# Planned Methodologies for Extending SHEDS-Multimedia version 3 (aggregate)

# to SHEDS-Multimedia version 4 (cumulative or aggregate)

**DRAFT**

**June 12, 2007**

**Disclaimer:** This report is currently undergoing EPA review and should not be considered final.

Prepared by: Graham Glen[1], Luther Smith[1], Valerie Zartarian[2], Casson Stallings[1], Kristin Isaacs[1], Jianping Xue[2]

[1] Alion Science and Technology

[2] US Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory

# TABLE OF CONTENTS

# INTRODUCTION

The Stochastic Human Exposure and Dose Simulation model for multimedia, multi-route/pathway chemicals (SHEDS-Multimedia), developed by EPA's Office of Research and Development (ORD), National Exposure Research Laboratory (NERL), is a state-of-science computer model for improving estimates of aggregate (single-chemical, multi-route/pathway) and cumulative (multi-chemical, multi-route/pathway) human exposure and dose. SHEDS-Multimedia is the EPA/ORD's principal model for simulating human exposures to a variety of multimedia, multipathway environmental chemicals such as pesticides, metals, and persistent bioaccumulative toxins. SHEDS-Multimedia version 3 is an aggregate residential (non-dietary) model focused on single chemical exposures from inhalation, dermal contact, and non-dietary ingestion. This model can help answer many useful questions related to populations' aggregate residential exposures for different multimedia chemicals and what factors and pathways are most important. It can be linked with other tools (e.g., dose estimation models, measurements) for reducing uncertainty in risk assessments.

Details of the SAS model implementation (Glen, 2007) are given in the SHEDS-Multimedia version 3 Technical Manual (Zartarian et al., 2007), and instructions for using the graphical user interface are given in the SHEDS-Multimedia version 3 User Manual (Stallings et al., 2007). SHEDS-Multimedia version 3 does not include the SHEDS-dietary module (Xue, 2007) that estimates exposure from eating food or drinking water. SHEDS-Multimedia version 4 will combine the residential and dietary components in the same model. Plans for SHEDS-Multimedia version 4 also include: extending version 3 to incorporate cumulative algorithms (multiple chemicals and their co-occurrence in space and time); a residential fugacity-based source-to-concentration module; an enhanced longitudinal diary assembly method; and a new sensitivity analysis option. Other improvements are also being considered.

This document describes the new methodologies and planned code changes for extending SHEDS-Multimedia version 3 to version 4. The following five sections, listed in no particular order, describe possible modifications to the existing version 3. Which of these methods and code changes will actually be included in version 4, will depend on end user and peer reviewer input, maximum anticipated impact for enhancing the science, available data, and available resources.

## References

Glen G (2007). SHEDS-Multimedia version 3 SAS code.

Stallings C, Zartarian VG, Glen G. (2007). SHEDS-Multimedia (Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway Chemicals) version 3 User's Manual, prepared for August 14-15, 2007 US EPA/OPP FIFRA Science Advisory Panel (SAP) meeting.

Xue J. (2007). SHEDS-dietary SAS code.

Zartarian VG, Glen G, Smith L, Xue J. (2007). SHEDS-Multimedia (Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway Chemicals) Technical Manual, prepared for August 14-15, 2007 US EPA/OPP FIFRA Science Advisory Panel (SAP) meeting.

# SECTION 1. Planned Methodology for Extending SHEDS-Multimedia to Address Multiple Chemicals

## Introduction

SHEDS-Multimedia version 3 simulates exposures of individuals to one chemical at a time. SHEDS-Multimedia version 4 will allow tracking exposures of individuals to multiple chemicals at the same time. Version 4 will be a cumulative version, but it could also be run in aggregate mode like version 3 for a single chemical. Unlike version 3 which has a single chemical focus, version 4 will have a "product formulation" orientation. Given that a single product may have multiple chemicals, certain chemicals will possibly be applied jointly quite often. In addition, a product-related co-occurrence priority system akin to the current version 3 co-occurrence approach (see the SHEDS-Multimedia version 3 Technical Manual) will likely be used to minimize the number of product combinations.

In version 3, the running exposures of the chemical are tracked in three carriers (air, surface residues, and dust/soil). However, the masses of the carriers themselves are not tracked. In version 4, the mass of each chemical and each carrier will be tracked (details below). "Carriers" are defined in SHEDS as chemical-containing substances that can be transferred onto or into the human body. Carriers to be included in version 4 are soil, dust, residue, air, food, and water. "Media" are defined in SHEDS as objects that hold the carriers before the carriers are transferred to the body. Examples of media are textured surfaces, smooth surfaces, and air in either treated or untreated rooms.

The basic operation of SHEDS-Multimedia will be unaffected by these changes, but the list of variables will be substantially longer. The current variables used for new exposure, running exposure, absorption, dose, and elimination will be replaced by vectors, with a numerical suffix indicating the position in the list of chemicals. Similarly, chemical-specific inputs will become vectors. The increase in the number of inputs may necessitate alterations to the current form of data entry in the model interface.

Some simplification is possible by assuming that the chemicals are not encountered independently, but through the process of contact with carriers. The focus on carriers will limit the number of input variables required, since the same values may be assumed to apply to all chemicals contained in a given carrier. For example, the user need only specify a bathing removal efficiency for soil, as opposed to a separate removal efficiency for each chemical in the soil . For exposure and absorption estimation, another simplifying assumption is that no chemical interactions occur.

## Tracking the carriers for the chemicals

As is the case in SHEDS-Multimedia version 3, the determination of chemical concentrations in various potential contact media will proceed independently of specific human diary events. For example, the post-application concentration option does not account for specific human activities at specific times. The output from the concentration modeling in version 4 will be a set of time series, one for each carrier and each chemical, for each potential contact medium. Suppose that for a given application scenario, a given model run uses 10 contact media, each potentially containing 1 carrier and 9 chemicals. Then a total of 100 concentration time series would be needed. This would be comprised of one time series for the carrier itself, and one for each chemical, in each contact medium. These would likely be (as they are now) hourly time series over the simulation period. One hourly time series for one year requires 80 Kb of storage in SHEDS-Multimedia version 3, so 100 time series would require less than 8 Mb, which is not excessive.

When a person comes into contact with a contaminated medium, a certain amount of one or more carriers will be transferred onto or into the body. For example, the person may pick up a given amount of soil on his/her skin. The chemicals present in the soil are transferred as well. The amount of new exposure mass to each chemical is given by the product of the amount of carrier transferred times the fractional chemical concentration in the carrier. This new carrier is added to the existing mass of the carrier, as are the new chemical exposures.

If dose is modeled, the absorption rates will be unique to each chemical, so the relative fractions of the chemicals in each carrier may change over time. These may also change due to encounters with new or different sources. Hence, the masses of all the chemicals must be tracked separately. When removal occurs, by hand washing, bathing, or hand-to-mouth transfer, the fraction of each carrier that is transferred will be calculated. These same fractions of the carrier-specific masses of each chemical will also be transferred or removed. Therefore, SHEDS-Multimedia version 4 will not require transfer or removal rates specific to each chemical, only to each distinct type of carrier. This should mitigate the proliferation of input variables.

