Species Sensitivity Distribution (SSD) Analysis Tool: User Guide

Version 1.0

Kristin Connors, Scott Belanger, Greg Carr, Christian Geneus

The Procter & Gamble Company

For further information, contact:

Kristin Connors (Connors.ka@pg.com)

**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.

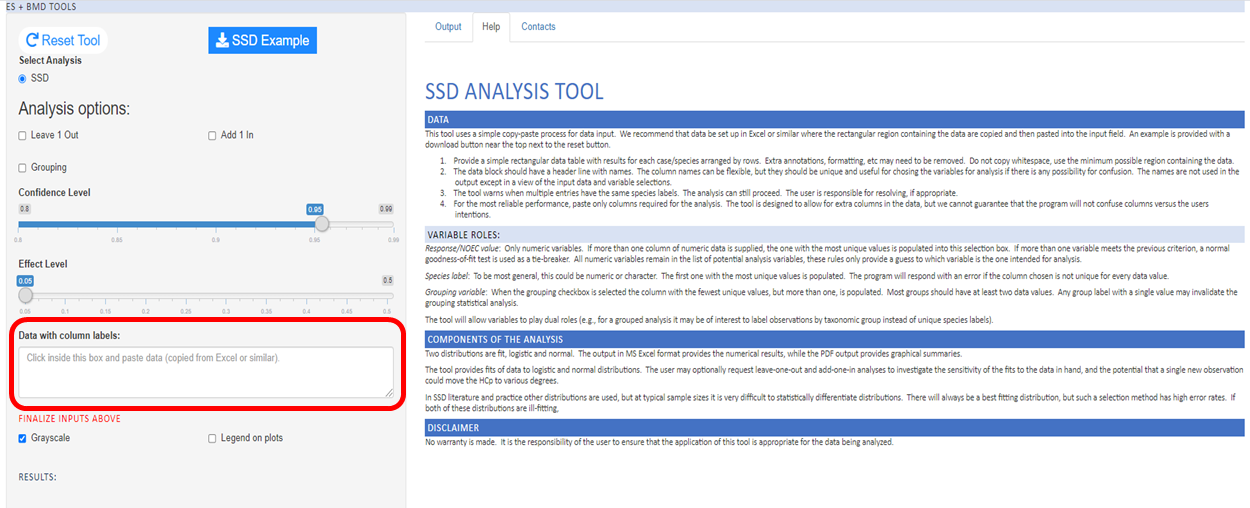
Typically, SSDs are displayed as a cumulative distribution function which takes the form of an S-curve. This is a convenient display that highlights the choice of the Hazardous Concentration X (HCx) where x is usually chosen to represent a small value that is interpreted to be protective of some large 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).

**Using SSD Analysis Tool**

**Step 1: Upload 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 paste into the input field.



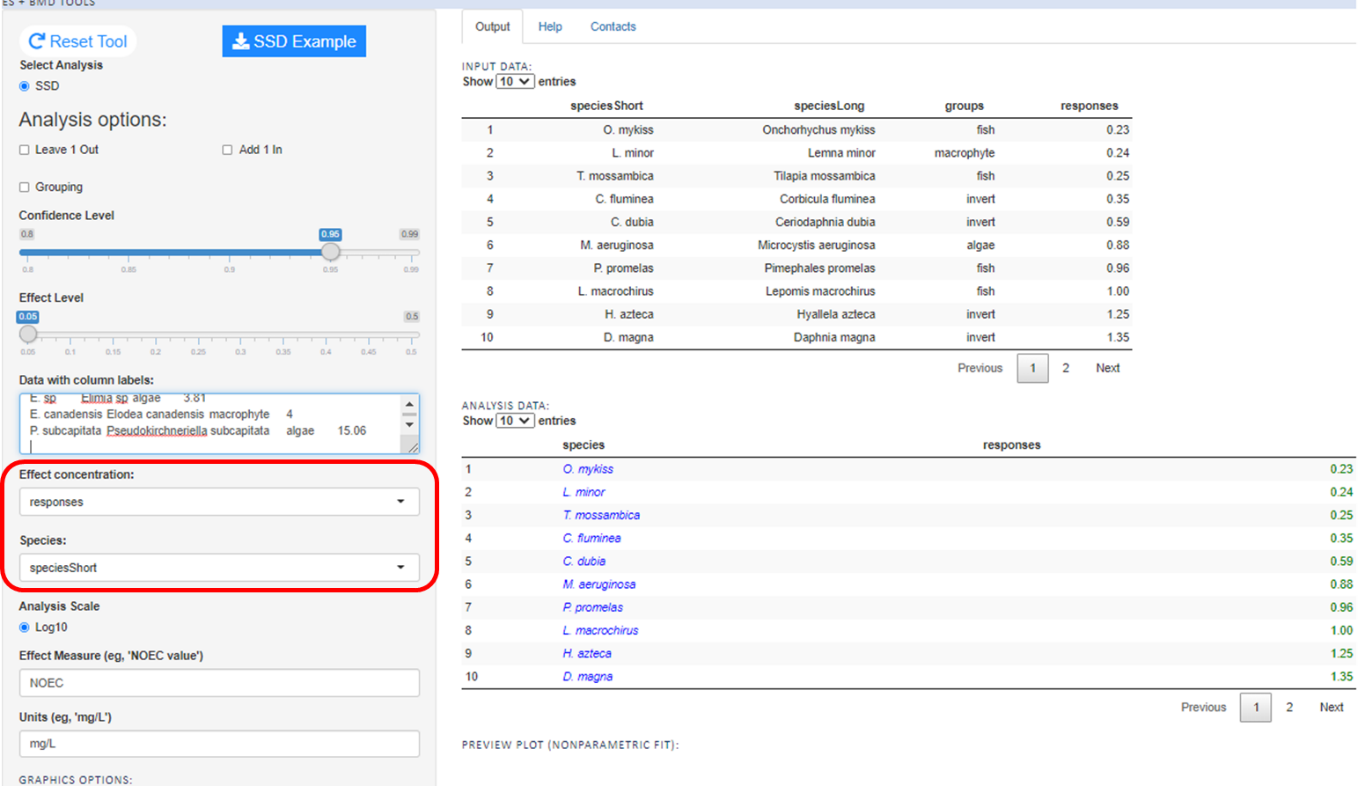
We suggest users upload the following information: Species name (Latin name, common name), Response information (Effect concentration), Test statistic (e.g., NOEC, ECx). The user may also want to include the trophic level for each species (e.g., Fish, Invertebrate, Algae, Macrophyte) to allow for additional grouping analysis (more details below). Extra annotations, formatting, etc. may need to be removed.  Do not copy whitespace, and use the minimum possible region containing the data.

