Environmental Toxicology and Chemistry—Volume 39, Number 9—pp. 1724–1736, 2020
Received: 22 January 2020 | Revised: 24 February 2020 | Accepted: 4 June 2020

Environmental Toxicology

Updated Multiple Linear Regression Models for Predicting Chronic Aluminum Toxicity to Freshwater Aquatic Organisms and Developing Water Quality Guidelines

David K. DeForest,[a,*] Kevin V. Brix,[b] Lucinda M. Tear,[a] Allison S. Cardwell,[c] William A. Stubblefield,[c] Eirik Nordheim,[d] and William J. Adams[e]

[a]Windward Environmental, Seattle, Washington, USA
[b]EcoTox, Miami, Florida, USA
[c]Oregon State University, Corvallis, Oregon, USA
[d]European Aluminium, Brussels, Belgium
[e]Red Cap Consulting, Lake Point, Utah, USA

Abstract: Multiple linear regression (MLR) models for predicting chronic aluminum toxicity to a cladoceran (*Ceriodaphnia dubia*) and a fish (*Pimephales promelas*) as a function of 3 toxicity-modifying factors (TMFs)—dissolved organic carbon (DOC), pH, and hardness—have been published previously. However, the range over which data for these TMFs were available was somewhat limited. To address this limitation, additional chronic toxicity tests with these species were subsequently conducted to expand the DOC range up to 12 mg/L, the pH range up to 8.7, and the hardness range up to 428 mg/L. The additional toxicity data were used to update the chronic MLR models. The adjusted R^2 for the *C. dubia* 20% effect concentration (EC20) model increased from 0.71 to 0.92 with the additional toxicity data, and the predicted R^2 increased from 0.57 to 0.89. For *P. promelas*, the adjusted R^2 increased from 0.87 to 0.92 and the predicted R^2 increased from 0.72 to 0.87. The high predicted R^2 relative to the adjusted R^2 indicates that the models for both species are not overly parameterized. When data for *C. dubia* and *P. promelas* were pooled, the adjusted R^2 values were comparable to the species-specific models (0.90 and 0.88 for *C. dubia* and *P. promelas*, respectively). This indicates that chronic aluminum EC20s for *C. dubia* and *P. promelas* respond similarly to variation in DOC, pH, and hardness. Overall, the pooled model predicted EC20s that were within a factor of 2 of observed in 100% of the *C. dubia* tests and 94% of the *P. promelas* tests. *Environ Toxicol Chem* 2020;39:1724–1736. © 2020 SETAC

Keywords: Aluminum; Bioavailability; Multiple linear regression models; Water quality guidelines

INTRODUCTION

The bioavailability of aluminum (Al) to aquatic organisms is strongly influenced by dissolved organic carbon (DOC), pH, and hardness (Gensemer and Playle 1999; Cardwell et al. 2018; Gensemer et al. 2018; Santore et al. 2018). Multiple linear regression (MLR) models were previously developed that predicted the toxicity of total Al to an alga, *Raphidocelis subcapitata* (formerly *Pseudokirchneriella subcapitata*), a cladoceran (*Ceriodaphnia dubia*), and the fathead minnow (*Pimephales promelas*) as a function of DOC, pH, and hardness (DeForest et al. 2018). The US Environmental Protection Agency (USEPA) used the MLR models for *C. dubia* and *P. promelas* to develop updated draft ambient water quality criteria for Al (US Environmental Protection Agency 2017).

Technical comments on the USEPA's draft Al criteria highlighted the need to expand toxicity testing over a broader range of DOC, pH, and hardness conditions, to increase the applicability of the models and the criteria (US Environmental Protection Agency 2020). Consequently, additional toxicity tests conducted in the present study have expanded the DOC range up to 12.3 mg/L, the pH range up to 8.7, and the hardness range up to 428 mg/L. These toxicity tests also provided additional information on how water chemistry parameters may interact to influence Al toxicity.

The new Al toxicity data for *C. dubia* and *P. promelas* and the updated MLR models were previously provided to the

This article includes online-only Supplemental Data.
* Address correspondence to DavidD@windwardenv.com
Published online 5 June 2020 Library in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/etc.4796

USEPA and used to develop its final ambient water quality criteria for Al (US Environmental Protection Agency 2018). The present study provides the new chronic Al toxicity data for *C. dubia* and *P. promelas* and updated MLR models for these species. In addition, updated Al water guidelines following USEPA and European Union approaches are provided. The MLR model recommended for criteria development in the present study (a pooled species model) differs from the MLR models selected by the USEPA in developing its final ambient water quality criteria for Al (individual species models). A comparison of how model selection influences ambient water quality criteria calculations is provided.

MATERIALS AND METHODS

Al toxicity tests

Chronic Al toxicity tests with *C. dubia* (7-d reproduction) and *P. promelas* (7-d survival and growth) were conducted at the Oregon State University Aquatic Toxicology Laboratory (Corvallis, Oregon, USA) following standard test methods (US Environmental Protection Agency 2002). Nine new toxicity tests were conducted for each species. Toxicity tests were conducted in reconstituted laboratory water using reagent-grade salts ($CaSO_4 \cdot 2H_2O$, $MgSO_4$, KCl, $NaHCO_3$) that were added to deionized water to achieve the desired hardness concentrations and ion compositions.

The DOC was added as Suwannee River Natural Organic Matter (NOM; catalogue #2R101N; obtained from the International Humic Substances Society) to achieve nominal DOC concentrations in each test water based on a composition of 48% carbon in the NOM.

The target pH was achieved by adding dilute NaOH or HCl to the test solution, along with a buffer to control the pH. The buffers (5 mM concentration for each) used were 2-(n-morpholino) ethanesulfonic acid, monohydrate (MES) for the pH 6.3 tests, 3-(4-morpholino) propane sulfonic acid (MOPS) for the pH 7 series of tests, and N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES; OmniPur) for the pH 8 series of tests. Although buffers were the primary method of pH control, an additional step of slightly manipulating the CO_2/air atmospheres in an "air-tight" enclosure was employed in certain tests to achieve pH control (Mount and Mount 1992). In 3 of the pH 6.3 tests, 0.4% CO_2 was added to the atmosphere. In the pH 7 tests, no manipulations to the headspace were employed, and in the pH 8.0, 8.2, and 8.8 tests, ultrapure air was added to the headspace to reduce CO_2 concentration and control pH.

The source of Al was reagent-grade aluminum nitrate nonahydrate, $Al(NO_3)_3 \cdot 9H_2O$ (CAS No. 7784-27-2; J.T. Baker). All tests were initiated following a 3-h aging period after mixing $Al(NO_3)_3 \cdot 9H_2O$ stock solutions with dilution water. The purpose of the 3-h aging period was to reduce the impact of transient, short-lived toxic effects caused by rapid changes in Al speciation, changes that are considered unrepresentative of Al exposures in most natural waters (Cardwell et al. 2018). This aging period was selected based on preliminary analytical measurements and toxicity tests that showed toxicity associated with speciation changes generally had stabilized by 3 h (Gensemer et al. 2018). Samples were collected for Al analysis from newly prepared waters (after the 3-h aging) at test initiation, during the tests, and from a composite of replicates at test termination. Samples were analyzed for total and dissolved Al using a Spectro Arcos inductively coupled plasma–optical emission spectrometer. Samples were analyzed according to USEPA method 200.7 (US Environmental Protection Agency 1994).

Concentration–response modeling was conducted using the USEPA's Toxicity Relationship Analysis Program (TRAP, Ver 1.30a). Total Al concentrations from "new" renewal waters were used in the modeling due to difficulty in consistently collecting a homogenized sample from the test chamber at the end of each 24-h exposure period. Completely homogenizing the sample within the test chamber (which contained both dissolved and precipitated Al) to collect an "old" sample would have negatively affected the organisms. Because the measurement concentrations were for total Al, there was no indication that samples from within the test chamber would have been significantly different from the renewal waters. Both 10 and 20% effect concentrations (EC10s and EC20s) were obtained from TRAP.

Additional details of the laboratory toxicity testing are provided in Supplemental Data 1 for *C. dubia* and Supplemental Data 2 and 3 for *P. promelas*.

Individual species models

The chronic Al toxicity dataset used to develop the *C. dubia* MLR models described in DeForest et al. (2018) comprised 23 tests with DOC concentrations ranging from 0.1 to 4 mg/L, pH ranging from 6.3 to 8.1, and hardness concentrations ranging from 9.8 to 123 mg/L as $CaCO_3$ (Supplemental Data 4, Table S1). The 9 new chronic toxicity tests expanded the DOC range to 12.3 mg/L, the pH range to 8.7, and the hardness range to 428 mg/L (Supplemental Data 4, Table S1). Two of the additional tests were also conducted at a circumneutral pH of 7.2, because most of the previous testing had been conducted at either pH approximately 6.3 or approximately 8 (Supplemental Data 4, Table S1). Two chronic *C. dubia* tests from McCauley et al. (1986) were excluded from the current MLR models due to unacceptable control performance. The final sample size for the *C. dubia* model was therefore 30.

