Farley, Kevin J.; Meyer, Joseph S.; Balistrieri, Laurie S.; De Schamphelaere, Karel A.C.; Iwasaki, Yuichi; Janssen, Colin R.; Kamo, Masashi; Lofts, Stephen; Mebane, Christopher A.; Naito, Wataru; Ryan, Adam C.; Santore, Robert C.; Tipping, Edward. 2015. **Metal mixture modeling evaluation project: 2. Comparison of four modeling approaches**. *Environmental Toxicology and Chemistry*, 34 (4). 741-753. 10.1002/etc.2820

Running head: MMME Comparison of Four Modeling Approaches

Corresponding author: Dr. Kevin J. Farley, Manhattan College

4513 Manhattan College Parkway, Riverdale, NY 10471 USA;

E-mail: Kevin.Farley@manhattan.edu;

METAL MIXTURE MODELING EVALUATION PROJECT: 2. COMPARISON OF FOUR MODELING APPROACHES

Kevin J. Farley,*† Joseph S. Meyer,‡ Laurie S. Balistrieri,§ Karel A.C. De Schamphelaere,‖ Yuichi Iwasaki,# Colin R. Janssen,‖ Masashi Kamo,†† Stephen Lofts,‡‡ Christopher A. Mebane,§§ Wataru Naito,†† Adam C. Ryan,‖‖ Robert C. Santore,‖‖ and Edward Tipping‡‡

† Department of Civil and Environmental Engineering, Manhattan College, 4513 Manhattan College Parkway, Riverdale, New York 10471 USA;

‡ ARCADIS U.S., Inc., 1687 Cole Boulevard, Suite 200, Lakewood, Colorado 80401 USA;

§ U.S. Geological Survey, School of Oceanography, University of Washington, Box 355351, Seattle, Washington 98195 USA;

‖ Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, Jozef Plateaustraat 22, B-9000 Gent, Belgium;

Department of Civil Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552 Japan;

†† Research Institute of Science for Safety and Sustainability, National Institute of Advanced Industrial Science and Technology, 16-1 Onogawa, Tsukuba, Ibaraki 305-8569 Japan;

‡‡ Centre for Ecology and Hydrology, Library Avenue, Bailrigg, Lancaster LA1 4AP UK;

§§ U.S. Geological Survey, 230 Collins Road, Boise, Idaho 83702 USA;

‖‖ HDR|HydroQual, 1304 Buckley Road, Suite 202, Syracuse, NY 13212 USA

* Corresponding author: Kevin.Farley@manhattan.edu

ABSTRACT

As part of the Metal Mixture Modeling Evaluation (MMME) project, models were developed by the National Institute of Advanced Industrial Science and Technology (Japan), the U.S. Geological Survey (USA), HDR|HydroQual, Inc. (USA), and the Centre for Ecology and Hydrology (UK) to address the effects of metal mixtures on biological responses of aquatic organisms. A comparison of the 4 models, as they were presented at the MMME Workshop in Brussels, Belgium (May 2012), is provided herein. Overall, the models were found to be similar in structure (free ion activities computed by WHAM; specific or non-specific binding of metals/cations in or on the organism; specification of metal potency factors and/or toxicity response functions to relate metal accumulation to biological response). Major differences in modeling approaches are attributed to various modeling assumptions (e.g., single versus multiple types of binding site on the organism) and specific calibration strategies that affected the selection of model parameters. The models provided a reasonable description of additive (or nearly additive) toxicity for a number of individual toxicity test results. Less-than-additive toxicity was more difficult to describe with the available models. Because of limitations in the available datasets and the strong inter-relationships among the model parameters (log K_M values, potency factors, toxicity response parameters), further evaluation of specific model assumptions and calibration strategies is needed.

Key words (<5 words): Biotic ligand model, Concentration addition, Metal bioavailability, Metal toxicity, WHAM-F_{TOX}

Note to the editor and reviewers: This is one of 11 manuscripts under consideration for an ET&C Special Section on Metal Mixtures. The Section includes an introduction, a technical background, a comparison of multiple modeling approaches, a lessons-learned manuscript, and seven manuscripts on specific modeling and interpretation approaches. While each manuscript should be able to stand alone, the individual manuscripts are interrelated and cross-reference each other. If another cross-referenced, submitted manuscript is essential to complete the review of the present manuscript, please request the other manuscript from the Corresponding Guest Editor, copying the handling editor. The Corresponding Guest Editor for the series is Eric Van Genderen (evangenderen@zinc.org). Any unpublished material provided to assist your review must also be treated in confidence.

INTRODUCTION

In regulatory applications, metal-mixture toxicity has generally been modeled by Toxic Unit (TU) or other additive approaches that are based on water-exposure concentrations [1]. Data reviews [2,3] have shown that additive approaches based on dissolved-metal concentrations are not always sufficient in predicting mixture toxicity. Rather, metal-mixture toxicity tests have shown a wide range of organism responses with no clear patterns in additive and non-additive behavior. As part of an effort to address metal-mixture toxicity, several quantitative models have been developed to evaluate responses of aquatic organisms to metal mixtures and ultimately to provide a priori predictions of toxicity. Two modeling frameworks have been considered for this purpose: the Biotic Ligand Model (BLM) (as first presented by Di Toro et al. [4]) and WHAM-F_{TOX} [5].

In the BLM framework, metal bioavailability is evaluated by considering competitive interactions of metals and cations for binding to dissolved organic matter (DOM) and inorganic ligands (e.g., HCO_3^-, CO_3^{2-}, Cl^-) using the Windermere Humic Aqueous Model (WHAM). Competitive binding of metals and cations is also assumed to occur at binding sites on or in biological organisms, which are referred to as the "biotic ligand(s)." The accumulation of metal on the biotic ligand is then correlated to the toxic response of the organism (e.g., using a logit response function). The BLM has been used by various investigators over the past decade to develop predictive models for acute and chronic toxicity in single-metal exposures [6,7]. More recently, several BLMs have been developed or revised for metal mixtures [8,9,10,11]. In these models, metals have either been assumed to exhibit similar joint action (with toxicity expressed in terms of concentration addition of metal accumulation on the biotic ligand) [8,9,10], or independent joint action (with toxicity expressed in terms of a multiplicative function of the responses to the individual metals) [11].

WHAM-F_{TOX} [5] was specifically developed to address the effects of metal mixtures on aquatic organisms. This approach uses WHAM to evaluate competitive interactions of metals and cations on DOM and inorganic ligands. In contrast to the BLM, WHAM-F_{TOX} does not explicitly consider competitive binding of metals and cations to a biotic ligand. Rather, the model assumes that non-specific accumulation of metabolically-active metals by the organism is proportional to metal concentrations predicted to accumulate on humic acid (HA, when exposed to the same exposure water (as calculated by WHAM). Accumulated metal is related to toxicity using the F_{TOX} function, which is obtained by multiplying the calculated humic-bound metal and proton concentrations (mmol g^{-1}) by cation-specific potency factors and then summing the results over all cations. The resulting F_{TOX} value is correlated to the toxic response of the organism

using a linear-threshold response function, in which the effect thresholds depend only on the organism (i.e., independent of water chemistry).

The Metal Mixture Modeling Evaluation (MMME) project was initiated to assess the current capabilities of BLM- and WHAM-F_{TOX}-type models in predicting metal-mixture toxicity and to promote the continued development of various modeling approaches [12]. As part of that initiative, models were developed/refined and tested by researchers from (1) the National Institute of Advanced Industrial Science and Technology (AIST) in Ibaraki, Japan; (2) the U.S. Geological Survey (USGS) in Seattle, Washington and Boise, Idaho, USA; (3) HDR|HydroQual, Inc. (HDR) in East Syracuse, New York, USA; and (4) the Centre for Ecology and Hydrology (CEH) of the Natural Environment Research Council in Lancaster, UK. The purpose of this paper is to summarize and compare the 4 models, as they were presented at the MMME Workshop in Brussels, Belgium in May 2012. Revised versions of the initial AIST, USGS, HDR and CEH models are presented in this issue [8,9,10,11,13]. Because the published version of the USGS model differed substantially from the 2012 version evaluated in this article and because the 2012 version of the USGS model is not otherwise available, it is included for reference in Supporting Information File SI-2.

MODEL DESCRIPTIONS

The 4 models presented at the MMME Workshop were based on previous developments of the BLM and WHAM-F_{TOX}. Although the models were based on differing frameworks, they shared many similarities in their overall structure. These similarities included:

- A chemical speciation calculation to compute the free ion activities of metals and major cations based on competitive binding to DOM and inorganic ligands using various versions of WHAM.
- An evaluation of competing binding of metals and major cations to 1 or more binding sites on an organism using conventional competitive equilibrium chemistry (by either considering biotic ligand(s) or using metal binding to HA as a surrogate for non-specific metal accumulation by the organism).
- A correlation of accumulated metal to toxicity using potency factors and/or toxicity-response functions.

However, the 4 models differed in the details of their formulations and in calibration procedures that were used by the different modeling groups in fitting the MMME project datasets. Those datasets are listed in Table 1; formulations of the 4 models are summarized in Table 2, and further details are provided below.

AIST model

The AIST model followed the BLM framework and considered a single biotic ligand as the binding site for all metals on the organism [8,14]. Free ion activities of metals and other cations were calculated using WHAM VII [15] . For this calculation, DOM was assumed to be 100% fulvic acid (FA) and was set directly equal to the reported dissolved organic carbon (DOC) concentration. Thus, conversion from DOC (mg L^{-1}) to the WHAM input for FA (g L^{-1}) was FA = 0.001 × DOC. Carbonate species were not included in the calculations. WHAM-calculated free ion activities were then used in computing competitive binding of metal(s) and major cations on the biotic ligand. Initial estimates of binding constants (log K_M values) for metals and major

cations to the biotic ligand were obtained from previous studies (e.g., for trout [16,17]; for *Daphnia magna* [18,19]; for *D. pulex* [20,21]).

Toxicity was expressed as a function of the fractional coverage of accumulated metal on the biotic ligand. Following the premise that toxicity is not caused by a metal-induced response but rather by the role of the metal in blocking Ca uptake sites [14], all metals were assumed to elicit equally potent toxicological responses when bound to the biotic ligand. According to this assumption, metals exhibit similar joint action and metal-mixture toxicity can be described by a concentration-additive approach based on the accumulation of total metal on the biotic ligand. Response functions considered in the AIST model included a 2-parameter logit (mortality) function for rainbow trout (*Oncorhynchus mykiss*), cutthroat trout (*O. clarkii*) and daphnids, and a 2-parameter linear (growth reduction) function for a freshwater alga (*Pseudokirchneriella subcapitata*). For example, the 2-parameter logit function is given as:

$$R = \frac{1}{1 + e^{-(a + b\,\theta_M)}} \qquad (1)$$

where R is the biological response (e.g., fractional mortality or growth reduction), a and b are the logit parameters, and θ_M is the fractional coverage of accumulated metals on the biotic ligand.