The data block should have a header line with names.  The column names can be flexible, but they should be unique and useful for choosing 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. Column names will be populated in output files and in figures.

The tool warns when multiple entries have the same species labels.  The analysis can still proceed.  The user is responsible for resolving, if appropriate. Typically, a geometric mean is taken for species with multiple entries. This would need to be completed outside the SSD Analysis Tool.

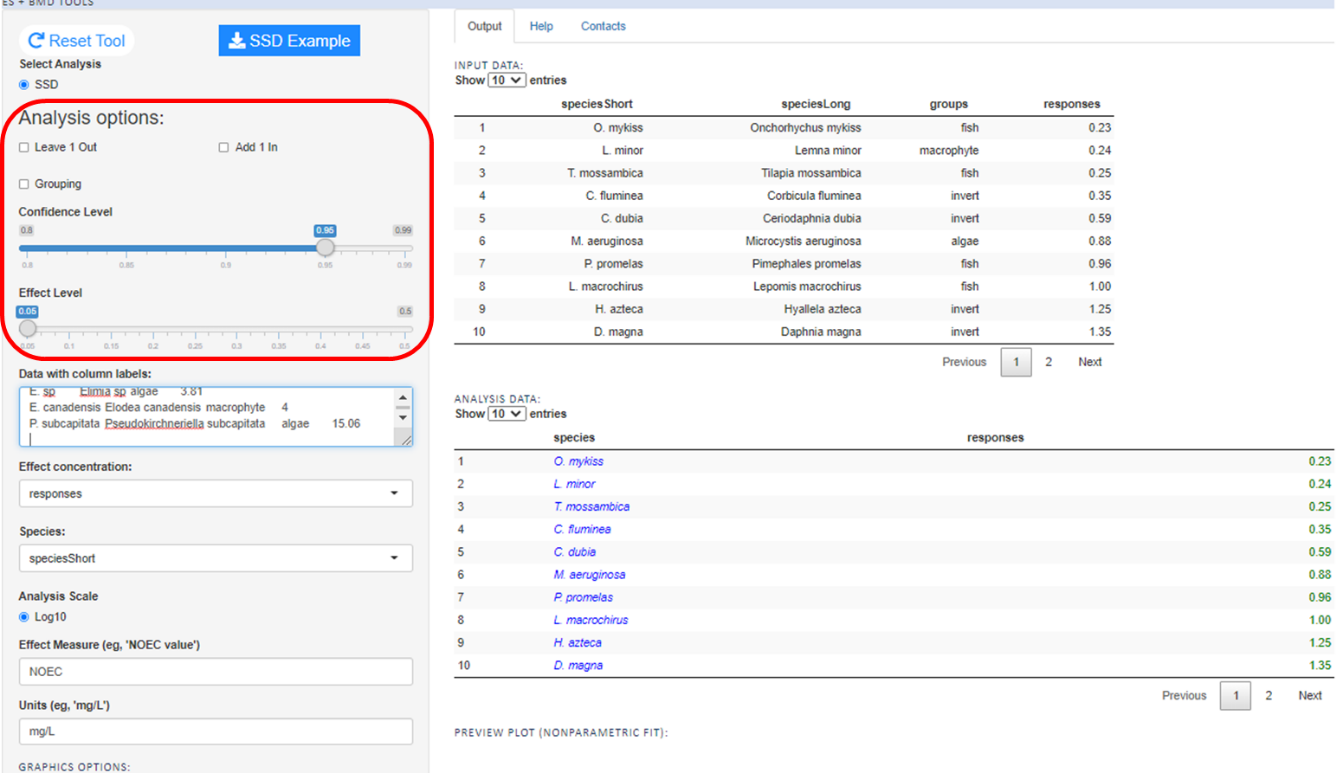
For the most reliable performance, paste only columns required for the analysis.  The tool is designed to allow for extra columns in the data, but we cannot guarantee that the program will not confuse columns versus the user’s intentions.