In the existing SHEDS-Multimedia version 3, the running exposures of the chemical are tracked in three carriers (air, surface residues, and dust/soil). However, the masses of the carriers themselves are not tracked. In version 4, the transfer and removal rates will pertain to the carrier, rather than to each chemical. The masses of the carrier substances may change even when the person is not near a source of the chemical. For example, a child may play in uncontaminated soil. Version 3 completely ignores such activity, since it does not add any chemical to the running exposures. In version 4, this would change the mass of the carrier (which is soil) on the hands, the body, and possibly in the GI tract. By mixing contaminated and uncontaminated soil on the skin, the concentrations of the chemicals in the dermal soil will change.

Since carriers can be contacted virtually anywhere, the model will have to make adjustments to the carrier masses on nearly all diary events, not just those where the chemical is present. In version 3, hand-to-mouth transfer adjustments occur on most diary events. In version 4, more variables are likely to change on every diary event.

Inhalation is the simplest pathway in SHEDS-Multimedia, because there is no carryover or persistence of running exposure from one event to the next. The relevant carrier is air. The volume of inhaled air (either total volume $V_E$ or alveolar volume $V_A$) depends on the nature of the activity and is important for determining the amount of the chemical entering the lungs. SHEDS-Multimedia version 3 tracks the inhaled chemical exposure, but not the inhaled volume of air. With the shift to the explicit tracking of carriers, it may be of interest to the modeler to track the time-integrated ventilation of air, analogously to tracking the soil, dust, and residue masses.

## Changes to exposure-related variables for version 4

SHEDS-Multimedia version 3 tracks each of seven variables for new exposure, running exposure, absorption, dose, and elimination. These are listed in the table below.

**Table 1. Exposure- and dose-related variables in SHEDS-Multimedia 3.**

| pathway | carrier | new exposure | Running exposure | absorption | dose | elimination |
|---|---|---|---|---|---|---|
| hands (H) | residue | expHR | ldgHR | absHR | dosHR | elmHR |
| hands(H) | matter | expHM | ldgHM | absHM | dosHM | elmHM |
| body (B) | residue | expBR | ldgBR | absBR | dosBR | elmBR |
| body (B) | matter | expBM | ldgBM | absBM | dosBM | elmBM |
| GI tract (G) | residue | expGR | ldgGR | absGR | dosGR | elmGR |
| GI tract (G) | matter | expGM | ldgGM | absGM | dosGM | elmGM |
| lungs (L) | air | expLA | ldgLA | absLA | dosLA | elmLA |

For version 4, the list of variables needs to be modified in the following ways:

1) it must identify the carrier separately from chemicals;
2) it must permit multiple chemicals;
3) it should identify dust and soil separately; and
4) it should allow for exposures from food and water carriers.

The general approach of using a three-letter prefix does not need to be changed. This is followed by a fourth letter indicating the pathway or body part (H, B, G, or L). The fifth letter indicates the carrier. The new list of carriers, including the dietary ones, would be A (air), R (residue), D (dust), S (soil), F (food), and W (water). The main change from the existing variable names is the addition of a numeric suffix. If the suffix is zero, then the variable indicates the mass of the carrier itself. If the suffix is 'n,' the variable indicates the mass of whichever chemical is at position 'n' on the list. Thus, the complete set of new exposure variables would be those in the following table.

**Table 2. Proposed variables for new exposure in SHEDS-Multimedia version 4.**

| Body part | Carrier type | Carrier amt. | Chem #1 | ... | Chem #n |
|---|---|---|---|---|---|
| hands (H) | residue | expHR0 | expHR1 | ... | expHRn |
| hands (H) | soil | expHS0 | expHS1 | ... | expHSn |
| hands (H) | dust | expHD0 | expHD1 | ... | expHDn |
| body (B) | residue | expBR0 | expBR1 | ... | expBRn |
| body (B) | soil | expBS0 | expBS1 | ... | expBSn |
| body (B) | dust | expBD0 | expBD1 | ... | expBDn |
| GI tract (G) | residue | expGR0 | expGR1 | ... | expGRn |
| GI tract (G) | soil | expGS0 | expGS1 | ... | expGSn |
| GI tract (G) | dust | expGD0 | expGD1 | ... | expGDn |
| GI tract (G) | food | expGF0 | expGF1 | … | expGFn |
| GI tract (G) | water | expGW0 | expGW1 | … | expGWn |
| lungs (L) | air | expLA0 | expLA1 | ... | expLAn |

Four more similar tables of variables would be needed, for running exposure, absorption, dose, and elimination. This makes a total of 60 rows (5 tables of 12 rows each). Hence, 60 variables are needed for carriers and 60 more for each chemical. Tracking N chemicals plus the carrier would require a total of (60 N + 60) variables. For N=10 this would be 660 variables per diary event, not counting the approximately140 other variables that are used in the calculations. With 800 variables of 8 bytes each (6.4 Kb) for each event, and perhaps 20,000 such events per person, the output for one person could occupy 130 Mb. A run of 5,000 persons would occupy 650 Gb, more than the size of current hard drives. It is therefore a practical necessity that the output be aggregated before proceeding to the next person.

## Other changes to the model

The expansion of the exposure and dose variable list is not the only change in going to a multiple chemical model. There will need to be additional input variables and changes in a number of other areas. Some of these are:

*New contact probabilities*: Probabilities for contact with carriers will be needed in virtually all possible locations. For example, a person may add to their dermal soil exposure in places without any contamination (chemicals) present. When indoors, residue and dust masses on individuals may be altered even when away from home.

*Maximum dermal loading*: Each carrier should have a maximum dermal loading, although perhaps some may be combined into a limit on the total for two or more carriers together. If a carrier exceeds its limit, a proportion will be removed, along with a similar proportion of all the chemicals it contains. This will avoid the need for chemical-specific maximum dermal loadings, and the logical difficulties that would result from such usage.

*Absorption*: If the multi-chemical exposure model is also used to estimate dose, it will need a large number of absorption rate inputs. In general, the rate will depend on the combination of body part, chemical, and carrier. With 12 combinations of body-part and carrier, then 10 chemicals would imply 120 absorption rate variables, requiring that an input distribution be specified for each one. Even if many of them were assigned the same distributions, this level of detail places an onerous burden on the user. Adding to this is the specification of the correlations between absorption rates (*e. g.*, an individual who is given a high rate for one chemical or pathway might have a high rate for some others as well).

*Hand-to-mouth transfer*: This process is driven by the transfer of the carrier, and all the chemicals embedded therein are transferred as well. Thus, only 3 carrier-to-person rates are needed (for soil, dust, and residues), and the multi-chemical aspect of the model does not lead to substantially more input data being required.

*Washing/bathing/dry removal*: Like hand-to-mouth transfer, these removal processes are based on transferring the carrier from the skin to another location. Thus, there should be one transfer rate per carrier, with all chemicals held by that carrier removed at the same rate. Again, the multi-chemical aspect does not change the number of inputs required.

## Summary

The extension of SHEDS-Multimedia to multiple chemicals involves relatively few changes to the model structure or the code. The principal shift is to a new emphasis on carriers. The number of inputs will increase in certain areas, particularly in the number of absorption rates needed when dose is modeled within SHEDS. The large number of new variables created for output will necessitate the calculation and retention of summary statistics. It may become more important to allow the user to specify which output data are desired, so that the rest may be eliminated. It is difficult at this time to estimate the change in the model run time. As an estimate based on prior experience, perhaps 1/3 of the run time for version 3 would need to be repeated for each chemical. Therefore, 10 chemicals might require about three times as much run time, as compared to just one chemical.