The AIST model was calibrated using 4 of the 6 MMME calibration datasets (Table 1). For each dataset, the model was fit to observed mortality (or growth-reduction) responses by adjusting the logit response-function parameters. For *D. magna* and algae, log K_M values were also adjusted, with a different set of log K_M values used for each organism (Supporting Information File SI-1, Tables S1 and S2). Metal-mixture toxicity was predicted using the calibrated log K_M values and the response-function parameters derived from single-metal exposure studies.

USGS model

The 2012 version of the USGS model also followed the BLM framework and considered a single biotic ligand as the binding site on the organism (see Supporting Information File SI-2). Free ion activities of metals and cations were calculated using WHAM VII and the following assumptions: DOM was specified as 2 times the reported DOC concentration, 65% of the DOM was assumed to be active, and DOM was considered to be composed of 10% HA and 90% FA. Thus, conversions from DOC (mg L^{-1}) to WHAM inputs for HA and FA (in g L^{-1}) were: HA = $2\times0.65\times0.1\times0.001\times$DOC and FA = $2\times0.65\times0.9\times0.001\times$DOC. In addition, carbonate species were included in the calculations (by specifying pH and alkalinity). For field-collected water samples, free ion activities of Al^{3+} and Fe^{3+} were assumed to be in equilibrium with amorphous iron and aluminum hydroxides using solubility relationships [22,23]. WHAM-calculated free ion activities were then used in computing competitive binding of metals and major cations on a single-site biotic ligand. Binding constants (log K_M values) for metals and major cations to the biotic ligand were determined from a re-evaluation of data from single-metal toxicity studies on rainbow and cutthroat trout (Supporting Information File SI-1, Table S3). The log K_M values remained constant and did not differ among biological species.

Toxicity was expressed as a function of the fractional coverage of accumulated metal on the biotic ligand. In contrast to the AIST model, metals were assumed to have different potencies when bound to the biotic ligand. This effect was included in the model by incorporating a *TOX* function to account for apparent differences in toxicities of the various metals. In this approach, the *TOX* function is conceptually similar to the F_{TOX} function [5]. However, 1 major difference is that the *TOX* function is related to the fractional coverage of metal on the biotic ligand, whereas F_{TOX} is expressed as a function of non-specific accumulation of metal on HA (in mmole g^{-1}).

Toxic response was then determined in 2 steps. First, a potency factor (α_i) was defined to account for the relative toxicity of different metals when bound to the biotic ligand. This factor was multiplied by the fraction coverage of metal on the biotic ligand (θ_i) to calculate the toxic potency of a specified metal, as defined by the *TOX* function:

$$TOX_i = \alpha_i \theta_i \quad (2)$$

The model was extended to metal mixtures using a concentration-addition type of approach. However, because each metal was considered to exhibit a different potency when bound to the biotic ligand, calculations were based on the summation of TOX_i values.

$$TOX = \sum_{i=1}^{n} TOX_i = \sum_{i=1}^{n} \alpha_i \theta_i \quad (3)$$

where *n* is the number of metals in the mixture. Second, a 3-parameter logit function was used to calculate biological response as a function of *TOX*.

$$R = \frac{1}{\left(1 + e^{-(\beta_1 + \beta_2 \log TOX)}\right)^{\beta_3}} \quad (4)$$

where *R* is the biological response (e.g., fractional mortality or growth reduction), and β_1, β_2, and β_3 are the logit parameters. The model was calibrated using 5 of the 6 MMME calibration datasets (Table 1). For each dataset, the model was fit to observed mortality (or growth-reduction) responses from single-metal and metal-mixture exposures by adjusting potency factors (α_i) and the 3 logit parameters (β_1, β_2, β_3; Supporting Information File SI-1, Table S4). The potency factors were assumed to be dependent only on the metal, and the logit parameters were considered to be organism-specific in the initial calibration of the model. However, it was necessary to consider organism-specific potency factors in fitting datasets for *P. subcapitata* growth.

178 *HDR model*

179 The HDR model extended the BLM approach by considering a separate biotic ligand for each
180 metal (i.e., a Cd-specific, Cu-specific, Pb-specific, and Zn-specific biotic ligand). All metals
181 could compete for all binding sites, but only 1 metal was considered to be toxicologically active
182 at a given biotic ligand. The remainder of the model followed BLM calculations for single-metal
183 exposures [24]. Within the HDR model, metal and cation binding to DOM and inorganic ligands
184 were calculated based on WHAM V [25] and the following assumptions: DOM was specified as
185 2 times the reported DOC concentrations, 100% of the DOM was assumed to be active, and
186 DOM was considered to be composed of 10% HA and 90% FA. Thus, conversions from DOC
187 (mg L^{-1}) to WHAM inputs for HA and FA (in g L^{-1}) were: HA = $2\times0.1\times0.001\times$DOC and FA =
188 $2\times0.9\times0.001\times$DOC. In addition, carbonate species were included in the calculations (by
189 specifying pH and alkalinity). Because HDR did not explicitly consider the effects of Al or Fe
190 binding, Al^{3+} and Fe^{3+} were not included in the model.

191 In addition to metal and cation binding to DOM and inorganic ligands, the HDR model also
192 included simultaneous calculations for metal and cation binding to each of the metal-specific
193 biotic ligands. For each biotic ligand, log K_M values for the toxicologically-active metal and
194 competing cations were obtained from previously-calibrated, single-metal BLMs [24]. For the
195 remaining metals that were not toxicologically-active at a given biotic ligand but could compete
196 for binding to the site, log K_M values were initially set equal to their log K_M values on their
197 toxicologically-active biotic ligands. For example, the log K_M value for Cd on the Cu-specific
198 biotic ligand was set equal to the log K_M value for Cd on the Cd-specific biotic ligand. However,
199 adjustments in some of the log K_M values were made during model calibration. The final log K_M
200 values for the HDR model are presented in Supporting Information File SI-1, Table S5.

Toxic response at each of the biotic ligands was correlated to the concentration of metal '*i*' on its toxicologically-active biotic ligand using a 2-parameter logit function:

$$R_i = \frac{1}{1 + e^{-(a_i + b_i \log v_i)}} \quad (5)$$

where R_i represents the biological response (e.g., mortality) due to metal '*i*'; a_i and b_i are the metal-specific logit parameters; and v_i is the concentration of metal '*i*' on its toxicologically-active biotic ligand (in nmol g^{-1}). For metal mixtures, the overall response (R) was determined by assuming independent joint action and expressing toxicity in terms of a multiplicative function of individual metal responses as:

$$R = 1 - \prod_{i=1}^{n} (1 - R_i) \quad (6)$$

where n is the number of metals in the mixture. This approach is also referred to as response addition (see [1] for further discussion).

The HDR model was calibrated using 3 of the 6 MMME calibration datasets (Table 1). The model was calibrated separately for multiple data series within a given dataset by adjusting logit intercepts (a_i) and slopes (b_i) to fit observed mortalities for single-metal exposures [11]. These analyses ultimately provided a global fit of the logit slope and a distribution of logit intercepts that were used to quantify the unexplained variance or uncertainty associated with the single-metal exposure data (Supporting Information File SI-1, Table S6). Metal-mixture toxicity was predicted using the log K_M values and calibrated logit parameters from single-metal exposures. The HDR model also considered uncertainty in single-metal toxicity predictions to generate response envelopes for metal-mixture exposures (see [11] for details). Log K_M values for Cu and Zn on the Cd-specific biotic ligand were subsequently adjusted to provide a better calibration of the model to the *D. magna* mortality data ("Index 4").

223 *CEH model*

224 The CEH model was based on WHAM-F_{TOX} [5,26]. In this calculation, WHAM VI [27] was
225 used to be consistent with previous applications of WHAM-F_{TOX} [5,28]. Dissolved organic
226 matter was specified as 2 times the reported DOC concentration, 65% of the DOM was assumed
227 to be FA, and the remaining 35% of the DOM was considered to be inert with respect to
228 metal/cation binding. Thus, the conversion from DOC (mg L^{-1}) to WHAM inputs for FA (in g L^{-1}) was: $FA = 2 \times 0.65 \times 0.001 \times DOC$. Carbonate species were included in the calculations (by
230 assuming that total carbonate concentrations were equal to the reported alkalinity). For field-
231 collected water samples, free ion activities of Al^{3+} and Fe^{3+} were assumed to be in equilibrium
232 with amorphous iron and aluminum hydroxides using solubility relationships [22,29]. A small
233 concentration of WHAM HA (e.g., 10^{-10} g/L) was also included in the calculation. The resulting
234 concentrations of metals and protons on the HA were then used in toxicity calculations. In these
235 calculations, metal concentrations on HA included metals that were specifically-bound to HA
236 functional groups as well as metals that were nonspecifically-bound by electrostatic interactions
237 and held in close proximity to the HA (i.e., in the Donnan Layer). However, only protons that
238 were specifically-bound to HA functional groups were included in toxicity predictions. Because
239 the CEH model considered a distribution of binding sites in WHAM HA (see [27]),
240 representative composite binding constants were calculated for illustrative purposes by taking the
241 weighted-averages of the K_M values (see Supporting Information File SI-1, Table S7 for details).
242 (Note that log K_M values were also modified for electrostatic corrections in WHAM-F_{TOX}.)

243 Toxic response was determined by assuming that concentrations of metabolically-active
244 metals and protons on or in the organism were proportional to their predicted concentrations on
245 WHAM HA in the same exposure water. Toxicity was then determined in 2 steps. First, a

potency factor was defined to relate the amounts of accumulated metals and protons to toxic effect using the F_{TOX} function:

$$F_{TOX} = \sum_{i=1}^{n+1} \alpha_i \nu_i \quad (7)$$

where α_i is the relative potency factor, ν_i is the concentration of metal and protons on humic acid (in mmol g^{-1}), n is the number of metals in the mixture, and $n+1$ is considered to account for proton toxicity. Second, a 2 parameter linear response function was defined to relate toxic response to the F_{TOX} function

$$\begin{aligned} R &= 0 && \text{for } F_{TOX} < F_{TOX-LT} \\ R &= \frac{F_{TOX} - F_{TOX-LT}}{F_{TOX-UT} - F_{TOX-LT}} && \text{for } F_{TOX-LT} \le F_{TOX} \le F_{TOX-UT} \\ R &= 1 && \text{for } F_{TOX} > F_{TOX-UT} \end{aligned} \quad (8)$$

where $F_{TOX\text{-}LT}$ represents the lower threshold for toxicity and $F_{TOX\text{-}UT}$ represents the threshold for the maximum toxic response.

