Simple equations for the calculation of free metal ion activities in natural surface waters
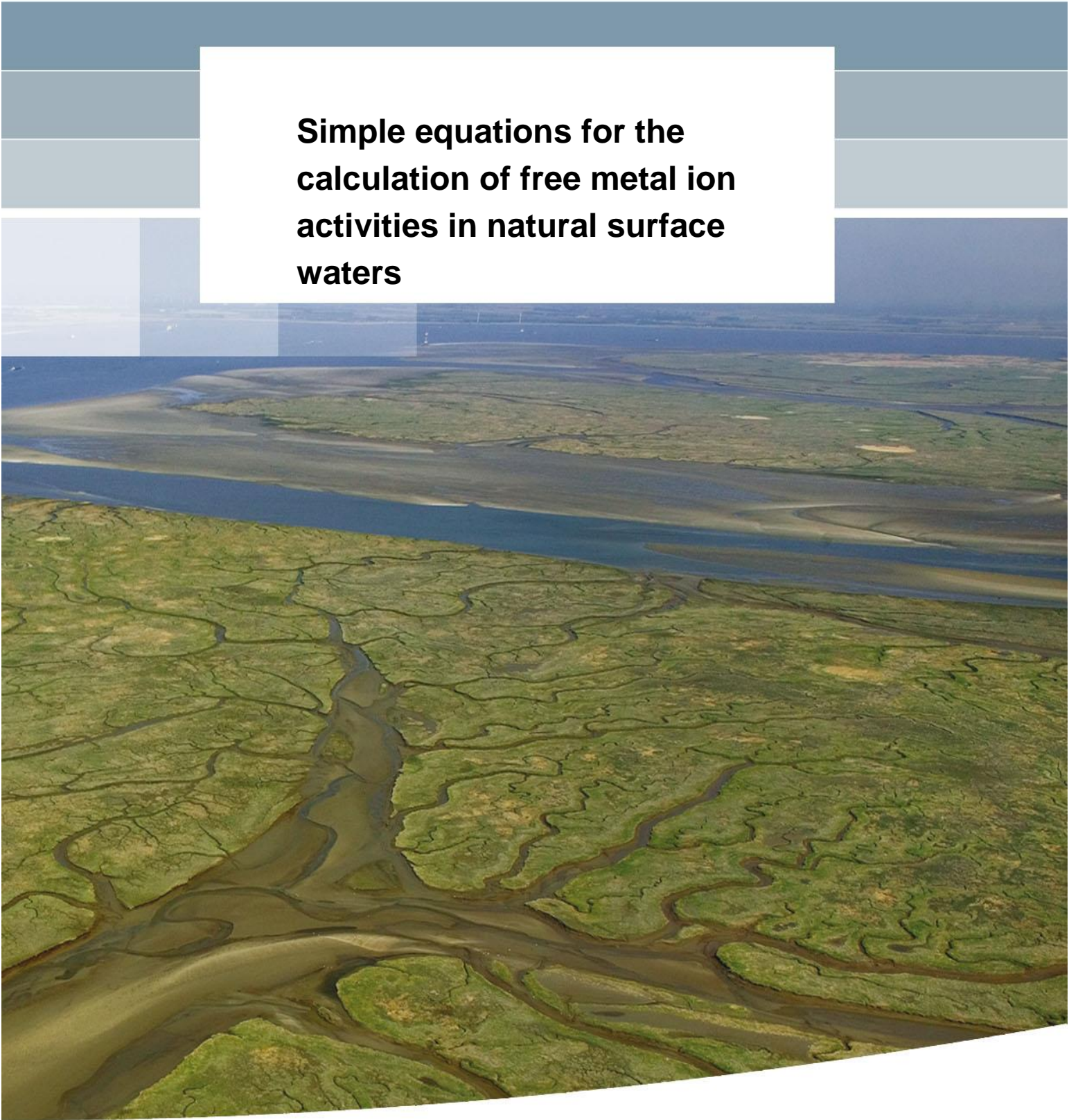
Simple equations for the calculation of free metal ion activities in natural surface waters

H.P. Bootsma
J.P.M. Vink

1210758-000

Deltares

Title
Simple equations for the calculation of free metal ion activities in natural surface waters
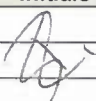
Summary

STOWA, the Dutch foundation for applied research and water management, introduced the concept of ecological key factors to determine the ecological status of a surface water system. One of these key factors is toxicity. A tool was developed that translates concentrations of multiple chemicals into a single value for toxic pressure. This tool needs free metal ion activities as an input variable, instead of total dissolved concentrations. This report presents easy-to-use equations in order to predict free metal ion activities in surface waters. Data from Dutch monitoring programmes were used to derive these equations.

The chemical speciation model WHAM7 was used to calculate free metal ion activities from total dissolved concentrations and common water characteristics, such as DOC, pH, and major cations and anions. Statistical multivariate regression was performed to derive equations that relate the free metal ion activity to commonly measured water characteristics. For 17 metals (Ag, Ba, Be, Cd, Co, Cr, Cu, Hg, La, Mn, Ni, Pb, Sn, Sr, U, V, Zn) relationships were derived using the thermodynamic constants provided by WHAM7 databases.

The predictive value of the equations is large (r^2>0.9 in all cases) and agrees well with the outcome of numerical speciation modelling performed with WHAM7. Therefore, the presented equations can be used for the reliable prediction of free ion concentrations, in order to predict toxic pressure in surface waters.

References
Bootsma, H. & Vink. J.P.M. (2016). Simple equations for the calculation of free metal ion activities in surface waters. Deltares report 1210758, Utrecht, The Netherlands.

Contents

1 Introduction

STOWA recently introduced the concept of ecological key factors to determine the ecological status of a surface water system. This methodology uses a maximum of eight key factors such as light conditions, organic load, and toxicity. The key factor *toxicity* is subdivided into two tracks: track 1 is called "Chemistry" and track 2 "Toxicology". Track 1 provides an estimation of the ecological effects as calculated from the concentration of chemicals in surface water. Track 2 subsequently provides the measured effect of the (unknown) mixture of chemical using bioassays. This report contributes to the Chemistry track, and focusses on the group of (heavy) metals and their speciation in surface waters.

The goal in the Chemistry track is to translate dissolved metal concentrations into predicted effects to aquatic biota. Earlier, a tool was developed to translate the concentrations of multiple compounds into a single value of toxic pressure. For the determination of dissolved heavy metal concentration, sampled water is filtrated over 0.45 µm and the total dissolved concentration is determined. This concentration is the sum of all metal salts, free ions, and metals bound to dissolved organic carbon (DOC). To relate the concentrations with toxic endpoints of bioassays (e.g., growth, reproduction, or mortality), so called bio-available concentrations are required. To this end, it is assumed that the free ion fraction is the dominant fraction that readily available for biologial uptake.

In order to improve the usability of the tool, its predictive performance, and its implementation as a risk indicator, the derivation of relatively simple equations that predict free ion concentrations is a primary condition. Such equations must meet three criteria:
1) Representatively is guaranteed via the use of a large number of data;
2) Predictive performance must be as high as statistically possible;
3) The number of input parameters is limited, and restricted to the common parameters measured in monitoring surveys.

2 Model description

The Windermere Humic Aqueous Model (WHAM; Tipping, 1994) is a numerical model that calculates the competitive reactions of protons and metals with natural organic matter in soils and water. In short, the primary two parts of the model consists of a chemical speciation model, and a humic ion binding model based on the NICA Donan concept (Kinninburgh et al., 1996). Speciation of metal salts and precipitates such as M-chlorides, M-carbonates, and M-sulphates, is calculated by thermodynamic equilibrium constants and is in fact straightforward. Solid-phase metal oxides are not included in the calculations, although WHAM7 provides this option. The rationale behind this is that solid phase concentrations are not desired parameters (and mostly not measured) to include in the simplified equations.

Association and dissociation with organic compounds is more complex, and dominant for some metals. Therefore, the next section describes these mechanisms in more detail.

For binding of metals to humic and fulvic acids, WHAM7 contains the Humic ion-binding model VII (Tipping, 1998; Tipping et al., 2011). Humic acid molecules are represented by homogeneous size spheres with a radius r. The humic acid molecule has a certain molecular mass M with a discrete number of heterogeneous binding sites n_A (sites per g), see figure 2.1.
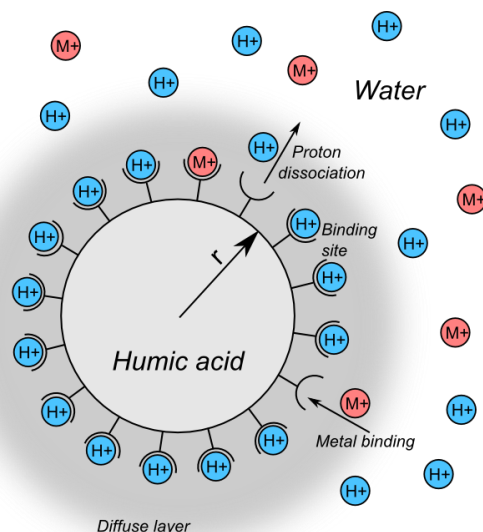

Figure 2.1. WHAM representation of proton and metal interaction with humic acid. Fulvic acid is schematized in analogy.

Binding to these sites is quantified by equilibrium constants. The proton dissociation reaction can be written as:

$$RH^Z = R^{Z-1} + H^+ \qquad (2.1)$$

where:
- R is the humic molecule
- Z is the net charge

The equilibrium quotient is defined as:

$$K_H(Z) = \frac{\left[R^{Z-1} \right] a_{H^+}}{\left[RH^Z \right]}$$

(2.2)

where:

- $K_H(Z)$ is the humic charge dependent quotient
- a_{H^+} is proton activity

Its value is given by:

$$K_H(Z) = K_H \exp(2P \cdot \log_{10}(I) \cdot Z)$$

(2.3)

where:

- K_H is the intrinsic dissociation constant
- P is an empirical constant for electrostatic interaction
- I is ionic strength

Similarly, binding of a metal ion is described using:

$$R^Z + M^z = RM^{Z+z}; \qquad K_H(Z) = \frac{\left[RM^{Z+z} \right]}{\left[R^Z \right] a_M} = K_M \exp(2P \cdot \log_{10}(I) \cdot Z)$$

(2.4)

where:

- M^z is the metal ion
- a_M is metal ion activity
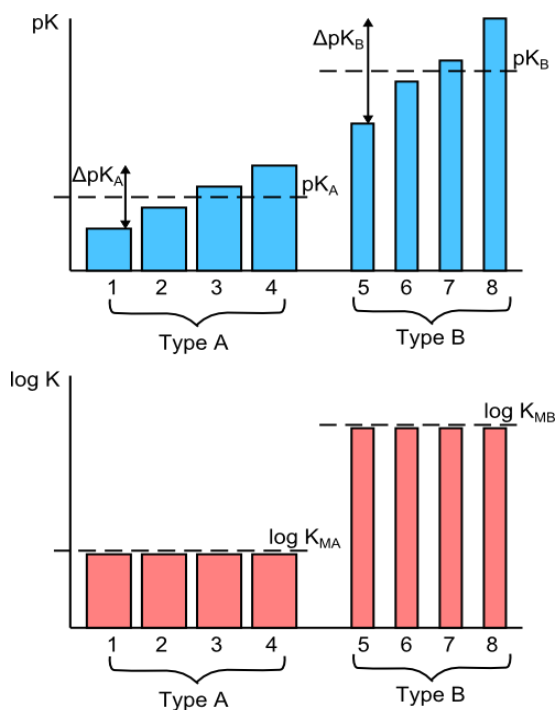- K_M is the intrinsic association constant for the metal ion

Figure 2.2. Representation of equilibrium constants for binding in WHAM7 of protons (top figure) and metals (bottom figure) to organic compounds. pK = -log K.

