Article (refereed)

Lofts, Stephen; **Tipping, Edward**. 2011 Assessing WHAM/Model VII against field measurements of free metal ion concentrations: model performance and the role of uncertainty in parameters and inputs. *Environmental Chemistry*, 8 (5). 501-516. 10.1071/EN11049

Research Paper (GEOSPEC Special Issue)

Assessing WHAM/Model VII against field measurements of free metal ion concentrations: model performance and the role of uncertainty in parameters and inputs

Stephen Lofts[A,B] *and Edward Tipping*[A]

[A]Centre for Ecology and Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, LA1 4AP, UK.

[B]Corresponding author. Email: stlo@ceh.ac.uk

Environmental context. The chemical speciation of metals in waters is of great importance in determining their transport, fate and effects in the environment. Modelling chemical speciation is valuable for making predictions about these effects. Here a model of metal speciation is tested against field data, and recommendations are made as to how both model and measurements might be improved in future.

Abstract A key question in the evaluation of chemical speciation models is: how well do model predictions compare against speciation measurements? To address this issue, the performance of WHAM/Model VII in predicting free metal ion concentrations in field samples has been evaluated. A statistical sampling method considering uncertainties in input measurements, model parameters and the binding activity of dissolved organic matter was used to generate distributions of predicted free ion concentrations. Model performance varied with the metal considered and the analytical technique used to measure the free ion. Generally, the best agreement between observation and prediction was seen for aluminium, cobalt, nickel, zinc and cadmium. Important differences in agreement between model and observations were seen, depending upon the analytical technique. In particular, concentrations of free ion determined with voltammetric techniques were largely over-predicted by the model. Uncertainties in model predictions varied among metals. Only for aluminium could discrepancies between observation and model could be explained by uncertainties in input measurements and model parameters. For the other metals, the ranges of model predictions were mostly too small to explain the discrepancies between model and observation. Incorporating the effects of uncertainty into speciation model predictions allows for more rigorous assessment of model performance.

EN11049

S. Lofts and E. Tipping

Running header: Chemical speciation model testing

Introduction

The modelling of equilibrium metal speciation in natural waters has assumed an increasing importance in recent years as the role of speciation in metal bioavailability and toxicity has been recognised. Advances in the understanding of the chemistry of natural components of waters, particularly organic matter (humic substances), have allowed the development of sophisticated metal–humic binding models, such as NICA–Donnan[1] and Humic Ion-Binding Model VI.[2] A predecessor

to Model VI, Model V,[3] is incorporated into the Biotic Ligand Model[4] of metal bioavailabilty. The evaluation of the performance of these models against field speciation data is thus becoming a research area of great importance. These models simulate the binding of protons and metal ions to humic substances (humic and fulvic acids), which are generally thought to be the dominant form of non-living organic matter in terrestrial and aquatic environments, and which are known to bind some metals strongly. The models are parameterised against laboratory data on the binding of protons and metals to isolated samples of humic substances, and thus their application to field data rests on the important assumption that the ion-binding properties of dissolved organic matter in the field are similar to those of isolated humics.

There are two key research questions when evaluating the performance of chemical speciation models:

1. How well do model predictions compare against speciation measurements on field samples?
2. How does uncertainty in model parameters and input variables influence the distribution of model predictions?

Evaluating model predictions of solid–solution metal partitioning in surface waters is fairly straightforward[5] as the separation of solid and solution phases can be readily done (e.g. by filtration). However, evaluation of speciation within the dissolved phase, which is important for bioavailability models, is more complex as field measurements of dissolved metal speciation are required. In recent years several techniques have been developed for measuring free metal ions in surface waters at environmental concentrations, such as the Donnan membrane technique (DMT),[6] a range of voltammetric techniques[7–13] and ion-exchange methods,[14] and there is a growing body of measured free metal ion concentration data. Several studies[7,15–18] have compared speciation calculations of free metal ion concentrations against measurements made by a variety of methods, however, to date there has been no systematic attempt to synthesise and model the data as a single exercise.

In assessing the ability of a humic ion-binding model to predict measured free ion concentrations in the field four factors must be considered:

- the degree to which the binding properties of isolated humic substances in the laboratory reflect the binding properties of dissolved organic matter (DOM) in the field;
- uncertainties in key input variables, such as pH, dissolved organic carbon (DOC) nature and concentration, dissolved metal and other solute concentrations, required by the model;
- uncertainties in the model parameters; and
- uncertainty in measured concentrations of free metal ions in the field.

The first factor is the most challenging to assess, as a thorough analysis would entail a comprehensive characterisation of field DOM, including the ion-binding properties of the humic

component. Therefore, past modelling efforts have taken a pragmatic approach whereby a portion of the measured DOM is assumed to have the ion-binding properties of humic acid, fulvic acid (FA), or a mixture of both,[6,19] reflecting the apparent 'binding activity' of the DOM. Using this approach, a 'best average' binding activity, based on the apparent binding activity of DOM from a range of field samples, has been derived.[19] The concentration of humic substances required to simulate free Al and Cu concentrations in freshwater samples is similar to, but on average somewhat lower than, the concentration of DOM.[19,20] The average DOM 'activity' is similar for separate studies involving Al and Cu, and so in the absence of similar data for the other metals it is reasonable to apply the same average 'activity' in studies involving speciation of the latter. This supports the hypothesis that the ion-binding properties of natural organic matter are similar to those of isolated humic substances. However, the optimal binding activity has also been shown to vary among field samples and thus represents a source of uncertainty if a 'best average' value is used for predictive work. For example, Kalis and coworkers[6] studied how varying the proportions of HA and FA comprising DOM influenced the prediction of the free ion by the NICA–Donnan model, and found that for copper the free ion prediction varied by four orders of magnitude, whereas for nickel, zinc, cadmium and lead the variation was up to an order of magnitude.

The effects on predictions of uncertainties in input variables and model parameters are amenable to analysis, provided suitable uncertainty ranges can be defined. Uncertainty in input variables arises from measurement error, and is present in all measurements used as inputs to the model. As the input data for a speciation problem will comprise multiple measurements including pH, temperature, and concentrations of major ions, dissolved organic carbon and metal, multiple sources of input variable uncertainty exist. Uncertainty in model parameters arises because the humic-binding models are parameterised using multiple datasets (where available) for metal binding to isolated humic substances, with the aim of producing 'best average' binding parameters. Parameterisation to multiple laboratory datasets gives multiple fitted values of the metal–humic binding constant which will exhibit a degree of variability, giving rise to uncertainties in the 'best average' values.

Little research has been done into the influences of uncertainty on the predictions of chemical speciation models, although methods for doing so have been presented (e.g. Anderson[21]). Given the complexity of speciation calculations, statistical sampling approaches (e.g. Monte Carlo analysis) have proved most popular in application (e.g. Groenenberg et al.[22]). Such approaches involve the generation of a sample set of an input or a parameter, by repeated sampling from a presumed statistical distribution of values, and the generation of a sample set of model outputs from these sample input sets. An advantage of this approach is that multiple sample sets, representing uncertainties in more than one input and/or parameter, may be simultaneously generated in order to study the combined effect on the distribution of model outputs.

Incorporating the ability to assess uncertainties into chemical speciation models would allow users to assess the main sources of uncertainty in their predictions, and would highlight those uncertainties to which the models are most sensitive, allowing focusing of future research efforts. The development of Humic Ion Binding Model VII[23] and the parallel development of an updated version of the WHAM model, has afforded the opportunity to extend WHAM to allow uncertainty assessments on calculations to be performed. In this paper we present the results of a comprehensive study of the performance of the newly updated WHAM, which we term WHAM/Model VII. We have collated available measurements of free metal ion concentrations in freshwaters along with associated measurements of pH, major ions and DOC where these are available. Uncertainty in the model predictions has been quantified using a Monte Carlo approach with uncertainties represented by distributions of selected model parameters and water quality input variables. Free ion measurements and model predictions are compared on the basis of goodness-of-fit and bias. Deviations between measurement and modelled outputs are assessed against the predicted ranges of variability in the free metal ion concentrations arising from the uncertainties considered.

Methods

A glossary of parameters and terms used herein is provided in the 'Glossary of terms and parameters used in this study' section.

WHAM/Model VII

Humic Ion-Binding Model VII[23] describes ion (proton and cationic metal) binding to HA and FA using a structured formulation of discrete binding site types to describe heterogeneity in ion binding strengths. The formulation of binding site types, and their interrelationships, in terms of relative abundances and binding strengths for protons and other cations, is done with the goal of adequately describing the binding of protons and cations to isolated HAs and FAs in laboratory studies using the minimum necessary set of parameters. Proton binding is simulated using the equilibrium:

$$R^- + H^+ = RH \quad (1)$$

where R represents a binding site. Eight sites are defined: four strong acid (Type A) with a total density of n_A moles per gram, and four weak acid (Type B) with a total density of n_B moles per gram. The total density of the B sites (n_B) is fixed to half the density of the A sites (n_A). Each of the Type A sites has the same density (i.e. $n_A/4$) and each Type B site also has the same density (i.e. $n_B/4$). The intrinsic proton binding strength to each site type is defined by a central p$K_{A/B}$ value and a spread factor, Δp$K_{A/B}$.