The model was calibrated using all 6 MMME calibration datasets (Table 1). For each dataset, the model was fit to observed mortality (or growth-reduction) responses from single-metal and metal-mixture exposures by adjusting relative potency factors (α_i) and the linear response parameters ($F_{TOX\text{-}LT}$, $F_{TOX\text{-}UT}$; Supporting Information File SI-1, Table S8). The relative potency factor for H^+ (α_H) was set equal to 1.0 in the calibration, effectively normalizing the potency of metals relative to that of H^+. Adjustments in $F_{TOX\text{-}LT}$ and $F_{TOX\text{-}UT}$ values were also examined in evaluations of the *D. magna* mortality data ("Index 4").

METHODS

In the present study, model performance was examined by first re-computing results for the 4 models using specifications described in the 'Model Descriptions' section. This step served as an independent check of results presented by the 4 modeling groups and also provided detailed model outputs that were subsequently used in model-model comparisons. Three of the larger calibration datasets were considered for this purpose, including:(i) "Index 8", which consisted of 114 test results for *P. subcapitata* growth in field-collected water samples spiked with Cd, Cu, Ni, and Zn (see Supporting Information File SI-3); (ii) "Index 4", which consisted of 561 test results for *D. magna* mortality in reconstituted laboratory water spiked with Cd, Cu, and Zn [30]; and (iii) "Index 6", which consisted of 369 test results for cutthroat and rainbow trout mortality in field-collected water samples spiked with Cd, Pb, and Zn [31]. Two additional datasets were considered for model validation: (i) "Index V-1", which consisted of 309 test results for *D. magna* mortality in reconstituted laboratory water spiked with Cd and Zn [30]; and (ii) "Index V-3", which consisted of 96 test results for rainbow trout survival in reconstituted laboratory water spiked with Cd, Cu, Zn [32].

Calculations for the 4 models were performed as follows: For the AIST model, free ion activities of metals and cations were calculated using WHAM VII [15,33]. The remainder of the calculation was performed in Excel, with fractional coverage of metal on the biotic ligand computed from the WHAM-calculated free ion activities and AIST-chosen log K_M values. Biological responses (e.g., mortality, growth reduction) were then determined as a function of the fractional coverage of metal(s) using the AIST response functions (Eqn. 1; Supporting Information File SI-1, Tables S1 and S2).

A similar procedure was followed for evaluation of the USGS model. Free ion activities of metals and cations were calculated using WHAM VII. The fractional coverage of metal on the

biotic ligand, the TOX function, and toxicity were computed in Excel using the WHAM-calculated free ion activities and the USGS-chosen log K_M values, potency factors (α_i), and response-function parameters (β_1, β_2, β_3) (Eqns. 2-4; Supporting Information File SI-1, Tables S3 and S4).

For evaluation of the HDR model, free ion activities were calculated using the TICKET model [34] with the WHAM V database. Accumulation of metals and cations on each of the biotic ligand sites (in nmol $g(wet)^{-1}$) were computed in Excel using the WHAM-calculated free ion activities and HDR-chosen log K_M values (Supporting Information File SI-1, Table S5). Toxic responses at each biotic ligand were determined using the HDR response-function parameters (a_i, b_i) (Eqns. 5, 6; Supporting Information File SI-1, Table S6).

Finally, for evaluation of the CEH model, free ion activities and concentrations of metals and cations on HA were computed using WHAM VI [27]. Because the concentration of specifically-bound protons on HA was not included in the model output of the commercially-available version of WHAM VI, concentrations of specifically-bound protons were estimated as the total number of proton binding sites minus the surface charge (in equivalents per gram) minus 2 times the summation of specifically-bound metals and cations. In this calculation, metals and divalent cations are assumed to primarily occupy bidentate and tridentate binding sites on the HA. Toxicity was then computed in Excel using the WHAM-calculated concentrations on HA and the CEH-determined potency factors (α_i) and response-function parameters ($F_{TOX\text{-}LT}$, $F_{TOX\text{-}UT}$) (Eqns. 7, 8; Supporting Information File SI-1, Tables S7 and S8).

Results for the 4 models were then plotted in comparable formats with mortality (or growth reduction) on the vertical axis and fractional coverage on the biotic ligand (θ_M) for the AIST model, *TOX* for the USGS model, and F_{TOX} for the CEH model on the horizontal axis. This

allowed metals for a large number of single-metal and metal-mixture exposure tests to be plotted and visually compared in a concise and convenient format. Unfortunately, this graphical format is not directly applicable to the HDR model because (i) toxicity response functions for individual metals were not considered to have common logit slopes, and (ii) metal-mixture toxicity was described by independent joint action using a response-additive approach. However, a reasonable comparison was provided by converting the HDR model-predicted responses (R) to equivalent TOX (TOX_{equiv}) values by rearranging Equation 5:

$$TOX_{equiv} = e^{-\left(\frac{2.303\,a}{b}\right)} \cdot \left[\frac{R}{1-R}\right]^{\left(\frac{2.303}{b}\right)} \quad (9)$$

In this equation, R is the fractional mortality (or growth reduction) for single-metal or metal-mixture exposures, a and b are the log-logit intercept and slope, and 2.303 is included because HDR used a mixed ln-log function in Equation 5 to describe toxicity. For subsequent calculations, TOX_{equiv} values were computed using response parameters (a, b) for Zn. This effectively normalized results for other metals to the toxicity of Zn. The TOX_{equiv} function is most appropriate for mortality (or growth reduction) responses near 50%. For responses near 0% or 100%, the TOX_{equiv} function is only a crude approximation for metals with log-logit response slopes that are different than Zn. Graphical comparisons of model coefficients and of model results for individual metal-mixture exposure tests were also prepared and analyzed.

RESULTS AND DISCUSSION

Observed mortality for *D. magna* and the corresponding AIST, USGS, HDR and CEH model-calibrated response curves are shown for single-metal (Figure 1A,C,E,G) and metal-mixture exposures (Figure 1B,D,F,H) following the graphical formats described above. Similar

comparisons for rainbow trout mortality are presented in Figure 2. In most cases, model-calibrated response curves described the central tendency of observed mortalities for both single-metal and metal-mixture exposures. However, scatter in the observed data around the model-calibrated response curves varied from model to model, with the USGS *D. magna* results (Figure 1C,D) and the AIST rainbow trout results (Figure 2A,B) showing the largest variations of observed mortality about the model-calibrated response curves. Observed mortality data was also relatively widely scattered around the HDR model-calibrated response curves, particularly for observations near 0% and 100% mortality (Figure 1E,F; Figure 2E,F). However, these differences in part can be attributed to the TOX_{equiv} approximation (Eqn. 9) in which all metals were assumed to have log-logit slopes similar to that of Zn. Finally, the observed mortality data appeared most closely aligned to the CEH model-calibrated response curves for both *D. magna* (Figure 1G,H) and rainbow trout (Figure 2G,H).

The graphical comparisons in Figures 1 and 2 provide an overview of how well the 4 models were calibrated to observed mortality for 2 of the larger calibration datasets. Additional model-model comparisons were made by examining model fits to individual-metal results. For example, USGS, HDR and CEH model results for *D. magna* mortality in a Cd-only toxicity test (Index 4, Cu-Cd #7-1, with 12.6 µg/L dissolved Cd) are presented in Figure 3A. The 3 models, which computed solution chemistry using 3 different versions of WHAM, predicted approximately 35% of total dissolved Cd was free Cd. However, the fractional coverage of Cd on the biotic ligand (or the HA surrogate in the CEH model) varied from 0.24% in the CEH model to 11% in the USGS model. Despite these large differences, the associated calibration of potency factors and/or response-function parameters resulted in similar predictions of mortality by the 3 models. This finding demonstrates the strong inter-relationship of log K_M values,

potency factors and response-function parameters, and underscores the latitude that exists in calibrating model parameters with toxicity datasets that include only measures of total dissolved metal concentrations, water chemistry (pH, major ions, alkalinity, DOC) and a select toxicological endpoint (e.g., mortality, growth reduction).

A similar example is given in Figure 3B for AIST, HDR, and CEH model results for *D. magna* mortality in a Cu-only toxicity test (Index 4, Cd-Cu #5-2, with 83.8 µg/L of dissolved Cu). In this case, model predictions of free Cu varied from 0.042% to 2.6% of the total dissolved Cu concentration. These differences were due to the version of WHAM that was used in the calculations and to assumptions for DOC composition and carbonate chemistry that were employed by the 3 modeling groups. Differences in model predictions for free Cu were reflected in differences for Cu accumulation on the biotic ligand (or the HA surrogate in the CEH model). Despite these large differences, calibration of potency factors and/or response-function parameters again resulted in similar predictions of mortality by the 3 models. This finding demonstrates that the WHAM calculation can also have a large effect on the final calibration of log K_M, potency factors, and response function parameters, but again the inter-relationships of the multiple calibration parameters allows for compensation of those differences to produce similar overall predictions among the different modeling approaches.

Based on results presented in Figure 3, it is difficult to evaluate calibration strategies in a simple step-by-step procedure. Rather, a more holistic view of the calibration process is needed (see comparison of model calibration parameters in Table 3). In all 4 models, log K_M values for the initial calibration were fixed based on previous studies and were not considered as adjustable parameters. Based on the remaining model parameters, the initial calibration of the AIST model appeared to be most constrained, with response parameters allowed to be adjusted only as a

379 function of the organism (Table 3). This was followed by the USGS model which allowed
380 potency factors (α_i) to be adjusted as a function of only the metal, and response parameters (β_1,
381 β_2, β_3) to be adjusted as a function of only the organism in its initial calibration. Additional
382 flexibility was considered in the initial calibration of the HDR model, which allowed toxicity
383 response parameters (a_i, b_i) to be adjusted as a function of both metal and organism. The CEH
384 model provided similar flexibility by allowing potency factors (α_i) to be adjusted as a function of
385 metal and organism, and by allowing small adjustments in response parameters ($F_{TOX\text{-}LT}$, $F_{TOX\text{-}UT}$)
386 as a function of only the organism.