The chronic Al toxicity dataset used to develop the *P. promelas* MLR models described in DeForest et al. (2018) was comprised of 22 tests with DOC concentrations ranging from 0.08 to 5 mg/L, pH ranging from 6.0 to 8.0, and hardness concentrations ranging from 10 to 127 mg/L (Supplemental Data 4, Table S2). The 9 additional chronic Al toxicity tests expanded the DOC range up to 11.6 mg/L, the pH range up to 8.1, and the hardness range up to 422 mg/L, and 2 additional tests were conducted at a circumneutral pH of 7.0 (Supplemental Data 4, Table S2).

This updated dataset was used to develop new MLR models following the methods described in DeForest et al. (2018) and Brix et al. (2020), which employed the Akaike Information

Criterion (AIC) and the Bayesian Information Criterion (BIC) to determine which combination of terms resulted in the most parsimonious models for predicting Al EC10s or EC20s (Burnham and Anderson 2004).

In DeForest et al. (2018), Al MLR models were developed following 2 approaches. In the first approach, ln(DOC), pH, and ln(Hard) were added to the model as independent variables. In the second approach, these same 3 parameters were added, as well as the following 4 terms: pH^2; ln(DOC) × pH; ln (DOC) × ln(Hard); and ln(Hard) × pH. Ultimately, MLR models that considered the pH^2 term and the 3 interaction terms were recommended because those terms are consistent with our current understanding of Al speciation and bioavailability. A negative pH^2 term would help account for decreasing Al bioavailability as pH increases from 6 to 7 and then increasing Al bioavailability as pH increases from pH 7 to pH 8. A negative ln (DOC) × pH term would characterize the mitigating effect of DOC on Al bioavailability, which tends to decrease as pH increases; a negative ln(DOC) × ln(Hard) term would reflect that the mitigating effect of DOC on Al bioavailability tends to decrease as hardness increases; and a negative ln(Hard) × pH term would reflect that the mitigating effect of hardness on Al bioavailability tends to decrease as pH increases.

The resulting models in DeForest et al. (2018) that considered these additional terms provided improved Al toxicity predictions. (Note that the terms were considered in the initial models, but the final models did not include all terms.) The updated MLR models we evaluated likewise considered the pH^2 term and the 3 interaction terms.

Pooled models

Pooled MLR models based on the combined *C. dubia* and *P. promelas* datasets were developed using the approach described for copper in Brix et al. (2017, 2020). In addition, the possibility of developing a pooled model using Al toxicity data for *C. dubia*, *P. promelas*, and the alga *R. subcapitata* was evaluated. The *R. subcapitata* dataset was previously described in Gensemer et al. (2018). In the pooled models, the combined data for 2 or more species were used to develop common slopes for parameters found to have an important influence on Al bioavailability and toxicity, with species-specific intercepts derived to account for differences in species' sensitivities. Pooled models were developed by specifying species-specific intercepts (a "species" term), plus a term for each of the independent variables and their interactions (ln [DOC], ln[Hard], pH, pH^2, ln[DOC] × pH, ln[DOC] × ln(Hard), and ln(Hard) × pH). The same procedures used to develop the individual species models, as described in DeForest et al. (2018), were then applied to the full model to select which intercepts and slopes to retain using AIC and BIC.

Predicted R^2

In addition to the model evaluation methods applied in DeForest et al. (2018), predicted R^2 values were calculated for both the previous MLR models and the updated MLR models described in the present study. The predicted R^2 is a form of leave-one-out cross-validation that summarizes a model's ability to predict new data (Allen 1971; Neter et al. 1990). Predicted R^2 is calculated by substituting the prediction sums of squares (PRESS) for the regression sum of squares (SS) in the formula for R^2 (SS regression/SS total). The PRESS is calculated as the sum of all the "deleted residuals" (d_i) rather than the "standard" residuals (e_i), where each d_i is calculated as the difference between the observed y_i and the y_i predicted for that observation based on a model including all other data except that observation. In most statistical packages, each d_i is calculated by weighting each e_i around the original regression by the record's leverage (h_{ii} from the "hat" matrix of the original model; Neter et al. 1990). The predicted R^2 will always be at least slightly lower than the adjusted R^2. A substantially lower predicted R^2 is an indication that the model may be overfit and/or is excessively reliant on individual data points.

Predicted and observed EC10/20 plots

It has been standard practice to compare predicted toxicity values from MLR models and biotic ligand models (BLMs) using 1:1 plots as predicted versus observed toxicity (predicted toxicity values on the *y*-axis and observed toxicity values on the *x*-axis). The predicted versus observed EC10/20 plots for Al in DeForest et al. (2018) were likewise plotted that way. In the present evaluation, we instead plot observed EC10/20s versus predicted EC10/20s because, as described in Piñeiro et al. (2008), the expected value of the slope of the observed versus predicted values regression = 1, whereas the slope of the predicted versus observed values regression = $R^2_{\text{predicted vs. observed}}$ (generally < 1). Although either form of the plots (predicted vs observed or observed vs predicted) provides a visual means for understanding how closely the predicted values match the observed values, moving the predicted values to the *x*-axis allows the expectation of the slope to be 1 and also provides the benefit of allowing the plot to be interpreted as showing how well predicted values explain the variance in the observed values. The form of the plot does not influence the adjusted and predicted R^2 values of the model from which the predicted values are derived.

Species and genus sensitivity distributions and 5% hazardous concentrations

Species and genus mean sensitivity distributions (SSDs and GSDs, respectively) for chronic Al toxicity were provided in DeForest et al. (2018), and were used to develop MLR-adjusted hazardous concentrations for 5% of the species (HC5s) that varied as a function of DOC, pH, and hardness. Chronic Al HC5s were developed following European Union and USEPA approaches (US Environmental Protection Agency 1985; European Chemicals Agency 2008). Key differences between these 2 approaches include: 1) use of EC10s in the European Union versus EC20s in the United States; 2) use of species-mean EC10s in the European Union versus genus-mean EC20s in the United States; 3) inclusion of algae and

plant data in the European Union SSD versus exclusion of algae and plant data in the US GSD; and 4) use of different statistical models for calculating the HC5. Although the United States does not include algae and plant data in determining the HC5, the sensitivity of algae and plant species relative to the HC5 is evaluated to ensure it is protective. The chronic Al SSD and GSD values were updated to include the new toxicity data for *C. dubia* and *P. promelas,* and calculations of HC5s were updated using the new MLR models.

RESULTS AND DISCUSSION

Ceriodaphnia dubia

The updated chronic Al MLR models for *C. dubia* had substantially higher adjusted R^2 values than the previous models. Based on the BIC-selected models, the adjusted R^2 increased from 0.685 to 0.925 for the EC10 model and from 0.710 to 0.925 for the EC20 model (Table 1). The adjusted R^2 for the AIC-selected models similarly increased from 0.719 to 0.929 for the EC10 model and from 0.726 to 0.928 for the EC20 model (Table 1). The predicted R^2 for the BIC-selected models decreased only slightly relative to the adjusted R^2—from 0.925 to 0.902 for the EC10 model and from 0.925 to 0.903 for the EC20 model, with similarly slight reductions observed for the AIC-selected models (Table 1). In the previous models, the predicted R^2 was 0.628 and 0.607 for the BIC-selected EC10 and EC20 models, respectively, and 0.572 and 0.607 for the AIC-selected EC10 and EC20 models, respectively (Table 1).

Overall, the AIC- and BIC-selected models performed similarly based on adjusted and predicted R^2. For the sample size of the *C. dubia* dataset ($n = 30$) in the present evaluation, BIC had a larger sample size penalty than AIC, which means the BIC-selected model will likely retain fewer parameters than the AIC-selected model. For the *C. dubia* EC10 and EC20 models, AIC and BIC both retained the following 4 terms: ln(DOC), pH, ln(Hard), and ln(Hard) × pH. The AIC-selected EC10 model also retained the pH^2 term, and the AIC-selected EC20 model also retained both the pH^2 and ln(DOC) × ln(Hard) terms (Table 1). Inclusion of the ln(DOC) × ln(Hard) term in the AIC-selected EC20 model but not the EC10 model raises a question about the importance of this term, because we would expect consistency between the EC10 and EC20 models—this is an indication that the AIC-selected EC20 model may be overparameterized.

In addition to the ln(DOC) × ln(Hard) term, another inconsistency in terms in the AIC- and BIC-selected models was the pH^2 term. This term was retained in the AIC-selected EC10 model but not the BIC-selected EC10 model, whereas it was retained in both the AIC- and BIC-selected EC20 models (Table 1). This inconsistency suggested that the pH^2 term may be an unnecessary parameter in the models where it was retained. To evaluate the pH^2 term further, we calculated variance inflation factors (VIFs) to measure the degree to which the variance of a covariate in a multivariate linear model is inflated by its correlation with other variables in the model. When both pH and pH^2 were included in the model, the VIFs for ln(DOC), ln(Hard), pH, and pH^2 were 1.27, 1.32, 813, and 816, respectively. A VIF of 10 indicates severe multicollinearity and almost complete correlation between a variable and a linear combination of other independent variables in the model (Zuur et al. 2010), so the VIFs >800 for the pH and pH^2 terms are, not surprisingly, strongly indicative of a complete correlation of these parameters. When the pH^2 term was excluded from the model, the pH VIF decreased to 1.13.

