Predicting Heavy Metal Partition Equilibrium in Soils: Roles of Soil Components and Binding Sites

The reactivity and bioavailability of heavy metals in soils are controlled by their binding to reactive soil components, including soil organic matter (SOM), metal (hydr)oxides, and clay minerals. In this study, we specifically investigated how soil components and SOM binding sites controlled metal partition at various chemistry conditions. We used the Windermere Humic Aqueous Model (WHAM 7) to predict the solid-solution partition and speciation of Cd, Cu, Ni, Pb, and Zn based on compiled literature data including 98 soil samples from five continents. Based on the root-mean-square-error (RMSE) values of logarithm of dissolved metal concentrations between model predictions and experimental results, WHAM 7 reasonably predicted metal partition equilibrium over a wide range of reaction conditions, with RMSE less than 0.5 for Cd and Zn, and less than 1.0 for the other three metals. Soil organic matter dominated metal binding at most acidic to neutral pH, clay minerals were significant at low pH, and iron (hydr)oxides might effectively compete with SOM for metal binding when pH was high. For all five heavy metals, WHAM 7 predicted the bidentate bindings were the dominant form of metal complexes, in which both complexes formed by two carboxylic sites and that by one carboxylic and another phenolic sites were major complexes. The formation of monodentate complexes and electrostatic outer-sphere complexes was significant at low pH, while tridentate complexes were only significant at high pH values. The modeling results help to accurately predict the environmental behavior of heavy metals in pH 3 to 7 soil environments.

Core Ideas

• WHAM 7 successfully predicted heavy metal partition in various soils.
• Major soil and solution parameters influenced the accuracy of model predictions.
• Soil organic matter dominated metal binding at most acidic to neutral pH in soils.
• WHAM 7 predicted bidentate binding as the dominant form of metal complexes.
parameters were derived, but little is known about the reaction mechanisms, especially the roles of different soil binding sites.

Over the past decades, computer software containing mechanistic chemical speciation models, such as WHAM (Windermere Humic Aqueous Model) (Tipping, 1994, 1998; Tipping and Hurley, 1992; Tipping et al., 2011), VisualMINEQ (Gustafsson, 2015; Gustafsson et al., 2003), ORCHESTRA (Objects Representing CHEmical Speciation and TRAnsport) (Meeussen, 2003), and ECOSAT (Equilibrium Calculation Of Speciation And Transport) (Weng et al., 2001), have been developed, which are capable of describing metal distribution between soil and solution phases. Those modeling tools are based on mechanistic descriptions of metal reactions with various reactive soil components with a number of surface complexation models (SCM) and implementation of component additivity approach for multiple soil components. Although models in those software tools differ in site distributions of soil components and model parameters, they have been relatively successful when applied to describe metal partitioning between soil and solution when SOM is the dominant adsorbent (Gustafsson et al., 2003; Tipping et al., 2003; Weng et al., 2001). Metal binding to SOM can be described with Model VII (Tipping et al., 2011), NICA-Donnan (Non Ideal Competitive Adsorption Donnan) model (Benedetti et al., 1995), and SHM (Stockholm Humic Model) (Gustafsson, 2001), and a number of surface complexation models have also been used to describe metal reactions with mineral phases, such as the CCM (constant capacitance model) (Lofts and Tipping, 1998), the DLM (diffuse layer model) (Dzombak and Morel, 1990), and the CD-MUSIC (Charge Distribution and Multisite Surface Complexation) model (Hiemstra and van Riemsdijk, 1996).

While chemical speciation models have shown promising success for predicting metal partitioning in soils, there are a few key factors affecting the accuracy of the model predictions, including the reactive organic matter (ROM), reactive metals in field contaminated soils, and cation competition, which has not been assessed systematically based on large datasets. Soil organic matter is one of the most important soil components controlling cation binding in soil, and numerous modeling studies have shown SOM played a dominant role in metal partitioning in soil (Duffner et al., 2014; Gustafsson et al., 2003; Ponizovsky et al., 2006a; Shi et al., 2007, 2013a; Tipping et al., 2003; Weng et al., 2001). Since the chemical speciation models were mainly developed based on the data derived from the model humic substances, SOM contains a variety of sub-components other than humic substances and the amount of ROM has to be quantified when applied to soil. Different methods have been used to quantify ROM, including base extractions (Dijkstra et al., 2004; Lumsdon, 2004), soil cation exchange capacity analysis (Weng et al., 2001) and model fitting (Gustafsson et al., 2003; Shi et al., 2007, 2013b). So far, no universal method has been established and the variations of ROM have been reported according to different studies, which is likely affected by the soils studied and the modeling methods used. Another important parameter, when predicting metal partitioning in field contaminated soils, is the reactive portion of the total heavy metals in soil since not all metals are labile in field samples. Different chemical extraction methods have been used to quantify the reactive metals in soil to get optimal model performance. In the most recent study by Groenenberg et al. (2017), isotopic dilution was considered to be conceptually the most sound approach, while the single dilute (0.43 M) nitric acid extraction was a good alternative method to determine geochemically reactive metals in soil except for high pH. In addition, cation competition is another important factor that may affect the accuracy of model predictions and different methods have been used to quantify important cations in soil that are capable of competing metal binding, including Ca$^{2+}$, Mg$^{2+}$, Al$^{3+}$, and Fe$^{3+}$ (Shi et al., 2013a; Tipping et al., 2003). All those factors should be carefully accounted for when applying the chemical speciation models to predict metal partitioning in soil.