After the data has been pasted into the “Data with column labels” box, the user must select the columns corresponding to “Effect concentration” and “Species”. Data in these columns will be used for the SSD analysis and figure generation.



**Step 2: Select analysis options**

The default “Effect level” is set to 0.05, corresponding to a calculation of an SSD HC5 value. Users can select a different percentile of interest by using the sliding bar.

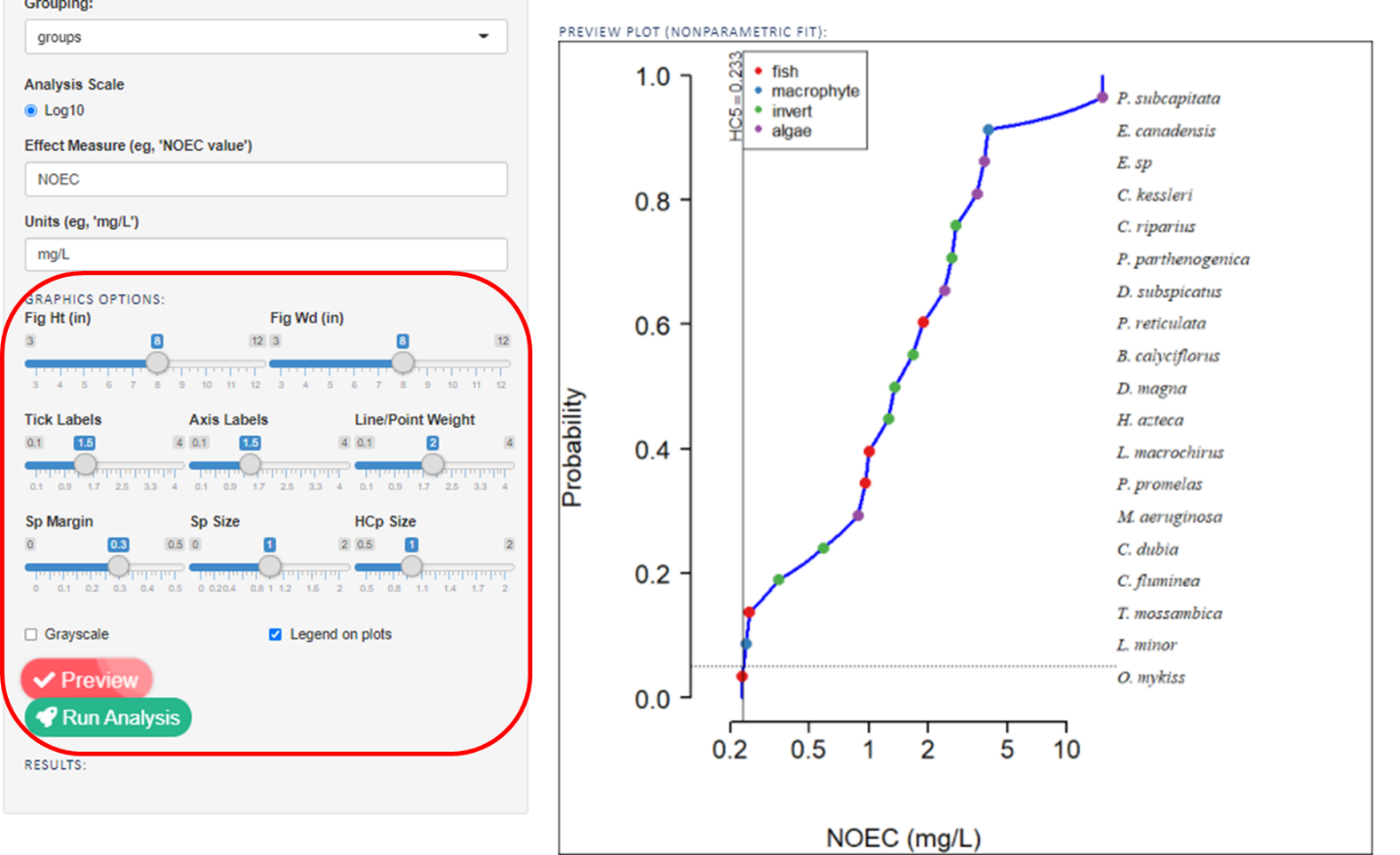


Three additional analysis options are available: leave-one-out, add-one-in, and grouping.

