Article (refereed) - postprint

Van Regenmortel, Tina; Nys, Charlotte; Janssen, Colin R.; Lofts, Stephen; De Schamphelaere, Karel A.C. 2017. **Comparison of four methods for bioavailability-based risk assessment of mixtures of Cu, Zn, and Ni in freshwater**. *Environmental Toxicology and Chemistry*, 36 (8). 2123-2138. 10.1002/etc.3746

Environmental Toxicology **Environmental Toxicology and Chemistry**
DOI 10.1002/etc.3746

COMPARISON OF FOUR METHODS FOR BIOAVAILABILITY-BASED RISK ASSESSMENT OF MIXTURES OF Cu, Zn AND Ni IN FRESHWATER

TINA VAN REGENMORTEL,*[a] CHARLOTTE NYS,[a] COLIN R. JANSSEN,[a] STEPHEN LOFTS,[b] and KAREL A.C. DE SCHAMPHELAERE[a]

[a]Faculty of Bioscience Engineering, Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University (UGent), Ghent, Belgium

[b]Centre for Ecology and Hydrology, Bailrigg, Lancaster, United Kingdom

* Address correspondence to tina.vanregenmortel@ugent.be

This article includes online-only Supplemental Data.

Submitted 19 September 2016; Returned for Revision 4 November 2016; Accepted 20 January 2017

Abstract: Although chemicals risk assessment is still mainly conducted on a substance-by-substance basis, organisms in the environment are typically exposed to mixtures of substances. Risk assessment procedures should therefore be adapted to fit these situations. Here, we compared four mixture risk assessment methodologies for risk estimations of mixtures of copper, zinc and nickel. We showed that using the log-normal species sensitivity distribution (SSD) instead of the best-fit distribution and sampling species sensitivities independently for each metal instead of based on inter-species correlations in metal sensitivity had little impact on risk estimates. Across four different monitoring datasets, we estimated between 0% and 52% of the samples are at risk, but only between 0% and 15% of the samples are at risk due to the mixture of metals and not due to any single metal individually. When examining a natural baseline database, we found that 10% of the samples was estimated to be at risk due to single metals or their mixtures when using the most conservative method (Concentration Addition applied directly to the SSD, i.e. CA_{SSD}). However, the issue of metal mixture risk at geochemical baseline concentrations becomes relatively small (2% of samples) when using a theoretically more correct method (CA applied to individual Dose Response Curves, i.e. CA_{DRC}). Finally, across the 4 monitoring datasets, we show the following order of conservatism for our 4 methods (from most to least conservative, with ranges of median margin of safety, MoS, relative to CA_{SSD}): CA_{SSD}>CA_{DRC} (MOS=1.17-1.25)>IA_{DRC} (MoS=1.38-1.60)>IA_{SSD} (MoS=1.48-1.72) (IA = Independent Action). Therefore, we suggest that these four methods can be used in a general tiered scheme for the risk assessment of metal mixtures in a regulatory context. In this scheme, the CA_{SSD} method could serve as a first (conservative) tier to identify situations with likely no potential risk at all, regardless of the method used ($SumTU_{HC5} < 1$) and the IA_{SSD} method to identify situations of potential risk, also regardless of the method used ($msPAF_{IA,SSD} > 0.05$). The CA_{DRC} and IA_{DRC} methods could be used for site-specific assessment for situations that fall in-between ($SumTU_{HC5} > 1$ and $msPAF < 0.05$). This article is protected by copyright. All rights reserved

Keywords: Mixture toxicity, Risk assessment, Concentration addition, Independent action

INTRODUCTION

In the environment, organisms are usually simultaneously exposed to a multitude of substances including pesticides, pharmaceuticals and metals [1]. Although risk assessment is still mainly performed on a single-substance basis, various approaches for the risk assessment of mixtures of chemicals have been proposed [2-6]. These approaches are mainly based on two fundamental concepts that predict the joint toxicity of substances in a mixture to single species based on each substance's individual effects: concentration addition (CA) and independent action (IA). Although these two concepts have originally been theorized and mathematically developed to predict mixture toxicity to different single species [2], they have also been applied directly to species assemblages, both real [7,8] and mathematical [3] assemblages. In the latter case this is done by applying the two models directly to species sensitivity distribution (SSD) curves, as explained by De Zwart and Posthuma [3]. These authors estimated the risks of chemical cocktails on species assemblages expressed as a "multi-substance potentially affected fraction of species" (msPAF). Although these methods are gaining increased interest in the field of mixture risk assessment, they also show an important limitation as the predictions made by the models of CA and IA are theoretically only consistent when applied to single species, i.e. dose-response curves (DRCs), and not when applied to communities, i.e. SSDs [5]. More recently, Backhaus and Faust [4] as well as Gregorio et al. [5] developed theoretically consistent methods, which apply CA or IA first to different single species separately and then combine all single-species information to calculate risk estimates for a species assemblage. However, these two approaches also show a few limitations. The approach by Backhaus and Faust [4] only uses the 'base set' of toxicity data for a substance. This base set, i.e. ECx values for algae, crustaceans and fish, is the minimum set of data required by REACH for the calculation of a Predicted No

Effect Concentration (PNEC) [9]. Although their approach can be applied to a broader array of substances (i.e. 'data-poor' substances) and can be extended to a higher number of species, the method applies subjective assessment factors to calculate the risk quotient for a mixture. Gregorio and colleagues [5] only evaluated their method using sets of species toxicity values that were randomly generated from species sensitivity distributions of sets of hypothetical substances, and they also assumed a range of possible slope values of dose-response curves for these species, as they argued that implementing the method on existing data was not conceivable with the typical amount of data available for a substance.

Fortunately, limited data availability is not an issue for several metals. Indeed, the effects of the metals Cu, Zn and Ni on single species have been studied extensively, which makes these metals 'data-rich' substances. More recently, metal mixture toxicity is receiving increased attention and study topics are focused both on influences of metal mixtures on single species [10,11] as well as on communities [7,8,12,13]. Because it is infeasible to examine the effects of all possible mixtures of substances on every natural community experimentally, estimations of risks by means of models such as the ones described above, are essential.

In the present study, we therefore aimed to evaluate differences in (ternary) mixture risk estimates between four methods (Table 1) using actual chronic toxicity data for Cu, Zn and Ni. To this end, available toxicity datasets were first extended with recently published toxicity data as well as with the slope values of the dose-response curves (*cf. Material and Methods*). Then, for the first time, we applied these four methods to four existing environmental monitoring datasets to estimate metal mixture risks, i.e. Dommel, Rhine, Austria and Flanders (VMM). The Dommel dataset represents a local industrial exposure scenario (i.e. historic pollution), while the Rhine, Austria and VMM datasets represent a regional mixed exposure scenario (i.e. a

combination of urban, industrial and agricultural pollution). In addition to the environmental monitoring datasets we investigated one dataset that contains high quality environmental geochemical baseline concentrations, i.e. natural background metal concentrations of freshwater surfaces across Europe, FOREGS (Forum of European Geological Surveys). By doing so, we aimed to answer three questions. (1) How big are the differences in risk estimates between the four methods when using actual chronic metal toxicity data and real monitoring datasets? (2) Is there a rank-order in risk estimates between the different methods? We expected that the CA_{SSD} method (Concentration Addition applied directly to the Species Sensitivity Distribution), (see Figure 3) is always the most conservative method among the four based on findings by Backhaus and Faust [4] who demonstrated this mathematically for an assemblage of 3 species. (3) If the CA_{SSD} method is the most conservative method, what is the Margin of Safety (MoS) provided by this method relative to the other three methods?

MATERIAL AND METHODS

A schematic overview of the methodology applied in the present study is given in Figure 1 and is explained step by step in the following paragraphs.

Monitoring data gathering

This manuscript focuses on four monitoring datasets, i.e. the Dommel, the Rhine, the VMM and Austria as well as a dataset with natural baseline concentrations in Europe, the FOREGS database. Extensive information on how these datasets were gathered and processed is given in the Supplementary Information S.1. Main results of all datasets are given in the main paper.

Data within the monitoring datasets was only retained when information on the major water-chemistry variables was present, i.e. Dissolved Organic Carbon (DOC), calcium, pH and

dissolved metal concentrations. When not present in the database, estimations of Na, Mg, K, Cl and SO_4 were based on reported regression relations with Ca concentrations [14]. In addition, alkalinity was estimated based on the pH value [15]. Although we acknowledge that the use of transfer functions (e.g. regressions) to estimate some water characteristics is not ideal, estimation of physico-chemical parameters was necessary (i.e. too little data otherwise). Within the different databases, certain metal concentrations were in some target water samples reported as below detection limit (DL) (see S.1 and S.9). Target water samples that included at least one metal that was reported as below the DL and for which $\sum_{i=1}^{n} \frac{DL}{HC5}$ was larger than one were not retained for data analysis (i.e. 0.3%, 8-10%, 0%, 16% and 0% of the target water samples for the Dommel, VMM, Rhine, Austria and FOREGS database, respectively (S.1)), because such samples would be categorized as 'at risk' while one or more metals would be below the DL, which would not be a meaningful result. For the remaining target water samples (i.e. those that were not removed by that filter), concentrations of metals that were reported to be below the DL were set equal to the DL/2. Although a more detailed investigation of the issue of non-detects is outside the scope of the present study (which was to compare and to rank four mixture risk assessment methods), we acknowledge that the presence of non-detect data is a reality for many monitoring datasets that needs careful consideration. For example, in cases with $\sum_{i=1}^{n} \frac{DL}{HC5}$ >1, it might be recommended to water quality managers to revisit these sampling locations and measure the metal concentrations with more precise equipment.

An overview of the monitoring data is given in Table 2. Monitoring data for sampling locations in the river Meuse tributary Dommel, the Netherlands, were obtained from Verschoor et al. who used the data for a previous study [16]. Monitoring data for Flanders (from now on referred to as "VMM") was gathered from the online database of the Flemish Environmental

Agency (VMM) (www.VMM.be). Monitoring data for the Rhine were gathered from the online database of the International Commision for the Protection of the Rhine (ICPR). Monitoring data for Austria was received from ARCHE (Assessing Risks of Chemicals, Ghent, Belgium). The FOREGS-EuroGeoSurveys Geochemical Baseline Database was obtained on the website of the Geological Survey of Finland and can also be found in Salminen et al. [17]. Additional information on the gathering of the monitoring data is given in the Supplementary Information (S.1).

Chronic toxicity databases

Databases containing chronic toxicity data, i.e. NOECs (No Observed Effect Concentrations) and EC10s (10% Effect Concentrations), of each of the three metals (Cu, Zn and Ni) were used for calculations. For the sake of simplicity as well as for all calculations, from now on 'EC10' will be used to specify both NOEC and EC10 values. Although there is a continuous debate on the use of NOEC vs. EC10 in literature [18-19], these measures are still being used as equivalents of each other in regulatory single metal risk assessments [20-22].

The following chronic toxicity databases were used as starting points for further calculations. The chronic Ni database was originally reported in the Nickel European Union Risk Assessment Report [20] and was recently updated by Nys et al. [23]. The chronic Zn database was reported in 2009 by Van Sprang et al. [14]. The chronic toxicity database of copper was originally reported in the European Union Risk Assessment Report (EU RAR) [22].

The toxicity databases that were used as starting points were updated as follows. For the three metals, a literature search was performed to update the databases with new toxicity data that were published after compilation of the databases. Particular attention was devoted to searching for data for species that were already represented for one or two metals, but not for all

three, as this was helpful for further calculations. Only data from chronic toxicity studies that reported measured metal concentrations (rather than just nominal) and the physico-chemistry of the test media that is important to account for bioavailability (e.g. pH, Ca and DOC concentrations) were included. Chronic toxicity data for three new species were added to the Zn database, i.e. the great pond snail *Lymnaea stagnalis* [24], the fatmucket clam *Lampsilis siliquoidea* [25] and a rotifer species *Brachionus calyciflorus* [24]. No new species were added to the chronic toxicity databases of copper and nickel.

