EPA Public Access

Author manuscript

Environ Sci Technol. Author manuscript; available in PMC 2018 July 18.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Environ Sci Technol. 2017 July 18; 51(14): 8158–8165. doi:10.1021/acs.est.7b01493.

Framework for Optimizing Selection of Interspecies Correlation Estimation Models to Address Species Diversity and Toxicity Gaps in an Aquatic Database

Adriana C. Bejarano^{*†}, Sandy Raimondo[‡], and Mace G. Barron[‡]

[†]Research Planning, Inc., 1121 Park Street, Columbia, SC 29201, USA

[‡]USEPA, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, Florida 32561

Abstract

The Chemical Aquatic Fate and Effects (CAFE) database is a tool that facilitates assessments of accidental chemical releases into aquatic environments. CAFE contains aquatic toxicity data used in the development of species sensitivity distributions (SSDs) and the estimation of hazard concentrations (HCs). For many chemicals, gaps in species diversity and toxicity data limit the development of SSDs, which may be filled with Interspecies Correlation Estimation (ICE) models. Optimization of ICE model selection and integration ICE-predicted values into CAFE required a multistep process that involved the use of different types of data to assess their influence on SSDs and HC estimates. Results from multiple analyses showed that SSDs supplemented with ICE-predicted values generally produced HC5 estimates that were within a 3-fold difference of estimates from measured SSDs (58%–82% of comparisons), but that were often more conservative (63%–76% of comparisons) and had lower uncertainty (90% of comparisons). ICE SSDs did not substantially underpredict toxicity (<10% of comparisons) when compared to estimates from measured SSD. The incorporation of ICE-predicted values into CAFE allowed the development of >800 new SSDs, increased diversity in SSDs by an average of 34 species, and augmented data for priority chemicals involved in accidental chemical releases.

TOC/Abstract Art

^{*}Corresponding Author. abejarano@researchplanning.com; phone: +1 803 254 0278; fax: +1 803 254 6445.

ASSOCIATED CONTENT

Supporting Information. Additional information is available in Supporting Information 1. This information includes a diagram summarizing the framework, a figure comparing HC5s between SSD pairs, a figure of ICE-predicted values from two approaches, and a figure comparing HC5s between SSD pairs from multiple surrogate species. Supporting Information 2 contains tables with details related to this work. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Keywords

aquatic toxicity; Chemical Aquatic Fate and Effects database; hazard concentrations; Interspecies Correlation Estimation; species sensitivity distributions

INTRODUCTION

Accidental chemical releases of potentially hazardous materials are a common occurrence in navigable waters of the United States, requiring rapid decisions to reduce risks to aquatic communities. The Chemical Aquatic Fate and Effects (CAFE) database, developed by the National Oceanic and Atmospheric Administration's Emergency Response Division (NOAA's ERD), is designed to provide rapid and unrestricted access to fate and effects data for hundreds of chemicals^{1, 2}. In CAFE, aquatic toxicity data are summarized as species sensitivity distributions (SSD), which are cumulative probability distributions of the relative sensitivity across species to a given chemical. Each SSD includes estimates of the 1st and 5th percentile hazard concentrations (HC), or concentrations assumed to be protective of 99% and 95% of the species, respectively. HC estimates could be used as levels of concern protective of a broad range of aquatic species and inform potential impacts associated with chemical releases.

While SSDs have been used in ecological risk assessments of chemicals (e.g., pesticides, metals)^{3–7}, they have seldom been used in assessments of chemical releases primarily due to data gaps in species diversity and aquatic toxicity^{1, 8}. Although CAFE has toxicity data for approximately 4,500 chemicals, three quarters of these chemicals, including priority chemicals for NOAA's ERD^{1, 8}, do not have enough data (<5 species) to develop SSDs. However, these data gaps may be filled with toxicity predictions from mathematical relationships, including Interspecies Correlation Estimation (ICE) models. These models generate toxicity predictions (median lethal concentration [LC50] and median effects concentration [EC50]) by extrapolating known toxicity values from a surrogate species to one or several predicted species. ICE models are developed by pairing chemical-specific toxicity data for each pair of species, with relationships derived based on data across multiple chemical classes or chemicals with a shared mode of action (MOA)^{9–14}.

Previous studies have demonstrated that SSDs developed with ICE-predicted values have HC estimates similar to SSDs developed from measured toxicity values, and that increasing species diversity in SSDs generally reduces uncertainty in HC estimates^{10, 14–19}. Thus, the integration of ICE-predicted values into CAFE could enhance its capabilities and increase confidence in environmental assessments related to accidental chemical releases. However, this integration requires a comprehensive evaluation to determine the influence of ICE-predicted values on the accuracy and uncertainty of SSDs and HC estimates. The objectives of the current research were to evaluate existing ICE models for aquatic species, identify which models generate the most accurate estimates of toxicity, and evaluate the integration of ICE-predicted values in CAFE. The goal of the current research was to address aquatic toxicity data gaps and increase species diversity in CAFE allowing the development of SSDs for an expanded number of chemicals of interest to spill response and ecological risk

assessment. The approach and framework has general applicability to aquatic toxicity databases, which are typically limited by species and chemical diversity.

