The Stochastic Human Exposure and Dose Simulation Model: High-Throughput

SHEDS-HT Beta Version 0.1.10

User and Technical Manual

April 2024

Prepared by: Kristin Isaacs,[1] Victoria Hull[1,2], and Graham Glen[3]

[1]US Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure

[2]Oak Ridge Associated Universities

[2]ICF International

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS AND DISCLAIMER

The United States Environmental Protection Agency through its Office of Research and Development funded and collaborated in the research and development of this software, in part under Contract EP-C-14-001 to ICF International. The model is publicly available in Beta version form. All input data used for a given application should be reviewed by the researcher so that the model results are based on appropriate data sources for the given application. This model, default input files, and R package are under continued development and testing. The model equations and approach are published in the peer-reviewed literature (Isaacs et al. Environ. Sci. Technol. 2014, 48, 12750-12759). The data included herein do not represent and should not be construed to represent any Agency determination or policy.

1 OVERVIEW

SHEDS-HT is an acronym for Stochastic Human Exposure and Dose Simulation – High Throughput. It is a screening model for estimating human exposure to a wide range of chemicals. The people in SHEDS-HT are simulated individuals who collectively form a representative sample of the target population in terms of age and gender, as chosen by the user. The scientific development of SHEDS-HT has been described in the peer-reviewed literature (Isaacs et al. 2014). The model is cross-sectional, with just one simulated day (24 hours) for each simulated person, although the selected day is not necessarily the same from one person to another. SHEDS-HT is stochastic, which means that many inputs are sampled randomly from user-specified distributions that are intended to capture variability. In the SHEDS series of models, e.g in SHEDS-Multimedia (Glen et al. 2012), variability and uncertainty are typically handled by a two-stage Monte Carlo process, but SHEDS-HT currently has a single stage and does not directly estimate uncertainty.

There are many thousand chemicals with registered chemical abstract service registration numbers (CAS numbers) in commercial use. The previous SHEDS models were designed to handle just one chemical (or just a few) per model run. Those SHEDS models were high-tier, performing detailed calculations that required much input data. By contrast, SHEDS-HT is a lower-tier "screening" model that requires relatively little input data for each chemical. This allows runs for thousands of chemicals to be practical.

SHEDS-HT is coded in R. The source code consists of four files, together less than 1 Mb. The default SHEDS-HT input files in the R package are about 4 Mb. Depending on run options, a model run takes between a minute to several hours. The two most important factors affecting the running time are the number of chemicals and the number of simulated people.

SHEDS-HT is a microenvironmental model. A "microenvironment" (abbreviated as "micro") in and refers to general information on location and activity that may be obtained from activity diaries. The person's home is divided into two micros, one while awake and the other while sleeping, since the potential for exposure in those cases is different. Each micro may contain several "media" which are generally surfaces, air, or food, and which are potential sites for the chemical to reside until encountered by the person.

This manual contains three sections:

1) A brief guide to installing and running the SHEDS-HT code
2) A description of the model design and its input and output files
3) A technical description of the SHEDS algorithms

2 INSTALLATION AND RUNNING SHEDS-HT

This section is a brief guide to installing and running the SHEDS-HT model (commonly called "SHEDS"). This is a screening model for estimating direct and near-field exposures to a large assortment of chemicals. The code is written in R. To run SHEDS, the user must have R installed on their computer. R is freely available. Optionally, the user may use R Studio as an interface; this may also be obtained for free. Input and output files are briefly discussed here; see the next section for details of file format and contents.

A "Quick Start Guide" to installing the model and running an example run is included with the SHEDS Distribution package. That guide walks the user through the procedures described in this section.

There are two ways to install and run SHEDS, as an R package, or as a stand-alone set of files. SHEDS was developed as the latter, and there may still be places in this manual that assume that is the case. The main differences when running SHEDS from an R package are that

a) the R functions are automatically available once the "library(ShedsHT)" statement is invoked in R, and

b) the .csv and .txt input files associated with SHEDS are stored in the library as .rda objects which must be unpacked. For example, if the SHEDS folder is chosen to be "C:\SHEDS", then

i) create C:/SHEDS/inputs and C:/SHEDS/output

ii) run setup("C:/SHEDS")

ii) run unpack()

Note that R does not handle backslashes well, as they are interpreted as escape codes for character strings. Be sure to use forward slashes in all paths and filenames used in R.

The first step takes place outside of SHEDS (and R). The "setup" function sets the working directory to the base SHEDS folder, and ensures that necessary packages are loaded. It should be run at the start of each R session. The unpack() function needs to be run just once for each SHEDS installation, because the unpacked .csv files are permanent, and survive outside the R session. Unpack() can be run again to overwrite the .csv files if they have been changed and the user wants the original versions back.

2.1 Organization of the SHEDS-HT Code and Files

The SHEDS-HT Code is written in R, and is included as an attachment. SHEDS-HT can be run as an R package or stand-alone R files. For further instructions about the SHEDS-HT R package,

please refer to the R package instruction manual called Sheds-HT.pdf. The following packages, with their dependencies, must be downloaded and installed:

- data.table (http://cran.r-project.org/web/packages/data.table/index.html)
- stringr (http://cran.r-project.org/web/packages/stringr/index.html)
- plyr (http://cran.r-project.org/web/packages/plyr/index.html)
- ggplot2 (http://cran.r-project.org/web/packages/ggplot2/index.html

The SHEDS code subroutines are included in 4 source files (not relevant when installing SHEDS from the ShedsHT package):

- ShedsHT.R: setup and main exposure routines.
- ReadData.R: subroutines for reading SHEDS-HT input files.
- Utility.R: subroutines for generating random samples from various probability distributions and summarizing results.
- Fugacity.R: subroutines for generating residential media concentrations using a reduced fugacity-based source-to-concentration module.

Additional functions are included in the ShedsHT R package. ExportDataTables.R has code for extracting the .csv and .txt files from the R package. Postprocess.R has code for two functions: filter_sources() and combine_output(). These functions are explained below.

The SHEDS-HT file structure is as follows (where … is the working directory):

- …/R/ SHEDS-HT code R files (if not installed as an R package)
- …/Inputs/ SHEDS-HT Input files
- …/Output/ Location where output exposure files are written

2.2 Running SHEDS-HT

The main steps in Running SHEDS-HT are as follows (assuming all input files are already prepared, with "unpack()" if necessary). Details are provided in the following sections.

1) Create a text file (called the "run file") defining the run settings, including input files
2) Call the **run** function, the main SHEDS-HT call.

Run files may have either a ".txt" or ".csv" extension, but if the latter is used, it must be explicitly included in the function call. If the user creates more than one run, it is best to create a uniquely-named run file for each run, to avoid confusion later. To ensure that all the keywords are correct, it is best to copy and edit an existing run file, saving it with a new name (in /inputs).

The run() command takes two arguments. The first argument is the name of the run file, and is required. The second argument is the SHEDS working directory, and is optional. It is best to check that R is using the SHEDS working directory by using "getwd()", and changing it (if necessary) using "setwd()". File names and paths must be in quotes. For example, to run the enclosed case study "prods_1c" using a working directory C:/SHEDS:

run("run_products","C:/SHEDS/")

As SHEDS runs, progress is written to the console. When some of the other required packages are loaded, messages may be written to the screen but these should not affect the running of the model (unless the required package is not available). SHEDS will continue to write messages until the run is complete, when the R cursor reappears.

Two useful utility functions included in the SHEDS package are "filter_sources" and "combine_output". The first of these may be used to edit any of the three "source file" types (that is, source_scen, source_chem, or source_vars) to select a subset of the sources on that file, as preparation for a new model run. The second function (combine_output) may be run on any output folder to combine selected population statistics for all chemicals into a single file. More details are provided in section 5.4 in the chapter on output.

To use filter_sources, the user supplies the arguments in.file and out.file, and either a list of types, or a list of source.ids. The arguments are not case sensitive. In.file is the name of the source file to be subsetted (that is, one of the files in the /inputs folder for SHEDS). Out.file may be any file name, and is automatically placed in the same folder. Types are specified using the "types=" keyword, and valid types are "product", "food", or "article". Combinations are specified as a list of strings, for example, c("product","article"). Ids are specified using "ids=" followed by a list of values from the "source.id" column of the input file.

2.3 Running the Illustrative Scenarios in the SHEDS package

The SHEDS R package contains several illustrative scenarios, each with a corresponding run file in the /inputs folder for SHEDS. The list is subject to change, but currently includes runs named

food_residues
cpdatv3max
artsandcrafts
other_sources

The example runs and data are illustrative only and with the exception of the "CPDatV3max" run the source chemical data are not intended to reflect any real data or model run results. The "run" files for these are found in the /inputs folder, and each has the

prefix “run_” before the run name. To run one of these, for example the last one, at the R prompt type

run(“run_other_sources”)

The run will create a new folder in the “output” folder of SHEDS, named by the run name (e.g. “other_sources”). The following is a summary of the scope of each illustrative scenario.

“food_residues” covers dietary exposure to all the food types on the food diaries input file (diet_diaries.csv), for one selected chemical, which is set to 1912_24_9 (atrazine). The user may change the chemical by altering the “chemical=” record on “run_foods_1c” before submitting the job.

“cpdatv3max” is a run of consumer product sources developed from V3 of EPA’s CPDat database. The source chemical file (source_chem_cpdatv3_max.csv) for this run contains empirical distributions for maximum weight fractions obtained from product MSDS sheets in CPDat, for products covered by SHEDS-HT (e.g., formulations). The individual product data are also included in this release of SHEDS-HT (in the Supplemental Data folder of the SHEDS-HT GitHub site), so that users can review concentration data for individual sources and make changes if desired. While these data are provided in good faith by EPA, continued curation of these data in CPDat and their translation to SHEDS-HT inputs are ongoing.

“Artsandcrafts” is an example of using the “filter_sources” function on the “source_chem_products” input file to select just consumer product arts and crafts sources ("P.AC.010.029", "P.AC.010.099", "P.AC.020.099", "P.AC.030.029", "P.AC.030.099", "P.AC.040.099", "P.AC.050.029"). The command to create this file was

filter_sources("source_chem_products.csv","source_chem_ac.csv",ids=c("P.AC.010.029", "P.AC.010.099", "P.AC.020.099", "P.AC.030.029", "P.AC.030.099", "P.AC.040.099", "P.AC.050.029"))

It is not necessary to filter the other two source files, because SHEDS processes only the sources that are present on all three source files.

“other_sources” is a small run that provides examples of how to set up the inputs for scenarios not found on the other example runs. These three scenarios are: food migration (that is, migration of chemical into food from packaging), articles, and “downthedrain”. The last of these evaluates the amount of chemical entering the wastewater system, but does not include a human exposure component.

The formats of the various input files are described later in this manual. The user can use the example runs provided with the R package as templates for creating their own runs or using their own data.

3 SHEDS-HT INPUT AND OUTPUT DATA AND FILES

3.1 Handling of Chemical Names and CAS Numbers

SHEDS-HT can handle thousands of chemicals in a single run. Data for these chemicals are spread across input files, so the code must match chemical identifiers from different files. Matching names is very difficult, as some chemicals have common names as well as official IUPAC names, and in many cases there are a variety of written forms for a given name. Another difficulty is the presence of commas in many chemical names, which may be mistaken as field separators, and may not be consistently placed on different files.

SHEDS-HT uses modified CAS numbers for chemical matching. A CAS number has three parts connected by hyphens. The first part ranges from 0 to 999999, the second has two digits, and the third is a single digit. Leading zeroes are always reported in the second and third parts, but are optional in the first part. For example, formaldehyde is usually 50-00-0, but could also be reported as 000050-00-0. SHEDS-HT disregards leading zeroes in the first part when matching CAS numbers.

One concern is that all input files for SHEDS are .csv files, which on many computers are opened in Excel by default. Excel may alter certain CAS numbers because it considers them to be dates. Thus, 2015-05-2 may be changed to 5/2/2015 automatically upon opening the file. Most CAS numbers are unaffected because they do not resemble dates. If the user does not notice that some have been changed and saves the file, those entries will not be recognizable as CAS numbers anymore.

In SHEDS, it is recommended that all CAS numbers be reported with underscores "_" instead of hyphens or dashes "-". Thus, formaldehyde is 50_00_0 in SHEDS-HT. Input files containing hyphens in CAS numbers will still work, but the hyphens are replaced by underscores internally to the model run. The CAS numbers (with underscores) are used as part of the output file names.

3.2 Understanding Sources, Scenarios, and Routes

SHEDS-HT models aggregate exposures from different chemical sources. Currently SHEDS-HT can model exposures to three main types of sources: consumer products (called "products" in SHEDS), consumer articles (called "articles"), and foods. The distinction is that products have specific times and frequencies of use. Articles are items that are present in the home continuously. For example, carpet is an article, but a carpet cleaning chemical is a product.

Exposure can occur from each of these types of sources via different use **scenarios**. A scenario can be thought of as a means by which the chemical and/or the source are used together to result in an exposure. The scenario defines the algorithm that is used to calculate the exposure, and also

defines what parameters SHEDS-HT expects. Route is the pathway by which the chemical enters the human body, and can be either ingestion, dermal, or inhalation. Each scenario in SHEDS can increment exposures by one or more of these routes.

Consumer product scenarios may also be characterized as direct or indirect; direct means exposure is associated with contact with the source during the product use (e.g. inhalation of hairspray during use), while indirect means application of a source to the residence resulting in subsequent exposure at a later time.

The sources present for a run, and the scenarios for which they are active are given on the source-scenario input file. Model input parameters that are specific to a particular **source** are given on the source-variables file; parameters that are specific to a source-chemical combination are given the source-chemicals file. Scenarios that are currently implemented in SHEDS-HT and the routes for which they generate exposures are given in Table 1.

Table 1. SHEDS-HT Exposure Scenarios

Scenario Description	Typical Associated Source Type	Keyword on Source-Scenario File	Routes
Direct application of chemical to skin (either purposefully or incidentally) during use of a consumer product	Product	Product.direct.dermal	Dermal, ingestion (via hand-to-mouth behaviors)
Inhalation of aerosol during use of a consumer product	Product	Product.direct.inhalationAerosol	Inhalation
Inhalation of vapor during use of a consumer product	Product	Product.direct.inhalationVapor	Inhalation
Incidental ingestion of a consumer product during use (not including hand-to-mouth)	Product	Product.direct.ingestion	Ingestion
Application of consumer product to environment leading to subsequent exposures (indirect)	Product	Product.indirect	Dermal, Inhalation, ingestion (via hand-to-mouth)
Down the Drain: A unique exposure scenario that does not result in exposure to the individual, but rather calculates the down the drain mass associated with use of a consumer product e.g. for input to other models	Product	Product.downthedrain	N/A

Scenario Description	Typical Associated Source Type	Keyword on Source-Scenario File	Routes
Emission of chemical from an article in the home (e.g. furnishings)	Article	Article.emission	Dermal, inhalation, ingestion (via hand-to-mouth)
Migration of chemical into food (which is then consumed) from packaging or other contact materials	Food	Food.migration	Ingestion
Consumption of food containing a known chemical residue	Food	Food.residue	Ingestion

Each of these scenarios has a set of one or more Source- and/or Chemical-specific parameters that must be defined on the Source-Variables and Source-Chemicals input files. These required parameters are listed in Table 2 and discussed below in the sections on the input files and the model exposure algorithms.

Table 2. Source- and Chemical-Specific Variables Required by Each Exposure Scenario

Scenario	Parameter Required on Source-Variable File (Source-Specific)	Parameter Required on Source-Chemical File (Chemical and Source-Specific)	Description
Product.direct.dermal	use.prev		Fraction of population using the source
	use.freq		Number of times per year the source is used
	mass		Mass in grams of product per use
	f.contact		Fraction of product mass contacting skin
	f.residual		Fraction of contacting mass remaining on skin after product use
		chem.prev	Likelihood (probability between 0-1) this source type contains the chemical of interest

Scenario	Parameter Required on Source-Variable File (Source-Specific)	Parameter Required on Source-Chemical File (Chemical and Source-Specific)	Description
		f.chemical	Fraction of product mass that is this chemical (if not zero)
Product.direct.inhalationAerosol	use.prev		Fraction of population using the source
	use.freq		Number of times per year the source is used
	mass		Mass in grams of product per use
	f.aerosol		The fraction of product mass that becomes an aerosol
	duration		The duration of product use in minutes
	volume		The volume into which the aerosol disperses during the duration
		chem.prev	Likelihood (probability between 0-1) this source type contains the chemical of interest
		f.chemical	Fraction of product mass that is this chemical (if not zero)
Product.direct.inhalationVapor	use.prev		Fraction of population using the source
	use.freq		Number of times per year the source is used
	mass		Mass in grams of product per use
	duration		The duration of product use in minutes
	volume		The volume into which the aerosol disperses during the duration
		chem.prev	Likelihood (probability between 0-1) this source type contains the chemical of interest

Scenario	Parameter Required on Source-Variable File (Source-Specific)	Parameter Required on Source-Chemical File (Chemical and Source-Specific)	Description
		f.chemical	Fraction of product mass that is this chemical (if not zero)
Product.direct.ingestion	use.prev		Fraction of population using the source
	use.freq		Number of times per year the source is used
	mass		Mass in grams of product per use
	f.ingested		Fraction of mass that is ingested
		chem.prev	Likelihood (probability between 0-1) that this source type contains the chemical of interest
		f.chemical	Fraction of product mass that is this chemical (if not zero)
Product.indirect	home.prev		Fraction of houses in which the product is used
	use.freq		Number of times per year the source is used
	mass		Mass in grams of product per use
		chem.prev	Likelihood (probability between 0-1) this source type contains the chemical of interest
		f.chemical	Fraction of product mass that is this chemical (if not zero)
Product.downthedrain	use.prev		Fraction of population using the source
	use.freq		Number of times per year the source is used
	mass		Mass in grams of product per use
	f.drain		Fraction going down the drain

Scenario	Parameter Required on Source-Variable File (Source-Specific)	Parameter Required on Source-Chemical File (Chemical and Source-Specific)	Description
		chem.prev	Likelihood (probability between 0-1) this source type contains the chemical of interest
		f.chemical	Fraction of product mass that is this chemical (if not zero)
Article_emission	home.prev		Fraction of homes with this article. May also be called "use.prev"
	f.area		Emission area of source as a fraction of the house floor area
		chem.prev	Probability that this chemical is present
		y0	Steady state gas concentration (ug/m3)
Food.migration		chem.prev	Probability that this chemical is present
		migration.conc	Concentration in food due to chemical migration from packaging (ug/g)
		contact	Currently this is a placeholder variable for a future function that calculates chemical migration from packaging to food. Always set to 1 for now.
Food.residue		residue	Concentration in food (ug/g)
		detects	Number of detected residues in data source (can be set to 1 along with nondetects=0 to model nonzero residues for all individuals)
		nondetects	Number of detected residues in data source

3.3 SHEDS-HT Input Files

SHEDS-HT requires twelve input files. However, many of these files can be considered defaults (e.g. U.S. population information) and are not typically changed between runs. Usually, the user will only need to change the Run file (which is a file containing a limited number of settings for the current run) and the Source-Scenarios, Source-Variables, and Source-Chemicals files. These three file define the chemical sources being modeled and provide the required information associated with them.

