EPA/600/B-24/158
May 2024

Web-based Interspecies Correlation Estimation

(Web-ICE) for Acute Toxicity: User Manual

Version 4.0

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

Crystal R. Lilavois, S. Alexis Nelson, and Sandy Raimondo

Released: May 2024

U.S Environmental Protection Agency
Office of Research and Development
Center for Environmental Measurement and Modeling
Gulf Ecosystem Measurement and Modeling Division
Gulf Breeze, FL 32561

Reference Web-ICE as:

Lilavois, C.R., S.A. Nelson, and S. Raimondo. 2024. Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity: User Manual. Version 4.0. U.S. Environmental Protection Agency, Office of Research and Development, Gulf Ecosystem Measurement and Modeling Division. Gulf Breeze, FL.

Disclaimer:
The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the US Environmental Protection Agency.

Version Control:
Web-ICE varies among versions by website structure and function and/or models resulting from updated databases. 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 1 for changes.

Contents

Introduction

Ecological Risk Assessments (ERA) aim to protect diverse ecological communities from chemical stressors through evaluation of available data on species sensitivity. Toxicity data are typically only available for some standard test species with limited or no data available for most species and taxa. The USEPA has prioritized the development and application of New Approach Methodologies (NAMs), defined as "any technology, methodology, approach, or combination that can provide information on chemical hazard and risk assessment to avoid the use of animal testing," to reduce the use of vertebrate animals for toxicity testing (USEPA 2021). Interspecies Correlation Estimation (ICE) models extrapolate acute toxicity to taxa with no acute data for a chemical of interest from the value of a surrogate species. Assumptions of ICE models are 1) they represent the relationship of inherent sensitivity between two species and this relationship is conserved across chemicals, mechanisms of action, and ranges of toxicity and 2) the nature of a contaminant that was tested on the surrogate reflects the nature of the contaminant in the predicted species (Raimondo et al. 2024). A suite of ICE models that can predict to a diversity of untested taxa is available through the Web-based Interspecies Correlation Estimation (Web-ICE) application (www3.epa.gov/webice).

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 sensitivity data (the surrogate species). ICE models are least square linear 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 vertebrates, invertebrates, and algae (EC/LC50) 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 surrogates for multiple 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., 5^{th}) 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 estimates toxicity to taxa representing threatened or endangered (listed) species using up to 25 surrogates. This module batch processes toxicity values for listed species from all species, genus, and family level models available for the selected listed species and the entered

surrogate(s). 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 January 31, 2022), 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 broad taxonomic groups (e.g., Fishes) or a particular species (e.g., Atlantic salmon, *Salmo salar*).

The Bulk ICE module is new with version 4.0 and produces batch predictions from species-, genus-, and family-level models using up to 25 surrogates. Genus and family ICE models are developed using genus or family mean acute values, respectively, and contain the same level of accuracy as species models. In previous versions of Web-ICE, genus and family models could only be accessed individually from the respective modules. Bulk ICE allows a user to generate predictions from models of all taxonomic levels and produces an excel-friendly output for evaluation of all possible estimates derived from the entered surrogates. These higher-level models may provide estimates for taxa not represented by species models with the available surrogate data and can be used to expand taxonomic diversity in acute datasets.

This manual provides step-by-step instructions for using Web-ICE v4.0, as well as information on the databases, model development, model validation, and updated user guidance on model selection and interpretation. This version differs from v3.3 with several changes of its functionality, including the new Bulk ICE module and calculation of Species Mean Acute Values (SMAVs) in the SSD module, as well as new and updated models for aquatic vertebrates and invertebrates (Appendix 1). The updated user guidance is based on recent uncertainty analyses conducted for low toxicity compounds (Raimondo et al. 2024). 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 ERA.

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 24, 48 or 96-hr EC/LC50 values based on death or immobility and are standardized by taxa (i.e., 24-h EC50 for fairy shrimp, 48-h EC50 for daphnids, 96-h LC50 for fish). 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., USEPA 2016).

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, USEPA 2012). 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 (USEPA) 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 (Raimondo et al. 2007). 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 linear 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 species 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 to be candidate for model development. 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, multiple toxicity values for the same chemical are represented by the geometric mean (Raimondo et al 2007, 2010, 2024). Only models with a significant relationship (p-value $\leq$ 0.05) are included in Web-ICE.

III. Model Validation

The uncertainty of each model was assessed using leave-one-out cross-validation (Raimondo et al. 2010, 2024). In this method, datapoints representing paired acute toxicity values for surrogate and predicted taxa were systematically removed one at a time from the original model and a submodel was develop from the remaining data. The submodel was used to estimate the toxicity value of the removed predicted taxa from the respective surrogate toxicity value. This method was 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 was used to determine the accuracy of the estimated toxicity value. The cross-validation success rate for each model is the proportion of removed data points that were predicted within 5-fold of the actual value. However, acute toxicity can be highly variable with inter-test variation spanning up to 3-orders of magnitude for a species and chemical (Hrovat et al. 2009, Schur et al. 2023). Acute data for aquatic vertebrates and invertebrates used in the ICE v4.0 database had a mean inter-test range of 11.6 with 80th and 90th percentiles of 5.0 and 10.8, respectively (Raimondo et al. 2024). As such, the cross-validation success rate indicates a highly conservative estimate of model accuracy. For more information on model validation, see Raimondo et al. (2007, 2010, 2024).

There is a strong relationship between taxonomic relatedness and prediction accuracy, with uncertainty increasing with larger taxonomic distance (Raimondo et al. 2007, 2010, 2024). For the models in this version, aquatic fish and invertebrates, models predict within 5-fold and 10-fold of the actual value with 85 and 95% certainty for species within the same order (Raimondo et al. 2024). 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. 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. The **Bulk ICE Modules** predict to all available taxa based on selected Model Level and Taxa of Interest for the surrogates entered. Modules are accessible either from the home page or from the navigation bar along the left side of the page (Figure 1).