When the pH^2 term was excluded from the *C. dubia* models, the adjusted and predicted R^2 only declined negligibly (Table 1). However, with exclusion of the pH^2 term, the ln(DOC) × ln(Hard) term was retained in both the AIC- and BIC-selected EC20 models but not the EC10 models. Due to this inconsistency, we evaluated one additional model iteration by excluding both the ln(DOC) × ln(Hard) term and pH^2 term. With exclusion of these 2 terms, the following 4 terms were retained in the AIC- and BIC-selected EC10 and EC20 models: ln(DOC), pH, ln(Hard), and ln(Hard) × pH (Table 1). The adjusted and predicted R^2 values for the AIC- and BIC-selected EC10 models were 0.925 and 0.902, respectively, and 0.913 and 0.886 for the AIC- and BIC-selected EC20 models (Table 1). Given the high R^2 and to avoid overparameterization of the model, we recommend *C. dubia* EC10 and EC20 models based on ln(DOC), pH, ln(Hard), and ln(Hard) × pH. In addition to the high R^2, 97% of predicted EC10s and EC20s (29 of 30) were within a factor of 2 of observed EC10/20s (Supplemental Data 4, Figure S1A and Figure 1A, respectively). Residuals were not homogeneously distributed over the full range of predicted Al bioavailability conditions; raw residuals tended to be positive at lower EC10/20s, negative at intermediate EC10/20s, and then positive at higher EC10/20s (Supplemental Data 4, Figure S2A and B). This pattern was not apparent when residuals were plotted against single TMFs (Supplemental Data 4, Figures S3A and B, S4A and B, and S5A and B).

When modeled EC20 predictions were plotted against empirical EC20s from a series of tests in which DOC, pH, or hardness were individually varied, EC20 predictions from the updated MLR model showed similar patterns compared with the previous MLR model from DeForest et al. (2018), but with a slightly stronger influence of DOC and hardness (i.e., steeper slopes; Figure 2). In addition, with removal of the pH^2 term in the present model, the relationship between pH and EC20s was now linear and appeared to adequately predict the data (Figure 2B).

The updated Al MLR EC10 and EC20 models for *C. dubia* are provided in Equations 1 and 2, respectively.

$$\ln(EC10) = -11.824 + 0.700 \times \ln(DOC) + 2.368 \times pH + 3.030 \times \ln(Hard) - 0.375 \times \ln(Hard) \times pH \quad (1)$$

$$\ln(EC20) = -9.272 + 0.619 \times \ln(DOC) + 2.021 \times pH + 2.581 \times \ln(Hard) - 0.307 \times \ln(Hard) \times pH \quad (2)$$

where ln(EC10) and ln(EC20) are total Al concentrations in units of μg/L, DOC is in mg/L, and Hard is in mg/L as $CaCO_3$.

TABLE 1: Comparison of previous and updated chronic aluminum multiple linear regression model statistics for *Ceriodaphnia dubia* and *Pimephales promelas*

Endpoint statistic	Dataset	No.	Terms excluded from model	Model	Adj. R^2	Pred. R^2	Slopes							
							ln(DOC)	pH	ln(Hard)	pH^2	ln(DOC) × pH	ln(DOC) × ln(Hard)	ln(Hard) × pH	Intercept
Ceriodaphnia dubia														
EC10	Previous	23	ln(DOC) × pH[a]	AIC	0.719	0.572	1.208	13.338	3.639	−0.737	ex	−0.155	−0.490	−51.420
				BIC	0.685	0.628	0.571	2.282	2.900	—	ex	—	−0.365	−11.055
	Updated	30	None	AIC	0.929	0.910	0.697	6.089	2.768	−0.269	—	—	−0.335	−32.273
				BIC	0.925	0.902	0.700	2.368	3.030	—	—	—	−0.375	−11.824
			pH^2	AIC	0.925	0.902	0.700	2.368	3.030	ex	—	—	−0.375	−11.824
				BIC	0.925	0.902	0.700	2.368	3.030	ex	—	—	−0.375	−11.824
			pH^2, ln(DOC) × ln(Hard)	AIC	0.925	0.902	0.700	2.368	3.030	ex	—	ex	−0.375	−11.824
				BIC	0.925	0.902	0.700	2.368	3.030	ex	—	ex	−0.375	−11.824
EC20	Previous	23	ln(DOC) × pH[a]	AIC	0.726	0.607	0.525	11.282	2.201	−0.663	ex	—	−0.264	−41.026
				BIC	0.710	0.571	0.502	10.709	0.378	−0.695	ex	—	—	−35.571
	Updated	30	None	AIC	0.928	0.901	0.294	6.020	1.924	−0.309	—	0.077	−0.206	−22.226
				BIC	0.925	0.903	0.615	7.155	2.219	−0.371	—	—	−0.251	−26.941
			pH^2	AIC	0.921	0.891	0.194	1.660	2.111	ex	—	0.102	−0.235	−6.960
				BIC	0.921	0.891	0.194	1.660	2.111	ex	—	0.102	−0.235	−6.960
			pH^2, ln(DOC) × ln(Hard)	AIC	0.913	0.886	0.619	2.021	2.581	ex	—	ex	−0.307	−9.272
				BIC	0.913	0.886	0.619	2.021	2.581	ex	—	ex	−0.307	−9.272
Pimephales promelas														
EC10	Previous	22	ln(DOC) × pH[a]	AIC and BIC	0.850	0.714	0.545	2.794	3.201	—	ex	—	−0.460	−12.009
	Updated	31	None	AIC and BIC	0.898	0.837	1.828	1.932	1.914	—	−0.193	—	−0.248	−6.700
EC20	Previous	22	ln(DOC) × pH[a]	AIC and BIC	0.874	0.724	0.503	3.131	3.443	—	ex	—	−0.494	−14.029
	Updated	31	None	AIC and BIC	0.923	0.873	2.209	2.041	1.862	—	−0.261	—	−0.232	−7.371

[a]See pp 85–86 in DeForest et al. (2018) for basis for excluding ln(DOC) × pH term in previous modeling.
DOC = dissolved organic carbon; EC10 = 10% effect concentration; EC20 = 20% effect concentration; ex = excluded; Hard = hardness; No. = sample size; — = slope not provided for AIC/BIC-selected model because it did not improve the model fit; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

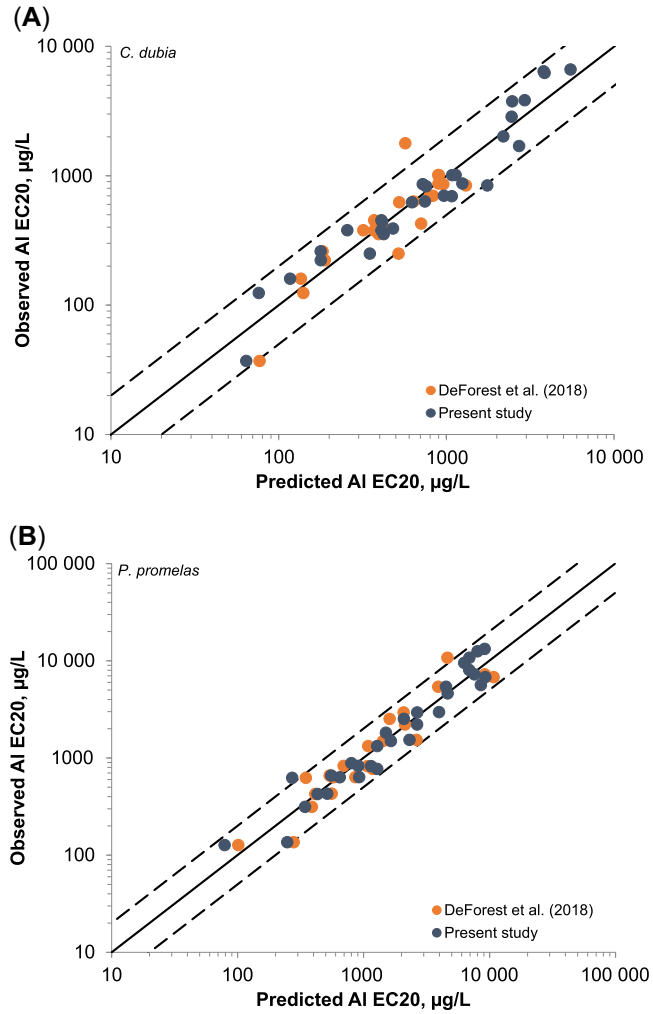

FIGURE 1: Comparison of multiple linear regression-predicted versus observed chronic total aluminum (Al) effect concentrations, 20% (EC20s) for (**A**) *Ceriodaphnia dubia* and (**B**) *Pimephales promelas* based on the previously developed models (DeForest et al. 2018) and updated models (present study). Two tests from McCauley et al. (1986) were excluded from the present study (see *Individual species models*). Solid line represents 1:1 agreement; dashed lines represent a factor of ±2 agreement between predicted and observed values.