387 The use of a more constrained or a more flexible calibration strategy had a significant
388 effect on the ability of the 4 models to describe mortality (or growth reduction) data (see Figures
389 1 and 2). To illustrate this point, USGS and CEH model results for growth reductions of *P.*
390 *subcapitata* at pH 6 are given in Figure 4. As shown in Figure 4A, the growth in single-metal
391 and metal-mixture exposures were poorly described by the initial calibration of the USGS model,
392 which was based on potency factors (α_i) that were determined from global fits to the MMME
393 calibration datasets. As shown by the USGS modeling group (Supporting Information File SI-2),
394 specification of a separate set of potency factors (α_i) for *P. subcapitata* at pH 6 produced a much
395 closer correspondence of the model-calculated response curve and observed growth reductions
396 (Figure 4B). By comparison, the CEH model still appears to provide a better description of the
397 observed growth reductions (Figure 4C). The reason is in part due to the inclusion of
398 specifically-bound protons in the CEH toxicity evaluation. For example, in the calculated
399 growth reductions for *P. subcapitata* at pH 6, proton toxicity accounted for 1.7±0.06 of the
400 computed F_{TOX} value (vertical gray bar in Figure 4C). This served to compress the effects of
401 metals to the right of the vertical gray bar. Subtracting the proton contribution from F_{TOX} and re-

plotting the model response curve and the observed growth reduction provided a different picture of the variability that may be associated with metal accumulation in the organism (Figure 4D). Therefore, excluding the extra factor of proton toxicity, which was included in the WHAM-F_{TOX} calibration, would likely result in some added variability of observed responses around the model-calculated response curve.

Next, the 4 models were further evaluated by comparing their final selection of model parameters. Log K_M comparisons for Cd, Pb, and Zn were based on AIST values for rainbow trout, USGS values that were previously determined from cutthroat trout and rainbow trout data, HDR values for the Zn-specific biotic ligand, and average log K_M values for the distribution of binding sites in the CEH model (Supporting Information File SI-1, Tables S1, S3, S5 and S7). From a chemical perspective, the log K_M values for the HDR and CEH models followed an expected increase in metal binding based on affinities of metals to oxygen donor groups on organic acids (Cd < Zn < Pb; Figure 5A). In contrast, the AIST and USGS models had larger binding constants for Cd that were similar to previously-reported log K_M values for Cd (e.g., [35,36,37]). This stronger binding of Cd to biological ligands was attributed to active Ca transport and ionic mimicry in fish gills [35]. Another possible explanation for larger Cd binding constants may be that Cd is binding to sulfur (and not oxygen) donor groups in the organism.

Toxicity parameters in the 4 models were also compared by combining potency factors (α_i) and response function parameters into a single measure of the lethal accumulation at 50% mortality (LA50) (see Supporting Information File SI-1, Tables S2, S4, S6, and S8). For the comparison, LA50 values were expressed in terms of percent coverage on the biotic ligand or on the surrogate HA binding sites for the WHAM-F_{TOX} model. The resulting LA50 values for rainbow trout exposed to Cd, Zn, and Pb ranged from 2 to 3% of the binding sites in the AIST

and USGS models (Figure 5B). In contrast, the LA50 values for the HDR and CEH models varied more (0.01% for Cd, 7% for Zn, and 2.9% for Pb for the HDR model; 0.05% for Cd, 9.8% for Zn, and 22% for Pb for the CEH model). For comparison, experimentally-derived LA50 values for Cd, Zn and Pb in rainbow trout studies have ranged from 10% to 64% coverage of strong binding sites on the gill [36,37,38,39]. These values represent the higher end of the model-calculated LA50 values and are not supportive of the very low LA50 values for Cd in the HDR model. However, there are some questions regarding the appropriateness of comparing experimentally-derived LA50 values (which are based on estimates for the density of strong binding sites that have been reported to vary as a function of water chemistry and the specific metal being examined) and model-calculated LA50 values (which are generally based on a binding site density that is considered to be constant across all water chemistries and metals). An alternative interpretation of experimentally-derived LA50 values is provided by considering the relative values of measured accumulations on a nmole per gram of fish gill basis. For example, 24-h LA50 measurements for rainbow trout have been reported as 1.1 nmole g^{-1} (ww) for Cd and 32.8 nmole g^{-1} (ww) for Pb [36]. This represents a difference of a factor of 30 in the Cd and Pb accumulations on the gill that would elicit a 50% mortality response and is consistent with the lower LA50 values for Cd that were computed by the HDR and CEH models. However, a more appropriate comparison for the CEH model-calculated LA50 values would be body-burden measurements for the various metals.

A final check on model calibration was performed by examining individual series of mixture toxicity test results for *D. magna* mortality (Index 4). For the Cu-Zn test series #5-4, increases in mortality were observed for *D. magna* exposed to a constant Cu concentration of 100 µg/L and increasing Zn concentrations (Figure 6A). This behavior is similar to an additive response curve

448 (see Figure 1 in Meyer et al. [1]). Model-calculated response curves for the AIST, HDR and
449 CEH models were consistent with the observed trend, with the HDR model corresponding most
450 closely to the observed data. Differences in the AIST, HDR and CEH model-calculated curves
451 for the Cu-Zn mixture can be attributed to the calibration of the models and not to the differences
452 in their formulations.

453 The Cd-Cu test series #20-3 showed very different behavior (Figure 6B), with observed
454 mortalities for *D. magna* exposed to a constant Cd concentration of 19.5 µg/L and increasing Cu
455 concentrations following a less-than-additive (Case 2) response curve (see Figure 1 in Meyer et
456 al. [1]). The AIST model (which is based on concentration addition) did not predict the observed
457 decrease in mortality as Cu concentrations increased. However, both the HDR model (which is
458 based on independent joint action) and the CEH model (which is based on a F_{TOX}-additive
459 approach) predicted a decrease and then an increase in *D. magna* mortality as the Cu
460 concentration increased. Additional adjustments in the both HDR and CEH model calibrations
461 were required to fit the observed response. For the HDR model, this consisted of increasing the
462 log K_M value for Cu binding to the Cd-biotic ligand sites by 4 log units. For the CEH model,
463 global values of $F_{TOX\text{-}LT}$ and $F_{TOX\text{-}UT}$ that were reported for *D. magna* were adjusted by
464 optimizing the $F_{TOX\text{-}LT}$ and $F_{TOX\text{-}UT}$ to the test series #20 data. Although the reported adjustments
465 in $F_{TOX\text{-}LT}$ and $F_{TOX\text{-}UT}$ were not large (global fit: 1.88 and 2.95; test series #20 fit: 2.61 and 3.18),
466 model responses to the single metal and metal-mixture exposure tests were sensitive to the
467 changes (see Supporting Information File SI-1, Figures S25 and S26). Therefore, questions still
468 remain about how to formulate and calibrate models to reproduce observed responses exhibiting
469 less-than-additive (Case 2) behavior.

In addition to model calibration evaluations, the calibrated models were used in a blind prediction of mortality for 2 validation studies (Index V-1; Index V-3). Comparisons of observed mortality and model-calculated response curves for the *D. magna* validation study (Index V-1) are presented in Figure 7. Model-data comparisons for single-metal exposures (Figure 7A,C,E,G) were comparable to the *D. magna* calibration results presented in Figure 1. Model-calculated response curves for metal mixtures tended to over-predict mortality by factors of 1-2 on the θ_M scale for the AIST model (Figure 7B), factors of 2-5 on the *TOX* scale for the USGS model (Figure 7D), approximately a factor of 2 on the TOX_{equiv} scale for the HDR model (Figure 7F), and approximately a factor of 2 on the F_{TOX} scale for the CEH model (Figure 7H).

Similar results for the rainbow trout validation study (Index V-3) are shown in Supporting Information File SI-1 (Figures S8, S15, S22 and S31). Model-data comparisons tended to show more variability for single-metal exposures. For mixtures, model-calculated response curves tended to over-predict trout mortality by a factor of 3 to 4 on the *TOX* scale for the USGS model. Model-calculated response curves for the HDR and CEH models were in closer agreement to observed mortality. The AIST model was not considered in the rainbow trout validation test because a rainbow trout log K_M value for Cu was not provided for the AIST model.

The overall results of the present study highlighted similarities and differences in 4 models that were developed to describe the effects of single-metal and metal-mixture exposures on biological response (e.g., mortality, growth reduction). The 4 models were calibrated to individual datasets that contained metal-exposure concentrations, water chemistry and biological response data. Because measurements of metal accumulation on a representative biological site were not available, independent calibration of log K_M, potency factors and response-function parameters were not possible. Despite these limitations, calibration of models to single-metal

493 exposure data often provided a reasonable basis for predicting metal-mixture toxicity. This was
494 particularly true for metal-mixtures exhibiting additive (or near additive) behavior. The ability
495 of the models to reproduce less-than-additive behavior posed a greater challenge. This less-than-
496 additive toxicity was predicted by the HDR model (which considered independent joint action,
497 but required substantial adjustment of some log K_M values to describe the observed behavior)
498 and by the CEH model (which considered F_{TOX} addition, but required adjustment of F_{TOX}
499 parameters to individual datasets).

500 These findings indicate that competitive interactions among metals add a level of complexity
501 to toxicity evaluations that will in all likelihood only be appreciated through continued model
502 development. The application of more complex geochemical models (with multiple biotic ligand
503 sites or distributions of log K_M binding sites) may be needed for this purpose. This has led to
504 revisions in the 4 modeling approaches, particularly for the AIST and USGS models that were
505 considered in the present study. Revised versions of the AIST, USGS, HDR and CEH models
506 are described in various papers in this issue (see [8,9,10,11,13]). Model calibration remains a
507 key issue for the various modeling approaches. Therefore, a further evaluation of specific model
508 assumptions and calibration strategies that were used by the 4 modeling groups is considered in
509 the following paper [40].

510

511 **SUPPORTING INFORMATION**

512 File SI-1: Modeling parameters and additional analyses (Tables S1 to S8, and Fig. S1 to S31).

513 File SI-2: USGS model description (2012 version).

514 File SI-3: Description of "Index 7" and "Index 8" data sets.

515

Acknowledgements

Funding for this work was provided by the Copper Alliance, International Zinc Association (IZA), Nickel Producers Environmental Research Association (NiPERA) and Rio Tinto. The authors are especially grateful to E. Van Genderen, W. Adams, R. Dwyer, E. Garman and J. Gorsuch for sharing their insights and for coordinating the MMME project study.

REFERENCES

1. Meyer JS, Farley KJ, Garman ER. 2014. Metal Mixture Modeling Evaluation project: 1. Background. *Environ Toxicol Chem* (this issue).
2. Norwood WP, Borgmann U, Dixon DG, Wallace A. 2003. Effects of metal mixtures on aquatic biota: A review of observations and methods. *Hum Ecol Risk Assess* 9:795-811.
3. Vijver MG, Elliott EG, Peijnenburg WJGM, De Snoo GR. 2011. Response predictions for organisms water-exposed to metal mixtures: A meta-analysis. *Environ Toxicol Chem* 30:1482-1487.
4. Di Toro DM, Allen HE, Bergman HL, Meyer JS, Paquin PR, Santore RC. 2001. Biotic ligand model of the acute toxicity of metals. 1. Technical basis. *Environ Toxicol Chem* 20:2383-2396.
5. Stockdale A, Tipping E, Lofts S, Ormerod SJ, Clements WH, Blust R. 2010. Toxicity of proton-metal mixtures in the field: Linking stream macroinvertebrate species diversity to chemical speciation and bioavailability. *Aquat Toxicol* 100:112-119.
6. 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.
7. Meyer JS, Clearwater SJ, Doser TA, Rogaczewski MJ, Hansen JA. 2007. *Effects of Water Chemistry on the Bioavailability and Toxicity of Waterborne Cadmium, Copper, Nickel, Lead, and Zinc to Freshwater Organisms.* SETAC Press, Pensacola, Florida, USA.
8. Iwasaki Y, Naito W, Kamo M. 2014. Testing an application of the biotic ligand model to predict effects of metal mixtures on rainbow trout survival. *Environ Toxicol Chem* (this issue).