Although metal partition equilibrium have been studied in multiple studies using chemical speciation models, one important issue, which is largely overlooked in previous studies, is the distribution of metals among different soil binding sites of major soil components, especially for SOM that has a wide distribution of metal binding sites. From the metal partition equilibrium aspect, different assumptions on metal distribution among different SOM binding sites may make little practical significance if the model can properly predict the metal distribution between soil and solution under certain conditions. However, as suggested by the recent studies (Shi et al., 2013b, 2016; Tian et al., 2017), metal binding to different sites may have significantly different reaction rates and the metal dissociation/desorption rate constants may vary a few orders of magnitude for different SOM binding sites. Therefore, it is essential to understand metal binding to various SOM binding sites for the prediction of the dynamic behavior of metals in soils, since the equilibrium assumption is usually rarely achieved in soil environments (Sparks, 1989). As a result, in addition to prediction of the macroscopic parameters (e.g., dissolved metal concentrations, equilibrium partition coefficient), it is essential to understand metal binding to various SOM binding sites for a better environmental risk assessment.

For typical soil minerals, metal binding to surface binding sites of minerals and formation of various surface complexes have been well elucidated at the molecular level, with the aid of the spectroscopic techniques. For SOM, only few studies looked at metal binding to specific SOM sites at the molecular level (Gustafsson et al., 2007, 2014; Karlsson et al., 2006; Xiong et al., 2013). Metal binding to SOM binding sites has been largely based on theoretical consideration on the chemical properties of SOM and the formation of various metal complexes is mainly hypothetical, due to the difficulty of experimentally quantifying SOM binding sites. Both carboxylic and phenolic sites are considered to be two most important SOM binding sites, and, through various combinations of those two types of sites, SOM can form multiple metal binding sites with distinct proton and metal binding constants (Kinniburgh et al., 1999; Tipping and Hurley, 1992). For metal reactions with SOM, two types of chemical
models have been developed based on two different assumptions of the binding site distribution of humic substances, a discrete site distribution or a continuous site distribution. Both WHAM and the SHM (Gustafsson, 2001) assume multiple discrete binding sites for both carboxylic and phenolic functional groups while the NICA-Donnan model uses a statistical function to describe the continuous site distribution for both carboxylic and phenolic groups (Benedetti et al., 1995). It is therefore interesting to quantitatively understand metal distribution among different SOM sites under different reaction chemistry conditions.

The objective of this study was to elucidate the roles of soil components and SOM binding sites on controlling metal partition equilibrium and the major factors affecting the performance of the predictive equilibrium model, based on the large data sets collected from multiple publications that studied metal partition equilibrium. The data involved five typical heavy metals (Cd, Cu, Ni, Pb, and Zn), and either field contaminated soils or laboratory metal spiked soils with a wide range of compositions. The data sets included a total of over 1700 metal partition equilibrium data at various chemistry conditions. A customized version of WHAM 7, which is able to output metal binding to various SOM binding sites, was used to conduct model calculations. We evaluated the performance of WHAM 7 for predicting metal partitioning in soils. We analyzed how different soil components and SOM binding sites control metal partitioning under various chemistry conditions, based on the large data of metal partition under various conditions. We hope that our modeling results will help to understand the roles of specific soil components and binding sites for metal binding, and therefore provide basis for more accurately predicting the dynamic behavior of heavy metals in soil environments.

MATERIALS AND METHODS

Soil Properties and Data Description

Data published from a total of ten studies (Almas et al., 2007; Dijkstra et al., 2004; Duffner et al., 2014; Gustafsson et al., 2003; Lumsdon, 2004; Ponizovsky et al., 2006b, 2008; Shi et al., 2007, 2013a; Weng et al., 2001), which contained all needed information for WHAM calculations, were collected and analyzed in this study. It includes 98 soil samples from five continents (America, Australia, Africa, Asia and Europe) with various soil types and properties. The major soil properties are summarized in Table 1. The soil organic carbon (SOC) content ranged from 0.18% to 47.9%. Different extraction methods were used to quantify Fe and Al minerals (summarized in Table 2). Either ammonium oxalate or ascorbate extraction was used to extract the amorphous Fe and Al in the soils, and dithionite bicarbonate citrate or dithionite extraction was used to extract the amorphous and crystalline Fe.

