

Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity: User Manual

Version 3.3



<http://www3.epa.gov/webice/>

Sandy Raimondo, Crystal R. Lilavois, and Mace G. Barron

U.S Environmental Protection Agency
Office of Research and Development
National Health and Environmental Effects Research Laboratory
Gulf Ecology Division
Gulf Breeze, FL 32561

Reference Web-ICE as:

Raimondo, S., C.R. Lilavois, and M.G. Barron. 2015. Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity: User Manual. Version 3.3, EPA/600/R-15/192, U. S. Environmental Protection Agency, Office of Research and Development, Gulf Ecology Division. Gulf Breeze, FL.

Disclaimer:

The information in this document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Approval does not signify that the content reflects the views of the Agency, nor does mention of trade names or products constitute endorsement or recommendation for use.

Note about Version Control:

Web-ICE updates may vary among versions by website structure and function, models resulting from updated databases, or both. A change in the first integer in the website version signifies an architectural change of the site. A change in the second integer represents a change to models in one or more modules. See Appendix for changes.

Contents

Contents.....	2
Introduction	3
Model Development	4
I. Toxicity Databases	4
II. Model Development.....	5
III. Model Validation	5
Using the Web-ICE Program.....	6
I. Working with Web-ICE Aquatic or Wildlife Modules.....	7
Selecting Model Taxa	7
Estimating Toxicity	8
II. The Species Sensitivity Distribution (SSD) Module	10
III. The Endangered Species Module.....	13
Producing a Endangered Species Toxicity Report.....	13
IV. Accessing Model Data & Chemical Information.....	15
Mode of Action (MOA)-specific models.....	15
Guidance for Model Selection and Use	16
I. Statistical Definitions	16
II. Selecting a Model with Low Uncertainty	17
Rules of Thumb.....	17
Surrogate Species Selection: An Example	18
III. Evaluating Model Predictions	18
IV. Selecting Predicted Toxicity Values for SSDs	18
V. Applying Web-ICE in Ecological Risk Assessment (ERA).....	19
Acknowledgements	20
References.....	21
Appendix 1. Number of Models by Version	23

Introduction

Information on the acute toxicity to multiple species is needed for the assessment of the risks to, and the protection of, individuals, populations, and ecological communities. However, toxicity data are limited for the majority of species, while standard test species are generally data rich. To address data gaps in species sensitivity, the Interspecies Correlation Estimations (ICE) application was developed by the U.S. Environmental Protection Agency (US EPA) and collaborators to extrapolate acute toxicity to taxa with little or no acute toxicity data for a chemical of interest, including threatened and endangered species (Asfaw et al. 2003). Web-based Interspecies Correlation Estimation (Web-ICE) provides interspecies extrapolation models for acute toxicity in a user-friendly internet platform.

ICE models estimate the acute toxicity (LC50/LD50) of a chemical to a species, genus, or family with no test data (the predicted taxon) from the known toxicity of the chemical to a species with test data (the commonly tested surrogate species). ICE models are least square regressions of the relationship between surrogate and predicted taxon based on a database of acute toxicity values: median effect or lethal water concentrations for aquatic species (EC/LC50; $\mu\text{g/L}$) and median lethal oral doses for wildlife species (LD50; mg/kg bodyweight). Experimental or estimated (e.g., quantitative structure activity relationship; QSAR) acute toxicity for a surrogate species may be used to estimate toxicity when there is an existing ICE model between the surrogate and taxa of interest (e.g., species-species; species-genus; species-family).

In addition to direct toxicity estimation, Web-ICE develops Species Sensitivity Distributions (SSDs) from multiple surrogate and predicted species. SSDs are cumulative distribution functions of toxicity values for multiple species and are used to estimate a hazard level [hazardous concentration (HC) or hazardous dose (HD)] that is protective of most test species (e.g., 95%) by estimating the concentration or dose at a corresponding percentile (e.g., 5th) of the distribution (de Zwart 2002). SSDs generated in Web-ICE are log-logistic cumulative distribution functions of toxicity developed from simultaneously estimated toxicity values to all predicted species available using up to 25 surrogates. The ICE-generated SSD hazard levels have been shown to be within an order of magnitude of measured HC5s (Dyer et al. 2006, Dyer et al. 2008) and HD5s (Awkerman et al. 2008, 2009) and provide additional information for ecological risk assessment.

The Web-ICE Endangered Species module simultaneously estimates toxicity to taxa representing threatened or endangered (listed) species using up to 25 tested surrogates. This module batch processes toxicity values for listed species from all species, genus, and family level models available for the selected listed species or taxa and entered surrogates. The classification of listed species was obtained from the US Fish and Wildlife Service Threatened and Endangered Species module of Environmental Conservation Online System (http://ecos.fws.gov/tess_public; Accessed August 2015), which was linked to Web-ICE species, genus, and family model databases for aquatic organisms (not currently available for algae) and wildlife. Users may predict to all available listed species within a broad taxonomic groups (e.g., Fishes) or a particular species (e.g., Atlantic salmon, *Salmo salar*).

This manual provides step-by-step instructions for using Web-ICE v3.3, as well as information on the databases, model development, model validation, and user guidance on model selection and interpretation. This version differs from v 3.2 in that it includes new and updated models for aquatic vertebrates and invertebrates based on an updated database (Appendix 1). It also provides updated user guidance based on new uncertainty analyses (Willming et al., in preparation). User guidelines outlined in the *Guidance for Model Selection and Use* section of this manual are recommended to ensure high confidence and low uncertainty in model predictions used in risk assessment.

Model Development

I. Toxicity Databases

Separate acute toxicity databases are maintained for aquatic animals (vertebrates and invertebrates), aquatic plants (algae), and wildlife (birds and mammals). Open-ended toxicity values (i.e. > 100 mg/kg or <100 mg/kg) and duplicate records among multiple sources are not included in any of the databases. Attributes for and the number of models developed from each database are listed for the different versions in Appendix 1.

The aquatic animal database is composed of 48 or 96-hr EC/LC50 values based on death or immobility. This database is described in detail in the Aquatic Database Documentation found on the [Download Model Data](#) page of Web-ICE and describes the data sources, normalization, and quality and standardization criteria (e.g., data filters) for data used in the models. Data used in model development adhered to standard acute toxicity test condition requirements of the ASTM International (2015a, and earlier editions) and the US EPA Office of Prevention, Pesticides, and Toxic Substances (e.g., US EPA 1996a).

The algal toxicity database is described in Appendix B of the Aquatic Database Documentation. Algae data are 72 or 96-hr EC50. Validity of each record was evaluated based on coherence to test acceptability criteria found within standard methods guidelines (ASTM International 2015b, OECD 1996, US EPA 1996b). Models derived from this database predict toxicity to a species or genus from a surrogate species or genus. Family level models were not developed for algae because there were limited families that had two or more species, which is a requirement for development of higher taxa models.