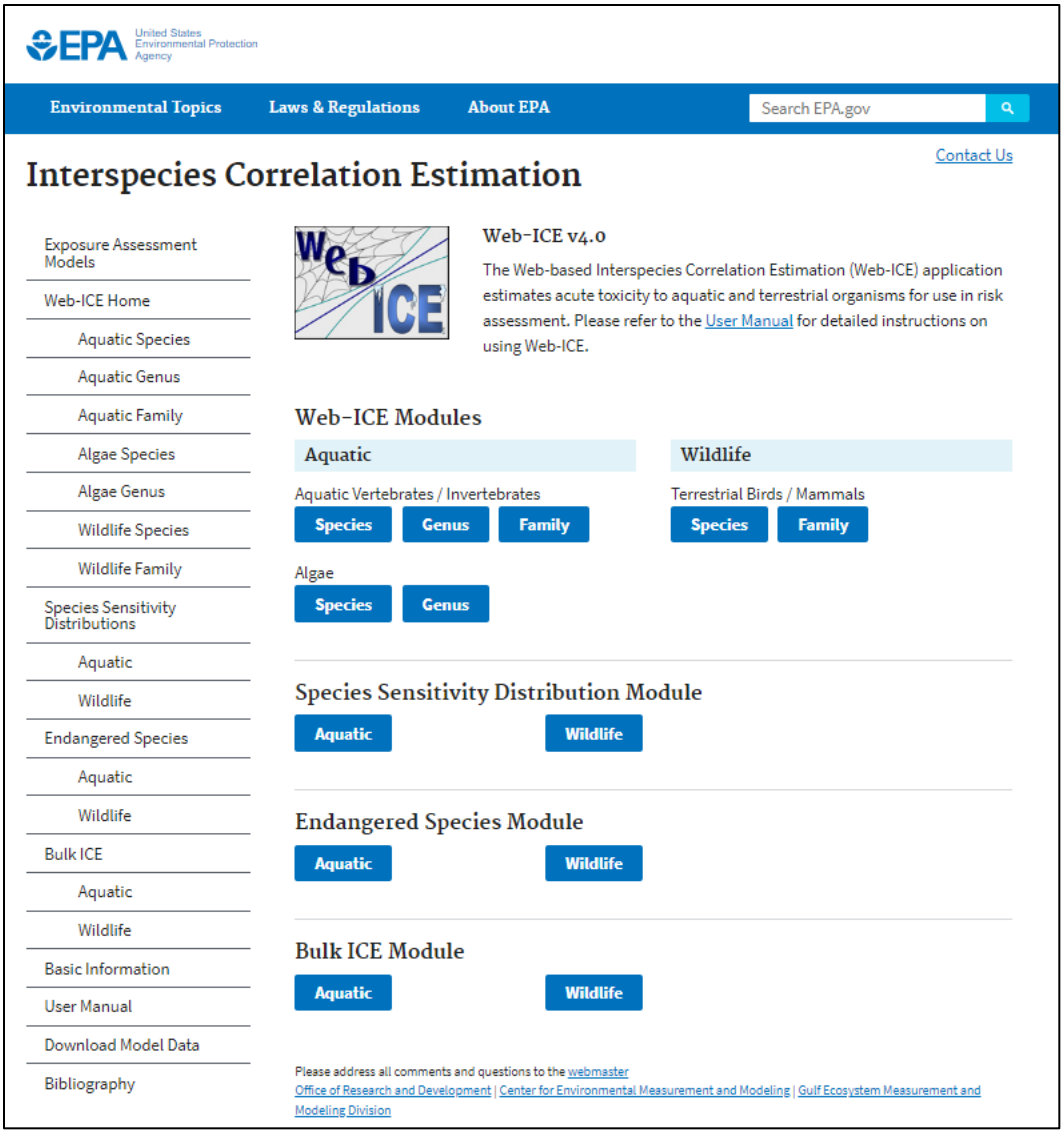

Figure 1. Home page of Web-ICE program.

I. Working with Web-ICE Aquatic or Wildlife Modules

Selecting Model Taxa

1. From either the home page or the 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, there may be a genus or family-level model available. The available models may be determined by browsing through the genus and family level modules or through the **Bulk ICE modules**.