With the exception of the run file (which is a .txt file), all input and output files for SHEDS-HT are CSV files, which have a ".csv" extension to their names. These may be opened for viewing or editing in Excel or most text editors. These CSV files contain one header record which contains the column (or variable) names, followed by any number of data records. All fields are comma separated, so it is bad practice for any of the character fields to contain commas. SHEDS requires 12 input files for a run; most of these are universal and do not need modification from one run to another. Table 3 lists these files along with a short description of their contents.

Table 3. Input Databases and Files for SHEDS-HT

File	Keyword in Run File	Description	Source for Default Information
Run File		List of job settings for this model run	User-defined
Activity Diary File	act.diary.file	Daily location (time spent in microenvironments) and activity (physical activity index PAI, a time-averaged metabolic equivalent) information for each age- and gender-specific CHAD diary	Calculated from USEPA's CHAD (McCurdy 2004; USEPA 2014)
Chemical Property File	chem.props.file	Chemical-specific property information by CAS number s	Chemical Properties from ORD's CompTox Dashboard where available; either OPERA model predictions or means or reported measured values; other values derived from EPI Suite (USEPA 2012).
Dietary Diary File	diet.diary,file	Daily mass of food group (by default crop group; see Table 5) consumed by individuals	Calculated from NHANES-WWEIA food diaries (USDA 2014)
Exposure Factor File	exp.factor.file	Distributions for various human exposure factors	See Section on Exposure Factor File

File	Keyword in Run File	Description	Source for Default Information
Fugacity File	fugacity.file	Distributions for house variables needed for fugacity modeling	See Section on Fugacity Input File
Media File	media.file	List of media in each micro that may contain chemicals	N/A
Physiology File	physiology.file	Body mass, height, and basal metabolic rate distributions or regressions by age/gender cohort	Developed for SHEDS-Multimedia from NHANES (Glen et al. 2012)
Population File	population.file	Number of individuals in US in each age age/gender cohort	2000 US Census (US Census Bureau 2000)
Source-Scenario File	source.scen.file	Lists of active exposure scenarios for each potential source	User-Defined
Source-Chemical File	source.chem.file	Distributions of variables dependent on both source and chemical	User-Defined
Source-Variable File	source.vars.file	Distributions of source-specific variables	User-Defined

3.3.1 Run File

The main SHEDS-HT input file is the "Run" file. This is a text file that defines the main run settings and assigns what versions of the other SHEDS-HT input files to use. To run SHEDS with the Run file "xxx.txt", type 'run("xxx")' at the R command prompt. This presumes that the file xxx.txt is located in the /inputs folder, with all the other input files.

Each Run file contains a header line, 12 settings, 11 file names, and an arbitrary number of chemical CAS numbers. Except for the header, all lines contain an equal sign "=", which separates a keyword on the left from a value on the right. The easiest way to ensure that the keywords are correct is to copy and edit a previous Run File, changing the values when necessary.

The **run.name** keyword specifies the name for a folder under /output that is to contain all the output from this run. If a folder with that name already exists, it is overwritten. Otherwise, it is created. The limitations on run.name are those of the operating system for valid folder names. The user should ensure that write permission is available to /output and any subfolders. Each of the twelve setting and the keywords are described below. The keywords are not case sensitive.

N.persons is the total number of simulated persons in this model run. Each will have one day of exposure calculated for each chemical in the run, although for some chemicals that particular day

may have zero exposure. Set n.persons to a minimum of two to ensure that all objects defined across persons are treated as vectors rather than scalars.

Person.output is a switch indicating whether output should be saved for every simulated person. If set to "yes", then a file with the suffix "all.csv" is created for each chemical, with a record for each person. If set to "no", then only summary files are created.

Source.output is a switch for computing and writing files containing source-specific data. If "yes", then files with the suffix "srcMeans.csv" are created for each chemical. Each contains all sources of that chemical, with the mean for dermal exposure, ingestion exposure, and inhalation dose, all in units of micrograms per day, averaged across all simulated persons.

Min.age is the lowest age in years to include in the simulation. Valid values are integers from 0 to 99.

Max.age is the highest age in years to include in the simulation. Valid values are integers from 0 to 99. It must be no smaller than min.age.

Genders indicates the genders to be modeled. Set to "F" for females only, "M" for males only, or either "MF" or "FM" to include both. The quote marks are not necessary.

Seasons specifies which seasons are to be included in the diary selection. Use "W" for winter (Dec, Jan, Feb), "P" for spring (Mar, Apr, May), "S" for summer (Jun, Jul, Aug), and "F" for fall (Sep, Oct, Nov). For all seasons use "WPSF", or any reordering of those letters.

Details is a switch for printing summary information on each chemical to the R log. For large runs the log may overflow causing loss of earlier information, so set details=0 for larger runs. For small runs, either 0 or 1 are reasonable.

Age.match.pct is a parameter for the allowed error in matching diary ages to simulated person ages. If age.match.pct=0 then exact matching is required for all persons. For larger values, the age of the simulated person defines the center of an age window extending from age*(100-age.match.pct) rounded down, to age*(100+age.match.pct) rounded up.

Run.seed is the initial setting for the random number seed. Other seeds are generated internally in SHEDS from this one. Repeating the same run with the same run.seed will produce identical results. Repeating the same run except for a change in run.seed will produce additional simulated persons that are independent of the first run, but may be combined with the earlier run to better define the output distributions.

Set.size is an internal setting for SHEDS. It defines vector and array lengths which depend on the number of persons. It is found in practice that when n.persons exceeds 20,000 or so, SHEDS slows down because storing and copying large arrays becomes burdensome. Therefore, the run is

broken internally into “sets”. A good set.size is 10,000 for current computers, but this may change with newer technology.

If a run consists of more than one set (that is, n.persons > set.size), then there will be additional output files for each chemical, labeled by set number. Problems in R may occur if set.size is set to 1, as in some cases scalars are handled differently than vectors.

The next 11 lines on the Run File are the names of the input files. All are assumed to be found in /inputs under the SHEDS installation. All files must be .csv files. The names shown on the Run file may optionally omit this extension; SHEDS will add it on if it is not there. Of course, the actual file names must include the .csv extension.

After the file names, the user may specify the chemicals to be considered in this run. If no chemicals are specified, then all chemicals found on both the source.chemicals and chem.props files are run. Otherwise, only the chemicals listed on the Run file are run (and only if they also appear on both source.chemicals and chem.props). Each chemical is listed on a separate line, with a separate keyword (which is “chemical = “). The values on the right of “=” are the CAS numbers, using underscores. For example, propane is 74_98_6.

3.3.2 Activity Diary File

The Activity Diary (ActDiaries) file contains a summary of the activity diaries from the Consolidated Human Activity Database (CHAD, US EPA 2014). It contains summaries of 24-hours of time-activity data from thousands of people. The information on each record consists of the following variables described in Table 4.

Table 4. Variables on the Activity Diary Input File

Parameter	Description
Age	Age in full, completed years, as an integer. Children below 12 months of age have age zero. In SHEDS the maximum age is set to 99.
Gender	“F” for female, or “M” for males.
Weekend	A flag, set to “1” for Saturday or Sunday diaries, or “0” otherwise. On a given model run, 2/7 of the simulated people selected at random will be assigned a weekend diary)
Season	A single-character variable indicating season of the year. The choices are “W” (winter), “P” (spring), “S” (summer), or “F” (fall). The mapping was 3 months per season, with December, January, and February being winter
Bath	”1” means a bath or shower was reported, “0” otherwise.

For every micro on the Media file with contact.p>0 (see Section 3.3.7 for definition), The activity dairy file has two corresponding fields, one with suffix “min” and the other with suffix

"pai". For example, "In.awk" is the micro name on the default Media file for "in home, awake", so the file has fields "In.awk.min" and "In.awk.pai". The first of these is the number of minutes (out of the 1440 on the diary day) spent in home while awake. The second is the average physical activity index (PAI) during that time. The PAI represents the rate of metabolic energy expenditure divided by the basal or resting rate. Sleep is assigned a PAI of 0.9 because it consumes less energy than resting while awake. Other PAI values should generally be between 1 and 4. PAI values above that are possible, but are difficult (or impossible) to sustain for long periods. SHEDS uses the PAI values to calculate breathing ventilation rates.

The current version of the default Activity Diaries File has 33,748 daily time-activity records.

3.3.3 Chemical Properties File

The default Chemical Properties file contains information on 2,423 CAS numbers. This default file was originally created by running EPA's Estimation Program Interface Suite (EPI Suite) models for all the CAS numbers recognized by EPI Suite (over 100,000 chemicals). In December 2017, where possible the EPA-Suite predictions were replaced with values predicted with ORD's OPERA models [] or means of measured values from ORD's Computational Toxicology Chemistry Dashboard. The exceptions are the variables Kp and fabs are defined and described below. This input file can be custom-made for any run to include only the chemicals of interest. The chemicals that are required to be on the file (and the column indicated with the following keywords) are:

- CAS: Chemical Abstracts Service Registration Number (CASRN). The file header for this variable can also be "Chem" or "Chemical"
- Name: name of the chemical to be used on the screen log. Names should be in quotes to avoid accidental interpretation of internal commas as field delimiters on the input file. The names appear only on the Chemical Properties file, so they do not need to be matched to any other input
- Kp: the skin permeability constant in which SHEDS uses to determine dermal absorption [cm/hr]. This is calculated using the ten Berge (2009) model.
- MW: molecular weight in [g/mol] of the target chemical.
- VP.Pa: vapor pressure in [Pa].
- Log.Kow: the base ten logarithm of the octanol-water partition coefficient
- Water.sol.mg.l: solubility in units of [mg/L]
- Half.sediment.hr: Half-life in sediment [hours], as predicted by the EPI Suite Level III Fugacity model. Used to estimate the decay of chemicals indoors.
- Half.air.hr: Half-life in air [hours], as predicted by the EPI Suite Level III Fugacity model. Used to estimate the decay of chemicals indoors.
- Fabs: absorption fraction for the ingestion route, calculated from O'Connor et al. (2013).

Other variables may appear on this file, as long as their names do not start with any of the above strings. The fields do not have to be in the above order on the input file. All the variable names are converted to lower case in the code, so are not case sensitive. A larger version of the Chemical Properties file (not included in the R package) has been produced, which contains many more variables from EPI Suite that may be used in future model algorithms.

3.3.4 Dietary Diaries File

The Dietary Diaries file contains daily intakes in [g/day] for various categories of food and water. Each record corresponds to one person for one day. The default diary file contains aggregated summaries of food consumption data from the NHANES-What We Eat In America Study (1999-2006, USDA 2014; NHANES-WWEIA). Food consumption in grams is given for each of the SHEDS-HT default food groups, for each person in the study. The variables on the file are the age, gender, body weight in [kg], and the intake quantities. The food groups are indicated by short labels (usually two characters). These are matched in SHEDS with distributions from the Source.chemicals input file. Food group "XX" on the diet.diaries file is matched to the source name beginning with "FOOD; XX:" on the Source.chemicals file. The user may use their own choice of food groups provided both input files are consistent.

The food groups used in the SHEDS-HT default file are given in Table 5. Note that the SHEDS-HT model is compatible with other food group definitions, and additional columns can be added to the file if residue or other food scenario information is available for different groups; the exposure algorithms may use different food groups but they will still be summed over scenarios within chemicals.

Table 5. SHEDS-HT default food groups (crop groups).

Crop Group Label in SHEDS	Crop Group Description
AS	Asparagus
BA	Banana
BF	Beef
BR	Brassica (Cole) Leafy Vegetables
BS	Berry and Small Fruit
BV	Bulb Vegetables
CF	Citrus Fruits
CG	Cereal Grains
CH	Chicken
CV	Cucurbit Vegetables

Crop Group Label in SHEDS	Crop Group Description
DP	Dairy Products
DW	Drinking Water (Direct, Indirect)
EF	Edible Fungi
EG	Egg
FI	Fish
FV	Fruiting Vegetables (except cucurbits)
GO	Goat
HO	Honey
HS	Herbs and Spices
LE	Leafy Vegetables (except Brassica vegetables)
LR	Leaves of Root and Tuber Vegetables
LV	Legume Vegetables (Succulent or dried)
MA	Mango
MG	Meat, game
NA.	Non-grass Animal Feeds (Forage, Fodder, Straw and Hay)*
OS	Oilseeds
OT	Other
PA	Papaya
PE	Peanut
PF	Pome Fruits
PI	Pineapple
PO	Poultry, Other
PR	Pork
RA	Rabbit
RS	Raisins
RT	Root and Tuber Vegetables
SF	Stone Fruits
SH	Sheep
TN	Tree Nuts

Crop Group Label in SHEDS	Crop Group Description
TU	Turkey

*Alfalfa is only food commodity in this group

Not all of these food groups are required to have data on the Source.chemicals file. SHEDS assumes that any food groups with no date do not contain any of the target chemical, so the exposure from that source is zero. Similarly, if any food groups on the Source.chemicals file do not appear here, SHEDS assumes that the consumption of that food group was zero, so again the exposure from that source is zero.

3.3.5 Exposure Factors Input File

This file contains distributions for various exposure factors. The variables on this file are given in Table 6.

Table 6. Variables on the Exposure Factors File

Parameter	Description
varname	name of the modeling variable being defined
form	type of distribution
par1-par4	parameters relevant to the specified form
lower.trun	lower truncation bound (minimum returned value)
upper.trun	upper truncation bound (maximum returned value)
resamp	flag: yes=resample, no=stack at truncation bounds
min.age	minimum age person using this distribution
max.age	maximum age person using this distribution
gender	gender using this distribution
season	seasons using this distribution
media	medium using this distribution

Currently, SHEDS requires distributions for the following variables in Table 7; default values for the input parameters and how they were derived are provided.

Table 7. Default SHEDS-HT Exposure Factor Input Parameters

Parameter	Description	Units	Dependencies	Value / Source in SHEDS-HT
avail.f	Fraction available for transfer from surfaces via touching	-		0.5 (assumed)
dermal.tc	Dermal transfer coefficient	cm^2/hr		Lognormal (250,1.93) Fit to daily aggregated SHEDS-MM results (Glen et al., 2012)
om.ratio	Object-to-mouth ratio (ratio of object-to-mouth exposure to indirect dermal exposure)	-	Cohort	Age 0-5: Triangle (0.001,0.004,0.002) Age 6-99: Triangle (0.0001,0.0004,0.0002) Based on fits for SHEDS-MM, which was parameterized based on Xue et al., (2010).
bath.p	Probability of bathing on a given day, if the CHAD diary did not record one	-		0.8 Based on SHEDS-MM (Glen et al., 2012)
handmouth.area.f	Hand-to-mouth area fraction	-	Cohort	Age 0-5: Beta (3.7,25) Age 6-10: Uniform (0.0.1) Age 11-99:Uniform (0,0.02) Based on SHEDS-MM
handmouth.freq	Hand-to-mouth frequency	#/hr	Cohort	Weibull (0.75,12.59) Based on (Xue et al. 2007)
handwash.freq	Handwashing frequency	#/day		Lognormal (3.6,2) Same as SHEDS-MM (Glen et al. 2012)
ingest.abs.f	Absorption fraction (ingestion)	-		No longer used (chemical-specific fraction now provided on chemical property input file).
inhal.abs.f	Absorption fraction (inhalation)	-		0.16 Assumed
rem.bath.f	Daily chemical removal fraction (bathing)	-		Beta (4,1.36) Fit to daily aggregated SHEDS-MM results
rem.handwash.f	Chemical removal fraction from a single handwashing event	-		Triangle (0.05, 0.5, 0.25) Same as SHEDS-MM (Glen et al. 2012)

Parameter	Description	Units	Dependencies	Value / Source in SHEDS-HT
rem.brushoff.f	Chemical removal fraction (brush-off)	-		Triangle (0.01, 0.25, 0.05) Daily fit to SHEDS-MM runs that assumed 1% per hour
rem.derm.abs.f	Absorption fraction (dermal)	-	Chemical	Initial distribution is Triangle (0.001, 0.01, 0.004). Sampled result is scaled by K_p (determined from EPI Suite) as described in eq. b23
urine.f	(Excretion fraction)	-	Chemical	Uniform(0.3, 0.7) Assumed

Variable names ending in ".f" are proper fractions which are bounded by zero and one. Names ending in ".p" are probabilities which are also bounded by zero and one. Other variables are bounded below by zero. More restrictive bounds may be specified using lower.trun and/or upper.trun. The Exp_factors file requires definitions for all persons that may be generated in the model run. If all persons are to sample from the same distribution, then that modeling variable needs just one line on the Exp_factors file, with the min.age, max.age, gender, season, and media fields left blank. A blank entry in any of these fields implies that this distribution applies to all cases of that field variable. The default file has age-specific distributions for two variables: om.ratio and handmouth.area.f. All input variables names in the file are not case sensitive (the code converts them all to lower case upon reading them).

The modeling variables serve the following purposes in SHEDS-HT:

Avail.f – This represents the fraction of chemical in or on surfaces that could potentially be transferred to the skin. SHEDS has a built-in chemical fugacity model that moves chemical between various compartments (either surfaces or air). Some of the chemical in "surface" compartments is bonded to (or trapped inside) the material rather than sitting on the surface. The variable avail.f represents a common factor applied to all sources. Related factors are found on the Source-variables file, which may be unique to each source.

Dermal.tc – This is a surface-to-skin transfer coefficient, with units of [cm^2/hr]. It is a net rate representing the effective surface area that transfers all of its available chemical in a unit time. For example, suppose the chemical concentration on the relevant surface is 10 ug/cm^2, avail.f = 0.5, dermal.tc = 200 cm^2/hr, and the duration of contact is 30 min. Then the amount of chemical transferred to the skin is 10 * 0.5 * 200 * 30 * (1/60), where the last factor converts minutes to hours. The result is a new dermal exposure of 500 ug of this chemical during this 30 minute period. The daily exposure is the sum of similar expressions over the list of chemical-containing media.

Om.ratio – This variable represents a unitless ratio of the new indirect ingestion exposure to new indirect dermal exposure. Indirect exposure refers to chemical that is encountered not during direct use of the source (or consumption of food), but is encountered after it has migrated away from the source. The built-in fugacity model handles these calculations. People (especially children) tend to place various objects (including parts of their hands) in contact with their mouth or lips. In the fugacity model, every surface and every object in the home will have some amount of chemical on it after it has had time to migrate, if any of that chemical is present at all in the home. SHEDS-Multimedia demonstrated that there was a strong correlation between indirect dermal exposure and indirect ingestion exposure for persons in the same age-gender cohort. To keep inputs to a minimum, SHEDS-HT codifies this explicitly. Setting the om.ratio distribution to a point value ensures that this correlation is 100%. Distributions with variability result in smaller correlation.

Bath.p – Like all ".p" variables, bath.p should be specified to have "binomial" form ("Bernoulli" is equivalent). The par1 value gives the probability of a "yes" or "1" outcome, while the remaining people are assigned a "no" or "0" outcome. Bath.p is a correction factor to adjust for the undercounting of bath or shower activities in the ActDiaries file. Bath.p is the fraction of persons not reporting a bath or shower on their diary who actually had one that day. If the diaries are assumed to be accurate, set bath.p to zero.