Generally, metal speciation in field contaminated soils may differ significantly from the laboratory metal-spiked soils. Not all metals in field contaminated soils are as reactive as that in the laboratory metal-spiked soils. Four of the ten studies used field-contaminated samples and soil samples were extracted at laboratory with different concentrations of HNO3 (i.e., 0.22, 0.43, and 2 M HNO3) to quantify the reactive portion of heavy metals. This portion of metals was considered to be the most reactive and its distribution between soil and solution may be properly predicted by chemical speciation models (Duffner et al., 2014; Weng et al., 2001). Metal partition equilibrium was then studied with either batch extraction or Donnan membrane technique as specified in each study. The rest of studies used uncontaminated soils and metal adsorption experiments were conducted to obtain adsorption isotherms or adsorption edges at various reaction conditions. Those studies involved five heavy metals, Cd, Cu, Ni, Pb and Zn, either as a mixture or single element as specified in Table 1.

Data at pH lower than 3.0 and higher than 7.0 were not analyzed in this work. When pH is too low, the Fe3+ and Al3+ activities cannot be reliably estimated which may significantly affect the accuracy of the model predictions (Shi et al., 2013a). While at pH > 7.0, the precipitation of metals may occur which is not considered in WHAM 7 calculations and some of the studies did not have the dissolved organic matter (DOM) data, which may significantly affect the model performance.

Brief Description of WHAM 7

Humic Ion-Binding Model VII for SOM and DOM

WHAM is a chemical speciation model with a few sub-models to calculate cation binding to humic substances and mineral phases (Tipping, 1994). WHAM 7 includes Model VII (Tipping et al., 2011), a discrete-site distribution model, for cation binding to humic acid (HA) and fulvic acid (FA). In WHAM 7, according to the binding mechanisms and metal binding constants with the carboxylic sites (A sites) and phenolic sites (B sites), metal complexation to HA or FA can be divided into two monodentate (denoted as A and B), six bidentate (denoted as AB strong, AB medium, AB weak, AA strong, AA medium, and AA weak), three tridentate complexes (denoted as AAB strong, AAB medium, and AAB weak), and one out-sphere complex (denoted as D) (Shi et al., 2016). The calculations of metal binding constants of specific metal complexes are as follows,

\[ \log K_{MB} = \log K_{MA} \times xK_B / yK_A \]  
\[ \log K_{MAA} = 2\log K_{MA} + x\Delta LK_2 \]  
\[ \log K_{MAB} = \log K_{MA} + \log K_{MB} + x\Delta LK_2 \]  
\[ \log K_{MAAB} = 2\log K_{MA} + \log K_{MB} + y\Delta LK_2 \]
### Table 1. Soil characteristics.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Reference</th>
<th>pH</th>
<th>SOC †</th>
<th>DOC †</th>
<th>Fe&lt;sub&gt;am&lt;/sub&gt; ‡</th>
<th>Fe&lt;sub&gt;am+cry&lt;/sub&gt; ‡</th>
<th>Al&lt;sub&gt;am&lt;/sub&gt; ‡</th>
<th>Clay</th>
<th>Metals</th>
<th>Metal source</th>
<th>Soil location</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Weng et al. (2001)</td>
<td>3.7–6.1</td>
<td>0.4–4.2</td>
<td>NM‡</td>
<td>2.5–33.9</td>
<td>9.5–74.1</td>
<td>NM</td>
<td>2.1–4.9</td>
<td>Cd, Cu, Ni, Pb, Zn</td>
<td>Field contaminated</td>
<td>Netherlands</td>
</tr>
<tr>
<td>D2</td>
<td>Gustafsson et al. (2003)</td>
<td>3.8–7.0</td>
<td>0.8–47.9</td>
<td>0–152.7</td>
<td>13–346</td>
<td>25–403</td>
<td>790–9</td>
<td>0–65</td>
<td>Cd, Cu, Pb, Zn</td>
<td>Laboratory spiked</td>
<td>Sweden</td>
</tr>
<tr>
<td>D3</td>
<td>Dijkstra et al. (2004)</td>
<td>5.0–6.7</td>
<td>1.5–2.4</td>
<td>13.3–229.9</td>
<td>1.1–53.9</td>
<td>11.8–82.6</td>
<td>2.2–48.2</td>
<td>&lt;1.0–4.8</td>
<td>Cd, Cu, Ni, Pb, Zn</td>
<td>Field contaminated</td>
<td>Netherlands</td>
</tr>
<tr>
<td>D4</td>
<td>Lumsdon (2004)</td>
<td>3.6–6.0</td>
<td>1.0–32.0</td>
<td>5.2–137.6</td>
<td>26.0–492.4</td>
<td>NM</td>
<td>68.2–365.8</td>
<td>0–8.0</td>
<td>Cd</td>
<td>Laboratory spiked</td>
<td>Scotland</td>
</tr>
<tr>
<td>D5</td>
<td>Ponizovsky et al. (2006b)</td>
<td>3.4–6.8</td>
<td>0.4–23.3</td>
<td>8–1188</td>
<td>3.6–289.4</td>
<td>NM</td>
<td>6.0–77.2</td>
<td>5–51</td>
<td>Cu</td>
<td>Laboratory spiked</td>
<td>Europe</td>
</tr>
<tr>
<td>D6</td>
<td>Almas et al. (2007)</td>
<td>5.1–6.6</td>
<td>2.0–7.7</td>
<td>13.8–815.3</td>
<td>43.1–188.0</td>
<td>NM</td>
<td>26.1–132.0</td>
<td>3.5–6.8</td>
<td>Cd, Cu, Zn</td>
<td>Field contaminated</td>
<td>Norway</td>
</tr>
<tr>
<td>D7</td>
<td>Shi et al. (2007)</td>
<td>4.2–6.4</td>
<td>0.2–7.2</td>
<td>NM</td>
<td>8.8–195.4</td>
<td>80.2–481.0</td>
<td>19.6–282.5</td>
<td>6–37</td>
<td>Cd</td>
<td>Laboratory spiked</td>
<td>USA</td>
</tr>
<tr>
<td>D8</td>
<td>Ponizovsky et al. (2008)</td>
<td>3.6–6.7</td>
<td>0.3–33.1</td>
<td>0–1137.5</td>
<td>10.9–318.7</td>
<td>NM</td>
<td>14.1–155.7</td>
<td>0.4–46.9</td>
<td>Ni</td>
<td>Laboratory spiked</td>
<td>Europe</td>
</tr>
<tr>
<td>D9</td>
<td>Shi et al. (2013a)</td>
<td>4.2–6.4</td>
<td>0.2–7.2</td>
<td>NM</td>
<td>8.8–195.4</td>
<td>80.2–481.0</td>
<td>19.6–282.5</td>
<td>6–37</td>
<td>Pb</td>
<td>Laboratory spiked</td>
<td>USA</td>
</tr>
<tr>
<td>D10</td>
<td>Duffner et al. (2014)</td>
<td>4.2–7.5</td>
<td>0.4–4.8</td>
<td>1.7–74.5</td>
<td>1.8–100.3</td>
<td>32.2–195.2</td>
<td>NM</td>
<td>1.5–7.2</td>
<td>Zn</td>
<td>Field contaminated</td>
<td>Saudi Arabia, Western Australia, Ethiopia, Netherlands, Turkey, Italy, Spain</td>
</tr>
</tbody>
</table>