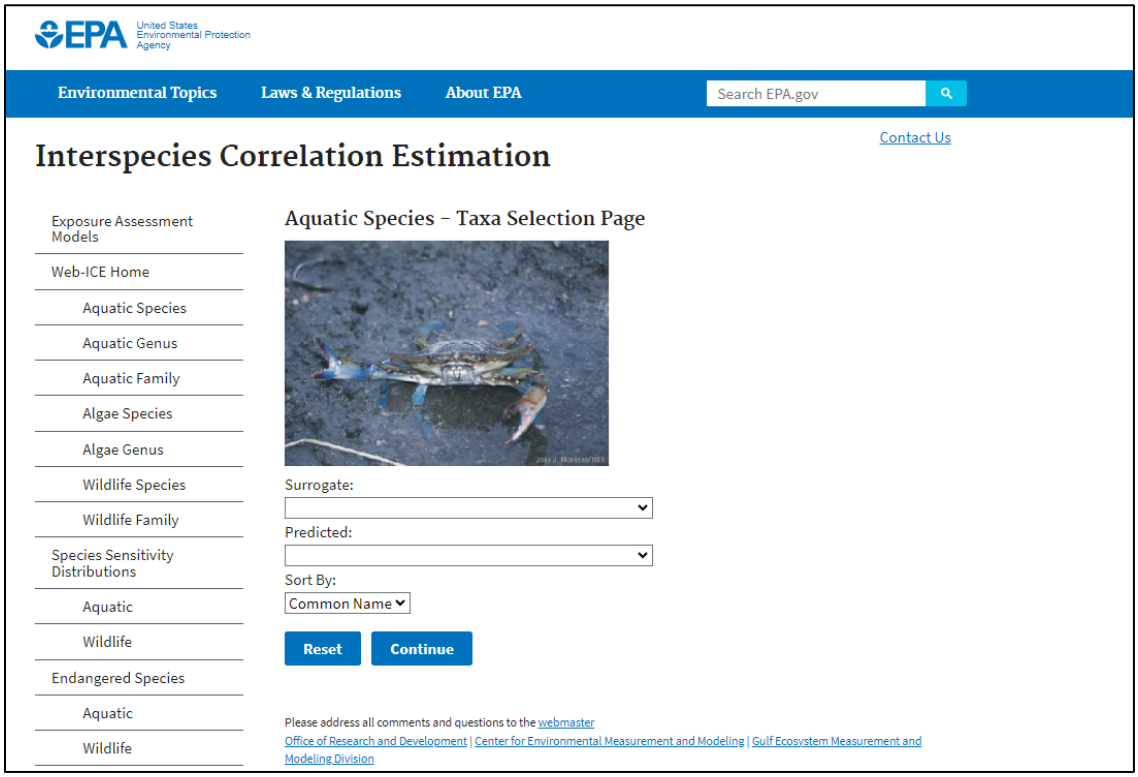

Figure 2. Taxa selection page for aquatic species.

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., $\log_{10}$(LC50) 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. If the user elects to use the toxicity value this error message will also appear below the Surrogate and Predicted Acute Toxicity values. See the *Guidance for Model Selection and Use* section of this manual for interpretation of input values beyond the model domain.

7. To select a different model, select the link to the desired module in the navigation bar.

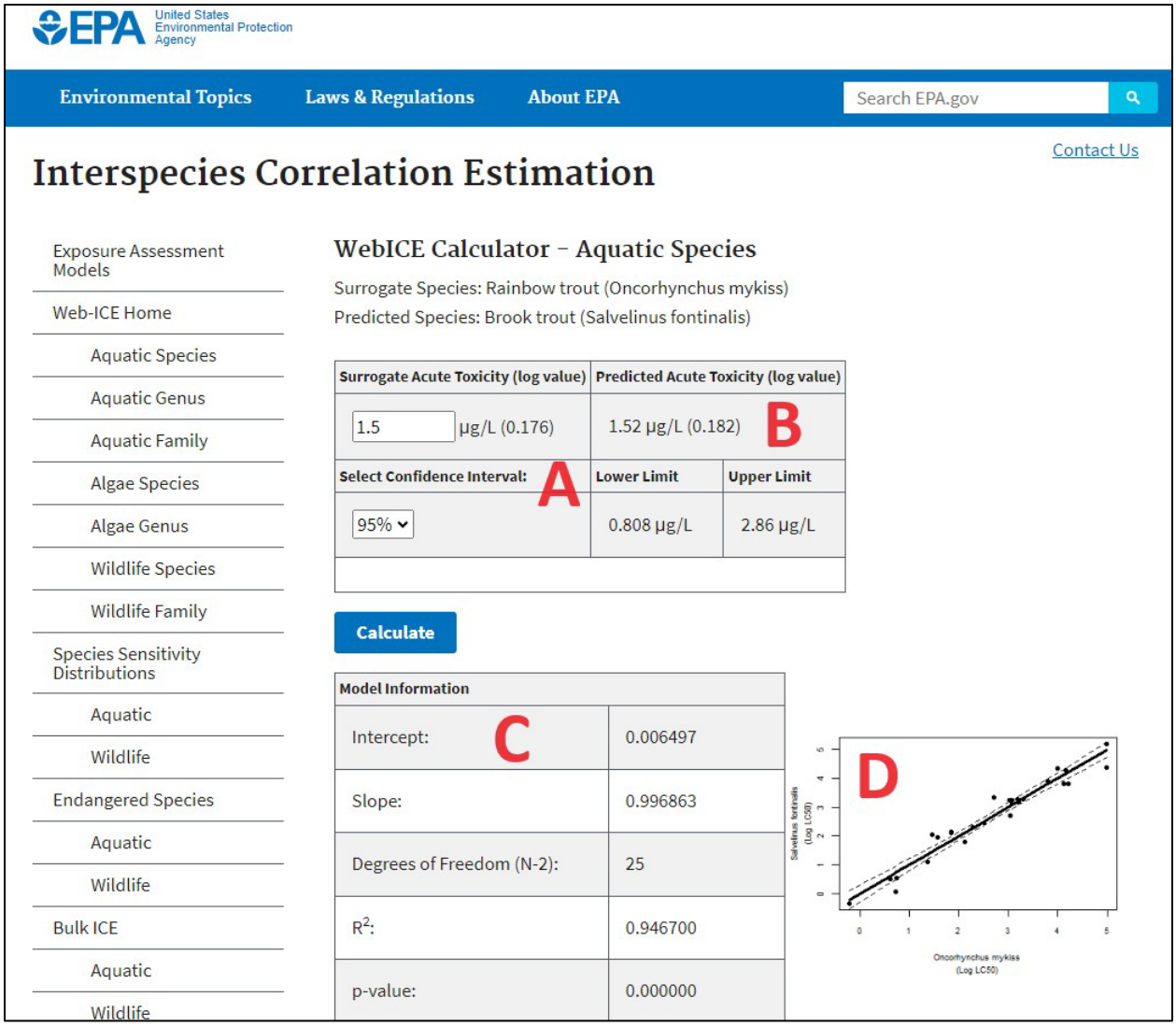

Figure 3. Calculator Page. A. Toxicity input box, B. Predicted toxicity value, C. Model information, D. Model graph.

II. The Species Sensitivity Distribution (SSD) Module

The SSD modules generate SSDs from Web-ICE toxicity values and 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 and calculates the Species Mean Acute Value (SMAV) as the geometric mean of all predicted values. The user may deselect predictions using the checkbox by each model to exclude values from the SSD, which subsequently recalculates the HC.

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., 5^{th}) of the described distribution.