Pimephales promelas

The updated chronic Al MLR models for *P. promelas* also had slightly higher adjusted R^2 values than the previous models, increasing from 0.850 to 0.898 for the EC10 model and from 0.874 to 0.923 for the EC20 model (Table 1). As for *C. dubia*, the predicted R^2 decreased only slightly relative to the adjusted R^2—from 0.898 to 0.837 for the EC10 model and from 0.923 to 0.873 for the EC20 model (Table 1). Ninety-four percent of predicted EC10s (29 of 31; Supplemental Data 4, Figure S1B) and 97% of predicted EC20s (30 of 31; Figure 1B) were within a factor of 2 of observed EC10/20 values.

No clear patterns in the residuals were observed over a wide range of predicted Al bioavailability conditions as expressed by the magnitudes of the EC10s and EC20s (Supplemental Data 4, Figure S6A and B) or relative to single independent variables (Supplemental Data 4, Figures S7A and B, S8A and B, and S9A and B).

As for *C. dubia*, Al EC20 predictions from the updated MLR model had similar patterns as the previous MLR model when plotted along with empirical EC20s from a series of tests in which DOC, pH, or hardness were individually varied. There was a stronger influence of DOC (28% steeper slope), a weaker influence of pH (8–12% shallower slopes), and a similar influence of hardness (2% shallower slope; Figure 3).

The terms retained in both the EC10 and EC20 MLR models were ln(DOC), pH, ln(Hard), ln(DOC) × pH, and ln(Hard) × pH (Table 1). DeForest et al. (2018) had ultimately excluded the ln(DOC) × pH term because data for tests with DOC concentrations >0.5 mg/L and pH 8 were limited to a single test for *P. promelas*. Using the updated *P. promelas* dataset that contains a broader range of DOC and pH conditions, the ln(DOC) × pH term was retained by AIC and BIC in the EC10 and EC20 models for *P. promelas*. However, the pH^2 term, which was retained in the previous *P. promelas* models, was not retained in the updated models (Table 1).

Thus, the *P. promelas* models differ from the recommended *C. dubia* models in that the former retains the ln(DOC) × pH term. It is difficult to speculate on the basis for these differences. However, the additional retained terms do not have a substantial influence on overall model performance. For example, when the ln(DOC) × pH term is not included in the *P. promelas* EC20 model, the adjusted R^2 decreases from 0.923 to 0.903. The negligible influence of these terms is further reflected in the pooled *P. promelas* and *C. dubia* models (see *Pooled models* section following), where neither the ln(DOC) × pH nor the pH^2 terms are retained in the recommended pooled EC10 and EC20 models.

The updated Al MLR EC10 and EC20 models for *P. promelas* are as follows:

$$\ln(EC10) = -6.700 + 1.828 \times \ln(DOC) + 1.932 \times pH + 1.914 \times \ln(Hard) - 0.193 \times \ln(DOC) \times pH - 0.248 \times \ln(Hard) \times pH \quad (3)$$

$$\ln(EC20) = -7.371 + 2.209 \times \ln(DOC) + 2.041 \times pH + 1.862 \times \ln(Hard) - 0.261 \times \ln(DOC) \times pH - 0.232 \times \ln(Hard) \times pH \quad (4)$$

where ln(EC10) and ln(EC20) are total Al concentrations in units of µg/L, DOC is in mg/L, and Hard is in mg/L as $CaCO_3$.

Pooled models

The pooled *C. dubia* and *P. promelas* MLR EC10 and EC20 models performed comparably to the species-specific *C. dubia* and *P. promelas* MLR models. Both AIC and BIC selected the same EC10 models, but different EC20 models. (Both models retained the ln[DOC], pH, ln[Hard], and ln[Hard] × pH terms, and AIC also retained the pH^2 and ln[DOC] × pH terms.) The adjusted R^2 for *C. dubia* and *P. promelas* using the pooled EC10 model were 0.916 and 0.875, respectively (Table 2). Based on the BIC-selected model, the adjusted R^2 values were

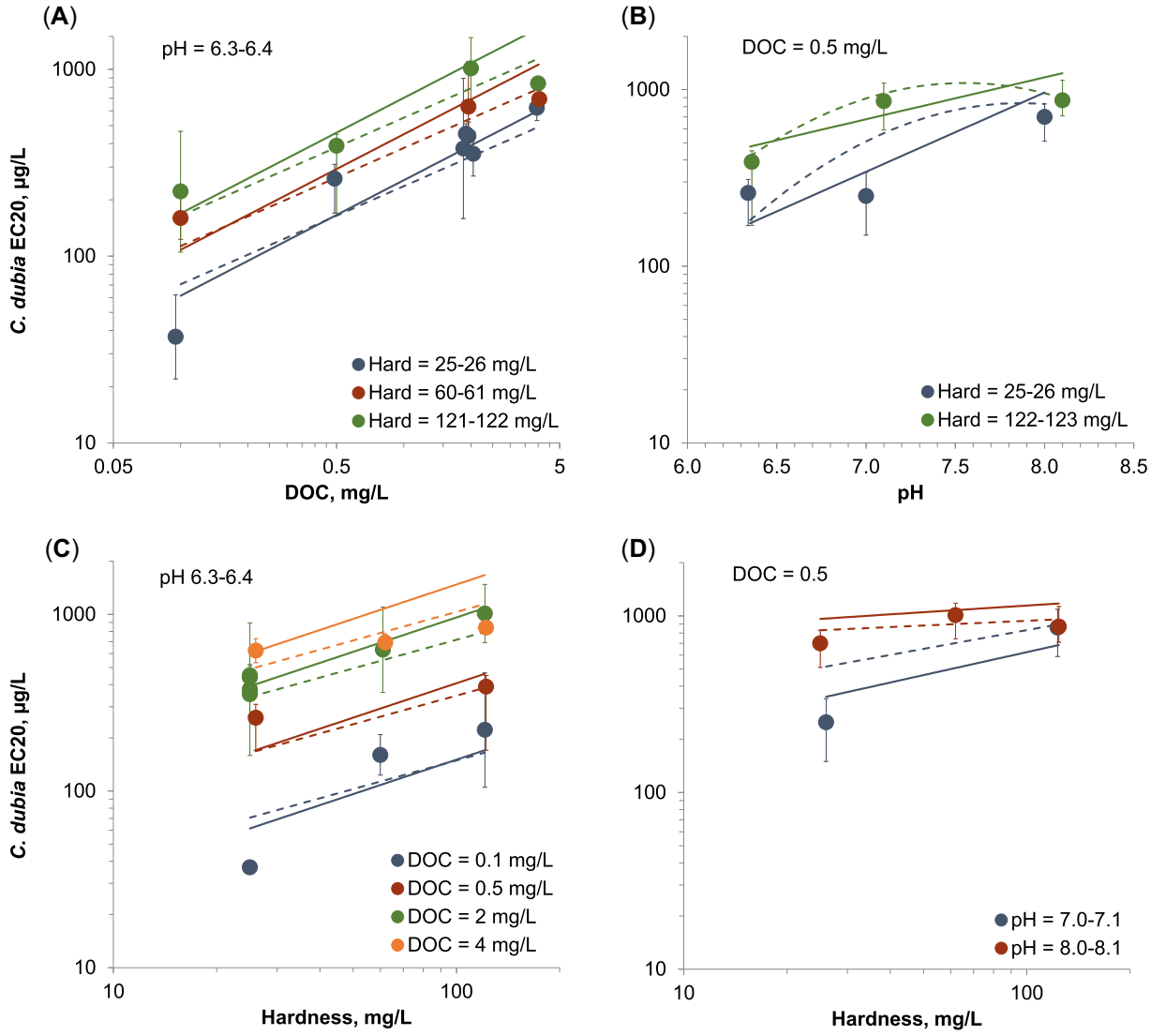

FIGURE 2: Observed and multiple linear regression (MLR)-predicted (regression lines) total aluminum (Al) effect concentrations, 20% (EC20s; ±95% confidence limits) for *Ceriodaphnia dubia* reproduction where one water chemistry parameter was varied. Dashed lines = previous MLR model (DeForest et al. 2018); solid lines = updated MLR model (present study). DOC = dissolved organic carbon.

0.895 for *C. dubia* and 0.881 for *P. promelas* (Table 2). The adjusted R^2 values from the AIC-selected model were comparable (Table 2). Because the BIC-selected EC20 model had fewer parameters than the AIC-selected model, but an almost identical adjusted R^2, the BIC-selected model is recommended as the pooled EC20 model. In addition, this EC20 model has the same parameters as the EC10 model (i.e., ln[DOC], pH, ln[Hard], and ln[Hard] × pH).