Updating the SHEDS-Multimedia GUI for multiple chemicals is expected to be straightforward. Since requiring the user to simultaneously enter all the required inputs for a multichemical run is not ideal, it is envisioned that tools will be created to allow the user to define and save a “library” of chemical-specific data. The data for each chemical will be stored as a text file; multiple files will be able to be loaded in SHEDS for a given run via the GUI. The chemical-specific files will have a standard format, and the user will be able to create them in one of two ways: either directly in a text editor, or by entering and saving the data in the SHEDS GUI (analogous to what is done now for a single-pollutant run). Utilities will be added to the GUI to save the entered data in the correct library format. In addition, individual library files could be loaded into the GUI for minor editing and then resaved. In this way the user could develop a large dataset of chemical-specific files to be used in different runs, eliminating the need to re-enter data each time it is needed.

# SECTION 2. Planned Methodology for Incorporating a Fugacity-Based Source-to-Concentration Module into SHEDS-Multimedia

Acknowledgment: *We gratefully acknowledge Dr. Debbie Bennett of UC-Davis and Dr. Tom McKone of LBNL for their assistance in developing a fugacity module for SHEDS. The SHEDS fugacity module is a reduced version of the residential fugacity model previously developed by Dr. Bennett and refined through collaborations and contracts with EPA/ORD.*

## Introduction

The SHEDS-Multimedia version 3 model estimates human exposure to chemicals in a residential setting. To achieve this, chemical concentrations are needed as functions of time, in a variety of potential contact media (e. g., textured surfaces, smooth surfaces, and air in either treated or untreated rooms). Version 3 includes three options for establishing concentrations:

(1) a simple decay/dispersion model;

(2) post-application distributions (<1 day, 1-7 days, 8-30 days, >31 days); and

(3) user-specified time series from measurement studies or an external model.

These options are discussed in detail in the SHEDS-Multimedia version 3 technical manual. The fugacity module planned for inclusion in SHEDS-Multimedia version 4 would provide a fourth method for determining indoor residential concentrations.

Fugacity is a well-accepted way of performing chemical mass balance in multicompartment models ( Mackay, 1991). “Fugacity” is a measure of the tendency of the chemical to leave a compartment. The system of equations describing the flows between compartments is solved to find the chemical mass in each compartment as a function of time. The proposed fugacity module can accommodate both particle-bound and vapor phase chemicals.

The SHEDS fugacity module is a reduced version of the indoor fugacity model developed by Bennett and Furtaw (2004). The SHEDS fugacity module is more detailed than the other options in version 3 and would require more inputs. It divides the house into treated and untreated areas, each having vinyl, carpet, air, and wall compartments. The output concentration time series for the different compartments will be used as contacted concentrations for simulated individuals in SHEDS.

## Compartments

As mentioned above, a simulated individual’s residence is divided into “treated” and “untreated” areas. The chemical is applied only in the treated area. The fraction of the simulated house which is in the treated area will depend on the nature of the chemical and its standard modes of usage. For some chemicals, the treated area may be the entire house.

Each area is divided into four compartments: "air," "walls," "carpet," and "vinyl." These terms should be interpreted rather loosely; for example, "vinyl" represents any hard flooring (e.g., wood, stone, linoleum). In current SHEDS terminology, the "smooth surface" media correspond to the vinyl in the fugacity module, and the "textured surface" media correspond to the carpet in the fugacity module.

The conceptual layout of the compartments is shown in the following diagram:

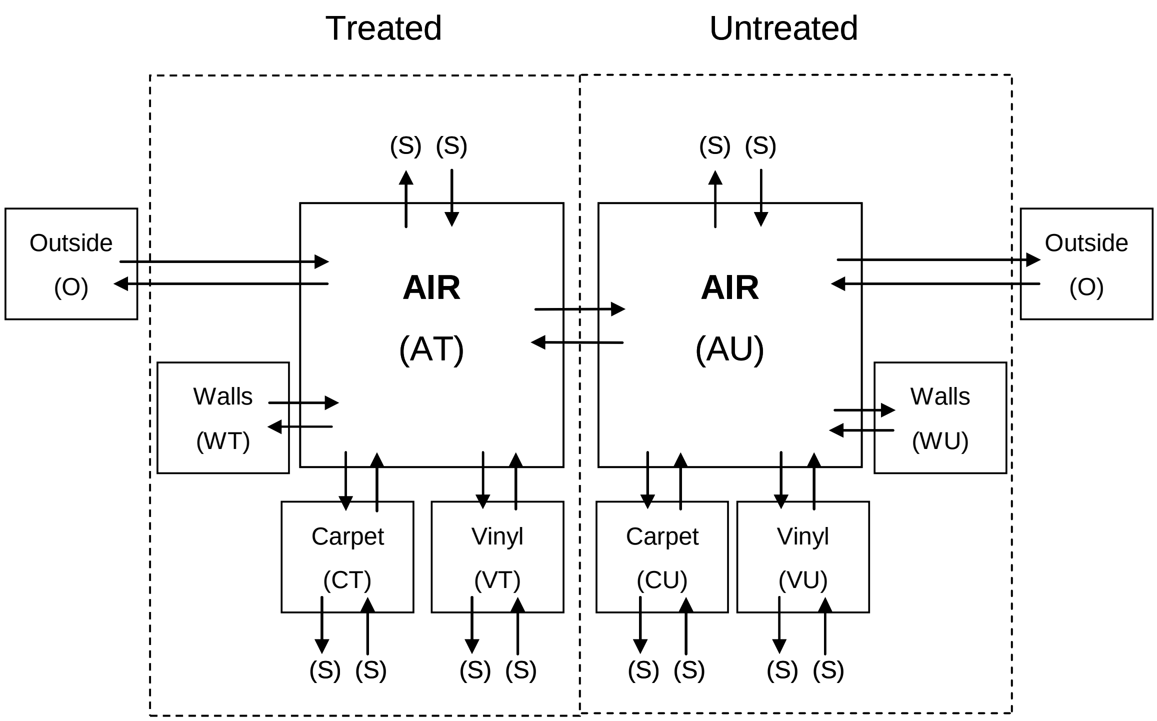


**Figure 1. Conceptual layout of SHEDS fugacity module compartments.**

The symbols in parentheses are the abbreviations given to each compartment in the fugacity module. The symbol "S" represents sources and/or sinks. The arrows in this diagram represent flows, either of the chemical itself (diffusive flows) or flows of particles which may carry the chemical (advective flows).

## Advective flows

Advection refers to the movement of chemical via the bulk transport of its surroundings. The chemical may either be attached to particles which physically move from one compartment to another, or the chemical may be in vapor phase. Advective transport therefore may arise from either particle flows or from the exchange of air parcels between two air compartments.

The proposed fugacity module considers two types of particles, distinguished by their size – above and below 10 microns. For each of the sizes, it is assumed that the amount of particles in each compartment is constant in time. Therefore, the total particle flows into and out of each compartment must balance.