Metal binding equilibria are defined as follows:

$$R_{mon}^- + M^{z+} = RM^{(z-1)+} \quad (2)$$

$$R_{bi}^{2-} + M^{z+} = RM^{(z-2)+} \quad (3)$$

$$R_{tri}^{3-} + M^{z+} = RM^{(z-3)+} \quad (4)$$

where R_{mon}, R_{bi} and R_{tri} represent mondentate, bidentate and tridentate binding sites. There are two intrinsic binding constants per metal, log K_{MA} and log K_{MB}, referring to monodentate binding to the Type A and Type B sites. The relationship between log K_{MA} and log K_{MB} for each metal is fixed, so only log K_{MA} needs to be fitted. Bidentate and tridentate metal binding sites are generated by calculating the proportions of single sites that are able to form pairs and triplets, assuming a spherical shape for the humic molecules. Metal binding strengths are calculated by summing the log $K_{MA/MB}$ values for the individual sites. Additional heterogeneity of binding strength is achieved by an additional metal-specific parameter (ΔLK_2) that is used to increase the binding strength of subsets of the bidentate and tridentate sites. A total of 9 % of the bidentate sites have the logarithms of their binding strength increased by ΔLK_2 and 0.9 % have increases of $2\Delta LK_2$. For tridentate sites the respective increases are $1.5\Delta LK_2$ and $3\Delta LK_2$.

The intrinsic proton and metal binding constants p$K_{A/B}$ and log $K_{MA/MB}$ are defined for a state of zero electrical charge of the humic substances. Development of charge, by ion binding and release, modifies the overall binding strengths because of attractive or repulsive electrostatic forces between the humic molecules and the binding ions. This effect is taken into account by empirical electrostatic terms that modify the intrinsic binding constants. Counterion accumulation adjacent to humic molecules is simulated by a Donnan model.

Besides Model VII, the WHAM code comprises submodels for ion binding to metal oxides[24] and a cation exchanger, and solution speciation. Ion binding to metal oxides is described using a surface complexation model. As with Model VII, relationships among binding site densities, and binding strengths, are fixed where this facilitates an adequate description of metal binding to oxides under laboratory conditions with a minimal parameter set. Proton and metal binding reactions are described using the following equilibria, where ≡XOH represents a surface hydroxy group, X being an atom of the metal comprising the metal oxide:

$$\equiv XOH_2^+ = \equiv XOH + H^+ \quad (5)$$

$$\equiv XOH = \equiv XO^- + H^+ \quad (6)$$

$$\equiv XOH + M^{n+} = \equiv XOM^{(n-1)+} + H^+ \quad (7)$$

for which the intrinsic binding constants are described respectively by parameters $K_{1,oxide}$, $K_{2,oxide}$ and $K_{MH,oxide}$. The surface sites are assumed to be homogeneous with respect to proton binding. Heterogeneity in metal binding is achieved using the expression

$$pK_{MH,oxide,y} = pK_{MH,oxide,y} + \Delta pK_{MH,oxide}$$

where $y = 0$ for 90.1 % of the sites, $y = 1$ for 9 % of the sites and $y = 2$ for 0.9 % of the sites. The term $\Delta pK_{MH,oxide}$ is a binding strength heterogeneity term specific to the oxide but common to all metals.

Four types of metal oxide are simulated: iron(III) oxide, aluminium oxide, manganese oxide and silica.

Electrostatic effects on binding strength are described by empirical electrostatic terms and counterion accumulation adjacent to the mineral surface is simulated by a Donnan model. The cation exchanger is defined by a fixed (negative) surface charge per unit mass. Cation binding by electrostatic accumulation is described by a Donnan model.

Equilibria among simple ions in solution are described by conventional mole balance equations. The precipitation of aluminium and iron(III) hydroxides is simulated by conventional solubility products. Alternatively, the solution chemistry of aluminium and iron(III) may be simulated by assuming equilibrium with the hydroxide. Here we use the previously quoted standard solubility product of 8.5 and enthalpy of –107 kJ mol^{-1} for aluminium(III) hydroxide.[25] A solubility product expression for of iron(III) hydroxide was derived by Lofts and co-workers[26] using WHAM/Model VI. We have updated these calculations using WHAM/Model VII (see Accessory publication for details). An enthalpy of iron(III) hydroxide solubility of –100.4 kJ mol^{-1} [27] was used. The model allows iron(III) and aluminium(III) hydroxide precipitates to have chemically active surfaces, which are simulated using the surface complexation model.

WHAM/Model VII may be used to simulate partitioning of metals between the particulate and solution phases in a freshwater, by specifying concentrations of chemically active substances (humic and FAs, metal oxides, cation exchanger) in the particulate phase. Concentrations of any of these binding substances may also be specified to be present in the solution phase (concentrations of these substances in the solution phase are defined as 'colloidal' in WHAM/Model VII). As WHAM/Model VII is a purely chemical model, the method used to distinguish the particulate and solution phases (e.g. filter size used to isolate particulate matter from the solution) is not important. Alternatively, the model may be used to simulate speciation in the filtrate phase alone, as is done in this study.

Consideration of uncertainty in WHAM/Model VII

To produce estimates of uncertainty in model output, WHAM/Model VII has been developed to allow uncertainty in parameters and measurements to be inputted and for measures of the resulting uncertainty to be outputted. Output uncertainties are calculated using a Monte Carlo method. Each input, x_i (parameter or input variable), for which uncertainty is to be considered, is assigned an absolute or relative uncertainty factor (σ_i or p_i). Sample sets of y values of each variable are then generated by random sampling. A perturbation factor q, generated by random sampling from a standard normal distribution, is used to calculate perturbed values of each variable. For variables assigned an absolute uncertainty the perturbed value is given by:

$$x'_{i,j} = x_i + q \cdot \sigma_i \quad (1)$$

For variables assigned a relative uncertainty the perturbed value is given by[24]:

$$x'_{i,j} = \exp\ \ln x_i + q \cdot \sqrt{\ln\ 1 + p_i} \quad (2)$$

such that for each variable a sample set of values $x_{i,1}$, $x_{i,2}$, $x_{i,3}$, ...$x_{i,y}$ is generated. The model is run y times, each run using one set of the sampled variables, producing for each output a population of values. A further run is done using the unperturbed set of inputs, to produce a baseline prediction. The advantages of using a Monte Carlo approach include the rapid and robust generation of many 'scenarios' (samples) by statistical sampling from parameter and input variable distributions, and the ability to impose different distributions of uncertainty on different input variables and parameters. When uncertainty in humic–metal binding strengths (log K_{MA}) is considered, a single perturbation factor q is applied per sample, so that the metal binding strengths are always all perturbed in the same direction and by the same proportion of their uncertainty factors. For metals having the same uncertainty factor the relative values of their log K_{MA} values are thus kept the same. This is done because we consider it reasonable to assume that the molecular variability in humic substances giving rise to variability in metal binding strengths is likely to affect the binding of all metals to a similar degree and in the same direction. The same restriction is also applied when calculating uncertainties in the parameter p$K_{MH,oxide}$ for metal binding to oxides.

Variability in the predicted free ion concentrations in theoretical calculations and modelling of measured free metal ions was quantified by calculating the 16–84 % interquartile range of predicted log free ion concentration ($Q_{16–84}$). The 2.5–97.5 % interquartile range ($Q_{2.5–97.5}$) was used to assess the importance of the variability in outputs in relation to the observations, and thus the overall role of uncertainty in explaining the discrepancies between observation and prediction. If the observed free ion concentration falls within this interquartile range there is at least a 95 % probability that the discrepancy between observation and prediction can be attributed to the uncertainties used to generate the spread of outputs. The $Q_{16\text{-}84}$ and $Q_{2.5\text{-}97.5}$ interquartile ranges are approximately equivalent to ±1 or ±2 standard deviations respectively around the mean of a normally distributed variable. Using these ranges, rather than calculating the standard deviation of the output, takes into account any deviation from normality of the output distribution, while providing a measure of the distribution that is consistent with the concept of the standard deviation as it applies to normally distributed data.

Speciation and uncertainty calculations

The speciation modelling presented in this study comprises theoretical calculations of dissolved metal speciation, including free metal ion concentrations, in field samples in which the free metal ion concentration has been measured. Theoretical uncertainty calculations were done to allow a systematic analysis and understanding of the overall contributions of different inputs to the total uncertainty. The metals for which calculations were done were Al, Co, Ni, Cu, Zn, Cd and Pb, representing the set of metals for which field data were subsequently modelled. All calculations

represented the dissolved phase of surface waters. For Al a soft water composition broadly representative of the field samples simulated (Table 1) was used and calculations were done for the pH range 4.0 to 6.5. A FA concentration of 10 mg L^{-1} was used to simulate dissolved organic carbon. For the remaining metals, a harder water composition (Table 1) and a pH range of 5.0 to 8.5 was used. In both sets of calculations the dissolved Ca was adjusted to give charge balance at each pH value simulated.

Field datasets containing metal speciation data (free ion measurements) were compiled from the peer-reviewed literature. The datasets used are listed in Table 2 along with measurement methods. For all metals except Al, the datasets obtained were based on a complete review of the available literature. Data were used if the measurements of the free ion concentration, dissolved metal, DOC and pH were available. Measurements where prior adjustment of the pH to a quoted value was done were accepted, whereas measurements made on samples amended with additional metal were not. For Al, data were taken from three sites monitored under the UK Acid Waters Monitoring Scheme.[30] Al measurements in these samples did not comprise direct measurements of free ion concentration; instead, [Al^{3+}] was estimated by the speciation of measured 'labile' Al,[31] taken to comprise the free ion and all Al bound to inorganic ligands.

