

Metal Bioavailability Modeling

Development of Empirical Bioavailability Models for Metals

Kevin V. Brix,^{a,*} David K. DeForest,^b Lucinda Tear,^b Willie Peijnenburg,^c Adam Peters,^d Ellie T. Middleton,^e and Russ Erickson^f

^aEcoTox, Miami, Florida, USA

^bWindward Environmental, Seattle, Washington, USA

^cRIVM, Bilthoven, The Netherlands

^dWCA Environment, Faringdon, Oxfordshire, United Kingdom

^eNiPERA, Durham, North Carolina, USA

^fUS Environmental Protection Agency, Duluth, Minnesota

Abstract: Recently, there has been renewed interest in the development and use of empirical models to predict metal bioavailability and derive protective values for aquatic life. However, there is considerable variability in the conceptual and statistical approaches with which these models have been developed. In the present study, we review case studies of empirical bioavailability model development, evaluating and making recommendations on key issues, including species selection, identifying toxicity-modifying factors (TMFs) and the appropriate environmental range of these factors, use of existing toxicity data sets and experimental design for developing new data sets, statistical considerations in deriving species-specific and pooled bioavailability models, and normalization of species sensitivity distributions using these models. We recommend that TMFs be identified from a combination of available chemical speciation and toxicity data and statistical evaluations of their relationships to toxicity. Experimental designs for new toxicity data must be sufficiently robust to detect nonlinear responses to TMFs and should encompass a large fraction (e.g., 90%) of the TMF range. Model development should involve a rigorous use of both visual plotting and statistical techniques to evaluate data fit. When data allow, we recommend using a simple linear model structure and developing pooled models rather than retaining multiple taxa-specific models. We conclude that empirical bioavailability models often have similar predictive capabilities compared to mechanistic models and can provide a relatively simple, transparent tool for predicting the effects of TMFs on metal bioavailability to achieve desired environmental management goals. *Environ Toxicol Chem* 2020;39:85–100. © 2019 SETAC

Keywords: Bioavailability; Metal; Multilinear regression; Water quality criteria; Model

INTRODUCTION

Considerable progress has been made in the last 30 yr in the development of metal bioavailability models for use in setting protective values for aquatic life (PVALs) and conducting aquatic risk assessments (Di Toro et al. 2001; Paquin et al. 2002; Niyogi and Wood 2004; Van Sprang et al. 2009; De Schampheleere and Janssen 2010; Farley et al. 2015; Brix et al. 2017a). Recently, a technical workshop, Bioavailability-Based Aquatic Toxicity Models for Metals, was organized to evaluate the state of the science regarding metal bioavailability models and their regulatory applications, particularly in setting PVALs (for simplicity, we define PVALs as synonymous with other terms such as “water quality criteria,” “water quality guidelines,” and

“environmental quality standards” that are used in various regulatory jurisdictions).

The workshop was organized into 5 working groups, each with specific objectives related to the incorporation of bioavailability in the derivation of PVALs. The objective of our working group and the subject of the present study was to explore the data needs and technical approaches for developing empirical bioavailability models such as multilinear regression (MLR) models (Brix et al. 2017a; DeForest et al. 2018). The other 4 working groups considered the history and scientific basis for incorporation of bioavailability in the assessment of freshwater metal toxicity (Adams et al. 2020), the development and use of mechanistic bioavailability models such as the biotic ligand model (BLM; Mebane et al. 2020), validation methods for metal bioavailability models (Garman et al. 2020), and the application of metal bioavailability models in PVAL derivation (Van Genderen et al. 2020).

Empirical bioavailability models are statistical models developed from experimental data collected to examine

This article includes online-only Supplemental Data.

* Address correspondence to kevinbrix@icloud.com

Published online December 2019 in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/etc.4570

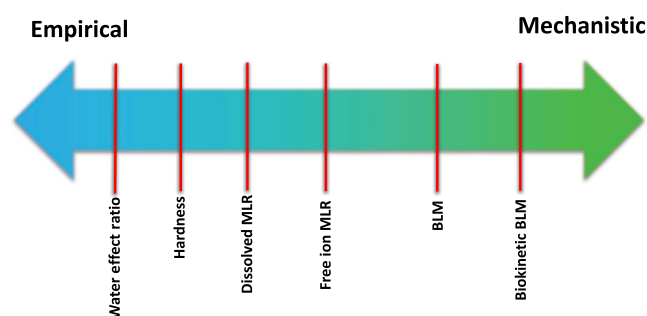


FIGURE 1: Empirical to mechanistic model continuum. BLM = biotic ligand model; MLR = multiple linear regression.

relationships between toxicity and one or more toxicity-modifying factors (TMFs; e.g., pH, calcium, dissolved organic carbon [DOC]). From a mechanistic perspective, TMFs can influence not only the uptake of metals (i.e., what some might consider true bioavailability) but also the physiological response of the organism in a way that makes it more or less sensitive to a metal (e.g., temperature; see Adams et al. [2020] for a more detailed discussion). Clear distinctions do not exist between empirical and mechanistic bioavailability models. Indeed, the range of metal bioavailability models that have been developed can be viewed as an empirical–mechanistic continuum (Figure 1). Models range from simple empirical ratios such as the water effect ratio (US Environmental Protection Agency 2001) to more simple linear and multilinear empirical models (Brix et al. 2017a) to quasi-mechanistic models such as the BLM (Di Toro et al. 2001) and fully mechanistic models such as the biokinetic BLM (Veltman et al. 2014). Many models are actually hybrids having both empirical and mechanistic components or being informed by mechanistic or empirical data in developing model structure (Deleebeeck et al. 2008; Van Regenmortel et al. 2015; Brix et al. 2017a; DeForest et al. 2018).

The earliest empirical bioavailability models were hardness-based PVALs developed by the US Environmental Protection Agency (1985). Similar approaches have been applied in other countries (HMSO 1989) and have continued to be used by the US Environmental Protection Agency (2016). With the development of models such as the BLM that are more mechanistically based and consider the effects of multiple TMFs on metal bioavailability (Di Toro et al. 2001), use of empirical models for setting PVALs has generally been considered less scientifically robust. However, interest in developing empirical bioavailability models that consider multiple TMFs for deriving PVALs has recently increased (Canadian Council of Ministers of the Environment 2016; US Environmental Protection Agency 2018) because they can be more transparent and easier to use than BLM-based approaches. Further, multiple-TMF empirical models are mathematically similar to, and simple extensions of, the widely accepted hardness-based PVALs for metals.

The advantages and disadvantages of empirical and mechanistic models may vary depending on the complexity and interactions of TMFs that modify bioavailability for a given metal, the specific model structure, the model's combination of empirical and mechanistic elements, data requirements and availability, intended model use, and other scientific, practical, and policy

considerations. Given all the factors that can influence model selection, our goal is not to recommend one approach over the other but rather to provide best practice guidance on how to develop empirical bioavailability models for metals once the decision is made to pursue this general modeling framework.

Development of empirical bioavailability models for use in PVAL derivation is an iterative process involving multiple steps: 1) identifying the biological species for model development, 2) identifying TMFs to be considered in model development, 3) determining TMF ranges to be covered by the models, 4) identifying and/or generating appropriate and sufficient data for model development, 5) deriving models for individual species, 6) developing pooled multispecies models for extrapolation to other species, and 7) incorporating models into species sensitivity distributions (SSDs). The present study considers each of these steps, describing how they have been addressed in past efforts and how they should be addressed in future efforts.

SPECIES SELECTION

Except when only a particular species is of concern, PVALs are based on data for a broad array of species representative of an aquatic community. Because models can be developed for only a small subset of these species, the selection of these species needs to consider which are necessary and sufficient for establishing the relationship of PVALs to TMFs.

Protective values focus on sensitive organisms, and the effects of TMFs can vary as a function of metal concentration such that the influence of TMFs can differ for sensitive and insensitive taxa (Santore et al. 2002). Consequently, although it is desirable to have data for species across the full range of metal sensitivity (particularly when the full SSD will be used to develop the PVAL), it is important that species sensitive at concentrations near the PVAL be evaluated to ensure that the model is properly parameterized at the PVAL.

We recommend that models be developed for a minimum of one fish, one invertebrate, and one alga/plant. Currently available data suggest that fish species generally respond to TMFs in a similar manner and that the physiological processes in fish with respect to metal uptake are generally similar across species (Niyogi and Wood 2004; Brix et al. 2016, 2017c). However, invertebrates have greater diversity in ion transport physiology, and differential responses to TMFs have been observed across invertebrate taxa (Schlekat et al. 2010; Esbaugh et al. 2012). As such, information for multiple invertebrate taxa (e.g., crustacean, insect, mollusk) would be desirable. Data on the response of algae and aquatic plants to TMFs are limited to a few species, but available information indicates that their responses to TMFs can differ substantially compared to fish and invertebrates (De Schamphelaere et al. 2005).

IDENTIFICATION OF TMFs

Selection of TMFs for empirical bioavailability models should include both existing empirical evidence as well as mechanistic considerations regarding regulation of metal bioavailability and

toxicity. The chemical speciation of metals in water is of central importance to bioavailability because different metal species will be more or less readily taken up by an organism; thus, those TMFs important to speciation need consideration. This includes information on mechanisms of metal uptake and toxicity, in determining both how metal speciation might affect uptake and how physicochemical variables of exposure water might otherwise affect uptake and toxicity. Attention should be given to correlations among TMFs that can make it difficult to separate the effects of individual TMFs and lead to erroneous model formulations if not considered.

Dissolved organic carbon, pH, and hardness are often identified as TMFs for divalent metals (Niyogi and Wood 2004). Water hardness is generally a function of Ca and Mg concentrations, although other elements can contribute. Calcium is often the primary component of hardness influencing metal bioavailability (Heijerick et al. 2002), although Mg can also influence bioavailability for some metals (Peters et al. 2011; Brix et al. 2017b). Use of water hardness as a TMF incorporates effects from both ions but does not consider variation in Ca:Mg ratios, which can substantially influence hardness-dependent metal toxicity (Welsh et al. 2000). Similarly, the chemical composition of DOC can vary, affecting its metal-binding capabilities and thus its effect on toxicity (Al-Reasi et al. 2011).

It is important to recognize, however, that DOC, pH, and hardness may not be important TMFs for some metals and/or taxonomic groups. Various forms of nitrogen and phosphorus are important TMFs for algae and plants (Lee and Wang 2001). For metals and metalloids that occur predominantly as oxyanions, chemically similar anions can affect bioavailability. Sulfate (SO_4^{2-}), for example, has been reported to be an important TMF for selenate (SeO_4^{2-}) for fish, invertebrates, and algae (Brix et al. 2001; DeForest et al. 2017).

Interestingly, perhaps the most ubiquitous TMF is temperature (Sokolova and Lannig 2008), but this is rarely considered in empirical or mechanistic bioavailability models (Erickson et al. 1987; Santore et al. 2018). Most model development (empirical and mechanistic) involves evaluating how well a model predicts laboratory toxicity test data where test temperatures are largely standardized to generate toxicity data that can be compared across chemicals. Consequently, temperature has not typically been identified as an important explanatory variable in normalizing metal toxicity data. However, application of these models to the field where temperature is much more variable could potentially lead to significant under- or overestimation of toxicity. This is a consideration that needs to be addressed in future bioavailability studies and models.