The copper database was additionally updated in the present study according to the chemistry found in the original peer reviewed publications and reports. All adaptations to the copper database and a description of why EC10 values were not retained can be found in the Supplementary Information (S.2). The final toxicity databases for Cu, Ni and Zn which include the physico-chemistry of the test media as well as the chronic toxicity data (i.e. EC10 values) can be found in the Supplementary Information (S.3). The database of Cu contains 133 chronic toxicity test results from in total 27 species. That of Ni contains 31 species (214 test results) and that of Zn contains 22 species (128 test results). The toxicity databases include 7 species for which data on all three metals is present, i.e. they have seven species in common, these species include the algae *Pseudokirchneriella subcapitata*, the cladocerans *Daphnia magna* and *Ceriodaphnia dubia*, the amphipod *Hyalella azteca*, the rotifer *Brachionus calyciflorus* and two fish species *Pimephales promelas* and *Oncorhynchus mykiss*. The effect concentrations of these seven species are evenly distributed within the toxicity databases, i.e. this set of seven species comprises both sensitive and less sensitive species to the different metals. For example, for an average water sample within the VMM database (pH 7.6, Ca 69.0 mg/L and DOC 5.9 mg/L), the

seven species reside between PAF values ranging from 0.04 to 0.89 for Cu, from 0.06 to 0.97 for Ni and from 0.05 and 0.82 for Zn.

In addition to the chronic toxicity data that were already present in the toxicity databases (i.e. EC10 values) we also needed the slope of the dose-response curves, to be able to apply one of the four mixture evaluations tools, i.e. the IA_{DRC} method. For this, we reviewed all literature present in the toxicity databases for all three metals. However, information on the slope of the curves was never reported explicitly in the peer-reviewed papers. Therefore, other methods were used to gather this information. An extensive overview of how slope values were retrieved based on the assumption of a log-logistic dose response curve (Equation 1) is given in the Supplementary information (S.4).

$$y = \frac{100}{1+\left(\frac{x}{EC50}\right)^{\beta}} \text{ or} \qquad \text{(Eq1)}$$

For certain EC10 values within the toxicity databases, no associated information on the slope of the dose-response curve could be retrieved. The percentage of EC10 values for which slope values could be retrieved, was equal to 87%, 84% and 85 % for Cu, Zn and Ni, respectively. Furthermore, this implies that for certain species within the database, no information could be gathered. The percentage of species for which at least one slope value could be retrieved was equal to 96%, 82% and 94% for Cu, Zn and Ni, respectively. The median slope value was equal to 3.8, 2.5 and 2.1 for Cu, Zn and Ni, respectively. 10^{th} and 90^{th} percentile values were equal to 1.9 and 10.9 for Cu, 1.1 and 9.6 for Zn and 1.4 and 7.1 for Ni, respectively. No correlation was found between slope values and the sensitivities of the species, i.e. species that are sensitive to a certain metal (low EC10) can show both low or high slope values

(Supplementary S.5). This is also clear from Figure 2 in which the distribution of the slopes for the different metals is shown. Because no correlations were found, slope values for the species generated (see further on) were sampled randomly from the log-logistic distribution fitted to the set of slope values for each metal (best fit distribution based on the Kolmogorov-Smirnov goodness-of-fit statistic [26-27]).

Bioavailability models and normalizations

Chronic toxicity of metals to aquatic organisms is influenced by water chemistry variables (e.g., pH, water hardness and DOC) due to bioavailability effects of metals. Biotic Ligand Models (BLM) were developed to account for this influence of water chemistry variables on metal toxicity [28]. Therefore, all chronic toxicity data from the three ecotoxicity databases (Cu, Zn and Ni) were normalized to the specific physico-chemistry of each individual water sample (i.e. target water sample) in each of the five monitoring databases before risks for the monitoring sites could be calculated. This was done in an identical way as is explained in Van Sprang et al. [14] for Zn, in the RAR for Cu [22] and in Nys et al. [23] for Ni. An overview of the process of normalization is also given in the Supplementary information (S.6). Normalizations for Zn and Cu were performed using BLM software [29] which incorporates the Windermere Humic Aqueous Model (WHAM) number V [30], while normalizations for Ni were performed using the 'chronic Ni bioavailability and normalization tool' [23] which incorporates the WHAM-Model VI [31].

SSD construction and HC_5 estimation

After normalization of the toxicity data within the three databases to the given target water samples, species sensitivity distribution (SSD) curves were constructed as explained in Van Sprang et al. [14]. The SSDs were fitted in two different ways. (1) The log-normal

distribution was used to construct the SSD for all target water samples and (2) five different parametric distributions (i.e. log-normal, log-gamma, log-logistic, log-exponential and log-weibull) were fitted and the best fitting distribution was determined based on the Kolmogorov-Smirnov goodness-of-fit statistic [26-27]. These two different distribution fittings were compared to examine whether the output based on a single default distribution (i.e. all log-normal) is comparable to the output using the best fit distribution, and whether or not extensive computational work (i.e. using best fit distributions) is redundant.

From these SSDs we calculated HC5 values for each of the single metals, i.e. hazardous concentrations for these metals that are assumed to protect 95% of the species within a community against adverse effects of exposure beyond their no-effect level (here EC10). Parameters of the various SSDs are reported in the Supplementary information S.7.

Toxic pressure (msPAF) calculations

All toxic pressures (expressed as msPAF) reported are on the basis of EC10 values and are given as fractions (ranging between 0 and 1), e.g. msPAF=0.5 means that 50% of the species are assumed to experience 10% effect or more by the mixture.

The toxic pressure of the metal mixture for the different target water samples within the monitoring databases was calculated with four different methods. The R code that was used to apply these methods can be found in the Supplementary Information (S.8).

A first method, and also the most simple approach (Figure 3), was proposed earlier by De Zwart and Posthuma [2] (i.e. their $msPAF_{CA}$ method, Table 1). In this approach, the CA model is applied directly to the SSDs. For this, the species are considered the 'ecological receptors' in an equivalent way as 'toxicological receptors' in individual organisms. Hence, the SSD curve of an individual substance (representing the fraction of species affected as a function of the

concentration of a substance) is considered the equivalent of the dose-response curve of a species, i.e. representing the % effect of the considered endpoint as a function of the concentration. Following this approach, a risk quotient (RQ) for a given chemical mixture can be calculated as follows [3]:

$$RQ_{\frac{PEC}{PNEC}} = \sum_{i=1}^{n} \frac{PEC_i}{PNEC_i} \qquad \text{(Equation 1)}$$

With PEC_i the Predicted Environmental Concentration and $PNEC_i$ the Predicted No Effect Concentration of substance *i*. However, the PNEC is under influence of a certain arbitrariness (i.e. choice of the safety factor applied to toxicity data for each individual substance *i* [3]), and an equivalent, but more general alternative, devoid of arbitrariness, can be formulated based on measured environmental concentrations (EC) and the HC_5:

$$RQ_{\frac{c_i}{HC5}} = SumTU_{HC5} = \sum_i \frac{[c_i]}{HC5_i} \qquad \text{(Equation 2)}$$

With $[c_i]$ the environmental concentration of a metal *i* and HC_5 the hazardous concentration of a metal *i* affecting 5% of the species within a community. According to this approach, which we will further on call the CA_{SSD} approach (Concentration Addition applied directly to the Species Sensitivity Distribution), the community is considered to contain exactly 5% of the species that are potentially affected under the mixture exposure when $RQ_{\frac{c_i}{HC5}} = SumTU_{HC5} = 1$, i.e. the toxic pressure expressed as the multisubstance potentially affected fraction of species ($msPAF_{CA,SSD}$) is equal to 0.05. When $RQ_{\frac{c_i}{HC5}} = SumTU_{HC5} > 1$, more than 5% of the species are potentially

affected. To evaluate whether a sample is at risk or not due to a mixture of metals, i.e. to calculate the $SumTU_{HC5}$, only information on the HC_5 of the metals is necessary, which makes this method the most simple one of the four methods considered (Figure 3).

Next to a Risk Quotient or $SumTU_{HC5}$ for a given mixture scenario, it is also possible to calculate the exact toxic pressure (expressed as $msPAF_{CA,SSD}$). This is done by solving Equation 3 for *x*, i.e. searching for *x* such that the $SumTU_{HCx}$ is exactly one, given the c_i for the three metals. This value of *x* is then the $msPAF_{CA,SSD}$ value of the water body. Calculating an exact toxic pressure therefore not only requires information on the HC_5 of each metal (as is the case for the $SumTU_{HC5}$ calculations), but also knowledge of the mean and standard deviation of the SSD distribution.

$$SumTU_{HCx} = \sum \frac{[c_i]}{HCx_i} = 1 \qquad \text{(Equation 3)}$$

This method for calculating the exact $msPAF_{CA,SSD}$ value is conceptually similar to that of De Zwart and Posthuma [3]. However, with our method we acknowledge that differences between slope values of SSDs may exist among metals, while the method by De Zwart and Posthuma assumes that the slopes of the SSDs are equal across chemicals.

A second approach is analogous to what Backhaus and Faust [4] call the RQ_{STU} approach (Table 1), which they applied for demonstrative purposes to a limited toxicity dataset containing three acute toxicity values (i.e. EC50 values for fish, *Daphnia* and algae) and which also makes use of a safety factor (i.e. assessment factor AF) (Equation 4).

$$RQ_{STU} = \max\left(\sum_{i=1}^{n} \frac{PEC_i}{EC50_{i,algae}}, \sum_{i=1}^{n} \frac{PEC_i}{EC50_{i,daphnids}}, \sum_{i=1}^{n} \frac{PEC_i}{EC50_{i,fish}}\right) \cdot AF \qquad \text{(Equation 4)}$$

The difference with the approach that we follow here is that we extended their methodology to a method for data-rich substances by using an SSD approach. In this approach, which we call the CA_{DRC} approach (Concentration Addition applied to individual Dose Response Curves before calculating the msPAF), the CA model is first applied to toxicity data (i.e. dose response data) of the individual species by calculating a $SumTU_{EC10}$ for each species *j* (Equation 5).

$$SumTU_{EC10,j} = \sum_{i=1}^{n} \frac{[c_i]}{EC10_{i,j}} \qquad \text{(Equation 5)}$$

With $[c_i]$ the environmental concentration of substance *i* and $EC10_{i,j}$ the 10% effect concentration for species *j* for a given substance *i*. Using Equation 4, a species *j* is considered 'affected' if the sum of toxic units relative to the EC10 (i.e. $SumTU_{EC10,j}$) across '*n*' substances exceeds 1. The toxic pressure (expressed as $msPAF_{CA,DRC}$) is then estimated as the fraction of species that at a given mixture exposure are predicted to have a $SumTU_{EC10} > 1$, as this implies that the species would experience an effect of >10% compared to a control (according to the CA concept). To calculate the toxic pressure with the CA_{DRC} method therefore requires information of all EC10 values within each SSD, as is the case for calculation of the $msPAF_{CA,SSD}$ value (Figure 3). An advantage of this method compared to the CA_{SSD} method is that we apply the CA concept on individual species, which is consistent with the original theory of CA [3,4].

A third method is grounded in the other important mixture toxicity concept, i.e. Independent Action (IA), and will be called the IA_{SSD} method (Independent Action applied directly to the Species Sensitivity Distribution). This method has first been proposed by De Zwart and Posthuma [3] (Table 1) and applies the IA model directly to the SSD (Equation 6).

$$msPAF_{IA,SSD} = 1 - \prod_{i=1}^{n}(1 - PAF_i) \qquad \text{(Equation 6)}$$

With PAF_i the potentially affected fraction of species as a result of substance *i*. Similar to the CA_{SSD} and CA_{DRC} method, to calculate the toxic pressure with the IA_{SSD} method requires information on the whole SSD of each metal (i.e. all EC10 values within each SSD) (Figure 3).

A final method has been proposed by Gregorio et al [5] (Table 1), and is the most complex method. This method will be referred to as the IA_{DRC} approach (Independent Action applied to individual Dose Response Curves before calculating the msPAF). For this approach, similar to the CA_{DRC} approach, the IA model is first applied to the dose-response data of the individual species, after which the SSD approach is used to calculate the $msPAF_{IA,DRC}$ value. In a first step, the effect on each individual species *j* due to each substance *i* in a given mixture is calculated following the IA concept (Equation 7).

$$E_j = 1 - \prod_{i=1}^{n}(1 - E_i) \qquad \text{(Equation 7)}$$

To this end it requires the full dose-response curve of each species, i.e. not only the EC10 value but also the slope of the dose-response curve. Subsequently, the toxic pressure (expressed as $msPAF_{IA,DRC}$) is estimated as the fraction of species that at a given mixture exposure is predicted to have more than 10% effect, i.e. $E_j > 0.1$. This method is the most complex approach as it requires not only the EC10 values per species and per substance, but also information on the slopes of the dose-response curves of each substance for each species that is in the toxicity database (Figure 3).