Generating an SSD

1. Under the SSD module, select either Aquatic or Wildlife.

2. On the Aquatic SSD **Taxa Selection Page**, first select for Animal or Algal Models. Next select your surrogate species from the drop-down menu and click **Add** to add the species as a surrogate. A maximum of 25 surrogates can be selected (Figure 4).

3. To remove a surrogate species from the list after it is added, click **Remove Species** next to the species name.

4. Enter the known toxicity for the surrogate species, click **Calculate SSD**.

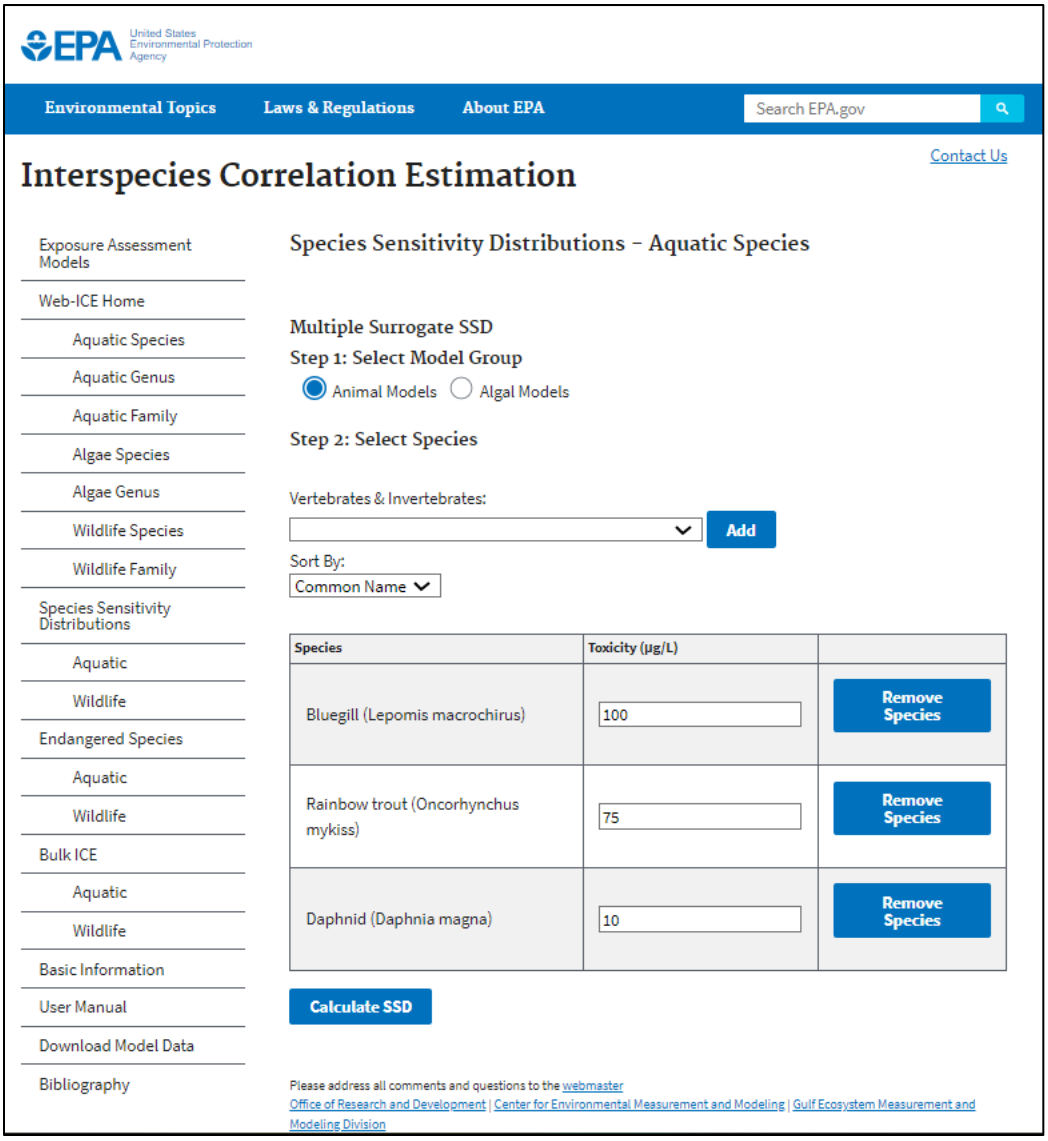

Figure 4. SSD taxa selection page.

Working with the SSD Output

The SSD output page is divided into four parts: the input values and HC calculation (Figure 5A), calculated results (Figure 5B), model filters (Figure 5C), and result export functions (Figure 5D).

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 5th percentile, or the HC5/HD5 (Figure 5A).

2. If multiple surrogates predict to the same species, the species mean acute value (SMAV) is calculated. This value is the geometric mean of all predictions for a species.

3. The warning "**Input toxicity is greater (less) than model maximum (minimum)"** indicates that a predicted value was generated from a surrogate species toxicity value that was outside the range of toxicity values used to generate that model. See the *Guidance for Model Selection and Use* section of this manual for interpretation of input values beyond the model domain.

4. The user can uncheck the box to the left of a predicted species to exclude that model and estimated toxicity from the SSD. The SMAV and HC/HD values are automatically recalculated based on these selections.

5. The estimated toxicity values may be sorted by a column of interest by selecting **Sort** below the column heading.

6. Predicted values can be filtered by inputing desired ranges for the lower and upper bounds for degrees of freedom, R^2, p-value, mean squared 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.

7. The user can generate an Excel-friendly output for either all predicted toxicity values or just the data selected by the check boxes for inclusion in the SSD by selecting the desired **Export to Excel** button (Figure 5D).

NOTE: Web-ICE output has the significant digits of the predicted values limited to improve the presentation of results in the user interface. In cases where the predicted value is < 0.0001, the SMAV appears as "0.0000" and the HC value appears as "NaN". If this occurs, the value can be excluded by deselecting it from the SSD output. If the user wishes to include the value, export the data to excel and replace the value with one calculated from the model-specific calculator pages. The SSD will need to be recalculated from the data from an external software package.