In addition to the TMFs already described, which may vary by taxon and metal, there are likely other TMFs associated with each species–metal combination that have relatively minor influences on metal toxicity. The relative importance and need to assess these “minor” TMFs will vary and can be assessed statistically and using best professional judgment. Importantly, we do not recommend blindly using the results of statistical tests to make decisions regarding inclusion/exclusion of TMFs and caution against overparameterizing models which can lead to a model that is not sufficiently generalized to other data sets to

which it may be applied. For example, Erickson et al. (1987) demonstrated that Na is a TMF for Cu toxicity to *Pimephales promelas*. However, Na reduces Cu toxicity only at concentrations >1 mM Na, and toxicity is reduced by slightly less than a factor of 2 in the presence of 4 mM Na. Given that approximately 65% of surface waters in the United States have <1 mM Na and approximately 90% of surface waters have <4 mM Na, inclusion of Na in an empirical model for *P. promelas* will have a relatively minor effect on toxicity predictions except in unusually saline freshwaters. Further discussion of how best to assess the inclusion or exclusion of TMFs can be found in later sections (see *Development of Species-Specific Models* and *Development of Pooled Species Models*).

TMF RANGES

For the development and application of empirical bioavailability models, various linked issues regarding TMF ranges need to be considered. First, it is important to consider how the minimum TMF range necessary for a statistically reliable model relates to the range of water quality conditions to which the model will be applied. Because empirical models should generally be used only within the range of water quality for which they have been evaluated, one should assess the collective scope of applicability for available models and how this range differs from the desired scope of applicability for a PVAL. After model development, recommendations should be made to indicate if PVALs need to be restricted in any way until more data are generated. It is important to specify the range of parameters to which a model applies to avoid misuse and misinterpretation of data outside the intended TMF range, especially if this leads to the PVAL not being protective.

Issues regarding TMF ranges are evident in the development and application of the early US Environmental Protection Agency (USEPA) hardness-dependent water quality criteria (WQC) for metals. To support adequate estimation of hardness slopes, the US Environmental Protection Agency (1985) recommended that hardness concentrations for tests for a given species span at least 100 mg L^{-1} , with the highest concentration at least 3-fold greater than the lowest concentration; but these minimum requirements might cover less than half of the range of hardness to which WQC might be applied. However, because multiple species satisfying these range requirements were combined into pooled models, the resultant models generally had broader ranges than these minimum data requirements. This broader range was still less than the hardness range to which WQC might apply though, so restrictions were necessary regarding WQC outside this range (e.g., US Environmental Protection Agency 2016).

More recently, with regard to minimum ranges needed for reliable model development, in developing MLR models for Cu and Zn, Brix et al. (2017a) and the Canadian Council of Ministers of the Environment (2016) required minimum ranges for hardness (100 mg L^{-1}), pH (1.5 units), and DOC (5 mg L^{-1}) for a given species. In developing an empirical bioavailability model for PVALs, consideration should be given to ranges of TMFs that reflect wide environmental relevance. The USEPA does not

provide specific guidance regarding this, but in the European Union the 5th to 95th percentiles of TMFs are recommended (European Commission 2011). For a model with 3 TMFs that are not significantly correlated, testing within the 5th to 95th percentiles for all of the TMFs will include only approximately 75% of the total range of water compositions, whereas testing within the 10th to 90th percentiles will only include approximately 50% of the total range of water compositions. Alternatively, if TMFs are correlated, testing at the 5th and 95th percentiles for multiple TMFs might involve water compositions that do not naturally occur in the environment. For example, a water with low pH (e.g., 6.5) and high hardness (e.g., 420 mg L⁻¹) is unlikely to occur naturally. However, it is important to recognize that regulatory applications will need to address unusual conditions, and in some cases, model formulation might benefit from testing certain extreme compositions. For example, the low-pH/high-hardness scenario just described is unlikely to occur naturally but can occur in effluent-dominated streams receiving wastewater from metal and other resource-extraction activities. Consequently, development of a model that includes these conditions is desirable for regulatory application.

Resolving issues of TMF ranges would benefit from more rigorous analysis of the water compositions to which PVALs should apply. As an example of a more comprehensive analysis for characterizing ranges of multiple TMFs, we evaluated a compilation of over 20 000 water chemistry data from the Water Quality Portal for several potential metal TMFs spanning 65 of the 84 ecoregions in the United States (A.C. Ryan, International Zinc Association, Syracuse, New York, USA, unpublished; Table 1). Figure 2 provides scatter plots of these data for pH versus hardness, DOC versus hardness, and pH versus DOC and shows the 5th and 95th percentiles of each TMF to illustrate the coverage provided by such limits. This figure also illustrates how the correlation between hardness and pH should be considered in selecting realistic combinations of these variables, but even when variables are not substantially correlated (DOC vs either pH or hardness), the utility of testing combinations at extreme percentiles of both variables that are unlikely to occur may be questionable. Because the >20 000 points in these plots are densely overlaid in the middle of the distributions, it is difficult to assess how well existing data for a given metal may cover the distribution of these variables and how to design experiments to generate new data to improve coverage (see section *Data Set Development*). To better illustrate suitable ranges for variables, the

joint probability distributions for pairs of these variables were characterized by kernel density estimation (KDE; Silverman 1986). Results from the KDE analysis were used to define constant-density contours enclosing 90 and 99% of the data

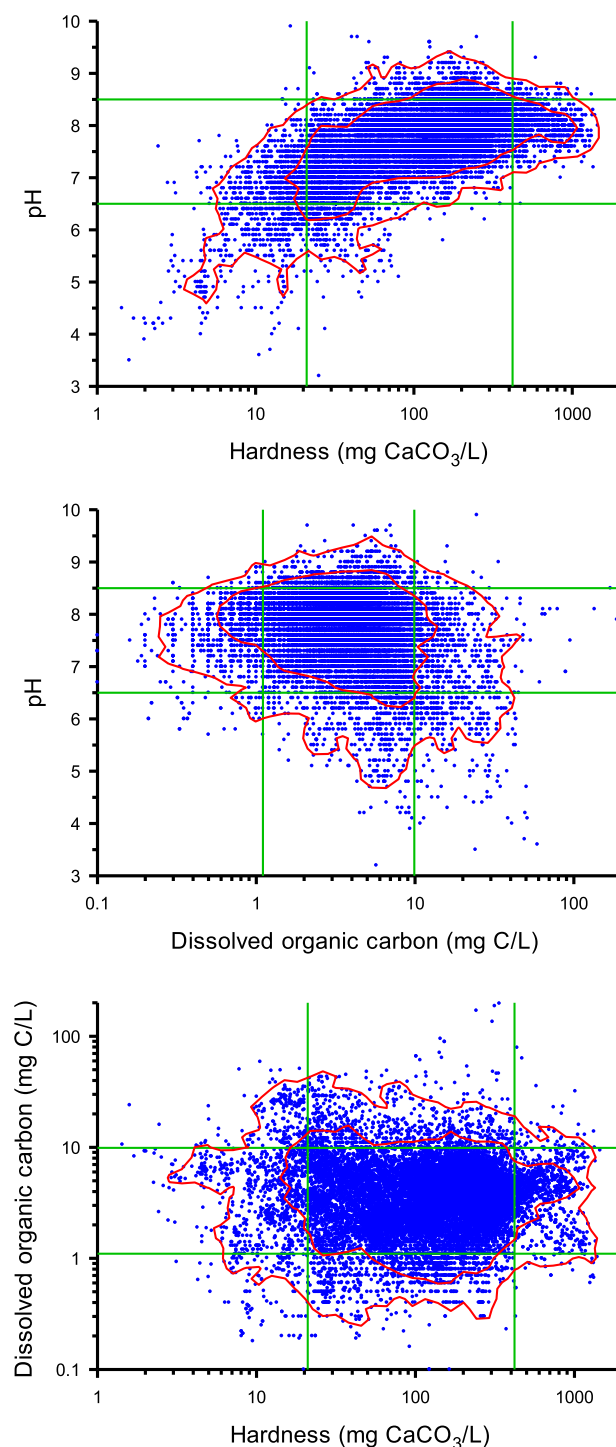


FIGURE 2: Pairwise relationships between hardness, pH, and dissolved organic carbon in natural waters occurring across the United States ($n = 22\,076$). Green lines represent 5th and 95th percentiles of each parameter. Red lines denote constant-density contours enclosing 90% and 99% of the data based on kernel density estimation (see Supplemental Data for analysis methods). Data from A.C. Ryan, International Zinc Association, Syracuse, New York, USA (unpublished).

TABLE 1: Summary distribution of select toxicity-modifying factors in the United States^a

Percentile	DOC	pH	Hardness	Ca	Mg	Na	Alkalinity
5th	1.1	6.5	21	5.4	1.3	1.6	17
10th	1.5	7.0	31	8.6	1.8	2.5	27
50th	3.5	7.8	136	37	9.7	14	103
90th	7.7	8.3	324	84	28	97	210
95th	9.9	8.5	420	105	38	160	240

^aAll units are in milligrams per liter except pH (A.C. Ryan, International Zinc Association, Syracuse, New York, USA, unpublished).
DOC = dissolved organic carbon.

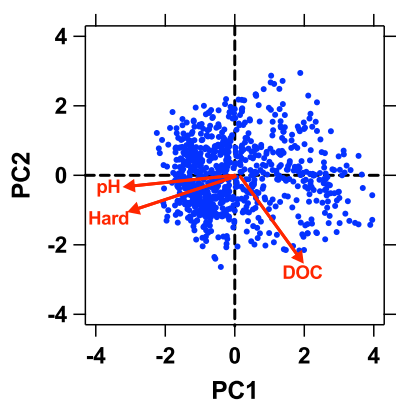


FIGURE 3: Example of a principal component analysis of toxicity-modifying factors within a specific geographical region. Hard = hardness; DOC = dissolved organic carbon; PC = principal component.

(red lines in Figure 2; see Supplemental Data for method details). These contours allow visualization of more appropriate parameter values for testing; in particular, the contours for the pH/hardness relationship illustrate how the rectangle defined by the 5th and 95th percentiles of the individual parameters includes some improbable parameter combinations, while excluding some important ones.

Another option for processing these data into useful descriptions of environmentally realistic combinations of TMFs is the use of principal component analysis (PCA; Jolliffe 2002). Scatterplots of the principal component scores from a PCA based on monitoring data for all TMFs considered can be used to characterize the space occupied by naturally occurring combinations of TMFs (Figure 3). The test set (i.e., the optimal set of TMF combinations) may be further refined by focusing on specific areas within the PCA plot to ensure that waters likely to have high bioavailability are tested. Similar to the KDE analysis, ranges developed from a PCA-based approach may be quite different for some parameters from predefined ranges (percentiles) for the univariate TMF. As for any method of selecting TMF ranges, combinations of TMFs selected for toxicity testing should not induce physiological stress to the test organisms.