The toxic pressure was calculated using the four above described approaches for all target water samples in the monitoring databases and the natural baseline database. A sample was defined to be at risk when the toxic pressure (expressed as msPAF) was higher than 0.05, which is equivalent to the typical protection goal for single substances, i.e. a maximum of 5% affected species at the HC_5 concentration. The percentage of samples that is predicted to be at risk was calculated for each database. Furthermore, it was examined which individual substances or combinations of substances contributed to the adverse effects.

Generalization of species

Two out of the four approaches listed above, i.e. CA_{DRC} and IA_{DRC}, require data on the individual species. If only the data present in the three chronic toxicity databases would be considered, it would be possible to predict mixture toxicity for only seven species. This is because only these seven species are represented in all three toxicity databases. As natural communities are composed of a multitude of species, the set of actual toxicity data was used to generate a set of hypothetical toxicity data for 20 000 hypothetical species (i.e. species sensitivities were sampled from the SSD) by applying methods to extrapolate unknown species sensitivity from known species sensitivity [33]. This was done in two ways, (1) by not taking into account inter-metal sensitivity correlations when sampling hypothetical species for a given target water sample and (2) by sampling the species based on the correlations found between the sensitivity of a species for one metal and its sensitivity for a second metal for a given target water sample. As the effects of water chemistry on chronic metal toxicity - as predicted with the bioavailability models used - depend on metal identity and species, inter-metal sensitivity correlations can be dependent on the water chemistry of the target water sample. The sampling method that accounted for inter-metal sensitivity correlations was executed using the method of

Iman and Conover [32]. This method is used to generate rank order correlated input distributions and is often applied in literature [33-34].

These two methods were performed and their output (i.e. msPAF values) was compared to examine whether sampling species randomly, i.e. a less computation-time demanding approach than sampling non-randomly, has an influence on the outcome of the risk estimates.

For the present study we chose to use the non-random sampling technique, and in that way we used 'full option' methods for our toxicity predictions and msPAF estimations. The R-codes for both options (random and non-random sampling) are given in the Supplementary information (S.8), so that other users can choose which method to use.

Margin of Safety

The CA_{SSD} method is the most simple method to implement and it is claimed to be a conservative method. By calculating the Margin of Safety (MoS) provided by the CA_{SSD} approach relative to the other methods, the following question can be answered: "By how-many-fold can the $SumTU_{HC5}$ in a given target water sample be raised until 'risk' (msPAF = 0.05) is just being predicted with each of the methods?" For MoS calculations, we start from a situation in which the metals are present at the concentrations and metal-metal ratios as reported in the databases. Then, the metal concentration of each metal is increased (keeping all metal concentration ratios constant) until the level where toxic pressure according to the different approaches equals 0.05. The $SumTU_{HC5}$ at this new combination of metal concentrations is then calculated and this value is equal to the MoS provided by the CA_{SSD} approach. Only those samples were examined which, according to the three different methods (CA_{DRC}; IA_{SSD} and IA_{DRC}) were not affected by the metal mixture (i.e. msPAF < 0.05), as an MoS calculation does not make sense for target water samples not falling in this category.

RESULTS AND DISCUSSION

Monitoring data

A detailed overview of the four monitoring datasets and the geochemical baseline dataset is given in the Supplementary material (S.9). An overview of the main physico-chemical variables and the dissolved metal concentrations is given in Table 3 for all datasets. Median Cu concentrations are similar across the monitoring databases and are on average 91% lower than the median bioavailability corrected HC_5 values in all four monitoring databases (Table 3). Median Zn and Ni concentrations differ more between monitoring databases than Cu. Median Zn and Ni concentrations are below the median bioavailability corrected HC_5 values in all four monitoring databases, and are on average 71% and 84% lower, respectively. For the geochemical baseline dataset (FOREGS) median Cu, Zn and Ni concentrations are below the median bioavailability corrected HC_5 values and are 95%, 93% and 87% lower, respectively.

SSD construction: log-normal or best-fit?

Probability distributions were fitted to the data using (1) log-normal distributions for all data and (2) using distributions that best fitted to the data. The log-normal distribution was the best-fit distribution in 29.2%, 1.8% and 3.9% of the samples for Cu, Zn and Ni respectively. The highest percentage of data was fitted with the log-logistic distribution, i.e. 33.6%, 93.1% and 73.1% of the target water samples for Cu, Zn and Ni, respectively. From the fitted distributions, HC_5 values (based on dissolved concentrations) per target water sample were estimated and 10^{th}, 50^{th} (median) and 90^{th} percentiles of the HC_5 for each monitoring database are given in Table 3. If the conventional log-normal distribution was fitted to all target water samples, median HC_5 values vary between 4.1 µg/L and 46.6 µg/L for Cu, 22.2 µg/L and 52.1 µg/L for Zn and 7.0 µg/L and 27.3 µg/L for Ni. Fitting the best-fit distribution to all target water samples gives

median HC_5 values that vary between 4.3 µg/L and 46.6 µg/L for Cu, 22.9µg/L and 47.9 µg/L for Zn and 6.9 µg/L and 27.3 µg/L for Ni. On average, the HC_5 values generated from log-normal distributions and best-fit distributions are 3.6% higher for Cu, 0.25% higher for Zn and 0.01% higher for Ni. Thus, using a single default distribution (i.e. all log-normal) for mixture toxic pressure estimations, which is computationally less demanding, seems justified.

Furthermore, as the msPAF values calculated based on log-normal and best-fit SSD distributions were similar (see further), the preference was given to only present data (in figures and tables) and conduct downstream data-analyses based on the log-normal species sensitivity distribution in this manuscript and report all results based on best fitting SSDs in the supplementary material.

Generalization of species: random or non-random?

Hypothetical species were generated in 2 ways: (1) by not taking into account inter-metal sensitivity correlations and (2) by taking into account inter-metal sensitivity correlations, which depended on the chemistry of the target water sample. When considering all monitoring datasets together, correlations between the sensitivity of species to Ni and Zn ranged from $r = -0.36$ to $r = 0.48$. However, none of these correlations were statistically significant ($p > 0.05$). Correlations between Ni and Cu ranged from $r = -0.9$ to $r = 0.22$, and only 6.6% were significant ($p < 0.05$). These significant correlations are strong negative correlations ($r < -0.6$), suggesting that – in these 6.6% cases - when a species is sensitive to Ni it is more likely to be less sensitive to Cu and vice versa. In addition, these negative correlations between Cu and Ni sensitivity are more likely to occur at low pH and positive correlations are more likely at high pH (Supplementary information S.9), i.e. at low pH, a species that is less sensitive to Ni is more likely to be more sensitive to Cu and vice versa. Such a correlation is not apparent with either DOC or Ca

concentrations, two other variables affecting metal bioavailability. Correlations between Zn and Cu ranged from r = -0.80 to r = 0.47, and of these only 0.06% were significant correlations ($p < 0.05$). Here too, we observed a trend of negative correlations between Zn and Cu sensitivity at low pH and positive correlations at high pH, and again no trend is apparent with either DOC or Ca concentrations (Supplementary S.9).

As median correlation coefficients between the sensitivity of species to one metal and their sensitivity to a second metal are rather low (see Supplementary S.9), one would expect that the results (i.e. msPAF values) when sampling species by not taking into account inter-metal sensitivity correlations is quite similar to results when sampling the species based on the correlations found between the sensitivities. Indeed, we found that differences in msPAF values between these two methods were small, i.e. on average 0.002 (stdev 0.005) difference in toxic pressure for the CA_{DRC} method and on average 0.001 (stdev 0.003) difference in toxic pressure for the IA_{DRC} method. Therefore, sampling species randomly appears a justifiable option to reduce computational time.

Risk calculations

Results for the most simple method to estimate risks due to metal mixtures, i.e. CA_{SSD}, are visualized in Figure 4. In this figure the (Sum)TU_{HC5} is given for every metal and for every monitoring database. For the Dommel dataset, median TU_{HC5} values are smaller than 1 for all three metals. However, 9% of the TU_{HC5}'s for Ni, 35% of the TU_{HC5}'s for Zn and 0.3% of the TU_{HC5} for Cu within the Dommel basin show a TU_{HC5} >1, indicating that there might be a risk due to the single metals at these sites. Adding up the TU_{HC5}'s gives a $SumTU_{HC5}$ that is indicative for the risk of a mixture of substances. Figure 4 shows that the median $SumTU_{HC5}$ for the Dommel lies above 1, indicating a risk due to the metal mixture or due to single metals in

more than half of the cases. Similar results are found for the VMM, Austria and FOREGS database (Figure 4). However, for these waters the median $SumTU_{HC5}$ lies below 1, indicating that less than half of the cases show a risk due to metal mixtures or single metals. For the Rhine, none of the TU_{HC5} or $SumTU_{HC5}$ values lie above 1, indicating no risk according to the CA_{SSD} method in this waterway.

More advanced methods to calculate the toxic pressure include the four methods described above in which the exact msPAF value of a sample is calculated. Table 4 shows the distribution of toxic pressure (expressed as msPAF values) for all four methods for the different monitoring datasets. For the results using the best-fit SSD calculations, the supplementary material (S.10) can be consulted. A toxic pressure > 0.05 indicates that the sample is affected by the metal mixture.

For the Dommel monitoring database, the median toxic pressure is above 0.05 only when using the CA_{SSD} method (Table 4), which suggests that the most simple method is the most conservative. The median toxic pressure is lowest (0.024) using the IA_{SSD} method, suggesting that this method is the most liberal (least conservative) method. The percentage of samples affected within the Dommel dataset ranges between 52% and 39% depending on the method used. Similar results were obtained for the best fit SSD (S.10), which shows that using the log-normal distribution by default does not have a large influence on the outcome of the toxic pressure calculations. The results suggest that almost half of the samples within the Dommel waterway in the Netherlands are at risk due to metal contamination. However, according to the CA_{SSD} method, 15% of the samples in the Dommel are affected by the mixture itself and not by any individual metal while the IA_{SSD} methods predicts only 3% of the samples to be affected by the mixture of metals itself (Table 4). When going into more detail (Table 5), we see that zinc

has a large effect individually, i.e. in 26.84% of the samples a risk due to Zn alone. Furthermore, in 8.48% of the samples a risk is caused by both zinc and nickel individually (i.e. the TU_{HC5} of zinc and of nickel is above 1). When examining the remaining samples that are affected by the mixture and not by any individual metal, 13.01% of the samples has an effect due to a binary combination of the metals, while 2.14% is affected due to a ternary combination of the metals. When examining the contribution of each metal to the $SumTU_{HC5}$ of these mixture effects, we see that zinc, which has the largest TU_{HC5} in 70.96% of the cases, is the largest contributor to the mixture effect.

Verschoor and colleagues [16] also investigated the mixture toxicity due to Cu, Zn and Ni in the Dommel waterways. These authors assessed the risk by calculating the multi-metal Risk characterization Ratios (RCR) (Equation 7), which is conceptually identical to our CA_{SSD} approach.

$$\sum RCR = \frac{[Cu]}{HC5_{Cu}} + \frac{[Ni]}{HC5_{Ni}} + \frac{[Zn]}{HC5_{Zn}} \qquad \text{(Equation 7)}$$

A similar percentage of affected samples was predicted by Verschoor et al [16], i.e. these authors found that 47% of the samples was at risk, while we found that 52% of the samples were affected (Table 4). In addition, when comparing annual mean RCR values (by Verschoor et al. [16]) with TU_{HC5} values (the present study), these were equal for Zn (i.e. 1.36) and Cu (i.e. 0.075), but not for Ni (i.e. 1.35 vs. 0.47) and therefore also not for the Zn-Cu-Ni mixture (2.79 vs. 1.91).

The difference between the results of Verschoor et al. [16] and our present study could be due to a number of factors. A first factor could be the different parameterization of the BLMs used for bioavailability normalisation. For Cu and Ni, Verschoor et al. [16] used the stability

constants describing the interactions at the biotic ligand from the original BLMs [35, 20], while we used the updated BLMs [36-39]. Furthermore, the choice of speciation software is also different between both studies. Whereas Verschoor et al. [16] used WHAM Model-VI for all speciation calculations, we used WHAM Model-V for Cu and Zn speciation calculations because the BLMs for Cu and Zn were originally calibrated and developed with WHAM Model-V. Finally, we used updated toxicity databases as well as validated BLMs that have been cross-validated for other species and that are currently used in regulatory environmental risk assessments of the three metals. However, despite the considerable differences in methods used by both studies, the differences in the % of samples calculated to be at risk as well as the differences in RCR are still relatively small.