EXPERIMENTAL SECTION

Approach Overview

A multistep process of developing and comparing SSDs was used to establish the technical basis for integrating ICE-predicted values into CAFE. This multistep process included the following: 1) screening of ICE models using simulated data as input concentrations, and comparisons of SSDs with values from surrogate and ICE-predicted species (hereafter ICE SSDs) based on model accuracy criteria; 2) screening of ICE models using empirical data as input concentrations, and comparisons of ICE SSDs based on model accuracy criteria as well as with SSDs containing only empirical data (hereafter measured SSDs); 3) an integration step that determined the optimal ICE model selection protocol by comparing measured SSDs to SSDs supplemented with ICE-predicted values (hereafter ICE supplemented SSDs); 4) assessment of data gains and improvements in CAFE from the incorporation of ICE-predicted values; and 5) an internal verification step that compared ICE SSDs to measured SSDs developed from robust datasets (Figure 1; see Supporting Information 1, SI1, Figure S1 for details).

ICE models were obtained from the USEPA¹², and augmented with models for MOA chemicals^{11, 13}, dispersant and petroleum products¹⁰ and non-polar aromatic hydrocarbons¹⁴. The combined dataset contained 3,634 ICE models for 190 unique surrogate species. Aquatic toxicity data from CAFE previously evaluated for data quality^{1, 20} were used to either develop measured SSDs or as input values in ICE models. Only empirical data (LC50 and EC50) that met the following standardization requirements were included^{1, 20}: toxicity data based on measured concentrations; tests performed under flow-through, static or static renewal conditions; and tests performed with individual chemicals containing 75% active ingredient purity.

Following the internal structure of CAFE, SSDs were developed for each unique chemical (by chemical abstracts service [CAS] number) and by exposure duration (24 h, 48 h, 72 h or 96 h). For each step of the multistep process, SSDs with at least 5 species were developed by fitting each data set to a lognormal distribution that was randomly resampled 2,000 times to estimate the mean SSD and the HC5 with and its 95% Confidence Interval (95% CI)⁸. This resampling approach employs a Bayesian framework and was preferred over classical approaches because it allows for uncertainty propagation as it takes into account intrinsic features of the original data, including sample size⁵. Measured SSDs used the geometric mean of all empirical values for each unique species⁵, and all data used to develop ICE supplemented SSDs were weighted equally regardless of source (i.e., empirical, ICE-predicted). All SSDs were assessed for goodness-of-fit ($\alpha=0.05$) using the Kolmogorov–Smirnov, Anderson–Darling, and Cramer–Von Mises test statistics^{8, 21} and only SSDs that passed goodness-of-fit tests were included in these analyses. Comparisons between pairs of SSDs were performed using a log-likelihood chi square statistic ($\alpha=0.05$)²². In some of cases, data limitations restricted comparisons of completely independent pairs of SSDs. To reduce violations of data independence, only pairs of SSDs that shared data were compared

if their datasets differed by at least 50% in the number of species. In the current study, and to err on the side of caution, it was assumed that the preferred models are those that produce the most conservative estimates.

Prior to each analysis, assessments were performed on the influence on SSDs of both input and predicted concentrations above 1-order of magnitude of the range of toxicity values originally used to develop each ICE model (hereafter model domain), which represent the range associated with reliable model predictions. Uncertainty in ICE SSDs and ICE supplemented SSDs was evaluated by comparing pairs of SSDs with and without surrogate and predicted values outside the model domain. All statistical analyses were performed using the R statistical platform (v. 2.1.3)²³ and associated packages^{24–26}.

ICE Model Screening with Simulated Data

The initial screening of ICE models involved a comparison between SSDs developed with predicted values from all available ICE models (hereafter ICE_{All}) and SSDs with predicted values from ICE models that met criteria associated with greater predictive accuracy (Mean Square Errors [MSE] <0.95, slope >0.6, adjusted coefficients of determination [adj-R²] >0.6)²⁷ (hereafter ICE_{Reduced}). These analyses included 3,082 models in the ICE_{All} model set and 2,584 models in the ICE_{Reduced} model set for 91 unique surrogate species.

In this analysis, model predictions were generated for a hypothetical chemical from eight simulated concentrations spanning from 1 µg/L to 10 g/L in 1 order of magnitude increments. This approach was used given that empirical data were not sufficient to develop ICE SSDs for each unique surrogate species in the set of ICE models. Each simulated value was used as input toxicity for each surrogate species in both the ICE_{All} and ICE_{Reduced} model sets resulting in toxicity estimates for all possible predicted species. For each surrogate species, ICE SSD were developed with simulated and predicted values from ICE_{All} or ICE_{Reduced} model sets, and pairs of ICE SSDs for the same surrogate species statistically compared only when datasets were sufficiently different.