Eight types of reactive sites are described, divided into types A and B (hence, there are four groups per type). Each group has a specific equilibrium constant. Type A groups represent relatively strong acids, such as carboxylic acids, and type B groups represent weaker acids such as phenolic acids. Per type, two parameters are required for proton dissociation: an intrinsic constant (pK_A, pK_B) and a distribution term (ΔpK_A, ΔpK_B). The model assumes twice as many type A sites as type B sites ($n_B = 0.5 \, n_A$). Within a type, each group has the same number of sites. For metals, no distribution term is used. See figure 2.2.

Monodentate binding sites may also combine to form bidentate or tridentate binding sites when they are close together, see figure 2.3. The equilibrium constant for a bidentate site composed of monodentate sites i and j is given by combining the values for the monodentate sites and adding an additional term, as:

$$\log K(i, j) = \log K(i) + \log K(j) + x \cdot \Delta LK_2 \tag{2.5}$$
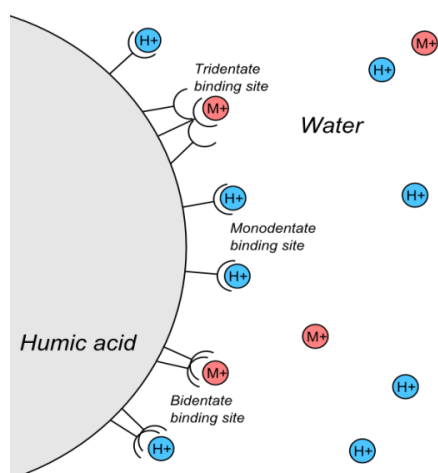

Figure 2.3. Multidentate binding sites in WHAM7.

Similarly, for tridentate sites (i, j, k):
$$\log K(i, j, k) = \log K(i) + \log K(j) + \log K(k) + y \cdot \Delta LK_2 \tag{2.6}$$

where:
- x, y are site-dependent sorption constants.
- ΔLK_2 is the additional increase by multidentate binding for the metal ion.

Multiple values of x and y make it possible to generate a range of multidentate binding strengths. The fraction of total binding sites that from bidentate and tridentate binding sites are denoted by fprB and fprT, respectively.

In conclusion, WHAM7 has nine input parameters specific to humic acid: nA, pKA, PKB, ΔPKA, ΔPKB, P, fprB, fprT, M, and r. To describe fulvic acid, different parameter values are used. Additionally, for each metal two input parameters are required: log KMA, log KMB, and ΔLK2. These values were fitted from experimental data, calculated from geometry or estimated from literature.

Simple equations for the calculation of free metal ion activities in natural surface waters

3 Methodology

Given the criteria described in chapter 1, the simplified equations should predict the free ion activities using only a limited amount of surface water characteristics. This means, in terms of WHAM7 inputs, that the following model assumptions are made:

1. Suspended solid matter (SPM) concentration is zero;
2. No particulate phases are present and all dissolved organic matter is assumed to be colloidal. Total DOC is assumed to consist half of humic acid, and half of fulvic acid;
3. Iron- and aluminium (oxy)hydroxide concentrations are 0, thereby excluding precipitates from solution.

An example input file is shown in appendix C.

3.1 Dataset

To perform multivariate analyses that would lead to reliable multiparameter equations, part of the methodology of Verschoor et al. (2012) was used. First, a large dataset was constructed containing realistic ranges of water characteristics in such a way that the set ensures good coverage of all possible combinations of selected water variables. These included pH, DOC, Ca, Na, Mg, Cl, SO_4, and CO_3. Lower and upper values were extracted from the dataset of Verschoor et al. (2012). Between the lower and upper bounds, a number of evenly spaced values have been generated for the factors. A dataset of "samples" is created using these values, all possible combinations (every single combination of values is a sample).

For total metal concentrations, the upper and lower bounds for the heavy metals were based on the total heavy metal concentrations measured in the year 2013, available on the Waterkwaliteitsportaal (Informatiehuis Water, 2015). For each metal, the 5th percentile value was used as the lower value, the 95th percentile for the upper value. The total concentrations of metals are assumed to be uncorrelated with the other factors: with six values between the upper and lower bound, a total of 19980 samples is used for each metal.

The methodology of data-construction may create unrealistic combinations. In reality, samples are electroneutral and some characteristics are correlated (e.g., pH and HCO_3^-). To remove unrealistic combinations, all samples that exceed electroneutrality by 10% are discarded from the dataset. This procedure is described in appendix A. The resulting dataset features 3330 samples.

3.2 Data handling and multivariate analyses

The constructed dataset was used as an input file for WHAM7. Free ion activities are calculated for each data record with all possible combinations. Multivariate regression was performed in R (R Core Team, 2015) to find relationships between the the total dissolved (measured) metal concentration, the water characteristics and the calculated free ion concentrations. A detailed elaboration on the multivariate analyses is given in Appendix A.

Table 1 shows the values from the dataset that were included in the WHAM7 calculations. The metals of interest were modelled separately, so multimetal interactions were disregarded. Metals are present in very low concentrations relative to the other variables. Consequently, we assumed their role in influencing the behaviour of other metals to be negligible; no site-competition occurs between metals. As explained earlier, metals from table 1 were included because they have been given the highest priority within the Chemistry track of the EFS tool, and because parameter values are present for those metals in the datasets of WHAM[1].

Table 1. Selected values for the WHAM7 calculations. Colloidal DOC is assumed to consist half of humic acid, and half of fulvic acid so that humic acid concentration = fulvic acid concentration = 0.5*DOC.

Parameter	Unit	Selected values
Temperature	C	10
pCO2	ppm	390
pH		6, 6.5, 7, 7.5, 8, 8.5
DOC	mg/L	1, 6.8, 12.6, 18.4, 24.2, 30
Na	µg/L	10000, 33000, 56000, 79000, 102000, 125000
Mg	µg/L	5000, 9000, 13000, 17000, 21000, 25000
Ca	µg/L	25000, 50000, 75000, 1e+05, 125000, 150000
Cl	µg/L	48000, 86000, 10000, 124000, 162000, 2e+05
SO_4	µg/L	10000, 38000, 66000, 94000, 122000
CO_3	µg/L	10000, 98000, 186000, 274000
Ag	µg/L	0.02, 0.316, 0.612, 0.908, 1.2, 1.5
Ba	µg/L	11, 32.8, 54.6, 76.4, 98.2, 120
Be	µg/L	0.02, 0.07, 0.12, 0.17, 0.22, 0.27
Cd	µg/L	0.04, 0.178, 0.316, 0.454, 0.592, 0.73
Co	µg/L	0.1, 0.699, 1.3, 1.9, 2.5, 3.1
Cr	µg/L	0.52, 1.14, 1.75, 2.37, 2.98, 3.60
Cu	µg/L	0.915, 2.03, 3.15, 4.27, 5.38, 6.5
Hg	µg/L	0.02, 0.044, 0.068, 0.092, 0.116, 0.14
La	µg/L	4.94, 16.2, 27.5, 38.8, 50.2, 61.5
Mn	µg/L	20, 138, 256, 374, 492, 610
Ni	µg/L	1.1, 5.08, 9.06, 13, 17, 21
Pb	µg/L	0.2, 1.02, 1.84, 2.66, 3.48, 4.3
Sn	µg/L	0.060, 0.167, 0.274, 0.381, 0.488, 0.595
Sr	µg/L	76, 197, 318, 438, 559, 680
U	µg/L	0.300, 0.869, 1.440, 2.010, 2.580, 3.150
V	µg/L	0.69, 1.85, 3.02, 4.18, 5.34, 6.51
Zn	µg/L	3.3, 16.8, 30.4, 43.9, 57.5, 71

---

[1] *Uranium (U) and vanadium (V) are only available in WHAM7 as oxide forms, UO_2 and VO. To include these elements in the analyses, it is assumed that all U and V is present as UO_2 and VO. However, for consistency with the other metals, the derived equations predict free metal ion concentrations (in µg/L) as if U and V were present as U(aq) and V(aq).*

Deltares

For most metals, linear relationships proved to be insufficient to provide a satisfactory fit. Instead, we attempted to describe binding as pH-dependent Freundlich isotherms:

$$Q = K \cdot Me_{total}^{n} \cdot pH^{b} \tag{3.1}$$

where:
- K is the empirical Freundlich constant
- Me_{total} is the total metal concentration
- n, b are empirical exponents

Linear regression was used after logarithmic transformation:

$$\log Q = \log K + n \log Me_{total} + b \log pH \tag{3.2}$$

However, inorganic complexation does not follow the same relationships, given the mono- and multidentate binding sites. Consequently, the fraction bound or complexed metal fraction could not be described in a similar way. Therefore, the conventional shape of the Freundlich isotherm was abandoned. Free ion activities were derived empirically as descriptors from the water characteristics described before.

In this way, free ion activities were described via power law functions:

$$\left(Me^{z+} \right) = K \cdot Me_{total}^{n} \cdot pH^{b} \dots etc. \tag{3.3}$$

Log transformation gives:

$$\log(Me^{z+}) = \log(K) + n \cdot \log(Me_{total}) + b \cdot pH \dots etc. \tag{3.4}$$

Log refers to logarithm with base 10.

One issue that arises by applying logarithmic forms of the equations is that an optimal fit is found using residuals of the logtransformed data. Effectively, this heavily discounts the residuals at the higher levels. For the purpose of estimating toxicity this discounting is undesirable: high (toxic) values are the most important. Ideally, regression is therefore performed without log transformation, necessitating non-linear regression. However, the non-linear regression algorithms have some downsides: they have to be provided with starting values or they will often not successfully find a fit (Elzhov et al., 2015; R Core Team, 2015). This also increases the difficulty of factor selection. Consequently, factors were chosen and coefficients were determined as follows:

1. A linear model is fitted, using all available factors.
2. Factors are eliminated from this linear model using stepwise regression.
3. The linear model equation is rewritten into a non-linear form.
4. Nonlinear regression is used to re-determine the coefficients, using the coefficients of the reduced linear model as starting point.

