	420
				53-7064	Packers and packagers, hand	16,850	4,250
0.30	0.30	0.10	0.90	53-7081	Refuse and recyclable material collectors	1,550	530
				53-7199	Material moving workers, all other	0	160
22,480	40,378	57,791	90,469		TOTALS - Atlanta	2,211,840	
11,341	22,041	23,868	43,828		TOTALS - Sacramento		853,950