The wildlife database includes 96-hr LD50 values for terrestrial birds and mammals collected from the open literature (Hudson et al. 1984; Shafer and Bowles 1985, 2004; Shafer et al. 1983; Smith 1987) and from datasets compiled by governmental agencies of the United States (US EPA) and Canada (Environment Canada; Baril et al. 1994; Mineau et al. 2001). Data were standardized by using only data for adult animals and chemicals of technical grade or formulations with $\geq 90\%$ active ingredient. Models derived from this database predict toxicity to a species or family from a surrogate species. Genus level models were not developed for wildlife

because there were limited genera that had two or more species, which is a requirement for development of higher taxa models.

II. Model Development

Models are only developed for species within the same database (i.e. there are no fish to algae models or algae to bird models). Models are least squares regressions such that:

$$\text{Log}_{10}(\text{predicted toxicity}) = a + b * \text{Log}_{10}(\text{surrogate toxicity})$$

where a and b are the intercept and slope of the line, respectively. Within a database, all species are paired with each other by common chemical. Three or more common chemicals per pair are required to develop a model. Genus and family-level models are similarly developed by pairing each surrogate species or genus (algae only) with each genus or family by common chemical. A genus or family requires unique toxicity values for two or more species within the taxon. In cases where a surrogate species is compared to its own genus or family, toxicity values of the surrogate are excluded from the values used to represent the higher taxonomic level. For species, genus, and family models in the primary **Web-ICE Calculator Modules** (See *Working with Web-ICE Aquatic or Wildlife Modules*), multiple toxicity values for the same species and chemical are represented by the geometric mean (Raimondo et al 2007, 2010). For genus and family models in the aquatic **Endangered Species module**, multiple toxicity values for the same species and chemical are represented by the most sensitive toxicity value for that chemical within the taxa (Willming et al. in review). Only models with a significant relationship ($p\text{-value} \leq 0.05$) are included in Web-ICE.

III. Model Validation

The uncertainty of each model is assessed using leave-one-out cross-validation. In this method, each pair of acute toxicity values for surrogate and predicted taxa are systematically removed from the original model (Raimondo et al. 2010). The remaining data are used to rebuild a model and estimate the toxicity value of the removed predicted taxa toxicity value from the respective surrogate toxicity value. This method is only used for models developed using 4 or more data points. To maintain uniformity among the large number of models contained within Web-ICE, the “N-fold” difference of each estimated and actual value is used to determine the accuracy of the estimated toxicity value. For aquatic species, inter-laboratory variation of acute toxicity test data for a given species and chemical can be as great as a 5-fold difference (Fairbrother 2008). For wildlife species, the average variability of toxicity measurements for a specific chemical and species is between 4.0 and 6.4-fold (Raimondo et al. 2007). Thus, a 5-fold difference is considered a good fit of predicted ICE values.

The cross-validation success rate for each model is the proportion of removed data points that are predicted within 5-fold of the actual value. If the removal of a xy data pair results in a model that is not significant at the $p < 0.05$ level, the replicate is not included in calculating the cross-validation success rate. This is typically only the case for models with low degrees of freedom (<8) and a p-value between 0.01 and 0.05 in the original model. There is a strong relationship between taxonomic distance and cross-validation success rate, with uncertainty increasing with larger taxonomic distance (Raimondo et al. 2007, 2010). For the models in this version, aquatic fish and invertebrates, models predict within 5-fold and 10-fold of the actual value with 92 and 98% certainty for surrogate and predicted taxa within the same family, and with 87 and 97% certainty within the same order (Raimondo, unpublished data). In wildlife species, models predict within 5-fold and 10-fold of the actual value with 90 and 97% certainty for surrogate and predicted taxa within the same order (Raimondo et al. 2007). Uncertainty analysis of algal ICE models is ongoing.

Using the Web-ICE Program

The **Web-ICE Modules** contain models that predict single acute toxicity values to aquatic vertebrate and invertebrate species, genera, and families; aquatic algae species and genera; and wildlife (terrestrial birds and mammals) species and families. The **Species Sensitivity Distribution Modules** use models for either terrestrial wildlife species or aquatic species. The aquatic module can combine toxicity values for vertebrates, invertebrates and algae. These modules batch process species-level toxicity from all entered surrogates to develop a cumulative probability distribution of toxicity data and generate a prescribed hazard level. The **Endangered Species Modules** predict multiple toxicity values to represent listed species using all available species, genus, or family level models for the entered surrogates. Modules are accessible either from the home page or from the blue navigation bar along the left side of the page (Figure 1).

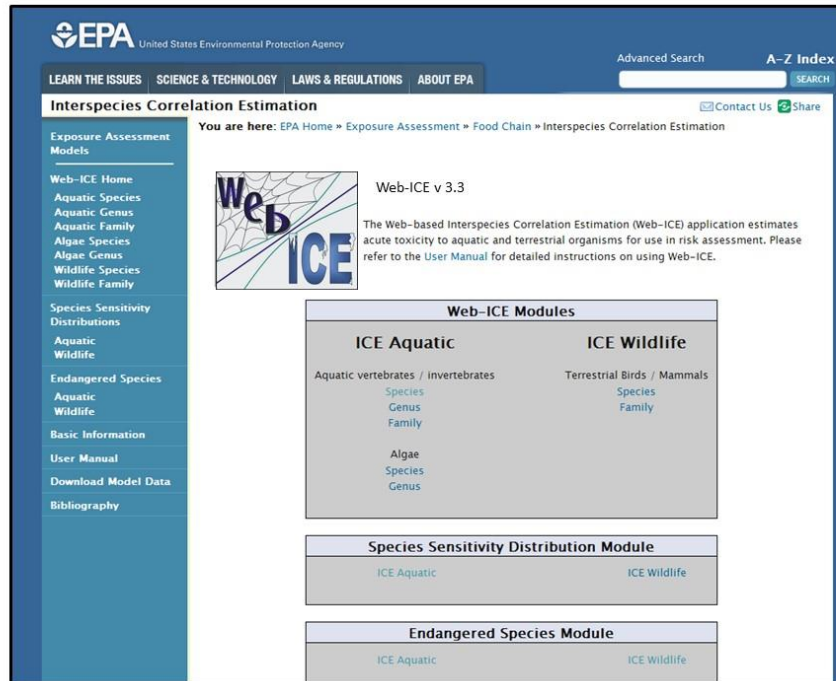


Figure 1. Home page of Web-ICE program.

