

RESPONSE TO THE EXTERNAL PEER REVIEW

OF THE

STOCHASTIC HUMAN EXPOSURE AND DOSE
SIMULATION FOR PARTICULATE MATTER
(SHEDS-PM)
Version 3.5

September 2010

TABLE OF CONTENTS

INTRODUCTION.....	1
RESPONSE TO PEER REVIEWER COMMENTS.....	2
GENERAL COMMENTS.....	2
MODEL INSTALLATION (CHARGE QUESTION #1)	7
EXAMPLE MODEL RUN IN USER GUIDE (CHARGE QUESTION #2)	8
POPULATION VARIABILITY IN PM _{2.5} EXPOSURE (CHARGE QUESTION #3).....	10
UNCERTAINTY IN PM _{2.5} EXPOSURE (CHARGE QUESTION #4).....	14
SPATIAL VARIABILITY IN PM _{2.5} EXPOSURE (CHARGE QUESTION #5)	16
SUMMARY ASSESSMENT (CHARGE QUESTION #6)	17
<i>Model Output and Algorithms</i>	17
<i>User Guide and Algorithm Descriptions</i>	18
<i>Graphical User Interface (GUI)</i>	22
PRIORITY RANKING FOR IMPROVEMENTS (CHARGE QUESTION #7)	22
REFERENCES	25
APPENDIX	26

Disclaimer

Information in this document has been funded by the United States Environmental Protection Agency under Contract No. EP-D-07-100 to Versar, Inc. It has been subjected to Agency peer and administrative review and has been approved for publication as an EPA document.

Introduction

EPA's National Exposure Research Laboratory (NERL) has developed a human exposure model for assessing the variability and uncertainty in population exposures to particulate matter, called the Stochastic Human Exposure and Dose Simulation for Particulate Matter (SHEDS-PM). SHEDS-PM simulates the time-series of inhalation exposure and dose for individuals that demographically represent a population of interest based on PM concentrations supplied as input to the model. The generation of the time-series involves stochastic processes utilizing numerical Monte-Carlo sampling to characterize the variability within an individual over time and between individuals across a population. Uncertainty in the model output is estimated by incorporating the knowledge- and/or measurement-based uncertainty associated with the inputs through multiple iterations of the model.

The first version of SHEDS-PM was developed to estimate the contribution of ambient PM from outdoor sources, as well as indoor sources of PM, to total personal exposure (Burke et al., 2001). The model was then integrated into a mechanistically consistent source-to-dose modeling framework for estimating the time series of PM exposure and dose for each simulated individual (Georgeopoulos et al., 2005). The latest version of the model, SHEDS-PM 3.5, is a user-friendly exposure modeling tool capable of broad application for PM exposure assessment. The model has a graphical user interface (GUI) for selecting inputs, defining model run scenarios, and analysis of model results. Required input databases (US Census demographic data and human activity diary data) are included in the model, and a detailed User Guide has been developed. SHEDS-PM 3.5 has been applied internally by NERL for various projects (e.g., Ozkaynak et al., 2009). An external peer review was required prior to release of the model to users outside of NERL.

The external peer review of SHEDS-PM 3.5 was completed in December 2009 under EPA Contract No. EP-D-07-100 to Versar, Inc. Versar coordinated the scientific and technical review of the SHEDS-PM 3.5 model by an independent panel of scientists with relevant expertise. Experts in PM exposure and human exposure modeling external to EPA were identified and contacted by Versar, and five were selected as the peer reviewers following an assessment of any potential conflict of interest. The peer reviewers selected were: Arlene Rosenbaum (ICF International), Barry Ryan (Emory University), Ira Tager and Fred Lurmann (University of California, Berkeley; Sonoma Technology, Inc.), Helen Suh (Harvard School of Public Health), and Cliff Weisel (Environmental & Occupational Health Sciences Institute (EOHSI)/UMDNJ).

Each reviewer was provided the SHEDS-PM 3.5 model and User Guide, along with a set of charge questions developed by NERL that required the reviewers to perform several different model runs to evaluate the technical performance of the model algorithms, verify the accuracy and completeness of the User Guide, and provide recommendations for future improvements.

Each reviewer provided Versar with a written report of their responses to the charge questions, and Versar consolidated the comments from the five reviewers into a final report.

This document provides NERL's response to the external peer reviewer comments and is structured into sections that directly correspond to the sections in the Versar Final Report, including general comments and comments for each of the seven charge questions. Within each section, reviewer comments are provided in tables grouped by topic. Each comment in a table has a reference number, the initials of the reviewer that provided the comment, and the page number of the Versar Final Report where the comment is located. The Versar Final Report is provided as the appendix to this document for reference.

Response to Peer Reviewer Comments

General Comments

Overall, the external peer reviewers provided positive comments in their general assessment of the SHEDS-PM 3.5 model. The reviewers commented that SHEDS-PM 3.5 is a state-of-the-science exposure modeling tool with a well-designed user interface for model run specification and exploration of model results. The reviewers also commented that the User Guide was clearly written, well-organized, and thorough. Table 1 below provides a listing of the general comments that did not require a response.

Table 1. General Comments by Peer Reviewers Not Requiring a Response

Ref. No.	Comment	Reviewer	Report Page No.
<i>Model Comments</i>			
1-1	SHEDS-PM is a state-of the science exposure modeling tool. Some advanced modeling features include estimation of dose as well as exposure, capability of performing 2-stage Monte Carlo sampling to estimate variability and uncertainty separately, and well designed GUIs that facilitate data input and results analysis with a wide range of output options. The GUIs makes the model extremely easy to implement and provides the capability to quickly construct graphs, plots, and maps, as well as to stratify results.	AR	p. 7
1-2	The SHED-PM model appears to be very complete and comprehensive, allowing both variability and uncertainty to be modeled. The model requires a large amount of input data, data that are unlikely to be available for many situations. However, that may not be problematic in that the large populations simulated, along with the numerous microenvironments allow the researcher to glean much useful information from the model results much of which should be generalizable to any other situations. Aside from the large database needed to run the model, the model specification is quite straightforward. Various individual microenvironments can be explored as can specific age groups, gender-specific exposures.	BR	p. 7
1-3	The SHEDS-PM model contains many of the features expected for a modern stochastic general population exposure model.	IT/FL	p. 8

Ref. No.	Comment	Reviewer	Report Page No.
1-4	The model appears to provide a state of the science approach for rapidly modeling distributions of exposures when the input data are available and can result in sound conclusions about exposures in many regions of the country.	CW	p. 9
1-5	Overall the SHEDS model was simple to use within the settings provided, i.e. all input files being provided. It appeared to generate valid data sets based on the input, with the few exceptions or questions noted below in the response to the charge questions. The framework of plots and summary tables that are available allow for a rapid examination of different trends in the data so that potential variations in the PM ambient concentration, exposure and dose can be easily compared as well as the levels present in various microenvironments. The mapping capacity provides a visual idea of the exposures across a region and can provide individual census tract information.	CW	p. 9
<i>User Guide Comments</i>			
1-5	The User Guide is well organized, well written, and easy to follow, with a few exceptions noted.	AR	p. 7
1-6	The User Guide for SHEDS-PM 3.5 is especially thorough and well written, providing clear and easy to follow instructions that are helpful in navigating the SHEDS-PM software. Further, the manual provides a nice introduction to the software, with background information and some references. The model software GUIs are also visually appealing and relatively easy to read.	HS	p. 7
1-7	The User Guide and companion papers describe the model and explain its use reasonably well.	IT/FL	p. 8
1-8	The User Guide is written clearly and in detail to provide the user with the necessary guidance to run the SHEDS model. There is sometime too much detail or redundancy, though that is better for those that need it and can be skipped over by individuals who have worked with this type of model previously.	CW	p. 9

In their general comments, the reviewers identified two areas needing improvement that were also noted by multiple reviewers in their comments addressing the specific charge questions. These include the level of guidance provided in the User Guide and the model run time. The reviewers commented in this section that more guidance was needed in the User Guide to provide additional context and explanation for the less experienced or knowledgeable user, and that model run times were much longer than expected, as shown in Table 2 below. Due to the general nature of these comments, we provide our responses in this section. However, additional reviewer comments on these topics can be found in tables in the sections below on the example model run and test scenarios (Charge Questions #2, #3 and #5).

The User Guide was developed and maintained at the level of detail needed for a user already familiar with human exposure modeling. Therefore, we agree with the reviewer comments on the need for additional guidance for less knowledgeable users. In response to Comment #2-1, the User Guide has been revised to provide more information about the various options and their impact on model results. These changes to the User Guide were focused on adding details that would aid the user in making appropriate decisions for their application, including relevant references. The GUI includes ‘Help’ pushbuttons that provide quick access to sections of the User Guide for each screen. It is unclear why the pushbuttons did not function for the reviewer

(since no other reviewer commented on this). In future versions of the model we will consider providing additional on-screen information and help modules. In response to Comment #2-2, the User Guide has been revised to incorporate information on the default input parameters for the microenvironment concentration equations, including how they were developed from measurement study data, and a discussion of important issues to consider when selecting appropriate input parameters for an application.

Table 2. General Comments on Additional Guidance Needed and Model Run Time

Ref. No.	Comment	Reviewer	Report Page No.
<i>Additional Guidance</i>			
2-1	The manual and especially the software, however, do assume a great deal of knowledge about exposure assessment, particulate behavior, and activity patterns on the part of the user, limiting its accessibility, usability, and interpretability of the results. To help in this regard, the software would benefit from direct linkages to the relevant sections of the manual, including not only the step-by-step instructions, but also relevant information about what the options mean, when they should choose between various options, and their implications for particulate exposures. To do so, targeted help modules and/or from further instruction imbedded on the screen would be helpful. [The “View User Guide” button did not work on my version. Similarly, the help screen buttons (when available) were not working.] Also, it would be helpful to include scientific links, citations or additional information in the manual and on the screen that can provide some guidance that will help people select and think about the different options.	HS	p. 8
2-2	Insufficient guidance is provided for the user regarding the process of selecting scientifically credible input data. For example, the data and methods to calculate microenvironmental concentrations are often critical for the results. It was disturbing to find that the test problems data and regression equations for the nonresidential microenvironments came from an unpublished reference (Zufall et al. submitted 2001). Many users are likely to use whatever data comes with the model without critically evaluating its suitability for their applications. We believe the user guide, for example, would benefit from the presentation and explanation of how residential mass balance model parameters and nonresidential microenvironmental concentrations estimating equations are selected for one or more regions of the U.S. (the results could become the basis for the model’s default parameter).	IT/FL	p. 8
<i>Model Run Time</i>			
2-3	One difficulty is the size of the files that must be manipulated and the time that takes to do the calculations. While the laptop I was using is hardly state of the art, is also not archaic. Yet the estimates of time were consistently underestimated by about 50%. Further, a trial scenario that takes 24 hours to perform does not make the best test of the system. The time associated with writing out data files for later use, coupled with the size of the files gives one pause. For example, in Scenario #2, writing the data to disk took in excess of three hours and ended up with an MS Excel file that exceeded 250 MB in size. If this program is to be useful as a tool for the typical exposure assessor, this process should be streamlined.	BR	p. 7

Regarding model run time, we acknowledge that the amount of time it takes to perform some types of model runs may be an issue for users. However, run time is an issue for population-based exposure models that simulate thousands of individuals over time in order to provide distributions of exposures that adequately represent population variability. In response to Comment #2-3 (and other comments in the following sections), we have improved model run times for SHEDS-PM in two ways. First, we have upgraded the model code to the latest version of Matlab® which provides substantial improvement in run time (approx. 20% reduction). Second, since the lung deposition algorithm is the largest contributor to the model run time, we have modified the code to allow the user to select whether to calculate deposited dose to the lung in a model run. Model run times can be reduced by nearly 40% if the user does not require estimates of deposited dose. In addition, more information about run times on a variety of computers has been provided in the User Guide so that a user can calculate an approximate run time based on their computer hardware and the number of individuals and days to be simulated.

In addition, two reviewers summarized their specific comments from the charge questions in this section. The individual reviewers' comments are provided below in Table 3. We provide responses to these summary comments in this section, but note where the more specific comments on these topics are located in the following sections on the charge questions. Table 3 also includes the remaining two comments from this section that point out specific problems encountered by one reviewer.

Table 3. Additional General Comments from Reviewers

Ref. No.	Comment	Reviewer	Report Page No.
<i>Summary Comments based on Charge Question Comments</i>			
3-1	In addition, it would be helpful for the user to be able to summarize the results in additional, more flexible ways without having to move to EXCEL or other platforms. For example, it would be great to be able to quality or data checks within the program or to construct specific regression models. Correspondingly, the program would benefit from improved ability to view input databases (for example for time/activity data) directly from the program and also the equations (or codes) used to generate various results. It was unclear whether the user could import measured activity or microenvironmental concentration databases, so that if measured data were available, the user could use these data instead of the provided distributional data. This information would transform the program from a “black box” program to one with increased flexibility and scientific rigor.	HS	p. 8
3-2	SHEDS, like other exposure models, provides a mathematical framework for exposure calculations. SHEDS-PM also contains a fair amount of pre-selected or embedded data (CHAD, US Census, etc.). The validity of exposure estimates derives from both the mathematical framework and the choice of data for particular applications. Since most modern exposure models share a common microenvironmental approach, the distinguishing element of exposure simulations is generally the choice of data rather than the model framework. We believe there are some limitations of the embed data (e.g., the CHAD data are out-of-date, the met assignments are based on incorrect estimates of oxygen utilization, and the geographic resolution of census tracts is too coarse to resolve the influence of important local sources, such as traffic).	IT/FL	p. 8

Ref. No.	Comment	Reviewer	Report Page No.
3-3	Confidence in models like SHEDS-PM comes from documented model evaluation, refinement and validation studies using field data. Model evaluation is common practice and essential for most complex mathematical models (e.g., EPA's Models-3 Community Multiscale Air Quality (CMAQ) Modeling System). Even if the different types of data and submodels selected for SHEDS are individually sound, the performance of the whole model against real-world data needs to be demonstrated for exposure scientists and epidemiologists to accept the model. Thus, the lack of one or more peer-reviewed, published model validation studies undermines the credibility of the SHEDS-PM model.	IT/FL	p. 9
3-4	Given the lack of validation of the model, the out of date activity data, incorrect estimates of oxygen utilization, and likely uncertainty and variability of dose estimates, we doubt that any creditable epidemiologist would use the current model to estimate individual-level exposure and dose or even distributions of exposure and dose for the general population. If EPA release the model in the near future, it is important to disclose the model's limitations and have a program to address them.	IT/FL	p. 9
<i>Specific Problems</i>			
3-5	Other issues relate to the fact that the database needed to run the requested analyses was initially omitted from the provided materials, resulting in some confusion as to whether the database was not provided or whether the database was imbedded in one of the database files. This confusion suggests that the databases and other information contained in each module should be more clearly delineated.	HS	p. 8
3-6	Further, running the SHEDS-PM often made other programs on my computer fail, requiring a hard reboot before these other programs could be used again. As a result, work in these other programs was lost. Some warning of this possibility should be provided prior to running the program.	HS	p. 8

In Comment #3-1, the reviewer recommends several improvements to the model that would add new features in different areas such as analysis of results, data checks, and input data. Specific comments by the reviewer on these topics can be found in the sections on the test scenarios and the summary assessment (Comments #6-5, #6-6, #8-5 and #9-18), as well as in the priority ranking of possible model improvements (see Charge Question #7). We agree with the reviewer that the suggested improvements would be valuable additions to the model, and the additional flexibility and advanced features would be beneficial for users. However, implementing these features would require significant programming changes, and currently our resources are focused on improving the scientific basis of the model algorithms and conducting model evaluations. Therefore, improvements such as those suggested in Comment #3-1 will be given higher priority in the future when other aspects of the model development are completed.

Comment #3-2 summarizes the reviewers criticisms of the input data embedded in the SHEDS-PM model. More specific reviewer comments on these topics can be found in the summary assessment section (Comments #9-10, #9-12, and #9-19). We have addressed the comment regarding the SHEDS-PM human activity database by updating it with the latest version of NERL's Consolidated Human Activity Database (CHAD) that includes approx. 10,000 new diaries from 1999-2003, many of which are for school-aged children. We disagree with the

reviewers criticism of the MET assignments used to estimate oxygen consumption, and are conducting research in collaboration with exercise physiologists to improve METS estimates specifically for children and the elderly. A more detailed response on this is provided in a later section addressing Comment #9-12. Regarding the criticism of the geographic resolution for the US Census data, we plan to add flexibility in the census unit resolution to the SHEDS-PM model when we develop the input database from the 2010 US Census data after it becomes available.

We agree with the reviewer comment on the importance of model evaluation (Comment #3-3). Evaluations of previous versions of SHEDS-PM using measurement study data have been conducted and the results presented at scientific conferences. Model evaluation studies using the current version of SHEDS-PM are underway using data from two NERL human exposure studies: the RTP PM Panel Study (Williams et al., 2003) and the Detroit Exposure and Aerosol Research Study (DEARS) (Williams et al., 2009). Journal manuscripts on both these projects are being developed to document the evaluation results for the peer-reviewed version of the SHEDS-PM model.

In response to Comment #3-4, our program is currently focused on addressing the limitations noted by the reviewers as described above for the two previous comments (e.g., model evaluation, improved METS estimates, and updated human activity data). Also, the model was not designed to estimate exposures for specific individuals, only population distributions of exposures. When the model is released, we intend to provide appropriate documentation and information to help users understand both the model's strengths and its limitations.

Two specific problems were also identified by one reviewer in this general comments section. Comment #3-5 has been addressed by providing more introductory information in the User Guide on the various input databases required by the SHEDS-PM model and the specific location of the databases provided as part of the model. This reviewer also had problems running other software at the same time as the SHEDS-PM model (Comment #3-6). We were not able to replicate this, but have added running the model with other software open to our QA testing.

Model Installation (Charge Question #1)

All reviewers were able to successfully install the model on their computers. Reviewers used both Windows XP and Vista operating systems. A few minor issues were encountered as noted in the comments in Table 4. In response to these comments, the model installation section of the User Guide was revised to show a figure of the pop-up window noted in Comment #4-1, to have clearer instructions regarding the administrative privileges needed for installation and the location of instructions for Vista operating systems (Comment #4-2), and a warning regarding automatic rebooting of computers after software updates (Comment #4-3).

Table 4. Comments by Peer Reviewers Regarding Installation of SHEDS-PM on a Computer

Ref. No.	Comment	Reviewer	Report Page No.
4-1	I encountered a pop-up window reading "extract census boundaries – one time only" that was not mentioned in the User Guide instructions. Noting this in the instructions will confirm to the user that this is not a problem.	AR	p. 10

Ref. No.	Comment	Reviewer	Report Page No.
4-2	Because my laptop runs Vista, I ran into a small problem installing the program. The User's Manual gives instructions for XP, working all the way through, then modifications for VISTA. My desktop computer at work runs XP, but access to administrative mode is restricted. Therefore all of my testing was done on my laptop. Because of the minor difficulty outlined above, I suggest a stronger statement in the User's Manual regarding Administrative Mode. Perhaps even a separate, albeit repetitive, set of instructions for XP, Vista, and now Windows 7, is in order. If your operating system is Windows XP, go here. If Windows Vista go to page, xx. Et cetera.	BR	p. 10
4-3	I did encounter a problem when running the program for the longer time period (overnight) in that my computers, as is the case for many, are scheduled to do updates of windows and other resident programs during the night. On both computers one of the updates required an automatic restart of the computer. This resulted in a loss of the results obtained from runs, which for a run that takes hours can be at least an annoyance. I therefore had to turn off the scheduled update options on my computer when running the 24+hour runs so as not to lose the results prior to my review of the analysis results. I suggest this be indicated in the installation section AND in other parts of the manual unless it can be fixed.	CW	p. 11

Example Model Run in User Guide (Charge Question #2)

The reviewers successfully completed the example model run described in the User Guide. Most of the reviewers commented that additional context and explanation was needed for the example model run. Individual reviewers' comments regarding this issue are provided in Table 5 below. We agree with the reviewers that this section of the User Guide should be improved by providing more of a tutorial for the first-time user, rather than the current "cookbook" approach for the example.

In response to Comments #5-1 through #5-4, we have improved this section of the User Guide in several ways. First, the overall structure of this section was changed to provide background information and context for each major step in the example. Second, additional explanation was included on the selection of model run inputs and various options, as well as references to more information. Third, instructions for creating a new PM concentration input file were provided in this section, in addition to the example PM input file provided with the model code. Fourth, we have specified a random number seed for the example to allow the user to confirm their output exactly matches the example. Finally, we have added text to the end of this section reiterating that this section provides one example, and that each application of the model requires attention to the preparation of model inputs and appropriate selection of model options for the application.

Comment #5-5 is a general comment on the SHEDS-PM GUIs, and not specific to the example model run. The reviewer noted the limitations of the SHEDS-PM options for analyzing the model results, similar to another reviewer comments above (Comment #3-1). We recognize the limitations, but the model allows the output to be exported for use in other software for additional analysis or graphing as concluded by the reviewer.

Table 5. Comments by Peer Reviewers on Example Model Run in User Guide

Ref. No.	Comment	Reviewer	Report Page No.
<i>Need for Additional Explanation</i>			
5-1	As written, the manual takes one through various sections of the input and running of the SHEDS-PM model. However, it is very “cook-book.” It tells you to press this button, select this option, etc., without going into any detail or supplying any information about what is being accomplished by pursuing that action. This is a failing of the document. While technically fulfilling the requested information about “...familiarize[ing] the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model” I would not know how to run a substantively different scenario that the one input given the information present at this time. It is satisfying to get a result and see that the system actually does produce (a lot of) data, it would be better if I felt as though I knew what I was doing a bit more. While I realize that the remainder of the Manual does indeed address the specifics of what each step means, it would be useful to give at least some context and explanation this point. For example, one could simply say, “... now we are going to take the data as input from an external file, and use it to perform a Monte Carlo simulation. Begin this by reading in the data. This is accomplished by...” and continue.	BR	p. 12
5-2	The example test run and its associated components were fine, although perhaps would have been better with more information about each step. For example, introductory information regarding what the example test run will teach, the processes involved, and the reasons for generating the output.	HS	p. 13
5-3	The example test run was easy to follow and produced output very similar to the User’s Guide. It might make sense to specify the random number seed(s) so that the user could confirm that the program calculated the exact expected results. It is important to emphasize that executing the program with pre-selected inputs is one of many steps needed to understand how exposure modeling should be carried out.	IT/FL	p. 13
5-4	The example, while providing a PM model structure does not require the user to construct the PM data file nor does it provide any guidance on the criteria for selecting the input values, rather the example just provides the values. This is fine for instruction on how to use the screens, which appears to be the focus of the example. There should be a section that provides insight into the PM input values. The description of the file structure for the PM data file provided in the Appendix of the manual is clearly written so should provide the needed directions.	CW	p. 14
<i>Analysis of Results</i>			
5-5	The GUI user interface is relatively easy to use, particularly since it was designed to be linked to the output generated by the SHEDS model so minimal keystrokes and decisions need to be made to see some very common type outputs that are of interest. The tradeoff for this specially develop output tool is its limited in options in the way the graphics are presented. However, that approach is acceptable since the data can be exported if more detailed analyses or graphics are desired.	CW	p. 13
<i>Additional Specific Issues</i>			
5-6	However, as I was experimenting with changing settings for the microenvironment factors , I noticed that whenever I pressed the “cancel” button, I got a message on the DOS screen reading “Error using ==> load; unable to read file mostRecentMicroEnvChoices: No such file or directory”, and the set up screen remained active. When I reset the values to the defaults and pressed the “OK” button the set up screen de-activated.	AR	p. 12

Ref. No.	Comment	Reviewer	Report Page No.
5-7	The default for penetration is normal distribution with mean of 0.97 and sd = 0.2, so a significant amount of the simulations operate with $P > 1.0$. Similarly, would deposition allow a negative value? How is this reconciled?	BR	p. 12
5-8	The Excel spreadsheet came through with formatting problems with the headers. I do not know if the is an XML translation problem, but it was a bit annoying.	BR	p. 12
5-9	The example run was adequate for an initial “tour” of the input screens, though the figures in the printed manual showing screen images are reduced to an extent that I found it difficult to read some of the numbers for comparison purposes.	CW	p. 13
5-10	On page 19, in item C “In-Vehicle Macroenvironment” (shouldn’t it be Microenvironment?) the value for MEAN is missing from the instructions, but since all other values and the default were 0, I used that. However, it should be added to the text.	CW	p. 13
5-11	One minor issue that occurred was if I tried to plot any data prior to pressing the RETREIVE button from the Data Analysis Screen an error message was displayed which continued to be displayed after retrieving the data unless I exited the GUI screen and restarted the data analysis. However, it did not require a new RUN so was not that time consuming, though a fix should be attempted.	CW	p. 14

The additional specific issues noted by the reviewers have been addressed as follows:

- The “Cancel” button has been fixed to function correctly (Comment #5-6)
- Upper and lower limits to the values randomly sampled for the mass balance equation parameters have been added to the text in the User Guide and the GUI for clarity (Comment #5-7)
- Headers in the Excel spreadsheet outputs have been formatted for easier reading (Comment #5-8)
- Figures in the User Guide showing GUI screens were enlarged (Comment #5-9)
- More details on vehicle microenvironment inputs were added to the User Guide (Comment #5-10)
- A catch was added for an attempt to plot before pressing “Retrieve” (Comment #5-11)

Population Variability in $PM_{2.5}$ Exposure (Charge Question #3)

The reviewers successfully completed the SHEDS-PM model run scenarios that estimated population variability in exposures to $PM_{2.5}$ as described in the peer review charge. Individual reviewers’ comments regarding these model runs are provided in Table 6 below.

In performing the three separate model runs for this charge question, reviewers commented that they verified the technical performance of the key model algorithms for simulating PM exposure variability across a population, including population demographics and activity diary assignment, merging of PM concentration data with diaries, calculation of microenvironmental PM concentrations, and the exposure/dose calculations. Some reviewers provided details on their verification of the model algorithms (Comments #6-1, #6-2, and #6-4), while others only confirmed that the model performed as expected for these scenario runs (Comments #6-3 and #6-5).

Table 6. Comments by Peer Reviewers on Population Variability Scenario

Ref. No.	Comment	Reviewer	Report Page No.
<i>Algorithm Verification and General Model Performance</i>			
6-1	<p>For scenario 1-1, the resulting frequency statistics for “gender”, “age”, and “employment status” matched closely with those in the census data base for the tract, and the “season_number” and diary data appeared to be correct. The exposure concentrations matched the air quality input data.</p> <p>For scenario 1-2, the diary data and air quality data were assigned correctly. The ambient ME concentrations compared directionally with the ambient input concentrations as expected. The ratios of indoor-to-input ambient concentrations and in-vehicle-to-input ambient concentrations match the ME factor input distributions for those MEs closely. For a selected individual the number of diary records matched for each of the diaries and the location codes were correctly assigned. For a selected individual ME concentrations in event data and daily data match. Hand-calculated PM exposure matched the values in the events file.</p> <p>For scenario 1-3, comparison plots looked similar to the examples in the instructions.</p>	AR	p. 15
6-2	I was able to complete all the tasks outlined and gradually became more familiar with the workings of the model during this run. I believe the multiple scenarios selected afforded “exercising” the model and displaying all of its most important features. In a few places I “went rogue” and began exploring some features that were not part of the specific challenges offered at the time. The program responded well and gave me better insight into the operations of the system. For example, I inspected several specific microenvironments with regard to the plots and statistics offered. This proved insightful not only with regard to the tuning of the model but also proved fruitful in gaining insight into the abilities of the software.	BR	p. 16
6-3	Yes [the model performed as expected]. Yes, they provide the basic analysis tools.	IT/FL	p. 17
6-4	<p>Model Run #1-1 performed as expected based on the instructions given.</p> <p>Model Run #1-2 performed as expected. It appears to me that the same CHAD diary ID is used for each individual daytype (weekend, Saturday, Sunday) within a season but different ones are used for each daytype. The CHAD diary IDs are different for different seasons. This results in 12 different diaries being used. Longitudinal assignment included 365 days for each simulated individual. The model outputs were comparable to the input distributions.</p> <p>Comparison with exposure/dose calculations: each row did have the correct location based on the CHAD code except when the CHAD code was U or X, which I expect means missing, it was assigned ALL_INDOOR. Time spent in each location for each record was correct (this is based on 12 CHAD diaries used for the year for one individual). Each row in the event file had the correct microenvironmental PM concentration based on the daily export file. The PM exposure for the diary events matched my hand (excel spreadsheet) calculations for individual events and the value for the exposure in the DAILY file matches the hand calculated sums for that day. Using the ventilation rate that was in the SHEDS output the internal dose matched my hand calculations each record and when summed for the day matched the value in the DAILY file. A linear regression calculation comparing the Ambient PM Concentration with the Indoor Air Concentration (determined by summing the Ambient and the non-</p>	CW	p. 17

Ref. No.	Comment	Reviewer	Report Page No.
	<p>Ambient Indoor PM concentration columns, since a total indoor air is not provided in the SHED output) did result in a regression that matched the input data. Home Mass Balance Calculation: the air exchange rates were different on different dates, though some were within .001 of other date, but none were identical. The summary statistics matched the input data. The distribution of each season is consistent with a log normal distribution. Plots of the AER vs the Indoor/Outdoor ratio for the total indoor and for the ambient portion of the indoor both show the expected distribution, as the AER goes up the I/O approaches 1, for the total Indoor from higher I/O ratios and for the ambient only from I/O below unity (see figures). A second run was done with cooking set to 0 resulting in a non-ambient concentration in the home of 0 while it was $9.53 \pm 18.78 \mu\text{g}/\text{m}^3$ in the run with cooking on. The in home ambient concentration for the two runs were $9.53 \pm 7.59 \mu\text{g}/\text{m}^3$ and $9.53 \pm 7.58 \mu\text{g}/\text{m}^3$ indicating that the runs produced similar results for the non affect air concentrations.</p> <p>Model Run #1-3: The output distributions look reasonable and comparable to the plots in Figure 8 with the exception of in-vehicle ambient levels which were much higher than the other microenvironments. The tables that can be generated do give insight into the variability of the PM exposure and dose by gender, age, employment status (the properties examined) and presumably other choices for different microenvironments and days of the week.</p>		
6-5	The model performed as expected, although it was awkward to export the data to EXCEL, open up other databases, and do comparisons. This need for multiple programs involves too many steps and makes the SHEDS-PM seem incomplete and not sufficient on its own. To be complete, the model should include instructions to perform frequency statistics and other data analysis summaries in EXCEL. Otherwise, the software should include these capabilities within the program.	HS	p. 17
6-6	From the analyses, it was difficult to identify human activity factors affecting population variability in PM exposures, although the impact of gender, employment, and age were discernible. This may be due to the fact that more sophisticated analyses are needed to examine impacts of time-activity patterns than were requested or possible with the package. In addition, run results showed indoor non-ambient exposures to be zero, which seemed unlikely; however, it was difficult to figure out whether the values were zero due to an error in the program set-up or to some other reason. The analysis would be greatly improved with a provision to perform more diagnostics and to see the program go through the program steps.	HS	p. 17
<i>Model Run Time</i>			
6-7	The model ran as expected, except that it took about 80 minutes instead of 45, even though no other programs were open.	AR	p. 15
6-8	In running Scenario #1-3, I ran into timing troubles again. I can reproduce the timing table I kept, but the bottom line was that it took in excess of four hours to do this run. I kept on checking back while doing other activities and missed the actual finish, but it was between 218 minutes, then there was an estimate of 13 minutes left, and 242 minutes, when the job had finished. The estimates tended to be too long near the beginning, and too short near the end.	BR	p. 16
<i>Additional Specific Issues</i>			
6-9	Although the instructions state that the intake dose should be hand-calculated from the METS value, it seems like it should actually be calculated from the ventilation rate, according to the intake dose equation on page 130. The values	AR	p. 15

Ref. No.	Comment	Reviewer	Report Page No.
	matched except for a factor of 1000. <u>Note:</u> The intake dose equation on page 130 of the User Guide is off by a factor of 1000, since the concentration units are ug/m3 and the ventilation rate units are L/min. It needs a conversion factor (10^{-3}) to convert ug/m3 to ug/L. (See specific comments below.)		
6-10	The User's Manual reads: "...click on Edit/View Model Run Inputs" The Push Button reads: " View/Edit Model Run Inputs ". The manual should reflect what is in the program to avoid confusion.	BR	p. 16
6-11	One "glitch" I noted occurred during some of the plotting. If one plots multiple microenvironments in the same plots, often the plots themselves plot "through" the legend making for both an untidy presentation and, occasionally, one that is difficult to read. This is doubtless due to fixed size considerations on the plots. I am not sure if this can be remedied either easily or at all, but it was annoying.	BR	p. 16
6-12	Another minor annoyance occurred in that one of the exposure calculated was very high, an unlikely, but somewhat expected, occurrence in any kind of simulation. This resulted in certain of the plots, most notably the box plots, becoming compressed and essentially unusable because of trying to plot this one unusual individual with an exposure in excess of $900 \mu\text{g}/\text{m}^3$. This may have been my random seed that got me this guy, but it will happen.	BR	p. 16
6-13	I had trouble getting the Daily Time Series to run. I kept on getting errors the precluded finishing so I gave up on trying to get that accomplished. I believe the errors looked like: Error using \rightarrow shedpm ('run_Callback') and then some numbers- probably error codes. But this may have been some other error.	BR	p. 16

One reviewer commented that exporting the model output to Excel and performing the algorithm verification steps outside of the program was awkward (Comment #6-5). To perform these steps within the model would require running the model in the Matlab software. The approach of using the model output files allowed the reviewers to work with whatever software they are most familiar with, rather than requiring the reviewers to have Matlab software and programming skills. We addressed one aspect of this comment by adding the option to produce frequency statistics on population demographics to the analysis of model results. Comment #6-6 from the same reviewer also recommends more diagnostic capability within the model software. We will consider additional improvements such as these for future versions of the model (see also response to related Comment #3-1).

Comments on the model run time estimates (Comments #6-7 and #6-8) were also addressed above in response to the General Comments on this topic (see also response to Comment #2-3).

The additional specific issues noted by the reviewers were addressed as follows:

- The correct conversion factor has been added to the intake dose equation in the User Guide (Comment #6-9)
- The User Guide has been changed to be consistent with the GUI (Comment #6-10)
- Instructions for expanding the plot size and adjusting the axis limits for better plot views have been made more frequent in the User Guide (Comments #6-11 and #6-12)
- Additional testing of the plots was conducted to correct errors (Comment #6-13)

Uncertainty in PM_{2.5} Exposure (Charge Question #4)

The reviewers successfully completed the SHEDS-PM model run scenario to examine uncertainty in the population variability of PM_{2.5} exposures as described in the peer review charge. Individual reviewers' comments regarding this model run are provided in Table 7 below.

Most reviewers commented that the model performed as expected for this scenario, with some issues noted. Two reviewers encountered error messages setting up the model run, but were able to figure out the problem after some effort (Comments #7-1 and #7-2). To address this, we have provided greater detail in the User Guide on the progression of steps for setting up a model run with uncertainty, added message boxes to prevent users from getting errors, and conducted additional testing to identify any coding issues. Additional explanation of the output from the uncertainty runs was added to the User Guide to address the part of Comment #7-2 regarding the ability to examine population PM exposures with and without uncertainty for the same model run (the option currently exists within the GUI for this analysis).

Table 7. Comments by Peer Reviewers on Population Variability and Uncertainty Scenario

Ref. No.	Comment	Reviewer	Report Page No.
<i>General Model Performance</i>			
7-1	This scenario performed more or less as I would have expected. I did get some error messages on input, but was able to complete the task by getting around them. I have little new to report in this section. I made use of most of the features in the Analyze Results GUI and explored the output from them. I found the plots interesting again and explored a number of aspects. These visual representations are of most interest and offer a good deal of insight. The amount of work that can be done in terms of data exploration using this tool is enormous. It is truly an amazing tool.	BR	p. 19
7-2	Yes [the model performed as expected], although with some difficulty. Initial runs resulted in error messages. Although the manual provided information to fix the problem, it took several attempts to reboot my computer and re-run program before the program would work. Once the model worked, it performed well. However, it would be helpful if the program would automatically estimate population PM exposures with and without uncertainty to examine the relative impacts of uncertainty in the various microenvironmental infiltration parameters on the exposure distribution.	HS	p. 20
7-3	Yes [model performed as expected]. Yes, they provide the basic analysis tools.	IT/FL	p. 20
7-4	Yes the model performed as expected based on the instructions. Examples of the output are provided below.	CW	p. 20
<i>Additional Specific Issues</i>			
7-5	For restaurants and bars when I tried to add values to a triangular distribution for ASC emission rate, I get an error in the DOS window "Undefined function or variable 'distChosen'". When I save the input window and re-open it the triangular distribution selection has reverted to uniform. Also the Burke et al 2001 article mentioned applying a random factor to whether there was smoking in restaurants. I could not figure out how to implement such a scheme from the User Guide.	AR	p. 19

Ref. No.	Comment	Reviewer	Report Page No.
7-6	However, the exposure values were about half the levels in the graphs provided in Figure 9 and when the same settings were run a second time. In addition, the 50 th percent value in the PM variability and Total Exposures are ~10 ug/m ³ and ~15 ug/m ³ , respectively, which is about twice what appears to be the values in the uncertainty plot. The second run appears to be more consistent with the expected values, though the 50 th percent of the variability plots of the total exposure (last set of figures) is ~18 ug/m ³ and 50 th percentile for the 50 th percent on the uncertainty plot is ~1/2 that value. I do not know if this is the correct comparison between the variability and the uncertainty plots, but if so this discrepancy needs to be evaluated.	CW	p. 20
7-7	I believe that it was in this scenario that I ran into an enormous delay in writing out a file. Like a previous comment, it was at a point when the data are to be written out to a file and the manual says "This may take a few minutes." A few minutes stretched into three hours and the file produce was just over 265 MB in size. This could be a problem. Most computers these days have hard disks that stretch out to 500 GB and more, so the space is not really a problem. However, someone running on an older computer or one that is packed with data may run into a problem. It should be relatively simple to calculate how large a file is likely to be then following that up with a look-see on the operational hard disk to ensure that there is room for it. A text box could give this advice. Further, the phrasing "may take a few minutes" needs some work. A reasonable estimate for the time to write can be made through the software examining the hardware of the computer upon which it is running- disk access speed, expected size of the file, perhaps some other statistics- and given to the user up front. The user could then decide whether to write out the data and go get dinner, or not write out the data. As an alternative approach, a more compressed form of the file could be generated and written out more quickly, and software used to decompress the file on re-input, etc. This would substantially reduce the frustration factor.	BR	p. 19

We have addressed the specific issues noted by the reviewers for this charge question as follows:

- The model code was modified to correct the error identified by the reviewer and allow for the selection of a triangular distribution for the smoking emission rate in the restaurant and bar microenvironments for uncertainty runs (Comment #7-5)
- Comparisons between percentiles for the variability and uncertainty plots were confirmed to be correct (Comment #7-6). The User Guide description of the graphical plots for uncertainty and variability distributions has been improved to provide additional clarification on how the results should be compared.
- Output file size and the time required to export a large file was an issue noted by one reviewer (Comments #7-7). The user can select which output variables to export to a file and subset the output data before exporting in order to create smaller more manageable files. A message box was added to inform the user when a large output file will be generated. The user has the option to cancel the exporting of the data and make a different selection to reduce the file size and export time. More detailed information was added to the User Guide on the potential for large output files to be generated. We will consider other options for speeding up the process in future versions of the model.