When considering the three monitoring databases other than the Dommel database, the samples in the VMM database are most at risk due to metals with between 23 – 27% of the samples affected (Table 4), depending on the method used. The Rhine is the least at risk, with none (0%) of the samples at risk due to metal contamination. The Austrian samples are situated in-between, with 5-8% of the waters affected (Table 4).

For the FOREGS database, between 7-10% of the waters, depending on the method used, are predicted to be affected (Table 4). The latter demonstrates that even for waters with assumed "natural geochemical baseline" concentrations of metals, a substantial number of water bodies is predicted to be at risk. This result is a well-known issue in metals risk assessment in general, which arises when natural background concentrations of metals, which can vary markedly between geologically different areas, are not taken into account in risk assessment procedures of metals. One way to deal with this issue, e.g. in a higher tier of risk assessment, could be to use the added risk approach [40], but this is beyond the scope of the current study.

When only examining those waters that are affected by the metal mixture and not by any single metal individually in the FOREGS database, between 0.3 and 4% of the samples is at risk depending on the method used (Table 4). This analysis not only shows that true mixture risks are relatively low for the FOREGS database, but also that the 'issue' of risk at geochemical background levels is higher when CA_{SSD} is used compared to the other methods used. Indeed, 4% of the samples is at risk when using the CA_{SSD} while only between 0.3% and 2% of the samples is at risk when using the other methods.

When generalized to all databases, we see that – when the conservative CA_{SSD} method is used – in approximately 1/3 of the samples that is predicted to be 'affected', a risk is predicted due to an actual mixture of metals (and not by any individual metal) (Table 4). However, when the theoretically more correct methods are used (i.e. CA_{DRC} and IA_{DRC}), 1.5 – 2.0 times less samples are affected by mixtures according to the CA_{DRC} method and 2.3 – 8.0 times less samples are affected by mixtures according to the IA_{DRC} method. The difference between mixtures risk between the two latter methods emphasizes the need to establish which model, i.e. CA or IA, is the 'best model' in predicting chronic metal mixture toxicity to individual aquatic species, such that a well-informed choice can be made between CA_{DRC} or IA_{DRC} in the actual metal mixture risk assessment implementation.

Ranking the methods and MoS calculations

A larger difference in % of samples that are at risk is found between the four methods proposed here, i.e. a difference of 3% (Austria) to 13% (Dommel) between the methods. At present, it is too early to conclude which method might be the proper approach. Indeed, these approaches arise from two major toxicity concepts, CA and IA. It is currently not known which of both models is the most appropriate. Moreover, results suggest that the 'most appropriate'

model may be dependent on the metal combination, the species tested, the water chemistry of the test medium and so forth [10, 41]. However, earlier research [4] demonstrated mathematically that for an assemblage of 3 species, the $RQ_{\frac{PEC}{PNEC}}$ approach (i.e. analogous to our CA_{SSD} approach) was always more conservative than the RQ_{TU} approach (analogous to our CA_{DRC} method). Nonetheless, we found here that at high toxic pressure (above 0.15) (expressed as $msPAF_{CA,SSD)}$ the CA_{SSD} approach is no longer the most conservative method as a higher conservatism is found for the CA_{DRC} approach (Figure 5 and Figure 6). Furthermore, for the IA_{DRC} approach, the value at which this shift occurs is higher (i.e. $msPAF_{CA,SSD}$ of 0.55) (Figure 5 and Figure 6). However, we see that at toxic pressure values (expressed as $msPAF_{CA,SSD)}$ below 0.15, there is 0% chance of finding a $msPAF_{CA,SSD}$ value smaller than $msPAF_{CA,DRC}$ or $msPAF_{IA,DRC}$ values (Figure 5). Therefore, at toxic pressures around 0.05, the CA_{SSD} method is the most conservative not only for assemblages of 3 species but also at the community level.

This is also clear from our calculations in which the MoS of the CA_{SSD} approach compared to the other methods was calculated. Figure 7 shows the MoS that the CA_{SSD} method provides relative to the three other methods for all monitoring databases. For the Dommel database, using the CA_{DRC} method, variability in the MoS is the lowest and the median MoS is equal to 1.17. MoS show higher variability and higher median values with the IA_{DRC} and IA_{SSD} method, i.e. 1.38 and 1.48, respectively. The CA_{SSD} method is thus a factor 1.17 to 1.48 more conservative than the other methods. This means that for example, 1.48 fold higher metal concentrations are needed to conclude 'risk' based on the IA_{SSD} approach than based on the CA_{SSD} approach. Even more, if CA (or IA) is a conservative estimator of mixture toxicity across all species, which is demonstrable in toxicity tests, then the simple CA_{SSD} method is on average a factor of 1.17 (or 1.48) conservative.

Our MoS calculations can be compared to the findings by Gregorio et al [5], who based their research on theoretical data sets generated for hypothetical substances. These authors showed that the use of concentration addition directly on SSD (i.e. our CA_{SSD} method) may lead to an overestimation or underestimation of the mixture concentration affecting 5% of the species depending on the standard deviation (*s*) of the SSD of the substances within the mixture. These results were found by calculating $D_{msPAF=5}$, which is the ratio of the mixture concentration affecting 5% of the species calculated with the $M2_{ssd,CA}$ method (i.e. our CA_{SSD} method) to the mixture concentration affecting 5% of the species calculated with the $M1_{sp,CA}$ method (i.e. our CA_{DRC} method). This $D_{msPAF=5}$ value is therefore the reciprocal of our MoS value (i.e. 1/MoS). For mixtures of substances with a steep SSD ($s \leq 0.55$), Gregorio et al [5] demonstrated a higher likelihood of underestimating the mixture concentration affecting 5% of the species when using CA_{SSD} relative to using CA_{DRC} (i.e. $D_{msPAF=5} < 1$). This is in compliance with our results. The mixtures of metals considered in our analysis also showed steep SSD's (i.e. mean *s*; Dommel $s = 0.37$, VMM $s = 0.39$, Rhine $s = 0.47$, Austria $s = 0.50$) and average MoS values for all monitoring databases were larger than 1 (Table 4). Results by Gregorio et al [5] based on hypothetical data for hypothetical substances are therefore confirmed with our results based on real toxicity data for real substances.

In general, our calculations show the following order of conservatism (from most conservative to most liberal):

$$CA_{SSD} > CA_{DRC} > IA_{DRC} > IA_{SSD}$$

This rank order indicates that these methods could be implemented in a tiered metal-mixtures risk evaluation scheme (Figure 8). As CA gives a more cautious risk estimate (at low msPAF values), the CA_{SSD} method could serve as a first (conservative) tier to identify situations with likely no risk of metal mixtures ($SumTU_{HC5}<1$). The IA_{SSD} method could be applied in a second tier to identify situations of risk regardless of the method used (msPAF > 0.05). The CA_{DRC} and IA_{DRC} methods could be used in a third tier for more detailed calculations for situations that fall in-between, e.g. as part of a weight-of-evidence approach. For situations for which the outcome is dependent on the method used (i.e. CA_{DRC} or IA_{DRC}), targeted research could be performed in a final tier.

Our MoS calculations also demonstrate the possibility for an intermediate tier between the proposed Tier 1 and 2. When the $SumTU_{HC5}$ is > 2 (Figure 7), risk is always predicted, independent of the method used. The intermediate tier could therefore implement a cut-off on the $SumTU_{HC5}$ value, above which risk is always predicted and thus avoid unneeded time and resource investment in the more complicated calculations. However, this case is so far only demonstrated for the 4 monitoring datasets examined here and should first be examined more thoroughly before this intermediate tier can be added to the tiered metal-mixtures risk evaluation scheme (Figure 8).

Strengths and weaknesses

The research conducted here shows certain strengths when compared to existing literature. We evaluated the use of four mixture risk assessment methodologies simultaneously. For this, we used available real toxicity data and monitoring datasets. In addition, we compared the influence of the use of the log-normal and the best-fit SSD on the risk estimations, as well as the influence of generating hypothetical species randomly versus non-randomly on the risk

estimations. However, certain weaknesses in our research also exist. An important obstacle for applying either method at this moment is that the underlying assumptions of the different methods needs to be tested and the degree of conservatism compared to community-level metal mixture toxicity effects needs to be investigated (e.g. based on mesocosm or field data). Another current limitation of our methods is the fact that calculations of mixture toxicity have been performed on the basis of dissolved metal concentrations, while possible interactions between metals at DOC sites have not been accounted for. This is due to the fact that in the validated bioavailability models, speciation calculations for the different metals are performed with different speciation models in the current BLMs (i.e. WHAM Model V for Cu and Zn vs. WHAM Model VI for Ni). Taking into account these interactions at DOC sites could result in higher predicted msPAF values. However, we expect that this would only influence the absolute msPAF values per method, but not the relative ranking of msPAF values among the different methods.

Research recommendations

Although the CA_{SSD} approach is the most conservative (at $msPAF_{CA,SSD}$ values < 0.15), is the most easy to implement and shows a high margin of safety, more research is needed to conclude whether more complex, liberal methods (i.e. CA_{DRC} or IA_{DRC}) might be more accurate in predicting the level of risk posed by mixtures of metals. This could for instance be examined by performing mesocosm experiments. The real ecological meaning of the msPAF has been a topic of research [42-44], but uncertainties remain. For instance, although it is assumed that the HC_5 value for a single substance is protective for 95% of the species within a community, it is not straightforward to predict what effects may occur in an actual community when exposed to a concentration equal to the HC5 of that substance. This uncertainty applies invariantly to mixtures

of substances and thus, it is not straightforward to predict what effects on natural communities may be expected when exposed to a metal mixture with a toxic pressure equal to any msPAF. The relation between msPAF and actual effects of metal mixtures on natural communities should be the subject of future research. The calculated toxic pressure (expressed as msPAF values) could possibly be applied in an absolute way if the toxic pressure can be correlated to ecological effects or in a relative way by ranking contaminated sites. Either way, we propose that applying a tiered metal mixture risk evaluation scheme in which the four methods described here are applied, might be a way forward to evaluate risks implied by mixtures of substances.

In addition, the perception exists that adding more metals to a mixture, even when metals are present at background concentrations, will result in risk predictions for a higher percentage of samples. However, when examining the results from the FOREGS database, we see that only a limited percentage of samples (i.e. up to 4%) is affected by a mixture of 3 metals. Even more, when applying theoretically more correct models (i.e. CA_{DRC} and IA_{DRC}) the issue of mixture toxicity is even lower, i.e. 0.5 to 2% of the samples is said to be at risk at background concentrations. However, more research is needed to establish whether CA or IA is the 'best' model to implement. Further research needed also includes the update of existing chronic metal bioavailability models to the same speciation model (e.g. WHAM model VII) to allow more consistent speciation-based computations of msPAF.

CONCLUSION

The present study examined the use of 4 mixture risk assessment methodologies that combine chronic toxicity data, bioavailability modelling, SSDs and CA or IA for ecological risk assessment by calculating the toxic pressure (expressed as msPAF values) based on measured concentrations of metals in 4 monitoring databases and 1 natural baseline database. The

percentage of samples predicted to be at risk differed between the methods used and were between 0 % (Rhine) and 52 % (Dommel) when using the most simple approach (CA_{SSD}). When only examining the samples that were at risk due to metal mixtures and not due to any individual metals, % of affected samples ranged between 0 % (Rhine) and 15 % (Dommel). The % of samples predicted to be affected also differed between the methods used, with a difference of 3 to 13% between methods.

In general, our calculations showed the following order of conservatism for the 4 methods (from most to least conservative): $CA_{SSD} > CA_{DRC} > IA_{DRC} > IA_{SSD}$. As the CA_{SSD} method, the most simple method to implement, was shown to be the most conservative method (below certain risk values), MoS values could be calculated. It was demonstrated that the CA_{SSD} method is a factor 1.17 to 1.48 more conservative that the other methods (based on the Dommel dataset). Finally, we suggest applying these four approaches in a general tiered scheme for the risk assessment of chemical mixtures in a regulatory context. In this scheme, the CA_{SSD} method could serve as a first (conservative) tier to identify situations with likely no potential risk at all, regardless of the method used ($SumTU_{HC5} < 1$) and the IA_{SSD} method to identify situations of potential risk, also regardless of the method used ($msPAF_{IA,SSD} > 0.05$). The CA_{DRC} and IA_{DRC} methods could be used for site-specific assessment for situations that fall in-between ($SumTU_{HC5} > 1$ and $msPAF < 0.05$).