Stepwise regression was performed using the stepAIC function (Venables & Ripley, 2002). Factors are included or excluded from the model equation in order of significance to the free ion activity. The Aikaike information criterion (AIC) (Aikaike, 1981) is used to determine whether model performance increases or decreases. The AIC is defined as:

$$AIC = p \cdot k - 2\ln(L) \tag{3.5}$$

where:

- p is a penalty factor for the number of factors in the model equation
- k is the number of factors
- L is the maximized value of the likelihood function for the model

Generally, a high value for P was chosen at values over a 1000 (default P is 2); there was no straightforward way of selecting P, generally a very high value was required to force the exclusion of factors. The goal was to create a good fit for the untransformed equation (3.4) , and a factor that is important for the log transformed equation (3.5) is not necessarily as important for the untransformed equation. For the final model equation, the factors to include were selected based on R^2_{adj} as description of fit and visual inspection; how to reduce the number of factors was informed by the stepwise regression, but not determined by it. Keeping in mind that in some cases (monitoring programmes) no anion concentrations are measured, regression was also performed while excluding anion concentrations (Cl, SO_4, CO_3).

4 Results and discussion

To illustrate the four step method outlined earlier, the following plots show the subsequent regression models for beryllium (Be). First, a linear model is produced using all the available factors, and fed into the stepAIC function to reduce the number of factors. The resulting fit is shown in figure 4.1. For log transformed activities, the fit is good. In contrast, the fit for the untransformed activities shows some scatter, especially for higher activities, see figure 4.2.

Rewriting the equation into a nonlinear form (such as equation 3.4) and using nonlinear regression to adjust the coefficients results in much better predictions of the free ion activity at high (more toxic) activities, see figure 4.4. Conversely, for log transformed activities the fit is now a little worse at low values, see figure 4.3. Since high activities are the most important, the coefficients for all metals are based on nonlinear fits.

The (untransformed) free metal ion activities are heavily skewed right, and the data points at lower activities overlap. To give an accurate impression of the distributions of the activities, marginal histograms were included in each plot. Plots for all metals with calculated versus predicted activities are presented in Appendix B.
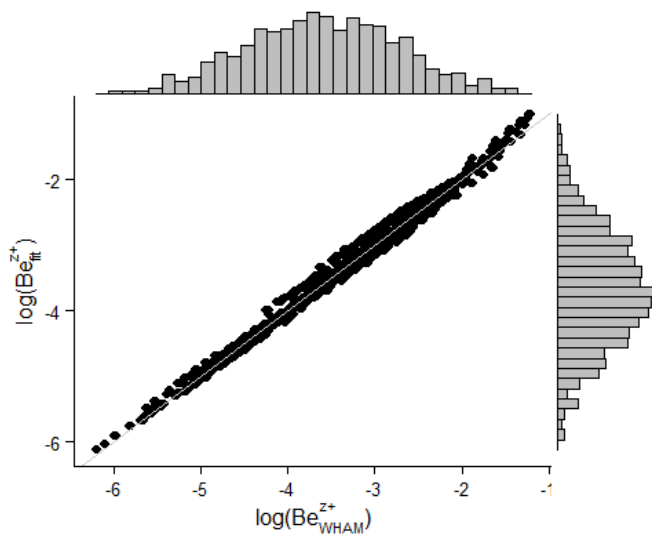

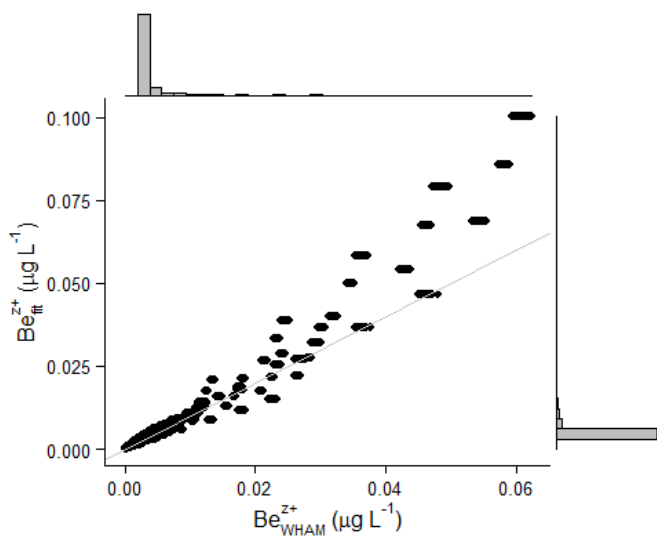*Figure 4.1. WHAM ion activities versus linearly fitted ion activities for Be, log transformed.*

Figure 4.2. WHAM ion activities versus linearly fitted ion activities for Be, untransformed.
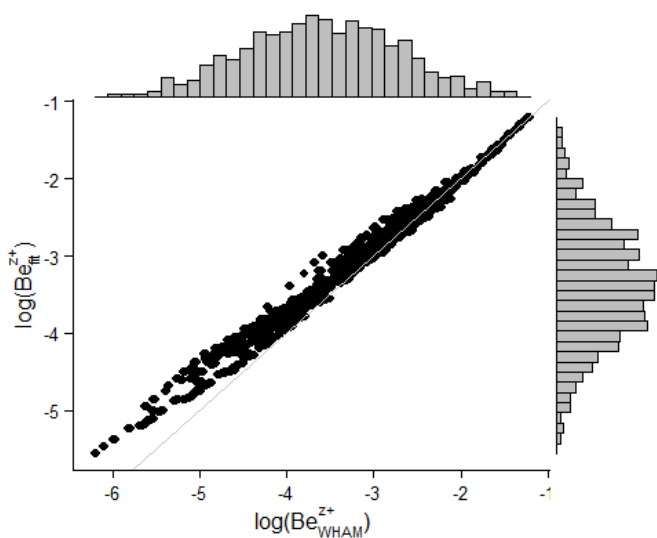

Figure 4.3. WHAM ion activities versus non linearly fitted ion activities for Be, log transformed.
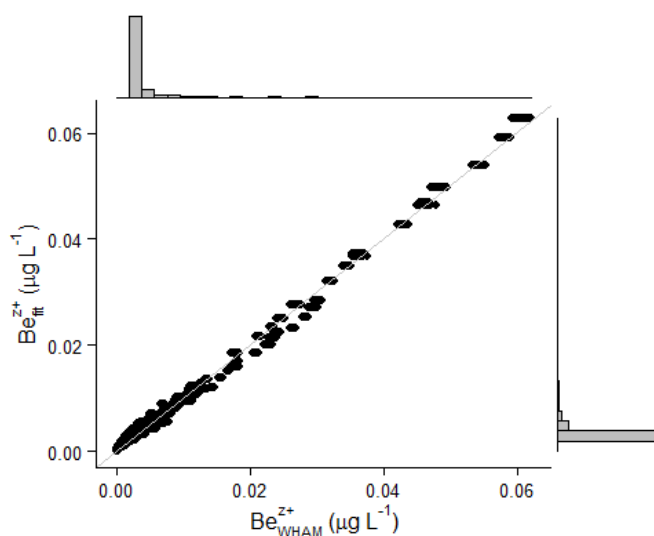
Figure 4.4. WHAM ion activities versus linearly fitted ion activities for Be, untransformed.

Table 2 shows the multivariate models for the free ion metal activities. The values of the coefficients appear plausible. Note however that the coefficients of multivariate models are statistically derived values and do not necessarily have a mechanistic explanation. DOC terms are conventionally negative, as more DOC results in more binding, and less free ions. Ca-relations are generally positive, which is explained by competition over binding sites to DOC.

For barium (Ba) and strontium (Sr), the effect of Ca is negative and binding to DOC does not play a role (no DOC terms are included). Instead, inorganic complexation also binds free ions and higher Ca levels lead to more complexation. This process also occurs for some of the other metals (the effect of Ca is seldom monotonic).

In Freundlich isotherms (such as equation (3.1)), the coefficient n of $\log(Metal_{total})$ is generally smaller than 1: proportionally, binding decreases as activities increase. The adsorbent becomes more saturated and a larger fraction of the total metal concentration is in free ion form. For the equation predicting free ion activities then, we would expect n > 1; tables 2 and 3 show that this is indeed the case. The value for mercury (Hg) is very high (>8), but combined with the values for the pH and DOC coefficients results in extremely low free ion activities; note that the highest value measured in the field in 2013 was 0.14 µg/L, at which the highest free ion activities calculated with WHAM7 is 6.4e-21 µg/L.

Some terms occur twice in the equations, e.g. DOC for silver (Ag). Inclusion of both forms (DOC and log(DOC)) did notably improve the fit in these cases. The relationship between the logarithm of the activity and e.g. DOC is apparently not (statistically) characterized well by either the linear term (DOC) or the logarithmic one (log(DOC)) for the conditions fed to WHAM7. The R^2 values indicate a good fit. This is also visible when fitted ion activities are plotted against the original activities as calculated by WHAM7. For detailed information on all compounds see the figures in appendix B.

Table 2. Equations for free ion activities, <u>anions included.</u> All variables in µg/l except DOC (mg/l).

Metal activity (M^{n+}) in µg/l	Model description	R^2
$\log(Ag_{ion})$	$= 2.16 + 1.07 * \log(Ag_{total}) -0.0441 * \log(DOC) -0.00925 * DOC -0.618 * \log(Cl)$	0.963
$\log(Ba_{ion})$	$= 0.291 + 1 * \log(Ba_{total}) -0.0923 * \log(Ca) -8.60e-07 * SO_4$	0.998
$\log(Be_{ion})$	$= 2.75 + 1.13 * \log(Be_{total}) -0.868 * \log(DOC) + 0.219 * \log(Ca) -0.733 * pH$	0.992
$\log(Cd_{ion})$	$= -0.00086 + 1.08 * \log(Cd_{total}) -0.211 * \log(DOC) + 0.0563 * \log(Ca) -0.0816 * pH$ $- 0.0195 * DOC$	0.985
$\log(Co_{ion})$	$= 0.843 + 1.01 * \log(Co_{total}) -0.161 * pH -0.00418 * DOC -1.22e-06 * CO_3$	0.969
$\log(Cr_{ion})$	$= -0.928 + 2.18 * \log(Cr_{total}) -2.32 * \log(DOC) + 0.651 * \log(Ca) -1.29 * pH$	1.000
$\log(Cu_{ion})$	$= -1.32 + 1.81 * \log(Cu_{total}) -2.05 * \log(DOC) + 0.380 * \log(Ca) -0.349 * pH$ $-3.78e-06 * CO_3$	0.991
$\log(Hg_{ion})$	$= -1.27 + 8.47 * \log(Hg_{total}) -2.99 * \log(DOC) -1.95 * pH$	0.998
$\log(La_{ion})$	$= -3.93 + 2.50 * \log(La_{total}) -1.98 * \log(DOC) + 0.529 * \log(Ca) -0.456 * pH$	0.992
$\log(Mn_{ion})$	$= 0.168 + 1.07 * \log(Mn_{total}) -0.0256 * \log(DOC) + 0.0117 * \log(Ca) -0.0928 * pH$ $-0.0102 * DOC$	0.986
$\log(Ni_{ion})$	$= 0.292 + 1.08 * \log(Ni_{total}) + 0.142 * \log(Ca) -0.198 * pH -0.0103 * DOC$ $-1.73e-06 * CO_3$	0.966
$\log(Pb_{ion})$	$= -0.551 + 1.18 * \log(Pb_{total}) -0.989 * \log(DOC) + 0.537 * \log(Ca) -0.469 * pH$ $-1.47e-06 * CO_3$	0.967
$\log(Sn_{ion})$	$= -0.44 + 2.04 * \log(Sn_{total}) -1.96 * \log(DOC) + 0.349 * \log(Ca) -1.31 * pH$	1.000
$\log(Sr_{ion})$	$= 0.473 + 1.00 * \log(Sr_{total}) -0.133 * \log(Ca)$	0.995
$\log(U_{ion})$	$= -0.044 + 2.01 * \log(U_{total}) -1.85 * \log(DOC) -0.985 * pH + 0.597 * \log(Ca)$ $-1.52e-06 * CO_3$	0.971
$\log(V_{ion})$	$= -0.619 + 2.13 * \log(V_{total}) -1.96 * \log(DOC) + 0.493 * \log(Ca) -0.726 * pH$	0.994
$\log(Zn_{ion})$	$= -0.229 + 1.0963 * \log(Zn_{total}) -0.0508 * \log(DOC) + 0.126 * \log(Ca) -0.116 * pH$ $-0.0130 * DOC -9.22e-07 * CO_3$	0.983

Ag = Silver; Ba = Barium; Be = Beryllium; Cd = Cadmium; Co = Cobalt; Cr = Chromium; Cu = Copper;
Hg = Mercury; La = Lanthanum; Mn = Manganese; Ni = Nickel; Pb = Lead; Sn = Tin; Sr = Strontium;
U = Uranium; V = Vanadium; Zn = Zinc.