Spatial Variability in PM_{2.5} Exposure (Charge Question #5)

The reviewers successfully completed the SHEDS-PM model run to examine spatial variability in PM_{2.5} exposure as described in the peer review charge. Individual reviewers' comments regarding this model run are provided in Table 8 below.

All reviewers commented that the model performed as expected for this scenario, with the main issue being that it took much longer than projected to complete the run (Comments #8-1, #8-2 and #8-4). This scenario required more than 24 hours of model run time, but was included in the peer review charge so that the reviewers performed a typical case study application of the model utilizing PM input concentration data that varied spatially and temporally (in this case, data from 5 different monitors for 1 year). We have significantly reduced the model run time, as described above in our response to the General Comments (see also response to related Comment #2-3). However, a SHEDS-PM simulation characterizing both spatial and temporal variability in PM exposures still requires many hours of model run time. Therefore, in response to the reviewer comment on the usefulness of the model given the long run times (Comment #8-4) we continue to look for ways to reduce the run time for these types of case study applications for future versions of the model. One such option is to take greater advantage of the multiple processors that have become common for computers. We plan to test the Matlab software capabilities for utilizing multiple processors and incorporate it into future versions of SHEDS-PM. In addition, improving the display and analysis of the model results using the map display GUI will be explored for future versions (Comments #8-3 and #8-5).

Table 8. Comments by Peer Reviewers on Spatial Variability Scenario

Ref. No.	Comment	Reviewer	Report Page No.
<i>General Model Performance</i>			
8-1	The model performed as expected, except that the simulation took approximately 36 hours rather than 24.	AR	p. 21
8-2	In general, yes [the model performed as expected.], but my comments below are most important. This was by far the most frustrating component of the review. The system hung about 1/3 of the way through. I had to restart my system and begin again. This time, it ran straight through, but took at least 36 hours to complete. And when it did finally complete and I went to perform the analyses requested, I found that I had looked at most of those features in earlier runs.	BR	p. 21
8-3	Yes, the system offered good insight into these areas [spatial and temporal variability in concentrations and exposures]. The system allowed adequate exploration of all effects. I found plotting the higher percentiles on the census tract most interesting and informative. The lower percentiles provided less insight. This was true no matter which of the parameter- ambient exposure, non-ambient exposure, does, etc. were being plotted.	BR	p. 21
8-4	Yes, [the model performed as expected,] although the instructions and GUI were not reliable regarding the approximate run time and the “estimated run time left”, respectively. Further, the usefulness of the model is greatly reduced given the long run times.	HS	p. 22

Ref. No.	Comment	Reviewer	Report Page No.
8-5	As with the other components, model results would be enhanced with more flexibility in the analysis, specifically so that analyses beyond summary statistics could be performed.	HS	p. 22
8-6	Yes, [the model performed as expected]. Yes, they provide the basic analysis tools.	IT/FL	p. 22
8-7	Yes, [the model performed as expected] as shown in the figures. Maps provide insight into spatial variation along with the specific values in individual census tract when the cursor is moved over it.	CW	p. 22
<i>Additional Specific Issues</i>			
8-8	Yes, although it took some “drilling down” to discover why 2 of the Philadelphia tracts showed some extremely high non-ambient concentrations. They turned out to be from the home ME, presumably from cooking. I obtained a maximum of 770 ug/m3, which may or may not be realistic. This led me to notice that the open-ended distributions are not given any artificial bounds. (See suggestions for other possible future improvements in #7 below)	AR	p. 21
8-9	We could not get the program to print maps centered on the page, regardless of the print setup instructions.	IT/FL	p. 22

The additional specific issues noted by the reviewers were addressed as follows:

- Upper limits for random samples of emission rate distributions for residential indoor sources have been implemented in the model code (Comment #8-8)
- The formatting problem with printing maps has been corrected (Comment #8-9)

Summary Assessment (Charge Question #6)

The reviewers were asked to provide a summary assessment of the SHEDS-PM model, specifically addressing whether the model output was technically correct and consistent with descriptions of the model algorithms in the User Guide, and whether those descriptions in the User Guide were clear and technically correct. In addition, the reviewers were asked to comment on the organization and usability of the GUIs. Individual reviewers’ summary assessment comments for these three topics are provided in Table 9 below.

Model Output and Algorithms Summary Assessment

Two reviewers commented that the model output was consistent with descriptions of the algorithms in the User Guide and appeared to be technically correct (Comments #9-1 and #9-5), while others stated they found nothing obviously incorrect or inconsistent with the algorithm descriptions (Comments #9-2, #9-3 and #9-4). In the latter two comments, the reviewers noted difficulty in determining whether the model outputs were technically correct, specifically due to stochastic sampling within the model (Comment #9-4) and limited user access to all calculation steps (Comment #9-3). In response to these latter comments, we acknowledge that the peer review charge should have clearly explained that verification of each algorithm calculation step can only be done while running the model within Matlab software currently, but the user should be able to check algorithm results using data in the output files. Additional data are now provided in the output files to allow users to verify more calculation steps for each simulated individual, particularly for algorithms that involve stochastic sampling of multiple inputs.

User Guide and Algorithm Descriptions Summary Assessment

The reviewers commented that the User Guide was clear and well organized, and the descriptions of the model algorithms were technically correct and represent the state of the science (Comments #9-6, #9-7, #9-8, #9-9 and #9-14), with one exception. A reviewer identified an error in the User Guide equations for intake dose calculation (Comments #9-1 and #9-6) which has been corrected.

Other reviewer comments to this charge question relate to issues previously noted in the General Comments section, specifically on the level of guidance provided in the User Guide (Comments #9-7, 9-8 and 9-13), and the need for updated human activity data in CHAD (Comment #9-10). As described above in the response to the General Comments on additional guidance needed (see Comments #2-1 and #2-2), the User Guide has been revised to provide more information to aid the user in making appropriate decisions for their application, including relevant references and information on how the default input parameters for the microenvironment concentration equations were developed from measurement study data. Two reviewers provided additional comments on this topic, one questioning whether a version of the model could be created that would be more accessible to the less sophisticated user wanting a simpler tool (Comment #9-7), and the other suggesting that a companion document on exposure modeling with more guidance and tutorials is needed (Comment #9-13). Both ideas are consistent with broadening the use and application of SHEDS-PM outside of EPA, and we plan to consider them in the future.

We have also updated CHAD with newer diaries, many of which are for school-aged children, as noted in the General Comments section (see Comment #3-2). In response to Comment #9-11 on selecting certain studies from CHAD and using a non-CHAD human activity database, these are not options in the model currently. The design of the SHEDS-PM model uses diary matching criteria to select appropriate diaries for each simulated individual, and analyses of the CHAD database support this approach (Graham and McCurdy, 2004; McCurdy and Graham, 2003). However, there may be certain applications where using a subset of the CHAD studies would be important, so we will plan to add this option to future versions of the model. Similarly, we will explore developing a pre-processing module for formatting human activity data for input to SHEDS-PM as the need for this develops with broader use and application of the model.

The two remaining comments on the User Guide addressed specific algorithms. In response to the comment that MET assignments based on kcal overestimate oxygen utilization compared to lean body mass (Comment #9-12), we agree that energy expenditure on a lean body mass basis is a more stable indicator of “work”. However, the MET assignments were developed from exercise physiology data for use with any basal metabolic rate metric (McCurdy, 2000), and population distributions of body composition/fitness level by age/gender needed to assign lean body mass are not currently available. We will plan to add flexibility in the breathing rate algorithm as more comprehensive data on lean body mass become available. We also agree that calculation of the “non-ambient” contribution in non-residential microenvironments could be improved by using the mass balance equation (Comment #9-15); however, little data exist for determining appropriate inputs for these non-residential locations. Adding the mass balance equation option to non-residential microenvironments can easily be added in a future version of the model as more data become available.

Table 9. Summary Assessment Comments by Peer Reviewers

Ref. No.	Comment	Reviewer	Report Page No.
<i>Model Outputs and Algorithms</i>			
9-1	The output generated was consistent with the descriptions of the algorithms in the User Guide, with the exception of the intake dose equation on page 130, as noted above and below in specific comments. They also appear to be technically correct.	AR	p. 23
9-2	I did not examine the technical contents for detailed mathematical errors. However, I saw nothing that gave me pause in the presentation.	BR	p. 23
9-3	The model output is consistent with the descriptions in the User Guide; however, it is not possible to assess its technical correctness as the model does not include user-administered quality control/assurance procedures nor does it display or make available the intermediate model steps and calculations.	HS	p. 24
9-4	None of the outputs were obviously inconsistent with expectation. Determination of whether they are technically correct is very difficult from the stochastic simulations.	IT/FL	p. 25
9-5	The model outputs appear to be technically correct and consistent with the algorithms in the User Guide.	CW	p. 26
<i>User Guide and Algorithm Descriptions</i>			
9-6	I found the descriptions of model components and algorithms in the User Guide to be clear and technically correct, with the exception of the discussion of the intake dose and its underlying components (see specific comments below). The algorithms generally represent the state of the science, although some modifications are suggested in #7 in addition to the ones already listed there.	AR	p. 23
9-7	There is a good deal of technical material there and I think it is presented in a more coherent fashion than in most presentations. For example, I am now plowing my way through the AERMOD series of programs (AERMET, AERSURFACE, etc.) and found this presentation much more rewarding - more like some of the technical appendixes in the documents I just mentioned. The Manual appears written for the exposure scientist who might use this model, rather than a technician looking for answers to a problem using a canned program. This is both a strength and a weakness. It is a strength because the user is likely to be sophisticated in exposure in general. It is a weakness, because the system may be less accessible to the “lay” audience. A decision will have to be made regarding the future direction of such a system. Will an effort be made to present this in a manner more accessible to a non-technical audience? If so, a re-write is in order. However, I would advise against modifying what is here. This is a sophisticated tool and should be used by those who are well versed in the science. Perhaps a “SHEDS Lite” could be developed that was less sophisticated in utilization for those wishing to use a simpler tool.	BR	p. 23
9-8	The User Guide is clear, well organized, and technically correct; however, it would be enhanced with more information about the state of the science, relevance and interpretation of various model components to exposure assessments (as noted in general comments).	HS	p. 24
9-9	The SHED-PM model is nicely packaged and includes many design features needed for state of the science exposure assessments.	IT/FL	p. 24

Ref. No.	Comment	Reviewer	Report Page No.
9-10	One of the problems with the current version of SHEDS is that the embedded CHAD database is outdated with respect to current activity patterns. Clear evidence of the strong temporal changes in activity patterns can be seen in comparison of 1981-82 with 2002-03 activity patterns Tables 16-49, 1650 of the Child-Specific Exposure Factors Handbook. The tables show the shift away from outdoor sports activity to indoor activities related to computers. This trend can be expected to be more pronounced now. The extent to which current estimates of exposure are biased due to this are unknown.	IT/FL	p. 24
9-11	CHAD includes data from different studies and the current model framework does not allow the user to easily select the portions of the CHAD data base that may be suitable for a given application. The user's guide also does not indicate how the user would specify an alternate (non-CHAD) time-activity database for use in model calculations	IT/FL	p. 25
9-12	Another problem is that the MET assignments do not reflect the full range of conversion data that are in the literature. Use of kcal overestimates oxygen utilization, since it includes body fat in the calculation. Ideally, estimates should be based on lean body mass. If such data are not available, then estimates of lean body mass for BMI along with error distributions should be provided. Users should be allowed to specify inputs provided that the following criteria are met: if the data are published, a citation needs to be provided; if data are unpublished, they must be available to the public; at a minimum the data should be specific to age, sex; estimates of error distributions need to be provided.	IT/FL	p. 25
9-13	More attention needs to be given the basis for selection of parameters for estimating microenvironmental concentrations. Care should be taken to carefully select the parameters given as the default values or sample problem values because these will likely be used without evaluation by many potential users. It is important to explain the process and types of data needed to select the parameters for various types of applications (and regions). In fact, there is probably a need for a companion document on exposure modeling that provides scientific guidance and tutorials.	IT/FL	p. 25
9-14	The User Guide is clear in its description of the modeling algorithms used and the combination the multiple microenvironments using the CHAD data base along with microenvironmental air concentration to generate distributions of exposure that include uncertainty estimates represent current state of science for performing exposure assessment. The inclusion of Mass Balance for estimating air concentrations in the indoor environment is a strong advance that potentially increases the potential to make the model output region specific if appropriate input factors are available.	CW	p. 25
9-15	One item that is not clear to me is the assignment of the non-ambient contributions to concentrations, exposure and dose. On page 128 it indicates for the linear regression equation and scaling factor approaches that if the $C_i/C_{\text{ambient}} < 1$ then the "non-ambient" is assumed to be zero. For particles there will be losses as the particles enter the buildings (as discussed in the mass-balance equation approach which will result in that ratio being less than unity in most conditions. While it is recognized that some of the microenvironments do not have enough data to employ the mass balance model, it may still be possible to provide a better estimate of the C_i/C_{ambient} ratio than unity for assigning a non-ambient component to the PM concentration and therefore exposure and dose than when using the current assumption	CW	p. 25

Ref. No.	Comment	Reviewer	Report Page No.
<i>Graphical User Interface (GUI)</i>			
9-16	The GUIs were organized well and very easy to use. I found the many output options to be the most useful, including the mapping and plotting options, as well as the ability to stratify the results. Especially useful additional features would be the ability to save the inputs and results from a simulation, and the ability to turn off dose calculations, as suggested in the list of possible future improvements below.	AR	p. 23
9-17	The GUI seems to be well organized logically once you understand what is being done. As I reported earlier, while the Manual is very complete, the “Getting Started” section is, in my opinion, too “cookbook-like” in that it tells you which buttons to press, but does not give insight into why you are pressing them. The details are supplied in later chapters, but even a brief gloss over of what is happening would add substantial insight. When you first bring the program up, it is pretty intimidating. I realize that the developers and users are long past that stage, but I am a pretty sophisticated software user, and I still felt overwhelmed and under-informed when I first went to use the system. A bit more explanation would be helpful.	BR	p. 23
9-18	The “view/edit model run input” GUI was very organized, clear and straightforward. All of the GUIs would be improved with working and targeted help functions. The “Microenvironment” and “Analyze Results” GUIs would be especially improved, with increased flexibility and ability to run more specialized analyses. The analysis GUIs are weak, allowing only summary type of analyses to be performed. While other statistical programs are available to run more sophisticated analyses, PM-SHEDS would be greatly enhanced with more sophisticated and/or flexible analysis tools.	HS	p. 24
9-19	The graphic user interface is well designed and provides for user control of many model inputs. Because “GUI input only” models inherently limit the user’s control of input parameters, we prefer designs where as many inputs as possible are read from input files (databases) rather than imbedded in the model, and where the model input files can be created from a GUI, preprocessors (where users can examine the outputs), or by a text editor. The User’s Guide and GUI are designed for a fairly unsophisticated user (perhaps at the expense of the flexibility and control more experienced users might want). For example, they don’t provide instructions for (1) how to input different time activity data or (2) how to use non-US census population data or the 2010 census data (when it becomes available) or census block or block group data instead of census tract data.	IT/FL	p. 24
9-20	Overall, the GUI interface was easy to use and allow for easy visualization of individual patterns across concentrations, exposure and dose of ambient and non-ambient sources as well as across different microenvironments. I found the ability to compare different microenvironments most useful if I wanted to better understand where exposures were occurring and how the exposures and times spent in different microenvironments varies across age, gender, season employment status, day type and smoking. The scatter plots provide some insight into underlying associations between different exposures. A mechanism to plot distributions of ratios directly of different outcomes and variables to complement the scatter plots might be worth considering.	CW	p. 25

Graphical User Interface (GUI) Summary Assessment

The reviewers commented that overall the GUI was well designed, logical, and easy to use (Comments #9-16, #9-17, #9-18, #9-19 and #9-20). Two of the reviewers also commented on the usefulness of the many options for analysis of results (Comments #9-16 and #9-20). However, two other reviewers recommended more flexibility in the input and analysis options (Comments #9-18 and #9-19). We conclude from these comments that the GUI structure is currently sufficient for allowing the user to easily select and change inputs, and to perform basic analysis of results. But the flexibility in user control of inputs and analysis of results suggested by the reviewers would certainly be desirable for more advanced users. In response, we will consider making the recommended improvements to the GUI in the future, as the use and application of SHEDS-PM expands.

One reviewer also repeated in this section the need to improve the introductory example in the User Guide (see Comment #5-1) by providing more background information on the GUI options (Comment #9-17). We agree, and as described above in response to Comment #5-1 we have revised that section of the User Guide to be more of a tutorial for the first-time user.

One reviewer noted some specific additional features related to the GUI that would be useful (Comment #9-16). These improvements are addressed more fully in the following section on the reviewers ranking of priority for model improvements. Lastly, one reviewer suggested adding the ability to plot ratios of different output variables to the GUI for analyzing results (Comment #9-20). We agree this specific feature would be useful, and will plan to include it in a future version.

Priority Ranking for Improvements (Charge Question #7)

Each reviewer provided their rankings of the relative priority for several possible improvements to the SHEDS-PM model listed in the charge question. Four different categories of improvements were included, including ease of use, specification of inputs, model algorithms, and new functionality. An overall average of the reviewer rankings was calculated to identify the improvements recommended as high priority by multiple reviewers. Reviewers also provided and ranked some additional improvements not specified in the charge question. Overall, the reviewers' priority rankings are consistent with our future model development plans and current research projects.

Three of the improvements listed in the charge question were consistently ranked high by the reviewers (average greater than 4.0). These include: (1) adding the capability to save user specified settings and recall output for multiple runs, (2) providing a log file that records all inputs specified for the model run that can be viewed and saved by user, and (3) adding an algorithm for estimating air exchange rate in the home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution.

We have modified the SHEDS-PM model to address the first two high priority improvements above. The model now has the capability to save the settings and recall output for multiple model runs, as well as have the input specifications for each model run easily accessible to users. Regarding the third highly ranked improvement, we currently have a research collaboration

underway to develop and evaluate an algorithm for estimating location-specific residential air exchange rates.

The ability to turn off the deposited dose calculations was also ranked as high priority (4 or 5) by three reviewers. The SHEDS-PM model has been modified to allow the user to select whether to calculate deposited dose to the lung in a model run, as noted in response to reviewer comments above (see Comment #2-3). Three of the five reviewers ranked adding more user options to the map view of output (e.g. for use in GIS software or Google Earth) as a high priority improvement. We will explore adding additional options for mapping of model results to a future version of the model.

Reviewers also ranked improving the longitudinal diary assembly algorithm, allowing user-specification of physiological parameter inputs, and allowing the mass balance option for any microenvironment as a priority (rankings of 3 or higher). Additional comments were provided on some of these improvements indicating that their lower rankings were primarily due to the lack of data available for them. We agree with these rankings and the general lack of existing data for them, but plan to incorporate these modifications in future versions as more data becomes available.

Other possible improvements listed in the charge question were either ranked lower priority in general, or had inconsistent priority rankings from the different reviewers. For example, adding the flexibility to use different census units (tracts, block groups, or blocks) in the model was ranked as a lower priority by three reviewers, but higher priority by the other two reviewers. We do plan to incorporate flexibility in the census unit resolution when the input database for the 2010 US Census data is developed in the future (see also Comment #3-2).

Three reviewers provided the following additional recommendations for improvements to the SHEDS-PM model, and ranked them as high priority. We include our responses below in italics:

- Add option to set bounds for open-ended parametric input distributions to avoid unrealistic selections (Priority=5 AR) – *Limits have been added to all distributions to prevent unrealistic values (see also response to Comments #5-7 and #8-8), and user control of these limits will be added to future versions of the model.*
- Allow user to specify re-sampling frequency for diaries and microenvironment factors instead of being fixed in code (Priority=5 AR) – *User control of sampling frequencies is an advanced feature, and therefore, will be considered for future versions of the model.*
- Consider use of Cluster-Markov algorithm for combining activity diaries since it accounts for diary similarities and transition probabilities (Priority=5 AR) – *Different approaches for longitudinal diary assembly are currently being evaluated for incorporation into SHEDS-PM, including the Cluster-Markov algorithm.*
- Incorporate census tract specific commuting time distributions (Priority=4 AR) – *Commuting time has not previously been considered in the algorithm for assigning a work census tract, however, improvements such as this will be considered in future modifications to SHEDS-PM to better characterize population mobility from commuting.*

- Upgrade mass balance algorithm to be dynamic (i.e., allow carryover from one time period to next) (Priority=4 AR) – *The current mass balance equation assumes equilibrium conditions due to the 24 hr integrated PM concentration data available for input to the model initially. Applications using hourly PM measurements or air quality model output have made the dynamic equation more appropriate and we plan to incorporate the necessary changes in the next version of the model.*
- Allow measured time/activity data or microenvironmental concentration data to be imported (Priority=5 HS) – *SHEDS-PM was designed for modeling population distributions of exposures, and so direct importing of study-specific time/activity data or microenvironment concentrations has not been considered an important design feature of the model. However, the user should develop microenvironmental input distributions from the most appropriate measurement data, and the User Guide now provides more information on this. As the use of the model expands, we will consider adding a preprocessor for developing input distributions to the model. In addition, NERL is currently developing an exposure model different from SHEDS-PM for estimating exposures for individuals in health studies where these data have been collected.*
- Allow microenvironment infiltration factors to vary by season (Priority=5 HS) – *Data to quantify variations by season for non-residential microenvironments are currently quite limited; however, this feature will be considered for future versions of the model.*
- Consider making the modeling system open source to encourage innovation and testing of new algorithms (Priority=3 IT/FL) – *We acknowledge the value of making the source code available to users, and we will explore ways to address this in the future.*

References

- Burke, J., M. Zufall, H. Özkaynak (2001) A population exposure model for particulate matter: Case study results for PM_{2.5} in Philadelphia, PA. *Journal of Exposure Science and Environmental Epidemiology*, **11**, 470-489.
- Georgopoulos, P., Wang, S., Vyas, V., Sun, Q., Burke, J., Vedantham, R., McCurdy, T., Özkaynak, H. (2005) A source-to-dose assessment of population exposures to fine PM and ozone in Philadelphia, PA, during a summer 1999 episode. *Journal of Exposure Science and Environmental Epidemiology*, **15**, 439-457.
- Graham, S. and T. McCurdy (2004) Developing meaningful cohorts for human exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, **14**, 23-43.
- McCurdy, T. (2000) Conceptual basis for multi-route intake dose modeling using an energy expenditure approach. *Journal of Exposure Analysis and Environmental Epidemiology*, **10**, 86-97.
- McCurdy, T. and S. Graham (2003) Using human activity data in exposure models: Analysis of discriminating factors. *Journal of Exposure Analysis and Environmental Epidemiology*, **13**, 294-317.
- Özkaynak, H., Frey, H.C., Burke, J., Pinder, R. (2009) Analysis of coupled model uncertainties in source-to-dose modeling of human exposures to ambient air pollution: A PM_{2.5} case study. *Atmospheric Environment*, **43**, 1641-1649.
- Williams, R., Suggs, J., Rea, A., Leovic, K., Vette, A., Croghan, C., Sheldon, L., Rodes, C., Thornburg, J., Ejire, A., Herbst, M., Sanders, W. (2003) The Research Triangle Park particulate matter panel study: PM mass concentration relationships. *Atmospheric Environment*, **37**, 5349-5363.
- Williams, R., Rea, A., Vette, A., Croghan, C., Whitaker, D., Stevens, C., McDow, S., Fortmann, R., Sheldon, L., Wilson, H., Thornburg, J., Phillips, M., Lawless, P., Rodes, C., Daughtrey, H. (2009) The design and field implementation of the Detroit Exposure and Aerosol Research Study. *Journal of Exposure Science and Environmental Epidemiology*, **19**, 643-659.

APPENDIX

COMMENTS SUMMARY REPORT

**EXTERNAL PEER REVIEW OF THE
“STOCHASTIC HUMAN EXPOSURE AND DOSE SIMULATION FOR PARTICULATE
MATTER (SHEDS-PM) VERSION 3.5”**

Prepared for:

**U.S. Environmental Protection Agency
Human Exposure and Atmospheric Science Division
National Exposure Research Laboratory
Research Triangle Park, NC 27711**

**EPA Contract No. EP-D-07-100
Task Order - 003**

Prepared by:

**Versar, Inc.
6850 Versar Center
Springfield, Virginia 22151**

Peer Reviewers:

**Frederick W. Lurmann, MS
Arlene S. Rosenbaum, MPH, PhD
P. Barry Ryan, PhD
Helen H. Suh, ScD
Ira B. Tager, MD, MPH
Clifford P. Weisel, PhD**

February 2010

THIS PAGE LEFT INTENTIONALLY BLANK

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY	A-i
II. INTRODUCTION.....	A-1
III. CHARGE TO THE PEER REVIEWERS.....	A-3
IV. GENERAL COMMENTS.....	A-7
V. RESPONSE TO CHARGE	A-10
VI. SPECIFIC COMMENTS.....	A-31
APPENDIX 1: MODEL SPECIFICATION DETAILS FOR SCENARIO RUNS.....	A-33
APPENDIX 2: PEER REVIEWER COMMENTS.....	A-51

THIS PAGE LEFT INTENTIONALLY BLANK

I. EXECUTIVE SUMMARY

Introduction

EPA's National Exposure Research Laboratory (NERL) has developed a human exposure model for assessing the variability and uncertainty in population exposures to particulate matter, called the Stochastic Human Exposure and Dose Simulation for Particulate Matter (SHEDS-PM). SHEDS-PM simulates the time-series of inhalation exposure and dose for individuals that demographically represent a population of interest based on PM concentrations supplied as input to the model. The generation of the time-series involves stochastic processes utilizing numerical Monte-Carlo sampling to characterize the variability within an individual over time and between individuals across a population. Uncertainty in the model output is estimated by incorporating the knowledge- and/or measurement-based uncertainty associated with the inputs through multiple iterations of the model.

The current version of the model, SHEDS-PM 3.5, is a user-friendly exposure modeling tool capable of broad application for PM exposure assessment. The model has a graphical user interface (GUI) for selecting inputs, defining model run scenarios, and analysis of model results. Required input databases (US Census demographic data and human activity diary data) are included in the model, and a detailed User Guide has been developed. An external peer review of SHEDS-PM 3.5 was required prior to public release of the model.

Under this task order, Versar Inc. coordinated a scientific and technical review of the SHEDS-PM 3.5 model by an independent panel of scientists with relevant expertise. The model was reviewed by Arlene Rosenbaum (ICF International), Barry Ryan (Emory University), Ira Tager and Fred Lurmann (University of California, Berkeley, Sonoma Technology, Inc.), Helen Suh (Harvard School of Public Health), and Cliff Weisel (Environmental & Occupational Health Sciences Institute (EOHSI)/UMDNJ).

Charge to the Peer Reviewers

Peer review charge questions were developed by EPA to evaluate the technical performance of the SHEDS-PM 3.5 model algorithms, verify the accuracy and completeness of the User Guide, and provide recommendations for future improvements. Peer reviewers were instructed to first read through the documentation provided, then install the model on their computer and perform the example model run as described in the User Guide. Once familiar with the model, the peer reviewers performed several different model runs using specific inputs and model features. Three different model scenarios were selected for the peer reviewers to examine the most common types of applications. These included applications of the SHEDS-PM model to estimate population variability in PM exposures, to characterize the uncertainty associated with the estimates of population variability in PM exposures, and to assess spatial variability in PM exposures due to concentration differences between monitors.

Peer Reviewer Comments

Each reviewer was required to provide a written report of their responses to the charge questions, and Versar consolidated their comments for each question into this summary report. The following general comments were also provided by each of the peer reviewers:

Arlene S. Rosenbaum

SHEDS-PM is a state-of the science exposure modeling tool. Some advanced modeling features include: (a) estimation of dose as well as exposure, (b) capability of performing 2-stage Monte Carlo sampling to estimate variability and uncertainty separately, and (c) well designed GUIs that facilitate data input and

results analysis with a wide range of output options. The GUIs makes the model extremely easy to implement and provides the capability to quickly construct graphs, plots, and maps, as well as to stratify results. The User Guide is well organized, well written, and easy to follow, with a few exceptions noted. The exercises selected for the review contained clear directions and demonstrated most of the features of the model. Some of the limitations of the model are addressed in the list of possible future improvements. Some additional ones and associated refinement suggestions are listed below in the “Other” section of possible future improvements.

P. Barry Ryan

The SHED-PM model appears to be very complete and comprehensive, allowing both variability and uncertainty to be modeled. The model requires a large amount of input data, data that are unlikely to be available for many situations. However, that may not be problematic in that the large populations simulated, along with the numerous microenvironments allow the researcher to glean much useful information from the model results much of which should be generalizable to any other situations. Aside from the large database needed to run the model, the model specification is quite straightforward. Various individual microenvironments can be explored as can specific age groups, gender-specific exposures.

One difficulty is the size of the files that must be manipulated and the time that takes to do the calculations. While the laptop I was using is hardly state of the art, is also not archaic. Yet the estimates of time were consistently underestimated by about 50%. Further, a trial scenario that takes 24 hours to perform does not make the best test of the system. The shorter duration tests are a better indicator of what the system can be done. The time associated with writing out data files for later use, coupled with the size of the files gives one pause. For example, in Scenario #2, writing the data to disk took in excess of three hours and ended up with an MSExcel file that exceeded 250 MB in size. If this program is to be useful as a tool for the typical exposure assessor, this process should be streamlined.

Helen H. Suh

It is clear that considerable work and thought has been put into the development of SHEDS-PM and the User Guide. The User Guide for SHEDS-PM 3.5 is especially thorough and well written, providing clear and easy to follow instructions that are helpful in navigating the SHEDSPM software. Further, the manual provides a nice introduction to the software, with background information and some references. The model software GUIs are also visually appealing and relatively easy to read. The manual and especially the software, however, do assume a great deal of knowledge about exposure assessment, particulate behavior, and activity patterns on the part of the user, limiting its accessibility, usability, and interpretability of the results. To help in this regard, the software would benefit from direct linkages to the relevant sections of the manual, including not only the step-by-step instructions, but also relevant information about what the options mean, when they should choose between various options, and their implications for particulate exposures. To do so, targeted help modules and/or from further instruction imbedded on the screen would be helpful. [The “View User Guide” button did not work on my version. Similarly, the help screen buttons (when available) were not working.] Also, it would be helpful to include scientific links, citations or additional information in the manual and on the screen that can provide some guidance that will help people select and think about the different options.

In addition, it would be helpful for the user to be able to summarize the results in additional, more flexible ways without having to move to EXCEL or other platforms. For example, it would be great to be able to quality or data checks within the program or to construct specific regression models. Correspondingly, the program would benefit from improved ability to view input databases (for example for time/activity data) directly from the program and also the equations (or codes) used to generate various results. It was unclear whether the user could import measured activity or microenvironmental concentration databases,

so that if measured data were available, the user could use these data instead of the provided distributional data. This information would transform the program from a “black box” program to one with increased flexibility and scientific rigor.

Other issues relate to the fact that the database needed to run the requested analyses was initially omitted from the provided materials, resulting in some confusion as to whether the database was not provided or whether the database was imbedded in one of the database files. This confusion suggests that the databases and other information contained in each module should be more clearly delineated. Further, running the SHEDS-PM often made other programs on my computer fail, requiring a hard reboot before these other programs could be used again. As a result, work in these other programs was lost. Some warning of this possibility should be provided prior to running the program.

Ira B. Tager/ Frederick W. Lurmann

The SHEDS-PM model contains many of the features expected for a modern stochastic general population exposure model. The User Guide and companion papers describe the model and explain its use reasonably well. We were able to install the software and run the model on the test problems without problems; however, we did not “stress test” the software to identify bugs or other potential problems.

SHEDS, like other exposure models, provides a mathematical framework for exposure calculations. SHEDS-PM also contains a fair amount of pre-selected or embedded data (CHAD, US Census, etc.). The validity of exposure estimates derives from both the mathematical framework and the choice of data for particular applications. Since most modern exposure models share a common microenvironmental approach, the distinguishing element of exposure simulations is generally the choice of data rather than the model framework. We believe there are some limitations of the embed data (e.g., the CHAD data are out-of-date, the met assignments are based on incorrect estimates of oxygen utilization, and the geographic resolution of census tracts is too coarse to resolve the influence of important local sources, such as traffic). Insufficient guidance is provided for the user regarding the process of selecting scientifically credible input data. For example, the data and methods to calculate microenvironmental concentrations are often critical for the results. It was disturbing to find that the test problems data and regression equations for the nonresidential microenvironments came from an unpublished reference (Zufall et al. submitted 2001). Many users are likely to use whatever data comes with the model without critically evaluating its suitability for their applications. We believe the user guide, for example, would benefit from the presentation and explanation of how residential mass balance model parameters and nonresidential microenvironmental concentrations estimating equations are selected for one or more regions of the U.S. the results could become the basis for the model’s default parameter).

Confidence in models like SHEDS-PM comes from documented model evaluation, refinement and validation studies using field data. Model evaluation is common practice and essential for most complex mathematical models (e.g., EPA’s Models-3 Community Multiscale Air Quality (CMAQ) Modeling System). Even if the different types of data and submodels selected for SHEDS are individually sound, the performance of the whole model against real-world data needs to be demonstrated for exposure scientists and epidemiologists to accept the model. Thus, the lack of one or more peer-reviewed, published model validation studies undermines the credibility of the SHEDS-PM model. Given the lack of validation of the model, the out of date activity data, incorrect estimates of oxygen utilization, and likely uncertainty and variability of dose estimates, we doubt that any creditable epidemiologist would use the current model to estimate individual-level exposure and dose or even distributions of exposure and dose for the general population. If EPA release the model in the near future, it is important to disclose the model’s limitations and have a program to address them.

Overall the SHEDS model was simple to use within the settings provided, i.e. all input files being provided. It appeared to generate valid data sets based on the input, with the few exceptions or questions noted below in the response to the charge questions. The framework of plots and summary tables that are available allow for a rapid examination of different trends in the data so that potential variations in the PM ambient concentration, exposure and dose can be easily compared as well as the levels present in various microenvironments. The mapping capacity provides a visual idea of the exposures across a region and can provide individual census tract information. The User Guide is written clearly and in detail to provide the user with the necessary guidance to run the SHEDS model. (There is sometime too much detail or redundancy, though that is better for those that need it and can be skipped over by individuals who have worked with this type of model previously.) The model appears to provide a state of the science approach for rapidly modeling distributions of exposures when the input data are available and can result in sound conclusions about exposures in many regions of the country.

Ranking of Possible Improvements

The reviewers were asked to rank a set of possible improvements in four different categories using a scale from 1 (low priority) to 5 (high priority). They were also asked to suggest and rank other specific improvements of their own. Based on an average of the reviewer rankings, the three highest priority improvements were: (1) adding the capability to save user specified settings and recall output for analysis for multiple runs, (2) create log file that records all inputs specified for the model run that can be viewed and saved by user, and (3) add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution.

II. INTRODUCTION

Introduction to SHEDS-PM

EPA's National Exposure Research Laboratory (NERL) has developed human exposure models for assessing the variability and uncertainty in population exposures to pollutants including particulate matter, air toxics, and pesticides. The Stochastic Human Exposure and Dose Simulation (SHEDS) models are physically-based probabilistic models that utilize two-dimensional Monte Carlo sampling of the input distributions to propagate the variability and uncertainty in the inputs through to the predicted exposure distributions.

The Stochastic Human Exposure and Dose Simulation for Particulate Matter (SHEDS-PM) simulates the time-series of inhalation exposure and dose for individuals that demographically represent a population of interest based on PM concentrations supplied as input to the model. The generation of the time-series involves stochastic processes utilizing numerical Monte-Carlo sampling to characterize the variability within an individual over time and between individuals across a population. Uncertainty in the model output is estimated by incorporating the knowledge- and/or measurement-based uncertainty associated with the inputs through multiple iterations of the model.

The first version of SHEDS-PM was developed to estimate the contribution of ambient PM from outdoor sources, as well as indoor sources of PM, to total personal exposure (Burke et al., 2001). Case study application of the model also produced the first exposure model input distributions for PM_{2.5} based on available data. Integration of the model into a mechanistically consistent source-to-dose modeling framework provided the foundation for the current model structure that estimates the time series of PM exposure and dose for each simulated individual (Georgeopoulos et al., 2005). A graphical user interface (GUI) driven version of the model was developed (SHEDS-PM Version 2.1) with GUIs for selecting inputs, defining the model run scenario, and analysis of model results. Also included in SHEDS-PM 2.1 were the required US Census and human activity databases, as well as a User Guide, to provide a user-friendly exposure modeling tool capable of broad application for PM exposure assessment. Additional algorithm development and user interface enhancements have continued, and resulted in the current version, SHEDS-PM 3.5, which requires external peer review prior to public release of the model.