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

Acknowledgment—The present study was funded by the Nickel Producers Environmental Research Association, the International Zinc Association and the Copper Alliance. The authors wish to thank K. Viaene, F. De Laender and F. Verdonck for their help with R coding and with

the statistics as well as M. Vijver and ARCHE for providing monitoring datasets used in the present study.

Data Availability—Data will be provided via Supplemental information with the manuscript

REFERENCES

1. EU Parliament and Council: Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC, 2006; http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:l21282

2. Backhaus T, Rofl Altenburger R, Arrhenius A, Blanck H, Faust M, Finizio A, Gramatica P, Grote M, Junghans M, Meyer W, Pavan M, Porsbring T, Scholze M, Todeschini R, Vighi M, Walter H, Grimme LH. 2003. The BEAM-project: Prediction and assessment of mixture toxicities in the aquatic environmental. *Continental Shelf Research* 23: 1757-1769

3. De Zwart D, Posthuma L. 2005. Complex mixture toxicity for single and multiple species: proposed methodologies. *Environmental Toxicology and Chemistry* 24(10): 2665-2676

4. Backhaus T, Faust, M. 2012. Predictive environmental risk assessment of chemical mixtures: a conceptual framework. *Environmental Science and Technology*, 46: 2564-2573

5. Gregorio V, Chèvre N, Junghans M. 2013. Critical issues in using the common mixture toxicity models concentration addition or response addition on species sensitivity distributions: a theoretical approach. *Environmental Toxicology and Chemistry*, 32(10): 2387-2395

6. European commission. 2012. Toxicity and assessment of chemical mixtures. European Commission, Brussels, Belgium [cited 15 January 2016]. Available online at: http://ec.europa.eu/health/scientific_committees/environmental_risks/docs/scher_o_155.pdf]

7. Balistrieri LS, Mebane CA, Schmidt TS, Kelle WB. 2015. Expanding metal mixture toxicity models to natural stream and lake invertebrate communities. *Environmental Toxicology and Chemistry.* 34(4): 761–776.

8. Mebane, C.A., T.S. Schmidt, and L.S. Balistrieri. 2016. Larval aquatic insect responses to cadmium and zinc in experimental streams. *Environmental Toxicology and Chemistry*. DOI: 10.1002/etc.3599

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

10. Nys C, Asselman J, Hochmuth J, Janssen CR, Blust R, Smolders E, De Schamphelaere KAC. 2015. Mixture toxicity of nickel and zinc to *Daphnia magna* is noninteractive at low effect sizes but becomes synergistic at high effect sized. *Environmental Toxicology and Chemistry* 34(5): 1091-1102

11. Nys C, Van Regenmortel T, Janssen CR, Blust R, Smolders E, De Schamphelaere KAC. 2016. Comparison of chronic mixture toxicity of Ni-Zn-Cu and Ni-Zn-Cu-Cd mixtures between *Ceriodaphnia dubia* and *Pseudkirchneriella subcapitata*. *Environmental Toxicology and Chemistry.* DOI: 10.1002/etc.3628

12. Iwasaki Y, Cadmus P, Clements WH. 2013. Comparison of different predictors of exposure for modelling impacts of metal mixtures on macroinvertebrates in stream microcosms. *Aquatic Toxicology* 132-133: 151-156

13. Richardson JS, Kiffney PM. 2000. Responses of a macroinvertebrate community from a pristine, southern British Columbia, Canada, stream to metals in experimental mesocosm. *Environmental Toxicology and Chemistry* 19(3): 736-743

14. Van Sprang PA, Verdonck FA, Van Assche F, Regoli, L, De Schamphelaere KAC. 2009. Environmental risk assessment of Zn in European freshwaters: a critical appraisal. *Science of the Total Environ*ment 407(20): 5373-5391.

15. Stumm W, Morgan JJ. *Aquatic Chemistry: chemical equilibria and rates in natural waters*; John Wiley & Sons: New York, 1996

16. Verschoor AJ, Vink JPM, Snoo, GR, Vijver MG. 2011. Spatial and temporal variation of watertype-specific no-effect concentrations and risks of Cu, Ni and Zn. *Environmental Science and Technology* 45: 6049-6056

17. Salminen R. (Chief-editor), Batista MJ, Bidovec M, Demetriades A, De Vivo B, De Vos W, Duris M, Gilucis A, Gregorauskiene V, Halamic J, Heitzmann P, Lima A, Jordan G, Klaver G, Klein P, Lis J, Locutura J, Marsina K, Mazreku A, O'Connor PJ, Olsson S.Å, Ottesen RT, Petersell V, Plant J.A, Reeder S, Salpeteur I, Sandström H, Siewers U, Steenfelt A, Tarvainen T, Geochemical Atlas of Europe. Part 1: Background Information, Methodology and Maps. Espoo, Geological Survey of Finland, 526 pages, 36 figures, 362 maps. 2005. http://weppi.gtk.fi/publ/foregsatlas/

18. Jager T. 2012. Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. Environmental Toxicology and Chemistry, 31: 228–229.

19. Chapman PM, Caldwell RS and Chapman PF. 1996. A warning: NOECs are inappropriate for regulatory use. Environmental Toxicology and Chemistry, 15: 77–79.

20. *Danish Environmental Protection Agency. European Union Risk Assessment Report (EU RAR) of nickel and nickel compounds prepared by the Danish Environmental Protection Agency (DEPA).* 2008. Available on: http://echa.europa.eu/documents/10162/cefda8bc-2952-4c11-885f-342aacf769b3. Last accessed: 3/3/2015. Last accessed: 07/07/2015.

21. Risk assessment zinc metal – environmental part. Final draft submitted to the European Chemicals Bureau (Ispra, Italy) by the National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands; 2006

22. EU, 2008. European Copper Institute. European Union risk assessment report: Voluntary risk assessment of copper, copper II sulphate pentahydrate, copper(I)oxide, copper(II)oxide, dicopper chloride tri-hydroxide. Brussels, Belgium

23. Nys C, Janssen CR, Van Sprang PA, De Schamphelaere KAC. The effect of pH on chronic aquatic Ni toxicity is dependent on the pH itself: extending the chronic Ni bioavailability models. Environmental Toxicology and Chemistry. DOI 10.1002/etc.3232

24. De Schamphelaere KAC, Janssen CR. 2010. Cross-phylum extrapolation of the *Daphnia magna* chronic biotic ligand model for Zn to the snail *Lymnaea stagnalis* and the rotifer *Brachionus calyciflorus*. *Science of the Total Environ*ment 408(22): 5414-5422.

25. Wang N, Ingersoll CG, Ivey CD, Hardesty DK, May WM, Augspurger T, Roberts AD, Van Genderen E, Barnhart MC. 2010. Sensitivity of early life stages of freshwater mussels (Unionidae) to acute and chronic toxicity of lead, cadmium, and Zn in water. *Environmental Toxicology and Chemistry* 29(9): 2053-2063.

26. Gan FF, Koehler KJ, Thompson J.C. 1991. Probability plots and distributions curves for assessing the fit of probability models. *The American Statistician* 45: 14-21

27. Stephens MA. Andersen-Darling test for goodness-of-fit. In: Kots, S; Johnson, N.L; editors. Encyclopedia of Statistical Sciences, vol 1. New York, NY, USA: John Wiley, 1982. 9 41.

28. Di Toro DM, Allen HE, Bergman H.L, Meyer JS, Paquin .R, Santore RC. 2009 Biotic ligand model of acute toxicity of metals. 1. Technical basis. *Environmental Toxicology and Chemistry* 20(10): 2383-2396

29. Biotic Ligand Model, version 2.1.2, HydroQual, Mahway, NJ, USA, http://hydroqual.com/wr_blm.html

30. Tipping E. 1994. WHAM – a chemical equilibrium model and computer code for waters, sediments and soils incorporating a discrete sie/electorstatic model of ion-binding by humic substances. *Computational Geosci*ences 20, 973-1023

31. Tipping E. 1998. Humic ion-Binding Model VI: an improved description of the interactions of protons and metal ions with humic substances. *Aquatic Geochemistry* 4, 3-48

32. Iman RL, Conover WJ. 1982. A distribution-free approach to inducing rank order correlation among input variables. *Commununications in Statistics-Simulations and Computation* 11(3), 311-334.

33. Verdonck F.2003. Geo-referenced probabilistic ecological risk assessment. PhD Thesis, Ghent University, Ghent, Belgium.

34. Cullen AC, Frey HC. 1999. Probablistic techniques in exposure assessment. A handbook for dealing with variability and uncertainty in models and inputs. Plenum, New York. 335.

35. De Schamphelaere KAC, Vasconcelos FM, Heijerick DG, Tack FM, Delbeke K, Allen HE, Janssen CR. 2003. Development and field validation of a predictive copper toxicity model for the green alga *Pseudokirchneriella subcapitata*. *Environmental Toxicology and Chemistry* 22, 2454–2465.

36. De Schamphelaere KAC, Janssen C.R. 2006. Bioavailability models for predicting copper toxicity to freshwater green microalgae as a function of water chemistry. *Environmental Science and Technology* 40:4515–4522.

37. Deleebeeck NME, De Schamphelaere KAC, Janssen CR. 2007. A bioavailability model predicting the toxicity of nickel to rainbow trout (*Oncorhynchus mykiss*) and fathead minnow

(*Pimephales promelas*) in synthetic and natural waters. *Ecotoxicology and Environmental Safety* 67: 1-13

38. Deleebeeck NME, De Schamphelaere KAC, Janssen CR. 2009. Effects of Mg^{2+} and H^{+} on the toxicity of Ni^{2+} to the unicellular green algae *Pseudokirchneriella subcapitata*: Model development and validation with surface waters. *Science of the Total Environment* 407: 1901-1914

39. Deleebeeck NME, De Schamphelaere KAC, Janssen CR. 2008. A novel method for predicting chronic nickel bioavailability and toxicity to Daphnia magna in artificial and natural waters. *Environmental Toxicology and Chemistry* 27(10): 2097-2107

40. Struijs J, van de Meent D, Peijnenburg WJGM, van den Hoop MAGT, Crommentuijn T. 1997. Added risk approach to derive maximum permissible concentrations for heavy metals: how to take natural background levels into account. *Ecotoxicology and Environmental Safety* 37(2): 112-118

41. Norwood WP, Borgmann U, Dixon DG, Wallace A. 2003. Effects of metal mixtures on aquatic biota: A review of observations and methods. *Human and Ecological Risk Assessment* 9: 795-811

42. Smetanova S, Blaha L, Liess M, Beketov MA. 2014. Do predictions from Species Sensitivity Distribtuions match with field data? *Environmental Pollution* 189: 126-133

43. Posthuma L, de Zwart D. 2006. Predicted effects of toxicant mixtures are confirmed by changes in fish species assemblages in Ohia, USA, Rivers. *Environmental Toxicology and Chemistry* 25: 1094-1105

44. Posthuma L, de Zwart D. 2012. Predicted mixture toxic pressure relates to observed fraction of benthic macrofauna species impacted by contaminant mixtures. *Environmental Toxicology and Chemistry* 31(9): 2175-2188.

Figure 1. Overview of the methodology used for the calculations in the present study. Hexagonal boxes represent different steps of data collection and data handling, rounded boxes represent calculations and rectangular boxes represent outcomes of calculations. Reference is given to tables, figures and supplementary material. ECx = x% Effect Concentration ; BLM = Biotic Ligand Model; SSD = Species Sensitivity Distribution; HC5 = hazardous concentration affecting 5% of the species within a community; CA= Concentration Addition; IA = Independent Action; msPAF = multi-substance potentially affected fraction. All msPAF values reported are on the basis of EC10 values.

Figure 2. Distribution of slope values of dose-response curves for the Zn (A), Cu (B) and Ni (C) chronic ecotoxicity database. Slope values for fish, invertebrates and algae are depicted as squares (red), triangles (blue) and diamonds (black), respectively.

Figure 3. Overview of 4 different methods combining two mixture toxicity concepts, concentration addition (CA) and independent action (IA) with species sensitivity distribution (SSD) functions to estimate toxic pressure expressed as multi-substance potentially affected fractions (msPAF) of species exposed to metal mixtures. For each method, the general mathematical function is given as well as the data required to calculate the msPAF value. The msPAF values reported are on the basis of EC10 values.