1. **“Leave-1-Out”** analysis can be used to probe the influence of individual points on the estimates of the HC5, by serially refitting the model after removing one of the data values from the full set. In this manner, the most influential data points can be identified. A counter-intuitive consequence is often that values in the mid-range of available toxicity measurements are highly influential with the reason being that when removed, the overall variance in remaining NOECs/ECx values is greater. This variance contributes to the fitted distribution and can be quite important when the overall SSD is flat (i.e., there is a wide distribution of toxicity data) and the number of values are fewer than 15. Note: this analysis is computationally intensive and requires additional time to complete.
2. **“Add-1-In”** analysis can be used to probe the stability of the SSD. In this sensitivity analysis, hypothetical data values are calculated that, if available and added to the SSD input, would decrease the HC5 by a factor of 2, 3, and 5, respectively (although any reduction can easily be calculated). The probability that the newly observed (hypothetical) toxicity value can be directly calculated based on the distribution estimated from the observed data values. This probability is defined as the probability of observing a toxicity value at least as extreme (or small) as the value that would give exactly the HC5/2, HC5/3, and HC5/5. It is also instructive to determine the relative ratio of the existing most sensitive toxicity value to the new hypothetical values. In practical terms, once an SSD is constructed from a fairly large toxicity data set, the ability to move the HC5 further left is increasingly difficult and can require in extreme cases new toxicity values that are several orders of magnitude below existing values and can be below any level of analytical detection (Belanger et al. 2019, Carr et al. 2019). Note: this analysis is computationally intensive and requires additional time to complete.
3. **Grouping** analysis performs a one-way analysis of variance assessment on grouped variable. This analysis can be useful to determine if there are substantial patterns of differences in sensitivity with respect to trophic level/taxonomic groups. 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.

**Step 3: Select graphical parameters**

The SSD Analysis Tool contains several features to create publication-quality figures. These options are purely for aesthetics and do not change the statistical analysis.



1. Provide the statistical endpoint (e.g., NOEC, EC50) value in the “Effect Measure (eg, ‘NOEC value’)” blank. This will be populated in figure legends and output files.
2. Provide the effect concentration units in the “Units (e.g., ‘mg/L’)” blank. This will be populated in figure legends and output files.
3. Adjust “Graphics Options” by sliding scales for figure height, weight, size of tick labels, axis labels, line/point weight, margin of graph allotted for species names, size of species names, and size of the HCp annotation.
4. Click “Preview” to view graphical output or to final any changes to graphical options before running the final analysis.

**Step 4: Interpreting data output**

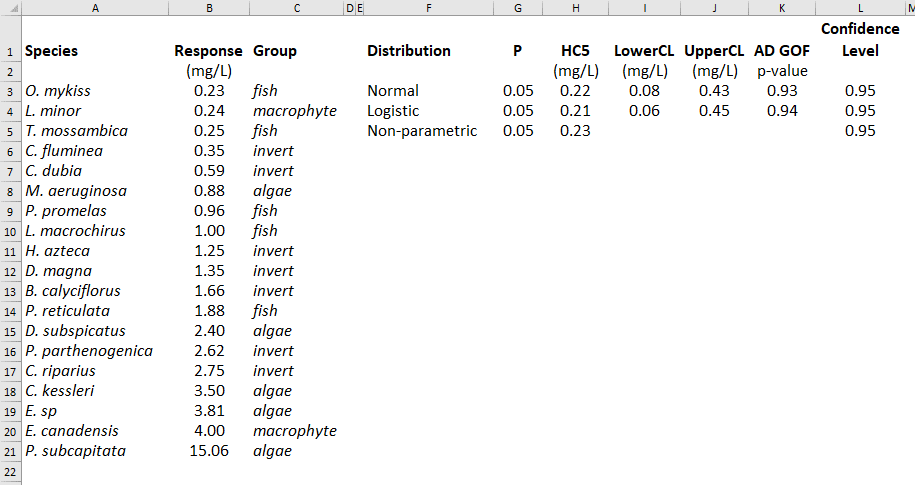
The SSD analysis tool fits log-logistic, log-normal, and non-parametric distributions.  The output in MS Excel format provides the numerical results, while the PDF output provides graphical summary