I. Working with Web-ICE Aquatic or Wildlife Modules

Selecting Model Taxa

1. From either the home page or the blue navigation bar, click the link for the module with which you will be working (Aquatic species, genus, or family; Algae species or genus; Wildlife species or family).
2. You will then be directed to a **Taxa Selection Page** (Figure 2) which will allow you to select your surrogate and predicted taxa for the model you want to use.
3. You may search for your surrogate and predicted taxa by either common name or scientific name by selecting the appropriate option in the **Sort by:** drop down menu. The default is set to common name (NOTE: Algae modules contain scientific names only).
4. From the drop down menus, select the surrogate species and predicted taxon. It does not matter which you select first; however, the second choice is limited to the models available for the taxon chosen first.
5. To change any of your selections, press **Reset** and start again.
6. Click **Continue** to be directed to the calculator page for toxicity estimation.

If there is not a model for your predicted species of interest, you will need to use a genus or family-level model to predict toxicity. The available models may be determined by browsing through the genus and family level modules, or by searching through the spreadsheets of model information available through the **Download Model Data** option on the blue navigation bar. The downloadable Microsoft Excel® spreadsheets provided for each Web-ICE module may be sorted by surrogate species or predicted taxa to identify available models.

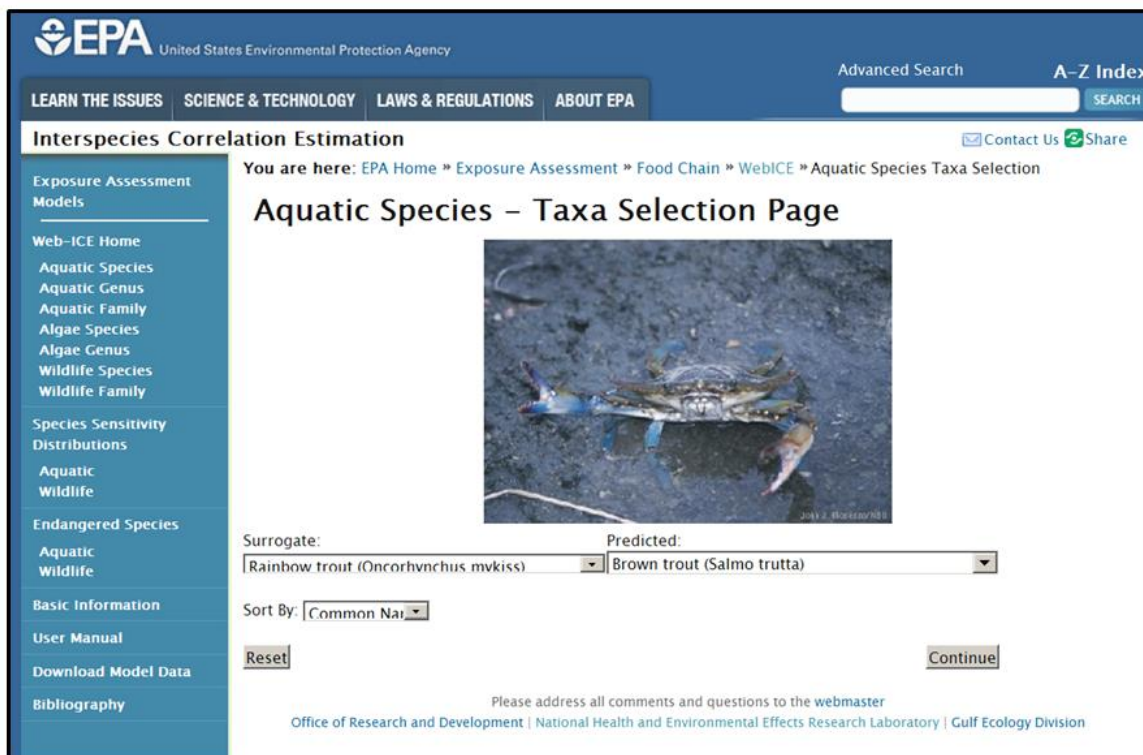


Figure 2. Taxa selection page for aquatic species.

Estimating Toxicity

The surrogate and predicted taxa selected from the previous page are listed at the top of a calculator page (Figure 3). This page is divided into four parts: input (Figure 3A), calculated results (Figure 3B), model statistics (Figure 3C), and model graphic (Figure 3D; not available for Algae). Please refer to the *Statistical Definitions* section of this manual for more information on model statistics. The graph shows the data (e.g., $\log_{10}(\text{LC}_{50})$ values) used to develop the model, the regression line (straight inner line), and 95% confidence intervals (curved outer lines). The surrogate and predicted taxa are labeled on the X and Y axes, respectively.

1. Enter the surrogate toxicity value in the box located under **Surrogate Acute Toxicity**.

2. Select your desired confidence interval (90, 95, or 99%) from the drop down menu located under **Select Confidence Interval** (Default is 95%).
3. Press **Calculate**
4. The calculated values will appear in the three boxes labeled **Predicted Acute Toxicity, Lower Limit** and **Upper limit**.
5. Log-transformed values of the surrogate and predicted toxicity values appear in parentheses in their respective boxes.
6. If the entered surrogate toxicity value is outside the range of values used to develop the model, a pop-up with the warning **“This value is outside the x-axis range for this model. Continue?”** will appear. The user may select **“OK”** to proceed to calculate the toxicity value or hit cancel to enter another value.
7. To select a different model, select the link to the desired module in the blue navigation bar.