9. Balistrieri LS, Mebane CA. 2014. Predicting the toxicity of metal mixtures. *Sci Total Environ* 466-467:788–799.
10. Balistrieri LS, Mebane CA. 2014. Predicting the toxicity of metal mixtures to benthic macroinvertebrate and zooplankton assemblages in streams and lakes. *Environ Toxicol Chem* (this issue).
11. Santore RC, Ryan AC. 2014. Technical basis of a multi-metal multi-biotic ligand model for assessing toxicity of metal mixtures. *Environ Toxicol Chem* (this issue).
12. Van Genderen E, Adams W, Dwyer R, Garman E, Gorsuch J. 2014. Modeling and interpreting biological effects of mixtures in the environment: Introduction to the Metal Mixture Modeling Evaluation project. *Environ Toxicol Chem* (this issue).
13. Tipping E, Lofts S. 2014. Testing WHAM-F_{TOX} with laboratory toxicity data for mixtures of metals (Cu, Zn, Cd, Ag, Pb). *Environ Toxicol Chem* (this issue).
14. Kamo M, Nagai T. 2008. An application of the biotic ligand model to predict the toxic effects of metal mixtures. *Environ Toxicol Chem* 27:1479-1487.
15. Lofts S. 2012. *User's guide to WHAM7*. NERC Centre for Ecology and Hydrology, Lancaster, UK.
16. Hollis L, McGeer JC, McDonald DG, Wood CM. 2000. Effects of long term sublethal Cd exposure in rainbow trout during soft water exposure: Implications for biotic ligand modelling. *Aquat Toxicol* 51:93-105.
17. Santore RC, Mathew 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* 133:271-285.

566 18. De Schamphelaere KAC, Janssen CR. 2002. A biotic ligand model predicting acute copper
567 toxicity for *Daphnia magna*: The effects of calcium, magnesium, sodium, potassium, and pH.
568 *Environ Sci Technol* 36:48-54.
569 19. Heijerick DG, De Schamphelaere KAC, Janssen CR. 2002. Predicting acute zinc toxicity for
570 *Daphnia magna* as a function of key water chemistry characteristics: Development and
571 validation of a biotic ligand model. *Environ Toxicol Chem* 21:1309-1315.
572 20. Clifford M, McGeer JC. 2010. Development of a biotic ligand model to predict the acute
573 toxicity of cadmium to *Daphnia pulex*. *Aquati Toxicol* 98:1-7.
574 21. Clifford M, McGeer JC. 2009. Development of a biotic ligand model for the acute toxicity of
575 zinc to *Daphnia pulex* in soft waters. *Aquat Toxicol* 91:26-32.
576 22. Tipping E. 2005. Modelling Al competition for heavy metal binding by dissolved organic
577 matter in soil and surface waters of acid and neutral pH. *Geoderma* 127:293-304.
578 23. Lofts S, Tipping E. 2011. Assessing WHAM/Model VII against field measurements of free
579 metal ion concentrations: model performance and the role of uncertainty in parameters and
580 inputs. *Environ Chem* 8:501-516.
581 24. HydroQual. 2007. The biotic ligand model Windows interface, Version 2.2.3: User's guide
582 and reference manual. HydroQual, Inc., Mahwah, NJ, USA.
583 25. Tipping E. 1994. WHAM - A chemical equilibrium model and computer code for waters,
584 sediments, and soils incorporating a discrete site/electrostatic model of ion-binding by humic
585 substances. *Comput Geosci* 20:973-1023.
586 26. Tipping E, Lofts S. 2013. Metal mixture toxicity to aquatic biota in laboratory experiments;
587 Application of the WHAM-FTOX model. *Aquat Toxicol* 142-143:114–122.

27. Tipping E. 1998. Humic Ion-Binding Model VI: An improved description of the interactions of protons and metal ions with humic substances. *Aquati Geochem* 4:3-48.

28. Stockdale A, Tipping E, Fjellheim A, Garmo ØA, Hildrew AG, Lofts S, Monteith DT, Ormerod SJ, Shilland EM. 2014. Recovery of macroinvertebrate species richness in acidified upland waters assessed with a field toxicity model. *Ecol Indic* 37:341-350.

29. Lofts S, Tipping E, Hamilton-Taylor J. 2008. The chemical speciation of Fe(III) in freshwaters. *Aquat Geochem* 14:337-358.

30. Meyer JS, Ranville JF, Pontasch M, Gorsuch JW, Adams WJ. 2014. Acute toxicity of binary and ternary mixtures of Cd, Cu, and Zn to *Daphnia magna*. *Environ Toxicol Chem* (this issue).

31. Mebane CA, Dillon FS, Hennessy DP. 2012. Acute toxicity of cadmium, lead, zinc, and their mixtures to stream-resident fish and invertebrates. *Environ Toxicol Chem* 31:1334-1348.

32. Naddy R, Stubblefield WA. 2014. The interactive toxicity of cadmium, copper, and zinc to *Ceriodaphnia dubia* and rainbow trout (*Oncorhynchus mykiss*). *Environ Toxicol Chem* (this issue).

33. 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.

34. Farley KJ, Rader KJ, Miller BE. 2008. Tableau input coupled kinetics equilibrium transport (TICKET) model. *Environ Sci Technol* 42:838-844.

35. Playle RC, Dixon DG, Burnison K. 1993. Copper and cadmium binding to fish gills: Estimates of metal-gill stability constants and modelling of metal accumulation. *Can J Fish Aquat Sci* 50:2678-2686.

610 36. Birceanu O, Chowdhury MJ, Gillis PL, McGeer JC, Wood CM, Wilkie MP. 2008. Modes of
611 metal toxicity and impaired branchial ionoregulation in rainbow trout exposed to mixtures of
612 Pb and Cd in soft water. *Aquat Toxicol* 89:222-231.
613 37. Niyogi S, Kent R, Wood CM. 2008. Effects of water chemistry variables on gill binding and
614 acute toxicity of cadmium in rainbow trout (*Oncorhynchus mykiss*): A biotic ligand model
615 (BLM) approach. *Comp Biochem Physiol C Toxicol Pharmacol* 148:305-314.
616 38. Niyogi S, Couture P, Pyle GG, McDonald DG, Wood CM. 2004. Acute cadmium biotic
617 ligand model characteristics of laboratory-reared and wild yellow perch (*Perca flavescens*)
618 relative to rainbow trout (*Oncorhynchus mykiss*). *Can J Fish Aquat Sci* 61:942-953.
619 39. Todd AS, Brinkman S, Wolf RE, Lamothe PJ, Smith KS, Ranville JF. 2009. An enriched
620 stable-isotope approach to determine the gill-Zn binding properties of juvenile rainbow trout
621 (*Oncorhynchus mykiss*) during acute zinc exposures in hard and soft waters. *Environ Toxicol*
622 *Chem* 28:1233-1243.
623 40. Farley KJ, Meyer JS. 2014. Metal Mixture Modeling Evaluation project: 3. Lessons learned
624 and steps forward. *Environ Toxicol Chem* (this issue).
625 41. Ingersoll CG, MacDonald DD, Besser JM, Brumbaugh WG, Ivey CD, Kemble NE, Kunz
626 JL, May TM, Wang N, Smorong DE. 2008. Sediment Chemistry, Toxicity, and
627 Bioaccumulation Data Report for the US Environmental Protection Agency – Department of
628 the Interior Sampling of Metal-Contaminated Sediment in the Tri-State Mining District in
629 Missouri, Oklahoma, and Kansas. Columbia Environmental Research Center, United States
630 Geological Survey. Administrative Report CERC-8335-FY07-20-12. Columbia, MO, USA.
631 42. Le TTY, Vijver MG, Hendriks AJ, Peijnenburg WJGM. 2013. Modeling toxicity of binary
632 metal mixtures (Cu2+-Ag+, Cu2+-Zn2+) to lettuce, Lactuca sativa, with the biotic ligand

633 model. Environmental Toxicology and Chemistry 32:137-143.
634 43. Carbonaro RF, Atalay YB, Di Toro DM. 2011. Linear free energy relationships for metal–
635 ligand complexation: Bidentate binding to negatively-charged oxygen donor atoms. *Geochim*
636 *Cosmochim Acta* 75:2499-2511.
637

Table 1. Datasets used in the Metal Mixture Modeling Evaluation project[a]

Index	Species	Metal mixture / water type	Endpoint	Number of exposures[b]	Source	Datasets used for model calibration[c]			
						AIST	USGS	HDR	CEH
1	*Hyalella azteca* *Lampsilis siliquoidea*	Cd-Cu-Ni-Pb-Zn / Porewater	28-d survival	2/60 2/38	[41]		✓	✓	✓
4	*Daphnia magna*	Cd-Cu, Cu-Zn / Lab	48-h survival	387/174	[30]	✓	✓	✓	✓
6	*Oncorhynchus mykiss*	Cd-Pb-Zn / Field	96-h survival	298/71	[31]	✓	✓	✓	✓
7	*Pseudokirchneriella subcapitata*	Field mixture / Field	72-h growth	7/28	Present Study[d]	✓	✓		✓
8	*P. subcapitata*	Cd-Cu-Ni-Zn / Field	72-h growth	102/12	Present Study[d]	✓	✓		✓
9	*Lactuca sativa*	Ag-Cu, Cu-Zn / Hydroponic	4-d root growth	36/202	[42]				✓
V-1	*D. magna*	Cd-Zn / Lab	48-h survival	132/177	[30]				
V-2	*D. magna*	Cd-Cu-Zn Lab	48-h survival	3/12	[30]				
V-3	*O. mykiss*	Cd-Cu-Zn Lab	96-h survival	72/24	[32]				

[a] See [12] for detailed descriptions.
[b] Single-metal or reference exposures / mixture exposures
[c] AIST = National Institute of Advanced Industrial Science and Technology, Japan; USGS = U.S. Geological Survey, USA; HDR = HDR|HydroQual, Inc., USA; CEH = Centre for Ecology and Hydrology, UK
[d] See Supplemental Information, File SI-3

638

Table 2. Summary of formulations for the AIST, USGS, HDR, and CEH metal-mixture-toxicity models[a]

	AIST	USGS	HDR	CEH
Solution chemistry	WHAM VII	WHAM VII	WHAM V	WHAM VI
Metal/cation binding to organisms	Competitive binding of metals/cations to a single BL site	Competitive binding of metals/cations to a single BL site	Competitive binding of metals/cations to multiple BL sites	Non-specific accumulation of metals / cations at a distribution of binding sites
Toxicity	Function of fractional coverage of metal on BL	Function of potency and coverage of each metal on BL ($TOX_i = \alpha_i\ \theta_i$)[b]	Function of potency and concentration of each metal on its toxicologically-relevant BL	Function of potency and concentration of protons and each metal on WHAM humic acid, assumed proportional to their binding on or in the organism ($F_{TOX} = \alpha_i\ \nu_i$)[c]
Toxic response	2-parameter logit (or linear) response function	3-parameter logit response function	2-parameter logit response function	2-parameter linear-threshold response function
Mixture response	Concentration additive	*TOX* additive[b]	Independent action	F_{TOX} additive[c]

[a] AIST = National Institute of Advanced Industrial Science and Technology, Japan; USGS = U.S. Geological Survey, USA; HDR = HDR|HydroQual, Inc., USA; CEH = Centre for Ecology and Hydrology, UK; BL = biotic ligand; WHAM V, VI, and VII = versions of Windermere Humic Aqueous Model.

