### **Species Sensitivity Distribution (SSD) Analysis TOOL**

Version 1.0

Kristin Connors, Scott Belanger, Greg Carr, Christian Geneus

The Procter & Gamble Company

### **Introduction**

### Species Sensitivity Distributions (SSD) are a statistical tool developed to describe the variation of species sensitivities to chemical exposure. SSDs are used for decision support in environmental protection and management. Use of SSDs in environmental management is flexible and wide-ranging including establishment of water quality criteria, prospective evaluations of chemical hazard, and as input to the evaluation of impacts of in the context of industrial products’ Life Cycle Assessment (LCA).

### No universal international guidance document on the development, application, and implementation of SSD methodology is presently available. Guidelines have been written at a high, overview level to allow flexibility by scientists and regulatory authorities in the use of SSD methodology. A single rigid form may not meet the needs for a regulatory authority and other aspects come to bear including choice of acceptable species and interpretation of the role of risk assessment in risk management. This may include how conservative (or liberal) an environmental assessment should be to meet environmental policies in a given region.

### The quality of a given SSD is tied to several factors, most of which are universally considered in regulatory applications across the globe. This includes discussion of the regulatory environment for which the assessment is conducted, data quality evaluations (identification of potential SSD input data), determinations of taxonomic diversity and breadth, statistical assessments, the parameter estimates from which the regulatory assessment is conducted and various additional considerations including mode of action assignment and the importance of specific ecotoxicity values below the 5th percentile of the distribution (Belanger et al. 2017). A detailed analysis of these considerations is beyond the scope of this user guide.

### Typically, SSDs are displayed as a cumulative distribution function which takes the form of an S-curve. This is a convenient display that highlights the choice of the Hazardous Concentration X (HCx) where x is usually chosen to represent a small value that is interpreted to be protective of some large fraction of species in an ecosystem (although literal interpretation of this is not advised). HC5NOEC (in other words, an HC5 that is derived from chronic NOEC toxicity data) is commonly used with an absolute interpretation. Assuming that sets of laboratory test species and sets of field species have a similar sensitivity distribution and exposure to a chemical, (only) 5% of the test and field species would be exposed above their species-specific NOEC at an ambient concentration equal to the HC5NOEC of that chemical, so that 95% of the tested species would be exposed below their NOEC. Protection of ecosystem structure at this level is assumed to imply protection of ecosystem function.

### When presenting SSD information under a given regulatory framework it is customary that the plot of the cumulative distribution, fitted curve, the actual HC5 and the upper and lower confidence limits are provided. Additional investigations can be completed to determine the sensitivity and stability of the SSDs (“leave-one-out” and “add-one-in” analysis, respectively).

### **DATA**

### This tool uses a simple copy-paste process for data input.  We recommend that data be set up in Excel or similar where the rectangular region containing the data are copied and then pasted into the input field.  An example is provided with a download button near the top next to the reset button.

### 1.     Provide a simple rectangular data table with results for each case/species arranged by rows.  Extra annotations, formatting, etc may need to be removed.  Do not copy whitespace, use the minimum possible region containing the data.

### 2.     The data block should have a header line with names.  The column names can be flexible, but they should be unique and useful for chosing the variables for analysis if there is any possibility for confusion.  The names are not used in the output except in a view of the input data and variable selections.

### 3.     The tool warns when multiple entries have the same species labels.  The analysis can still proceed.  The user is responsible for resolving, if appropriate.

### 4.     For the most reliable performance, paste only columns required for the analysis.  The tool is designed to allow for extra columns in the data, but we cannot guarantee that the program will not confuse columns versus the users intentions.

### **VARIABLE ROLES:**

### Response/NOEC value:  Only numeric variables.  If more than one column of numeric data is supplied, the one with the most unique values is populated into this selection box.  If more than one variable meets the previous criterion, a normal goodness-of-fit test is used as a tie-breaker.  All numeric variables remain in the list of potential analysis variables, these rules only provide a guess to which variable is the one intended for analysis.

### Species label:  To be most general, this could be numeric or character.  The first one with the most unique values is populated.  The program will respond with an error if the column chosen is not unique for every data value.

### Grouping variable:  When the grouping checkbox is selected the column with the fewest unique values, but more than one, is populated.  Most groups should have at least two data values.  Any group label with a single value may invalidate the grouping statistical analysis.

### The tool will allow variables to play dual roles (e.g., for a grouped analysis it may be of interest to label observations by taxonomic group instead of unique species labels).

### **COMPONENTS OF THE ANALYSIS:**

### The SSD analysis tool fits log-logistic and log-normal distributions.  In SSD literature other distributions are occasionally used, but at typical sample sizes it is very difficult to statistically differentiate distributions.

### There will always be a best fitting distribution, but such a selection method has high error rates. We recommend evaluating distributions based on Anderson-Darling Goodness of Fit statistic, and if both distributions provide a suitable fit, select the more conservative HCx value as the SSD output. The output in MS Excel format provides the numerical results, while the PDF output provides graphical summaries.

### The user may optionally request leave-one-out and add-one-in analyses to investigate the sensitivity of the fits to the data in hand, and the potential that a single new observation could move the HCx to various degrees. The user may also chose to complete a grouping analysis to determine if there are substantial patterns of differences in sensitivity with respect to trophic level/taxonomic groups.

### For more information, please consult the User Guide

### **DISCLAIMER:**

### No warranty is made.  It is the responsibility of the user to ensure that the application of this tool is appropriate for the data being analyzed.