The pooled Al MLR models provided a similar level of accuracy in EC10 and EC20 predictions for *C. dubia* and *P. promelas* as the species-specific MLR models. For *C. dubia*, the percentage of predicted EC10s and EC20s within a factor of 2 of observed based on the pooled model was unchanged relative to the *C. dubia*-specific model for EC10s (97%) and was improved for EC20s (increase from 97 to 100%; Figure 4A; EC20 model). For *P. promelas*, the percentage of predicted EC10s and EC20s within a factor of 2 of observed decreased from 94 to 87% for EC10s and decreased from 97 to 94% for EC20s (Figure 4B; EC20 model). No clear patterns were discerned in the residuals over a wide range of predicted Al bioavailability conditions as expressed by the magnitudes of the EC10s and EC20s (Supplemental Data 4, Figures S2C and D and S6C and D). Regarding patterns of residuals relative to single independent variables, there were weak patterns for *C. dubia*, with EC10s and EC20s tending to be underpredicted at high DOC and high hardness, but no clear patterns with relation to pH (Supplemental Data 4, Figures S3C and D, S4C and D, and S5C and D). For *P. promelas*, there was similarly a weak pattern of EC10s and EC20s tending to be underpredicted at low DOC and low hardness (Supplemental Data 4, Figures S7C and D, S8C and D, and S9C and D).

The pooled Al MLR models, with *C. dubia*- and *P. promelas*-specific intercepts, are as follows:

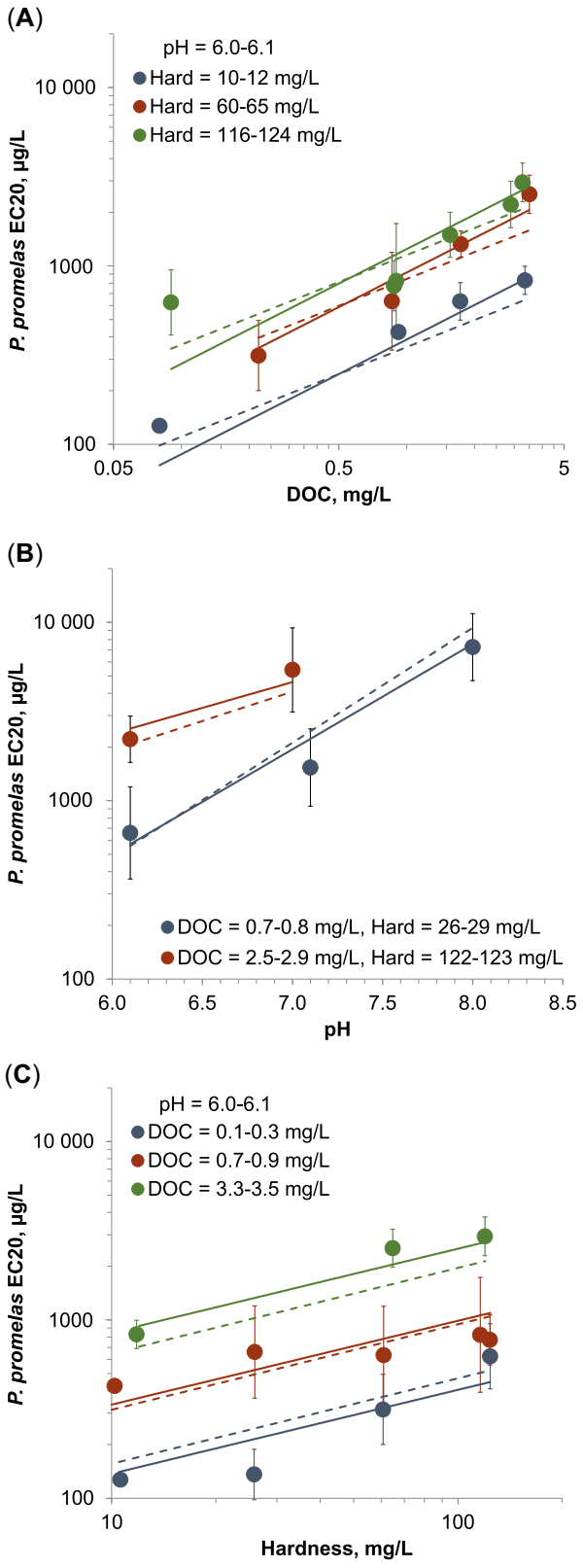

FIGURE 3: (**A–C**) Observed and multiple linear regression (MLR)-predicted (regression lines) total aluminum (Al) effect concentrations, 20% (EC20s; ±95% confidence limits) for *Pimephales promelas* biomass where one water chemistry parameter was varied. Dashed lines = previous MLR model (DeForest et al. 2018); solid lines = updated MLR model (present study). DOC = dissolved organic carbon.

C. dubia

$$\ln(EC10) = -8.474 + 0.660 \times \ln(DOC) + 1.969 \times pH + 2.216 \times \ln(Hard) - 0.278 \times \ln(Hard) \times pH \quad (5)$$

$$\ln(EC20) = -8.479 + 0.604 \times \ln(DOC) + 1.983 \times pH + 2.163 \times \ln(Hard) - 0.265 \times \ln(Hard) \times pH \quad (6)$$

P. promelas

$$\ln(EC10) = -7.424 + 0.660 \times \ln(DOC) + 1.969 \times pH + 2.216 \times \ln(Hard) - 0.278 \times \ln(Hard) \times pH \quad (7)$$

$$\ln(EC20) = -7.395 + 0.604 \times \ln(DOC) + 1.983 \times pH + 2.163 \times \ln(Hard) - 0.265 \times \ln(Hard) \times pH \quad (8)$$

Pooled MLR models that included the alga *R. subcapitata* did not perform as well as the pooled models based on *C. dubia* and *P. promelas*. Although application of the BIC-selected pooled model for EC10s resulted in adjusted R^2 values of 0.895 and 0.833 for *C. dubia* and *P. promelas*, respectively, the adjusted R^2 was just 0.354 for *R. subcapitata*. The same patterns were observed for the EC20 model, which resulted in adjusted R^2 values of 0.887, 0.846, and 0.336 for *C. dubia*, *P. promelas*, and *R. subcapitata*, respectively. For comparison, the adjusted R^2 values for the *R. subcapitata*-specific EC10 and EC20 models were 0.940 and 0.956, respectively (DeForest et al. 2018). Accordingly, it is recommended that the pooled model be limited to *C. dubia* and *P. promelas*.

SSDs and GSDs and HC5s

The pooled *C. dubia* and *P. promelas* MLR models were used to adjust the sensitivities of invertebrates and fish (EC10s and EC20s) to a range of DOC, pH, and hardness conditions; the *R. subcapitata* MLR model presented in DeForest et al. (2018) was used to adjust the sensitivities of algae and plants. Based on chronic toxicity data compiled for the present evaluation and following the USEPA approach for criteria development, an example of the GSD for invertebrates and fish based on EC20s adjusted to a DOC concentration of 2 mg/L, pH of 7, and hardness of 75 mg/L resulted in a total Al HC5 of 409 µg/L (Figure 5A). Following the European Union approach, an example of the species sensitivity distribution for algae/plants, invertebrates, and fish based on EC10s adjusted to the same DOC, pH, and hardness conditions resulted in a total Al HC5 of 243 µg/L (Figure 5B). The most sensitive genus was *Salvelinus* in Figure 5A, and the most sensitive species in Figure 5B is *S. fontinalis* (brook trout), based on an early life stage toxicity test (Cleveland et al. 1989). Accordingly, the differences in HC5s between the USEPA and European Union approaches, at least for these DOC, pH, and hardness conditions, were primarily related to differences in EC20s versus EC10s.

TABLE 2: Chronic aluminum multiple linear regression model statistics for pooled *Ceriodaphnia dubia* and *Pimephales promelas* dataset

Endpoint statistic	No.	Terms excluded from model	Model	Adj. R^2		Slopes							Intercepts	
				C. dubia	*P. promelas*	ln(DOC)	pH	ln(Hard)	pH^2	ln(DOC) × pH	ln(DOC) × ln(Hard)	ln(Hard) × pH	*C. dubia*	*P. promelas*
EC10	61	None	AIC and BIC	0.916	0.875	0.660	1.969	2.216	—	—	—	–0.278	–8.474	–7.424
EC20	61	None	AIC	0.899	0.878	1.333	4.338	1.643	–0.193	–0.114	—	–0.182	–15.670	–14.520
			BIC	0.895	0.881	0.604	1.983	2.163	—	—	—	–0.265	–8.479	–7.395

DOC = dissolved organic carbon; EC10 = 10% effect concentration; EC20 = 20% effect concentration; Hard = hardness; No. = sample size; — = slope not provided for AIC/BIC-selected model because it did not improve the model fit; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