ICE Model Screening with Empirical Data

A second screening analysis followed the same methodology previously described, with the exception that empirical data for a variety of chemicals were used as input toxicity for each surrogate in both the ICE_{All} and ICE_{Reduced} model sets. These analyses included 1,870 models in the ICE_{All} model set and 1,342 models in the ICE_{Reduced} model set for 55 unique surrogate species. For each chemical, ICE SSDs were developed for each surrogate species and their predicted values from ICE_{All} or ICE_{Reduced} model sets, and pairs of ICE SSDs for the same surrogate species statistically compared only when datasets were sufficiently different. Similar comparisons were also made to categorize each individual surrogate species based on how close estimates from their ICE SSD were to those from measured SSDs.

Integration of ICE-Predicted Data with CAFE

The process for evaluating the integration of ICE-predicted values into CAFE focused on chemicals with generally limited data (< 10 species), and involved the development of

measured and ICE supplemented SSDs. The latter were developed from single or multiple surrogate species using ICE models selected through the screening analyses. In small datasets, the use of single surrogate species to develop ICE supplemented SSDs has the disadvantage of having all predicted values derived from the same surrogate species. While this issue is unavoidable for chemicals with limited toxicity data, combining predicted values across multiple surrogate species is the preferred approach. However, this approach may produce multiple predicted values for the same species, which were dealt with via the evaluation of two approaches: 1) the selection of the ICE-predicted values with the narrowest confidence intervals (i.e., best models) around the prediction; and 2) using the geometric mean of ICE-predicted values for the same species. Outcomes from these evaluations were used in all subsequent analyses.

Assessments of goodness-of-fit of all SSDs were made as previously described, with the addition that SSDs failing goodness-of-fit tests were further examined for the potential influence of extreme outliers²⁶. For each chemical, pairs of SSDs (i.e., measured SSDs vs. ICE supplemented SSDs from single or multiple surrogate species) were statistically compared only when datasets were sufficiently different.

Assessment of CAFE's Data Gains and Improvements

An assessments of data gains and improvements in CAFE from the incorporation of ICE-predicted values were based on the number of additional SSDs developed through this integration. Chemicals with data gains are those for which SSDs were not previously attainable due to limited toxicity data for at least 5 species, and chemicals with data improvements are those for which SSDs were available, but with a relatively small species diversity (5–10 species). A second assessment focused on data gains and improvements for chemicals identified as priority for NOAA's ERD based on several criteria (i.e., chemicals involved in incidents, potentially toxic, reasonably water soluble, and shipped in bulk generally in large quantities; 210 chemicals)^{1, 8}, as well as for a subset of chemicals (78 unique CAS numbers) involved in incidents reported to the National Response Center database (2000–2014) and by NOAA's ERD (2003–2014)¹.

Verification of ICE SSD Estimates

A final analysis involved a verification of estimates from ICE SSDs versus measured SSDs from chemicals in CAFE with relatively robust datasets (>10 species). This verification was needed to provide a sense of the dependability of SSD estimates obtained from ICE-predicted values. Assessments of goodness-of-fit of all SSDs were made as previously described, and included assessment of extreme outliers. For each chemical, pairs of SSDs (i.e., measured SSD vs. ICE SSDs) were statistically compared.

RESULTS

ICE Model Screening with Simulated Data

The initial screening analyses identified 34 of 221 SSDs as being influenced by values outside the model domain, which were then removed from further analyses (details in SI1). ICE models were removed from further analyses when ICE SSDs did not pass all goodness-

of-fit tests. This criterion resulted in the exclusion of 160 ICE models for 7 unique surrogate species, mostly involving MOA models (hereafter, ICE_{MOA}) (Supporting information 2, SI2). When ICE SSDs from both ICE_{All} and ICE_{Reduced} models were available for the same surrogate species (1,000 comparisons), HC5 estimates were comparable (within a 3-fold difference in 79% of cases) with ICE SSDs from ICE_{All} models generally producing smaller HC5 estimates (more conservative; 62% of cases) (Figure 2A). Statistical comparison were only possible for a small number of SSDs with substantially different datasets (data from 5 to 14 additional species). In most cases (82%; 46 of 56 comparisons) pairs of ICE SSDs from ICE_{All} and ICE_{Reduced} models were not statistically significantly different, and HC5 estimates were generally (67% of cases) within a 3-fold difference. These initial results indicated that because of the marginal impacts on the fit of SSDs, there may not be real gains in using a reduced set of models for the purpose of developing SSDs. Based on these results, 3,222 ICE models for 106 unique surrogate species were used in further analyses.