DATA SET DEVELOPMENT

Data set development for empirical bioavailability models will vary depending on existing data and on assessment goals. The models in Table 2 were developed using data from individual laboratory studies in which one or more TMFs were systematically varied whereas others were held constant, from multiple studies that collectively cover a range of TMFs, and from combinations of both approaches. Model development sometimes has relied solely on existing data but at other times has included generation of new supplementary data to comprehensively address model needs. The amount of data used in model development has varied depending on their source and quality. This section will describe some of these data sets to identify noteworthy issues in their development and then discuss aspects of experimental design for generating new data to augment available data sets.

For a single TMF, the number and range of data points should be sufficient to define the shape of the relationship and to encompass the range of applicability desired for the PVAL. Organisation for Economic Co-operation and Development (2007) guidance recommends a minimum of 5 data points per independent variable for univariate regression analyses. If suitably spaced and satisfying considerations regarding the TMF range (see previous section), 5 data points would support developing an empirical bioavailability model for a single TMF, including assessing the need for a nonlinear model; and even fewer points might suffice depending on prior knowledge regarding the form of the model equation, the quality of the data, and the assessment needs.

However, attention must also be given to how the data were obtained. If the data come from multiple studies, with no single study covering a substantial range of the TMFs, differences in sensitivity across the studies could affect results, and more data would be necessary to test for study effects and provide confidence in a final model. Even if the data come from a single study in which a single TMF was systematically varied, care must be taken that the TMF was not confounded by other factors that might modify toxicity. For example, several studies on the effect of hardness on Ag toxicity, in which hardness was varied using chloride salts, were confounded by chloride complexation of Ag until this effect was identified (Bury et al. 1999). A model resulting from such data would only be applicable to waters in which hardness was similarly correlated with chloride.

For multiple TMFs, the data set must consider not only the separate data issues for each TMF but also the possibility of interactions among the TMFs. Brix et al. (2017a), for example, used all available data for a given species to define relationships between Cu toxicity and multiple TMFs (hardness, DOC, and pH). This involved large data sets with broad coverage for each TMF. Although this data set was able to support estimation of TMF interactions, correlations and unevenness in the data coverage can distort model parameters, so methods described in that study for assessing model fit should be used to assure that model assumptions are met and that fit is adequate for any study-related TMF or test species subsets within the larger data set. Tools for addressing such data set limitations are discussed in the section *Development of Species-Specific Models*, some of which were employed by Brix et al. (2017a).

In contrast, DeForest et al. (2018) based the development of a model for Al toxicity to the alga *Pseudokirchneriella subcapitata* on a single study by Gensemer et al. (2018), which tested toxicity in a complete factorial design with 3 levels for each TMF (hardness, pH, DOC). Although the TMF ranges were more restricted than would be desirable for WQC, this data set still provided broad coverage of the TMFs and established clear relationships for model development, with model predictions nearly always being within the error bounds of observed values. Even with just 3 levels for each TMF, the total of 27 TMF combinations and the consistency of effects observed across the entire matrix argue for the sufficiency of this particular data set and the potential strengths of the factorial study design for testing across multiple TMFs (including reduction of cross-study variance).

TABLE 2: Summary of empirical metal bioavailability models

Metal/species	Endpoint	Model ^a	n	Adjusted R ²	Reference
Aluminum					
<i>Ceriodaphnia dubia</i>	Chronic	$\ln(\text{EC20}) = -41.026 + (0.525 \times \ln[\text{DOC}]) + (2.201 \times \ln[\text{Hard}]) + (11.282 \times \text{pH}) - (0.663 \times \text{pH}^2) - (0.262 \ln[\text{Hard}] \times \text{pH})$	23	0.73	DeForest et al. (2018), US Environmental Protection Agency (2017)
<i>Pimephales promelas</i>	Chronic	$\ln(\text{EC20}) = -14.029 + (0.503 \times \ln[\text{DOC}]) + (3.443 \times \ln[\text{Hard}]) + (3.131 \times \text{pH}) - (0.494 \ln[\text{Hard}] \times \text{pH})$	22	0.87	DeForest et al. (2018), US Environmental Protection Agency (2017)
<i>Pseudokirchneriella subcapitata</i>	Chronic	$\ln(\text{EC20}) = -61.952 + (1.678 \times \ln[\text{DOC}]) + (4.007 \times \ln[\text{Hard}]) + (17.019 \times \text{pH}) - (1.020 \times \text{pH}^2) - (2.04 \ln[\text{DOC}] \times \text{pH}) - (0.556 \ln[\text{Hard}] \times \text{pH})$	27	0.96	DeForest et al. (2018), US Environmental Protection Agency (2017)
Copper					
<i>Pimephales promelas</i>	Acute	$\log(\text{LC50}) = -0.308 + (0.192 \times \text{pH}) + (0.136 \text{pH} \times \log \text{DOC})$	21	0.92 ^b	Welsh et al. (1993, 1996)
<i>Pimephales promelas</i>	Acute	$\log(\text{LC50}) = -0.856 + (0.166 \times \text{pH}) + (0.095 \text{pH} \times \log \text{DOC}) + (0.237 \times \text{Ca})$	18	0.82 ^b	Welsh et al. (1993, 1996)
<i>Pimephales promelas</i>	Acute	$\log(\text{LC50}) = -1.003 + (0.195 \times \text{pH}) + (0.104 \text{pH} \times \log \text{DOC}) + (0.211 \times \text{Ca})$	17	0.82 ^b	Welsh et al. (1993, 1996)
<i>Pimephales promelas</i>	Acute	$\log(\text{LC50}) = -1.003 + (0.195 \times \text{pH}) + (0.104 \text{pH} \times \log \text{DOC}) + (0.211 \times \text{Ca})$	18	0.79 ^b	Welsh et al. (1993, 1996)
<i>Pimephales promelas</i>	Acute	$\log(\text{LC50}) = -0.426 + (0.061 \times \text{pH}) + (0.005 \text{pH} \times \log \text{Color}) + (0.116 \times \text{Hard})$	17	0.89 ^b	Welsh et al. (1993, 1996)
<i>Daphnia magna</i>	Acute	$\text{EC50}_{\text{Cu}^{2+}} (\text{nM}) = 308 + (42.6 \times \text{Ca} [\text{mM}]) - (41.3 \times \text{pH})$	25	0.53 ^b	De Schampelaere et al. (2002)
<i>Daphnia magna</i>	Chronic	$\text{NOEC} = -160.5 + (7.652 \times \text{DOC} [\text{mg/L}]) + (25.50 \times \text{pH})$	35	0.75 ^b	De Schampelaere and Janssen (2004)
<i>Daphnia magna</i>	Chronic	$\text{EC50} (\mu\text{g/L}) = -212.4 + (10.41 \times \text{DOC} [\text{mg/L}]) + (34.36 \times \text{pH})$	35	0.76 ^b	De Schampelaere and Janssen (2004)
<i>Pomacea paludosa</i>	Acute	$\log(\text{LC50}) = 0.540 + (0.008 \times \text{Age} [\text{d}]) + (0.024 \times \text{DOC}) + (0.120 \times \text{pH})$	14	0.92 ^b	Rogevich et al. (2008)
<i>Ceriodaphnia dubia</i>	Acute	$\ln(\text{LC50}) = -9.535 + (6.703 \times \ln[\text{DOC}]) + (0.144 \times \ln[\text{Hard}]) + (1.511 \times \text{pH}) - (0.776 \ln[\text{DOC}] \times \text{pH})$	87	0.78	Brix et al. (2017a)
<i>Daphnia magna</i>	Acute	$\ln(\text{LC50}) = -4.005 + (0.947 \times \ln[\text{DOC}]) - (0.254 \times \ln[\text{Hard}]) + (0.628 \times \text{pH}) + (0.101 \ln[\text{Hard}] \times \text{pH})$	307	0.87	Brix et al. (2017a)
<i>Daphnia obtusa</i>	Acute	$\ln(\text{LC50}) = -6.245 + (4.224 \times \ln[\text{DOC}]) + (0.139 \times \ln[\text{Hard}]) + (1.131 \times \text{pH}) - (0.353 \ln[\text{DOC}] \times \text{pH}) - (0.171 \ln[\text{DOC}] \times \ln[\text{Hard}])$	53	0.89	Brix et al. (2017a)
<i>Daphnia pulex</i>	Acute	$\ln(\text{LC50}) = -9.932 + (0.6931 \times \ln[\text{DOC}]) + (0.172 \times \ln[\text{Hard}]) + (1.502 \times \text{pH}) - (0.782 \ln[\text{DOC}] \times \text{pH})$	35	0.92	Brix et al. (2017a)
<i>Pimephales promelas</i>	Acute	$\ln(\text{LC50}) = -6.744 + (1.620 \times \ln[\text{DOC}]) + (1.065 \times \ln[\text{Hard}]) + (0.925 \times \text{pH}) - (0.241 \ln[\text{DOC}] \times \ln[\text{Hard}])$	206	0.80	Brix et al. (2017a)
<i>Daphnia magna</i>	Chronic	$\ln(\text{EC20}) = 0.200 + (0.848 \times \ln[\text{DOC}]) + (0.235 \times \ln[\text{Hard}]) + (0.172 \times \text{pH})$	77	0.87	Brix et al. (2017a)
Pooled model	Acute	$\ln(\text{LC50}) = \text{Intercept}_{\text{species}^i} + (0.786 \times \ln[\text{DOC}]) + (0.582 \times \ln[\text{Hard}]) + (0.966 \times \text{pH})$	688	0.81	Brix et al. (2017a)
Lead					
<i>Ceriodaphnia dubia</i>	Acute	$\ln(\text{LC50}) = -6.994 + (0.589 \times \ln[\text{Ca}]) + (0.992 \times \ln[\text{DOC}]) + (0.745 \times \text{pH})$	23	0.82	Esbaugh et al. (2011)
<i>Pimephales promelas</i>	Acute	$\ln(\text{LC50}) = 0.932 + (0.784 \times \ln[\text{DOC}]) + (0.443 \times \ln[\text{ionic strength} (\text{mM})])$	25	0.63	Esbaugh et al. (2011)
<i>Ceriodaphnia dubia</i>	Chronic	$\ln(\text{EC50}) = -5.168 + (1.001 \times \ln[\text{DOC}]) + (0.322 \times \ln[\text{TCO}_2] (\mu\text{M})) + (0.371 \times \ln[\text{Na}]) (\mu\text{M})$	22	0.55	Esbaugh et al. (2012)
<i>Lymnaea stagnalis</i>	Chronic	$\ln(\text{EC50}) = -2.919 + (1.089 \times \ln[\text{DOC}]) (\mu\text{M})$	7	0.82	Esbaugh et al. (2012)
<i>Philodina rapida</i>	Chronic	$\ln(\text{EC50}) = 19.376 - (3.018 \times \text{pH}) + (1.503 \times \ln[\text{Ca}]) (\mu\text{M})$	6	0.92	Esbaugh et al. (2012)
Zinc					
<i>Daphnia pulex</i>	Acute	$\ln(\text{LC50}) = 3.1\% + (0.284 \times \ln[\text{DOC}]) + (0.845 \times \ln[\text{Hard}])$	25	0.58	Canadian Council of Ministers of the Environment (2016)
<i>Daphnia magna</i>	Acute	$\ln(\text{LC50}) = 3.083 + (0.191 \times \ln[\text{DOC}]) + (0.865 \times \ln[\text{Hard}])$	7	0.97	Canadian Council of Ministers of the Environment (2016)
<i>Daphnia</i> spp. Pooled	Acute	$\ln(\text{EC50}) = 3.224 + (0.240 \times \ln[\text{DOC}]) + (0.833 \times \ln[\text{Hard}])$	32	0.81	Canadian Council of Ministers of the Environment (2016)
<i>Oncorhynchus mykiss</i>	Chronic	$\ln(\text{LC10}) = 7.041 + (0.999 \times \ln[\text{DOC}]) + (0.886 \times \ln[\text{Hard}]) - (0.937 \times \text{pH})$	14	0.79	Canadian Council of Ministers of the Environment (2016)

^aUnless otherwise noted, units for metals are micrograms per liter and units for DOC, hardness, and calcium are milligrams per liter.
^bAdjusted R² was calculated from the R² based on sample size (n) and the number of parameters (k) included in the model: Adj. R² = $1 - \left(\frac{(1 - R^2)(n - 1)}{n - k - 1} \right)$.
 Models based only on hardness or with adjusted R² < 0.50 are excluded.