MS Excel file: “Data Listing” tab

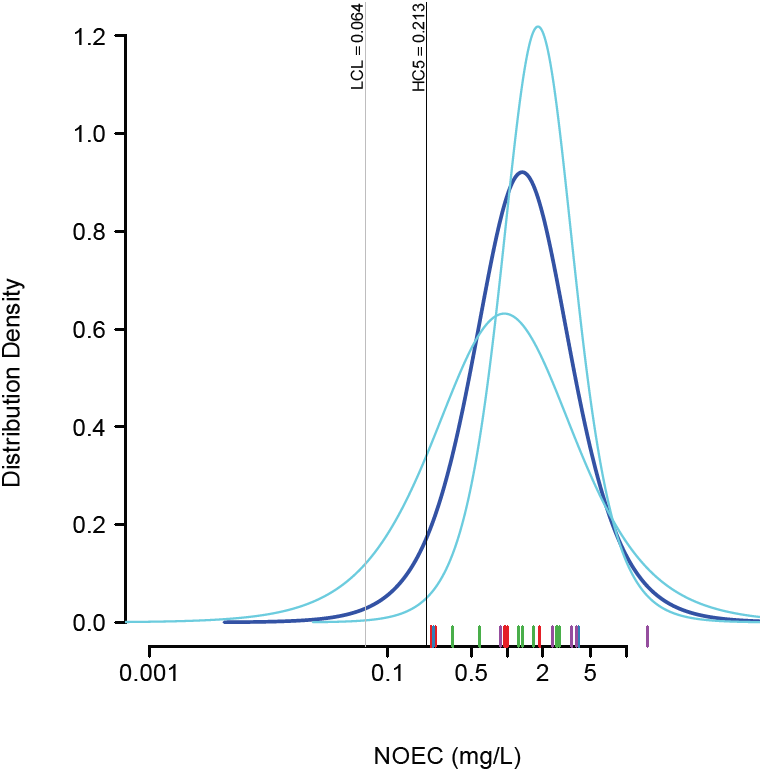
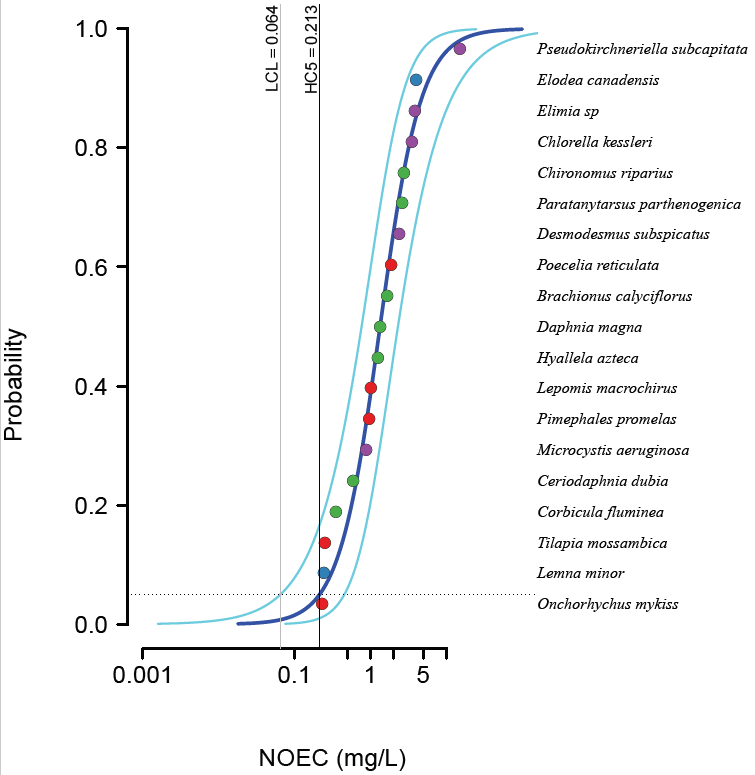
This tab provides a summary of the data uploaded into the SSD Analysis Tool. The HC5 values and upper and lower 95% confidence intervals are reported for each distribution. The upper and lower 95% confidence intervals are determined using non-parametric bootstrap. The general concept can be formulated as follows: 1) We generate random values based on resample (with replacement) from the data, 2) fit these values to the fitted distributions, 3) for each of the fitted values, we estimate the HC5, 4) repeat the process for a total of 10,000 permutations to determine the variance of the HC5 values.

The Anderson-Darling Goodness of Fit (AD GOF) value is provided. Anderson-Darling is a non-parametric test used to compare how well a data set fits a probability distribution model (e.g., log-normal or log-logistic). The null hypothesis is that the data follows the selected distributed (e.g., log-normal). Failure to reject the null hypothesis (a non-statistically significant AD GOF p-value), indicates the data plotted with this distribution is appropriate for this SSD analysis.

In this example below, the AD GOF statistic for both the normal and logistic distributions suggest either distribution provides a suitable fit and either distribution model would be defensible. Given that the log-logistic distribution results in a slightly more conservative HC5 in this example, the log-logistic results would be selected as the final SSD output.



The pdf output contains the graphical representation of the distribution.