[b] *TOX* = toxicity-response function in USGS model; α_i = potency factor for metal *i*; θ_i = proportion of BL sites occupied by metal *i* (# of sites occupied / # total sites).

[c] F_{TOX} = toxicity-response function in CEH model; α_i = potency factor for proton or metal *i*; ν_i = concentration of BL sites occupied by protons or metal *i* (mmol g^{-1} humic acid).

639

Table 3. Summary of parameters used in the AIST, USGS, HDR, and CEH metal-mixture-toxicity models[a]

	AIST	USGS	HDR	CEH
Binding constants ($log\ K_M$)	f(metal, organism)[b]	f(metal)[c]	f(metal)[b]	f(metal)[c]
Metal potency factors (α_i)	n/a	f(metal)[d]	n/a	f(metal, organism)
Proton potency factor (α_H)	n/a	n/a	n/a	f(organism)
Response parameters (β_1, β_2, β_3, F_{TOX-LT}, F_{TOX-UT}, etc.)	f(organism)	f(organism)	f(metal, organism)	f(organism)

[a] AIST = National Institute of Advanced Industrial Science and Technology, Japan; USGS = U.S. Geological Survey, USA; HDR = HDR|HydroQual, Inc., USA; CEH = Centre for Ecology and Hydrology, UK.

[b] log K_M values for the AIST and HDR models were taken from previously-calibrated, single-metal biotic ligand models. Additional adjustments of log K_M values were made during their studies.

[c] log K_M values were held constant in the USGS and CEH models and were determined as follows:
USGS: from a re-evaluation of single-metal toxicity data for cutthroat and rainbow trout;
CEH: from previous calibration for WHAM VI (a version of the Windermere Humic Aqueous Model), using WHAM humic acid as a surrogate for non-specific accumulation of protons and metabolically-active metals by the organism.

[d] In the initial calibration of the USGS model, metal potency factors were considered to be a function of only the metal. A separate set of potency factors was required for the final calibration of the algal dataset.

640

641

642

643 **Figure Captions**

644 Figure 1. Model-data comparisons for the effects of single-metal and metal-mixture
645 exposures on mortality of *Daphnia magna* in laboratory water with spiked metals. Mortality
646 is shown as a function of fractional coverage of metal on the biotic ligand (θ_M) for the AIST
647 model (panels **A**, **B**); *TOX* for the USGS model (panels **C**, **D**); TOX_{equiv} for the HDR model
648 (panels **E**, **F**); and F_{TOX} for the CEH model (panels **G**, **H**). Observed responses (open
649 symbols) are compared to the model-calculated response curve for mortality (continuous
650 line). Dashed lines represent plus/minus a factor of 2 in the concentration at which a model-
651 calculated response occurs. See text for calibration procedures used in each model. Data
652 from Meyer et al. [30].

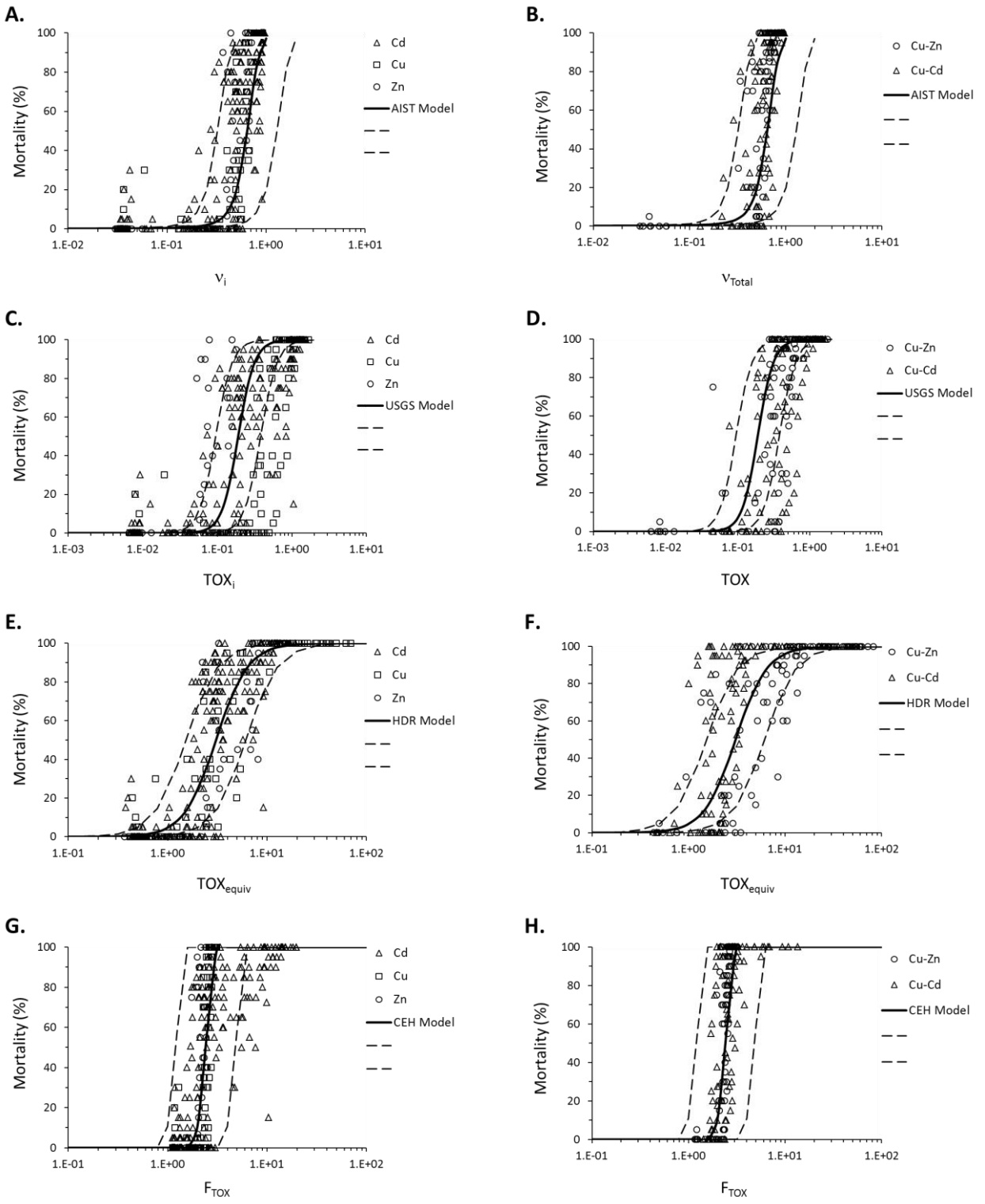
653 Figure 2. Model-data comparisons for the effects of single-metal and metal-mixture
654 exposures on mortality of rainbow trout (*Oncorhynchus mykiss*) in field-collected water with
655 spiked metals. Mortality is shown as a function of fractional coverage of metal on the biotic
656 ligand (θ_M) for the AIST model (panels **A**, **B**); *TOX* for the USGS model (panels **C**, **D**);
657 TOX_{equiv} for the HDR model (panels **E**, **F**); and F_{TOX} for the CEH model (panels **G**, **H**).
658 Observed responses (open symbols) are compared to the model-calculated response curve for
659 mortality (continuous line). Dashed lines represent plus/minus a factor of 2 in the
660 concentration at which a model-calculated response occurs. See text for calibration
661 procedures used in each model. Data from Mebane et al. [31].

662 Figure 3. Comparison of model-calculated responses for free metal as a percentage of the
663 total dissolved metal, percent accumulated metal on the biotic ligand (or on the WHAM
664 humic acid surrogate for generalized binding on or in organisms in the CEH model), and
665 percent mortality of *Daphnia magna* for: (**A**) 12.6 µg/L total dissolved Cd (from Index 4, Cu-

666 Cd #7-1); and (**B**) 83.8 µg/L total dissolved Cu (from Index 4, Cd-Cu #5-2). Data from
667 Meyer et al. [30]; indexes are described in Van Genderen et al. [12].
668 Figure 4. Model-data comparisons for the effects of single-metal and metal-mixture
669 exposures on growth reduction of *Pseudokirchneriella subcapitata* at pH 6.0 in field-
670 collected water with spiked metals. Growth reduction is shown for (**A**) USGS model with
671 model parameters from a global calibration to all datasets; (**B**) USGS model with model
672 parameters from a calibration to *P. subcapitata* pH 6.0 data; (**C**) CEH model; and (**D**) CEH
673 model with the baseline effect of H^+ removed from F_{TOX}. Observed responses for single-
674 metal (open symbols) and metal-mixture exposures (closed symbols) are compared to the
675 model-calculated response curve for mortality (continuous line). Dashed lines represent
676 plus/minus a factor of 2 in the concentration at which a model-calculated response occurs.
677 See text for calibration procedures used in each model
678 Figure 5. Comparison of (**A**) metal binding affinity to binding sites on or in the organism
679 (log K_M); and (**B**) lethal accumulations for 50% mortality (LA50) for rainbow trout
680 (*Oncorhynchus mykiss*) based on the AIST, USGS, HDR and CEH model calibrations.
681 Average log K_M values are given for the CEH model for illustrative purposes. Metals are
682 arranged according to their expected affinity to bind to oxygen donor groups [43].
683 Figure 6. AIST, HDR and CEH model-data comparisons for the effects of metals on
684 mortality of *Daphnia magna* in laboratory water with spiked metal concentrations. Mortality
685 is shown: (**A**) as a function of total dissolved Zn with fixed total dissolved Cu concentrations
686 (test series #5-4); and (**B**) as a function of total dissolved Cu with fixed total dissolved Cd
687 concentrations (test series #20-3). CEH model-calculated response curves are based on
688 $F_{TOX\text{-}LT}$ and $F_{TOX\text{-}UT}$ values that were optimized to the individual test series ($F_{TOX\text{-}LT}$ = 2.30,