ICE Model Screening with Empirical Data

The second screening analyses identified 13 of 128 SSDs as being influenced by values outside the model domain, which were then removed from further analyses (details in SI1). When ICE SSDs from both ICE_{All} and ICE_{Reduced} models were available for the same surrogate species (328 comparisons), HC5 estimates were comparable (within a 3-fold difference in 89% of cases) with ICE SSDs from ICE_{All} models producing smaller HC5 estimates (more conservative; 94% of cases) (Figure 2B). Statistical comparison were only possible for a small number of SSDs with substantially different datasets (data from 5 to 14 additional species). In most cases (92%; 23 of 25 comparisons) pairs of ICE SSDs from ICE_{All} and ICE_{Reduced} models were not statistically significantly different, and in all cases HC5 estimates were within a 3-fold difference. These results further indicate that, in most cases, there may not be real gains in using a reduced set of ICE models for the purpose of developing SSDs, and that a 3-fold difference may be a reasonable cutoff for determining the adequacy of estimates obtained from ICE SSDs.

Similar analyses were made for ICE SSDs from ICE_{MOA} models. For a total of 34 chemicals with assigned MOA, comparisons were made between HC5 estimates from measured SSDs versus estimates from ICE SSDs from generic ICE_{All} or ICE_{Reduced} models, or ICE_{MOA} models for all available surrogate species (91 measured SSDs; 411 comparisons). ICE SSDs from generic ICE_{All} and ICE_{Reduced} models, and ICE_{MOA} models produced HC5 estimates closest to the measured SSD value in 54%, 23% and 23% of all comparisons, respectively, but in most cases (69%) HC5 estimates from ICE SSDs from the generic ICE_{All} models were within a 3-fold difference to estimates from other models. These results suggest that there may not be real gains in using ICE_{MOA} models to develop SSDs.

Further assessments based on fold-difference of HC5 estimates between ICE SSDs from ICE_{All} models and measured SSDs were made to determine which surrogate species generate the most accurate estimates of aquatic toxicity (SI1, Box S1; SI2), leading to the classification of 32 surrogate species as best surrogates (see below). Only ICE models selected through screening analyses were used in further analyses.

Integration of ICE-Predicated Data with CAFE

Analyses within the integration step identified 110 of 1,081 SSDs as being influenced by values outside the model domain, which were then removed from further analyses (details in SI1).

SSDs from single surrogate species—A total of 1,179 ICE supplemented SSDs were developed for 469 chemicals with limited empirical data (<5 species) using single surrogate species. In the large majority of cases (95%; 1,119 SSDs) ICE supplemented SSDs passed all goodness-of-fit tests, but potential extreme outliers were found in 212 SSDs, including those 60 SSDs that failed goodness-of-fit tests. In 53 of those 60 SSD cases, outliers were introduced by ICE models with *P. reticulata* as the surrogate species, but their removal resulted in SSDs passing goodness-of-fit tests. Because insufficient empirical data were available to develop measured SSDs, comparisons were not made with ICE supplemented SSDs as part of this analysis. However, comparisons were made between measured SSDs for 68 chemicals with limited, but sufficient empirical data (5–9 species), and the same SSDs supplemented with ICE-predicted values. Comparison of 212 pairs of SSDs showed that HC5 estimates were within a 3-fold difference in 58% of cases with ICE supplemented SSDs producing smaller HC5 estimates (more conservative) and narrower confidence intervals (lower uncertainty) in 63% and 90% of cases, respectively (SI1, Figure S2; SI2). Pairs of SSDs with substantially different datasets (142 comparisons, data from 5 to 72 additional species) were statistically significantly different in 51% of cases.

When comparisons were made for 56 chemicals with relatively limited toxicity data (2–8 species) that had at least two candidate surrogate species categorized as best surrogates (113 SSDs total), in 56% of these cases HC5 estimates had overlapping confidence intervals (Figure 3; SI2). ICE supplemented SSDs with *Americamysis bahia* as the surrogate species generally produced the most conservative HC5 estimates across chemicals. HC5 estimates from measured SSDs (5–8 species; 33 comparisons) had in 79% of cases overlapping 95% CI with HC5 estimates from at least one surrogate species, and in all cases confidence intervals around the HC5 estimate were wider than those from ICE supplemented SSDs. These results suggest that ICE models could add species diversity (34±23 species; range: 1–78 species) to SSDs for chemicals with limited empirical data resulting in HC5 estimates with generally narrower confidence intervals (lower uncertainty) than the measured estimates.

SSDs from multiple surrogate species—Combining predicted values from multiple surrogate species is advantageous as it increases species diversity in SSDs. By combining empirical (<10 species) and ICE-predicted values across multiple surrogate species, sufficient data were available to develop SSDs for 1,028 unique chemical-exposure duration combinations (521 chemicals), which in most cases (98%; 1,014 SSDs) passed all goodness-of-fit tests. Extreme outliers introduced by ICE models were found in 14 SSDs, but their removal resulted in SSDs passing all goodness-of-fit tests.

The two approaches for dealing with multiple predicted values for the same species produced comparable estimates (overlapping HC5s and 95% CIs), with little overall impact

on SSDs (details in SI1 and Figure S3). Based on these results, and consistent with data processing in CAFE, multiple toxicity records for the same ICE-predicted species were handled using the geometric mean approach.