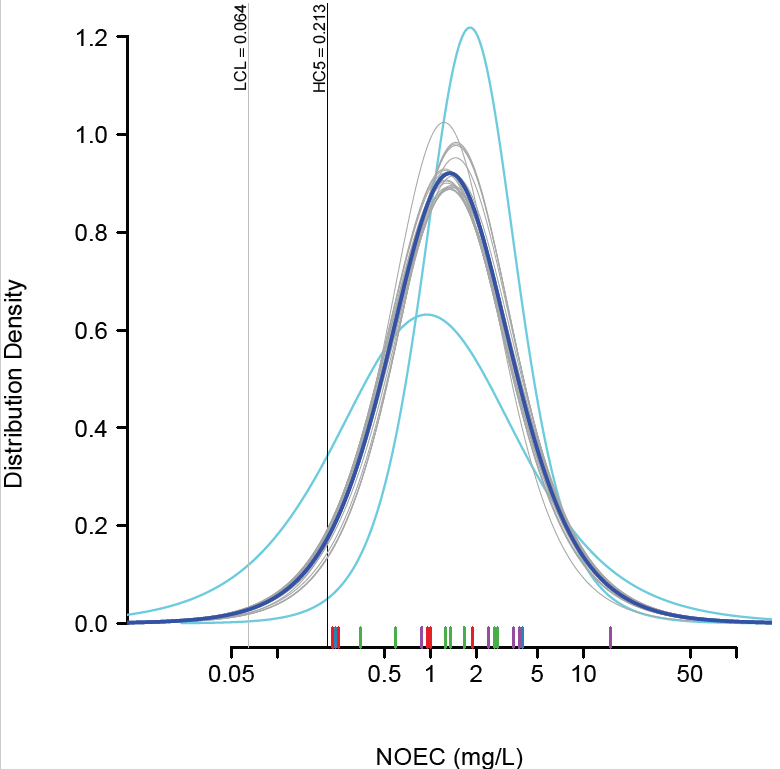
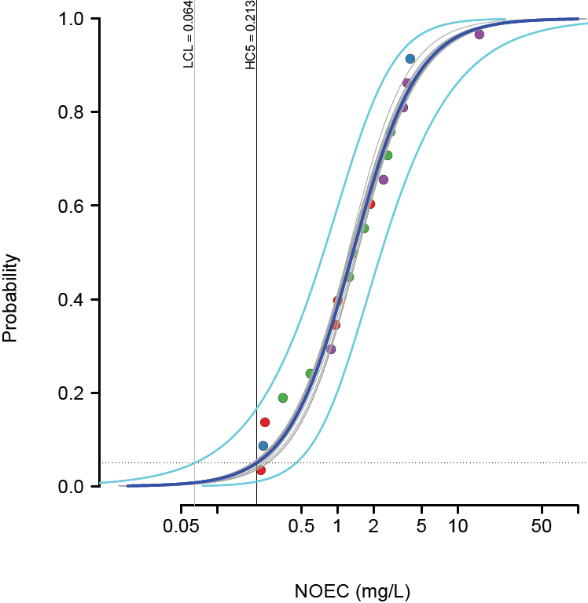
SSD cumulative toxicity distribution (left) and log-logistic distribution (right) for the simulated case example. The HC5 and lower 95%ile confidence limit of the HC5 are given in solid vertical lines. The overall 95%ile confidence limits for the distribution are given in light blue in both graphs

These follow outputs will be provided only if selected during the “Analysis options” step.

**Leave-One-Out**

In the leave-one-out simulations, an individual species is removed from the dataset and the statistical distributions are recreated. In the table below, each row represents the species that is being *removed* from analysis. The resulting HCx, 95% confidence interval, mean, and standard deviation are then calculated. In the simulated example below, the response values ranged from 0.23-15.06 mg/L. Removal of one of the four most sensitive taxa *increased* the HC5.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Species.Out | Responses (mg/L) | X | HCx (mg/L) | LowerCL (mg/L) | UpperCL (mg/L) |
| NONE |  | 0.05 | 0.221 | 0.078 | 0.425 |
| Onchorhychus mykiss | 0.23 | 0.05 | 0.266 | 0.094 | 0.503 |
| Lemna minor | 0.24 | 0.05 | 0.264 | 0.093 | 0.500 |
| Tilapia mossambica | 0.25 | 0.05 | 0.261 | 0.092 | 0.497 |
| Corbicula fluminea | 0.35 | 0.05 | 0.245 | 0.084 | 0.474 |
| Ceriodaphnia dubia | 0.59 | 0.05 | 0.227 | 0.075 | 0.446 |
| Microcystis aeruginosa | 0.88 | 0.05 | 0.217 | 0.071 | 0.431 |
| Pimephales promelas | 0.96 | 0.05 | 0.215 | 0.070 | 0.428 |
| Lepomis macrochirus | 1 | 0.05 | 0.215 | 0.070 | 0.427 |
| Hyallela azteca | 1.25 | 0.05 | 0.211 | 0.069 | 0.421 |
| Daphnia magna | 1.35 | 0.05 | 0.210 | 0.069 | 0.419 |
| Brachionus calyciflorus | 1.66 | 0.05 | 0.209 | 0.068 | 0.415 |
| Poecelia reticulata | 1.88 | 0.05 | 0.208 | 0.068 | 0.413 |
| Desmodesmus subspicatus | 2.4 | 0.05 | 0.207 | 0.068 | 0.410 |
| P. parthenogenica | 2.62 | 0.05 | 0.207 | 0.068 | 0.409 |
| Chironomus riparius | 2.75 | 0.05 | 0.207 | 0.069 | 0.409 |
| Chlorella kessleri | 3.5 | 0.05 | 0.208 | 0.070 | 0.408 |
| Elimia sp | 3.81 | 0.05 | 0.209 | 0.070 | 0.408 |
| Elodea canadensis | 4 | 0.05 | 0.209 | 0.071 | 0.408 |
| R. subcapitata | 15.06 | 0.05 | 0.243 | 0.094 | 0.436 |