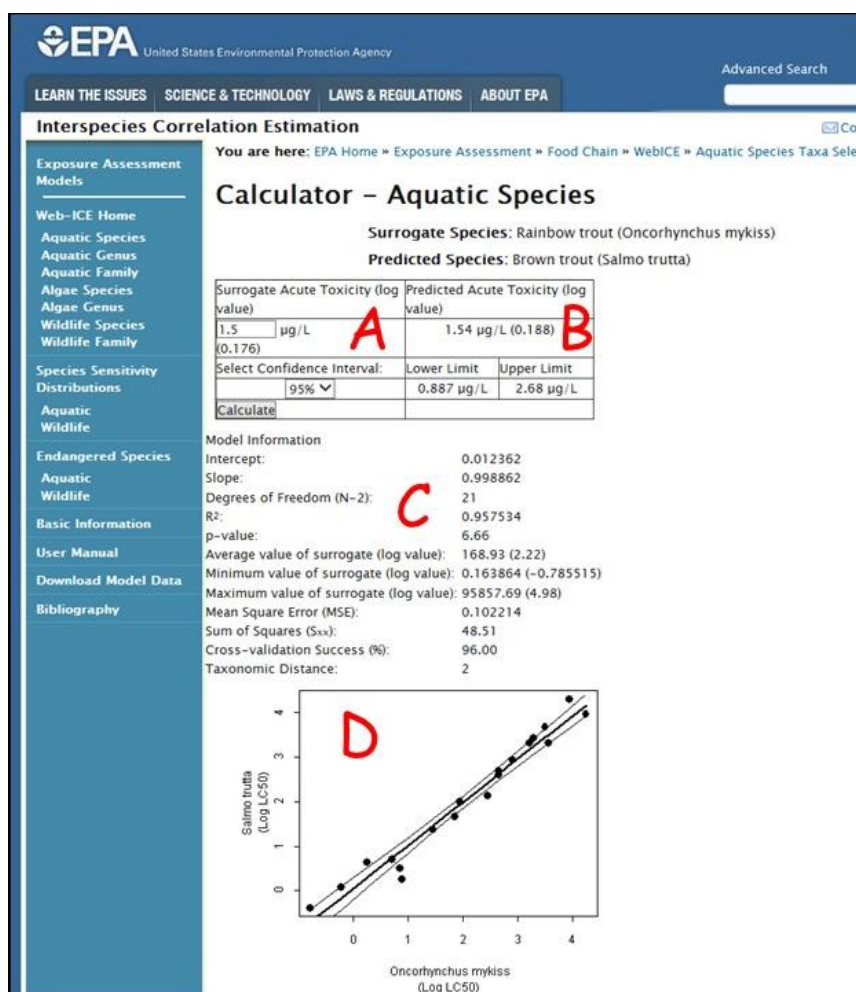


Figure 3. Calculator Page. A. Toxicity input box, B. Predicted toxicity output, C. Model information, D. Model graph.

II. The Species Sensitivity Distribution (SSD) Module

The SSD modules generate SSDs from Web-ICE toxicity values estimated from one or more surrogate species (up to 25), which simultaneously estimate toxicity to all possible predicted species with existing Web-ICE models. The SSD is initially generated using all estimated toxicity values and the entered toxicity values of the surrogate species. If multiple surrogates are used and a predicted value is estimated for one of the surrogate species, Web-ICE uses the entered value for that species and excludes the predicted value(s) from the SSD. If more than one surrogate predicts a toxicity value to the same species, Web-ICE lists all predicted values with the default noted by the radio button. The default is the predicted value with the smallest confidence interval, but the user may select a different value, as described below.

Web-ICE uses the SSD described by the logistic distribution function of de Zwart (2002):

$$F(C) = 1/(1 + \exp ((\alpha - C) / \beta))$$

The \log_{10} -transformed environmental concentration (or dose) of the evaluated chemical is represented by C, the parameter α is the sample mean of the \log_{10} -transformed toxicity values and β is defined as $\sqrt{3/\pi} * \sigma$, where σ is the standard deviation of the \log_{10} -transformed toxicity values. The HC/HD level is the percentile of interest (e.g., 5th) of the described distribution.

Corresponding SSDs are also developed from the upper and lower confidence limits of the predicted toxicity values and are used to calculate the upper and lower bounds of the HC/HD value at a given percentile¹. For example, the lower bound of the HC5 is calculated as the 5th percentile of the SSD developed from the estimated lower confidence limit of each predicted toxicity value. Similarly, the upper bound of an HC5 is calculated as the 5th percentile of the SSD developed from the estimated upper limit of each predicted toxicity value.

Generating an SSD:

1. Under the SSD module, select either Aquatic or Wildlife.
2. On the SSD taxa selection page, select your surrogate species from the drop down menu and click **Add** to add the species as a surrogate. There are separate drop down menus for vertebrates/invertebrates and algae; however if both are selected they will be combined into the same SSD. A maximum of 25 total surrogates can be selected (Figure 4).

¹ It should be noted that the standard approach to calculate confidence intervals of the logistic distribution (de Zwart 2002) cannot be computed using the java script based platform of Web-ICE. The approach applied here uses the uncertainty in model predictions to estimate the uncertainty of the HC5 estimation. Large or unrealistic confidence intervals may result if toxicity values with exceptionally large confidence intervals are included in the SSD. If this occurs, toxicity values should be re-evaluated based on the *Guidance for Model Selection and Use* in this manual.

3. To remove a surrogate species from the list after it is added, click **Remove** next to the species name.
4. Enter the known toxicity for the surrogate species, click **Calculate SSD**.

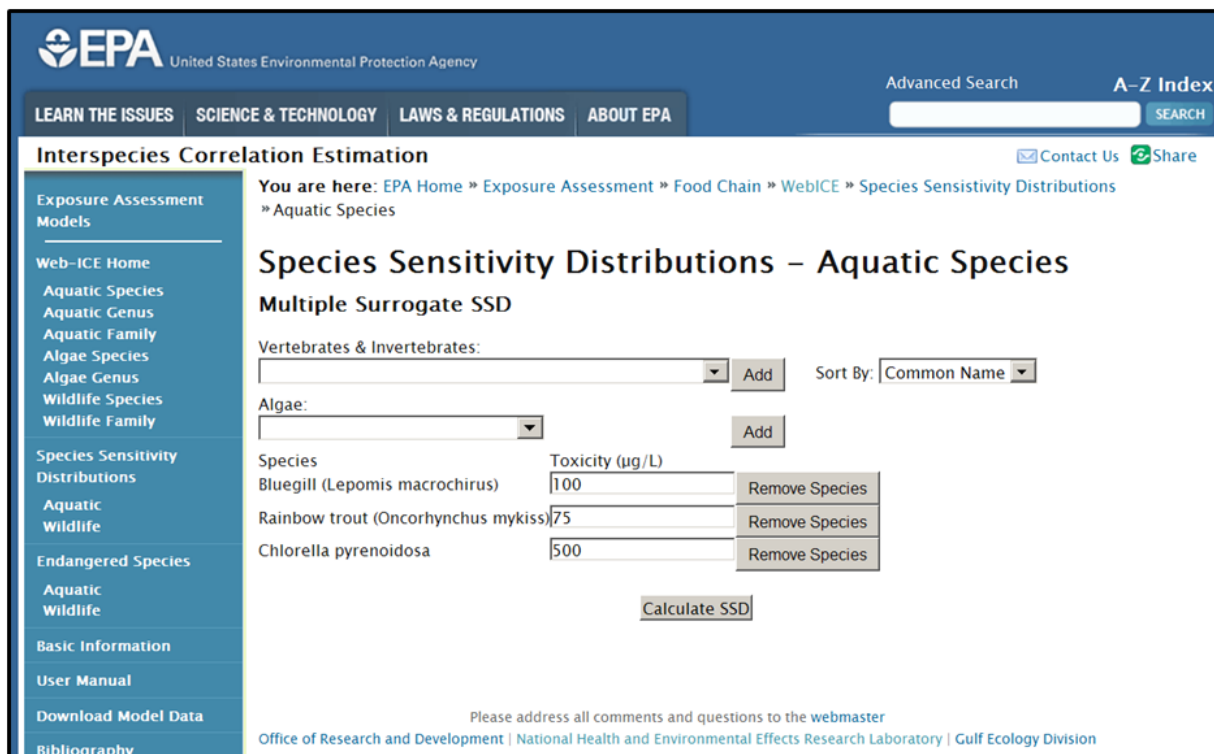


Figure 4. SSD taxa selection page.