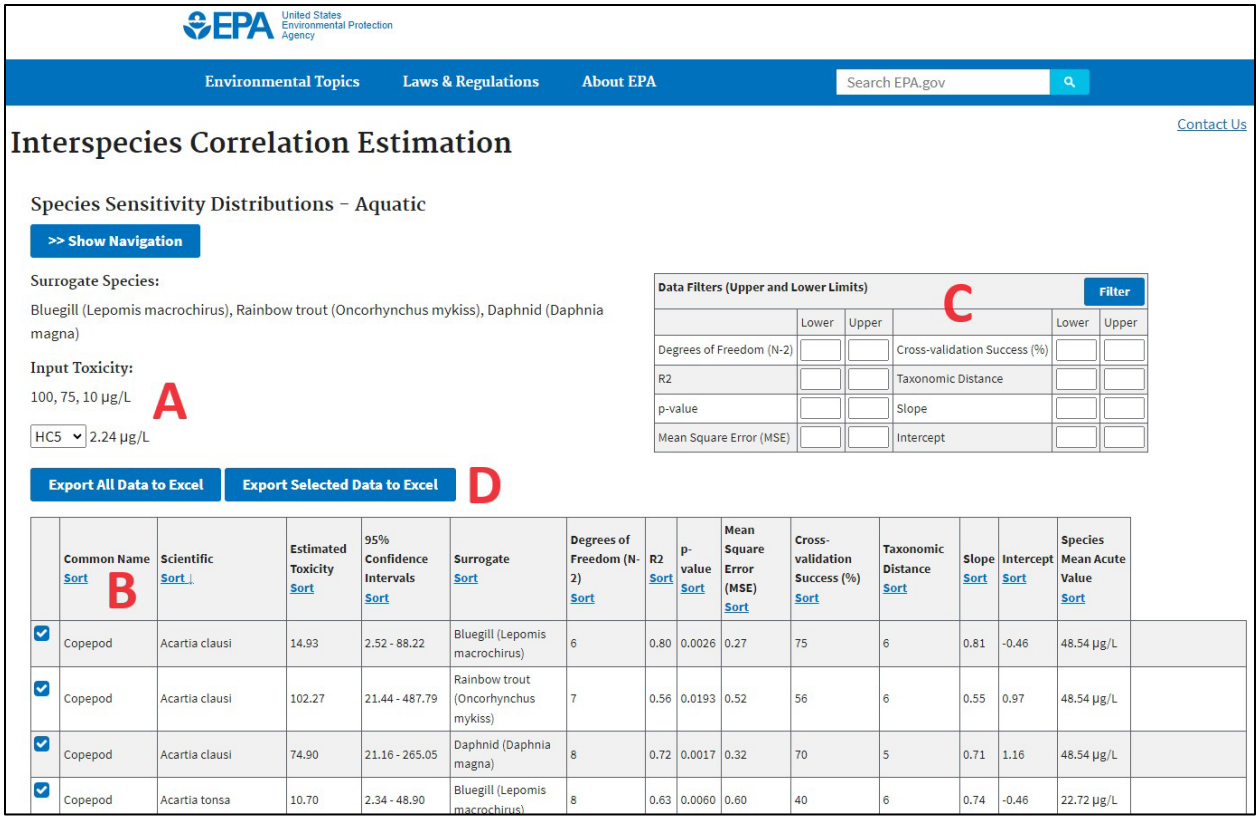

Figure 5. SSD output page. A. Surrogate species input values and hazard level, B. Predicted toxicity values, C. Model data filters, D. Excel-friendly output options.

III. The Endangered Species Module

The **Endangered Species Modules** allow users to select for either a particular species (e.g., Atlantic salmon, *Salmo salar),* or a broad taxonomic group (e.g., fishes). All models available in this module represent species listed as threatened or endangered under the U.S. Endangered Species Act. The entered Surrogate toxicity value is used to predict at all model levels (i.e., species, genus, family) available for the selected species or taxa. Species level models within this module are limited to some aquatic animals, while most listed species included in this module are represented by genus and family level models.

Producing an Endangered Species Toxicity Report

NOTE: The following steps must be completed in the order described below. If changes to a selection are needed, please first use the ***Reset this Form*** *button to refresh the page before continuing.*

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) in Step 1 and/or a particular species of interest from the drop-down menu in Step 2 (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 Species** 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 **Sort** below the column heading.

8. Predicted values can be filtered by inputing desired ranges for the lower and upper bounds for degrees of freedom, R^2, p-value, mean squared 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 **Export to Excel** button.