Figure 4. Toxic Units (TU_{HC5}) for Ni, Zn and Cu for the different target water samples of the Dommel (A), VMM (B), Rhine (C), Austria (D) and FOREGS (E) dataset. $SumTU_{HC5}$ shows the summation of the TU_{HC5}'s according to the CA_{SSD} method using the log-normal SSD distribution. The horizontal line indicates a TU_{HC5} or $SumTU_{HC5}$ of 1. Results are represented as box plots: median values are given in bold, bottom and top of the box plots give the 25^{th} and 75^{th}

percentile. Bottom and top of the error bars represent the 5^{th} and 95^{th} percentile, open circles are outliers.

Figure 5. Percentage of samples for which the $msPAF_{CA,SSD}$ value is smaller than the $msPAF_{CA,DRC}$ value (top) and for which $msPAF_{CA,SSD}$ value is smaller than $msPAF_{IA,DRC}$ value (bottom), for different categories of $msPAF_{CA,SSD}$ values (for the Dommel, VMM, Austria and FOREGS databases combined).

Figure 6. Comparison of the toxic pressure (expressed as msPAF) according to the CA_{SSD} versus CA_{DRC} method (left graphs) and according to the CA_{SSD} versus IA_{DRC} method (right graphs) for the Dommel (A), the VMM (B), Austria (C) and FOREGS (D) database. The msPAF values reported are on the basis of EC10 values.

Figure 7. Representation of the Margin of Safety (MoS), i.e. the $SumTU_{HC5}$ corresponding to a msPAF of 0.05 for the Dommel (A), the VMM (B), Rhine (C) and Austria (D) database, for the different methods: CA_{DRC}; IA_{DRC} and IA_{SSD}. Results are represented as box plots: median values are given in bold, bottom and top of the box plots give the 25^{th} and 75^{th} percentile. Bottom and top of the error bars represent the 5^{th} and 95^{th} percentile, asterisks are outliers.

Figure 8. Possible tiered metal mixture risk evaluation scheme. A sample is defined to be at risk when the toxic pressure (expressed as msPAF) was higher than 0.05 (or $SumTU_{HC5} > 1$), which is equivalent to the typical protection goal for single substances, i.e. a maximum of 5% affected species at the HC_5 concentration. The msPAF values reported are on the basis of EC10 values.[a] Unless very strong synergisms at low effect levels. [b] unless very strong antagonisms at low effect levels.

Tables

Table 1. Four different approaches to calculate the toxic pressure expressed as multisubstance potentially affected fractions (msPAF[a]) of species that are described in this study, and terminology of equivalent or analogous approaches used by De Zwart and Posthuma [2], Backhaus and Faust [3] and Gregorio et al [5]. [a] The msPAF values reported are on the basis of EC10 values.

This study	De Zwart and Posthuma	Backhaus and Faust	Gregorio et al.
CA_{SSD}	CA	$RQ_{PEC/PNEC}$	$M2_{ssd,CA}$
CA_{DRC}	NI	RQ_{STU}	$M1_{sp,CA}$
IA_{SSD}	RA or IJA	NI	$M2_{ssd,IA}$
IA_{DRC}	NI	NI	$M1_{sp,IA}$

CA = Concentration Addition; SSD = Species Sensitivity Distribution, RQ = Risk Quotient, PEC = Predicted Environmental Concentration; PNEC = Predicted No Effect Concentration; DRC = Dose-Response Curve, NI = Not Included; RA = Response Addition; IJA = Independent Joint Action; STU = Sum of Toxic Units

Table 2. Overview of the monitoring databases used in this study

Database	Exposure scenario	Time period	Number of samples	Number of sampling locations
Dommel	Industrial (historic pollution)	2007-2010	3176	97
VMM	Regional mixed[a]	2012	155	48
Rhine	Regional mixed[a]	2010-2011	209	53
Austria	Regional mixed[a]	2006	2138	249
FOREGS	Natural background	1998-2001	784	784

[a] i.e. a combination of urban, industrial and agricultural pollution

Table 3. Physico-chemical parameters (pH, Dissolved Organic Carbon and Ca concentration) and dissolved metal concentrations (Nickel, Zinc and Copper) of the different monitoring databases. In addition, HC_5 values (hazardous concentration affecting 5% of the species within a community, beyond their no-effect level (here EC10)) for the different monitoring datasets (log-normal SSD and best-fit SSD).

		Dommel	VMM	Rhine	Austria	FOREGS
pH		7.1 (6.5 - 7.6)[a]	7.6 (7.0 - 8.0)	8.0 (7.8 - 8.2)	8.0 (7.6 - 8.3)	7.7 (6.4 - 8.3)
DOC[b] (mg/L)		9.4 (5.5 - 15.0)	7.7 (5.2 – 15.1)	2.4 (1.7 - 3.4)	1.6 (0.7 - 4.5)	5.3 (1.0 - 17.1)
Ca (mg/L)		41.4 (31.0 - 57.0)	84.0 (26.4 - 146.0)	67.0 (50.0 - 110.4)	45.9 (18.8-80.0)	40.3 (2.8 - 118.2)
Ni (µg/L)		8.3 (0.8 - 29.0)	2.5 (2.0 - 11.0)	1.1 (0.5 - 2.0)	0.5 (0.03 - 1.9)	1.9 (0.4 - 4.7)
Zn (µg/L)		28 (3.5 - 98.0)	15.0 (5.0 - 66.0)	2.8 (1.0 - 5.1)	1.9 (0.4 - 7.8)	2.7 (1.0 - 9.8)
Cu (µg/L)		2.1 (0.5 - 4.6)	1.0 (1.0 - 4.0)	1.6 (0.8 - 2.3)	0.5 (0.4 - 1.6)	0.9 (0.3 - 2.3)
Ni HC_5	log-normal	27.3 (18.1-39.4)	20.6 (14.9-32.8)	7.9 (5.3-18.9)	7.0 (3.9-14.4)	14.8 (4.1-39.3)
	best-fit	27.3 (18.2-39.3)	22.1 (16.0-31.3)	7.9 (5.2-17.2)	6.9 (3.8-14.7)	14.6 (3.9-38.1)
Zn HC_5	log-normal	42.8 (27.4-67.9)	52.1 (27.9-92.9)	24.4 (19.1-36.5)	22.2 (13.2-40.4)	36.2 (14.2-100.0)
	best-fit	42.6 (27.3-67.8)	47.9 (27.0-81.9)	25.5 (19.0-40.3)	22.9 (12.9-40.4)	37.8 (14.1-95.3)
Cu HC_5	log-normal	46.6 (19.0-78.8)	39.5 (24.3-82.0)	12.5 (7.3-21.7)	4.1 (1.9-13.3)	19.6 (3.4-74.3)
	best-fit	46.6 (19.0-78.8)	39.4 (24.3-78.5)	13.4 (7.2-23.6)	4.3 (2.2-13.1)	19.7 (3.7-73.9)

[a] Values reported are median values, 10th and 90th percentiles are given in between parentheses.

[b] DOC = Dissolved Organic Carbon

Table 4. Toxic pressure expressed as multisubstance potentially affected fraction of species (msPAF[a]) for the Dommel, VMM, Rhine, Austria and FOREGS database obtained with the different methods (CA_{SSD}; CA_{DRC}; IA_{SSD} and IA_{DRC}) when SSDs are fitted with log-normal distributions. The percentage of affected samples is given per method. Furthermore, median Margin of Safety (MoS) values provided by the CA_{SSD} approach for the other methods are given. NA = not applicable, [a] the msPAF values reported are on the basis of EC10 values

	Dommel				VMM				Rhine			
	CA_{SSD}	CA_{DRC}	IA_{SSD}	IA_{DRC}	CA_{SSD}	CA_{DRC}	IA_{SSD}	IA_{DRC}	CA_{SSD}	CA_{DRC}	IA_{SSD}	IA_{DRC}
median msPAF	0.054	0.038	0.024	0.027	0.009	0.004	0.003	0.003	0.006	0.002	0.002	0.002
10^{th} percentile msPAF	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	<0.001	<0.001	<0.001
90^{th} percentile msPAf	0.423	0.466	0.342	0.364	0.227	0.2201	0.177	0.185	0.012	0.0005	0.003	0.004
% samples affected (msPAF>0.05)	52	46	39	44	27	25	23	23	0	0	0	0
% samples affected by mixture of metals and not by any of the individual metals	15	10	3	5	7	4	2	3	0	0	0	0
MoS provided by the CA_{SSD} approach	NA	1.17	1.48	1.38	NA	1.18	1.57	1.46	NA	1.25	1.72	1.60

	Austria				FOREGS			
	CA_{SSD}	CA_{DRC}	IA_{SSD}	IA_{DRC}	CA_{SSD}	CA_{DRC}	IA_{SSD}	IA_{DRC}
median msPAF	0.004	0.001	0.001	0.001	0.004	0.001	0.001	0.001
10^{th} percentile msPAF	<0.001	<0.001	<0.001	<0.001	<0.001	0	<0.001	<0.001
90^{th} percentile msPAf	0.035	0.023	0.016	0.017	0.052	0.039	0.031	0.033
% samples affected (msPAF>0.05)	8	6	5	5	10	8	7	7
% samples affected by mixture of metals and not by any of the individual metals	3	2	0.2	0.6	4	2	0.4	0.5
MoS provided by the CA_{SSD} approach	NA	1.21	1.52	1.45	NA	1.22	1.52	1.43

Table 5. Percentage of samples that is not affected and percentage that is affected (msPAF value > 0.05) by a mixture of Cu, Zn and/or Ni according to the CA_{SSD} method for the Dommel database.

	Percentage			
No effect	48.33			
Effect	51.67			
	Individual metal effects	36.52		
		Only Zinc[a]	26.84	
		Only Nickel[a]	0.91	
		Only Copper[a]	0.09	
		Both Zinc and Nickel[b]	8.48	
		Both Zinc and Copper[b]	0.16	
	Mixture effects	15.15		
		Binary combinations[c]	13.01	
		Ternary combination[d]	2.14	
			Shows the largest TU[e]	Percentage
			Zn	77.96
			Ni	22.04
			Cu	0

[a] The Toxic Unit of zinc, nickel or copper is above 1
[b] The Toxic Unit of all mentioned metals is above 1
[c] At least one of the possible binary combinations (i.e.Zn&Ni, Zn&Cu, Ni&Cu) shows an effect
[d] The ternary combination (but none of the 3 possible binary combinations) shows an effect
[e] For each metal the percentages of samples is given in which that metal has the largest Toxic Unit in the sample affected by a binary or ternary combination, i.e.in which that metal is the largest contributor to the toxic effect

Figure 1.