Handmouth.area.f – Hand-to-mouth transfer is one of several competing processes for removing chemical from the skin. The rate of hand-to-mouth transfer depends on several exposure factors (specifically handmouth.area.f, handmouth.freq, rem.handwash.f, and handwash.freq), as well as on the dermal exposure and the relative sizes of the other removal pathways. Children may place one or more entire fingers in the mouth, while others may just touch or lick their fingertips, or bite their nails. The larger the contact area, the larger is the fraction of the dermal exposure that may be transferred to the mouth. Handmouth.area.f represents the fractional surface area of both hands that is licked or placed in the mouth.

Handmouth.freq – This is the number of hand-to-mouth contacts per hour, used in conjunction with handmouth.area.f. Handmouth.freq represents the average rate over the waking hours of the day. The amount of chemical transfer is not linear in handmouth.freq, but is fitted to results from SHEDS-Multimedia. Regardless of the handmouth frequency, one cannot remove more than the amount of dermal exposure. Each successive contact transfers less chemical, because the hand loading is reduced by prior hand-to-mouth transfers.

Handwash.freq – This is the number of hand washes per day. The user may enter a continuous distribution (for example, the default is lognormal) and the mean can be non-integer. The result is rounded probabilistically to a whole number. For example, if 4.75 is randomly generated for a given person, it is either rounded to 4 (with 25% probability) or 5 (with 75% probability) for the given simulation day.

Ingest.abs.f – This represents the fraction of ingested chemical that is absorbed from the GI tract into the bloodstream.

Inhal.abs.f – This is the fraction of the chemical mass in inhaled air that passes into the blood. The remainder is assumed to be exhaled.

Rem.bath.f – This is the fraction of daily dermal exposure that is removed by bath or shower events.

Rem.handwash.f – This is the fraction of daily dermal exposure that is removed by hand washing.

Rem.brushoff.f – This is the fraction of daily dermal exposure that is removed by friction, skin flaking, or other processes not involving washing, bathing, or hand mouthing.

Rem.derm.abs.f – This is the fraction of daily dermal exposure that is removed by being absorbed into the body. No distinction is made between absorption into subcutaneous fat or into the blood. See the section on the modeling of dermal removal processes for more details.

Urine.f – This is the fraction of the total absorption (that is, from ingestion, inhalation, and dermal absorption) that is excreted in the urine. It is used to calculate chemical excretion in micrograms per day, for comparison to measurement data. If the excreted chemical is a metabolite of the original, then differences in molecular weight should be included, so urine.f represents the ratio of excreted metabolite mass to original (parent) chemical absorption.

In many cases, the literature provides only point estimates for these variables. SHEDS allows (and expects) that there should be variability between individuals. The user should choose whether to define such variables as point values (using the "point" distribution), or whether some estimate of variance (such as an expert-opinion coefficient of variation, or CV) be used. If the user chooses an open-ended distributional form such as the Normal or Lognormal, then set upper and/or lower truncation bounds to prevent the generation of unreasonable values.

3.3.6 Fugacity Input File

The fugacity input file contains distributions for 21 variables used in the modeling of chemical dispersion inside the house of each simulated person. Each of these variables is sampled once per person. For details on how these variables are used, see the section on fugacity modeling. The columns on the input file are varname, units, form, par1, par2, par3, par4, lower.trun, upper.trun, resamp, and descrip. The units and descrip are user information and are not used by the code. The other columns are standard for input distributions, as described in the Exposure Factors File above. This file allows only one distribution per variable, which applies to all houses. The "varname" must match the name of the corresponding modeling variable (see below).

The fugacity model is simplified down to two compartments, air and floor. The air represents the total air volume of the heated portion of the house, where the air is free to circulate and become well mixed. The floor is a proxy for surfaces in general. Houses with more floor area will have more surface area as well. The model does not distinguish between various types of surfaces like carpet, hard floors, tables, chairs, and so on. To do so would require far more input variables to determine chemical flows, and would also require information on personal behavior to determine the amount of contact with each type of surface. That is beyond the scope of the model.

This version of the fugacity model can handle two types of chemical application: and instantaneous release designated to occur at t=0, and a continual constant release. The former is typical of a short-duration single usage of some product, while the latter is typical of (say) emissions from carpets or furniture. A product that is used or released very frequently may be better modeled as a constant source.

In this model, chemical may flow in or out of the house by air exchange, from the air to the floor by diffusion or by particle deposition, or from the floor to the air by diffusion or resuspension of particles. The variables in the file and their default values are given in Table 8. See the Model Code and Algorithm section for a description of the full fugacity-based fate and transport model.

Table 8. Variables on the Fugacity Input File

Parameter Name	Description	Units	Default Value in SHEDS-HT‡
aer.out	Air exchange rate for rooms with outdoors	1/d	Lognormal(11.9, 1.7); Assumed
area.sur	Total floor area of the house	m^2	Lognormal(130, 1.8); Assumed
height	Height of walls	m	Uniform(2.44,3); Assumed
lg.carb.f	Organic carbon fraction for large particles	-	Normal (0.15,0.01); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.clean.air	Cleaning removal rate for large particles in air e.g. from air filters on HVAC systems, or electrostatic filters.	1/d	Uniform (0.03, 0.5); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.clean.sur	Cleaning removal rate for large particles on surfaces, e.g. from vacuuming or mopping.	1/d	Uniform (0.035, 0.045) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.depos	Air-to-floor large particle deposition rate	m/d	Normal (387,20) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed

Parameter Name	Description	Units	Default Value in SHEDS-HT‡
lg.load.air	Loading of large particles in air	µg/m^3	Uniform (2.2, 2.5) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.load.sur	Loading of large particles on surfaces	µg/cm^2	Uniform (11.5, 28) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.resus	Surface-to-air large particle resuspension rate	1/d	Uniform (0.0015, 0.0017) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
sm.carb.f	Organic carbon fraction for small particles	-	Normal (0.3,0.03); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.clean.air	Cleaning removal rate for small particles in air, e.g. from air filters on HVAC systems, or electrostatic filters.	1/d	Uniform (0.018,0.22); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.clean.sur	Cleaning removal rate for small particles on surfaces, e.g. from vacuuming or mopping.	1/d	Uniform (0.035, 0.045) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.depos	Air-to-floor small particle deposition rate	m/d	Normal (11,1) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.load.air	Loading of small particles in air	µg/m^3	Uniform (15, 25) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.load.sur	Loading of small particles on surfaces	µg/cm^2	Uniform (6, 14.5) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.resus	Surface-to-air small particle resuspension rate	1/d	Uniform (0.00072, 0.00082) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
temp	Indoor temperature	K	Normal (296,2); Assumed
thick.bou	Boundary layer thickness over surfaces	m	Uniform (0.025,0.0275); Assumed
thick.sur	Effective thickness of surfaces	m	Normal (0.0098,0.002); Assumed

‡Normal distributions reported as (mean, SD); lognormal as (geometric mean, geometric standard deviation); uniform as (min, max)

3.3.7 Media File

The Media File lists the potential exposure media and relates them to the microenvironments (micros) on the Activity Diaries file. Each micro has a time duration and a ventilation rate (expressed as a physical activity index, or PAI) for each person-day on the ActDiaries file. The default file contains all the currently anticipated media for holding chemical. However, exposure pathways have not been built into SHEDS yet for the micros other than In.awk and In.slp. The column labeled "contact.p" refers to the overall contact probability with the given medium when in that micro. Media with contact.p=0 are deleted from further consideration. For example, In.awkPet is on the default file because exposure from chemicals on pets in the home was part of earlier SHEDS models, and may be introduced into SHEDS-HT in the future. Setting contact.p to be non-zero on this file would only ensure that the input file information is retained; it would not change the model output unless the user added code that referred to those variables. For pets, contact.p should not be set to one because that would imply that the person is always near a pet when in that micro. For air and surfaces, contact.p=1 because the person is always breathing and always sitting, lying down, and/or touching surfaces, and all surfaces can contain chemicals.

The micros in SHEDS are:

- In.awk Indoors at home, awake
- In.slp Indoors at home, sleeping
- In.wrk Indoors at work or school
- In.oth Indoors at other locations
- In.veh Inside a motorized vehicle
- Out.hm Outdoors at or near home
- Out.wrk Outdoors at or near work or school
- Out.oth Outdoors at other locations

The user cannot change the above list without recoding the ActDiaries input file to correspond with the new micros. The SHEDS_HT R package contains a reduced set of activity diaries which record only the in.awk and in.slp times. The full activity diary set contains times for all the above micros.

The user can add or delete media within each micro. If the user wanted (for example) to define two surface types in the home, then contact.p values should be lower (say, 0.5 each). That would reduce the contact duration in the relevant exposure equation by that factor (in this example, hard surfaces for half the time awake in the home, and soft surfaces for the other half). Of course, the exposure equations would have to be modified to utilize the new media names.

3.3.8 Physiology File

This file contains distributions and parameters relating to human physiology. There is a header line followed by 200 data records, one for each year of age (0-99) for females, followed by the same for males. The parameters in this file are given in Table 9.

Table 9. Parameters on the Physiology File

Parameter	Description	Ages
wgtmean	Mean value of natural logarithm of body weight (kg)	All
wgtstdev	Standard deviation of natural logarithm of body weight (kg)	All
hgtmean	Mean height (cm)	0-19
hgtstdev	Standard deviation of height (cm)	0-19
hgtslope	Slope of height versus log(weight) regression	20-99
hgtinter	Intercept of height versus log(weight) regression	20-99
hgtresid	Standard deviation of residual of height-weight regression	20-99
bmrslope	Slope of bmr versus weight regression	All
bmrinter	Intercept of bmr versus weight regression	All
bmrresid	Standard deviation of residual of bmr-weight regression	All
gender	“F” for females or “M” for males	All
age	Age in full years (0-99)	

SHEDS bases height on age from ages 0-19, and on body weight for persons age 20+. Weight is in kilograms (kg), height in centimeters (cm), and basal metabolic rate (BMR) in megajoules per day (MJ/day). SHEDS calculates a basal alveolar ventilation “bva” rate from BMR. The “bva” values are modified by the physical activity index (PAI) for each micro, as read from the activity diaries, to determine the micro-specific ventilation rates. Height is used to determine body surface area (BSA). Weight is used in the BSA calculation and in the determination of dose on a unit body weight basis.

3.3.9 Population Input File

The population file contains a header line plus 100 data rows, one each for ages 0-99. The three entries on each row are age, #males, and #females. The default file contains counts for the total U.S. population from the 2010 census. SHEDS converts the data on this input file to fractions of the total population. The purpose of this file is to assign age and gender randomly using probabilities consistent with the general population. Therefore, a SHEDS run of any size would statistically be a representative sample of the total U.S. population.

If SHEDS is run for a limited age range, then the counts for ages outside that range are set to zero, so the run will randomly generate ages within the allowed range consistent with the overall population distribution in that range.

If the user is targeting a different but known population, then this file should be replaced by one containing the appropriate data.

3.3.10 Source-Scenarios Input File

The Source-Scenarios file is one of the main input files that customizes the SHEDS-HT run. This file lists the chemicals sources to be included in the run and the scenarios to be run for each. SHEDS currently has nine scenarios (see Table 1), each of which has a column on the Source-Scenarios file. The first column contains the source type: one of food, product, or article. The second column is the source.ID, which is a brief label used for matching data to other input files. Each line on this file must have a unique source.ID, but otherwise they are arbitrary, and the user may add new ones. The third column is a description of the source. This is not used by the model (the source.ID is always used internally). The fourth column is a flag indicating whether the source in considered to be indoors (1) or outdoors (0), which may impact the inhalation.

There is one column for each exposure scenario. The source type must match the first part of the scenario name. But not all appropriate scenarios need to be active. A "1" indicates that the scenarios is active, or "0" means that the combination of source and scenario will not be evaluated. To eliminate sources from the model run, either delete their records in this file, or else set all eight columns to 0.

Customizing the source-scenarios file allows focused model runs without needing to alter any of the other input files.

3.3.11 Source-Chemicals File

This file contains distributions for modeling variables that depend on both the source and the specific chemical. Currently, this means the modeling variables chem.prev and f.chemical for non-food sources, and detects, nondetects, and residue for food residue sources. The columns on this file are described in Table 10.

Table 10. Columns on the Source-Chemicals File

Parameter	Description
source.ID	Arbitrary, but matched with Source from other files
source.description	For user only, not used by SHEDS
variable	One of "chem.prev", "f.chemical", "detects", "nondetects", or "residue"

Parameter	Description
chemical	The CAS number, with underscores in place of hyphens
units	For the user's reference (not used by SHEDS)
gender	Male (M), female (F), or both (B)
min.age	Minimum age in full years
max.age	Maximum age in full years
form	The type or shape of the distribution
mean	The mean value for this variable
cv	The coefficient of variation (standard deviation / mean) for this variable

The recognized distributional forms are Bernoulli, binomial, beta, discrete, empirical, exponential, gamma, lognormal, normal, point, probabilistic, and triangle. These divide into two categories. The ones that return a set of discrete values are Bernoulli, binomial, discrete, empirical, point, and probabilistic. The continuous distributions are beta, exponential, gamma, lognormal, normal, point, and triangle. A "point" may be used when either type is expected.

3.3.12 Source-Variables Input File

This file is similar in format to the Source-Chemicals file, except that the variables in this file apply to all chemicals found in a given source type. Therefore, the column labeled "chemical" is not present. The options for the "variable" column on this file are given in Table 11.

Table 11. List of modeling variables on the source-variables file

Parameter	Description
duration	Time in minutes (per use)
f.contact	Fraction of product mass that contacts skin
f.residual	Fraction of contacting mass that remains on skin after use
mass	Mass of product (per use)
use.freq	Frequency of product use (# per year)
use.prev	Fraction of population that use this product
area	Surface area of product application
volume	Average air volume into which product spreads during use duration
f.aerosol	Fraction of product mass that becomes aerosolized
f.ingested	Fraction of product mass that is (accidentally) ingested

3.4 Output files

SHEDS-HT produces three types of output files. Each is chemical-specific, and the CAS number is part of the filename. A large run with thousands of chemicals will produce thousands of output files. The three file types are Statistics, Source Means, and All Individuals. Statistics files are run for each set of simulated people, and contain means, standard deviations, and quantiles for various exposure and dose variables. The Source Means file are produced for each set and for the run as a whole, and contain population means for each exposure source. The All Individuals file contains demographic, exposure, and dose variables for all simulated persons in the run. However, source-specific information is not included on the All Individuals file.

3.4.1 The All Individuals File

There is one All Individuals file per chemical. It has one record for each simulated person, and covers all the sets (if there is more than one). The variables on this file are described in Table 12.

Table 12. Parameters in the All Individuals Output File

Parameter	Description
person	Sequential numbering from 1 to the total number modeled
gender	“F” for female or “M” for male
age	Age in full years, rounded down
season	“P” (spring), “S” (summer), “F” (fall), or “W” (winter)
weekend	1 = Saturday or Sunday, 0 = other day of week
weight	Body weight in kilograms (kg)
exp.dermal	Dermal exposure (ug/day)
exp.inhal	Inhalation exposure (ug/m3), averaged over the day
exp.ingest	Ingestion exposure (ug/day)
exp.diet	The dietary part of the ingestion exposure (ug/day)
exp.nondiet	The non-dietary part of the ingestion exposure (ug/day)
exp.drain	The chemical mass entering the water disposal system (g/day)
dose.inhal.ug	The inhalation intake dose (ug/day)
dose.intake.ug	The sum of exp.ingest + dose.inhal.ug + abs.dermal.ug (ug/day)
dose.intake.mgkg	Dose.intake.ug converted to units of (mg/kg/day)
abs.dermal.ug	Dermal absorbed dose (ug/day)

Parameter	Description
abs.inhal.ug	Inhalation absorbed dose (ug/day)
abs.ingest.ug	Ingestion absorbed dose (ug/day)
abs.tot.ug	Sum of abs.dermal.ug + abs.inhal.ug + abs.ingest.ug in (ug/day)
abs.tot.mgkg	Abs.tot.ug converted to units of (mg/kg/day)
urine.tot.ug	Chemical excreted in urine (ug/day)
exp.window	Airborne chemical escaping the house (ug/day)

The "all" file can be used to calculate statistics for any of the variables, either across the whole population or across any selected subgroup.

3.4.2 The Statistics File

There is one Statistics file for each chemical, for each set of simulated people. When the number of simulated people becomes large, SHEDS breaks the calculations into "sets" which are small enough to avoid bogging down the computer. The stats file reports statistics on the current set. Apart from the mean, the other statistics cannot be properly computed for the entire run from the collection of set-specific stats. For example, the overall 90th percentile cannot be found from the 90th percentiles on the various sets (except approximately, say by averaging them). The proper way to compute overall stats requires reading the All Individuals file and recalculating the desired statistics.

The Statistics file reports the mean standard deviation, and the following quantiles: 0.005, 0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.975, 0.99, and 0.995. These statistics are reported for the variables in Table 13.

Table 13. Exposure and Dose Metrics Summarized in the Statistics File.

Parameter	Description
exp.dermal	Dermal exposure (ug/day)
exp.ingest	Ingestion exposure (ug/day)
exp.inhal	Inhalation exposure (ug/m3), averaged over the day
dose.inhal	The inhalation intake dose (ug/day)
dose.intake	The sum of exp.ingest + dose.inhal.ug +abs.dermal.ug, converted to (mg/kg/day)
abs.dermal.ug	Dermal absorbed dose (ug/day)
abs.ingest.ug	Ingestion absorbed dose (ug/day)
abs.inhal.ug	Inhalation absorbed dose (ug/day)

Parameter	Description
abs.tot.ug	Sum of abs.dermal.ug + abs.inhal.ug + abs.ingest.ug in (ug/day)
abs.tot.mgkg	Abs.tot.ug converted to units of (mg/kg/day)
ddd.mass	Chemical mass going down the drain (g/day)

3.4.3 The Source Means File

The third type of output file from SHEDS-HT is the "source means" file. This file provides the mean value across the population for various exposure variables, by each source given on the Source-Scenarios file. This file lists all the source types for the chemical being modeled, plus a "total" which is the sum of all sources. The eight variables on this file are:

- exp.dermal Dermal exposure (ug/day)
- exp.ingest Ingestion exposure (ug/day)
- exp.inhal Inhalation exposure (ug/m3), averaged over the day
- dose.inhal The inhalation intake dose (ug/day)
- f.dermal dermal exposure as a fraction of house chemical mass (-)
- f.ingest ingestion exposure as a fraction of house chemical mass (-)
- f.inhal inhaled dose as a fraction of house chemical mass (-)
- mean.mass chemical mass in house (average over persons) (ug/day)

4 SHEDS-HT CODE DESIGN AND ALGORITHMS

SHEDS is a stochastic human exposure model. It randomly samples values for human characteristics such as age and gender, selects activity and diet diaries from databases, and assigns values for product use, frequency, and amount based on user-specified distributions. The goal is to capture the full spectrum of exposures, including high-use days, typical days, and days with little or no exposure, in the correct relative frequencies. SHEDS is cross-sectional in the sense that a single day is modeled for each person, although it is not necessarily the same day as for other persons. SHEDS is designed to be able to process many chemicals in a single run. Test runs have been successful with more than 2400 chemicals run simultaneously.

Every simulated person in SHEDS is independent, meaning that randomly sampled values are assigned to them without any connection to the values assigned to anyone else. Thus, output from runs sharing the same inputs (except the random number seed) can be combined into a larger output. The limitation here is that only person-level output can be so combined, or means across persons. It is not possible to combine summary statistics such as population percentiles.