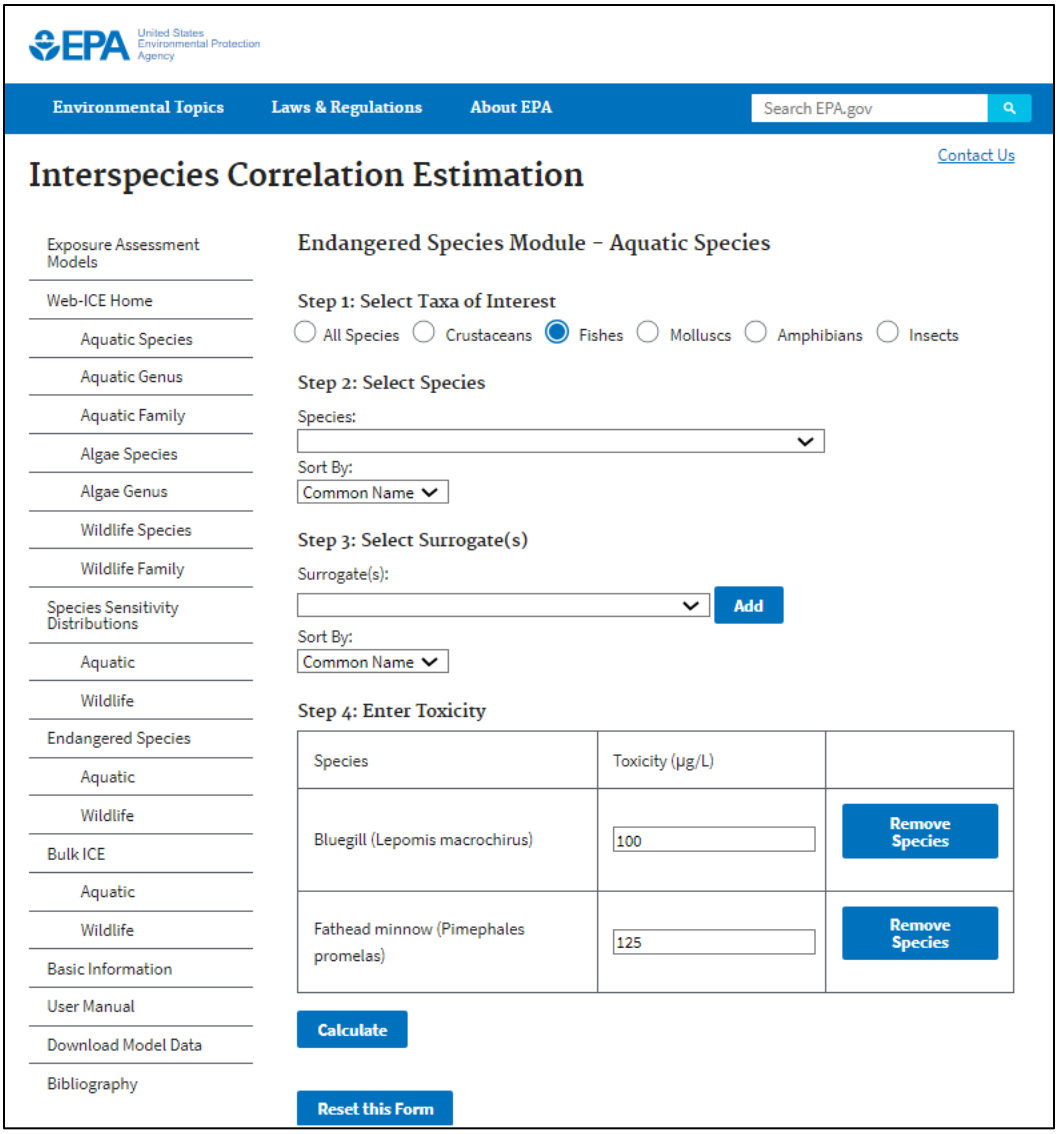

Figure 6. Taxa selection page for the Endangered Species module.

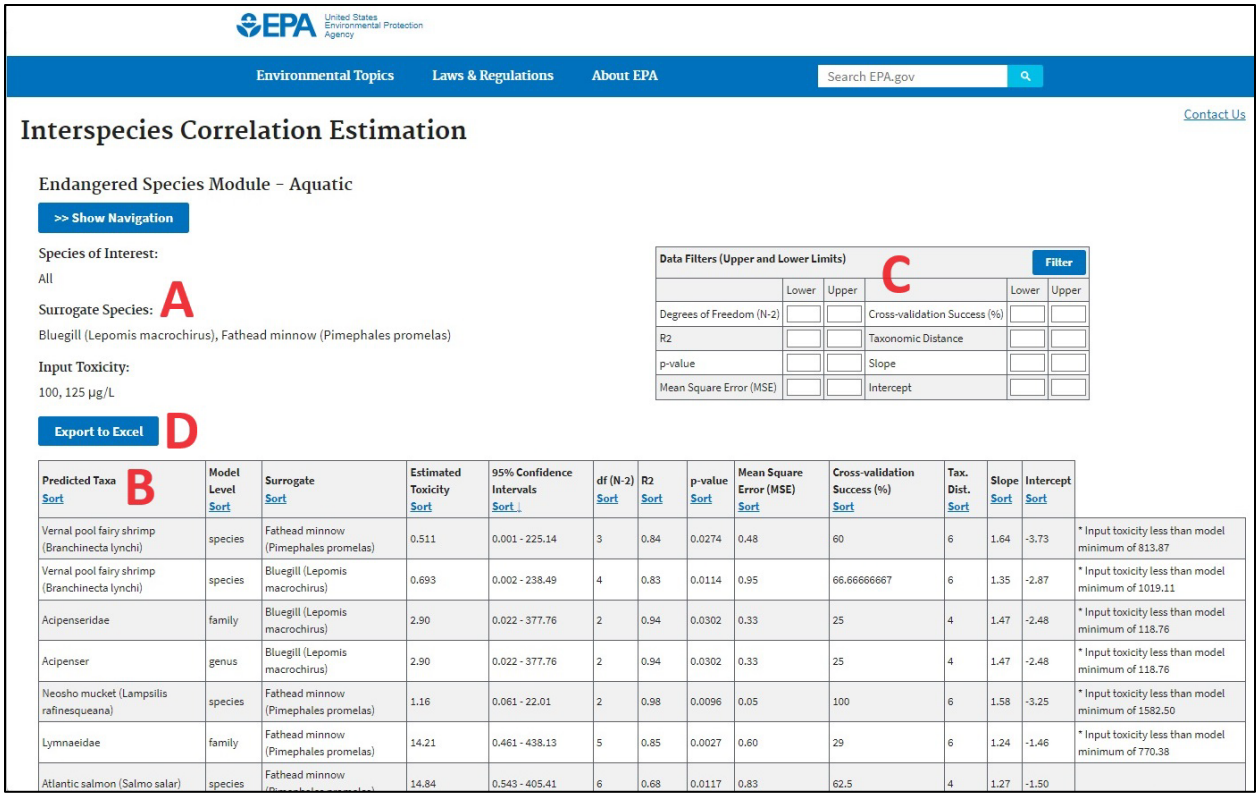

Figure 7. Endangered species predicted toxicity report. A. Surrogate species input values, B. Predicted toxicity values, C. Model data filters, D. Excel-friendly output options.