Working with the SSD Output

1. On the SSD output page, the HC/HD level (e.g., 1st percentile, 5th percentile) may be changed from the drop down box. The hazard level is automatically recalculated if the level is changed. The default is the HC/HD5 (Figure 5A).
2. If multiple surrogates predict to the same species, all predicted values are shown, but only one can be included in the SSD. By default, Web-ICE includes the one with the smallest confidence interval, but this can be changed by the user by selecting the radio button of the desired value (Figure 5B). The HC/HD value is automatically recalculated.
3. The warning “**Input toxicity is greater (less) than model maximum (minimum)**” indicates if a predicted value was generated from a surrogate species toxicity value that was outside the range of toxicity values used to generate that model.

- The user can unmark the box to the left of a predicted species to exclude it entirely from the SSD, which is automatically recalculated. (NOTE: See *Selecting Predicted Toxicity Values for SSDs* in the *Guidance for Model Selection and Use* section below for guidance on removing estimated toxicity values).
- The estimated toxicity values may be sorted by a column of interest by selecting the **sort** tab below the heading.
- Predicted values can be filtered by inputting desired ranges for the lower and upper bounds for degrees of freedom, R^2 , p-value, mean square error, cross-validation success rate, taxonomic distance, slope, or intercept in the **Data Filters** box (Figure 5C). Open ended ranges are allowed by only inputting a lower or upper limit.
- The user can generate an Excel-friendly output for either all predicted toxicity values or just the data selected by the radio buttons for inclusion in the SSD by selecting the desired **Provide Copy-friendly Output** tab (Figure 5D).

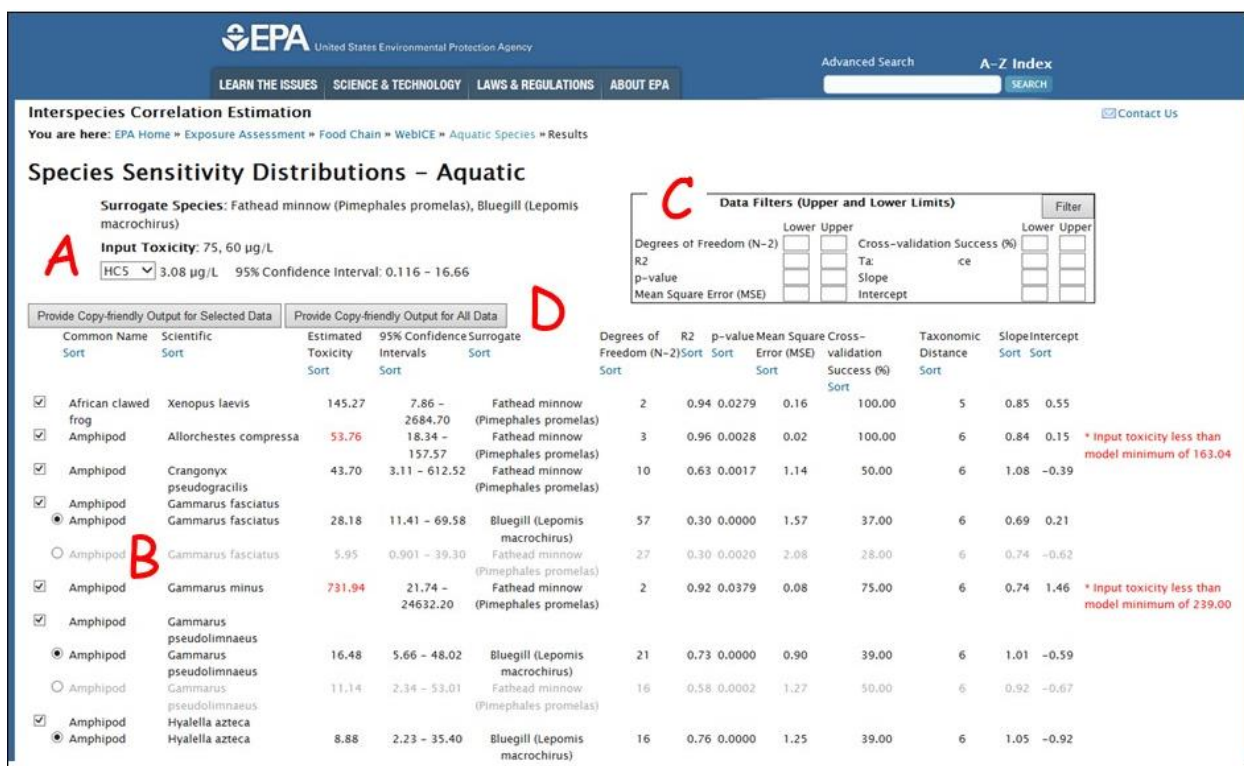


Figure 5. SSD output page. A. Surrogate species input values and hazard level, B. Predicted toxicity values, C. Model data filters, D. Excel-friendly output options.

III. The Endangered Species Module

Producing an Endangered Species Toxicity Report

1. Under the **Endangered Species** module, select either Aquatic or Wildlife.
2. On the taxa selection page, select either the broad taxa of interest using the radio buttons (e.g., all species, fishes) or a particular species of interest from the drop down menu (Figure 6).
3. Select your surrogate species from the drop down menu and click **Add** to add the species as a surrogate. A maximum of 25 species can be selected.
4. To remove a surrogate species from the list after it is added, click **Remove** next to the species name.
5. Enter the toxicity for the surrogate species, click **Calculate**.
6. The output page provides the estimated toxicity for each predicted taxa, the model level (e.g., species), surrogate, and model information (Figure 7).
7. The user may sort the ICE-estimated toxicity values by each column by selecting the **sort** tab below the column heading.
8. Predicted values can be filtered by inputting desired ranges for the lower and upper bounds for degrees of freedom, R^2 , p-value, mean square error, cross-validation success rate, taxonomic distance, slope, or intercept in the **Data Filters** box. Open ended ranges are allowed by only inputting a lower or upper limit.
9. The user can generate an Excel-friendly output by clicking on the **Provide Copy-friendly Output** option.

Figure 6. Taxa selection page for listed species.