The amount of chemical carried by the particles depends on the fugacity. For each compartment, the particles are assumed to be at the same fugacity as the rest of the compartment. However, different compartments will generally be at different fugacities. Therefore, even though the particle inflows and outflows may balance, the amount of chemical carried in the two directions will generally be different. Over time, the tendency is for the compartments to move toward equal fugacities; however, the model assumes neither an equilibrium nor a steady-state condition

The flows on the diagram marked with (S) indicate sources and sinks. Particle sources are generally associated with human activities such as track-in of dirt, cooking, or use of aerosol products. Apart from the designated applications, it is assumed that these particle sources do not contain any of the chemical.

Particle sinks include cleaning, mopping, and vacuuming of floors, and filtration of air. While in practice the cleaning of floors is occasional, in this model it is treated as a continual process (otherwise, the particle levels could not be kept constant over time).

The amount of chemical attached to each particle is directly proportional to the fugacity of the compartment that the particle is leaving. Since the rates of particle flows are constant in time, each advective chemical flow is proportional to the fugacity of the relevant "from" compartment.

## Diffusive flows

Diffusive flow takes place without any visible macroscopic motion such as air currents or particle motion. It is a significant transport mechanism only between the surface and air compartments. Thus, there are only 12 diffusive flows in the module. These co-exist with advective flows between the same compartments. Diffusion may dominate if there are relatively few particles moving between the compartments.

The amount of diffusive chemical transport is also proportional to the fugacity of the compartment. Hence, the total outflow of chemical from each compartment is proportional to the current fugacity of that compartment. In the absence of inflows, this would result in the first-order exponential decay of chemical concentration. With inflows present the situation is more complicated. At least initially, the low fugacity compartments (such as those in the untreated room) will experience a net gain of chemical.

## Using the fugacity module in SHEDS-Multimedia

The fugacity module is one option that would be available for the source-to-concentration module in SHEDS-Multimedia version 4. It is compatible with either the SHEDS model-specified dates or user-specified dates methods for selecting application dates (see the SHEDS-Multimedia version 3 Technical and User Manuals). In either case, the fugacity module would be run each time a new application occurs, for each person. For example, if a SHEDS version 4 model run consisted of 5,000 persons and each averaged ten applications, the fugacity module would be called 50,000 times during this run.

The initial amount of chemical present in each compartment immediately after an application could be determined directly as SHEDS model inputs or via a partitioning algorithm specific to each application type. The latter option is not currently available in the SHEDS fugacity module and would need to be developed.

Applications in SHEDS are effectively instantaneous. Thereafter, until the next application, the chemical mass in each compartment would change in accordance with the fugacity module. Even under the user-specified dates option, where all persons have the same application dates, the fugacity module would have to be run separately for each person, since they have different house characteristics. At the time of the next application, the current concentrations remaining from the prior application are added to the "new" concentrations resulting from the next application. This continues until the end of the simulation period, when modeling on the next person is started.

The fugacity module only applies to the indoor media. If lawn or garden applications are used, then those concentrations would have to be calculated using one of the existing methods (decay/dispersion, post-application, or time series; see the SHEDS-Multimedia version 3 Technical Manual). Hence, when the fugacity module is incorporated into SHEDS version 4, the option of using one source-to-concentration method indoors and another outdoors will be needed. Currently, SHEDS version 3 requires both indoor and outdoor media to be modeled using the same approach.

SHEDS requires hourly time series of concentrations in each medium. The fugacity module calculates concentrations as analytic functions of time. These can be quickly evaluated at a large number of points in time with no difficulty.

Extending the fugacity module to multiple chemicals would be straightforward with respect to updating the model code. As far as the SHEDS user is concerned, the main issue with using the fugacity module will be the identification of appropriate input values (see next subsection), which will have to be entered in an updated version of the SHEDS Multimedia GUI. The chemical-independent fugacity inputs will require the addition of several new data input screens. The chemical-specific inputs will be handled in the same manner as other chemical-specific inputs to SHEDS (see Section 1 of this document on extending SHEDS to multiple chemicals).

## Inputs for the fugacity module

In its current form, the SHEDS fugacity module requires distributions for 45 input variables, as well as the initial chemical mass in each compartment. As stated above, the initial chemical concentrations could be determined either directly from existing SHEDS inputs or via new algorithms based on product label information. In the latter case, additional inputs would be needed (e.g., application rates, surface areas treated, and variables that determine partitioning of the chemical among compartments). Some of the other 45 variables are also currently used as inputs for SHEDS-Multimedia. However, 35-40 inputs have no equivalents in SHEDS-Multimedia version 3.

There are both general and chemical-specific inputs to the fugacity module. The general inputs are not chemical-dependent; as examples, these include air exchange rate, cleaning rate, house dimensions, and particle deposition and resuspension rates. In some cases, input values have been developed for non-chemical specific variables (Bennett and Canales, 2004). The chemical-specific inputs to the model are listed in Table 1; distributions would need to be developed for each chemical modeled in a SHEDS simulation.

**Table 1. Chemical-specific inputs for the SHEDS fugacity module.**

| Variable | Description | Units |
|---|---|---|
| decay_a | chemical decay rate in air | (1/d) |
| decay_c | chemical decay rate in carpet | (1/d) |
| decay_v | chemical decay rate in vinyl | (1/d) |
| decay_w | chemical decay rate in walls | (1/d) |
| Dair | diffusion coefficient in air | ($m^2$/d) |
| Kow | octanol-water partition coefficient | (-) |
| vapor | vapor pressure | (Pa) |
| solub | Solubility | (mole/$m^3$) |
| mw | molecular weight | (g/mol) |

When SHEDS-Multimedia is modified to support multiple chemicals, then this information will be needed for each specific compound being considered. If the fugacity module is also used to determine initial concentrations after applications, then additional chemical-dependent inputs will be needed. The fugacity module calculates total chemical mass, including both dislodgeable (e.g., amount that can be dislodged from a floor to a hand or body) and embedded mass in the walls, carpet, and vinyl media. Note that it will be important to consider matching the correct type of surface loading with the corresponding transfer factor (i.e., transfer efficiency or transfer coefficient) when developing inputs for modeling dermal exposure.

.

Requesting distributions for all the additional fugacity module inputs may be burdensome. Various types of sensitivity analyses can be applied to explore further model simplification and to identify key inputs. These include Sobol's method (discussed in Section 4 of this document) and percentile scaling (see the SHEDS-Multimedia version 3 Technical Manual). With assistance from Lawrence Berkeley National Laboratory(LBNL) and UC-Davis via EPA/ORD/NERL's University Partnership Agreement with LBNL, this is planned for chlorpyrifos and permethrin. Less important inputs could be assigned pre-set distributions to bypass the need for direct user input.

## Current status of the fugacity module

The equations for the fugacity module described above have been programmed in SAS. The program runs very quickly and should not add too much time to a SHEDS run, despite the large number of times it must be run. A preliminary sensitivity analysis has been run on this model. However, the results of this analysis are not definitive, since they depend on the choice of distributions for the input variables, and these have not been finalized.

The predecessor of the SHEDS fugacity module is a more complex fugacity model developed for EPA and described in Bennett and Furtaw (2004). That model contained inputs for six different ranges of particle size. Some effort has been made to convert the inputs for that model into distributions for comparable inputs for SHEDS. The simplified SHEDS module can be compared with the original published fugacity model to confirm that the relevant effects have been retained.