Concentrations of major dissolved species and other physicochemical parameters required for speciation calculations were obtained from the same literature source as the free metal ion concentrations, where possible. These included the temperature at the time of analysis, and concentrations of dissolved Na^{+}, Mg^{2+}, Al^{3+}, K^{+}, Ca^{2+}, Fe^{3+}, Cl^{-}, NO_3^{-}, SO_4^{2-} and F and alkalinity or CO_3^{2-}. Where these measurements were not available, they were sourced from alternative literature, or were estimated (see Tables A1 and A2 in the Accessory publication). If measurements of carbonate or alkalinity were not available, equilibrium with atmospheric CO_2 was assumed, using a partial pressure of 38.50 Pa (equivalent to 3.80×10^{-4} partial pressure). To check the sensitivity of the model to the chosen partial pressure we simulated the free metal ion concentrations in a freshwater medium at partial pressures between 3.0×10^{-4} and 3.0×10^{-3}. The calculated log free ion concentrations varied by no more 0.13.

In line with previous work,[19] the portion of DOC chemically active with respect to metal binding was represented by FA only and an 'activity factor', F_{FADOC}, was used to convert from measured DOC into model input FA. In this study we have used an F_{FADOC} value of 1.27, representing 63.5 % of the DOC being 'active' on average. This value was calculated from the geometric mean of F_{FADOC} values for a set of UK waters found by Bryan and coworkers (Fig. 1).[19] Where estimation of Al concentrations was not possible, dissolved Al was estimated by assuming that the solution was in equilibrium with aluminium(III) hydroxide. Dissolved iron(III) chemistry was simulated by allowing the precipitation of iron(III) hydroxide to occur and by allowing the precipitate to have a chemically

active surface. Iron(III) oxide was used as the model substance to simulate the surface chemistry of the precipitate, assuming one mole of precipitated iron(III) to form 90 g of oxide.[32]

Uncertainties considered in the speciation calculations are shown in Table 3. The uncertainty in log K_{MA} for Model VII was set to the standard deviation of the fitted value[23] and for consistency the uncertainty in pK_{MH} values for metal binding to iron(III) oxide was set to the same value. The uncertainty in the solubility product of iron(III) hydroxide was derived from the calculations of the solubility product expression (see Accessory publication for details). For consistency the uncertainty in the solubility product of aluminium(III) hydroxide was set to the same value. For the theoretical calculations, separate model runs were done considering uncertainty in:

1. F_{FADOC} only;
2. F_{FADOC} and measurements;
3. parameters; and
4. F_{FADOC}, measurements and parameters.

For the calculations of free metal in the field datasets, the same runs were done except for 'F_{FADOC} only'. A sample size of 1999 was used to generate output distributions.

Results

Modelling results: theoretical uncertainty calculations

Calculations for Al are shown for pH 5.0 and pH 6.0 (Fig. 2). At pH 5.0, precipitation of aluminium(III) hydroxide was predicted not to occur in the majority of the Monte Carlo samples. By contrast, at pH 6.0, precipitation of aluminium(III) hydroxide was predicted to occur in all the Monte Carlo samples. Thus, at pH 6.0 calculated Al activities were controlled by the pH and the solubility of aluminium(III) hydroxide. Where only uncertainty in F_{FADOC} was considered, the calculated activity of Al^{3+} in equilibrium with the hydroxide was constant. Therefore, the only possible source of variability in calculated concentrations of [Al^{3+}] at pH 6.0 was variation in the calculated ionic strength, which was negligible. Hence the variability in calculated [Al^{3+}] (expressed as the interquartile range $Q_{16–84}$) at pH 6.0 was also negligible when only uncertainty in F_{FADOC} was considered. At pH 5.0, in the absence of precipitation control, variability in [Al^{3+}] was controlled by the variability in input FA concentration imposed by the uncertainty in F_{FADOC}. Therefore, at this pH the variability as a result of uncertainty in F_{FADOC} was notably larger than at pH 6.0. This was also seen when uncertainties in measurements were considered along with the uncertainty in F_{FADOC}. Variability in [Al^{3+}] was similar at both pH 0.0 and pH 6.0 when uncertainty in parameters was considered, and in both cases was greater than the variability seen when uncertainties in measurements and F_{FADOC} were considered. At both pH values, the uncertainties in the log K_{MA} values for Al^{3+} and the competing ions Mg^{2+} and Ca^{2+} contributed to the variability. At pH 6.0, the uncertainty in the solubility product of Al hydroxide also contributed, resulting in a slightly larger

interquartile range. The variability obtained by combining different sources of uncertainty was clearly less than additive (on a log scale). Combining uncertainties attributable to F_{FADOC}, measurements and parameters produced similar variability at both pH values (1.28 at pH 5.0 and 1.20 at pH 6.0). The effect of combining uncertainties on the resulting variability was clearly less than additive. This was particularly clear at pH 5.0.

Fig. 3 shows the variabilities in predicted free ion concentration for Co, Ni, Cu, Zn, Cd and Pb at pH 7.0. For Co and Ni, uncertainty in input measurements dominates the overall uncertainty, but uncertainty in F_{FADOC} does not contribute greatly to the measurement uncertainty. For Zn and Cd, the contributions of parameter uncertainty and of F_{FADOC} to measurement uncertainty are large relative to the other metals, particularly for Zn. The variability seen when considering measurements and parameters together is consistently less than additive. For Cu and Pb, uncertainties in parameters clearly make a larger contribution to the total variability than measurement uncertainties, and variability attributable to uncertainty in F_{FADOC} is approximately half that resulting from measurements overall. The relative importance of the different sources of uncertainty did not vary with pH (data not shown).

These calculations indicate that the sources of uncertainty in calculations are generally similar to those found by Groenenberg and coworkers[22] using the NICA–Donnan model. They found that uncertainties in binding predictions for Cu and Pb were largely attributable to uncertainty in the metals' binding affinity, which is consistent with our observation that parameter uncertainty (including uncertainty in the main metal binding parameter log K_{MA}) was more important than input measurement uncertainty for these metals. They also found that uncertainties in F_{FADOC} and in binding parameters produced comparable interquartile ranges of uncertainty for Cd^{2+}. This does not agree with our findings, where parameter uncertainty has a somewhat larger effect than uncertainty in F_{FADOC}.

Modelling results: field data

Model fits and the variability in outputs are given in Tables 3 and 4. The basic fit is characterised by the bias and the root mean squared error (RMSE) in log $[M^{z+}]$.

Including a chemically active iron(III) oxide phase had a notable effect on the predicted speciation of lead. Free lead ion concentrations predicted, assuming precipitated iron(III) hydroxide to be surface active, were predicted to be up to almost two orders of magnitude lower than the concentrations predicted if precipitated iron(III) hydroxide was assumed to be not surface active (Fig. 4). This is in accordance with past modelling predictions[5] and with experimental observation (e.g. Lyvén et al.[33]) Negligible effects were seen for the other metals.

Observed and predicted concentrations of $[Al^{3+}]$ (Fig. 5) had mean biases under 0.5 orders of magnitude for all three datasets. Dataset Al-01 had all observations predicted to within one order of magnitude and 52 of the 55 observations predicted to within 0.5 orders of magnitude. Dataset Al-02

had 179 of 180 observations predicted to within one order of magnitude and 162 observations predicted to within 0.5 orders of magnitude, and a small absolute bias (Table 5). Values in dataset Al-03 were not quite so well predicted with a relatively large negative bias. Of the observations, 141 of 167 were predicted to within one order of magnitude. The number of observations predicted to within half an order of magnitude (59) was low relative to Al-01 and Al-02. Most observations fall within the interquartile range $Q_{2.5–97.5}$, suggesting that uncertainty in input variables and parameters can largely account for the discrepancies between observation and prediction.

Observed concentrations of [Co^{2+}] (Fig. 6) were all predicted to within one order of magnitude and six of the seven observations were predicted to within half an order of magnitude. One of the seven observations fell within the $Q_{2.5–97.5}$ interquartile range.

Observed and predicted concentrations of [Ni^{2+}] (Fig. 7) showed that when [Ni^{2+}] > 10^{-8} M the model predictions were consistently within one order of magnitude of the observations with a mean bias of 0.06 log units. At lower concentrations, overestimation of observed [Ni^{2+}] was greater (mean bias 0.82 log units). This tendency to overestimation has been noted previously[34,35] when using Model VI. Overall, 43 of 54 observations were predicted to within one order of magnitude, and 21 to within 0.5 orders of magnitude.

Observed and predicted concentrations of [Cu^{2+}] (Fig. 8) showed that agreement between observation and prediction was better at higher free copper (>10^{-12} M) than at lower concentrations. Overall 33 of 133 observations were predicted to within one order of magnitude and 21 to within 0.5 orders of magnitude. The data for which [Cu^{2+}] < 10^{-12} M have all been measured using voltammetric techniques. Measured [Cu^{2+}] ranges were as low as 10^{-16} M yet the model did not predict any concentrations to be below 10^{-13} M. Mean biases for four of the five datasets obtained using competitive ligand exchange–adsorptive cathodic stripping voltammetry (CLE-AdCSV) (Cu-02 to Cu-05) showed overestimation of measured concentrations by at least two orders of magnitude on average. By contrast, the concentrations in dataset Cu-01, also measured by CLE-AdCSV, were underestimated by the model. Seven of nine CLE-AdCSV measurements in dataset Cu-03 with observed [Cu^{2+}] in the range 10^{-12} to 10^{-9} M, were relatively well predicted, to within an order of magnitude. These measurements were done in soft, acidic-to-circumneutral waters (pH 5.3–7.5; dissolved Ca 42–85 mg L^{-1}) of low [DOC] (0.4–1.1 mg C L^{-1}) whereas the remaining measurements were done in waters of higher pH (samples were adjusted to pH 7.9), hardness (dissolved Ca 0.1–0.4 mg L^{-1}) and [DOC] (0.9–5.8 mg C L^{-1}).