Comparisons were made between measured SSDs for 131 chemicals with limited, but sufficient empirical data (5–10 species), and the same SSDs supplemented with ICE-predicted values from multiple surrogates. Comparison of 212 pairs of SSDs showed that HC5 estimates were within a 3-fold difference in 82% of cases with ICE supplemented SSDs producing smaller HC5 estimates (more conservative) and narrower confidence intervals (lower uncertainty) in 76% and 92% of cases, respectively (Figure 4; SI1, Figure S4; SI2). Pairs of SSDs with substantially different datasets (199 comparisons, data from 6 to 99 additional species) were statistically significantly different in 41% of cases. These results suggest that ICE-predicted values from multiple surrogate species: 1) could facilitate the development of SSDs for chemicals with limited toxicity data, generally without introducing extreme outliers; 2) could add species diversity (45 ± 26 species; range: 5–105 species) to SSDs while producing HC5 estimates with narrower confidence intervals (lower uncertainty); and 3) may improve estimates from SSDs with limited species diversity, possibly generating conservative values.

Assessment of CAFE's Data Gains and Improvements

A total of 4,565 toxicity records (geometric means) from CAFE for 386 aquatic species and 1,278 chemicals were used to generate ICE-predicted values. While for half of these chemicals ICE models produced little or no new data, ICE models generated 39,562 predicted values allowing the development of 806 new SSDs and the improvement of 212 existing SSDs (Figure 5A). Unique chemical-exposure duration combination gained a wide number of ICE-predicted species (range: 1–99 species) with an average of 34 species. Most ICE-predicted values were generated from generic ICE models (87%), followed by ICE models for nonpolar aromatic hydrocarbons (10%) and petroleum products (2%).

Assessments focusing on priority chemicals (210 chemicals) and chemicals involved in spill incidents (78 unique CAS numbers) showed that ICE-predicted values allowed the development of 142 new SSDs and the improvement of 56 existing SSDs (Figure 5B), with an average increase from 3 to 34 species per SSD. With the incorporation of ICE-predicted values, data would be available in CAFE to develop SSDs for 33 of those 210 priority chemicals, and for 66 of those 78 chemicals involved in spill incidents.

Verification of ICE SSD Estimates

A final analysis compared ICE SSDs to measured SSDs for chemicals with relatively robust datasets (>10 species). All SSDs passed goodness-of-fit tests, and did not appear to contain extreme outliers. Statistical comparison of 72 pairs of SSDs with completely independent datasets showed that HC5 estimates were within a 3-fold difference in 58% of cases with ICE SSDs producing smaller HC5 estimates (more conservative) and narrower confidence intervals (lower uncertainty) in 46% and 69% of cases, respectively (Figure 6; SI2). In 71% of cases (51 of 72 comparisons), ICE SSDs were not statistically significantly different from measured SSD, and in only 6 of 21 comparisons that were statistically significantly

different, HC5 estimates ICE SSDs were underpredicted (>20-fold difference). These results combined with previous findings confirm that ICE models produce reliable ICE-predicted data that could be used to supplement empirical data, while partially addressing aquatic toxicity and diversity data gaps.

DISCUSSION

One of the greatest challenges in evaluating potential risks of exposure and impacts to aquatic communities following an accidental chemical release is access to reliable information on the chemical of concern. During a release, rapid decisions and assessments of potential risks are often necessary, but because of limited data, these can be based on incomplete knowledge and have great uncertainty. Although CAFE provides access to a comprehensive aquatic toxicity dataset^{1, 20}, enhancements of its capabilities require systematic evaluations of complementary data sources. In the current study, ICE models were evaluated as a potential source for addressing CAFE's gaps in aquatic toxicity and species diversity.

There are several important findings emerging from the analyses of the current study. Evaluations of the influence on SSDs of data falling outside the range of toxicity values used to develop ICE models showed that those values, in the large majority of cases (89% of 1,430 SSDs), do not have appreciable effects on the fit of SSDs. This finding is important because it suggests that ICE models are robust enough to accommodate surrogate and predicted data that are beyond the range associated with reliable model predictions. Assessments were also made based on ICE model accuracy following guidance recommended to assist ICE model selection^{9, 27}. The present and previous studies (based on more stringent criteria than those used here)^{10, 14} indicated that the inclusion of predicted values from ICE models with lower predictive accuracy had marginal to no influences on SSDs. This conclusion is likely due to the dilution of error associated with ICE model prediction by the uncertainty of SSD model parameterization and the variability in toxicity values across species in the SSD^{10, 14, 19}. As a result, SSDs with and without predicted data from ICE models with lower predictive accuracy produced comparable HC5 estimates supporting earlier conclusions^{10, 14} that there may not be real gains in using a reduced set of ICE models. Hence, a general recommendation from the current study is the inclusion of all available ICE models when developing ICE SSDs or ICE supplemented SSDs. In addition, the use of a larger number of ICE models results in greater species diversity in SSDs that generally have lower uncertainty and often produce conservative HC5 estimates.