In practice, SHEDS computes details for many persons simultaneously (this number is the "set.size" parameter on the Run file). In the following discussion, only one person is explicitly considered, with the understanding that the same rules apply to all.

4.1 Steps in a SHEDS Model Run

The main steps in a SHEDS run are:

1) Read the input files
2) Create and populate internal arrays
3) Define the personal demographic and media variables
4) Add the personal exposure factors
5) Set all source-specific variables
6) Start loop over all chemicals
7) Start loop over all sources of that chemical
8) Evaluate exposure and intake dose for each source-scenario combination
9) End loop over sources
10) Evaluate removal pathways
11) Summarize the current chemical and write output
12) End loop over chemicals
13) For additional sets of people, return to step 3.

For a run with multiple sets of persons, steps 3 onwards are repeated for each set. Post-processing of results is not automatic in SHEDS, and requires a separate program.

4.1.1 Read the Input Files

Step (1) in a SHEDS run is implemented by the following code from the "run" function in the main code module:

```
specs        <- read.run.file(run.file)
act.diaries  <- read.act.diaries(specs$act.diary.file,specs)
chem.props   <- read.chem.props(specs$chem.props.file,specs)
specs        <- update.specs(specs,chem.props)
diet.diaries <- read.diet.diaries(specs$diet.diary.file,specs)
exp.factors  <- read.exp.factors(specs$exp.factor.file)
fug.vars     <- read.fug.inputs(specs$fugacity.file)
media        <- read.media.file(specs$media.file)
media.sur    <- tolower(media$media[media$type!="air"])
media.air    <- tolower(media$media[media$type=="air"])
physio       <- read.phys.file(specs$physiology.file)
pop          <- read.pop.file(specs$population.file,specs)
source.scen  <- read.source.scen.file(specs$source.scen.file)
source.chem  <- read.source.chem.file(specs$source.chem.file, source.scen$src,specs)
specs        <- update.specs(specs,source.chem)
source.scen  <- source.scen[source.scen$src %in% source.chem$src]
source.vars  <- read.source.vars.file(specs$source.vars.file,source.scen)
```

SHEDS has a "read" function for each input file, and all are found in the ReadData.R module. These are listed in (near) alphabetical order, both in the calling code (shown here) and in the ReadData module. The information from the RunFile is stored in "specs". In general, each input file becomes an R object, usually a data table.

4.1.2 Create and Populate Internal Arrays

Step (2) above is implemented using the following code from the Run function:

```
act.pools    <- act.diary.pools(act.diaries,specs)
diet.pools   <- diet.diary.pools(diet.diaries,specs)
gen.facs     <- gen.factor.tables(exp.factors)
med.facs     <- med.factor.tables(exp.factors,media.sur)
sv           <- set.pars(source.vars)
scv          <- set.pars(source.chem)
sets         <- ceiling(specs$n.persons/specs$set.size)
```

See Section 4.3 below for details on the activity and dietary diary assignment.

The gen.facs object is a table of indices for each combination of gender, season, and age range for the general exposure factors (those that cannot have any media dependence). SHEDS contains eleven such variables (bath.p, handmouth.area.f, handmouth.freq, handwash.freq, ingest.abs.f, inhal.abs.f, rem.bath.f, rem.brushoff.f, rem.derm.abs.f, rem.handwash.f, and urine.f). Since a standard run has two genders and four seasons, this file contains at least 88 rows (eight per variable), and will have more rows if some variables have multiple distributions for different age ranges. For each combination of variable, age, gender, and season, gen.facs contains a **row**

variable which is the record on the exp.factors file for the appropriate distribution. If, for example, a given distribution applies to both genders and all seasons, then the eight records on gen.facs for that variable and age range will all have the same value for the variable **row** This avoids making eight copies of the same distribution.

The med.facs object is similar to gen.facs, but covers the exposure factors that may be media-specific. Med.facs has eight rows for each combination of variable and age range, for each of the surface media in the run that have contact.p greater than zero. Surface media are ones that have type "Srf" on the Media input file. Med.facs has a column called "row" that indicates which row of the exp.factors data set is to be used for each combination of variable, age range, and medium.

The set.pars function is called twice per run, once to process the source.variables file and once for the source.chemicals file. These files have the same format, but the latter has distributions that are specific to source-chemical combinations, while the former applies to any chemical for a given source. Set.pars alters the distributions for the distributional parameters expected internally by SHEDS, given the ones on the input file.

Set.pars alters the "use.prev" and "home.prev" variables from point values of probability to Bernoulli distributions, sampled once per person. Lognormal distributions are also modified. First, if the coefficient of variation (CV) is zero, then the distribution type is changed to "point" to avoid numerical problems. Second, SHEDS uses the geometric mean (GM) and geometric standard deviation (GSD), whereas the input file reports the arithmetic mean and CV. In SHEDS, CV is the ratio of the standard deviation to the mean (it is not a percentage):

$$CV = \frac{std.dev.}{mean} \tag{1}$$

Where:

$mean$	=	arithmetic mean of distribution
$std.dev.$	=	arithmetical standard deviation

The relationships for a lognormal distribution are:

$$GM = \frac{mean}{\sqrt{1 + CV^2}} \tag{2}$$

$$GSD = e^{\sqrt{\log(1+CV^2)}} \tag{3}$$

Where:

GM	=	geometric mean
CV	=	coefficient of variation

GSD = geometric standard deviation

For the normal distribution, set.pars sets Par2 to be the standard deviation, which is the product of the mean and CV. At present, only the point, Bernoulli, empirical, normal, and lognormal distributions are supported by the set.pars function, which means that the source.variables and source.chemicals input files can only use those functions. By contrast, the exp.factors input file does not report mean and CV, but instead expects the correct parameters used by SHEDS for all distributions.

Step (3) in a SHEDS run defines the demographic and media variables for the set of simulated people. This is achieved using the "select.people" and "add.media" functions.

4.2 Define the Personal Demographic and Media Variables

The process of creating simulated people begins with the "select.people" function. The first steps are to assign gender and age. The population arrays contain the data from the input population file, except that any ages and genders not being modeled have their counts set to zero. The total remaining numbers of males and females are used as relative probabilities for the respective genders. Once gender is determined, age is similarly chosen using the counts for each age within that gender as relative probabilities.

The season is assigned next. There are four possible seasons, winter (W), spring (P), summer (S), and fall (F). Each season represents three calendar months, with winter being December, January, and February. The user chooses which of the four are allowed in the run. All allowed seasons are equally likely to be selected, regardless of the number of diaries belonging to each season. The "weekend" setting determines whether an activity diary from a weekday or a weekend day is selected, with weekend being chosen with probability 2/7.

The physiological variables are then assigned. These are body weight (kg), height (cm), basal metabolic rate (MJ/day), basal ventilation rate (m^3/day), and skin surface area (cm^2). The natural logarithm of body weight is assumed to have a normal distribution with parameters (mean and standard deviation) that are specific to each age/gender combination. These parameters come from the physiology input file. Height has two defining relationships. One is a normal distribution with parameters specific to each age and gender. The second is a linear regression on body weight with slope, intercept, and residual parameters read from the physiology input file. The larger of the two values becomes the assigned height.

Basal metabolic rate (bmr) measures the energy expenditure of the body when resting. It is determined using a regression on body weight. Basal ventilation rate uses a hard-coded relationship with bmr:

$$bva = bmr \times 0.166 \times 19.63 \times 0.001 \times 1440 \times (0.20 + 0.01 \times u) \quad (4)$$

Where:

bva	=	basal ventilation rate [m^3/day]
bmr	=	basal metabolic rate [megajoules/day]
0.166	=	conversion factor from [kcal/min] to [megajoules/day]
19.63	=	ratio of inhaled air volume to consumed oxygen volume [-]
0.001	=	conversion factor [m^3 /liter]
1440	=	conversion factor [min/day]
$(0.20 + 0.01 \times u)$	=	ranges uniformly between 0.20 and 0.21

The uniform distribution u(0.20,0.21) represents the energy conversion factor in units of (liters of oxygen / kcal) (EPA 2012) The factor of 19.63 was derived by Joumard et al. (1981) and is used in other EPA air inhalation models (OAQPS 2012).

Body surface area has three formulas, one for ages <= 5, one for ages 6-19, and one for ages 20 or more. These are hard coded into SHEDS-HT and are based on thte values derived for SHEDS-Multimedia (Glen et al. 2012).

4.2.1 Activity and Dietary Diary Assignment

Each person is assigned an activity diary and a diet (or food) diary. The activity diaries are indexed by age, gender, season, and weekend/weekday, and contain the time duration and the average PAI in each micro. PAI is the time-averaged MET value, where MET is the multiple of the basal metabolic rate specific to each activity. If all the micros were present on the activity diary, the time durations would always sum to 1,440 minutes, which is one full day. Micros may be omitted from this file if no exposure scenarios are active for them.

The time duration in each micro affects the exposure, and the PAI affects the inhaled dose, if inhalation is a relevant pathway for the particular combination of source and chemical under consideration.

The activity diaries are numbered when they are read from the input file. A vector of lists called “act.pools” is created for each combination of gender, weekend, season, and age. There are two gender pools (“M” and “F”), two weekend pools (0=weekday, 1=weekend), four season pools (“W”,”P”, “S”, or “F”), and 100 age pools (0-99). There are 1600 pools in total (2*2*4*100) for all combinations of these variables.

There may not be enough diaries to populate every age category sufficiently, so the age does not need to match exactly. The input setting AgeMatchPct sets the maximum allowed difference between target age and diary age, as a percentage of the target age. For example AgeMatchPct=20 means that a 30 year old simulated person may be assigned a diary with age

ranging from 24 to 36 (since 20% of 30 is 6 years). This also means that each diary belongs to several pools for nearby ages. As an example, the list of acceptable diary numbers for a 35-year-old female on a summer weekend is given by the list act.pools[["F1S35"]] or act.pools$F1S35. The "empirical" distribution in SHEDS selects one item from a list with equal probability for each.

The diet diaries are handled similarly to the activity diaries, except there are no weekend or season index variables. The pools are by gender and age. For example, diet.pools[["M42"]] or diet.pools$M42 return the list of diary numbers appropriate for males age 42.

The "select.people" function combines gender, age, season, weekend, weight, height, bmr, bva, surface area, activity diary id, and diet diary id into an object called "pd" for personal and diary variables.

4.2.2 Generating Media Variables

After the demographic variables have been determined on the "pd" data set, SHEDS adds the media variables. There is one such variable for each potential exposure medium in each micro that has contact.p>0, as listed on the "media" input file. That variable is the contact time, which is the time duration in that micro from the activity diary, multiplied by contact.p. Contact.p represents the likelihood that the person is in contact with the medium while in that micro. For indoor air, contact.p should be set to one because one is always breathing the same air. Outdoors, this might not be the case, as one would have to be sufficiently close to a source to be "breathing" it. For surfaces, the user may decide how to separate the factors contact.p and dermal.tc, as the latter would typically account for some non-contact time, at least implicitly.

Each of these media variables has the prefix "dur." followed by the name of the medium. The default media input file produces dur.in.awkair, dur.in.awksrf, and dur.in.slpair. These are added to the "pd" object to produce the "pdm" object.

4.3 Assigning Exposure Factors

Step (4) in the SHEDS algorithm is the evaluation of the exposure factors that are common to all chemicals and sources, but specific to each simulated person. The variables are listed and described in the earlier section on the exp.factors input file. This is one of the more complicated parts of SHEDS, as each variable may optionally have multiple distributions, depending on age, gender, season, and/or exposure media. Age ranges are specified using the minimum and maximum ages covered by the given distribution. For example, handmouth.area.f is age-specific on the current default input file. The code allows any of the exposure factor variables to depend on age, gender, or season. However, the code currently limits the media-specific variables to just avail.f, dermal.tc, and om.ratio. The default exposure factors file does not actually provide

media-specific distributions for any of these three, so in each case the same distribution is used for all media.

The complicated part of the evaluation is that no explicit loop over persons is used, yet different distributions apply to different people. This is achieved by the functions add.factors and eval.factors.

The add.factors function sets up and calls the eval.factors function. For efficiency, the factors are evaluated for all persons in the set. There are two loops, one over the media-specific factors listed in med.facs, and the other over the general factors in gen.facs. In each loop, first the "pdm" data set containing demographic variables is joined with the records from med.facs (or gen.facs) pertaining to one modeling variable, with the join performed by gender and season. Only one of eight records on med.facs (or gen.facs) is selected for each person. If there are multiple age ranges defined for that variable, then the joined data set has multiple records for each person (one for each age range). The next line of code selects the records for which each person's age is between the min.age and max.age settings. If the input file has no overlapping age ranges (which it should not), then one record is selected for each person. This data set "p" now contains a "row" variable which indicates which row of the exp.factors file applies to each person, for this variable.

The eval.factors function actually selects the random values from the appropriate distributions. The "row" variable from the "p" data set, the vector of uniform random samples "q" for the given variable, and the exp.factor data set are supplied to eval.factors, which then loops over the records on the exp.factors data set. The persons whose "row" variable corresponds to the current record have their "q" value transformed. On each loop over the exp.factors records, the persons who do not have "row" equal to the current record number are skipped through subsetting both the "q" and "z" vectors. When all exp.factors rows have been processed, the "z" vector should contain a result for each person, and is returned to the add.factors function.

After evaluating all the exp.factor variables, add.factors sets the number of hand washes and the bath variable for each person. The hand washes starts with the handwash.freq variable, which is one of the ones on exp.factors. This is taken to be the long-term average for that particular person. SHEDS assumes that each person varies the number of hand washes from day to day, and the distribution for this is uniform with a range up to the square root of handwash.freq on either side of the mean. Thus, if a person has handwash.freq=4, then the number on the given day may range from 2 to 6. Or if the mean is 9, the range is 6 to 12. The numbers do not have to be integers, these examples were chosen for convenience. The uniform range is sampled and the result is then rounded probabilistically to an integer. This type of rounding does not change the expected value, because it may either round up or down. For example, if the sampled value was 4.25, it is rounded up 25% of the time to 5, and down 75% of the time to 4. The result is always an integer, and in this example its expected value is (5*0.25+4*0.75) = 4.25. The bath variable is a logical variable, either true or false. If bath.mins from the activity diary is not zero, then bath is true. Also, if the Bernoulli variable bath.p evaluated to 1 for the given person, then bath is true. If

both are zero then bath is false. In SHEDS, “bath” includes showers – essentially any full-body washing that will remove most of the dermal loading.

4.4 Generating Source-Specific Variables

Step (5) is the evaluation of the source-specific variables. A “source” is a product type, for example toothpaste. Certain variables such as the prevalence of users in the population, the frequency of use, and the mass used, will be common to all chemicals found in that source type. Effectively, source variables create correlation between chemical exposures. Thus, a person with a high usage for a given source will tend to have high exposure to all chemicals found in that source. This is important when summing exposures across chemicals belonging to the same chemical class, at the level of individual persons. Earlier versions of SHEDS treated each chemical as independent, so the source variables were re-evaluated for each one.

On the input file, the source variables are effectively in list format. Each source has distributions for the variables it requires. Different exposure scenarios require different variables. As a practical matter, SHEDS uses a three-dimensional array called “srcdata” to store the values of these variables: one index for the source type, another for the simulated person, and a third for the specific modeling variable. Sources not used by any of the chemicals modeled in the run are dropped. For sources that are kept, all possible modeling variables have array elements set aside for them, whether they will be needed for that source type, or not. All elements of srcdata are set to zero initially, so the unused elements simply remain at zero.

The following FOR loop evaluates the elements of srcdata:

```
for (s in 1:n.sdist) {
  q <- runif(n.per)
  srcdata[vpos(sv[s]$src,src.list),,vpos(sv[s]$varname,svar.list)] <-
    distrib(sv[s]$form,sv[s]$par1,sv[s]$par2,sv[s]$par3,sv[s]$par4,
      sv[s]$lower.trun,sv[s]$upper.trun,'y',n=n.per,q=q,v=sv[s]$values)
}
```

Here n.sdist is the total number of distributions defined on the source.variables file. These distributions are stored in the “sv” object. The loop over these uses the index “s” to keep track of the current position. On each loop the vector “q” is loaded with n.per uniform random samples, where n.per is the number of persons in the set. The “vpos” function locates a particular name in a list, and is used twice here. The first use locates the source corresponding to distribution “s” (that is, sv[s]$src) in the list of sources “src.list”. For example, row 59 of the source.variables file (s=59) may correspond to source number 7. The “s” index is likely to increase much faster than the source index because there are multiple distributions for each source. The second use of vpos locates the variable type in the list of possible source variables. These uses of vpos mean that the source.variables file does not need to be in any particular order, and the elements of srcdata will be filled as the distributions are encountered. Incidentally, the two commas between the vpos

functions indicate that the second array index takes on all possible values. This is the index for the person number.

The “distrib” function is found in the Utility module. It will either return random samples from a distribution (if “q” is not supplied as an input argument), or transform the supplied “q” vector using the given distributional parameters. Only one value may be supplied for each of the distributional parameters (that is, for form, par1-par4, lower.trun, upper.trun, and resample). Here resample is hard-coded to “yes”, although it would be easy to allow this to be read from the input file. The argument “n” is the number of samples desired, and “q” is the vector of quantiles to be evaluated. The “v” argument is only used for empirical distributions. In such cases, the other parameters are not used and “v” contains a list of the possible values, each of which is equally likely to be chosen (when “q” is not specified”). In the above usage, distrib returns n.per values for a single distribution (that is, one source and one modeling variable) on each call. The above loop occupies about 1/3 of the total execution time for a SHEDS run which models just one chemical. In a multichemical run, the above loop is still evaluated just once, while each chemical must repeat the set of exposure calculations which follow, so the time for the source variable evaluations becomes insignificant.

4.5 Loop over Chemicals

Step (6) in the SHEDS algorithm is a loop over all the chemicals being considered in this run. If no chemicals were listed on the run file, this is the entire set of chemicals found on both the source.chemicals file and the chem.props file. This can be thousands of chemicals, and takes a long time to run (several hours, at least). If any chemicals are listed on the Run file, then just those chemicals are run (as long as they are also on the other two files). As mentioned earlier, SHEDS uses the CAS numbers with underscores in place of hyphens as the chemical name. This is also incorporated into the filenames for chemical-specific output files.

The chemicals are modeled independently of each other with two exceptions: the same set of persons is used for all chemicals, and the same set of source variables is used. For these reasons, exposures to different chemicals may be correlated across persons.

SHEDS effectively considers each chemical alone, in that no chemical interactions are considered. However, the demographic data and product use data are common across chemicals, so the results are likely to exhibit correlation in exposure across the set of simulated people. For example, if two chemicals are commonly found together in the same products, then the people with high exposure to one will likely have high exposure to the other.

For each chemical, SHEDS identifies the sources (that is, products, articles, and foods) that contain that chemical, and loops over them. For each source, the active exposure scenarios as indicated on the source-scenario file are evaluated. For a given chemical, there may be anywhere from just one to more than one hundred sources. Once all the sources have been

evaluated for a given chemical, post-processing of exposure occurs, along with two types of summarization. If the run file indicates that person.output=1, then a sum over sources, by person, is written to the output folder. If source.output=1, then a mean of persons is calculated for each source.

For each scenario switch, if it is active then a specialized exposure routine is called, and the results are accumulated in the srcMeans and cb data sets. Details for each exposure scenario are given in the next chapter.