DOC = dissolved organic carbon; ECx = x% effect concentration; LCx = x% lethal concentration.

Empirical bioavailability models should be based on existing data but generate, as needed, new data that will improve model performance and applicability while minimizing testing efforts to reduce the use of resources and test animals. When existing data are found to be inadequate, new systematic experiments should be conducted. New data can come from a single factorial design in which all combinations of low and high values for each TMF are tested (e.g., 5th and 95th percentiles in Figure 2) or include TMF combinations that have not been adequately tested in previous studies. For 3 TMFs, this constitutes 8 treatment combinations at the corners of a cube in the 3-dimensional TMF space. Although a minimal study with just these 8 treatments could be conducted (perhaps including a treatment with intermediate values for all TMFs, i.e., the center of the cube), this would not support evaluating nonlinearities and interactions or meet reasonable standards for the number of data points per variable (Organisation for Economic Co-operation and Development 2007). Including an intermediate value for each TMF in a full factorial design would entail 27 tests (as in the Gensemer et al. [2018] algal tests). A less data-demanding option would be to include the TMF combinations at the center of each face of the cube, thereby testing 3 levels of each TMF at the intermediate values for the other 2. This would involve 15 total tests rather than the 27 tests for the full factorial design and satisfy the Organisation for Economic Co-operation and Development recommendation that a minimum of 3 data points over an adequate range is needed to discriminate between linear and nonlinear responses. Ultimately, model performance will dictate whether the number of TMF combinations tested is sufficient, and initial model development will provide feedback regarding additional data generation.

Simple factorial designs like those just described have the drawback that the corners of the factorial cube could represent unlikely combinations of TMFs (see section *TMF Ranges*). To avoid this problem and to incorporate environmental realism, we recommend investigating the distribution of actual field data (Figure 2) to select more realistic combinations. The joint distributions of variables, as in Figure 2, can be assessed (e.g., by KDE) to describe surfaces in the TMF space that enclose a certain percentage (e.g., 95%) of the data. The experimental design could then test effects at systematic points within the joint space of all TMFs and allow both main effects and interactions to be estimated. The same approach could be conducted via PCA, as discussed in the *TMF Ranges* section, to select TMF combinations that reflect the interactions of the TMFs embodied in the principal components (Figure 3).

DEVELOPMENT OF SPECIES-SPECIFIC MODELS

Developing empirical bioavailability models is usually an iterative process involving both statistical and qualitative evaluation of issues such as model performance, validity of model assumptions, and perceived costs and benefits involved in development and application of a model. For example, what kind of evidence should be required to justify the use of linear versus

nonlinear models, the inclusion of interaction terms in a model, or use of data from a study that did not vary all water quality parameters? How should models be compared to assess the benefits of one over the other? Can mechanistic knowledge of metal bioavailability (speciation, uptake pathways, and toxicity mechanisms) inform empirical model development?

Given differences among metals and species in the mechanisms of toxicity and availability of data to develop empirical models; a lack of formal and prescriptive criteria for model development, and the varieties of models to consider, it is not feasible to describe a fixed process for developing empirical bioavailability models. However, based on previous experience, it is possible to suggest some particularly useful tools for testing assumptions and evaluating model performance. These should be incorporated into the iterative model development process and used to develop an awareness of the strengths and limitations of a model that will affect how it is applied.

What form of model should be used and how to decide?

The first steps of model development should involve exploratory analyses. Graphical analyses should be used liberally to develop an understanding of distributions and relationships among variables and help identify unusual cases that may affect model results. Graphical analyses are a critical component in every stage of model development because they reveal adequacies and inadequacies of the data and the model, whether statistical performance criteria are capturing important aspects of model fit, whether assumptions of the model and the mechanistic conceptual framework are consistent, and numerous other issues that can be overlooked without visualizing the data and model. All of the tools of exploratory data analysis (e.g., boxplots, cumulative frequency distributions, color-coding data by subsets such as species, region, study) should be considered for helping understand the adequacy of the model structure and the data (see Garman et al. [2020] for examples of these approaches).

Based on current experience with metal bioavailability models, standard linear models should generally be considered first. The USEPA guidelines for deriving aquatic life WQC (US Environmental Protection Agency 1985) consider the linear relationships of log metal effect concentration (EC) versus log hardness that were recognized at the time and provide a methodology for regression of the data to develop metal WQC as a function of hardness. Subsequent efforts have addressed additional TMFs, such as pH and DOC (Welsh et al. 1996; Rogevich et al. 2008; Esbaugh et al. 2011, 2012; Fulton and Meyer 2014; Brix et al. 2017a; DeForest et al. 2018). In Canada, an MLR-based draft water quality guideline for Zn has been issued (Canadian Council of Ministers of the Environment 2016), and the USEPA recently released a final aquatic life WQC for Al (US Environmental Protection Agency 2018) primarily based on the MLR model of DeForest et al. (2018).

Methods for assessing performance of linear models are highly developed (Harrell 2015). Within the linear framework, it is important to determine the need for transformation of any of the dependent or independent variables to create linearity in

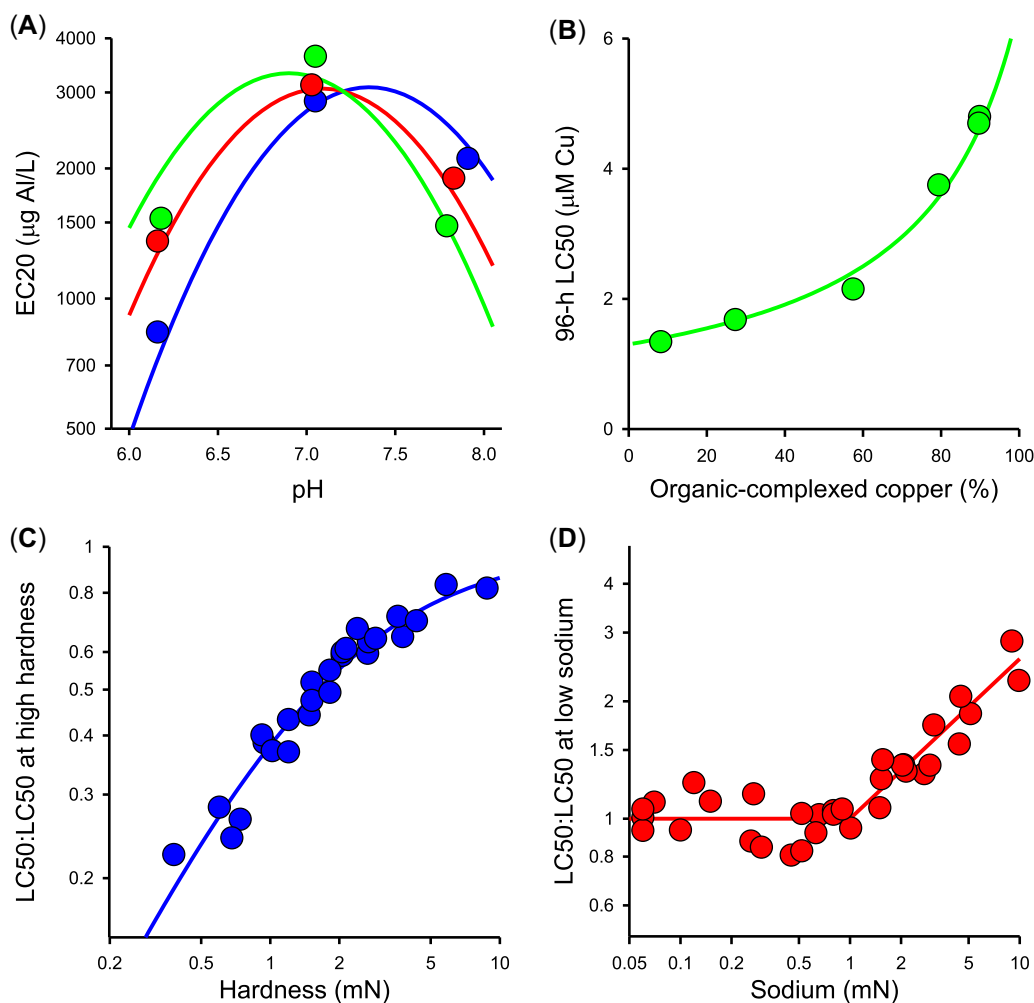


FIGURE 4: Examples of nonlinear relationships between toxicity-modifying factors and metal toxicity. (A) Relationship between pH and Al toxicity to *Raphidocelis subcapitata* (dissolved organic carbon = 4 mg L⁻¹, hardness = 22 [blue], 60 [red], and 120 [green] mg L⁻¹ (DeForest et al. 2018)). (B) Relationship between organic-complexed Cu and Cu toxicity to *Pimephales promelas* (Erickson et al. 1987). (C) Relationship between hardness and Cu toxicity to *P. promelas* (Erickson et al. 1987). (D) Relationship between Na concentrations and Cu toxicity to *P. promelas* (Erickson et al. 1987). EC20 = 20% effect concentration; LC50 = median lethal concentration.

the response and homogenize variance. It is also important to assess collinearity among independent variables via correlation tables, variance inflation factors (Zuur et al. 2010), unexpected signs of coefficients, significance of individual parameters and the model as a whole (p values), and model performance (e.g., adjusted R^2 , Akaike and Bayesian Information Criteria [AIC/BIC]; Burnham and Anderson 2004). Within each of these arenas, comparisons of model performance criteria across models and of the distribution of residuals are the main tools for evaluating different model formulations.