Earlier work^{10, 14, 18, 19} assessed the accuracy of HC5 estimates between ICE SSD and measured SSDs based on 5-fold differences, which is the difference in acute toxicity values found during inter-laboratory comparisons with the same species (generally between 3 and 5)^{28, 29}. In the current study, a 3-fold difference between HC5 estimates was found to represent a level below which pairs of SSDs may not be statistically different. While it is always advisable to make these determinations based on results from statistical analyses, a 3-fold difference could be used as a conservative and defensible cutoff for determining the adequacy of estimates obtained from ICE SSDs relative to those from empirical data.

The current study also indicated that the use of multiple surrogate species to develop SSDs is preferred over SSDs developed from single surrogate species. The multiple surrogate approach may be advantageous because an increase in species diversity of SSDs generally results in estimates with lower uncertainty (narrower bounds), thus potentially improving HC estimates for chemicals with limited empirical data (e.g., <10 species). However, if data are only available to develop SSDs from single surrogates, selection of ICE models could be based on surrogate species that generate the most accurate estimates of aquatic toxicity and identified in the current study as best or adequate surrogates.

While previous studies have recommended between five and fifteen species to develop accurate SSDs and associated estimates^{15, 18, 30}, CAFE requires a minimum of five species. However, the addition of ICE-predicted values could allow for these recommendations to be incorporated. Despite these potential enhancements to the capabilities of CAFE, there are taxonomic groups that continue to be underrepresented and under-tested such as aquatic macrophytes and corals. As a result, HC values may be more protective of a wider number of crustaceans and fish species as these are the dominant taxonomic groups in CAFE. In addition, the use of HC values as a protective level of the aquatic assemblage as a whole assumes that the sensitivity of untested aquatic species from underrepresented taxonomic groups fall within the range of sensitivities of those tested. When these types of concerns exist, the HC1 could be used as a more protective level, but caution should be used as uncertainty in SSDs estimates increases towards the end tails of the distributions¹. In the past, the ecological relevance of SSDs (i.e., lack of representative community structure) and their applicability in risk assessment has been debated³¹. Although addressing this concern is outside the scope of the current study, increasing species diversity in SSDs with the addition of ICE-predicted values may allow for the selection of species that better reflect the composition of the aquatic community of interest (e.g., based on taxonomic relatedness, functional groups, trophic composition). This same approach could be used when assessing site-specific impacts of chemical releases into aquatic environments.

The current study also provides general insights into the importance, implications and uncertainties in the current understanding of species sensitivity for a variety of chemicals. For example, the U.S. aquatic life ambient water quality criteria for ammonia was only recently updated based on new information on the greater sensitivity of freshwater molluscs³². Similarly, the suite of available ICE models may be missing sensitive species, and the development of additional ICE models will require the collection of aquatic toxicity data for additional sensitive taxa. Furthermore, the fact that ICE_{MOA} did not perform as well as generic ICE models suggests that data for chemicals with specific MOA are also needed to improve predictions. Thus, the development of new ICE models and improved predictive accuracy of existing ones require additional data collection.

As shown here and elsewhere^{10, 14–16, 18, 19}, ICE supplemented SSDs generally produce smaller (more conservative) HC5 estimates with lower uncertainty (narrower bounds) around the estimate than those from measured SSDs, adding confidence in the use of ICE models in augmenting aquatic toxicity data. In addition, ICE SSDs and measured SSDs generally produced comparable estimates, with ICE SSDs under-predicting aquatic toxicity in only a very few instances. Despite these findings, the selection of ICE models and the use

of ICE-predicted values to augment SSDs should be considered on a case-by-case basis as it would ultimately depend on the specific goals of the study or assessment. In the interim, the use of ICE-predicted data should be considered when empirical data are of insufficient quantity and quality, and when all necessary statistical analyses are undertaken to ensure that estimates are scientifically defensible. Larger discussions on the use of predictive modeling, including ICE models, in risk assessment are outside the scope of the current study, but warrant further debate and consideration by the scientific community.

The framework presented in the current study for optimizing selection of ICE models could potentially increase the capabilities of CAFE through the addition of predicted aquatic toxicity values for many chemicals. These efforts also demonstrated that the use of ICE models is a valid approach for generating predicted data and increasing species diversity, thus facilitating the development of SSDs for chemicals with limited toxicity data. While CAFE may continue to have limitations in aquatic toxicity data and species diversity for many chemicals, the incorporation of ICE-predicted values substantially addresses data limitations potentially enhancing its future capabilities. Enhancing CAFE through the use of predictive models may benefit its users by expanding data availability for chemicals and species of local, national and international importance, and of interest to spill response and ecological risk assessment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support of this research was received by A.C. Bejarano from NOAA's ERD. We thank A. Beasley for reviewing a draft of this manuscript, and M. Willming and C. Lilavois for compiling ICE models. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of NOAA's ERD or the USEPA. This publication does not constitute an endorsement of any commercial product.