† SOC, soil organic carbon; DOC, dissolved organic carbon.
‡ am, amorphous; cry, crystalline.
§ NM, not measured.

### Table 2. Detailed experimental information.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Reference</th>
<th>Methods to determine major soil and solution parameters</th>
<th>Background electrolyte</th>
<th>Solid to liquid ratio, g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Weng et al. (2001)</td>
<td>2 M HNO₃ Solution concentrations</td>
<td>NM†</td>
<td>Ammonium oxalate Oxalic acid</td>
</tr>
<tr>
<td>D2</td>
<td>Gustafsson et al. (2003)</td>
<td>0.5 M NH₄OAc/0.016 M EDTA</td>
<td>0.1 M BaCl₂ extraction</td>
<td>0.1 M BaCl₂ extraction</td>
</tr>
<tr>
<td>D3</td>
<td>Dijkstra et al. (2004)</td>
<td>0.43 M HNO₃</td>
<td>Solution concentrations</td>
<td>NM</td>
</tr>
<tr>
<td>D4</td>
<td>Lumsdon (2004)</td>
<td>Negligible natural concentration</td>
<td>1 M NH₄ acetate</td>
<td>NM</td>
</tr>
<tr>
<td>D5</td>
<td>Ponizovsky et al. (2006b)</td>
<td>Boiled aqua regia</td>
<td>Solution concentrations</td>
<td>Solution concentrations</td>
</tr>
<tr>
<td>D6</td>
<td>Almas et al. (2007)</td>
<td>0.2 M HNO₃</td>
<td>Solution concentrations</td>
<td>Solution concentrations</td>
</tr>
<tr>
<td>D7</td>
<td>Shi et al. (2007)</td>
<td>Negligible natural concentration</td>
<td>Boiled aqua regia</td>
<td>Solution concentrations</td>
</tr>
<tr>
<td>D8</td>
<td>Ponizovsky et al. (2008)</td>
<td>Boiled aqua regia</td>
<td>Solution concentrations</td>
<td>Solution concentrations</td>
</tr>
<tr>
<td>D9</td>
<td>Shi et al. (2013a)</td>
<td>Negligible natural concentration</td>
<td>0.43 M HNO₃</td>
<td>Solution concentrations</td>
</tr>
<tr>
<td>D10</td>
<td>Duffner et al. (2014)</td>
<td>Boiled aqua regia</td>
<td>Solution concentrations</td>
<td>Solution concentrations</td>
</tr>
</tbody>
</table>

† NM, not measured or not mentioned.
‡ am, amorphous; cry, crystalline.
§ DCB, dithionite-biocarbonate-citrate.
other B sites, respectively, and $K_{\text{MAAB}}$ represents metal binding constants of tridentate complexes with two A and one B sites. $pK_A$ and $pK_B$ are the average values of the negative logarithms of the proton dissociation constants of the A and B sites, respectively. Values $x=0, 1, 2$ and $y=0, 1, 5, 3$ correspond to the weak, medium and strong binding strength of the bidentate and tridentate complexes, respectively. $\Delta L K_2$ is a spread factor for the formation of the bidentate and tridentate complexes. The weak, medium, and strong binding sites consist of 90.1, 9, and 0.9% of the total binding sites, respectively (Tipping et al., 2011).