IV. Bulk ICE Module

The **Bulk ICE Modules** allow a user to generate all available models for the surrogate(s) entered. A user may also choose to narrow the search by specifying a particular Model Level (e.g., species) or selecting a Taxa of Interest (e.g., Amphibian). Selecting only species-level models in this module will produce the same results as the **SSD Module**; however, the **Bulk ICE module** allows the user to similarly predict to all genus and/or family-level models available for the surrogate species.

Producing a Bulk ICE Toxicity Report

*NOTE: The following steps must be completed in the order described below. If changes to a selection are needed, please first use the **Reset this Form** button to refresh the page before continuing.*

1. Under the **Bulk ICE** module, select either Aquatic or Wildlife.

2. On the **Taxa Selection Page**, select the Model Level in Step 1 and Taxa of Interest in Step 2 (Figure 8).

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 Species** 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, surrogate, model information and Mean Acute Values (Figure 7).

7. The user may sort the ICE-estimated toxicity values by each column by selecting **Sort** below the column heading.

8. Predicted values can be filtered by inputing desired ranges for the lower and upper bounds for degrees of freedom, R^2, p-value, mean squared 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 **Export to Excel** button.

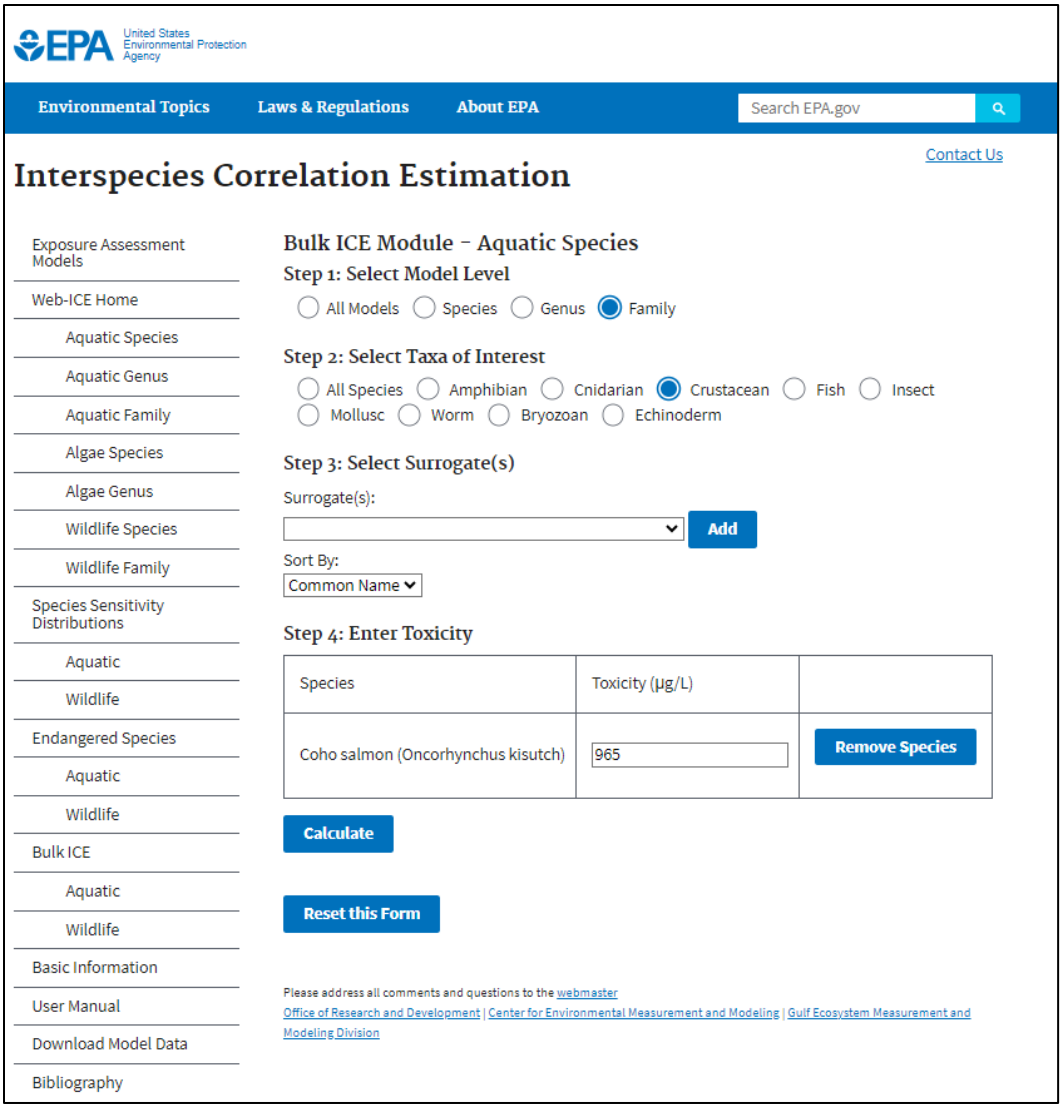

Figure 8. Taxa selection page for Bulk ICE Module.