Predicted Taxa Sort	Model Level Sort	Surrogate Sort	Estimated Toxicity Sort	95% Confidence Intervals Sort	df (N-2) Sort	R2 Sort	p-value Sort	Mean Square Error (MSE) Sort	Cross-validation Success (%) Sort	Tax. Dist. Sort	Slope Sort	Intercept Sort
Shortnose sturgeon (Acipenser brevirostrum)	species	Fathead minnow (Pimephales promelas)	4.70	0.249 - 88.86	3	0.92	0.0094	0.30	40.00	4	1.24	-1.43
Leon springs pupfish (Cyprinodon bovinus)	species	Fathead minnow (Pimephales promelas)	63.77	37.43 - 108.64	2	0.99	0.0043	0.00	100.00	4	0.67	0.65
Spotfin Chub (Erimonax monachus)	species	Fathead minnow (Pimephales promelas)	12.63	1.37 - 116.41	3	0.83	0.0292	0.30	60.00	2	1.00	-0.60
Fountain darter (Etheostoma fonticola)	species	Fathead minnow (Pimephales promelas)	16.12	5.58 - 46.55	1	0.99	0.0180	0.00	na	4	0.96	-0.42
Bonytail (Gila elegans)	species	Fathead minnow (Pimephales promelas)	60.40	10.62 - 343.39	4	0.74	0.0275	0.24	67.00	2	0.69	0.59
Cape Fear shiner (Notropis mekistocholas)	species	Fathead minnow (Pimephales promelas)	19.47	1.70 - 222.15	3	0.78	0.0440	0.36	60.00	2	0.93	-0.30
Cutthroat trout (Oncorhynchus clarkii)	species	Fathead minnow (Pimephales promelas)	15.49	6.70 - 35.79	24	0.79	1.02	0.39	81.00	4	0.94	-0.40
Apache trout (Oncorhynchus gilae)	species	Fathead minnow (Pimephales promelas)	11.63	1.18 - 114.08	3	0.79	0.0437	0.31	60.00	4	0.88	-0.42

Figure 7. Listed species predicted toxicity report

IV. Accessing Model Data & Chemical Information

A list of chemicals in the aquatic and wildlife databases is available for download using the **Chemicals in Aquatic** and **Chemicals in Wildlife** links on the **Download Model Data** option on the blue navigation bar. In the **Chemicals in Aquatic** file, the chemical CAS number and associated toxicity values used in each model are provided. The **Chemicals in Wildlife** file contains the number of species present for each chemical. The acute data used to develop the ICE models for wildlife and algae are not available due to proprietary rights of some information.

Models for all Web-ICE aquatic and wildlife modules are available as a downloadable Microsoft Excel® spreadsheet on the **Download Model Data** page. The data spreadsheets include model parameters (R^2 , p-value, df, intercept, slope, standard error of the slope, Sxx, and MSE), general model information (taxonomic distance, cross-validation success rate), descriptive statistics (average, minimum, and maximum values of the surrogate species), and critical t-values used to calculate 90, 95, and 99% confidence intervals (t90, t95, t99). These spreadsheets provide all of the information that is needed to generate Web-ICE toxicity estimates and confidence intervals, as well as facilitate the selection of the most robust models and predictions.

Using model data provided, users may calculate toxicity as:

$$\text{Predicted toxicity} = 10^{(\text{intercept} + \text{slope} * \log_{10}(\text{surrogate toxicity}))}$$

And confidence intervals as:

$$\begin{aligned} \text{Lower bound} &= 10^{(\log_{10}(\text{predicted}) - t_{1-\alpha} * \sqrt{[\text{MSE} * (1/n + (\log_{10}(x) - x.\text{ave})^2 / \text{Sxx})]})} \\ \text{Upper bound} &= 10^{(\log_{10}(\text{predicted}) + t_{1-\alpha} * \sqrt{[\text{MSE} * (1/n + (\log_{10}(x) - x.\text{ave})^2 / \text{Sxx})]})} \end{aligned}$$

Where x is the untransformed value of surrogate toxicity, x.ave is the average value of log-transformed surrogate toxicity values, Sxx is the sum of squared errors of the surrogate, MSE is the mean square error, n is the sample size, and $t_{1-\alpha}$ is the value of the t distribution corresponding to the desired level of confidence (i.e., 90, 95, 99%).

Mode of Action (MOA)-specific models

ICE models have been developed using chemicals of just one MOA and are provided on the **Download Model Data** page. These models may be used to improve predictions of models with large taxonomic distance, but may offer limited improvement of predictions for species pairs that are closely related (Raimondo et al. 2010). The suite of MOA-specific models differs from models developed using all MOAs and may include some models for species pairs that were not significant using all data, or may not include models for species pairs that were developed using all chemicals. Currently, MOA-specific models are not currently accessible with the Web-ICE user interface, limiting the use of these models to calculations performed from data in the spreadsheets external to Web-ICE.

Guidance for Model Selection and Use

I. Statistical Definitions

Several statistics are provided with the models and may be used to evaluate the accuracy and precision of the estimated value. The following provides basic definitions of model statistics:

Intercept - The \log_{10} value of the predicted taxon toxicity when the \log_{10} of the surrogate species toxicity is zero.

Slope - The regression coefficient represents the change in \log_{10} value of the predicted taxon toxicity for every change in \log_{10} value of the surrogate species toxicity.

Degrees of Freedom (df, N - 2) - The number of data points used to build the model minus two. The df is related to statistical power; in general, the higher the df, the more information is used to develop the model.

R² - The proportion of the data variance that is explained by the model. The closer the R² value is to one, the more robust the model is in describing the relationship between the predicted and surrogate taxa.

p-value - The significance level of the linear association and the probability that the linear association was a result of random data. Models with lower p-values are more significant. Model p-values of < 0.0001 are reported as 0.00000.

Average value of the surrogate - The average of all toxicity values for the surrogate species used to develop the model. The first number is the actual value and the number in parentheses is the log-transformed value.

Minimum value of the surrogate - The lowest toxicity value for the surrogate species used to develop the model. The first number is the actual value and the number in parentheses is the log-transformed value.

Maximum value of the surrogate - The largest toxicity value for the surrogate species used to develop the model. The first number is the actual value and the number in parentheses is the log-transformed value.

Mean Square Error (MSE) - An unbiased estimator of the variance of the regression line.

Sum of Squares (Sxx) - Sum of squared errors of the surrogate.

Cross-validation Success - The percentage of removed data points that were predicted within 5-fold of the actual value in the leave-one-out cross-validation. Models with a cross-validation success of “na” are those that either had $df = 1$ or where no significant models were developed when data points were removed. Higher cross-validation success is an indication of greater model robustness.

Taxonomic Distance - The taxonomic relationship between the surrogate and predicted taxa. Two taxa within the same genus have taxonomic distance of 1; within the same family = 2; within the same order = 3; within the same class = 4; within the same phylum = 5; within the same kingdom = 6; across kingdoms = 7 (algal models only, plants vs. cyanobacteria).

II. Selecting a Model with Low Uncertainty

Rules of Thumb

Model attributes, such as taxonomic distance of the predicted and surrogate species, model parameters, and cross-validation success rate, should be used to select models with low uncertainty. The following criteria should be used as a guide to select more accurate predictions. These values are intended to be for guidance purposes only; predicted values should be evaluated holistically using best professional judgement.