**Cation Exchange Model for Clay Minerals**

The non-specific cation exchange model is used to calculate metal sorption to clay minerals (Lofts and Tipping, 1998), while specific adsorption on clay is neglected in WHAM 7. The default site density ($1 \times 10^{-4}$ eq per gram) for the clay minerals in WHAM 7 was used. Detailed information about the cation exchange model can be found in the previous published literature (Lofts and Tipping, 1998).

**Surface Complexation Models for (Hydr)oxides**

In this study, metal binding to Fe and Al (hydr)oxides was modeled by the same SCM, which only considers amorphous metal oxides and the OH groups of the metal oxide surface reacted with both protons and metal ions (Lofts and Tipping, 1998). The surface complexation reactions was modified by an electrostatic reaction that depends on the net surface charge. The metal ions formed monodentate complexes. To account for the site heterogeneity of the oxide surface, the model further divided the surface sites into three groups according to their binding affinity, which had 90.1, 9, and 0.9% of the total binding sites. The specific surface areas are 600 and 400 m$^2$ g$^{-1}$ for Fe and Al (hydr) oxides, respectively.

**Model Input Parameters**

The major WHAM model input parameters include total heavy metal concentrations, ROM in particulate and colloidal organic matter, mineral components including Fe and Al (hydr) oxides and clay minerals, and solution chemistry conditions (e.g., pH, cations, and anions).

For laboratory spiked soil samples, the total metal concentrations were input as used in the experiments (plus the metal concentrations in original soils when available). The metal concentrations in original soils are usually much lower than that spiked in the laboratory. For the field contaminated soils, however, only the portion of metals extracted by the specific extraction methods was input in WHAM 7. Most of those extraction methods differed from the acid concentrations as specified in Table 2. To quantify the ROM in SOM, we adapted the relationship between SOC and ROC (reactive soil organic carbon) derived in our previous study (Shi et al., 2013a, 2013b). The ROC values for all soils were calculated by linear interpolation of the ROC values of soils determined in the previous study. Generally, organic carbon contents in SOM may vary among soils. In this study, we continued to use the approach used in previous studies to convert organic carbon to organic matter (Ponizovsky et al., 2006b; Shi et al., 2007, 2013a), in which the ROC was converted to ROM by multiplying a factor of 1.8. For ROM, it was input as 84% HA and 16% FA in WHAM. For DOM, it was assumed that 65% of DOM was present as FA (Tipping et al., 2003). Since DOM compositions may vary (Vink et al., 2017), we also tested model performance by assuming 30% of DOM was FA. For Fe minerals, both crystalline and amorphous Fe (hydr)oxides are responsible for metal binding. In the previous study (Dijkstra et al., 2004), the specific surface area of the crystalline Fe (hydr)oxides was assumed to be 100 m$^2$ g$^{-1}$, 1/6 of the amorphous Fe (hydr)oxides’ specific surface area, while site density of both crystalline and amorphous Fe (hydr)oxides were the same. Therefore, in WHAM 7 calculations, the sum of amorphous Fe (hydr)oxides and 1/6 crystalline Fe (hydr)oxides was input as the total reactive amorphous Fe (hydr)oxides. For comparison, model performance without considering crystallized iron (hydr)oxides was also tested. For Al (hydr)oxides, only the amorphous Al oxides were input in WHAM 7. The clay content was input as determined.

Cation competition from typical cations including Ca$^{2+}$, Mg$^{2+}$, Al$^{3+}$ and Fe$^{3+}$ was considered. For both Ca$^{2+}$ and Mg$^{2+}$, the solution concentrations, when available, were input in WHAM 7 calculations. In other cases, the BaCl$_2$ extractable Ca and Mg were input as total Ca and Mg in WHAM 7 calculations. For Al$^{3+}$ competition, the Al$^{3+}$ ion activity was assumed to be controlled by the solubility of Al hydroxide with log $K_{\text{so}} = 8.3$ at pH ≥ 5.5,

$$\log \{\text{Al}^{3+}\} = -3\text{pH} + 3\log \text{Al} + 2.7$$  \[5\]

and, when pH < 5.5, we used the linear regression equation between Al$^{3+}$ ion activity and pH to calculate the Al$^{3+}$ ion activity (Shi et al., 2013a),

$$\log \{\text{Al}^{3+}\} = -1.28\text{pH} - 1.27$$  \[6\]