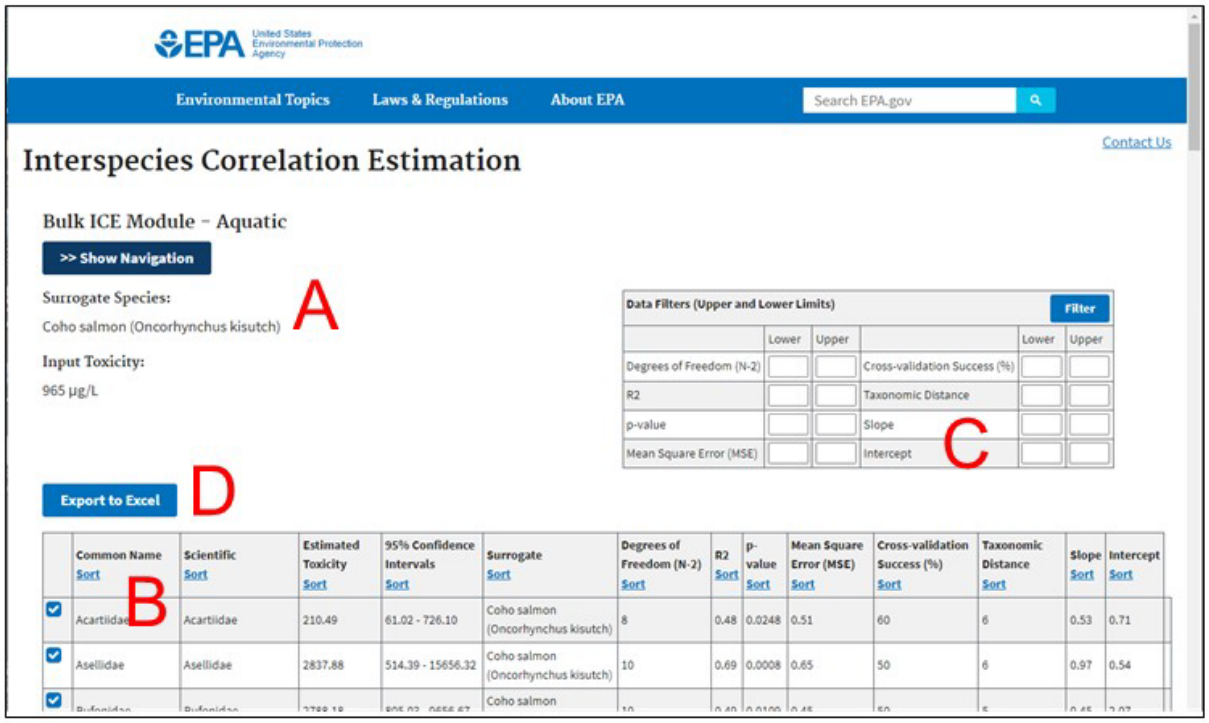

Figure 9. Bulk ICE predicted toxicity report. A. Surrogate species input values, B. Predicted toxicity values, C. Model data filters, D. Excel-friendly output options.

V. 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 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 the information that is needed to calculate 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:

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

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 squared 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 grouped by Mode of Action (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. Conversely, species pairs for which models were developed using all chemicals may not have significant models for some or all MOAs. Currently, MOA-specific models are not 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 robustness 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.

R^2 – The proportion of the data variance that is explained by the model. The closer the R^2 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 reported as μg/L 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 reported as μg/L 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 reported as μg/L and the number in parentheses is the log-transformed value.

Mean Squared 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.

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. User Guidance for Selecting Models with Low Uncertainty

Model parameters, such as MSE, should be used to select models with low uncertainty. The following criteria should be used as a guide to select robust models associated with more accurate predictions:

1. Relatively low mean squared error (MSE) ($\leq$ 0.95)
2. High R^2 value ($\geq$ 0.6)
3. Robust slope (0.6 – 1.4; 0.66 – 1.33 when μg/L input is beyond model domain)

4. Confidence interval range of 2 orders of magnitude between lower and upper limit

For most Web-ICE predictions, surrogate toxicity should be entered as µg/L. However, when the input value is beyond the range of surrogate values used to develop a model (*i.e.*, reported as mg/L), a user may choose to either enter the values as µg/L and allow the model to extrapolate beyond its domain or enter the toxicity as a "scaled" value (i.e., enter and estimate the value as mg/L). In these cases, a modified slope rule of 0.66-1.33 should be used for model selection. An uncertainty analysis of input values beyond the model domain found no difference in prediction accuracy when using scaled values, input as µg/L beyond the model domain, or conventional model use (i.e., input as µg/L within the model domain) for models with slope between 0.66–1.33 (Raimondo et al. 2024).

The best estimations generally occur for surrogate and predicted taxa that are closely related and with low MSE (Raimondo et al. 2007, 2010, 2024). 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.

Acknowledgements

Special thanks to Walter Schwab (CGI Federal Inc.) for constructing and revising the website. We appreciate Elizabeth George, Janet Nestlerode, Leah Oliver, and Kimberly Salinas for review of this user manual and website.

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.

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 (LD_{50}) in avian risk assessment. Technical Report No. 216. Canada Wildlife Service, Headquarters.

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.

Hrovat, M., H. Segner, and S. Jeram. 2009. Variability of in vivo fish acute toxicity data. Regulatory Toxicol. Parmacol. 54: 294-300.

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.

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.

Raimondo, S., C.R. Lilavois, and S.L. Nelson. 2024. Uncertainty analysis and updated user guidance for Interspecies Correlation Estimation (ICE) models and low toxicity compounds. Integr. Environ. Assess. Manage. https://doi.org/10.1002/ieam.4884

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.
USEPA. 2016. Ecological Effects Test Guidelines. OSCPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test. EPA 712-C-16-007. Washington DC.
USEPA. 2012. Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity. EPA 712-C-006.
USEPA. 2021. New Approach Methods Work Plan (v2). U.S. Environmental Protection Agency, Washington, DC. EPA/600/X-21/209.

Appendix 1. Data and Models by Version

Web-ICE 4.0:

	Attributes			Number of models		
Database	Records	Species	Chemicals	Species	Genus	Family
Aquatic animals	10,737	478	1708	2286	1074	1363
Algae	1647	69	457	58	44	0
Wildlife	4329	156	951	560	0	292

Web-ICE 3.3 – Release June 2016:

	Attributes			Number of models		
Database	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:

	Attributes			Number of models		
Database	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:

	Attributes			Number of models		
Database	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.

Web-ICE 2.0 – Release August 2007:

	Attributes			Number of models		
Database	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:

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