Exploratory data analysis or evaluation of deviations from linear models can indicate nonlinear relationships between ECs and TMFs. For example, DeForest et al. (2018) reported that aluminum EC20s for the alga *Raphidocelis subcapitata* were higher at pH 7 than at both pH 6 and 8 (Figure 4A). The authors attributed this nonlinearity to pH-dependent changes in Al speciation and addressed it by including a pH $\times$ pH (i.e., pH²) interaction term in the model, providing a parabolic shape to the relationship with a maximum within the pH 7.0 to 7.5 range. This

model is mathematically linear relative to the set of independent variables (individual TMFs and their interactions) and can therefore still be addressed by linear statistical tools. However, inclusion of a squared term should only be undertaken when data convincingly demonstrate a maximum or minimum within an environmentally relevant range for the TMF because if there is only a reduction in slope as a TMF increases, a parabolic shape might be erroneously inferred when asymptotic behavior would be more appropriate. Furthermore, such a parabolic shape makes extrapolation beyond the range of the data from which it was derived particularly uncertain. For similar reasons, higher-order polynomials should be avoided in favor of models that functionally address observed nonlinearities.

Sometimes relationships between ECs and TMFs are inherently nonlinear and unsuited to linear statistical analysis tools. Erickson et al. (1987) modeled the relationships between Cu toxicity to fathead minnows (*P. promelas*) and several TMFs. Log ECs varied linearly with some TMFs over the range tested (e.g., pH, temperature), but the relationships for other TMFs were

inherently nonlinear. Nonlinear models that were applied included a joint toxicity relationship for Cu complexed and not complexed by DOC (Figure 4B), an asymptotic relationship with declining slope for the effect of hardness (Figure 4C), and a linear relationship of log EC to log [Na] but with a threshold below which Na had no effect (Figure 4D). When one or more of the relationships of the EC to TMFs is inherently nonlinear, the selected equations should, as simply as possible, incorporate important attributes of the data, for which Erickson et al. (1987) provide some examples (e.g., linearity of log EC vs log hardness at low hardness but declining slopes at high hardness; Figure 4C). Analyses like those described in the present study for linear models should be conducted to establish that deviations from linearity are significant and to test the relative merits of different possible nonlinear models. Formulating nonlinear models often requires ECs at several TMF values well distributed over a broad range, so when data are limited, a linear model will often have to serve as the best possible approximation.

How are model residuals distributed and how do these distributions compare across models?

Residual analysis is a primary tool for revealing interactions among parameters and should be conducted for all independent variables in the model (e.g., species, DOC, hardness, pH, interactions) as well as the dependent variable using

a variety of approaches to ensure that important patterns are detected. In addition to scatterplots of predicted versus observed values, quantile–quantile (QQ) plots of the residuals, scatterplots of residuals versus predicted values and each independent variable, and plots of residuals of one model versus residuals of another (when comparing models) can all provide slightly different perspectives on residual distributions and lead to insights about how a model might be refined (e.g., Figure 5).

When investigating residuals, homogeneity of variance and normality should be evaluated to verify that model assumptions have been met. It is especially important to investigate patterns that indicate environmental conditions where the model does not fit well. This will identify areas where data for water quality conditions or species are not adequate to develop strong model relationships and assess the need for additional model terms (interactions, other variables, etc.). Patterns of residuals should be investigated to help identify inadequately parameterized effects of variables already in the model (e.g., need for interactions) as well as effects of variables that were not considered for which data are available. Residual plots should also be used to evaluate whether individual or a few data points may be creating a statistically significant effect that is inconsistent with our understanding of chemical and physiological mechanisms.

The distribution of residuals for important subsets of the data should be explored to understand how differences among

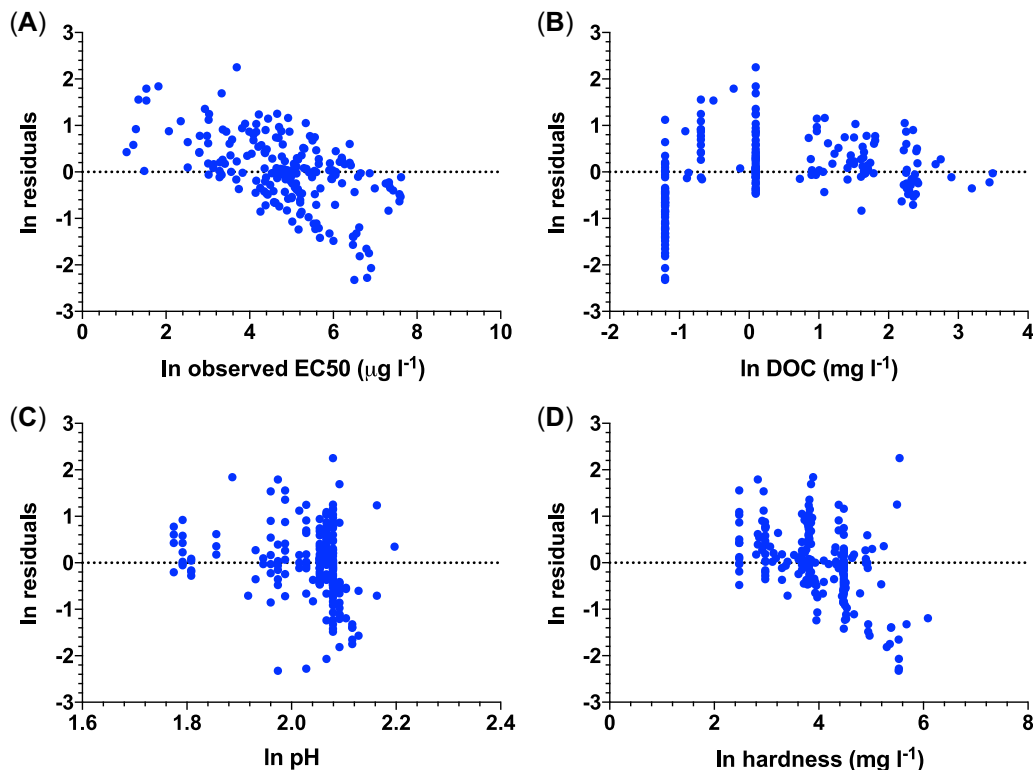


FIGURE 5: Residuals for *Pimephales promelas* based on the pooled Cu multiple linear regression in Brix et al. (2017a) plotted as a function of (A) observed median lethal concentrations (LC50s), (B) DOC, (C) pH, and (D) hardness. Note the similar structure in model residuals as a function of LC50 and hardness. This suggests that the pooled model is overpredicting the effect of hardness on Cu toxicity to *P. promelas*. DOC = dissolved organic carbon; EC50 = median effect concentration.

these subsets affect overall model performance and how well the subsets are predicted by the model. For example, the effects of individual species on a pooled model or the effects of individual studies on a species-specific model should be investigated. Boxplots and QQ plots of residuals for different subsets of the data and color-coding points in residual plots by data subset are all useful tools for understanding how the model may be influenced by specific data and how well it describes the overall data set.

Statistical versus practical considerations

We have focused on describing potential mathematical structures and the array of statistical techniques and tools for developing empirical bioavailability models for metals. A robust use of statistical tools and methods is certainly important in model development. However, we also advocate consideration of practical issues rather than dogmatic acceptance of results from statistical tests in model development. For example, both Brix et al. (2017a) and DeForest et al. (2018) tested for effects of interaction terms (e.g., pH $\times$ DOC) in their empirical models for Cu and Al given the strong mechanistic underpinnings for these interactions and used AIC and BIC to statistically assess inclusion of interactions. For Cu, Brix et al. (2017a) observed that species-specific models with interaction terms provided modest improvement of some toxicity predictions (adjusted R^2 values increased by 0.00–0.15) and balanced these modest improvements in species-specific model performance against the advantages of a pooled model without interactions for PVAL derivation. In contrast, DeForest et al. (2018) observed large improvements in Al MLR performance when interaction terms were included in a species-specific model for *R. subcapitata* (adjusted R^2 increased by $\sim$ 0.60) but only modest improvement for *Ceriodaphnia dubia* (adjusted R^2 increased $\sim$ 0.08) and *P. promelas* (adjusted R^2 increased $\sim$ 0.03). Despite only modest improvements for *C. dubia* and *P. promelas*, DeForest et al. (2018) concluded that evidence regarding the underlying mechanisms for these interactions for Al and the strength of their significance for a species with adequate data to test for their effects justified their retention in the species-specific models.

Empirical models based on free cationic metal (hybrid models)

Although empirical models can be informed by a mechanistic understanding of bioavailability, the model structure is dictated not by this understanding but rather by statistical tests of the observed relationship of effect concentrations to the TMFs. In contrast, current mechanistic models such as the BLM structure are based on the mechanistic assumption that toxicity is elicited by the binding of the free metal cation (or possibly a hydroxyl complex of the metal) to a biological receptor (biotic ligand), with competitive binding of other nontoxic cations (e.g., Ca^{2+} , H^+) reducing the binding and consequent effects of the toxic metal (Di Toro et al. 2001). The structure of the BLM

thus must involve 1) a speciation model that determines the free toxic metal cation activity, and thereby dictates the impact of ligands (e.g., DOC) in the exposure water on bioavailability and hence toxicity, and 2) equations for the competitive interactions of various cations at the biotic ligand. Toxicity data are used to calibrate certain parameters in this model structure but do not determine the model structure.

However, in some cases, deviations from the assumptions of the BLM have been observed. Notably, Deleebeeck et al. (2008) reported that the effect of pH on chronic Ni toxicity to *Daphnia magna* did not adhere to the mechanism of competitive H^+ binding. Instead, an empirical equation was used to describe the pH effect. This resulted in a model that combined certain mechanistic aspects of the BLM with empirical equations, thereby raising the possibility of hybrid models that have a mechanistic foundation but are more explicitly informed by empirical data.

Another possibility for combining empirical and mechanistic tools is to base empirical models on the free ion activity of the metal instead of dissolved metal, as is the case with the empirical models thus far developed. Such an approach would use a speciation model to address the effect of complexation by DOC and inorganic ligands, thereby incorporating a major element of the BLM mechanistic framework but allow for the treatment of other TMFs as empirical. De Schamphelaere et al. (2002) provide an example of this approach for acute Cu toxicity to *D. magna*, and Erickson et al. (1987) used free Cu measurements to address the effect of DOC in a similar manner for *P. promelas*.

Such free metal activity-based empirical models would lose the advantage of being expressed as a simple function of measured dissolved metal, requiring instead first applying a speciation model. However, this complication might be alleviated by developing an empirical relationship between free metal activity, DOC, and pH. This would be analogous to the Bio-met and M-BAT tools (Peters et al. 2016) used to estimate the output of BLM programs as a function of a limited number of TMFs.