4.6 Exposure Scenario Algorithms

There are several possible exposure scenarios in SHEDS, and it is not difficult to add more. The currently supported exposure scenarios are: Food.residue, Food.migration, Product.direct.dermal, Product.direct.ingestion, Product.direct.inhalationAerosol, Product.direct.inhalationVapor, Product.downthedrain, Product.indirect, and article.emission. Each of these is described in turn.

4.6.1 Exposure Scenario "Food.residue"

The food.residue scenario covers ingestion exposure (that is, eating and drinking) to chemicals left on the food as residue. For example, produce may have some residual pesticide or other chemical that is not completely washed off. SHEDS divides foods (including drinks) into multiple user-defined categories. There are two input files that are used by this scenario, and the same food category definitions must be used on each. The diet.diary file reports daily food consumption by food category for a large number of individuals, each labeled by age and gender cohort. For all categories the units are grams per day. The second input file is the source.chemicals file, which (among other things) reports the distribution of chemical concentration in each food category, for each chemical. See the section on input file format for details of these files, including the default food categories.

The relevant variables for the food residue scenario are the following:

Source.variables:
    none
Source.chemicals:

detects	Number of samples with this chemical present
nondetects	Number of samples with chemical not detected
residue	Non-zero detected chemical concentrations [μg/g]

As an example of the data on the source.chemicals file, consider the following:

FOOD; CF: CITRUS FRUITS , detects , 10605-21-7 , - , Point , 5 , .
FOOD; CF: CITRUS FRUITS , nondetects , 10605-21-7 , - , Point , 0 , .

FOOD; CF: CITRUS FRUITS , residue , 10605-21-7 , μg/g , Lognormal , 0.0172 , 0.517931569

These records do not need to appear consecutively. The first five characters "FOOD;" on each line indicate that this record is for a food type. The next set of letters (before the colon) is the label for that type which should match a column header on the food consumption file. In this case it is "CF". The description "CITRUS FRUITS" is not relevant to SHEDS. The second item on each row (after the first comma) indicates the variable. Each food item should have three variables: detects, nondetects, and residue. Detects and nondetects should be point values, their relative sizes determine the likelihood of finding the chemical in that food type. In the above example, there are no nondetects, which means that likelihood is 100%. It is calculated as

$$likelihood = \frac{detects}{detects + nondetects} \quad (5)$$

The "detects" should be positive. If it is zero, there is no chance of exposure and that combination of food type and chemical should be dropped from the file. Following the variable type is the chemical CAS number. All chemicals in the run have their data on the same input file. Here the CAS numbers use hyphens, but in SHEDS these are replaced by underscores. The next item on each record is the units. This is not used by SHEDS. Detects and nondetects are just counts, but the residue has units of micrograms per gram (which is the same as parts per million by weight). Warning: if the user enters data using different units here, SHEDS will not notice and will assume it is still micrograms per gram.

The last three fields on each line are the distribution type, the mean, and the CV (coefficient of variation), which is the ratio of the standard deviation to the mean. In this example, the type is lognormal, but the reported value 0.0172 is still the arithmetic mean of the data, not the geometric mean or the mean of the underlying normal distribution. For point values the CV may either be zero, blank, or a dot, which is a placeholder for an unused numeric variable.

The residue indicates the distribution of chemical concentrations among the "detects". The current version of SHEDS assumes no chemical is present in the nondetects. If there is a nonzero value for CF on the selected food consumption diary, SHEDS generates a uniform random sample "q" and compares it to the likelihood. If the "q" value is larger than the likelihood, that food type is considered to be uncontaminated with that chemical. If the random sample is lower than the likelihood, the residue is sampled and the dietary exposure in [μg] is

$$exp.diet = consumption \times residue \quad (6)$$

Where:

$exp.diet$	=	dietary exposure from food residues [μg/day]
$consumption$	=	amount of this food type consumed [g/day]
$residue$	=	chemical residue in food [μg/g]

SHEDS could either loop over all the food types on the selected consumption diary, or loop over all the food types for that chemical on the source.chemicals file. The latter is chosen for two reasons. First, for most chemicals there are fewer foods containing that chemical than the number of types on a food diary (so the loop is smaller). Second (and most important), for a given chemical the source.chemicals list is fixed, but the food types on the consumption diaries may differ by person. SHEDS processes all the persons in the set in a single pass through the code, so the loop must consider all the possible sources for any person, not just the sources that apply to a given person.

4.6.2 Exposure Scenario "Food.migration"

This is similar to food.residue, except that the chemical originates in food packaging and migrates into the food before being consumed. This version of SHEDS supports a simple calculation which requires a "migration concentration" in units of [μg/g], to be specified for each type of food. In effect, this is handled by SHEDS just like a "residue concentration". The difference is that while the food.residue pathway uses detects and nondetects, the migration pathway uses chem.prev. The user may set this using the same likelihood calculation as for food.residue.

Source.variables:
none

Source.chemicals:

chem.prev	Fraction of food samples with this chemical present
migration.conc	Non-zero detected chemical concentrations [μg/g]
contact	Currently set to 1.0. Intended for future use.

The exposure for this scenario is

$$exp.diet = consumption \times migration.conc \times contact \quad (7)$$

Where:

$exp.diet$	=	dietary exposure from food residues [μg/day]
$consumption$	=	amount of this food type consumed [g/day]
$migration.conc$	=	amount of chemical that has migrated from the food packaging [μg/g]
$contact$	=	food-packaging contact factor (should be set to 1) [-]

It is anticipated that SHEDS in the future may support an internal calculation of chemical flux from packaging into food, so SHEDS would determine the "migration concentration" from other input variables.

4.6.3 Exposure Scenario "Product.direct.dermal"

This is the direct dermal contact exposure scenario, meaning contact with a product during its use. This is distinct from indirect dermal contact, which occurs after use has concluded. Dermal exposure from permanent articles such as furniture and carpets is also not included.

The relevant variables for the Product.direct.dermal scenario are the following:

Source.variables:

use.prev	Fraction of population using this source
use.freq	Number of times per year this source is used
mass	Mass in grams of product per use
f.contact	Fraction of product mass contacting skin
f.residual	Fraction of contacting mass remaining on skin after product use

Source.chemicals:

chem.prev	Likelihood this source type contains the chemical of interest
f.chemical	Fraction of product mass that is this chemical (if not zero)

All of the above variables are defined as distributions. Use.prev can be defined as a Bernoulli, binomial, or point distribution; in all cases the Par1 value is the fraction of the population that "uses" (that is, exposed to) this source at least once per year. Use.prev may be made age and/or gender dependent, if the modeler so chooses. After evaluating the use.prev distribution, each person has a value of 1 (user) or 0 (non-user) for this variable.

Use.freq is the number of times per year the source is used. For some it may be only once (per year), and for others it may be a thousand or more (for things like soap). SHEDS determines a number of uses on the simulated day using the probabilistic rounding function (called p.round in SHEDS). This either rounds up or down to the nearest whole number, using the (1-distance) as the statistical weight. The key feature of p.round is that the mean number of events over a large number of persons is unchanged by rounding, regardless of the initial value. In SHEDS the number of uses on the given day is

$$\#\ uses\ on\ given\ day = p.round\left(\frac{use.freq}{365}\right) \quad (8)$$

Use.freq itself does not have to be an integer – it can be sampled from a continuous distribution. If use.freq<365, then some people will not use that source on the given exposure modeling day, even though that person is designated as a "user" by the use.prev test. If use.freq=365, then every person passing the use.prev test will use that source. If use.freq>365, everyone passing the

use.prev test will use the source at least once, and some will use it more than once on that day. Multiple uses add their exposures together.

Mass is the product mass (in grams) released during each use. Chem.prev is similar to use.prev, it is the likelihood that the source contains the given chemical, and may be defined as a Bernoulli, binomial, or point distribution, in all cases with Par1 being the likelihood. For each person, chem.prev evaluates to either 1 (chemical present in source) or 0 (chemical absent). F.chemical measures the fraction of product mass that is the chemical of interest, in the sources with chem.prev=1. For example, if f.chemical evaluates to 0.15, then 15% of the product mass is the given chemical.

The remaining two variables f.contact and f.residual can be used in more than one way. F.contact is intended to be the fraction of the product mass than contacts the skin. For example, when painting, only a small fraction of the paint will get on the skin, so only that fraction of the chemical mass in the paint counts as dermal exposure. F.residual represents the amount that remains on the skin for a substantial time after product use. For products like soap, f.contact may be set to one, but f.residual is small since most of it is rinsed off. Dermal absorption is a slow process (generally with a much longer time scale than product use), so a chemical must remain on the skin after use has concluded or else the five removal processes do not have time to take effect.

The dermal exposure for the Product.direct.dermal scenario is given by

$$exp.derm = use.p \times p.round\left(\frac{use.f}{365}\right) \times mass \times chem.p \times f.chem \times f.con \times f.res \times 1E6 \quad (9)$$

Where:

$exp.derm$	=	direct dermal exposure [µg/day]
$use.p$	=	fraction of population who are users [-]
$use.f$	=	use frequency [#/yr]
365	=	conversion factor [days/yr]
$mass$	=	product mass released per use [g]
$chem.p$	=	likelihood of chemical being in this product [-]
$f.chem$	=	ratio of chemical mass to product mass (if not zero)[-]
$f.con$	=	fraction of product contacting skin [-]
$f.res$	=	residual fraction remaining on skin after use [-]
$1E6$	=	conversion from [g] to [µg]

4.6.4 Exposure Scenario "Product.direct.ingestion"

This is the direct ingestion scenario during product use, excluding food and drinks. This scenario is nonzero for very few household sources, most examples are related to the mouth, as with toothpaste, mouthwash, and lip balm, and the ingestion is accidental. The relevant variables are:

Source.variables:

use.prev	Fraction of population using this source
use.freq	Number of times per year this source is used
mass	Mass in grams of product per use
f.ingested	Fraction of product mass that is ingested (provided on chemical property input file).

Source.chemicals:

chem.prev	Likelihood this source type contains the chemical of interest
f.chemical	Fraction of product mass that is this chemical (if not zero)

The only variable not also used for the product.direct.dermal scenario is f.ingested, which represents the fraction that is (generally inadvertently) swallowed. The exposure is given by

$$exp.ingest = use.p \times p.round\left(\frac{use.f}{365}\right) \times mass \times chem.p \times f.chem \times f.ingest \times 1E6 \quad (10)$$

Where:

$exp.ingest$	=	direct dermal exposure [µg/day]
$use.p$	=	fraction of population who are users [-]
$use.f$	=	use frequency [#/yr]
365	=	conversion factor [days/yr]
$mass$	=	product mass released per use [g]
$chem.p$	=	likelihood of chemical being in this product [-]
$f.chem$	=	ratio of chemical mass to product mass (if not zero)[-]
$f.ingest$	=	fraction of product mass ingested [-]
$1E6$	=	conversion from [g] to [µg]

4.6.5 Exposure Scenario "Product.direct.inhalationAerosol"

This is one of two direct inhalation scenarios, used for aerosols. The other one is Product.direct.inhalationVapor, used for inhalation of vapor. Product.direct.inhalationAerosol is used for chemicals with low vapor pressure that are generally bound to airborne particles or liquid droplets. The relevant variables are

Source.variables:

use.prev	Fraction of population using this source
use.freq	Number of times per year this source is used
mass	Mass in grams of product per use
f.aerosol	Fraction of product mass remaining in aerosol form [-]
duration	Duration of product use [min]
volume	Volume into which the aerosol disperses during use [m^3]

Source.chemicals:

chem.prev	Likelihood this source type contains the chemical of interest
f.chemical	Fraction of product mass that is this chemical (if not zero)

SHEDS calculates an air concentration (for users) as follows:

$$air.conc = mass \times chem.prev \times f.chemical * f.aerosol/volume \quad (11)$$

Where:

$air.conc$	=	chemical concentration as airborne aerosol [g/m^3]
$mass$	=	product mass released per use [g]
$chem.prev$	=	likelihood of chemical being in this product [-]
$f.chemical$	=	ratio of chemical mass to product mass (if not zero) [-]
$f.aerosol$	=	fraction remaining in aerosol form unitless]
$volume$	=	volume into which the aerosol disperses during use [m^3]

Two limitations are imposed. First, the volume cannot be less than 24 m^3, which is the size of a typical room. The reason is that while the aerosol may initially be sprayed into a smaller volume, it will quickly mix with the rest of the air in the room. For very short durations (one minute or less) this might not be a good assumption, but product use durations typically range from 5 to 30 minutes, or more. Within a room the air is well mixed within five minutes. For outdoor products this minimum volume is 480 m^3. The second limitation is that air.conc cannot exceed 1.2 g/m^3, which is 0.1% of air density. Exposure to air concentrations above this for a prolonged period may comprise basic breathing and oxygen uptake, so is not reasonable. If exceptions are identified, these limitations may be modified.

Once air.conc has been determined, the exposure from the Product.direct.inhalationAerosol scenario is

$$exp.aerosol = use.prev \times p.round\left(\frac{use.freq}{365}\right) \times air.conc \times duration \times \frac{1E6}{1440} \quad (12)$$

Where:

$exp.aerosol$	=	inhalation exposure from aerosols [μg/day]
$use.prev$	=	fraction of population who are users [-]
$use.freq$	=	use frequency [#/yr]
365	=	conversion factor [days/yr]
$air.conc$	=	defined in previous equation [g/m3]
$duration$	=	duration of product use [min]
1440	=	conversion factor [min/day]
$1E6$	=	conversion factor [μg/g]

In SHEDS, inhalation exposure is defined as the average air concentration encountered over the simulation period, which is one day. Hence, the air concentration "air.conc" is multiplied by the number of uses and by the time fraction of usage, which is (duration/1440).

SHEDS also calculates an inhaled dose, which is the mass of chemical entering the lungs. It is given by the product of the exposure and the breathing ventilation rate in [m^3/day]:

$$dose.aerosol = exp.aerosol \times bva \times pai \quad (13)$$

Where:

$dose.aerosol$	=	inhalation applied dose from aerosols [μg/day]
$exp.aerosol$	=	inhalation exposure from aerosols [μg/m^3]
bva	=	basal ventilation rate [m^3/day]
pai	=	micro-specific average multiple of basal rate [-]

Here bva is the basal ventilation rate assigned to each person, based on their age, gender, and body weight. The "pai" factor represents the average MET value over the relevant activities. A possible enhancement to SHEDS would be to incorporate source-specific (as opposed to micro-specific) MET values on the input files.

4.6.6 Exposure Scenario "Product.direct.inhalationVapor"

This is the second of the two direct inhalation scenarios. It is meant for chemicals with relatively high vapor pressures. The relevant variables are

Source.variables:

use.prev	Fraction of population using this source
use.freq	Number of times per year this source is used
mass	Mass in grams of product per use
duration	Duration of product use [min]

volume	Volume into which the aerosol disperses during use [m^3]
Source.chemicals:	
chem.prev	Likelihood this source type contains the chemical of interest
f.chemical	Fraction of product mass that is this chemical (if not zero)

Also, vapor pressure in Pascals and the molecular weight in grams per mole are required. These are automatically extracted from the chemical properties input file, so the user does not need to specify them on the source files.

SHEDS calculates an evaporation fraction "vapfrac" which is set equal to the ratio of chemical vapor pressure to air pressure, but capped at one. An "evaporation mass" is determines as

$$air.conc = mass \times chem.prev \times f.chemical \times vapfrac/volume \quad (14)$$

Where:

$air.conc$	=	chemical concentration in air [g/m^3]
$mass$	=	product mass released per use [g]
$chem.prev$	=	likelihood of chemical being in this product [-]
$f.chemical$	=	ratio of chemical mass to product mass (if not zero) [-]
$vapfrac$	=	ratio of vapor pressures [-]
$volume$	=	volume into which the aerosol disperses during use [m^3]

The volume into which this mass evaporates is randomly sampled from the "volume" distribution from the source.variables file, with a minimum of 24 cubic meters for indoor products (a typical room volume), or 480 cubic meters for outdoor products. The air.conc is capped at 0.1% of air density, or 1.2 g/m^3. The exposure is

$$exp.vapor = use.prev \times p.round\left(\frac{use.freq}{365}\right) \times air.conc \times duration \times \frac{1E6}{1440} \quad (15)$$

Where:

$exp.vapor$	=	inhalation exposure from vapor [µg/day]
$use.prev$	=	fraction of population who are users [-]
$use.freq$	=	use frequency [#/yr]
365	=	conversion factor [days/yr]
$air.conc$	=	defined in previous equation [g/m^3]
$duration$	=	duration of product use [min]
1440	=	conversion factor [min/day]

$1E6$ = conversion factor [µg/g]

The inhaled dose is

$$dose.vapor = exp.vapor \times bva \times pai \quad (16)$$

Where:
$dose.vapor$ = inhalation applied dose from vapor [µg/day]
$exp.vapor$ = inhalation exposure from vapor [µg/m^3]
bva = basal ventilation rate [m^3/day]
pai = micro-specific average multiple of basal rate [-]

4.6.7 Exposure Scenario "Product.downthedrain"

This scenario is unique in SHEDS in that no human exposure is determined. Instead, the amount of chemical being washed into the sewer system is the focus of this scenario. The relevant variables are

Source.variables:
- use.prev — Fraction of population using this source
- use.freq — Number of times per year this source is used
- mass — Mass in grams of product per use
- f.drain — Fraction of product mass going into waste system

Source.chemicals:
- chem.prev — Likelihood this source type contains the chemical of interest
- f.chemical — Fraction of product mass that is this chemical (if not zero)

The chemical mass going down the drain is

$$ddd.mass = use.prev \times p.round\left(\frac{use.freq}{365}\right) \times chem.prev \times f.chem \times mass \times f.drain \quad (17)$$

Where:
$ddd.mass$ = chemical mass "down the drain" [g/day]
$use.prev$ = fraction of population who are users [-]

$use.freq$	=	use frequency [#/yr]
365	=	conversion factor [days/yr]
$chem.prev$	=	likelihood that this product contains this chemical [-]
$f.chem$	=	ratio of chemical mass to product mass (if not zero) [-]
$mass$	=	mass of product released per use [g]
$f.drain$	=	fraction of mass entering waste system [-]

4.6.8 Exposure Scenario "Product.indirect"

This is one of two indirect exposure scenarios in SHEDS. Product.indirect is designed for products with a definite time of usage when the chemical is released. A fugacity-based model is incorporated which calculates chemical flow rates. The concept is that once chemical is released, it will flow into all air and surface compartments to some extent, before dissipating. Hence all family members may receive exposure long after product use, even in locations not directly treated. The required source variables are:

Source.variables:

home.prev	Fraction houses in which this source is used
use.freq	Number of times per year this source is used
mass	Mass in grams of product per use

Source.chemicals:

chem.prev	Likelihood this source type contains the chemical of interest
f.chemical	Fraction of product mass that is this chemical (if not zero)

The fugacity model is a reduced version of the SHEDS-Fugacity model. The full model has ten compartments, but in SHEDS-HT only the air and surface compartments remain. A version with both treated and untreated surfaces has been tested, but while the chemical concentrations can be very different, there is no information on which surfaces the simulated people will contact. With a single day-long time step, both types of surfaces will be contacted on a single step, so an average surface concentration would have to be determined anyway. It is simpler to model a single average surface in the first place.