Description of SHEDS-PM

The overall structure of the SHEDS-PM model has been described in detail by Burke et al. (2001) and Georgopoulos et al. (2005). Briefly, required input databases for the model include census demographic data, human activity pattern data, and PM concentration data. In addition, data for exposure factors used in the model algorithms are input as distributions that characterize the variability and uncertainty in the available data for each exposure factor. The core model algorithms include (a) generation of individuals for the simulation that demographically represent the user-specified population, (b) assignment of appropriate human activity diary records to each simulated individual for the time period of the input PM concentration data, (c) selection of appropriate values from the exposure factor input distributions for each individual (e.g. gender/age-specific values), (d) calculation of PM concentrations in each location the individual spends time in (e.g., outdoors, indoors, in vehicles) based on the input PM concentrations, and (e) calculation of each simulated individual's PM exposure and dose profile for the time period using the PM concentration, time spent, and activity level-specific inhalation rate in each location.

Daily-averaged PM exposure and total daily PM dose for each simulated individual are produced as output from the model, as well as the model estimated population distribution of daily PM exposure and dose (variability) and the uncertainty associated with the model-estimated distributions. In addition, the

contribution of PM from ambient or outdoor air is calculated separately from the contribution of PM from other sources (e.g., cooking) in the model algorithms. This separation is maintained throughout the exposure and dose calculations, producing results for the daily-averaged exposure and total daily dose due to PM from outdoor sources (ambient PM exposure and dose) versus that due to indoor PM sources (non-ambient PM exposure and dose) for each simulated individual as well as the population distribution.

Motivation for Peer Review

The SHEDS-PM model has been developed, applied, and evaluated within NERL's air pollution research program. For example, SHEDS-PM has recently been applied in an analysis of coupled model uncertainty (Özkaynak et al., 2009). Currently, efforts are underway within NERL to evaluate the output from the SHEDS-PM model using PM data from NERL's human exposure studies conducted in Raleigh, NC and Detroit, MI. Additional research efforts with academic collaborators currently include application of the SHEDS-PM model to estimate exposures for use in time series epidemiological studies of the health effects of PM.

The purpose for this peer review was to provide a scientific and technical review of the SHEDS-PM model by an independent panel of scientists with relevant expertise. The reviewers were tasked with evaluating the usability and functionality of the model for use both within NERL's research program and the broader research community. Under this task order, Versar Inc. arranged the peer review of the SHEDS-PM model

Peer Reviewers:

This document was reviewed by:

- 1) Arlene S. Rosenbaum, MPH
ICF International
Rohnert Park, CA 94928
- 2) P. Barry Ryan, PhD
Emory University
Atlanta, Georgia 30322
- 3) Ira B. Tager, MD, MPH
University of California, Berkeley
Berkeley, CA 94720

and

Frederick W. Lurmann, MS
Sonoma Technology, Inc.
Petaluma, CA 94954

- 4) Helen H. Suh, PhD
Harvard School of Public Health
Boston, MA 02215
- 5) Clifford P. Weisel, PhD
Environmental & Occupational Health Sciences Institute (EOHSI)/UMDNJ
Piscataway, NJ 08854

III. CHARGE TO THE PEER REVIEWERS

This review of the SHEDS-PM model will involve (a) installing the model on a computer, and setting up and performing a short example model run using input data provided and instructions in the User Guide, (b) conducting 3 different model runs using specific inputs and settings to test different features of the model, and (c) providing written comments to address charge questions on model performance for the 3 model run scenarios, completeness and accuracy of the User Guide and documentation, and recommendations for future improvements.

First, read through the documentation provided for the SHEDS-PM model which includes the User Guide and three published journal manuscripts. Then follow the procedures described for the items outlined below and provide answers for each of the questions, being as specific and detailed as possible.

1) Install the SHEDS-PM model software program.

Install the model on a computer with Windows XP or later operating system using the file provided (*'EPA SHEDS-PM 3.5 Installation.EXE'*) and following instructions in Section 2 of the User Guide.

- a) Did you encounter any problems using the self-installing executable program to set up the model on your computer?

If yes, please describe the problem, the type of computer used, the operating system release number, the location the model was installed on the computer (e.g. 'C:\Program Files' or other drive), and whether the User Guide provided information to help correct the issue.

- b) Do you have any suggestions for improving the User Guide section on the model installation procedures (Section 2)?

2) Perform SHEDS-PM Example Test Run.

Set up and run the example described in Section 3 of the User Guide using the PM concentration input file provided in the 'Data' directory (*'philaPM2008.csv'*). Display and export the model results as described in Section 3.

- a) Does the example test run provide a sufficient introduction to familiarize the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model?

3) Perform Scenario #1 (Population Variability).

This scenario demonstrates a typical SHEDS-PM application to estimate the variability in exposures to ambient PM_{2.5} for the population of an urban metropolitan area. The PM_{2.5} concentration input file includes daily, 24-hour average PM_{2.5} concentrations for 1 year from a monitor located in an urban area. A representative population from census tracts near the monitor is simulated, and includes all ages and both genders. This scenario defines several microenvironments with different infiltration characteristics for ambient PM_{2.5} (indoor PM sources are not included in this scenario). Analysis of the model results focuses on options available for displaying the output to characterize the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Follow the procedures outlined in **Appendix 1** for specifying the model inputs and analyzing the results for Scenario #1. Provide comments on the following for this scenario:

- a) Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?
- b) Do the options for analysis of model results provide the user with sufficient information to understand the population variability in PM exposures and the impact of human activities?

4) Perform Scenario #2 (Population Variability with Uncertainty).

This scenario demonstrates a SHEDS-PM application to characterize the uncertainty associated with the model estimates of population variability in ambient PM_{2.5} exposures. The same input PM_{2.5} concentration data and population demographics as Scenario #1 are used. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model. Analysis of the model results focuses on displaying the estimated uncertainty in the population distribution of exposure to ambient PM_{2.5}.

Follow the procedures outlined in **Appendix 1** for specifying the model inputs and analyzing the results for Scenario #2. Provide comments on the following for this scenario:

- a) Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?
- b) Do the options for analysis of model results provide the user with sufficient information to understand the predicted uncertainty in the population variability of PM exposures?

5) Perform Scenario #3 (Spatial Variability).

This scenario demonstrates a SHEDS-PM application for understanding the spatial variability in PM_{2.5} exposures. The PM_{2.5} concentration input file includes PM_{2.5} input concentrations for multiple monitoring locations within an urban area. Commuting is included to account for time spent outside the home census tract when individuals are at work. A representative population for each monitor is simulated. Analysis of the model results focuses on options available for displaying the output to understand the spatial variability in PM exposures due to concentration differences between monitors.

Follow the procedures outlined in **Appendix 1** for specifying the model inputs and analyzing the results for Scenario #3. Provide comments on the following for this scenario:

- a) Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?
- b) Do the options for analysis of model results provide the user with sufficient information to understand the impact of spatial and temporal variability in PM concentrations on the modeled distributions of PM exposures?

6) Provide summary assessment.

Please provide comments on the following:

- a) The organization and usability of the user interface (GUIs), which features or options were most useful, and whether additional features or options are needed
- b) Whether the descriptions of the model components and algorithms in the User Guide are sufficiently clear, technically correct, and represent the current state of the science for performing exposure assessments
- c) Whether the output generated by the model are technically correct and consistent with descriptions of the algorithms in the User Guide

7) Rank priority for possible future improvements.

Several possible improvements to the SHEDS-PM model are listed below. Please provide a number ranking for the relative priority that should be given to each improvement, using a scale from 1 (low priority) to 5 (high priority).

Improving ease of use:

- ___ Create log file that records all inputs specified for the model run that can be viewed and saved by user
- ___ Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)
- ___ Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)
- ___ Provide more information on error messages to help users identify the reason for the error for common problems
- ___ Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)

Allowing additional user specification of inputs:

- ___ Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)
- ___ Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)

Improving/refining model algorithms:

- ___ Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of “unique” diary being used repeatedly for an individual)
- ___ Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)
- ___ Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)
- ___ Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)
- ___ Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

Adding new functionality:

- ___ Option for using mapping tool to select census tracts for simulation based on a map
- ___ Add more user options to map view of output (e.g. for use in GIS software or Google Earth)

Other:

___ Please describe: _____

8) Open comments (optional)

Please provide any additional comments that you wish to on the SHEDS-PM model.

IV. GENERAL COMMENTS

Arlene S. Rosenbaum

SHEDS-PM is a state-of the science exposure modeling tool. Some advanced modeling features include:

- Estimation of dose as well as exposure
- Capability of performing 2-stage Monte Carlo sampling to estimate variability and uncertainty separately
- Well designed GUIs that facilitate data input and results analysis with a wide range of output options.

The GUIs makes the model extremely easy to implement and provides the capability to quickly construct graphs, plots, and maps, as well as to stratify results.

The User Guide is well organized, well written, and easy to follow, with a few exceptions noted below.

The exercises selected for the review contained clear directions and demonstrated most of the features of the model.

Some of the limitations of the model are addressed in the list of possible future improvements. Some additional ones and associated refinement suggestions are listed below in the “Other” section of possible future improvements.

P. Barry Ryan

The SHED-PM model appears to be very complete and comprehensive, allowing both variability and uncertainty to be modeled. The model requires a large amount of input data, data that are unlikely to be available for many situations. However, that may not be problematic in that the large populations simulated, along with the numerous microenvironments allow the researcher to glean much useful information from the model results much of which should be generalizable to any other situations.

Aside from the large database needed to run the model, the model specification is quite straightforward. Various individual microenvironments can be explored as can specific age groups, gender-specific exposures.

One difficulty is the size of the files that must be manipulated and the time that takes to do the calculations. While the laptop I was using is hardly state of the art, is also not archaic. Yet the estimates of time were consistently underestimated by about 50%. Further, a trial scenario that takes 24 hours to perform does not make the best test of the system. The shorter duration tests are a better indicator of what the system can be done. The time associated with writing out data files for later use, coupled with the size of the files gives one pause. For example, in Scenario #2, writing the data to disk took in excess of three hours and ended up with an MSExcel file that exceeded 250 MB in size. If this program is to be useful as a tool for the typical exposure assessor, this process should be streamlined.

Helen H. Suh

It is clear that considerable work and thought has been put into the development of SHEDS-PM and the User Guide. The User Guide for SHEDS-PM 3.5 is especially thorough and well written, providing clear and easy to follow instructions that are helpful in navigating the SHEDSPM software. Further, the manual provides a nice introduction to the software, with background information and some references. The

model software GUIs are also visually appealing and relatively easy to read. The manual and especially the software, however, do assume a great deal of knowledge about exposure assessment, particulate behavior, and activity patterns on the part of the user, limiting its accessibility, usability, and interpretability of the results. To help in this regard, the software would benefit from direct linkages to the relevant sections of the manual, including not only the step-by-step instructions, but also relevant information about what the options mean, when they should choose between various options, and their implications for particulate exposures. To do so, targeted help modules and/or from further instruction imbedded on the screen would be helpful. [The “View User Guide” button did not work on my version. Similarly, the help screen buttons (when available) were not working.] Also, it would be helpful to include scientific links, citations or additional information in the manual and on the screen that can provide some guidance that will help people select and think about the different options.

In addition, it would be helpful for the user to be able to summarize the results in additional, more flexible ways without having to move to EXCEL or other platforms. For example, it would be great to be able to quality or data checks within the program or to construct specific regression models. Correspondingly, the program would benefit from improved ability to view input databases (for example for time/activity data) directly from the program and also the equations (or codes) used to generate various results. It was unclear whether the user could import measured activity or microenvironmental concentration databases, so that if measured data were available, the user could use these data instead of the provided distributional data. This information would transform the program from a “black box” program to one with increased flexibility and scientific rigor.

Other issues relate to the fact that the database needed to run the requested analyses was initially omitted from the provided materials, resulting in some confusion as to whether the database was not provided or whether the database was imbedded in one of the database files. This confusion suggests that the databases and other information contained in each module should be more clearly delineated. Further, running the SHEDS-PM often made other programs on my computer fail, requiring a hard reboot before these other programs could be used again. As a result, work in these other programs was lost. Some warning of this possibility should be provided prior to running the program.

Ira B. Tager/ Frederick W. Lurmann

The SHEDS-PM model contains many of the features expected for a modern stochastic general population exposure model. The User Guide and companion papers describe the model and explain its use reasonably well. We were able to install the software and run the model on the test problems without problems; however, we did not “stress test” the software to identify bugs or other potential problems.

SHEDS, like other exposure models, provides a mathematical framework for exposure calculations. SHEDS-PM also contains a fair amount of pre-selected or embedded data (CHAD, US Census, etc.). The validity of exposure estimates derives from both the mathematical framework and the choice of data for particular applications. Since most modern exposure models share a common microenvironmental approach, the distinguishing element of exposure simulations is generally the choice of data rather than the model framework. We believe there are some limitations of the embed data (e.g., the CHAD data are out-of-date, the met assignments are based on incorrect estimates of oxygen utilization, and the geographic resolution of census tracts is too coarse to resolve the influence of important local sources, such as traffic). Insufficient guidance is provided for the user regarding the process of selecting scientifically credible input data. For example, the data and methods to calculate microenvironmental concentrations are often critical for the results. It was disturbing to find that the test problems data and regression equations for the nonresidential microenvironments came from an unpublished reference (Zufall et al. submitted 2001). Many users are likely to use whatever data comes with the model without critically evaluating its suitability for their applications. We believe the user guide, for example, would benefit from the presentation and explanation of how residential mass balance model parameters and

nonresidential microenvironmental concentrations estimating equations are selected for one or more regions of the U.S. the results could become the basis for the model's default parameter).

Confidence in models like SHEDS-PM comes from documented model evaluation, refinement and validation studies using field data. Model evaluation is common practice and essential for most complex mathematical models (e.g., EPA's Models-3 Community Multiscale Air Quality (CMAQ) Modeling System). Even if the different types of data and submodels selected for SHEDS are individually sound, the performance of the whole model against real-world data needs to be demonstrated for exposure scientists and epidemiologists to accept the model. Thus, the lack of one or more peer-reviewed, published model validation studies undermines the credibility of the SHEDS-PM model.

Given the lack of validation of the model, the out of date activity data, incorrect estimates of oxygen utilization, and likely uncertainty and variability of dose estimates, we doubt that any creditable epidemiologist would use the current model to estimate individual-level exposure and dose or even distributions of exposure and dose for the general population. If EPA release the model in the near future, it is important to disclose the model's limitations and have a program to address them.

Clifford P. Weisel

Overall the SHEDS model was simple to use within the settings provided, i.e. all input files being provided. It appeared to generate valid data sets based on the input, with the few exceptions or questions noted below in the response to the charge questions. The framework of plots and summary tables that are available allow for a rapid examination of different trends in the data so that potential variations in the PM ambient concentration, exposure and dose can be easily compared as well as the levels present in various microenvironments. The mapping capacity provides a visual idea of the exposures across a region and can provide individual census tract information.

The User Guide is written clearly and in detail to provide the user with the necessary guidance to run the SHEDS model. (There is sometime too much detail or redundancy, though that is better for those that need it and can be skipped over by individuals who have worked with this type of model previously.)

The model appears to provide a state of the science approach for rapidly modeling distributions of exposures when the input data are available and can result in sound conclusions about exposures in many regions of the country.

V. RESPONSE TO CHARGE

Question No. 1 **Install the SHEDS-PM Model Software Program**

Install the model on a computer with Windows XP or later operating system using the file provided ('EPA SHEDS-PM 3.5 Installation.EXE') and following instructions in Section 2 of the User Guide.

- a) Did you encounter any problems using the self-installing executable program to set up the model on your computer?

If yes, please describe the problem, the type of computer used, the operating system release number, the location the model was installed on the computer (e.g. 'C:\Program Files' or other drive), and whether the User Guide provided information to help correct the issue.

- b) Do you have any suggestions for improving the User Guide section on the model installation procedures (Section 2)?

Arlene S. Rosenbaum

- a) No problems installing.
- b) I encountered a pop-up window reading "extract census boundaries - one time only" that was not mentioned in the User Guide instructions. Noting this in the instructions will confirm to the user that this is not a problem.

P. Barry Ryan

- a) Because my laptop runs Vista, I ran into a small problem installing the program. The User's Manual gives instructions for XP, working all the way through, then modifications for VISTA.

My desktop computer at work runs XP, but access to administrative mode is restricted. Therefore all of my testing was done on my laptop (1.8 GHz, T5550 Processor with 3 GB of memory. 320 GB hard drive, WiFi 802.11g Networking).

- b) Because of the minor difficulty outlined above, I suggest a stronger statement in the User's Manual regarding Administrative Mode. Perhaps even a separate, albeit repetitive, set of instructions for XP, Vista, and now Windows 7, is in order. If your operating system is Windows XP, go here. If Windows Vista go to page, xx. Et cetera.

Helen H. Suh

- a) No. Installation was straight-forward with no identified problems.
- b) The User Guide was clear and comprehensive in model installation procedures.

Ira B. Tager/ Frederick W. Lurmann

- a) The program installation went smoothly on two Windows XP computers. The information in the User Guide was clear and sufficient to install the program.

- b) No, the installation was comparable to other Windows software.

Clifford P. Weisel

- a) No problems were encountered with installing the software on two computers, one running Window XP and a second running VISTA when following the directions provided.

- b) User Guide section is clear

I did encounter a problem when running the program for the longer time period (overnight) in that my computers, as is the case for many, are scheduled to do updates of windows and other resident programs during the night. On both computers one of the updates required an automatic restart of the computer. This resulted in a loss of the results obtained from runs, which for a run that takes hours can be at least an annoyance. I therefore had to turn off the scheduled update options on my computer when running the 24+hour runs so as not to lose the results prior to my review of the analysis results. I suggest this be indicated in the installation section AND in other parts of the manual unless it can be fixed.

Question No. 2

Perform SHEDS-PM Example Test Run

Set up and run the example described in Section 3 of the User Guide using the PM concentration input file provided in the 'Data' directory ('philaPM2008.csv'). Display and export the model results as described in Section 3.

- a) Does the example test run provide a sufficient introduction to familiarize the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model?

Arlene S. Rosenbaum

Yes. However, as I was experimenting with changing settings for the microenvironment factors, I noticed that whenever I pressed the "cancel" button, I got a message on the DOS screen reading "'Error using ==> load; unable to read file mostRecentMicroEnvChoices: No such file or directory", and the set up screen remained active. When I reset the values to the defaults and pressed the "OK" button the set up screen de-activated.

P. Barry Ryan

I have a series of specific comments noted at each of several points along the process here. One overarching comment begs a solution, however. As written, the manual takes one through various sections of the input and running of the SHEDS-PM model. However, it is very "cook-book." It tells you to press this button,, select this, option, etc., without going into any detail or supplying any information about what is being accomplished by pursuing that action. This is a failing of the document. While technically fulfilling the requested information about "...familiarize[ing] the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model" I would not know how to run a substantively different scenario that the one input given the information present at this time. It is satisfying to get a result and see that the system actually does produce (a lot of) data, it would be better if I felt as though I knew what I was doing a bit more. While I realize that the remainder of the Manual does indeed address the specifics of what each step means, it would be useful to give at least some context and explanation this point. For example, one could simply say, "... now we are going to take the data as input from an external file, and use it to perform a Monte Carlo simulation. Begin this by reading in the data. This is accomplished by..." and continue.

Specific Comments:

*Example Test Run:
Section 3-o.*

<i>Example Test Run: Section 3-o.</i>	The bar progresses very slowly on my system. It is not several seconds, but rather 2-3 minutes before the data checking was completed.
<i>Microenvironmental Section 2.iii</i>	The default for penetration is normal distribution with mean of 0.97 and sd = 0.2, so a significant amount of the simulations operate with P>1.0 Similarly, would deposition allow a negative value? HOW is this reconciled?
2.viii	The parameters here are indistinguishable from the Restaurant scenario.
c.	At this point, the whole process is a big "black-box." What is going on?
Section 3.3	Here we make our first real "run." The manual estimates suggest a 45-minute time to complete. Further, it says that if you don't hit these marks, it recommends getting a faster computer. This is not

	<p>especially helpful. Often, one cannot just run you and get a new computer just to run an EPA model. Here is a brief table of m experience given my computer as described above:</p> <p>After a 2 minute setup, estimate is 108 minute remaining.</p> <p>Dropped to 100 minutes after an addition 1 minute</p> <p>Dropped to 85 after an additional minute</p> <p>Dropped to 82 after an additional minute</p> <p>Back to 86 after an additional 3 minutes</p> <p>After 20 minutes, estimate is 72 minutes.</p> <p>After 25 minutes, estimate is 65 minutes</p> <p>After 35 minutes, estimate is 46 minutes.</p> <p>Left to do a bit of housework.</p> <p>Returned after 115 minutes from start and the process was complete.</p>
Comment	<p>I really like the “Current Status” window that is constantly updated.</p> <p>This type of feedback give assurance that one is not stuck in a loop.</p>
3.4 4.c.	The Microsoft table did not appear. I had to go get it.
Page 23 Figure 16	The Excel spreadsheet came through with formatting problems with the headers. I do not know if the is an XML translation problem, but it was a bit annoying.
Page 23 Figure 17	I experimented with a number of other plotting combinations as well. These are quite useful as representations of the data.
Page 24 6.b.	I suggest a “browse” option be put in here.
Page 24 6.c.	Where it says “”will take a few minutes” it took 20 minutes and produced a 165MB file. More warning and a better estimate is warranted here.

Helen H. Suh

The example test run and its associated components were fine, although perhaps would have been better with more information about each step. For example, introductory information regarding what the example test run will teach, the processes involved, and the reasons for generating the output.

Ira B. Tager/ Frederick W. Lurmann

The example test run was easy to follow and produced output very similar to the User’s Guide. It might make sense to specify the random number seed(s) so that the user could confirm that the program calculated the exact expected results. It is important to emphasize that executing the program with pre-selected inputs is one of many steps needed to understand how exposure modeling should be carried out.

Clifford P. Weisel

The example run was adequate for an initial “tour” of the input screens, though the figures in the printed manual showing screen images are reduced to an extent that I found it difficult to read some of the numbers for comparison purposes. On page 19, in item C “In-Vehicle Macroenvironment” (shouldn’t it be Microenvironment?) the value for MEAN is missing from the instructions, but since all other values and the default were 0, I used that. However, it should be added to the text. The GUI user interface is relatively easy to use, particularly since it was designed to be linked to the output generated by the SHEDS model so minimal keystrokes and decisions need to be made to see some very common type outputs that are of interest. The tradeoff for this specially develop output tool is its limited in options in the way the graphics are presented. However, that approach is acceptable since the data can be exported if more detailed analyses or graphics are desired.

One minor issue that occurred was if I tried to plot any data prior to pressing the RETREIVE button from the Data Analysis Screen an error message was displayed which continued to be displayed after retrieving

the data unless I exited the GUI screen and restarted the data analysis. However, it did not require a new RUN so was not that time consuming, though a fix should be attempted.

The example, while providing a PM model structure does not require the user to construct the PM data file nor does it provide any guidance on the criteria for selecting the input values, rather the example just provides the values. This is fine for instruction on how to use the screens, which appears to be the focus of the example. There should be a section that provides insight into the PM input values. The description of the file structure for the PM data file provided in the Appendix of the manual is clearly written so should provide the needed directions.

Question No. 3
Perform Scenario #1 (Population Variability)

This scenario demonstrates a typical SHEDS-PM application to estimate the variability in exposures to ambient PM_{2.5} for the population of an urban metropolitan area. The PM_{2.5} concentration input file includes daily, 24-hour average PM_{2.5} concentrations for 1 year from a monitor located in an urban area. A representative population from census tracts near the monitor is simulated, and includes all ages and both genders. This scenario defines several microenvironments with different infiltration characteristics for ambient PM_{2.5} (indoor PM sources are not included in this scenario). Analysis of the model results focuses on options available for displaying the output to characterize the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Follow the procedures outlined in the **Appendix** for specifying the model inputs and analyzing the results for Scenario #1. Provide comments on the following for this scenario:

- a) Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?
- b) Do the options for analysis of model results provide the user with sufficient information to understand the population variability in PM exposures and the impact of human activities?

Arlene S. Rosenbaum

- a) The model ran as expected, except that it took about 80 minutes instead of 45, even though no other programs were open.

My PC is a Dell Latitude laptop running Windows XP Professional ver 2002, service pack 3, with Intel Core 2 Duo CPU T5500@1.66GHz processors.

For scenario 1-1 the resulting frequency statistics for “gender”, “age”, and “employment status” matched closely with those in the census data base for the tract, and the “season_number” and diary data appeared to be correct. The exposure concentrations matched the air quality input data

For scenario 1-2 the diary data and air quality data were assigned correctly. The ambient ME concentrations compared directionally with the ambient input concentrations as expected. The ratios of indoor-to-input ambient concentrations and in-vehicle-to-input ambient concentrations match the ME factor input distributions for those MEs closely. For a selected individual the number of diary records matched for each of the diaries and the location codes were correctly assigned. For a selected individual ME concentrations in event data and daily data match. Hand-calculated PM exposure matched the values in the events file.

Although the instructions state that the intake dose should be hand-calculated from the METS value, it seems like it should actually be calculated from the ventilation rate, according to the intake dose equation on page 130. The values matched except for a factor of 1000. Note: The intake dose equation on page 130 of the User Guide is off by a factor of 1000, since the concentration units are ug/m³ and the ventilation rate units are L/min. It needs a conversion factor (10⁻³) to convert ug/m³ to ug/L. (See specific comments below.)

For scenario 1-3 comparison plots looked similar to the examples in the instructions.

- b) Yes

P. Barry Ryan

- a) The User's Manual reads: "...click on **Edit/View Model Run Inputs**...." The Push Button reads: "**View/Edit Model Run Inputs**". The manual should reflect what is in the program to avoid confusion.

I was able to complete all the tasks outlined and gradually became more familiar with the workings of the model during this run.

- b) I believe the multiple scenarios selected afforded "exercising: the model and displaying all of its most important features. In a few places I "went rogue" and began exploring some features that were not part of the specific challenges offered at the time. The program responded well and gave me better insight into the operations of the system. For example, I inspected several specific microenvironments with regard to the plots and statistics offered. This proved insightful not only with regard to the tuning of the model but also proved fruitful in gaining insight into the abilities of the software.

Comments:

I do not have much specific to say about this scenario. My main thoughts in running it were to gather acumen and skill in modifying the parameters of the input and evaluating what came out. In this regard, the software seems quite complete. I perhaps spent less time comparing my results to the tables in the Appendix than I should have, but I found it more interesting to "play" with the program to determine what kinds of output were available, what parameters could be modified and the effects such modification would have on the output, and exploring graphical forms of outputs, e.g., pie charts, scatter plots, etc. In this regard, I think I went a bit out of sequence and hence got a bit frustrated later on with the long time scales needed to complete some of the tasks.

I enjoyed this section of the evaluation more than the others, perhaps because of my curiosity and the exploration done.

One "glitch" I noted occurred during some of the plotting. If one plots multiple microenvironments in the same plots, often the plots themselves plot "through" the legend making for both an untidy presentation and, occasionally, one that is difficult to read. This is doubtless due to fixed size considerations on the plots. I am not sure if this can be remedied either easily or at all, but it was annoying.

Another minor annoyance occurred in that one of the exposure calculated was very high, an unlikely, but somewhat expected, occurrence in any kind of simulation. This resulted in certain of the plots, most notably the box plots, becoming compressed and essentially unusable because of trying to plot this one unusual individual with an exposure in excess of $900 \mu\text{g}/\text{m}^3$. This may have been my random seed that got me this guy, but it will happen.

In running Scenario #1-3, I ran into timing troubles again. I can reproduce the timing table I kept, but the bottom line was that it took in excess of four hours to do this run. I kept on checking back while doing other activities and missed the actual finish, but it was between 218 minutes, then there was an estimate of 13 minutes left, and 242 minutes, when the job had finished. The estimates tended to be too long near the beginning, and too short near the end.

I had trouble getting the Daily Time Series to run. I kept on getting errors the precluded finishing so I gave up on trying to get that accomplished. I believe the errors looked like: Error using \rightarrow shedprn ('run_Callback') and then some numbers- probably error codes. But this may have been some other error.

Helen H. Suh

- a) The model performed as expected, although it was awkward to export the data to EXCEL, open up other databases, and do comparisons. This need for multiple programs involves too many steps and makes the SHEDS-PM seem incomplete and not sufficient on its own. To be complete, the model should include instructions to perform frequency statistics and other data analysis summaries in EXCEL. Otherwise, the software should include these capabilities within the program.
- b) From the analyses, it was difficult to identify human activity factors affecting population variability in PM exposures, although the impact of gender, employment, and age were discernible. This may be due to the fact that more sophisticated analyses are needed to examine impacts of time-activity patterns than were requested or possible with the package. In addition, run results showed indoor non-ambient exposures to be zero, which seemed unlikely; however, it was difficult to figure out whether the values were zero due to an error in the program set-up or to some other reason. The analysis would be greatly improved with a provision to perform more diagnostics and to see the program go through the program steps.

Ira B. Tager/ Frederick W. Lurmann

- a) Yes.
- b) Yes, they provide the basic analysis tools.

Clifford P. Weisel

[A summary of the output results, including charts and tables, can be found in Appendix 2, brief responses to the question can be found below]

- a) Model Run 1-1 The model for the Scenario #1 performed as expected based on the instructions given.

Model Run #1-2 Yes, model performed as expected.

It appears to me that the same CHAD diary ID is used for each individual daytype (weekend, Saturday, Sunday) within a season but different ones are used for each daytype. The CHAD diary IDs are different for different seasons. This results in 12 different diaries being used. Longitudinal assignment included 365 days for each simulated individual.

The model outputs were comparable to the input distributions. In examining this I had concerns that I may have overwritten the in-vehicle column when I was manipulating the data for other purposes before preparing the graph, so I ran the simulation again and verified for that run the in-vehicle were greater than the input PM concentration.

Comparison with exposure/dose calculations: Each row did have the correct location based on the CHAD code except when the CHAD code was U or X, which I expect means missing, it was assigned ALL_INDOOR. The locations were only ALL_INDOOR; ALL_OUTDOOR or ALL_IN VEHICLE, no sub locations were specified in SHED. Time spent in each location for each record was correct. (this is based on 12 CHAD diaries used for the year for one individual.) Each row in the event file had the correct microenvironmental PM concentration based on the daily export file. The PM exposure for the diary events matched my hand (excel spreadsheet) calculations for individual events and the valued for the exposure in the DAILY File matches the hand calculated sums for that day. Using the ventilation rate that was in the SHED output the internal dose matched my hand calculations each record and when summed for the day matched the value in the DAILY file. A Linear regression calculation comparing the Ambient PM Concentration with the Indoor air Concentration (determined by summing the Ambient and the non-Ambient Indoor PM concentration columns, since a total indoor air is not provided in the SHED output) did result in a regress that matched the input data.

Since I calculated the indoor air concentration by summing the ambient indoor air concentration with the non-ambient indoor air concentration columns, it does not make sense to me to “Confirm that the non-ambient PM contribution is calculated as described in Appendix D” which starts with the Indoor Air concentration and subtracts the Ambient component.

Home Mass Balance Calculation. The air exchange rates were different on different dates, though some were within .001 of other date, but none were identical.

The summary statistics matched the input data. The distribution of each season is consistent with a log normal distribution

Plots of the AER vs the Indoor/Outdoor ratio for the total indoor and for the ambient portion of the indoor both show the expected distribution, as the AER goes up the I/O approaches 1, for the total Indoor from higher I/O ratios and for the ambient only from I/O below unity. (see figures)

Confirm non-ambient contribution from cooking in home microenvironment. A second run was done with cooking set to 0 resulting in a non-ambient concentration in the home of 0 while it was $9.53 \pm 18.78 \mu\text{g}/\text{m}^3$ in the run with cooking on. The in home ambient concentration for the two runs were $9.53 \pm 7.59 \mu\text{g}/\text{m}^3$ and $9.53 \pm 7.58 \mu\text{g}/\text{m}^3$ indicating that the runs produced similar results for the non affect air concentrations.

Model Run 1-3

The output distributions look reasonable and comparable to the plots in Figure 8 as below [see Appendix 2] with the exception of in-vehicle ambient levels which were much higher than the other microenvironments.

- b) The tables that can be generated do give insight into the variability of the PM exposure and dose by gender, age, employment status (the properties examined) and presumably other choices for different microenvironments and days of the week.

Question No. 4
Perform Scenario #2 (Population Variability with Uncertainty)

This scenario demonstrates a SHEDS-PM application to characterize the uncertainty associated with the model estimates of population variability in ambient PM_{2.5} exposures. The same input PM_{2.5} concentration data and population demographics as Scenario #1 are used. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model. Analysis of the model results focuses on displaying the estimated uncertainty in the population distribution of exposure to ambient PM_{2.5}.

Follow the procedures outlined in the **Appendix** for specifying the model inputs and analyzing the results for Scenario #2. Provide comments on the following for this scenario:

- a) Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?
- b) Do the options for analysis of model results provide the user with sufficient information to understand the predicted uncertainty in the population variability of PM exposures?

Arlene S. Rosenbaum

- a) No. For restaurants and bars when I tried to add values to a triangular distribution for ASC emission rate, I get an error in the DOS window “Undefined function or variable ‘distChosen’”. The when I save the input window and re-open it the triangular distribution selection has reverted to uniform.

Also the Burke et al 2001 article mentioned applying a random factor to whether there was smoking in restaurants. I could not figure out how to implement such a scheme from the User Guide.

- b) Yes.

P. Barry Ryan

- a) This scenario performed more or less as I would have expected. I did get some error messages on input, but was able to complete the task by getting around them.
- b) I have little new to report in this section. I made use of most of the features in the Analyze Results GUI and explored the output from them. I found the plots interesting again and explored a number of aspects. These visual representations are of most interest and offer a good deal of insight.

Because of the timing problems I had in Scenario #3 (see below), I had to set this project aside for a period of 4-5 days, and then return to it. Hence, some of my recollections may be a bit in error. Nevertheless, I forge ahead. I believe that it was in this scenario that I ran into an enormous delay in writing out a file. Like a previous comment, it was at a point when the data are to be written out to a file and the manual says “This may take a few minutes.” A few minutes stretched into three hours and the file produce was just over 265 MB in size. This could be a problem. Most computers these days have hard disks that stretch out to 500 GB and more, so the space is not really a problem. However, someone running on an older computer or one that is packed with data may run into a problem. It should be relatively simple to calculate how large a file is likely to be then following that up with a look-see on the operational hard disk to ensure that there is room for it. A text box could give this advice. Further, the phrasing “may take a few minutes” needs some work. A reasonable estimate for the time to write can be made through the software examining the hardware of the computer upon which it is running- disk access speed, expected size of the file, perhaps some other

statistics- and given to the user up front. The user could then decide whether to write out the data and go get dinner, or not write out the data. As an alternative approach, a more compressed form of the file could be generated and written out more quickly, and software used to decompress the file on re-input, etc. This would substantially reduce the frustration factor.

All of these things being said, the amount of work that can be done in terms of data exploration using this tool is enormous. It is truly an amazing tool.

Helen H. Suh

- a) Yes, although with some difficulty. Initial runs resulted in error messages that asked that I consult with Janet Burke. Although the manual provided information to fix the problem, it took several attempts to reboot my computer and re-run program before the program would work. Once the model worked, it performed well.
- b) Yes. However, it would be helpful if the program would automatically estimate population PM exposures with and without uncertainty to examine the relative impacts of uncertainty in the various microenvironmental infiltration parameters on the exposure distribution.

Ira B. Tager/ Frederick W. Lurmann

- a) Yes.
- b) Yes, they provide the basic analysis tools.

Clifford P. Weisel

[A summary of the output results, including charts and tables, can be found in Appendix 2, brief responses to the question can be found below]

- a) Yes the model performed as expected based on the instructions. Examples of the output are provided below. However, the exposure values were about half the levels in the graphs provided in Figure 9 and when the same settings were run a second time. In addition, the 50th percent value in the PM variability and Total Exposures are ~10 ug/m³ and ~15 ug/m³, respectively, which is about twice what appears to be the values in the uncertainty plot. The second run appears to be more consistent with the expected values, though the 50th percent of the variability plots of the total exposure (last set of figures) is ~18 ug/m³ and 50th percentile for the 50th percent on the uncertainty plot is ~1/2 that value. I do not know if this is the correct comparison between the variability and the uncertainty plots, but if so this discrepancy needs to be evaluated.
- b) Yes, though I prefer the plot style in Burke et al 2001 figure 4.

Question No. 5
Perform Scenario #3 (Spatial Variability)

This scenario demonstrates a SHEDS-PM application for understanding the spatial variability in PM_{2.5} exposures. The PM_{2.5} concentration input file includes PM_{2.5} input concentrations for multiple monitoring locations within an urban area. Commuting is included to account for time spent outside the home census tract when individuals are at work. A representative population for each monitor is simulated. Analysis of the model results focuses on options available for displaying the output to understand the spatial variability in PM exposures due to concentration differences between monitors.

Follow the procedures outlined in the **Appendix** for specifying the model inputs and analyzing the results for Scenario #3. Provide comments on the following for this scenario:

- a) Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?
- b) Do the options for analysis of model results provide the user with sufficient information to understand the impact of spatial and temporal variability in PM concentrations on the modeled distributions of PM exposures?

Arlene S. Rosenbaum

- a) The model performed as expected, except that the simulation took approximately 36 hours rather than 24.

My PC is a Dell Latitude laptop running Windows XP Professional ver 2002, service pack 3, with Intel Core 2 Duo CPU T5500@1.66GHz processors.

- b) Yes, although it took some “drilling down” to discover why 2 of the Philadelphia tracts showed some extremely high non-ambient concentrations. They turned out to be from the home ME, presumably from cooking. I obtained a maximum of 770 ug/m³, which may or may not be realistic. This led me to notice that the open-ended distributions are not given any artificial bounds. (See suggestions for other possible future improvements in #7 below)

P. Barry Ryan

- a) In general, yes, but my comments below are most important.

This was by far the most frustrating component of the review. I did not notice the expected time for this run until two days before the due date for the report- now several days in the past. But, I figured, I have two days- I run the scenario and even if it takes 30 hours I will still have plenty of time. So off I went. The software chugged along for a period of time and I finally went to bed, expecting the system to take care of itself and complete its task while I slept. But while I slept, something bad happened. I do not know what. The system hung about 1/3 of the way through. It appeared to be still running in the morning and it took me a few minutes to realize that it was constantly displaying the same tract, individual, etc. I had to restart my system and begin again. This time, it ran straight through, but took at least 36 hours to complete. And when it did finally complete and I went to perform the analyses requested, I found that I had looked at most of those features in earlier runs. So, I was unable to complete my task on time, and had other priorities scheduled for the intervening few days.

- b) Yes. The system offered good insight into these areas.