To further assess the utility of such free ion activity-based empirical models, we reanalyzed the data set used by Brix et al. (2017a) to develop dissolved acute Cu MLR models for 6 species (*C. dubia*, *D. magna*, *Daphnia obtusa*, *Daphnia pulex*, *Oncorhynchus mykiss*, *P. promelas*; Supplemental Data). Free ion Cu (Cu^{2+}) concentrations were estimated using the Windermere humic aqueous model VII (Tipping et al. 2011). We compared adjusted R^2 for the final stepwise linear regression models selected from 3 full models: 1) toxicity [Cu dissolved] = DOC + Ca + pH, 2) toxicity [Cu^{2+}] = Ca + pH, and 3) toxicity [Cu^{2+}] = Ca. All combinations of arithmetic and log-transformed concentrations of dependent and independent variables were attempted. In addition, the log–log versions of models 1 and 2 were rerun substituting hydrogen ion concentration (as $10^{-\text{pH}}$) for pH and again substituting hardness for Ca.

In all cases, the log–log models had the highest adjusted R^2 , and model 1 (dissolved Cu) had the highest adjusted R^2 of the 3 models (Table 3), with adjusted R^2 being 0.05–0.60 lower for the 2 models based on Cu^{2+} activity compared to the dissolved

TABLE 3: Adjusted R^2 of dissolved and free-ion activity multiple linear regression models for Cu

Model	<i>Ceriodaphnia dubia</i>	<i>Daphnia magna</i>	<i>Daphnia obtusa</i>	<i>Daphnia pulex</i>	<i>Oncorhynchus mykiss</i>	<i>Pimephales promelas</i>
1	0.635	0.814	0.816	0.789	0.625	0.749
2	0.043	0.470	0.239	0.738	0.498	0.515
3	0.012	0.468	0.239	0.738	0.301	0.494

Model 1: $\ln[\text{dissolved Cu}] = \ln(\text{DOC}) + \ln(\text{Ca}) + \text{pH}$.

Model 2: $\ln[\text{Cu}^{2+}] = \ln(\text{Ca}) + \text{pH}$.

Model 3: $\ln[\text{Cu}^{2+}] = \ln(\text{Ca})$.

Toxicity data sets from Brix et al. (2017a).

Cu models. This analysis indicates that empirical models based on dissolved metal (at least for Cu) perform substantially better than free ion–based models. This may be the result of some Cu complexes (e.g., CuOH^+) being bioavailable to varying degrees, as has been postulated in previous studies (De Schampelaere et al. 2002). Whether these findings for Cu can be generalized to other metals may warrant further investigation.

DEVELOPMENT OF POOLED SPECIES MODELS

Typically, species-specific empirical bioavailability models for a given metal will be developed for multiple species but will not be available for all of the species needing consideration in PVAL development. Rather than having to choose from among individual species-specific models to apply to species without models (Schlekat et al. 2010), it may be possible to use data/models from multiple species to develop a pooled model useful for a broader set of species. The term “pooled model” has been used to describe approaches where data from multiple species have been pooled and approaches where coefficients from species-specific models have been pooled (e.g., averaged) to develop an all-species model. These approaches are not equivalent, and we discuss them further later in this section.

A pooled model developed from a diverse representation of aquatic taxa is generally desirable for PVAL development. Being based on more data, such a model can increase confidence in applying the model to other species in the SSD, including the hypothetical “5th percentile” organism that is typically the basis for PVALs. Further, pooled models will often encompass a broader range of TMFs than individual species-specific models, thereby allowing bioavailability normalizations to be performed for a greater range of conditions.

Pooled models may be developed from a phylogenetically diverse set of species (e.g., across phyla), by categories of taxonomic groups (e.g., algae, invertebrates, and fish), or by more refined classifications of organisms (e.g., crustaceans, mollusks). How species-specific models or the data on which they are based are, or are not, pooled will depend on the data available for a metal, metal-specific characteristics, and consideration of PVAL protectiveness over a broad range of water chemistry conditions. Some important considerations are whether and how relationships between TMFs and toxicity are influenced by the magnitude of metal concentrations (i.e., sensitive vs insensitive species) and whether TMFs act similarly

across taxonomically diverse species. We provide several conceptual examples of these issues and then discuss specific methodologies for developing pooled models.

Conceptual considerations in developing pooled models

Historically, the USEPA and other regulatory bodies have used pooled hardness models in deriving hardness-based PVALs for metals (US Environmental Protection Agency 1985; Canadian Council of Resources and Environment Ministers 1987; HMSO 1989). In such models, a common (pooled) slope is assumed for all species, but each species has its own intercept; this entails including species identity as a factor in the analysis to estimate separate intercepts while imposing a single shared slope. Similar approaches have now been used to consider multiple TMFs. For example, Brix et al. (2017a) first developed acute Cu MLR models for the cladocerans *C. dubia*, *D. magna*, *D. obtusa*, and *D. pulex* and the fathead minnow (*P. promelas*). These individual species models included DOC, hardness, and pH as TMFs and performed well, with adjusted R^2 ranging from 0.63 to 0.86. Although 4 of the models were for closely related daphnid species, the influence of DOC, hardness, and pH on Cu toxicity to the fathead minnow was generally consistent with the daphnids. Thus, a pooled MLR model was developed based on data for all 5 species, and the adjusted R^2 remained relatively consistent when the pooled model was applied to each of the 5 species, with a range of 0.56 to 0.86. In this case, the authors recommended a pooled MLR for PVAL development.

For some metals, TMFs do not influence metal toxicity consistently across taxa; and in these cases, more than one model is needed. In the European Union, the decision was made to develop separate algae, invertebrate, and fish BLMs for application to SSDs (European Commission 2010). For some metals, such as Cu and Zn, this decision appears to be reasonable because algae respond to TMFs differently from fish and invertebrates (which respond similarly). A similar example is the previously described empirical bioavailability models developed for Al (DeForest et al. 2018), in which effects of DOC, hardness, and pH on the chronic toxicity of Al to an alga (*R. subcapitata*), a cladoceran (*C. dubia*), and a fish (*P. promelas*) differed substantially in terms of the parameters included in the models. As such, it was not reasonable to develop a pooled model. Furthermore, the relative sensitivities of algae, invertebrates, and fish in the chronic Al SSD varied substantially depending on water chemistry conditions. Fish,

for example, were most sensitive in conditions of low pH and hardness, but invertebrates became more sensitive with increasing pH and hardness. In this case, it was recommended that multiple MLR models be applied to species in the SSD, to account for changes in the relative sensitivity of species with changing water chemistry conditions. Similarly, bioavailability normalizations for Zn in Europe find invertebrates to be the most sensitive species under relatively acidic conditions, but plants and algae are more sensitive under neutral to alkaline pH conditions (Van Sprang et al. 2009).

We generally recommend the conceptual considerations described previously for determining whether to pool or not pool empirical models. However, other considerations and rationales have been used to address this issue. For example, the Canadian Council of Ministers of the Environment (2016) evaluated species-specific models for Zn based on precision and level of protection. Model precision was evaluated as the percentage of data for that species predicted by the model to be within a factor of 2 of observed toxicity values. Level of protection was evaluated as the percentage of all data (i.e., across species) that were protected by a given species-specific model. Hence, the Canadian Council of Ministers of the Environment did not require species-specific models to have similar slopes. Instead, emphasis was placed on selecting a species-specific model that was protective of nearly all data in the toxicity data set on the premise that differing responses between species are not important as long as the model is protective. The council found that 98% of data for all species under all conditions tested were above the value predicted by the model for *D. magna*, and this model was selected for PVAL derivation (Canadian Council of Ministers of the Environment 2016). Although this approach eliminated the need for SSD normalization (see section *SSD Normalization*), it also reduced the environmental range over which the influence of TMFs on PVALs could be predicted.

Statistical approaches to pooled model development

Different methods have been proposed and implemented for pooling empirical bioavailability models. As mentioned, the vocabulary and terminology involved with specifying species-specific and pooled models can become confusing. Linear models, specifically analysis of covariance (ANCOVA), provide a complete framework within which species-specific and pooled models can be specified using the same data set.

As a simple example, assume a toxicity data set containing 3 species and concentration data for 3 TMFs (DOC, hardness, and pH). The variable names in the data set are *sp* (species), *ltox* (ln toxicity), *ldoc* (ln DOC), *lhard* (ln hardness), *ph* (pH). A “pooled” model that does not consider any species-specific effects would be specified as follows:

$$ltox = a + (b \times ldoc) + (c \times lhard) + (d \times ph) \quad (1)$$

This model fits one intercept and one slope for each TMF across all of the species. That is, it assumes that the same

intercept and slopes apply equally to all species (see Supplemental Data for example R code).

A pooled model that normalizes to species but pools across slopes would be specified as follows:

$$ltox = (a \times sp) + (b \times ldoc) + (c \times lhard) + (d \times ph) \quad (2)$$

This model fits a separate intercept for each species to account for different species sensitivities and assumes that the same slopes apply across all species (parallel slopes). An early univariate example is that described in the USEPA guidelines (US Environmental Protection Agency 1985), in which ANCOVA is used to test for differences in the slope of the relationship between hardness and toxicity for multiple species.

For a model with species-specific intercepts, the model output will be affected by the kind of “contrasts” that are specified (see Supplemental Data for example R code). Contrasts are similar to a priori multiple comparisons; the contrast specific for the model determines what hypotheses are tested. Whereas the contrast and resulting output affect how species-specific coefficients are calculated from the output, the final coefficients are not affected by the contrast.

It is possible to specify and test a wide variety of linear contrasts, and the topic of linear contrasts is beyond the scope of the present study. Commonly used contrasts include “deviance” contrasts, which compare each group to the mean of all groups, and “treatment” contrasts, which compare each group (Institute for Digital Research and Education 2019).

For the example above, specifying deviance contrasts would test the null hypothesis that each species' intercept is equal to the mean of all the species' intercepts (this average will not always equal the intercept in the pooled model, in which no species effects were specified). The ANOVA output from the model will report whether any of these species-specific intercepts differ from the mean. The regression output (usually just the model summary) will report the mean intercept as the intercept along with a line for each species (each contrast) that includes the difference between the species' intercept and the mean intercept, as well as the standard error and *p* value of the difference. Note, the intercept for the third species must be calculated as the difference between the mean intercept and the total of all of the species-specific differences (see Table 1 in Supplemental Data 2 for example output and Table S11 in Brix et al. 2017a for examples of calculating species-specific intercepts from the regression output of pooled models with species-specific intercepts using deviance contrasts).

Specifying treatment contrasts tests the null hypothesis that each species' intercept is equal to a reference species' intercept. The regression output from the model will report the reference species' intercept as the intercept along with a line for each species that includes the difference between the species' intercept and the reference intercept. The standard error and *p* value for this difference will also be reported.

Finally, a fully “species-specific” model, with intercepts and slopes calculated separately for each species, would be

specified by adding an interaction between species and each TMF in the model, in addition to the interaction specified in the previous model with the intercept:

$$l_{tox} = (a \times sp) + (b \times doc) + (c \times lhard) + (d \times ph) \\ + (e \times ldoc \times sp) + (f \times lhard \times sp) + (g \times ph \times sp) \quad (3)$$

As for the models above with a species-specific intercept term, the output of this “species-specific model” could be used to calculate the coefficients for each species for each TMF and the significance of the difference between species’ coefficient and the coefficient for the TMF specified by the contrast (e.g., mean or reference group). Likewise, the calculation of each species’ coefficient will depend on the kind of contrast that was specified for the model. For each TMF, the output table would include a slope term for each TMF that would equal either the mean slope (using deviance contrasts) or the slope of a reference group (using treatment contrasts) along with rows for differences between species-specific slopes and the slope to which they are being compared (see Table S10 in Brix et al. 2017a for an example of calculating species-specific slopes and intercepts from a model with deviance contrasts). The coefficients calculated from the output table will be identical to the coefficients calculated by running a separate MLR for each species. Using the ANCOVA framework simply allows testing of null hypotheses of interest.