In addition to incorporating the fugacity module into SHEDS-Multimedia version 4, predictions from the module can be compared to pyrethroid results from an ORD/NERL test house study that has been conducted in Research Triangle Park, North Carolina. This study was designed in part to evaluate the original fugacity model, but will be useful in evaluating the simplified SHEDS module.

## Stand-alone applications of the fugacity module

The application of the fugacity module in SHEDS is configured to simulate results applicable to a specific modeled individual having a specific chemical usage pattern. An alternate application of the fugacity module is as a stand-alone model for prediction of typical chemical concentrations in residential settings resulting either from retention of a chemical from a single application or build-up from multiple applications. As noted in the above discussion, multiple applications present no difficulty for executing the module. Each new application simply adds chemical to the amount remaining from the last application; thus new “initial conditions” are established and the system of equations solved again.

One approach to this could be to sample all model inputs over a spectrum of reasonable values in an attempt to characterize the distribution of chemical concentrations. These distributions may have utility in other chemical exposure models. For example, the distribution of chemical concentrations many months post-application may be used as typical background concentrations.

## Summary

The incorporation of the proposed fugacity module into SHEDS-Multimedia should be straightforward. Both models are written in SAS, and the fugacity module requires no special types of SAS software beyond those already used in SHEDS-Multimedia. The distributions for the fugacity inputs can be specified using the same format as for current SHEDS-Multimedia inputs. However, if the option of using the fugacity module to establish the initial chemical concentrations is desired, then the algorithms would still need to be developed. Internally, the outputs from the fugacity module itself consist of coefficients and terms in a set of analytical equations. These can be easily applied at hourly intervals to generate the concentrations required for SHEDS exposure calculations. For those who do not wish to utilize the fugacity module, the current options for source-to-concentration in SHEDS-Multimedia would still be available.

## References

Bennett, D. H. and R. Canales, 2004, Exposure-related Data and Models to Refine NERL's Stochastic Human Exposure and Dose Simulation (SHEDS) Model for Pesticides and Air Toxics. Task 1 Report: Indoor Residential Fugacity-based Fate and Transport Model. Prepared under Task Order 23, Contract Number 68-D-99-011 for the National Exposure Research Laboratory, Office of Research and Development, U. S. Environmental Protection Agency.

Bennett, D. H. and E. J. Furtaw, Jr., 2004, Fugacity-based Indoor Residential Pesticide Fate Model, *Environ. Sci. Technol.* 38:2142-2152.

Mackay, D., 1991, *Multimedia Environmental Models: The Fugacity Approach*. Lewis Publishers, Inc., Chelsea, MI.

US EPA ARCHIVE DOCUMENT

# SECTION 3. Planned Methodology for New Longitudinal Diary Assembly Option in SHEDS-Multimedia

## Introduction

SHEDS-Multimedia version 3 requires the construction of human activity diaries that cover the entire simulation period of a model run. This period is often several months, a year, or even longer. The human activity diaries are drawn from EPA's CHAD (Consolidated Human Activity Database; McCurdy et al., 2000; http://www.epa.gov/chadnet1), which typically includes just one day (24 hours) of activities from each person. A "longitudinal" diary is one that covers the same person over a long period of time. While the SHEDS modeling period may be of user-specified duration, it is assumed in this section to be one year, to provide a concrete example.

A new choice for longitudinal diary assembly is planned for SHEDS version 4, providing the user with a more detailed option for assembling year-long diaries. This method requires a few additional inputs to be designated by the user, but allows for more control over the properties of the assembled diaries. The existing method in SHEDS-Multimedia version 3 that uses eight one-day diaries (one weekend and one weekday from each of four seasons) will be retained in SHEDS-Multimedia version 4 as the default option (see SHEDS-Multimedia version 3 Technical Manual). This method, which may be appropriate for many applications of the model, has undergone a previous SAP review (August 30, 2002) and requires no additional input by the SHEDS user.

In addition to the existing inputs, the new diary assembly method will require the user to:

1) select the diary property most relevant to exposure for the current application;

2) specify the D statistic, which relates the within-person and between-person variances for this diary property; and

3) specify the 1-day lag autocorrelation in this diary property.

These three steps are explained in greater detail in the following sections. All of the other steps in the diary assembly method are handled automatically by the program. The new method is detailed in Glen et. al. (2007) and has already been implemented in EPA's APEX and SHEDS-Air Toxics models. It has been programmed in SAS for the latter application, so implementation in SHEDS-Multimedia would be straightforward.

## Within-person variance $\sigma_w^2$ and between-person variance $\sigma_b^2$

When randomly drawing multiple one-day diaries from multiple individuals that are intended to represent a single individual's behavior over time, the modeler faces a dilemma. If a small number of diaries are drawn for each individual to cover a long simulation period, then each diary must be re-used many times; that is, each diary must be used on many different dates in the simulation to represent the individual's behavior. While this creates repetitive or habitual behavior patterns, it also narrows the behavioral space and lessens the within-person activity variability. This within-person variability is summarized by the $\sigma_w^2$ statistic.

Using many different one-day diaries would address these last two concerns by broadening the simulated individual's behavioral space and increasing the within-person variability. However, this approach would exacerbate other problems. In particular, any two persons belonging to the same cohort will draw their diaries from the same diary pools, and the samples will tend to converge to the same overall average behavior of the cohort. For example, if a set of six-year-old children are modeled for one year each and 365 different diaries were drawn for each child, there would be almost no variation in average time outdoors from child to child. This is also true for any other property of the diaries, such as travel time, time spent exercising, and so on. The differences in such quantities from one child to another are summarized by the between-person variance $\sigma_b^2$.

If all the diaries are drawn from the same pool, it may be shown that regardless of the number of diaries drawn per person, the sum of $\sigma_w^2$ and $\sigma_b^2$ is the same, namely, the total variance in the diary pool itself. Thus, an increase in one of these quantities requires a decrease in the other. One can assess the relative size of $\sigma_w^2$ and $\sigma_b^2$ in an activity study, provided it records a sufficient number of days from each person to allow good estimates of both variances.

In practice, this trade-off almost always demands that relatively few one-day diaries be used for each person. So while the overall variances $\sigma_w^2$ and $\sigma_b^2$ may be acceptable, the narrow range of behavior for any one person tends to cause other problems. For example, any activity that happens at all will happen many times, since each diary is re-used many times. There is an absence of activities that happen only once or just a few times over the year.

## The concept behind the new method

The new method allows the modeler to apportion the total variance into $\sigma_w^2$ and $\sigma_b^2$ by specifying the D statistic, defined to be

$$D = \sigma_b^2 / (\sigma_w^2 + \sigma_b^2).$$

D pertains to the population as a whole and is bounded by zero and one. A value of zero implies all persons have the same average behavior, whereas a value of one implies the greatest possible difference in mean behavior that is consistent with the total variance.

Most of the existing random-draw methods of diary selection assume that all diaries that are suitable (meaning they are from the correct age-gender cohort and match the chosen day-type) are equally likely to be chosen, and that any subsequent draws are independent of prior draws. The new method drops both these assumptions by assigning each simulated person a "target behavior," and then preferentially sampling diaries to produce the target behavior. The method assigns target behaviors and executes the preferential sampling based on the value of D specified by the modeler. If not executed carefully, preferential sampling can result in behavioral biases, where some diaries are consistently drawn more often than others. The method contains internal rules for this sampling that ensure that over a large number of simulated persons, all available diaries in each diary pool will be sampled nearly uniformly.