After omitting anion concentrations from the multivariate analyses, the equations alter in stoichiometry: both water variables and model parameter values change. However, general performances of the equations do not decline significantly. The effect of including anions in the equations is therefore somewhat limited. It should be noted however that this is valid for the ranges of anion concentrations mentioned in Table 1. Outside this range, lower performances of the equations may be expected.

An exception is found with silver (Ag). The goodness of fit drops notably from 0.963 to 0.702. The explanation is found in the complexation with Cl, which plays an important role in Ag-speciation. Na and Mg concentrations are correlated with Cl and were included to improve the fit.

Table 3. Equations for free ion activities, <u>anions excluded</u>. All variables in µg/l except DOC (mg/l).

Metal activity (M^{2+}) in µg/l	Model description	R^2
$\log(Ag_{ion})$	$= -0.473 + 1.06 * \log(Ag_{total}) -0.0364 * \log(DOC) -0.00804 * DOC -1.85e\text{-}05 * Mg$ $-2.20e\text{-}06 * Na$	0.702
$\log(Ba_{ion})$	$= 0.493 + 1 * \log(Ba_{total}) -0.142 * \log(Ca)$	0.992
$\log(Be_{ion})$	$= 2.75 + 1.13 * \log(Be_{total}) -0.868 * \log(DOC) + 0.219 * \log(Ca) -0.733 * pH$	0.992
$\log(Cd_{ion})$	$= -0.000856 + 1.08 * \log(Cd_{total}) -0.211 * \log(DOC) + 0.0563 * \log(Ca) -0.0816 * pH$ $-0.0195 * DOC$	0.985
$\log(Co_{ion})$	$= 1.08 + 1.01 * \log(Co_{total}) -0.192 * pH -0.00416 * DOC -1.38e\text{-}06 * Ca$	0.942
$\log(Cr_{ion})$	$= -0.928 + 2.18 * \log(Cr_{total}) -2.32 * \log(DOC) + 0.651 * \log(Ca) -1.29 * pH$	1.000
$\log(Cu_{ion})$	$= -7.33 + 1.791 * \log(Cu_{total}) -2.04 * \log(DOC) + 1.80 * \log(Ca) -0.336 * pH$ $-1.48e\text{-}05 * Ca$	0.951
$\log(Hg_{ion})$	$= -1.27 + 8.47 * \log(Hg_{total}) -2.99 * \log(DOC) -1.95 * pH$	0.998
$\log(La_{ion})$	$= -3.93 + 2.50 * \log(La_{total}) -1.98 * \log(DOC) + 0.529 * \log(Ca) -0.456 * pH$	0.992
$\log(Mn_{ion})$	$= 0.168 + 1.07 * \log(Mn_{total}) -0.0256 * \log(DOC) + 0.0117 * \log(Ca) -0.0928 * pH$ $-0.0102 * DOC$	0.986
$\log(Ni_{ion})$	$= -2.65 + 1.07 * \log(Ni_{total}) + 0.834 * \log(Ca) -0.191 * pH -0.0102 * DOC$ $-7.07e\text{-}06 * Ca$	0.949
$\log(Pb_{ion})$	$= 0.979 + 1.18 * \log(Pb_{total}) -0.985 * \log(DOC) + 0.265 * \log(Ca) -0.515 * pH$	0.952
$\log(Sn_{ion})$	$= -0.44 + 2.04 * \log(Sn_{total}) -1.96 * \log(DOC) + 0.349 * \log(Ca) -1.31 * pH$	1.000
$\log(Sr_{ion})$	$= 0.473 + 1.00 * \log(Sr_{total}) -0.133 * \log(Ca)$	0.995
$\log(U_{ion})$	$= 1.51 + 2.01 * \log(U_{total}) -1.85 * \log(DOC) -1.04 * pH + 0.33 * \log(Ca)$	0.959
$\log(V_{ion})$	$= -0.619 + 2.13 * \log(V_{total}) -1.96 * \log(DOC) + 0.493 * \log(Ca) -0.726 * pH$	0.994
$\log(Zn_{ion})$	$= 0.758 + 1.10 * \log(Zn_{total}) -0.0509 * \log(DOC) -0.0489 * \log(Ca) -0.147 * pH$ $-0.0129 * DOC$	0.967

As noted earlier, the numerical results of the simplified equations show good correlation with calculated values from the chemical speciation model WHAM7 (see figures 4 and appendix B). In order to provide some insight in the absolute values of free ion activities (fia), and their relative contribution to total dissolved concentrations of each metal, Figures 5.1 and 5.2 were constructed.

In Figure 5.1, free ion activities were calculated using the simplified equations from Table 2, including anions. These calculations serve as an example and are only valid for the composition at which they were computed, namely at pH =7; DOC =5 mg/l; and CO_3 =10 mg/l. Other characteristics were Na=10 mg/l; Mg=5 mg/l; Cl=48 mg/l; SO_4 =10 mg/l. In all cases, the total metal concentration (C-Tot) was 2.5 µg/l.
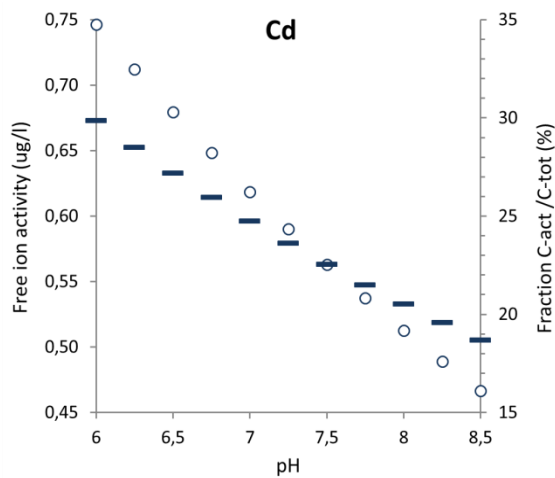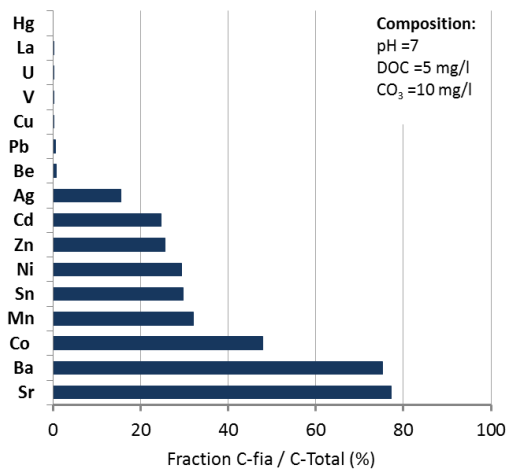
Figure 5.1 (left) Free ion fraction as percentage of total concentration for 17 ranked metals.

Figure 5.2 (right) Calculated Cd free ion activities as a function of pH (o), and the effect on the free ion fraction (–).

Typically, the contribution of free ion activities to total concentrations is below 1% for Hg, La, U, V, Cu, Pb, and Be. These elements are known to be complexed and/or sorbed in colloidal matrices (DOC, salts). Significant amounts of fia occurs for Ag, Cd, Zn, Ni, Sn, Mn, which show a contribution ranging from 15 to 32% (at the water composition described above). Co, Ba, and Sr have relatively high fia-contributions, ranging from 48 to 77%.

Figure 5.2 shows an example of cadmium speciation, in relation to pH. Free ion activities (presented by open circles) drop as a result of increase of pH, resulting from decreased H^+ competition on sorption sites. The fraction C-act/C-Tot (represented by horizontal bars) reacts in analogy.

Very little data exist on the actual measurement of free ion concentrations at different water composition. One of the few studies in which these quantitative data are provided is by Vink (2009). Fia was measured for Cd, Cu, Ni, Pb, and Zn for 6 natural Dutch surface waters and sediment pore waters, and it was concluded that the contribution of the free ion concentration to the total dissolved concentration is relatively large for Cd and Zn, and relatively small for Cu and Pb, which coincides with the findings presented in figure 5.1. However, it is worthwhile to verify, or validate, the simplified equations with actual measurements of fia at various water compositions.

5 Conclusions

Simple and user-friendly equations were derived for 17 metals to predict the activities of heavy metal ions as a function of total dissolved metal concentrations and major, commonly monitored water characteristics. In the majority of cases, DOC and pH are required input parameters.

The predictive performance of the equations are high (R^2>0.9). To improve applicability of the equations for existing and future monitoring data, two sets of the equations are presented: one set including anion concentrations (Cl, SO_4, CO_3), and one set without anions.

Equations were derived for 17 metals for which thermodynamic data were available in the WHAM7 database. For this reason, no equations could be derived for thallium and boron, which are both high priority metals. Of the second priority metals lithium, molybdenum, titanium, tungsten, antimony, and arsenic, no numerical routines are provided by WHAM7. For these metals, alternatives have to be explored, either by fitting parameters to data from literature or by performing targeted speciation measurements for these compounds.

Some form of uncertainty analysis is necessary. The assumption that the results of the modelling routines are completely accurate may a-priori be questioned. Although WHAM7 validation studies were performed to compare calculations with field measurements (e.g. Tipping et al., 2011), this is limited to a few metals. In recent years, various analytical methods to actually measure free ion concentrations in natural waters were developed (e.g., Temminghoff et al., 2000; Vink et al., 2002; Zhang, 2004) and reviewed (e.g., Van Leeuwen et al., 2005; Apte et al., 2005). Comparing the presented simplified equations with such operational methods may be a necessary step towards uncertainty analysis.

6 References

Akaike, H. (1981). Likelihood of a model and information criteria. Journal of econometrics, 16(1), 3–14.

Apte, S.C., Batley, G.E., Bowles, K.C., Brown, P.L. (2005). A comparison of copper speciation measurement with the toxic resposes of three sensitive freshwater organisms. Environ Chem. 2:320-330.