ABBREVIATIONS

CAFE	the Chemical Aquatic Fate and Effects Database
HC5	5 th percentile hazard concentrations
ICE	Interspecies Correlation Estimation
LC50 and EC50	Median Lethal and Effects concentrations, respectively
MOA	mode of toxic action
MSE	Mean Square Error
SSD	species sensitivity distributions
95% CI	95% Confidence Interval

References

1. Bejarano AC, Farr JK, Jenne P, Chu V, Hielscher A. The chemical aquatic fate and effects database (CAFE), a tool that supports assessments of chemical spills in aquatic environments. *Environ. Toxicol. Chem.* 2016; 35(6):1576–1586. [PubMed: 26497000]
2. NOAA/ERD. The Chemical Aquatic Fate and Effects (CAFE) Database. National Oceanic and Atmospheric Administration, Emergency Response Division, Office of Response and Restoration; Seattle, WA:
3. Solomon KR, Baker DB, Richards RP, Dixon KR, Klaine SJ, La Point TW, Kendall RJ, Weisskopf CP, Giddings JM, Giesy JP. Ecological risk assessment of atrazine in North American surface waters. *Environ. Toxicol. Chem.* 1996; 15(1):31–76.
4. Van den Brink PJ, Blake N, Brock TC, Maltby L. Predictive value of species sensitivity distributions for effects of herbicides in freshwater ecosystems. *Hum. Ecol. Risk. Assess.* 2006; 12(4):645–674.
5. Posthuma, L., Suter, GW., Traas, TP. *Species Sensitivity Distributions in Ecotoxicology*. Lewis Publisher; Boca Raton, FL: 2002.
6. van Straalen NM. Threshold models for species sensitivity distributions applied to aquatic risk assessment for zinc. *Environ. Toxicol. Pharm.* 2002; 11(3–4):167–172.
7. Hose G, Van den Brink P. Confirming the species-sensitivity distribution concept for endosulfan using laboratory, mesocosm, and field data. *Arch. Environ. Contam. Toxicol.* 2004; 47(4):511–520. [PubMed: 15499502]
8. Bejarano AC, Farr JK. Development of short acute exposure hazard estimates: A tool for assessing the effects of chemical spills in aquatic environments. *Environ. Toxicol. Chem.* 2013; 32(8):1918–1927. [PubMed: 23625642]
9. Raimondo S, Mineau P, Barron M. Estimation of chemical toxicity to wildlife species using interspecies correlation models. *Environ. Sci. Technol.* 2007; 41(16):5888–5894. [PubMed: 17874802]
10. Bejarano AC, Barron MG. Development and practical application of petroleum and dispersant interspecies correlation models for aquatic species. *Environ. Sci. Technol.* 2014; 48(8):4564–4572. [PubMed: 24678991]
11. Raimondo S, Jackson CR, Barron MG. Influence of taxonomic relatedness and chemical mode of action in acute interspecies estimation models for aquatic species. *Environ. Sci. Technol.* 2010; 44(19):7711–7716. [PubMed: 20795664]
12. Raimondo, S., Jackson, CR., Barron, MG. Version 3.3. Office of Research and Development, U. S. Environmental Protection Agency; Gulf Breeze, FL: Web-based Interspecies Correlation Estimation (Web-ICE) for acute toxicity: user manual. EPA/600/R-15/192. Accessed from: <https://www3.epa.gov/ceampubl/fchain/webice/index.html>
13. Barron MG, Lilavois CR, Martin TM. MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development. *Aquat. Toxicol.* 2015; 161(0):102–107. [PubMed: 25700118]
14. Bejarano AC, Barron MG. Aqueous and tissue residue-based interspecies correlation estimation models provide conservative hazard estimates for aromatic compounds. *Environ. Toxicol. Chem.* 2016; 35(1):56–64. [PubMed: 26184086]
15. Dyer SD, Versteeg DJ, Belanger SE, Chaney JG, Mayer FL. Interspecies correlation estimates predict protective environmental concentrations. *Environ. Sci. Technol.* 2006; 40(9):3102–3111. [PubMed: 16719118]
16. Dyer SD, Versteeg DJ, Belanger SE, Chaney JG, Raimondo S, Barron MG. Comparison of species sensitivity distributions derived from interspecies correlation models to distributions used to derive water quality criteria. *Environ. Sci. Technol.* 2008; 42(8):3076–3083. [PubMed: 18497169]
17. Raimondo S, Vivian DN, Delos C, Barron MG. Protectiveness of species sensitivity distribution hazard concentrations for acute toxicity used in endangered species risk assessment. *Environ. Toxicol. Chem.* 2008; 27(12):2599–2607. [PubMed: 18699704]
18. Awkerman JA, Raimondo S, Barron MG. Development of species sensitivity distributions for wildlife using interspecies toxicity correlation models. *Environ. Sci. Technol.* 2008; 42(9):3447–3452. [PubMed: 18522132]