Comments:

The system allowed adequate exploration of all effects. I found plotting the higher percentiles on the census tract most interesting and informative. The lower percentiles provided less insight. This was true no matter which of the parameter- ambient exposure, non-ambient exposure, does, etc.- were being plotted.

Helen H. Suh

- a) Yes, although the instructions and GUI were not reliable regarding the approximate run time and the “estimated run time left”, respectively. Further, the usefulness of the model is greatly reduced given the long run times.
- b) As with the other components, model results would be enhanced with more flexibility in the analysis, specifically so that analyses beyond summary statistics could be performed.

Ira B. Tager/ Frederick W. Lurmann

- a) Yes.
- b) Yes, they provide the basic analysis tools. We could not get the program to print maps centered on the page, regardless of the print setup instructions.

Clifford P. Weisel

[Figures can be found in Appendix 2, brief responses to the question can be found below]

- a) Yes, as shown in the figures.
- b) Maps provide insight into spatial variation along with the specific values in individual census tract when the cursor is moved over it.

Question No. 6
Provide Summary Assessment

Please provide comments on the following:

- a) The organization and usability of the user interface (GUIs), which features or options were most useful, and whether additional features or options are needed
- b) Whether the descriptions of the model components and algorithms in the User Guide are sufficiently clear, technically correct, and represent the current state of the science for performing exposure assessments
- c) Whether the output generated by the model are technically correct and consistent with descriptions of the algorithms in the User Guide

Arlene S. Rosenbaum

- a) The GUIs were organized well and very easy to use. I found the many output options to be the most useful, including the mapping and plotting options, as well as the ability to stratify the results.

Especially useful additional features would be (a) the ability to save the inputs and results from a simulation and (b) the ability to turn off dose calculations, as suggested in the list of possible future improvements below. Some other possible future improvements are suggested below.

- b) I found the descriptions of model components and algorithms in the User Guide to be clear and technically correct, with the exception of the discussion of the intake dose and its underlying components (see specific comments below).

The algorithms generally represent the state of the science, although some modifications are suggested in #7 in addition to the ones already listed there.

- c) The output generated was consistent with the descriptions of the algorithms in the User Guide, with the exception of the intake dose equation on page 130, as noted above and below in specific comments. They also appear to be technically correct.

P. Barry Ryan

- a) The GUI seems to be well organized logically once you understand what is being done. As I reported earlier, while the Manual is very complete, the “Getting Started” section is, in my opinion, too “cookbook-like” in that it tells you which buttons to press, but does not give insight into why you are pressing them. The details are supplied in later chapters, but even a brief gloss over of what is happening would add substantial insight. When you first bring the program up, it is pretty intimidating. I realize that the developers and users are long past that stage, but I am a pretty sophisticated software user, and I still felt overwhelmed and under-informed when I first went to use the system. A bit more explanation would be helpful.
- b) I did not examine the technical contents for detailed mathematical errors. However, I saw nothing that gave me pause in the presentation. There is a good deal of technical material there and I think it is presented in a more coherent fashion than in most presentations. For example, I am now plowing my way through the AERMOD series of programs (AERMET, AERSURFACE, etc.) and found this presentation much more rewarding- more like some of the technical appendixes in the documents I just mentioned. The Manual appears written for the exposure scientist who might use this model, rather than a technician looking for answers to a problem using a canned program. This is both a strength and a weakness. It is a strength because the user is likely to be sophisticated in exposure in

general. It is a weakness, because the system may be less accessible to the “lay” audience. A decision will have to be made regarding the future direction of such a system. Will an effort be made to present this in a manner more accessible to a non-technical audience? If so, a re-write is in order. However, I would advise against modifying what is here. This is a sophisticated tool and should be used by those who are well versed in the science. This may sound elitist; if so, so be it. Perhaps a “SHEDS Lite” could be developed that was less sophisticated in utilization for those wishing to use a simpler tool.

- c) I believe I covered this in the above comment.

Helen H. Suh

- a) The “view/edit model run input” GUI was very organized, clear and straight-forward. All of the GUIs would be improved with working and targeted help functions. The “Microenvironment” and “Analyze Results” GUIs would be especially improved, with increased flexibility and ability to run more specialized analyses. The analysis GUIs are weak, allowing only summary type of analyses to be performed. While other statistical programs are available to run more sophisticated analyses, PM-SHEDS would be greatly enhanced with more sophisticated and/or flexible analysis tools.
- b) The User Guide is clear, well organized, and technically correct; however, it would be enhanced with more information about the state of the science, relevance and interpretation of various model components to exposure assessments (as noted in general comments).
- c) The model output is consistent with the descriptions in the User Guide; however, it is not possible to assess its technical correctness as the model does not include user-administered quality control/assurance procedures nor does it display or make available the intermediate model steps and calculations.

Ira B. Tager/ Frederick W. Lurmann

- a) The graphic user interface is well designed and provides for user control of many model inputs. Because “GUI input only” models inherently limit the user’s control of input parameters, we prefer designs where as many inputs as possible are read from input files (databases) rather than imbedded in the model, and where the model input files can be created from a GUI, preprocessors (where users can examine the outputs), or by a text editor. The User’s Guide and GUI are designed for a fairly unsophisticated user (perhaps at the expense of the flexibility and control more experienced users might want). For example, they don’t provide instructions for (1) how to input different time activity data or (2) how to use non-US census population data or the 2010 census data (when it becomes available) or census block or block group data instead of census tract data.
- b) The SHED-PM model is nicely packaged and includes many design features needed for state of the science exposure assessments. Several shortcomings are worth noting.
 - 1) One of the problems with the current version of SHEDS is that the embedded CHAD database is outdated with respect to current activity patterns. Clear evidence of the strong temporal changes in activity patterns can be seen in comparison of 1981-82 with 2002-03 activity patterns Tables 16-49, 1650 of the Child-Specific Exposure Factors Handbook. The tables show the shift away from outdoor sports activity to indoor activities related to computers. This trend can be expected to be more pronounced now. The extent to which current estimates of exposure are biased due to this are unknown.

- 2) CHAD includes data from different studies and the current model framework does not allow the user to easily select the portions of the CHAD data base that may be suitable for a given application. The user's guide also does not indicate how the user would specify an alternate (non-CHAD) time-activity database for use in model calculations.
 - 3) Another problem is that the MET assignments do not reflect the full range of conversion data that are in the literature. Use of kcal overestimates oxygen utilization, since it includes body fat in the calculation. Ideally, estimates should be based on lean body mass. If such data are not available, then estimates of lean body mass for BMI along with error distributions should be provided. Users should be allowed to specify inputs provided that the following criteria are met:
 - i. If the data are published, a citation needs to be provided.
 - ii. If data are unpublished, they must be available to the public
 - iii. At a minimum the data should be specific to age, sex
 - iv. Estimates of error distributions need to be provided
 - 4) More attention needs to be given the basis for selection of parameters for estimating microenvironmental concentrations. Care should be taken to carefully select the parameters given as the default values or sample problem values because these will likely be used without evaluation by many potential users. It is important to explain the process and types of data needed to select the parameters for various types of applications (and regions). In fact, there is probably a need for a companion document on exposure modeling, that provides scientific guidance and tutorials.
- b) None of the outputs were obviously inconsistent with expectation. Determination of whether they are technically correct is very difficult from the stochastic simulations.

Clifford P. Weisel

- a) Overall, the GUI interface was easy to use and allow for easy visualization of individual patterns across concentrations, exposure and dose of ambient and non-ambient sources as well as across different microenvironments. I found the ability to compare different microenvironments most useful if I wanted to better understand where exposures were occurring and how the exposures and times spent in different microenvironments varies across age, gender, season employment status, day type and smoking. The scatter plots provide some insight into underlying associations between different exposures. A mechanism to plot distributions of ratios directly of different outcomes and variables to complement the scatter plots might be worth considering.
- b) The User Guide is clear in its description of the modeling algorithms used and the combination the multiple microenvironments using the CHAD data base along with microenvironmental air concentration to generate distributions of exposure that include uncertainty estimates represent current state of science for performing exposure assessment. The inclusion of Mass Balance for estimating air concentrations in the indoor environment is a strong advance that potentially increases the potential to make the model output region specific if appropriate input factors are available.

One item that is not clear to me is the assignment of the non-ambient contributions to concentrations, exposure and dose. On page 128 it indicates for the linear regression equation and scaling factor approaches that if the $C_i/C_{\text{ambient}} < 1$ then the "non-ambient" is assumed to be zero. For particles there will be losses as the particles enter the buildings (as discussed in the mass-balance equation approach which will result in that ratio being less than unity in most conditions. While it is recognized that some of the microenvironments do not have enough data to employ the mass balance model, it may still be possible to provide a better estimate of the C_i/C_{ambient} ratio than unity for assigning a non-ambient component to the PM concentration and therefore exposure and dose than when using the current assumption.

- c) The model outputs appear to be technically correct and consistent with the algorithms in the User Guide. The example provided showed the location of the monitors, but they are not in the figures from my runs.

Question No. 7

Rank Priority for Possible Future Improvements

Several possible improvements to the SHEDS-PM model are listed below. Please provide a number ranking for the relative priority that should be given to each improvement, using a scale from 1 (low priority) to 5 (high priority).

Improving ease of use:

- ☐ Create log file that records all inputs specified for the model run that can be viewed and saved by user
- ☐ Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)
- ☐ Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)
- ☐ Provide more information on error messages to help users identify the reason for the error for common problems
- ☐ Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)

Allowing additional user specification of inputs:

- ☐ Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)
- ☐ Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)

Improving/refining model algorithms:

- ☐ Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of "unique" diary being used repeatedly for an individual)
- ☐ Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)
- ☐ Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)
- ☐ Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)
- ☐ Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

Adding new functionality:

- ☐ Option for using mapping tool to select census tracts for simulation based on a map
- ☐ Add more user options to map view of output (e.g. for use in GIS software or Google Earth)

Other:

- ☐ Please describe: _____

Results for Question No. 7						
Reviewer Ranking						Possible improvement
AR	PBR	HS	IT/FL	CW	Average	
Improving ease of use:						
3	5 ^m	4	5 ^a	4	4.2	Create log file that records all inputs specified for the model run that can be viewed and saved by user
5	5 ⁿ	3	5/2 ^b	5	4.3	Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)
5	4 ^o	3	5 ^c	1	3.6	Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)
1	3 ^p	5	3 ^d	3	3	Provide more information on error messages to help users identify the reason for the error for common problems
3	1 ^q	4	5 ^e	2	3	Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)
Allowing additional user specification of inputs:						
3	3	4	4 ^f	3	3.4	Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)
3	3 ^r	4	3 ^g	4	3.4	Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)
Improving/refining model algorithms:						
5	3 ^s	3	3 ^h	4	3.6	Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of “unique” diary being used repeatedly for an individual)
5	3 ^t	4	2 ⁱ	3	3.4	Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)
2	2 ^u	2	5 ^j	1	2.4	Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)
4	2 ^v	2	5 ^k	2	3	Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)
5	2 ^w	4	5 ^l	5	4.2	Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

Results for Question No. 7						
Reviewer Ranking						Possible improvement
AR	PBR	HS	IT/FL	CW	Average	
Adding new functionality:						
2	2 ^x	2	1	2	1.8	Option for using mapping tool to select census tracts for simulation based on a map
3	4 ^y	2	5	4	3.6	Add more user options to map view of output (e.g. for use in GIS software or Google Earth
Other (Reviewer Specified):						
NA	NA	5	NA	NA	NA	An important model improvement would be to allow the user to import measured time/activity or microenvironmental concentration databases for use in model calculations. These measured data would reduce uncertainty in estimated exposure distributions. In addition, the model would be enhanced if the infiltration factors for the different microenvironments could vary by season. Given the sometimes long model run times, the ability to perform preliminary or crude exposure assessments (possibly by using fixed values for certain steps) may be important to allow the user to compare among different model options and to decide final model run parameters.
NA	NA	NA	3	NA	NA	EPA should consider making the modeling system open source to encourage innovation and testing of new algorithms. This would also provide transparency that can enhance its credibility. Going open source could help build a community of knowledgeable developers and users that could expand the software platform to other pollutants and regions, and subject the software to more testing.
5	NA	NA	NA	NA	NA	It appears that there is not an option to artificially bound open-ended parametric distributions (e.g., normal), If this is correct, adding such an option should be considered, to avoid unrealistic selections.
5	NA	NA	NA	NA	NA	Allowing the user to specify the re-sampling frequency for dairies and ME factors should be considered, instead of hard-wiring the model to re-sample dairies seasonally and ME factors daily.
5	NA	NA	NA	NA	NA	If a more sophisticated algorithm for combining activity dairies from CHAD in longitudinal simulations is added, the Cluster-Markov algorithm used in HAPEM and in a special version of APEX should be considered. The Cluster-Markov algorithm samples dairies daily taking into account diary similarities and diary-to-diary transition probabilities.
4	NA	NA	NA	NA	NA	Incorporating consideration of tract-specific commuting time distributions, available from US Census data, should be considered.
4	NA	NA	NA	NA	NA	Upgrading the mass balance algorithm to be dynamic (i.e., allow carryover from one time period to the next) instead of equilibrium should be considered.

Footnotes: AR = Arlene Rosenbaum, PBR = P. Barry Ryan, HS = Helen Suh, IT/FL = Ira Tager and Frederick Lurmann, CW = Clifford Weisel

- a. It is not appropriate to release the model without having a log file that shows which inputs were used for a particular simulation because this is essential for quality assurance of individual simulations and, based on experience, crucial for large batches of simulations.
- b. Saving the user's setting can help with consistency in multiple runs (high priority). Most users will use other software for comparison of outputs from multiple runs so this is a low priority.
- c. This should be easy to implement and worthwhile given that the model's long run times.
- d. We did not test the model enough to encounter errors so it is difficult to evaluate this option.
- e. Perhaps include options for age of housing stock and frequency of window openings and air conditioner use.
- f. Allow user specification from GUI or input file or database.
- g. This feature is scientifically desirable but only useful if studies are conducted to collect and analyze sufficient supporting data for credible specification of these parameters in different types of applications.
- h. While this would be desirable, it would be justifiable only if we had more data on true longitudinal activity data—i.e., the relation between any given day's activity to any other day corrected for season, age, sex. While newer methods for assignment of activity take in to account autocorrelation in activity patterns, the databases for estimation are generally quite small (e.g. only 163 children from southern California for whom 48 observations/child are available in Glenn, G., et al. JESEE, 2008). Currently available longitudinal database cannot be assumed to represent the broad spectrum of subjects (children and adults) and the myriad environments in which they carry out outdoor activities.
- i. Absent data from more subjects from different climates with longer time series of activities, it is not clear that there is any benefit from increased sophistication.
- j. Given the data present in Özkaynak, et al. (Figure 5, Atmos Environ 2009), this would be an absolute necessity. These data show considerable uncertainty over the percentiles of exposure such that any dose estimates, independent of the uncertainties and variability of the estimates on their own, are suspect from the start.
- k. At least one recent publication (Wu et al 2009 Atmos Environ 43, 1962–1971.) suggest census block groups or blocks are needed to capture the extremes of the exposure distributions for traffic related PM.
- l. This is a high priority because published data indicate window position and air conditioning use, both of which are related to temperature, as well as building age have large influences on residential air exchange rates.
- m. I believe that this could be implemented easily and has a great deal of utility in software QA. Hence I place a high priority on it.
- n. Notwithstanding the database storage requirements, this should be a high priority as well. I did find it frustrating not to be able to return to a different scenario and retest something I discovered in a later version.
- o. This is a high priority, but not as high as the first two. Speed of calculation is important, however, and this could help tremendously in this regard.
- p. Again, important, but of lower priority. At least an error number could be implemented and printed out, with a table to identify the error type.
- q. This is far less important in my view than any of the others.
- r. Both of these are of interest, but would likely require a substantial amount of work. Requiring mass-balance among a number of compartments is problematic and leads to restriction on input. Physiological parameters may be of higher priority, but are further down the road.
- s. I put this at a mid-level of priority. Including more variability on the diaries is generally good, but it is not clear if such data exist. An alternative strategy is sampling without replacement from the current list to ensure that the same "unique" diary is not used over and over.
- t. I put this at mid-level priority in that it may be hard to do and would require a lot of information that may not be readily available.
- u. I am not even sure I understand what would have to be done, much less the degree of difficulty for implementation. A low priority.
- v. I put this at lower priority because it increases the scale of the data input size and likely slows down the calculation process substantially. As it already takes a while for these simulations to run, making them more detailed may not be a great use of resources.
- w. This is, once again, a lot of work for not so much benefit and is, therefore, a lower priority.
- x. I find this to be of low priority because of the need for very sophisticated data at the map site. This is unlikely to occur frequently as it costs a lot of money to generate the data.
- y. Using a GIS version of an aerial view instead of the census tract maps would add more interest and should be relatively easily done.

Question No. 8 (optional)
Open Comments

Please provide any additional comments that you wish to on the SHEDS-PM model.

Arlene S. Rosenbaum

No additional comments provided.

P. Barry Ryan

No additional comments provided.

Helen H. Suh

No additional comments provided.

Ira B. Tager/ Frederick W. Lurmann

No additional comments provided.

Clifford P. Weisel

No additional comments provided.

VI. SPECIFIC COMMENTS

Arlene S. Rosenbaum

Page 126: (corrections in bold)

“ If **When** the activity is preparation of food, ~~and~~ **if** the diary event has a ‘Y’ for gas stove use during the event, then the total duration of the diary event is used for t_{cook} . Otherwise, a factor is randomly generated to account for food preparation activities that do not generate PM. The factor is a random number between 0 and 1.”

Page 130: (corrections in bold)

$$IDose_{ij} = C_i \cdot Ve_{nj} \cdot t_{ij} / 1000$$

where:

$IDose_{ij}$ = inhaled PM dose during activity j while in microenvironment i (μg)
 C_i = PM concentration for microenvironment i ($\mu\text{g}/\text{m}^3$)
 Ve_{nj} = exhaled ventilation rate for individual n during activity j ($\text{L}_{\text{air}}/\text{min}$)
 t_{ij} = duration of activity j while in microenvironment i (min)

Page 130

$$VO_{2nj} = METS_j \cdot BMR_n \cdot EE_{to} VO_{2n} / BM_n$$

This equation and its subsequent development on page 131 are unclear, especially with respect to the measurement units. The measurement units of each term should be presented immediately following the equation.

P. Barry Ryan

I have placed these comments inline above.

Helen H. Suh

None.

Ira B. Tager/ Frederick W. Lurmann

None.

Clifford P. Weisel

I did encounter a problem when running the program for the longer time period (overnight) in that my computers, as is the case for many, are scheduled to do updates of windows and other resident programs during the night. On both computers one of the updates required an automatic restart of the computer. This resulted in a loss of the results obtained from runs, which for a run that takes hours can be at least an annoyance. I therefore had to turn off the scheduled update options on my computer when running the 24+hour runs so as not to lose the results prior to my review of the analysis results. I suggest this be indicated in the installation section AND in other parts of the manual unless it can be fixed.

The Push Button for “View/Edit Model Run Inputs” is some times listed as “Edit/View Model Run Inputs” in the text (e.g. page 27 and 28) rather than “View/Edit Model Run Inputs” as it appears on the screen.

APPENDIX 1:
MODEL SPECIFICATION DETAILS FOR SCENARIO RUNS

Overview

Instructions for the example model run scenarios to be performed are provided below. For each scenario, the specific goals and approach for the model run are provided along with an approximate run time. The model input specifications for the GUI are listed, along with steps for obtaining and analyzing model results. Initial runs will take only a few minutes to execute the run and analyze results. Later runs may take several hours to generate output.

Because SHEDS-PM is a stochastic model, different results will be obtained with each model run using the same input specifications. To produce the same results, the seed for random number generation can be set, so the same sequence of random numbers is produced. If the seed is fixed, the same individual characteristics and the same CHAD IDs are assigned, and the same exposure and dose outputs are produced when the same input specifications are used for the model runs. The reviewer may want to set the seed if working with the model on different computers to obtain the same output. Instructions for setting the random seed are location in Section 5 of the SHEDS-PM User Guide (page 92).

Daily $PM_{2.5}$ concentrations from a monitoring site in Detroit, MI during 2005 are provided in the file 'Detroit Daily PM_{25} 2005 (Linwood).xls'. The location of the Linwood monitoring site is shown in Figure 1. The time series of $PM_{2.5}$ concentrations is shown in Figure 2. Initial scenario runs will use the census tract where the monitoring site is located (Census Tract ID #26163522300), and then expand to include the surrounding census tracts as shown in Figure 3.

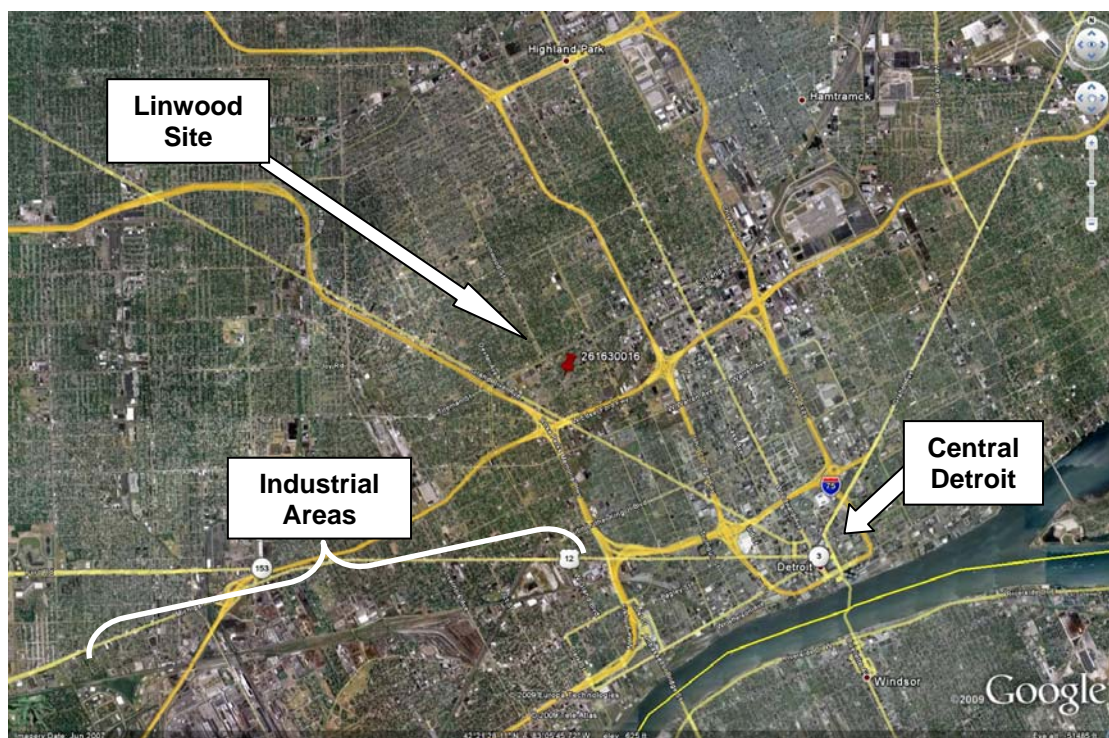


Figure 1. Location of Linwood $PM_{2.5}$ Monitoring Site in Detroit, MI.

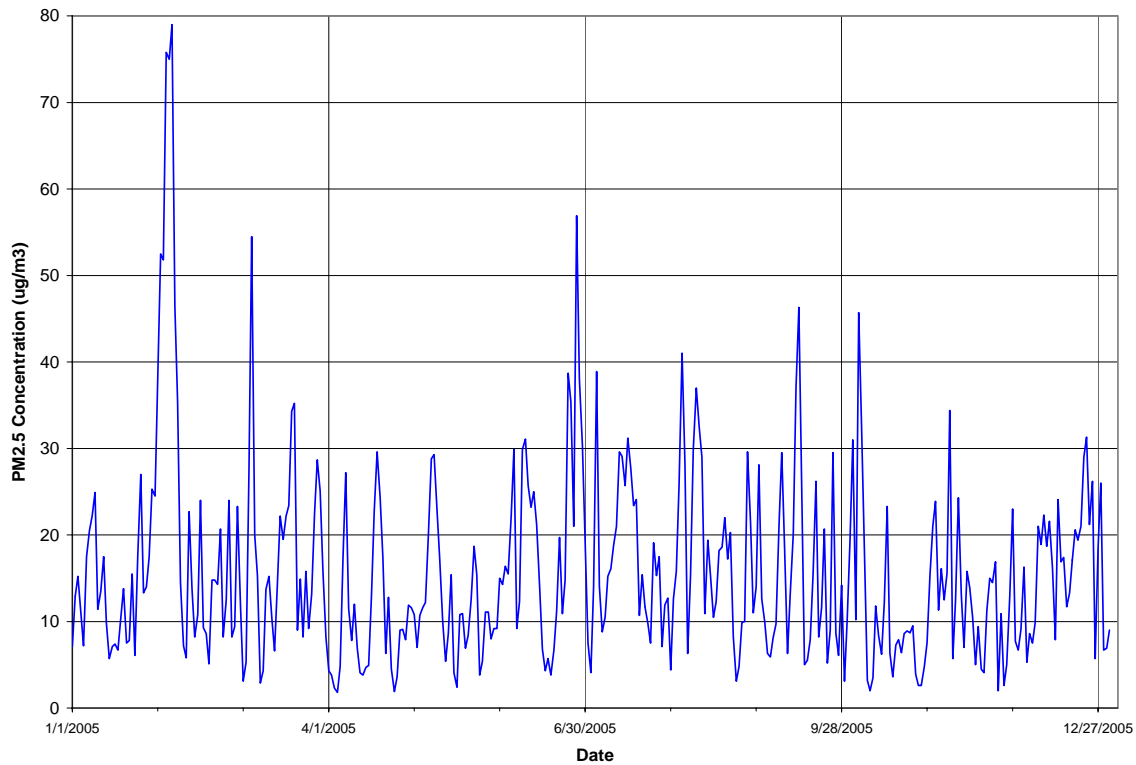


Figure 2. Daily PM_{2.5} Concentrations (µg/m³) at Linwood Site in Detroit, MI during 2005.

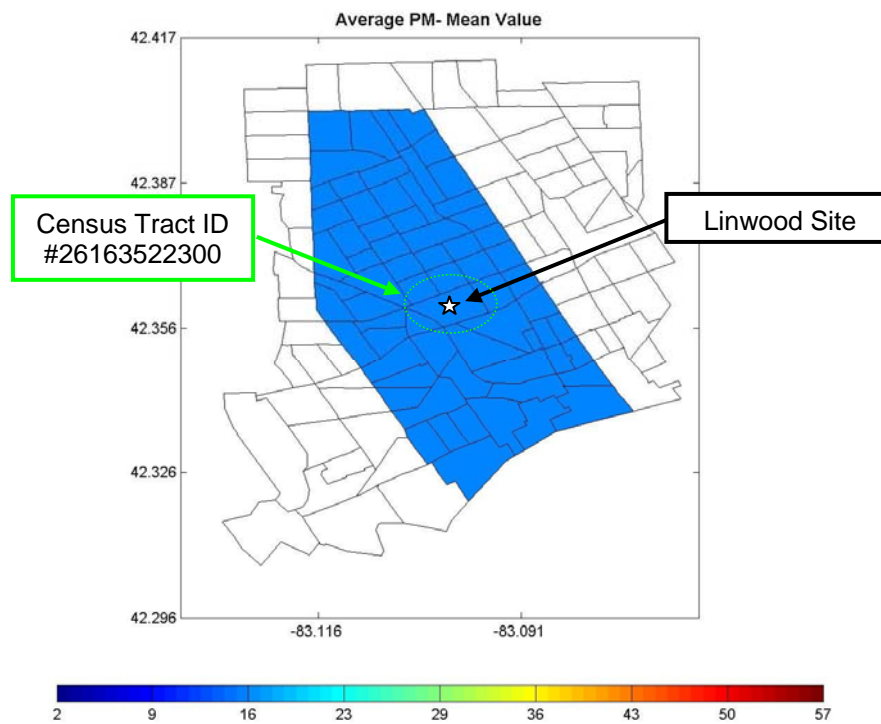


Figure 3. SHEDS-PM Map View of Census Tracts for Scenario Displaying Average Daily PM_{2.5} Concentration (µg/m³)

The US Census input data for Census Tract ID# 26163522300 have been extracted from the SHEDS-PM input database for comparison with the model output from Scenario #1. Demographic proportions for gender and age are provided in Table 1, and employment proportions (for 16 yrs old and older) are provided in Table 2.

Table 1. Gender/Age Demographic Proportions from SHEDS-PM US Census Database

	Male	Female								
All	0.459	0.541								
Age	Male	Female		Age	Male	Female		Age	Male	Female
0	0.005	0.006		40	0.012	0.010		80	0.003	0.004
1	0.008	0.007		41	0.009	0.005		81	0.001	0.006
2	0.007	0.007		42	0.007	0.009		82	0.001	0.003
3	0.007	0.006		43	0.006	0.008		83	0.004	0.005
4	0.005	0.005		44	0.009	0.006		84	0.001	0.003
5	0.008	0.009		45	0.009	0.010		85	0.002	0.002
6	0.009	0.008		46	0.007	0.005		86	0	0.002
7	0.007	0.007		47	0.005	0.006		87	0	0.002
8	0.007	0.010		48	0.008	0.006		88	0	0.002
9	0.009	0.009		49	0.008	0.009		89	0.001	0.001
10	0.009	0.016		50	0.008	0.003		90	0	0.001
11	0.007	0.011		51	0.006	0.006		91	0.001	0.002
12	0.008	0.006		52	0.006	0.005		92	0.002	0
13	0.008	0.006		53	0.005	0.003		93	0	0
14	0.008	0.005		54	0.003	0.005		94	0	0.001
15	0.005	0.006		55	0.004	0.008		95	0	0
16	0.003	0.006		56	0.005	0.003		96	0	0
17	0.005	0.007		57	0.004	0.003		97	0	0
18	0.005	0.008		58	0.005	0.006		98	0	0
19	0.006	0.006		59	0.003	0.003		99	0	0
20	0.005	0.009		60	0.004	0.005		100	0	0
21	0.005	0.007		61	0.004	0.003		101	0	0
22	0.006	0.008		62	0.005	0.006		102	0	0
23	0.008	0.010		63	0.002	0.005				
24	0.004	0.006		64	0.003	0.005				
25	0.006	0.006		65	0.003	0.005				
26	0.006	0.007		66	0.003	0.004				
27	0.003	0.007		67	0.002	0.005				
28	0.005	0.006		68	0.004	0.005				
29	0.003	0.006		69	0.004	0.003				
30	0.005	0.005		70	0.006	0.007				
31	0.006	0.008		71	0.003	0.006				
32	0.006	0.006		72	0.005	0.008				
33	0.008	0.009		73	0.004	0.005				
34	0.011	0.008		74	0.003	0.003				
35	0.007	0.005		75	0.003	0.003				
36	0.006	0.009		76	0.004	0.005				
37	0.006	0.007		77	0.002	0.006				
38	0.004	0.008		78	0.004	0.006				
39	0.004	0.005	79	0.004	0.007					

Table 2. Employment Proportions from SHEDS-PM US Census Database (age 16 yrs & older)

	Total		Male	Female
Unemployed	0.624		0.435	0.565
Employed	0.376		0.496	0.504
			Male	Female
Unemployed			0.593	0.650
Employed			0.407	0.350

	Age	Male	Female
Unemployed	16 to 19	0.030	0.084
	20 to 21	0.036	0
	22 to 24	0.056	0.063
	25 to 29	0.052	0.022
	30 to 34	0.011	0.037
	35 to 44	0.148	0.098
	45 to 54	0.200	0.118
	55 to 59	0.039	0.036
	60 to 61	0.009	0.029
	62 to 64	0.071	0.097
	65 to 69	0.107	0.053
	70 to 74	0.082	0.121
	75 plus	0.159	0.242
Employed	16 to 19	0.163	0.040
	20 to 21	0	0.059
	22 to 24	0.044	0.094
	25 to 29	0.076	0.131
	30 to 34	0.204	0.209
	35 to 44	0.292	0.311
	45 to 54	0.128	0.072
	55 to 59	0.049	0
	60 to 61	0	0
	62 to 64	0	0.029
	65 to 69	0	0.024
	70 to 74	0.044	0.029

Note: The reviewer can choose other census tracts for this comparison, but must calculate the proportions from the data in the SHEDS-PM US Census database for those tracts (see important information in Appendix C of User Guide before opening the database).

The model runs for the scenarios below test the key elements of the model and a limited number of options. The SHEDS-PM model currently has many options and features that can be explored in addition to those specified. The reviewer is encouraged to perform additional tests with different input specifications.

Scenario #1: Population Variability

This model scenario will utilize the PM_{2.5} concentration input file for Detroit described above in three different model runs. The first model run (#1-1) will generate a representative population from the census tract where the monitor is located (includes all ages and both genders) for comparison with the various input data sets (US Census demographics, CHAD activity diaries, PM concentration inputs). The second model run (#1-2) will focus on the specifications for the different microenvironments for comparison with the description of the algorithms for calculating PM concentrations in the microenvironments. The third model run (#1-3) will generate a simulation population representative for the area near the monitoring site and focus on the analysis of the model results for characterizing the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Model Run #1-1

Goals: Confirm population demographics assignment, activity diary assignment, and PM concentration merging

Approach: Cross-sectional simulation with 1,000 individuals simulated for the census tract where monitor is located (ID #26163522300); Microenvironment settings produce PM concentrations same as input data

Approx. Model Run Time: 5 minutes

GUI specifications:

- **Model Run Inputs GUI** (see Section 4.3 of User Guide for detailed instructions)
 - Input Data: Select 'PM Concentration' pushbutton and on PM Concentration File Information GUI:
 1. Select 'Excel Spreadsheet (*.xls)' option, and locate 'Detroit Daily PM25 2005 (Linwood).xls' file.
 2. Select '24-hour Average' option
 3. Select start hour as '12 AM'
 4. Select 'Monitor ID included', then 'No' for Monitor-Tract ID matching file
 5. Select 'No' for Temperature Included in File
 6. Leave particle size distribution options as defaults
 7. Leave missing value symbols as '0' (no missing data in file)
 - Simulation Type: Cross-sectional
 - Stage: 1-Stage(Variability)
 - Census Tracts: Select 'Census Tracts' pushbutton and on Census Tract Selection GUI select 'Michigan' for State, and 'Wayne County', then locate and choose census tract ID number '26163522300' from the list, click on the 'Linwood' monitor ID box and press 'OK'
 - Population: Select 'Fixed Value Per Tract' and enter '1000' for Individuals Per Tract
 - Day Type: Select 'Include Both'
 - Gender: Select 'Both Genders'
 - Age: Enter '0' and '102'
 - Seasons: Select 'Define Seasons' pushbutton and select 'Default' pushbutton
 - Activity Diary Match Criteria: Select 'Employment' and 'Season'
 - Commuting: none
- **Define Microenvironments GUI** (see Section 4.4 of User Guide for detailed instructions)
 - Leave as defaults (All Outdoor, All Indoor, All In-vehicle; Scaling factor, fixed value=1.0 for each)
- **Output Options GUI** (see Section 4.5 of User Guide for detailed instructions)
 - Leave as default (no 'Event Time Series Data')
- Select 'Run' pushbutton to perform the simulation for Model Run #1-1.

- When model run has finished, export the *daily* data through the Analyze Results GUI (see Section 4.7 of User Guide for detailed instructions).
- The daily data output file can be limited to the output variables shown below and collated into one Excel spreadsheet:

Comparison with input data:

The daily data output file (example shown below in Figure 4) contains the data needed to compare the model output with the input data and specifications as follows:

- Compare frequency statistics for ‘Gender’, ‘Age’, and ‘Employment_Status’ columns with the SHEDS-PM US Census database proportions for census tract ID #26163522300 provided above in Tables 1 and 2 for gender/age and employment, respectively.

Note:

- Only individuals age 16 or older from the SHEDS output should be used to compare with employment proportions in Table 2.
- Compare ‘Season_Number’ with month for each ‘PM_date’
- Confirm that different CHAD activity diaries are assigned to each simulated individual

Figure 4. Example SHEDS-PM Daily Data Output File for Model Run #1-1.

- Confirm that the CHAD diaries are correctly assigned by comparing 'Associated_CHAD_ID' and diary matching criteria (gender/age employment, season) with data for CHAD ID in 'Data' table of CHAD database (see important information in Appendix B before opening CHAD database).