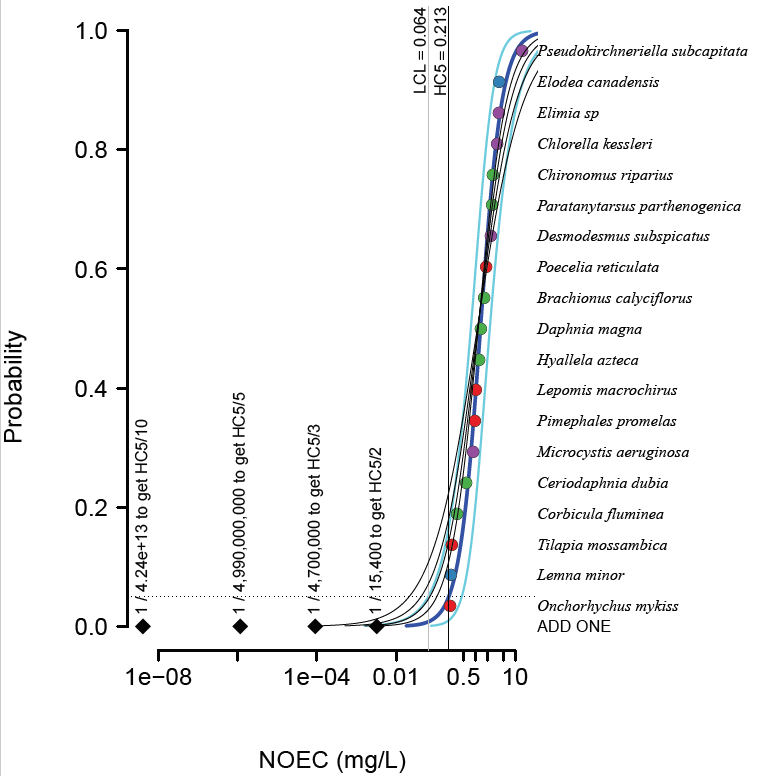
The resulting distributions are plotting graphically in light gray. These distributions are overlaid behind the original SSD cumulative toxicity distribution (left) and log-logistic distribution (right).

**Add-One-In**

In the add-one-in simulation, hypothetical data values are calculated that if available and added to the SSD input would decrease the HC5 by a factor of 2, 3, and 5, respectively. It is also instructive to determine the relative ratio of the existing most sensitive toxicity value to the new hypothetical values.

From add-one-in simulation below, it is readily apparent that adding additional species has relatively little influence on reducing the SSD. To reduce the HC5 by a factor of 2 (HC5 =0.106), a new toxicity value of 0.0032 mg/L would need to be generated. The chance of this value being generated, given the assumed log-logistic distribution of the initial 19 ecotoxicity data points, is 1 in 15,400.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| distribution | target | targetValue (mg/L) | addOneValue (mg/L) | addOneProb | oneIn | oneInFormated |
| logis | HC5/2 | 0.106 | 0.003233 | 6.498E-05 | 15,400 | 1 / 15,400 |
| logis | HC5/3 | 0.071 | 0.000090 | 2.130E-07 | 4,700,000 | 1 / 4,700,000 |
| logis | HC5/5 | 0.043 | 0.00000116 | 2.003E-10 | 4,990,000,000 | 1 / 4,990,000,000 |
| logis | HC5/10 | 0.021 | 0.0000000041 | 2.356E-14 | 4.24e+13 | 1 / 4.24e+13 |



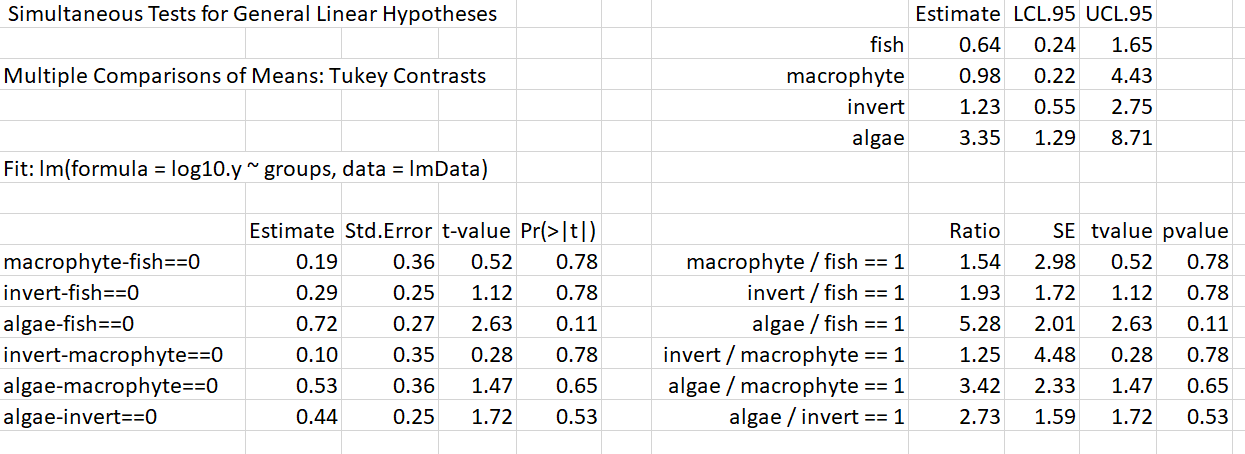
Graphically, the location of these new hypothetical data points are plotted as black diamonds.