An example of using ANCOVA in a multivariate model (DOC, hardness, and pH) to test for differences among species in slopes and intercepts was recently described for Cu in Brix et al. (2017a). Deviance contrasts were used to test for differences between species slopes and intercepts and the mean slope and intercept. In the multivariate case, the effect of pooling species data with statistically different slopes on the final predicted toxicity value is not as straightforward as in a simple linear regression and will depend on the relative importance of each water quality variable on model predictions and the number of variables within a model that are different from those in the pooled model. In the Cu example, the authors decided to ignore differences in slopes for individual TMFs in several cases and pooled data from multiple species-specific models. They justified this approach by demonstrating that despite statistically significant differences in slope for single TMFs in the species-specific model, the pooled model was comparable to the species-specific model in predicting the toxicity data set for that species.

An alternative approach for developing a pooled model is to simply average TMF model coefficients (e.g., slopes) for individual species. These mean coefficients may then be used to assemble an MLR model. The acceptability of this pooled MLR model can then be evaluated relative to its performance in predicting toxicity to the individual species that were used to develop the model. The averaging of model coefficients provides an equal weighting to each individual species tested, regardless of the number of tests performed on each species, and avoids weighting the final model toward those species which have been most extensively tested. A comparison of the

TABLE 4: Comparison of dissolved organic carbon, hardness, and pH slopes based on averaging of individual species models versus pooled data model^a

Species	Slopes			n	Adjusted R ²
	ln(DOC)	ln(Hardness)	pH		
<i>Ceriodaphnia dubia</i>	0.625		0.877	87	0.63
<i>Daphnia magna</i>	0.941	0.503	1.042	307	0.86
<i>Daphnia obtusa</i>	0.843	0.233	0.670	53	0.82
<i>Daphnia pulex</i>	0.875	0.406	0.838	35	0.81
<i>Pimephales promelas</i>	0.670	1.030	0.855	206	0.76
Mean slopes	0.791	0.543	0.856	688	0.80
Pooled model slopes	0.786	0.582	0.966	688	0.81

^aIndividual species models and pooled model from Brix et al. (2017a). DOC = dissolved organic carbon.

coefficient averaging approach from individual species models to the pooled approach described in Brix et al. (2017a) demonstrates that, at least for that particular data set, the 2 approaches are comparable in terms of response to TMFs and overall performance (Table 4).

While simply averaging slopes for a TMF across species is an attractive option, care must be taken in this type of analysis. First, all of the species-specific models should have the same model structure in terms of TMFs and interaction terms considered. It is possible that a species-specific model will not have all of the TMFs included in the models for other species. If this is because data are unavailable for that species, then the model should be excluded from the averaging analysis. In contrast, if the TMF term is missing because the slope was insignificant, then it should be included in the analysis and a slope of zero should be used for averaging purposes. However, both considerations are automatically taken care of in a formal pooled analysis, which should generally be the preferred method for developing a pooled model.

SSD NORMALIZATION

The derivation of a bioavailability-normalized PVAL usually requires that the intrinsic sensitivity of each individual species included in the toxicity database can be corrected to the specific water chemistry conditions to which it is applied. When a single pooled bioavailability model is derived, SSD normalization is straightforward because the same model is applied to all species. However, when this is not the case, how bioavailability models available for selected species should be applied to other species must be addressed.

One approach has been to select the most appropriate model for normalization of species data to specific water chemistry conditions. In this approach, model application to a species is based on the species-specific model that best predicts the response to TMFs using information such as adjusted R² and other goodness-of-fit indicators previously described (e.g., QQ plots, residual plots). Schlekat et al. (2010) used this type of approach in applying Ni BLMs to species for which BLM

parameters were not available. For empirical models, this approach might be appropriate if the species data set is too small to develop a species-specific model but large enough to evaluate the performance of existing species-specific models in normalizing toxicity data for this new species. When adequate data are not available, it is not possible to assess the performance of an empirical model derived for another species, and alternative approaches may be required to pragmatically identify an appropriate model for data normalization.

When model selection cannot be definitively based on model performance, most regulatory jurisdictions have taken an approach of basing model selection on phylogenetic relatedness. This approach is sometimes used as a scientific rationale for model selection in the absence of data on which to make a more informed decision, whereas in other cases it is used somewhat more arbitrarily. For example, European bioavailability-based approaches which have been applied to date have generally required different models to normalize data for algae/plants, invertebrates, and vertebrates, regardless of any differences or similarities which may exist between these groups (European Commission 2010). As discussed, we recommend that, when scientifically supportable, application of a pooled model is preferable to this approach.

In the United States, where algae/plants have not historically been included in SSDs, it has generally been possible to use a single bioavailability model for SSD normalization, although occasional differences between animal models have been observed. In Europe, where it is required that algae/plants be included in SSDs, multiple models have been required. Relatively consistent and significant differences in the response to TMFs have been observed between algae/plants and animals, particularly with respect to the effect of pH on toxicity (De Schamphelaere et al. 2003, 2005). As a result, it is generally not appropriate to apply the same normalization model to plants as would be used for animals.

Finally, not all regulatory approaches for deriving PVALs require that the entire SSD be normalized for each water chemistry condition of interest, and this may also influence decisions regarding the use of species-specific models or strategies for pooling models. For example, in deriving its hardness-based metal criteria, the USEPA first normalizes the SSD to a common hardness condition and then defines the sensitivity of the 5th percentile organism as the intercept. The hardness slope and that intercept can then be used to calculate the criterion for any hardness without needing to normalize the SSD for every hardness condition. This same type of approach was used to derive a MLR-based PVAL for Cu following USEPA guidelines (Brix et al. 2017a), which results in a considerably simpler calculation of the PVAL than approaches which require that all of the data in the SSD be normalized to calculate the PVAL. In this case, a simplified approach was possible because a single pooled bioavailability model was applied to all species in the SSD, so the relative sensitivity of each species is always the same regardless of water chemistry conditions. This allowed the sensitivity parameter in the bioavailability model to be set based on the 5% hazard concentration derived from the SSD rather than on the sensitivity

of any individual species. This simplified approach would not work if more than one bioavailability model is applied to the SSD; in those cases, the SSD would need to be normalized for each water chemistry condition of interest, as was the case with Al (DeForest et al. 2018; US Environmental Protection Agency 2018).

CONCLUSION AND PERSPECTIVE

Empirical bioavailability models are a useful tool in risk assessment and for deriving PVALs for metals (and potentially other classes of toxicants). More mechanistic models such as the BLM may provide a more scientifically rigorous approach to accounting for the effects of TMFs on metal bioavailability and may require less initial testing to formulate a new model but may not be more accurate in their predictions. Moreover, these models may be more complex than is needed to achieve the desired environmental management goals in some regulatory settings. In these cases, empirical bioavailability models can provide a simpler, more transparent tool for predicting the effects of TMFs on metal bioavailability and often have similar predictive capabilities.

As with all models, empirical bioavailability models must be developed by rigorous assessment of model performance using both visual plotting and statistical assessment tools to check for model bias and accuracy. We recommend taking advantage of available toxicity data and the same theoretical understanding used to develop mechanistic models such as the BLM to make informed decisions about which TMFs to include in empirical bioavailability models. Model validation using independent data sets and/or comparison to other models such as the BLM is also critical.

Although we have found that empirical bioavailability model performance is generally comparable to more mechanistic models, the temptation often arises to include more complexity in the model to improve performance. For example, in the present study we investigated the potential of using free ion rather than dissolved metal concentrations to develop a hybrid model for predicting metal bioavailability. Although this example did not improve model performance, inclusion of other variables may. In such cases, comparison of final models using different fit statistics, such as R^2 , AIC, and BIC, can be helpful for assessing whether improved model performance justifies increased model complexity. On the other hand, as we discussed for temperature, there may be TMFs that are important in applying models to field conditions but are not typically included in models because they are not addressed in typical laboratory testing.

Supplemental Data—The Supplemental Data are available on the Wiley Online Library at DOI: 10.1002/etc.4570.

Acknowledgment—We acknowledge the financial support of the SETAC technical workshop Bioavailability-Based Aquatic Toxicity Models for Metals, including the US Environmental Protection Agency, the Metals Environmental Research Associations, Dow Chemical Company, Newmont Mining Company,

Rio Tinto, Umicore, and Windward Environmental. We also thank the SETAC staff for their support in organizing the workshop.

Data Availability Statement—Data, associated metadata, and calculation tools are available from the corresponding author (kevinbrix@icloud.com).