Note:

- Age in CHAD database for CHAD ID should be from the same age group as the simulated individual (see page 117 in Appendix D of User Guide for age group definitions)
- Season should match the Month column for the CHAD diary, and day type should match the Weekday column for the CHAD diary (if not missing)
- Employment status should match the Employed column in the CHAD diary (if not missing)
- Compare 'Average_Input_PM_Concentration' data to PM concentration for 'PM_Date' in the input file 'Detroit Daily PM25 2005 (Linwood).xls'
- Confirm that microenvironment calculation scheme 'Scaling Factor' defaults produce microenvironmental concentrations and exposures equivalent to the input PM concentration (e.g. 'Average_Total_PM_Concentration' and 'Total_Exposure' are the same as 'Average_Input_PM_Concentration')

Model Run #1-2

Goals: Confirm microenvironmental PM concentration calculations, activity diary and PM concentration data merging for longitudinal runs, and exposure/dose calculations

Approach: Longitudinal simulation with 10 individuals simulated for the census tract where monitor is located (ID #26163522300); Test different microenvironment calculation schemes and inputs

Approx. Model Run Time: 5 minutes per test run

GUI specifications:

- **Model Run Inputs GUI**
 - Input Data: Same as Model Run #1-1 above
 - Simulation Type: Longitudinal
 - Stage: 1-Stage(Variability)
 - Census Tracts: Same as Model Run #1-1 above
 - Population: Select 'Fixed Value Per Tract' and enter '10' for Individuals Per Tract
 - Gender: Select 'Both Genders'
 - Age: Enter '0' and '102'
 - Seasons: Select 'Define Seasons' pushbutton and select 'Default' pushbutton
 - Activity Diary Match Criteria: Select 'Employment' and 'Season'
 - Commuting: none
- **Define Microenvironments GUI** (see Section 4.4 of User Guide for detailed instructions)
 - All Outdoor: Leave as defaults (Scaling factor, fixed value=1.0)
 - All Indoor: Select Scaling Factor and change distribution type to 'Normal'; enter Mean=0.6 and Std Dev=0.1
 - All In-vehicle: Select Scaling Factor and change distribution type to 'Uniform'; enter Min=1 and Max=1.2
- **Output Options GUI**
 - Select 'Event Time Series Data'

- Select 'Run' pushbutton to perform the simulation for Model Run #1-2.
- When model run has finished, export the *daily* data through the Analyze Results GUI. The daily data output file can be limited to the output variables shown below (and not collated) for the check of microenvironmental calculations:

Individual Characteristics:	Daily Characteristics	Daily Totals	Microenvironment-Specific Totals
<input checked="" type="checkbox"/> Individual Number	<input checked="" type="checkbox"/> Associated CHAD ID	<input checked="" type="checkbox"/> Mean PM Concentration	<input checked="" type="checkbox"/> Duration
<input type="checkbox"/> Uncertainty Number	<input checked="" type="checkbox"/> Date	<input type="checkbox"/> Exposure	<input checked="" type="checkbox"/> Mean PM Concentration
<input checked="" type="checkbox"/> Age	<input checked="" type="checkbox"/> Season	<input type="checkbox"/> Intake Dose	<input type="checkbox"/> Exposure
<input checked="" type="checkbox"/> Gender	<input checked="" type="checkbox"/> Day type	<input type="checkbox"/> Deposited Dose	<input type="checkbox"/> Intake Dose
<input checked="" type="checkbox"/> Tract ID	<input checked="" type="checkbox"/> Julian Date (Year-to-Date number)		<input type="checkbox"/> Deposited Dose
<input type="checkbox"/> Smoking Status	<input type="checkbox"/> Number of Cigs (by individual)	Output File Type: <input checked="" type="radio"/> Excel Spreadsheet <input type="radio"/> Comma-separated File (.csv)	
<input type="checkbox"/> ETS Exposure in Residence	<input type="checkbox"/> Number of Cigs (by others)	Filename & Location C:\Program Files\EPA SHEDS-PM 3.5\Results\Daily Data Export 5223 long 10 Micro Conc.xls	
<input checked="" type="checkbox"/> Employment Status	<input checked="" type="checkbox"/> Average PM Concentration		
<input type="checkbox"/> Work Tract	<input checked="" type="checkbox"/> Air Exchange Rate		
<input type="checkbox"/> Collate Individual Characteristics with Person-Day Details			

Comparison with input data:

The daily data output file (example shown below in Figure 5) contains the data needed to compare the model output with the input data and specifications as follows:

- Confirm that the CHAD activity diaries are assigned to each simulated individual for this longitudinal simulation as follows:
 - 3 different CHAD diaries for a season (weekday, Saturday, Sunday)
 - A different set of 3 CHAD diaries are assigned for each season
 - Confirm that the longitudinal assignment of PM concentration data included 365 days for each simulated individual, and the PM concentration data were assigned correctly by comparing 'Average_Input_PM_Concentration' data for each date to PM concentration data in the input file 'Detroit Daily PM25 2005 (Linwood).xls'

Comparison with microenvironment input distributions:

The daily data output file (example shown below in Figure 5) contains the data needed to compare the model output with the microenvironment input distributions as follows:

- Confirm that the 'Mean_Ambient_Concentration_in_All_Outdoor' column is the same as the 'Average_Input_PM_Concentration' column for dates when 'Time_Spent_in All_Outdoor' column contains a value. Scatter plot shown in Figure 6 below.
- Confirm that the 'Mean_Ambient_Concentration_in_All_Indoor' column is less than the 'Average_Input_PM_Concentration' column for dates when 'Time_Spent_in All_Indoor' column contains a value. Confirm that ratio of columns changes for each date (sampled from distribution for each date). Scatter plot shown in Figure 6 below.
- Confirm that the 'Mean_Ambient_Concentration_in_All_In-Vehicle' column is greater than the 'Average_Input_PM_Concentration' column for dates when 'Time_Spent_in In-Vehicle' column contains a value. Confirm that ratio of columns changes for each date (sampled from distribution for each date). Scatter plot shown in Figure 6 below.

Individual Number	Associated CHAD ID	PM Date	Season	Day Type	Julian Date	Average Input PM Concentration	Air Exchange Rate	Average Total PM Concentration	Average Ambient PM Concentration	Average Non-Ambient PM Concentration	Time Spent in All Outdoor	Time Spent in All Indoor	Time Spent in All In-Vehicle	Mean Ambient Concentration in All Outdoor	Mean Ambient Concentration in All Indoor	Mean Ambient Concentration in All In-Vehicle	Mean Non-Ambient Concentration in All In-Vehicle
1	CAC06366A	01/01/05	1	Saturday	1	7.100		4.739	4.739	0.000	60	1380		7.100	4.612		0.000
2	CAC06524A	01/02/05	1	Sunday	2	12.900		9.797	9.797	0.000	390	1050		12.900	8.008		0.000
3	NHA11374A	01/03/05	1	Weekday	3	15.200		11.366	11.366	0.000	60	1295	85	15.200	10.770	15.620	0.000
4	NHA11374A	01/04/05	1	Weekday	4	11.100		8.076	8.076	0.000	60	1295	85	11.100	7.415	13.110	0.000
5	NHA11374A	01/05/05	1	Weekday	5	7.200		5.147	5.147	0.000	60	1295	85	7.200	4.727	8.312	0.000
6	NHA11374A	01/06/05	1	Weekday	6	17.500		9.700	9.700	0.000	60	1295	85	17.500	8.260	20.354	0.000
7	NHA11374A	01/07/05	1	Weekday	7	20.500		15.475	15.475	0.000	60	1295	85	20.500	14.551	22.297	0.000
8	CAC06366A	01/08/05	1	Saturday	8	22.200		13.659	13.659	0.000	60	1380		22.200	13.197		0.000
9	CAC06524A	01/09/05	1	Sunday	9	24.900		18.518	18.518	0.000	390	1050		24.900	14.836		0.000
10	NHA11374A	01/10/05	1	Weekday	10	11.400		7.642	7.642	0.000	60	1295	85	11.400	7.042	11.953	0.000
11	NHA11374A	01/11/05	1	Weekday	11	13.500		8.281	8.281	0.000	60	1295	85	13.500	7.246	16.033	0.000
12	NHA11374A	01/12/05	1	Weekday	12	17.500		9.136	9.136	0.000	60	1295	85	17.500	7.906	17.811	0.000
13	NHA11374A	01/13/05	1	Weekday	13	9.800		6.115	6.115	0.000	60	1295	85	9.800	5.384	11.591	0.000
14	NHA11374A	01/14/05	1	Weekday	14	5.700		3.559	3.559	0.000	60	1295	85	5.700	3.225	5.946	0.000
15	CAC06366A	01/15/05	1	Saturday	15	7.100		4.093	4.093	0.000	60	1380		7.100	3.931		0.000
16	CAC06524A	01/16/05	1	Sunday	16	7.400		6.358	6.358	0.000	390	1050		7.400	5.758		0.000
17	NHA11374A	01/17/05	1	Weekday	17	6.700		3.987	3.987	0.000	60	1295	85	6.700	3.502	7.559	0.000
18	NHA11374A	01/18/05	1	Weekday	18	10.200		7.563	7.563	0.000	60	1295	85	10.200	7.131	10.681	0.000
19	NHA11374A	01/19/05	1	Weekday	19	13.800		9.914	9.914	0.000	60	1295	85	13.800	9.078	16.251	0.000
20	NHA11374A	01/20/05	1	Weekday	20	7.500		3.684	3.684	0.000	60	1295	85	7.500	3.047	8.298	0.000
21	NHA11374A	01/21/05	1	Weekday	21	7.800		4.833	4.833	0.000	60	1295	85	7.800	4.305	8.711	0.000
22	CAC06366A	01/22/05	1	Saturday	22	15.500		10.153	10.153	0.000	60	1380		15.500	9.864		0.000
23	CAC06524A	01/23/05	1	Sunday	23	6.100		4.826	4.826	0.000	390	1050		6.100	4.092		0.000
24	NHA11374A	01/24/05	1	Weekday	24	17.600		9.780	9.780	0.000	60	1295	85	17.600	8.467	19.316	0.000
25	NHA11374A	01/25/05	1	Weekday	25	27.000		18.931	18.931	0.000	60	1295	85	27.000	17.529	29.188	0.000
26	NHA11374A	01/26/05	1	Weekday	26	13.300		7.588	7.588	0.000	60	1295	85	13.300	6.479	15.871	0.000
27	NHA11374A	01/27/05	1	Weekday	27	14.000		9.226	9.226	0.000	60	1295	85	14.000	8.357	15.629	0.000
28	NHA11374A	01/28/05	1	Weekday	28	17.400		10.856	10.856	0.000	60	1295	85	17.400	9.884	17.728	0.000
29	CAC06366A	01/29/05	1	Saturday	29	25.300		15.287	15.287	0.000	60	1380		25.300	14.746		0.000
30	CAC06524A	01/30/05	1	Sunday	30	24.500		18.302	18.302	0.000	390	1050		24.500	14.727		0.000
31	NHA11374A	01/31/05	1	Weekday	31	38.200		26.863	26.863	0.000	60	1295	85	38.200	24.541	42.339	0.000
32	NHA11374A	02/01/05	1	Weekday	32	52.500		42.233	42.233	0.000	60	1295	85	52.500	39.708	61.760	0.000
33	NHA11374A	02/02/05	1	Weekday	33	51.800		30.224	30.224	0.000	60	1295	85	51.800	26.648	56.117	0.000
34	NHA11374A	02/03/05	1	Weekday	34	75.800		50.560	50.560	0.000	60	1295	85	75.800	46.126	83.047	0.000

Figure 5. Example SHEDS-PM Daily Data Output File for Model Run #1-2.

- Calculate the ratio of the 'Mean_Ambient_Concentration...' columns to the 'Average_Input_PM_Concentration' column for dates when 'Time_Spent_in...' columns contains a value (Excel formula with 'if' condition such as =IF(O3>0,O3/G3,"")) for column O and row 3 can be used). Compare summary statistics (mean, standard deviation, minimum, maximum) to input distributions specified for run.

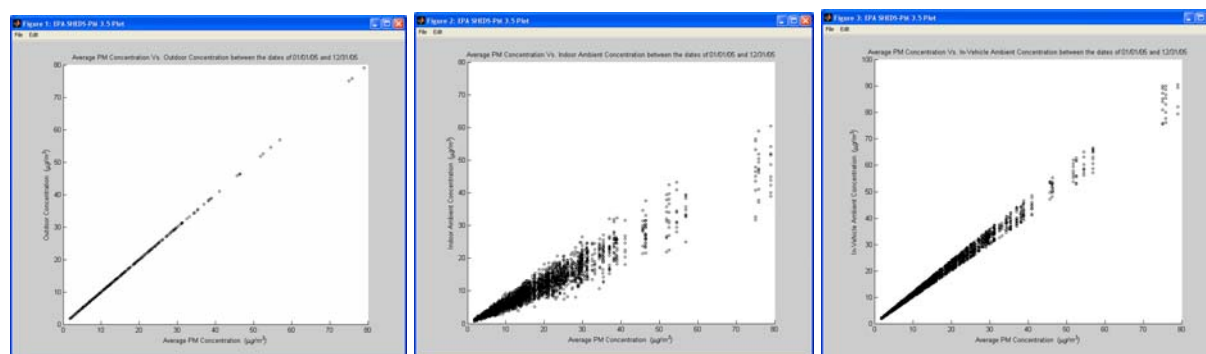


Figure 6. SHEDS-PM Scatter Plots of Microenvironmental PM Concentration vs. Input PM Concentration for Model Run #1-2 Input Distributions.

Comparison with exposure/dose calculations:

Export the *event* data for one simulated individual through the Analyze Results GUI (see instructions on pages 70-71 of User Guide for selecting data for one individual). The event data output file (example shown below in Figure 7) contains the data needed to compare the model output with the exposure and dose calculations described in Appendix D as follows:

- Each row in the event data output file corresponds to a diary record from the assigned CHAD ID with the 'Location', or SHEDS microenvironment, and time spent in the microenvironment. Confirm that the number of diary records matches the data for the CHAD ID in the 'Diary' table of the SHEDS-PM CHAD database, and that the SHEDS microenvironment codes are correctly assigned (see Table B1 of Appendix B).
- Each row in the event data output file contains the calculated microenvironmental PM concentration. Confirm that the microenvironmental PM concentrations are the same as in the daily data export file for this model run.
- Calculate PM exposure for dairy events using the time spent in the microenvironment (minutes), the calculated microenvironmental PM concentration, and the total averaging time in minutes (1440 minutes) according to the equation in Appendix D (page 129). Compare hand-calculated PM exposure with data for each diary record in event data file.
- Calculate PM intake dose using the ventilation rate (based on the METS value assigned) according to the equation in Appendix D (page 130). Compare hand-calculated PM intake dose with data for each diary record in event data file.
- Confirm the daily data are calculated from the event data by totaling all the diary records for the simulated individual for PM exposure and dose.

IndivNumber	Tract_ID	Date	Location	Time_in_Microenv (minutes)	Input_PM_Concentration (mu_g/m3)	Total_Micro-ambient PM_Concentration (mu_g/m3)	Non-ambient PM_Concentration (mu_g/m3)	Total_Exp (mu_g/m3)	Ambient_Exp (mu_g/m3)	Non-ambient_Exp (mu_g/m3)	Vent_Rate	METS	Total_Intake Dose (mu_g)	Ambient_Intake Dose (mu_g)	Non-ambient_Intake Dose (mu_g)	Total_Deposited Dose (mu_g)	Ambient_Deposited Dose (mu_g)	Non-ambient_Deposited Dose (mu_g)
1	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	3.596139	1.03634	0.995093	0.995093	0.995093	0.860373	0.860373	0.860373
2	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	3.547886	1.001	0.981889	0.981889	0.981889	0.848043	0.848043	0.848043
3	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	4.695077	1.001	1.299102	1.299102	1.299102	1.140591	1.140591	1.140591
4	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	4.252647	1.001	1.175756	1.175756	1.175756	1.027716	1.027716	1.027716
5	26163522300	01/01/05	All Indoor	30	7.1	4.611857	4.611857	0.09608	0.09608	0.09608	4.081912	1.001	0.596139	0.596139	0.596139	0.493369	0.493369	0.493369
6	26163522300	01/01/05	All Indoor	30	7.1	4.611857	4.611857	0.09608	0.09608	0.09608	6.085074	1.84007	0.841905	0.841905	0.841905	0.74765	0.74765	0.74765
7	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	3.787499	1.001	1.048044	1.048044	1.048044	0.909112	0.909112	0.909112
8	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	3.663623	1.001	1.013622	1.013622	1.013622	0.877606	0.877606	0.877606
9	26163522300	01/01/05	All Indoor	30	7.1	4.611857	4.611857	0.09608	0.09608	0.09608	4.141367	1.001	0.572902	0.572902	0.572902	0.499666	0.499666	0.499666
10	26163522300	01/01/05	All Indoor	30	7.1	4.611857	4.611857	0.09608	0.09608	0.09608	7.509565	1.69318	1.037745	1.037745	1.037745	0.928602	0.928602	0.928602
11	26163522300	01/01/05	All Indoor	30	7.1	4.611857	4.611857	0.09608	0.09608	0.09608	9.963952	2.34111	1.378957	1.378957	1.378957	1.246312	1.246312	1.246312
12	26163522300	01/01/05	All Indoor	15	7.1	4.611857	4.611857	0.04804	0.04804	0.04804	10.069350	2.5	0.751227	0.751227	0.751227	0.661113	0.661113	0.661113
13	26163522300	01/01/05	All Indoor	15	7.1	4.611857	4.611857	0.04804	0.04804	0.04804	3.760924	1.001	0.260173	0.260173	0.260173	0.225595	0.225595	0.225595
14	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	4.28919	1.001	1.187014	1.187014	1.187014	1.037173	1.037173	1.037173
15	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	3.336434	1.001	0.923229	0.923229	0.923229	0.794323	0.794323	0.794323
16	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	4.195881	1.02005	1.161048	1.161048	1.161048	1.013236	1.013236	1.013236
17	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	4.304893	1.04837	1.191213	1.191213	1.191213	1.041044	1.041044	1.041044
18	26163522300	01/01/05	All Indoor	5	7.1	4.611857	4.611857	0.016013	0.016013	0.016013	9.652212	2.5	0.222573	0.222573	0.222573	0.201009	0.201009	0.201009
19	26163522300	01/01/05	All Outdoor	55	7.1	7.1	7.1	0.271181	0.271181	0.271181	7.145462	1.59681	2.790303	2.790303	2.790303	2.492627	2.492627	2.492627
20	26163522300	01/01/05	All Outdoor	5	7.1	7.1	7.1	0.024653	0.024653	0.024653	8.86991	1.8671	0.314527	0.314527	0.314527	0.26304	0.26304	0.26304
21	26163522300	01/01/05	All Indoor	30	7.1	4.611857	4.611857	0.09608	0.09608	0.09608	5.924442	1.51302	0.81960	0.81960	0.81960	0.727146	0.727146	0.727146
22	26163522300	01/01/05	All Indoor	10	7.1	4.611857	4.611857	0.032027	0.032027	0.032027	6.490422	1.3296	0.299329	0.299329	0.299329	0.26547	0.26547	0.26547
23	26163522300	01/01/05	All Indoor	15	7.1	4.611857	4.611857	0.04804	0.04804	0.04804	9.170182	1.97896	0.834374	0.834374	0.834374	0.717153	0.717153	0.717153
24	26163522300	01/01/05	All Indoor	60	7.1	4.611857	4.611857	0.192161	0.192161	0.192161	10.435997	2.29010	2.887757	2.887757	2.887757	2.613166	2.613166	2.613166
25	26163522300	01/01/05	All Indoor	10	7.1	4.611857	4.611857	0.032027	0.032027	0.032027	9.026937	2.09999	0.416298	0.416298	0.416298	0.374859	0.374859	0.374859
26	26163522300	01/01/05	All Indoor	50	7.1	4.611857	4.611857	0.160134	0.160134	0.160134	3.734786	1.001	0.861215	0.861215	0.861215	0.746401	0.746401	0.746401
27	26163522300	01/01/05	All Indoor	45	7.1	4.611857	4.611857	0.144121	0.144121	0.144121	4.67093	1.001	0.969375	0.969375	0.969375	0.850022	0.850022	0.850022
28	26163522300	01/01/05	All Indoor	15	7.1	4.611857	4.611857	0.04804	0.04804	0.04804	6.320209	1.58689	0.437218	0.437218	0.437218	0.388836	0.388836	0.388836
29	26163522300	01/01/05	All Indoor	15	7.1	4.611857	4.611857	0.04804	0.04804	0.04804	7.569943	1.53399	0.522981	0.522981	0.522981	0.468104	0.468104	0.468104
30	26163522300	01/01/05	All Indoor	15	7.1	4.611857	4.611857	0.04804	0.04804	0.04804	0.252094	1.9559	0.570062	0.570062	0.570062	0.512406	0.512406	0.512406
31	26163522300	01/01/05	All Indoor	20	7.1	4.611857	4.611857	0.064054	0.064054	0.064054	6.500321	1.56367	0.599571	0.599571	0.599571	0.533784	0.533784	0.533784
32	26163522300	01/01/05	All Indoor	5	7.1	4.611857	4.611857	0.016013	0.016013	0.016013	7.85273	1.70491	0.181078	0.181078	0.181078	0.162288	0.162288	0.162288

Figure 7. Example SHEDS-PM Event Data Output File for Model Run #1-2.

Additional microenvironment calculation options:

Repeat this model run with different microenvironment concentration input selections and distributions, as follows:

- Compare output for Linear Regression calculation option (see Appendix D of User Guide for information on linear regression calculation on page 127)
 - Run model with linear regression option
 - Select linear regression calculation scheme for 'All Indoor'
 - Enter equation parameters (for example: slope=0.85, intercept=8, residual std. dev=4)
 - Run model and export daily data
 - Confirm linear regression calculation by estimating regression parameters (slope, intercept, residual distribution) using daily data. Compare to input parameters.
 - Confirm non-ambient PM contribution is calculated as described in Appendix D
- Compare output for Home Mass Balance calculation option (see Appendix D of User Guide for information on mass balance calculation on page 125)
 - Run model with mass balance option for Home microenvironment
 - Select mass balance calculation scheme for 'Home'
 - Select default input parameters on Mass Balance GUI
 - Run model and export daily data
 - Confirm air exchange rates are different for each date. Calculate summary statistics for air exchange rate, compare to input distribution for each season. Confirm different distributions for each season.
 - Calculate ambient PM indoor/outdoor ratio and plot vs. air exchange rate. Confirm relationship is as expected.
 - Confirm non-ambient contribution from cooking in home microenvironment

Model Run #1-3

Goals: Simulate population variability in PM_{2.5} exposure/dose, and examine effect of human activities on variability

Approach: Longitudinal simulation of 1% of population for multiple census tracts around tract where monitor is located using monitor ID-census tract ID matching file, and typical microenvironment selections and input distributions for variability

Approx. Model Run Time: 3 hours

GUI specifications:

- Model Run Inputs GUI
 - Input Data: Select 'PM Concentration' pushbutton and on PM Concentration File Information GUI:
 1. Select 'Excel Spreadsheet (*.xls)' option, and locate 'Detroit Daily PM25 2005 (Linwood).xls' file.
 2. Select '24-hour Average' option
 3. Select start hour as '12 AM'
 4. Select 'Monitor ID included', then 'Yes' for Monitor-Tract ID matching file. Locate 'Detroit Linwood CT IDs.xls' file.
 5. Select 'No' for Temperature Included in File
 6. Leave particle size distribution options as defaults
 7. Leave missing value symbols as '0' (no missing data in file)
 - Simulation Type: Longitudinal
 - Stage: 1-Stage(Variability)

- Census Tracts: Select 'All Tracts' pushbutton
 Population: Select 'Percent of Tract Population' and enter '1' for Percent of Tract Pop.
 Day Type: Select 'Include Both'
 Gender: Select 'Both Genders'
 Age: Enter '0' and '102'
 Seasons: Select 'Define Seasons' pushbutton and select 'Default' pushbutton
 Activity Diary Match Criteria: Select 'Employment' and 'Season'
 Commuting: none
- **Define Microenvironments GUI**
 Select same settings as example test run in User Guide (pages 18-19)
 - **Output Options GUI**
 Leave as default (no 'Event Time Series Data')
 - Select 'Run' pushbutton to perform the simulation for Model Run #1-3.

Analysis of population variability:

When model run has finished, perform the following comparisons using the Analyze Results GUI (see Section 4.7 of User Guide for detailed instructions):

- Compare output distributions for entire simulated population with those for different demographic groups such as males vs. females, children vs. adults vs. elderly, employed vs. unemployed. Either of the following can be done:
 - Use the options on the top of the Analyze Results GUI to subset different groups and generate plots and/or tables for each group to compare. Figure 8 below provides an example using percentile plots (y-axis limits changed to be the same) of exposures and box plots of time spent in microenvironments (Doers Only) comparing males/females/all individuals.
 - Export daily data to an output file and import file into a statistical analysis package to perform comparisons

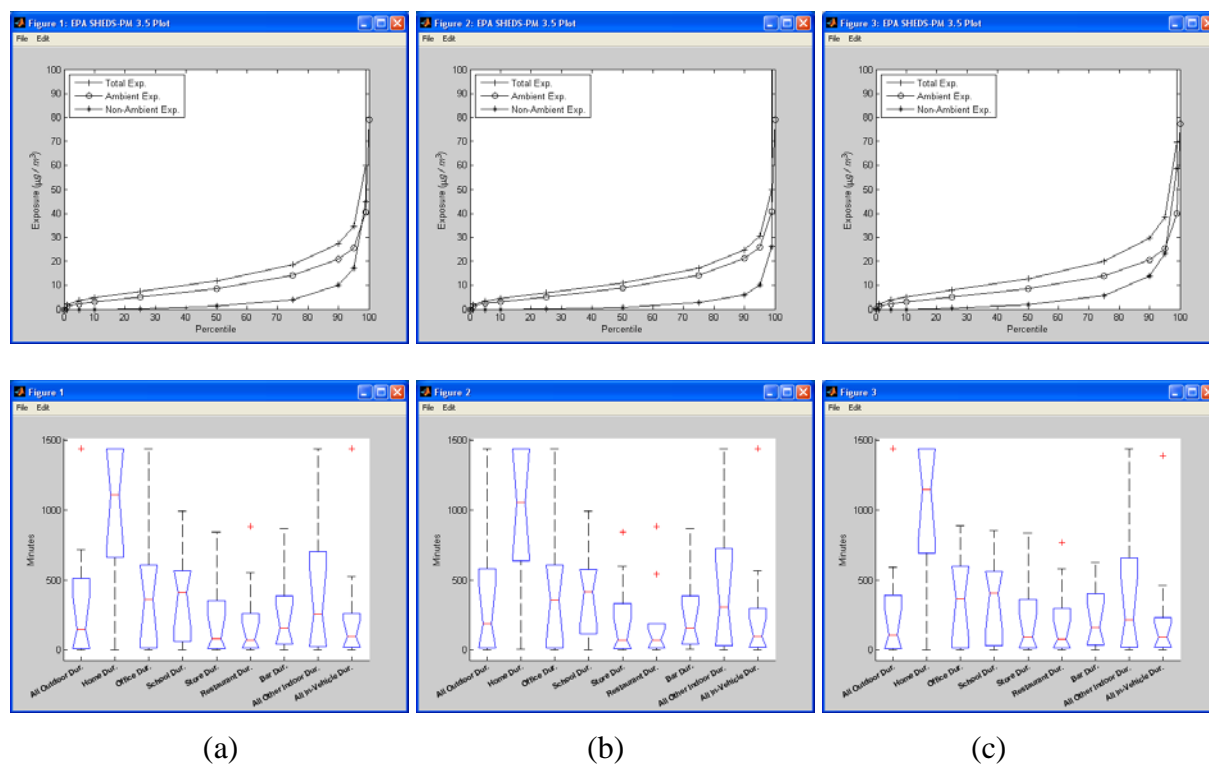


Figure 8. SHEDS-PM Percentile Plots of Exposures (top row) and Box Plots of Time in Microenvironments for (a) All, (b) Males, and (c) Females.

Scenario #2: Population Variability with Uncertainty

This model scenario will utilize the PM_{2.5} concentration input file for Detroit described above and typical model specifications for an example run with uncertainty estimation. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model.

Goals: Estimate uncertainty in the population variability for PM_{2.5} exposure/dose
Approach: Longitudinal simulation with 50 individuals simulated for the census tract where monitor is located (ID #26163522300); typical microenvironment selections and input distributions; 10 uncertainty iterations
Approx. Model Run Time: 1 hour

GUI specifications:

- **Model Run Inputs GUI**
 - Input Data: Select 'PM Concentration' pushbutton and on PM Concentration File Information GUI:
 1. Select 'Excel Spreadsheet (*.xls)' option, and locate 'Detroit Daily PM25 2005 (Linwood).xls' file.
 2. Select '24-hour Average' option
 3. Select start hour as '12 AM'
 4. Select 'Monitor ID included', then 'No' for Monitor-Tract ID matching file
 5. Select 'No' for Temperature Included in File
 6. Leave particle size distribution options as defaults
 7. Leave missing value symbols as '0' (no missing data in file)
 - Simulation Type: Longitudinal
 - Stage: 2-Stage(Uncertainty)
 - Census Tracts: Select 'Census Tracts' pushbutton and on Census Tract Selection GUI select 'Michigan' for State, and 'Wayne County', then locate and choose census tract ID number '26163522300' from the list, click on the 'Linwood' monitor ID box and press 'OK'
 - Population: Select 'Fixed Value Per Tract' and enter '50' for Individuals Per Tract
 - Day Type: Select 'Include Both'
 - Gender: Select 'Both Genders'
 - Age: Enter '0' and '102'
 - Seasons: Select 'Define Seasons' pushbutton and select 'Default' pushbutton
 - Activity Diary Match Criteria: Select 'Employment' and 'Season'
 - Commuting: none
- **Define Microenvironments GUI** (see Section 4.4 of User Guide for detailed instructions on input distributions for 2-Stage(Uncertainty) runs)
 - Select same microenvironments as example test run in User Guide (pages 18-19)
 - Leave default for All Outdoor
 - Leave default for Home Mass Balance
 - For all other indoor microenvironments see uncertainty distributions in Table 3 of Burke, *et al* (2001).
- **Output Options GUI**
 - Leave as default (no 'Event Time Series Data')
- Select 'Run' pushbutton to perform the simulation for Scenario #2.

When model run has finished, explore the data through the Analyze Results GUI (see Section 4.7 of User Guide for detailed instructions). Figure 9 below shows example for percentile plots for the uncertainty distributions for Total, Ambient, and Non-ambient Exposures for Scenario #2.

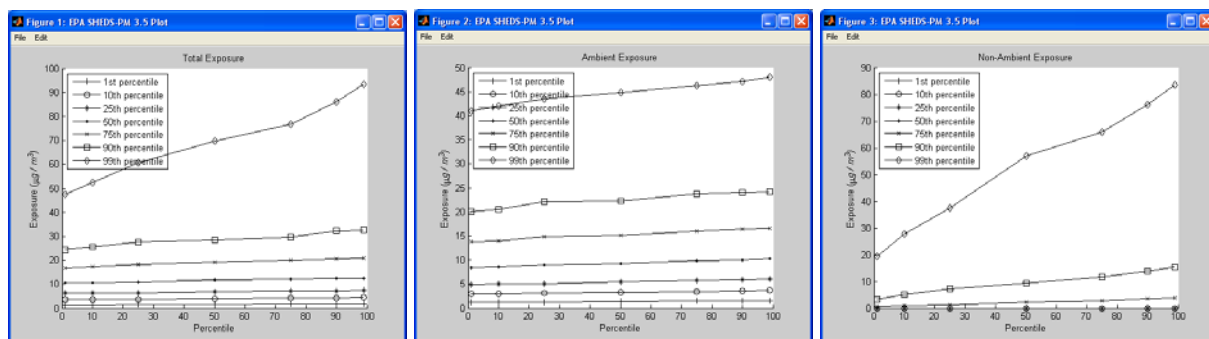


Figure 9. Example SHEDS-PM Percentile Plots of Uncertainty Distributions for PM Exposures from Scenario #2.

Scenario #3: Spatial Variability

This model scenario will utilize the PM_{2.5} concentration input file provided with the SHEDS-PM installation package (philaPM2008.csv) which contains daily concentration data for 5 monitors across Philadelphia, PA during 2008 (see page 14 of User Guide for more information). This scenario involves simulating exposures for a representative population living in Philadelphia County. Figure 10 provides a view of the spatial variability in the input PM_{2.5} concentrations. Commuting is included to account for time spent in census tracts that are different than the home census tract when individuals are at work. Analysis of the model results focuses on the spatial variability in PM exposures due to concentration differences between monitors.

- Goals:* Simulate population variability in PM_{2.5} exposure/dose using input data with spatial variability
- Approach:* Longitudinal simulation of 1% of population for all census tracts in Philadelphia county, and typical microenvironment selections and input distributions for variability
- Approx. Model Run Time:* 24 hours

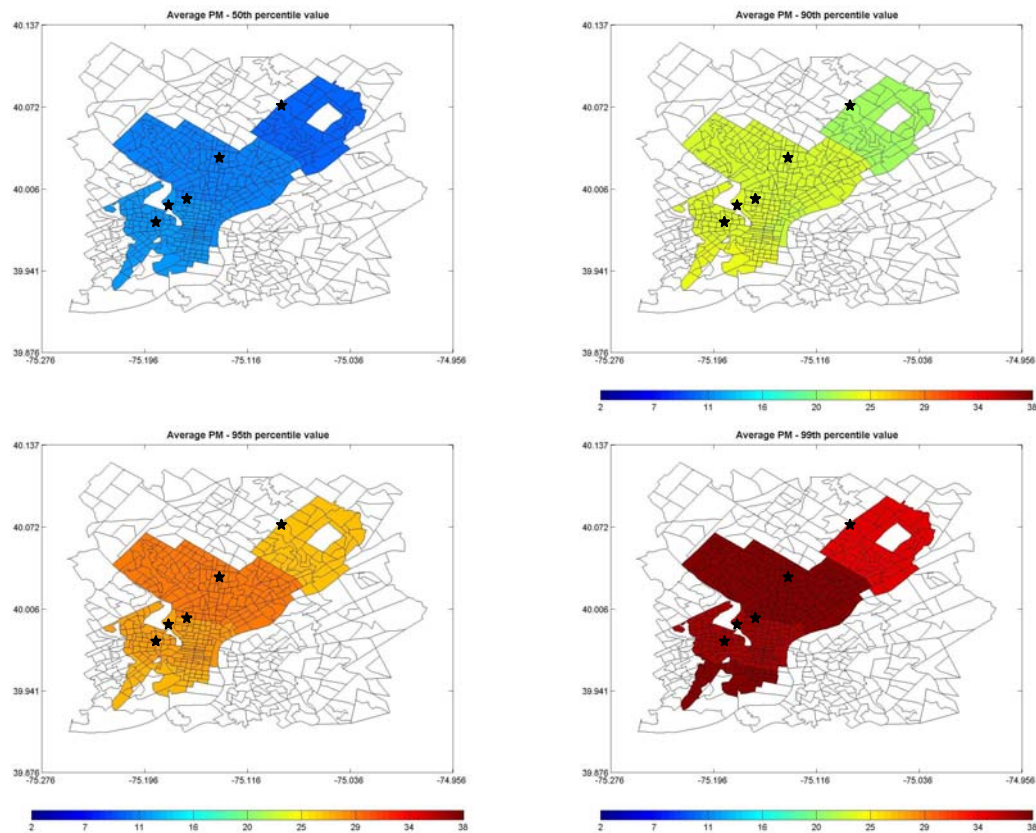


Figure 10. SHEDS-PM Map View of Input PM_{2.5} Concentrations for Scenario #3.

GUI specifications:

- **Model Run Inputs GUI**

Input Data: Select 'PM Concentration' pushbutton and on PM Concentration File Information GUI:

1. Select 'Text file' option, and locate 'philaPM2008.csv' file in the 'Data' subdirectory.
2. Select '24-hour Average' option
3. Select start hour as '12 AM'
4. Select 'Tract ID included'
5. Select 'No' for Temperature Included in File
6. Leave particle size distribution options as defaults
7. Leave missing value symbols as '0' (no missing data in file)

Simulation Type: Longitudinal

Stage: 1-Stage(Variability)

Census Tracts: Select 'All Tracts' pushbutton

Population: Select 'Percent of Tract Population' and enter '1' for Percent of Tract Pop.

Gender: Select 'Both Genders'

Age: Enter '0' and '102'

Seasons: Select 'Define Seasons' pushbutton and select 'Default' pushbutton

Activity Diary Match Criteria: Select 'Employment' and 'Season'

Commuting: Select 'Include Commuting'

- **Define Microenvironments GUI**

Select same microenvironments and input parameters as example test run in User Guide (pages 18-19)

- **Output Options GUI**

Leave as default (no 'Event Time Series Data')

- Select 'Run' pushbutton to perform the simulation for Scenario #3.

When model run has finished, explore the data through the 'Map View' on the Analyze Results GUI (see Section 4.7 of User Guide for detailed instructions).

APPENDIX 2: PEER REVIEWER COMMENTS

Arlene S. Rosenbaum, MPH, PhD

Peer Review Comments on EPA's Draft Model

Stochastic Human Exposure and Dose Simulation for Particulate Matter (SHEDS-PM) Version 3.5

Arlene S. Rosenbaum
ICF International
Nov 30, 2009

I. GENERAL IMPRESSIONS

SHEDS-PM is a state-of the science exposure modeling tool. Some advanced modeling features include:

- Estimation of dose as well as exposure
- Capability of performing 2-stage Monte Carlo sampling to estimate variability and uncertainty separately
- Well designed GUIs that facilitate data input and results analysis with a wide range of output options.

The GUIs makes the model extremely easy to implement and provides the capability to quickly construct graphs, plots, and maps, as well as to stratify results.

The User Guide is well organized, well written, and easy to follow, with a few exceptions noted below.

The exercises selected for the review contained clear directions and demonstrated most of the features of the model.

Some of the limitations of the model are addressed in the list of possible future improvements. Some additional ones and associated refinement suggestions are listed below in the "Other" section of possible future improvements.

II. RESPONSE TO CHARGE QUESTIONS

1) Install the SHEDS-PM model software program.

Install the model on a computer with Windows XP or later operating system using the file provided ('EPA SHEDS-PM 3.5 Installation.EXE') and following instructions in Section 2 of the User Guide.

a) Did you encounter any problems using the self-installing executable program to set up the model on your computer?

If yes, please describe the problem, the type of computer used, the operating system release number, the location the model was installed on the computer (e.g. 'C:\Program Files' or other drive), and whether the User Guide provided information to help correct the issue.

Comments:

No problems installing.

- b) *Do you have any suggestions for improving the User Guide section on the model installation procedures (Section 2)?*

Comments:

I encountered a pop-up window reading "extract census boundaries - one time only" that was not mentioned in the User Guide instructions. Noting this in the instructions will confirm to the user that this is not a problem.

2) *Perform SHEDS-PM Example Test Run.*

Set up and run the example described in Section 3 of the User Guide using the PM concentration input file provided in the 'Data' directory ('philaPM2008.csv'). Display and export the model results as described in Section 3.

- a) *Does the example test run provide a sufficient introduction to familiarize the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model?*

Comments:

Yes. However, as I was experimenting with changing settings for the microenvironment factors, I noticed that whenever I pressed the "cancel" button, I got a message on the DOS screen reading ""Error using ==> load; unable to read file mostRecentMicroEnvChoices: No such file or directory", and the set up screen remained active. When I reset the values to the defaults and pressed the "OK" button the set up screen de-activated.

3) *Perform Scenario #1 (Population Variability).*

This scenario demonstrates a typical SHEDS-PM application to estimate the variability in exposures to ambient PM_{2.5} for the population of an urban metropolitan area. The PM_{2.5} concentration input file includes daily, 24-hour average PM_{2.5} concentrations for 1 year from a monitor located in an urban area. A representative population from census tracts near the monitor is simulated, and includes all ages and both genders. This scenario defines several microenvironments with different infiltration characteristics for ambient PM_{2.5} (indoor PM sources are not included in this scenario). Analysis of the model results focuses on options available for displaying the output to characterize the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #1. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

The model ran as expected, except that it took about 80 minutes instead of 45, even though no other programs were open.