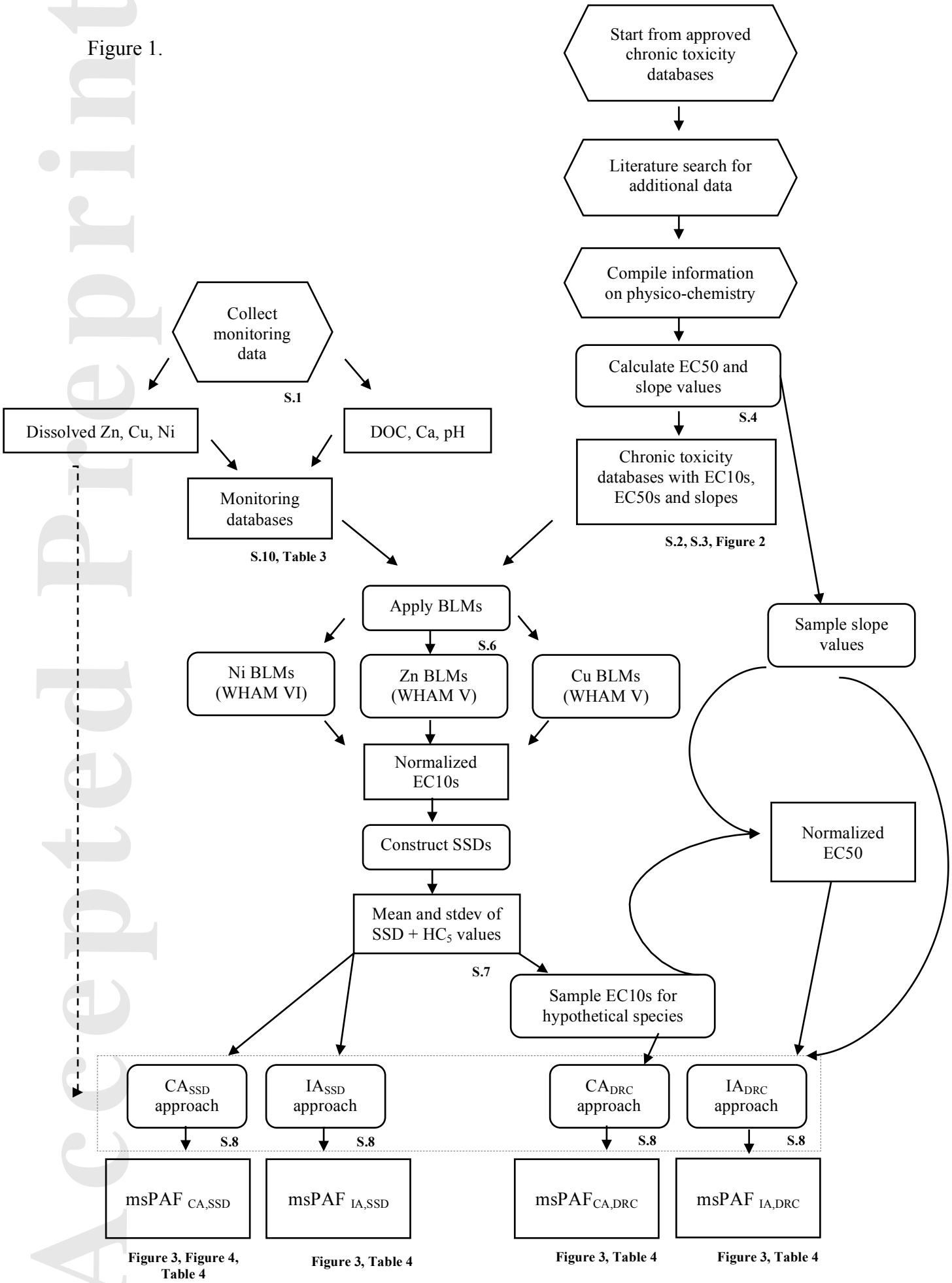


Figure 2.

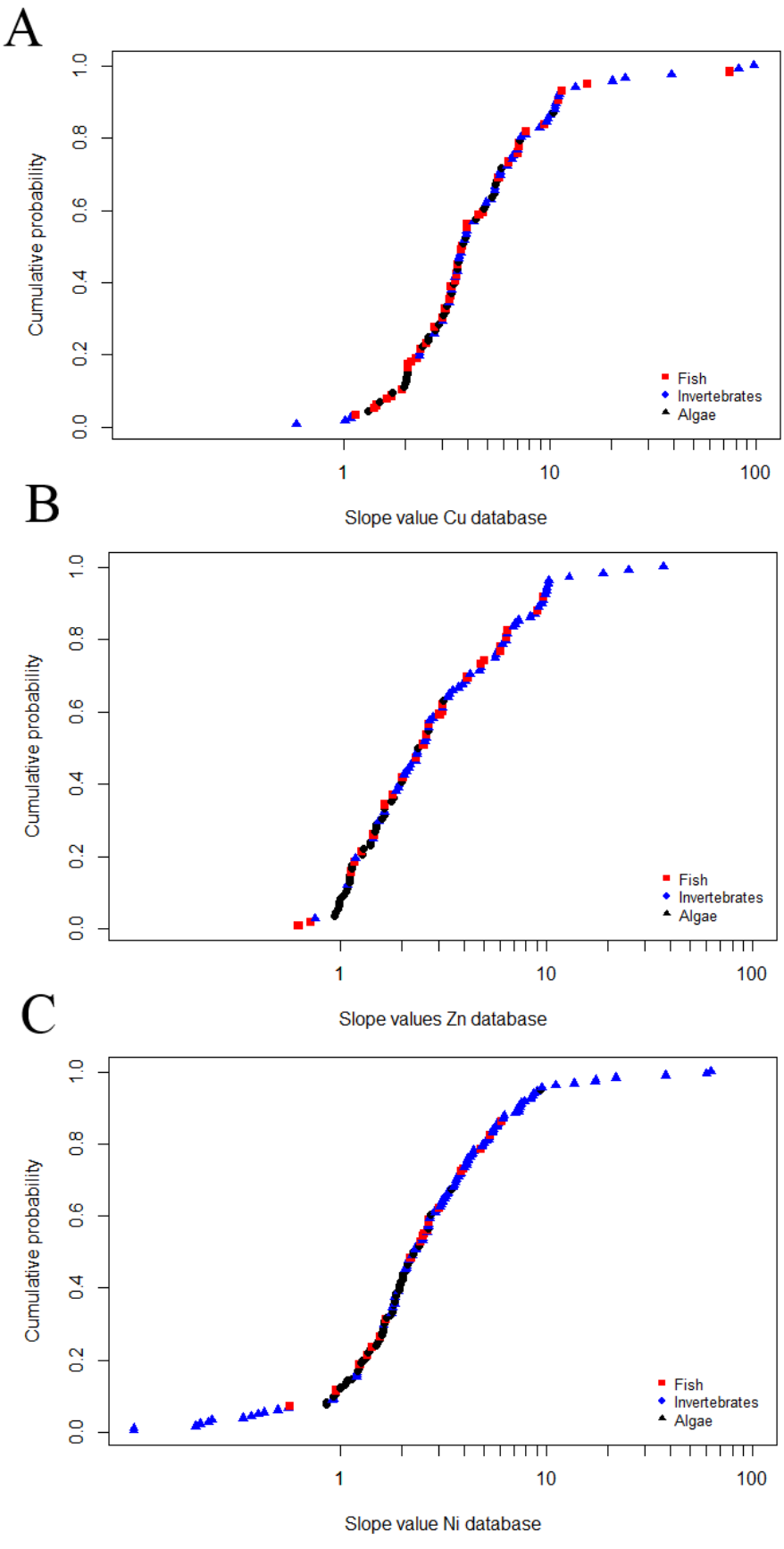


Figure 3.

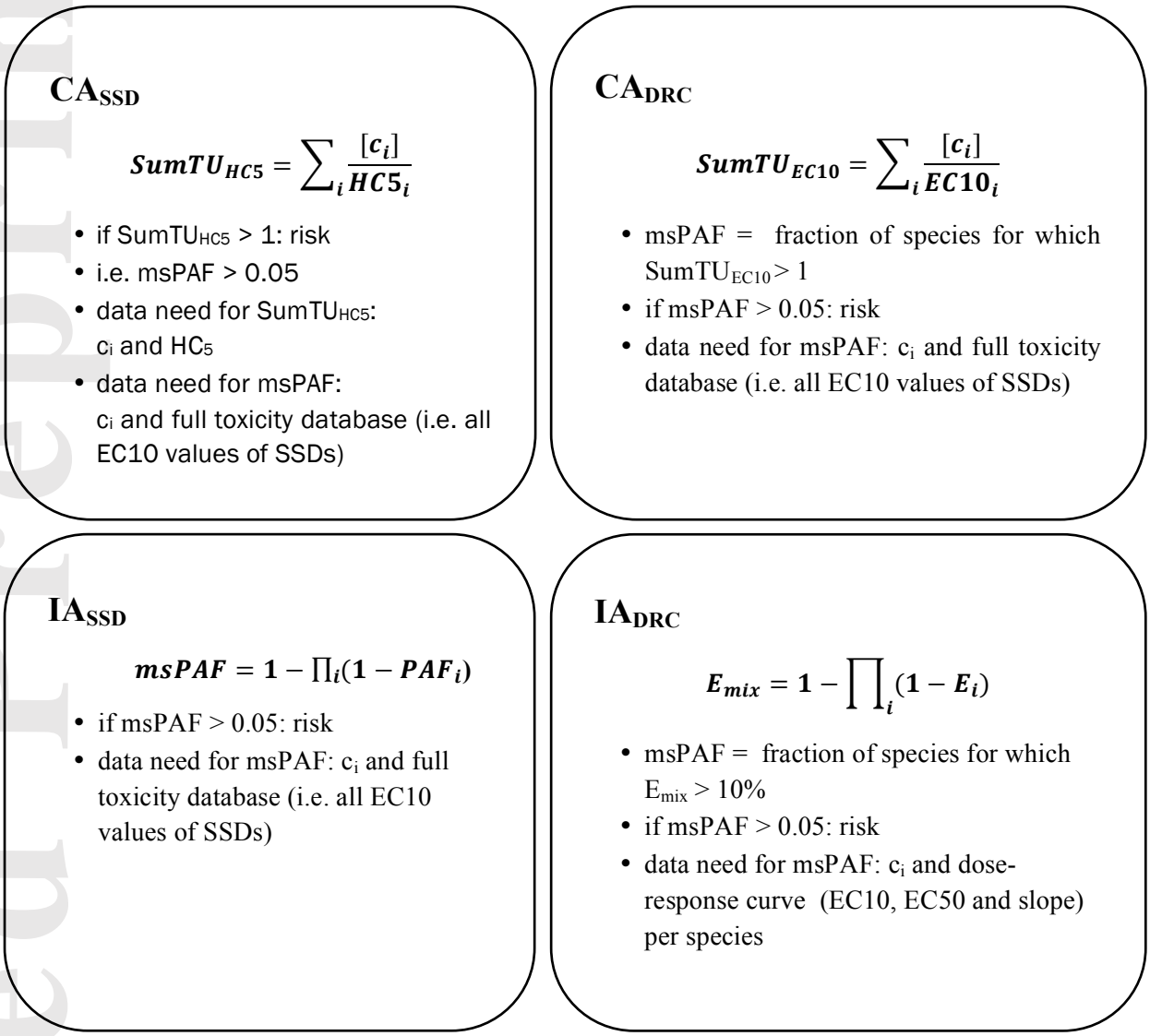
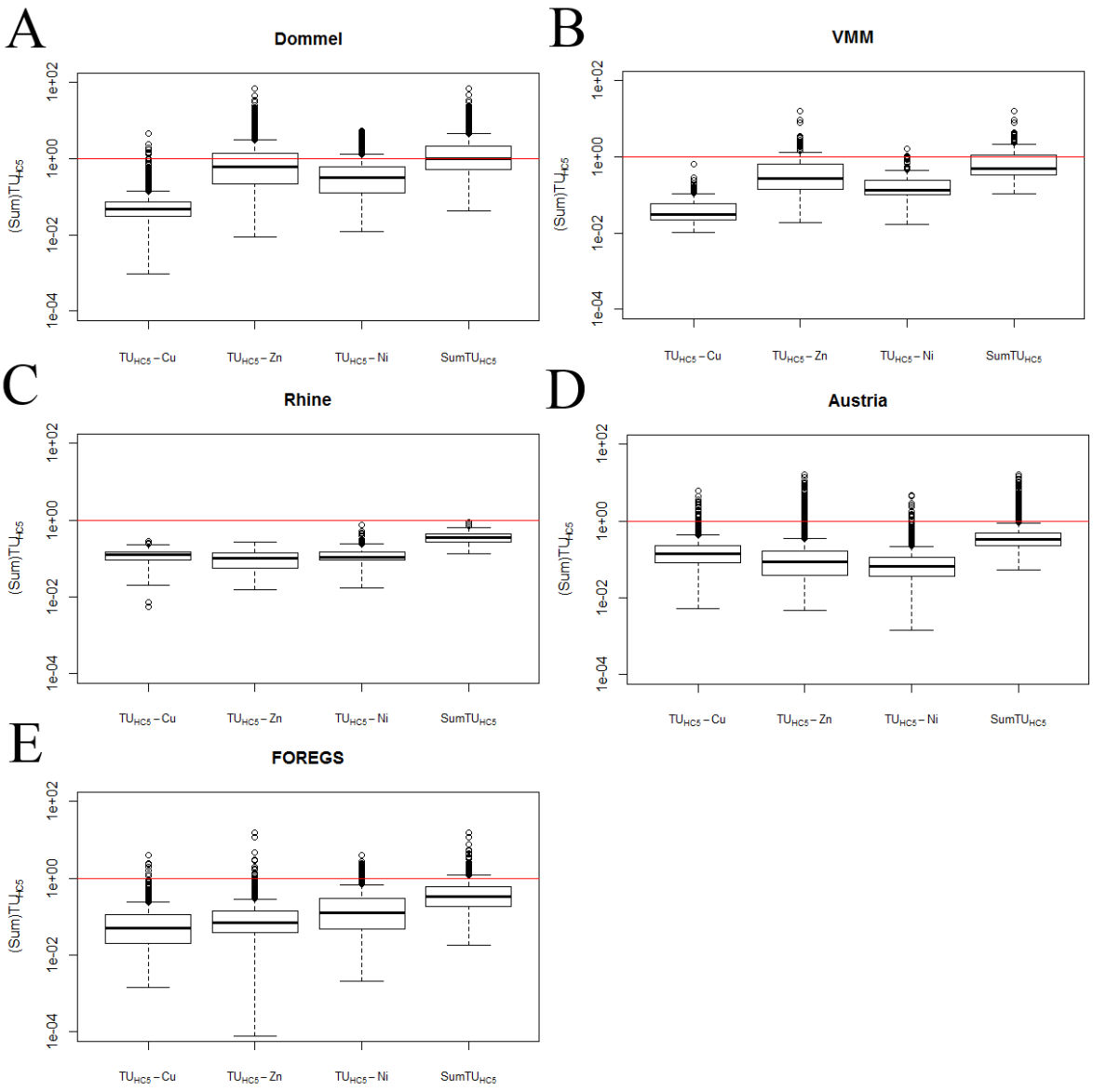


Figure 4.