The other measurement techniques used generally resulted in closer agreement between observation and prediction. For measurements using the DMT (Cu-06b, Cu-07 and Cu-08), bias between observation and prediction was fairly low, but there was notable scatter between observation and prediction when [Cu^{2+}] < 10^{-9} M. Dataset Cu-06a (permeation liquid membrane) was

underestimated on average. Dataset Cu-09, obtained using an ion exchange column method (IET), showed relatively close agreement between observation and prediction when $[Cu^{2+}] > 10^{-8}$ M (mean bias = –0.12 log units). At lower free copper concentrations agreement was relatively poor (mean bias = –2.22 log units).

Observed and predicted concentrations of $[Zn^{2+}]$ (Fig. 9) showed that 72 of 84 observations were predicted to within an order of magnitude and 45 to within half an order of magnitude. Some patterns can be observed depending upon measurement technique. Observations obtained by CLE-AdCSV (Zn-01, Zn-02, Zn-03) were mostly overestimated by the model (mean bias 0.5 log units). Observations obtained by DMT (Zn-04, Zn-05) were correlated fairly well with prediction when $[Zn^{2+}] > 10^{-7}$ M (RMSE 0.12 log units) although with a relatively small bias to underestimation (mean –0.26 log units). For $[Zn^{2+}] < 10^{-7}$ M, observations obtained by DMT became more scattered with respect to predictions (RMSE 0.26 log units). Observations obtained by IET (Zn-06) were predicted relatively well when $[Zn^{2+}] > 10^{-7}$ M (mean bias –0.14 log units) but were not so well predicted at lower concentrations (mean bias –0.91 log units).

Observed and predicted concentrations of $[Cd^{2+}]$ (Fig. 10) showed patterns generally similar to those for zinc. Of 85 measurements, 72 were predicted to within an order of magnitude and 51 to within half an order of magnitude. Measurements made using voltammetric techniques (Cd-01, Cd-02) were consistently overpredicted. Measurements made using DMT were simulated relatively well (RMSE 0.06 log units, mean bias 0.08 log units) when $[Cd^{2+}] > 10^{-10}$M but were overpredicted at lower concentrations (mean bias 0.68 log units). Observations made using IET were all simulated to within one order of magnitude. Two of three measurements made using a permeation liquid membrane (PLM) were predicted to within one order of magnitude.

Table 5 compares bias and error in the predictions made with and without assuming binding of lead to iron(III) oxide. Allowing precipitated iron(III) hydroxide to have an active surface either reduced the predicted free ion or made little difference to the prediction. Overall, bias and RMSE were reduced by allowing precipitated iron(III) hydroxide to have an active surface. Of the individual datasets, the fit was most improved for Pb-01, for which the binding to iron(III) oxide is most important in determining the predicted free ion (Fig. 4) although the bias and RMSE remained large.

Observed and predicted concentrations of $[Pb^{2+}]$, assuming binding of lead to iron(III) oxide (Fig. 11), showed somewhat similar patterns to Cu. Measurements made using competitive ligand exchange–differential pulse cathodic stripping voltammetry (CLE-DPCSV) were consistently underpredicted. Measurements made by DMT were predicted with low bias but with notable scatter (21 of 45 observations were not predicted to within one order of magnitude). Two of three measurements made by PLM were predicted to within an order of magnitude. Overall, 33 of 76

observations were predicted to within one order of magnitude and 19 to within 0.5 orders of magnitude.

Model uncertainty

Model uncertainty results (Tables 4, 5) summarise the spread of output values (quantified as the $Q_{16–84}$ interquartile range of log free ion concentration) obtained in each dataset when considering uncertainty in (a) input varaibles and F_{FADOC} only, (b) parameters only, and (c) input variables, F_{FADOC} and parameters together. The influence of the different categories of uncertainty on the variability of the output free ion concentrations is not additive, consistent with the findings of the theoretical calculations. Where uncertainties in inputs, F_{FADOC} and parameters are considered together, the spread of output values is smaller than the sum of the corresponding ranges calculated when the uncertainties are considered in separate simulations. With the exception of Cu, considering input uncertainty only mostly results in a greater spread of output predictions than does considering parameter uncertainty only. Parameter uncertainty resulted in a greater spread of output predictions in one of three Al datasets, and in two of five Pb datasets. For Cu, considering parameter uncertainty resulted in spreads of output values that were consistently greater than or equal to the spreads obtained when uncertainties in inputs and F_{FADOC} were considered. On average, the spreads of output values (i.e. the breadth of the $Q_{16–84}$ range) followed the trend Al ≈ Cu > Pb > Ni ≈ Zn ≈ Cd.

A key aspect of assessing uncertainty is whether discrepancies between observations and predictions can generally be attributed to uncertainty in the input variables and model parameters, i.e. if the observation falls within the range of predicted values then the discrepancy is explained. The $Q_{2.5–97.5}$ interquartile range of the predictions, when all uncertainties were considered, was used to assess this (Tables 3, 4).

The majority of observations of [Al^{3+}] (88 %) fell within the $Q_{2.5–97.5}$ range and so generally the discrepancies can be attributed to the influence of uncertainties. However, for the other metals, the proportions of observations falling within this range were lower, ranging from 0 % for Co to 42 % for Cd, and 22 % when all five metals were considered together. Thus, in general the discrepancies between observation and prediction could not be attributed to uncertainty in the model predictions.

Discussion

As noted in the Introduction, the following factors must be considered when the performance of WHAM/Model VII is assessed against field observations:

- the influences of uncertainties in model inputs variables and parameters on the spread of model predictions;
- the extent to which the metal binding properties of humic substances represent those of field DOM; and

- the accuracy and precision of measured free metal ion concentrations.

This study has explicitly tackled the first of these factors, by incorporating the consideration of uncertainty in input variables and parameters into the model to allow a spread of predictions that can be compared to measured values. The second factor was partly tackled by incorporating uncertainty in F_{FADOC} into calculations. We have shown that with the exception of Al, discrepancies between measurements and model predictions can largely not be explained by uncertainties in input variables or model parameters, assuming that such uncertainties have been reasonably quantified. Two factors thus remain as possible sources of discrepancy between model and measurement: the extent to which isolated humic substances represent field DOM, and the precision and accuracy of free ion measurements.

In this study we have simulated metal speciation assuming DOM to comprise a mixture of FA and an undefined, 'inert' material having no ion-binding properties. No other organic ligands are considered. Non-humic, metal-binding organic ligands could influence speciation if present in sufficient quantities to compete with FAs. This is not, however, a structural issue with the model, but a limitation of the ability to characterise and model all the significant individual components of DOM. Some researchers have suggested a significant role for non-humic ligands in controlling freshwater metal speciation. For example, Xue and Sigg[36] suggested that the strong binding and low free ion concentrations of copper and cadmium found in some Swiss lakes were attributable to small amounts of ligands having stronger binding than humic substances, Rozan and Benoit[37] suggested that copper speciation in rivers of southern New England was controlled in part by sulphide and ethylenediaminetetraacetic acid (EDTA), and Baken et al.[38] found that EDTA was an important control on metal speciation in anthropogenically affected waters, particularly for Ni. Thus, where the model overestimates the observed free ion it is reasonable to postulate that this may be attributable to unaccounted-for strong ligands. Clearly there is a need to further investigate the occurrence and metal-binding properties of such ligands. Methods to distinguish different types of DOM in waters exist (e.g. Murphy et al.[39]) and research on structural identification of individual DOM components (e.g. Woods et al.[40]) is ongoing. Parallel application of such methods in speciation studies may assist in identifying and characterising non-humic components of DOM and their role in metal binding.

We now turn to the precision and accuracy of measurements of the free metal ion. Of the two, precision (i.e. reproducibility) is the easier to quantify. We did not consider precision when analysing the agreement between model and prediction, because it was not quantified in all the studies and so a consistent analysis of its role was not feasible. More emphasis is needed on quantifying the precision of speciation measurements, as it is a key aspect of the assessment both of the method and of the comparative performance of speciation models. Measurement accuracy is more challenging to assess, particularly for the complex methods used to measure free metal ions. Method validation in simple matrices under laboratory conditions may not hold in the more complex chemical environment

presented by a field sample. The accuracy of speciation methods has accordingly been questioned in the literature. For example, van Leeuwen and Town[41] have questioned the results of competitive ligand exchange voltammetry, arguing that equilibrium between the natural and added competing ligands is not necessarily attained within the experimental timeframe, and that this introduces bias into the reported free ion concentrations by underestimation of the actual equilibrium value. In this case, the argument against the accuracy of the method is consistent with the model predictions in that WHAM tends to predict a higher free metal ion than is measured by voltammetry. Comparing the model with those datasets not obtained using voltammetry produces decreases in bias for Cu, Cd and Pb, and RMSE for Cu, Zn, Cd and Pb (Table 6). However, this cannot be taken as definitive evidence that voltammetry gives inaccurate measurements, as this would mean assuming that the model predictions are accurate. More systematic evaluation of the accuracy of speciation methods is needed. A useful first step would be to move towards standardisation of methods, to make their intercomparison more reliable.[42] Intercomparison among methods should provide insights not obtainable by comparison solely with models that are based on key a-priori assumptions about the system under study. This is not to say, however, that models do not have a useful role to play in the evaluation of measurement methods. Models provide an internally consistent set of predictions against which to compare multiple measurement methods and gain additional information on the relative performance of different methods under different chemical conditions. The collation of literature data presented here shows that comparative measurements are rare, being confined to one study. Unsworth and coworkers[16] compared DMT and PLM measurements of free ionic Cu and Pb in three samples and of Cd in two samples. This study is a beginning, but as the number of samples is limited there is currently little scope to systematically assess the relative performance of the methods.