My PC is a Dell Latitude laptop running Windows XP Professional ver 2002, service pack 3, with Intel Core 2 Duo CPU T5500@1.66GHz processors.

For scenario 1-1 the resulting frequency statistics for “gender”, “age”, and “employment status” matched closely with those in the census data base for the tract, and the “season_number” and diary data appeared to be correct. The exposure concentrations matched the air quality input data

For scenario 1-2 the diary data and air quality data were assigned correctly. The ambient ME concentrations compared directionally with the ambient input concentrations as expected. The ratios of indoor-to-input ambient concentrations and in-vehicle-to-input ambient concentrations match the ME factor input distributions for those MEs closely. For a selected individual the number of diary records matched for each of the diaries and the location codes were correctly assigned. For a selected individual ME concentrations in event data and daily data match. Hand-calculated PM exposure matched the values in the events file.

Although the instructions state that the intake dose should be hand-calculated from the METS value, it seems like it should actually be calculated from the ventilation rate, according to the intake dose equation on page 130. The values matched except for a factor of 1000. Note: The intake dose equation on page 130 of the User Guide is off by a factor of 1000, since the concentration units are ug/m³ and the ventilation rate units are L/min. It needs a conversion factor (10⁻³) to convert ug/m³ to ug/L. (See specific comments below.)

For scenario 1-3 comparison plots looked similar to the examples in the instructions.

- a) *Do the options for analysis of model results provide the user with sufficient information to understand the population variability in PM exposures and the impact of human activities?*

Comments:

Yes

4) Perform Scenario #2 (Population Variability with Uncertainty).

This scenario demonstrates a SHEDS-PM application to characterize the uncertainty associated with the model estimates of population variability in ambient PM_{2.5} exposures. The same input PM_{2.5} concentration data and population demographics as Scenario #1 are used. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model. Analysis of the model results focuses on displaying the estimated uncertainty in the population distribution of exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #2. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

No. For restaurants and bars when I tried to add values to a triangular distribution for ASC emission rate, I get an error in the DOS window “Undefined function or variable

‘distChosen’’. The when I save the input window and re-open it the triangular distribution selection has reverted to uniform.

Also the Burke et al 2001 article mentioned applying a random factor to whether there was smoking in restaurants. I could not figure out how to implement such a scheme from the User Guide.

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the predicted uncertainty in the population variability of PM exposures?*

Comments:

Yes

5) Perform Scenario #3 (Spatial Variability).

This scenario demonstrates a SHEDS-PM application for understanding the spatial variability in PM_{2.5} exposures. The PM_{2.5} concentration input file includes PM_{2.5} input concentrations for multiple monitoring locations within an urban area. Commuting is included to account for time spent outside the home census tract when individuals are at work. A representative population for each monitor is simulated. Analysis of the model results focuses on options available for displaying the output to understand the spatial variability in PM exposures due to concentration differences between monitors.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #3. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

The model performed as expected, except that the simulation took approximately 36 hours rather than 24.

My PC is a Dell Latitude laptop running Windows XP Professional ver 2002, service pack 3, with Intel Core 2 Duo CPU T5500@1.66GHz processors.

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the impact of spatial and temporal variability in PM concentrations on the modeled distributions of PM exposures?*

Comments:

Yes, although it took some “drilling down” to discover why 2 of the Philadelphia tracts showed some extremely high non-ambient concentrations. They turned out to be from the home ME, presumably from cooking. I obtained a maximum of 770 ug/m³, which may or may not be realistic. This led me to notice that the open-ended distributions are not given any artificial bounds. (See suggestions for other possible future improvements in #7 below)

6) Provide summary assessment.

Please provide comments on the following:

- a) *The organization and usability of the user interface (GUIs), which features or options were most useful, and whether additional features or options are needed*

Comments:

The GUIs were organized well and very easy to use. I found the many output options to be the most useful, including the mapping and plotting options, as well as the ability to stratify the results.

Especially useful additional features would be (a) the ability to save the inputs and results from a simulation and (b) the ability to turn off dose calculations, as suggested in the list of possible future improvements below. Some other possible future improvements are suggested below.

- b) *Whether the descriptions of the model components and algorithms in the User Guide are sufficiently clear, technically correct, and represent the current state of the science for performing exposure assessments*

Comments:

I found the descriptions of model components and algorithms in the User Guide to be clear and technically correct, with the exception of the discussion of the intake dose and its underlying components (see specific comments below).

The algorithms generally represent the state of the science, although some modifications are suggested in #7 in addition to the ones already listed there.

- c) *Whether the output generated by the model are technically correct and consistent with descriptions of the algorithms in the User Guide*

Comments:

The output generated was consistent with the descriptions of the algorithms in the User Guide, with the exception of the intake dose equation on page 130, as noted above and below in specific comments. They also appear to be technically correct.

7) Rank priority for possible future improvements.

Several possible improvements to the SHEDS-PM model are listed below. Please provide a number ranking for the relative priority that should be given to each improvement, using a scale from 1 (low priority) to 5 (high priority).

Improving ease of use:

- ___3___ Create log file that records all inputs specified for the model run that can be viewed and saved by user
- ___5___ Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)
- ___5___ Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)
- ___1___ Provide more information on error messages to help users identify the reason for the error for common problems

- __3__ Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)

Allowing additional user specification of inputs:

- __3__ Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)
- __3__ Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)

Improving/refining model algorithms:

- __5__(see comment below) __Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of “unique” diary being used repeatedly for an individual)
- __5 (see comment below) __Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)
- __2__ Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)
- __4__ Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)
- __5__ Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

Adding new functionality:

- __2__ Option for using mapping tool to select census tracts for simulation based on a map
- __3__ Add more user options to map view of output (e.g. for use in GIS software or Google Earth)

Other:

Please describe

__5__ It appears that there is not an option to artificially bound open-ended parametric distributions (e.g., normal). If this is correct, adding such an option should be considered, to avoid unrealistic selections.

__5__ Allowing the user to specify the re-sampling frequency for dairies and ME factors should be considered, instead of hard-wiring the model to re-sample diaries seasonally and ME factors daily.

__5__ If a more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations is added, the Cluster-Markov algorithm used in HAPEM and in a special version of APEX should be considered. The Cluster-Markov algorithm samples diaries daily taking into account diary similarities and diary-to-diary transition probabilities.

__4__ Incorporating consideration of tract-specific commuting time distributions, available from US Census data, should be considered.

_4__ Upgrading the mass balance algorithm to be dynamic (i.e., allow carryover from one time period to the next) instead of equilibrium should be considered.

Open comments (optional)

Please provide any additional comments that you wish to on the SHEDS-PM model.

Comments:

None.

III. SPECIFIC OBSERVATIONS

Provide specific observations, corrections, or comments on the document, mentioning page, paragraph, and/or line number.

Page 126.

“ If When the activity is preparation of food, and if the diary event has a ‘Y’ for gas stove use during the event, then the total duration of the diary event is used for t_{cook} . Otherwise, a factor is randomly generated to account for food preparation activities that do not generate PM. The factor is a random number between 0 and 1.”

Page 130:

$$IDose_{ij} = C_i \cdot Ve_{nj} \cdot t_{ij} / 1000$$

where:

IDose _{ij}	=	inhaled PM dose during activity j while in microenvironment i (µg)
C _i	=	PM concentration for microenvironment i (µg/m ³)
Ve _{nj}	=	exhaled ventilation rate for individual n during activity j (L _{air} /min)
t _{ij}	=	duration of activity j while in microenvironment i (min)

Page 130

$$VO_{2nj} = METS_j \cdot BMR_n \cdot EEtoVO_{2n} / BM_n$$

This equation and its subsequent development on page 131 are unclear, especially with respect to the measurement units. The measurement units of each term should be presented immediately following the equation.

P. Barry Ryan, PhD

Peer Review Comments on EPA's Draft Model

Stochastic Human Exposure and Dose Simulation for Particulate Matter (SHEDS-PM) Version 3.5

P. Barry Ryan
Emory University
30 November 2009

I. GENERAL IMPRESSIONS

The SHED-PM model appears to be very complete and comprehensive, allowing both variability and uncertainty to be modeled. The model requires a large amount of input data, data that are unlikely to be available for many situations. However, that may not be problematic in that the large populations simulated, along with the numerous microenvironments allow the researcher to glean much useful information from the model results much of which should be generalizable to any other situations.

Aside from the large database needed to run the model, the model specification is quite straightforward. Various individual microenvironments can be explored as can specific age groups, gender-specific exposures.

One difficulty is the size of the files that must be manipulated and the time that takes to do the calculations. While the laptop I was using is hardly state of the art, is also not archaic. Yet the estimates of time were consistently underestimated by about 50%. Further, a trial scenario that takes 24 hours to perform does not make the best test of the system. The shorter duration tests are a better indicator of what the system can be done. The time associated with writing out data files for later use, coupled with the size of the files gives one pause. For example, in Scenario #2, writing the data to disk took in excess of three hours and ended up with an MSEXcel file that exceeded 250 MB in size. If this program is to be useful as a tool for the typical exposure assessor, this process should be streamlined.

II. RESPONSE TO CHARGE QUESTIONS

1) Install the SHEDS-PM model software program.

Install the model on a computer with Windows XP or later operating system using the file provided ('EPA SHEDS-PM 3.5 Installation.EXE') and following instructions in Section 2 of the User Guide.

a) Did you encounter any problems using the self-installing executable program to set up the model on your computer?

If yes, please describe the problem, the type of computer used, the operating system release number, the location the model was installed on the computer (e.g. 'C:\Program Files' or other drive), and whether the User Guide provided information to help correct the issue.

Comments:

Because my laptop runs Vista, I ran into a small problem installing the program. The User's Manual gives instructions for XP, working all the way through, then modifications for VISTA.

My desktop computer at work runs XP, but access to administrative mode is restricted. Therefore all of my testing was done on my laptop (1.8 GHz, T5550 Processor with 3 GB of memory. 320 GB hard drive, WiFi 802.11.g Networking).

- b) *Do you have any suggestions for improving the User Guide section on the model installation procedures (Section 2)?*

Comments:

Because of the minor difficulty outlined above, I suggest a stronger statement in the User's Manual regarding Administrative Mode. Perhaps even a separate, albeit repetitive, set of instructions for XP, Vista, and now Windows 7, is in order. If your operating system is Windows XP, go here. If Windows Vista go to page, xx. Et cetera.

2) Perform SHEDS-PM Example Test Run.

Set up and run the example described in Section 3 of the User Guide using the PM concentration input file provided in the 'Data' directory ('philaPM2008.csv'). Display and export the model results as described in Section 3.

- a) *Does the example test run provide a sufficient introduction to familiarize the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model?*

Comments:

I have a series of specific comments noted at each of several points along the process here. One overarching comment begs a solution, however. As written, the manual takes one through various sections of the input and running of the SHEDS-PM model. However, it is very "cook-book." It tells you to press this button,, select this, option, etc., without going into any detail or supplying any information about what is being accomplished by pursuing that action. This is a failing of the document. While technically fulfilling the requested information about "...familiarize[ing] the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model" I would not know how to run a substantively different scenario that the one input given the information present at this time. It is satisfying to get a result and see that the system actually does produce (a lot of) data, it would be better if I felt as though I knew what I was doing a bit more. While I realize that the remainder of the Manual does indeed address the specifics of what each step means, it would be useful to give at least some context and explanation this point. For example, one could simply say, "... now we are going to take the data as input from an external file, and use it to perform a Monte Carlo simulation. Begin this by reading in the data. This is accomplished by..." and continue.

Specific Comments.

Example Test Run:
Section 3-o.

<i>Example Test Run: Section 3-o.</i>	The bar progresses very slowly on my system. It is not several seconds, but rather 2-3 minutes before the data checking was completed.
<i>Microenvironmental</i>	The default for penetration is normal distribution with mean of 0.97

Section 2.iii	and sd = 0.2, so a significant amount of the simulations operate with $P > 1.0$. Similarly, would deposition allow a negative value? HOW is this reconciled?
2.viii	The parameters here are indistinguishable from the Restaurant scenario.
c.	At this point, the whole process is a big “black-box.” What is going on?
Section 3.3	Here we make our first real “run.” The manual estimates suggest a 45-minute time to complete. Further, it says that if you don’t hit these marks, it recommends getting a faster computer. This is not especially helpful. Often, one cannot just run you and get a new computer just to run an EPA model. Here is a brief table of my experience given my computer as described above: After a 2 minute setup, estimate is 108 minute remaining. Dropped to 100 minutes after an addition 1 minute Dropped to 85 after an additional minute Dropped to 82 after an additional minute Back to 86 after an additional 3 minutes After 20 minutes, estimate is 72 minutes. After 25 minutes, estimate is 65 minutes After 35 minutes, estimate is 46 minutes. Left to do a bit of housework. Returned after 115 minutes from start and the process was complete.
Comment	I really like the “Current Status” window that is constantly updated. This type of feedback give assurance that one is not stuck in a loop.
3.4 4.c.	The Microsoft table did not appear. I had to go get it.
Page 23 Figure 16	The Excel spreadsheet came through with formatting problems with the headers. I do not know if the is an XML translation problem, but it was a bit annoying.
Page 23 Figure 17	I experimented with a number of other plotting combinations as well. These are quite useful as representations of the data.
Page 24 6.b.	I suggest a “browse” option be put in here.
Page 24 6.c.	Where it says “will take a few minutes” it took 20 minutes and produced a 165MB file. More warning and a better estimate is warranted here.

3) **Perform Scenario #1 (Population Variability).**

This scenario demonstrates a typical SHEDS-PM application to estimate the variability in exposures to ambient PM_{2.5} for the population of an urban metropolitan area. The PM_{2.5} concentration input file includes daily, 24-hour average PM_{2.5} concentrations for 1 year from a monitor located in an urban area. A representative population from census tracts near the monitor is simulated, and includes all ages and both genders. This scenario defines several microenvironments with different infiltration characteristics for ambient PM_{2.5} (indoor PM sources are not included in this scenario). Analysis of the model results focuses on options available for displaying the output to characterize the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #1. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

The User's Manual reads: "...click on **Edit/View Model Run Inputs**..." The Push Button reads: "**View/Edit Model Run Inputs**". The manual should reflect what is in the program to avoid confusion.

I was able to complete all the tasks outlined and gradually became more familiar with the workings of the model during this run.

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the population variability in PM exposures and the impact of human activities?*

I believe the multiple scenarios selected afforded "exercising" the model and displaying all of its most important features. In a few places I "went rogue" and began exploring some features that were not part of the specific challenges offered at the time. The program responded well and gave me better insight into the operations of the system. For example, I inspected several specific microenvironments with regard to the plots and statistics offered. This proved insightful not only with regard to the tuning of the model but also proved fruitful in gaining insight into the abilities of the software.

Comments:

I do not have much specific to say about this scenario. My main thoughts in running it were to gather acumen and skill in modifying the parameters of the input and evaluating what came out. In this regard, the software seems quite complete. I perhaps spent less time comparing my results to the tables in the Appendix than I should have, but I found it more interesting to "play" with the program to determine what kinds of output were available, what parameters could be modified and the effects such modification would have on the output, and exploring graphical forms of outputs, e.g., pie charts, scatter plots, etc. In this regard, I think I went a bit out of sequence and hence got a bit frustrated later on with the long time scales needed to complete some of the tasks.

I enjoyed this section of the evaluation more than the others, perhaps because of my curiosity and the exploration done.

One "glitch" I noted occurred during some of the plotting. If one plots multiple microenvironments in the same plots, often the plots themselves plot "through" the legend making for both an untidy presentation and, occasionally, one that is difficult to read. This is doubtless due to fixed size considerations on the plots. I am not sure if this can be remedied either easily or at all, but it was annoying.

Another minor annoyance occurred in that one of the exposure calculated was very high, an unlikely, but somewhat expected, occurrence in any kind of simulation. This resulted in certain of the plots, most notably the box plots, becoming compressed and essentially unusable because of trying to plot this one unusual individual with an exposure in excess of $900 \mu\text{g}/\text{m}^3$. This may have been my random seed that got me this guy, but it will happen.

In running Scenario #1-3, I ran into timing troubles again. I can reproduce the timing table I kept, but the bottom line was that it took in excess of four hours to do this run. I kept on checking back while doing other activities and missed the actual finish, but it was between 218 minutes, then there was an estimate of 13 minutes left, and 242 minutes, when the job had finished. The estimates tended to be too long near the beginning, and too short near the end.

I had trouble getting the Daily Time Series to run. I kept on getting errors that precluded finishing so I gave up on trying to get that accomplished. I believe the errors looked like: Error using → shedprn ('run_Callback') and then some numbers- probably error codes. But this may have been some other error.

4) Perform Scenario #2 (Population Variability with Uncertainty).

This scenario demonstrates a SHEDS-PM application to characterize the uncertainty associated with the model estimates of population variability in ambient PM_{2.5} exposures. The same input PM_{2.5} concentration data and population demographics as Scenario #1 are used. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model. Analysis of the model results focuses on displaying the estimated uncertainty in the population distribution of exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #2. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*
This scenario performed more or less as I would have expected. I did get some error messages on input, but was able to complete the task by getting around them.
- b) *Do the options for analysis of model results provide the user with sufficient information to understand the predicted uncertainty in the population variability of PM exposures?*

Comments:

I have little new to report in this section. I made use of most of the features in the Analyze Results GUI and explored the output from them. I found the plots interesting again and explored a number of aspects. These visual representations are of most interest and offer a good deal of insight.

Because of the timing problems I had in Scenario #3 (see below), I had to set this project aside for a period of 4-5 days, and then return to it. Hence, some of my recollections may be a bit in error. Nevertheless, I forge ahead. I believe that it was in this scenario that I ran into an enormous delay in writing out a file. Like a previous comment, it was at a point when the data are to be written out to a file and the manual says "This may take a few minutes." A few minutes stretched into three hours and the file produced was just over 265 MB in size. This could be a problem. Most computers these days have hard disks that stretch out to 500 GB and more, so the space is not really a problem. However, someone running on an older computer or one that is packed with data may run into a problem. It should be relatively simple to calculate how large a file is likely to be then following that up with a look-see on

the operational hard disk to ensure that there is room for it. A text box could give this advice. Further, the phrasing “may take a few minutes” needs some work. A reasonable estimate for the time to write can be made through the software examining the hardware of the computer upon which it is running- disk access speed, expected size of the file, perhaps some other statistics- and given to the user up front. The user could then decide whether to write out the data and go get dinner, or not write out the data. As an alternative approach, a more compressed form of the file could be generated and written out more quickly, and software used to decompress the file on re-input, etc. This would substantially reduce the frustration factor.

All of these things being said, the amount of work that can be done in terms of data exploration using this tool is enormous. It is truly an amazing tool.

5) Perform Scenario #3 (Spatial Variability).

This scenario demonstrates a SHEDS-PM application for understanding the spatial variability in PM_{2.5} exposures. The PM_{2.5} concentration input file includes PM_{2.5} input concentrations for multiple monitoring locations within an urban area. Commuting is included to account for time spent outside the home census tract when individuals are at work. A representative population for each monitor is simulated. Analysis of the model results focuses on options available for displaying the output to understand the spatial variability in PM exposures due to concentration differences between monitors.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #3. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

In general, yes, but my comments below are most important.

Comments:

This was by far the most frustrating component of the review. I did not notice the expected time for this run until two days before the due date for the report- now several days in the past. But, I figured, I have two days- I run the scenario and even if it takes 30 hours I will still have plenty of time. So off I went. The software chugged along for a period of time and I finally went to bed, expecting the system to take care of itself and complete its task while I slept. But while I slept, something bad happened. I do not know what. The system hung about 1/3 of the way through. It appeared to be still running in the morning and it took me a few minutes to realize that it was constantly displaying the same tract, individual, etc. I had to restart my system and begin again. This time, it ran straight through, but took at least 36 hours to complete. And when it did finally complete and I went to perform the analyses requested, I found that I had looked at most of those features in earlier runs. So, I was unable to complete my task on time, and had other priorities scheduled for the intervening few days.

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the impact of spatial and temporal variability in PM concentrations on the modeled distributions of PM exposures?*

Yes. The system offered good insight into these areas.

Comments:

The system allowed adequate exploration of all effects. I found plotting the higher percentiles on the census tract most interesting and informative. The lower percentiles provided less insight. This was true no matter which of the parameter- ambient exposure, non-ambient exposure, does, etc.- were being plotted.

6) Provide summary assessment.

Please provide comments on the following:

- a) *The organization and usability of the user interface (GUIs), which features or options were most useful, and whether additional features or options are needed*

Comments:

The GUI seems to be well organized logically once you understand what is being done. As I reported earlier, while the Manual is very complete, the “Getting Started” section is, in my opinion, too “cookbook-like” in that it tells you which buttons to press, but does not give insight into why you are pressing them. The details are supplied in later chapters, but even a brief gloss over of what is happening would add substantial insight. When you first bring the program up, it is pretty intimidating. I realize that the developers and users are long past that stage, but I am a pretty sophisticated software user, and I still felt overwhelmed and under-informed when I first went to use the system. A bit more explanation would be helpful.

- b) *Whether the descriptions of the model components and algorithms in the User Guide are sufficiently clear, technically correct, and represent the current state of the science for performing exposure assessments*

Comments:

I did not examine the technical contents for detailed mathematical errors. However, I saw nothing that gave me pause in the presentation. There is a good deal of technical material there and I think it is presented in a more coherent fashion than in most presentations. For example, I am now plowing my way through the AERMOD series of programs (AERMET, AERSURFACE, etc.) and found this presentation much more rewarding- more like some of the technical appendixes in the documents I just mentioned. The Manual appears written for the exposure scientist who might use this model, rather than a technician looking for answers to a problem using a canned program. This is both a strength and a weakness. It is a strength because the user is likely to be sophisticated in exposure in general. It is a weakness, because the system may be less accessible to the “lay” audience. A decision will have to be made regarding the future direction of such a system. Will an effort be made to present this in a manner more accessible to a non-technical audience? If so, a re-write is in order. However, I would advise against modifying what is here. This is a sophisticated tool and should be used by those who are well versed in the science. This may sound elitist; if so, so be it. Perhaps a “SHEDS Lite” could be developed that was less sophisticated in utilization for those wishing to use a simpler tool.

- c) *Whether the output generated by the model are technically correct and consistent with descriptions of the algorithms in the User Guide*

Comments:

I believe I covered this in the above comment.

7) Rank priority for possible future improvements.

Several possible improvements to the SHEDS-PM model are listed below. Please provide a number ranking for the relative priority that should be given to each improvement, using a scale from 1 (low priority) to 5 (high priority).

Improving ease of use:

- 5 Create log file that records all inputs specified for the model run that can be viewed and saved by user

I believe that this could be implemented easily and has a great deal of utility in software QA. Hence I place a high priority on it.

- 5 Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)

Notwithstanding the database storage requirements, this should be a high priority as well. I did find it frustrating not to be able to return to a different scenario and retest something I discovered in a later version.

- 4 Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)

This is a high priority, but not as high as the first two. Speed of calculation is important, however, and this could help tremendously in this regard.

- 3 Provide more information on error messages to help users identify the reason for the error for common problems

Again, important, but of lower priority. At least an error number could be implemented and printed out, with a table to identify the error type.

- 1 Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)

This is far less important in my view than any of the others.

Allowing additional user specification of inputs:

- 3 Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)

- 3 Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)

Both of these are of interest, but would likely require a substantial amount of work. Requiring mass-balance among a number of compartments is problematic and leads to restriction on input. Physiological parameters may be of higher priority, but are further down the road.

Improving/refining model algorithms:

- 3 Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of “unique” diary being used repeatedly for an individual)

I put this at a mid-level of priority. Including more variability on the diaries is generally good, but it is not clear if such data exist. An alternative strategy is

sampling without replacement from the current list to ensure that the same “unique” diary is not used over and over.

- 3 Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)

I put this at mid-level priority in that it may be hard to do and would require a lot of information that may not be readily available.

- 2 Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)

I am not even sure I understand what would have to be done, much less the degree of difficulty for implementation. A low priority.

- 2 Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)

I put this at lower priority because it increases the scale of the data input size and likely slows down the calculation process substantially. As it already takes a while for these simulations to run, making them more detailed may not be a great use of resources.

- 2 Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

This is, once again, a lot of work for not so much benefit and is, therefore, a lower priority.

Adding new functionality:

- 2 Option for using mapping tool to select census tracts for simulation based on a map

I find this to be of low priority because of the need for very sophisticated data at the map site. This is unlikely to occur frequently as it costs a lot of money to generate the data.

- 4 Add more user options to map view of output (e.g. for use in GIS software or Google Earth)

Using a GIS version of an aerial view instead of the census tract maps would add more interest and should be relatively easily done.

Other:

____ Please describe: _____

8) Open comments (optional)

Please provide any additional comments that you wish to on the SHEDS-PM model.

III. SPECIFIC OBSERVATIONS

Provide specific observations, corrections, or comments on the document, mentioning page, paragraph, and/or line number.

I have placed these comments inline above.

Helen H. Suh, ScD

Peer Review Comments on EPA's Draft Model
Stochastic Human Exposure and Dose Simulation for Particulate Matter
(SHEDS-PM) Version 3.5

Helen H. Suh
Department of Environmental Health
Harvard School of Public Health

November 30, 2009

I. GENERAL IMPRESSIONS

It is clear that considerable work and thought has been put into the development of SHEDS-PM and the User Guide. The User Guide for SHEDS-PM 3.5 is especially thorough and well written, providing clear and easy to follow instructions that are helpful in navigating the SHEDS-PM software. Further, the manual provides a nice introduction to the software, with background information and some references. The model software GUIs are also visually appealing and relatively easy to read. The manual and especially the software, however, do assume a great deal of knowledge about exposure assessment, particulate behavior, and activity patterns on the part of the user, limiting its accessibility, usability, and interpretability of the results. To help in this regard, the software would benefit from direct linkages to the relevant sections of the manual, including not only the step-by-step instructions, but also relevant information about what the options mean, when they should choose between various options, and their implications for particulate exposures. To do so, targeted help modules and/or from further instruction imbedded on the screen would be helpful. [The "View User Guide" button did not work on my version. Similarly, the help screen buttons (when available) were not working.] Also, it would be helpful to include scientific links, citations or additional information in the manual and on the screen that can provide some guidance that will help people select and think about the different options.

In addition, it would be helpful for the user to be able to summarize the results in additional, more flexible ways without having to move to EXCEL or other platforms. For example, it would be great to be able to quality or data checks within the program or to construct specific regression models. Correspondingly, the program would benefit from improved ability to view input databases (for example for time/activity data) directly from the program and also the equations (or codes) used to generate various results. It was unclear whether the user could import measured activity or microenvironmental concentration databases, so that if measured data were available, the user could use these data instead of the provided distributional data. This information would transform the program from a "black box" program to one with increased flexibility and scientific rigor.

Other issues relate to the fact that the database needed to run the requested analyses was initially omitted from the provided materials, resulting in some confusion as to whether the database was not provided or whether the database was imbedded in one of the database files. This confusion suggests that the databases and other information contained in each module should be more clearly delineated. Further, running the SHEDS-PM often made other programs on my computer fail, requiring a hard reboot before these other programs could be used again. As a result, work in these other programs was lost. Some warning of this possibility should be provided prior to running the program.

II. RESPONSE TO CHARGE QUESTIONS

Provide narrative responses to each of the eight charge questions below.

1) *Install the SHEDS-PM model software program.*

Install the model on a computer with Windows XP or later operating system using the file provided ('EPA SHEDS-PM 3.5 Installation.EXE') and following instructions in Section 2 of the User Guide.

- a) *Did you encounter any problems using the self-installing executable program to set up the model on your computer? If yes, please describe the problem, the type of computer used, the operating system release number, the location the model was installed on the computer (e.g. 'C:\Program Files' or other drive), and whether the User Guide provided information to help correct the issue.*

Comments:

No. Installation was straight-forward with no identified problems.

- b) *Do you have any suggestions for improving the User Guide section on the model installation procedures (Section 2)?*

Comments:

The User Guide was clear and comprehensive in model installation procedures.

2) *Perform SHEDS-PM Example Test Run.*

Set up and run the example described in Section 3 of the User Guide using the PM concentration input file provided in the 'Data' directory ('philaPM2008.csv'). Display and export the model results as described in Section 3.

- a) *Does the example test run provide a sufficient introduction to familiarize the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model?*

Comments: The example test run and its associated components were fine, although perhaps would have been better with more information about each step. For example, introductory information regarding what the example test run will teach, the processes involved, and the reasons for generating the output.

3) *Perform Scenario #1 (Population Variability).*

This scenario demonstrates a typical SHEDS-PM application to estimate the variability in exposures to ambient PM_{2.5} for the population of an urban metropolitan area. The PM_{2.5} concentration input file includes daily, 24-hour average PM_{2.5} concentrations for 1 year from a monitor located in an urban area. A representative population from census tracts near the monitor is simulated, and includes all ages and both genders. This scenario defines several microenvironments with different infiltration characteristics for ambient PM_{2.5} (indoor PM sources are not included in this scenario). Analysis of the model results focuses

on options available for displaying the output to characterize the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #1. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

The model performed as expected, although it was awkward to export the data to EXCEL, open up other databases, and do comparisons. This need for multiple programs involves too many steps and makes the SHEDS-PM seem incomplete and not sufficient on its own. To be complete, the model should include instructions to perform frequency statistics and other data analysis summaries in EXCEL. Otherwise, the software should include these capabilities within the program.

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the population variability in PM exposures and the impact of human activities?*

Comments:

From the analyses, it was difficult to identify human activity factors affecting population variability in PM exposures, although the impact of gender, employment, and age were discernible. This may be due to the fact that more sophisticated analyses are needed to examine impacts of time-activity patterns than were requested or possible with the package. In addition, run results showed indoor non-ambient exposures to be zero, which seemed unlikely; however, it was difficult to figure out whether the values were zero due to an error in the program set-up or to some other reason. The analysis would be greatly improved with a provision to perform more diagnostics and to see the program go through the program steps.

4) *Perform Scenario #2 (Population Variability with Uncertainty).*

This scenario demonstrates a SHEDS-PM application to characterize the uncertainty associated with the model estimates of population variability in ambient PM_{2.5} exposures. The same input PM_{2.5} concentration data and population demographics as Scenario #1 are used. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model. Analysis of the model results focuses on displaying the estimated uncertainty in the population distribution of exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #2. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

Yes, although with some difficulty. Initial runs resulted in error messages that asked that I consult with Janet Burke. Although the manual provided information to fix the problem, it

took several attempts to reboot my computer and re-run program before the program would work. Once the model worked, it performed well.

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the predicted uncertainty in the population variability of PM exposures?*

Comments:

Yes. However, it would be helpful if the program would automatically estimate population PM exposures with and without uncertainty to examine the relative impacts of uncertainty in the various microenvironmental infiltration parameters on the exposure distribution.

5) Perform Scenario #3 (Spatial Variability).

This scenario demonstrates a SHEDS-PM application for understanding the spatial variability in PM_{2.5} exposures. The PM_{2.5} concentration input file includes PM_{2.5} input concentrations for multiple monitoring locations within an urban area. Commuting is included to account for time spent outside the home census tract when individuals are at work. A representative population for each monitor is simulated. Analysis of the model results focuses on options available for displaying the output to understand the spatial variability in PM exposures due to concentration differences between monitors.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #3. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

Yes, although the instructions and GUI were not reliable regarding the approximate run time and the “estimated run time left”, respectively. Further, the usefulness of the model is greatly reduced given the long run times.

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the impact of spatial and temporal variability in PM concentrations on the modeled distributions of PM exposures?*

Comments:

As with the other components, model results would be enhanced with more flexibility in the analysis, specifically so that analyses beyond summary statistics could be performed.

6) Provide summary assessment.

- a) *The organization and usability of the user interface (GUIs), which features or options were most useful, and whether additional features or options are needed*

Comments:

The “view/edit model run input” GUI was very organized, clear and straight-forward. All of the GUIs would be improved with working and targeted help functions. The “Microenvironment” and “Analyze Results” GUIs would be especially improved, with increased flexibility and ability to run more specialized analyses. The analysis GUIs are

weak, allowing only summary type of analyses to be performed. While other statistical programs are available to run more sophisticated analyses, PM-SHEDS would be greatly enhanced with more sophisticated and/or flexible analysis tools.

- b) *Whether the descriptions of the model components and algorithms in the User Guide are sufficiently clear, technically correct, and represent the current state of the science for performing exposure assessments*

Comments:

The User Guide is clear, well organized, and technically correct; however, it would be enhanced with more information about the state of the science, relevance and interpretation of various model components to exposure assessments (as noted in general comments).

- c) *Whether the output generated by the model are technically correct and consistent with descriptions of the algorithms in the User Guide*

Comments:

The model output is consistent with the descriptions in the User Guide; however, it is not possible to assess its technical correctness as the model does not include user-administered quality control/assurance procedures nor does it display or make available the intermediate model steps and calculations.

7) Rank priority for possible future improvements.

Several possible improvements to the SHEDS-PM model are listed below. Please provide a number ranking for the relative priority that should be given to each improvement, using a scale from 1 (low priority) to 5 (high priority).

Improving ease of use:

- ☐_4_ Create log file that records all inputs specified for the model run that can be viewed and saved by user
- ☐_3_ Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)
- ☐_3_ Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)
- ☐_5_ Provide more information on error messages to help users identify the reason for the error for common problems
- ☐_4_ Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)

Allowing additional user specification of inputs:

- ☐_4_ Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)
- ☐_4_ Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)

Improving/refining model algorithms:

- __3__ Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of “unique” diary being used repeatedly for an individual)
- __4__ Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)
- __2__ Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)
- __2__ Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)
- __4__ Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

Adding new functionality:

- __2__ Option for using mapping tool to select census tracts for simulation based on a map
- __2__ Add more user options to map view of output (e.g. for use in GIS software or Google Earth)

Other:

- __5__ An important model improvement would be to allow the user to import measured time/activity or microenvironmental concentration databases for use in model calculations. These measured data would reduce uncertainty in estimated exposure distributions. In addition, the model would be enhanced if the infiltration factors for the different microenvironments could vary by season. Given the sometimes long model run times, the ability to perform preliminary or crude exposure assessments (possibly by using fixed values for certain steps) may be important to allow the user to compare among different model options and to decide final model run parameters.

8) Open comments (optional)

Please provide any additional comments that you wish to on the SHEDS-PM model.

Comments:

None.

III. SPECIFIC OBSERVATIONS

Provide specific observations, corrections, or comments on the document, mentioning page, paragraph, and/or line number.

Ira B. Tager, M.D., M.P.H./
Fred Lurmann, MS

Peer Review Comments on EPA's Draft Model
Stochastic Human Exposure and Dose Simulation for Particulate Matter
(SHEDS-PM) Version 3.5

Ira B. Tager, M.D., M.P.H.
Professor of Epidemiology
University of California, Berkeley
Berkeley, CA

Fred Lurmann, MS
Manager of Exposure Assessment Studies
Sonoma Technology, Inc.
Petaluma, CA

November 30, 2009

I. GENERAL IMPRESSIONS

The SHEDS-PM model contains many of the features expected for a modern stochastic general population exposure model. The User Guide and companion papers describe the model and explain its use reasonably well. We were able to install the software and run the model on the test problems without problems; however, we did not “stress test” the software to identify bugs or other potential problems.

SHEDS, like other exposure models, provides a mathematical framework for exposure calculations. SHEDS-PM also contains a fair amount of pre-selected or embedded data (CHAD, US Census, etc.). The validity of exposure estimates derives from both the mathematical framework and the choice of data for particular applications. Since most modern exposure models share a common microenvironmental approach, the distinguishing element of exposure simulations is generally the choice of data rather than the model framework. We believe there are some limitations of the embed data (e.g., the CHAD data are out-of-date, the met assignments are based on incorrect estimates of oxygen utilization, and the geographic resolution of census tracts is too coarse to resolve the influence of important local sources, such as traffic). Insufficient guidance is provided for the user regarding the process of selecting scientifically credible input data. For example, the data and methods to calculate microenvironmental concentrations are often critical for the results. It was disturbing to find that the test problems data and regression equations for the nonresidential microenvironments came from an unpublished reference (Zufall et al. submitted 2001). Many users are likely to use whatever data comes with the model without critically evaluating its suitability for their applications. We believe the user guide, for example, would benefit from the presentation and explanation of how residential mass balance model parameters and nonresidential microenvironmental concentrations estimating equations are selected for one or more regions of the U.S.. the results could become the basis for the model's default parameter).

Confidence in models like SHEDS-PM comes from documented model evaluation, refinement and validation studies using field data. Model evaluation is common practice and essential for most complex mathematical models (e.g., EPA's Models-3 Community Multiscale Air Quality (CMAQ) Modeling System). Even if the different types of data and submodels selected for SHEDS are individually sound, the performance of the whole model against real-world data

needs to be demonstrated for exposure scientists and epidemiologists to accept the model. Thus, the lack of one or more peer-reviewed, published model validation studies undermines the credibility of the SHEDS-PM model.

Given the lack of validation of the model, the out of date activity data, incorrect estimates of oxygen utilization, and likely uncertainty and variability of dose estimates, we doubt that any creditable epidemiologist would use the current model to estimate individual-level exposure and dose or even distributions of exposure and dose for the general population. If EPA release the model in the near future, it is important to disclose the model's limitations and have a program to address them.

II. RESPONSE TO CHARGE QUESTIONS

1) Install the SHEDS-PM model software program.

Install the model on a computer with Windows XP or later operating system using the file provided ('EPA SHEDS-PM 3.5 Installation.EXE') and following instructions in Section 2 of the User Guide.

- a) Did you encounter any problems using the self-installing executable program to set up the model on your computer?*

If yes, please describe the problem, the type of computer used, the operating system release number, the location the model was installed on the computer (e.g. 'C:\Program Files' or other drive), and whether the User Guide provided information to help correct the issue.

Comments:

The program installation went smoothly on two Windows XP computers. The information in the User Guide was clear and sufficient to install the program.

- b) Do you have any suggestions for improving the User Guide section on the model installation procedures (Section 2)?*

Comments:

No, the installation was comparable to other Windows software.

2) Perform SHEDS-PM Example Test Run.

Set up and run the example described in Section 3 of the User Guide using the PM concentration input file provided in the 'Data' directory ('philaPM2008.csv'). Display and export the model results as described in Section 3.

- a) Does the example test run provide a sufficient introduction to familiarize the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model?*

Comments:

The example test run was easy to follow and produced output very similar to the User's Guide. It might make sense to specify the random number seed(s) so that the user could confirm that the program calculated the exact expected results. It is important to

emphasize that executing the program with pre-selected inputs is one of many steps needed to understand how exposure modeling should be carried out.