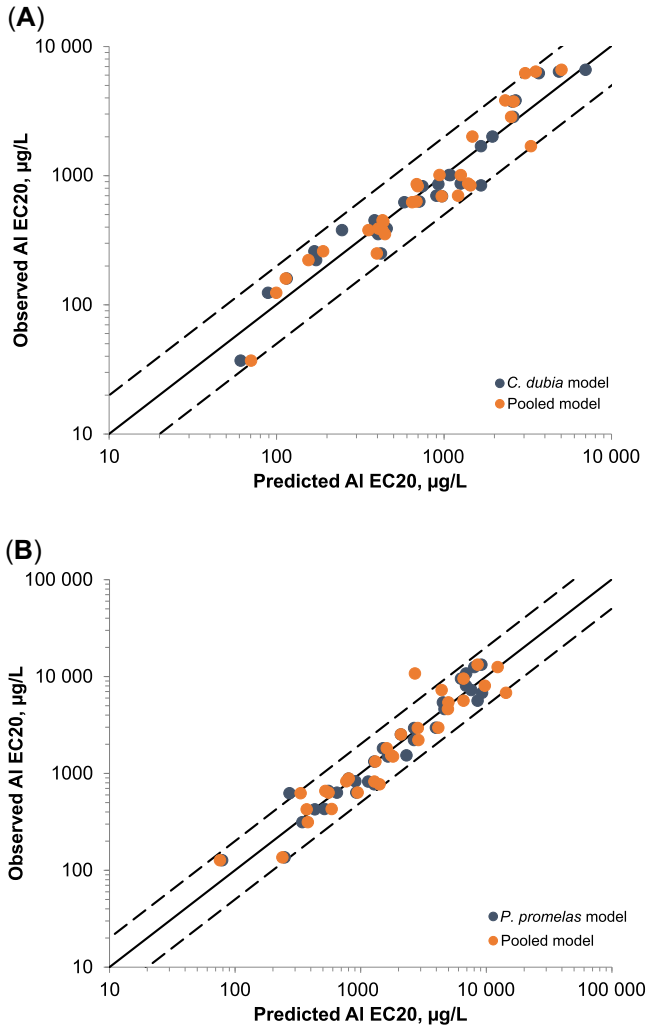

FIGURE 4: Comparison of multiple linear regression-predicted chronic total aluminum (Al) effect concentrations, 20% (EC20s) for (**A**) *Ceriodaphnia dubia* and (**B**) *Pimephales promelas* based on the individual species and pooled species models (both from present study). Solid line represents 1:1 agreement; dashed lines represent a factor of ±2 agreement.

Over a broad range of DOC, pH, and hardness conditions, the resulting HC5s can vary over 2 orders of magnitude. For example, based on a DOC range of 1 to 12 mg/L, a pH range of 6.0 to 8.5, and a hardness range of 10 to 400 mg/L, total Al HC5s based on the USEPA criteria approach for criteria development ranged from 37 to 4600 µg/L (Figure 6). A spreadsheet for calculating MLR-normalized HC5s for DOC, pH, and hardness conditions of interest, based on both the European Union and USEPA approaches, is provided in Supplemental Data 5.

Comparison of HC5s with the USEPA's chronic aluminum criteria

The USEPA used the MLR models described in DeForest et al. (2018) to develop draft MLR-based Al criteria (US Environmental Protection Agency 2017). The additional Al toxicity data, the updated species-specific *C. dubia* and *P. promelas*

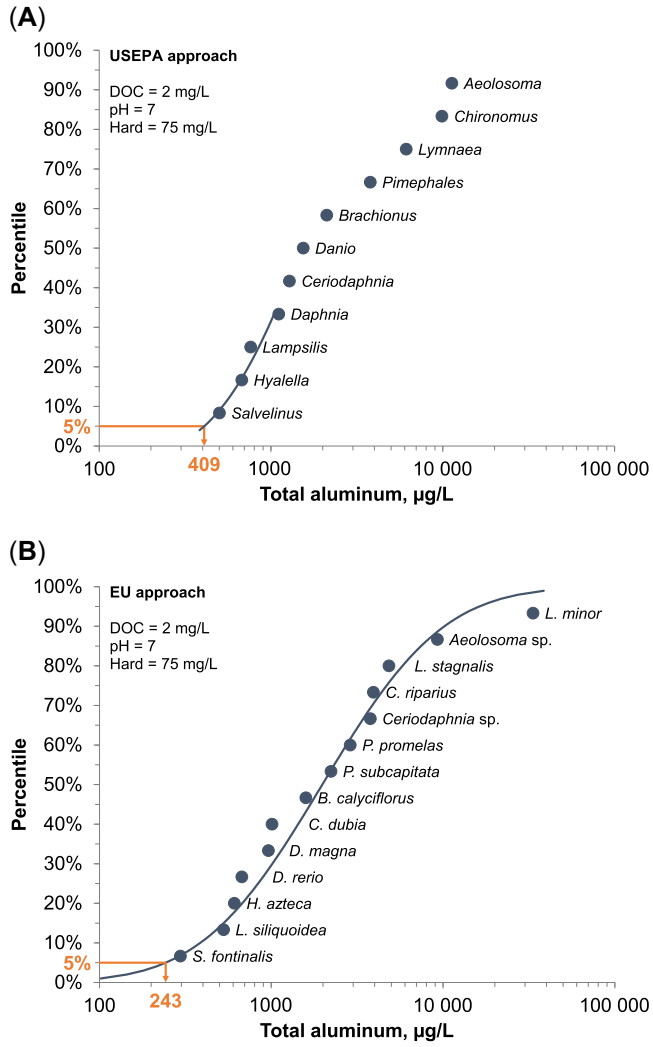

FIGURE 5: (**A**) Genus sensitivity distribution for invertebrates and fish based on chronic total aluminum (Al) effect concentrations, 20% (EC20s; US Environmental Protection Agency [USEPA] approach) and (**B**) species sensitivity distributions for algae/plants, invertebrates, and fish based on chronic total Al EC10s (European Union [EU] approach) adjusted to a dissolved organic carbon (DOC) of 2 mg/L, a pH of 7, and hardness of 75 mg/L. The EC10s and EC20s were adjusted using the pooled multiple linear regression model slopes. The hazardous concentration for 5% of the species (HC5) in (**A**) was based on triangular distribution fit to 4 lowest genus mean chronic values; the HC5 in (**B**) was based on log-normal distribution fit to all species mean chronic values. *L. stagnalis* = *Lymnaea stagnalis*; *C. riparius* = *Chironomus riparius*; *P. promelas* = *Pimephales promelas*; *P. subcapitata* = *Pseudokirchneriella subcapitata*; *B. calyciflorus* = *Brachionus calyciflorus*; *C. dubia* = *Ceriodaphnia dubia*; *D. magna* = *Daphnia magna*; *D. rerio* = *Danio rerio*; H. azteca = *Hyalella azteca*; *L. siliquoidea* = *Lampsilis siliquoidea*; *S. fontinalis* = *Salvelinus fontinalis*.

models, and the pooled *C. dubia* and *P. promelas* models were then evaluated by the USEPA in developing the final MLR-based Al criteria (US Environmental Protection Agency 2018). The *C. dubia* dataset in those evaluations included the 2 toxicity tests from McCauley et al. (1986), which were subsequently excluded in our analysis due to poor control performance (see *Materials and Methods*). In developing its final Al criteria, the USEPA decided to use the individual *C. dubia* and *P. promelas* models (including the McCauley et al. [1986] data) rather than the pooled *C. dubia* and *P. promelas* models. This decision was based on patterns in the residuals of the pooled model that were not observed in the individual species models. For example, the USEPA commented that EC20 predictions for *C. dubia* using the pooled model decreased with increasing pH and increased with increasing DOC and hardness, whereas no residual patterns were observed for EC20 predictions from the *C. dubia* species-specific model. Exclusion of the 2 tests from McCauley et al. (1986) in the present study does not change the overall patterns of residuals in the species-specific *C. dubia* model versus the pooled model (Supplemental Data, Figures S3–S5).

Species-specific models will obviously perform better for the species on which they are based compared with a pooled model. However, there are advantages to using a pooled model for criteria development over species-specific models that may outweigh the reduced performance of pooled models in predicting toxicity for a given species (Brix et al. 2020). Specifically, use of a pooled model ensures that all taxa in the GSD will respond to TMFs in the same way. Alternatively, if species-specific models are used to modify the GSD on a taxa-specific basis (e.g., separate models for invertebrates and fish as in the USEPA criteria for Al), then relatively small differences in the way these models respond to TMFs can change the relative ranking of taxa in the GSD and have unexpected consequences on the 5th percentile of the GSD. Whether these apparent differences reflect fundamental differences in the way taxa respond to TMFs or are simply an artifact of variability in organism response and/or differences in data availability between species can be difficult to discern.