The details of the fugacity model are provided in another chapter. Here, the emphasis is on how the input variables from the source.variables file are used. The prevalence is determined by the "home.prev" variable. If this is not defined on source.variables, then "use.prev" is used instead. This allows backward compatibility with older input files created before home.prev was added to SHEDS. The applied chemical mass is

$$app.mass = home.prev \times mass \times chem.prev \times f.chemical \quad (18)$$

Where:

$app.mass$	=	chemical mass applied at time of last use [g]
$home.prev$	=	fraction of houses in which this product is used [-]
$mass$	=	mass of product released per use [g]
$chem.prev$	=	likelihood that this product contains this chemical [-]
$f.chemical$	=	ratio of chemical mass to product mass (if not zero) [-]

Initially this mass is distributed evenly over the entire floor surface area

$$app.rate.sur = app.mass \,/\, floor.area \tag{19}$$

Where:

$app.rate.sur$	=	chemical loading on surfaces [g/m^2]
$app.mass$	=	chemical mass applied at time of last use [g]
$floor.area$	=	floor area of house [m^2]

A small amount of chemical is initially placed in the air compartment

$$app.rate.air = app.rate.sur \tag{20}$$

Where:

$app.rate.air$	=	chemical concentration in air [µg/m^3]
$app.rate.sur$	=	chemical loading on surfaces [g/m^2]

For a room that is 3 meters tall, the mass in air is only 3 millionths of the mass on the floor, because of the difference in units. Having a small amount in the air avoids some numerical problems. The amount specified here is arbitrary, but if this initial distribution is not correct, it should quickly redistribute to the proper value due to the fugacity imbalance.

The use.freq variable is not used in the way it is for the direct scenarios, because it does not matter whether the simulated day is a usage day, or not. In the indirect scenarios, if the product is used at all, then some chemical will remain on all subsequent days. The important variable is the number of days since the product was last used, since this affects the amount of time the chemical has to disperse, decay, and be removed from the house. This is sampled as

$$time = uniform * 365/use.freq \tag{21}$$

Where:

$time$ = time since last product use [days]
$uniform$ = uniform random sample between 0 and 1 [-]
365 = conversion factor [days/yr]
$use.freq$ = frequency of product use [#/yr]

Here, (365/use.freq) is the average interval between product uses. While these events might not be evenly spaced, the above equation provides variability while capturing the correct frequency. Since the simulated day is chosen at random, the time delay can be anywhere from zero to the full interval after the last use.

The fugacity model uses the initial air and floor concentrations, the house and chemical variables, and the time delay (dates) to determine the final air and surface concentrations. See the section on The Fugacity-Based Indoor Fate and Transport Model for details. Once the air and surface concentrations at the correct time have been found, the exposure is

$$exp.inhal = air.conc(time) \times duration/1440 \quad (22)$$

Where:
$exp.inhal$ = inhalation exposure concentration [ug/m^3]
$air.conc(time)$ = fugacity-based air concentration at "time" days after product use [ug/m^3]
$duration$ = time spent in home on selected day [min]
1440 = conversion factor [min/day]

The inhalation dose is given by

$$dose.inhal = exp.inhal \times bva \times pai \quad (23)$$

Where:
$dose.inhal$ = inhalation dose [µg/day]
$exp.inhal$ = inhalation exposure concentration [µg/m^3]
bva = basal ventilation rate [m^3/day]
pai = physical activity index [-]

Both indirect scenarios allow for dermal and ingestion exposure, as well. The dermal exposure uses the transfer coefficient method to determine contact with surfaces:

$$exp.dermal = sur.conc \times TC \times avail.f \times duration/60 \quad (24)$$

Where:

$exp.dermal$	=	indirect dermal exposure from product use [µg/day]
$sur.conc$	=	average chemical loading on surfaces [µg/m^2]
TC	=	dermal transfer coefficient [m^2/hr]
$avail.f$	=	fraction of chemical available for transfer [-]
$duration$	=	time at home while awake [min]
60	=	conversion factor [min/hr]

Finally, ingestion of chemical from mouthing toys and other objects on contaminated surfaces may be included by specifying a non-zero value for the "om.ratio" variable on the exp.factors file. If this option is used, the indirect ingestion exposure is

$$exp.ingest = exp.dermal \times om.ratio \quad (25)$$

Where:

$exp.ingest$	=	indirect ingestion exposure from product use [µg/day]
$exp.dermal$	=	indirect dermal exposure from product use [µg/day]
$om.ratio$	=	object-mouthing ratio [-]

Om.ratio may vary by surface type (it is one of the media-specific exposure factors), or it may be set to a single distribution that applies to all surface media. In the latter case, variables for each surface are created automatically in SHEDS.

4.6.9 Exposure Scenario "Article.emission"

The article.emission exposure scenario applies to articles, which are chemical-containing items that exist permanently in the home. Typical examples are furniture and carpets. While these may occasionally wear out and be replaced, it is reasonable to assume that the replacement will contain similar chemicals.

Since the articles exist indefinitely, time is not a factor. The chemical concentrations and flows will reach equilibrium, in which new chemical outflow from the articles replaces any losses in the air and surfaces in the home.

The required variables on the source files are:

Source.variables:

home.prev	Fraction houses in which this source is used
f.area	Ratio of article surface area to house floor area

Source.chemicals:

chem.prev	Likelihood this source type contains the chemical of interest
y0	Equilibrium clean air concentration near article surface

The variable "y0" effectively measures the rate of chemical flux from the article, and depends on the type of material in the article and the chemical properties. The article calculations use y0 because it is relatively simple to measure on a variety of articles without destructive analysis.

The other variables needed for the y0 scenario are found on the fugacity input file, the chemical properties file, or are derived from those variables. These include $aer.out$, $vol.air$, $Kp.sm$, *sm.load.air, Kp.lg, lg.load.air, h.y0, µgmol, z.air, z.floor, area.sur, thick.sur,* and *z.sur*.

The first step is to calculate a particulate partition coefficient as

$$particles = Kp.sm \times sm.load.air + Kp.\lg \times lg.load.air \quad (26)$$

Where:

$particles$	=	mass ratio of chemical on airborne particles to vapor form [-]
$Kp.sm$	=	the chemical carrying capacity for small particles [$m^3/\mu g$]
$sm.load.air$	=	loading of small particles in air [$\mu g/m^3$]
$Kp.\lg$	=	the chemical carrying capacity for large particles [$m^3/\mu g$]
$lg.load.air$	=	loading of large particles in air [$\mu g/m^3$]

The carrying capacities Kp. sm and Kp. lg are calculated from the chemical properties as

$$Kp.sm = 1.6625\text{E-}12 \times Kow \times Rgas \times T \times sm.carb.f \times solub/vapor \quad (27)$$

$$Kp.lg = 1.6625\text{E-}12 \times Kow \times Rgas \times T \times lg.carb.f \times solub/vapor \quad (28)$$

Where:

$Kp.sm$	=	the chemical carrying capacity for small particles [$m^3/\mu g$]
$Kp.lg$	=	the chemical carrying capacity for large particles [$m^3/\mu g$]
1.6625E-12	=	regression constant [$m^3/\mu g$]
Kow	=	octanol-water partition coefficient [-]
$Rgas$	=	universal gas constant = 8.314 [Pa m^3 / (mol K)]
T	=	indoor temperature [K]
$sm.carb.f$	=	organic carbon fraction (by mass) for small particles [-]
$lg.carb.f$	=	organic carbon fraction (by mass) for large particles [-]
$vapor$	=	vapor pressure of chemical [Pa]
$solub$	=	solubility of chemical [mol/m^3]

The modified loss rate $Qstar$ is defined as

$$Qstar = aer.out \times vol.air \times (1 + particles)/24 \quad (29)$$

Where:

$Qstar$	=	modified air exchange rate [m^3/hr]
$aer.out$	=	air exchange rate with outdoors [#/day]
$vol.air$	=	air volume of house [m^3]
$particles$	=	mass ratio of chemical on airborne particles to vapor form [-]
24	=	conversion factor [hr/day]

Next, find the average gas-phase concentration in the house, given by

$$y = y0 \,/\left(1 + \frac{Qstar}{(h.y0 \times f.area \times a.floor)}\right) \quad (30)$$

Where:

y	=	average vapor phase air concentration in house [µg/m^3]
$y0$	=	vapor phase air concentration near article surface [µg/m^3]
$Qstar$	=	modified air exchange rate with outdoors [m^3/hr]
$h.y0$	=	diffusivity in air [m/hr]
$f.area$	=	ratio of article surface area to floor area [-]
$a.floor$	=	floor area of the house [m^2]

For large articles, or if Qstar were small enough, y would be close to y0. For small articles or large Qstar, the average house concentration y is much less than y0.

The total air concentration air.conc is the sum of y plus the particle-bound air concentration. The latter is assumed to be at the same fugacity as the gas-phase concentration, so the total becomes

$$conc.air = y \times (1 + particles) \quad (31)$$

Where:

$conc.air$	=	chemical concentration in air [µg/m^3]
y	=	vapor phase concentration [µg/m^3]
$particles$	=	mass ratio of chemical on airborne particles to vapor form [-]

At the same fugacity, the concentration on the surfaces is given by

$$conc.sur = \frac{conc.air \times z.sur \times thick.sur}{z.air} \quad (32)$$

Where:

$conc.sur$	=	chemical concentration on surfaces [μg/m^2]
$conc.air$	=	chemical concentration in air [μg/m^3]
$thick.sur$	=	effective thickness of surface compartments [m]
$z.sur$	=	fugacity capacity of surfaces [mol/(Pa-m^3)]
$z.air$	=	fugacity capacity of air [mol/(Pa-m^3)]

Here z.sur and z.air are the fugacity capacities of the surface and air compartments. For air,

$$z.air = 1/(Rgas \times T) \quad (33)$$

Where:

$z.air$	=	fugacity capacity of air [mol/(Pa-m^3)]
$Rgas$	=	universal gas constant = 8.314 [Pa m^3 / (mol K)]
T	=	indoor temperature [K]

For the surfaces, the value depends on vapor pressure and surface material. The fit used here is

$$z.sur = z.air \times \frac{82500}{(vapor)^{0.65}} \quad (34)$$

Where:

$z.air$	=	fugacity capacity of air [mol/(Pa-m^3)]
$z.sur$	=	fugacity capacity of surfaces [mol/(Pa-m^3)]

The factor $\frac{82500}{(vapor)^{0.65}}$ requires vapor pressure in Pa, and results from a log-log regression model.

Once conc.air and conc.sur are determined, the Article_emission scenario calculates exposure and dose in a similar way as in Product.indirect, namely

$$exp.inhal = conc.air \times duration/1440 \quad (35)$$

$$dose.inhal = exp.inhal \times bva \times pai \quad (36)$$

$$exp.dermal = conc.sur \times TC \times avail.f \times dur.awake/60 \quad (37)$$

$$exp.ingest = exp.dermal \times om.ratio \quad (38)$$

Where:

$exp.inhal$	=	inhalation exposure concentration [ug/m^3]
$conc.air$	=	fugacity-based equilibrium air concentration [ug/m^3]
$duration$	=	time spent in home on selected day [min]
1440	=	conversion factor [min/day]
$dose.inhal$	=	inhalation dose [µg/day]
$exp.inhal$	=	inhalation exposure concentration [µg/m^3]
bva	=	basal ventilation rate [m^3/day]
pai	=	physical activity index [-]
$exp.dermal$	=	indirect dermal exposure from product use [µg/day]
$conc.sur$	=	average chemical loading on surfaces [µg/m^2]
TC	=	dermal transfer coefficient [m^2/hr]
$avail.f$	=	fraction of chemical available for transfer [-]
$dur.awake$	=	time spent at home while awake [min]
60	=	conversion factor [min/hr]
$exp.dermal$	=	indirect dermal exposure from product use [µg/day]
$om.ratio$	=	object-mouthing ratio [-]

4.7 Post-Exposure calculations

After all the exposure scenarios have been evaluated for all sources of a given chemical, the "post.exposure" function is called to evaluate the removal, absorption, and excretion processes. The methods depend on the route of exposure.

For inhalation, the exposure scenarios determine both an exposure and an inhaled dose. The units for exposure are the same as for concentration (micrograms per cubic meter). It is calculated as the time-weighted average over the 24 hour period being simulated. The inhaled dose has units of micrograms per day, and is given by the product of concentration, breathing ventilation rate, and exposure duration, summed over the inhalation sources. The inhaled dose represents the amount of chemical entering the lungs. An absorbed dose for inhalation is calculated by multiplying the inhaled dose by an absorption fraction specific to inhalation. In principle, this fraction should be chemical-dependent, but the current version uses a fraction common to all chemicals.

The ingestion exposure is the sum of two terms: one from the exposure scenarios and the other from the hand-to-mouth transfer of dermal exposure (see below). The absorbed dose for ingestion is the product of the ingestion exposure and an ingestion absorption fraction. As for

inhalation, this absorption fraction should in principle be chemical-specific, but the current version of SHEDS uses a value common to all chemicals.

The tradeoff in using absorption fractions common to all chemicals is that some chemical-specificity is lost, but the data requirements are greatly reduced. An important goal of SHEDS-HT is to be able to produce exposure estimates for a very large number of chemicals, with readily available input data. Chemical-specific absorption fractions could be used if they could be estimated from the chemical properties already available. This has been done for the dermal pathway.

The dermal exposure pathway is the most elaborate of the three, because it utilizes the concept of competing removal processes. Dermal exposure is defined in SHEDS as the amount of chemical contacting the skin. Dermal exposure arises from contact with surfaces (an air-to-skin pathway has not yet been implemented in SHEDS). In principle, "contact" exists only when the chemical has the possibility of interacting with the skin, so it must refer to the chemical in a thin layer close to the skin.

SHEDS uses the concept of "dermal loading" to quantify dermal exposure. Dermal loading measures the chemical burden on (or in) the skin surface. In the SHEDS model, when surfaces are contacted, some of the surface chemical transfers onto the skin, and will remain there until removed by one of the five competing removal processes. The loading calculation is detailed in the section on the Product.direct.dermal scenario.

The five removal processes are hand washing, bathing and showering (together), hand-to-mouth transfer, dermal absorption, and brushoff. Hand washing depends on the number of hand washes and the hand washing removal efficiency (rem.handwash.f). Bathing is the term for full body washing, and includes showers and swimming. It is a logical variable in SHEDS, and it either happens or not on the simulated day. If this happens, the removal term is sampled from the distribution for rem.bath.f. Hand-to-mouth transfer is a function of handmouth.area, handmouth.freq, and handwash.freq. Dermal absorption is sampled from rem.derm.abs.f. Brushoff represents removal by contact with clothes, other surfaces, shedding of skin flakes, and other processes not involving water or ingestion, and is sampled from rem.brushoff.f.

The five samples for the removal process rates are random samples from distributions, given on the exp.factors input file. Even if these distributions have been scaled so that their means sum to one, the random samples for any given person will not. It is not possible to remove more chemical than is present. Therefore, once the five samples have been evaluated for a given person, they are each converted to fractional amounts by dividing each one by the sum of the five. The dermal loading is then apportioned using these fractions.

For hand washing, bathing, and brushoff, the removed chemical effectively disappears and no longer matters. For dermal absorption, it becomes the absorbed dermal dose. For hand-to-mouth transfer, it is added to the ingestion exposure as a non-dietary term.

The bath and brushoff terms are evaluated by directly sampling the corresponding variable on the exp.factors file (that is, rem.bath.f for bath, and rem.brushoff.f for the brushoff process). The other three removal terms are composites of multiple factors.

The hand washing removal term was found in an analysis of SHEDS Multimedia output to increase more slowly as the number of hand washes increased. For a large population sample in a run for permethrin, the fraction of dermal loading removed on a daily basis was plotted against the number of hand washes per day, and the result were found empirically to be fit by

$$rem.hand.wash = rem.handwash.f \times \left(1 - e^{\left(\frac{-num.hand.washes}{1.6}\right)}\right) \quad (39)$$

Where:

$rem.hand.wash$	=	hand washing removal term [-]
$rem.handwash.f$	=	variable from exp.factors input file [-]
$num.hand.washes$	=	number of hand washing events that day [-]

If the number of hand washes per day is zero, then nothing is removed. As the number of hand washes becomes large, the amount removed is basically independent of the number of hand washes, and is given by rem.handwash.f. This value is less than one because the dermal loading includes chemical elsewhere on the body, which cannot be removed by hand washing. For smaller numbers of hand washes the results are not linear. However, alternate forms of this relationship could be defined and used without changing the nature of the SHEDS model.

The dermal absorption removal term is based on two factors, the exp.factor variable rem.derm.abs.f and the permeability factor Kp:

$$rem.absorb = rem.derm.abs.f \times Kp \,/0.208 \quad (40)$$

Where:

$rem.absorb$	=	dermal absorption term[-]
$rem.derm.abs.f$	=	variable from exp.factors input file [-]
Kp	=	dermal permeability from chem.props file [cm/hr]
0.208	=	permeability of permethrin [cm/hr]

The factor rem.derm.abs.f is common to all chemicals, and for historical reasons in SHEDS is assumed to be scaled to the dermal absorption of permethrin. Chemicals with higher Kp values are more readily absorbed. Therefore, the rem.derm.abs.f samples are multiplied by the Kp for the chemical under consideration and divided by the Kp for permethrin. This introduces chemical specificity into the dermal absorption process.

The hand-to-mouth removal term is based on fitting data obtained in runs of the SHEDS Multimedia model for permethrin. The function is

$$rem.mouth = hm.area \times sqrt(hm.freq) \times hw.eff \times \exp\left(-\frac{nhw}{11}\right) \quad (41)$$

Where:

$rem.mouth$	=	hand-mouthing term[-]
$hm.area$	=	handmouth.area.f variable from exp.factors input file [-]
$hm.freq$	=	handmouth.freq variable from exp.factors input file [#/hr]
$hw.eff$	=	rem.handwash.f variable from exp.factors input file [-]
nhw	=	number of hand washes on given day [#/day]

Note: The hm.area variable is the fraction of the total hand area (both hands) that contacts the mouth.

This equation is the result of a fit to data from SHEDS-Multimedia runs. The first term reflects the potential for transfer (more hand area in the mouth means more transfer). The second term is non-linear, reflecting the reduced loading on the hands after each hand-mouth contact. The third term hw.eff measures the difficulty of removing the chemical from the skin (here, saliva is assumed to have the same efficiency as water). The final term is nonlinear, and decreases with the number of hand washes.

There are three absorbed doses in SHEDS: for inhalation, ingestion, and dermal absorption. Each is expressed in two set of units: in micrograms per day and in (mg/kg/day). The latter is obtained from the former by dividing the former by (1000* body weight), where the body weight is in kg, and the factor of 1000 converts from micrograms to milligrams. In the R session, and on the output files, the "microgram" doses have the suffix "ug", and have been evaluated as absorbed chemical mass on the simulated day, so effectively are rates with units of micrograms per day.

4.8 Performing the Chemical Summary

After the post-exposure calculations, the results for the chemical are summarized and written to output files, before proceeding to the next chemical. This is done primarily for memory reasons, as the data requirements become onerous for a sizeable modeled population, and it is necessary to re-use many of the arrays for the next chemical.