In the new method, a new random draw is made for every day in the simulation. Thus, a one-year longitudinal diary would be comprised of potentially 365 different diaries. The D statistic affects the width of the diary selection probability peak around the target behavior, with a low D giving a broad peak and a high D giving a sharp, narrow peak. Depending on the width of this peak and the number of diaries in the pool, some diaries may be selected multiple times, but others may be selected just once or not at all.

The choice of D for a given application is dependent on the key variable of interest. Glen et al. (2007) estimated values of D for time spent outdoors and time spent in vehicles from a longitudinal activity study of children in California. Note that these values are specific to these diary properties and the participants in that study. There are other studies that could potentially be used to calculate D that have not been analyzed. Further field studies would be useful in establishing D values for other key variables and target populations (e.g., different age groups). The user should decide if adequate data exist to support a particular choice of D for his or her application.

## The selection of the key variable and the use of rankings

The prior section explains qualitatively how the within-person variance and between-person variance can be controlled by adjusting the width of the probability peak around the target mean behavior. To implement this in practice, one first needs to define a measure of similarity so that one can identify how close each of the available one-day diaries is to the target behavior. In an exposure model, "similar" diaries should be ones that have a similar potential for exposure. Differences in activities that have little or no potential to influence exposure are not relevant here. The modeler must assign a "key variable" that estimates the exposure potential for each diary, given the nature of the chemical and its forms of application. For example, when running a lawn scenario, the amount of time spent outdoors at home on each one-day diary would be a good measure of the likelihood of exposure to lawn chemicals. If all the application types were indoors, then a different diary variable would be relevant.

The key variable may be defined in any way at all, so long as it can be given a specific numeric value on each diary in the database. The diaries are then grouped by age-gender cohort and by day-type; each such grouping is called a "pool." The program then sorts each pool according to the values of the key variable. Hence, the diaries in each pool are ranked by their potential for exposure. Each diary is given an "x-score" which indicates its position in the ranked list. For example, a diary which is 25% of the way up its list has an x-score of 0.25. The x-scores are basically just percentiles divided by 100, and so are bounded by zero and one.

While the key variable itself could have almost any distribution, which might also differ from pool to pool, the x-scores are always uniformly distributed within each pool. The program automatically calculates the distribution of the diary selection probability peak that matches a particular within-person variance in the x-scores, and also calculates the distribution of preferred target behavior in x-scores that matches a given between-person variance in the x-scores. These distributions depend only on the value of the D statistic -- not on the choice of key variable, the definitions of the cohorts or pools, or the number of diaries in the database.

For each person simulated, a target mean behavior is randomly selected using the distribution appropriate for the specified D statistic. Then a set of x-scores (one for each day in the simulation) is selected from another distribution centered on this target mean behavior. Over a large enough sample of persons so that the stochastic variation is small, the within-person and between-person variance in these x-scores will match the values indicated by the D statistic. These x-scores could then be mapped directly to one-day diaries. However, they first may be re-ordered to induce autocorrelation.

## Autocorrelation and the re-ordering of selected x-scores

In addition to targeting the within-person and between-person variances through setting the D statistic, the new diary assembly method optionally allows targeting of the day-to-day autocorrelation. This is a measure of the tendency for similar diaries to occur on consecutive days. This could be of interest to the exposure modeler if the concentration time series were strongly episodic, for example. In the diary assembly, a positive autocorrelation indicates a tendency for diaries with x-scores near each other to be used on consecutive days, while a negative autocorrelation indicates a tendency for dissimilar x-scores to be used on consecutive days. Some preliminary values of A have been derived from the same data that were used to estimate D (Glen et al., 2007). If there is no clear tendency in either direction, or if the ordering of days does not affect the exposure statistics, then one can either set the target autocorrelation to zero or else omit the re-ordering step entirely.

## Assigning one-day diaries to x-scores

Once the model has generated the daily sequence of x-scores for a given person, it selects corresponding one-day diaries from the database. Each day of the simulation requires drawing from a particular diary pool, based on the day-type. The x-score assigned to that day indicates the position in the sorted list of diaries belonging to that pool.

## Implementing the new diary assembly method in SHEDS-Multimedia

The new diary assembly method has been programmed in SAS and incorporated in another model in the SHEDS family (SHEDS-Air Toxics). It would be straightforward to add this method as an alternative to the current diary assembly method used in SHEDS-Multimedia version 3. Beyond variables used in the existing method, the new method requires only three additional inputs: the specification of the key variable, and the target values for D and the autocorrelation A.

## Discussion

The diary assembly method may have less direct importance in SHEDS-Multimedia than in some other exposure models, since some factors that may drive exposure are not found on the diaries – for example, chemical usage patterns and contact probabilities. Nevertheless, the diaries still influence the exposure, particularly outdoors where there is a larger variation between persons. The population distribution for exposure is driven by the distributions of the various factors (like outdoor time) that influence exposure. Therefore, improvement in constructing longitudinal diaries to better reflect these factors should lead to better estimates of exposure.

## Summary

The new diary assembly method can be implemented in SHEDS-Multimedia with minimal difficulty. A full description of the method has been accepted for publication (Glen et al., 2007). The required code already exists in SAS in another SHEDS model, and relatively little modification would be needed to incorporate it into SHEDS-Multimedia version 4. A small amount of additional data would need to be specified for model inputs. For users of SHEDS-Multimedia version 4 who do not wish to utilize this new approach, they would still be able to choose the current 8-diary approach in SHEDS-Multimedia version 3.

## References


Glen, G., L. Smith, K. Isaacs, T. McCurdy, and J. Langstaff, 2007, A New Method of Longitudinal Diary Assembly for Human Exposure Modeling, accepted by *Journal of Exposure Science and Environmental Epidemiology*.

McCurdy, T., G. Glen, L. Smith, and Y. Lakkadi, 2000, The National Exposure Research Laboratory's Consolidated Human Activity Database. *J Expo Anal Environ Epidemiol* 10:566-78.

# SECTION 4. Planned Methodology for Utilizing Sobol's Method for Sensitivity Analysis with SHEDS-Multimedia

## Introduction

The SHEDS-Multimedia model is a sophisticated physically-based probabilistic model with numerous input variables. Most of these inputs can be assigned distributions from which values are randomly sampled for each person simulated. In some cases, multiple random samples are assigned to each person corresponding to different times in the simulation period; for example, one variable may be sampled daily and another sampled on an hourly basis.

The primary purpose of sensitivity analysis is to determine which model inputs are most influential in determining the values of the model output. This knowledge may be applied in several ways. It can be used to set priorities for the allocation of resources to obtain better input data or the model itself might be simplified by pre-setting or eliminating unimportant inputs. Additionally, understanding the importance of various factors can help an exposure modeler estimate the efficacy of potential exposure reduction strategies.

There are many methods of sensitivity analysis, each providing a somewhat different view of how the output responds to the input. The EPA SHEDS models have utilized both "one-at-a-time" and "multivariate" sensitivity analysis methods. In this discussion, "input" refers to a variable that has different values either between or within simulated individuals.