The scenario just described is occurring for some water chemistry conditions using the species-specific model approach adopted by the USEPA (US Environmental Protection Agency 2018) for the revised Al criteria (Figure 6). Note that in these plots (Figure 6), the GSD used for the pooled model differs slightly from the GSD in the USEPA approach (US Environmental Protection Agency 2018), so comparisons should focus on comparing the response with changing TMFs rather than the absolute values of estimated HC5s. Two important observations arise from the plots in Figure 6. First, the only fundamental difference in response between the USEPA models and the pooled model in the present analysis is the effect of pH on DOC. In the USEPA models, the effect of DOC on Al toxicity declines with increasing pH such that at pH 8 there is effectively no DOC effect (Figure 6A–C). This pattern is not observed in the pooled model derived in the present study, where the DOC effect is independent of pH. These differences are caused by the inclusion of a negative pH^2 term in the *C. dubia* model and a negative DOC × pH term in the *P. promelas* model, whereas both terms were excluded from the pooled model. Available pH-dependent DOC data for *P. promelas* suggest that the pooled model is more appropriate (Figure 3B), whereas the available data for *C. dubia* are unclear (Figure 2B).

The other important observation is that when deriving chronic USEPA criteria or HC5s using the *C. dubia* and *P. promelas* species-specific models, several counterintuitive

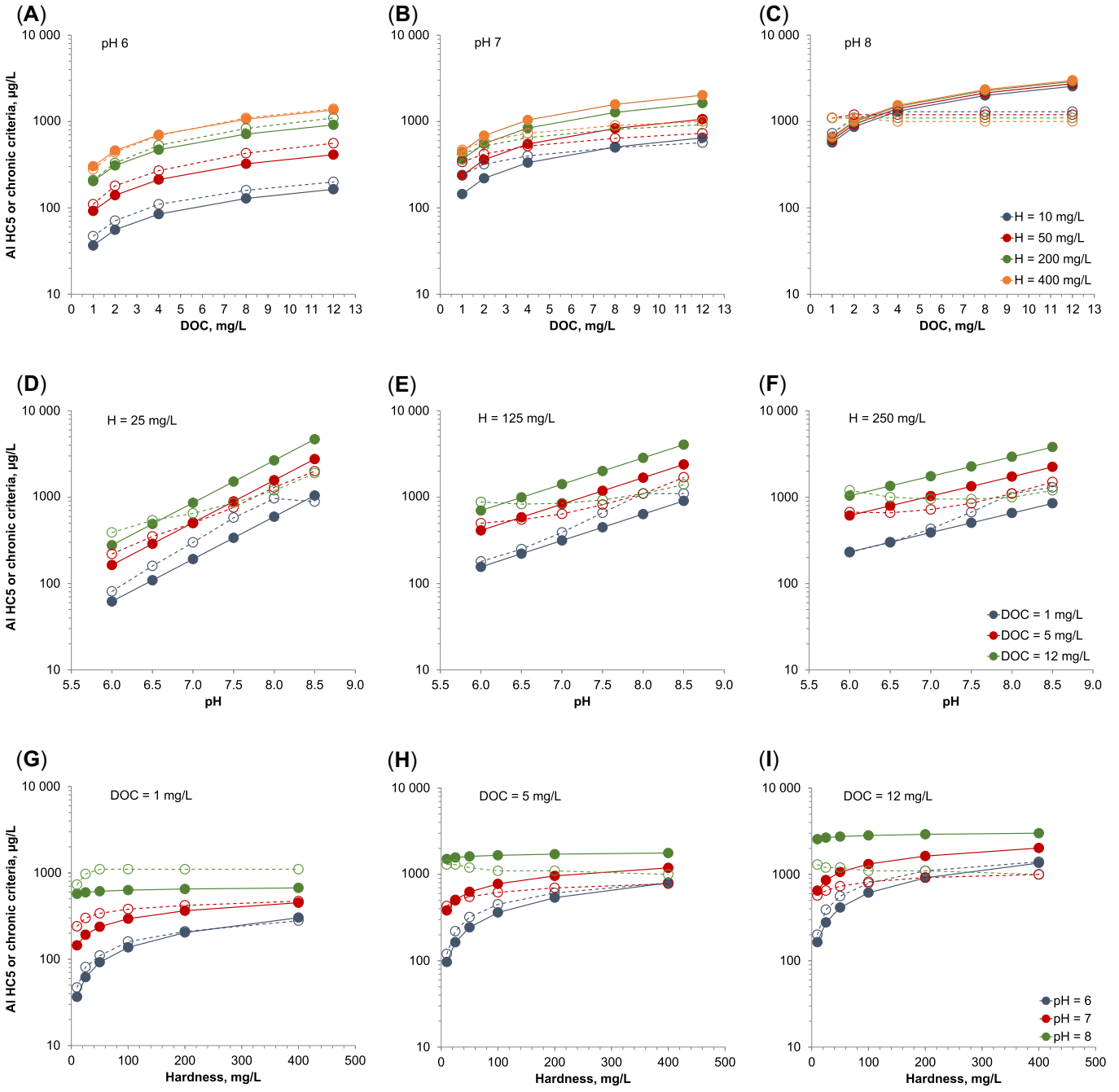

FIGURE 6: Total aluminum (Al) hazardous concentrations for 5% of the species (HC5s; present study: filled circles and solid lines) and chronic criteria (US Environmental Protection Agency 2018: open circles and dashed line) as a function of dissolved organic carbon (DOC) concentration (**A–C**), pH (**D–F**), and hardness (**G–I**). (**A–C**) Hardness of 10, 50, 200, and 400 mg/L. (**D–F**) Dissolved organic carbon of 1, 5, and 12 mg/L. (**G–I**) pH of 6, 7, and 8. H = hardness.

patterns can be observed as a function of the TMFs. For example, at a hardness of 125 mg/L, chronic USEPA criteria for high DOC waters (12 mg/L) are relatively constant with increasing pH, whereas in moderate DOC waters (5 mg/L), the chronic USEPA criteria increase continuously over a pH range of 6.0 to 8.5 (Figure 6E). The net result of these differential responses is that at pH 8.5, the chronic USEPA criterion for a water with 5 mg/L DOC is higher than for a water with 12 mg/L DOC. A similar juxtaposition is observed in high DOC and pH waters as a function of hardness (Figure 6I). There are no toxicity data that support either of these patterns.

The patterns just described appear to be the result of small differences in the way the species-specific models for *C. dubia* and *P. promelas* respond to the TMFs. These differences cause changes in the ranking of taxa in the GSD as a function of the TMFs. In the case of the USEPA criteria, in which the 5th percentile of the GSD is based on the distribution of the 4 most sensitive taxa, reranking of the 4 most sensitive taxa can have a

large influence on the slope of the distribution, which in turn can cause the criteria to change in a direction inconsistent with the toxicity data under some conditions.

For example, under the water chemistry scenarios just described (pH of 8.5 and hardness of 125 mg/L), *Salmo*, *Salvelinus*, and *Daphnia* are the 3 most sensitive taxa at a DOC of both 5 and 12 mg/L (Figure 7). Predicted EC20s for the fish taxa *Salmo* and *Salvelinus* change negligibly between a DOC of 5 and 12 mg/L, whereas the predicted *Daphnia* (an invertebrate) EC20 increases by 1.7-fold. The fourth most sensitive genus is *Lampsilis* (a bivalve) at a DOC of 5 mg/L, but it is replaced with *Danio* (fish) at a DOC of 12 mg/L. This rearrangement in relative ranking decreases the distribution slope at the higher DOC, which results in extrapolation to a lower 5th percentile (Figure 7A). This issue can be more pronounced for chemicals with smaller toxicity datasets, such as Al, where the 5th percentile is extrapolated beyond the sensitivity of the most sensitive taxa.

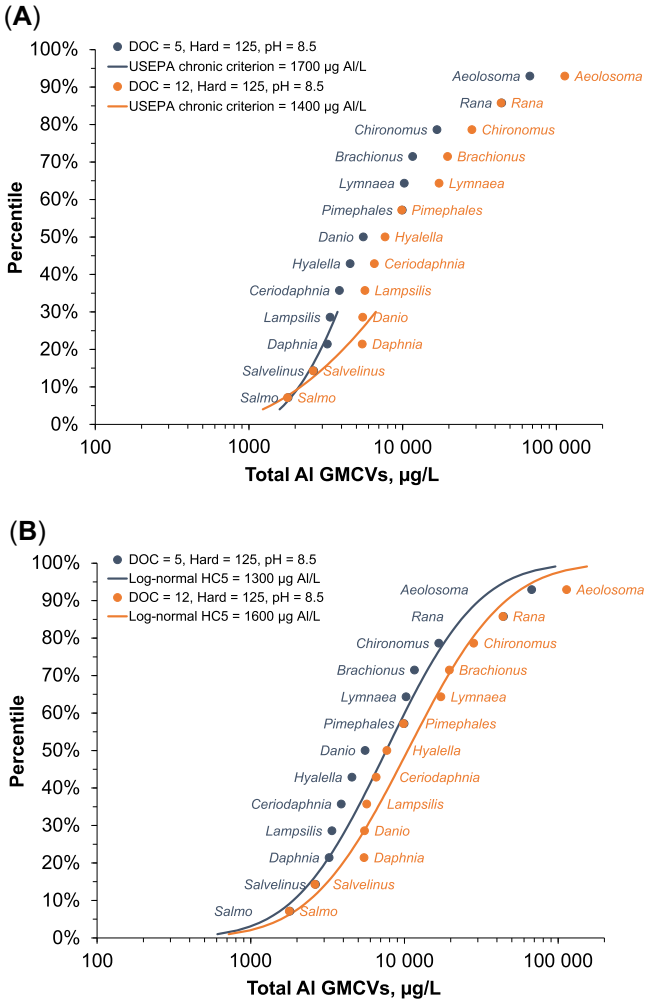

FIGURE 7: Genus sensitivity distributions based on chronic total aluminum (Al) effect concentrations, 20% (EC20s) from US Environmental Protection Agency (2018) adjusted to a pH of 8.5, a hardness of 125 mg/L, and a dissolved organic carbon (DOC) of 5 and 12 mg/L. The EC20s were adjusted using the individual *Ceriodaphnia dubia* and *Pimephales promelas* model slopes (*C. dubia* for invertebrates; *P. promelas* for fish). (**A**) The hazardous concentration for 5% of the species (HC5) was based on triangular distribution fit to 4 lowest genus mean chronic values following US Environmental Protection Agency (1985). (**B**) The HC5 was based on log-normal distribution to all genus mean chronic values.