If the observations obtained using voltammetry are excluded, the bias between the model predictions and observations is below 0.5 orders of magnitude for all the metals except Ni (Tables 5, 6). Bias above 0.5 orders of magnitude in individual datasets, e.g. Al-03 and Cd-04, could be reduced by model 'calibration', such as adjustment of F_{FADOC} within plausible ranges. A recent modelling study of Al speciation data from the UK Acid Waters Monitoring Network[43] showed that the data could be simulated using plausible general values of F_{FADOC}, one for streams and one for lakes, similar in magnitude to the central value adopted here.

Speciation models are increasingly being incorporated into risk assessment models such as the Biotic Ligand Model (BLM).[44] Accuracy and precision of speciation model predictions are thus important for the accuracy and precision of the predictions of such models. Judgements as to what constitutes 'reasonable' or 'acceptable' accuracy and precision in model predictions are contingent on the purpose of the specific model. For example, the purpose of the BLM is to describe the variability in metal toxicity across varying water compositions. The BLM has been shown to accomplish this purpose for several metals and organisms (e.g. De Schamphelaere and Janssen[45] and De

Schamphelaere et al.[46]). This suggests that the humic-binding submodel in the BLM is performing well under the conditions of the metal toxicity tests, although bias in the model predictions will be compensated for to some extent if the BLM is calibrated using data generated in waters containing DOM. The performance of the BLM is qualitatively consistent with the observation, in the current study, that WHAM predictions of the free metal ion tend to agree with observations best when the latter are relatively high, where toxic effects are most likely to be seen under laboratory conditions (~10^{-8} M, 10^{-9} M, 10^{-7} M, 10^{-10} M and 10^{-9} M for Ni, Cu, Zn, Cd and Pb). It is feasible that future extension of the BLM to chronic effects of sensitive freshwater organisms may extend the range of its application to lower free ion concentrations, where agreement between observation and model prediction is sometimes poorer. This possibility needs to be borne in mind in future development of biotic ligand models. Ideally, the influence of uncertainty in WHAM predictions should be incorporated into biotic ligand models so that the propagation of uncertainties into predictions of toxicity can be assessed. Similar considerations apply to other environmental models having a speciation component, such as CHUM[47] and TICKET-UWM.[48]

A full consideration of the performance of WHAM, or any other model, can only ideally be done if the precision of the observations and variability in the model predictions can be quantified and the accuracy of the observations can be relied upon. In this study we have made a first quantification of the uncertainty (i.e. the variability) of the model predictions and a comparison of predictions with available observations. A possible next step would be to extend the calculation of uncertainty to include solution complexation. This has been previously done for actinides,[49] where complexation to inorganic ligands such as carbonate may be important. Looking to the future, we would recommend the following research steps to progress the understanding and assessment of metal speciation measurement and modelling in freshwaters:

- A more comprehensive approach to speciation measurements, including the simultaneous testing of multiple methods and parallel characterisation of DOM structure and composition, in as wide a range of waters as possible.
- Explicit and routine consideration of the influence of model uncertainty on the outputs of geochemical and toxicity models, and assessment of the implications for specific model applications such as risk assessment.

Conclusions

- Discrepancies between observations and predictions of free Al were generally attributable to uncertainties in input variables and model parameters, however, this was not the case for Co, Ni, Cu, Zn, Cd and Pb.
- Generally, the best agreements between measurements and predictions were seen for Al, Co, Ni, Zn and Cd. Agreements for Cu and Pb were dependent upon the analytical method used to

measure the free ion: measurements made using voltammetry were largely overestimated by the model. Generally, the model agreed best with measurements made using the DMT, IET and PLM methods.

- More parallel comparisons of multiple analytical methods and modelling are needed to better understand the reasons for the observed patterns and discrepancies between observations and model predictions. More comprehensive characterisation of the different components of DOM is needed.
- More emphasis needs to be placed generally on quantifying the precision (uncertainty) of measurements and model predictions, as this is essential for assessing the accuracy of model predictions.

Glossary of terms and parameters used in this study

n_A, density of Type A binding sites on humic or fulvic acid

$K_{A/B}$, equilibrium constant for proton binding to Type A or Type B sites on humic or fulvic acid (Eqn 1)

$K_{MA/MB}$, equilibrium constant for metal binding to Type A or Type B sites on humic or fulvic acid (Eqn 2)

ΔLK_2, heterogeneity term used to generate strong metal binding sites on humic or fulvic acid

$K_{1,oxide,}$ equilibrium constant for protonation of metal oxide surface site (Eqn 5)

$K_{2,oxide,}$ equilibrium constant for deprotonation of metal oxide surface site (Eqn 6)

$K_{MH,oxide}$, equilibrium constant for metal-proton exchange at metal oxide surface site (Eqn 7)

$\Delta pK_{MH,oxide}$, heterogeneity term used to generate strong metal binding sites on metal oxide surface

$x_{i,}$ central value of input variable or parameter for which uncertainty is to be considered

$x'_{i,j,}$ perturbed value of input variable or parameter, used for calculation of speciation in a single Monte Carlo sample

q, perturbation factor used to generate $x'_{i,j}$ values, generated by random sampling from a standard normal distribution

$\sigma_{i,}$ uncertainty factor for input variable or parameter having absolute uncertainty

$p_{i,}$ uncertainty factor for input variable or parameter having relative uncertainty

F_{FADOC}, factor relating observed dissolved organic carbon concentrations to fulvic acid concentrations for the purposes of speciation calculation. A value of two indicates that DOC is 100% active with respect to ion binding

$Q_{16–84,}$ the interquartile range between the 16th and 84th percentiles of the distribution of calculated free metal ion concentrations

$Q_{2.5–97.5,}$ the interquartile range between the 2.5th and 97.5th percentiles of the distribution of calculated free metal ion concentrations

bias, the mean of absolute deviations between observed and calculated log [free ion] in a dataset. A positive bias indicates that on average predicted concentrations exceed observed ones, and vice versa.

Accessory publication

Additional information on the sources of pH, DOC and major solute concentrations used in the field study modelling, and the calculation of the solubility express ion for iron(III) hydroxide, is given in the Accessory publication (see http://www.publish.csiro.au/?act = view_file&file_id=EN11049_AC.pdf).

Acknowledgements

The authors thank K. Delbeke, C. E. Schlekat, I. Schoeters, J. Meyer, R. L. Dwyer, D. S. Smith and R. C. Santore for their comments and suggestions on the report from which this paper was developed. They also thank Don Monteith for providing aluminium speciation data from the UK Acid Waters Monitoring Network, and Claude Fortin and Peter Campbell for providing the raw data for datasets Cu-09, Zn-06 and Cd-06. This work was funded by the following organisations: International Copper Association (ICA), International Council on Mining and Minerals (ICMM), International Lead Zinc Research Organization (ILZRO), Nickel Producers Environmental Research Association (NiPERA), Cobalt Development Institute (CDI), Rio Tinto Minerals, International Chromium Development Association (ICDA), International Molybdenum Association (IMOA), European Aluminium Association (EAA), UK Natural Environment Research Council (NERC).

References

[1] M. F. Benedetti, C. J. Milne, D. G. Kinniburgh, W. H. van Riemsdijk, L. K. Koopal, Metal ion binding to humic substances: application of the nonideal competitive adsorption model. Environ. Sci. Technol. 1995, 29, 446.

[2] E. Tipping, Humic Ion Binding Model VI: an improved description of the interactions of protons and metal ions with humic substances. Aquat. Geochem. 1998, 4, 3.

[3] E. Tipping, WHAM – a chemical-equilibrium model and computer code for waters, sediments, and soils incorporating a discrete site electrostatic model of ion-binding by humic substances. Comput. Geosci. 1994, 20, 973.

[4] D. N. Di Toro, H. E. Allen, H. L. Bergman, J. S. Meyer, P. R. Paquin, R. C. Santore, Biotic ligand model of the acute toxicity of metals. 1. Technical basis. Environ. Toxicol. Chem. 2001, 20, 2383.

[5] S. Lofts, E. Tipping, Solid-solution metal partitioning in the Humber Rivers: application of WHAM and SCAMP. Sci. Total Environ. 2000, 251–252, 381.

[6] E. J. J. Kalis, L. Weng, F. Dousma, E. J. M. Temminghoff, W. H. van Riemsdijk, Measuring free metal ion concentrations in situ in natural waters using the Donnan membrane technique. Environ. Sci. Technol. 2006, 40, 955.

[7] J. Qian, H. B. Xue, L. Sigg, A. Albrecht, Complexation of cobalt by natural ligands in freshwater. Environ. Sci. Technol. 1998, 32, 2043.