Christensen, J.B., Botma, J.J., Christensen, Th. H. (1999). Complexation of Cu and Pb by DOC in polluted groundwater: a comparison of experimental data and predictions by speciation models WHAM and MINTEQA2. Water Research 33/15:3231-3238.

Elzhov, T. V., Mullen, K. M., Spiess, A.-N., & Bolker, B. (2015). minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in minpack, plus support for bounds [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=minpack.lm (R package version 1.1-9)

Inforfmatiehuis Water (2015). Waterkwaliteitsportaal, Waterkwaliteitsmetingen – Chemie. Retrieved from: https://www.waterkwaliteitsportaal.nl/Beheer/Rapportage/Bulkdata.

Kinniburgh D.G., Milne, C.J., Benedetti, M.F., Pinheiro, J.P., Filius, J., Koopal, L.K., Van Riemsdijk, W.H. (1996). Metal ion binding by humic acid: application of the NICA-Donnan model. Environ Sci Technol 30:1687-1698.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Tipping, E. (1994). WHAM: a chemical equilibrium model and computer code for waters, sediments and soil incorporating a discrete site/electrostatic model of ion-binding by humic substances. Computers & Geosciences 20/6:973-1023.

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(1), 3-47.

Tipping, E., Lofts, S., & Sonke, J. (2011). Humic ion-binding model vii: a revised parameterisation of cation-binding by humic substances. Environmental Chemistry, 8(3), 225–235.

Lofts, S., & Tipping, E. (1998). An assemblage model for cation binding by natural particulate matter. Geochimica et Cosmochimica Acta, 62(15), 2609-2625.

Temminghoff, E., Plette, S., Van Eck, R., Van Riemsdijk, W. (2000) Determination of the chemical speciation of trace metals in aqueous systems by the Wageningen Donnan Membrane Technique. Analytica Chim. Acta 417:149-157

Van Leeuwen, H.P., Town, R.M. Buffle, J., Cleven, R., Davison, W., Puy, J., Van Riemsdijk, W., Sigg, L. (2005). Dynamic speciation analysis and bioavailability of metals in aquatic systems. Environ. Sci. Technol. 22:8545-8556.

Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S (Fourth ed.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4 (ISBN 0-387-95457-0)

Verschoor, A. J., Vink, J. P., & Vijver, M. G. (2012). Simplification of biotic ligand models of Cu, Ni, and Zn by 1-, 2-, and 3-parameter transfer functions. Integrated environmental assessment and management, 8(4), 738–748.

Vink, J.P.M. (2002). Measurement of heavy metal speciation over redox gradients in natural water-sediment interfaces and implications for uptake by benthic organisms. Environ. Sci. Technol. 36:5130-5138.

Vink, J.P.M. (2009). The origin of speciation. Environ. Poll. 157:519-527.

Zhang, H. (2004) In-situ speciation of Ni and Zn in freshwaters: comparison between DGT measurements and speciation models. Environ. Sci. Technol. 38: 1421-1427.
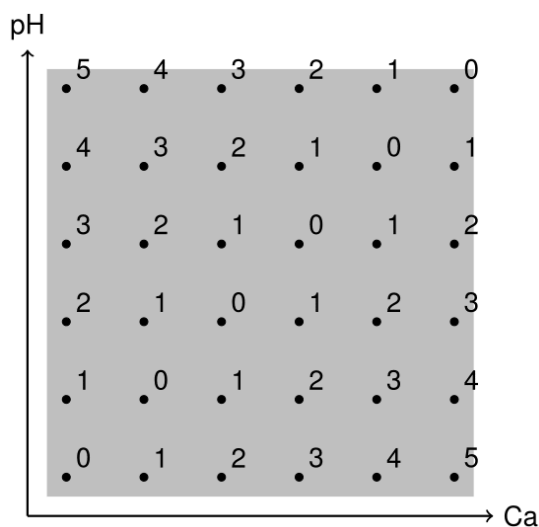
A WHAM input creation

First, all the possible factor levels are combined to create a dataset of `samples'. All samples that exceed electroneutrality by 10% are discarded. To further constrain the dataset to (more) realistic combinations of factor levels, a simple algorithm discards certain samples.

The algorithm selects two parameters. In the next example, Ca concentration and pH are chosen. If all combinations of values are present in the data, correlation is 0 and a plot of the two parameters is as follows:

pH

Ca
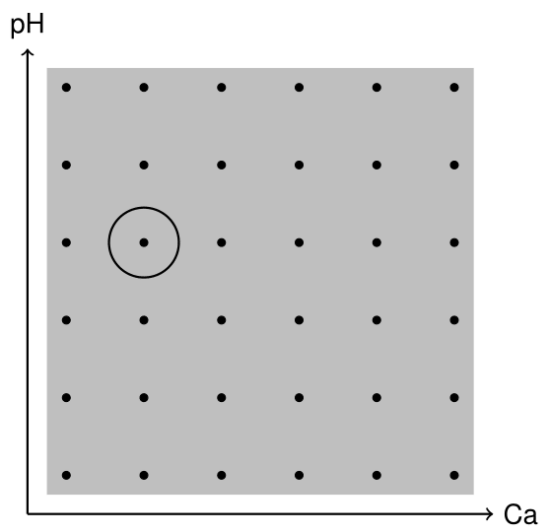
The gray background marks undiscarded points. Note the encircled point. Counting from left to right, its Ca rank is 2. Counting from bottom to top, its pH rank is 4, so that the absolute difference in rank is |4 - 2| = 2. Similarly, the absolute rank difference for each point is calculated:

pH

```
 5    4    3    2    1    0
 •    •    •    •    •    •

 4    3    2    1    0    1
 •    •    •    •    •    •

 3    2    1    0    1    2
 •    •    •    •    •    •

 2    1    0    1    2    3
 •    •    •    •    •    •

 1    0    1    2    3    4
 •    •    •    •    •    •

 0    1    2    3    4    5
 •    •    •    •    •    •
```

Ca

For each point the absolute rank difference is shown. To create positive correlation, the points with the highest absolute rank difference can be discarded, e.g. those points with a difference of 5 or 4:
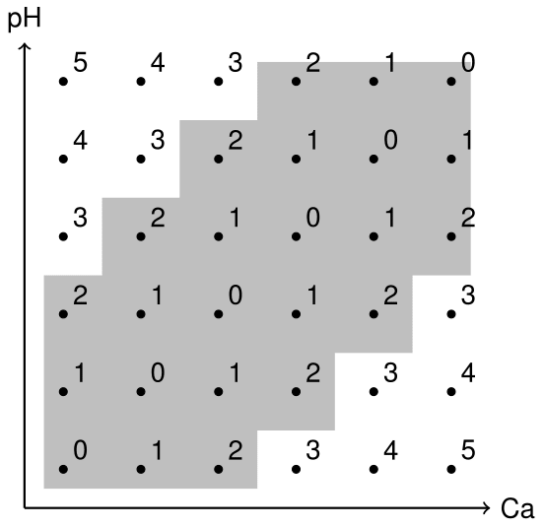
If more correlation is required, the points with a rank difference of e.g. 3 or 2 can be discarded as well:
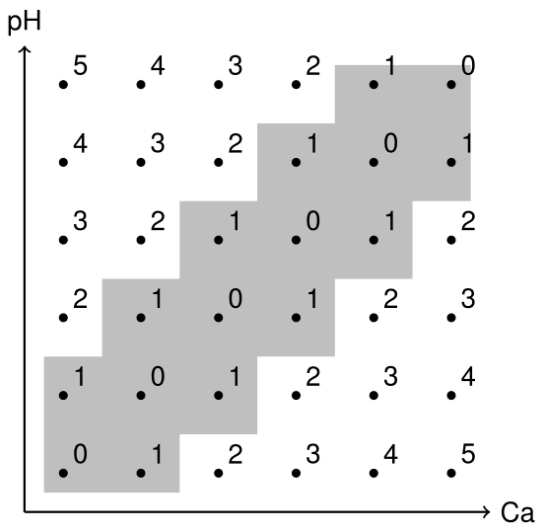
When the correlation goal is reached, the algorithm moves on to the next two factors and repeats the steps until the correlation goal for those two factors is also reached. It then moves to the next two factors, et cetera, until the correlation goal for every combination of two factors has been reached.

By providing some lenient correlation goals for every combination of two factors, correlations of the generated and measured dataset agree to an acceptable level (see table 4).

This method effectively assumes linearity between factors and homoscedasticity. This is not too problematic, keeping the objective in mind: to provide a somewhat constrained input space for WHAM 7. Factor space is also fairly coarsely discretised with only six points.

The reduced dataset contains 3330 samples. The total metal concentrations are assumed to be uncorrelated with the other factors. Choosing six metal concentration levels and combining with the generated dataset results in a dataset with 19980 samples. These conditions are then simulated in WHAM 7.

Table A1 Correlation between factors in dataset of Verschoor et al. (2012), and in the generated dataset.

var1	var2	measured correlation	generated correlation
DOC	pH	-0.1	0.05
DOC	Ca	0.04	0.06
DOC	Mg	0.19	0.07
DOC	Na	0.12	0.06
DOC	Cl	0.11	0.05
DOC	SO_4	-0.09	0.04
DOC	HCO_3	0.08	0.05
pH	Ca	0.59	0.72
pH	Mg	0.45	0.59
pH	Na	0.45	0.54
pH	Cl	0.5	0.55
pH	SO_4	0.03	0.32
pH	HCO_3	0.6	0.66
Ca	Mg	0.67	0.74
Ca	Na	0.47	0.56
Ca	Cl	0.57	0.6
Ca	SO_4	0.26	0.44
Ca	HCO_3	0.92	0.84
Mg	Na	0.71	0.68
Mg	Cl	0.82	0.7
Mg	SO_4	0.52	0.58
Mg	HCO_3	0.61	0.65
Na	Cl	0.92	0.79
Na	SO_4	0.41	0.43
Na	HCO_3	0.55	0.58
Cl	SO_4	0.37	0.31
Cl	HCO_3	0.6	0.54
SO4	HCO_3	0.09	0.21