For Fe$^{3+}$ competition, the Fe$^{3+}$ ion activity was assumed to be controlled by the solubility of Fe hydroxide with log $K_{\text{so}} = 2.7$,

$$\log \{\text{Fe}^{3+}\} = -3\text{pH}$$  \[7\]

**Data Analysis**

To assess the model performance, model calculated total dissolved metals were compared with experimentally determined values. The RMSEs were computed for each heavy metal. The error analysis was used to evaluate the most important model parameters affecting the model performance. We further assessed the roles of each soil components in controlling metal partitioning at various reaction pH. We specifically analyzed how each of SOM binding sites controlled metal partitioning as a function of both reaction pH and adsorbed metal concentrations on each SOM binding site.
Prediction of Heavy Metal Partition Equilibrium

The predictive capability of WHAM 7 for heavy metal partitioning between soil and solution was assessed based on the dissolved metal concentrations ($C_w$). The comparison between model predicted and experimentally determined $C_w$ is shown in Fig. 1, for all five heavy metals. For all 1737 data points, the RMSE value of $\log C_w$ is 0.57, which indicates that WHAM 7 can reasonably predict the partition of heavy metals between soil and solution over a wide range of conditions, with the default model parameters of WHAM 7. Cd has the most data points among five heavy metals, with a total of 846 data points, and the RMSE value of $\log C_w$ is 0.32, showing the best model performance among five heavy metals. Pb has the second largest data sets (562 data points) but with the largest RMSE value of $\log C_w$ (0.83). For Cu, Ni and Zn, they have much smaller amount of data points and their model performance is between that of Cd and Pb, based on the RMSE values of $\log C_w$ (0.68 for Cu, 0.51 for Ni, and 0.45 for Zn).

The effects of iron (hydr)oxides and DOM compositions on model performance were also assessed based on $\log C_w$ values. When no crystallized iron (hydr)oxides were considered in model calculations, there was little impact on the model performance for all five metals, with the changes of RMSE values of $\log C_w$ less than 0.05. Decreasing the percent of FA from 65 to 30% may affect the model performance, with improved model predictions for some studies but worse predictions for others (e.g., changes of RMSE values of $\log C_w$ about 0.1 to 0.2). The overall model performance including all studies, however, did not change significantly, indicating no systematic impact of DOM compositions on model performance. Nevertheless, an accurate determination of dissolved organic carbon (DOC) concentrations and the compositions of DOM will help to accurately predict metal partitioning in soils.

We further conducted error analysis based on $\log C_w$ to examine the major impact factors accounting for the observed errors between WHAM 7 predictions and experimental results. Analyzed factors include major soil properties, such as concentrations of SOM and iron (hydr)oxides, and solution chemistry conditions, such as pH, dissolved Ca and Mg concentrations, Fe$^{3+}$ and Al$^{3+}$ activities, etc. Generally, we did not find any correlation between the major soil properties and the residuals of $\log C_w$ between experimental results and model predictions ($\Delta \log C_w$) for any of the five heavy metals. Among the major solution parameters, total dissolved Ca concentrations and pH appeared to be two main parameters correlated to observed $\Delta \log C_w$. For Cd, Cu, Pb, and Zn, it was found that there was a negative correlation between $\Delta \log C_w$ and the logarithm of total dissolved Ca concentrations (Fig. 2), highlighting the importance of Ca competition for heavy metal binding to soil.
that $\Delta \log C_w$ was negatively correlated with the logarithm of total dissolved Ca concentrations for each group of data (results not shown). With the increase of total dissolve Ca concentrations, WHAM 7 predictions switched from underprediction to overprediction of total dissolved metal concentrations compared with the experimental data. Therefore, an accurate measurement of total reactive Ca concentrations in soil and Ca binding to soil are important for accurate model predictions. For Pb, there was a positive correlation between $\Delta \log C_w$ and solution pH. For Ni, we did not find any correlations between $\Delta \log C_w$ and the major soil and solution parameters that we analyzed.

The application of chemical speciation models including WHAM to predict metal partitioning between soil and solution has been extensively studied, as reviewed by Groenenberg and Lofts (2014). Among typical heavy metals, Pb usually showed the poorest model performance as shown in multiple studies (e.g., Gustafsson et al., 2003; Weng et al., 2001; Xiong et al., 2013), which was attributed to multiple factors such as underestimation of the strong Pb binding sites in SOM (Gustafsson et al., 2003), difference between SOM and generic humic acids used in developing chemical speciation models (Xiong et al., 2013), underestimation of Pb binding to mineral phases (e.g., Al and Fe (hydr) oxides and Mn oxides) (Groenenberg and Lofts, 2014), and the presence of Pb colloids (Sjostedt et al., 2018). In addition to specific soil and solution parameters causing errors of model predictions, which has been discussed in details in each individual study, there are also other important processes that may affect the accuracy of model predictions. Current assemblage models including WHAM use a component additivity approach and the interactions between organic matter and mineral phases are ignored, which may affect model performance (Christl and Kretzschmar, 2001; Vermeer et al., 1999). Furthermore, the molecular level investigation of DOM has shown that the fractionation of DOM due to interactions with mineral phases may significantly affect its compositions and reactivity (Lv et al., 2016). Therefore, further research should consider incorporating those important environmental processes in modeling metal behavior in soil.