[8] S. Meylan, N. Odzak, R. Behra, L. Sigg, Speciation of copper and zinc in natural freshwater: comparison of voltammetric measurements, diffusive gradients in thin films (DGT) and chemical equilibrium models. Anal. Chim. Acta 2004, 510, 91.

[9] H. B. Xue, A. Oestreich, D. Kistler, L. Sigg, L., Free cupric ion concentrations and Cu complexation in selected Swiss lakes and rivers. Aquat. Sci. 1996, 58, 69.

[10] A. Plöger, E. Fischer, H.-P. Nirmaier, L. M. Laglera, D. Monticelli, C. M. G. van den Berg, Lead and copper speciation in remote mountain lakes. Limnol. Oceanogr. 2005, 50, 995.

[11] H. B. Xue, L. Sigg, Free cupric ion concentration and CuII speciation in a eutrophic lake. Limnol. Oceanogr. 1993, 38, 1200.

[12] H. B. Xue, L. Sigg, Zinc speciation in lake waters and its determination by ligand-exchange with EDTA and differential-pulse anodic-stripping voltammetry. Anal. Chim. Acta 1994, 284, 505.

[13] J. Cao, H. B. Xue, L. Sigg, Effects of pH and Ca competition on complexation of cadmium by fulvic acids and by natural organic ligands from a river and a lake. Aquat. Geochem. 2006, 12, 375.

[14] C. Fortin, P. G. C. Campbell, An ion-exchange technique for free-metal ion measurements (Cd2+, Zn2+): applications to complex aqueous media. Int. J. Environ. Anal. Chem. 1998, 72, 173.

[15] Y. Gopalapillai, C. L. Chakrabarti, D. R. S. Lean, Assessing toxicity of mining effluents: equilibrium- and kinetics-based metal speciation and algal bioassay. Environ. Chem. 2008, 5, 307.

[16] E. R. Unsworth, K. W. Warnken, H. Zhang, W. Davison, F. Black, J. Buffle, J. Cao, R. Cleven, J. Galceran, P. Gunkel, E. Kalis, D. Kistler, H. P. van Leeuwen, M. Martin, S. Noël, Y. Nur, N. Odzak, J. Puy, W. van Riemsdijk, L. Sigg, E. J. M. Temminghoff, M.-L. Tercier-Waeber, S. Topperwien, R. M. Town, L. Weng, H. B. Xue, Model predictions of metal speciation in freshwaters compared to measurements by in situ techniques. Environ. Sci. Technol. 2006, 40, 1942.

[17] Y. Gopalapillai, I. I. Fasfous, J. D. Murimboh, T. Yapici, P. Chakraborty, C. L. Chakrabarti, Determination of free nickel ion concentrations using the ion exchange technique: application to aqueous mining and municipal effluents. Aquat. Geochem. 2008, 14, 99.

[18] E. J. M. Temminghoff, A. C. C. Plette, R. van Eck, W. H. Van Riemsdijk, Determination of the chemical speciation of trace metals in aqueous systems by the Wageningen Donnan Membrane technique. Anal. Chim. Acta 2000, 417, 149.

[19] S. E. Bryan, E. Tipping, J. Hamilton-Taylor, Comparison of measured and modelled copper binding by natural organic matter in freshwater. Comp. Biochem. Physiol. C 2002, 133, 37.

[20] E. Tipping, C. Woof, M. A. Hurley, Humic substances in acid surface waters; modelling aluminium binding, contribution to ionic charge-balance, and control of pH. Water Res. 1991, 25, 425.

[21] G. M. Anderson, Error propagation by the Monte Carlo method in geochemical calculations. Geochim. Cosmochim. Acta 1976, 40, 1533.

[22] J. E. Groenenberg, G. F. Koopmans, R. N. J. Comans, Uncertainty analysis of the nonideal competitive adsorption–Donnan model: effects of dissolved organic matter variability on predicted metal speciation in soil solution. Environ. Sci. Technol. 2010, 44, 1340.

[23] E. Tipping, S. Lofts, J. E. Sonke, Humic Ion-Binding Model VII: a revised parameterisation of cation-binding by humic substances. Environ. Chem. 2011, 8, 225.

[24] S. Lofts, E. Tipping, An assemblage model for cation binding by natural particulate matter. Geochim. Cosmochim. Acta 1998, 62, 2609.

[25] E. Tipping, C. Rey-Castro, S. E. Bryan, J. Hamilton-Taylor, AlIII and FeIII binding by humic substances in freshwaters, and implications for trace metal speciation. Geochim. Cosmochim. Acta 2002, 66, 3211.

[26] S. Lofts, E. Tipping, J. Hamilton-Taylor, The chemical speciation of FeIII in freshwaters. Aquat. Geochem. 2008, 14, 337.

[27] X. Liu, F. J. Millero, The solubility of iron hydroxide in sodium chloride solutions. Geochim. Cosmochim. Acta 1999, 63, 3487.

[28] D. T. Monteith, C. D. Evans, The United Kingdom Acid Waters Monitoring Network: a review of the first 15 years and introduction to the special issue. Environ. Pollut. 2005, 137, 3.

[29] C. T. Driscoll, A procedure for the fractionation of aqueous aluminium in dilute acid waters. Int. J. Environ. Anal. Chem. 1984, 16, 267.

[30] D. A. Dzombak, F. M. M. Morel, Surface Complexation Modeling: Hydrous Ferric Oxide 1990 (Wiley: New York).

[31] B. Lyvén, M. Hassellov, D. R. Turner, C. Haraldsson, K. Andersson, Competition between iron- and carbon-based colloidal carriers for trace metals in a freshwater assessed using flow field-flow fractionation coupled to ICPMS. Geochim. Cosmochim. Acta 2003, 67, 3791.

[32] J. A. B. Bass, R. Blust, R. T. Clarke, T. A. Corbin, W. Davison, K. A. C. De Schamphelaere, C. R. Janssen, E. J. J. Kalis, M. G. Kelly, N. T. Kneebone, A. J. Lawlor, S. Lofts, E. J. M. Temminghoff, S. A. Thacker, E. Tipping, C. D. Vincent, K. W. Warnken, H. Zhang, Environmental Quality Standards for trace metals in the aquatic environment. Environment Agency Science Report – SC030194 2008 (Environment Agency: Bristol, UK).

[33] L. Van Laer, E. Smolders, F. Degryse, C. Janssen, K. A. C. De Schamphelaere, Speciation of nickel in surface waters measured with the Donnan membrane technique. Anal. Chim. Acta 2006, 578, 195.

[34] H. B. Xue, L. Sigg, Comparison of the complexation of Cu and Cd by humic or fulvic acids and by ligands observed in lake waters. Aquat. Geochem. 1999, 5, 313.

[35] T. F. Rozan, G. Benoit, Geochemical factors controlling free Cu ion concentrations in river water. Geochim. Cosmochim. Acta 1999, 63, 3311.

693 [36] S. Baken, F. Degryse, L. Verheyen, R. Merckx, E. Smolders, Metal complexation properties of freshwater
694 dissolved organic matter are explained by its aromaticity and by anthropogenic ligands. Environ. Sci. Technol.
695 2011, 45, 2584.

696 [37] K. R. Murphy, C. A. Stedmon, T. D. Waite, G. M. Ruiz, Distinguishing between terrestrial and
697 autochthonous organic matter sources in marine environments using fluorescence spectroscopy. Mar. Chem.
698 2008, 108, 40.

699 [38] G. C. Woods, M. J. Simpson, P. J. Koerner, A. Napoli, A. J. Simpson, HILIC-NMR: toward the
700 identification of individual molecular components in dissolved organic matter. Environ. Sci. Technol. 2011, 45,
701 3880.

702 [39] H. P. van Leeuwen, R. M. Town, Kinetic limitations in measuring stabilities of metal complexes by
703 competitive ligand exchange-adsorptive stripping voltammetry (CLEAdSV). Environ. Sci. Technol. 2005, 39,
704 7217.

705 [40] M. Filella, R. Town, J. Buffle, Speciation in freshwaters, in Chemical speciation in the environment (Eds.
706 A. M. Ure, C. M. Davidson) 1995, pp. 169–200 (Chapman & Hall: Glasgow).

707 [41] E. Tipping, H. T. Carter, Aluminium speciation in streams and lakes of the UK Acid Waters Monitoring
708 Network, modelled with WHAM. Sci. Total Environ. 2011, 409, 1550.

709 [42] D. M. Di Toro, H. E. Allen, H. L. Bergman, J. S. Meyer, P. R. Paquin, R. C. Santore, Biotic ligand model
710 of the acute toxicity of metals. 1. Technical basis. Environ. Toxicol. Chem. 2001, 20, 2383.

711 [43] K. A. C. De Schamphelaere, C. R. Janssen, Development and field validation of a biotic ligand model
712 predicting chronic copper toxicity to Daphnia magna. Environ. Toxicol. Chem. 2004, 23, 1365.

713 [44] K. A. C. De Schamphelaere, S. Lofts, C. R. Janssen, Bioavailability models for predicting acute and
714 chronic toxicity of zinc to algae, daphnids, and fish in natural surface waters. Environ. Toxicol. Chem. 2005, 24,
715 1190.

716 [45] E. Tipping, A. J. Lawlor, S. Lofts, L. Shotbolt, Simulating the long-term chemistry of an upland UK
717 catchment: heavy metals. Environ. Pollut. 2006, 141, 139.