**Grouping:**

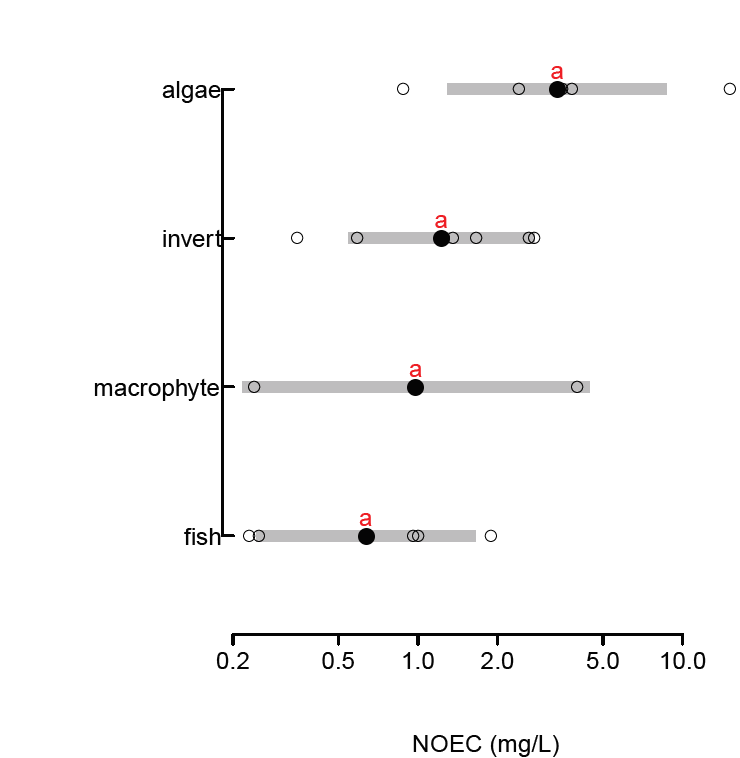
This analysis can be useful to determine if there are substantial patterns of differences in sensitivity with respect to trophic level/taxonomic groups. If one trophic level/taxon is consistently the most sensitive to the compound of interest, an SSD containing only members of this trophic level/taxon may be warranted. An ANOVA analysis is used to determine if there is a significant difference between group means. These results are displayed in the Excel output. In the example below, the significance value is 0.11 (p > 0.05) thus we can conclude that there is no significant difference between the group means.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Analysis of Variance Table | | |  |  |  |
| Response: log10.y | |  |  |  |  |
|  | Df | SumSq | MeanSq | F-value | Pr(>F) |
| groups | 3 | 1.36 | 0.45 | 2.41 | 0.11 |
| Residuals | 15 | 2.83 | 0.19 |  |  |

In the case where a significant difference between the group means is present (i.e., p ≤0.05), a Tukey Contrasts post-hoc test is used to compare all taxonomic groups. For each pair of groups, the difference between group means, the standard error of the difference, and the significance level of that difference are given in the MS Excel output under the “Group Analysis” tab. This analysis is reported both as the difference between two groups and as the ratio between groups. In the example below, all pairwise comparisons are not statistically significant (p>0.05).



Group comparisons are also displayed graphically. Results are displayed by their group (y-axis). Individual experimental results are depicted by open dots. The bold dot represents the mean value for each group. The gray shaded area represents the 95% confidence interval. The letter above the bold dot represents which groups are similar (or not statistically significantly different) to one another. In this example, there is no significant difference between the groups.



**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

**Citations:**

Belanger, S. E., M. Barron, P. Craig, S. Dyer, M. Galay-Burgos, M. Hamer, S. Marshall, L. Posthuma, S. Raimondo and P. Whitehouse. 2017. Future needs and recommendations in the development of Species Sensitivity Distributions: Estimating toxicity thresholds for aquatic ecological communities and assessing impacts of chemical exposures. Integrated Environmental Assessment and Management, 13(4):664-667.

Belanger, S.E., G. J. Carr. 2019. SSDs Revisited: Part II – Practical considerations in the development and use of application factors applied to Species Sensitivity Distribution. Environmental Toxicology and Chemistry, 38(7):1526-1541.

Carr, G. J., S. E. Belanger. 2019. SSDs Revisited: Part I – A framework for sample size guidance on Species Sensitivity Distribution Analysis. Environmental Toxicology and Chemistry, 38(7):1514-1525.

ECHA (European Chemicals Agency). 2008. Guidance on information requirements and chemical safety assessment Chapter R.10: Characterization of dose [exposure]‐response for environment. May, 2008. 65 p. Downloaded from: <http://echa.europa.eu/reach_en.asp>.