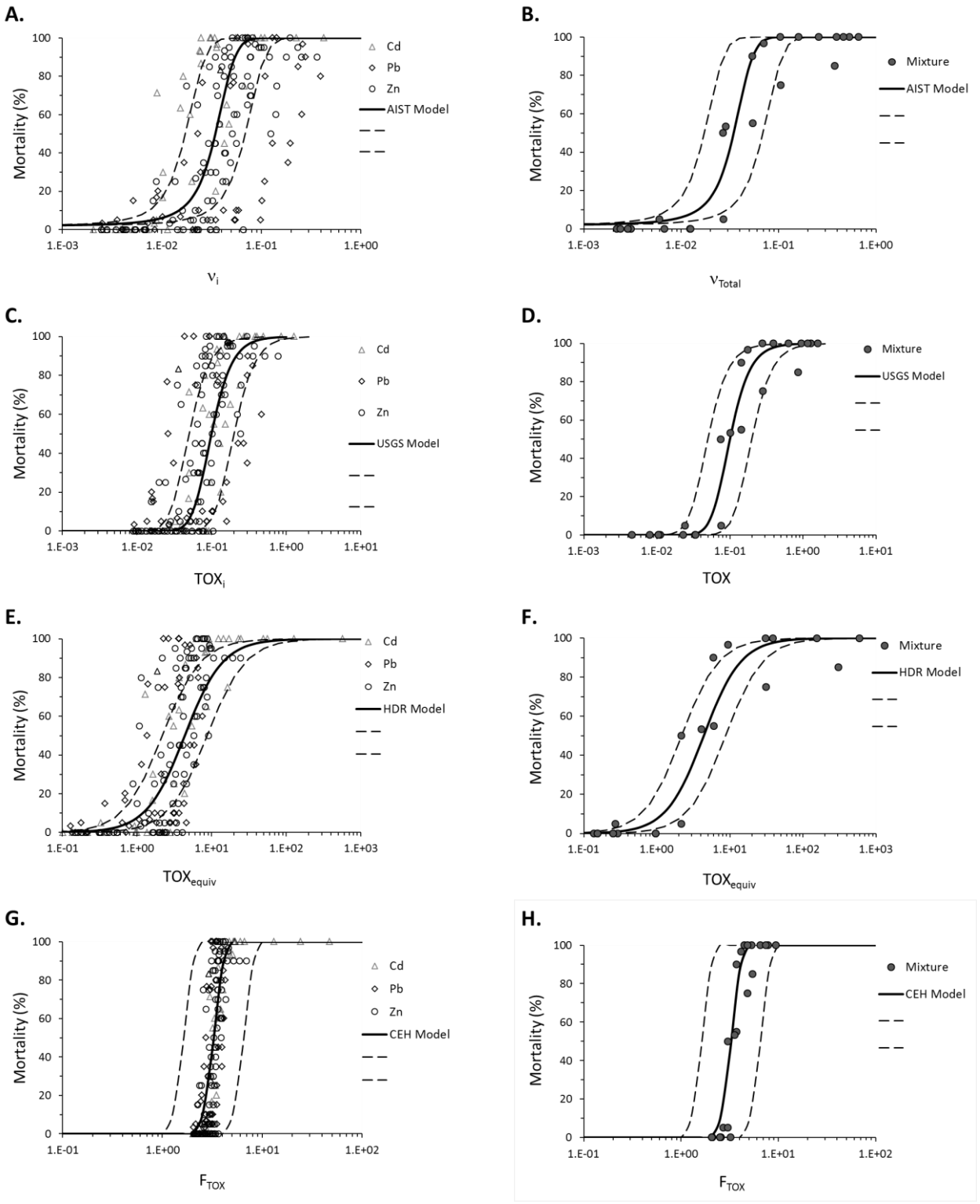
689 $F_{TOX\text{-}UT}$ = 2.71 for test series #5; $F_{TOX\text{-}LT}$ = 2.61, $F_{TOX\text{-}UT}$ = 3.18 for test series #20). Observed
690 responses for metal-mixture exposures (closed symbols) are compared to the model-
691 calculated response curve for mortality (dotted, short dashed and continuous lines). Data
692 from Meyer et al. [30]; indexes are described in Van Genderen et al. [12].
693 Figure 7. Model validation for the effect of single-metal and metal-mixture exposures on
694 mortality of *Daphnia magna* in laboratory water with spiked metals. Mortality is shown as a
695 function of fractional coverage of metal on the biotic ligand (θ_M) for the AIST model (panels
696 **A**, **B**); *TOX* for the USGS model (panels **C**, **D**); TOX_{equiv} for the HDR model (panels **E**, **F**);
697 and F_{TOX} for the CEH model (panels **G**, **H**). Observed responses (open symbols) are
698 compared to the model-calculated response curve for mortality (continuous line). Dashed
699 lines represent plus/minus a factor of 2 in the concentration at which a model-calculated
700 response occurs. Data from Meyer et al. [30].
701

702



703

704 Figure 1.



705

706 Figure 2.

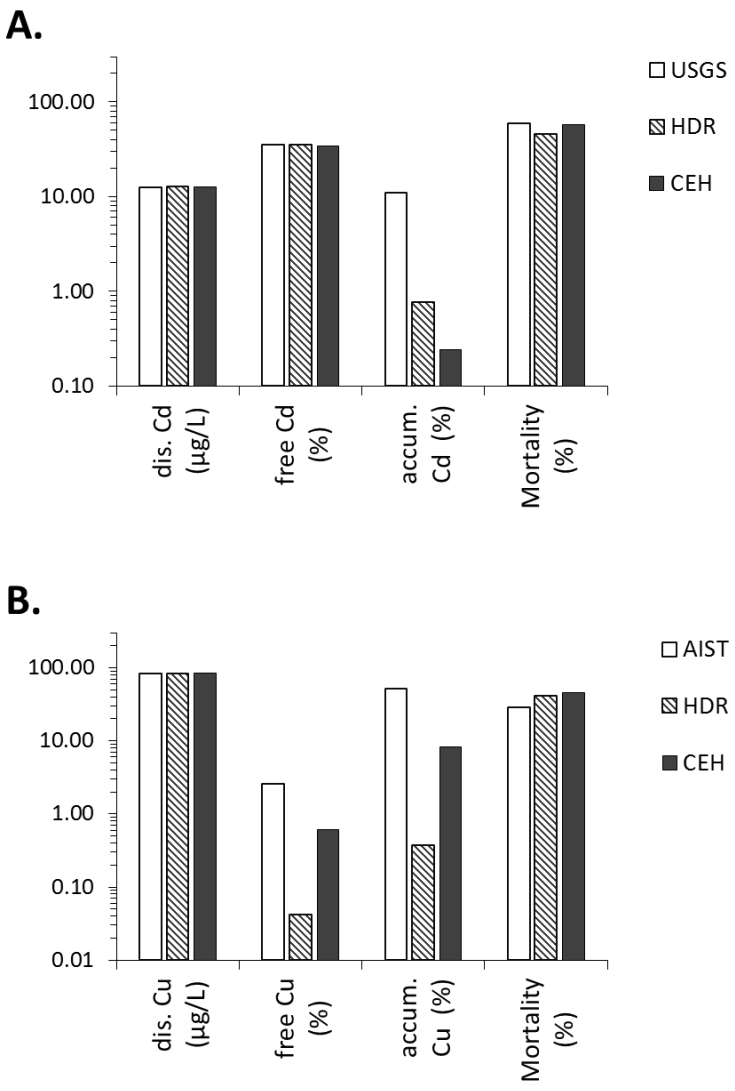

A.
100.00
10.00
1.00
0.10
dis. Cd (µg/L)
free Cd (%)
accum. Cd (%)
Mortality (%)
USGS
HDR
CEH
B.
100.00
10.00
1.00
0.10
0.01
dis. Cu (µg/L)
free Cu (%)
accum. Cu (%)
Mortality (%)
AIST
HDR
CEH

Figure 3.

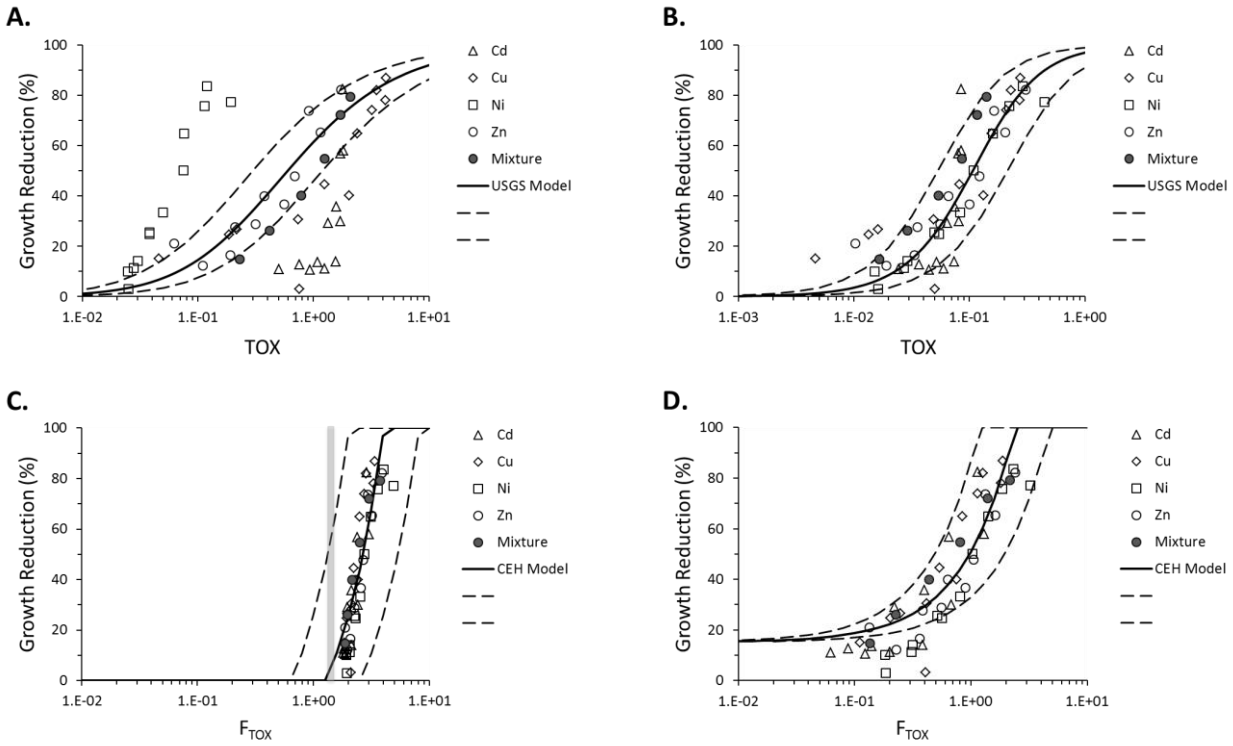

A.
Growth Reduction (%)
TOX
Cd
Cu
Ni
Zn
Mixture
USGS Model
B.
Growth Reduction (%)
TOX
Cd
Cu
Ni
Zn
Mixture
USGS Model
C.
Growth Reduction (%)
F_TOX
Cd
Cu
Ni
Zn
Mixture
CEH Model
D.
Growth Reduction (%)
F_TOX
Cd
Cu
Ni
Zn
Mixture
CEH Model

Figure 4.