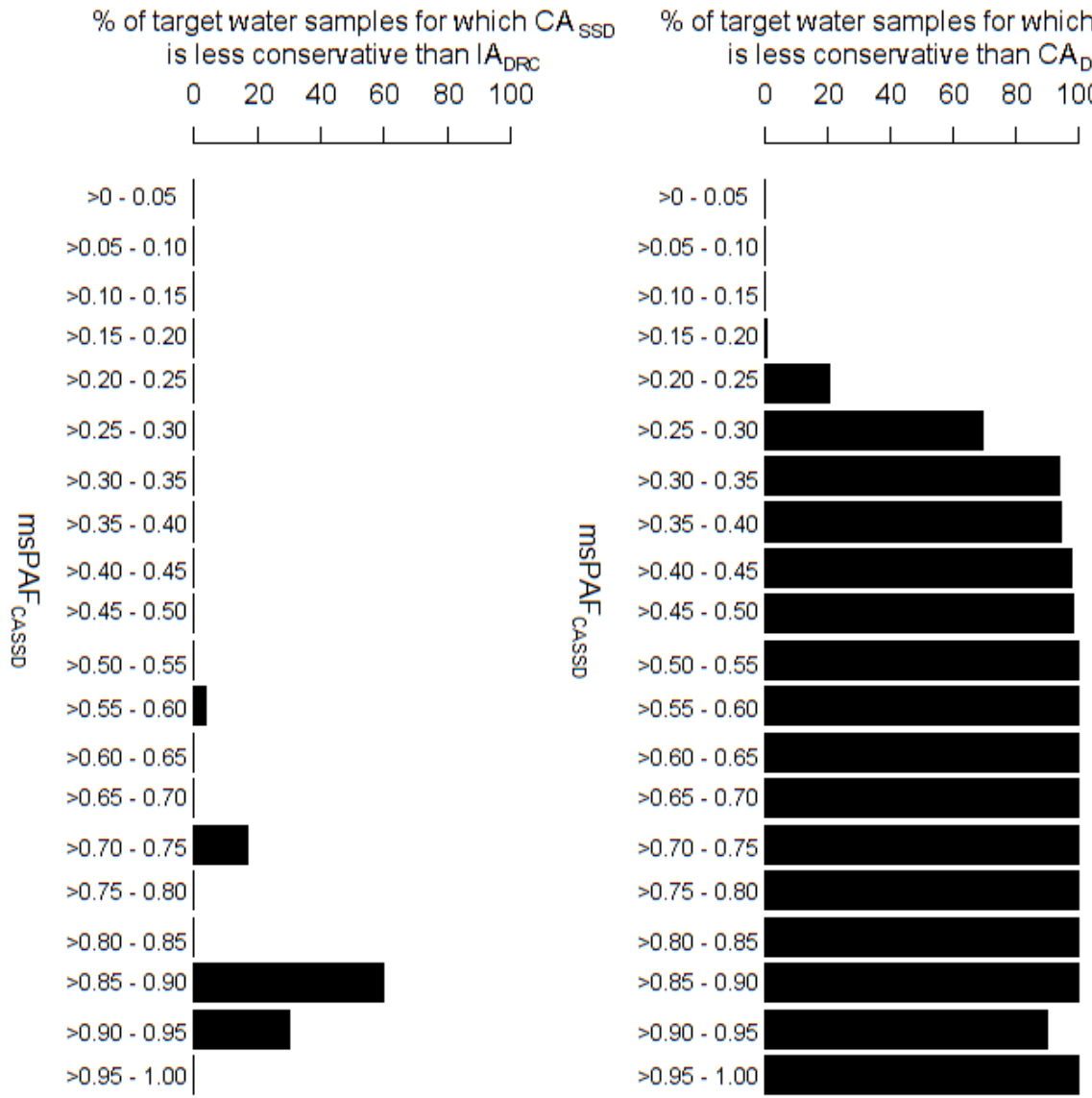


Figure 5.

Figure 6.

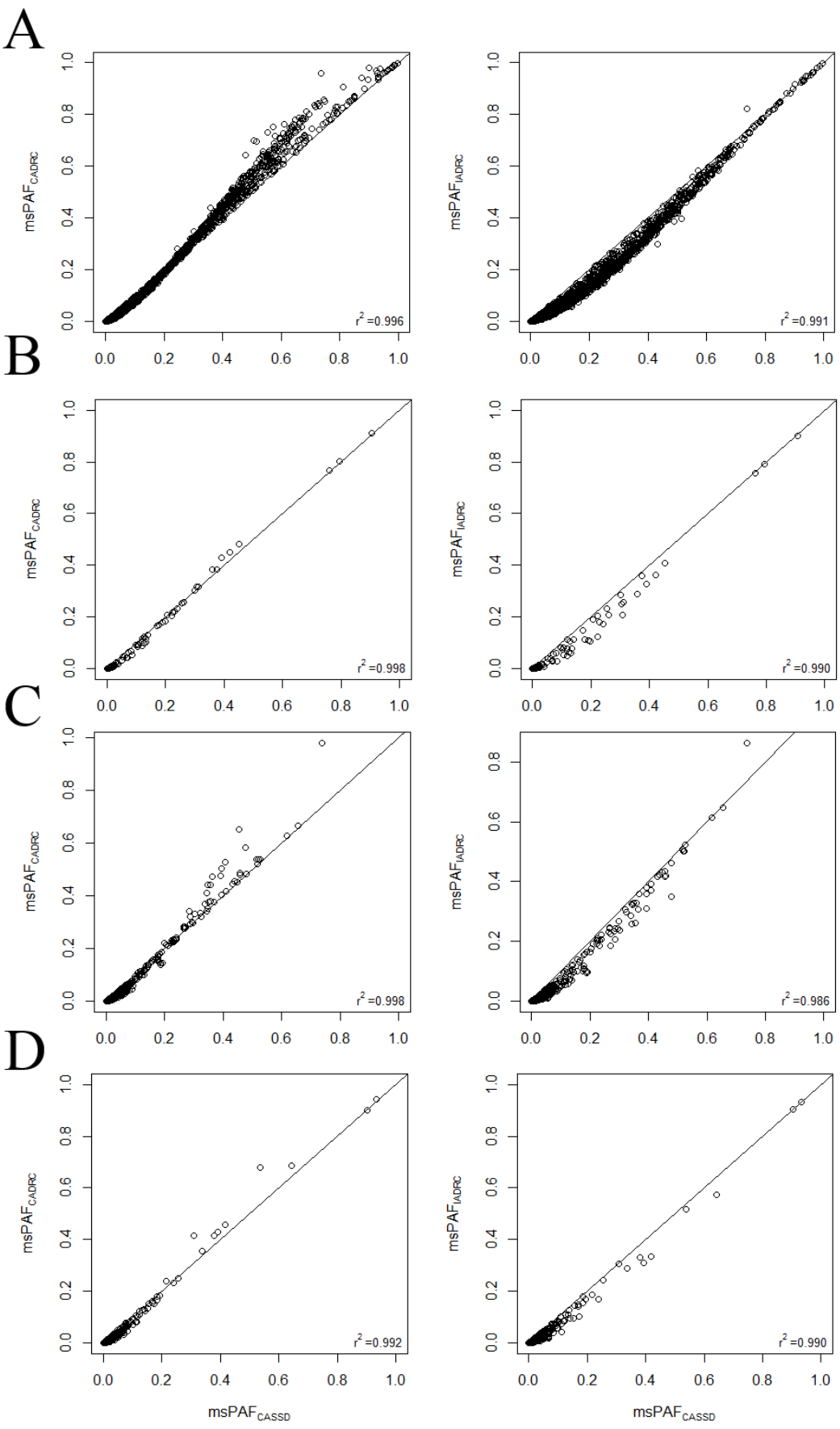


Figure 7.

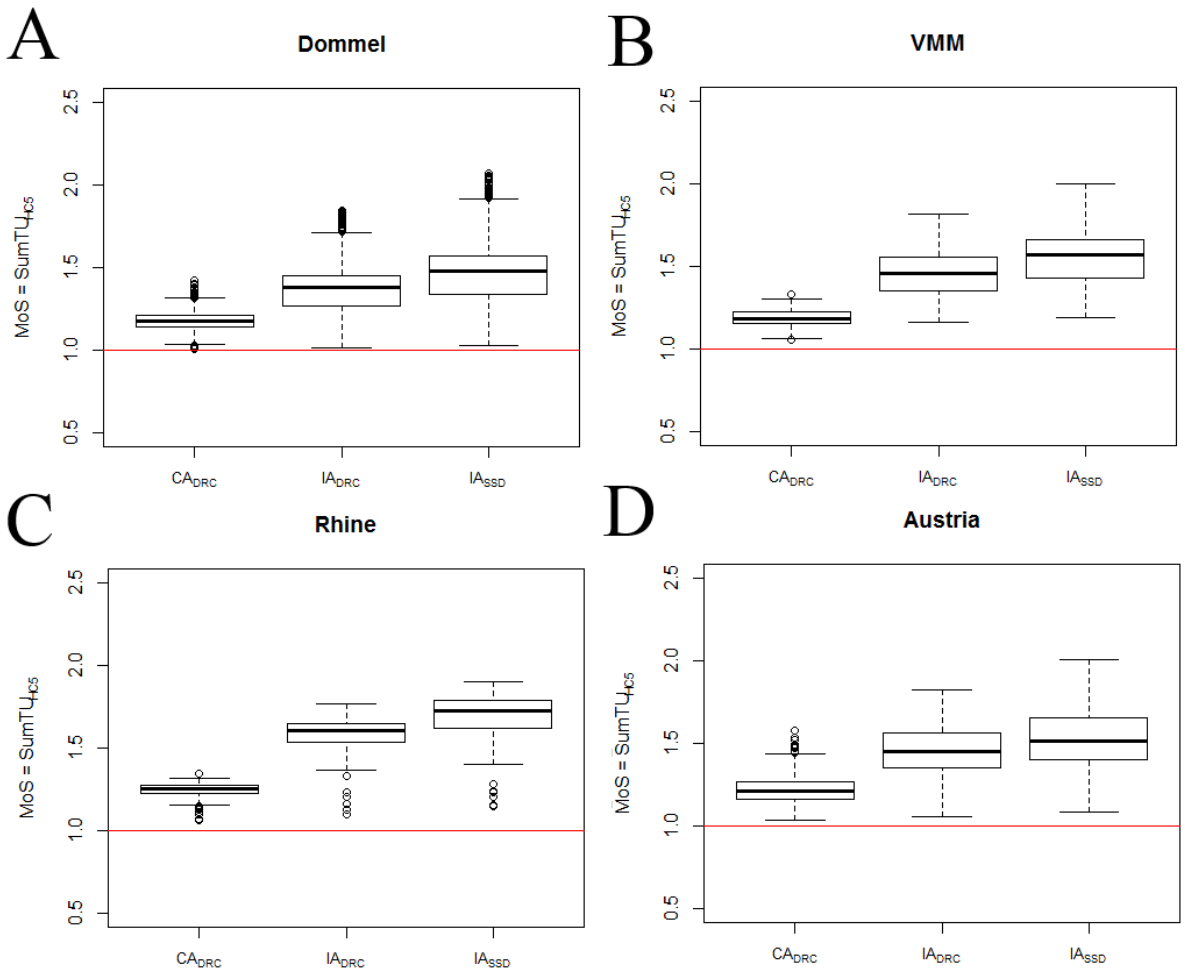


Figure 8.