718 [46] K. J. Farley, R. F. Carbonaro, C. J. Fanelli, R. Costanzo, K. J. Rader, D. M. Di Toro, TICKET-UWM: a
719 coupled kinetic, equilibrium, and transport screening model for metals in lakes. Environ. Toxicol. Chem. 2011,
720 30, 1278.

721 [47] F. H. Denison, J. Garnier-Laplace, The effects of database parameter uncertainty on uranium(VI)
722 equilibrium calculations. Geochim. Cosmochim. Acta 2005, 69, 2183.

723 [48] J. W. Guthrie, N. M. Hassan, M. S. A. Salam, I. I. Fasfous, C. A. Murimboh, J. Murimboh, C. L.
724 Chakrabarti, D. C. Grégoire, Complexation of Ni, Cu, Zn, and Cd by DOC in some metal-impacted freshwater
725 lakes: a comparison of approaches using electrochemical determination of free-metal-ion and labile complexes
726 and a computer speciation model, WHAM V and VI. Anal. Chim. Acta 2005, 528, 205.

[49] C. Fortin, Y. Couillard, B. Vigneault, P. G. C. Campbell, Determination of Free Cd, Cu and Zn concentrations in lake waters by in situ diffusion followed by column equilibration ion-exchange. Aquat. Geochem. 2010, 16, 151.

Manuscript received 13 April 2011, accepted 26 July 2011

732 **Table 1. Water compositions used for theoretical calculations**

733 M = Co, Ni, Cu, Zn, Cd, Pb; FA = fulvic acid

Determinand	Composition for calculation of uncertainty in $[Al^{3+}]$	Composition for calculation of uncertainty in $[M^{2+}]$
Temperature (K)	293	293
FA (mg L^{-1})	10.0	5.0
Dissolved Na (M)	1.0×10^{-4}	5.0×10^{-4}
Dissolved Mg (M)	5.0×10^{-5}	5.0×10^{-4}
Dissolved Al (M)	5.0×10^{-6} [A]	1.0×10^{-5} [A]
Dissolved K (M)	1.0×10^{-5}	5.0×10^{-5}
Dissolved Ca (M)[B]	5.1×10^{-5}–1.3×10^{-4}	4.7×10^{-4}–1.7×10^{-3}
Dissolved Fe^{III} (M)	1.0×10^{-6} [A]	1.0×10^{-6} [A]
Dissolved Co (M)	–	1.0×10^{-9}
Dissolved Ni (M)	–	5.0×10^{-9}
Dissolved Cu (M)	–	2.0×10^{-8}
Dissolved Zn (M)	–	1.0×10^{-7}
Dissolved Cd (M)	–	1.0×10^{-9}
Dissolved Pb (M)	–	1.0×10^{-8}
Dissolved Cl (M)	5.0×10^{-4}	1.0×10^{-3}
Dissolved NO_3 (M)	5.0×10^{-6}	5.0×10^{-4}
Dissolved SO_4 (M)	1.0×10^{-4}	5.0×10^{-4}
pCO_2 (Pa)	36.50	36.50

734 [A]$Al(OH)_3$ and $Fe(OH)_3$ allowed to precipitate using the parameters given in the main text.

735 [B]Adjusted to maintain charge balance across the pH range simulated.

736

737 **Table 2. Field datasets used for WHAM/Model VII testing**

Metal	Measurement method	*n*	Code	Reference
Al	Cation exchange column + speciation	55	Al-01	[28]
	Cation exchange column + speciation	180	Al-02	[28]
	Cation exchange column + speciation	167	Al-03	[28]
Co	CLE-DPCSV[A]	7	Co-01	[7]
Ni	IET[B]	3	Ni-01	[17]
	DMT[C]	3	Ni-02	[16]
	DMT	7	Ni-03	[6]
	DMT	35	Ni-04	[32]
	DMT	6	Ni-05	[35]
Cu	CLE-AdCSV[D]	2	Cu-01	[48]
	CLE-DPCSV	14	Cu-02	[8]
	CLE-DPCSV	38	Cu-03	[9]
	CLE-CSV[E]	5	Cu-04	[10]
	CLE-CSV	15	Cu-05	[11]
	PLM[F]	2	Cu-06a	[16]
	DMT	3	Cu-06b	[16]
	DMT	7	Cu-07	[6]
	DMT	35	Cu-08	[32]
	IET	10	Cu-09	[49]
Zn	CLE-DPASV[G]	4	Zn-01	[48]
	CLE-DPASV	12	Zn-02	[8]
	CLE-DPASV	6	Zn-03	[12]
	DMT	5	Zn-04	[6]
	DMT	34	Zn-05	[32]
	IET	24	Zn-06	[49]
Cd	CLE-AdCSV	4	Cd-01	[48]
	CLE-DPASV	11	Cd-02	[13]
	PLM	3	Cd-03a	[16]
	DMT	2	Cd-03b	[16]
	DMT	6	Cd-04	[6]
	DMT	35	Cd-05	[32]
	IET	24	Cd-06	[49]
Pb	CLE-DPCSV	28	Pb-01	[10]
	PLM	2	Pb-02a	[16]
	DMT	3	Pb-02b	[16]
	DMT	7	Pb-03	[6]
	DMT	35	Pb-04	[32]

738 [A]Competitive ligand exchange–differential pulse cathodic stripping voltammetry.

739 [B]Ion exchange column method.

740 [C]Donnan membrane technique.

741 [D]Competitive ligand exchange–adsorptive cathodic stripping voltammetry.

742 [E]Competitive ligand exchange–cathodic stripping voltammetry.

743 [F]Permeation liquid membrane.

744 [G]Competitive ligand exchange–differential pulse anodic stripping voltammetry.

745

Table 3. Uncertainties in WHAM/Model VII inputs used in this study

Variable	Uncertainty type	σ_i or p_i
Input variables		
Temperature (K)	Absolute	1
pCO_2	Relative	0.025
pH	Absolute	0.05
Colloidal fulvic acid (g L^{-1})	Relative	0.09
Solute concentrations[A] (M)	Relative	0.025
Parameters – Model VII		
$\log K_{MA}$	Absolute	0.3
Parameters – iron(III) oxide		
$pK_{MH,oxide}$	Absolute	0.3
Solution speciation		
$\log K_{SO,Fe(OH)3}$	Absolute	0.7
$\log K_{SO,Al(OH)3}$	Absolute	0.7

[A]Solutes comprise dissolved metals, major ions and alkalinity.

Table 4. Bias, error (root mean square error, RMSE) and variability in WHAM predictions of field speciation of Al, Co, Ni, Cu, Zn and Cd, and numbers of observations falling within the interquartile range 2.5–97.5 % of the predictions

For variability ($Q_{16–84}$), the interquartile range of the predictions of log free metal ion is 16–84 %. Input variables include uncertainty in F_{FADOC}. For observations within $Q_{2.5–97.5}$, the interquartile range of the predictions of log free metal ion is 2.5–97.5 %. Bold entries refer to bias, error and variability in predictions for each metal as a whole.

Metal	Dataset	n_{obs}	Mean observed log[M^{z+}]	Mean predicted log[M^{z+}]	Bias	RMSE	Variability ($Q_{16–84}$)		Observations	
							Input variables	Parameters	All	within $Q_{2.5–97.5}$
Al	Al-01	55	–6.03	–6.23	0.20	0.29	0.68	0.44	0.82	55
	Al-02	180	–7.69	–7.84	0.16	0.29	0.56	0.56	0.82	173
	Al-03	167	–6.01	–6.73	0.72	0.81	0.60	0.64	0.88	127
	All Al	**402**	**–6.76**	**–7.14**	**0.38**	**0.54**	**0.60**	**0.58**	**0.84**	**355**
Co	Co-01	7	–9.74	–9.36	0.33	0.40	0.14	0.0024	0.14	0
Ni	Ni-01	3	–6.65	–6.24	0.41	0.46	0.14	0.0086	0.15	0
	Ni-02	3	–9.54	–9.24	0.30	0.46	0.18	0.058	0.19	1
	Ni-03	7	–9.35	–8.48	0.86	0.95	0.18	0.030	0.19	0
	Ni-04	35	–8.75	–8.00	0.75	0.92	0.16	0.052	0.17	1
	Ni-05	6	–8.01	–7.89	0.12	0.30	0.18	0.084	0.20	2
	All Ni	**54**	**–8.67**	**–8.02**	**0.65**	**0.83**	**0.16**	**0.050**	**0.18**	**4**
Cu	Cu-01	2	–8.86	–10.71	–1.84	1.84	0.44	0.56	0.76	0
	Cu-02	14	–13.62	–10.29	3.34	3.39	0.44	0.46	0.62	0
	Cu-03	39	–13.44	–10.89	2.55	2.73	0.40	0.62	0.76	4
	Cu-04	5	–14.82	–10.62	4.30	4.30	0.32	0.76	0.84	0
	Cu-05	15	–15.05	–11.47	3.58	3.97	0.46	0.64	0.80	0
	Cu-06a	3	–10.37	–11.71	–1.33	1.48	0.38	0.62	0.68	1
	Cu-06b	3	–10.36	–11.71	–1.34	1.53	0.38	0.62	0.68	1
	Cu-07	7	–10.46	–10.90	–0.44	1.19	0.44	0.58	0.76	2
	Cu-08	35	–10.19	–10.11	0.078	1.19	0.38	0.48	0.64	12
	Cu-09	10	–8.02	–9.47	–1.44	1.76	0.48	0.58	0.78	3
	All Cu	**133**	**–12.00**	**–10.87**	**1.76**	**2.62**	**0.40**	**0.54**	**0.68**	**23**
Zn	Zn-01	4	–7.15	–7.42	–0.27	0.57	0.26	0.28	0.40	4
	Zn-02	12	–8.31	–7.52	0.80	0.84	0.15	0.046	0.15	0
	Zn-03	5	–8.68	–8.28	0.40	0.41	0.15	0.054	0.16	0
	Zn-04	4	–8.00	–7.27	0.73	0.84	0.18	0.11	0.22	1
	Zn-05	34	–6.44	–6.59	0.15	0.33	0.21	0.15	0.28	16
	Zn-06	25	–7.44	–8.20	–0.76	0.84	0.30	0.34	0.45	2
	All Zn	**84**	**–7.26**	**–7.38**	**–0.12**	**0.65**	**0.22**	**0.19**	**0.30**	**23**
Cd	Cd-01	4	–9.61	–8.91	0.69	0.81	0.24	0.19	0.30	3
	Cd-02	11	–12.00	–10.85	1.16	1.26	0.22	0.19	0.28	2
	Cd-03a	3	–11.22	–10.57	0.65	1.10	0.20	0.13	0.26	0
	Cd-03b	2	–11.61	–10.64	0.97	1.12	0.20	0.13	0.26	0
	Cd-04	5	–11.19	–10.00	1.19	1.27	0.17	0.060	0.18	0
	Cd-05	35	–9.29	–9.11	0.19	0.33	0.18	0.092	0.22	20
	Cd-06	25	–9.54	–9.90	–0.36	0.42	0.28	0.26	0.38	11
	All Cd	**85**	**–9.96**	**–9.71**	**0.27**	**0.71**	**0.22**	**0.16**	**0.28**	**36**