3) Perform Scenario #1 (Population Variability).

This scenario demonstrates a typical SHEDS-PM application to estimate the variability in exposures to ambient PM_{2.5} for the population of an urban metropolitan area. The PM_{2.5} concentration input file includes daily, 24-hour average PM_{2.5} concentrations for 1 year from a monitor located in an urban area. A representative population from census tracts near the monitor is simulated, and includes all ages and both genders. This scenario defines several microenvironments with different infiltration characteristics for ambient PM_{2.5} (indoor PM sources are not included in this scenario). Analysis of the model results focuses on options available for displaying the output to characterize the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #1. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

Yes

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the population variability in PM exposures and the impact of human activities?*

Comments:

Yes, they provide the basic analysis tools.

4) Perform Scenario #2 (Population Variability with Uncertainty).

This scenario demonstrates a SHEDS-PM application to characterize the uncertainty associated with the model estimates of population variability in ambient PM_{2.5} exposures. The same input PM_{2.5} concentration data and population demographics as Scenario #1 are used. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model. Analysis of the model results focuses on displaying the estimated uncertainty in the population distribution of exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #2. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

Yes

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the predicted uncertainty in the population variability of PM exposures?*

Comments:

Yes, they provide the basic analysis tools.

5) Perform Scenario #3 (Spatial Variability).

This scenario demonstrates a SHEDS-PM application for understanding the spatial variability in PM_{2.5} exposures. The PM_{2.5} concentration input file includes PM_{2.5} input concentrations for multiple monitoring locations within an urban area. Commuting is included to account for time spent outside the home census tract when individuals are at work. A representative population for each monitor is simulated. Analysis of the model results focuses on options available for displaying the output to understand the spatial variability in PM exposures due to concentration differences between monitors.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #3. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

Yes

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the impact of spatial and temporal variability in PM concentrations on the modeled distributions of PM exposures?*

Comments:

Yes, they provide the basic analysis tools. We could not get the program to print maps centered on the page, regardless of the print setup instructions.

6) Provide summary assessment.

Please provide comments on the following:

- a) *The organization and usability of the user interface (GUIs), which features or options were most useful, and whether additional features or options are needed*

Comments:

The graphic user interface is well designed and provides for user control of many model inputs. Because “GUI input only” models inherently limit the user’s control of input parameters, we prefer designs where as many inputs as possible are read from input files (databases) rather than imbedded in the model, and where the model input files can be created from a GUI, preprocessors (where users can examine the outputs), or by a text editor. The User’s Guide and GUI are designed for a fairly unsophisticated user (perhaps at the expense of the flexibility and control more experienced users might want). For example, they don’t provide instructions for (1) how to input different time activity data or (2) how to

use non-US census population data or the 2010 census data (when it becomes available) or census block or block group data instead of census tract data.

- b) Whether the descriptions of the model components and algorithms in the User Guide are sufficiently clear, technically correct, and represent the current state of the science for performing exposure assessments*

Comments:

The SHED-PM model is nicely packaged and includes many design features needed for state of the science exposure assessments. Several shortcomings are worth noting.

- 1) One of the problems with the current version of SHEDS is that the embedded CHAD database is outdated with respect to current activity patterns. Clear evidence of the strong temporal changes in activity patterns can be seen in comparison of 1981-82 with 2002-03 activity patterns Tables 16-49, 1650 of the Child-Specific Exposure Factors Handbook. The tables show the shift away from outdoor sports activity to indoor activities related to computers. This trend can be expected to be more pronounced now. The extent to which current estimates of exposure are biased due to this are unknown.
- 2) CHAD includes data from different studies and the current model framework does not allow the user to easily select the portions of the CHAD data base that may be suitable for a given application. The user's guide also does not indicate how the user would specify an alternate (non-CHAD) time-activity database for use in model calculations.
- 3) Another problem is that the MET assignments do not reflect the full range of conversion data that are in the literature. Use of kcal overestimates oxygen utilization, since it includes body fat in the calculation. Ideally, estimates should be based on lean body mass. If such data are not available, then estimates of lean body mass for BMI along with error distributions should be provided. Users should be allowed to specify inputs provided that the following criteria are met:
 - a. If the data are published, a citation needs to be provided.
 - b. If data are unpublished, they must be available to the public
 - c. At a minimum the data should be specific to age, sex
 - d. Estimates of error distributions need to be provided
- 4) More attention needs to be given the basis for selection of parameters for estimating microenvironmental concentrations. Care should be taken to carefully select the parameters given as the default values or sample problem values because these will likely be used without evaluation by many potential users. It is important to explain the process and types of data needed to select the parameters for various types of applications (and regions). In fact, there is probably a need for a companion document on exposure modeling, that provides scientific guidance and tutorials.

- c) Whether the output generated by the model are technically correct and consistent with descriptions of the algorithms in the User Guide*

Comments:

None of the outputs were obviously inconsistent with expectation. Determination of whether they are technically correct is very difficult from the stochastic simulations.

- 7) Rank priority for possible future improvements.**

Several possible improvements to the SHEDS-PM model are listed below. Please provide a number ranking for the relative priority that should be given to each improvement, using a scale from 1 (low priority) to 5 (high priority).

Improving ease of use:

- _5_ Create log file that records all inputs specified for the model run that can be viewed and saved by user*

It is not appropriate to release the model without having a log file that shows which inputs were used for a particular simulation because this is essential for quality assurance of individual simulations and, based on experience, crucial for large batches of simulations.

- _5/2_ Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)*

Saving the user's setting can help with consistency in multiple runs (high priority). Most users will use other software for comparison of outputs from multiple runs so this is a low priority.

- _5_ Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)*

This should be easy to implement and worthwhile given that the model's long run times.

- _3_ Provide more information on error messages to help users identify the reason for the error for common problems*

We did not test the model enough to encounter errors so it is difficult to evaluate this option.

- _5_ Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)*

Perhaps include options for age of housing stock and frequency of window openings and air conditioner use.

Allowing additional user specification of inputs:

- _4_ Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)*

Allow user specification from GUI or input file or database.

- _3_ Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)*

This feature is scientifically desirable but only useful if studies are conducted to collect and analyze sufficient supporting data for credible specification of these parameters in different types of applications.

Improving/refining model algorithms:

- _3_ Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of "unique" diary being used repeatedly for an individual)*

While this would be desirable, it would be justifiable only if we had more data on true longitudinal activity data—i.e., the relation between any given day's activity to any other day corrected for season, age, sex. While newer methods for assignment of activity take in to account autocorrelation in activity patterns, the databases for estimation are generally quite small (e.g. only 163 children from southern California for whom 48 observations/child are available in Glenn, G., et al. JESEE, 2008). Currently available longitudinal database cannot be assumed to represent the broad spectrum of subjects (children and adults) and the myriad environments in which they carry out outdoor activities.

_2__Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)

Absent data from more subjects from different climates with longer time series of activities, it is not clear that there is any benefit from increased sophistication.

*_5__Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)
Given the data present in Özkaynak, et al. (Figure 5, Atmos Environ 2009), this would be an absolute necessity. These data show considerable uncertainty over the percentiles of exposure such that any dose estimates, independent of the uncertainties and variability of the estimates on their own, are suspect from the start.*

*_5__Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)
At least one recent publication (Wu et al 2009 Atmos Environ 43, 1962–1971.) suggest census block groups or blocks are needed to capture the extremes of the exposure distributions for traffic related PM.*

_5__Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

This is a high priority because published data indicate window position and air conditioning use, both of which are related to temperature, as well as building age have large influences on residential air exchange rates.

Adding new functionality:

_1__Option for using mapping tool to select census tracts for simulation based on a map
_5__Add more user options to map view of output (e.g. for use in GIS software or Google Earth)

Other:

_3__Please describe: EPA should consider making the modeling system open source to encourage innovation and testing of new algorithms. This would also provide transparency that can enhance its credibility. Going open source could help build a community of knowledgeable developers and users that could expand the software platform to other pollutants and regions, and subject the software to more testing..

8) *Open comments (optional)*

Please provide any additional comments that you wish to on the SHEDS-PM model.

Comments:

None

III. SPECIFIC OBSERVATIONS

Provide specific observations, corrections, or comments on the document, mentioning page, paragraph, and/or line number.

Comments:

None

Clifford P. Weisel, PhD

Peer Review Comments on EPA's Draft Model

Stochastic Human Exposure and Dose Simulation for Particulate Matter (SHEDS-PM) Version 3.5

Clifford P. Weisel, Professor, University of Medicine and Dentistry of New Jersey, Environmental and Occupational Health Sciences Institute, November 30, 2009

I. GENERAL IMPRESSIONS

Provide overall impressions (approximately 1/2 page in length) addressing the accuracy of information presented, clarity of presentation, and soundness of conclusions.

Comments:

Overall the SHEDS model was simple to use within the settings provided, i.e. all input files being provided. It appeared to generate valid data sets based on the input, with the few exceptions or questions noted below in the response to the charge questions. The framework of plots and summary tables that are available allow for a rapid examination of different trends in the data so that potential variations in the PM ambient concentration, exposure and dose can be easily compared as well as the levels present in various microenvironments. The mapping capacity provides a visual idea of the exposures across a region and can provide individual census tract information.

The User Guide is written clearly and in detail to provide the user with the necessary guidance to run the SHEDS model. (There is sometime too much detail or redundancy, though that is better for those that need it and can be skipped over by individuals who have worked with this type of model previously.)

The model appears to provide a state of the science approach for rapidly modeling distributions of exposures when the input data are available and can result in sound conclusions about exposures in many regions of the country.

II. RESPONSE TO CHARGE QUESTIONS

Provide narrative responses to each of the eight charge questions below.

1) Install the SHEDS-PM model software program.

Install the model on a computer with Windows XP or later operating system using the file provided ('EPA SHEDS-PM 3.5 Installation.EXE') and following instructions in Section 2 of the User Guide.

a) Did you encounter any problems using the self-installing executable program to set up the model on your computer?

If yes, please describe the problem, the type of computer used, the operating system release number, the location the model was installed on the computer (e.g.

'C:\Program Files' or other drive), and whether the User Guide provided information to help correct the issue.

Comments:

No problems were encountered with installing the software on two computers, one running Window XP and a second running VISTA when following the directions provided.

- b) *Do you have any suggestions for improving the User Guide section on the model installation procedures (Section 2)?*

Comments:

User Guide section is clear

I did encounter a problem when running the program for the longer time period (overnight) in that my computers, as is the case for many, are scheduled to do updates of windows and other resident programs during the night. On both computers one of the updates required an automatic restart of the computer. This resulted in a loss of the results obtained from runs, which for a run that takes hours can be at least an annoyance. I therefore had to turn off the scheduled update options on my computer when running the 24+hour runs so as not to lose the results prior to my review of the analysis results. I suggest this be indicated in the installation section AND in other parts of the manual unless it can be fixed.

2) Perform SHEDS-PM Example Test Run.

Set up and run the example described in Section 3 of the User Guide using the PM concentration input file provided in the 'Data' directory ('philaPM2008.csv'). Display and export the model results as described in Section 3.

- a) *Does the example test run provide a sufficient introduction to familiarize the user with the SHEDS-PM model structure, graphical user interface (GUIs), and type of output generated by the model?*

Comments:

a) The example run was adequate for an initial "tour" of the input screens, though the figures in the printed manual showing screen images are reduced to an extent that I found it difficult to read some of the numbers for comparison purposes. On page 19, in item C "In-Vehicle Macroenvironment" (shouldn't it be Microenvironment?) the value for MEAN is missing from the instructions, but since all other values and the default were 0, I used that. However, it should be added to the text. The GUI user interface is relatively easy to use, particularly since it was designed to be linked to the output generated by the SHEDS model so minimal keystrokes and decisions need to be made to see some very common type outputs that are of interest. The tradeoff for this specially develop output tool is its limited in options in the way the graphics are

presented. However, that approach is acceptable since the data can be exported if more detailed analyses or graphics are desired.

One minor issue that occurred was if I tried to plot any data prior to pressing the RETREIVE button from the Data Analysis Screen an error message was displayed which continued to be displayed after retrieving the data unless I exited the GUI screen and restarted the data analysis. However, it did not require a new RUN so was not that time consuming, though a fix should be attempted.

The example, while providing a PM model structure does not require the user to construct the PM data file nor does it provide any guidance on the criteria for selecting the input values, rather the example just provides the values. This is fine for instruction on how to use the screens, which appears to be the focus of the example. There should be a section that provides insight into the PM input values. The description of the file structure for the PM data file provided in the Appendix of the manual is clearly written so should provide the needed directions.

3) **Perform Scenario #1 (Population Variability).**

This scenario demonstrates a typical SHEDS-PM application to estimate the variability in exposures to ambient PM_{2.5} for the population of an urban metropolitan area. The PM_{2.5} concentration input file includes daily, 24-hour average PM_{2.5} concentrations for 1 year from a monitor located in an urban area. A representative population from census tracts near the monitor is simulated, and includes all ages and both genders. This scenario defines several microenvironments with different infiltration characteristics for ambient PM_{2.5} (indoor PM sources are not included in this scenario). Analysis of the model results focuses on options available for displaying the output to characterize the effect of population variability in human activities on exposure to ambient PM_{2.5}.

Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #1. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

Comments:

Model Run 1-1

- a. The model for the Scenario #1 performed as expected based on the instructions given.

The following is a summary of the output results I obtained.

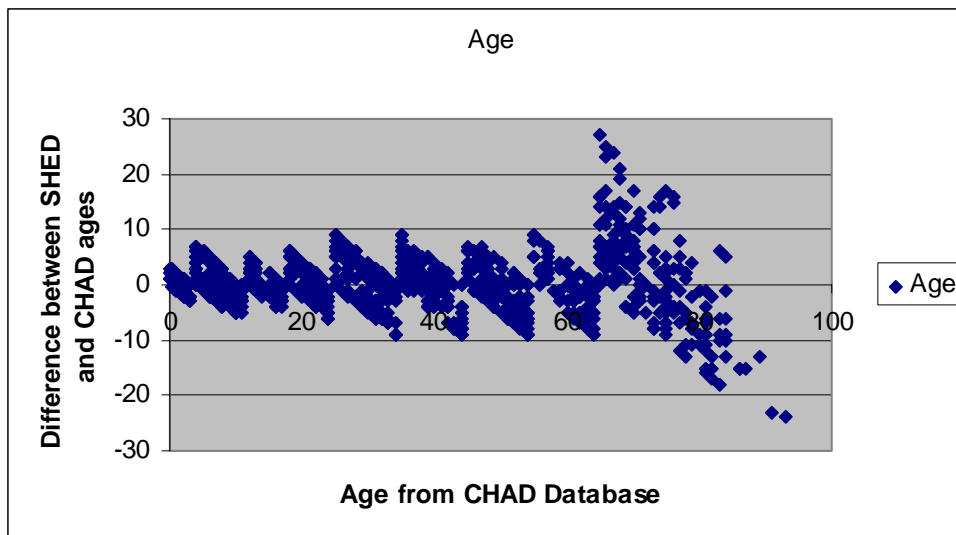
Comparison with input data

- 1) Compare Frequency Statistics for Gender Age and Employment Status – The results were similar, though not identical for the Gender, Age and Employment

Status since different seed numbers were for the random number generator (see attached table for comparisons)

- 2) The season numbers associated with dates were confirmed to be:
Season 1: 12/1 to 2/28; Season 2: 3/2 to 5/31; Season 3: 6/1 to 8/31 and
Season 4: 9/1 to 11/30
- 3) Different CHAD activities were assigned to each individual with the total minutes for each individual adding to 1440 representing the number of minutes in a day.
- 4) Gender, age and employment matched the criteria from the CHAD diary ID.
- 5) Ave Input PM Conc for each day matched input data file
- 6) Comparison of SHEDs data file for Associated _CHAD_ID for diary matching criteria in CHAD database.

Age – The ages stay within the age groups as specified in Appendix D. The graph below show the difference between the ages in the two data base with the age listed in the CHAD data base. It is noted that cutoffs were used to assign an age rather a distribution or distance from the mid-point of an age group, so that someone at age 15 could be assigned age between 10 and 15 but not 16. I realize this is a conscious choice but should be reviewed to make sure it represents the best approach.



All genders were correct.

For Employment. The CHAD database included an X in the employed/unemployed column for a subset of individuals, most of who were under 16 so did not apply to this analysis. However, several were 16 or older. In those cases three individuals (WAS94423A CIN00918A, NHA18412A) were marked employed and four were marked unemployed (NHW12580A, WAS85660A, NHA11808A, NHA15380A) by SHEDS. My understanding is the 'X' signifies a missing value and as per the information provided in the manual missing employment is assigned in a random fashion based on the probability distribution. The assignments are consistent with the numbers employed and not

employed though this is too small of a population to actually determine if the assignments are being made to reflect the population employment status.

For Weekday/Saturday/Sunday – all were assigned correctly

For Season – 10 of the 1000 were assigned an incorrect season based Dec-Feb Season 1, Mar-May #2, Jun-Aug #3 and Sep-Nov #4. The ones incorrectly assigned were NHA14696A, NHW17540A, NHW11902A, NHW15010A, NHW16501A, NHW19317A, WAS19768A, WAS19768A, NHW12609A, CIN40829C. They included all seasons, though not all months.

Average_Input_PM_Concentraitons were the same as the input file

Average_Total_PM_Concentrations were the same as the Total_Exposure and the Average_Input_PM_Concentrations

	Males	Females						
all	0.479	0.521						
age	Males	Females	age	Males	Females	age	Males	Females
0	0.006	0.008	40	0.023	0.023	80	0.004	0.013
1	0.021	0.012	41	0.021	0.006	81	0.004	0.010
2	0.013	0.013	42	0.010	0.006	82	0.002	0.000
3	0.023	0.010	43	0.008	0.010	83	0.008	0.010
4	0.017	0.010	44	0.013	0.008	84	0.000	0.002
5	0.004	0.013	45	0.031	0.012	85	0.002	0.000
6	0.025	0.010	46	0.017	0.012	86	0.000	0.000
7	0.015	0.010	47	0.010	0.013	87	0.000	0.006
8	0.008	0.019	48	0.021	0.021	88	0.000	0.002
9	0.004	0.017	49	0.006	0.017	89	0.004	0.004
10	0.017	0.025	50	0.019	0.008	90	0.000	0.002
11	0.013	0.015	51	0.019	0.015	91	0.000	0.006
12	0.010	0.004	52	0.015	0.006	92	0.004	0.002
13	0.017	0.010	53	0.010	0.004	93	0.000	0.000
14	0.013	0.015	54	0.008	0.010	94	0.000	0.000
15	0.006	0.010	55	0.004	0.012	95	0.000	0.000
16	0.004	0.008	56	0.006	0.004	96	0.000	0.000
17	0.008	0.017	57	0.010	0.008	97	0.000	0.000
18	0.006	0.015	58	0.017	0.004	98	0.000	0.000
19	0.013	0.012	59	0.008	0.006	99	0.000	0.000
20	0.010	0.008	60	0.013	0.012	100	0.000	0.000
21	0.008	0.015	61	0.013	0.008			
22	0.021	0.008	62	0.008	0.010			
23	0.025	0.015	63	0.000	0.012			
24	0.019	0.021	64	0.004	0.012			
25	0.023	0.013	65	0.015	0.010			
26	0.010	0.019	66	0.006	0.004			
27	0.004	0.017	67	0.002	0.010			
28	0.010	0.012	68	0.008	0.006			
29	0.002	0.006	69	0.006	0.012			
30	0.004	0.012	70	0.008	0.013			
31	0.017	0.013	71	0.013	0.010			
32	0.023	0.017	72	0.006	0.029			
33	0.013	0.027	73	0.023	0.010			
34	0.027	0.013	74	0.008	0.004			
35	0.013	0.013	75	0.004	0.006			
36	0.015	0.021	76	0.006	0.015			
37	0.008	0.013	77	0.004	0.017			
38	0.008	0.013	78	0.008	0.010			
39	0.025	0.000	79	0.008	0.008			

	Total	Male	Female			Male	Female
Unemployed	0.587	0.458	0.542	Unemployed	Age		
Employed	0.413	0.500	0.500		16-19	0.014	0.079
					20-21	0.042	0.000
					22-24	0.079	0.063
Unemployed		0.566	0.607		25-26	0.047	0.043
Employed		0.434	0.393		30-34	0.005	0.012
					35-44	0.154	0.063
					45-54	0.220	0.166
					55-59	0.065	0.067
					60-61	0.056	0.040
					62-64	0.028	0.055
					65-69	0.084	0.059
					70-74	0.070	0.123
					>75	0.136	0.229

Employed	16-19	0.073	0.043
	20-21	0.000	0.073
	22-24	0.085	0.043
	25-26	0.085	0.146
	30-34	0.238	0.244
	35-44	0.220	0.262
	45-54	0.171	0.116
	55-59	0.049	0.000
	60-61	0.000	0.000
	62-64	0.000	0.018
	65-69	0.000	0.037
	70-74	0.079	0.018
	>75	0.000	0.000

Model Run #1-2

Yes, model performed as expected.

The longitudinal assignment of the PM Concentration included 365 days for each individual and the PM concentration were assigned correctly (Ave_Input _PM_Conc. matched the input file data)

Scatter plots generated are given below and match the examples provided in Figure 6 and show consistency for other comparisons.

In this scenario run, a total of 12 different CHAD diaries should be used for each simulated individual since diary matching by season was selected. The 12 diaries include 3 different diaries for daytype (weekday, Saturday, and Sunday) times 4 different seasons specified using the default definition of seasons.

For #1 below, confirm that each simulated individual has the same 3 CHAD diaries repeated within a season - one used on weekdays, one used for Saturdays, and one used for Sundays.

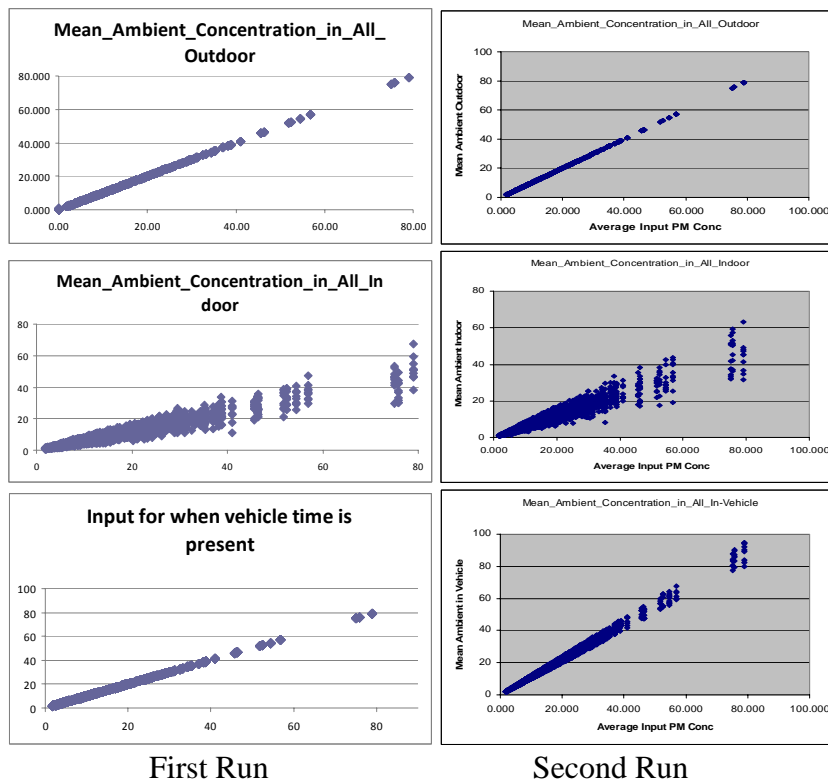
For #2 below, confirm that each simulated individual has one set of 3 CHAD diaries repeated for Season 1, another

set of 3 CHAD diaries repeated for Season 2, a third set for Season 3 and a fourth set for Season 4.

It appears to me that the same CHAD diary ID is used for each individual daytype (weekend, Saturday, Sunday) within a season but different ones are used for each daytype. The CHAD diary IDs are different for different seasons. This results in 12 different diaries being used.

Longitudinal assignment included 365 days for each simulated individual.

The model outputs were comparable to the input distributions, as per the following graphs except that the in-vehicle was identical to the input data not greater (Left set of figures). In examining this I had concerns that I may have overwritten the in-vehicle column when I was manipulating the data for other purposes before preparing the graph, so I ran the simulation again and verified for that run the in-vehicle were greater than the input PM concentration (Right set of figures).



Comparison of ratios XXX to input values with Input Distributions showing agreement

	Outdoor	Indoor	In-vehicle
Mean	1	0.598644	1.100153
Standard Deviation	0	0.100293	0.057509
Count	2385	3649	2908
Minimum	1	0.214989	1.00005

Maximum	1	0.89475	1.199986
Input		Normal	
Distributions	Fixed value	mean 0.6	Uniform Min
	of 1	Std Dev 0.1	1, Max 1.2

Comparison with exposure/dose calculations: (Excel spreadsheets used to evaluate these are attached)

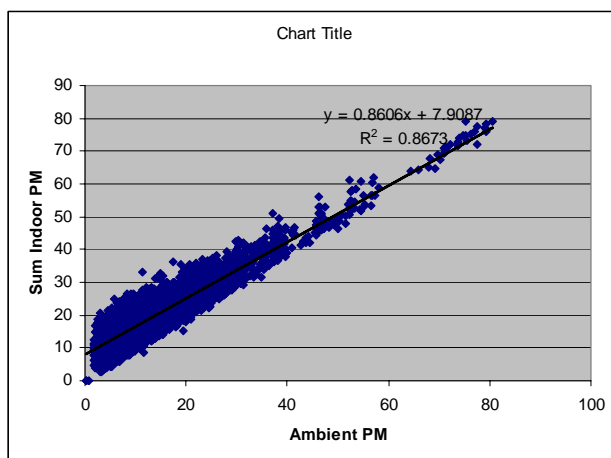
Each row did have the correct location based on the CHAD code except when the CHAD code was U or X, which I expect means missing, it was assigned ALL_INDOOR. The locations were only ALL_INDOOR; ALL_OUTDOOR or ALL_IN VEHICLE, no sub locations were specified in SHED. Time spent in each location for each record was correct. (this is based on 12 CHAD diaries used for the year for one individual.)

Each row in the event file had the correct microenvironmental PM concentration based on the daily export file.

The PM exposure for the diary events matched my hand (excel spreadsheet) calculations for individual events and the value for the exposure in the DAILY File matches the hand calculated sums for that day.

Using the ventilation rate that was in the SHED output the internal dose matched my hand calculations each record and when summed for the day matched the value in the DAILY file.

A Linear regression calculation comparing the Ambient PM Concentration with the Indoor air Concentration (determined by summing the Ambient and the non-Ambient Indoor PM concentration columns, since a total indoor air is not provided in the SHED output) did result in a regress that matched the input data as shown in the following figure.



Since I calculated the indoor air concentration by summing the ambient indoor air concentration with the non-ambient indoor air concentration columns, it does not make sense to be to “Confirm that the non-ambient PM contribution is calculated as described in Appendix D” which starts with the Indoor Air concentration and subtracts the Ambient component.

Home Mass Balance Calculation

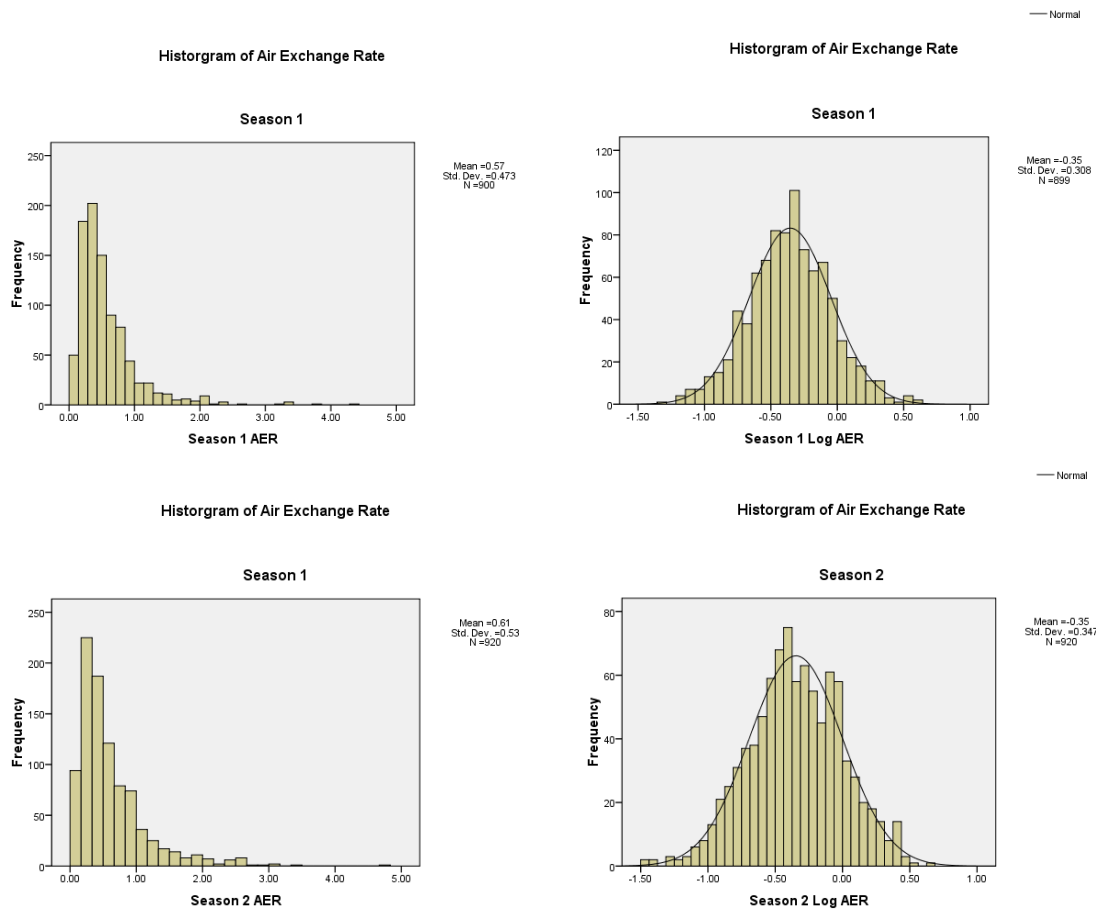
The air exchange rates were different on different dates, though some were within .001 of other date, but none were identical.

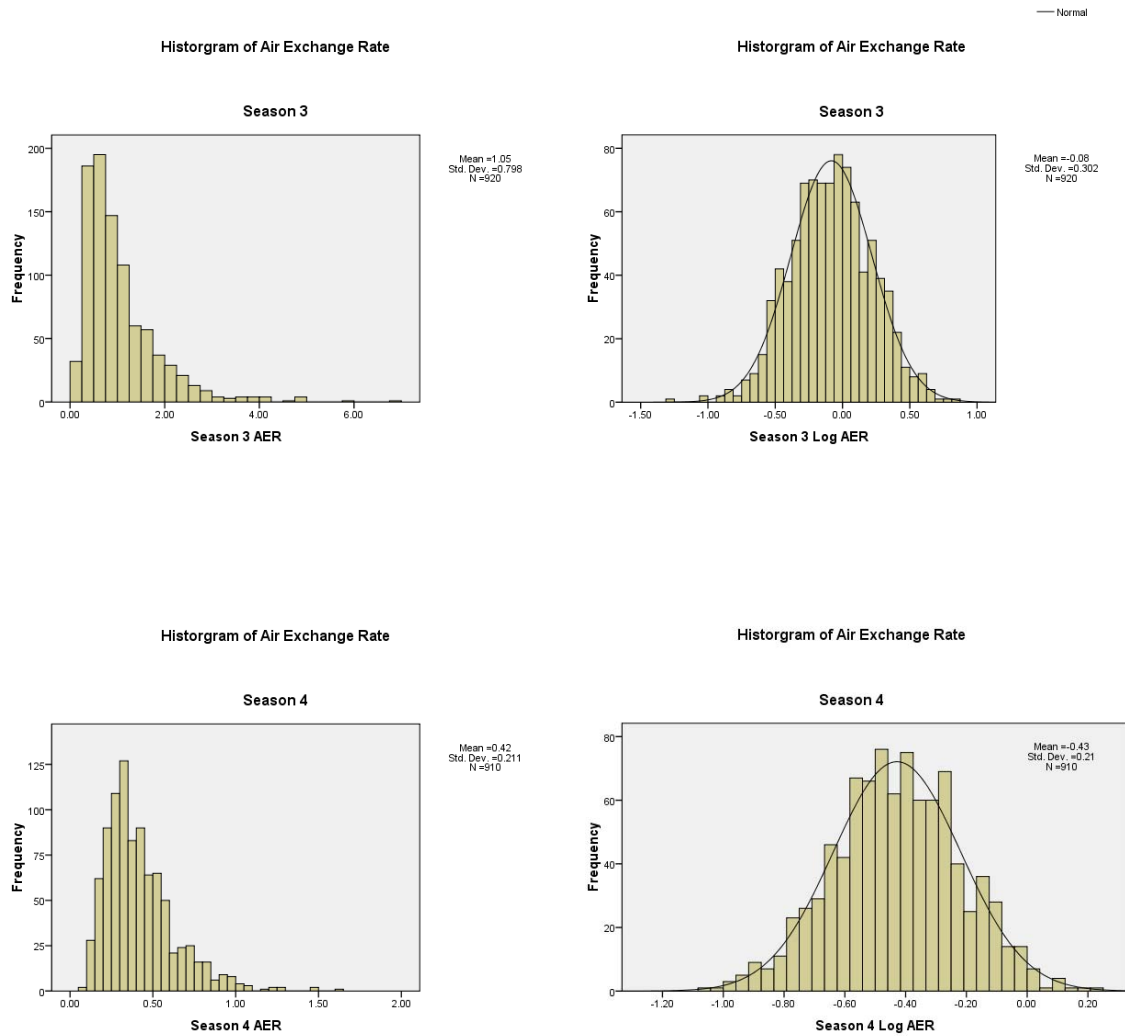
The summary statistics matched the input data. The distribution of each season is consistent with a log normal distribution (Left side figure AER, right side log transformed AER).

:

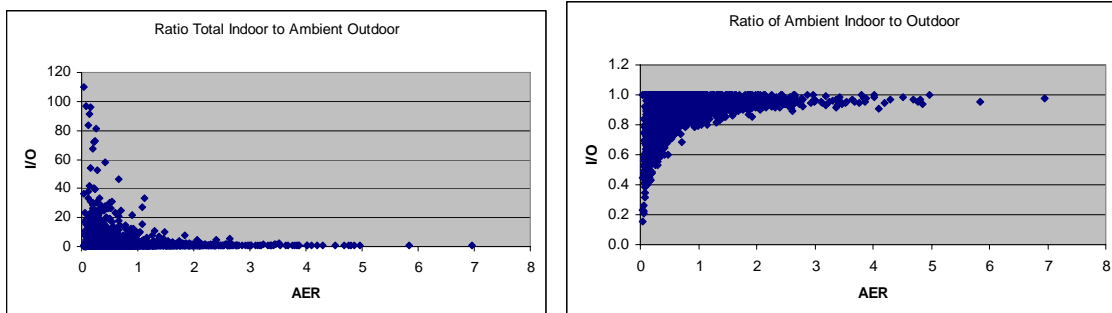
Season--

>	1	2	3	4
Mean	0.571	0.615	1.051	0.419
Std Dev	0.474	0.532	0.800	0.211





Plots of the AER vs the Indoor/Outdoor ratio for the total indoor and for the ambient portion of the indoor both show the expected distribution, as the AER goes up the I/O approaches 1, for the total Indoor from higher I/O ratios and for the ambient only from I/O below unity. (see figures)

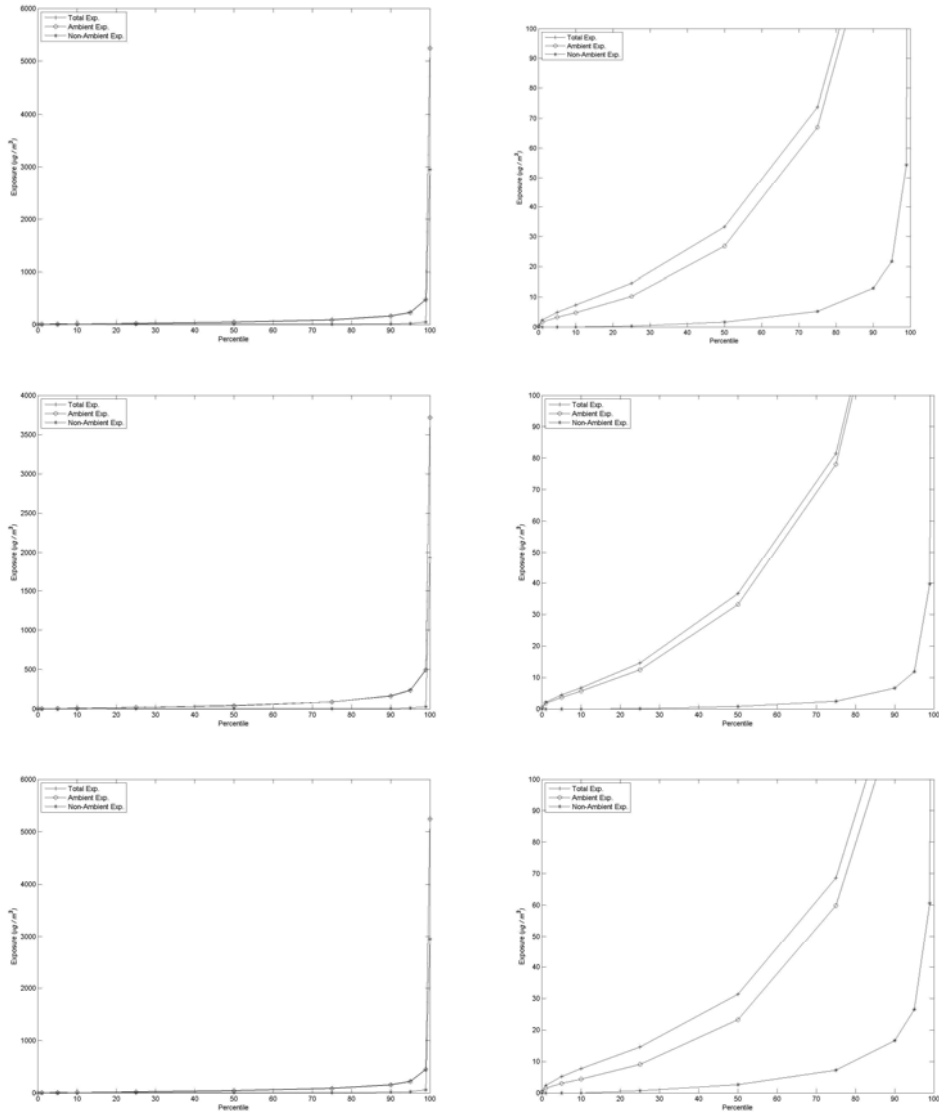


Confirm non-ambient contribution from cooking in home microenvironment.