The USEPA approach for calculating the 5th percentile based on the sensitivity of the 4 most sensitive taxa can also contribute to this issue, because the 5th percentile is driven by relative changes to just 4 taxa. If a model is fit to the entire GSD (a log-normal model in this example), the 5th percentiles at DOC concentrations of 5 and 12 mg/L are 1300 and 1600 mg/L, respectively, which are more consistent with the expected pattern based on the toxicity data (Figure 7B). Overall, the use of multiple models (e.g., separate models for invertebrates and fish) can sometimes result in unexpected criteria or HC5 patterns, and these effects can be compounded by issues such as the amount of toxicity data available and the method for estimating the 5th percentile.

The scenario just described provides a good example of why, if available data are supportive, a pooled model is preferred over species-specific models. However, this does not mean that there are not scenarios similar to the one just described where species-specific models should be maintained. In particular, in regulatory settings where algae are included in the SSD, both MLR models (DeForest et al. 2018) and BLMs (De Schamphelaere et al. 2005) for algae typically differ substantially from invertebrate and fish models. In these cases, it is entirely possible that differential responses to TMFs across broad taxonomic groups could lead to patterns similar to those just discussed.

Lastly, we note that recently developed methods for comparing the performance of different models for predicting metals toxicity as a function of varying TMFs could also have been applied to the comparison of individual and pooled species models (Brix et al. 2020; Garman et al. 2020). The primary components of the model comparison metrics from those studies are based on adjusted and predicted R^2 values and slopes of the residuals for TMFs. In the present evaluation, it is expected that the individual species models would perform better than the pooled species models based on these metrics; in fact, part of the USEPA's basis for selecting the individual species models for Al criteria development was related to patterns in residuals. However, the present study highlights additional factors that should be considered in selecting an appropriate model for criteria development. For Al it was observed that some unexpected patterns in criteria as a function of TMFs were observed when individual species models were used; such patterns appear to be an artifact of how invertebrate and fish species shift within the GSD with varying TMFs. Thus, factors in addition to those related to model performance should be considered when models for water quality criteria or guideline development are evaluated.

SUMMARY AND CONCLUSIONS

The updated chronic Al toxicity datasets for *C. dubia* and *P. promelas* expand the range of DOC, pH, and hardness conditions from which MLR models can be developed and to which

the models may be applied. The updated dataset also results in higher adjusted and predicted R^2 values compared with the models described in DeForest et al. (2018). Furthermore, the updated dataset supports development of a pooled MLR model for *C. dubia* and *P. promelas* that has comparably high adjusted and predicted R^2 values compared with the species-specific MLR models. The pooled model also allows for consistent responses to TMFs across all invertebrate and fish taxa, eliminating unexpected patterns in estimated HC5s as a function of TMFs that do not appear to be statistically supported.

Supplemental Data—The Supplemental Data are available on the Wiley Online Library at https://doi.org/10.1002/etc.4796.

Acknowledgment—We thank the Aluminium REACH Consortium for supporting additional Al toxicity testing and updated model development. We also thank the 3 anonymous reviewers for providing helpful comments that improved the manuscript.

Author Contribution Statement—A. Cardwell, W. Stubblefield, E. Nordheim, and W. Adams conceived and designed the experimental approach for the toxicity testing. A. Cardwell and W. Stubblefield conducted the toxicity testing. L. Tear conducted the MLR modeling. D. DeForest and K. Brix provided oversight of the MLR modeling and interpretation of results. D. DeForest was the primary author of the manuscript with support from K. Brix and L. Tear. A. Cardwell and W. Stubblefield developed the toxicity data reports included as Supplemental Data. A. Cardwell, W. Stubblefield, E. Nordheim, and W. Adams all reviewed and provided comments on the manuscript.

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

REFERENCES

Allen DM. 1971. Mean square error of prediction as a criterion for selecting variables. *Technometrics* 13:469–475.

Brix KV, DeForest DK, Tear L, Grosell M, Adams WJ. 2017. 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, DeForest DK, Tear L, Peijnenburg W, Peters A, Traudt E, Erickson R. 2020. Development of empirical bioavailability models for metals. *Environ Toxicol Chem* 39:85–100.

Burnham KP, Anderson DR. 2004. Multimodel interference: Understanding AIC and BIC in model selection. *Sociol Method Res* 33:261–304.

Cardwell AS, Adams WJ, Gensemer RW, Nordheim E, Santore RC, Ryan AC, Stubblefield WA. 2018. Chronic toxicity of aluminum, at a pH of 6, to freshwater organisms: Empirical data for the development of international regulatory standards/criteria. *Environ Toxicol Chem* 37:36–48.

Cleveland L, Little EE, Wiedmeyer RH, Buckler DR. 1989. Chronic no-observed-effect concentrations of aluminum for brook trout exposed in low-calcium, dilute acidic water. In Lewis TE, ed, *Environmental Chemistry and Toxicology of Aluminum*. Lewis, Chelsea, MI, USA, pp 229–245.

DeForest DK, Brix KV, Tear LM, Adams WJ. 2018. Multiple linear regression models for predicting chronic aluminum toxicity to freshwater aquatic organisms and developing water quality guidelines. *Environ Toxicol Chem* 37:80–90.

De Schamphelaere 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.

European Chemicals Agency. 2008. Characterisation of dose [concentration]-response for environment. *Guidance on Information Requirements and Chemical Safety Assessment*. Helsinki, Finland.

Garman ER, Meyer JS, Bergeron CM, Blewett TA, Clements WH, Elias MC, Farley KJ, Gissi F, Ryan AC. 2020. Validation of bioavailability-based toxicity models for metals. *Environ Toxicol Chem* 39:101–117.

Gensemer RW, Gondek JC, Rodriquez PH, Arbildua JJ, Stubblefield WA, Cardwell AS, Santore RC, Ryan AC, Adams WJ, Nordheim Eirik. 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.

Gensemer RW, Playle RC. 1999. The bioavailability and toxicity of aluminum in aquatic environments. *Crit Rev Environ Sci Technol* 29:315–450.

McCauley DJ, Brooke LT, Call DJ, Lindberg CA. 1986. Acute and chronic toxicity of aluminum to *Ceriodaphnia dubia* at various pH's. Testing laboratory: Center for Lake Superior Environmental Studies, University of Wisconsin-Superior, Superior, WI, USA. Owner company: Battelle Memorial Research Institute, Columbus, OH, USA. Study no. F-4114(8834) -411, work assignment no. 45, task order no. 4. Report date: 1986-11-01.

Mount DR, Mount DI. 1992. A simple method of pH control for static and static-renewal aquatic toxicity tests. *Environ Toxicol Chem* 11:609–614.

Neter J, Wasserman W, Kutner MH. 1990. *Applied Linear Statistical Models*, 3rd edition. Irwin, Homewood, IL, USA.

Piñeiro G, Perelman S, Guerschman JP, Paruelo JM. 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol Model* 216:316–322.

Santore RC, Ryan AC, Krogland 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.

US Environmental Protection Agency. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. PB85-227049. US Environmental Protection Agency, National Technical Information Service, Springfield, VA.

US Environmental Protection Agency. 1994. Determination of metals and trace elements in water and wastes by inductively coupled plasma-atomic emission spectrometry in methods for the determination of metals in environmental samples, Supplement 1, Method 200.7 EPA/600/R-94/111. National Exposure Research Laboratory (NERL), Washington, DC.

US Environmental Protection Agency. 2002. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. EPA-821-R-02-013. Washington, DC.

US Environmental Protection Agency. 2017. Draft aquatic life ambient water quality criteria for aluminum 2017. EPA-822-P-17-001. Washington, DC.

US Environmental Protection Agency. 2018. Final aquatic life ambient water quality criteria for aluminum 2018. EPA-822-R-18-001. Washington, DC.

US Environmental Protection Agency. 2020. EPA response to public comments on 2017 draft aluminum ambient water quality criteria (2020). EPA-822-R-20-004. Washington, DC.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 1:3–14.