1. Relatively low mean square error (MSE) (≤ 0.95)
2. High R^2 value (> 0.6)
3. High slope (≥ 0.6)
4. Close taxonomic distance (≤ 4)
5. Narrow confidence intervals (one order of magnitude between lower and upper limit)

The best estimations generally occur for surrogate and predicted taxa that are within the same genus, family, or order and with low MSE (Raimondo et al. 2007, 2010). The cross-validation success rate is a conservative estimate of model uncertainty and should not be interpreted as an exact estimate of model error. Additionally cross-validation success rate may be misleading for models of small sample sizes due to potentially significant changes in models rebuilt when data points are removed, and the exclusion validated points predicted from rebuilt models that were not statistically significant. However, this metric may be used as weight of evidence to identify robust models.

Surrogate Species Selection: An Example

In an example of how to select a suitable model, Raimondo et al. (2007) outlined a selection procedure to find an appropriate surrogate species for estimating the toxicity of a chemical to red-winged blackbird. In the example, toxicity for the chemical of interest was available for northern bobwhite, mallard, Japanese quail, fulvous whistling duck, common grackle, and house sparrow, making them all potential surrogates. The common grackle and house sparrow have the closest taxonomic distance (2, same family; 3, same order); the other potential surrogates in this example have a taxonomic distance of 4 (same class). Of the grackle and house sparrow, both have similar MSE (~ 0.13), however house sparrow has a higher model R^2 (0.84), higher cross-validation success rate (95%), and greater degrees of freedom (107) and is the best surrogate for red-winged blackbird in this example. The grackle would also provide good surrogacy, with high R^2 (0.65), high cross-validation success rate (93%) and good degrees of freedom (54). If neither of these species were available surrogates, Japanese quail ($R^2 = 0.79$, MSE = 0.15, df = 135, cross-validation success rate = 91%) would be the next best surrogate, followed by northern bobwhite ($R^2 = 0.63$, MSE = 0.23, df = 45, cross-validation success rate = 85%) and mallard ($R^2 = 0.48$, MSE = 0.34, df = 80, cross-validation success rate = 79%).

III. Evaluating Model Predictions

Uncertainty of model predictions may be evaluated by assessing the confidence intervals around the prediction and the value of the input data relative to the data used to generate the model. Uncertainty may be evaluated by reviewing the confidence intervals calculated with the predicted value. Narrow confidence intervals represent higher confidence that the model fits through the range of datapoints for the entered surrogate species toxicity. If the surrogate toxicity value entered into an ICE model is outside the range of surrogate toxicity data used to generate the model, the warning **“This value is outside the x-axis range for this model. Continue?”** will appear to alert the user. This warning alone does not indicate low confidence in the model estimate, but should be used in conjunction with the calculated confidence intervals to evaluate the model prediction. For example, if the upper and lower bounds of the confidence interval are several orders of magnitude from the predicted value, caution should be used in applying the ICE estimate in risk assessment.

IV. Selecting Predicted Toxicity Values for SSDs

The SSD modules of Web-ICE automatically predict toxicity values from all available models for the selected surrogate species simultaneously. The user has the discretion to remove predicted toxicity values from the SSD to either customize the SSD for a particular taxa (e.g., birds only, fish only), or to remove predicted toxicity values with large confidence intervals. If an estimated toxicity value was derived from an input value that was outside of the range of surrogate species data used to generate the model from which it was predicted, a warning appears next to the value indicating the

maximum or minimum value of the model. This warning alone does not indicate low confidence in the model estimate, but should be used in conjunction with the calculated confidence intervals to evaluate the model prediction.

V. Applying Web-ICE in Ecological Risk Assessment (ERA)

Web-ICE was developed to support both chemical hazard assessment and ecological risk assessment (ERA) by providing a method to estimate acute toxicity to specific taxa, such as listed species, or to a larger number of taxa (species, genera, families) with known uncertainty. Potential applications of acute toxicity values generated by Web-ICE include the problem formulation phase of an ERA to screen for contaminants of potential concern and in the analysis phase to characterize effects to a larger number of species. The estimation of species-specific toxicity values using Web-ICE is recommended as an alternative to safety factors typically applied when extrapolating toxicity or risks to taxa without chemical and species-specific toxicity data (NRC 2013). Another potential application of the chemical and taxon-specific acute toxicity estimates generated from ICE models includes input into existing exposure and risk models (e.g., TREX; EPA 2005). Web-ICE generated toxicity values may also be used in the analysis of uncertainty and variability in toxicity to ecological receptors in both screening level and baseline or Tier II ERAs.

In addition to taxa-specific ICE models, Web-ICE can be used to generate SSDs and estimated 1st, 5th or 10th percentile values of the cumulative distribution of species-specific toxicity values. These percentile values, expressed as the hazard concentration (e.g., HC5) or hazardous dose (e.g., HD5), provide an estimate of toxicity at a prescribed level of species protection with known uncertainty. Hazard concentrations could be used in ERA in place of species-specific toxicity values or as a component of the uncertainty analysis (Dyer et al. 2008, Awkerman et al. 2008, 2009, Barron et al. 2012).

Acknowledgements

For database development, the authors would like to thank Deborah Vivian (US EPA, GED); Sonny Mayer (US EPA, retired); Thomas Steeger and Brian Montague (US EPA, Office of Pesticide Programs); Don Rodier (US EPA, retired); Pierre Mineau, Alain Baril and Brian Collins (National Wildlife Research Centre, Environment Canada); Chris Russom and Teresa Norberg-King (US EPA, Mid-Continent Ecology Division); Christopher Ingersoll and Ning Wang (Columbia Environmental Research Center, U.S. Geological Survey); Thomas Augspurger (US Fish and Wildlife Service); Scott Dyer, Scott Belinger, and Jessica Brill (Procter and Gamble). Special thanks to Wally Schwab and Derek Lane (Computer Sciences Corporation) for constructing the website, and to Carl Litzinger (US EPA, Gulf Ecology Division) and David Owens (Computer Sciences Corporation) for their facilitation of website development. Also, thanks to our support personnel: Nicole Allard, Christel Chancy, Anthony DiGirolamo, Laura Dobbins, Brandon Jarvis, Sarah Kell, Larissa Lee, Nathan Lemoine, Marion Marchetto, Cheryl McGill, Michael Norberg, Hannah Rutter, and Alice Watts. Peer review and beta testing of the website were contributed by Larry Goodman, Michael Murrell, Raymond Wilhour, Susan Yee, Jill Awkerman, and Kimberly Nelson (US EPA, Gulf Ecology Division), Rick Bennet and Dale Hoff (US EPA, Mid-Continent Ecology Division), Glen Thursby (US EPA, Atlantic Ecology Division), and Anne Fairbrother (Exponent).