REFERENCES

- Adams W, Blust R, Dwyer R, Mount D, Nordheim E, Rodriguez PH, Spry D. 2020. Bioavailability assessment of metals in freshwater environments: A historical review. *Environ Toxicol Chem* 39:48–59.
- Al-Reasi HA, Wood CM, Smith DS. 2011. Physiochemical and spectroscopic properties of natural organic matter (NOM) from various sources and implications for ameliorative effects on metal toxicity to aquatic biota. *Aquat Toxicol* 103:179–190.
- Brix KV, DeForest DK, Tear L, Grosell M, Adams WJ. 2017a. Use of multiple linear regression models for setting water quality criteria for copper: A complementary approach to the biotic ligand model. *Environ Sci Technol* 51:5182–5192.
- Brix KV, Schlekert CE, Garman ER. 2017b. The mechanisms of nickel toxicity in aquatic environments: An adverse outcome pathway analysis. *Environ Toxicol Chem* 36:1128–1137.
- Brix KV, Tellis MS, Cremazy A, Wood CM. 2016. Characterization of the effects of binary metal mixtures on short-term uptake of Ag, Cu, and Ni by rainbow trout (*Oncorhynchus mykiss*). *Aquat Toxicol* 180:236–246.
- Brix KV, Tellis MS, Cremazy A, Wood CM. 2017c. Characterization of the effects of binary metal mixtures on short-term uptake of Cd, Pb, and Zn by rainbow trout (*Oncorhynchus mykiss*). *Aquat Toxicol* 193:217–227.
- Brix KV, Volosin JS, Adams WJ, Reash RJ, Carlton RG, McIntyre DO. 2001. Effects of sulfate on the acute toxicity of selenate to freshwater organisms. *Environ Toxicol Chem* 20:1037–1045.
- Burnham KP, Anderson DR. 2004. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol Methods Res* 33:261–303.
- Bury NR, Galvez F, Wood CM. 1999. Effects of chloride, calcium, and dissolved organic carbon on silver toxicity: Comparison between rainbow trout and fathead minnows. *Environ Toxicol Chem* 18:56–62.
- Canadian Council of Ministers of the Environment. 2016. Scientific criteria document for the development of the Canadian water quality guidelines for the protection of aquatic life: Zinc. Draft. Winnipeg, MB, Canada.
- Canadian Council of Resources and Environment Ministers. 1987. Canadian water quality guidelines. Winnipeg, MB, Canada.
- DeForest DK, Brix KV, Elphick JR, Rickwood CJ, deBruyn AMH, Tear LM, Giron G, Hughes SA, Adams WJ. 2017. Lentic, lotic, and sulfate-dependent waterborne selenium screening guidelines for freshwater systems. *Environ Toxicol Chem* 36:2503–2513.
- DeForest DK, Brix KV, Tear LM, Adams WJ. 2018. Multiple linear regression (MLR) models for predicting chronic aluminum toxicity to freshwater aquatic organisms and developing water quality guidelines. *Environ Toxicol Chem* 37:80–90.
- Deleebeeck NME, De Schampelaere KAC, Janssen CR. 2008. A novel method for predicting chronic nickel bioavailability and toxicity to *Daphnia magna* in artificial and natural waters. *Environ Toxicol Chem* 27:2097–2107.
- De Schampelaere KAC, Heijerick DG, Janssen CR. 2002. Refinement and field validation of a biotic ligand model predicting acute copper toxicity to *Daphnia magna*. *Comp Biochem Physiol C Toxicol Pharmacol* 133:243–258.
- De Schampelaere KAC, Janssen CR. 2010. Cross-phylum extrapolation of the *Daphnia magna* chronic biotic ligand model for zinc to the snail *Lymnaea stagnalis* and the rotifer *Brachionus calyciflorus*. *Sci Total Environ* 408:5414–5422.
- De Schampelaere KAC, Lofts S, Janssen CR. 2005. Bioavailability models for predicting acute and chronic toxicity of zinc to algae, daphnids, and fish in natural surface waters. *Environ Toxicol Chem* 24:1190–1197.
- De Schampelaere K, Vasconcelos FM, Heijerick D, Tack FMG, Delbeke K, Allen HE, Janssen CR. 2003. Development and field validation of a predictive copper toxicity model for the green alga *Pseudokirchneriella subcapitata*. *Environ Toxicol Chem* 22:2454–2465.
- Di Toro DM, Allen HE, Bergman HL, Meyer JS, Paquin PR, Santore RC. 2001. A biotic ligand model of the acute toxicity of metals. I. Technical basis. *Environ Toxicol Chem* 20:2383–2396.
- Erickson RJ, Benoit DA, Mattson VR. 1987. A prototype toxicity factors model for site-specific water quality criteria. US Environmental Protection Agency, Duluth, MN.
- Esbaugh AJ, Brix KV, Mager EM, De Schampelaere KAC, Grosell M. 2012. Multi-linear regression analysis, preliminary biotic ligand modeling, and cross species comparison of the effects of water chemistry on chronic lead toxicity in invertebrates. *Comp Biochem Physiol C Toxicol Pharmacol* 155:423–431.
- Esbaugh AJ, Brix KV, Mager EM, Grosell M. 2011. Multi-linear regression models predict the effects of water chemistry on acute lead toxicity to *Ceriodaphnia dubia* and *Pimephales promelas*. *Comp Biochem Physiol C Toxicol Pharmacol* 154:137–145.
- European Commission. 2010. Nickel and its compounds. Environmental quality standards sheet. Danish Environmental Protection Agency, Copenhagen, Denmark.
- European Commission. 2011. Common implementation strategy for the water framework directive (2000/60/EC). Guidance document no. 27. Technical guidance for deriving environmental quality standards, European communities, Brussels, Belgium.
- Farley KJ, Meyer JS, Balistreri LS, De Schampelaere KAC, Iwasaki Y, Janssen CR, Kamo M, Lofts S, Mebane CA, Naito W, Ryan AC, Santore RC, Tipping E. 2015. Metal mixture modeling evaluation project: 2. Comparison of four modeling approaches. *Environ Toxicol Chem* 34:741–753.
- Fulton BA, Meyer JS. 2014. Development of a regression model to predict copper toxicity to *Daphnia magna* and site-specific copper criteria across multiple surface-water drainages in an arid landscape. *Environ Toxicol Chem* 33:1865–1873.
- Garman ER, Meyer JS, Bergeron CM, Blewett TA, Clements W, Elias M, Farley KJ, Gissi F, Ryan A. 2020. Validation of bioavailability-based toxicity models for metals. *Environ Toxicol Chem* 39:101–117.
- Geneser RW, Gondek JC, Rodriguez PH, Arbildua JJ, Stubblefield WA, Cardwell AS, Santore RC, Ryan AC, Adams WJ, Nordheim E. 2018. Evaluating the effects of pH, hardness, and dissolved organic carbon on the toxicity of aluminum to freshwater aquatic organisms under circumneutral conditions. *Environ Toxicol Chem* 37:49–60.
- Harrell F. 2015. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer, New York, NY, USA.
- Heijerick DG, De Schampelaere KAC, Janssen CR. 2002. Predicting acute zinc toxicity for *Daphnia magna* as a function of key water chemistry characteristics: Development and validation of a biotic ligand model. *Environ Toxicol Chem* 21:1309–1315.
- HMSO. 1989. Statutory instruments. 1989 No. 2286. Water, England and Wales. The surface water (Dangerous Substances) (Classification) regulations 1989. Her Majesty's Stationery Office, London, UK.
- Institute for Digital Research and Education. 2019. University of California, Los Angeles, Los Angeles, CA, USA. [cited 2019 July 30]. Available from: <https://stats.idre.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/>
- Jolliffe IT. 2002. *Principal Component Analysis*. Springer-Verlag, New York, NY, USA.
- Lee WY, Wang WX. 2001. Metal accumulation in the green macroalga *Ulva fasciata*: Effects of nitrate, ammonium and phosphate. *Sci Total Environ* 278:11–22.
- Mebane CA, Chowdhury MJ, De Schampelaere KAC, Lofts S, Paquin PR, Santore RC, Wood CM. 2020. Metal bioavailability models: Current status, lessons learned, considerations for regulatory use, and the path forward. *Environ Toxicol Chem* 39:60–84.
- Niyogi S, Wood CM. 2004. Biotic ligand model, a flexible tool for developing site-specific water quality guidelines for metals. *Environ Sci Technol* 38:6177–6192.
- Organisation for Economic Co-operation and Development. 2007. Guidance document on the validation of (quantitative) structure–activity relationship [(Q)SAR] models. Series on Testing and Assessment, No. 69. ENV/JM/MONO(2007)2. Paris, France.
- Paquin PR, Gorsuch JW, Apte SC, Batley GE, Bowles KC, Campbell PGC, Delos CG, Di Toro DM, Dwyer RL, Galvez F, Geneser RW, Goss GG, Hogstrand C, Janssen CR, McGeer JC, Naddy RB, Playle RC, Santore RC, Schneider U, Stubblefield WA, Wood CM, Wu KB. 2002. The biotic

- ligand model: A historical overview. *Comp Biochem Physiol C Toxicol Pharmacol* 133:3–36.
- Peters A, Lofts S, Merrington G, Brown B, Stubblefield WA, Harlow K. 2011. Development of biotic ligand models for chronic manganese toxicity to fish, invertebrates, and algae. *Environ Toxicol Chem* 30:2407–2415.
- Peters A, Schlekot CE, Merrington G. 2016. Does the scientific underpinning of regulatory tools to estimate bioavailability of nickel in freshwaters matter? The European-wide environmental quality standard for nickel. *Environ Toxicol Chem* 35:2397–2404.
- Rogevich EC, Hoang TC, Rand GM. 2008. The effects of water quality and age on the acute toxicity of copper to the Florida apple snail, *Pomacea paludosa*. *Arch Environ Contam Toxicol* 54:690–696.
- Santore RC, Matthew R, Paquin PR, Di Toro DM. 2002. Application of the biotic ligand model to predicting zinc toxicity to rainbow trout, fathead minnow, and *Daphnia magna*. *Comp Biochem Physiol C Toxicol Pharmacol* 133:271–287.
- Santore RC, Ryan AC, Kroglund F, Rodriguez PH, Stubblefield WA, Cardwell AS, Adams WJ, Nordheim E. 2018. Development and application of a biotic ligand model for predicting the chronic toxicity of dissolved and precipitated aluminum to aquatic organisms. *Environ Toxicol Chem* 37:70–79.
- Schlekot CE, Van Genderen EJ, De Schamphelaere KAC, Antunes PMC, Rogevich EC, Stubblefield WA. 2010. Cross-species extrapolation of chronic nickel biotic ligand models. *Sci Total Environ* 408:6148–6157.
- Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, New York, NY, USA.
- Sokolova IM, Lannig G. 2008. Interactive effects of metal pollution and temperature on metabolism in aquatic ectotherms: Implications of global climate change. *Climate Res* 37:181–201.
- Tipping E, Lofts S, Sonke JE. 2011. Humic ion-binding model VII: A revised parameterisation of cation-binding by humic substances. *Environ Chem* 8:225–235.
- US Environmental Protection Agency. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. Duluth, MN.
- US Environmental Protection Agency. 2001. Streamlined water-effect ratio procedure for discharges of copper. Washington, DC.
- US Environmental Protection Agency. 2016. Aquatic life ambient water quality criteria for cadmium—2016. Washington, DC.
- US Environmental Protection Agency. 2018. Final aquatic life ambient water quality criteria for aluminum 2018. Washington, DC.
- Van Genderen EJ, Stauber JL, Delos C, Eignor D, Gensemer RW, McGeer J, Merrington G, Whitehouse P. 2020. Best practices for derivation and application of thresholds for metals using bioavailability-based approaches. *Environ Toxicol Chem* 39:118–130.
- Van Regenmortel T, Janssen CR, De Schamphelaere KAC. 2015. Comparison of the capacity of two biotic ligand models to predict chronic copper toxicity to two *Daphnia magna* clones and formulation of a generalized bioavailability model. *Environ Toxicol Chem* 34:1597–1608.
- Van Sprang PA, Verdonck FAM, Van Assche F, Regoli L, De Schamphelaere KAC. 2009. Environmental risk assessment of zinc in European freshwaters: A critical appraisal. *Sci Total Environ* 407:5373–5391.
- Veltman K, Hendriks AJ, Huijbregts MAJ, Wannaz C, Jolliet O. 2014. Toxicokinetic toxicodynamic (TKTD) modeling of Ag toxicity in freshwater organisms: Whole-body sodium loss predicts mortality across aquatic species. *Environ Sci Technol* 48:14481–14489.
- Welsh PG, Lipton J, Chapman GA. 2000. Evaluation of water-effect ratio methodology for establishing site-specific water quality criteria. *Environ Toxicol Chem* 19:1616–1623.
- Welsh PG, Parrott JL, Dixon DG, Hodson PV, Spry DJ, Mierle G. 1996. Estimating acute copper toxicity to larval fathead minnow (*Pimephales promelas*) in soft water from measurements of dissolved organic carbon, calcium, and pH. *Can J Fish Aquat Sci* 53:1263–1271.
- Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 1:3–14.