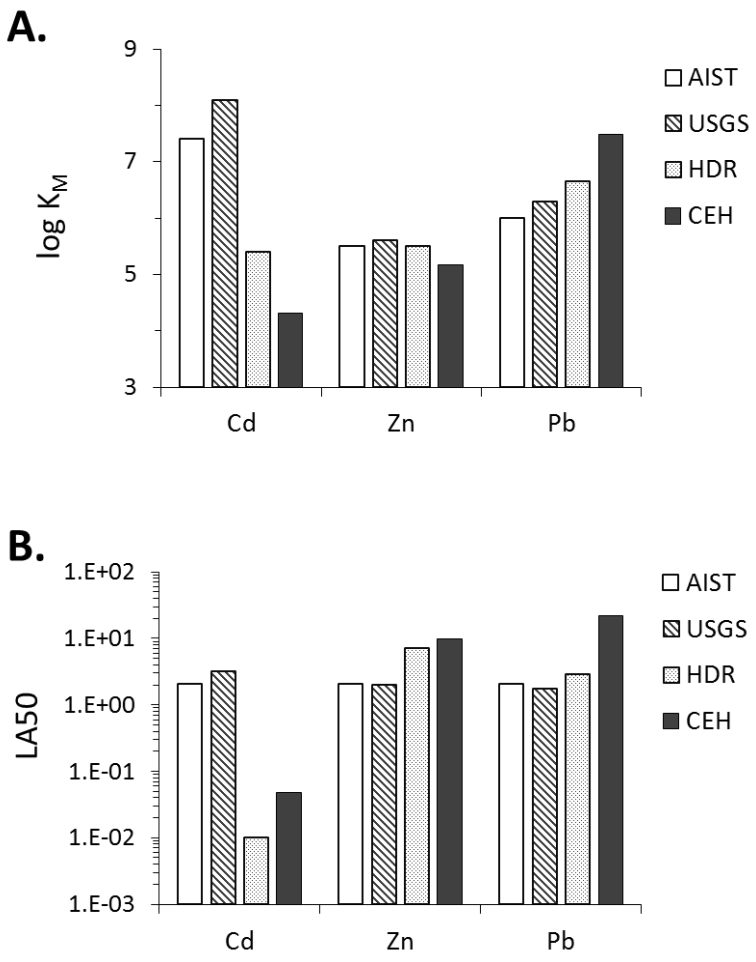

A.
9
7
5
3
log K_M
Cd
Zn
Pb
AIST
USGS
HDR
CEH
B.
1.E+02
1.E+01
1.E+00
1.E-01
1.E-02
1.E-03
LA50
Cd
Zn
Pb
AIST
USGS
HDR
CEH

Figure 5.

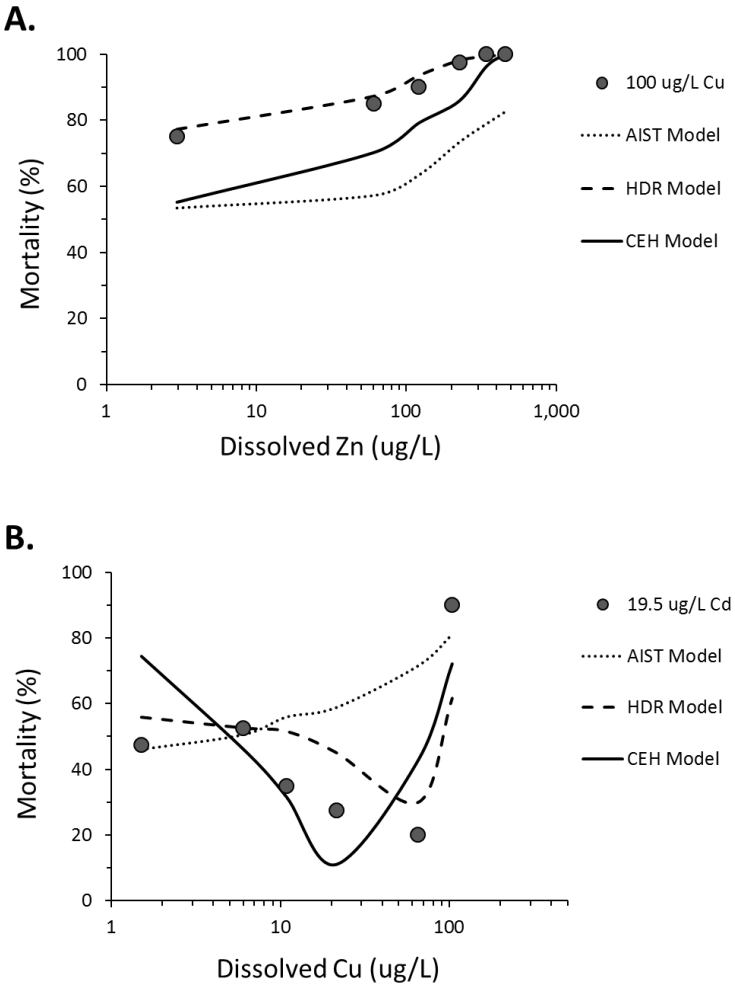

A.
100
80
60
40
20
0
Mortality (%)
1
10
100
1,000
Dissolved Zn (ug/L)
100 ug/L Cu
AIST Model
HDR Model
CEH Model
B.
100
80
60
40
20
0
Mortality (%)
1
10
100
Dissolved Cu (ug/L)
19.5 ug/L Cd
AIST Model
HDR Model
CEH Model

Figure 6.

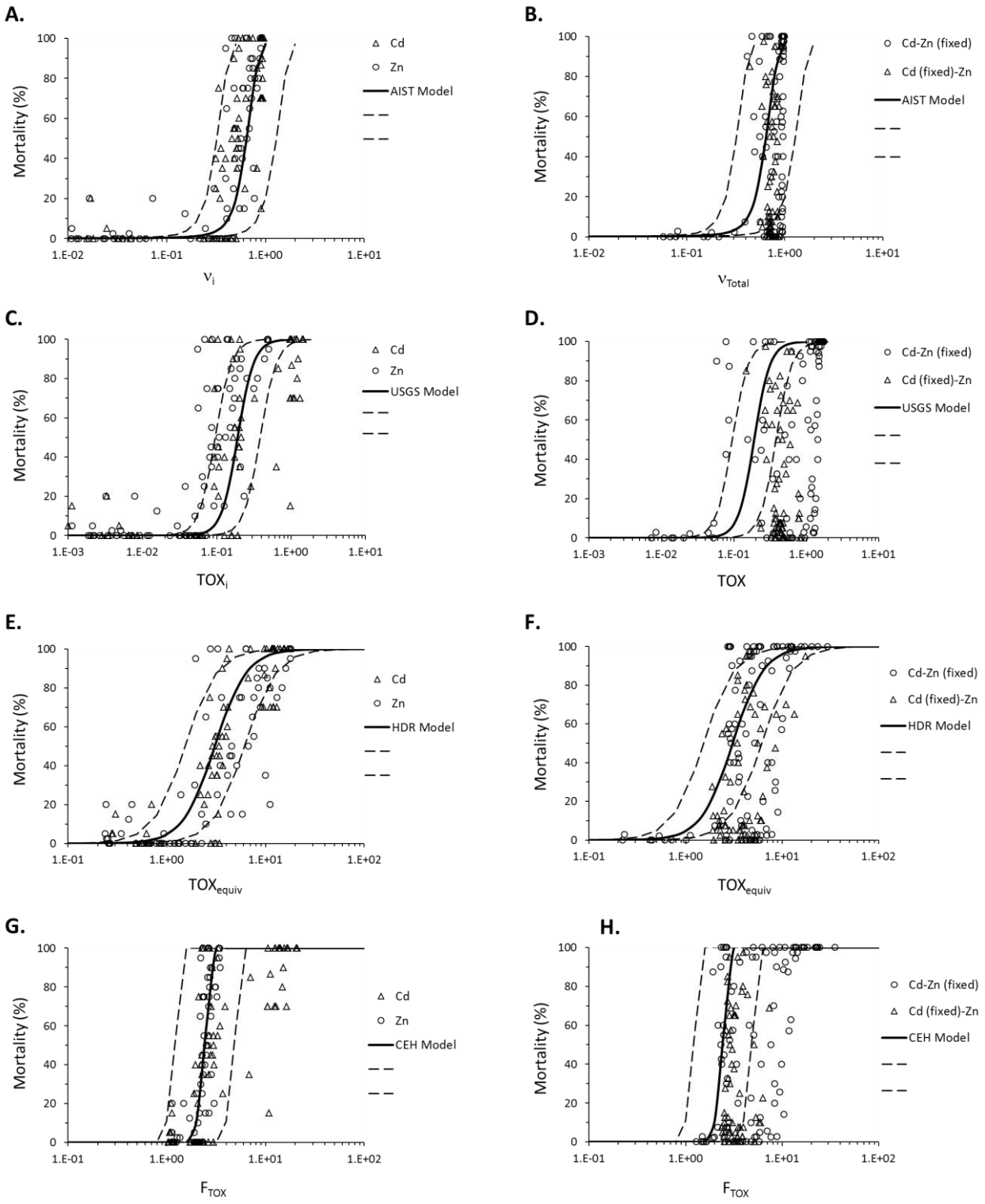

A.
Mortality (%)
Cd
Zn
AIST Model
ν_i
B.
Cd-Zn (fixed)
Cd (fixed)-Zn
AIST Model
ν_{Total}
C.
USGS Model
TOX_i
D.
USGS Model
TOX
E.
HDR Model
TOX_{equiv}
F.
HDR Model
TOX_{equiv}
G.
CEH Model
F_{TOX}
H.
CEH Model
F_{TOX}

Figure 7.