Table 5. Bias, error (root mean square error, RMSE) and variability in WHAM predictions of field speciation of Pb, and numbers of observations falling within the interquartile range 2.5–97.5 % of the predictions

For variability ($Q_{16–84}$), the interquartile range of the predictions of log free metal ion is 16–84 %. Input variables include uncertainty in F_{FADOC}. For observations within $Q_{2.5–97.5}$, the interquartile range of the predictions of log free metal ion is 2.5–97.5 %. Bold entries refer to bias, error and variability in predictions for each metal as a whole.

Dataset	n_{obs}	Mean observed log[M^{z+}]	Mean predicted log[M^{z+}]	Bias	RMSE	Variability ($Q_{16–84}$)			Observations within $Q_{2.5–97.5}$
						Input variables	Parameters	All	
Not considering binding by iron(III) oxide									
Pb-01	28	–13.20	–10.56	2.64	2.83	0.28	0.36	0.42	1
Pb-02a	3	–11.49	–10.70	–0.53	0.53	0.22	0.16	0.56	0
Pb-02b	3	–11.30	–10.70	–0.72	0.72	0.22	0.16	0.56	0
Pb-03	7	–11.90	–11.40	0.50	0.93	0.26	0.20	0.32	1
Pb-04	35	–9.81	–9.38	0.46	1.12	0.34	0.36	0.48	5
All Pb	**76**	**–11.38**	**–10.11**	**1.17**	**1.69**	**0.30**	**0.33**	**0.45**	**7**
Considering binding by iron(III) oxide									
Pb-01	28	–13.20	–11.48	1.72	1.96	0.26	0.20	0.54	2
Pb-02a	3	–11.49	–12.51	–1.02	1.09	0.20	0.22	0.30	0
Pb-02b	3	–11.30	–12.51	–1.21	1.29	0.20	0.22	0.30	0
Pb-03	7	–11.90	–11.77	0.13	0.85	0.22	0.22	0.26	1
Pb-04	35	–9.81	–9.50	0.31	1.14	0.26	0.35	0.46	6
All Pb	**76**	**–11.38**	**–10.68**	**0.70**	**1.48**	**0.26**	**0.28**	**0.46**	**9**

Table 6. Bias and scatter (root mean squared error, RMSE) in WHAM predictions of field speciation of Cu, Zn, Cd and Pb, for all free ion observations and for observations not obtained using voltammetry

Parameter		Copper		Zinc		Cadmium		Lead	
		All	Not including voltammetry	All	Not including voltammetry	All	Not including voltammetry	All	Not including voltammetry
n_{obs}		133	58	84	63	85	70	76	48
Mean log[M^{z+}]	observed	–12.00	–9.87	–7.26	–6.94	–9.96	–9.66	–11.38	–10.31
Mean log[M^{z+}]	predicted	–10.87	–10.26	–7.38	–7.27	–9.71	–9.56	–10.68	–10.21
Bias		1.76	–0.39	–0.12	–0.17	0.27	0.11	0.70	0.11
RMSE		2.62	1.32	0.65	0.56	0.71	0.48	1.48	1.10

Fig. 1. Distribution of estimated dissolved organic carbon (DOC) 'activity', from Bryan and coworkers.[26] The DOC 'activity' (F_{FADOC}) is the ratio of model-optimised fulvic acid to measured DOC. The dashed lines represent the mean, mean + 1 standard deviation and mean – 1 standard deviation of log F_{FADOC}. The Anderson–Darling test indicated no significant departure from normality at a significance level of $P = 0.05$.

Fig. 2. Theoretical calculations of variability in free ion concentrations of Al: calculated free concentrations and $Q_{16–84}$ ranges for the chemical conditions given in Table 3 and pH 5.0 (closed symbols), pH 6.0 (open symbols), for four uncertainty scenarios. 1 = uncertainty in F_{FADOC} only, 2 = uncertainty in F_{FADOC} and input variables, 3 = uncertainty in parameters only, 4 = uncertainty in F_{FADOC}, input variables and parameters together.

Fig. 3. Theoretical calculations of variability in free ion concentrations of Co, Ni, Cu, Zn, Cd and Pb: calculated free concentrations and $Q_{16–84}$ ranges for the chemical conditions given in Table 3 and pH 7.0, for four uncertainty scenarios. 1 = uncertainty in F_{FADOC} only, 2 = uncertainty in F_{FADOC} and input variables, 3 = uncertainty in parameters only, 4 = uncertainty in F_{FADOC}, input variables and parameters together.

Fig. 4. Comparison of Pb^{2+} concentrations calculated assuming iron(III) oxide to be an active binding phase with concentrations calculated assuming iron(III) oxide to be inert with respect to ion binding. FeOx = iron(III) oxide.

Fig. 5. Comparison of Al^{3+} concentrations calculated by speciation of inorganic monomeric Al and by speciation of total monomeric Al, for all Al datasets. Error bars indicate the $Q_{16–84}$ range predicted by WHAM/Model VII.

Fig. 6. Comparison of observed Co^{2+} concentrations with predictions of WHAM/Model VII, for dataset Co-01. Error bars indicate the $Q_{16–84}$ range predicted by WHAM/Model VII.

Fig. 7. Comparison of observed Ni^{2+} concentrations with predictions of WHAM/Model VII, for all Ni datasets. Vertical error bars indicate the $Q_{16–84}$ range predicted by WHAM/Model VII, horizontal error bars represent ±1 standard deviation of measurements, where quoted.

Fig. 8. Comparison of observed Cu^{2+} concentrations with predictions of WHAM/Model VII, for measurements made by voltammetry (top), by Donnan membrane technique (DMT) (middle) and by Permeation liquid membrane (PLM) or ion exchange column method (IET) (bottom). Vertical error bars indicate the $Q_{16–84}$ range predicted by WHAM/Model VII, horizontal error bars represent ±1 standard deviation of measurements, where quoted. CLE-AdCSV: competitive ligand exchange adsorptive cathodic stripping voltammetry; CLE-

DPCSV: competitive ligand exchange differential pulse cathodic stripping voltammetry; CLE-CSV: competitive ligand exchange cathodic stripping voltammetry.

Fig. 9. Comparison of observed Zn^{2+} concentrations with predictions of WHAM/Model VII, for measurements made by voltammetry (top) and by Donnan membrane technique (DMT) and ion exchange column method (IET) (bottom). Vertical error bars indicate the $Q_{16–84}$ range predicted by WHAM/Model VII, horizontal error bars represent ±1 standard deviation of measurements, where quoted. CLE-DPASV: competitive ligand exchange differential pulse anodic stripping voltammetry.

Fig. 10. Comparison of observed Cd^{2+} concentrations with predictions of WHAM/Model VII, for measurements made by voltammetry, permeation liquid membrane (PLM), and ion exchange column method (IET) (top), and by Donnan membrane technique (DMT) (bottom). Vertical error bars indicate the $Q_{16–84}$ range predicted by WHAM/Model VII, horizontal error bars represent ±1 standard deviation of measurements, where quoted. CLE-AdCSV: competitive ligand exchange adsorptive cathodic stripping voltammetry; CLE-DPCSV: competitive ligand exchange differential pulse cathodic stripping voltammetry.

Fig. 11. Comparison of observed Pb^{2+} concentrations with predictions of WHAM/Model VII, for measurements made by voltammetry and permeation liquid membrane (PLM) (top), and by Donnan membrane technique (DMT) (bottom). Vertical error bars indicate the $Q_{16–84}$ range predicted by WHAM/Model VII, horizontal error bars represent ±1 standard deviation of measurements, where quoted. CLE-DPCSV: Competitive ligand exchange differential pulse cathodic stripping voltammetry.