**Metal Distribution among Different Soil Components**

The model-predicted heavy metal distribution among different soil components is presented in Fig. 3. The metals Cd, Cu, Ni, and Zn showed a similar pattern, in which SOM and clay minerals dominated metal sorption at low pH and then SOM became dominant at intermediate and high pH values, especially at pH > 5.5. Note that, at low pH, the overall metal adsorption was small and the majority of metals were in the solution. With the increase in pH, the portion of Cd, Cu, Ni, and Zn adsorbed on Fe (hydr) oxides increased and the adsorption on clay minerals decreased. The portion of heavy metals adsorption on Al (hydr)oxides was always much lower than that on Fe (hydr)oxides.

For Cu, SOM was the most dominant adsorbent across the pH range we studied, mainly due to the strong binding ability of Cu to SOM, which was consistent with the previous studies on Cu partitioning in soil (Gustafsson et al., 2003; Weng et al., 2001). The strength of Cu binding to Fe (hydr)oxides is usually considered to be moderate, stronger than Cd, Ni and Zn but weaker than Pb (Tiberg and Gustafsson, 2016; Tiberg et al., 2013), which generally cannot effectively compete with SOM for Cu binding at acidic pH in most soils. For Pb, the distribution of Pb among SOM, metal (hydr)oxides, and clay minerals is highly dependent on the reaction pH. At low pH values, both SOM and clay minerals were significant on Pb binding. With the increase in pH, the role of Fe (hydr)oxides became more important. When pH > 5.0, Fe (hydr)oxides can effectively compete with SOM for Pb binding, and, when pH > 6.0, more Pb may bind to Fe (hydr)oxides than to SOM. This binding behavior is consistent with the strong binding ability of Pb to Fe (hydr) oxides, which was also reported in previous studies (Gustafsson et al., 2011; Shi et al., 2013a).
The overall model predictions of WHAM 7, compared with multiple previous studies as listed in Table 1, generally show similar performance to some studies (e.g., Shi et al., 2007, 2013a; Weng et al., 2001) when the chemical speciation models were used without any further adjustment, but show larger deviations than those studies in which model parameter optimization was used (e.g., Almas et al., 2007; Gustafsson et al., 2003). Metal distribution among different soil components, which was reviewed by Groenenberg and Lofrs (2014), is highly dependent on pH. At acidic to neutral pH, SOM was considered to be the most dominant adsorbent in soil (Duffner et al., 2014; Gustafsson et al., 2003; Weng et al., 2001, 2004), clay minerals may also significantly contribute to metal binding at low pH (Gustafsson et al., 2011; Shi et al., 2013a), and iron and aluminum oxides became more important when pH is greater than 7 (Dijkstra et al., 2009; Fest et al., 2005; Gustafsson et al., 2011; Izquierdo et al., 2013; Marzouk et al., 2013; Shi et al., 2013a). Our model predictions are consistent with those previous observations. Furthermore, the dominance of SOM on metal binding in soils has been reported in spectroscopic studies (Fan et al., 2016; Karlsson et al., 2006; Shi et al., 2012; Strawn and Sparks, 2000), while the importance of mineral phases was found to be dependent on the soil compositions and reaction conditions (Peng et al., 2018; Shi et al., 2012). However, metal speciation in field contaminated soils with high metal concentrations may be significantly different from that in laboratory metal-spiked soils, which may contain different metal minerals formed during contamination processes (Degryse et al., 2011; Khaokaew et al., 2011, 2012; Roberts et al., 2002). Cautions should be taken when applying chemical speciation models to those soils.

Metal Distribution among SOM Binding Sites

As shown in our previous studies, metal binding to different sites of natural organic matter may have different adsorption and desorption rates (Peng et al., 2018; Shi et al., 2016). Therefore, it is essential to understand metal binding to different SOM sites under various reaction conditions. Figure 4 shows metal distribution on different binding sites of SOM as a function of pH and adsorbed metal concentrations on SOM for Cd and Pb. For Cd, bidentate complexes dominated at intermediate and high pH while the formation of the outer-sphere complexes was significant at low pH (e.g., pH < 5) (Fig. 4A). The monodentate complexes may be significant only at high Cd concentrations while the tridentate sites have small quantity and require two A sites and another B site to form metal complexes in WHAM 7. Therefore, at much lower metal concentrations and high pH, it is expected that tridentate sites may be significant in controlling metal binding to SOM. Among all individual bidentate and monodentate binding sites, three binding sites, AB weak, AA weak, and A sites, accounted for the majority of Cd binding to SOM sites.