Sobol (1990; English translation, 1993) describes a multivariate method that is capable of determining both "main" and "total" effects for each input, as discussed below. The main (or first-order) effect measures the importance of a single input without considering any interactions with other inputs. The total effect equals the main effect plus all interactions involving the given input variable.

In addition to the ability to address direct and interaction influences among the inputs, Sobol's method has several other features that make it a useful expansion of the sensitivity capabilities of SHEDS. Specifically, it can handle categorical and other non-numeric inputs, and it is suitable for examining aspects of the model (such as diary assembly) that are not easily handled by other sensitivity methods.

## Stochastic variation

Like other probabilistic models, SHEDS-Multimedia uses random sampling to simulate the inherent stochastic variation in the processes being modeled. Consequently, the model output can change from run to run, even if all the user settings remain the same. The stochastic variation acts as "noise" which could mask or suppress the complex relationship between input and output.

The stochastic variation can be reduced by increasing the sample size; that is, by taking averages over larger numbers of cases. In SHEDS-Multimedia, this corresponds to generating more persons in each model run. A general rule of thumb is that the stochastic variation scales as $N^{-1/2}$; this implies that to halve the stochastic variation, four times as many persons must be analyzed. In practice, very large numbers of persons are not feasible, so one is faced with finding a sensitivity analysis approach where the stochastic variation, while still present, does not overwhelm the other (deterministic) variation in output that is due to changes in inputs.

## Existing sensitivity analysis methods in SHEDS

SHEDS-Multimedia currently does not have any sensitivity analysis methods available through the user interface. However, several sensitivity analysis methods used in SHEDS-Wood (Xue et al., 2006) have been tested in SHEDS-Multimedia using batch run capabilities with the SAS code. These include a "percentile scaling method," in which one input is set to a given high (or low) percentile of its usual distribution, while all other inputs are set to their median values. A set of persons is run using these fixed settings, and the mean output over these persons is determined. Many persons are averaged because there are random elements (such as diary selection and determination of contact events) that are not controlled directly by input settings, so there is still some stochastic variation present in the results. A comparison of the mean output at the high and low settings of a given input gives an indication of the importance of that input in determining the mean of the output.

Another existing sensitivity analysis method involves recording the mean setting for each input variable and the relevant output variable for each person, during an otherwise typical model run. All inputs are allowed to vary simultaneously. At the end of the run, a correlation analysis is performed between the various input variables and the output. Either Pearson or Spearman correlation statistics may be used. Correlations near zero indicate input variables that have little influence on exposure. Larger correlations, either positive or negative, indicate more influential inputs.

The final method currently used in SHEDS models is a stepwise regression of the output variable on the collective set of inputs. As with the correlation methods, the analysis unit for this purpose is each simulated person. Each person is assigned an exposure statistic that summarizes his or her time series. Input variables that are resampled within the individual are averaged over their sampled values.

## Sobol's method of sensitivity analysis

Sobol's method is a variance-based approach to sensitivity analysis, which is useful for application to stochastic models. Sobol (1990; English translation 1993) discusses applications to 'nonlinear mathematical models,' but none of the examples discussed therein are stochastic. However, Sobol himself suggested using stochastic sampling techniques for the input values being used even for analytical functions (non-stochastic models). The goal of Sobol's method is to quantify the direct and interaction contributions of each input to the variance of the output.

The requirements for the use of Sobol's method on a stochastic model are: (1) that the model has a single output variable of interest, (2) that this output has one value for each repetition of the calculation, (3) that this output variable is a deterministic function of the values assigned to the input variables, (4) that each repetition is independent of other repetitions, (5) that each input is randomly sampled once per repetition, and (6) that all inputs are independent.

While some stochastic models meet all these conditions, SHEDS-Multimedia needs some minor alterations. Condition (1) simply means that each output variable of interest must be analyzed separately. If the output is multi-valued (like a time-series), then condition (2) can be met by creating a summary statistic, possibly (but not necessarily) a mean. Condition (4) is met by SHEDS-Multimedia, since each person is independent of all others. As discussed later, condition (5) can easily be relaxed without changing the basic method. Condition (6) can be problematic for two reasons -- correlated and conditional inputs. Thus, correlation of inputs should not be used with Sobol's method. However, the greater difficulty is that some inputs in SHEDS are conditional on other inputs; for example, the distribution for body weight is conditional on age. Methods for dealing with this issue in SHEDS are discussed later. Finally, condition (3) is also problematic. To resolve it, all the random numbers used in the model must be re-interpreted as being values assigned to specific input variables.

Sobol's method partitions the variance of the output from a model into terms representing single inputs, pairs of inputs, groups of three inputs, and so on. Thus, it is one of the analysis techniques based on variance decomposition. Sobol's method is characterized by first randomly creating two sets of sample values for each input variable for each simulated individual. Here these are referred to as the A and B sets. One then performs model runs with various combinations of these values. When the model runs are complete, the variance in the model output is examined to determine sensitivity to each input parameter. In brief, the output variances associated with the various combinations of inputs are divided by the total output variance to estimate the main and total effects for each input.

Sobol's method can only evaluate the influence of inputs that are sampled for each person from random distributions. Models may also have other inputs or settings that are fixed by the user when requesting a model run. These are either constants, or may systematically change through the model run. While constants may be important in determining the mean value for the output variable, they cannot explain variation from one repetition to another and are not the focus of a variance-based method.

## Non-numeric inputs (categorical inputs)

Unlike many other sensitivity analysis methods, Sobol's method does not perform any mathematical functions (like calculations of mean, variance, or correlation) on any of the inputs, only on the model outputs. Therefore, there is no problem if some of the inputs are discrete or even categorical, as long as no dependencies exist among the inputs. For example, suppose that in a human exposure model input #1 is gender. Then the Sobol indices $S_1$ and $T_1$ give the main and total effects of gender on model output. However, many exposure models are structured so that other input distributions are functions of gender. Accounting for this dependence is addressed in the following section.

## Intentional correlation or dependence among input variables

Many stochastic models, including SHEDS-Multimedia, have inputs whose distributions depend on the settings of some other input variable(s). For example, the body weight distribution for children depends on age. There are two approaches that could be used in such cases. One method is to group the related inputs, effectively treating them as a single input. In implementation, this means that all the variables grouped as a single input would use their A values (or their B values) without any mixing. The result is that the main and total effects of this group can be determined, although not the contributions of individual inputs within the group. The other approach is to redefine the input variables to eliminate the dependence. For example, in the age-body weight example, replace body weight with a variable measuring the percentile of body weight relative to one's cohort, as an input variable. Then all ages have the same distribution (uniform 0-100) for this new input variable. The exposure model will have to convert this percentile to an actual body weight at some point, but this is a deterministic calculation that is interpreted as part of the model structure. Sobol's method can then evaluate the influence of age and the influence of body weight percentile, as these are now independent input variables. However, one must be careful in the interpretation of the results.

## Multiple input samples per model repetition

Many of the input variables in human exposure models are sampled not just once per model repetition, but multiple times. For example, each repetition may consist of a year-long time series of exposure for a single person. A given input might be sampled once for the entire model run or on a seasonal, monthly, weekly, daily, hourly, or even a diary event basis.

The sampling frequency can be determined *a priori* and does not depend on other variables. Therefore, the number of samples to be drawn for each input is the same for each repetition (person). For example, a variable that is to be sampled hourly should have one random draw for each hour in the simulation period, regardless of whether or not that variable is actually utilized every hour. Then the A and B sets can be populated without reference to which person is being considered. The point is that these input values must be re-used on other persons in other model runs.