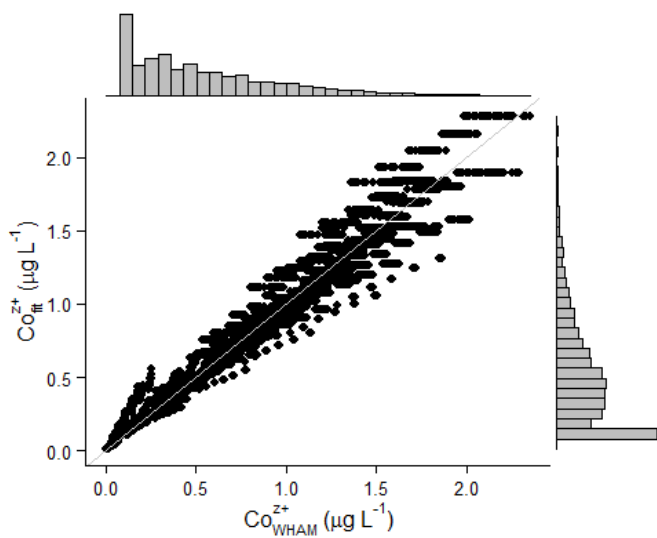# B Comparison of calculated versus predicted activities
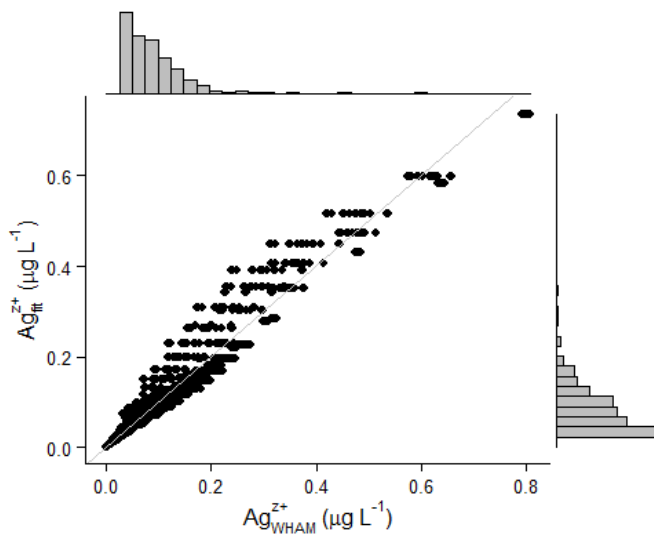
Figure B.1 WHAM ion activities versus non linearly fitted ion activities for Ag.
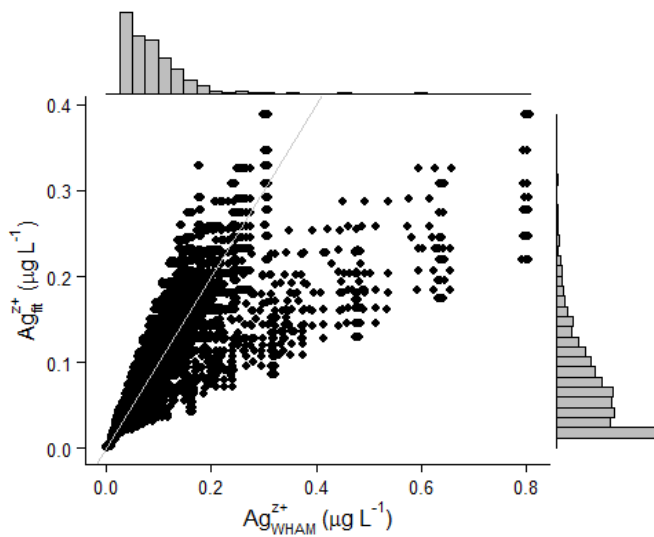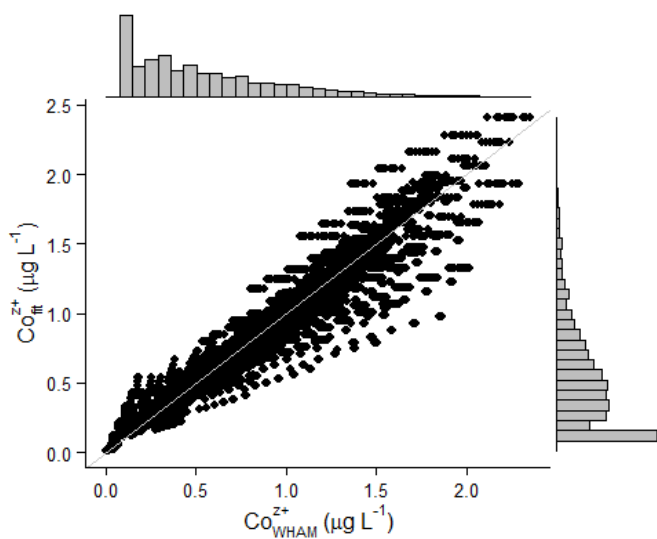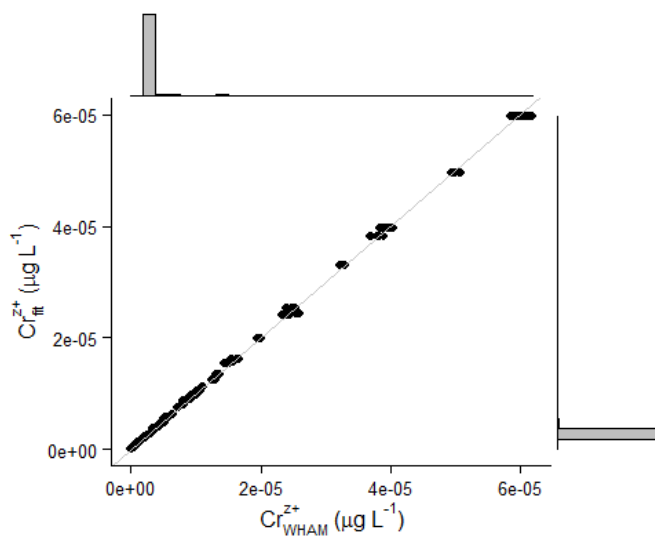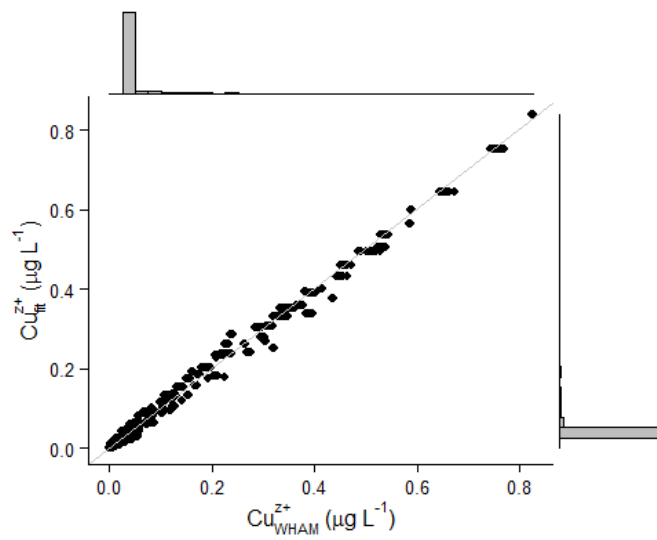

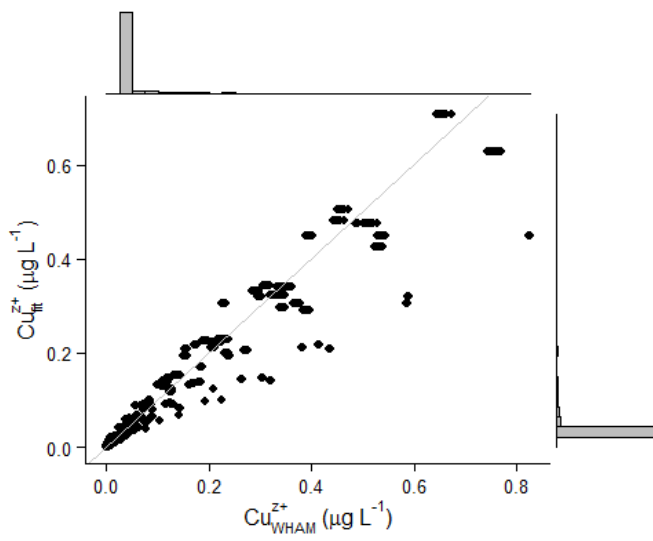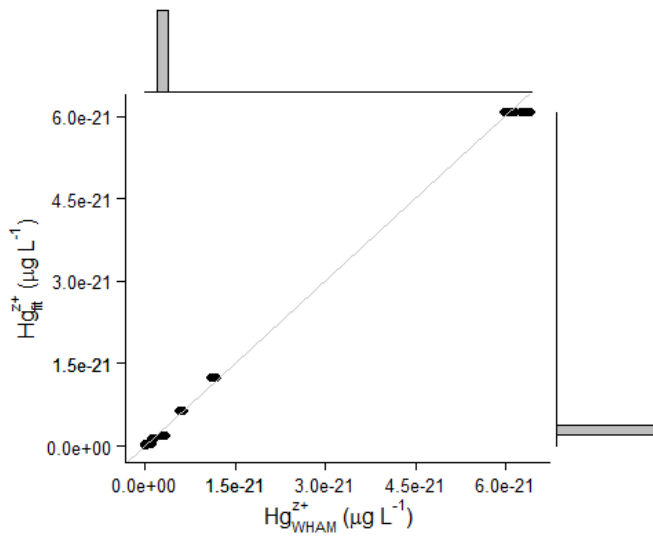*Figure B.2 WHAM ion activities versus non linearly fitted ion activities for Ag, anions excluded.*
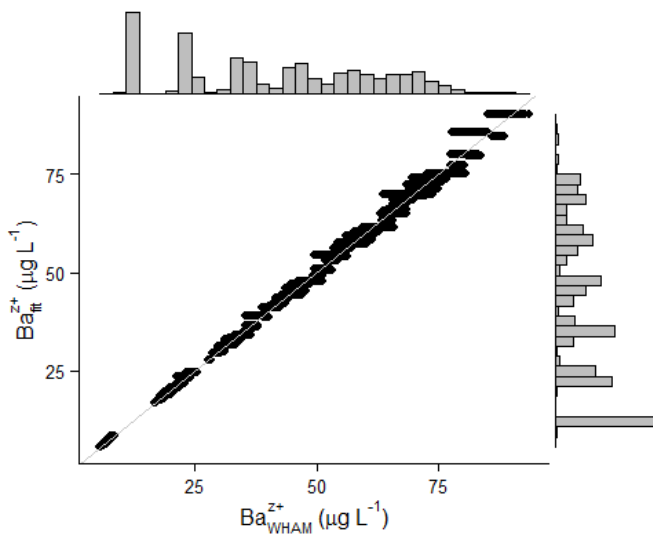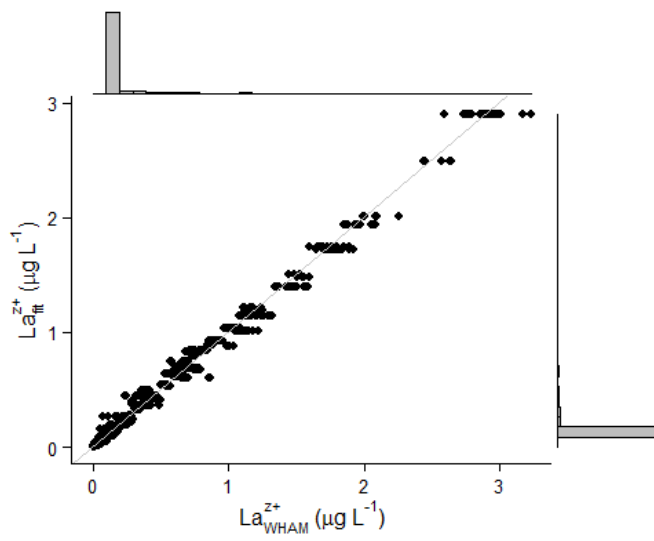
Figure B.3 WHAM ion activities versus non linearly fitted ion activities for Ba.
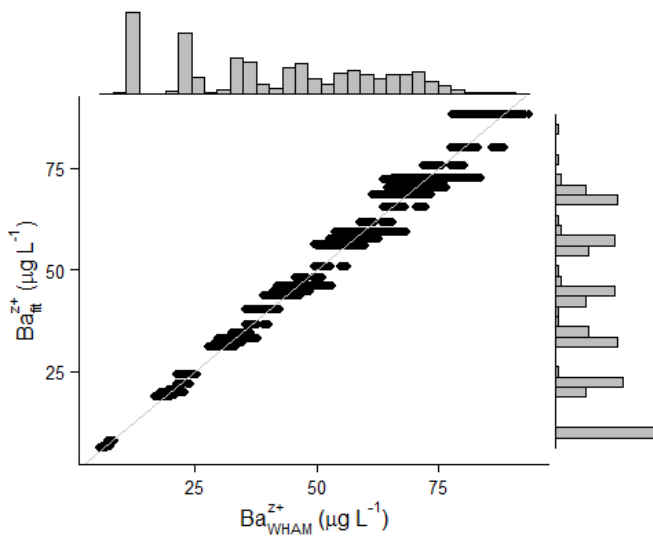
Figure B.4 WHAM ion activities versus non linearly fitted ion activities for Ba, anions excluded.
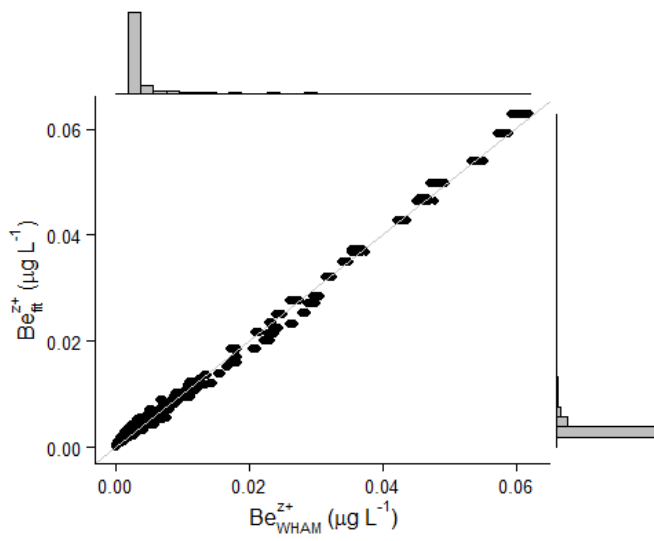
Figure B.5 WHAM ion activities versus non linearly fitted ion activities for Be.
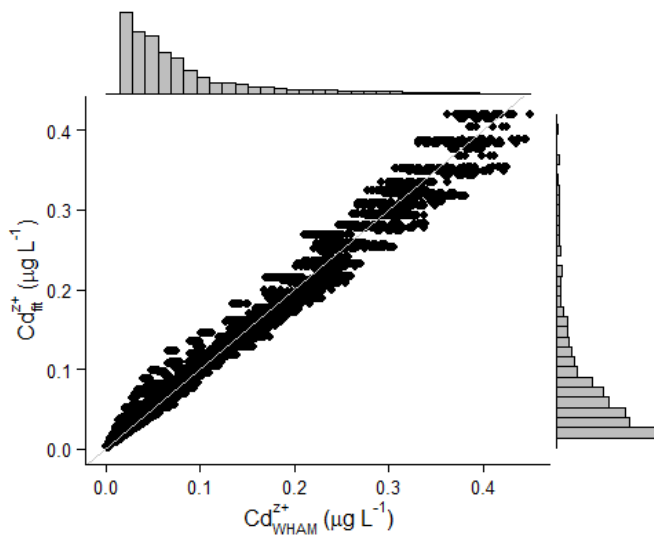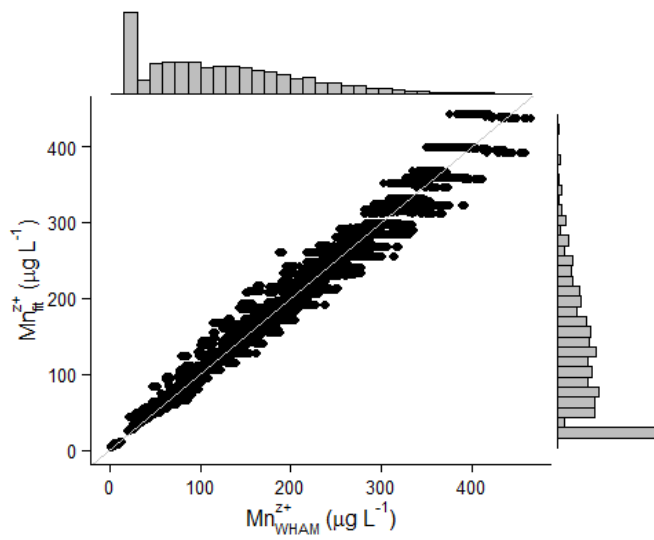