A second run was done with cooking set to 0 resulting in a non-ambient concentration in the home of 0 while it was $9.53 \pm 18.78 \mu\text{g}/\text{m}^3$ in the run with cooking on. The in home ambient concentration for the two runs were $9.53 \pm 7.59 \mu\text{g}/\text{m}^3$ and $9.53 \pm 7.58 \mu\text{g}/\text{m}^3$ indicating that the runs produced similar results for the non affect air concentrations.

Model Run 1-3

The output distributions look reasonable and comparable to the plots in Figure 8 as below with the exception of in-vehicle ambient levels which were much higher than the other microenvironments:

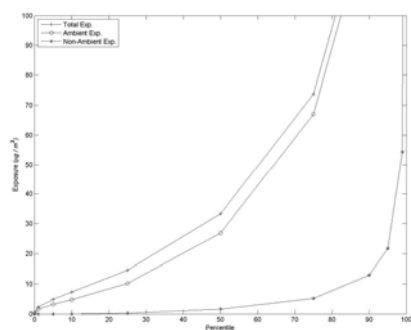
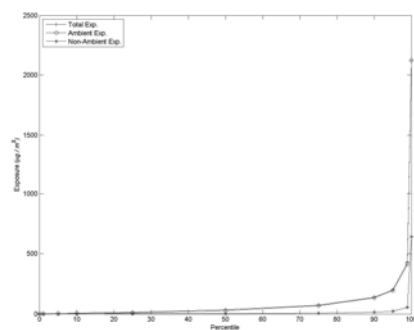
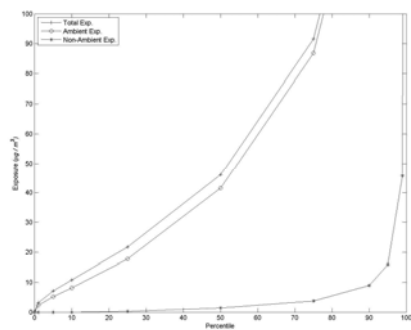
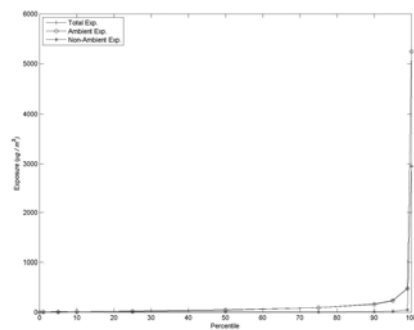
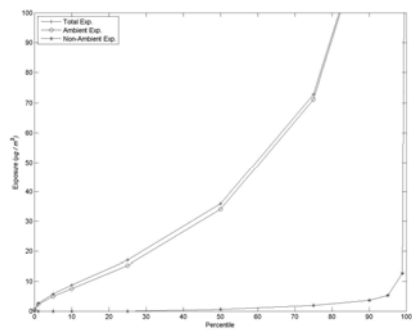
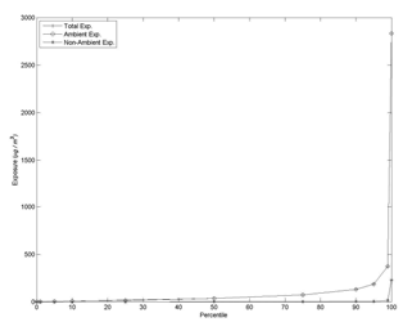


Distribution by gender – each with full data on left and scale expanded on right,

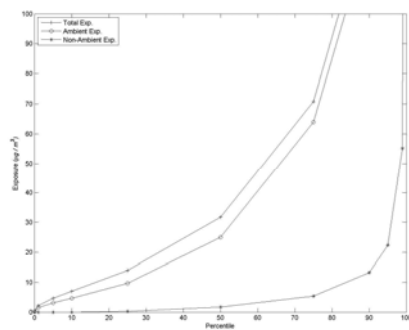
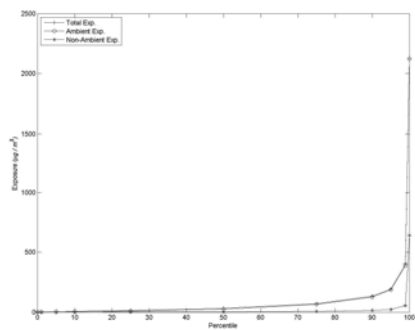
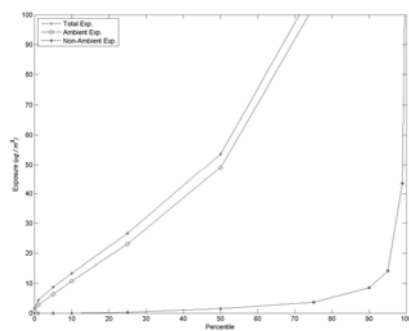
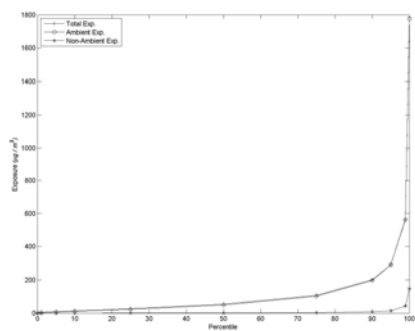
Top all data

Middle Males

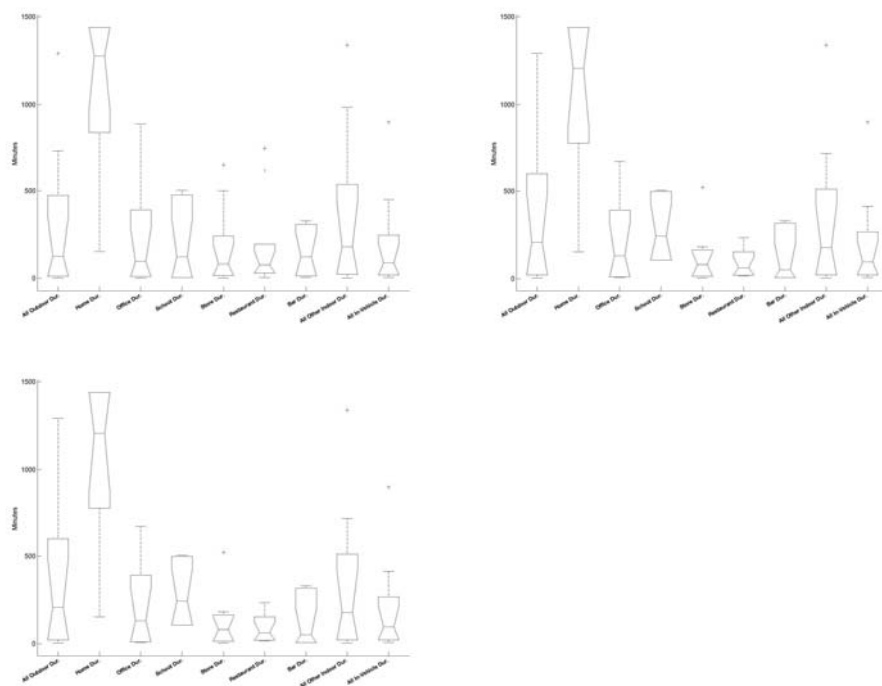
Bottom Females



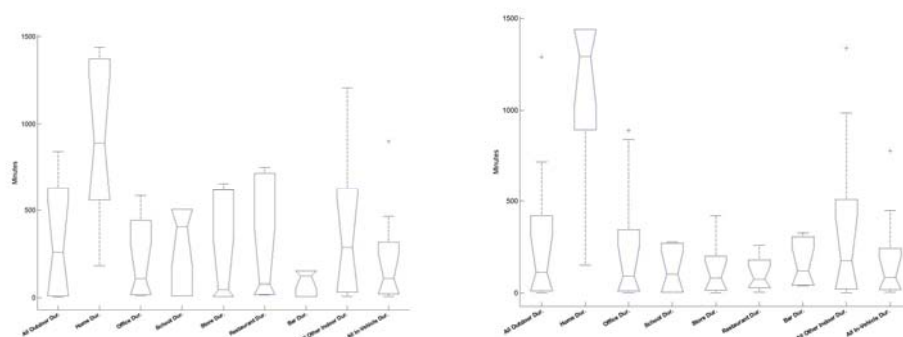
Distribution by age – each with full data on left and scale expanded on right,
 Top age 0-18
 Middle age 19-65
 Bottom age >65



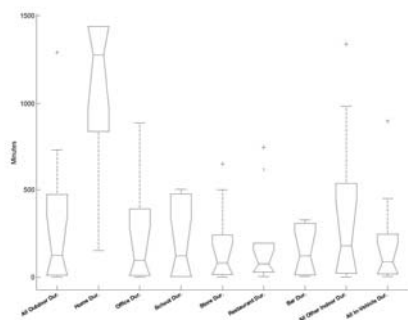
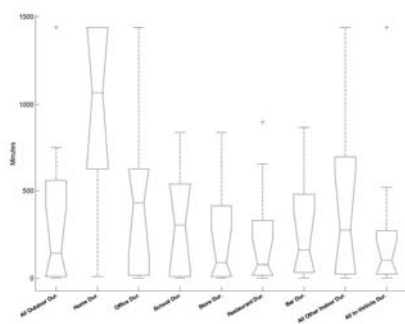
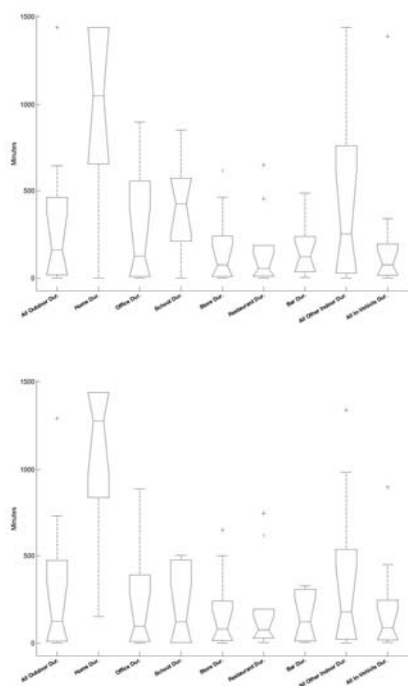
Distribution by employment – each with full data on left and scale expanded on right,
 Top Employed
 Bottom Unemployed



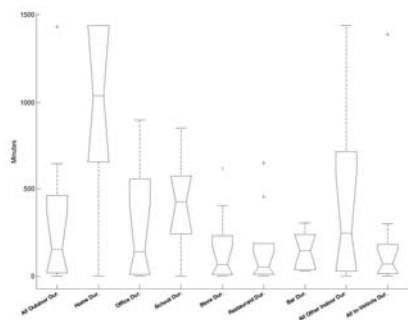
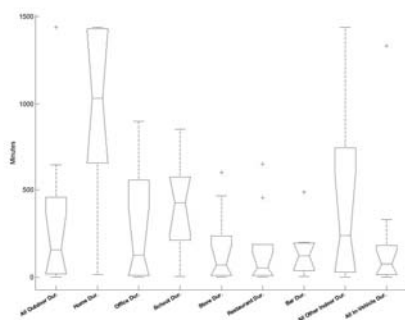
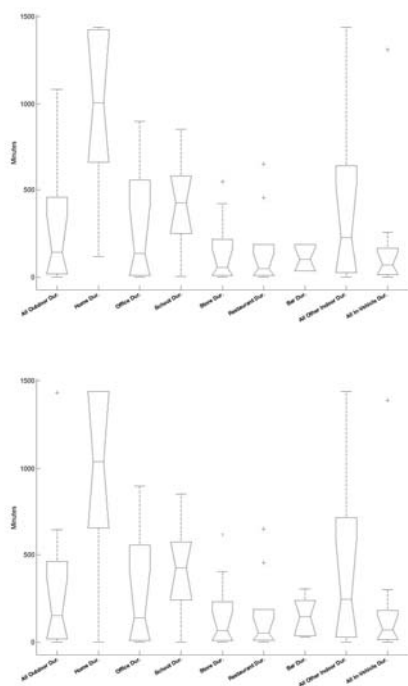
Time spent in microenvironments, top left - all subjects, top right – males, bottom left - females



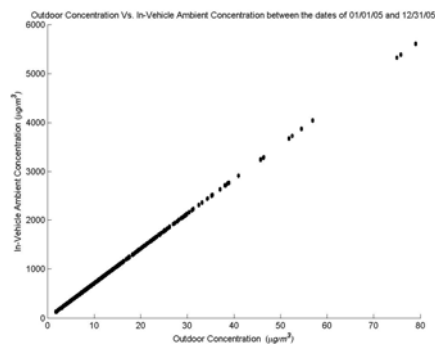
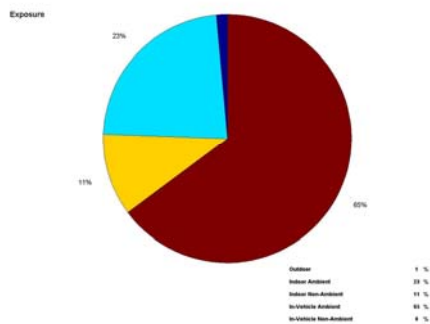
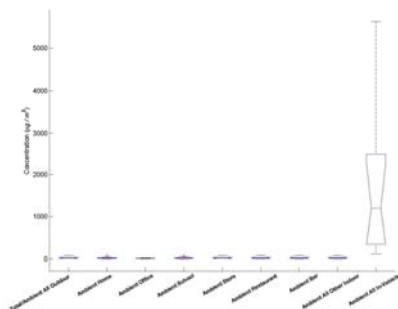
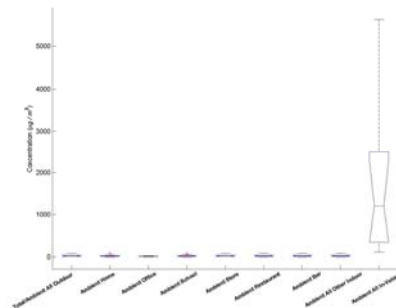
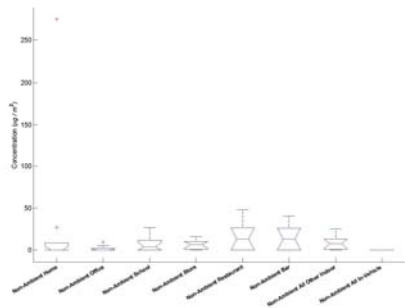
Time spend in microenvironment, left – employed, right – unemployed



Time spent in microenvironments, top left – age 0- 18, top right – age 19 - 65, bottom left – age > 66-102



Time spent in microenvironments, top left – weekday, top right – Saturday, bottom left – Sunday



Top Right Non-ambient PM Concentration by microenvironment – All data
 Top Left Ambient PM Concentration by microenvironment – All data
 Middle Ambient PM Concentration by microenvironment – ages 0-18
 Bottom Left Pie Chart of Exposure all Data
 Bottom Right Scatter plot of ambient air concentrations outside vs in-vehicle

b) Do the options for analysis of model results provide the user with sufficient information to understand the population variability in PM exposures and the impact of human activities?

Comments:

The tables that can be generated do give insight into the variability of the PM exposure and dose by gender, age, employment status (the properties examined) and presumably other choices for different microenvironments and days of the week.

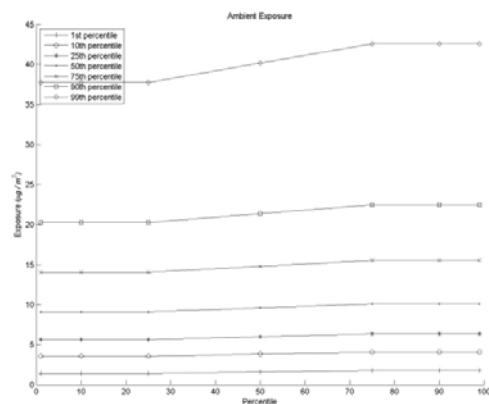
4) *Perform Scenario #2 (Population Variability with Uncertainty).*

This scenario demonstrates a SHEDS-PM application to characterize the uncertainty associated with the model estimates of population variability in ambient PM_{2.5} exposures. The same input PM_{2.5} concentration data and population demographics as Scenario #1 are used. This scenario involves specifying uncertainty distributions for the microenvironment infiltration parameters which are sampled during multiple iterations of the model. Analysis of the model results focuses on displaying the estimated uncertainty in the population distribution of exposure to ambient PM_{2.5}.

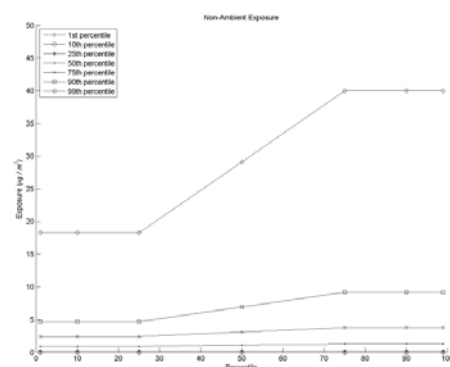
Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #2. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

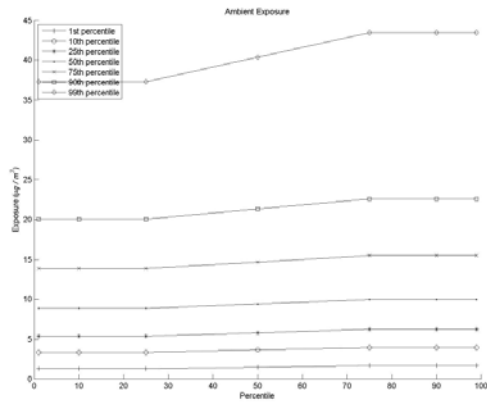
Yes the model performed as expected based on the instructions. Examples of the output are provided below. However, the exposure values were about half the levels in the graphs provided in Figure 9 and when the same settings were run a second time. In addition, the 50th percent value in the PM variability and Total Exposures are ~10 ug/m³ and ~15 ug/m³, respectively, which is about twice what appears to be the values in the uncertainty plot. The second run appears to be more consistent with the expected values, though the 50th percent of the variability plots of the total exposure (last set of figures) is ~18 ug/m³ and 50th percentile for the 50th percent on the uncertainty plot is ~1/2 that value. I do not know if this is the correct comparison between the variability and the uncertainty plots, but if so this discrepancy needs to be evaluated.



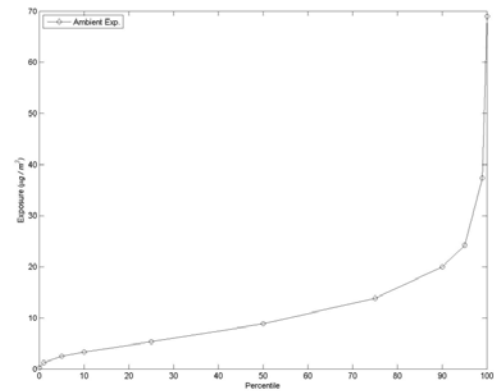
Ambient Exposure



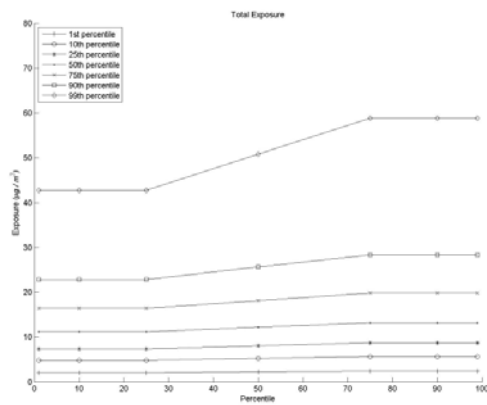
Non-ambient exposure



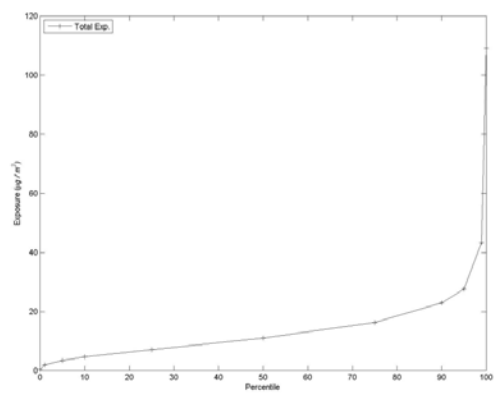
PM Concentration



PM Concentration Variability

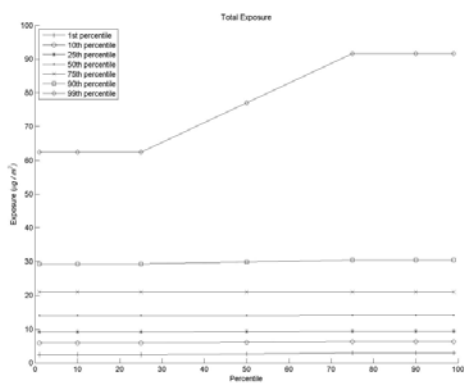


Total Exposure

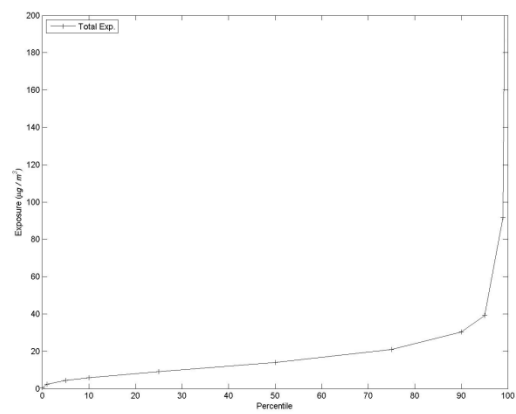


Total exposure variability

SECOND RUN OUTPUTS



Total Exposure 2nd run



Total Exposure Variability 2nd run

Comments:

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the predicted uncertainty in the population variability of PM exposures?*

Comments:

Yes, though I prefer the plot style in Burke et al 2001 figure 4.

5) Perform Scenario #3 (Spatial Variability).

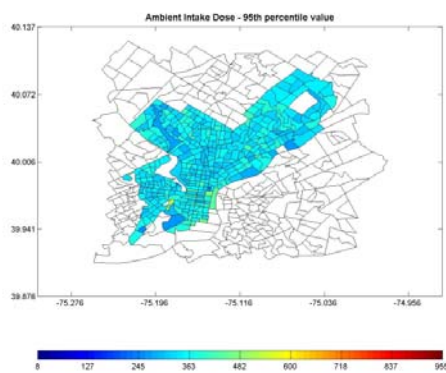
This scenario demonstrates a SHEDS-PM application for understanding the spatial variability in PM_{2.5} exposures. The PM_{2.5} concentration input file includes PM_{2.5} input concentrations for multiple monitoring locations within an urban area.

Commuting is included to account for time spent outside the home census tract when individuals are at work. A representative population for each monitor is simulated. Analysis of the model results focuses on options available for displaying the output to understand the spatial variability in PM exposures due to concentration differences between monitors.

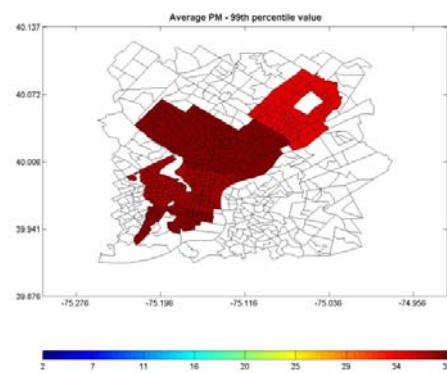
Follow the procedures outlined in the Appendix for specifying the model inputs and analyzing the results for Scenario #3. Provide comments on the following for this scenario:

- a) *Did the model perform as expected based on the instructions in the Appendix and information in the User Guide?*

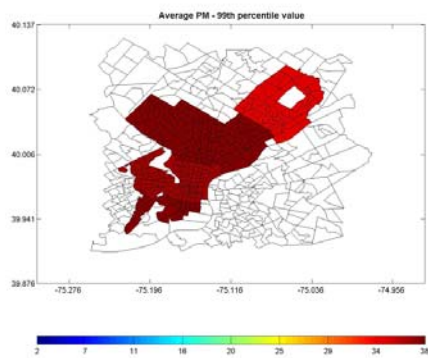
Yes, as shown in the figures below:



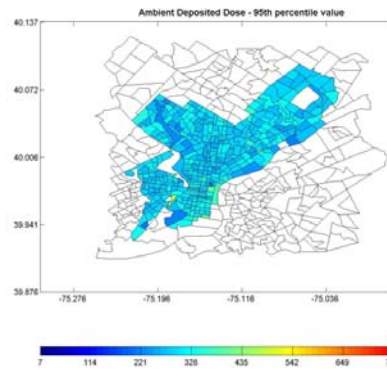
Ambient Intake Dose 90th



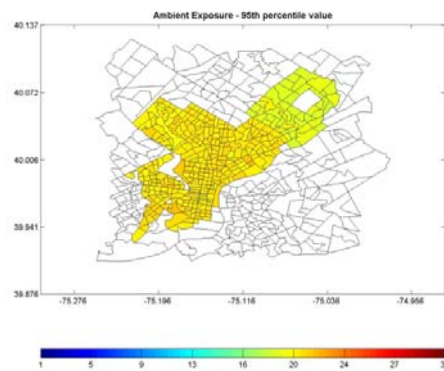
Average PM 99th



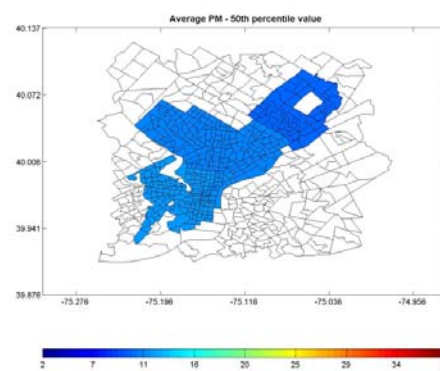
Average PM 99th



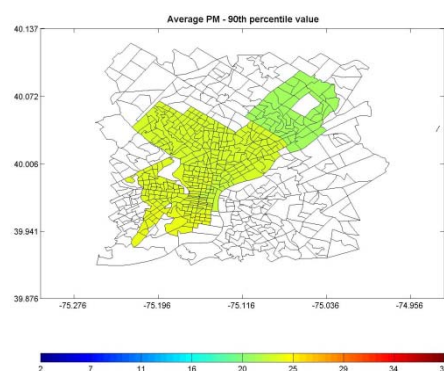
Ambient Deposited Dose 95th



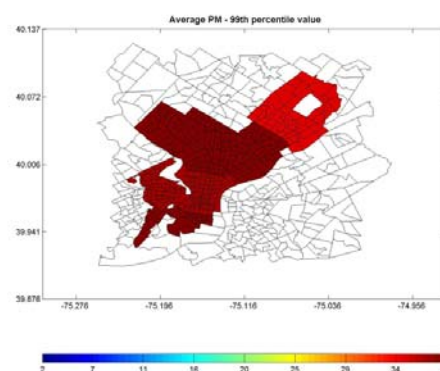
Ambient Exposure 90th



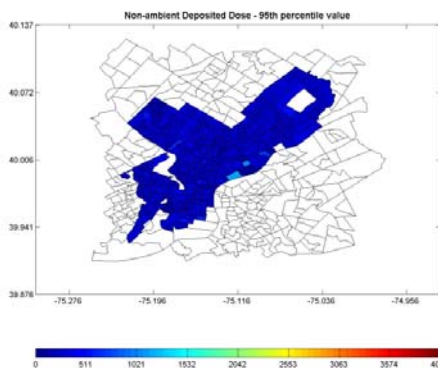
Average PM 50th



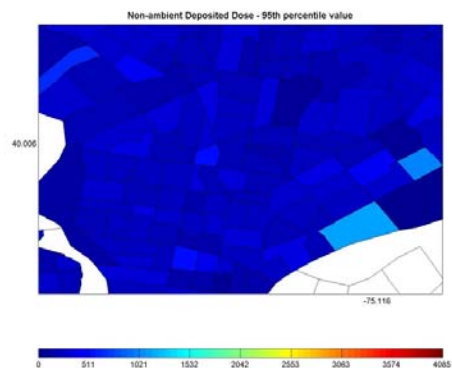
Average PM 90th



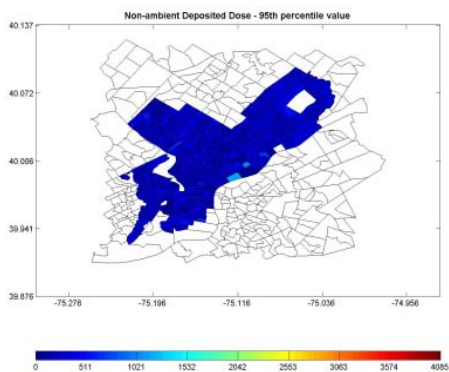
Average PM 99th



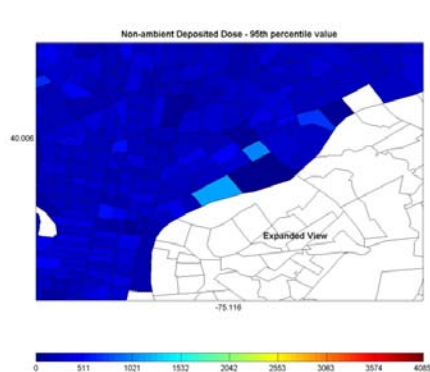
Non ambient deposited dose 95th
expanded



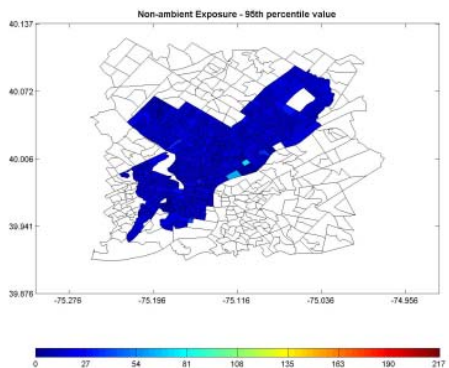
Non-ambient Deposited Dose 95th



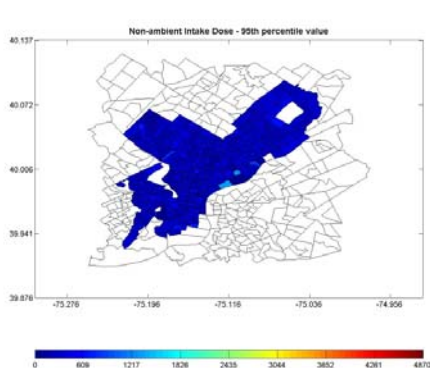
Non ambient deposited dose 95th



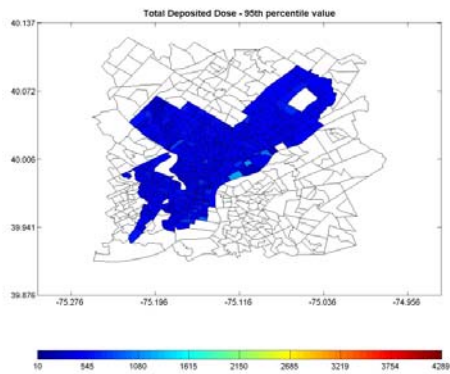
Non ambient 95th expanded



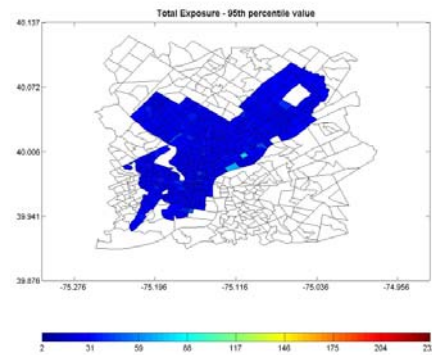
Non ambient exposure 90th



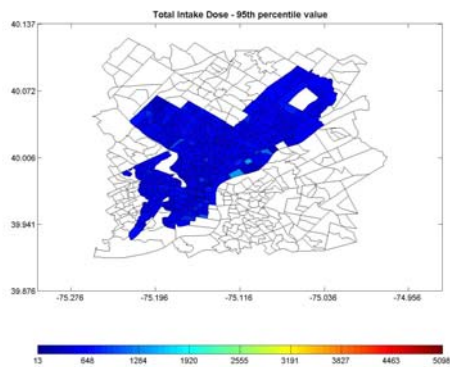
Non ambient Intake dose 90th



Total deposited dose 95th



Total exposure 90th



Total intake dose 90th

- b) *Do the options for analysis of model results provide the user with sufficient information to understand the impact of spatial and temporal variability in PM concentrations on the modeled distributions of PM exposures?*

Comments:

Maps provide insight into spatial variation along with the specific values in individual census tract when the cursor is moved over it.

6) Provide summary assessment.

Please provide comments on the following:

- a. *The organization and usability of the user interface (GUIs), which features or options were most useful, and whether additional features or options are needed*

Comments:

Overall, the GUI interface was easy to use and allow for easy visualization of individual patterns across concentrations, exposure and dose of ambient and non-ambient sources as well as across different microenvironments. I found the ability to compare different microenvironments most useful if I wanted to better understand

where exposures were occurring and how the exposures and times spent in different microenvironments varies across age, gender, season employment status, day type and smoking. The scatter plots provide some insight into underlying associations between different exposures. A mechanism to plot distributions of ratios directly of different outcomes and variables to complement the scatter plots might be worth considering.

- b. *Whether the descriptions of the model components and algorithms in the User Guide are sufficiently clear, technically correct, and represent the current state of the science for performing exposure assessments*

Comments:

The User Guide is clear in its description of the modeling algorithms used and the combination the multiple microenvironments using the CHAD data base along with microenvironmental air concentration to generate distributions of exposure that include uncertainty estimates represent current state of science for performing exposure assessment. The inclusion of Mass Balance for estimating air concentrations in the indoor environment is a strong advance that potentially increases the potential to make the model output region specific if appropriate input factors are available.

One item that is not clear to me is the assignment of the non-ambient contributions to concentrations, exposure and dose. On page 128 it indicates for the linear regression equation and scaling factor approaches that if the $C_i/C_{\text{ambient}} < 1$ then the “non-ambient” is assumed to be zero. For particles there will be losses as the particles enter the buildings (as discussed in the mass-balance equation approach which will result in that ratio being less than unity in most conditions. While it is recognized that some of the microenvironments do not have enough data to employ the mass balance model, it may still be possible to provide a better estimate of the C_i/C_{ambient} ratio than unity for assigning a non-ambient component to the PM concentration and therefore exposure and dose than when using the current assumption.

- c) *Whether the output generated by the model are technically correct and consistent with descriptions of the algorithms in the User Guide*

Comments:

The model outputs appear to be technically correct and consistent with the algorithms in the User Guide. The example provided showed the location of the monitors, but they are not in the figures from my runs.

7) *Rank priority for possible future improvements.*

Several possible improvements to the SHEDS-PM model are listed below. Please provide a number ranking for the relative priority that should be given to each improvement, using a scale from 1 (low priority) to 5 (high priority).

Improving ease of use:

- _4__ Create log file that records all inputs specified for the model run that can be viewed and saved by user
- _5__ Add capability to save user specified settings and recall output for analysis for multiple runs (only data for most recent run is available for analysis in current version of model)
- _1__ Add capability to turn off dose calculations (to decrease model run time when user is only interested in estimating exposures and not dose)
- _3__ Provide more information on error messages to help users identify the reason for the error for common problems
- _2__ Provide more default values for locations-specific parameters of home mass balance equation (i.e. air exchange rates, home volumes)

Allowing additional user specification of inputs:

- _3__ Add GUI screen for user specification of physiological parameter distributions (e.g. age/gender specific basal metabolic rates, lung parameters, METS distributions)
- _4__ Allow selection of mass balance option for any microenvironment (currently limited to home microenvironment only)

Improving/refining model algorithms:

- _4__ Add more diary sampling to current longitudinal diary algorithm to include a pool of diaries for each simulated individual rather than a fixed set of diaries (to reduce impact of “unique” diary being used repeatedly for an individual)
- _3__ Add more sophisticated algorithm for combining activity diaries from CHAD in longitudinal simulations that uses correlation in activities day-to-day for each individual (requires development of default values and guidance to users in addition to code modifications)
- _1__ Add uncertainty to deposited dose algorithm (requires development of uncertainty distributions for parameters of dose equations in addition to code modifications)
- _2__ Add flexibility to use census tracts, block groups, or blocks (requires expanding census input databases for population demographics)
- _5__ Add algorithm for estimating air exchange rate in home mass balance equation that depends on home characteristics and daily temperature instead of sampling from a distribution

Adding new functionality:

- _2__ Option for using mapping tool to select census tracts for simulation based on a map
- _4__ Add more user options to map view of output (e.g. for use in GIS software or Google Earth)

Other:

____Please describe:

8) Open comments (optional)

Please provide any additional comments that you wish to on the SHEDS-PM model.

Comments:

III. SPECIFIC OBSERVATIONS

Provide specific observations, corrections, or comments on the document, mentioning page, paragraph, and/or line number.

I did encounter a problem when running the program for the longer time period (overnight) in that my computers, as is the case for many, are scheduled to do updates of windows and other resident programs during the night. On both computers one of the updates required an automatic restart of the computer. This resulted in a loss of the results obtained from runs, which for a run that takes hours can be at least an annoyance. I therefore had to turn off the scheduled update options on my computer when running the 24+hour runs so as not to lose the results prior to my review of the analysis results. I suggest this be indicated in the installation section AND in other parts of the manual unless it can fixed.

The Push Button for “View/Edit Model Run Inputs” is some times listed as “Edit/View Model Run Inputs” in the text (e.g. page 27 and 28) rather than “View/Edit Model Run Inputs” as is appears on the screen.