An exception to this rule is found in inputs that are sampled once per diary event, since the number of events varies from person to person. The implementation of Sobol's method would be simplified if event-based sampling were eliminated in SHEDS.

## Contribution of activity diary selection to output variance

One input that has not been evaluated in prior sensitivity analyses is the effect of assigning a particular longitudinal activity diary to each simulated individual. There were two main difficulties. First, the pools of available diaries differ from one person to another, so one could not separate the effect of diary selection from other factors such as age or gender. Second, the diary choice was not quantifiable on a numeric scale, and so, could not be used in correlation or regression analysis.

With Sobol's method and the proposed new method of assembling longitudinal diaries (see Section 3 of this document), both problems disappear. As discussed earlier, the way to resolve the age-gender dependence is to replace the "diary selection" variable with another variable that is not dependent on age and gender, and yet provides enough information to uniquely identify which diaries are to be used. In the new diary assembly method, the vector of 'x' scores meets these conditions, after rearranging to induce autocorrelation. Such vectors can be arbitrarily swapped between individuals without regard to any personal demographic information, and yet this vector (along with the pool definitions that are common to all the evaluations) provides enough information to identify the specific daily diaries. This vector is already intended to measure the potential of each one-day diary for exposure; therefore, it should capture most of the effects of diary selection on the output.

## Application in SHEDS-Multimedia

The implementation of Sobol's method in SHEDS-Multimedia requires two main alterations to the model. One change is that all random determinations must be re-expressed as independent input variables, as in the examples of "percentile body weight" or "vector of x-scores" given above. Thus, the list of "inputs" to SHEDS-Multimedia will change significantly. The second alteration relates to the actual generation of random numbers.

The current SHEDS-Multimedia version 3 code generally produces random numbers whenever they are needed, without tracking the random number seeds. For Sobol's method the random numbers must be reproducible, transferable from person to person, and generated independently of the choice of random numbers for all other input variables. This requires careful tracking of random number seeds. Suppose a total of 100 input variables are identified once the inputs have been re-defined to be independent. Each of these 100 inputs would have a specified sampling frequency. If the model were run with a sample size of 1000 persons, then the A and B data sets would each consist of 100 lists, each with 1000 random number seeds. Using a given seed and the sampling frequency, a list of actual random numbers can be reproduced whenever it is needed.

Once the SHEDS-Multimedia code is reorganized to permit Sobol's method, the same code can easily be used for standard model runs as well. Instead of reading the seeds needed for a model run (100,000 of them in the above example) from an input file, the code would simply generate a set of new seeds at random. This would correspond to the way the model is structured at present.

Once coded, the full determination of both main and total effects for N inputs requires that the model be run a total of (2N+2) times. The number of inputs has not been definitely determined, as some re-definition is required, but a reasonable estimate is somewhere around 120 for a single-chemical model. A total of 242 model runs would be time-consuming (possibly a week or more on a standard personal computer), but is quite comparable to the time required for a 2-stage or uncertainty run.

## Summary

Sobol's method of sensitivity analysis provides significantly more information than the current alternatives, but at a cost of requiring some reorganization of the model code and the redefinition of some of the model input variables. Sobol's method has the capability to assess the influence of categorical or non-numeric inputs, unlike the correlation and regression methods previously employed. Similarly, it accounts for non-linear response to inputs. Sobol's method would be a useful supplement to the existing sensitivity analysis methods used for the SHEDS-Multimedia version 3 model.

## References

Sobol, I. M., 1990, Sensitivity estimates for nonlinear mathematical models (in Russian), *Matematicheskoe Modelirovanie* 2:112-118. Re-published in English (1993) in *Mathematical Modeling and Computational Experiment* 1: 407-414.

Xue, J., V. G. Zartarian, H. Ozkaynak, W. Dang, G. Glen, L. Smith, and C. Stallings, 2006, A probabilistic arsenic exposure assessment for children who contact chromated copper arsenate (CCA)-treated playsets and decks, Part 2: Sensitivity and uncertainty analyses, *Risk Anal*. 26:533-41.

# SECTION 5. Planned Additional Changes for SHEDS-Multimedia Code and GUI

The following list of items has been discussed as possible modifications to the existing version 3 of the SHEDS-Multimedia model, depending on available data and resources. These are in addition to any changes to code and interface that would be required with respect to implementing the other major model modifications described in this document (multiple chemicals, fugacity, new diary assembly method, Sobol's method).

The list has been roughly divided into "Code," "Interface," "Input," and "Output" topics. However, many of the individual items do not divide neatly into these categories, but rather overlap them. For example, allowing user-defined age ranges on inputs is listed as an "Input" item, but pursuing this will require changes to the diary preprocessing (so this might also have been listed as a "Code" item). Beyond this example, no cross-referencing between or among categories is done here. The purpose of this list is to simply collect all such items in a single place. The list appears in no particular order.

## Code items

- consider capability to allow simulated individuals to age (including < 1 year olds)
- consider memory/space saving alternatives, particularly as regards multiple chemical scenarios
- allow separate values for transfer coefficients and transfer efficiencies for smooth versus textured surfaces
- calculate margins of exposure (MOEs)
- implement ability to repeat runs with same stream of random numbers
- allow reuse of same diaries for sensitivity analysis
- allow sensitivity analysis by pathway
- allow distribution shapes to change, but retain same mean and variance
- incorporate exposures for the applicators of chemicals
- consider option to put confidence intervals on variability distributions
- add capability to allow correlation of sampled values for single variables for a simulated individual (in addition to current capability to correlate different variables)
- incorporate the SHEDS-dietary module into SHEDS-Multimedia code; consider optimal approach/variables for matching CHAD diaries and food consumption diaries (e.g., age, gender, season, weekday, region, race, METS/caloric intake)
- to enhance longitudinal simulations with more realistic inputs, develop approaches to partition the variance of important SHEDS inputs into their major components (e. g., inter-personal, intra-personal, season, gender) and utilize this information in assembling the inputs and diaries for model runs

- make necessary code changes for enhanced exposure-dose model linkage

## Interface items

- implement Help files/buttons
- improve windows/screens (e. g., minimize scrolling, more bread crumbs)
- allow display of uncertainty input clouds
- allow visualization of probability vectors
- allow sensitivity/uncertainty runs via the interface, subject to computational and/or run time constraints
- accommodate the inclusion of SHEDS dietary module
- make the View button a true browse mode (that is, no editing allowed)
- more informative responses to input errors

## Input items

- allow user-defined input age categories
- develop capability to accept empirical distributions, possibly via cut/paste from Excel
- allow user to more finely specify input distributions via the interface (e.g., age-specific, indoor/outdoor, seasonal)
- allow correlation of "has lawn" and "has garden" probabilities
- allow user to specify sampling frequency (event, daily, …) of variables
- allow use of non-CHAD activity diaries

## Output items

- more readability and flexibility on output graphs (e. g., log-linear axis, choose X vs. Y axis, improved boxplots, specification of numeric format)
- print message if all values are zero for a requested plot
- allow user to view/graph sensitivity and/or uncertainty outputs, subject to computational and/or run time constraints
- add start and stop time stamps to log