The "fexp" object (the name stands for "final exposure") contains all the information on the current chemical, for the current set of simulated persons. If the person.output switch on the run.file was set to yes, then the relevant parts of fexp are written to an output file with suffix "_all" after the CAS number in the file name. Otherwise a small file containing summary

statistics is produced for that set. This summary file contains statistics on the variables exp.dermal, exp.ingest, exp.inhal, dose.inhal, dose.intake, abs.dermal.ug, abs.ingest.ug, abs.inhal.ug, abs.tot.mgkg, and ddd.mass. The statistics for each variable are the mean, standard deviation, and the following quantiles: 0.005, 0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.975, 0.99, and 0.995. The current version of SHEDS computes these statistics separately for each set, for which the filename has suffix "_setN" for the Nth set.

If source.output=yes, SHEDS creates a file for each set and chemical listing the mean exposures across the persons in the set, for each source. This file has suffix "_srcMeans" after the CAS and set number, and contains dermal, ingestion, and inhalation exposures, and the inhalation intake dose. All are in units of micrograms per day, except for inhalation exposure which is in micrograms per cubic meter. SHEDS also creates a summary file of source means across all sets, with suffix "_srcMeans" after the CAS number. This file has the same variables and units as the similar files for each set.

If person.output=yes, SHEDS calculates statistics across all the sets, reported in one file for each chemical with suffix "_allstats" after the CAS number. These statistics are the same ones reported for each set, namely, the mean, standard deviation, and a list of quantiles. If person.output=no then quantiles cannot be calculated, so this file is not produced.

5 OUTPUT

SHEDS writes all the output files from each run to a single folder under the general /output folder for the SHEDS installation. This folder has the same name as the run.name given on the input run.file. If this is the same as in a previous run, the old output is overwritten.

There is a set of output files for each chemical in the run. The three files types are as follows:

Stats	Selected percentiles for several exposure and dose variables
All	Exposure and dose variables for each simulated person
SrcMeans	Average exposure and dose from each source

"Stats" is always written. "All" is written when person.output=yes. "SrcMeans" is written when source.output=yes.

If a run contains a large number of simulated persons, it is faster to break it into "sets" using the set.size option on the run.file. Suppose the run has n.persons=50,000 and set.size=10,000 (do not put commas in large numbers in the input files; write these are 50000 and 10000). SHEDS then breaks the model run into five parts, modeling 10,000 persons at a time. The Stats and SrcMeans files are summaries for 10,000 persons, so there would be five output files of each

type, with suffixes like “set1”, “set2”, and so on. Each is produced and written when the set is complete.

The “All” file type has one record for every simulated person, covering all sets. Hence, in the above example it would have 50,000 records. If the All file exists (that is, if person,output=yes), then at the end of the run, SHEDS computes another Stats file (called “allstats”) and another SrcMeans file (called “all_SrcMeans”), for all the persons in the run.

5.1 The Stats Output Files

This file type is always created. There is one Stats file for each set of persons, for each chemical. For example, for chemical 50_00_0 (formaldehyde), these file names would be “CAS_50_00_0_set1stats.csv”, “CAS_50_00_0_set2stats.csv”, and so on. If the All file is created, then at the end of the run, an overall Stats file called “CAS_50_00_0_allstats.csv” is generated. All these files are in the subfolder with the run.name for the run, under the /output folder. Table 14 lists the variables contained in the Stats output files.

Table 14. Variables on the Stats output files

Name	Units	Description
statistic	[-]	Selected statistics of the population distribution
cohort	[-]	Population subgroup
exp.dermal	µg/day	Total dermal exposure across scenarios
exp.ingest	µg/day	Total ingestion exposure across scenarios
exp.inhal	µg/m^3	Total inhalation exposure (average air concentration) across scenarios
dose.inhal	µg/day	Total inhaled dose across scenarios
dose.intake	mg/kg/day	(abs.dermal+exp.ingest+dose.inhal), converted to (mg/kg/day)
abs.dermal.ug	µg/day	Amount absorbed through the skin
abs,ingest.ug	µg/day	Amount absorbed from the GI tract
abs.inhal.ug	µg/day	Amount absorbed from the lungs
abs.tot.ug	µg/day	Sum of abs.dermal, abs.ingest, and abs.inhal
abs.tot.mgkg	mg/kg/day	Same as abs.tot.ug, converted to standard dose units
ddd.mass	g/day	Chemical mass going down the drain

The Stats file consist of a set of tables, one for each of the following population cohorts:

Total (all persons)

Males
Females
Females_Repro (ages 16-49 inclusive)
Age 0-5
Age 6-11
Age 12-19
Age 20-65
Age 66+

Tables are omitted if the run does not contain anyone in that cohort. The population statistics that are reported are the mean, standard deviation, and the following percentiles:

0.005, 0.011, 0.025, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50,0.60, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 0.97.5, 0.99, 0.99.5

Each set of Stats will cover the number of persons given by the set.size for the run, except that the last set may have fewer (if the number of persons is not a multiple of the set size).

5.2 The All-persons Output Files

This file contains one record for each simulated person. It is not broken into sets, but there is still a different All file for each chemical. For example, for formaldehyde, the All file name is "CAS_50_00_0_all.csv". The variables on the All file are given in Table 15.

Table 15. Variables on the All output files

Name	Units	Description
person	[-]	Sequential index
gender	[-]	M=male, or F=female
age	yr	Age in years, rounded down (0-99)
season	[-]	W=winter, P=spring, S=summer, F=fall
weekend	[-]	1=Saturday or Sunday, 0=Monday-Friday
weight	kg	Body weight
exp.dermal	µg/day	Total dermal exposure across scenarios
exp.inhal	$\mu g/m^3$	Total inhalation exposure (average air concentration) across scenarios
exp.ingest	µg/day	Total ingestion exposure across scenarios
exp.diet	µg/day	Ingestion exposure from food scenarios
exp.nondiet	µg/day	Ingestion exposure from product and article scenarios

exp.drain	g/day	Chemical mass going down the drain
dose.inhal	µg/day	Total inhaled dose across scenarios
dose.intake	mg/kg/day	(abs.dermal+exp.ingest+dose.inhal), converted to (mg/kg/day)
abs.dermal.ug	µg/day	Amount absorbed through the skin
abs,ingest.ug	µg/day	Amount absorbed from the GI tract
abs.inhal.ug	µg/day	Amount absorbed from the lungs
abs.tot.ug	µg/day	Sum of abs.dermal, abs.ingest, and abs.inhal
abs.tot.mgkg	mg/kg/day	Same as abs.tot.ug, converted to standard dose units
urine.tot.ug	µg/day	Chemical mass excreted in urine
exp.window	µg/day	Airborne chemical leaving house by air exchange

5.3 The SrcMeans Output File

This set of files is produced if source.output=yes. For each chemical, there is one file for each set, and an overall file is produced as well if the All file is created. The variables in the SrcMeans file are given in Table 16.

Table 16. Variables on theSrcMeans output files

Name	Units	Description
[-]	[-]	Source index number
src.names	[-]	The source.id variable from the source input files
exp.dermal	µg/day	Average dermal exposure across persons
exp.ingest	µg/day	Average ingestion exposure across persons
exp.inhal	$\mu g/m^3$	Average inhalation exposure across persons
dose.inhal	µg/day	Average inhaled dose across persons
f.dermal	[-]	Average dermal exposure as fraction of total chemical in house
f.ingest	[-]	Average ingestion exposure as fraction of total chemical in house
f.inhal	[-]	Average inhalation dose as fraction of total chemical in house
mean.mass	[ug]	Average chemical mass across houses from this source

5.4 Combining Statistics across Chemicals

The usual SHEDS output files are separate for each chemical. A function called "combine_output" in the Postprocess,R module will combine selected information on the Stats files so that information on all chemicals is in the same file. The arguments to combine_output are

run.name	The name of the output folder to process
out.file	The name of the summary file to create (default="SHEDSOutFCS.csv")
metrics	List of statistics to include on this file

The metrics must be present on the Stats file. One possible list for metrics is

metrics = c("5%", "50%", "75%", "95%", "99%", "mean", "sd")

Each selected statistic is quoted, and the percentiles include the "%" sign. The list is presented in usual R format, using the "c" function to combine a comma-separated list of items.

The combine_output function extracts the requested statistics for each cohort, from all the "Stats" files in the designated folder, and writes them to the out.file (also in the same folder). The out.file has one record for each combination of chemical and cohort. The file has eleven columns for each requested metric, plus two for the statistic and cohort. Each data column has a name composed from the modeling variable and the metric.

5.5 Visualizing Outputs

There are four options for visualization functions in Plots.R, One function, allstats.variable.rank.plot, plots data from the allstats output files for all chemicals; two functions, puc.rank.plot and puc.boxplot, plot data from the srcMean output files from each run; the last function, puc.dtxsid, plots data from the srcMean output file for up to 6 chemicals. All four functions have a log transformed y-axis with a constant of 1e-10 added to the dependent variable,

5.5.1 all.stats.variable.rank.plot

The function allstats.variable.rank.plot creates a scatter plot with output.variable on the y axis and ascending chemical rank on the x axis. User can color points by cohort and assign percentiles to be point shape. The default color is black for the "Total" cohort, but the user can include as many cohorts as they wish as long as they are calculated in the SHEDS-HT run.

run.name default	The name of the output folder to process
output.variable	Any **one** output variable. Default = abs.tot.mgkg (mg/kg/day).
metrics	List of metrics to plot. Default = c("50%, "95%, "mean)
cohort.col	List of cohorts to plot. Default = "Total"
label_chem	Include chemical label. Default = F

Both the metrics and cohort.col arguments can take any number of metrics or cohorts as long as it is in the format of comma-separated character strings. For example, the argument cohort.col = c("Total", "Male", "Female") will plot points for the cohorts "Total," "Male," and "Female," with each point having a different color.

5.5.2 puc.rank.plot

puc.rank.plot is a scatter plot with the desired srcMeans variable on the y axis and ascending chemical rank on the x axis, for all product sources. Points will be colored by the higher-level PUC. Since this function pulls data from srcMean, every data point will be the mean calculated for that variable.

run.name default	The name of the output folder to process
output.variable	Any **one** output variable. Default = abs.tot.mgkg
label_chem	Include chemical label. Default = F

6 REFERENCES

Glen G, Zartarian V, Smith L, and Xue J. The Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway, Chemicals (SHEDS-Multimedia): Residential Module. SHEDS-Residential version 4. Technical Manual. Office of Research and Development, US EPA. 2012. Available at: http://www.epa.gov/heasd/research/sheds/user_information.html.

Isaacs KK, Glen WG, Egeghy P, Goldsmith MR, Smith L, Vallero D, Brooks R, Grulke CM, Özkaynak H. SHEDS-HT: an integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources. Environ Sci Technol. 2014 Nov 4;48(21):12750-9.

Joumard, R, M. Chiron, R. Vidon, M. Maurin, and J-M Rouzioux. 1981. “Mathematical Models of the Uptake of Carbon Monoxide on Hemoglobin at Low Carbon Monoxide Levels.” Environmental Health Perspectives. Vol. 41, pp. 277 - 289 (October 1981).

McCurdy, T.; Glen, G.; Smith, L.; Lakkadi Y. The National Exposure Research Laboratory's Consolidated Human Activity Database. J. Exp. Anal. Environ. Epidemiol. 2000, 10, 566-578.

O'Connor, A. M.A. Huijbregts, A.M. Ragas, A.J. Hendriks. Predicting the oral uptake efficiency of chemicals in mammals: combining the hydrophilic and lipophilic range. Toxicol. Appl. Pharmacol., 266 (1) (2013 Jan 1), pp. 150-156.

ten Berge, W. A simple dermal absorption model: derivation and application Chemosphere, 75 (2009), pp. 1440-1445.

US Census Bureau. 2000 Census of Population and Housing. https://www.census.gov/. 2000.

US Department of Agriculture (USDA). What We Eat in America, NHANES 1999-2006; Agricultural Research Service, Beltsville Human Nutrition Research Center, Food Surveys Research Group and U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics: Hyattsville, MD; http://www.cdc.gov/nchs/nhanes/wweia.htm, 2014a.

US Environmental Protection Agency (EPA). Exposure Factors Handbook 2011 Edition. EPA/600/R-09/052F; United States Environmental Protection Agency: Washington, DC, 2011.

US Environmental Protection Agency (EPA). Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA. Available at: http://www.epa.gov/oppt/exposure/pubs/episuite.htm, 2012.

Office of Air Quality Planning and Standards (OAQPS) US EPA. Total Risk Integrated Methodology (TRIM) Air Pollutants Exposure Model Documentation (TRIM.Expo / APEX, Version 4) Volume II: Technical Support Document. Available at: https://www.epa.gov/sites/production/files/2013-08/documents/apex45_usersguide_vol2_aug2012_0.pdf

US Environmental Protection Agency (EPA). The Consolidated Human Activity Database. http://www.epa.gov/heasd/chad.html, 2014.

Xue, J.; Zartarian, V.; Wang, S. W.; Liu, S. V.; Georgopoulos, P. Probabilistic modeling of dietary arsenic exposure and dose and evaluation with 2003-2004 NHANES data. Environ. Health Perspect. 2010, 118, 345-350.

Xue J, Zartarian V, Moya J, Freeman N, Beamer P, Black K, Tulve N, Shalat S. A meta-analysis of children's hand-to-mouth frequency data for estimating nondietary ingestion exposure. Risk Anal. 2007 Apr;27(2):411-20.

APPENDIX A: THE FUGACITY-BASED INDOOR FATE AND TRANSPORT MODEL

The SHEDS-HT model includes a simplified version of the SHEDS-Fugacity model. This is described in detail in the supplementary material to Isaacs et al. (2014). The original SHEDS-Fugacity model has ten compartments, while in SHEDS-HT this has been reduced to two: air and surfaces.

The SHEDS-Fugacity model was originally designed for one chemical at a time, and all the inputs were together on the same file. This is no longer feasible, as SHEDS-HT may have thousands of chemicals in the same run. Hence, the inputs are now split between two files. The chemical-independent variables are house properties, and come from the Fugacity input file. The chemical-dependent variables are from the Chemical Properties input file.

Table A-1. Fugacity-based model chemical-independent variables

Name	Description	Units	Value
aer.out	Air exchange rate for rooms with outdoors	1/d	Lognormal(11.9, 1.7); Assumed
area.sur	Total floor area of the house	m^2	Lognormal(130, 1.8); Assumed
height	Height of walls	m	Uniform(2.44,3); Assumed
lg.carb.f	Organic carbon fraction for large particles	-	Normal (0.15,0.01); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.clean.air	Cleaning removal rate for large particles in air e.g. from air filters on HVAC systems, or electrostatic filters.	1/d	Uniform (0.03, 0.5); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.clean.sur	Cleaning removal rate for large particles on surfaces, e.g. from vacuuming or mopping.	1/d	Uniform (0.035, 0.045) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.depos	Air-to-floor large particle deposition rate	m/d	Normal (387,20) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.load.air	Loading of large particles in air	$\mu g/m^3$	Uniform (2.2, 2.5) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed

lg.load.sur	Loading of large particles on surfaces	µg/cm^2	Uniform (11.5, 28) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
lg.resus	Surface-to-air large particle resuspension rate	1/d	Uniform (0.0015, 0.0017) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 larger particle sizes, variability assumed
sm.carb.f	Organic carbon fraction for small particles	-	Normal (0.3,0.03); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.clean.air	Cleaning removal rate for small particles in air, e.g. from air filters on HVAC systems, or electrostatic filters.	1/d	Uniform (0.018,0.22); Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.clean.sur	Cleaning removal rate for small particles on surfaces, e.g. from vacuuming or mopping.	1/d	Uniform (0.035, 0.045) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.depos	Air-to-floor small particle deposition rate	m/d	Normal (11,1) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.load.air	Loading of small particles in air	µg/m^3	Uniform (15, 25) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.load.sur	Loading of small particles on surfaces	µg/cm^2	Uniform (6, 14.5) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
sm.resus	Surface-to-air small particle resuspension rate	1/d	Uniform (0.00072, 0.00082) Mean estimated from values reported in Bennett and Furtaw (2004) for 3 smaller particle sizes, variability assumed
temp	Indoor temperature	K	Normal (296,2) ; Assumed
thick.bou	Boundary layer thickness over surfaces	m	Uniform (0.025,0.0275); Assumed
thick.sur	Effective thickness of surfaces	m	Normal (0.0098,0.002); Assumed

Table A-2. Fugacity-based model chemical -dependent variables

Name	Description	Units	Value
mw	Molecular weight	g/mol	Molecular weight
vapor	Vapor pressure	Pa	Minimum of 1E-12 Pa, Maximum 1E5 Pa
log.kow	Octanol-water partition coefficient	[-]	Kow = 10^(log.kow) used in calculations
Water.sol.mg.l	Solubility	mg/liter	Converted to mol/m3 for use in calculations Solub (mol/m3) = water.sol.mg.l / mw
Half.sediment.hr	Half-life in sediment	1/hr	Converted to daily decay rate on surfaces Decay.sur = 24*log(2)/half.sediment.hr
Half.air.hr	Half-life in air	1/hr	Converted to daily decay rate on surfaces Decay.air = 24*log(2)/half.air.hr

The fugacity model considers two particle sizes, with generic names "small" and "large". Particles may be airborne or deposited on surfaces. In this model, all masses are in micrograms [µg], and lengths are in meters [m]. For some input and output purposes, surface loadings may be expressed in [µg/cm^3].

A.1 Particle Masses and Cleaning Rates

Volume of air compartment [m^3]:

$$vol.air = area.sur \times height$$

Small and large particle masses in air [µg]:

$$sm.mass.air = vol.air \times sm.load.air$$

$$lg.mass.air = vol.air \times lg.load.air$$

Small and large particle masses on surfaces (µg); 1E4 factor is for unit conversion (cm^2 to m^2):

$$sm.mass.sur = area.sur \times 1E4 \times sm.load.sur$$

$$lg.mass.sur = area.sur \times 1E4 \times lg.load.sur$$

Cleaning rates from surfaces for large and small particles (1/day). These are sampled from the input distributions, but then constrained to force particle mass balance as below:

$$sm.clean.sur \geq sm.depos \times \frac{sm.load.air}{sm.load.sur} - sm.resus$$

$$lg.clean.sur \geq lg.depos \times \frac{lg.load.air}{lg.load.sur} - lg.resus$$

Cleaning rates from air for large and small particles (1/day). These are sampled from the input distributions, but then constrained to force particle mass balance as below:

$$sm.clean.air \geq \frac{sm.resus \times sm.mass.sur - sm.depos \times sm.load.air \times area.sur}{sm.mass.air}$$

$$lg.clean.air \geq \frac{lg.resus \times lg.mass.sur - lg.depos \times lg.load.air \times area.sur}{lg.mass.air}$$

A.2 Partition Coefficients

Conversion factor between µg and moles (µg/mol):

$$ug.mol = 1E6 \times MW$$

Air/surface partition coefficient (derived from the mean of the expressions for vinyl and carpet from Bennett and Furtaw, 2004):

$$z.sur = z.air \times \frac{82500}{vapor^{0.65}}$$

Air/particle partition coefficient for small and large particles (m^3/µg), where the universal gas constant is $R_{gas} = 8.314 \frac{\text{Pa}-\text{m}^3}{\text{mol}-\text{K}}$:

$$sm.kp = 1.6625\text{E-}12 \times K_{ow} \times sm.carb.f \times solub \times R_{gas} \times temp/vapor$$

$$lg.kp = 1.6625\text{E-}12 \times K_{ow} \times lg.carb.f \times solub \times R_{gas} \times temp/vapor$$

A.3 Fugacity Capacities and Fugacity to Mass Ratios

Fugacity capacity of air ([mol/(Pa m^3)]:

$$z.air = \frac{1}{R_{gas} \times temp}$$

Ratio of mass to fugacity (ZV, in µg/Pa) for bulk (vapor phase) air:

$$zvb.air = z.air \times vol.air \times ug.mol$$

Ratio of mass to fugacity (ZV, in µg/Pa) for surfaces:

$$zvb.sur = z.sur \times area.sur \times thick.sur \times ug.mol$$

Unit capacities of large and small particles (fugacity capacity per unit mass, in 1/Pa):

$$sm.cap = z.air \times sm.kp \times ug.mol$$

$$lg.cap = z.air \times lg.kp \times ug.mol$$

Ratio of mass to fugacity (ZV, in µg/Pa) for small and large particles in the air compartment:

$$sm.zv.air = zvb.air \times sm.kp \times sm.load.air$$

$$lg.zv.air = zvb.air \times lg.kp \times lg.load.air$$

Ratio of mass to fugacity (ZV, in µg/Pa) for small and large particles in the surface compartment:

$$sm.zv.sur = sm.cap \times sm.mass.sur$$

$$lg.zv.sur = lg.cap \times lg.mass.sur$$

Ratio of mass to fugacity (ZV, in µg/Pa) and fugacity to mass (iZV, Pa/ µg) for entire air and surface compartments (sum of phase-specific ZV):

$$zv.air = zvb.air + sm.zv.air + lg.zv.air$$

$$zv.sur = zvb.sur + sm.zv.sur + lg.zv.sur$$

A.4 Chemical Flows

The previous sections have defined the static properties of the surface and air compartments. In this section, the dynamics of chemical flow are considered.