Figure B.6 WHAM ion activities versus non linearly fitted ion activities for Cd.

Figure B.7 WHAM ion activities versus non linearly fitted ion activities for Co.

Figure B.8 WHAM ion activities versus non linearly fitted ion activities for Co, anions excluded.

Figure B.9 WHAM ion activities versus non linearly fitted ion activities for Cr.

Figure B.10 WHAM ion activities versus non linearly fitted ion activities for Cu.

Figure B.11 WHAM ion activities versus non linearly fitted ion activities for Cu, anions excluded
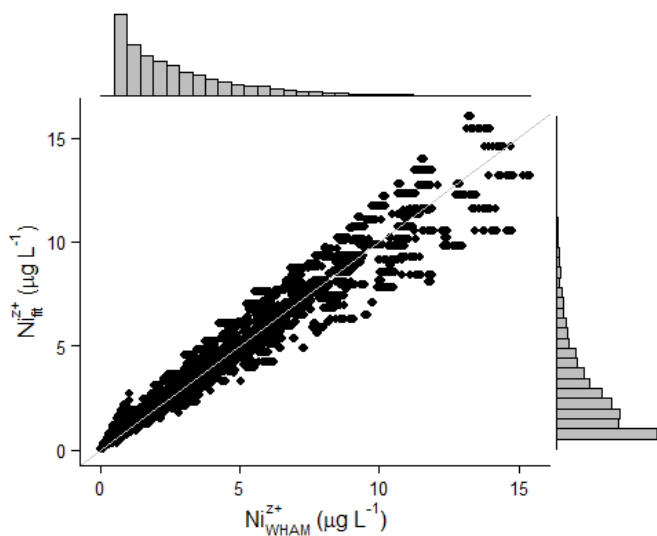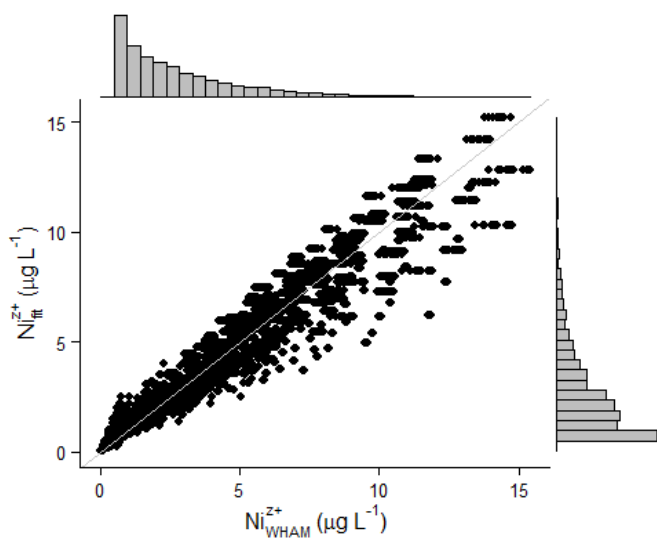

Figure B.12 WHAM ion activities versus non linearly fitted ion activities for Hg.

Figure B.13 WHAM ion activities versus non linearly fitted ion activities for La.

Figure B.14 WHAM ion activities versus non linearly fitted ion activities for Mn.

Figure B.15   WHAM ion activities versus non linearly fitted ion activities for Ni.

Figure B.16 WHAM ion activities versus non linearly fitted ion activities for Ni, anions excluded.
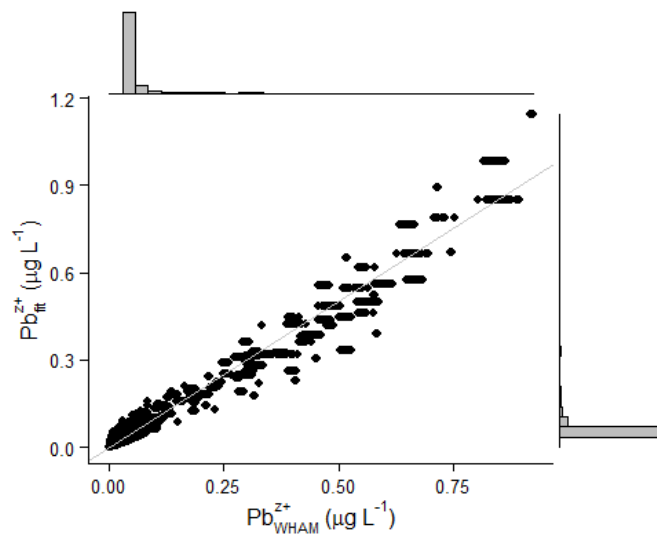
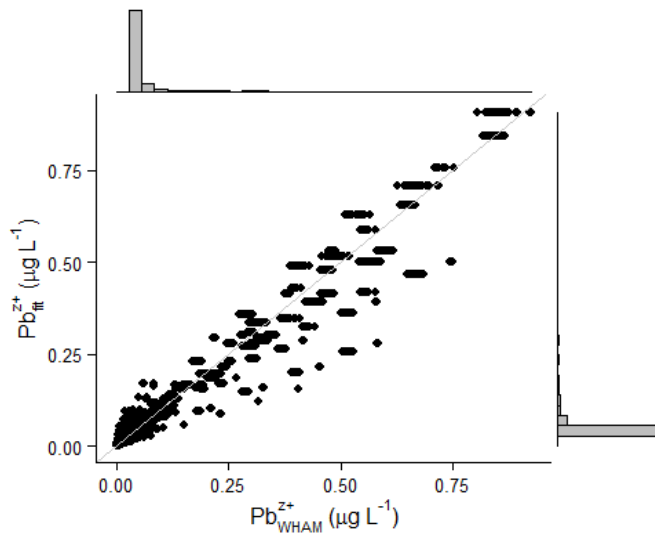*Figure B.17 WHAM ion activities versus non linearly fitted ion activities for Pb.*

Figure B.18 WHAM ion activities versus non linearly fitted ion activities for Pb, anions excluded.
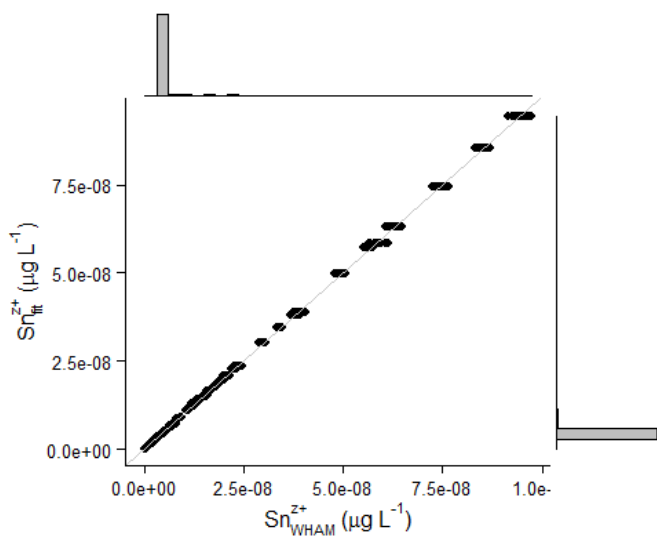
Figure B.19 WHAM ion activities versus non linearly fitted ion activities for Sn.