References

- ASTM International. 2015a. Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphibians. E 729-96(2014). West Conshohocken, PA, <http://www.astm.org/Standards/E729.htm>.
- ASTM International. 2015b. Standard guide for conducting static toxicity tests with microalgae. ASTM E1218 - 04e1 (2014). ASTM International, West Conshohocken, PA, <http://www.astm.org/Standards/E1218.htm>.
- Asfaw, A., M. R. Ellersieck, and F. L. Mayer. 2003. Interspecies Correlation Estimations (ICE) for acute toxicity to aquatic organisms and wildlife. II. User Manual and Software. EPA/600/R-03/106. U.S. Environmental Protection Agency, National health and Environmental Effects Research Laboratory, Gulf Ecology Division, Gulf Breeze, FL. 14 p.
- Awkerman, J., S. Raimondo, and M.G. Barron. 2008. Development of Species Sensitivity Distributions for wildlife using interspecies toxicity correlation models. *Environ. Sci. Technol.* 42 (9): 3447-3452.
- Awkerman, J., S. Raimondo, and M.G. Barron. 2009. Estimation of wildlife hazard levels using interspecies correlation models and standard laboratory rodent toxicity data. *J Toxicol Environ Health, Part A.* 72: 1604-1609.
- Baril, A., B. Jobin, P. Mineau, and B. T. Collins. 1994. A consideration of inter-species variability in the use of the median lethal dose (LD₅₀) in avian risk assessment. Technical Report No. 216. Canada Wildlife Service, Headquarters.
- Barron, M. G., C. R. Jackson, J. A. Awkerman. 2012. Evaluation of in silico development of aquatic toxicity species sensitivity distributions. *Aquat Toxicol.* 116-117: 1-7.
- De Zwart, D. 2002. Observed regularities in species sensitivity distributions for aquatic species. In *Species Sensitivity Distributions in Ecotoxicology*, L. Posthuma, G.W. Suter, T.P. Traas, Eds. Lewis Publishers, Boca Raton, FL. pp133-154.
- Dyer, S. D., D. J. Versteeg, S. E. Belanger, J. G. Chaney, and F. L. Mayer. 2006. Interspecies correlation estimates predict protective environmental concentrations. *Environ. Sci. Technol.* 40: 3102-3111.
- Dyer, S. D., D. J. Versteeg, S. E. Belanger, J. G. Chaney, S. Raimondo and M. G. Barron. 2008. Comparison of Species Sensitivity Distributions Derived from Interspecies Correlation Models to Distributions used to Derive Water Quality Criteria. *Environ. Sci. Technol.* 42: 3076-3083.
- Fairbrother, A. 2008. Risk Management Safety Factor. In. *Encyclopedia of Ecology*, vol. 4. S. E. Jørgensen and B. D. Fath (eds.). Elsevier publishing. pp. 3062-3068.
- Hudson, R. H., R. K. Tucker, and M. A. Haegele. 1984. Handbook of toxicity of pesticides to wildlife. U.S. Fish and Wildlife Service, Resource Publ. 153, Washington D.C. 90 p.
- Mineau, P., A. Baril, B. T. Collins, J. Duffe, G. Joerman, and R. Luttik. 2001. Pesticide acute toxicity reference values for birds. *Rev. Environ. Contam. Toxicol.* 170: 13-74.
- National Research Council (NRC). 2013. Assessing risks to endangered and threatened species from pesticides. National Academies Press, Washington, D.C. 175 pp.
- OECD (Organization for Economic Cooperation and Development). 1996. OECD Guidelines for the Testing of Chemicals. Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Paris, France 26p.

- Raimondo, S., P. Mineau, and M. G. Barron. 2007. Estimation of chemical toxicity in wildlife species using interspecies correlation models. *Environ. Sci. Technol.* 41: 5888-5894.
- Raimondo, S., C.R. Jackson, M.G. Barron. 2010. Influence of taxonomic relatedness and chemical mode of action in acute interspecies estimation models for aquatic species. *Environ Sci Technol.* 44: 7711-7716.
- Shafer, E. W. Jr. and W. A. Bowles Jr. 1985. Acute oral toxicity and repellency of 933 chemicals to house and deer mice. *Arch. Environ. Contam. Toxicol.* 14: 111-129.
- Shafer, E. W. Jr. and W. A. Bowles Jr. 2004. Toxicity, repellency or phototoxicity of 979 chemicals to birds, mammals and plants. Research Report No. 04-01. United States Department of Agriculture, Fort Collins, CO. 118 p.
- Shafer, E. W. Jr., W. A. Bowles Jr. and J. Hurlbut, . 1983. The acute oral toxicity, repellency and hazard potential of 998 chemicals to one or more species of wild and domestic birds. *Arch. Environ. Contam. Toxicol.* 12: 355-382.
- Smith, G. J. 1987. Pesticide use and toxicology in relation to wildlife: organophosphorus and carbamate compounds. Resource Publication 170. United States Department of the Interior, Washington, DC. 171 p.
- US EPA. 1996a. Ecological Effects Test Guidelines. OPPTS 850.1075 Fish Acute Toxicity Test, Freshwater and Marine. EPA 712-C-96-118. Washington DC.
- US EPA. 1996b. Ecological Effects Test Guidelines OPPTS 850.5400, Algal Toxicity, Tiers I and II. EPA 712-C-96-164, 11p.
- US EPA. 2005. TREX: Terrestrial Residue EXposure model. Office of Pesticide Programs. U.S. Environmental Protection Agency.
http://www.epa.gov/oppefed1/models/terrestrial/trex_usersguide.htm#content4

Appendix 1. Number of Models by Version

Web-ICE 3.3 - Release June 2016:

Database	Attributes			Number of models		
	Records	Species	Chemicals	Species	Genus	Family
Aquatic animals	8632	316	1499	1550	854	887
Algae	1647	69	457	58	44	0
Wildlife	4329	156	951	560	0	292

Web-ICE 3.2 - Release April 2013:

Database	Attributes			Number of models		
	Records	Species	Chemicals	Species	Genus	Family
Aquatic animals	5501	180	1266	780	289	374
Algae	1647	69	457	58	44	0
Wildlife	4329	156	951	560	0	292

Web-ICE 3.1 - Release January 2010:

Database	Attributes			Number of models		
	Records	Species	Chemicals	Species	Genus	Family
Aquatic animals*	5501	180	1266	780	289	374
Wildlife	4329	156	951	560	0	292

* Aquatic models were reduced between versions 2.0 and 3.1 due to increased data standardization criteria between versions. Data standardization was increased to ensure model relationships were reflective of inherent species sensitivity with minimal influence of extraneous variables (e.g. life stage, test conditions). See the database documentation for details on standardization (<http://v26265ncay507.aa.ad.epa.gov/webICE/ICE%20Aquatic%20DB%20documentation.pdf>)

Web-ICE 2.0 - Release August 2007:

Database	Attributes			Number of models		
	Records	Species	Chemicals	Species	Genus	Family
Aquatic animals	4706	217	695	1074	481	526
Wildlife	4329	156	951	560	0	292

Web-ICE 1.1 - Release July 2007:

Database	Attributes			Number of models		
	Records	Species	Chemicals	Species	Genus	Family
Wildlife	4329	156	951	560	0	292