The cleaning rate constants in units of [1/day] are given by

$$cln.air = \frac{(sm.zv.air \times sm.clean.air \; + \; lg.zv.air \times lg.clean.air)}{zv.air}$$

$$cln.sur = \frac{(sm.zv.sur \times sm.clean.sur \; + \; lg.zv.sur \times lg.clean.sur)}{zv.sur}$$

The chemical flow onto surfaces by particle deposition is the sum of the following two terms:

$$sm.dep = \frac{area.sur \; \times \; sm.load.air \; \times \; sm.depos \; \times \; sm.cap}{zv.air}$$

$$lg.dep = \frac{area.sur \; \times \; lg.load.air \; \times lg.depos \; \times lg.cap}{zv.air}$$

$$dep = sm.dep + lg.dep$$

The chemical flow from surfaces to air by particle resuspension is

$$res = \frac{(sm.mass.sur \times sm.resus \; + \; lg.mass.sur \times lg.resus)}{zv.sur}$$

Chemical may also flow between air and surfaces by diffusive flow. The diffusive transfer factor Yaf is adapted from Bennett and Furtaw (2004):

$$Yaf = min\left(\frac{diffus.air * z.air}{thick.bou} \; , \; \frac{0.0135}{vapor\text{^}0.32}\right)$$

$$diff.air \; = \frac{ug.mol \; \times \; area.sur \; \times \; Yaf}{zv.air}$$

$$diff.sur \; = \frac{ug.mol \; \times \; area.sur \; \times \; Yaf}{zv.sur}$$

The initial chemical masses in the surface and air compartments are

$$m0.air = vol.air \times app.rate.air$$

$$m0.sur = area.sur \times app.rate.sur$$

SHEDS allows constant chemical source terms, although these are not used for the current scenarios. These terms are:

$$src.air = c.out.air \times aer.out \times vol.air + c.src.air \times vol.air$$

$$src.sur = c.src.sur \times area.sur$$

Here, c.out.air is the chemical concentration [μg/m^3] in the outdoor air entering the house, c.src.air is the chemical emission rate into the indoor air [μg/ m^3/day], and c.src.sur is the chemical emission rate for surfaces [μg/m^2/day]. These variables could be added to the input files, if needed.

A.5 Construction of Jacobian Matrix

The equation for the system of masses in each compartment is

$$\frac{\mathrm{d}\mathbf{M}(\mathbf{t})}{\mathrm{dt}} = -\boldsymbol{J}\,\boldsymbol{M}(t) + \boldsymbol{S}$$

Here **M**(t) is a two component vector for the chemical mass in each compartment as a function of time. Compartment #1 is air, and #2 is the surfaces. The Jacobian matrix **J** is the 2x2 matrix of chemical flow coefficients. The vector **S** has elements (src.air, src.sur), but in the current applications of SHEDS is always set to zero. The elements of **J** are:

$$J_{1,1} = aer.out + decay.air + cln.air + dep + diff.air$$

$$J_{1,2} = -(res + diff.sur)$$

$$J_{2,1} = -(dep + diff.air)$$

$$J_{2,2} = decay.sur + cln.sur + res + diff.sur$$

A.6 Solution of the Model

As given above, the equation for the system of masses in each compartment is

$$\frac{\mathrm{d}\mathbf{M}(\mathbf{t})}{\mathrm{dt}} = -\boldsymbol{J}\,\boldsymbol{M}(t) + \boldsymbol{S}$$

Separate **M**(t) into a time-dependent part $\boldsymbol{MT}(t)$ and a time-constant part $\boldsymbol{MC}$:

$$\boldsymbol{M}(t) = \boldsymbol{MT}(t) + \boldsymbol{MC}$$

Then the differential equation separates into

$$\frac{\mathrm{d}\mathbf{MT}(\mathrm{t})}{\mathrm{dt}} = -\,\mathbf{J}\,\mathbf{MT}(\mathrm{t})$$

and

$$\boldsymbol{J}\,\boldsymbol{MC} = \boldsymbol{S}$$

These equations are simple enough that the solutions may be given explicitly. Let the initial masses in the air and floor compartments be m0.air and m0.sur, respectively, in units of micrograms. Then

$$r = \sqrt{(\mathrm{J}_{11})^2 + 4\,(\mathrm{J}_{12}\mathrm{J}_{21}) - 2\,(\mathrm{J}_{11}\mathrm{J}_{22}) + (\mathrm{J}_{22})^2}$$

$$m.c.air = \frac{\mathrm{J}_{22} \times \mathrm{S}_1 - \mathrm{J}_{12} \times \mathrm{S}_2}{\mathrm{J}_{11} \times \mathrm{J}_{22} - \mathrm{J}_{12} \times \mathrm{J}_{21}}$$

$$m.c.sur = \frac{\mathrm{J}_{11} \times \mathrm{S}_2 - \mathrm{J}_{21} \times \mathrm{S}_1}{\mathrm{J}_{11} \times \mathrm{J}_{22} - \mathrm{J}_{12} \times \mathrm{J}_{21}}$$

$$m.t0.air = m0.air - m.c.air$$

$$m.t0.sur = m0.sur - m.c.sur$$

The eigenvalues and eigenvectors of the Jacobian matrix are

$$\lambda_1 = (\mathrm{J}_{11} + \mathrm{J}_{22} + \mathrm{r})\,/2$$

$$\lambda_2 = (\mathrm{J}_{11} + \mathrm{J}_{22} - \mathrm{r})\,/2$$

$$U_1 = \begin{pmatrix} U_{1air} \\ U_{1sur} \end{pmatrix} = \begin{pmatrix} r + \mathrm{J}_{11} - \mathrm{J}_{22} \\ 2\,\mathrm{J}_{21} \end{pmatrix}$$

$$U_2 = \begin{pmatrix} U_{2air} \\ U_{2sur} \end{pmatrix} = \begin{pmatrix} r - \mathrm{J}_{11} + \mathrm{J}_{22} \\ -2\,\mathrm{J}_{21} \end{pmatrix}$$

The above eigenvectors are not scaled to unit length. The determinant of the eigenvalue matrix $U=(U_1,U_2)$ is:

$$det(U) = U_{1air} \times U_{2sur} - U_{1sur} \times U_{2air} = -4 \times r \times J_{21}$$

In general, the initial state is a mix of the two eigenstates with amounts given by the constantsk_{1a}, k_{2a}, k_{1f}, and k_{2f}. Each eigenstate decays exponentially at a rate that depends on its eigenvalue:

$$mt.air(t) = k1.air \times e^{-\lambda_1 t} + k2.air \times e^{-\lambda_2 t}$$

$$mt.sur(t) = k1.sur \times e^{-\lambda_1 t} + k2.sur \times e^{-\lambda_2 t}$$

The constants in the above equations are given by

$$k1.air = \frac{U_{1air} \times U_{2sur} \times m.t0.air - U_{1air} \times U_{2air} \times m.t0.sur}{\det(U)}$$

$$k2.air = \frac{-U_{1sur} \times U_{2air} \times m.t0.air + U_{1air} \times U_{2air} \times m.t0.sur}{\det(U)}$$

$$k1.sur = \frac{U_{1sur} \times U_{2sur} \times m.t0.air - U_{1sur} \times U_{2air} \times m.t0.sur}{\det(U)}$$

$$k2.sur = \frac{-U_{1sur} \times U_{2sur} \times m.t0.air + U_{1air} \times U_{2sur} \times m.t0.sur}{\det(U)}$$

These equations determine the time-dependent mass **MT**(t). The total chemical mass in each compartment is then **MT**(t)+**MC**:

$$mass.air(t) = mc.air + k1.air \times e^{-\lambda_1 t} + k2.air \times e^{-\lambda_2 t}$$

$$mass.sur(t) = mc.sur + k1.sur \times e^{-\lambda_1 t} + k2.sur \times e^{-\lambda_2 t}$$

These equations may be evaluated at any time t>0, where t is the time in days since the product application. In SHEDS-HT, these equations are implemented in the last few lines of the get.fug.concs() function in the Fugacity.R file.

APPENDIX B: SOURCE.ID VALUES IN THE CPDATV3 ILLUSTRATIVE SCENARIO

SHEDS uses "source.id" to link source-specific information across the three "source" input files. While "source.id" effectively arbitrary (is not hard-coded into SHEDS), it is useful to have a standard list, so that new input files may be prepared without the danger of incompatible definitions among the three files.

A four-part coding system has been developed for this purpose. The parts are

1) One initial letter.
2) A two-letter code for general category.
3) A three-digit code for "type".
4) A three-digit code for specific refinements of the type.

These four parts are conventionally separated by periods for clarity, but a compressed form which omits the periods can also be used.

The initial letter may be set to "A" for articles, "F" for foods, or "P" for products. Currently, only the products have been assigned explicit codes for all the levels, for use with EPA's Product Composition Database. The general categories for products are:

Table B-1. Two-letter codes for general categories

Abbreviation	Description
AC	Arts and crafts
AP	Automotive products
HM	Home maintenance
HO	Home office
IH	Inside the home
LY	Lawn and yard
PC	Personal care
PE	Pesticides
PT	Pet-related products

The "types" refer to products for specific purposes. For example, "P.PC.870" is toothpaste. The final three-digit code breaks this down into more specialized groupings. In all cases, omitting the final three digits is equivalent to having ".000" at the end, and it means all examples are included. In the above case, both "P.PC.870" and "P.PC.870.000" refer to all varieties of

toothpaste. In this case, there are specific sub-types "P.PC.870.007" for gel toothpaste, and "P.PC.870.099" for "other toothpaste", which means non-gel in this context. Since "other toothpaste" is not clearly defined in isolation, it is preferable to either run just the ".000" refined types, or to run all the non-zero refined types.

The ".000" category should not be combined with non-zero categories in the same run, because this double-counts products. For example, running "P.PC.870.000" and "P.PC.870.007" together would result in each person who uses gel toothpaste being exposed twice, once to each of these product codes.

The code for product type (the first 3-digit code) is not universal. For example, the code "P.AC.010" represents "arts and crafts adhesive", but "P.AP.010" represents "antifreeze". The unique product types on the current list are:

Table B-2. Codes for product types

Code	Description	Code	Description
P.AC.010	arts and crafts adhesive	P.LY.070	lawnmower fluids
P.AC.020	arts and crafts cleaner	P.LY.080	mulch
P.AC.030	arts and crafts finish	P.LY.090	pool chemicals
P.AC.040	arts and crafts paint	P.LY.100	potting soil
P.AC.050	body paint	P.LY.110	surface deicer
P.AC.060	bubble solution	P.LY.120	trees
P.AC.070	craft kit	P.PC.010	acne spot treatment
P.AC.080	crayons	P.PC.020	aftershave
P.AC.090	dye	P.PC.030	baby lotion
P.AC.100	fabric paints and sealers	P.PC.040	baby oil
P.AC.110	finger paint	P.PC.050	baby powder
P.AC.120	flocking	P.PC.060	baby shampoo
P.AC.130	fogger	P.PC.070	baby wash
P.AC.140	glaze	P.PC.080	baby wipes
P.AC.150	gun cleaner	P.PC.090	bar soap
P.AC.160	pens and markers	P.PC.100	bath oil
P.AC.170	play dough	P.PC.110	bath paints/crayons
P.AP.010	antifreeze	P.PC.120	bath salts
P.AP.020	auto air freshener	P.PC.130	bite relief
P.AP.030	auto fluids and additives	P.PC.140	blush/bronzer
P.AP.040	auto lubricant	P.PC.150	body adhesive
P.AP.050	auto paint	P.PC.160	body care set
P.AP.060	auto refrigerant	P.PC.170	body oil
P.AP.070	boat cleaner	P.PC.180	body powder
P.AP.080	boat engine fluids	P.PC.190	body scrub

Code	Description	Code	Description
P.AP.090	body cleaner	P.PC.200	body wash
P.AP.100	body repair	P.PC.210	body wipes
P.AP.110	body wax	P.PC.220	bubble bath
P.AP.120	degreaser	P.PC.230	clipper lubricant/cleaner
P.AP.130	detailing	P.PC.240	contact care
P.AP.140	motor oil	P.PC.250	cosmetic tool cleaner
P.AP.150	windows/windshield	P.PC.260	denture adhesive
P.HM.010	adhesive remover	P.PC.270	denture cleaner
P.HM.020	caulk/sealant	P.PC.280	deodorant
P.HM.030	concrete	P.PC.290	depilatory
P.HM.040	corrosion protection	P.PC.300	diaper cream
P.HM.050	degreaser	P.PC.310	dry shampoo
P.HM.060	finish	P.PC.320	ethnic hair care
P.HM.070	grout sealer	P.PC.330	eye cream
P.HM.080	lock deicer	P.PC.340	eye drops
P.HM.090	lubricant	P.PC.360	eye liner
P.HM.100	mortar or grout	P.PC.370	eye makeup
P.HM.110	multipurpose adhesive	P.PC.380	eye products
P.HM.120	paint	P.PC.390	eye shadow
P.HM.130	paint cleaner	P.PC.400	face cleansing wipes
P.HM.140	paint texture	P.PC.410	face cream/moisturizer
P.HM.150	paint thinner	P.PC.420	face mask
P.HM.160	patch and repair	P.PC.430	face powder
P.HM.170	plumbing	P.PC.440	face scrub
P.HM.180	primer	P.PC.450	face wash
P.HM.190	putty or filler	P.PC.460	foot care
P.HM.200	refrigerant	P.PC.470	foundation/concealer
P.HM.210	roof	P.PC.480	fragrance
P.HM.220	septic system	P.PC.490	glitter
P.HM.230	spray foam	P.PC.500	hair bleach
P.HM.240	stain	P.PC.510	hair color
P.HM.250	stripper	P.PC.520	hair color activator
P.HM.260	surface sealer	P.PC.530	hair color developer
P.HM.270	welding	P.PC.540	hair color toner
P.HM.280	wood adhesive	P.PC.550	hair conditioner
P.HO.010	printer ink	P.PC.560	hair conditioning treatment
P.HO.020	printer toner	P.PC.570	hair relaxer
P.HO.030	white out	P.PC.580	hair spray
P.IH.010	air freshener	P.PC.590	hair styling
P.IH.020	automatic dishwashing additive	P.PC.600	hand sanitizer

Code	Description	Code	Description
P.IH.030	automatic dishwashing detergent	P.PC.610	hand soap
P.IH.040	bathroom cleaner	P.PC.620	hand wipes
P.IH.050	bleach	P.PC.630	hand/body lotion
P.IH.060	candles	P.PC.640	lice shampoo
P.IH.070	carpet cleaner	P.PC.650	liniment
P.IH.080	carpet deodorizer	P.PC.660	lip balm
P.IH.090	dish soap	P.PC.670	lip color
P.IH.100	disinfectant	P.PC.680	lip gloss
P.IH.110	drain	P.PC.690	lip liner
P.IH.120	dry cleaner	P.PC.700	makeup primer
P.IH.130	dryer sheets	P.PC.710	makeup remover
P.IH.140	electronics cleaner	P.PC.720	makeup set
P.IH.150	fabric deodorizer	P.PC.730	mascara
P.IH.160	fabric protectant	P.PC.740	mouthwash
P.IH.170	fabric softener	P.PC.750	nail adhesive
P.IH.190	floor cleaner	P.PC.760	nail polish
P.IH.200	floor polish	P.PC.770	nail polish remover
P.IH.210	glass cleaner	P.PC.780	nail products
P.IH.220	hand cleaner	P.PC.790	scalp treatment
P.IH.230	heavy duty cleaner	P.PC.800	self-tanner
P.IH.240	houseplant care	P.PC.810	sexual wellness
P.IH.250	lamp oil/lighter fluid	P.PC.820	shampoo
P.IH.260	laundry detergent	P.PC.830	shaving cream
P.IH.270	laundry fragrance	P.PC.840	sunscreen
P.IH.280	laundry stain remover	P.PC.850	teeth whitener
P.IH.290	laundry starch	P.PC.860	toner
P.IH.300	lime remover	P.PC.870	toothpaste
P.IH.310	metal polish	P.PC.880	waxing
P.IH.320	oven cleaner	P.PE.010	animal repellent
P.IH.330	shoe polish or protectant	P.PE.020	fungicide
P.IH.340	surface cleaner	P.PE.030	insect repellent
P.IH.350	upholstery cleaner	P.PE.040	insecticide
P.IH.360	wood polish	P.PE.050	rodenticide
P.LY.010	cleaner	P.PT.010	aquarium
P.LY.020	garden care	P.PT.020	cat litter
P.LY.030	garden fertilizer	P.PT.030	other pet treatments
P.LY.040	grill/camping fuel	P.PT.040	pesticide
P.LY.050	herbicide	P.PT.050	pet shampoo
P.LY.060	lawn fertilizer	P.PT.060	pet stain cleaner

The final 3-digit code for refined product type is intended to be universal. The currently assigned values are:

Table B-3. Three-digit codes for refined product types

Code	Description	Code	Description
001	acne	019	permanent
002	algaecide	020	pet
003	chlorinating	021	pet|spray
004	dandruff	022	ph control
005	exterior	023	powder
006	exterior|spray	024	professional
007	gel	025	shock
008	gel|spray	026	skin
009	interior	027	skin|spray
010	interior|spray	028	solid
011	leave-in	029	spray
012	leave-in|spray	030	temporary
013	liquid	031	temporary|spray
014	mousse	099	other
015	mousse|spray	100	children
016	oil or diffuse	101	children|spray
018	other|spray		