Figure B.20 WHAM ion activities versus non linearly fitted ion activities for Sr.

Simple equations for the calculation of free metal ion activities in natural surface waters
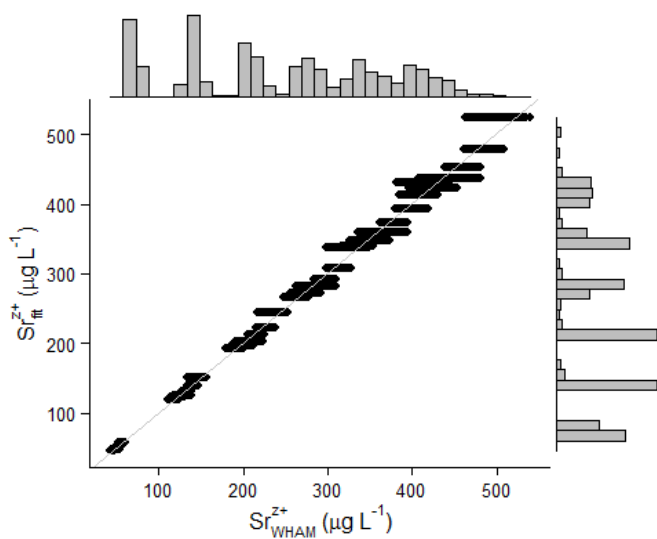
Figure B.21 WHAM ion activities versus non linearly fitted ion activities for U.

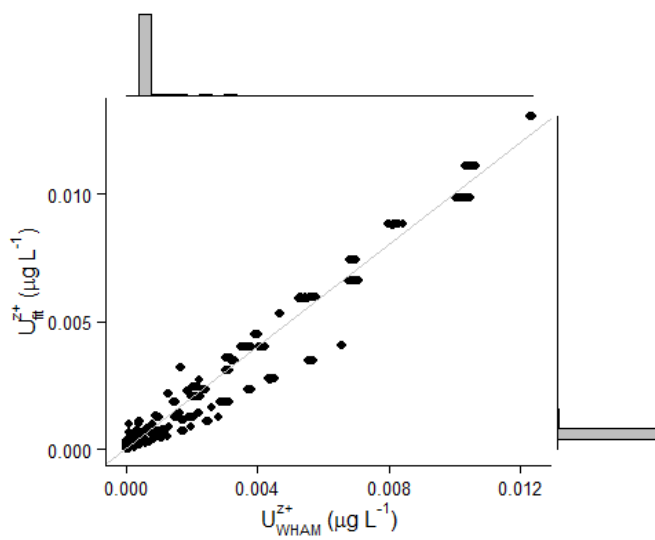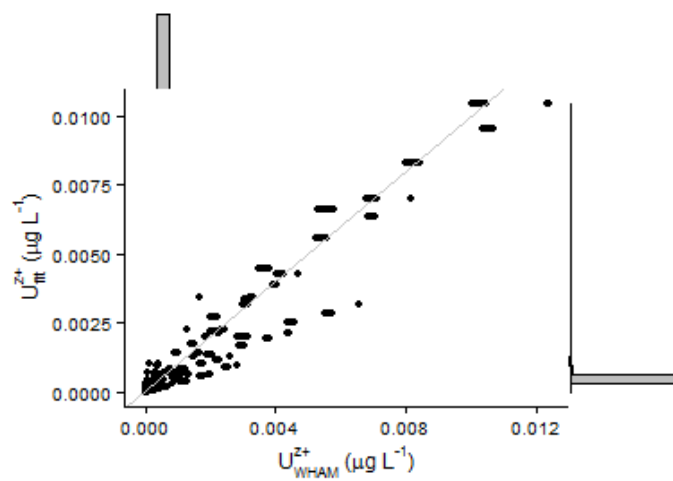*Figure B.22   WHAM ion activities versus non linearly fitted ion activities for U, anions excluded.*
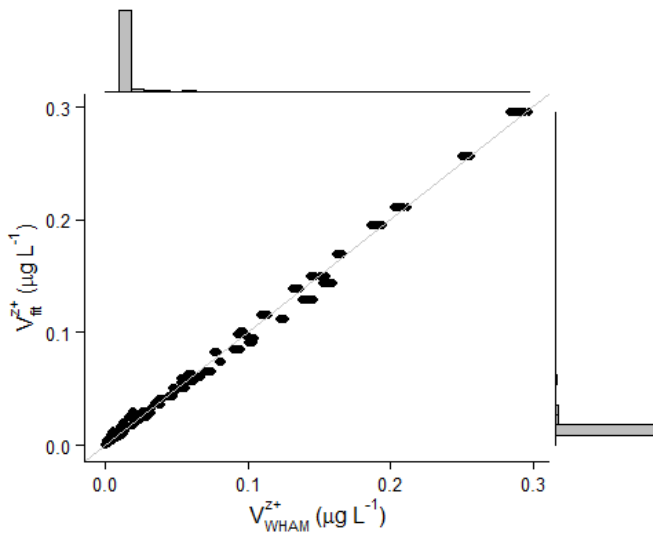
Figure B.223 WHAM ion activities versus non linearly fitted ion activities for V.
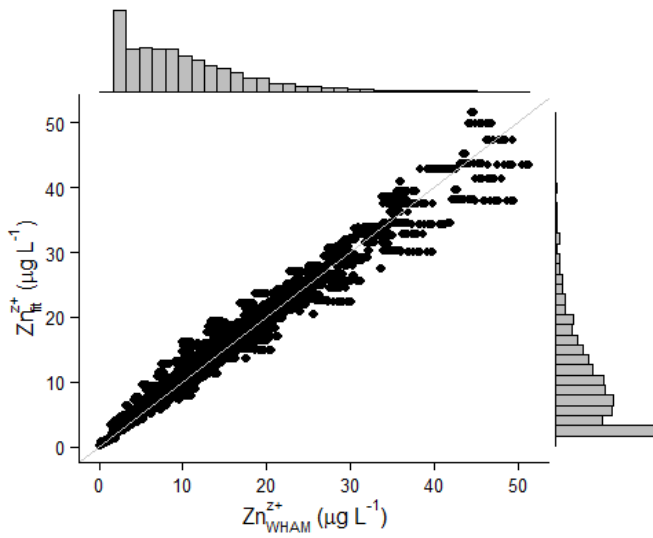
Figure B.234 WHAM ion activities versus non linearly fitted ion activities for Zn.

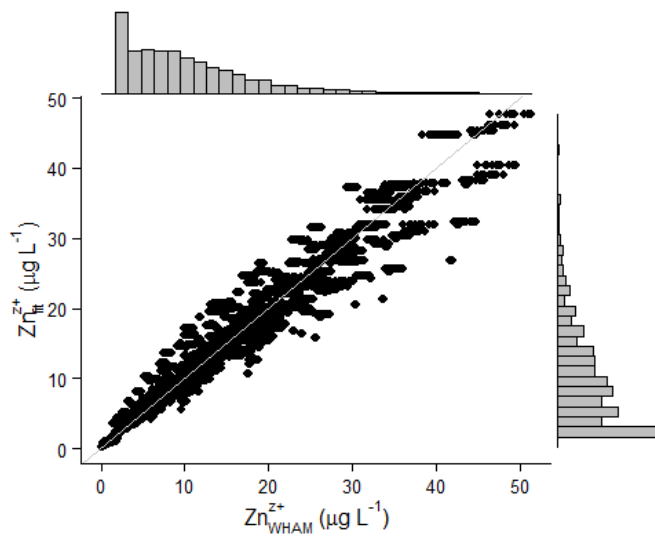*Figure B.245 WHAM ion activities versus non linearly fitted ion activities for Zn, anions excluded.*

C WHAM7 Input Example

Screen captures of WHAM7 input file for silver (Ag), displaying the model settings. See the next page for the ASCII form of the input file.

Figure C.1   Screen capture WHAM input file grid.

Figure C.2   Screen capture of options for WHAM input file.

Sample input for Ag, with 10 input lines (starting at "Ag,0,10,390,…,2e-05"). The actual file extends down to 19980 input lines. The input for the other metals is identical, except for the metal and its concentrations.

```
Description (File: Ag.wi7)
Database,default.db7
Backup,True
Colour,False
Phases,2
Name, Particulate?, Colloidal?
ha,No,Yes
fa,No,Yes
Number of Solutes, 7
Number of Data Lines, 19980
Operating Precision, .01
Consider uncertainty in measurements, No
Consider uncertainty in parameters, No
Uncertainty SD, 1
Number of samples, 1999
Allow FeOH3 pptn, No
Allow AlOH3 pptn, No
Control Fe by FeOH3, No
Control Al by AlOH3, No
FeOH3 has active surface, No
Binding phase for precipitated FeOH3, Binding phase for precipitated FeOH3
Conversion factor for precipitated FeOH3, 0
Precipitated FeOH3 is particulate, no
AlOH3 has active surface, No
Binding phase for precipitated AlOH3,
Conversion factor for precipitated AlOH3, 0
Precipitated AlOH3 is particulate, no
Adjustable Component, 999
Is Fixed, No
Activity Correction,dh
Na,Mg,Ca,Cl,SO4,CO3,Ag
 3 , 4 , 7 , 502 , 504 , 505 ,21 t,t,t,t,t,t,t
Description,SPM,Temperature,pCO2,pH,ha,fa,Na,Mg,Ca,Cl,SO4,CO3,Ag
 ,g/l,deg C,ppm, ,mg/l,mg/l,mg/l,mg/l,mg/l,mg/l,mg/l,mg/l,mg/l
 , , , , , , , , , , , ,
Ag,0,10,390,6,0.5,0.5,10,5,25,48,10,10,2e-05
Ag,0,10,390,6,3.4,3.4,10,5,25,48,10,10,2e-05
Ag,0,10,390,6,6.3,6.3,10,5,25,48,10,10,2e-05
Ag,0,10,390,6,9.2,9.2,10,5,25,48,10,10,2e-05
Ag,0,10,390,6,12.1,12.1,10,5,25,48,10,10,2e-05
Ag,0,10,390,6.5,0.5,0.5,10,5,25,48,10,10,2e-05
Ag,0,10,390,6.5,3.4,3.4,10,5,25,48,10,10,2e-05
Ag,0,10,390,6.5,6.3,6.3,10,5,25,48,10,10,2e-05
Ag,0,10,390,6.5,9.2,9.2,10,5,25,48,10,10,2e-05
Ag,0,10,390,6.5,12.1,12.1,10,5,25,48,10,10,2e-05
```

Notes