19. Awkerman JA, Raimondo S, Jackson CR, Barron MG. Augmenting aquatic species sensitivity distributions with interspecies toxicity estimation models. *Environ. Toxicol. Chem.* 2014; 33(3): 688–695. [PubMed: 24214839]
20. NOAA/ERD. Version 1.1. National Oceanic and Atmospheric Administration, Emergency Response Division, Office of Response and Restoration; Seattle, WA: 2015. Chemical Aquatic Fate and Effects (CAFE) Database: User Manual. p 40 + Appendices
21. D'Agostino, RB., Stephens, MA. Goodness-of-fit techniques. Marcel Dekker, NY and Basel: 1986.
22. Piegorsch, WW., Bailer, AJ. Statistics for environmental biology and toxicology. Vol. 4. Chapman and Hall; London, UK: 1997.
23. R Development Core Team. R: A Language and Environment for Statistical Computing, Version 3.13.2015-03-09. Vienna, Austria: 2015.
24. Gross, J., Ligges, U. nortest: Tests for Normality. R package version 1.0–3. 2015. <http://CRAN.R-project.org/package=nortest>
25. Delignette-Muller ML, Dutang C. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software.* 2015; 64(4):1–34.
26. van der Loo, MPJ. extremevalues: a package for outlier detection in univariate data, R package version 2.0. The Hague, Netherlands: 2010.
27. Willming MM, Lilavois CR, Barron MG, Raimondo S. Acute toxicity prediction to threatened and endangered species using Interspecies Correlation Estimation (ICE) models. *Environ. Sci. Technol.* 2016; 50(19):10700–10707. [PubMed: 27585402]
28. Rand, G., Wells, P., McCarty, L. Chapter 1. Introduction to Aquatic Toxicology. In: Rand, G., editor. Fundamentals of aquatic toxicology effects, environmental fate, and risk assessment. Taylor and Francis Publishers, North Palm Beach, Florida, USA. 1995. p. 3-67.
29. Fairbrother, A. Risk management safety factor. In: Jørgensen, S., Fath, B., editors. Encyclopedia of ecology: ecotoxicology. Vol. 4. Elsevier; Amsterdam: 2008. p. 3062-3068.
30. Stephen, CE., Mount, DL., Hansen, DJ., Gentile, JR., Chapman, GA., Brungs, WA. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. US EPA, Office of Water Regulations and Standards Criteria; Washington DC: 1985. p. PB85-227049.
31. Forbes VE, Calow P. Species sensitivity distributions revisited: a critical appraisal. *Hum. Ecol. Risk. Assess.* 2002; 8(3):473–492.
32. USEPA. Water Quality Criteria for Ammonia - Freshwater 2013. U.S. Environmental Protection Agency; Washington, DC: 2013. EPA 822-R-13-001

Figure 1.
 Overview of steps used to optimize ICE model selection and integration of ICE-predicted values in CAFE. Each step within the framework, builds on the outcomes of the preceding steps. See Figure S1 for details.

Figure 2.

Comparison between HC5 estimates derived from ICE SSDs using all available models (ICE_{All}) and only models that met the specified accuracy criteria ($ICE_{Reduced}$). Models were based on simulated data (A) and empirical data (B). The solid line represents the 1:1 line (equal toxicity), while the dashed lines represent a 3-fold difference between HC5 estimates. Gray symbols show pairs of SSDs that were statistically compared. Data points above the 1:1 line indicate that HC5 estimates from ICE SSD with ICE_{All} models are more conservative.

Figure 3.

HC5 estimates and 95% CIs from single surrogate species SSDs for 56 unique chemical-exposure duration combinations (x-axis, not shown) with limited toxicity data (2–8 records) and at least two surrogate species categorized as best surrogates. Symbol size represents the additional number of ICE-predicted species (smallest=1, largest=53). Black symbols in insert show HC5 estimates from measured SSDs. See SI2 for details.

Figure 4.

HC5 estimates and 95% CIs from multiple surrogate species SSDs for 212 unique chemical-exposure duration combinations (x-axis, not shown) with limited toxicity data (1–10 records). Symbol size represents the additional number of ICE-predicted species (smallest=5, largest=105). Black symbols show HC5 estimates from measured SSDs. See SI2 for details.

Figure 5.

Data gains and improvements resulting from the incorporation of ICE-predicted values into CAFE: (A) assessment based on 2,320 unique chemical-exposure duration combinations where symbol size represents the number of gained ICE-predicted species (smallest=1, largest=99); and (B) assessment based on species diversity for a subset of chemicals (198 unique chemical-exposure duration combinations). Red lines indicate the minimum required number of species needed to develop measured SSDs in CAFE.

Figure 6. Comparison of HC5 estimates and 95% CIs derived from measured SSDs (n=11–64 species) to ICE SSDs from multiple surrogate species (n=7–120 species). The solid line represents the 1:1 line (equal toxicity), and the dashed lines represent a 3-fold difference between estimates.