Different from Cd, both Pb bidentate and monodentate complexes dominated Pb binding to SOM while the formation of both tridentate complexes and outer-sphere complexes was generally small (Fig. 4B), due to the much larger metal binding constants of Pb compared with Cd. In WHAM 7, Pb has a \( \log K_{MA} \) (logarithm of the average of the metal binding constant for A sites) value of 2.37 for HA and 2.15 for FA, while Cd
has a log"K_{MA}" value of 1.67 for HA and 1.51 for FA. Low Pb concentrations favored the formation of bidentate complexes and the monodentate complexes were the most significant at high Pb concentrations and low pH values. Among those individual bidentate and monodentate sites, AB weak, AA weak, A and B sites, were four most important sites accounting for Pb binding.

Not shown here, the distribution of Ca, Ni and Zn among various SOM binding sites was, to different extent, similar to what we observed for both Cd and Pb, with bidentate binding dominating and the contribution of monodentate binding and outer-sphere complexes slightly varying among metals. Although numerous studies have been done on predicting metal partitioning using chemical speciation models, very few focused on the roles of each specific binding sites on controlling metal binding (Gustafsson and van Schaik, 2003; Shi et al., 2016; Tipping and Hurley, 1992; Xiong et al., 2013). In WHAM 7, Model VII predicted that the formation of bidentate complexes dominated metal binding to SOM in most conditions, which is consistent with previous studies with WHAM (Shi et al., 2016; Tipping and Hurley, 1992). Our results also suggest that, while different SOM binding sites may be formed via different combinations of carboxylic and/or phenolic sites, only a few binding sites dominated metal binding even across a wide range of experimental conditions analyzed in this study. The relative significance of those individual sites may differ according to specific reaction conditions.

While recent spectroscopic studies have supported the formation of metal bidentate complexes in SOM (Karlsson et al., 2006; Xiong et al., 2013), the formation of different bidentate binding sites of SOM in WHAM 7 is hypothetical. In WHAM, bidentate sites are formed by combining two monodentate sites, which uses an additive approach to calculate binding constants for bidentate complexes based on the binding constants of the two monodentate sites forming the bidentate complexes (Tipping and Hurley, 1992). It is usually difficult, if possible, to experimentally determine the values of metal binding constants for each individual SOM binding sites, complicating the assessment of the model performance at the molecular level. The magnitude and distribution of metal binding constants used in WHAM were evaluated based on the LFERs (linear free energy relationship) for metal–ligand complexation (Atalay et al., 2013; Carbonaro et al., 2011; Carbonaro and Di Toro, 2007). Consistency has been found between binding constants calculated from the bidentate LFERs and those used in WHAM, but there are also some significant differences in metal binding to those minor amount of strong binding sites and sites containing neutral nitrogen functional groups (Atalay et al., 2013; Carbonaro et al., 2011). Further research is thus desired to more accurately describe metal binding to individual SOM binding sites.

As briefly reviewed in the Introduction section, chemical speciation models with different assumptions on site distributions and binding properties of soil binding sites have shown success in predicting metal partition equilibrium with reasonable accuracy, but metal binding to different binding sites may result in significantly different sorption/desorption rates (Shi et al., 2016; Tian et al., 2017). Since those individual SOM binding sites may control the kinetic behavior of metals at different time scales, it is essential to accurately predict the distribution of metal ions on different soil binding sites. In soil environments, reaction equilibrium may not reach due to the frequent changes of environmental conditions and the dynamic cycling of SOM and Fe minerals. As discussed in the Introduction section, different SOM binding sites may have significantly different reaction rates with heavy metals. To more accurately predict the dynamic behavior of heavy metals in soil environments, it is essential to relate the reaction equilibrium to the kinetics of heavy metal reactions with each individual soil binding site. The results presented in this study have shown that WHAM 7 may provide a basis for further research for this aspect.

CONCLUSIONS

WHAM 7 has been successfully applied to predict heavy metal partitioning between soil and solution over a wide range of reaction chemistry conditions, with reasonable accuracy based on the analysis of over 1700 literature data. While multiple factors may account for the deviations between model predictions and experimental results, solution pH and Ca concentrations are the most important factors explaining the observed errors. SOM, Fe (hydr)oxides and clay minerals are three most important reactive soil components for metal binding, and their relative contributions are highly dependent on reaction pH and specific metals. Based on WHAM 7 calculations, among multiple SOM binding sites, bidentate sites dominate metal binding and those abundant bidentate sites formed either through two carboxylic sites or through one carboxylic and another phenolic sites are most significant ones. Monodentate binding may be significant when total metal concentrations are high while tridentate binding is only significant at low metal concentrations and high pH. It is desired, at the molecular level, to verify metal binding to various SOM sites in the future. Our modeling results have highlighted the importance of specifically considering each individual soil binding site when predicting the fate and transport of heavy metals in soil environments.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Project No. 41573090), Guangdong Innovative and Entrepreneurial Research Team Program (No. 2016ZT06N569), Science and Technology Planning Project of Guangdong Province (2016B020242004), and the Thousand Talent Program for Young Outstanding Scientists of China. The authors declare no conflicts of interest.

REFERENCES


